

Inferring Binary Properties from Gravitational-Wave Signals

Javier Roulet¹ and Tejaswi Venumadhav^{2,3}

¹TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California, USA; email: jroulet@caltech.edu

²Department of Physics, University of California, Santa Barbara, California, USA

³International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore, India

ANNUAL
REVIEWS **CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Nucl. Part. Sci. 2024. 74:207–32

First published as a Review in Advance on
June 19, 2024

The *Annual Review of Nuclear and Particle Science*
is online at nucl.annualreviews.org

<https://doi.org/10.1146/annurev-nucl-121423-100725>

Copyright © 2024 by the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.



Keywords

gravitational waves, black holes, Bayesian statistics

Abstract

This review provides a conceptual and technical survey of methods for parameter estimation of gravitational-wave signals in ground-based interferometers such as Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo. We introduce the framework of Bayesian inference and provide an overview of models for the generation and detection of gravitational waves from compact binary mergers, focusing on the essential features that are observable in the signals. Within the traditional likelihood-based paradigm, we describe various approaches for enhancing the efficiency and robustness of parameter inference. This includes techniques for accelerating likelihood evaluations, such as heterodyne/relative binning, reduced-order quadrature, multibanding, and interpolation. We also cover methods to simplify the analysis to improve convergence, via reparameterization, importance sampling, and marginalization. We end with a discussion of recent developments in the application of likelihood-free (simulation-based) inference methods to gravitational-wave data analysis.

Contents

1. INTRODUCTION	208
2. FORMALISM	209
2.1. Bayesian Inference	209
2.2. Likelihood	209
2.3. Gravitational-Wave Signals	211
2.4. Model Selection	217
2.5. Representation of the Posterior	218
3. LIKELIHOOD-BASED METHODS	218
3.1. Likelihood Acceleration	219
3.2. Degeneracy and Parameterization	222
3.3. Importance Sampling	224
3.4. Marginalization	225
4. LIKELIHOOD-FREE METHODS	227
5. CONCLUSION	229

1. INTRODUCTION

Gravitational waves, predicted by general relativity over 100 years ago, are perturbations in the metric of spacetime sourced by accelerating massive compact objects. Their existence was first established observationally by indirect means, as being responsible for the decay in the orbit of a binary pulsar (1). In 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO) made the first direct detection, sensing the perturbation due to a passing gravitational wave along the length of the detector (2). By now, the LIGO and Virgo interferometers have observed signals from over a hundred mergers of compact binaries hosting black holes or neutron stars, establishing gravitational waves as a major field of research in observational astrophysics (3–5). These detections allow researchers to probe mechanisms of formation and evolution of massive compact binaries over cosmic history and, furthermore, enable empirical tests of fundamental questions in physics, such as the equation of state of matter at supranuclear densities and the validity of general relativity in the dynamic, strong-field regime (6). Extracting these insights requires constructing physical models of the signals and contrasting them with the data to constrain their associated parameters, which are the subjects of this review.

Compared to other astrophysical systems, modeling gravitational waves from binary mergers is, in principle, remarkably clean. Black holes being the simplest objects in nature, a black hole binary is completely described with only 17 parameters. With general relativity, we have a complete theory of its dynamics, and the gravitational radiation emitted propagates unobstructed across the Universe. This simplicity makes the problem of parameter estimation well suited for Bayesian inference. We review all this in Section 2. In practice, considerable complications arise. Modeling waveforms with sufficient accuracy is challenging and requires insights from perturbation theory in various regimes as well as from expensive numerical simulations. The signals are weak and their parameters exhibit significant degeneracy, limiting the validity of analytic approximations and calling for sophisticated inference algorithms (7). These can be classified into two broad classes: likelihood-based and likelihood-free, which are covered in Sections 3 and 4, respectively.

In this review, we focus on methods for parameter estimation in second-generation detectors and the intuition behind them. We touch briefly on some results only insofar as they help illustrate concepts. Although many of these ideas apply more generally to next-generation ground-based or space-based detectors, we do not discuss those.

2. FORMALISM

In this section, we formulate the problem of Bayesian parameter estimation in mathematical terms. We provide models for the measurement process of gravitational waves and their generation by compact binary coalescences.

2.1. Bayesian Inference

The task at hand is to measure the unknown parameters θ of the physical system, in our case a binary merger, given the data d . In the framework of Bayesian inference, this measurement is fully described by the posterior probability density $p(\theta | d)$. By virtue of Bayes' theorem, the posterior can be expressed as

$$p(\theta | d) = \frac{\pi(\theta)\mathcal{L}(d | \theta)}{\mathcal{Z}(d)}. \quad 1.$$

Here, $\pi(\theta)$ is the prior probability density, $\mathcal{L}(d | \theta)$ is the likelihood of the data, and $\mathcal{Z}(d)$ is the evidence. The prior is the probability density we assign to the parameter values θ before the data are analyzed; for us this corresponds to a model of the astrophysical distribution of gravitational-wave sources. The likelihood constitutes a model of the measurement process, including the generation of the signal at the source, the response of the detector to the signal, and the noise in the detector data. The evidence, or partition function, is the normalization constant that makes the posterior integrate to 1 over θ :

$$\mathcal{Z}(d) = \int d\theta \pi(\theta)\mathcal{L}(d | \theta). \quad 2.$$

It is mainly relevant to model selection, which we review in Section 2.4.

2.2. Likelihood

The likelihood is the crucial point for inference via Equation 1, as it links the data and the parameters. We assume that the data are the combination of a signal b and noise n , i.e.,

$$d(t) = b(t) + n(t). \quad 3.$$

We have to solve the inverse problem: Start with given data and infer the partition between the signal b and noise n . To make progress, we need to know the distribution of the noise $n(t)$. Two common simplifying assumptions are that the noise is stationary and Gaussian in nature, with zero mean. This means that we can completely describe its statistics in terms of its autocorrelation function $C(\tau)$, which is defined by

$$\langle n(t)n(t + \tau) \rangle = C(\tau). \quad 4.$$

The $\langle \rangle$ on the left-hand side denotes an expectation value over the distribution that n is drawn from. The right-hand side is a function of only the lag τ due to the stationary nature of the noise. We define the two-sided power spectral density (PSD) of the noise as the Fourier transform of the autocorrelation:

$$S_2(f) = \int_{-\infty}^{\infty} d\tau C(\tau)e^{-i2\pi f\tau}. \quad 5.$$

The autocorrelation function $C(\tau)$ approaches zero in the limit of large lags τ , so the integral in Equation 5 converges. The data are real-valued, and hence the autocorrelation $C(\tau)$ and the PSD $S_2(f)$ are real-valued and even functions of their arguments. It is convenient to define the one-sided PSD as $S(f) = S_2(f) + S_2(-f) = 2 S_2(f)$ for frequencies $f > 0$. **Figure 1** shows examples of

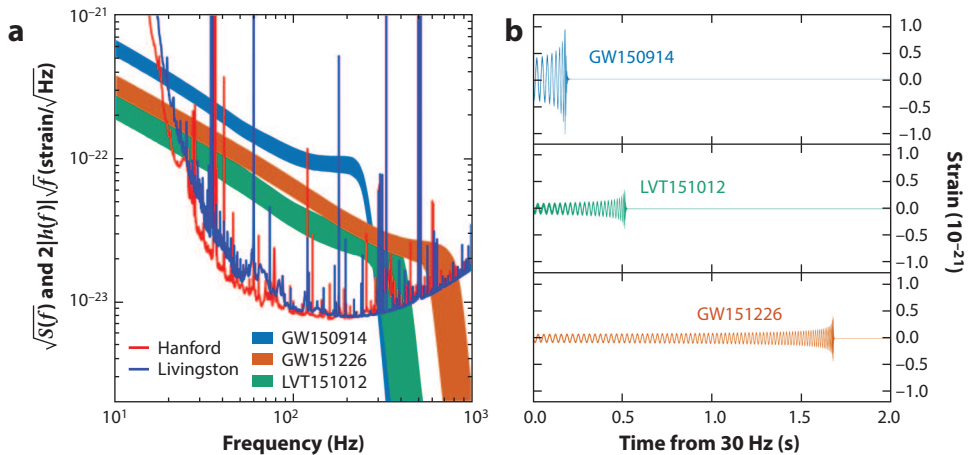


Figure 1

(a) Noise amplitude spectral density $\sqrt{S(f)}$ of the Laser Interferometer Gravitational-Wave Observatory (LIGO) Hanford and Livingston detectors, and (modified) strain of the first three events detected. The detectors are broadband and most sensitive around 100 Hz. The signal amplitude scales as $|b| \propto f^{-7/6}$ at low frequencies (inspiral) and terminates at a frequency of $\sim M^{-1}$ (merger; see Section 2.3.2). (b) Strains in the time domain. The three systems have decreasing mass from top to bottom. Figure adapted from Reference 8 (CC BY 3.0).

the PSD and strain from three signals detected during the first observing run of Advanced LIGO (8).

We work with finite data segments sampled at an interval Δt , such that the Nyquist frequency $1/(2\Delta t)$ encompasses all relevant features of the signal of interest. Given a data segment of duration $T = N\Delta t$, its discrete Fourier transform is defined as

$$\tilde{d}(f_m) = \Delta t \sum_{n=0}^N d(n\Delta t) e^{-i2\pi n\Delta t f_m}, \text{ where} \quad (6)$$

$$f_m = \frac{m}{T}, \quad -N/2 + 1 \leq m \leq N/2. \quad (7)$$

If the PSD $S(f)$ is well behaved and the segment is long, it is approximately true that

$$\langle \tilde{n}(f_m) \tilde{n}^*(f_{m'}) \rangle \approx \frac{T}{2} S(f_m) \delta_{m,m'}, \quad (8)$$

where $\delta_{m,m'}$ is the Kronecker delta function. Given Equation 8 for the noise, we can write down the following expression for the likelihood of the data under the hypothesis that the signal has parameters θ , called the Whittle likelihood:

$$\mathcal{L}(d | \theta) = \mathcal{L}(n \equiv d - b(\theta)) \approx \prod_{0 < m < N/2} \frac{2}{\pi T S(f_m)} \exp \left(-\frac{2|\tilde{d} - \tilde{b}(\theta)|^2}{T S(f_m)} \right). \quad (9)$$

We only let the product run over positive frequencies $f_m > 0$, because the negative frequency components are related to them by complex conjugation (the data and noise are real-valued). To get Equation 9, we interpreted the real and imaginary parts of the noise Fourier components, $\tilde{n}(f_m)$, as real-valued Gaussian random variables with variance equal to half that of the right-hand side of Equation 8. The zero and Nyquist frequency components are purely real-valued, and

including their contribution would change the general form of Equation 9, so we have omitted them for simplicity.

The natural logarithm of the likelihood takes the form

$$\log \mathcal{L}(d | \theta) \approx \sum_{m>0} \left[\log \left(\frac{2}{\pi T S(f_m)} \right) - \frac{2|\tilde{d} - \tilde{b}(\theta)|^2}{T S(f_m)} \right] \quad 10.$$

$$\approx (\text{term independent of } \theta) - \frac{1}{2} \int_0^{f_{\max}} df \frac{4|\tilde{d} - \tilde{b}(\theta)|^2}{S(f)} \quad 11.$$

$$= (\text{term independent of } \theta) - \frac{1}{2} \langle d - b(\theta) | d - b(\theta) \rangle, \quad 12.$$

where we have defined the inverse-noise-weighted inner product

$$\langle A(t) | B(t) \rangle = 4\Re \int_0^{f_{\max}} df \frac{\tilde{A}(f) \tilde{B}^*(f)}{S(f)}. \quad 13.$$

We can proceed by expanding the terms in the inner product in Equation 12 and collecting those that depend on the signal $b(\theta)$. The other terms constitute the likelihood of the data in the case in which the signal is absent, so we can write

$$\log \frac{\mathcal{L}(d | \theta)}{\mathcal{L}(d | \text{no signal})} = \langle d | b(\theta) \rangle - \frac{1}{2} \langle b(\theta) | b(\theta) \rangle. \quad 14.$$

Note that the noise within the data is not square integrable, so the integral in the inner product in Equation 12, and consequently the log likelihood in Equation 9, does not approach a finite limit as the duration $T \rightarrow \infty$ (this makes sense, as the likelihood is the probability of the data and the data change as we extend T). The signal $\tilde{b}(f)$, especially when restricted to the sensitive band, is square integrable, so the likelihood ratio of Equation 14 does have a finite limit as $T \rightarrow \infty$.

2.3. Gravitational-Wave Signals

In this section, we give a description of gravitational waves emitted by compact binaries. We emphasize how various features observable in the signals relate to different physical effects.

2.3.1. Parameter space. Quasi-circular (i.e., noneccentric) binary black hole mergers are characterized completely by fifteen parameters: eight “intrinsic” and seven “extrinsic.” The two component masses and six spin components constitute the intrinsic parameters. Three extrinsic parameters specify the location of the source relative to Earth (right ascension, declination, and distance), three its orientation (inclination, polarization, and orbital phase), and one the time of merger.

Eccentric binaries have two additional parameters: the eccentricity and the argument of periapsis. Gravitational radiation is efficient at dissipating eccentricity (9), thus, only binaries that form at close separation and merge promptly can have a measurable eccentricity (10). In contrast, binaries formed at great orbital separation that later merge owing to gravitational-wave emission are expected to have quasi-circular orbits. Finally, if the merger involves neutron stars or other matter objects, additional parameters are required to specify their internal structure.

2.3.2. Waveform models. Waveform models predict the gravitational-wave signal given the parameters of the source. In principle, this involves solving the dynamics of the compact objects dictated by general relativity (plus matter physics if the objects are not black holes) as they inspiral, merge, and ring-down to a single object, and computing the gravitational waves radiated in the

process. This problem has not been solved exactly, but three main approximate methods have been developed over the past decades: numerical relativity, the post-Newtonian expansion, and black hole perturbation theory. They are complementary to each other because they have different regimes of validity and computational cost.

Numerical relativity directly simulates a spacetime and evolves it according to general relativity; it represents the gold standard in terms of accuracy. However, it is prohibitively expensive for direct use in parameter estimation, as a single simulation may take months to run. Numerical relativity is currently applicable for the late inspiral (a few tens of orbits), merger, and ringdown for moderate mass ratio signals ($q \equiv m_2/m_1 \gtrsim 1/10$). Multiple groups have supplied catalogs of numerical relativity waveforms (11–14; see 15 for a review on numerical relativity).

The post-Newtonian expansion is a low-velocity, weak-field approximation in powers of v/c . It is accurate in the early inspiral but breaks down close to merger as the gravitational field and velocity become large. It features closed-form analytic solutions in both the time and frequency domains. For the case of aligned-spin systems with quadrupole emission (and using relativist's units $G = c = 1$), the frequency-domain strain signal is given by (16, 17)

$$h(f) \propto Q(\text{angles}) \frac{\mathcal{M}^{5/6}}{D} f^{-7/6} e^{i\Phi(f)}, \quad 15.$$

$$\Phi(f) \approx 2\phi_0 - 2\pi f t_0 - \frac{3}{128 \eta v^5} [1 + a_2 v^2 + a_3 v^3 + \mathcal{O}(v^4)], \quad 16.$$

$$a_2 = \frac{55}{9} \eta + \frac{3715}{756}, \quad 17.$$

$$a_3 = \frac{113}{3} \chi_{\text{eff}} - \frac{38}{3} \eta (\chi_1 + \chi_2) \cdot \hat{\mathbf{L}} - 16\pi \quad 18.$$

(we truncated the expansion in v , but more terms are known). The parameters above include the following:

$$v = (\pi M f)^{1/3} \quad (\text{orbital velocity}), \quad 19.$$

$$M = m_1 + m_2 \quad (\text{total mass}), \quad 20.$$

$$\eta = \frac{m_1 m_2}{(m_1 + m_2)^2} \quad (\text{symmetric mass ratio}), \quad 21.$$

$$\mathcal{M} = \eta^{3/5} M \quad (\text{chirp mass}), \quad 22.$$

$$\chi_{\text{eff}} = \frac{m_1 \chi_1 + m_2 \chi_2}{m_1 + m_2} \cdot \hat{\mathbf{L}} \quad (\text{effective aligned spin}), \quad 23.$$

$$\chi_i = \frac{S_i}{m_i^2}, \quad i \in \{1, 2\} \quad (\text{dimensionless spin}). \quad 24.$$

Due to the expansion of the Universe, gravitational waves get redshifted as they propagate. As a result, the measurable quantities are “redshifted” detector-frame masses $m_{\text{det}} = (1+z)m_{\text{source}}$ and luminosity distance $D_L = (1+z)D_{\text{comoving}}$, where z is the redshift (18). This is a consequence of the absence of any absolute mass scale in general relativity. The effect of cosmological redshift is to rescale the time by $(1+z)$, which is indistinguishable from observing waves from a source with higher mass at a greater distance by the same factor in the absence of cosmological expansion. The source-frame mass is unobservable; it is inferred indirectly by using a cosmological model to relate the luminosity distance and the redshift. Thus, the above formulas use detector-frame

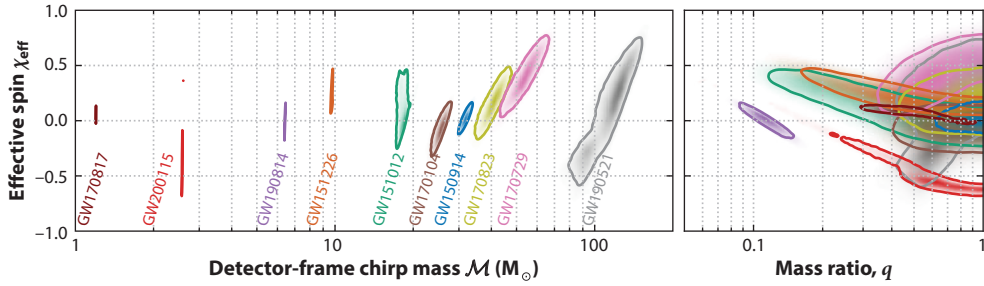


Figure 2

Estimated parameters for an arbitrarily selected subset of signals. For low-mass systems, the detector-frame chirp mass is very well measured, and the mass ratio and effective spin are degenerate. For high-mass events, the mass and effective spin are degenerate. [Some posteriors, e.g., GW200115, differ from published results (3). The reasons are that for this figure we used a broader spin prior, uniform in χ_{eff} , and a newer waveform approximant, IMRPhenomXODE (21).]

masses and luminosity distance. If the source and observer are moving with respect to each other, there is an additional Doppler shift and aberration (19).

Per Equation 15, during the inspiral the amplitude of the waveform scales with $f^{-7/6}$, as can also be seen in **Figure 1**. The amplitude profile starts to deviate from this simple power law as the orbital velocity becomes comparable with c , at a frequency scale given by the innermost stable circular orbit (ISCO) at which the objects plunge—for a test mass, $f_{\text{ISCO}} = (6^{3/2}\pi M)^{-1} \approx 4.4 \text{ kHz } (M/M_{\odot})^{-1}$. After the merger, the emission cuts off rather abruptly as the remnant relaxes to a Kerr black hole. The ISCO and merger frequencies are affected by the component spins: They increase if spins are aligned with the orbit and decrease if they are antialigned (see 20 for a fitting function). Measuring the merger frequency therefore constrains a combination of the total mass and the aligned spins (see the heavier events in the left panel of **Figure 2**).

For the phase, the leading dependence on source parameters is $\propto 1/(\eta v^5) \propto (Mf)^{-5/3}$, therefore the first quantity we can estimate from the phase evolution is $M^{-5/3}$. For low-mass systems, this is a large value, and hence the chirp mass is very well measured, as can also be seen in **Figure 2**. The next two corrections incrementally involve the mass ratio and aligned spins. These terms have smaller, and similar, powers of v , so they do not evolve so differently over the detector frequency band. This correlates the measured coefficients, causing a degeneracy between the mass ratio and effective spin, which is prominent at low masses (16, 17, 22). The degeneracy is exacerbated by the fact that the 1.5 PN coefficient a_3 itself combines the mass ratio and spins (23). By Equation 19, given a frequency range at which the detector operates, the post-Newtonian expansion is more accurate for lower-mass systems, as they have smaller v .

A third approach is black hole perturbation theory, which treats the metric as a perturbation of a background Schwarzschild or Kerr spacetime (24). For example, the self-force approximation is an expansion in the ratio between the curvature scales of the perturbation and background; for binary black holes it becomes an expansion in powers of the mass ratio (which in the limit approaches a test mass in a geodesic orbit). Other expansions also exist. Although black hole perturbation theory is much more recent than the other approaches, it is maturing rapidly and starting to inform current models (25–27).

Parameter estimation requires a stringent balance of computational efficiency and accuracy for waveform models because state of the art methods require $\sim 10^5$ – 10^7 waveform evaluations for each event. For this purpose, different families of approximants have been developed that combine elements of the different approaches mentioned above. Some approximant families are

the *effective one-body* (EOB) [for which two subgroups exist: SEOB (28–31) and TEOBResumS (32–35)], the *phenomenological* IMRPhenom (21, 36–40), and *numerical relativity surrogates* (41–43).

2.3.3. Extrinsic parameters. The distance D , polarization ψ , sky location $\hat{\mathbf{n}}$, geocenter time of arrival $t_\oplus$, inclination ι , and orbital phase ϕ affect the waveform in an easily computable way. The gravitational waves radiated by the source in a given direction (ι, ϕ) can be expressed in terms of two polarizations, $b^+(t)$ and $b^\times(t)$. These, or their Fourier transforms, are the output of a waveform approximant. The response to this signal measured by a detector k is given by the linear combination,

$$h_k(t) = F_k^+(\hat{\mathbf{n}}, \psi) b^+(t - t_k) + F_k^\times(\hat{\mathbf{n}}, \psi) b^\times(t - t_k), \quad 25.$$

$$t_k(t_\oplus, \hat{\mathbf{n}}) = t_\oplus - \hat{\mathbf{n}} \cdot \mathbf{r}_k / c. \quad 26.$$

The antenna responses F^+ and $F^\times$ have known forms in terms of trigonometric functions (44). The arrival time at the detector, t_k , depends on its location $\mathbf{r}_k$ due to the finite speed of gravitational waves c . As we explain in more detail in Section 2.3.5, the inclination determines the polarization content of the wave; for aligned-spin, quadrupole waveforms Equation 25 reduces to

$$\tilde{h}_k(f) = \left[\frac{1 + \cos^2 \iota}{2} F_k^+(\hat{\mathbf{n}}, \psi) - i \cos \iota F_k^\times(\hat{\mathbf{n}}, \psi) \right] e^{-i2\pi f t_k(t_\oplus, \hat{\mathbf{n}})} e^{i2\phi} \frac{1}{D} \tilde{b}_0(f), \quad 27.$$

where $\tilde{b}_0$ only depends on intrinsic parameters. In particular, waves from face-on or -off sources ($\cos \iota = \pm 1$) are circularly polarized, and those from edge-on sources ($\cos \iota = 0$) are linearly polarized.

2.3.4. Spin-orbit precession. If the spins of the compact objects are aligned with the orbital angular momentum, the system has a reflection symmetry about the orbital plane. The evolution preserves this symmetry, and hence the orbital plane stays fixed. If the spins are misaligned, this symmetry is broken and the spin and orbital angular momenta exhibit precession. For most configurations, the direction of the total angular momentum J is approximately conserved and L , S_1 , and S_2 precess around it (45). This is because over the course of a precession cycle, the components of the radiated angular momentum that are perpendicular to J approximately average out, thus J decreases in magnitude but has a fixed direction. In general the two spins precess at different rates, causing an additional nutation of L (46). This picture is justified early in the inspiral, where there is a clean separation of timescales: Orbits are much faster than precession cycles, which in turn are much faster than the dissipation timescale at which the system loses energy to gravitational waves. As the binary approaches merger, these timescales become comparable with each other and are on the order of the mass of the remnant. The behavior just described can break if at some point the spin approximately cancels the orbital angular momentum, for example, for a binary with unequal masses (thus, small values of L) and large values of antialigned primary spin. Then, $J \ll L$ and a small change to J significantly affects its direction. In this case the binary tumbles, a process known as transitional precession (45). Richer spin resonances can also arise from the interplay between precession and nutation (46).

Phenomenologically, compared to aligned-spin systems, precession introduces amplitude modulations in the waveform, along with corrections to the phase evolution (both secular and oscillatory). One way to understand these effects is by introducing the noninertial coprecessing frame, which rotates in such a way that the orbital angular momentum direction $\hat{L}$ appears fixed (47). In this frame, the gravitational-wave emission resembles that of an aligned-spin system—with some late-inspiral corrections to the orbital phase due to spin-orbit and spin-spin coupling, which

involve the in-plane spins. The coprecessing and inertial frames are related by a time-dependent rotation at the (slow) precessional timescale. A binary emits gravitational waves preferentially perpendicular to the orbital plane. Depending on the location of the observer relative to the source, the direction of maximum emission may alternatively point toward or away from the observer as the orbital plane wobbles. When this happens, the observed amplitude of the wave shows modulations on the precession timescale.

The effects of precession on the waveform can be described equivalently in terms of precession harmonics (48). For a given radiation multipole (ℓ, m') in the coprecessing frame, the instantaneous rotation connecting to the inertial frame will leave ℓ invariant and mix the m , manifesting as $2\ell + 1$ harmonics with $-\ell \leq m \leq \ell$. In particular, the dominant coprecessing mode $(\ell, m') = (2, 2)$ yields five harmonics in the inertial frame. Each precession harmonic receives a factor $e^{im\alpha(t)}$ from the time-dependent rotation, where α is the precession angle, so their frequencies differ by the precession frequency $\Omega_p = \dot{\alpha}$. In this picture, the amplitude modulations arise from the beating of the harmonics. The harmonics form a power series in $\tan(\beta/2)$, where β is the angle between $\mathbf{J}$ and $\mathbf{L}$. For most configurations β is small, and then the leading effect of precession is well approximated by the first two harmonics.

2.3.5. Higher-order multipoles. The gravitational radiation emitted by a localized source can be decomposed in multipoles that govern the angular dependency of the wave in the far-field (49). Because the metric perturbation is a tensor, the usual (scalar) spherical harmonics are not adequate to describe it. We say that a function ${}_sf(\theta, \phi)$ on the sphere has spin s if it transforms according to ${}_sf'(\theta, \phi) = {}_sf(\theta, \phi)e^{-is\varphi}$ under a rotation by φ about $\hat{\mathbf{n}}(\theta, \phi)$. The spin-weighted spherical harmonics ${}_sY_{\ell m}(\theta, \phi)$ are an orthonormal basis for such functions. We can construct a quantity with spin -2 out of the gravitational-wave polarizations as $b_+ - ib_\times$ and decompose it using spin-weighted spherical harmonics with $s = -2$:

$$b_+ - ib_\times = \sum_{\ell \geq 2} \sum_{m=-\ell}^{\ell} {}_{-2}Y_{\ell m}(t, \phi) b_{\ell m}. \quad 28.$$

Expressions for ${}_{-2}Y_{\ell m}$ and $b_{\ell m}$ can be found, e.g., in Reference 50; here, we simply point out a few relevant properties and special situations in which some modes vanish. For this discussion, we assume nonprecessing binaries (or, alternatively, use the coprecessing frame), and we align the z -axis with the orbital angular momentum.

Gravitational radiation has $\ell \geq 2$: The $\ell = 0$ (monopole) and $\ell = 1$ (dipole) multipoles are absent due to conservation of energy and momentum, respectively. Inspiral binaries emit dominantly in the quadrupole mode $(\ell, |m|) = (2, 2)$, although other multipoles—usually referred to as higher modes or higher harmonics—are generically present and have in fact been detected in some of the signals (51, 52). Depending on the source parameters and detector frequency response, the next modes in importance are usually $(3, 3)$ and $(4, 4)$.

Just like scalar harmonics, spin-weighted harmonics are of the form

$${}_sY_{\ell m}(t, \phi) = {}_sA_{\ell m}(t)e^{im\phi}, \quad {}_sA_{\ell m} \in \mathbb{R}. \quad 29.$$

All ${}_{-2}Y_{\ell m}$ with $|m| \neq 2$ vanish at the poles; this can be understood because the metric perturbation must transform with $e^{i2\varphi}$ under a rotation by φ , which matches ϕ or $-\phi$ at the two poles. Furthermore, modes $b_{\ell m}$ with odd m vanish for equal-mass binaries; in that case an azimuthal rotation by π becomes a symmetry and only modes with $e^{im\pi} = 1$ survive. Hence, the most favorable systems for observing higher modes are inclined (edge-on), unequal-mass binaries.

Another useful property is that, during the inspiral, a mode (ℓ, m) undergoes m cycles per orbit: Its phase is approximately $\Phi_{\ell m}(t) \approx m \Phi_{\text{orb}}(t)$ or, in the frequency domain (38),

$$\Phi_{\ell m}(f) \approx m \Phi_{\text{orb}}\left(\frac{f}{m}\right). \quad 30.$$

Thus, in general higher modes have higher frequencies, which can affect their detectability due to the uneven detector response to different frequencies. For high-mass systems, which merge at a low orbital frequency, the quadrupole may fall largely outside the detector band, whereas the higher modes do not get as suppressed (53). The simple scaling in Equation 30 breaks as the binary merges and relaxes, or rings down, to a Kerr hole. The ringdown can be described in terms of a superposition of quasi-normal modes with characteristic frequencies, damping times, and angular patterns of radiation. These frequencies are determined by the mass and spin of the remnant and depend strongly on ℓ as well as m (54).

In the context of parameter estimation, there is good motivation to model higher modes. By virtue of Equation 28, when multiple modes are present the signal depends in a nuanced way on the inclination and phase (via $_{-2}Y_{\ell m}$) as well as intrinsic parameters (via $b_{\ell m}$). Notably, quadrupole-only waveforms exhibit degeneracies between the effective spin and mass ratio and between the distance and inclination. The relative amplitude of the higher harmonics depends explicitly on the mass ratio and the inclination, breaking these degeneracies and allowing measurement of those parameters separately. This is illustrated in **Figure 3** for the event GW190412 (51): Including higher modes (and precession) in the model helps constrain the inclination and in turn the luminosity distance, which is otherwise degenerate. Constraining the mass ratio allows measurement of the component masses individually beyond the chirp mass; for example, the

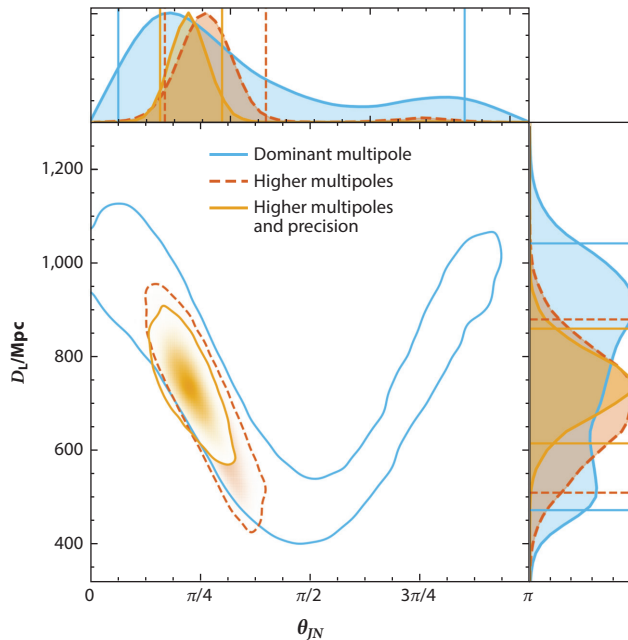


Figure 3

Posterior for the inclination and distance of GW190412 under increasingly refined waveform models. As the effects of higher multipoles and precession are sequentially added, the uncertainty on the inclination, and thereby on the distance, becomes reduced. Figure adapted from Reference 51 (CC BY 4.0).

secondary object in GW190814 has a remarkable mass $m_2 = 2.59^{+0.09}_{-0.08} M_\odot$, which is heavier than any neutron star and lighter than any black hole otherwise measured (52). Improving the distance constraints can potentially aid efforts to search for electromagnetic counterparts by ruling out certain galaxies as possible hosts. With future detector networks, it may even allow singling out the host galaxy without the need for an electromagnetic counterpart, enabling among other things accurate measurement of the Hubble parameter (55).

For precessing systems, we distinguish between coprecessing frame and inertial frame m numbers. In both cases, the waveform is measured in the inertial frame, but the decomposition into modes is different, with different desirable properties: Modes with the same coprecessing frame m have a smooth amplitude and phase obeying Equation 30, whereas the inertial frame (ℓ, m) determine how the waveform transforms under rotations per Equation 28. The two frames are related by a slowly varying (in the precession timescale) 3D rotation. The ℓ number is invariant under rotations so, in the adiabatic regime, it is the same in both frames. For nonprecessing (aligned-spin) signals, the two frames coincide.

2.3.6. Eccentricity. As mentioned earlier, gravitational waves tend to circularize the orbit during the inspiral; furthermore, in general relativity it is challenging to even define eccentricity because it is gauge dependent. As a result, there are multiple definitions in use that satisfy the appropriate Keplerian limit (56). One practical definition is in terms of the waveform at infinity (57; see 58 for an open-source implementation). Phenomenologically, eccentricity excites emission in all harmonics of the orbital frequency. Modeling eccentric systems is a topic of active research: Recent developments include the incorporation of higher harmonics (59, 60) and misaligned spins (61).

2.3.7. Tides. If one or both of the merging compact objects are not black holes (e.g., neutron stars), the orbital dynamics are affected. Each star is distorted by the tidal field from the companion, extracting energy from the orbit, and additionally inducing a quadrupole moment that adds to that of the binary itself, enhancing gravitational-wave emission. These two effects are similar in importance, and both tend to accelerate the inspiral, especially toward the end as the separation between the objects becomes comparable with their size.

To linear order, the induced quadrupole is proportional to the tidal field, $Q_{ij} = -\lambda \varepsilon_{ij}$. The proportionality constant λ is called tidal deformability and depends on the equation of state and mass of the star. A dimensionless deformability can be constructed as $\Lambda = \lambda/m^5 \sim (R/m)^5$, where R and m are the star's radius and mass, respectively. The leading-order change to the waveform phase in Equation 16 is only at 5th post-Newtonian order: $\Delta\Phi_{\text{tidal}}(f) = (117/256\eta)\tilde{\Lambda}v^5(f)$, where the parameter $\tilde{\Lambda}$ is a combination of the component masses and deformabilities, which for identical objects reduces to $\Lambda_1 = \Lambda_2 = \tilde{\Lambda}$ (62, 63). Although the phase correction is suppressed by a high power of v/c , its prefactor is large for neutron stars because $(R/m)^5 \gg 1$, rendering this effect potentially measurable. The dimensionless deformability is larger, thus easier to measure, for less massive neutron stars. The binary neutron star merger GW170817 was constrained to have $\tilde{\Lambda} \lesssim 800$ (64), which together with the chirp mass determination informs the neutron stars' equation of state and radius (65, 66; see 67 for a detailed review).

2.4. Model Selection

Sometimes we may have multiple competing models, or hypotheses $\mathcal{H}_1, \mathcal{H}_2, \dots$, to interpret the data. For example, we may consider different distributions of black hole spins (encoded in the prior), or different theories of gravity (encoded in the likelihood). Bayesian statistics offers a way to assign a “posterior probability” to the models themselves in order to quantify how well they explain

the observations. In this context, the Bayesian evidence $\mathcal{Z}(d | \mathcal{H})$ is nothing but the likelihood for the whole model $\mathcal{H}$ (implicit in Equation 2 in the form of the prior and likelihood), aggregated over the range of predictions $\mathcal{H}$ makes by marginalizing over its internal parameters θ . For any two models, the ratio of these probabilities is the odds,

$$\mathcal{O}_2^1 \equiv \frac{p(\mathcal{H}_1 | d)}{p(\mathcal{H}_2 | d)} = \frac{\pi(\mathcal{H}_1)}{\pi(\mathcal{H}_2)} \cdot \frac{\mathcal{Z}(d | \mathcal{H}_1)}{\mathcal{Z}(d | \mathcal{H}_2)} = \frac{\pi(\mathcal{H}_1)}{\pi(\mathcal{H}_2)} \cdot \frac{\int d\theta_1 \pi(\theta_1 | \mathcal{H}_1) \mathcal{L}(d | \theta_1, \mathcal{H}_1)}{\int d\theta_2 \pi(\theta_2 | \mathcal{H}_2) \mathcal{L}(d | \theta_2, \mathcal{H}_2)}. \quad 31.$$

If $\mathcal{O}_2^1 \gg 1$, we say that model $\mathcal{H}_1$ is favored over $\mathcal{H}_2$ and vice versa. Note that the parameter spaces of the two models need not be the same. The part of the odds that depends on the data is the ratio of evidences, called the Bayes factor between the two models. As a practical note, model selection is one application in which it is important that the normalization is consistent across all the models compared. Another caveat to keep in mind is that the odds do not provide a goodness-of-fit test: Both models might be bad at describing the data even if the odds favor one over the other. See Reference 68 for a review with emphasis on Bayesian inference. Algorithms for computing the evidence $\mathcal{Z}$ are comprehensively reviewed in Reference 6.

2.5. Representation of the Posterior

As discussed in Section 2.3.1, binary mergers are characterized by $\gtrsim 15$ parameters. This dimensionality is prohibitively large for the straightforward approach of tiling the parameter space with a dense grid and recording the posterior probability everywhere.

One efficient way of describing the high-dimensional posterior distribution is by reporting samples drawn randomly from it, $\{\theta_i\} \sim p(\theta | d)$. Samples from the distribution are very versatile, because they allow Monte Carlo estimation of integrals of the form

$$\int d\theta p(\theta) f(\theta) \approx \frac{1}{N} \sum_{\theta_i \sim p}^N f(\theta_i) \quad 32.$$

for arbitrary functions f . A simple example is a histogram of the samples to estimate a marginal distribution: The counts in each bin can be expressed in the form of Equation 32 by choosing $f(\theta) = 1$ inside the bin and 0 outside. Other summary statistics like the mean or quantiles of the distribution are expressible in this form as well. Such integrals also arise in population inference for evaluating the likelihood of models of the astrophysical distribution.

A second way of representing the posterior is by fitting a parametric functional form. For gravitational-wave applications the posterior usually has a complicated structure, motivating the use of flexible approaches such as mixture models (69, 70) or normalizing flows (71–73; see Section 4). Those models also allow the easy generation of samples, so the two methods can be combined.

3. LIKELIHOOD-BASED METHODS

A large class of parameter estimation methods rely on the ability to evaluate the likelihood function given a choice of parameters. This requires an explicit model for the noise, which nearly always is assumed to be stationary and Gaussian, leading to the Whittle likelihood (Equation 14).

General-purpose algorithms can generate samples from a distribution given the ability to evaluate the probability density at arbitrary points. These samplers explore the parameter space efficiently by using the result of previous evaluations to guide new proposals. Two such algorithms frequently used in gravitational-wave inference are nested sampling (74) and Markov Chain Monte Carlo (MCMC) (75, 76; overviewed, e.g., in 6 and 77).

Although sampling algorithms make parameter estimation tractable, in general the process is computationally expensive and not always robust. In this section, we review several methods developed to improve the speed and convergence of sampling algorithms in the specific context of gravitational waves. In likelihood-based parameter estimation, most of the computational effort is generally spent evaluating the likelihood. To improve this, some techniques aim at making each individual likelihood evaluation faster, whereas others aim at decreasing the number of likelihood evaluations required for convergence. We review these approaches in Sections 3.1 and 3.2, respectively. Finally, other methods try to simplify the sampled distribution by correcting (reweighting) samples taken from an approximating distribution (Section 3.3) or marginalizing some parameters (Section 3.4).

Codes for likelihood-based inference of gravitational-wave sources include `LALInference` (77), `BayesWave` (78), `BAYESTAR` (79), `RIFT` (80), `Bilby` (81, 82), `PyCBC Inference` (83), `bajes` (84), `QuickCBC` (85), `cogwheel` (86), `simple-pe` (87), `VARAHA` (88), and `jim` (89).

3.1. Likelihood Acceleration

We now discuss methods to decrease the computational cost of each likelihood evaluation. Each evaluation involves generating a frequency-domain waveform b and computing the inner products with the data $\langle d | b \rangle$ and with itself $\langle b | b \rangle$. Spelling out the log-likelihood (Equation 14), the following integrals need to be computed:

$$\log \mathcal{L}(d | \theta) = \Re 4 \int df \frac{d(f)b^*(f; \theta)}{S(f)} - \frac{1}{2} 4 \int df \frac{|b(f; \theta)|^2}{S(f)}. \quad 33.$$

The straightforward approach is to replace these integrals by a Riemann sum. Unfortunately, the integrands in Equation 33 are highly oscillatory functions of the frequency, so we should evaluate them on a great number of frequencies in order to resolve their features. The maximum frequency must contain the support of the whitened template $b(f)/\sqrt{S(f)}$ (usually several hundred hertz), whereas the frequency resolution is determined by the duration of the signal and of the noise correlations ($\Delta f \lesssim 1/\tau$, where τ is the longest of these timescales). Due to narrow-band spectral lines, noise correlations typically span several seconds, whereas, depending on the mass of the system, the signal duration can be from a fraction of a second for heavy binary black holes to several minutes for light neutron stars.

Next, we review methods to rearrange Equation 33 so that the integrals can still be evaluated accurately at a coarser frequency resolution. Importantly, these algorithms require evaluating the waveform at an arbitrary set of frequencies—the speed-up comes from the reduced number of evaluations. Phenomenological waveform approximants and analytic post-Newtonian models have closed-form expressions and are well suited for this. By contrast, time-domain models such as the EOB family or numerical relativity surrogates do not benefit from evaluating the waveform at fewer frequencies.

3.1.1. Heterodyned likelihood/relative binning. The heterodyne (90), or relative binning (91), algorithm allows the integration of oscillatory functions—such as those encountered in likelihood computation—at low resolution, in the vicinity of a reference integrand function. The key observation is that waveforms supported by the posterior should be similar to each other, because they all fit the data well. We can therefore pick one reference waveform $b_0(f)$, e.g., by maximizing the likelihood, and describe any similar $b(f; \theta)$ in terms of a smooth departure from it. The reference waveform is computed at full resolution but only once, whereas the smooth corrections are computed for multiple θ but at low frequency resolution. The frequency resolution required,

and thereby the computational cost, turns out to be independent of the duration of the signal (91). Several variants of the algorithm exist (90–92); borrowing heavily from these ideas, we present our own formulation.

Let us decompose the strain waveform into harmonics with smooth phase evolution,

$$b(f) = \sum_m b_m(f), \quad 34.$$

where each mode b_m is obtained by summing the coprocessing frame modes over all ℓ of interest—by Equation 30, modes with the same m have a similar phase evolution. We may then express Equation 33 “mode by mode” as

$$\log \mathcal{L}(d \mid \theta) = \sum_m \Re 4 \int df \frac{d(f)}{S(f)} \cdot b_m^*(f; \theta) - \frac{1}{2} \sum_{m, m'} 4 \int df \frac{1}{S(f)} \cdot b_m(f; \theta) b_{m'}^*(f; \theta). \quad 35.$$

We have written each integral in Equation 35 in the form

$$I(\theta) = \int df G(f) \cdot H(f; \theta). \quad 36.$$

Our assumption is that, though both G and H may be highly oscillatory, we are only interested in values of θ for which $H(f; \theta)$ is close to some known $H_0(f)$ in the sense that the ratio $H(f; \theta)/H_0(f)$ is smooth (this requirement motivates the mode-by-mode decomposition) (92). Concretely, we approximate this ratio by a spline $s(f; \theta)$, interpolating it at a coarsely spaced set of N_{coarse} nodes $\{f_i\}$:

$$\frac{H(f; \theta)}{H_0(f)} \approx s(f; \theta), \quad 37.$$

with exact equality at the nodes. It is convenient to write $s(f; \theta)$ as a combination of N_{coarse} basis splines $\{s_i(f)\}$, defined by

$$s_i(f_j) = \delta_{ij} \quad \forall i, j \quad 38.$$

at the nodes $\{f_j\}$, where δ_{ij} is the Kronecker delta. Indeed, in this basis the spline coefficients are simply the values to interpolate:

$$s(f; \theta) = \sum_{i=1}^{N_{\text{coarse}}} \frac{H(f_i; \theta)}{H_0(f_i)} s_i(f). \quad 39.$$

Crucially, Equation 39 separates the (fine) frequency dependence from the parameter dependence. We assume that we can evaluate these coefficients directly without the need to compute a full waveform. Using Equations 37 and 39, we can rewrite Equation 36 as

$$I(\theta) = \int df G(f) H_0(f) \frac{H(f; \theta)}{H_0(f)} \approx \int df G(f) H_0(f) s(f; \theta) = \sum_i w_i H(f_i; \theta), \quad 40.$$

where the parameter-independent weights

$$w_i = \frac{1}{H_0(f_i)} \int df G(f) H_0(f) s_i(f) \quad 41.$$

are computed ahead of time via a Riemann sum.

Applying Equations 40 and 41 to Equation 35, the log-likelihood is computed as

$$\log \mathcal{L}(d | \theta) \approx \sum_{i,m} \Re\{u_{im} b_m(f_i; \theta)\} - \frac{1}{2} \sum_{i,m,m'} \Re\{v_{imm'} b_m(f_i; \theta) b_{m'}^*(f_i; \theta)\}, \quad 42.$$

$$u_{im} = \frac{1}{b_m^{0*}(f_i)} 4 \int df \frac{d(f) b_m^{0*}(f)}{S(f)} s_i(f), \quad 43.$$

$$v_{imm'} = \frac{1}{b_m^0(f_i) b_{m'}^{0*}(f_i)} 4 \int df \frac{b_m^0(f) b_{m'}^{0*}(f)}{S(f)} s_i(f). \quad 44.$$

Note that the second sum is symmetric in m, m' , which further cuts down computations.

Heterodyning achieves a two-fold speedup: Both the waveform evaluation and the sum over frequencies are performed on the sparse interpolation nodes $\{f_i\}$, whereas in the Riemann sum approach the two operations are performed at full resolution.

The choice of coarse frequencies $\{f_i\}$ is made with the heuristic of keeping the integration errors around each f_i bounded and similar to each other. Several algorithms have been used: based on the functional form of the post-Newtonian expansion (91) or a numerical estimate of the Fisher matrix (93), using adaptive refinement via bisection (92), or progressively and monotonically adding frequencies to the set (94). As a ballpark estimate, a few hundred frequencies are typically needed, taking on the order of a millisecond to evaluate a waveform with the IMRPhenomXPHM approximant.

3.1.2. Reduced-order quadrature. Reduced-order quadrature is another method to decrease the number frequencies at which the waveform, and integration quadrature, are evaluated. The method shares many similarities with heterodyning/relative binning, but has the advantage that no reference waveform is required, at the expense of a lower efficiency.

The first step is to build a reduced-order model of the waveform as a function of physical parameters. We approximate the waveform by a model of the form

$$H(f; \theta) \approx \sum_{i=1}^{N_{\text{coarse}}} c_i(\theta) e_i(f), \quad 45.$$

where, again, N_{coarse} is small compared to the number of frequencies in a Riemann sum. The functions $\{e_i(f)\}$ form a basis for a vector space of waveforms, whereas the functions c_i describe the coefficients that yield the specific waveform of interest (or, rather, its projection on this vector space). Building this decomposition is the challenging part; once this has been done, the frequency integrals appearing in the likelihood are evaluated in much the same way as in Section 3.1.1:

$$I(\theta) \approx \sum_i w_i c_i(\theta), \quad \text{where} \quad w_i = \int df G(f) e_i(f). \quad 46.$$

The weights w_i are computed at the start of the analysis, once the data are known.

The reduced-order model contains no information about the data, so it must be accurate globally. An important benefit is that the model is built offline, before any data are collected, and reused for many events. In general, strategies to build the reduced-order model are based on a large collection of example waveforms that tile the target parameter space. One such strategy is a “greedy” algorithm: Start with a random waveform as the first basis element, then, define each subsequent element as the waveform that is worst described as a linear combination of the current bases up to that point (95–97). Another natural way to choose the basis is through a singular value decomposition of the example waveforms (98–100). The basis is kept as small as possible while guaranteeing that the waveforms are reproduced with sufficient accuracy.

To obtain the coefficients $\{c_i(\theta)\}$, two main strategies exist. If we have the ability to evaluate the waveform model at arbitrary frequencies, then we can choose N_{coarse} interpolation nodes $\{f_j\}$ and

solve for the coefficients $\{c_i\}$ in terms of $\{H(f_j)\}$ —again, similarly to Section 3.1.1. These interpolation nodes may be chosen with another greedy algorithm: For each j , pick f_j as the frequency that maximizes the residual after interpolating $e_j(f)$ at the previous frequency nodes $\{f_{<j}\}$ with a linear combination of the previous basis elements $\{e_{<j}(f)\}$ (95–97). More optimal choices for the interpolation nodes are explored in Reference 100. The second strategy is to build fitting formulas for $\{c_i(\theta)\}$ directly, e.g., with interpolating splines (98, 99) or neural networks (101–103). This approach is particularly useful if we lack a closed-form expression for the waveform as a function of frequency (see the discussion at the beginning of Section 3.1).

3.1.3. Multibanding. Gravitational-wave inspirals spend most of the time at low frequencies, whereas high-frequency emission is concentrated toward the end of the inspiral. Multibanding (104, 105) exploits this particular morphology to evaluate waveforms on fewer frequencies while preserving accuracy. Heuristically, the frequency resolution required to accurately represent the waveform is inversely proportional to its duration, and because the high-frequency portion of the signal lasts a shorter time it can be described at coarser frequency resolution.

The frequency domain is partitioned into bands,

$$b(f) = \sum_b b_b(f), \quad 47.$$

and in each band b the waveform is evaluated at a different resolution $\delta f_b \sim 1/\tau_b$. The amplitude and phase of the waveform can then be interpolated to the fine frequency resolution of the original, regular grid (104). However, in that case the speedup comes exclusively from evaluating the waveform model at fewer frequencies—the frequency integral is still performed at full resolution. To accelerate the quadrature over frequencies as well, the data can be similarly separated into frequency bands at different resolution (106).

To choose the bands and resolutions, Vinciguerra et al. (104) designed a scheme based on the leading post-Newtonian expression for the time to merger as a function of frequency and chirp mass, whereas García-Quirós et al. (105) used general error bounds for the Taylor series.

Unlike the heterodyne/relative binning method described in Section 3.1.1, the computational cost of multibanding does increase with waveform duration, so the achieved speedup is more modest. One comparative advantage is that a reference waveform is not required, only a target chirp-mass range.

3.1.4. Interpolation. Finally, another option for accelerating the likelihood evaluation is to construct an interpolant up front, which then allows bypassing the expensive computation of the waveform. Likelihood interpolation is used by the RIFT code (80), which employs Gaussian process interpolation to approximate the marginalized likelihood on the space of intrinsic parameters. Another example is provided by Pathak et al. (107, 108); in this case they use radial basis function interpolation to compute the matched-filtering timeseries and waveform amplitude, from which the likelihood is calculated.

3.2. Degeneracy and Parameterization

An approach complementary to accelerating likelihood evaluations is to ensure that the parameter space is explored in an efficient way, so that fewer waveform evaluations are needed to achieve convergence. Generally, stochastic sampling performs best when the distribution is unimodal and approximately Gaussian, which in practice is never the case with gravitational-wave posteriors due to the ubiquitousness of degeneracy. Degeneracies arise when different values of the parameters produce a similar waveform within measurement uncertainty. When this happens, the posterior does not have a sharp peak; rather, it stretches along the degenerate directions in parameter space

if the degeneracy is continuous, or becomes multimodal if the degeneracy is discrete. Degeneracies are especially problematic if they produce tight nonlinear correlations involving multiple parameters or if they create disconnected modes; both cases are hard to explore with general-purpose methods and turn out to be relevant to gravitational waves.

We can use to our advantage our analytical knowledge of the posterior by predicting how degeneracies will appear and devising schemes that deal with them. Let us, then, understand how degeneracies arise in gravitational-wave posteriors—a comprehensive discussion can be found in Reference 87. Especially for low-mass systems, the inspiral determines the first few coefficients of the post-Newtonian expansion of the waveform phase (Equation 16). Typically, two independent quantities are measured, which roughly correspond to the chirp mass and a combination of the mass ratio and aligned spins. Higher modes, when present, may alleviate the mass-ratio–effective-spin degeneracy. Occasionally, a measurement of precession can constrain the amplitude and phase of the subleading precession harmonic (87, 109, 110). All in all, even for the most favorable sources at most four or five combinations (often less) out of the eight intrinsic parameters get meaningfully constrained. The remaining seven extrinsic parameters control the amplitude, phase, and time of arrival at each detector.¹ In principle this allows the measurement of up to three independent parameters per detector, although in practice it is fewer because the two LIGO interferometers have similar orientations. Thus, extrinsic parameters exhibit significant degeneracy as well, especially when few detectors observe the signal. What is more, multiple disconnected modes often appear in the sky location, inclination, and orbital phase, because those enter the likelihood through oscillatory trigonometric functions that have multivalued inverses (111).

3.2.1. Continuous degeneracies. A solution against continuous degeneracies is to redefine the coordinate system so that the parameters become uncorrelated. The general heuristic is to identify the observables that are constrained by the data and use those as coordinates.

For example, most codes use the chirp mass as a coordinate rather than the individual masses (77), which is well constrained because it controls the leading coefficient in the post-Newtonian expansion of the phase. More generally, a systematic way of identifying these observables is by constructing the Fisher matrix from derivatives of the log likelihood, which gives an estimate of the inverse covariance (16). This is demonstrated in References 112 and 113, where the authors use two linear combinations of the first three phase coefficients as coordinates. They choose these linear combinations as the principal components of the Fisher matrix, as this best removes the correlations. The reason for using only the two leading principal components is that the subsequent components are in fact more constrained by the prior than by the likelihood (e.g., the spins must satisfy $|\chi| < 1$ and the symmetric mass ratio $\eta \leq 1/4$), and in that regime Fisher analysis is a poor description of the posterior (7).

For the in-plane spins, the observables are the amplitude and phase of the precession modulations. These effects are naturally expressed in a coordinate system whose z -axis is set by the approximately constant $\hat{J}$ and whose x -axis is set using the direction of propagation toward the observer (114); other choices of axes produce more correlated posteriors.

Extrinsic parameters are mostly constrained by the observable amplitude, phase, and arrival time of the wave at each detector. The time of arrival at one detector constrains a combination of the geocenter arrival time and the sky location; it can be used directly as a coordinate. The arrival time at a second detector further confines the source to a ring in the sky, centered on the

¹This is strictly true for quadrupolar, aligned spin waveforms. The viewing angle affects the relative contribution from each harmonic mode and, for precessing systems, the magnitude of the amplitude modulations (87). In the majority of cases, these effects are subdominant.

axis containing the two detectors. That axis can be used to define a polar coordinate system in which correlations are minimized (86, 115). The amplitude of the signal measured at one detector determines a combination of the chirp mass, distance, inclination, and sky location—the chirp distance (116)—that again is a good coordinate. Finally, the phase of the wave at the detector constrains a combination of the orbital phase, polarization, and sky location that can also be used as a well-measured coordinate (86).

Eccentric systems require two additional parameters. A convenient choice is the initial eccentricity and relativistic anomaly at an orbit-averaged reference frequency (60).

When changing the coordinate system it is necessary to update the prior, which transforms according to the Jacobian determinant: $p(y) = p(x)|\partial x/\partial y|$. This can limit the flexibility of the reparameterization approach whenever the Jacobian becomes intractable. The ideas outlined above are also domain specific, requiring a sound physical intuition about the measurement process. Some recent approaches effectively change coordinates in an automated way by training normalizing flows on the fly as the sampler progresses in its exploration of the parameter space (89, 117). Normalizing flows can be regarded as flexible coordinate changes that simplify the structure of the distribution and have simple Jacobians; we discuss these further in Section 4.

3.2.2. Discrete degeneracies. Approximate discrete symmetries of the posterior may cause multimodality. Some approximate symmetries are predictable: For example, a quadrupole-only waveform is invariant under a shift of π in the orbital phase. Similarly, under a simultaneous shift in phase and polarization by $\pi/2$, both the strain (h_+ , $h_\times$) and the antenna response (F_+ , $F_\times$) change sign, again leaving the measured signal invariant. For signals dominant in the two LIGO detectors, due to near-alignment of the detectors a flip in $\cos\iota$ and a rotation by π in the sky location about the axis containing the two LIGO detectors become approximate symmetries (86, 111). Alternatively, multimodality can be identified automatically with a preliminary exploration and kernel density estimation (118).

For discrete degeneracies, it is beneficial to connect the modes in some way, so that the sampler can easily explore all of them. This can be done by designing custom jumps for the sampler to propose (77, 78, 82) or by sampling a modified distribution in which the modes of the posterior are folded together (86). Custom jump proposals can handle a greater variety of jumps but require modifying the sampler, and so far, jump proposals have only been implemented for sampling methods that use MCMC. In this case it is important to ensure that the jump proposals preserve detailed balance. Conversely, folding can work with general-purpose samplers; however, it requires modifying the sampled distribution and introduces an additional postprocessing step to reconstruct the original posterior.

3.3. Importance Sampling

Importance sampling is a Monte Carlo method that allows the study of one distribution while sampling from another (119). The idea is to use samples $\{\theta_i\}$ from a distribution q as if they had been drawn from a different distribution p by assigning weights $w_i = p(\theta_i)/q(\theta_i)$. We can see how this works by multiplying and dividing by $q(\theta)$ in Equation 32:

$$\int d\theta p(\theta) f(\theta) = \int d\theta q(\theta) \frac{p(\theta)}{q(\theta)} f(\theta) \approx \frac{1}{N} \sum_{\theta_i \sim q}^N w_i f(\theta_i). \quad 48.$$

Abstracting away the arbitrary function f , the weighted samples represent the distribution p . In many applications, neither p nor q need to be normalized.

Importance sampling is subject to pitfalls if the importance distribution q is not chosen adequately. We can see that the weights are badly conditioned if $q \ll p$ in some region. If the

weights are too imbalanced, we may end up with a few samples dominating the estimation. A way to quantify this is through the effective sample size,

$$n_{\text{eff}} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}, \quad 49.$$

defined in this way so that the variance of the estimator (Equation 48) is $\text{Var}_q(f)/n_{\text{eff}}$. A low n_{eff} is a diagnostic that the scheme is failing and more samples, or a better choice of q , are needed. As a general rule, the proposal distribution q should have heavier tails than the nominal distribution p so that the weights do not diverge.

Importance sampling can be advantageous when it is much easier to sample from the importance distribution than from the nominal distribution. For example, based on the consideration that higher modes usually have a subdominant effect on the posterior but significantly increase the cost of waveform evaluation, Payne et al. (120) implemented a scheme in which posterior samples are drawn using a quadrupole-only waveform model and reweighted using a model that includes higher modes. If the sampler requires many evaluations of the posterior per sample produced, this approach can provide a speedup, because the more expensive model only has to be evaluated on the accepted samples. Using a similar logic, importance sampling was applied in References 121 and 122 to include eccentricity for samples initially obtained with a quasi-circular waveform model, or in Reference 123 to account for detector calibration uncertainty. Another example is DINGO-IS (124), which reweights samples obtained with machine-learning methods (reviewed in Section 4). This allows verification and correction of the machine-learned posterior and, furthermore, computation of the Bayesian evidence because in this application the proposal is normalized.

Another application in which importance sampling is useful is in population inference, where we are interested in fitting models of the astrophysical distribution to the full catalog of observations. This requires evaluating integrals of the form of Equation 48 over a large set of distributions p , which we cannot afford to sample every time.

A different regime in which importance sampling is beneficial is when the support of $f \cdot p$ is very different from the support of p in Equation 48. In that case, we may use a proposal matched to $f \cdot p$ to concentrate the samples around the region where the integrand is important. This has been used to marginalize over time of arrival and sky location in References 125 and 126, as we discuss in Section 3.4.

3.4. Marginalization

A more elaborate technique is to marginalize the distribution over some of the parameters, thereby reducing the parameter space dimensionality and simplifying the task for the sampler. We know the dependence of the posterior on extrinsic parameters analytically, which makes them well suited for conditioning and marginalizing without requiring additional calls to the waveform approximant. Let us partition the parameter space $\theta = (\kappa, \mu)$ into parameters κ that are kept and μ that are marginalized over; then the marginal posterior is

$$p(\kappa \mid d) = \int d\mu p(\kappa, \mu \mid d). \quad 50.$$

To see how this helps, suppose that we have an efficient way of computing $p(\kappa \mid d)$ and also of sampling from the conditional posterior $p(\mu \mid \kappa, d)$ —such that the cost of these operations is comparable with that of evaluating $p(\kappa, \mu \mid d)$, which is typically dominated by the cost of computing a waveform. Then, we can divide the problem into first obtaining posterior samples for κ and later completing them with μ taken from the conditional posterior to undo the marginalization.

I.e., $\{\theta_i\} = \{\kappa_i, \mu_i\}$ with $\kappa_i \sim p(\kappa | d)$ and $\mu_i \sim p(\mu | \kappa_i, d)$. The advantage over sampling the full posterior directly is that the κ -space is lower dimensional and possibly better behaved if the μ parameters exhibit complicated structure.

The BAYESTAR pipeline first used this approach to obtain a low-latency posterior for the source location, conditioned on intrinsic parameters and marginalized over the time and orientation under certain approximations, to enable follow-up of transient electromagnetic counterparts (79). Pankow et al. (125) developed a scheme to marginalize all extrinsic parameters by Monte Carlo integration while gridding over the intrinsic, subsequently implemented in the software RIFT (80). Islam et al. (126) implemented a fast algorithm to perform the extrinsic marginalization for quadrupolar, aligned-spin sources. Beyond this, several parameter estimation codes allow marginalization of a subset of the extrinsic parameters, such as orbital phase, distance, time, and/or polarization.

One very general integration method is Monte Carlo via importance sampling (125, 126), overviewed in Section 3.3. One draws multiple samples of extrinsic parameters, evaluates their likelihood, and estimates the marginalized posterior using Equation 48. The challenge is to find a proposal distribution that is at the same time feasible to sample and well matched to the posterior. To inform time and sky location proposals, Pankow et al. (125) perform adaptive Monte Carlo, whereas Islam et al. (126) use estimates of the arrival time at each detector. Monte Carlo naturally allows for sampling the conditional posterior by drawing from the same weighted samples.

Due to their particular functional form, some parameters are easier to marginalize. Integration over orbital phase ϕ can be done analytically for quadrupolar signals, which satisfy $b(\phi) = e^{i2\phi} b(\phi=0)$. Then, $\langle b | b \rangle$ is independent of ϕ , and $\langle d | b \rangle = |\langle d | b \rangle| \cos[2(\phi - \phi_0)]$, where the $\langle d | b \rangle$ inner product is defined as in Equation 13 except without taking the real part, and ϕ_0 is a constant. The marginalized posterior becomes

$$\int_0^{2\pi} \frac{d\phi}{2\pi} \exp\left(|\langle d | b \rangle| \cos[2(\phi - \phi_0)] - \frac{1}{2} \langle b | b \rangle\right) = e^{-\langle b | b \rangle / 2} I_0(|\langle d | b \rangle|), \quad 51.$$

where I_0 is a modified Bessel function of the first kind (77). To undo the marginalization, given the other parameters, we can sample ϕ by noting that the conditional posterior (the integrand in Equation 51) follows a von Mises distribution for 2ϕ , for which efficient routines exist. For waveforms with higher modes there is no analytical solution, so a different integration method like a quadrature may be used.

To marginalize the distance, we use that $b(D) = b_1/D$ with $b_1 = b(D=1)$. Hence, $\langle d | b \rangle = \langle d | b_1 \rangle / D$ and $\langle b | b \rangle = \langle b_1 | b_1 \rangle / D^2$. The marginalized posterior,

$$\int_0^\infty dD \pi(D) \exp\left(\frac{\langle d | b_1 \rangle}{D} - \frac{\langle b_1 | b_1 \rangle}{2D^2}\right), \quad 52.$$

only depends on $\langle d | b_1 \rangle$ and $\langle b_1 | b_1 \rangle$, so it can be tabulated in advance and evaluated with 2D interpolation in terms of these two quantities. To better condition the interpolation, in practice it is convenient to use other combinations instead of $\langle d | b_1 \rangle$ and $\langle b_1 | b_1 \rangle$ as coordinates. Also, a good quantity to tabulate is the logarithm of the ratio of the marginal likelihood to an analytical approximation, as this has a smaller dynamic range (see 79). Simultaneous marginalization over phase and distance can be done for quadrupolar signals with a similar interpolation if replacing the likelihood by the phase-marginalized likelihood (Equation 51) and changing the inputs to $|\langle d | b_1 \rangle|$ and $\langle b_1 | b_1 \rangle$.

The arrival time can be marginalized by quadrature. This requires the timeseries of $\langle d | b e^{-i2\pi ft} \rangle$ in the vicinity of the event, which can be computed efficiently, e.g., by a pruned fast Fourier transform or the heterodyne method of Section 3.1.1.

4. LIKELIHOOD-FREE METHODS

The methods outlined in the preceding section all use the same basic framework; they repeatedly query the data's likelihood (and prior) at well-chosen sets of parameters for each event. In recent years, the development and availability of specialized computing hardware has enabled a different class of approaches to inference, which are collectively called likelihood-free or simulation-based methods (127, 128). The various approaches under this umbrella share a few common themes:

1. *Likelihood*: The functional form of the data's likelihood $\mathcal{L}(d | \theta)$ is not directly used; indeed, it may not even be known or tractable. Instead, the likelihood is viewed as a distribution for the data that may be easily sampled from (recall the perspective of the likelihood as a model of the data generation process mentioned in Section 2.1).
2. *Posterior*: The traditional approach using methods like MCMC yields samples from the posterior distribution rather than the functional form itself. In contrast, likelihood-free methods typically fit functional forms to the posterior $p(\theta | d)$, viewed as a function that takes both θ and d as inputs. It is still advantageous to be able to draw samples from the posterior, e.g., to easily marginalize nuisance parameters and compute weighted integrals via Monte Carlo (see Section 2.5), so functional representations that are easy to sample from are convenient.
3. *Cost amortization*: The cost of fitting a functional form for the posterior is formidable and is an extra cost that is absent in the conventional approach. However, this fit is typically performed ahead of time by simulating many data sets (see point 1 above), and once fresh data come in, the fit can be reused with negligible computational overhead. In this sense, the costs of performing the fit are amortized over the many occasions we have to perform parameter inference, to the point where the average computational cost of any single inference can be very low.

For points 1 and 2 above, it is interesting to consider the inversion of perspective in relation to the conventional approach, which samples from the posterior but works with a functional form for the likelihood.

Let us assume we have a flexible parameterization of functions $q_\phi(\theta | d)$ with some adjustable internal parameters ϕ , and we want to perform the fit in point 2 above. For a given ϕ , we can interpret $q_\phi(\theta | d)$ as a function that takes in data d and outputs a probability distribution over θ —our approximation to the posterior. Our task is to optimize ϕ to make this approximation as good as we can. The starting point for this fit is to define a loss function to minimize. The conventional loss function is the Kullback–Leibler (KL) divergence between the true posterior distribution and the fitted form, i.e.,

$$D_{\text{KL}}(p(\theta | d) || q_\phi(\theta | d)) = \int d\theta p(\theta | d) \ln \frac{p(\theta | d)}{q_\phi(\theta | d)} \quad 53.$$

$$= (\text{term independent of } \phi) - \int d\theta p(\theta | d) \ln q_\phi(\theta | d). \quad 54.$$

It makes sense to minimize the KL divergence, owing to its properties of being nonnegative and equal to 0 if and only if the two distributions coincide. We want the fit to perform well for a range of parameters and data and not just one realization, so the actual loss function is averaged over all the possibilities, i.e.,

$$L(\phi) = - \int dd p(d) \int d\theta p(\theta | d) \ln q_\phi(\theta | d) \quad 55.$$

$$= - \int d\theta \pi(\theta) \int dd \mathcal{L}(d | \theta) \ln q_\phi(\theta | d). \quad 56.$$

In the second line, we have rewritten the joint probability $p(\boldsymbol{\theta}, d)$ in a way that makes it straightforward to simulate the data generation process: We draw signal parameters $\boldsymbol{\theta}_j$ from the prior $\pi(\boldsymbol{\theta})$ and then compute the waveforms and add noise to generate synthetic strain data realizations d_j , where $j \in \{1, \dots, N\}$ indexes the samples. With many simulations, we can evaluate the integral in Equation 56 using Monte Carlo:

$$L(\phi) = -\frac{1}{N} \sum_{\boldsymbol{\theta}_j, d_j} \ln q_\phi(\boldsymbol{\theta}_j | d_j). \quad 57.$$

Given this evaluation of the loss function, one can numerically optimize it to find the best parameters ϕ for the function approximation (69). The practical challenge is to propose a family of models q_ϕ that is flexible enough to express any posterior that the data may produce while keeping Equation 57 sufficiently amenable to minimization algorithms.

Typically, this challenge is addressed with some use of deep neural networks—functions with an arbitrary number of inputs and outputs that achieve a high expressiveness by combining thousands of simple, tunable nonlinear components. A common concern is the extremely great dimensionality of the data d ; direct “brute force” optimization of the parameters of a network representing the posterior conditioned on such a great number of inputs is not advisable. Existing implementations differ in the reduced representation of the data used for the fits and the exact parameterization of the “model” posterior $q_\phi(\boldsymbol{\theta} | d)$.

In the initial work of Chua & Vallisneri (69), the dimensionality of the data was reduced by projecting it into the lower-dimensional space spanned by basis functions such as the ones used in reduced-order modeling in Equation 45. The resulting coefficients were used to train simple representations of the marginal posteriors of a subset of the parameters.

Subsequent work that built on this (71–73) used the technique of normalizing flows to increase the flexibility of the posterior representation and achieved successful parameter estimations in the full parameter space for quasi-circular binary mergers using the likelihood-free method. A normalizing flow conditioned on the data d (or some reduced representation of it) is a mapping between abstract variables $\mathbf{u}$ and the desired physical parameters $\boldsymbol{\theta}$; i.e., $\boldsymbol{\theta} = f_{\phi, d}(\mathbf{u})$, such that if $\mathbf{u}$ are distributed according to some simple distribution of choice (e.g., the standard multivariate normal), under the transformation f , the physical variables $\boldsymbol{\theta}$ end up being distributed according to the posterior of interest. A (tractable) normalizing flow is a very powerful representation of the posterior, because it makes both sampling and density evaluation trivial. Indeed, the flow directly produces samples of $\boldsymbol{\theta}$ given samples of $\mathbf{u}$, which can be cheaply generated for simple distributions. The posterior density in terms of $\boldsymbol{\theta}$ is related to the analytical density in the $\mathbf{u}$ space by the determinant of the Jacobian of the flow. Mathematically,

$$q_\phi(\boldsymbol{\theta} | d) = p_0 \left(\mathbf{u} \equiv f_{\phi, d}^{-1}(\boldsymbol{\theta}) \right) \left| \frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}} \right|, \quad 58.$$

where p_0 is the simple base distribution over $\mathbf{u}$ (e.g., multivariate normal). Thus, flows with a known Jacobian determinant $|\partial \mathbf{u} / \partial \boldsymbol{\theta}|$ are desirable. An elegant way to construct expressive yet tractable normalizing flows is by composing multiple simple flows: $f_{\phi, d}(\mathbf{u}) = f_N \circ \dots \circ f_1(\mathbf{u})$, where each element f_i may have parameters that are the output of a neural network that takes the (reduced) data as inputs. With this construction, the Jacobian determinant is the product of the individual ones.

Another successful implementation (70) used a variational autoencoder (129), which finds an appropriate reduction of the data as an ingredient in the construction of the model posterior $q_\phi(\boldsymbol{\theta} | d)$ (as opposed to the coefficients of the reduced-order basis used in 69).

Likelihood-free inference also works for inferring only a subset of the physical parameters; in that case, we simply ignore the nuisance parameters after the training data are generated. In

contrast, within likelihood-based inference one must implement an explicit marginalization of the model, as discussed in Section 3.4.

Finally, it has been noted that likelihood-free inference does not require making assumptions about Gaussianity or stationarity of the noise so long as we are able to simulate it. One way to simulate detector noise empirically under weak assumptions is to train the model on actual data with synthetic signals injected (73). Codes for likelihood-free inference of gravitational-wave sources include PERCIVAL (69), VItamin (70) and DINGO (73).

5. CONCLUSION

Virtually all implications of gravitational-wave data for physics and astrophysics—ranging from multimessenger astronomy to evolution of massive stellar binaries or tests of theories of gravity—rely on parameter estimation. There is ample opportunity for the broad astrophysics community to get involved in these endeavors, as these data are periodically released (after an initial proprietary period) through a Gravitational Wave Open Science Center (<https://gwosc.org>), along with software and other resources (130).

In this article, we have reviewed the theoretical foundations and state-of-the-art methods for retrieving parameters of gravitational-wave signals. We introduced the basics of gravitational-wave detection, describing how the different parameters (chirp mass, mass ratio, aligned and in-plane spins, and extrinsic parameters) influence the qualitative features of the signal (phase evolution, amplitude modulations, and higher harmonics).

Equipped with this information, we have described an array of methods for achieving robust and efficient parameter estimation. Within the traditional paradigm of likelihood-based inference, we discussed algorithms for accelerating likelihood evaluation (heterodyne/relative binning, reduced-order quadrature, multibanding, and interpolation) and, complementarily, improving convergence by simplifying the distribution (via reparameterization, approximating distributions, or marginalization of nuisance parameters). We have also covered the emerging field of likelihood-free inference, where advancements in hardware infrastructure, as well as algorithms for fitting very flexible parametric distributions, allow for inference based on a large number of examples of synthetic data rather than explicit likelihood evaluations. Although computationally intensive, this training is performed offline, subsequently allowing inference in real time as new data are collected.

Due to space constraints, we have not covered nonstationarity and non-Gaussianity. Literature on gravitational-wave data analysis under weaker assumptions includes References 131–136. In principle, likelihood-free inference can also deal with these issues naturally, by training on synthetic injections made on real data.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGEMENTS

We thank Katerina Chatziioannou for insightful comments on the manuscript, and Eliot Finch and Jonathan Thompson for helpful discussion. J.R. acknowledges support from the Sherman Fairchild Foundation. T.V. acknowledges support from National Science Foundation grants 2012086 and 2309360, the Alfred P. Sloan Foundation through grant number FG-2023-20470, the Binational Science Foundation through award number 2022136, and the Hellman Family Faculty Fellowship.

LITERATURE CITED

1. Taylor JH, Fowler LA, McCulloch PM. *Nature* 277(5696):437–40 (1979)
2. Abbott BP, et al. *Phys. Rev. Lett.* 116(6):061102 (2016)
3. Abbott R, et al. *Phys. Rev. X* 13(4):041039 (2023)
4. Nitz AH, et al. *Astrophys. J.* 946(2):59 (2023)
5. Olsen S, et al. *Phys. Rev. D* 106(4):043009 (2022)
6. Christensen N, Meyer R. *Rev. Mod. Phys.* 94(2):025001 (2022)
7. Vallisneri M. *Phys. Rev. D* 77(4):042001 (2008)
8. Abbott BP, et al. *Phys. Rev. X* 6(4):041015 (2016)
9. Peters PC, Mathews J. *Phys. Rev.* 131(1):435–40 (1963)
10. Samsing J. *Phys. Rev. D* 97(10):103014 (2018)
11. Boyle M, et al. *Class. Quantum Gravity* 36(19):195006 (2019)
12. Healy J, Lousto CO. *Phys. Rev. D* 102(10):104018 (2020)
13. Ferguson D, et al. arXiv:2309.00262 [gr-qc] (2023)
14. Hamilton E, et al. *Phys. Rev. D* 109(4):044032
15. Lehner L, Pretorius F. *Annu. Rev. Astron. Astrophys.* 52(1):661–94 (2014)
16. Cutler C, Flanagan EE. *Phys. Rev. D* 49(6):2658–97 (1994)
17. Poisson E, Will CM. *Phys. Rev. D* 52(2):848–55 (1995)
18. Krolak A, Schutz BF. *Gen. Relativ. Gravit.* 19(12):1163–71 (1987)
19. Bonvin C, et al. *Mon. Not. R. Astron. Soc.* 525(1):476–88 (2023)
20. Ajith P, et al. *Phys. Rev. Lett.* 106(24):241101 (2011)
21. Yu H, et al. *Phys. Rev. D* 108(6):064059 (2023)
22. Baird E, Fairhurst S, Hannam M, Murphy P. *Phys. Rev. D* 87(2):024035 (2013)
23. Ng KKY, et al. *Phys. Rev. D* 98(8):083007 (2018)
24. Pound A, Wardell B. In *Handbook of Gravitational Wave Astronomy*, ed. C Bambi, S Katsanevas, KD Kokkotas. Singapore: Springer. https://doi.org/10.1007/978-981-15-4702-7_38-1 (2021)
25. Islam T, et al. *Phys. Rev. D* 106(10):104025 (2022)
26. Wardell B, et al. *Phys. Rev. Lett.* 130(24):241402 (2023)
27. van de Meent M, et al. *Phys. Rev. D* 108(12):124038 (2023)
28. Bohé A, et al. *Phys. Rev. D* 95(4):044028 (2017)
29. Cotesta R, et al. *Phys. Rev. D* 98(8):084028 (2018)
30. Ossokine S, et al. *Phys. Rev. D* 102(4):044055 (2020)
31. Ramos-Buades A, et al. *Phys. Rev. D* 108(12):124037 (2023)
32. Nagar A, et al. *Phys. Rev. D* 98(10):104052 (2018)
33. Nagar A, et al. *Phys. Rev. D* 102(2):024077 (2020)
34. Gamba R, Akçay S, Bernuzzi S, Williams J. *Phys. Rev. D* 106(2):024020 (2022)
35. Nagar A, et al. *Phys. Rev. D* 108(12):124018 (2023)
36. Khan S, et al. *Phys. Rev. D* 93(4):044007 (2016)
37. Khan S, Ohme F, Chatziioannou K, Hannam M. *Phys. Rev. D* 101(2):024056 (2020)
38. García-Quirós C, et al. *Phys. Rev. D* 102(6):064002 (2020)
39. Pratten G, et al. *Phys. Rev. D* 103(10):104056 (2021)
40. Thompson JE, et al. *Phys. Rev. D* 109(6):063012 (2024)
41. Blackman J, et al. *Phys. Rev. Lett.* 115(12):121102 (2015)
42. Varma V, et al. *Phys. Rev. D* 99(6):064045 (2019)
43. Varma V, et al. *Phys. Rev. Res.* 1(3):033015 (2019)
44. Whelan JT. *The geometry of gravitational wave detection*. https://dcc.ligo.org/public/0106/T1300666/003/Whelan_geometry.pdf (2013)
45. Apostolatos TA, Cutler C, Sussman GJ, Thorne KS. *Phys. Rev. D* 49(12):6274–97 (1994)
46. Gangardt D, et al. *Phys. Rev. D* 103(12):124026 (2021)
47. Schmidt P, Hannam M, Husa S, Ajith P. *Phys. Rev. D* 84(2):024046 (2011)
48. Fairhurst S, et al. *Phys. Rev. D* 102(2):024055 (2020)
49. Thorne KS. *Rev. Mod. Phys.* 52(2):299–339 (1980)

50. Blanchet L. *Living Rev. Relativ.* 17(1):2 (2014)
51. Abbott R, et al. *Phys. Rev. D* 102(4):043015 (2020)
52. Abbott R, et al. *Astrophys. J. Lett.* 896(2):L44 (2020)
53. Mills C, Fairhurst S. *Phys. Rev. D* 103(2):024042 (2021)
54. Berti E, Cardoso V, Will CM. *Phys. Rev. D* 73(6):064030 (2006)
55. Borhanian S, et al. *Astrophys. J.* 905(2):L28 (2020)
56. Loutrel N, Liebersbach S, Yunes N, Cornish N. *Class. Quantum Gravity* 36(2):025004 (2018)
57. Mora T, Will CM. *Phys. Rev. D* 69(10):104021 (2004)
58. Shaikh MA, et al. *Phys. Rev. D* 108(10):104007 (2023)
59. Liu X, Cao Z, Zhu ZH. *Class. Quantum Gravity* 39(3):035009 (2022)
60. Ramos-Buades A, Buonanno A, Gair J. *Phys. Rev. D* 108(12):124063 (2023)
61. Liu X, Cao Z, Zhu ZH. arXiv:2310.04552 [gr-qc] (2023)
62. Flanagan EE, Hinderer T. *Phys. Rev. D* 77(2):021502 (2008)
63. Favata M. *Phys. Rev. Lett.* 112(10):101101 (2014)
64. Abbott BP, et al. *Phys. Rev. Lett.* 119(16):161101 (2017)
65. Abbott BP, et al. *Phys. Rev. Lett.* 121(16):161101 (2018)
66. Raithel CA, Özel F, Psaltis D. *Astrophys. J.* 857(2):L23 (2018)
67. Chatzioannou K. *Gen. Relativ. Gravit.* 52(11):109 (2020)
68. Thrane E, Talbot C. *Publ. Astron. Soc. Aust.* 36:e010 (2019)
69. Chua AJK, Vallisneri M. *Phys. Rev. Lett.* 124(4):041102 (2020)
70. Gabbard H, et al. *Nat. Phys.* 18(1):112–17 (2021)
71. Green SR, Simpson C, Gair J. *Phys. Rev. D* 102(10):104057 (2020)
72. Green SR, Gair J. *Mach. Learn. Sci. Technol.* 2(3):03LT01 (2021)
73. Dax M, et al. *Phys. Rev. Lett.* 127(24):241103 (2021)
74. Skilling J. *Bayesian Anal.* 1(4):833–59 (2006)
75. Metropolis N, et al. *J. Chem. Phys.* 21(6):1087–92 (1953)
76. Hastings WK. *Biometrika* 57(1):97–109 (1970)
77. Veitch J, et al. *Phys. Rev. D* 91(4):042003 (2015)
78. Cornish NJ, Littenberg TB. *Class. Quantum Gravity* 32(13):135012 (2015)
79. Singer LP, Price LR. *Phys. Rev. D* 93(2):024013 (2016)
80. Lange J, O’Shaughnessy R, Rizzo M. arXiv:1805.10457 [gr-qc] (2018)
81. Ashton G, et al. *Astrophys. J. Suppl. Ser.* 241(2):27 (2019)
82. Ashton G, Talbot C. *Mon. Not. R. Astron. Soc.* 507(2):2037–51 (2021)
83. Biwer CM, et al. *Publ. Astron. Soc. Pac.* 131(996):024503 (2019)
84. Breschi M, Gamba R, Bernuzzi S. *Phys. Rev. D* 104(4):042001 (2021)
85. Cornish NJ. *Phys. Rev. D* 103(10):104057 (2021)
86. Roulet J, et al. *Phys. Rev. D* 106(12):123015 (2022)
87. Fairhurst S, et al. *Phys. Rev. D* 108(8):082006 (2023)
88. Tiwari V, Hoy C, Fairhurst S, MacLeod D. *Phys. Rev. D* 108(2):023001 (2023)
89. Wong KWK, Isi M, Edwards TDP. *Astrophys. J.* 958(2):129
90. Cornish NJ. arXiv:1007.4820 [gr-qc] (2010)
91. Zackay B, Dai L, Venumadhav T. arXiv:1806.08792 [astro-ph.IM] (2018)
92. Leslie N, Dai L, Pratten G. *Phys. Rev. D* 104(12):123030 (2021)
93. Cornish NJ. *Phys. Rev. D* 104(10):104054 (2021)
94. Narola H, et al. arXiv:2308.12140 [gr-qc] (2023)
95. Antil H, et al. *J. Sci. Comput.* 57(3):604–37 (2013)
96. Canizares P, Field SE, Gair JR, Tiglio M. *Phys. Rev. D* 87(12):124005 (2013)
97. Smith R, et al. *Phys. Rev. D* 94(4):044031 (2016)
98. Pürrer M. *Class. Quantum Gravity* 31(19):195010 (2014)
99. Cotesta R, Marsat S, Pürrer M. *Phys. Rev. D* 101(12):124040 (2020)
100. Morrás G, Nuño Siles JF, García-Bellido J. *Phys. Rev. D* 108(12):123025 (2023)
101. Chua AJK, Galley CR, Vallisneri M. *Phys. Rev. Lett.* 122(21):211101 (2019)

102. Khan S, Green R. *Phys. Rev. D* 103(6):064015 (2021)
103. Thomas LM, Pratten G, Schmidt P. *Phys. Rev. D* 106(10):104029 (2022)
104. Vinciguerra S, Veitch J, Mandel I. *Class. Quantum Gravity* 34(11):115006 (2017)
105. García-Quirós C, Husa S, Mateu-Lucena M, Borchers A. *Class. Quantum Gravity* 38(1):015006 (2020)
106. Morisaki S. *Phys. Rev. D* 104(4):044062 (2021)
107. Pathak L, Reza A, Sengupta AS. *Phys. Rev. D* 108(6):064055 (2023)
108. Pathak L, Munishwar S, Reza A, Sengupta AS. *Phys. Rev. D* 109(2):024053 (2024)
109. Fairhurst S, Green R, Hannam M, Hoy C. *Phys. Rev. D* 102(4):041302 (2020)
110. Green R, et al. *Phys. Rev. D* 103(12):124023 (2021)
111. Singer LP, et al. *Astrophys. J.* 795(2):105 (2014)
112. Morisaki S, Raymond V. *Phys. Rev. D* 102(10):104020 (2020)
113. Lee E, Morisaki S, Tagoshi H. *Phys. Rev. D* 105(12):124057 (2022)
114. Farr B, Ochsner E, Farr WM, O'Shaughnessy R. *Phys. Rev. D* 90(2):024018 (2014)
115. Romero-Shaw IM, et al. *Mon. Not. R. Astron. Soc.* 499(3):3295–319 (2020)
116. Brady PR, Fairhurst S. *Class. Quantum Gravity* 25(10):105002 (2008)
117. Karamanis M, et al. *Mon. Not. R. Astron. Soc.* 516(2):1644–53 (2022)
118. Farr B, Kalogera V, Lijten E. *Phys. Rev. D* 90(2):024014 (2014)
119. Owen AB. <https://artowen.su.domains/mc/> (2013)
120. Payne E, Talbot C, Thrane E. *Phys. Rev. D* 100(12):123017 (2019)
121. Romero-Shaw IM, Lasky PD, Thrane E. *Mon. Not. R. Astron. Soc.* 490(4):5210–16 (2019)
122. Romero-Shaw I, Lasky PD, Thrane E. *Astrophys. J. Lett.* 921(2):L31 (2021)
123. Payne E, et al. *Phys. Rev. D* 102(12):122004 (2020)
124. Dax M, et al. *Phys. Rev. Lett.* 130(17):171403 (2023)
125. Pankow C, Brady P, Ochsner E, O'Shaughnessy R. *Phys. Rev. D* 92(2):023002 (2015)
126. Islam T, Roulet J, Venumadhav T. arXiv:2210.16278 [gr-qc] (2022)
127. Cranmer K, Brehmer J, Louppe G. *PNAS* 117(48):30055–62 (2020)
128. Beaumont MA. *Annu. Rev. Stat. Appl.* 6:379 (2019)
129. Kingma DP, Welling M. arXiv:1312.6114 [stat.ML] (2022)
130. Abbott R, et al. *Astrophys. J. Suppl. Ser.* 267(2):29 (2023)
131. Chatziioannou K, et al. *Phys. Rev. D* 100(10):104004 (2019)
132. Cornish NJ. *Phys. Rev. D* 102(12):124038 (2020)
133. Talbot C, Thrane E. *Phys. Rev. Res.* 2(4):043298 (2020)
134. Biscoveanu S, Haster CJ, Vitale S, Davies J. *Phys. Rev. D* 102(2) (2020)
135. Talbot C, Thrane E, Biscoveanu S, Smith R. *Phys. Rev. Res.* 3(4):043049 (2021)
136. Zackay B, et al. *Phys. Rev. D* 104(6):063034 (2021)