Constraining astrophysical parameter bias due to glitches in future ground-based gravitational-wave detectors

Final Report for the 2026 LIGO SURF Program

Samantha Walker*
Barnard College, Columbia University, New York, NY 10027, USA

Mentors: Sophie Bini and Jake Summers
LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

(Dated: October 25, 2026)

Transient detector noise artifacts, or "glitches," that coincide in time and frequency with gravitational-wave signals can bias the inferred astrophysical parameters of compact binary coalescences, particularly the black hole spins. We investigate how this bias evolves as detector sensitivity increases by simulating binary black hole signals based on GW200129 and GW191109, contaminated by glitches, under projected sensitivities for LIGO A$^\sharp$ and Cosmic Explorer (CE). We consider two limiting scenarios in which either the glitch amplitude or the glitch signal-to-noise ratio (SNR) is held fixed relative to present-day detectors, and perform Bayesian parameter estimation on each simulated data set. We find that glitches injected with fixed SNR produce progressively smaller absolute biases in the effective spin parameters as the signal SNR increases, whereas glitches injected with fixed amplitude continue to induce biases of comparable magnitude despite substantially narrower posterior distributions. Lowering the minimum analysis frequency from 20 to 14 Hz further reduces the fixed-SNR glitch bias to a negligible level. Adopting a signal SNR of 150 as the threshold above which bias from a fixed-SNR glitch (with glitch SNR $\sim$ 6–8) becomes negligible, we simulate a population of binary black hole mergers and find that only 0.03% of detections in an A$^\sharp$ network and 2–5% in a CE network exceed it, suggesting that most future detections would still require mitigation if they overlapped a comparable glitch. These results show that the impact of glitches in future detectors depends critically on how glitch populations scale with detector sensitivity, and provide insight into when computationally expensive glitch-mitigation methods remain necessary.

I. INTRODUCTION

Gravitational waves (GWs) are traveling-wave perturbations of spacetime generated by an accelerating mass-energy quadrupole moment. Since 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO), Virgo Interferometer, and Kamioka Gravitational Wave Detector (KAGRA) have detected over 300 gravitational-wave events from compact binary coalescences (CBCs) to date [1].

For well-modeled sources such as CBCs, the properties of the source can be inferred from the GW signal using GW waveform models and Bayesian inference in a process known as parameter estimation (PE). For binary black hole (BBH) coalescences, these parameters include the masses of the black holes (BHs), their spins, the sky position, and the luminosity distance to the source, among other parameters. Accurate parameter estimation is crucial to characterizing GW sources, enabling population studies, tests of general relativity, and multi-messenger astronomy [2]. Measuring the BH spins is of particular interest, as they provide insight into the origin and evolution of the binary system, which may help address open questions regarding binary stellar evolution and the existence of BBH subpopulations [3–5].

The broad characteristics of the BH spins are encoded in two parameters: the effective inspiral spin $\chi_{\rm eff}$ and the effective precession spin $\chi_{\rm p}$, which describe the effective BH spin aligned with the orbital angular momentum and the effective BH spin parallel to the orbital plane, respectively [6]. BBHs whose progenitor stars formed and evolved together as a binary are more likely to have spins aligned with the orbital angular momentum, whereas BBHs formed through dynamical interactions are instead expected to exhibit nearly isotropic spin orientations [5]. Consequently, systems with a confidently measured negative value of $\chi_{\rm eff}$ or a large value of $\chi_{\rm p}$ are particularly informative for constraining BBH formation channels.

Among the BBH events observed to date, GW191109_010717 (hereafter GW191109) and GW200129_065458 (hereafter GW200129) have emerged as notable candidates for such spin configurations [7]. GW191109 has been inferred to have a negative value of $\chi_{\rm eff}$, while GW200129 exhibits evidence for a large value of $\chi_{\rm p}$ [7, 8].

However, these conclusions may be biased by transient detector noise, known as "glitches." Typically, the process of parameter estimation assumes as a baseline a transient signal and stationary Gaussian noise [9]. Glitches violate this model, and must be accounted for separately to minimize their influence. If a glitch is suspected, one method is to model it as a separate component of the data and subtract the best fitting estimate before proceeding with analysis [10]. A more robust approach is

* sfw2117@barnard.edu

joint CBC–glitch inference, in which the properties of the CBC signal and the glitch are inferred simultaneously. This allows the analysis to marginalize over multiple possible glitch realizations rather than relying on a single best-fit subtraction [11].

Despite these mitigation strategies, the issue of glitch-induced bias cannot always be resolved completely. The interpretation of both GW200129 and GW191109 has been shown to depend sensitively on the specific mitigation approach used, leaving the evidence for anti-aligned or precessing spin inconclusive [7, 10, 12–15]. Beyond glitches detectable in the data, Udall *et al.* [16] showed that low signal-to-noise-ratio (SNR) glitches can still significantly bias spin parameter estimates, even below the threshold of detectability.

These results highlight that glitch mitigation is not merely a technical consideration, but an important component of reliable GW parameter estimation. While existing methods can reduce the impact of glitches, the extent to which residual or undetected glitches may bias astrophysical inferences is not fully understood. This question will become increasingly important as detector sensitivity improves and the number of observed CBCs grows. Across all LIGO observing runs, about 300 events have been observed in roughly 3.5 years of total observing time [1, 17]. In comparison, LIGO A$^\sharp$, the final planned upgrade to the current LIGO detectors, is expected to observe thousands of GW events per year of observation [18, 19]. Following this upgrade, the upcoming third-generation detector Cosmic Explorer (CE) is expected to observe $\sim 10^5$ CBCs per year of observation [20, 21].

Currently, parameter estimation for a single source can take days, or longer if a glitch overlaps with the signal, which occurred in about 36% of cases in GWTC-5 [22]. Moreover, the process of joint CBC-glitch inference is not fully automated and must be manually assessed at each step. With the increased detection rates of third-generation detectors, it will no longer be practical to treat every glitch-contaminated event with the same level of manual intervention. Additionally, with more events available for population studies and increasingly precise measurements of individual sources, even relatively small systematic biases could affect our interpretation of the underlying astrophysical population. It is therefore crucial to characterize how glitches influence parameter estimation under the conditions expected for future detectors, and to determine when such biases may become significant.

In this work, we investigate how glitch-induced parameter bias evolves with increasing detector sensitivity. We perform two injection studies based on GW200129 and GW191109, injecting simulated BBH signals together with a single Morlet-Gabor wavelet glitch, using glitch configurations that Udall *et al.* [16] showed strongly bias $\chi_{\rm p}$ and $\chi_{\rm eff}$, respectively. Holding the source fixed, we increase the detector sensitivity from that of LIGO during its third observing run (O3) to that of A$^\sharp$ and CE, considering two limiting scenarios in which either the glitch amplitude or the glitch SNR is held fixed. We find that fixed-SNR glitches produce raw biases that decrease as the signal SNR increases and eventually become negligible, whereas fixed-amplitude glitches produce roughly constant raw biases that become increasingly significant relative to the narrowing posteriors, in accordance with our expectations [16]. Finally, we use these results to estimate the fraction of future detections for which glitch mitigation may be unnecessary. In our fixed-SNR runs, where the glitch SNR is held at its O3 value ($\rho_{\rm g} \approx 8$ for GW200129 and $\rho_{\rm g} \approx 6$ for GW191109, inducing initial biases of $\delta\chi_{\rm p} \approx 0.58$ and $\delta\chi_{\rm eff} \approx -0.56$, respectively), the raw bias in the focused spin parameter falls below ≈ 0.15 once the signal SNR exceeds $\rho \approx 150$. Using a simulated population of BBH mergers, we find that only 0.03% of detections in an A$^\sharp$ network and 2–5% in a CE network exceed this SNR, so most future detections may still be susceptible to glitches of this kind.

The remainder of this paper is organized as follows. Section II reviews detector sensitivity, the expected scaling of glitch-induced bias, and our glitch model. Section III describes the data generation and parameter estimation procedure. Section IV presents the bias results for both glitch scaling scenarios, and Section V examines the effect of lowering the minimum frequency. In Section VI, we estimate the fraction of future detections that may still require glitch mitigation based on our results, and Section VII summarizes our conclusions and future work. The appendices examine waveform systematics, bias in parameters other than $\chi_{\rm eff}$ and $\chi_{\rm p}$, and the detailed injection parameters.

II. BACKGROUND

A. Characterizing detector sensitivity

Assuming stationary Gaussian noise, the sensitivity of a GW detector is reflected in its power spectral density (PSD) or amplitude spectral density (ASD) function, where the ASD is the square root of the PSD. The PSD and ASD are functions of frequency and are measured in units of Hz^{-1} and $\mathrm{Hz}^{-1/2}$ respectively, with lower values indicating greater sensitivity. Figure 1 illustrates the ASD for LIGO detectors during O3 and the projected ASD for LIGO A$^\sharp$ detectors and CE. In order to simulate increased sensitivity of the detector, we vary the PSD throughout our analysis.

The PSD also defines the relationship between "colored" and "whitened" strain data. Colored strain refers to strain whose frequency content is shaped by the detector's frequency-dependent noise. Whitening divides out this frequency dependence through the relation

$$\tilde{h}_{\rm white}(f) = \frac{\tilde{h}(f)}{\sqrt{S_n(f)}}, \tag{1}$$

where $\tilde{h}(f)$ is the Fourier transform of the colored strain

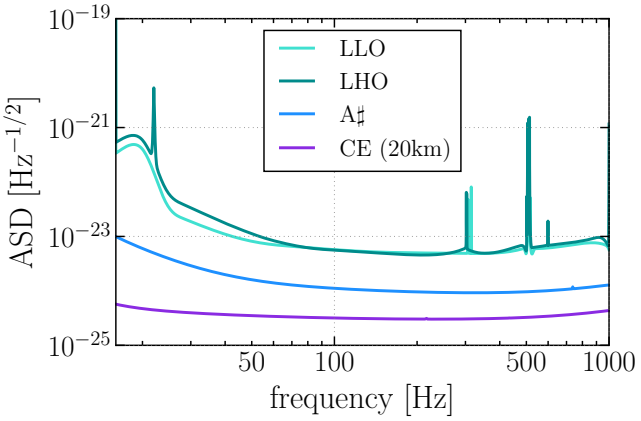


FIG. 1. Amplitude spectral density (ASD) as a function of frequency for the LIGO Livingston Observatory (LLO) and LIGO Hanford Observatory (LHO) during the third observing run (O3), shown in green and teal respectively, along with the projected sensitivities of LIGO A$^\sharp$ (blue) and the 20 km Cosmic Explorer (CE) detector (purple).

and $S_n(f)$ is the PSD. This suppresses frequency components in regions of high detector noise and emphasizes those in regions of low detector noise.

B. Quantifying parameter bias

The GW strain amplitude is fixed by the physics of the source, but with increased detector sensitivity, we expect the noise to decrease, thus increasing the signal SNR. However, it is unclear how the glitches will be impacted. In this paper, we consider two limiting cases: 1) the glitch amplitude remains constant and 2) the glitch SNR remains constant, relative to glitches observed in current detectors.

To quantify the parameter bias in these two cases, we consider two different measures, the first being the raw or absolute bias, which is defined as

$$\delta\theta^\alpha \equiv \hat{\theta}^\alpha - \theta_0^\alpha, \tag{2}$$

where $\hat{\theta}^\alpha$ are the maximum-likelihood parameters and θ_0^α are the injection parameters. The second is the standardized bias, which is defined as $\delta\theta^\alpha/\sigma_\alpha$ where σ_α is the standard deviation of the parameter θ^α.

Based on an approximate Fisher matrix scaling argument outlined in Udall *et al.* [16], we expect the two glitch scaling cases to affect the bias differently. Our simulations aim to test whether these expectations hold.

Consider an increase in detector sensitivity which scales the BBH signal SNR from $\rho \to \kappa\rho$. In the case of fixed glitch SNR, the raw bias is expected to scale approximately as

$$\delta\theta^\alpha \propto 1/\kappa. \tag{3}$$

As the standard deviation follows the same scaling, the standardized bias then remains constant with increasing SNR.

Conversely, for the fixed glitch amplitude case, the raw bias is predicted to remain approximately constant, leaving the standardized bias to scale accordingly with the standard deviation:

$$\delta\theta^\alpha/\sigma_\alpha \propto \kappa. \tag{4}$$

Thus, in the fixed-SNR case, absolute bias is expected to decrease, while the statistical deviation of the maximum likelihood parameters from truth will remain the same. In the fixed-amplitude case, the absolute bias is expected to remain the same, while the statistical deviation from truth will increase. Note that these scalings assume the validity of the linear-signal approximation, which generally applies in the high-SNR regime [16, 23–25]. Our simulations use a variety of signal SNR values to test whether these expectations hold.

C. Modeling the glitch

Glitches can be modeled as one or more Morlet-Gabor wavelets, given by the expression

$$\Psi(t) = Ae^{-(t-t_c)^2/\tau^2}\cos(2\pi f_c(t-t_c)+\phi_0), \tag{5}$$

where $\tau = Q/(2\pi f_c)$ and A is the amplitude, Q is the quality factor, t_c and f_c are the central time and frequency, and ϕ_0 is the phase offset [26].

For the fixed amplitude case, the glitch form will be modeled identically between current and future detectors. For the fixed SNR case, the amplitude of the glitch must be scaled accordingly. Here, we derive the amplitude scaling relation used to fix the SNR of a glitch.

The SNR of an arbitrary GW strain time series $h(t)$ is defined as

$$\rho = \left[4\int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)}\,df\right]^{1/2}, \tag{6}$$

where $\tilde{h}(f)$ is the Fourier transform of $h(t)$, and $S_n(f)$ is the PSD of the detector. When h is the noise-free signal waveform, Eq. (6) gives the optimal SNR, the SNR that would be recovered by a filter perfectly matched to the signal. Throughout, ρ denotes the optimal network SNR of the BBH signal, obtained by adding the single-detector SNRs in quadrature, and $\rho_{\rm g}$ denotes the SNR of the glitch.

The Fourier transform of Equation (5) is given by

$$\begin{aligned}\tilde{\Psi}(f) = &\frac{\sqrt{\pi}A\tau}{2}e^{-2\pi i f t_c}\\ &\times\left(e^{i\phi_0-\pi^2\tau^2(f-f_c)^2}+e^{-i\phi_0-\pi^2\tau^2(f+f_c)^2}\right).\end{aligned} \tag{7}$$

Assuming a narrow wavelet relative to changes in the PSD, we can take $S_n(f) \approx S_n(f_c)$ over the width of the glitch. Neglecting exponentially suppressed terms, Equation (6) becomes

$$\rho_{\rm g}^2 \approx \frac{\pi A^2 \tau^2}{S_n(f_c)} \int_0^\infty e^{-2\pi^2\tau^2(f-f_c)^2}\, df. \tag{8}$$

Evaluating the Gaussian integral and simplifying yields the SNR approximation for a Morlet-Gabor wavelet:

$$\boxed{\rho_{\rm g} \approx \frac{A\sqrt{Q}}{\sqrt{2\sqrt{2\pi} f_c S_n(f_c)}}.} \tag{9}$$

Thus, to maintain a constant glitch SNR, we rescale the glitch amplitude following

$$A' = A\sqrt{\frac{S_n'(f_c)}{S_n(f_c)}}, \tag{10}$$

where primed quantities denote the modified glitch amplitude and PSD, while all other parameters remain fixed.

III. METHODS

A. Data generation procedure

We perform two injection studies inspired by the events GW200129 and GW191109, using glitch parameters that have been shown to cause substantial bias to the spins [16]. We consider three main detector configurations, LIGO O3, LIGO A$^\sharp$, and CE, whose ASDs are pictured in Figure 1. As the future site of CE has not yet been chosen, we assume the location of two 20 km CE detectors to be at the present sites of LIGO Livingston Observatory (LLO) and LIGO Hanford Observatory (LHO). Thus, all configurations are identical in location and orientation, with only their sensitivity varied.

To study a range of signal SNRs within each detector regime, we uniformly scale the detector PSD while preserving its spectral shape. This is equivalent to rescaling the source distance while keeping the detector-frame chirp mass fixed. The scaling predictions of Section II B assume exactly this kind of uniform rescaling, so they apply directly within each detector regime. However, between the three detector PSDs used, the spectral shape also changes, which can alter posterior widths and biases beyond the simple dependence on SNR (e.g., through improved low-frequency sensitivity), and thus the predictions should be regarded as approximate across regimes.

We generated six full sets of data frames with scaled PSDs in addition to the three original detector configurations: four created with a scaled A$^\sharp$ PSD and two created with a scaled CE PSD, with the baseline PSD chosen based on the SNR regime, for a total of 9 sets of generated data for each event. Each of the detector configuration cases and their corresponding network SNR are summarized in Table I. Hereafter, we will refer to the runs performed with the original (unscaled) detector PSD as the "principal" runs, referring to O3*, A$^\sharp$*, and CE* in Table I.

We use Bilby [27] to simulate stationary Gaussian noise colored by the PSD of the chosen detector configuration, generating an independent noise realization for each of the two detector locations. For the O3 configuration, we use separate PSDs for the Hanford and Livingston detectors, estimated from detector data surrounding the corresponding real event [7]. For the A$^\sharp$ and CE configurations, only projected design sensitivities are available, so both detector locations share the same PSD.

For each configuration, we produce 4 seconds of data with a sampling frequency of $f_{\rm samp} = 1024$ Hz, minimum frequency of $f_{\rm min} = 20$ Hz, and maximum frequency of $f_{\rm max} = 448$ Hz. Then, we inject a CBC signal using the waveform model NRSur7dq4 [28] with injection parameters inspired by either GW200129 or GW191109 [7], listed in Appendix C.

For the detector located at the LLO site, we save a copy of the uncontaminated data before injecting a glitch, modeled as a single Morlet-Gabor wavelet. The glitch amplitude and SNR for the LIGO O3 configuration are preset to $A = 3.47 \times 10^{-22}$ and $\rho_{\rm g} = 7.96$ for GW200129, and $A = 3.14 \times 10^{-22}$ and $\rho_{\rm g} = 6.00$ for GW191109. For the other configurations, these parameters vary according to whether we are modeling fixed glitch amplitude or fixed glitch SNR. In the case of fixed glitch SNR, the amplitude is scaled with Equation (10), otherwise, the glitch form is not modified and the SNR scales accordingly as the detector noise decreases. The full set of amplitude and SNR values for each scaling case and configuration are listed in Table I. The other glitch parameters are kept constant between detector configurations and are listed in Table II. The glitch parameters were chosen to create substantial bias in $\chi_{\rm p}$ for GW200129 and in $\chi_{\rm eff}$ for GW191109, as found in Udall *et al.* [16].

The choice of glitch SNR ($\rho_{\rm g} = 7.96$ for GW200129 and $\rho_{\rm g} = 6.00$ for GW191109) was motivated by the following three considerations. First, glitches with $\rho_{\rm g} \lesssim 5$ cannot be reliably identified by current detection tools [29], so glitches near this threshold may go unmitigated entirely. Second, glitches with $\rho_{\rm g} \sim 6$–8 are among the most common in current detectors [30]. Third, louder glitches that are identified and subtracted leave behind residuals whose SNR is set by the complexity of the glitch model rather than the original glitch SNR, typically $\rho_{\rm res} \sim 3$–7 [16], so even well-mitigated loud glitches can leave power comparable to the glitches we consider.

Figure 2 shows the components of the simulated LLO data including a glitch for both events in two of the detector configurations: O3* and A$^\sharp$* (see Table I). The noise, BBH signal, and glitch are plotted separately; the strain used in each run is the sum of the noise, the signal, and (for the glitch runs) one of the two glitch variants

		Fixed glitch SNR		Fixed glitch amplitude	
Detector	**Signal SNR**	Glitch amplitude	Glitch SNR	Glitch amplitude	Glitch SNR
		GW200129			
O3*	25	3.47	8	3.47	8
A$^\sharp$	40	2.26	8	3.47	12
A$^\sharp$	75	1.21	8	3.47	23
A$^\sharp$	104	0.87	8	3.47	32
A$^\sharp$*	139	0.65	8	3.47	42
A$^\sharp$	173	0.52	8	3.47	53
CE	294	0.30	8	3.47	93
CE	504	0.17	8	3.47	160
CE*	671	0.13	8	3.47	213
		GW191109			
O3*	16	3.14	6	3.14	6
A$^\sharp$	25	2.23	6	3.14	8
A$^\sharp$	46	1.19	6	3.14	16
A$^\sharp$	64	0.86	6	3.14	22
A$^\sharp$*	85	0.64	6	3.14	29
A$^\sharp$	106	0.51	6	3.14	37
CE	188	0.23	6	3.14	83
CE	322	0.14	6	3.14	142
CE*	430	0.10	6	3.14	189

TABLE I. Detector configurations and signal network SNR of the injected event for each set of runs, alongside the corresponding glitch amplitude and SNR under the fixed-SNR and fixed-amplitude scenarios (SNR held fixed at, or amplitude held fixed at, the LIGO O3 value respectively). Amplitude in units of $\times 10^{-22}$. Configurations marked * use the original (unscaled) PSD, all other configurations uniformly scale the corresponding detector PSD while preserving its spectral shape.

Parameter	**GW200129**	**GW191109**
$t_c - t_{\rm det}$ (s)	-0.053	-0.045
f_c (Hz)	43.56	33.58
Q	10.71	8.73
ϕ_0 (rad)	0.0	π

TABLE II. Parameters of the Morlet-Gabor wavelet glitch, given by Equation (5), injected into the detector located at the LLO site for all events and detector configurations. The time offset $t_c - t_{\rm det}$ is the central time of the glitch wavelet minus the BBH coalescence time in the LLO detector frame. The glitch amplitude and SNR are not included as they vary between configurations, and are instead listed in Table I.

depending on the scaling case.

B. Parameter estimation

We perform parameter estimation on the generated data frames with Bilby [27] using the Dynesty sampler [31] and the default LIGO-Virgo-KAGRA priors (uniform in masses, spin magnitudes, merger time, and coalescence phase, isotropic in spin and binary orientations and sky location) [22]. NRSur7dq4 is the waveform model used for inference, matching what is used for data generation. NRSur7dq4 interpolates between numerical-relativity waveforms, making it typically the most accurate waveform model for high-mass signals [28].

Probability distributions for the values of each BBH parameter, known as "posteriors," are produced from these Bilby runs. Due to increased detector sensitivity, we expect the width of the posteriors to decrease, scaling with $1/\rho$ where ρ is the SNR of the signal, or equivalently scaling with the square root of the detector PSD for a fixed signal, following the linear signal approximation [24].

For each of the detector configurations we perform a Bilby run with uncontaminated data in both LHO and LLO as a control, and then another run on the same data with the addition of the glitch in the LLO data. For the A$^\sharp$ and CE configurations, two such glitch runs are performed, one for each glitch scaling case. The Bilby run structure is summarized in Figure 3.

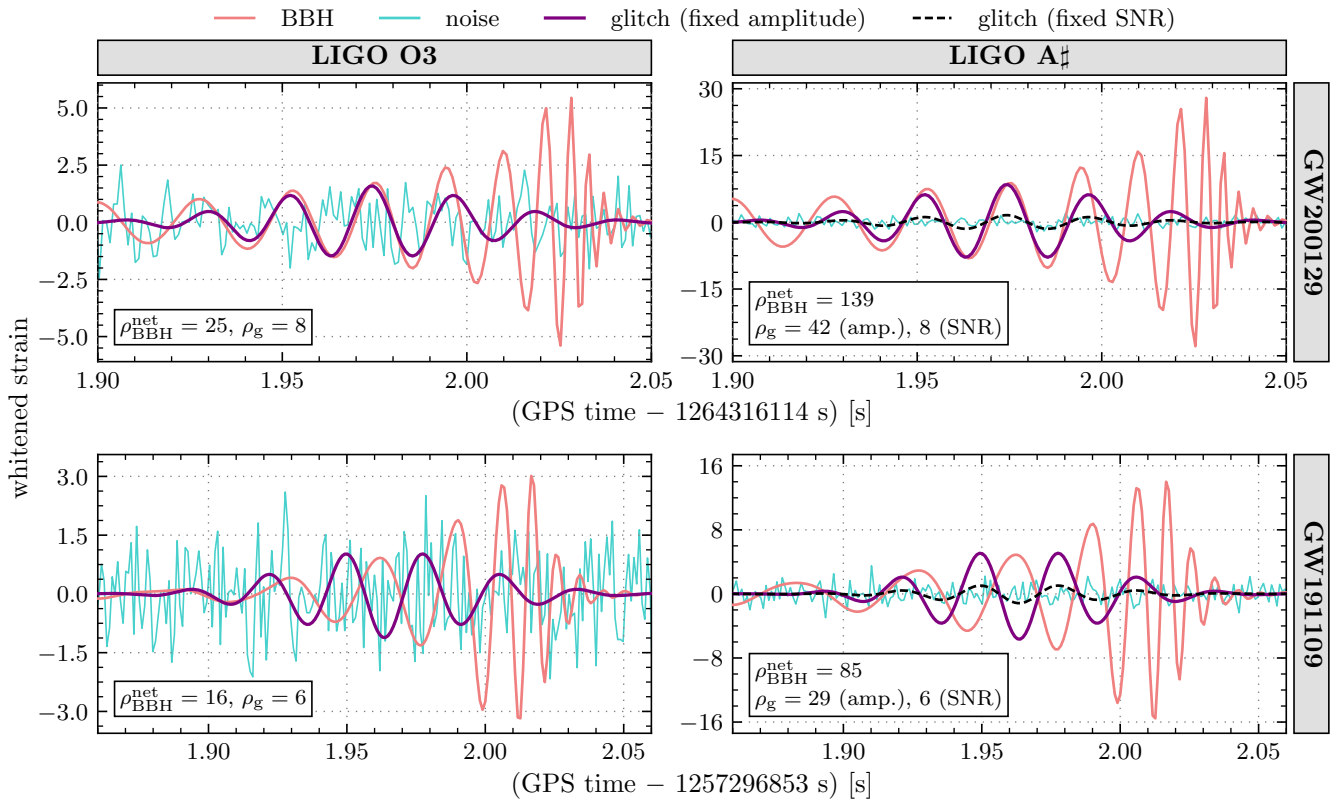


FIG. 2. Whitened LIGO Livingston strain for each component of the simulated data, for two detector configurations: LIGO O3 (left column; O3* in Table I) and LIGO A$^\sharp$ (right column; A$^\sharp$* in Table I). The top row shows the GW200129-like event and the bottom row the GW191109-like event. In every panel the BBH signal is drawn in red and the Gaussian noise in light blue. In the O3 panels, the glitch (purple) uses the parameters given in Table II. The A$^\sharp$ panels show both glitch scaling cases: the fixed-amplitude glitch (purple) keeps the same strain amplitude as in O3, while the fixed-SNR glitch (black dashed) has its amplitude reduced so that its SNR matches the O3 value. Because the strain is whitened, the noise has roughly unit variance in both configurations, so the lower noise level of A$^\sharp$ appears instead as a louder BBH signal and a louder fixed-amplitude glitch. The box in the lower left of each panel gives the network SNR of the BBH signal, $\rho^{\rm net}_{\rm BBH}$, and the SNR of the glitch, $\rho_{\rm g}$, listed for the fixed-amplitude (amp.) and fixed-SNR (SNR) cases in the A$^\sharp$ panels. Note that for all runs, the glitch is injected only into the Livingston data, so its network SNR is its Livingston SNR. All SNRs are rounded to the nearest integer.

IV. PARAMETER ESTIMATION RESULTS

Using the results from the Bilby runs, we test the predictions from Section II B for our two simulated signals, GW200129-like and GW191109-like. For each signal, we plot the glitch-induced bias in the focused parameter ($\chi_{\rm p}$ for GW200129, $\chi_{\rm eff}$ for GW191109) for each of the runs listed in Table I against signal SNR under both glitch-scaling cases. For an examination of the bias in other parameters, see Appendix B. In general, increasing detector sensitivity did not shift the glitch's impact onto parameters that were unbiased in O3: parameters biased in the O3 run became similarly or less biased in the fixed-SNR case, and kept a similar pattern of bias in the fixed-amplitude case.

The scaling predictions of Section II B are derived for the maximum-likelihood point, as the bias in Eq. (2) is formulated in terms of the maximum-likelihood parameters rather than any other point estimate. In practice, however, the posterior median and its 90% credible interval are typically the reported point estimate. The two are expected to coincide in the high-SNR, linear-signal-approximation regime, but may diverge otherwise, so we show both throughout: the bias of the maximum-likelihood point, fit against the predicted scaling, and the bias of the median, the point estimate that would actually be reported for a real event. Fits are performed by ordinary least squares (OLS) against the maximum-likelihood points only, as a visual check of the predicted scaling. Because the maximum-likelihood point is a single sample with no associated uncertainty of its own, we show the error in bias as the posterior's 90% credible interval, centered on the median; this indicates the width of the posterior rather than a sampling uncertainty on the maximum-likelihood point itself.

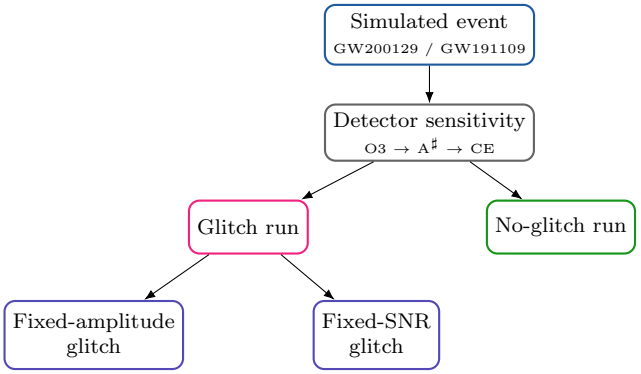


FIG. 3. Structure of the BILBY runs listed in Table I. For each simulated event, the CBC signal is injected and recovered with NRSUR7DQ4 at $f_{\rm min} = 20$ Hz. We hold the source parameters fixed and vary the detector sensitivity from O3 to A♯ to CE, so the signal SNR increases with sensitivity. At each sensitivity, we perform a glitch-contaminated run and a clean baseline run. Glitch runs are further split into fixed-amplitude and fixed-SNR scaling scenarios, generated from the same noise realization. The two scenarios coincide at O3, where the glitch has its original SNR, so O3 serves as the common anchor and the scenarios differ only for A♯ and CE.

A. Fixed-SNR glitch case

The glitch was held at SNR ≈ 8 for the GW200129-like signal and SNR ≈ 6 for the GW191109-like signal. For both, we observe a decrease in absolute bias in $\chi_{\rm p}$ and $\chi_{\rm eff}$, respectively, as the signal SNR increases. Figures 4 and 5 show the raw and standardized bias as a function of network SNR; the OLS fit to Equation (3) is overlaid as a gray dashed line. Both figures show rough agreement with the trend predicted for the raw bias, which decreases substantially with increased SNR. Outliers at low SNRs are expected, as the scaling relations were derived assuming the high-SNR limit. For GW191109, the standardized bias remains roughly constant as expected, with the lowest-SNR point (corresponding to the O3 configuration) the primary outlier. However, the standardized bias progressively decreases for GW200129. This could be either a reflection of the raw bias decreasing more rapidly than expected, or an indication that the posterior widths are decreasing more slowly than expected.

To resolve this, we plotted the posterior widths against the signal SNR ρ for the glitch and no-glitch runs in Figure 6. We expect the standard deviation σ to scale as $1/\rho$ [24], and fit a power law model to the data using OLS to test this expectation. Due to the log-log axes of the graph, the expected scaling manifests as a slope of -1. For the GW191109-like event we obtain a best-fit slope of -0.94 for the glitch runs and -0.99 for the no-glitch runs, in general agreement with the expectation. For the GW200129-like event, we instead observe a slope of -0.78 for the glitch runs and -0.77 for the no-glitch runs, indicating that the posterior widths are decreasing more slowly than expected and explaining some of the progressive decrease in standardized bias seen in the lower panel of Figure 4. The consistency in slope between the glitch and no-glitch runs for each simulated event suggests that the presence of the glitch is not driving this behavior, despite some marginal differences in individual GW200129-like runs observed in Figure 6.

B. Fixed-amplitude glitch case

The bias results are plotted in Figures 7 and 8 for the fixed-amplitude case. The glitch amplitude was held at $A = 3.47 \times 10^{-22}$ for GW200129 and $A = 3.14 \times 10^{-22}$ for GW191109. As the signal SNR increases, the posteriors narrow but do not move closer to the truth value or to the no-glitch posterior. Both scaling predictions are approximately followed: the raw bias stays roughly constant, so the standardized bias grows as the posteriors narrow. The main deviation is a slight increase in raw bias for the GW200129-like event.

V. LOWERING THE MINIMUM FREQUENCY

In addition to improved sensitivity in the current band, future detectors will extend sensitivity to lower frequencies, with projected low-frequency cutoffs of $f_{\rm min} \approx 10$ Hz for A♯ and $f_{\rm min} \approx 5$ Hz for CE. The analyses in Section IV use $f_{\rm min} = 20$ Hz for every detector configuration, so they capture the effect of the improved PSD but not this additional low-frequency data. Here we test how including it changes the glitch-induced bias.

Lowering $f_{\rm min}$ adds inspiral cycles to the observed signal, which tightens constraints on the source parameters [32]. However, certain glitches, such as scattered-light glitches, can extend into lower frequencies and have an increased impact with these frequencies taken into account. Whether the bias decreases therefore depends on which effect dominates.

To test this for our particular setup, we regenerate the A♯* fixed-SNR frame for the GW200129-like event (Table I) with $f_{\rm min} = 14$ Hz, both with and without the glitch. The value of 14 Hz is the lowest starting frequency the NRSUR7DQ4 waveform model supports for this system, above the projected A♯ cutoff, but is sufficient to demonstrate the effect.

We perform this test only for the fixed-SNR scenario. As shown in Section IV B, improving detector sensitivity does nothing to reduce the raw bias from a fixed-amplitude glitch, and because the posteriors narrow as sensitivity increases, the same raw bias becomes increasingly significant, reaching $\sim 100\sigma$ in A♯ and CE. Such a glitch cannot be left in the data, and with SNRs far above current detection thresholds (see Table I), it would also be easy to identify. The question of whether mitigation is necessary therefore only remains open for the fixed-SNR glitches, whose SNR stays near or below the detection threshold as detectors improve. The same scenario also

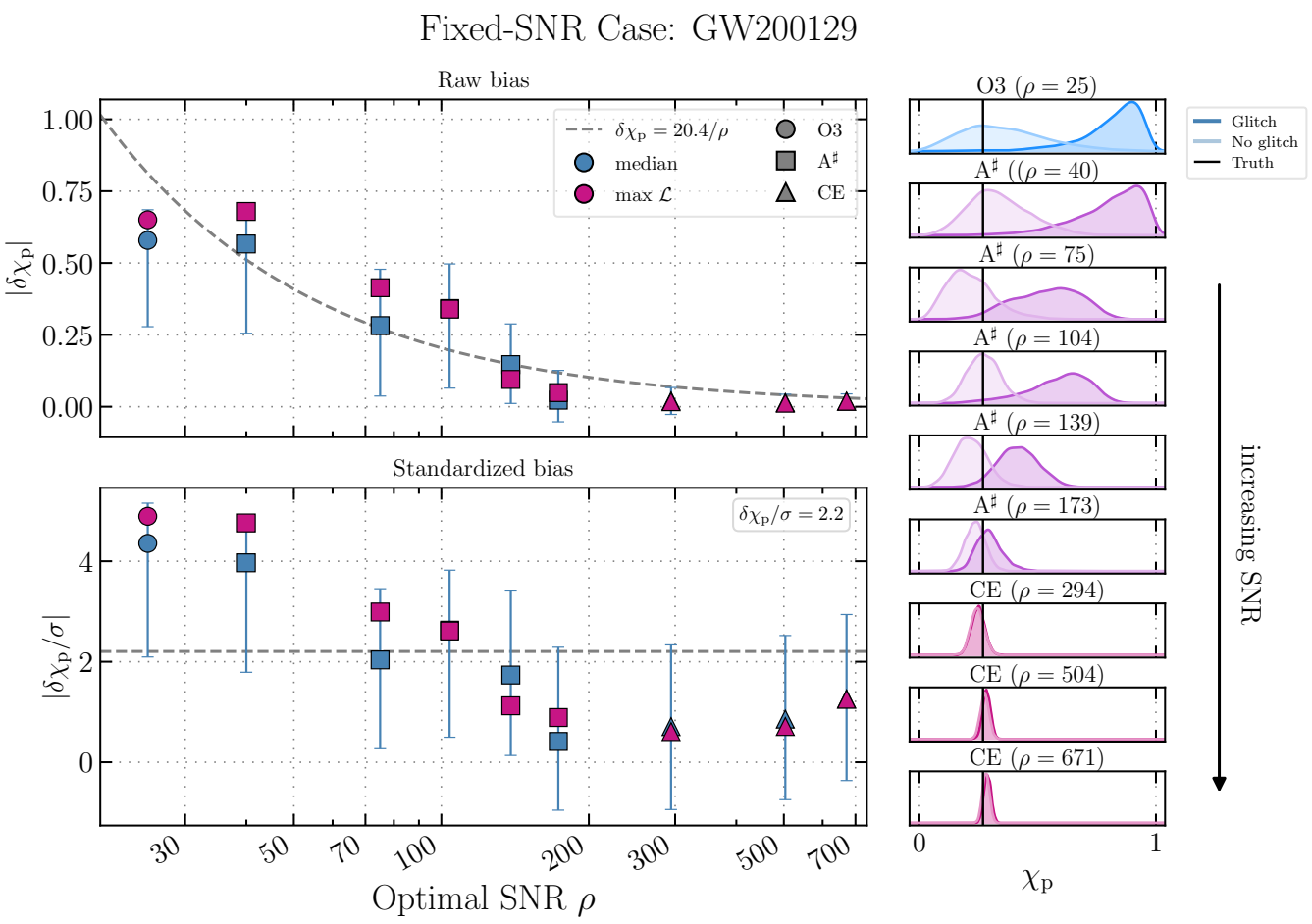


FIG. 4. Bias trends and marginalized χ_p posteriors for the fixed-SNR glitch scaling case and GW200129-like event. The top left panel shows the raw biases for the fixed-SNR glitch runs for all detector configurations. The expected trend given by Equation (3) is fit using the ordinary least squares (OLS) method to the data and drawn as a gray dashed line. Similarly, the bottom left panel shows the same points but in terms of standardized bias. The corresponding expected trend is fit to the data points using OLS with a dashed gray line. The right panels display the marginalized posteriors for χ_p for each of the runs plotted left, with the no-glitch runs in lighter shading and the truth value drawn as a black vertical line.

describes the residual left after a high-SNR glitch is subtracted; because subtraction is never perfect, it leaves a lower-SNR residual whose SNR is set by the accuracy of the glitch model rather than by detector sensitivity [16]. Our fixed-SNR results therefore apply both to glitches that are never mitigated and to what remains of those that are.

Figure 9 shows a comparison between the glitch and no-glitch posteriors for $f_{\rm min} = 20$ Hz and $f_{\rm min} = 14$ Hz. As expected, the no-glitch posterior narrows when $f_{\rm min}$ is lowered. The glitch-affected posterior also narrows, but more notably, shifts back toward the injected value. For this case, lowering $f_{\rm min}$ makes the glitch-induced bias negligible, recovering essentially the measurement that would have been made without the glitch. A glitch with more low-frequency power relative to the signal could behave differently, so we do not expect this result to hold universally.

This result has a direct implication for our main analysis. Because we held $f_{\rm min} = 20$ Hz fixed across all detectors to isolate the effect of the PSD, the fixed-SNR biases reported in Section IV are likely overestimates for A♯ and CE, whose extended low-frequency sensitivity would further suppress the glitch's impact. We discuss this further in Section VI, where we estimate the signal SNR above which glitches of this kind no longer meaningfully bias spin inference.

VI. WHEN IS GLITCH MITIGATION NO LONGER NEEDED?

As discussed in Section I, current glitch mitigation pipelines are likely unsustainable for the increased event detection rates expected for future detectors. Thus, it is important to determine what proportion of future events will actually require glitch mitigation for accurate parameter estimation, in order to minimize the associated

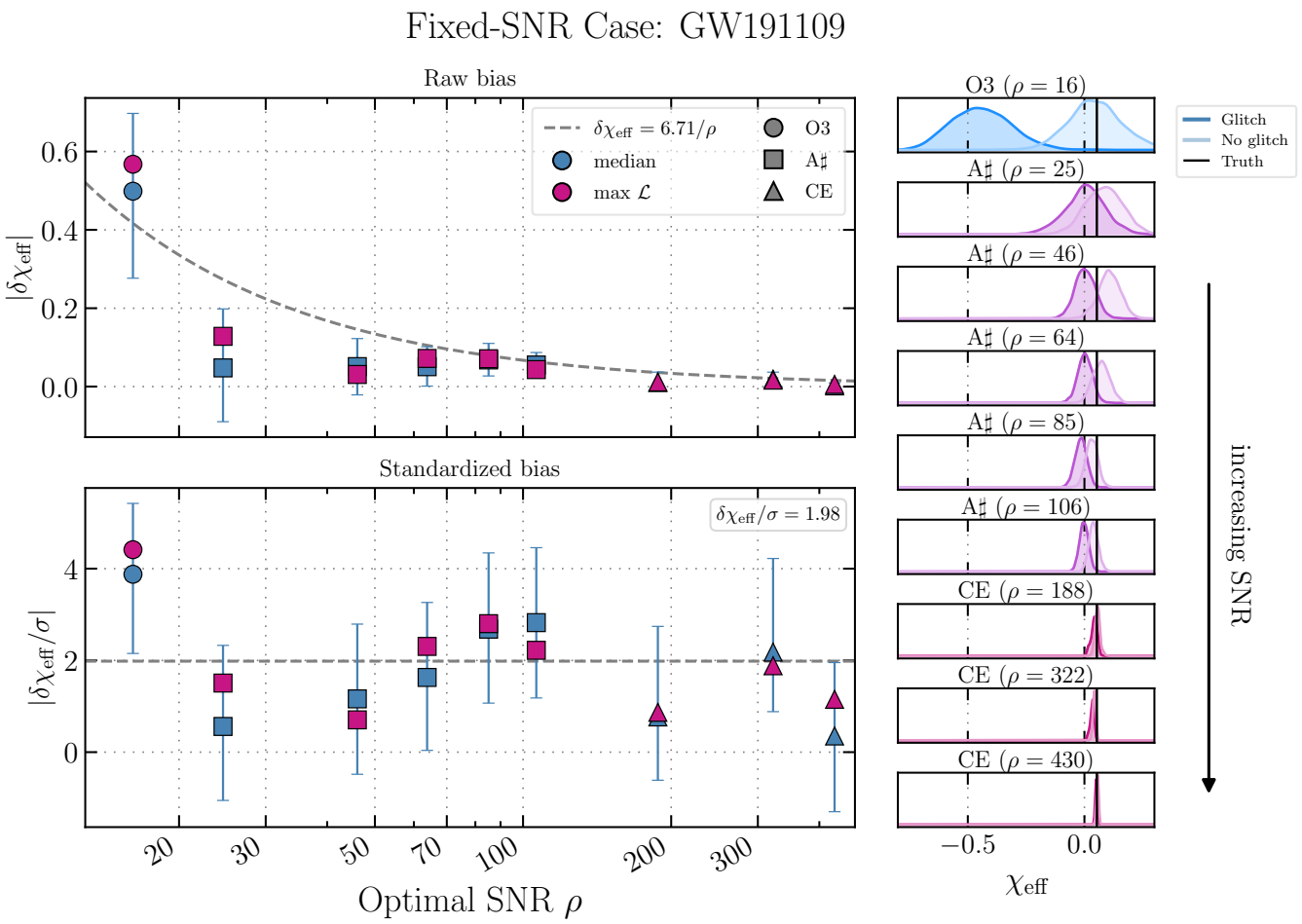


FIG. 5. Same as Figure 4, but for the GW191109-like event.

computational costs and manual labor involved. In this section, we discuss our results from Sections IV and V to determine a minimum signal SNR above which our fixed-SNR glitch runs experience negligible parameter bias. Then, we simulate one year of BBH mergers drawn from the GWTC-5 population model [33] and compute their network SNRs in several future detector networks. This lets us estimate how many detected events per year fall below this cutoff and would therefore still require glitch mitigation if a glitch similar to those we simulated overlapped the signal.

In the fixed-glitch-SNR scenario, the raw bias in $\chi_{\rm eff}$ and $\chi_{\rm p}$ decreases as the signal SNR increases (Sections IV and V). Here, we consider the bias negligible when the raw bias falls below ≈ 0.15. For the GW200129-like event, with injected $\chi_{\rm p} = 0.269$, this corresponds to a recovered value in the range $0.12 \lesssim \chi_{\rm p} \lesssim 0.42$, and for the GW191109-like event, with injected $\chi_{\rm eff} = 0.053$, to the range $-0.10 \lesssim \chi_{\rm eff} \lesssim 0.20$. Our runs satisfied this requirement for signal SNRs of $\rho \gtrsim 150$; thus, we adopt $\rho = 150$ as the SNR above which glitch mitigation is no longer needed. Below, we outline the population model and discuss results for three example detector networks: two A$^\sharp$ detectors with $f_{\rm min} = 10\,$Hz, two 20 km CE detectors with $f_{\rm min} = 5\,$Hz, and a 40 km CE detector at Hanford paired with a 20 km CE detector at Livingston with $f_{\rm min} = 5\,$Hz.

A. Simulated BBH population

We generate a synthetic BBH population using the maximum-likelihood hyperparameters of the population analysis for the DEFAULT model of GWTC-5 [33]. The primary mass follows a broken power law with two Gaussian peaks, tapered at low masses by a Planck smoothing function, and the mass ratio follows a power law with a smoothed lower cutoff on the secondary mass. Spin magnitudes are drawn independently for each black hole from a Gaussian truncated to $[0, 1]$. Spin tilts are drawn from a mixture of an isotropic component and a truncated Gaussian in $\cos\theta_i$, and the azimuthal spin angles ϕ_{12} and ϕ_{JL} are uniform on $[0, 2\pi)$.

Redshifts are drawn on $0 \leq z \leq 20$ from

$$p(z) \propto \frac{\psi(z)}{1+z}\frac{dV_c}{dz}, \tag{11}$$

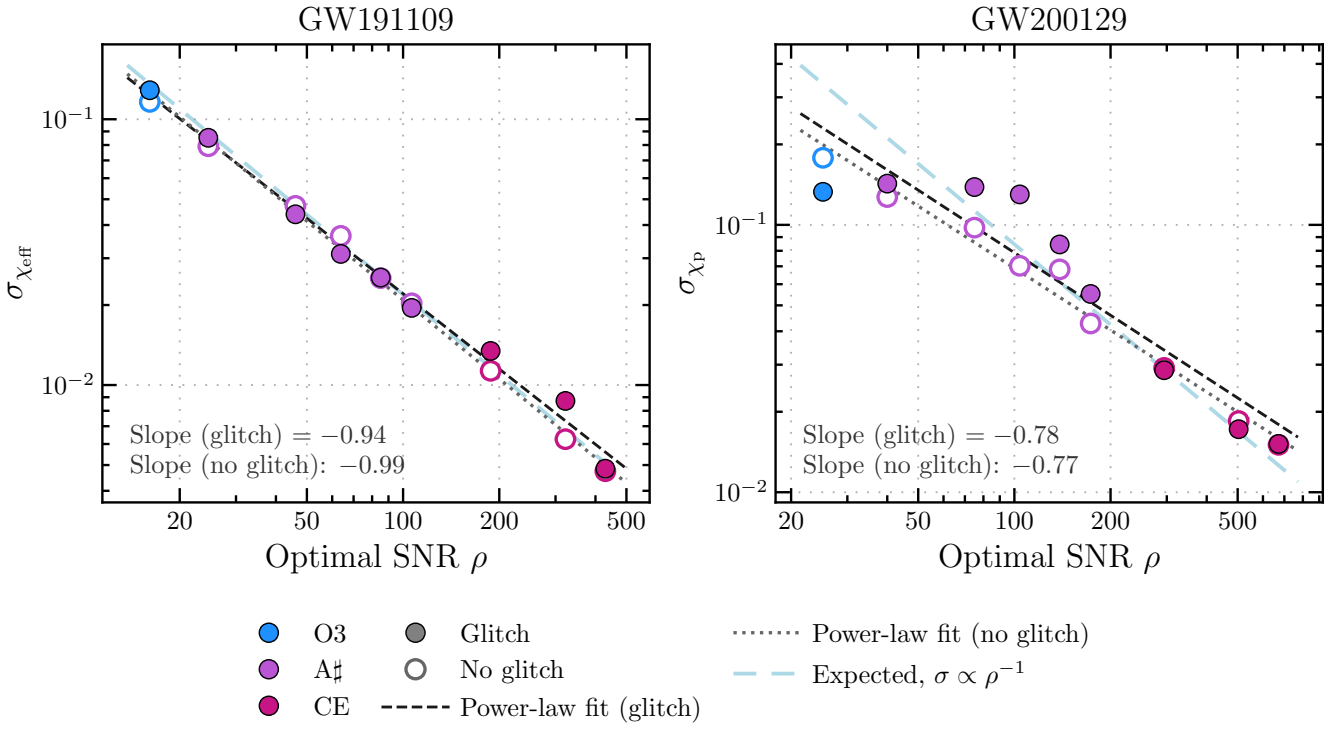


FIG. 6. Posterior widths of $\chi_{\rm eff}$ for GW191109 (left) and $\chi_{\rm p}$ for GW200129 (right), plotted against the signal optimal SNR ρ, for the fixed-SNR glitch runs (filled circles) and the corresponding no-glitch runs (open circles). Blue points indicate the original O3 run, and purple and magenta points correspond to runs in the A$^\sharp$ and CE regimes respectively. Power-law fits to the glitch and no-glitch points are shown as dark gray dashed and gray dotted lines, with slopes given in the top right corner of each panel: −0.94 (glitch) and −0.99 (no glitch) for GW191109, and −0.78 and −0.77 for GW200129. The light blue dashed line shows the expected $\sigma \propto \rho^{-1}$ scaling, fit to the glitch data points.

where dV_c/dz is the differential comoving volume in the Planck 2018 cosmology [34] and the factor $(1+z)^{-1}$ accounts for cosmological time dilation. We take $\psi(z)$ to be the Madau–Dickinson star formation rate density [35] convolved with a distribution of delay times between 20 Myr and 10 Gyr, refit to the same functional form,

$$\psi(z) \propto \frac{(1+z)^{\alpha}}{1+[(1+z)/(1+z_p)]^{\alpha+\beta}}, \tag{12}$$

with $\alpha = 1.801$, $\beta = 3.492$, and $z_p = 1.833$. The merger rate density is normalized to $\mathcal{R}_0 = 36.8\,\mathrm{Gpc^{-3}\,yr^{-1}}$ at $z = 0$, so the expected number of mergers in an observing time T is

$$\bar{N} = \mathcal{R}_0 T \int_0^{20} \frac{\psi(z)}{\psi(0)} \frac{1}{1+z} \frac{dV_c}{dz}\, dz, \tag{13}$$

and the realized number of mergers is a Poisson draw with mean $\bar{N}$. For $T = 1$ yr this gives $\sim 5\times10^4$ mergers. Extrinsic parameters are drawn isotropically: sky locations are uniform on the sphere, $\cos\theta_{JN}$ is uniform on $[-1,1]$, and the polarization angle and coalescence phase are uniform. Coalescence times are uniform over the year, so the Earth's rotation is sampled across events.

To account for uncertainty in the local merger rate, we repeat the simulation with $\mathcal{R}_0$ set to 27.5 and 49.4 $\mathrm{Gpc^{-3}\,yr^{-1}}$, the 5th and 95th percentiles of the rate inferred in Ref. [33]. All other hyperparameters are held at their maximum-likelihood values, so the resulting band reflects rate uncertainty only. We expect this to be the dominant source of uncertainty in the total number of events. The number of low-SNR events, however, is also sensitive to the redshift evolution of the merger rate, since such sources lie predominantly at high redshift; we do not propagate this uncertainty here. Uncertainties in the mass and spin hyperparameters are expected to have a smaller effect.

B. Network SNR distributions

For each simulated merger we compute the network optimal SNR, ρ, by evaluating Eq. (6) for each detector between $f_{\rm min}$ and $f_{\rm max}$ and adding the single-detector SNRs in quadrature. Signals are generated from the sampled masses, spins, and extrinsic parameters in BILBY [27], using a reference frequency of 20 Hz, a sampling rate of 1024 Hz, and $f_{\rm max} = 448$ Hz for consistency with the parameter estimation analyses of Section III. However, we use the IMRPHENOMXPHM waveform model [36] rather than NRSUR7DQ4 [28]. Due to restrictions in

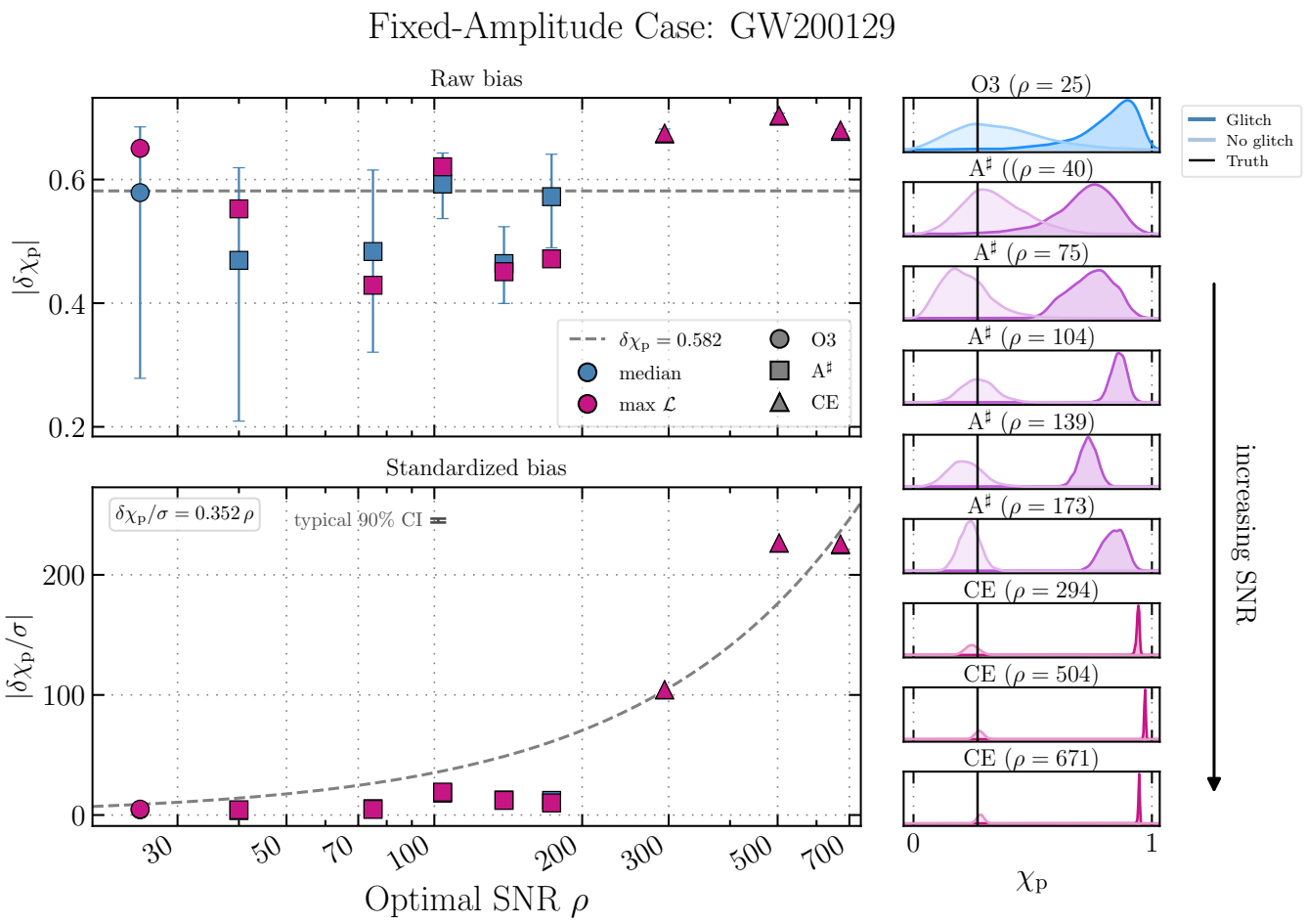


FIG. 7. Bias trends and χ_p posteriors for the fixed-amplitude glitch scaling case and GW200129-like event. The top left panel shows the raw biases for the fixed-amplitude glitch runs for all detector configurations. The expected trend outlined in Section II B is fit to the data and drawn as a gray dashed line. Similarly, the bottom left panel shows the same points but in terms of standardized bias. The corresponding expected trend given by Equation (4) is fit to the data points using OLS with a dashed gray line. In the lower panel, the representative error bar shows the typical (median) 90% posterior credible interval in units of the posterior standard deviation. Since this interval is small relative to the scale of the plot and nearly identical for every run ($\approx \pm 1.6\sigma$, as expected for an approximately Gaussian posterior), it is shown once rather than on each point. The right panels display the posteriors for each of the runs plotted left, with the no-glitch runs in lighter shading and the truth value drawn as a black vertical line.

its range, NRSur7dq4 covers only the final ~ 20 orbits before merger, so it cannot model the full inspiral down to $f_{\rm min} = 5$–$10\,$Hz for most of the simulated population. IMRPhenomXPHM does not share this limitation, and is fast enough to evaluate for the full population.

We consider three two-detector networks, each placed at the locations and orientations of LIGO Hanford and LIGO Livingston: two A♯ detectors with $f_{\rm min} = 10\,$Hz, two 20 km CE detectors with $f_{\rm min} = 5\,$Hz, and a 40 km CE detector at Hanford paired with a 20 km CE detector at Livingston with $f_{\rm min} = 5\,$Hz. We take $\rho \geq 8$ as the detection threshold and assume a 100% duty cycle.

C. Fraction of events requiring mitigation

Figure 10 shows the cumulative number of BBH signals per year above a given ρ for each of the three simulated networks. The events which exceed an SNR of 150 lie to the right of the red dashed line. For the A♯ network, only 4 events out of ∼12,000 exceeded this cutoff. For the CE networks, there were more than a thousand events with sufficiently high SNR; however, this only constituted 2–5% of detected events. A complete summary of the results for the median merger rate is given in Table III.

This result suggests that the majority of future detected signals will not have sufficient SNR to neglect glitch mitigation. However, several caveats make this fraction a conservative estimate. First, only events that

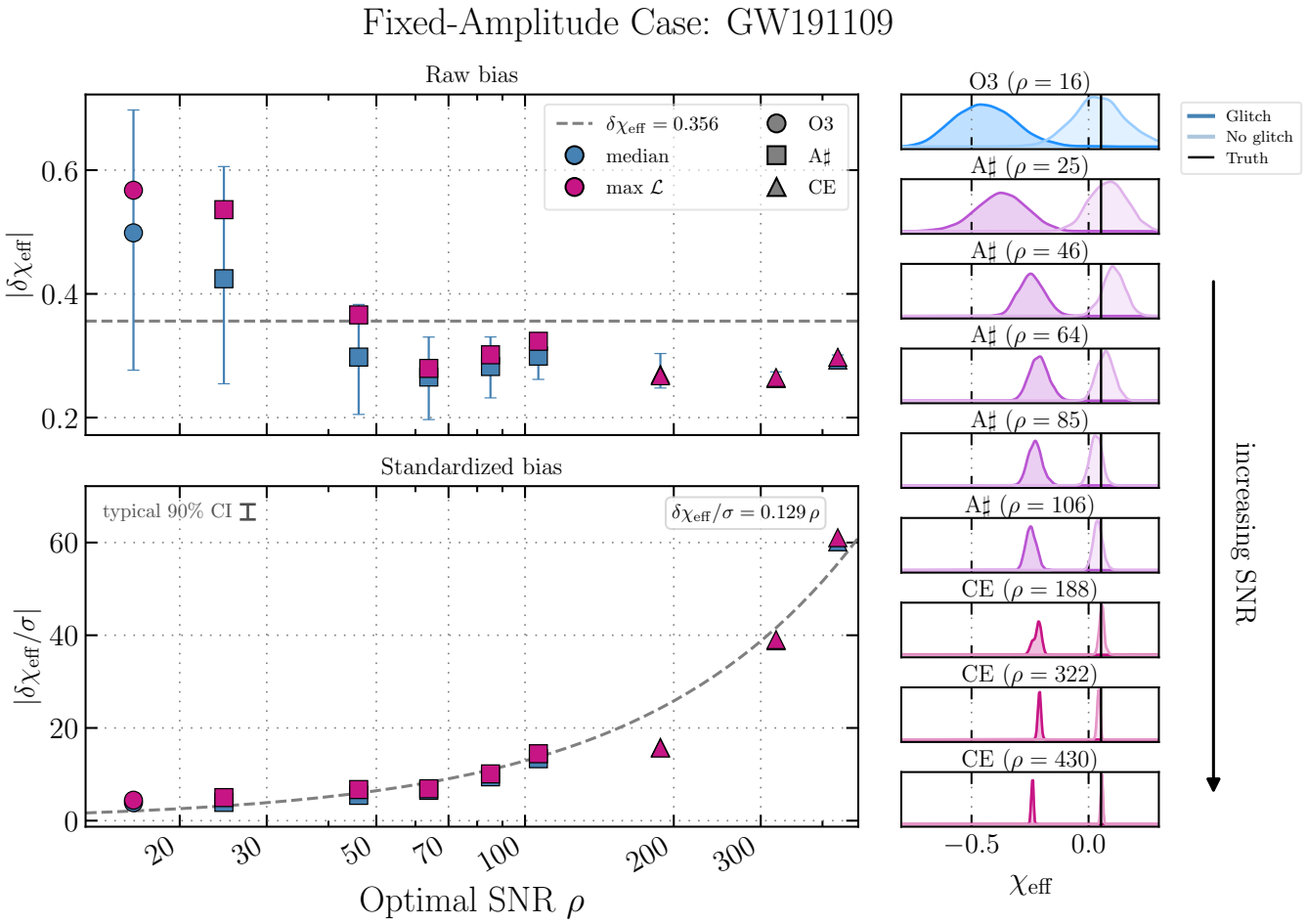


FIG. 8. Same as Figure 7, but for the GW191109-like event.

Network	Detectable	$\rho \geq 150$	Fraction
2×A♯	12,089	4	0.03%
2×20 km CE	50,438	1,228	2.4%
20 km + 40 km CE	52,105	2,786	5.3%

TABLE III. Expected detectable BBH signals per year at the median GWTC-5 merger rate, and the number and fraction above the mitigation-negligible cutoff ($\rho \geq 150$).

actually overlap a glitch require mitigation. Currently, $\sim$ 64% of detected events require no glitch mitigation regardless of the signal SNR. Second, the SNR cutoff of 150 was determined from our analysis which assumed a minimum detector frequency of 20 Hz. As demonstrated in Section V, a lower minimum frequency can greatly reduce the impact of a glitch for glitches similar to those we modeled. Thus, we expect that many events below our cutoff may not require mitigation even if they overlap with a glitch. However, glitches can take many forms, and not all have the same effect. Our results hold for the simplest glitch form, but may not apply to more complex cases. We discuss this further in the following section.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrated how the impact of a single Morlet-Gabor wavelet glitch evolves with increased detector sensitivity under two glitch scaling cases. We found that, when the glitch SNR is held constant, the absolute bias decreases and eventually becomes negligible as the detector becomes more sensitive. In contrast, we found that when the glitch amplitude is held constant, the bias remains roughly the same even with greatly increased detector sensitivity, although the parameter estimates reduce in uncertainty.

For the fixed-SNR glitches ($\rho_{\rm g} \approx 6$–8), the raw bias in the most affected spin parameter fell below ≈ 0.15 once the signal SNR exceeded $\rho \approx 150$, which we adopted as the threshold above which glitch mitigation is no longer needed. Applying this threshold to a simulated BBH population based on GWTC-5, we estimated that only 0.03% of detected events in an A♯ network, and 2.4–5.3%

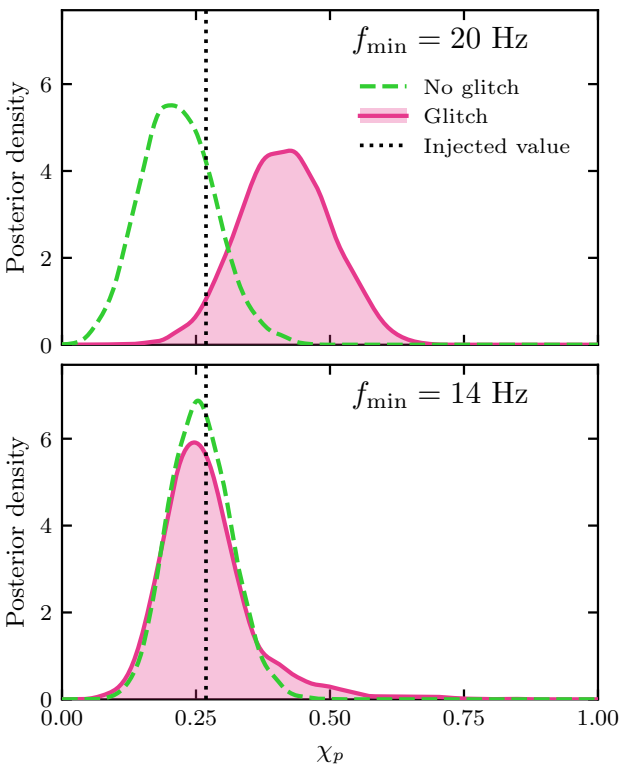


FIG. 9. Marginalized $\chi_{\rm p}$ posteriors for the GW200129 A$^{\sharp}$* ($\rho = 139$) fixed-glitch-SNR run. The top panel shows the glitch run posterior filled in pink and no-glitch run posterior dashed in green for the original set of runs performed with $f_{\rm min} = 20$ Hz. The bottom panel shows the same set of runs conducted with the minimum detector frequency lowered to $f_{\rm min} = 14$ Hz. The injected truth value is shown as a black dashed line.

in the CE networks we considered, exceed it. This estimate is conservative. Our main analysis held $f_{\rm min} = 20$ Hz fixed for every detector configuration, whereas future detectors will be sensitive to lower frequencies. When we lowered $f_{\rm min}$ to 14 Hz for the A$^{\sharp}$ configuration of the GW200129-like event, the glitch-induced bias became negligible, suggesting that the threshold for detectors operating at their projected low-frequency cutoffs may be considerably lower than $\rho \approx 150$. However, these results rely on our analysis in Section IV, which only examines the impact of a single Morlet-Gabor wavelet glitch, and a glitch with more low-frequency power relative to the signal may not show the same improvement.

Our future work will involve conducting similar analysis, except with a more physically motivated glitch model designed for scattered-light glitches [37]. In particular, we plan to model our glitch after the scattered-light glitch which impacted the real event GW191109 and test whether increasing detector sensitivity and fixing the glitch SNR improves parameter estimates for this common class of glitches, whose form is significantly more complex than that of a single wavelet. Additionally, this will test whether lowering the minimum detector frequencies still improves the accuracy of parameter estimation even if the glitch extends further into the lower-frequency range.

VIII. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) Research Experiences for Undergraduates (REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship (LIGO SURF) program, and the California Institute of Technology Student-Faculty Programs.

Appendix A: Waveform systematics

To perform simulated event injection as well as parameter estimation, we require a choice of waveform model, or a prediction of the form of the GW signal given a set of binary parameters. In principle, the waveform can be obtained by numerically solving Einstein's equations for a BBH system. However, numerical-relativity (NR) simulations are computationally expensive, making it impractical to simulate every possible binary across the full parameter space. Instead, waveform approximants provide computationally efficient predictions of the GW signal across the parameter space.

Two broad classes of waveform approximants have emerged. The first is phenomenological modeling, which constructs analytical representations of the waveform and calibrates them against NR simulations. The second is NR surrogate modeling, which interpolates between a finite set of NR waveforms, allowing waveforms to be generated at points in parameter space where a full NR simulation has not been performed. Phenomenological models such as IMRPhenomXPHM demonstrate accuracy, broad applicability, and computational efficiency [36]. NR surrogates such as NRSur7dq4 are typically more accurate within their more limited training range [28], but at greater computational cost.

However, these models are only approximants, and waveform mismodeling can introduce systematic errors in the inferred parameters, known as "waveform systematics" [38]. Any real GW signal will differ to some degree from the approximant used to analyze it, and because each approximant is constructed differently, the nature of this discrepancy depends on the model. Different approximants therefore also disagree with one another. This allows us to study waveform systematics without using a true signal: by injecting a signal with one approximant and recovering it with another, the injection model plays the role of the true signal, and the mismatch between the two models mimics the mismatch between a real signal and the model used to analyze it. Comparing our PE results with matched injection and recovery models against results performed instead with mismatched models thus

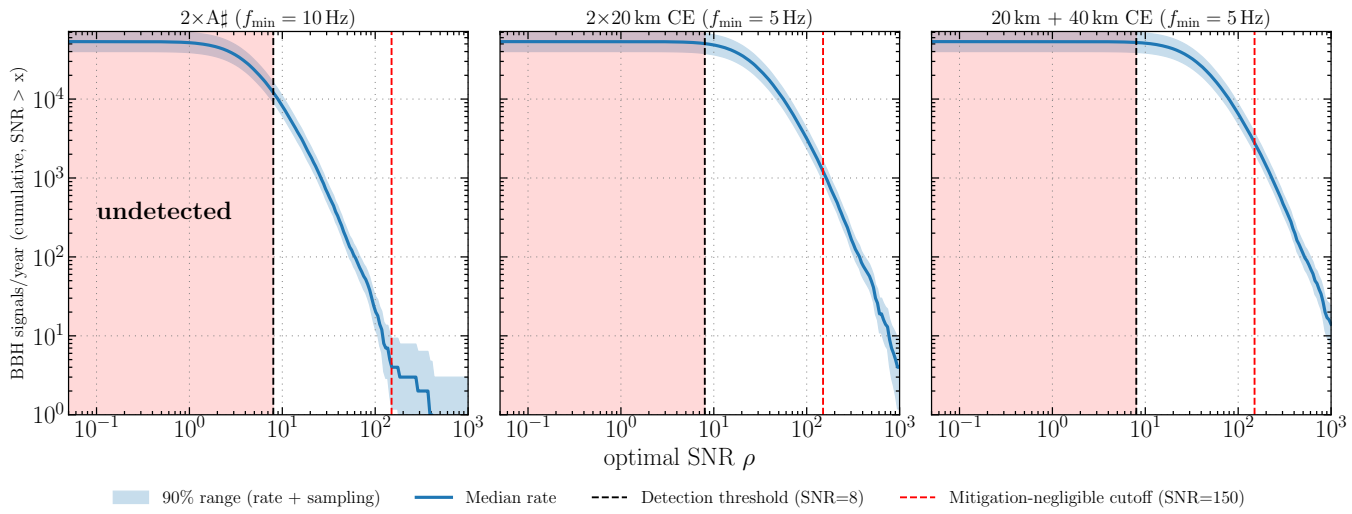


FIG. 10. BBH signals per year with SNR $> x$ for three simulated detector networks: two LIGO A$^\sharp$ detectors with $f_{\rm min} = 10$ Hz (left), two 20 km CE detectors with $f_{\rm min} = 5$ Hz (center), and one 20 km and one 40 km CE, both with $f_{\rm min} = 5$ Hz (right). The solid blue line gives the BBH signals/year for the median merger rate ($\mathcal{R}_0 = 36.8\ \mathrm{Gpc^{-3}\,yr^{-1}}$) found in Ref. [33]. The light blue shaded region shows the 90% confidence interval, combining in quadrature the 5th–95th percentile merger rate range ($\mathcal{R}_0 = 27.5$–$49.4\,\mathrm{Gpc^{-3}\,yr^{-1}}$) with the binomial uncertainty from the finite number of simulated events, which dominates at high SNR. A red dashed line is drawn at $\rho = 150$, representing the chosen SNR cutoff. The red region indicates mergers with an SNR < 8, which we consider undetected.

allows us to analyze an additional bias contribution from waveform systematics.

The results above use NRSur7dq4 for both CBC injection and PE recovery. Since NRSur7dq4 interpolates directly between NR simulations, it is the better proxy for the true signal, so we keep it as the injection model and additionally perform recovery with IMRPhenomXPHM, comparing the resulting posteriors. We perform this test with the GW200129-like signal for the fixed-glitch-SNR case.

Figure 11 shows the posteriors for the O3* and A$^\sharp$* runs when NRSur7dq4 is used for waveform recovery, matching injection, and when IMRPhenomXPHM is instead used, introducing waveform systematic errors. For both the O3 and A$^\sharp$ runs, the no-glitch posteriors largely overlap for both recovery methods, indicating agreement. However, the O3 glitch-run posteriors differ greatly between recovery methods. This indicates that the glitch pushes the waveform to a region of parameter space (in this case, the region of high precession spin) where waveform models are more discrepant. However, as seen in the bottom panel, the glitch-run posteriors overlap for the A$^\sharp$ runs, demonstrating that this discrepancy declines with reduced glitch impact. The same runs were performed for the CE* configuration; however, the small waveform mismodeling errors cause the XPHM posteriors to miss the injection truth values entirely due to the extremely narrow posterior width. This reflects the increased relative importance of waveform systematics as detectors become more sensitive [39].

Overall, these results show that unmitigated glitches can worsen errors due to waveform systematics. This effect may be reduced with increased detector sensitivity for a fixed-SNR glitch, however more analysis is needed to investigate this further.

Appendix B: Bias in other parameters

For our results, we focused on $\chi_{\rm p}$ for GW200129 and $\chi_{\rm eff}$ for GW191109, as these were the intended biased parameters of the original O3 glitch configuration [16]. However, there is no guarantee that the parameters which experience the most bias remain the same as we scale the Gaussian noise and the glitch. To check this, we tracked how the standardized bias changed across a few key parameters for the three principal runs. Figure 12 shows the standardized bias for these three runs for the GW200129-like and GW191109-like event for both glitch scaling cases. The parameter which we focused on for the main analysis in this paper is highlighted in yellow. Unlike in the main text, here we define the raw bias as the difference between the posterior medians of the glitch and no-glitch runs, rather than the difference between the glitch posterior maximum-likelihood value and the injection truth value. We divide this bias by the standard deviation of the glitch posterior to obtain our standardized bias. The main-text definition measures the total error relative to the truth, including the offset from Gaussian noise and any inaccuracy in the no-glitch recovery itself; here we instead seek the component of the error attributable to the glitch alone. Because both runs share the same noise realization, this choice cancels the offset due to Gaussian noise and isolates the shift caused

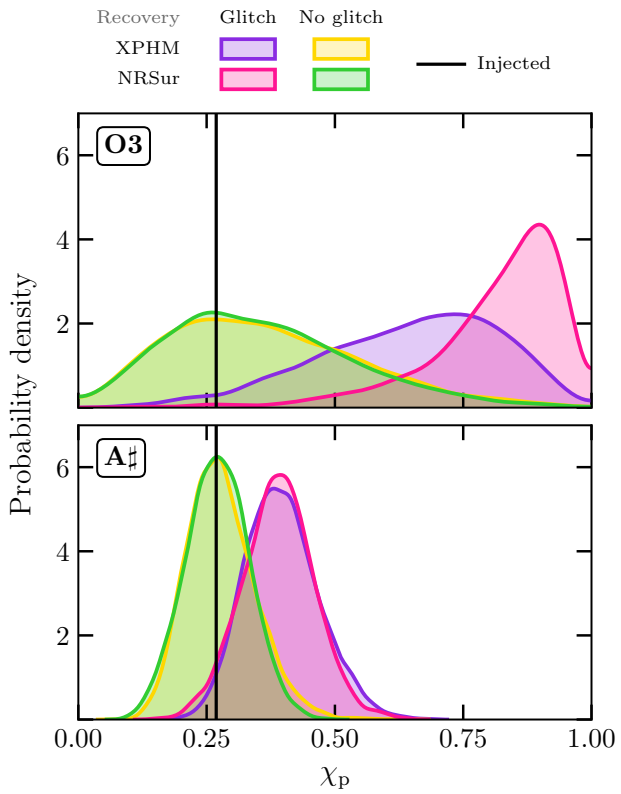


FIG. 11. Marginalized $\chi_{\rm p}$ posteriors for the GW200129 fixed-SNR glitch ($\rho_{\rm g} = 8$) scaling case for the principal runs O3* (top panel; $\rho_{\rm BBH} = 25$) and A♯* (bottom panel; $\rho_{\rm BBH} = 139$). For all runs, the signal was injected with the NRSur7dq4 waveform model. The posteriors in purple (glitch run) and yellow (no-glitch run) were recovered instead with IMRPhenomXPHM. The posteriors in pink (glitch run) and green (no-glitch run) were also recovered with NRSur7dq4. The injected truth value is drawn in black.

by the glitch, i.e., the difference between what would be measured with and without it.

Although the glitch also biases some parameters other than the focused ones, none of these biases grows appreciably with detector sensitivity in the fixed-SNR case, so the effective spin parameters remain the most relevant measure of the glitch's impact across detector generations. The largest such bias occurs for the GW191109-like event, where D_L is strongly underestimated in the O3 run, exceeding even the bias in $\chi_{\rm eff}$. However, it falls below 1σ for the A♯ and CE runs, while the $\chi_{\rm eff}$ bias persists and is largest for A♯. For the GW200129-like event, $\chi_{\rm p}$ is the most biased parameter in both the O3 and A♯ fixed-glitch-SNR runs, and the remaining parameters show no significant trends.

In the fixed-amplitude case, the standardized biases grow substantially from O3 to CE, mainly reflecting the sharp decrease in posterior width. Since the raw bias is expected to remain roughly constant in this scaling case, the pattern of biases across parameters in the CE run largely mirrors that of the O3 run, amplified by the narrower posteriors. For example, D_L shows a prominent standardized bias for the GW191109-like event in both the O3 run and the fixed-amplitude CE run. The most notable exception is M_c for the GW191109-like event, whose fixed-amplitude CE bias is large relative to the other parameters compared with its O3 bias.

Overall, in neither scaling case does increasing detector sensitivity shift the glitch's impact onto parameters that were unaffected in O3. Parameters biased in the O3 run either become similarly or less biased, as in the fixed-SNR case, or remain biased with a similar pattern across parameters, as in the fixed-amplitude case. This supports our focus on $\chi_{\rm eff}$ and $\chi_{\rm p}$ in the main text, although we note that this comparison covers only the three principal runs and the five parameters shown in Figure 12.

Appendix C: Detailed injection parameters

Table IV lists the parameters of the two simulated BBH signals, which are injected and recovered with NRSur7dq4 as described in Section III. They correspond to the ModerateMass-ModerateSpins (GW200129-like) and HighMass-ModerateSpins (GW191109-like) configurations of Udall *et al.* [16]. The same BBH signal parameters are used in every detector configuration; only the PSD and the glitch amplitude vary between runs (Tables I and II).

[1] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), GWTC-5.0: An introduction to version 5.0 of the Gravitational-Wave Transient Catalog (2026), arXiv:2605.27223 [gr-qc].
[2] N. Christensen and R. Meyer, Parameter estimation with gravitational waves, Rev. Mod. Phys. **94**, 025001 (2022).
[3] I. Mandel and A. Farmer, Merging stellar-mass binary black holes, Phys. Rep. **955**, 1 (2022).
[4] A. Ray, S. Mukherjee, M. Zevin, and V. Kalogera, On the astrophysical origin of binary black hole subpopulations: A tale of three channels? (2026), arXiv:2603.17987 [astro-ph.HE].
[5] W. M. Farr, S. Stevenson, M. Coleman Miller, I. Mandel, B. Farr, and A. Vecchio, Distinguishing spin-aligned and isotropic black hole populations with gravitational waves, Nature **548**, 426 (2017).
[6] P. Ajith *et al.*, Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins, Phys. Rev. Lett. **106**, 241101 (2011).
[7] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), GWTC-3: Compact binary coalescences observed by LIGO and

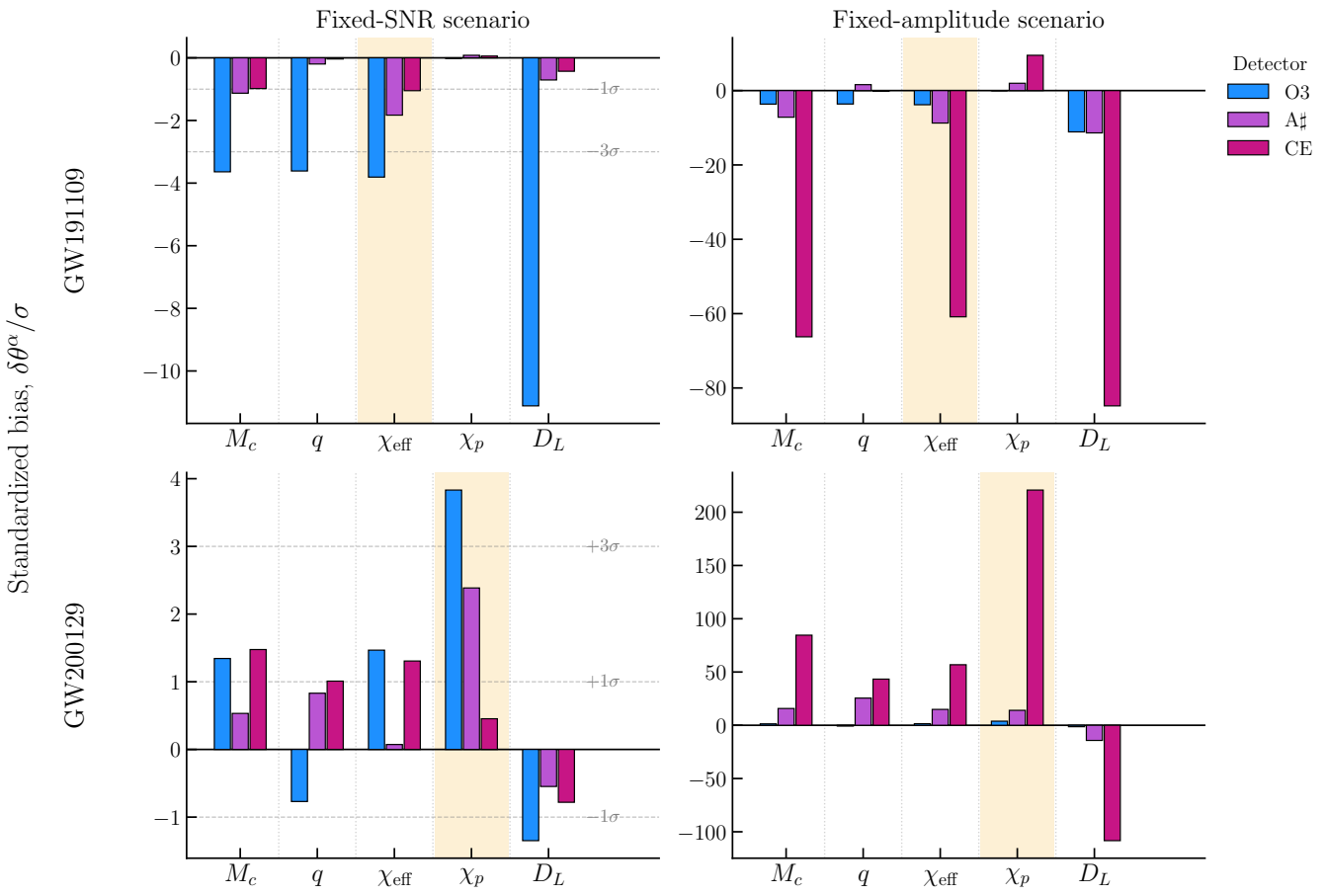


FIG. 12. Standardized bias by parameter for each detector, shown for both simulated events (rows: GW191109, GW200129) and both glitch-scaling cases (columns: fixed-SNR, fixed-amplitude). The parameters shown are the chirp mass M_c, the mass ratio q, the effective spin parameters $\chi_{\rm eff}$ and $\chi_{\rm p}$, and the luminosity distance D_L. The standardized bias is measured as the difference between the medians of the glitch and no-glitch runs, divided by the standard deviation of the glitch-run posterior. The standardized bias for three select runs are shown in each panel: O3 in blue, A$^\sharp$ in purple, and CE in pink (O3*, A$^\sharp$*, and CE* in Table I). The parameter which we focus on for each particular simulated event in this paper is highlighted in yellow.

Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[8] T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O'Shaughnessy, and R. Smith, Analysis of GWTC-3 with fully precessing numerical relativity surrogate models, Phys. Rev. D **112**, 044001 (2025).

[9] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, Publ. Astron. Soc. Aust. **36**, e010 (2019), [Erratum: Publ. Astron. Soc. Aust. **37**, e036 (2020)].

[10] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run, Classical Quantum Gravity **39**, 245013 (2022).

[11] S. Ghonge, J. Brandt, J. M. Sullivan, M. Millhouse, K. Chatziioannou, J. A. Clark, T. Littenberg, N. Cornish, S. Hourihane, and L. Cadonati, Assessing and mitigating the impact of glitches on gravitational-wave parameter estimation: A model agnostic approach, Phys. Rev. D **110**, 122002 (2024).

[12] R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, Antialigned spin of GW191109: Glitch mitigation and its implications, Phys. Rev. D **111**, 024046 (2025).

[13] E. Payne, S. Hourihane, J. Golomb, R. Udall, R. Udall, D. Davis, and K. Chatziioannou, Curious case of GW200129: Interplay between spin-precession inference and data-quality issues, Phys. Rev. D **106**, 104017 (2022).

[14] R. Macas, A. Lundgren, and G. Ashton, Revisiting the evidence for precession in GW200129 with machine learning noise mitigation, Phys. Rev. D **109**, 062006 (2024).

[15] M. Hannam *et al.*, General-relativistic precession in a black-hole binary, Nature **610**, 652 (2022).

[16] R. Udall, S. Bini, K. Chatziioannou, D. Davis, S. Hourihane, Y. Lecoeuche, J. McIver, and S. Miller, Inferring the spins of merging black holes in the presence of data-quality issues, Phys. Rev. D **113**, 042005 (2026).

[17] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Open data from LIGO, Virgo, and KAGRA through the second part

Parameter	GW200129	GW191109
Masses		
M_c ($M_\odot$)	33.06	63.45
q	0.779	0.792
m_1 ($M_\odot$)	43.10	81.99
m_2 ($M_\odot$)	33.57	64.96
Spins		
a_1	0.405	0.363
a_2	0.144	0.461
θ_1 (rad)	0.725	2.136
θ_2 (rad)	1.162	0.652
ϕ_{12} (rad)	1.424	3.064
ϕ_{JL} (rad)	4.884	0.181
$\chi_{\rm eff}$	0.195	0.053
$\chi_{\rm p}$	0.269	0.307
Extrinsic		
D_L (Mpc)	1292	1302
θ_{JN} (rad)	2.741	2.213
ψ (rad)	1.583	2.924
ϕ_c (rad)	4.665	2.738
α (rad)	5.916	4.138
δ (rad)	−0.351	−0.611
$t_{\rm geo}$ (GPS s)	1264316116.0067	1257296855.0000

TABLE IV. Injected parameters of the simulated BBH signals: chirp mass M_c, mass ratio q, component masses $m_{1,2}$, spin magnitudes $a_{1,2}$, spin tilts $\theta_{1,2}$, spin azimuthal angles ϕ_{12} and ϕ_{JL}, luminosity distance D_L, inclination θ_{JN}, polarization angle ψ, coalescence phase ϕ_c, right ascension α, declination δ, and geocenter coalescence time $t_{\rm geo}$. Masses are in the detector frame, spins are defined at a reference frequency of 20 Hz, and all angles are in radians. The component masses, $\chi_{\rm eff}$, and $\chi_{\rm p}$ are derived from the other parameters and are listed for reference.

of the fourth observing run (2026), arXiv:2605.27090 [gr-qc].
[18] A. Effler, P. Fritschel, E. D. Hall, M. Landry, A. Lazzarini, B. O'Reilly, D. H. Reitze, D. Shoemaker, and C. Torrie, A roadmap for the LIGO observatories in the era of Cosmic Explorer, LIGO Laboratory Technical Report No. LIGO-M2300107 (2023).
[19] P. Fritschel, K. Kuns, J. Driggers, A. Effler, B. Lantz, D. Ottaway, S. Ballmer, K. Dooley, R. X. Adhikari, M. Evans, B. Farr, G. Gonzalez, P. Schmidt, and S. Raja, Report of the LSC Post-O5 Study Group, LIGO Scientific Collaboration Technical Report No. LIGO-T2200287 (2022).
[20] M. Evans *et al.*, A Horizon Study for Cosmic Explorer: Science, observatories, and community (2021), arXiv:2109.09882 [astro-ph.IM].
[21] D. Reitze *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, Bull. Am. Astron. Soc. **51**, 035 (2019).
[22] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), GWTC-5.0: Observations from the second part of the fourth LIGO-Virgo-KAGRA observing run and updates to the Gravitational-Wave Transient Catalog (2026), arXiv:2605.27225 [gr-qc].
[23] C. Cutler and M. Vallisneri, LISA detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms, Phys. Rev. D **76**, 104018 (2007).
[24] C. Cutler and É. E. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform?, Phys. Rev. D **49**, 2658 (1994).
[25] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, Phys. Rev. D **77**, 042001 (2008).
[26] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32**, 135012 (2015).
[27] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[28] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, Phys. Rev. Res. **1**, 033015 (2019).
[29] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: A tool to characterize transient noise in gravitational-wave detectors, SoftwareX **12**, 100620 (2020).
[30] J. Glanzer *et al.* (LIGO), LIGO detector characterization in the second and third parts of the fourth observing run (2026), arXiv:2608.12193 [astro-ph.IM].
[31] J. S. Speagle, dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).
[32] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Directly comparing GW150914 with numerical solutions of Einstein's equations for binary black hole coalescence, Phys. Rev. D **94**, 064035 (2016).
[33] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), GWTC-5.0: Population properties of merging compact binaries (2026), arXiv:2605.27226 [astro-ph.HE].
[34] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020), [Erratum: Astron. Astrophys. **652**, C4 (2021)].
[35] P. Madau and M. Dickinson, Cosmic star formation history, Annu. Rev. Astron. Astrophys. **52**, 415 (2014).
[36] G. Pratten *et al.*, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103**, 104056 (2021).
[37] R. Udall and D. Davis, Bayesian modeling of scattered light in the LIGO interferometers, Appl. Phys. Lett. **122**, 094103 (2023).
[38] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Effects of waveform model systematics on the interpretation of GW150914, Classical Quan-

tum Gravity **34**, 104002 (2017).

[39] M. Pürrer and C.-J. Haster, Gravitational waveform accuracy requirements for future ground-based detectors, Phys. Rev. Res. **2**, 023151 (2020).