Curious case of GW200129: Interplay between spin-precession inference and data-quality issues

Ethan Payne⬤, Sophie Hourihane⬤, Jacob Golomb⬤, Rhiannon Udall⬤, Derek Davis⬤, and Katerina Chatziioannou⬤

Department of Physics, California Institute of Technology, Pasadena, California 91125, USA and
LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA

Measurement of spin-precession in black hole binary mergers observed with gravitational waves is an exciting milestone as it relates to both general relativistic dynamics and astrophysical binary formation scenarios. In this study, we revisit the evidence for spin-precession in GW200129 and localize its origin to data in LIGO Livingston in the 20–50 Hz frequency range where the signal amplitude is lower than expected from a nonprecessing binary given all the other data. These data are subject to known data quality issues as a glitch was subtracted from the detector's strain data. The lack of evidence for spin-precession in LIGO Hanford leads to a noticeable inconsistency between the inferred binary mass ratio and precessing spin in the two LIGO detectors, something not expected from solely different Gaussian noise realizations. We revisit the LIGO Livingston glitch mitigation and show that the difference between a spin-precessing and a nonprecessing interpretation for GW200129 is smaller than the statistical and systematic uncertainty of the glitch subtraction, finding that the support for spin-precession depends sensitively on the glitch modeling. We also investigate the signal-to-noise ratio ∼7 trigger in the less sensitive Virgo detector. Though not influencing the spin-precession studies, the Virgo trigger is grossly inconsistent with the ones in LIGO Hanford and LIGO Livingston as it points to a much heavier system. We interpret the Virgo data in the context of further data quality issues. While our results do not disprove the presence of spin-precession in GW200129, we argue that any such inference is contingent upon the statistical and systematic uncertainty of the glitch mitigation. Our study highlights the role of data quality investigations when inferring subtle effects such as spin-precession for short signals such as the ones produced by high-mass systems.

I. INTRODUCTION

GW200129_065458 (henceforth GW200129) is a gravitational wave (GW) candidate reported in GWTC-3 [1]. The signal was observed by all three LIGO-Virgo detectors [2,3] operational during the third observing run (O3) and it is consistent with the coalescence of two black holes (BHs) with source-frame masses $34.5^{+9.9}_{-3.2}$ $M_\odot$ and $28.9^{+3.4}_{-9.3}$ $M_\odot$ at the 90% credible level. Though the masses are typical within the population of observed events [4], the event's signal-to-noise-ratio (SNR) of $26.8^{+0.2}_{-0.2}$ makes it the loudest binary BH (BBH) observed to date. Additionally, it is one of the loudest triggers in the Virgo detector with a detected SNR of 6–7 depending on the detection pipeline [1]. The signal temporally overlapped with a glitch in the LIGO Livingston detector, which was subtracted using information from auxiliary channels [5]. The detection and glitch mitigation procedures for this event are recapped in Appendix A 1.

The interpretation of some events in GWTC-3 was impacted by waveform systematics, with GW200129 being one of the most extreme examples. As part of the catalog, results were obtained with the IMRPhenomXPHM [6] and

SEOBNRv4PHM [7] waveform models using the parameter inference algorithms BILBY [8,9] and RIFT [10] respectively. Both waveforms correspond to quasicircular binary inspirals and include high-order radiation modes and the effect of relativistic spin-precession arising from interactions between the component spins and the orbital angular momentum. All analyses used the glitch-subtracted LIGO Livingston data. The IMRPhenomXPHM result was characterized by large spins and a bimodal structure with peaks at ∼0.45 and ∼0.9 for the binary mass ratio. The SEOBNRv4PHM results, on the other hand, pointed to more moderate spins and near equal binary masses. Both waveforms, however, reported a mass-weighted spin aligned with the Newtonian orbital angular momentum of $\chi_{\rm eff} \sim 0.1$, and thus the inferred large spins with IMRPhenomXPHM corresponded to spin components in the binary orbital plane and spin-precession. Such differences between the waveform models are not unexpected for high SNR signals [11]. Waveform systematics are also likely more prominent when it comes to spin-precession, as modeling prescriptions vary and are not calibrated to numerical relativity simulations featuring spin-precession [6,7,12]. Data quality issues could further lead to evidence

for spin-precession [13]. Due to differences in the inference algorithms and waveform systematics, GWTC-3 argued that definitive conclusions could not be drawn regarding the possibility of spin-precession in this event [1].

Stronger conclusions in favor of spin-precession [14] and a merger remnant that experienced a large recoil velocity [15] were put forward by means of a third waveform model. NRSur7dq4 [16] is a surrogate to numerical relativity simulations of merging BHs that is also restricted to quasicircular orbits and models the effect of high-order modes and spin-precession. This model exhibits the smallest mismatch against numerical relativity waveforms, sometimes comparable to the numerical error in the simulations. It is thus expected to generally yield the smallest errors due to waveform systematics [16]. This fact was exploited in Hannam *et al.* [14] to break the waveform systematics tie and argue that the source of GW200129 exhibited relativistic spin-precession with a primary component spin magnitude of $\chi_1 = 0.9^{+0.1}_{-0.5}$ at the 90% credible level.

During a binary inspiral, spin-precession is described through post-Newtonian theory [17,18]. Spin components that are not aligned with the orbital angular momentum give rise to spin-orbit and spin-spin interactions that cause the orbit to change direction in space as the binary inspirals, e.g., [19–28]. The emitted GW signal is modulated in amplitude and phase, and morphologically resembles the beating between two spin-aligned waveforms [29] or a spin-aligned waveform that has been "twisted-up" [21,22]. As the binary reaches merger, numerical simulations suggest that the direction of peak emission continues precessing [30]. Parameter estimation analyses using NRSur7dq4 find that spins and spin-precession can be measured from merger-dominated signals for certain spin configurations [31], however the lack of analytic understanding of the phenomenon means that it is not clear how such a measurement is achieved.

The main motivation for this study is to revisit GW200129 and attempt to understand how spins and spin-precession can be measured from a heavy BBH with a merger-dominated observed signal. In Sec. II we use NRSur7dq4 to conclude that the evidence for spin-precession originates exclusively from the LIGO Livingston data in the 20–50 Hz frequency range, where the inferred signal amplitude is lower than what a spin-aligned binary would imply given the rest of the data. This range coincides with the known data quality issues described in Appendix A 1 and first identified in GWTC-3 [1]. LIGO Hanford is consistent with a spin-aligned signal, causing an inconsistency between the inferred mass ratio q and precession parameter χ_p inferred from each LIGO detector separately. By means of simulated signals, we argue that such $q - \chi_p$ inconsistency is unlikely to be caused solely by the different Gaussian noise realizations in each detector at the time of the signal, rather pointing to remaining data quality issues beyond the original glitch-subtraction [1].

We also re-analyze the LIGO Livingston data above 50 Hz, (while keeping the original frequency range of the LIGO Hanford data) and confirm that all evidence for spin-precession disappears.

In the process, we find that the Virgo trigger, though consistent with a spin-aligned BBH, is *inconsistent* with the signal seen in the LIGO Hanford and LIGO Livingston detectors. Specifically, the Virgo data are pointing to a much heavier BBH that merges ∼20 ms earlier than the one observed by the LIGO detectors. We discuss Virgo data quality considerations in Sec. III within the context of a potential glitch that affects the inferred binary parameters if unmitigated. As a consequence, we do not include Virgo data in the sections examining spin-precession unless otherwise stated. The Virgo-LIGO inconsistency can be resolved if we use BayesWave [32–34] to simultaneously model a CBC signal and glitches with CBC waveform models and sine-Gaussian wavelets respectively [35,36]. The Virgo data are now consistent with the presence of both a signal that is consistent with the one in the LIGO detectors and an overlapping glitch with SNR ∼ 4.6.

In Sec. IV we revisit the LIGO Livingston data quality issues and compare the original glitch-subtraction based on GWSUBTRACT [5,37] that uses information from auxiliary channels and the glitch estimate from BayesWave that uses only strain data. Though the CBC model used in BayesWave does not include the effect of spin-precession, we show that differences between the reconstructed waveforms from a nonprecessing and spin-precessing analysis for GW200129 are *smaller* than the statistical uncertainty in the glitch inference. Such differences can therefore not be reliably resolved in the presence of the glitch and its subtraction procedure. The two glitch estimation methods give similar results within their statistical errors, however GWSUBTRACT yields typically a lower glitch-amplitude. We conclude that any evidence for spin-precession from GW200129 is contingent upon the systematic and statistical uncertainties of the LIGO Livingston glitch subtraction. Given the low SNR of the LIGO Livingston glitch and the glitch modeling uncertainties, we can at present not conclude whether the source of GW200129 exhibited spin-precession or not.

In Sec. V we summarize our arguments that remaining data quality issues in LIGO Livingston cast doubt on the evidence for spin-precession. Besides data quality studies (i.e., spectrograms, glitch modeling, auxiliary channels), our investigations are based on comparisons between different detectors as well as different frequency bands of the same detector. We propose that similar investigations in further events of interest with exceptional inferred properties could help alleviate potential contamination due to data quality issues.

II. THE ORIGIN OF THE EVIDENCE FOR SPIN-PRECESSION

Our main goal is to pinpoint the parts of the GW200129 data that are inconsistent with a nonprecessing binary and

understand the relevant signal morphology. Due to different orientations, sensitivities, and noise realizations, different detectors in the network do not observe an identical signal. The detector orientation, especially, affects the signal polarization content and thus the degree to which spin-precession might be measurable in each detector. Motivated by this, we begin by examining data using different detector combinations.

We perform parameter estimation using the NRSur7dq4 waveform and examine data from each detector separately (left panel) as well as the relation between the LIGO and the Virgo data (right panel) and show posteriors for select intrinsic parameters in Fig. 1. Analysis settings and details are provided in Appendix A 2 and in all cases we use the same LIGO Livingston data as GWTC-3 [1] where the glitch has been subtracted. Though we do not expect the posterior distributions for the various signal parameters inferred with different detector combinations to be identical, they should have broadly overlapping regions of support. If the triggers recorded by the different detectors are indeed consistent, any shift between the posteriors should be at the level of Gaussian noise fluctuations.

The left panel shows that the evidence for spin-precession arises primarily from the LIGO Livingston data, whereas

the precession parameter χ_p posterior is much closer to its prior when only LIGO Hanford or Virgo data are considered. A similar conclusion was reached in Hannam *et al.* [14]. There is reasonable overlap between the two-dimensional distributions that involve the chirp mass $\mathcal{M}$, the mass ratio q, and the effective spin χ_{eff} inferred by the two LIGO detectors, as expected from detectors that observe the same signal under different Gaussian noise realizations. The discrepancy between the spin-precession inference in the two LIGO detectors, however, is evident in the $q - \chi_p$ panel. The two detectors lead to non-overlapping distributions that point to either unequal masses and spin-precession (LIGO Livingston), or equal masses and no information for spin-precession (LIGO Hanford).

Besides an uninformative posterior on χ_p, the left panel points to a bigger issue with the Virgo data: inconsistent inferred masses. The right panel examines the role of Virgo in more detail in comparison to the LIGO data. Due to the lower SNR in Virgo, the intrinsic parameter posteriors are essentially identical between the HL and the HLV analyses. The lower total SNR means that the Virgo-only posteriors will be wider, but they are still expected to overlap with the ones inferred from the two LIGO detectors. However, this
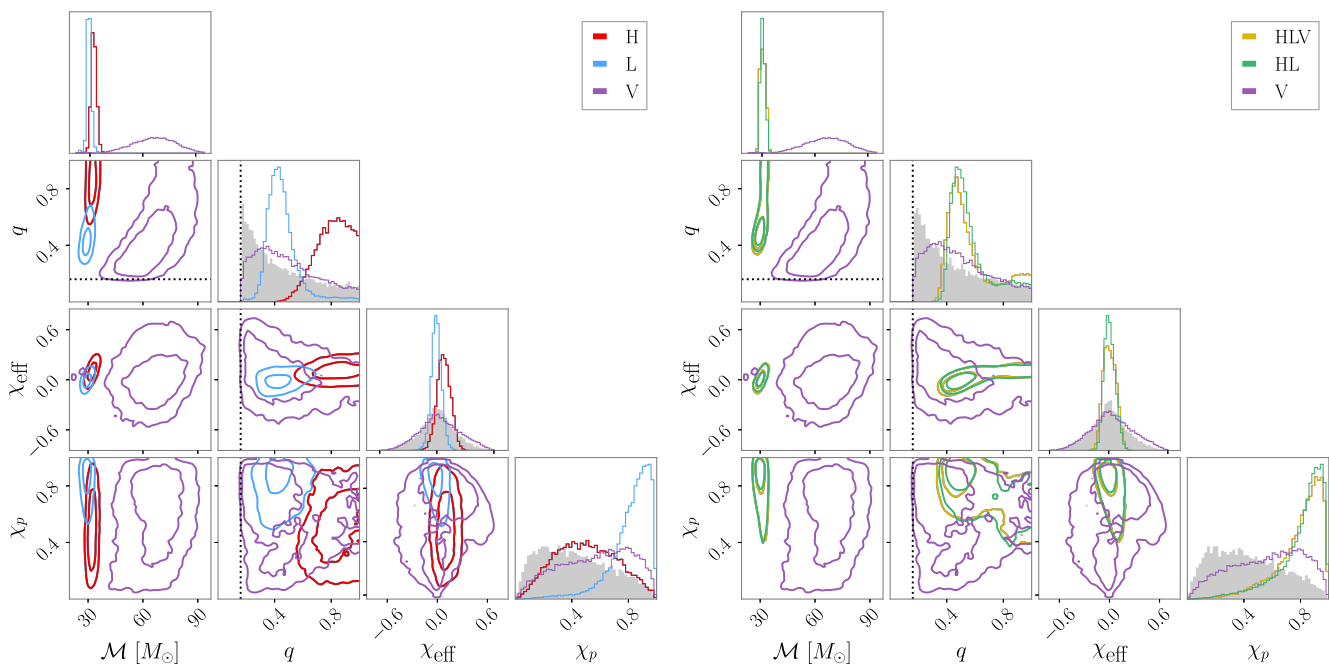

FIG. 1.   One- and two-dimensional marginalized posteriors for select intrinsic binary parameters: detector frame chirp-mass $\mathcal{M}$, mass ratio q, effective spin χ_{eff}, and processing spin χ_p. See Table I for analysis settings and Appendix A 2 for detailed parameter definitions. Two-dimensional panels show 50% and 90% contours. The black dashed line marks the minimum bound of $q = 1/6$ in NRSur7dq4's region of validity. Shaded regions shows the prior for $q, \chi_{\text{eff}}, \chi_p$. The $\mathcal{M}$ prior increases monotonically to the maximum allowed value (see Appendix A 2 for details on choices of priors). Left panel: comparison between analyses that use solely LIGO Hanford (red; H), LIGO Livingston (blue; L), and Virgo (purple; V) data. Right panel: comparison between analyses of all three detectors (yellow; HLV), only LIGO data (green; HL) and only Virgo data (purple; V). The evidence for spin-precession originates solely from the LIGO Livingston data as the other detectors give uninformative χ_p posteriors. Additionally, the binary masses inferred based on Virgo only are inconsistent with those from the LIGO data.
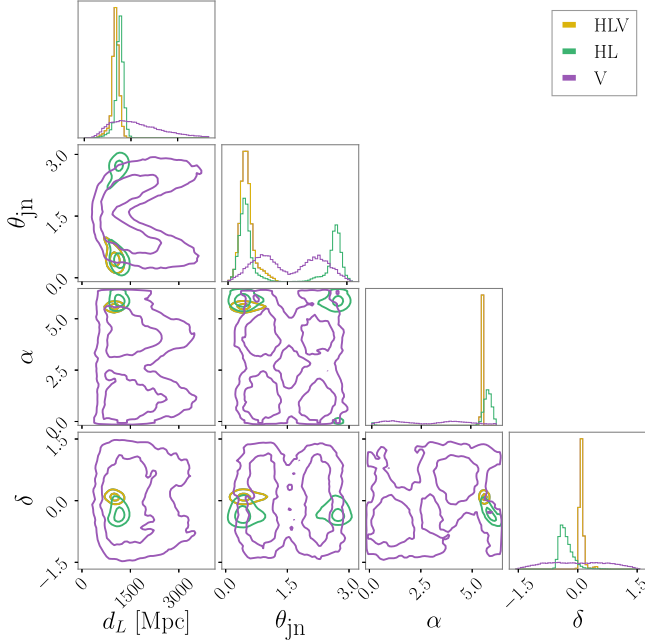
FIG. 2. Similar to the right panel of Fig. 1 but for select extrinsic parameters: luminosity distance d_L, angle between total angular momentum and line of sight $\theta_{\rm jn}$, right ascension α, and declination δ. For reference, the median optimal SNR for each run is HLV: 27.6, HL: 26.9, V: 6.7.

is not the case for the mass parameters as is most evident from the two-dimensional panels involving the chirp mass. While the LIGO data are consistent with a typical binary with (detector-frame) chirp mass $30.3^{+2.5}_{-1.6}$ $M_\odot$ at the 90% credible level, the Virgo data point to a much heavier binary with $66.7^{+19.7}_{-22.6}$ $M_\odot$ at the same credible level.

The role of Virgo data on the inferred binary extrinsic parameters is explored in Fig. 2. In general, Virgo data have a larger influence on the extrinsic than the intrinsic parameters as the measured time and amplitude helps break existing degeneracies. The extrinsic parameter posteriors show a large degree of overlap. The Virgo distance posterior does not rail against the upper prior cut off, suggesting that this detector does observe some excess power. The HL sky localization also overlaps with the Virgo-only one, though the latter is merely the antenna pattern of the detector that excludes the four Virgo "blind spots." We use the HL results to calculate the projected waveform in Virgo and calculate the 90% lower limit on the signal SNR to be 4.2. This suggests that given the LIGO data, Virgo should be observing a signal with at least that SNR at the 90% level.

In order to track down the cause of the discrepancy in the inferred mass parameters, we examine the Virgo strain data directly. Figure 3 shows the whitened time-domain reconstruction (left panel) and the spectrum (right panel) of the signal in Virgo from a Virgo-only and a full 3-detector
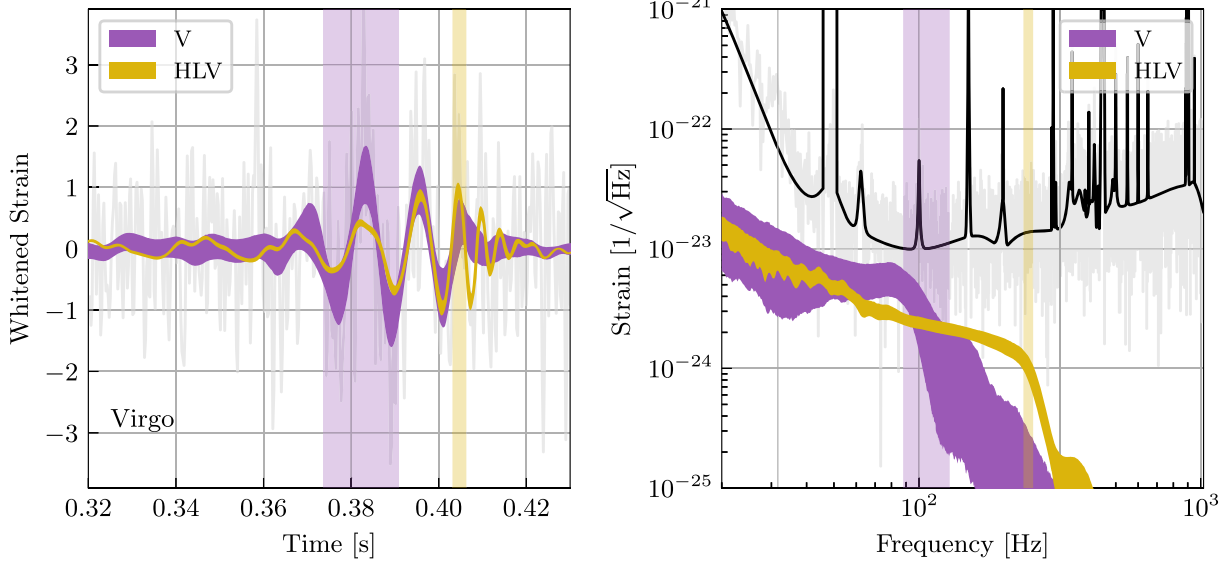

FIG. 3. 90% credible intervals for the whitened time-domain reconstruction (left) and spectrum (right) of the signal in Virgo from a Virgo-only (purple; V) and a full 3-detector (yellow; HLV) analysis, see Table I for analysis settings. The data are shown in gray and the noise PSD in black. The time on the left plot is relative to GPS 1264316116. The high value of the PSD at ∼50 Hz was imposed due to miscalibration of the relevant data [1]. Vertical shaded regions at each panel correspond to the 90% credible intervals of the merger time (left; defined as the time of peak strain amplitude) and merger frequency (right; approximated via the dominant ringdown mode frequency as computed with QNM [38], merger remnant properties were computed with surfinBH [39]). The Virgo data point to a heavier binary that merges ∼20 ms earlier than the full 3-detector results that are dominated by the LIGO detectors.
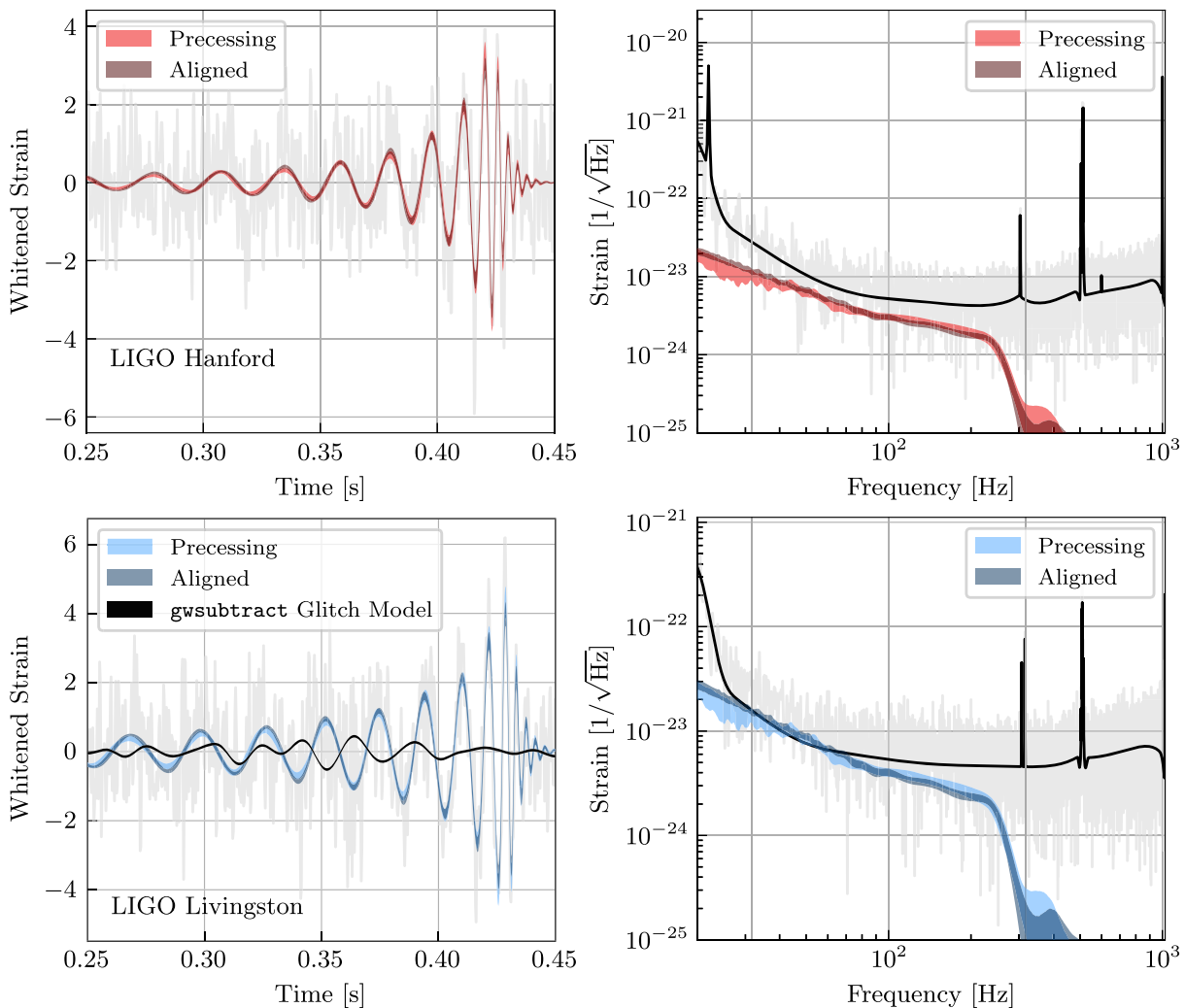
FIG. 4.   Whitened time-domain reconstruction (left) and spectrum (right) of GW200129 in LIGO Hanford (top) and LIGO Livingston (bottom). Shaded regions show the 90% credible intervals for the signal using a spin-precessing (light blue and red) and a spin-aligned (dark blue and red) analysis based on NRSur7dq4, see Table I for run settings. In gray we show the analyzed data where the GWSUBTRACT estimate for the glitch (black line) has already been subtracted. The black line in the right panels is the noise PSD. The glitch overlaps with the part of the inferred signal where the spin-aligned amplitude is on average larger than the spin-precessing one.

analysis. Compared to Figs. 1 and 2, here we only consider a 3-detector analysis as the reconstructed signal in Virgo inferred from solely LIGO data would not be phase-coherent with the data, and thus would be uninformative. Given the higher signal SNR in the two LIGO detectors, the signal reconstruction morphology in Virgo is driven by them, as evident from the intrinsic parameter posteriors from the right panel of Fig. 1.

The two reconstructions in Fig. 3 are morphologically distinct. The 3-detector inferred signal is dominated by the LIGO data and resembles a typical "chirp" with increasing amplitude and frequency. This signal is, however, inconsistent with the Virgo data as it underpredicts the strain at $t \sim 0.382$ s in the left panel. The Virgo-only inferred signal matches the data better by instead placing the merger at earlier times to capture the increased strain at $t \sim 0.382$ s as

shown by the shaded vertical region denoting the merger time. Rather than a "chirp," the signal is dominated by the subsequent ringdown phase with an amplitude that decreases slowly over ~2 cycles. As also concluded from the inferred masses in Fig. 1, the Virgo data point to a heavier binary with lower ringdown frequency (vertical regions in the right panel).

Despite these large inconsistencies, the issues with the Virgo data do not affect our main goal, which is identifying the origin of the evidence for spin-precession. In order to avoid further ambiguities for the remainder of this section we restrict to data from the two LIGO detectors unless otherwise noted. In Fig. 1 we concluded that LIGO Livingston alone drives this measurement and here we attempt to further zero in on the data that support spin-precession by comparing results from a spin-precessing and

a spin-aligned analysis with NRSur7dq4, see Appendix A 2 for details. Figure 4 shows the whitened time-domain reconstruction (left panel) and the spectrum (right panel) in LIGO Hanford (top) and LIGO Livingston (bottom). The two reconstructions remain phase-coherent, however there are some differences in the inferred amplitudes, with the spin-aligned amplitude being slightly larger at ~30–50 Hz and slightly smaller for $\gtrsim 100$ Hz. Comparison to the estimate for the glitch that was subtracted from the data based on information from auxiliary channels with GWSUBTRACT shows that the glitch overlaps with the part of the signal where the spin-precessing amplitude is smaller than the spin-aligned one. The glitch subtraction and data quality issues are therefore related to the evidence for spin-precession.

We confirm that the low-frequency data in LIGO Livingston (in relation to the rest of the data) are the sole source of the evidence for spin-precession, by carrying out analyses with a progressively increasing low-frequency cutoff in LIGO Livingston only, while leaving the LIGO Hanford data intact. Figure 5 shows the effect on the posterior for χ_p, q, and χ_{eff}. When we use the full data bandwidth, $f_{\text{low}}(L) = 20$ Hz, we find that q and χ_p are correlated and their two-dimensional posterior appears similar to the combination of the individual-detector posteriors from Fig. 1. However, as the low-frequency cutoff in LIGO Livingston is increased and the data affected by the glitch are removed, the posterior progressively becomes more consistent with an equal-mass binary and χ_p approaches its prior. By $f_{\text{low}}(L) = 50$ Hz, χ_p is similar to its prior and further increasing $f_{\text{low}}(L)$ has a marginal effect. This confirms that *given all the other data*, the LIGO Livingston data in 20–50 Hz drive the inference for spin-precession.

The signal network SNR (i.e., the SNR in both detectors added in quadrature) is given in the legend for each value of the low frequency cutoff. By $f_{\text{low}}(L) = 50$ Hz where all evidence for spin-precession has been eliminated, the SNR reduction is only 1.5 units, suggesting that the large majority of the signal is consistent with a nonprecessing origin. This might also suggest that χ_p inference is not degraded solely due to loss of SNR, as the latter is very small. The χ_{eff} posterior is generally only minimally affected, with a small shift to higher values driven by the $q - \chi_{\text{eff}}$ correlation [40]. We have verified that these conclusions are robust against re-including the Virgo data (using their full bandwidth).

The above analysis is *not* on its own an indication of data quality issues in LIGO Livingston, but we now turn to an observation that might be more problematic: the $q - \chi_p$ inconsistency between LIGO Hanford and LIGO Livingston identified in Fig. 1. In order to examine whether such an effect could arise from the different Gaussian noise realizations in each detector, we consider simulated signals. We use 100 posterior samples obtained from analyzing
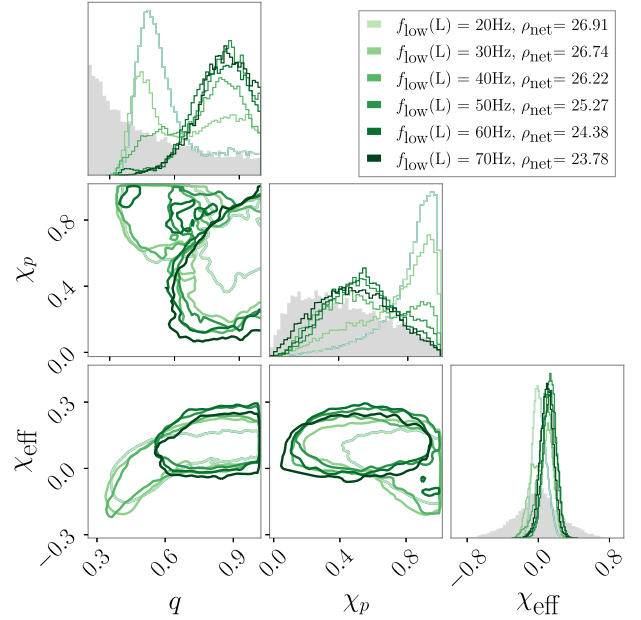

FIG. 5.   One- and two-dimensional marginalized posterior for the mass ratio q, the precession parameter χ_p, and the effective spin parameter χ_{eff} for analyses using a progressively increasing low-frequency cutoff in LIGO Livingston but all the LIGO Hanford data, see Table I for details. The median network SNR for each value of the frequency cutoff is given in the legend. Contours represent 90% credible regions and the prior is shaded in gray. As the glitch-affected data are removed from the analysis, the posterior approaches that of an equal-mass binary and becomes uninformative about χ_p. This behavior does not immediately indicate data quality issues and we only use this increasing-$f_{\text{low}}(L)$ analysis to isolate the data which contribute the evidence of spin-precession when compared to the rest of the data to within 20–50 Hz.

solely the LIGO Livingston data, make simulated data that include a noise realization with the same noise PSDs as GW200129, and analyze data from the two LIGO detectors separately. To quantify the degree to which the LIGO Hanford and LIGO Livingston posteriors overlap, we compute the Bayes factor for overlapping posterior distributions relative to if the two distributions do not overlap [41,42],

$$\mathcal{B}^{\text{overlapping}}_{\text{not overlapping}} = \iint d\chi_p dq \frac{p_L(\chi_p, q|d) p_H(\chi_p, q|d)}{\pi(\chi_p, q)}, \quad (1)$$

where we compute the overlap within the $q - \chi_p$ plane, $p_L(\chi_p, q|d)$ and $p_H(\chi_p, q|d)$ are the LIGO Livingston and LIGO Hanford posteriors, and $\pi(\chi_p, q)$ is the prior. While evaluating this quantity is subject to sizeable sampling uncertainty for events where the two distributions are more distinct (i.e., the case of GW200129), we find $\mathcal{O}(5/100)$ injections have a similar overlap as GW200129 (Fig. 1). Figure 6 shows a selection of $q - \chi_p$ posteriors for 10
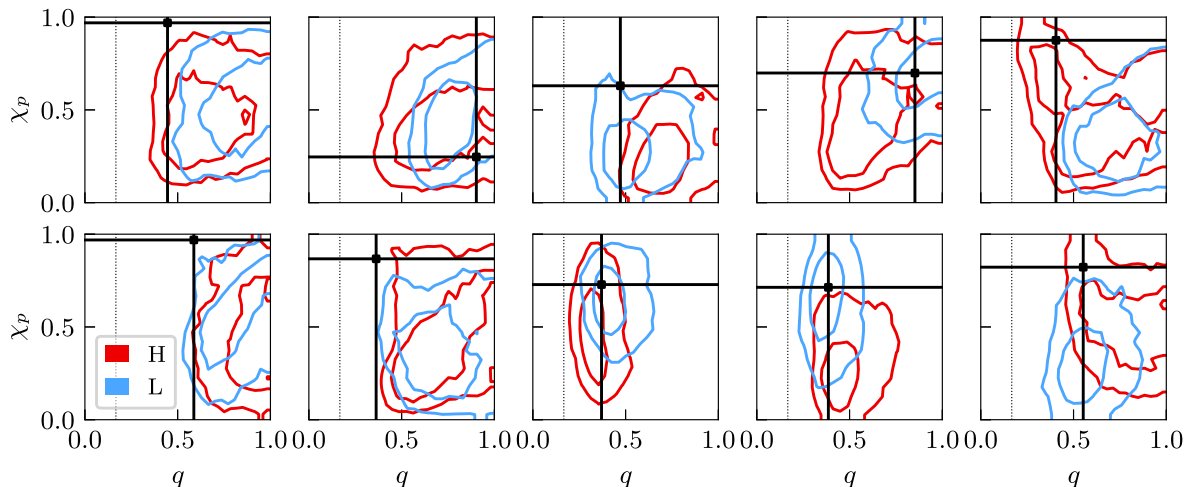
FIG. 6. 90% contours for the two-dimensional marginalized posteriors for the mass ratio q and the precessing parameter χ_p obtained from analyzing data from each LIGO detector separately for 10 simulated signals. The signal parameters are drawn from the posterior for GW200129 when using LIGO Livingston data only and true values are indicated by black lines. Due to the spin priors disfavoring large χ_p, the injected value is outside the two-dimensional 90% contour in some cases. We only encounter an inconsistency between LIGO Hanford (red; H) and LIGO Livingston (blue; L) as observed for GW200129 in Fig. 1 in $\mathcal{O}(5/100)$ injections.

injections as inferred from each detector separately. The posteriors typically overlap, though they are shifted with respect to each other as expected from the different noise realizations.

We conclude that the evidence for spin-precession originates exclusively from the LIGO Livingston data that overlapped with a glitch. This causes an inconsistency between the LIGO Hanford and LIGO Livingston that we typically do not encounter in simulated signals in pure Gaussian noise. This inconsistency suggests that there might be residual data quality issues in LIGO Livingston that were not fully resolved by the original glitch sub-traction. Though inconsequential for the spin-precession investigation, we also identify severe data quality issues in Virgo. Before returning to the investigation of spin-precession, we first examine the Virgo data in detail in Sec. III and argue that they should be removed from subsequent analyses. We reprise the spin-precession inves-tigations in Sec. IV.

III. DATA QUALITY ISSUES: VIRGO

Having established that the Virgo trigger is coincident but not fully coherent with the triggers in the two LIGO detectors, we explore potential reasons for this discrepancy. Figure 7 shows a spectrogram of the data in each detector centered around the time of the event. A clear chirp morphology is visible in the LIGO detectors but not in Virgo, though this might also be due to the low SNR of the Virgo trigger. Within a few seconds of the trigger, however, a number of other glitches are also present in Virgo, mostly assigned to scattered light. We estimate the SNR of the Virgo trigger without assuming it is a CBC signal (i.e., without using a CBC model) through Omicron [43] and

BayesWave using its glitch model that fits the data with sine-Gaussian wavelets, see Table II for run settings.[1] The former finds a matched-filter Omicron SNR[2] of 7.0, while the latter finds an optimal SNR of 7.3 for the median glitch reconstruction.

Given the prevalence of glitches, the first option is that the Virgo trigger is actually a detector glitch that happened to coincide with a signal in the LIGO detectors. To estimate the probability that such a coincidence could happen by chance, we consider the glitch rate in Virgo. In O3, the median rate of glitches in Virgo was 1.11/min, with significant variation versus time [1]. When we consider the hour of data around the event, the rate of glitches with Omicron SNR > 6.5 is 10.2/min. Most of the glitches in Virgo at this time are due to scattered light [46–50]. While Fig. 7 shows that there are scattered light glitches in the Virgo data near the time of GW200129, the excess power from these glitches are concentrated at frequencies < 30 Hz. To account for the excess power corresponding to GW200129 in Virgo, there must be a different type of glitch present in the data. The rate of glitches at frequencies similar to the signal is much lower; using data from four days around the event, the rate of glitches with frequency 60–120 Hz is only 0.06/hr. Given this rate, we calculate the probability that a glitch occurred in Virgo within a 0.06 s window (roughly corresponding to twice the light-travel time between the LIGO detectors and Virgo) around a trigger in

---

[1]The BayesWave analyses described here does not concurrently marginalize over the PSD uncertainty.

[2]The SNR reported by Omicron is normalized so that the expectation value of the SNR is 0, rather than $\sqrt{2}$ [43]. To highlight this difference, we use the phrase "Omicron SNR" whenever a reported result uses this normalization.
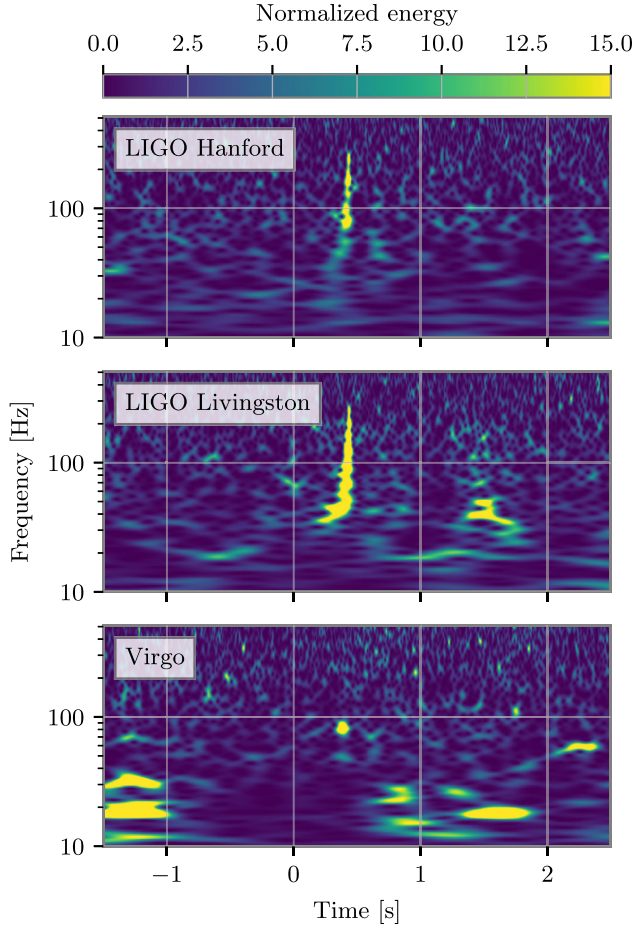
FIG. 7. Spectrogram of the data in each detector, plotted using plotted using the Q-transform [44,45]. Listed times are with respect to GPS 1264316116. Besides the clear chirp morphology in LIGO, there is visible excess power ∼1 s after the signal in LIGO Livingston. Virgo demonstrates a high rate of excess power, though most is due to scattered light and concentrated at frequencies < 30 Hz. The excess power in Virgo that is coincident with GW200129 does not have a chirp morphology.
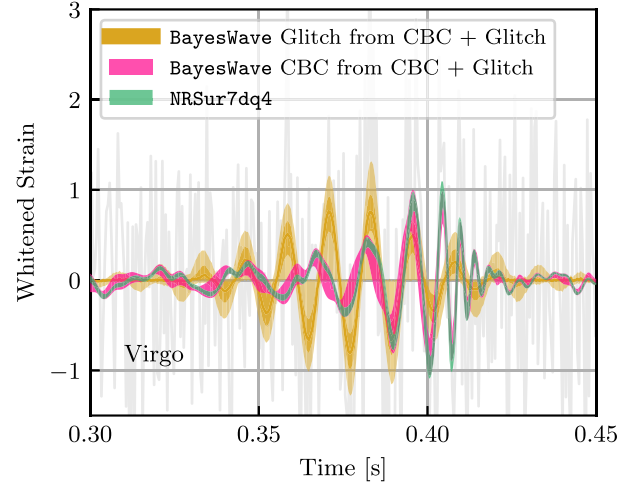

FIG. 8. Whitened time-domain reconstruction of the signal in Virgo obtained after analysis of data from all three detectors relative to GPS 1264316116. Shaded regions correspond to 90% and 50% (where applicable) credible intervals. Green corresponds to the same 3-detector result obtained with NRSur7dq4 as Fig. 3, while pink and gold correspond to the CBC and glitch part of the "CBC + glitch" analysis with BayesWave. See Tables I and II for run settings. The two CBC reconstructions largely overlap, suggesting that the lack of spin-precession in BayesWave's analysis does not affect the reconstruction considerably. A glitch overlapping with the signal is, however, recovered.

the LIGO detectors. We find that if glitches at any frequency are considered, the probability of coincidence per event is $\mathcal{O}(0.01)$, and if only glitches with similar frequencies are considered, the same probability is $\mathcal{O}(10^{-5})$.

Another option is that the Virgo trigger is a combination of a genuine signal and a detector glitch. We explore this possibility using BayesWave [32–34] to simultaneously model a potential CBC signal that is coherent across the detector network and overlapping glitches that are incoherent [35,36]. In this "CBC + glitch" analysis, BayesWave models the CBC signal with the IMRPhenomD waveform [51,52] and glitches with sine-Gaussian wavelets. Details about the models and run settings are provided in Appendix A 3. An important caveat here is that IMRPhenomD does not include the effects of higher-order modes and spin-precession. A concern is, therefore, that the CBC model could fail to model precession-induced

modulations in the signal amplitude and instead assign them to the glitch model. This precise scenario is tested in Hourihane *et al.* [36] where the analysis was shown to be robust against such systematics. Below we argue that the same is true here for the Virgo data, especially since they are consistent with a spin-aligned binary as shown in Fig. 1.

Figure 8 compares BayesWave's reconstruction in Virgo with the one obtained with the NRSur7dq4 analysis from Fig. 3 that ignores a potential glitch but models spin-precession and higher order modes. All results are obtained using data from all three detectors. The CBC reconstruction from BayesWave with IMRPhenomD is consistent with the one from NRSur7dq4 to within the 90% credible level at all times. This is unsurprising given Fig. 1 that shows that Virgo data are consistent with a spin-aligned BBH. Crucially, there is no noticeable difference between the two CBC reconstructions for times when the inferred glitch is the loudest. This suggests that the lack of higher-order modes and spin-precession in IMRPhenomD does not lead to a noticeable difference in the signal reconstruction and could thus not account for the inferred glitch. The differences between the inferred signals using IMRPhenomD and NRSur7dq4 are much smaller than the amount of incoherent power present in Virgo. In fact, the glitch reconstruction is larger than the signal at the 50% credible level, though not at the 90% level. This result suggests that a potential explanation for the trigger in Virgo is a combination of a signal consistent with the one in the LIGO detectors and a glitch.
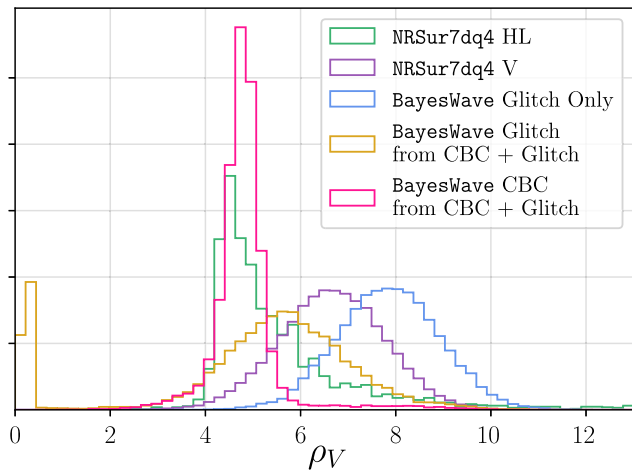
FIG. 9. Comparison of optimal SNR estimates for Virgo from different analyses. In green is the posterior for the expected SNR in Virgo from just the LIGO data using the NRSur7dq4 waveform (HL analysis of Fig. 1), while purple corresponds to the SNR from an analysis of the Virgo data only (V analysis of Fig. 1). The CBC and glitch SNR posterior from BayesWave's full "CBC + glitch" model (Fig. 8) are shown in pink and orange respectively. Part of the latter is consistent with zero, which corresponds to no glitch (as also seen from the 90% credible interval in Fig. 8). The SNR posterior from a "glitchOnly" BayesWave is shown in blue.

Figure 9 summarizes the various SNR estimates for the excess power in Virgo. We plot an estimate of the SNR in Virgo suggested by LIGO data; in other words it is the SNR that is consistent with GW200129 as observed by LIGO. In comparison, we also show the SNR from a Virgo-only analysis and the SNR from BayesWave's "glitchOnly" analysis that models the excess power with sine-Gaussian wavelets without the requirement that it is consistent with a CBC. The fact that the SNR inferred from HL data is smaller than the other two again suggests that the Virgo trigger is not consistent with the one seen by LIGO and contains additional power. BayesWave's "CBC + glitch" analysis is able to separate the part of the trigger that is consistent with a CBC and recovers a CBC SNR that is consistent to the one inferred from LIGO only. The "remaining" power is assigned to a glitch with SNR $\sim$ 4.6 (computed through the median BayesWave glitch reconstruction).

Based on the glitch SNR calculated by the BayesWave "CBC + glitch" model, we revisit the probability of overlap with a signal based on the SNR distribution of Omicron triggers. Since the lowest SNR recorded in Omicron analyses is 5.0, we fit the SNR distribution of glitches with Omicron SNR > 5.0 with a power-law and extrapolate to SNR 4.6. We find that the rate of glitches with frequencies similar to the one in Fig. 8 with SNR > 4.6 is 0.31/min and the probability of overlap with a signal in Virgo is $\mathcal{O}(10^{-3})$. Given the 60 events from GWTC-3 that

were identified in Virgo during O3, the overall chance of at least one glitch of this SNR overlapping a signal is $\mathcal{O}(0.1)$.

The above studies suggest that the most likely scenario is that the Virgo trigger consists of a signal and a glitch. However, due to the low SNR of both, this interpretation is subject to sizeable statistical uncertainties and we therefore do not attempt to make glitch-subtracted Virgo data. Such data would be extremely dependent on which glitch reconstruction we chose to subtract, for example the median or a fair draw from the BayesWave glitch posterior. For these reasons and due to its low sensitivity, we do not include Virgo data in what follows.

IV. DATA QUALITY ISSUES: LIGO LIVINGSTON

The data quality issues in LIGO Livingston were identified and mitigated in GWTC-3 [1] through use of information from auxiliary channels [5,37] and the GWSUBTRACT pipeline as also described in Appendix A 1. The comparison of Figs. 1 and 6, however, suggest that residual data quality issues might remain, as the two LIGO detectors result in inconsistent inferred $q - \chi_p$ parameters beyond what is expected from typical Gaussian noise fluctuations. Here we revisit the LIGO Livingston glitch with BayesWave and again model both the CBC and potential glitches. This analysis offers a point of comparison to GWSUBTRACT as it uses solely strain data to infer the glitch instead of auxiliary channels. Additionally, BayesWave computes a posterior for the glitch, rather than a single point
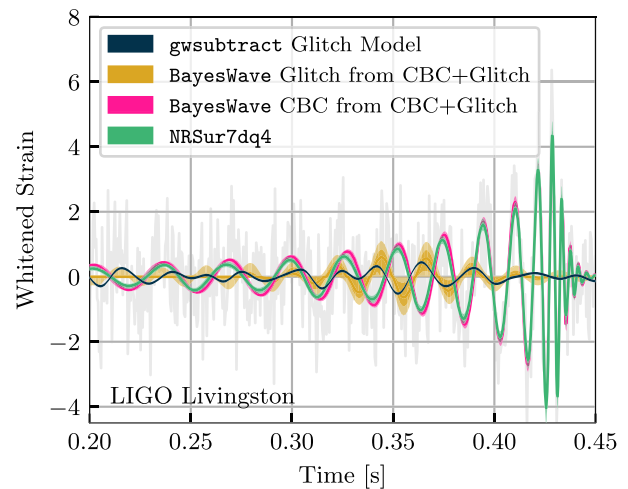

FIG. 10. Whitened time-domain reconstruction of the data in LIGO Livingston obtained after analysis of data from the two LIGO detectors. Shaded regions correspond to 90% and 50% (where applicable) credible intervals and gray gives the original data without any glitch mitigation. Green corresponds to the same 2-detector result obtained with NRSur7dq4 as Fig. 4, while pink and gold correspond to the CBC and glitch part of the joint "CBC + glitch" analysis with BayesWave. The black line shows an estimate for the glitch obtained through auxiliary channels. All analyses use only LIGO data.

estimate, and thus allows us to explore the statistical uncertainty of the glitch mitigation. In all analyses involving BayesWave we use the original LIGO Livingston data without any of the data mitigation described in Appendix A 1.

Figure 10 shows BayesWave's CBC and glitch reconstructions in LIGO Livingston compared to the one based on the NRSur7dq4 (from glitch-mitigated data) and the glitch model computed with GWSUBTRACT. All analyses use data from the two LIGO detectors only. Unsurprisingly, now, the CBC reconstructions based on IMRPhenomD and NRSur7dq4 do not fully overlap around $t = 0.3$ s, though they are consistent during the signal merger phase. This is expected from the fact that LIGO Livingston supports spin-precession as well as Fig. 4. However, this difference is *smaller* than the statistical uncertainty in the inferred glitch from BayesWave (yellow) and well as differences between the BayesWave and the GWSUBTRACT glitch estimates. This suggests that even though the BayesWave glitch estimate might be affected by the lack of spin-precession in its CBC model, this effect is smaller than the glitch uncertainty.

We also model the signal as a superposition of coherent wavelets in addition to the incoherent glitch wavelets using BayesWave [32–34]. This approach has been previously utilized for glitch subtraction [1]. However, we do not recover strong evidence for a glitch overlapping the signal in LIGO Livingston when running with this "signal + glitch" analysis. The "signal + glitch" analysis attempts to describe both the signal and the glitch with wavelets and hence it is significantly less sensitive than the "CBC + glitch" model. In the data of interest, both the signal and the glitch whitened amplitudes are $\sim 1\sigma$ and as such they are difficult to separate using coherent and incoherent wavelets. Given that we know (based on the auxiliary channel data) that there is some non-Gaussian noise in LIGO Livingston, we find that the "signal + glitch" analysis is not sensitive enough for our data.

The large statistical uncertainty in the glitch reconstruction (yellow bands in Fig. 10) implies that the difference between the spin-precessing and nonprecessing interpretation of GW200129 cannot be reliably resolved. To confirm this, we select three random samples from the
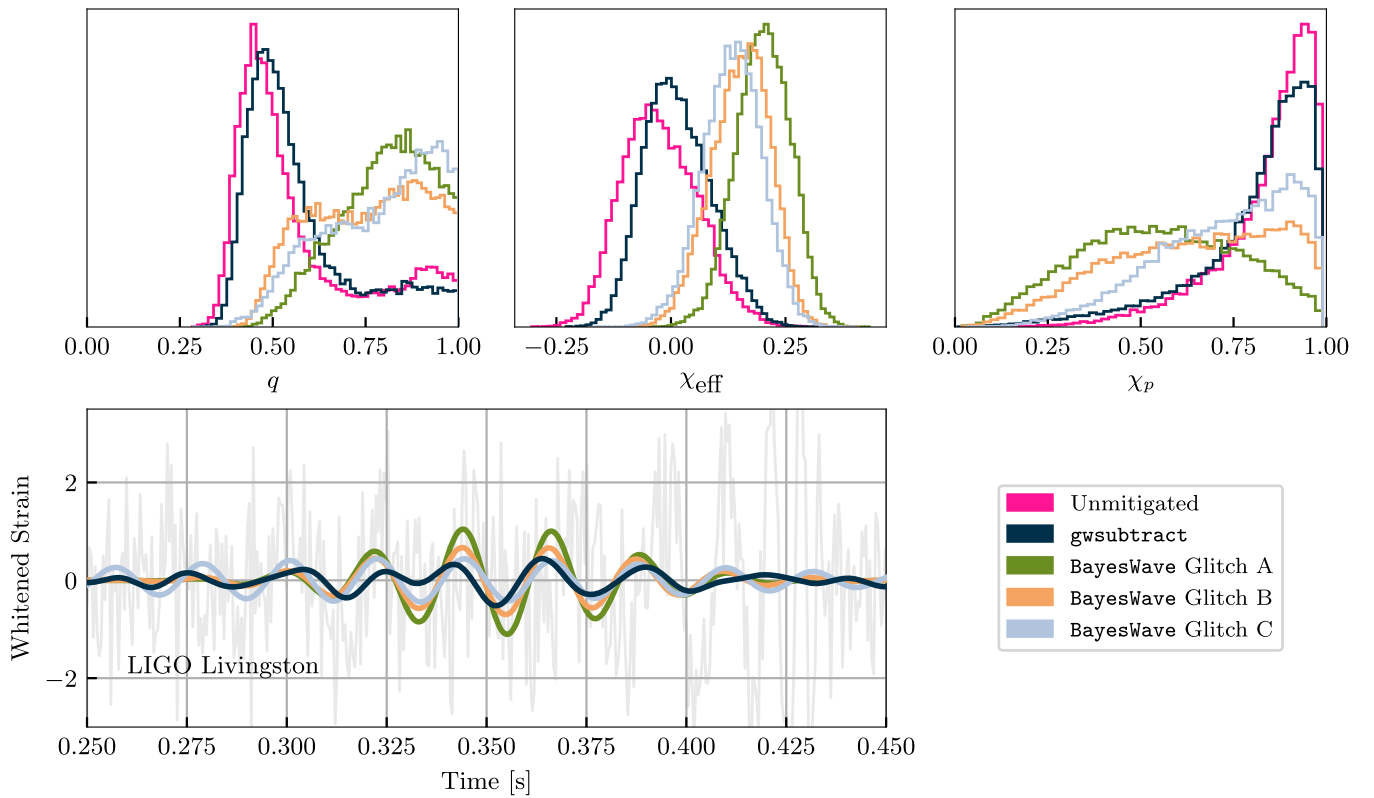

FIG. 11.   Bottom: Whitened, time domain reconstructions of various glitch reconstructions subtracted from LIGO Livingston data. The green line corresponds to the glitch reconstruction obtained from auxiliary data using GWSUBTRACT. The rest are glitch posterior draws from the BayesWave "CBC + glitch" analysis on HL unmitigated data. Top: Marginalized posterior distributions corresponding to parameter estimation performed with the NRSur7dq4 waveform model on HL data where each respective glitch realization was subtracted from LIGO Livingston (same colors). Pink corresponds to the original data without any glitch subtraction. Larger glitch reconstruction amplitudes roughly lead to less informative χ_p posteriors and eliminate the $q - \chi_p$ inconsistency between LIGO Hanford and LIGO Livingston.
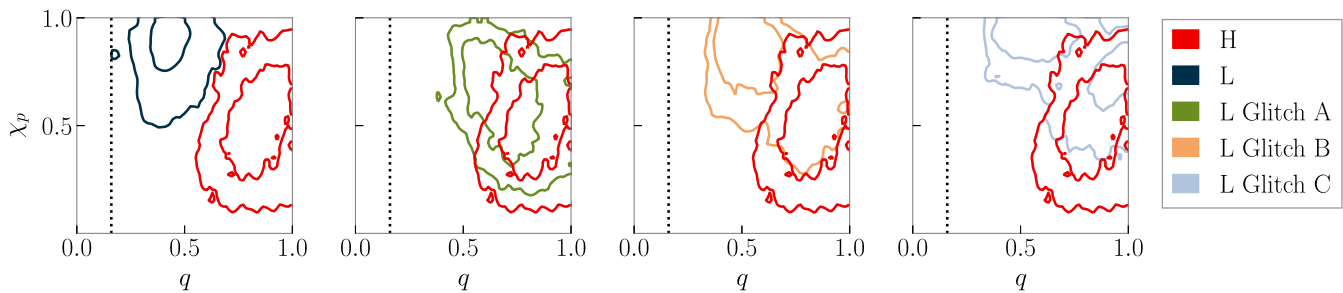
FIG. 12. Two-dimensional posterior distributions for χ_p and q (50% and 90% contours) from single-detector parameter estimation runs. The far left panel shows the same tension as the LIGO Hanford and LIGO Livingston data plotted in Fig. 1 when using the GWSUBTRACT estimate for the glitch. Subsequent figures show inferred posterior distributions using data where the same three different BAYESWAVE glitch models as Fig. 11 have been subtracted. These results show less tension between the two posterior distributions.

glitch posterior of Fig. 10, subtract them from the unmitigated LIGO Livingston data, and repeat the parameter estimation analysis with NRSur7dq4. The BayesWave glitch-subtracted frames and associated NRSur7dq4 parameter estimation results are available in [53]. For reference, we also analyze the original unmitigated data (no glitch subtraction whatsoever). Figure 11 confirms that the spin-precession evidence depends sensitively on the glitch subtraction. The original unmitigated data and the GWSUBTRACT subtraction yield the largest evidence for spin-precession, but this is reduced -or completely eliminated- with different realizations of the BayesWave glitch model. In general, larger glitch amplitudes lead to less support for spin-precession, suggesting that the evidence for spin-precession is increased when the glitch is *undersubtracted*.

Figure 12 compares the corresponding $q - \chi_p$ posterior inferred from LIGO Hanford and LIGO Livingston separately under each different estimate for the glitch. Each of the 3 BayesWave glitch draws results in single-detector posteriors that fully overlap, thus resolving the inconsistency seen in $q - \chi_p$ when using the GWSUBTRACT glitch estimate. Due to the lack of spin-precession modeling in the "CBC + glitch" analysis of Fig. 10, however, we cannot definitively conclude that any one of the new glitch-subtracted results is preferable. The 3 Bayewave glitch draws result in different levels of support for spin-precession, it is therefore possible that GW200129 is still consistent with a spin-precessing system. We do conclude, though, that the evidence for spin-precession is contingent upon the large statistical uncertainty of the glitch subtraction.

As a further check of whether the lack of spin-precession in BayesWave's CBC model could severely bias a potential glitch recovery, we revisit the 10 simulated signals from Fig. 6 and analyze them with the "CBC + glitch" model. These signals are consistent with GW200129 as inferred from LIGO Livingston data only, and thus exhibit the largest amount of spin-precession consistent with the signal. In all cases we find that the glitch part of the "CBC + glitch" model has median and 50% credible intervals that are consistent with zero at all times. This again confirms

that the differences between the spin-precessing and the spin-aligned inferred signals in Fig. 10 is smaller than the uncertainty in the glitch. This tests suggests that the glitch model is not strongly biased by the lack of spin-precession, however it does not preclude small biases (within the glitch statistical uncertainty); it is therefore necessary but not sufficient.

As a final point of comparison between BayesWave's glitch reconstruction that is based on strain data and the GWSUBTRACT glitch reconstruction based on auxiliary channels, we consider a *different* glitch in LIGO Livingston approximately 1s after the signal, see Fig. 7. Studying this glitch offers the advantage of direct comparison of the two glitch reconstruction methods without contamination from the CBC signal and uncertainties about its modeling. We analyze the original data with no previous glitch mitigation around that glitch using BayesWave's glitch model and plot the results in Fig. 13. For the GWSUBTRACT reconstruction we also include 90% confidence intervals, as described in Appendix A 1.

The two estimates of the glitch are broadly similar but they do not always overlap within their uncertainties. The main disagreement comes from the sharp data "spike" at $t = 1.43$ s that is missed by GWSUBTRACT, but recovered by BayesWave. The reason is that the maximum frequency considered by GWSUBTRACT was 128 Hz and thus cannot capture such a sharp noise feature [5]. Away from the "spike," the two glitch estimates are approximately phase-coherent. On average BayesWave recovers a larger glitch amplitude as the GWSUBTRACT result typically falls on BayesWave's lower 90% credible level.

Figures 10 and 13 broadly suggest that BayesWave recovers a higher-amplitude glitch. Figure 11 shows that the evidence for spin-precession is indeed reduced, the LIGO Hanford-LIGO Livingston inconsistency is alleviated (Fig. 12), and the LIGO Livingston data become more consistent across low and high frequencies (Fig. 5) if the glitch was originally undersubtracted. However, due to the low SNR of the glitch and other systematic uncertainties it is not straightforward to select a "preferred" set of
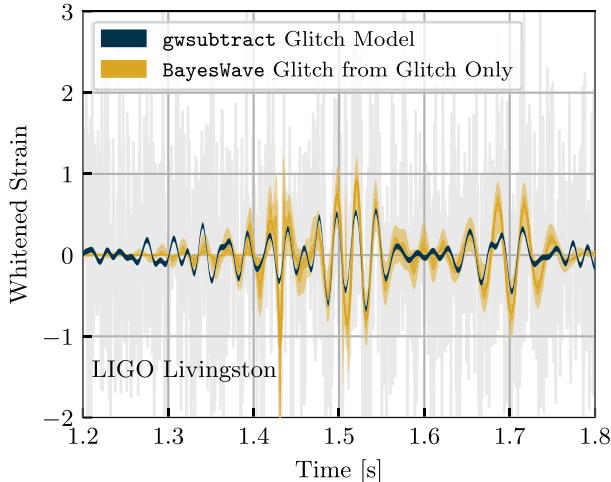
FIG. 13. Comparison between the two glitch reconstruction and subtraction methods for a glitch in LIGO Livingston ~ 1 s after GW200129, see the middle panel of Fig. 7. We plot the original data with no glitch mitigation (gray), the glitch reconstruction obtained from auxiliary channels with 90% confidence intervals (black), and the 50% and 90% credible intervals for the glitch obtained with BayesWave that uses only the strain data (gold).

glitch-subtracted data. All studies, however, indicate that the statistical uncertainty of the glitch amplitude is larger than the difference between the inferred spin-precessing and spin-aligned signals.

V. CONCLUSIONS

Though it might be possible to infer the presence of spin-precession and large spins in heavy BBHs, our investigations suggest that in the case of GW200129 any such evidence is contaminated by data quality issues in the LIGO Livingston detector. In agreement with [14] we find that the evidence for spin-precession originates exclusively from data from that detector. However, we go beyond this and also demonstrate the following.

(1) The evidence for spin-precession in LIGO Livingston is localized in the 20–50 Hz band in comparison to the rest of the data, precisely where the glitch overlapped the signal. Excluding this frequency range from the analysis, we find that GW200129 is consistent with an equal-mass BBH with an uninformative χ_p posterior; it is thus similar to the majority of BBH detections [4,54,55]. However, the fact that there is no evidence for spin-precession if $f_{\text{low}}(L) > 50$ Hz is not on its own cause for concern as it might be due to Gaussian noise fluctuations or the precise precessional dynamics of the system.

(2) LIGO Hanford is not only uninformative about spin-precession (which again could be due to Gaussian noise fluctuations or the lower signal SNR in that detector), but it also yields an *inconsistent* $q - \chi_p$ posterior compared to LIGO Livingston. Using

simulated signals, we find that the latter, i.e., the $q - \chi_p$ inconsistency, is larger than $\mathcal{O}(95\%)$ of results expected from Gaussian noise fluctuations.

(3) Given the LIGO Livingston glitch's low SNR, the statistical uncertainty in modeling it is *larger* than the difference between a spin-precessing and a nonprecessing analysis for GW200129. Inferring the presence of spin-precession requires reliably resolving this difference, something challenging as we found by using different realizations of the glitch model from the BayesWave glitch posterior. Crucially, any evidence for spin-precession in GW200129 depends sensitively on the glitch model and priors employed.

(4) Given the large statistical uncertainty in modeling the glitch, evidence for systematic differences between BayesWave and GWSUBTRACT that use strain and auxiliary data respectively is tentative. However, the BayesWave estimate typically predicts a larger glitch amplitude, which would reduce the evidence for spin-precession and alleviate the tension between LIGO Hanford and LIGO Livingston. Additionally, we do not recover any support for a glitch when injecting spin-precessing signals from the LIGO Livingston-only posterior distribution into Gaussian noise. This indicates that BayesWave is unlikely to be strongly biasing the glitch recovery due to its lack of spin-precession.

Overall, given the uncertainty surrounding the LIGO Livingston glitch mitigation, we cannot conclude that the source of GW200129 was spin-precessing. We do not conclude the opposite either, however. Though we obtain tentative evidence that the glitch was undersubtracted, we can at present not estimate how much it was undersubtracted by due to large statistical and potential systematic uncertainties. It is possible that some evidence for spin-precession remains, albeit reduced given the glitch statistical uncertainty.

In addition, we verify that this uncertainty in the glitch modeling is larger than uncertainty induced by detector calibration. We repeat select analyses in Appendix A 2 and confirm that the inclusion of uncertainty in the calibration of the gravitational-wave detectors negligibly impacts the spin-precession inference, as expected. Indeed, the glitch impacts the data at a level comparable to the signal strain, cf., Fig. 10, whereas the calibration uncertainty within 20 to 70 Hz is only $\sim 5\%$ in amplitude and $5°$ in phase [56]. Therefore, the glitch in LIGO Livingston's data dominates over uncertainties about the data calibration.

Though not critical to the discussion and evidence for spin-precession, we also identified data quality issues in Virgo. The inconsistency between Virgo and the LIGO detectors is in fact more severe than the one between the two LIGO detectors, however the Virgo data do not influence the overall signal interpretation due to the low-signal SNR in Virgo. Nonetheless, we argue that the most

likely explanation is that the Virgo data contain both the GW200129 signal and a glitch.

These conclusions are obtained with NRSur7dq4, which is expected to be the more reliable waveform model including spin-precession and higher-order modes in this region of the parameter space [14,16]. We repeated select analyses with IMRPhenomXPHM which also favored a spin-precessing interpretation for GW200129 [1]. We found largely consistent but not identical results between NRSur7dq4 and IMRPhenomXPHM, suggesting that there are additional systematic differences between the two waveform models. Appendix B shows some example results. Nonetheless, our results are directly comparable to the ones of [14,15] as they were obtained with the same waveform model.

Our analysis suggests that extra caution is needed when attempting to infer the role of subdominant physical effects in the detected GW signals, for example spin-precession or eccentricity. Low-mass signals are dominated by a long inspiral phase that in principle allows for the detection of multiple spin-precession cycles or eccentricity-induced modulations. However, the majority of detected events, such as GW200129, have high masses and are dominated by the merger phase. The subtlety of the effect of interest and the lack of analytical understanding might make inference susceptible not only to waveform systematics, but also (as argued in this study) potential small data quality issues.

Indeed, Fig. 11 shows that a difference in the glitch amplitude of $< 0.5\sigma$ can make the difference between an uninformative χ_p posterior and one that strongly favors spin-precession. This also demonstrates that low-SNR glitches are capable of biasing inference of these subtle physical effects. Low-SNR departures from Gaussian noise have been commonly observed by statistical tests of the residual power present in the strain data after subtracting the best-fit waveform of events [57–59]. If indeed such low-SNR glitches are prevalent, they might be individually indistinguishable from Gaussian noise fluctuations. Potential ways to safeguard our analyses and conclusions against them are (i) the detector and frequency band consistency checks performed here, (ii) extending the BayesWave "CBC + glitch" analysis to account for spin-precession and eccentricity while carefully accounting for the impact of glitch modeling and priors especially for low-SNR glitches, (iii) and modeling insight on the morphology of subtle physical effects of interest such as spin-precession and eccentricity in relation to common detector glitch types.

ACKNOWLEDGMENTS

APPENDIX A: ANALYSIS DETAILS

In this Appendix we provide details and settings for the analyses presented in the main text. All data are obtained via the GW Open Science Center [70]. Throughout we use geometric units, $G = c = 1$.

1. Detection and Glitch-subtracted data

GW200129 was identified in low latency [71] by GstLAL [72,73], cWB [74], PyCBC Live [75,76], MBTA Online [77], and SPIIR [78]. The quoted false alarm rate of the signal in low latency was approximately 1 in 10^{23} years, making this an unambiguous detection. Below we recap the detection and glitch mitigation process from [1].

Multiple data quality issues were identified in the data surrounding GW200129. As a part of the rapid response procedures, scattered light noise [49,79] was identified in the Virgo data, as seen in Fig. 7 in the frequency range 10–60 Hz. These glitches did not overlap the signal, and no mitigation steps were taken with the Virgo data. During offline investigations of the LIGO Livingston data quality, a malfunction of the 45 MHz electro-optic modulator system [80] was found to have caused numerous glitches in the days surrounding GW200129. To help search pipelines differentiate these types from glitches, a data quality flag was generated for this noise source [81]. These data quality vetoes are used by some pipelines to veto any candidates

identified during the data quality flag time segments [82]. The glitches from the electro-optic modulator system directly overlapped GW200129, meaning that the time of the signal overlapped the time of the data quality flag.

Although clearly an astrophysical signal, the data quality issues present in LIGO Livingston introduced additional complexities into the estimation of the significance of this signal [1]. Due to the data quality veto, the signal was not identified in LIGO Livingston by the PyCBC [83,84] MBTA [85], and cWB [74] pipelines. PyCBC was still able to identify GW200129 as a LIGO Hanford—Virgo detection, but the signal was not identified by MBTA due to the high SNR in LIGO Hanford and cWB due to postproduction cuts. The GstLAL [86,87] analysis did not incorporate data quality vetoes in its O3 analyses and was therefore able to identify the signal in all three detectors.

The excess power from the glitch directly overlapping GW200129 in LIGO Livingston was subtracted before estimation of the signal's source properties [1,5] using the GWSUBTRACT algorithm [37]. This method relies on an auxiliary sensor at LIGO Livingston that also witnesses glitches present in the strain data. The transfer function between the sensor and the strain data channel is measured using a long stretch of data by calculating the inner product of the two time series with a high-frequency resolution and then averaging the measured value at nearby frequencies to produce a transfer function with lower frequency resolution [88]. This transfer function is convolved with the auxiliary channel time series to estimate the contribution of this particular noise source to the strain data. Therefore, the effectiveness of this subtraction method is limited by the accuracy of the auxiliary sensor and the transfer function estimate. This tool was previously used for broadband noise subtraction with the O2 LIGO dataset [37], but this was the first time it was used for targeted glitch subtraction. Additional details about the use of GWSUBTRACT for the GW200129 glitch subtraction can be found in Davis *et al.* [5].

The GWSUBTRACT glitch model does not include a corresponding interval that accounts for all sources of statistical errors as is done by BayesWave. However, a confidence interval based on only uncertainties due to random correlations between the auxiliary channel and the strain data can be computed. For the GW200129 glitch model, this interval is ± 0.022 in the whitened strain data [5]. Additional systematic uncertainties due to time variation in the measured transfer function and effectiveness of the chosen auxiliary channel are expected to be present but are not quantified. The relative size of these uncertainties is dependent on the specific noise source that is being modeled and chosen auxiliary channel.

2. BILBY parameter estimation analyses

Quasicircular BBHs are characterized by 15 parameters, divided into eight intrinsic and seven extrinsic parameters.

Each component BH has source frame mass m_i^s, $i \in \{1, 2\}$. In the main text we mainly use the corresponding detector frame (redshifted) masses $m_i = (1 + z)m_i^s$, where z is the redshift, as we are interested in investigating data quality issues and detector frame quantities better relate to the signal as observed. Each component BH also has dimensionless spin vector $\vec{\chi}_i$, and χ_i is the magnitude of this vector. We also use parameter combinations that are useful in various contexts: total mass $M = m_1 + m_2$, mass ratio $q = m_2/m_1 < 1$, chirp mass $\mathcal{M} = (m_1 m_2)^{3/5}(m_1 + m_2)^{-1/5}$ [89–91], effective orbit-aligned spin parameter [92–94]

$$\chi_{\mathrm{eff}} = \frac{\vec{\chi}_1 \cdot \vec{L} + q\vec{\chi}_2 \cdot \vec{L}}{1 + q}, \tag{A1}$$

where $\vec{L}$ is the Newtonian orbital angular momentum, and effective precession spin parameter [23,95]

$$\chi_p = \max \left(\chi_{1\perp}, q\chi_{2\perp} \frac{3q + 4}{4q + 3} \right), \tag{A2}$$

where $\chi_{1\perp}$ is the $\vec{\chi}_i$ component that is perpendicular to $\vec{L}$. The remaining parameters are observer dependent, and hence referred to as extrinsic. The right ascension α and declination δ designate the location of the source in the sky, while the luminosity distance to the source is d_L. The angle between total angular momentum and the observer's line of sight is θ_{jn}; for systems without perpendicular spins it reduces to the inclination ι, the angle between the orbital angular momentum and observer's line of sight. The time of coalescence t_c is the geocenter coalescence time of the binary. The phase of the signal ϕ is defined at a given reference frequency, and the polarization angle ψ completes the geometric description of the sources position and orientation relative to us; neither of these are used directly in this work.

Parameter estimation results are obtained with PARALLEL BILBY [8,9,96] using the nested sampler, DYNESTY [97]. The numerical relativity surrogate, NRSur7dq4 [16], is used for all main results due to its accuracy over the regime of highly-precessing signals. Its space of validity is limited by the availability of numerical simulations [98] to $q > 1/4$ and component spin magnitudes $\chi < 0.8$, though it maintains reasonable accuracy when extrapolated to $q > 1/6$ and $\chi < 1$ [16].

The majority of our analyses use the publicly released strain data, including the aforementioned glitch subtraction in LIGO Livingston [5], and noise power spectral densities (PSDs) [1]. The exception to the publicly released data was the construction of glitch-subtracted strain data using BayesWave for LIGO Livingston, as discussed in Sec. IV. We do not incorporate the impact of uncertainty about the detector calibration as the SNR of the signal is far below the anticipated regime where calibration uncertainty is

TABLE I.   Table of BILBY runs and settings. All analyses use 4 s of data, and a sampling rate of 4096 Hz. Columns correspond to the main text figures each analysis appears in, the waveform model, the detector network used (H: LIGO Hanford, L: LIGO Livingston, V: Virgo), the type of glitch mitigation in LIGO Livingston, and the low-frequency cutoff of the analysis. Figure 6 also presents results for a set of 10 injections drawn from the LIGO Livingston only posterior distribution with $f_{\text{low}}(L) = 20$ Hz. These analyses use the same settings as above with $f_{\text{low}}(L) = 20$ Hz.

Figure(s)	Waveform model	Detector network	Glitch mitigation	f_{low} (Hz)
1, 12	NRSur7dq4	H	GWSUBTRACT	20
1, 12	NRSur7dq4	L	GWSUBTRACT	20
1, 2, 3	NRSur7dq4	V	GWSUBTRACT	20
1, 2, 3, 8	NRSur7dq4	HLV	GWSUBTRACT	20
1, 2, 4, 10, 11, 14	NRSur7dq4	HL	GWSUBTRACT	20
4	NRSur7dq4 spin-aligned	HL	GWSUBTRACT	20
5	NRSur7dq4	HL	GWSUBTRACT	$\{20, 30, 40, 50, 60, 70\}$ in L, 20 in H
11	NRSur7dq4	HL	No mitigation	20
11	NRSur7dq4	HL	BayesWave fair draws	20
12	NRSur7dq4	L	BayesWave fair draws	20
14	IMRPhenomXPHM	HL	GWSUBTRACT	20

non-negligible [99–102]. Furthermore, we confirm that including marginalization of calibration uncertainty does not qualitatively change the recovered posterior distributions or our main conclusions by also directly repeating select runs.

As is done in GWTC-3 [1], we choose a prior that is uniform in detector frame component masses, while sampling in chirp mass and mass ratio. The mass ratio prior bounds are 1/6 and 1, where we utilize the extrapolation region of NRSur7dq4. Since NRSur7dq4 is trained against numerical relativity simulations which typically have a short duration, only a limited number of cycles are captured before coalescence. With a reduced signal model duration, our analysis is restricted to heavier systems so that the model has content spanning the frequencies analyzed (20 Hz and above). We therefore enforce an additional constraint on the total detector-frame mass to be greater than 60 $M_\odot$. We verify that our posteriors reside comfortably above this lower bound. The luminosity distance prior is chosen to be uniform in comoving volume. The prior distribution on the sky location is isotropic with a uniform distribution on the polarization angle. Finally, for most analyses, the prior on the spin distributions is isotropic in orientation and uniform in spin magnitude up to $\chi = 0.99$. For the spin-aligned analyses, a prior is chosen on the aligned spin to mimic an isotropic and uniform spin magnitude prior. These settings and data are utilized in conjunction with differing GW detector network configurations and minimum frequencies in LIGO Livingston. The differences between runs and their corresponding figures are presented in Table I.

3. BayesWave CBC and glitch analyses

BayesWave [32–34] is a flexible data analysis algorithm that models combinations of coherent generic signals, glitches, Gaussian noise, and most recently, CBC signals that appear in the data [35,36,103]. To sample from the multidimensional posterior for all the different models, BayesWave uses a "Gibbs sampler" which cycles between sampling different models while holding the parameters of the nonsampling model(s) fixed.

For this analysis, we mainly use the CBC and glitch models (a setting we refer to as "CBC + glitch"). The CBC model parameters (see Appendix A 2) are sampled via a fixed-dimension Markov Chain Monte Carlo sampler (MCMC) using the priors described in Wijngaarden *et al.* [103]. The glitch model is based on sine-Gaussian wavelets and samples over both the parameters of each wavelet (central time, central frequency, quality factor, amplitude,

TABLE II.   Table of BayesWave runs and settings. All analyses use 4 s of data, a low-frequency cut-off of $f_{\text{low}} = 20$ Hz, a sampling rate of 2048 Hz, and the IMRPhenomD waveform when the CBC model is used. Furthermore, all analyses use the original strain data without the glitch mitigation described in Sec. A 1. Columns correspond to the main text figures each analysis appears in, the BayesWave models that are used, and the detector network (H: LIGO Hanford, L: LIGO Livingston, V: Virgo). While not plotted in any figure, we also performed "CBC + glitch" analyses on injections into the HL detector network as a glitch background study on GW200129-like sources, see Sec. IV.

Figure(s)	Models	Detector network
8, 9	CBC + glitch	HLV
10, 11	CBC + glitch	HL
9	Glitch	V
13	Glitch	L

phase [32]) and the number of wavelets via a transdimensional or reverse-jump MCMC. In some cases, we also make use of solely the glitch model (termed "glitchOnly" analyses) that assumes no CBC signal and the excess power is described only with wavelets. The differences between runs and the figures in which they appear are presented in Table II.

Though BayesWave typically marginalizes over uncertainty in the noise PSD [33], in this work we use the same fixed PSD as the BILBY runs for more direct comparisons. Additionally, we use identical data as Appendix A 2 for the LIGO Hanford and Virgo detectors. However, when it comes to LIGO Livingston we use the original (i.e., "unmitigated," without any glitch subtraction) data in order to independently infer the glitch. We do not marginalize over uncertainty in the detector calibration.

APPENDIX B: SELECT RESULTS WITH
IMRPhenomXPHM

In this Appendix, we present select results obtained with the IMRPhenomXPHM [6] waveform model that also resulted in evidence for spin-precession in GWTC-3 [1]. Even though IMRPhenomXPHM and NRSur7dq4 both support spin-precesion, in contrast to SEOBNRv4PHM, there are still noticeable systematic differences between them. Figure 14 shows that while NRSur7dq4 and IMRPhenomXPHM generally have overlapping regions of posterior support, IMRPhenomXPHM shows slightly more preference for higher q and less support for extreme precession when
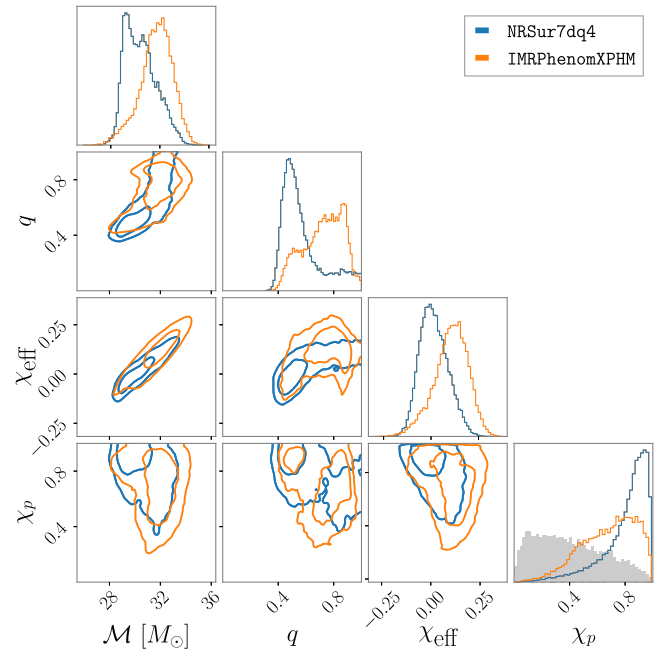

FIG. 14. Similar to Fig. 1, using data from LIGO Livingston and LIGO Hanford. The comparison shows slight tension between results when using NRSur7dq4 and IMRPhenomXPHM, though qualitatively IMRPhenomXPHM also seems to support the evidence for spin-precession.

compared to NRSur7dq4. Waveform systematics are expected to play a significant role in GW200129's inference (e.g., Refs. [1,14,104]), which motivates utilizing NRSur7dq4 for all of our main text results.

[1] The LIGO Scientific, the Virgo, and the KAGRA Collaborations, arXiv:2111.03606.
[2] J. Aasi *et al.* (LIGO Scientific Collaborations), Classical Quantum Gravity **32**, 074001 (2015).
[3] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[4] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA Collaborations), arXiv:2111.03634.
[5] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, arXiv:2207.03429.
[6] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021).
[7] S. Ossokine *et al.*, Phys. Rev. D **102**, 044055 (2020).
[8] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[9] I. M. Romero-Shaw *et al.*, Mon. Not. R. Astron. Soc. **499**, 3295 (2020).
[10] D. Wysocki, R. O'Shaughnessy, J. Lange, and Y.-L. L. Fang, Phys. Rev. D **99**, 084026 (2019).
[11] M. Pürrer and C.-J. Haster, Phys. Rev. Res. **2**, 023151 (2020).
[12] G. Pratten, P. Schmidt, R. Buscicchio, and L. M. Thomas, Phys. Rev. Res. **2**, 043096 (2020).
[13] G. Ashton, S. Thiele, Y. Lecoeuche, J. McIver, and L. K. Nuttall, Classical Quantum Gravity **39**, 175004 (2022).
[14] M. Hannam, C. Hoy, J. E. Thompson, S. Fairhurst, and V. Raymond, arXiv:2112.11300.
[15] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C.-J. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale, Phys. Rev. Lett. **128**, 191102 (2022).
[16] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. Res. **1**, 033015 (2019).
[17] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994).
[18] L. E. Kidder, Phys. Rev. D **52**, 821 (1995).
[19] A. Buonanno, Y.-b. Chen, and M. Vallisneri, Phys. Rev. D **67**, 104025 (2003); **74**, 029904(E) (2006).
[20] A. Buonanno, Y.-b. Chen, Y. Pan, and M. Vallisneri, Phys. Rev. D **70**, 104003 (2004); **74**, 029902(E) (2006).

[21] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Phys. Rev. D **84,** 024046 (2011).

[22] P. Schmidt, M. Hannam, and S. Husa, Phys. Rev. D **86,** 104063 (2012).

[23] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113,** 151101 (2014).

[24] K. Chatziioannou, A. Klein, N. Cornish, and N. Yunes, Phys. Rev. Lett. **118,** 051101 (2017).

[25] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Phys. Rev. D **95,** 104004 (2017).

[26] D. Gerosa, M. Kesden, U. Sperhake, E. Berti, and R. O'Shaughnessy, Phys. Rev. D **92,** 064016 (2015).

[27] M. Kesden, D. Gerosa, R. O'Shaughnessy, E. Berti, and U. Sperhake, Phys. Rev. Lett. **114,** 081103 (2015).

[28] A. Ramos-Buades, P. Schmidt, G. Pratten, and S. Husa, Phys. Rev. D **101,** 103014 (2020).

[29] S. Fairhurst, R. Green, C. Hoy, M. Hannam, and A. Muir, Phys. Rev. D **102,** 024055 (2020).

[30] R. O'Shaughnessy, L. London, J. Healy, and D. Shoemaker, Phys. Rev. D **87,** 044038 (2013).

[31] S. Biscoveanu, M. Isi, V. Varma, and S. Vitale, Phys. Rev. D **104,** 103018 (2021).

[32] N. J. Cornish and T. B. Littenberg, Classical Quantum Gravity **32,** 135012 (2015).

[33] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91,** 084034 (2015).

[34] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, Phys. Rev. D **103,** 044006 (2021).

[35] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, Phys. Rev. D **103,** 044013 (2021).

[36] S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, Phys. Rev. D **106,** 042006 (2022).

[37] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, Classical Quantum Gravity **36,** 055011 (2019).

[38] L. C. Stein, J. Open Source Software **4,** 1683 (2019).

[39] V. Varma, D. Gerosa, L. C. Stein, F. Hébert, and H. Zhang, Phys. Rev. Lett. **122,** 011101 (2019).

[40] C. Cutler and E. E. Flanagan, Phys. Rev. D **49,** 2658 (1994).

[41] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, arXiv:1807.07062.

[42] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, Astrophys. J. Lett. **874,** L2 (2019).

[43] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, SoftwareX **12,** 100620 (2020).

[44] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Classical Quantum Gravity **21,** S1809 (2004).

[45] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban, SoftwareX **13,** 100657 (2021).

[46] T. Accadia et al., Classical Quantum Gravity **27,** 194011 (2010).

[47] A. Longo, S. Bianchi, W. Plastino, N. Arnaud, A. Chiummo, I. Fiori, B. Swinkels, and M. Was, Classical Quantum Gravity **37,** 145011 (2020).

[48] A. Longo, S. Bianchi, G. Valdes, N. Arnaud, and W. Plastino, Classical Quantum Gravity **39,** 035001 (2022).

[49] F. Acernese et al. (Virgo Collaboration), arXiv:2205.01555.

[50] S. Soni et al. (LIGO Collaboration), Classical Quantum Gravity **38,** 025016 (2020).

[51] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93,** 044006 (2016).

[52] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93,** 044007 (2016).

[53] E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, https://zenodo.org/record/7259655# .Y1sPY-zMKAk.

[54] B. P. Abbott et al. (LIGO Scientific, Virgo Collaborations), Astrophys. J. Lett. **882,** L24 (2019).

[55] R. Abbott et al. (LIGO Scientific, Virgo Collaborations), Astrophys. J. Lett. **913,** L7 (2021).

[56] L. Sun et al., Classical Quantum Gravity **37,** 225008 (2020).

[57] B. Abbott et al. (LIGO Scientific, Virgo Collaborations), Phys. Rev. D **100,** 104036 (2019).

[58] R. Abbott et al. (LIGO Scientific, Virgo Collaborations), Phys. Rev. D **103,** 122002 (2021).

[59] R. Abbott et al. (LIGO Scientific, Virgo, KAGRA Collaborations), arXiv:2112.06861.

[60] https://www.gw-openscience.org.

[61] https://alliancecan.ca.

[62] J. D. Hunter, Comput. Sci. Eng. **9,** 90 (2007).

[63] C. R. Harris et al., Nature (London) **585,** 357 (2020).

[64] The Pandas Development Team, pandas-dev/pandas: PANDAS, 10.5281/zenodo.3509134 (2020).

[65] W. McKinney, Data structures for statistical computing in PYTHON, *Proceedings of the 9th* PYTHON *in science conference* (2010), pp. 56–61, 10.25080/Majora-92bf1922-00a

[66] P. Virtanen et al. (SciPy 1.0 Contributors), Nat. Methods **17,** 261 (2020).

[67] LIGO Scientific and Virgo Collaborations, BILBY, https:// git.ligo.org/lscsoft/bilby (2018).

[68] LIGO Scientific, Virgo Collaborations, LALSuite, https:// git.ligo.org/lscsoft/lalsuite (2018).

[69] LIGO Scientific and Virgo Collaborations, BayesWave, https://git.ligo.org/lscsoft/bayeswave (2018).

[70] R. Abbott et al. (LIGO Scientific, Virgo Collaborations), SoftwareX **13,** 100658 (2021).

[71] LIGO Scientific and Virgo Collaborations, GCN **26926** (2020), https://gcn.gsfc.nasa.gov/gcn3/26926.gcn3.

[72] C. Messick et al., Phys. Rev. D **95,** 042001 (2017).

[73] C. Hanna et al., Phys. Rev. D **101,** 022003 (2020).

[74] S. Klimenko et al., Phys. Rev. D **93,** 042004 (2016).

[75] A. H. Nitz, T. Dal Canton, D. Davis, and S. Reyes, Phys. Rev. D **98,** 024050 (2018).

[76] T. Dal Canton, A. H. Nitz, B. Gadre, G. S. Cabourn Davies, V. Villa-Ortega, T. Dent, I. Harry, and L. Xiao, Astrophys. J. **923,** 254 (2021).

[77] T. Adams, D. Buskulic, V. Germain, G. M. Guidi, F. Marion, M. Montani, B. Mours, F. Piergiovanni, and G. Wang, Classical Quantum Gravity **33,** 175012 (2016).

[78] Q. Chu et al., Phys. Rev. D **105,** 024023 (2022).

[79] T. Accadia *et al.*, Classical Quantum Gravity **27**, 194011 (2010).

[80] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Classical Quantum Gravity **33**, 134001 (2016).

[81] D. Davis *et al.*, Data quality vetoes applied to the analysis of LIGO data from the third observing run, Technical Report No. DCC-T2100045, LIGO, 2021.

[82] D. Davis *et al.* (LIGO Collaboration), Classical Quantum Gravity **38**, 135014 (2021).

[83] A. H. Nitz, T. Dent, T. D. Canton, S. Fairhurst, and D. A. Brown, Astrophys. J. **849**, 118 (2017).

[84] G. S. Davies, T. Dent, M. Tápai, I. Harry, C. McIsaac, and A. H. Nitz, Phys. Rev. D **102**, 022004 (2020).

[85] F. Aubin *et al.*, Classical Quantum Gravity **38**, 095004 (2021).

[86] S. Sachdev *et al.*, arXiv:1901.08580.

[87] K. Cannon *et al.*, arXiv:2010.05082.

[88] B. Allen, W.-s. Hua, and A. C. Ottewill, arXiv:gr-qc/9909083.

[89] E. Poisson and C. M. Will, Phys. Rev. D **52**, 848 (1995).

[90] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman, Phys. Rev. Lett. **74**, 3515 (1995).

[91] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[92] E. Racine, Phys. Rev. D **78**, 044021 (2008).

[93] L. Santamaria *et al.*, Phys. Rev. D **82**, 064016 (2010).

[94] P. Ajith *et al.*, Phys. Rev. Lett. **106**, 241101 (2011).

[95] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015).

[96] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Mon. Not. R. Astron. Soc. **498**, 4492 (2020).

[97] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).

[98] M. Boyle *et al.*, Classical Quantum Gravity **36**, 195006 (2019).

[99] S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch, Phys. Rev. D **85**, 064034 (2012).

[100] E. Payne, C. Talbot, P. D. Lasky, E. Thrane, and J. S. Kissel, Phys. Rev. D **102**, 122004 (2020).

[101] S. Vitale, C.-J. Haster, L. Sun, B. Farr, E. Goetz, J. Kissel, and C. Cahillane, Phys. Rev. D **103**, 063016 (2021).

[102] R. Essick, Phys. Rev. D **105**, 082002 (2022).

[103] M. Wijngaarden, K. Chatziioannou, A. Bauswein, J. A. Clark, and N. J. Cornish, Phys. Rev. D **105**, 104019 (2022).

[104] Q. Hu and J. Veitch, Phys. Rev. D **106**, 044042 (2022).