GW240925 and GW250207: Astrophysical Calibration of Gravitational-wave Detectors Supplemental Material

(The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration)

(Dated: May 7, 2026)

Supplemental materials to the discovery of GW240925_005809 (GW240925) and GW250207_115645 (GW250207).

CALIBRATION TECHNIQUES AND MONITORING

Since the detector arm length L is known to high precision, to reconstruct the strain we must accurately measure the differential arm length change $\Delta L_{\rm free}$. However, as $\Delta L_{\rm free}$ is suppressed by feedback control, the detectors only measure the residual differential arm displacement $\Delta L_{\rm res}$. In Laser Interferometer Gravitational-Wave Observatory (LIGO), this is converted to a digital error signal $d_{\rm err}$ via the sensing transfer function C, such that $d_{\rm err} = C\Delta L_{\rm res}$ (all quantities are denoted in frequency domain) [1–4]. The error signal is processed by a set of digital filters D to generate the digital control signal $d_{\rm ctrl} = Dd_{\rm err}$, which is sent through the actuation transfer function A to produce the analog control displacement $\Delta L_{\rm ctrl} = Ad_{\rm ctrl}$. The total free differential arm length change is given by $\Delta L_{\rm free} = \Delta L_{\rm res} + \Delta L_{\rm ctrl}$. Defining the detector response function as [5]

$$R = \frac{1}{C} + AD, \tag{1}$$

the strain is reconstructed as [6]

$$d = \frac{R\, d_{\rm err}}{L}. \tag{2}$$

KAGRA adopts an approach closely aligned with that of LIGO [7], and Virgo follows similar procedures, with minor differences in conventions and technical implementations [8, 9].

The timing for the detectors' real-time control systems and recorded data, provided by Global Positioning System (GPS) receivers located at the corner and end stations of each detector, serves as a critical absolute reference. Cross-checks between multiple GPS clocks and a local atomic clock estimate the timing accuracy to be better than ± 1 µs [10, 11].

Calibration errors in the modeled response function $R^{(\rm model)}(f;t)$ may result from imperfect estimates of model parameters, uncompensated time dependencies, or missing features present in the true detector response. The probability distribution of the correction factor $\eta(f;t)$ is evaluated hourly throughout the observing run by incorporating both in-situ regular measurements taken when the detector is not observing (typically once or twice per week) and real-time measurements within a one-hour window [11]. This distribution is constructed numerically using 10^3 realizations of $R(f;t)$ [2, 4]. The resulting uncertainty envelope is used to construct the calibration prior employed in signal parameter estimation for all candidates whose trigger time lies closest to the given hour $t_{\rm near}$ [12].

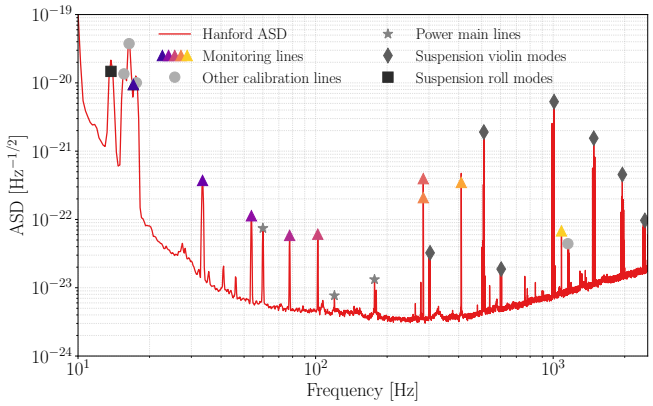

FIG. 1. ASD of the LIGO Hanford detector one day before the arrival of GW240925. Colored triangles indicate the calibration monitoring lines used for real-time tracking of the detector response. Dots, stars, diamonds and squares mark the additional calibration tracking lines for time-dependent correction factors, power mains, suspension violin harmonics and suspension roll modes, respectively.

In addition to the procedures adopted during the early observing runs [1–3], a real-time monitoring system was introduced in Virgo during the third observing run (O3) [8] and in LIGO during fourth observing run (O4) [11] to improve the evaluation of $\eta(f;t)$ in the low-latency calibrated data. This system operates by continuously injecting monochromatic calibration lines at discrete frequencies (nine lines for LIGO) via the photon calibrator [13–16]. These lines enable direct measurement of the detector response at their respective frequencies. Figure 1 shows the locations and relative strengths of these lines in the amplitude spectral density (ASD) of Hanford detector data. Two closely spaced lines at 283.91 Hz and 284.01 Hz are injected into the X- and Y-end test masses, respectively, to monitor the photon-calibrator absolute references [11]. For Virgo, absolute calibration was further improved through the deployment of a Newtonian calibrator, which provides a precise reference at low frequencies [17]. The detector response to these injected lines is continuously measured in real time, allowing direct inference of $\eta(f;t)$ at those frequencies.

Figures 2 and 3 show the real-time calibration monitoring measurements for the Hanford detector response function around the times of GW240925 and GW250207, respectively [11]. These figures show the estimated amplitude and phase errors derived from the monitoring lines at their respective frequencies. These diagnostics help assess the calibration accuracy during the periods of interest.

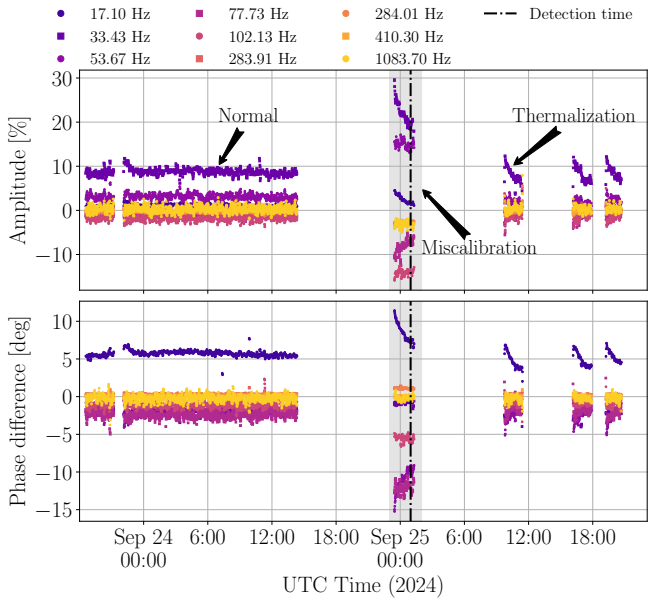

FIG. 2. Real-time calibration monitoring for Hanford around the time of GW240925 (spanning ~ 2 days). The top and bottom panels show the estimated amplitude and phase of the calibration error, $\delta\mathcal{A}$ and $\delta\phi$, respectively, expressed through the complex correction factor η. The shaded region highlights the miscalibrated interval.

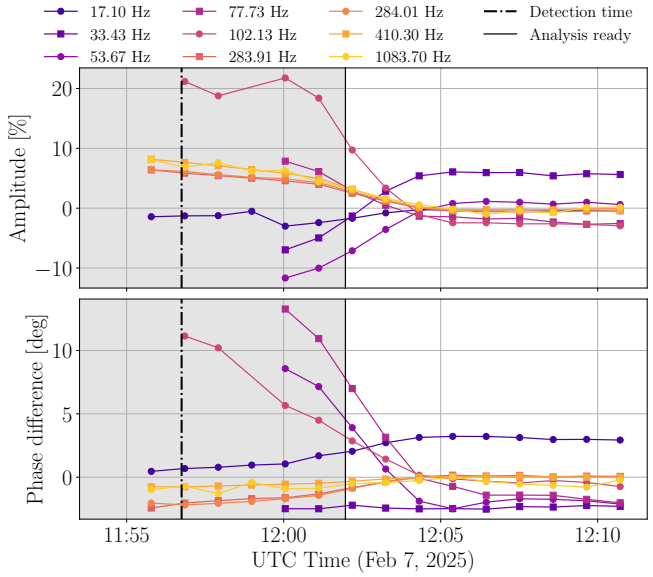

FIG. 3. Similar to Fig. 2, real-time calibration monitoring for Hanford around the time of GW250207 (zoomed in to the detection and observing starting times). The shaded region marks the interval during which the detector had not yet reached analysis-ready observing status.

For GW240925, Fig. 2 spans ~ 2 days to illustrate three representative periods of detector behavior: (i) a normal period on the day before the signal, during which the measured calibration amplitude and phase errors are within $|\delta\mathcal{A}| \lesssim 10\%$ and $|\delta\phi| \lesssim 6$ deg; (ii) a miscalibrated interval, where the measurements deviate significantly from $\delta\mathcal{A} = 0$ and $\delta\phi = 0$ deg, indicating a known calibration issue, and (iii) several short-duration lock stretches lasting $\lesssim 2$ h, where low-frequency evolution due to *thermalization* effects (time-varying lensing by the test masses due to differential heating by the laser beam profile) is evident, e.g., in the amplitude at 33.43 Hz and in the phase at 17.10 Hz.

For GW250207, Fig. 3 is zoomed in around the detection. The shaded region marks the interval when the detector had not yet reached observing status and was undergoing thermalization, as indicated by larger deviations from $\delta\mathcal{A} = 0$ and $\delta\phi = 0$ deg at various frequencies. The calibration is best modeled when the interferometer is fully thermalized and operating in a steady state. Several monitoring lines were also not yet recording measurements during this period.

With real-time monitoring and hourly evaluation of systematic errors, similarly elevated calibration errors to those seen around these two detections are not observed in other observing-mode stretches of O4 [18].

EVENT VALIDATION

Detector data can be impacted by noise of instrumental or environmental origin [19–21]. Noise artifacts can manifest in several ways, such as shorter-duration, broadband non-Gaussian features known as *glitches* [22, 23], and narrowband features known as *spectral lines* [24]. Data quality (DQ) vetting around candidate events involves multiple checks [25], such as event validation, the DQ Report (DQR) [26, 27], and (when needed) glitch subtraction [28–30].

GW240925 was detected while Livingston was observing with binary neutron star inspiral range (a conventional measure of detector sensitivity [31–33]) of ~ 168 Mpc, Virgo was observing with range ~ 51 Mpc, and Hanford was observing with an estimated range of ~ 147 Mpc, neglecting its calibration uncertainty. During event validation, some components of the DQR were flagged for investigation. At both Hanford and Livingston, excess noise was identified in an auxiliary channel that monitors environmental noise [34]; however, this noise occurred at sufficiently high frequencies to be well separated from the gravitational-wave (GW) signal band, and we found no evidence that it coupled into the GW strain during the times of interest. In addition, a glitch was identified in Livingston data 1.4–1.7 s after the signal in the frequency range 25–50 Hz. The glitch occurred sufficiently far from the signal that it does not impact the inference of source properties [35]. Thus, glitch mitigation was deemed unnecessary. Overall, no DQ issues impacting the observation were found for any of the detectors.

Approximately 80 s of Hanford strain data preceding the time of GW240925 were initially marked as non-observing, as the interferometer was briefly taken out of observing mode to make a digital configuration change to a control loop that is a part of the arm-length stabilization system. This stabilization

system is inactive when the detector is in a low-noise state and only used during lock acquisition. The configuration-change procedure had no impact on the strain data, which remain valid for analysis during that interval [36].

GW250207 was detected while Livingston and Virgo were observing with binary neutron star inspiral ranges of $\sim$ 168 Mpc and $\sim$ 51 Mpc, respectively; Hanford formally entered observing mode with a range of $\sim$ 149 Mpc approximately five minutes later. During event validation, one DQR task that monitors auxiliary channels for excess noise overlapping the signal [37] was flagged for investigation. As with GW240925, no DQ issues were identified for any of the detectors, as no excess noise was found in the strain data during the time of interest.

ADDITIONAL SEARCH-ANALYSIS RESULTS

Multiple search pipelines are used to search for compact binary coalescence (CBC) signals [12], for both low-latency, online analyses and higher-latency, offline analyses. Online searches prioritize rapid identification of candidates for prompt alerts to the astronomical community [38, 39], whereas offline searches are performed later to provide more comprehensive results. The offline searches can analyze recalibrated data, incorporate updated DQ information, and use more computationally expensive algorithms [12]. While online searches evaluate candidate significance using background data collected only prior to the candidate, offline searches can use data from both before and after the candidate time. As a result, marginal candidates can vary in significance between online and offline analyses, while highly significant candidates generally remain so [40]. High-signal-to-noise ratio (SNR) candidates are typically recovered as significant across all pipelines [12, 39].

GW240925 was identified in low-latency by cWB [41–44], GstLAL [45–53], MBTA [54–56] and SPIIR [57]. The PyCBC search pipeline [58] did not report the candidate in low latency because there were insufficient contiguous observing-mode data to identify a Hanford trigger, and it only considers single-detector candidates with potential electromagnetic counterparts (a duration > 7 s) [12, 58]. Nevertheless, PyCBC did recover a trigger with consistent SNR and timestamp in Livingston.

GW250207 was identified in low latency by GstLAL and SPIIR. The other low-latency search pipelines did not report the candidate because they did not account for Virgo triggers when constructing multi-detector coincidences, and they were configured to either require at least two-detector coincidence (for cWB [43, 44]) or only report single-detector candidates likely to accompany electromagnetic emissions [12] (e.g., a detector-frame chirp mass below $7M_\odot$ for MBTA [56]). Nevertheless, both MBTA and PyCBC recovered a trigger with consistent SNR and timestamp in Livingston.

Full offline search results for second part of O4 (O4b) will be presented in GWTC-5.0, with results for third part of O4 (O4c) to follow in GWTC-6.0 [33]. In Table I, we present the currently available offline results for GW240925, analyzed with the GstLAL [45–49, 51–53], MBTA [54–56] and PyCBC [59–61] matched-filter pipelines using the low-latency, online calibrated (C00) data. To assess the potential impact of miscalibration on the search results, we also analyzed the offline calibrated data (C01) data with the MBTA and PyCBC pipelines, and found that the SNR and 90% credible area of the sky localization differed by only a few percent relative to the C00 results. This suggests that the Hanford miscalibration did not significantly affect the search results for GW240925, which is consistent with expectations [6, 62, 63].

TABLE I. Offline search results for GW240925. The table lists the SNR in the LIGO Hanford and Livingston detectors and the FAR reported by each pipeline. The FARs are capped at $< 1 \times 10^{-5}$ yr^{-1} to ensure a consistent limit across pipelines.

Pipeline	$\rho_{\rm H}$	$\rho_{\rm L}$	FAR/yr^{-1}
GstLAL	16.8	25.5	$< 1 \times 10^{-5}$
MBTA	17.3	25.8	$< 1 \times 10^{-5}$
PyCBC	17.3	25.7	3.95×10^{-5}

PARAMETER-ESTIMATION METHODS AND RESULTS

To characterize each signal, we use the Dynesty nested sampler [64], as implemented in the Bilby Bayesian inference library [65, 66], to obtain samples from the posterior probability distributions of both the source and calibration parameters. Changing these parameters alters the waveform: for example, varying the chirp mass changes the frequency evolution of the inspiral [67–69]. Hence, we calculate the posterior probability distribution for the input parameters by assessing how well the corresponding waveform matches the data [70, 71].

Both signals were observed in the Hanford, Livingston and Virgo detectors, and we coherently analyze data from all three [70] (except when investigating the potential to constrain calibration with Hanford data alone). For GW240925, we analyze 32 s of data employing a sampling frequency of 4096 Hz, with a frequency range from 20 Hz to 1792 Hz. For GW250207, we analyze 8 s of data employing a sampling frequency of 1024 Hz, with a frequency range of 20 Hz to 448 Hz. For both signals, estimates of the detectors' power spectral densities (PSDs) are obtained using BayesWave [72–74].

The strain data consist of the GW signal and detector noise, $d^{(\rm corr)} = h(\boldsymbol{\theta}) + n^{(\rm corr)}$, where $h(\boldsymbol{\theta})$ is the waveform corresponding to parameters $\boldsymbol{\theta}$, and $n^{(\rm corr)}$ is the residual noise after applying the calibration correction. To account for calibration uncertainties in the inference, we apply the calibration correction factor η to the waveform model rather than the detector data:

$$d = \frac{d^{(\rm corr)}}{\eta} = \frac{1}{\eta}\left(h + n^{(\rm corr)}\right) = \frac{h}{\eta} + n, \tag{3}$$

where $n = n^{(\mathrm{corr})}/\eta$ is the effective noise without applying the calibration correction. Since the likelihood is a function of d, rather than $d^{(\mathrm{corr})}$, and the PSD characterizes n, not $n^{(\mathrm{corr})}$, applying the calibration correction to the waveform is the more convenient approach [12].

We parametrize the calibration uncertainty in the inference as

$$\frac{1}{\eta(f;t_{\mathrm{near}})} = [1+\delta\mathcal{A}'(f)]\exp[i\delta\phi'(f)], \tag{4}$$

where $\delta\mathcal{A}'(f)$ and $\delta\phi'(f)$ are the amplitude and phase deviation parameters, respectively [75]. The frequency-dependent amplitude and phase corrections are modeled using cubic splines [12, 70, 76]. The deviation parameters are related to the calibration errors defined in Eq. (1) of the main paper as

$$\delta\phi'(f) = -\delta\phi(f;t_{\mathrm{near}}); \quad 1+\delta\mathcal{A}'(f) = \frac{1}{1+\delta\mathcal{A}(f;t_{\mathrm{near}})}. \tag{5}$$

The in-situ measured calibration uncertainty estimates provide the distribution of $\{[1+\delta\mathcal{A}(f;t_{\mathrm{near}})], \delta\phi(f;t_{\mathrm{near}})\}$ at discrete frequencies. These are converted to $\{[1+\delta\mathcal{A}'(f)], \delta\phi'(f)\}$ and interpolated to a set of 10 frequency nodes with log-uniform spacing over the analysis frequency range for each detector. Gaussian priors are then set on $\delta\mathcal{A}'(f)$ and $\delta\phi'(f)$ at each node frequency, with medians and standard deviations determined from the in-situ measurements.

For the analysis of GW240925, we use three different inspiral–merger–ringdown (IMR) waveform models, all of which include the effects of spin precession and higher-order multipole moments, while neglecting orbital eccentricity: the frequency-domain IMRPhenomXPHM [77, 78] and IMRPhenomXPNR [79, 80], and the time-domain SEOBNRv5PHM [81, 82], which additionally includes equatorial asymmetric contributions to multipole moments. For the analysis of GW250207, we use the same three models as for GW240925, and additionally include NRSur7dq4 [83], a time-domain numerical-relativity surrogate model. This model cannot be used for GW240925 due to constraints on the mass range over which it can generate waveforms down to 20 Hz. We also perform an analysis with the frequency-domain IMRPhenomXAS waveform [84], which assumes aligned spins, thereby excluding orbital-plane precession effects, and includes only the contribution from the dominant $(2,2)$ multipole. This waveform is included for comparison with results from the search pipelines, which use waveforms incorporating similar physics. Our overall parameter estimates for both signals are obtained by combining results from a set of the different waveform models that include the effects of spin precession with equal weight (excluding IMRPhenomXAS) [12, 70]; since IMRPhenomXPNR supersedes IMRPhenomXPHM, the latter is excluded from the combination to avoid double-counting the results obtained with this waveform family.

For both signals, we adopt default priors for the source parameters: a uniform prior on detector-frame component masses, with bounds on the detector-frame chirp mass wide enough to avoid truncating the posterior; a uniform prior on spin magnitudes with isotropic spin directions; uniform priors on the reference phase and time; a uniform distribution on the polarization angle; a uniform distribution over the sphere for binary inclination; a uniform distribution over the sky for sky position, and a distance prior corresponding to a uniform distribution in comoving time and volume [12, 66, 85]. For analyses using IMRPhenomXAS, which neglect spin precession, we adopt a prior on the spin magnitudes that gives the same distribution for spin components aligned with the orbital angular momentum as the prior used in the other analyses. For analyses of Hanford data where we assume a wide calibration prior, we do not use frequency-dependent in-situ measurements to set uncertainties for $\delta\mathcal{A}(f;t_{\mathrm{near}})$ and $\delta\phi(f;t_{\mathrm{near}})$, but instead assume zero-mean Gaussian uncertainties with frequency-independent standard deviations of 20% and 20 deg, respectively. These calibration uncertainty priors are sufficiently broad to encompass the miscalibrations present at the times of the two signals, as illustrated in Fig. 2 and Fig. 3.

Results from the various analyses of the two signals are given in Table II, with the inferred component masses shown in Fig. 4. In addition to the analyses described in the main text, we also provide results for GW240925 assuming perfect calibration in all three detectors, to assess potential biases arising from neglecting calibration uncertainties. In Fig. 5, we show the inferred spin parameters for GW250207, where the inclusion of calibration uncertainty leads to a difference for χ_{p} but not χ_{eff}, illustrating how some signal properties and more sensitive to the calibration than others.

In contrast to the significant differences found for GW250207 discussed in the main text and shown in Fig. 5, for GW240925, neglecting calibration uncertainty leads to small shifts in most intrinsic parameters. For example, the detector-frame chirp mass is $(1+z)\mathcal{M} = 7.368^{+0.011}_{-0.009} M_\odot$ without incorporating calibration uncertainty, compared to $7.371^{+0.012}_{-0.010} M_\odot$ when using the wide calibration prior. The luminosity distance is inferred as $D_{\mathrm{L}} = 369^{+54}_{-156}$ Mpc when neglecting calibration uncertainty, shifted from 348^{+69}_{-160} Mpc with the wide calibration prior. The most significant impact is seen in the sky area [69], which increases from 22 deg^2 without incorporating calibration uncertainty to 60 deg^2 with the wide calibration prior, as illustrated in Fig. 6.

The source-mass distributions for GW240925 exhibit bimodality, as visible in the two-dimensional distributions of Fig. 4. The detector-frame masses have a unimodal distribution, but the distance distribution has two modes, which impacts the inferred source masses [33, 88]. The distance distribution has one mode at larger values (favored by the prior) corresponding to face-on inclinations, and another at smaller values corresponding to edge-on inclinations [89–91]. For edge-on inclinations, the signal is dominated by a single polarization component [67, 92, 93]. While the two LIGO detectors are approximately aligned, and so have similar sensitivities to signal polarization [33, 94, 95], Virgo has a different sensitivity [96, 97]. Hence with edge-on inclinations,

TABLE II. Inferred source properties for GW240925 and GW250207. We report median values with 90% symmetric credible intervals from various analyses for various parameters [33], and the 90% credible area for the sky location. Four GW240925 analyses are presented: using the initial calibration of the data and assuming all three detectors are perfectly calibrated (C00 no uncertainty); assuming a wide, uninformative prior for Hanford calibration and the in-situ measured uncertainty envelopes for the other detectors (C00 wide); assuming the in-situ measured uncertainty envelopes for all detectors (C00 envelope), and using recalibrated data assuming the in-situ measured uncertainty envelopes for all detectors (C01 envelope). Two GW250207 analyses are presented: assuming all three detectors are perfectly calibrated (C00 no uncertainty), and assuming a wide, uninformative prior for Hanford calibration and the in-situ measured uncertainty envelopes for the other detectors (C00 wide). The analyses that neglect calibration uncertainty are expected to yield biased results. Parameters that evolve throughout the inspiral are quoted at a reference frequency of 20 Hz. All results are computed assuming a standard cosmology with $H_0 = 67.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ [12, 86].

	GW240925				GW250207	
Parameter	C00 no uncertainty	C00 wide	C00 envelope	C01 envelope	C00 no uncertainty	C00 wide
Primary mass $m_1/M_\odot$	$9.0^{+1.8}_{-1.0}$	$9.0^{+1.8}_{-1.0}$	$9.0^{+1.7}_{-1.0}$	$9.0^{+2.0}_{-1.0}$	$35.4^{+1.4}_{-1.4}$	$35.2^{+1.7}_{-1.7}$
Secondary mass $m_2/M_\odot$	$6.9^{+0.9}_{-1.1}$	$7.0^{+0.8}_{-1.2}$	$7.0^{+0.8}_{-1.2}$	$7.0^{+0.8}_{-1.3}$	$30.5^{+1.3}_{-1.2}$	$30.6^{+1.5}_{-1.8}$
Total mass $M/M_\odot$	$16.0^{+0.7}_{-0.4}$	$16.1^{+0.6}_{-0.5}$	$16.1^{+0.6}_{-0.4}$	$16.1^{+0.7}_{-0.4}$	$66.0^{+0.8}_{-1.1}$	$65.9^{+1.0}_{-1.7}$
Chirp mass $\mathcal{M}/M_\odot$	$6.83^{+0.21}_{-0.07}$	$6.86^{+0.22}_{-0.09}$	$6.85^{+0.22}_{-0.08}$	$6.85^{+0.22}_{-0.08}$	$28.61^{+0.35}_{-0.46}$	$28.57^{+0.44}_{-0.77}$
Final mass $M_\mathrm{f}/M_\odot$	$15.2^{+0.7}_{-0.4}$	$15.3^{+0.7}_{-0.5}$	$15.3^{+0.6}_{-0.4}$	$15.3^{+0.7}_{-0.4}$	$62.8^{+0.7}_{-1.0}$	$62.7^{+1.0}_{-1.6}$
Effective inspiral spin χ_eff	$0.02^{+0.07}_{-0.02}$	$0.03^{+0.07}_{-0.02}$	$0.02^{+0.07}_{-0.02}$	$0.02^{+0.08}_{-0.02}$	$-0.00^{+0.03}_{-0.04}$	$0.00^{+0.03}_{-0.04}$
Effective precession spin χ_p	$0.22^{+0.31}_{-0.18}$	$0.26^{+0.37}_{-0.21}$	$0.25^{+0.39}_{-0.20}$	$0.25^{+0.37}_{-0.20}$	$0.13^{+0.17}_{-0.07}$	$0.06^{+0.15}_{-0.05}$
Final spin χ_f	$0.69^{+0.01}_{-0.03}$	$0.69^{+0.02}_{-0.03}$	$0.69^{+0.02}_{-0.02}$	$0.69^{+0.02}_{-0.03}$	$0.68^{+0.01}_{-0.01}$	$0.69^{+0.01}_{-0.01}$
Luminosity distance D_L/Mpc	369^{+54}_{-156}	348^{+69}_{-160}	355^{+63}_{-165}	356^{+61}_{-162}	175^{+42}_{-28}	187^{+121}_{-52}
Redshift z	$0.08^{+0.01}_{-0.03}$	$0.07^{+0.01}_{-0.03}$	$0.08^{+0.01}_{-0.03}$	$0.08^{+0.01}_{-0.03}$	$0.04^{+0.01}_{-0.01}$	$0.04^{+0.03}_{-0.01}$
Network SNR ρ	$31.34^{+0.10}_{-0.14}$	$31.44^{+0.16}_{-0.20}$	$31.50^{+0.11}_{-0.16}$	$31.96^{+0.11}_{-0.15}$	$68.66^{+0.05}_{-0.08}$	$68.91^{+0.08}_{-0.11}$
Hanford SNR ρ_H	$17.58^{+0.09}_{-0.14}$	$17.75^{+0.18}_{-0.25}$	$17.82^{+0.09}_{-0.13}$	$17.95^{+0.08}_{-0.13}$	$48.55^{+0.10}_{-0.11}$	$48.81^{+0.08}_{-0.11}$
Livingston SNR ρ_L	$25.98^{+0.11}_{-0.13}$	$25.99^{+0.12}_{-0.14}$	$25.99^{+0.11}_{-0.13}$	$26.46^{+0.11}_{-0.13}$	$47.90^{+0.08}_{-0.10}$	$48.00^{+0.05}_{-0.07}$
Virgo SNR ρ_V	$1.9^{+0.1}_{-0.5}$	$1.9^{+0.2}_{-0.8}$	$1.9^{+0.2}_{-0.8}$	$1.9^{+0.2}_{-0.7}$	$8.2^{+0.2}_{-0.4}$	$8.1^{+0.3}_{-0.6}$
Sky area Ω/deg^2	22	60	29	25	5	20

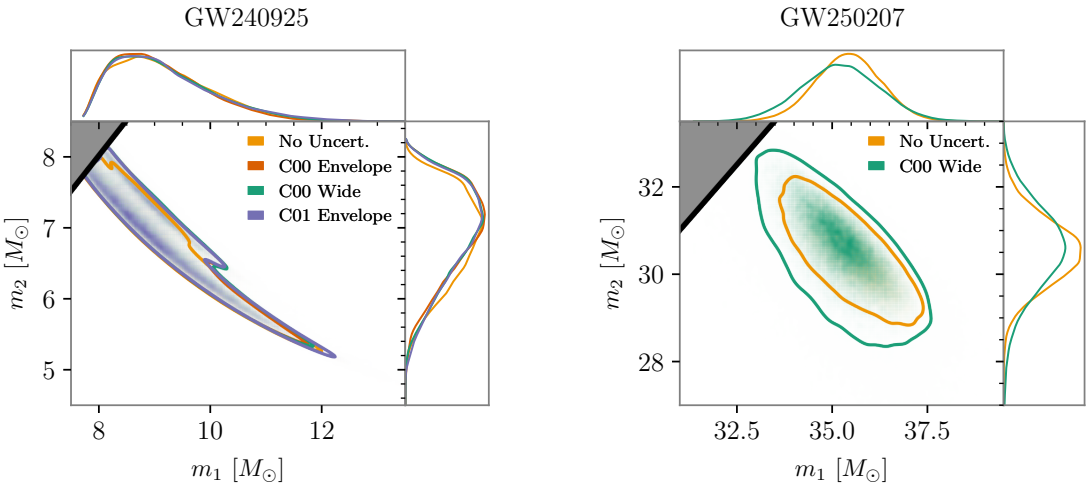

FIG. 4. Inferred component masses for GW240925 (left) and GW250207 (right). Results are shown for analyses that neglect calibration uncertainty (yellow) and that adopt a wide prior on the Hanford calibration (green), for both signals. For GW240925, we additionally show results using the in-situ measured calibration priors applied to the Hanford miscalibrated C00 (orange) and recalibrated C01 (purple) data. The component masses follow the convention that $m_1 \geq m_2$. Two-dimensional plots show the joint posterior probability densities with 90% credible contours; one-dimensional plots show the corresponding marginalized posteriors.

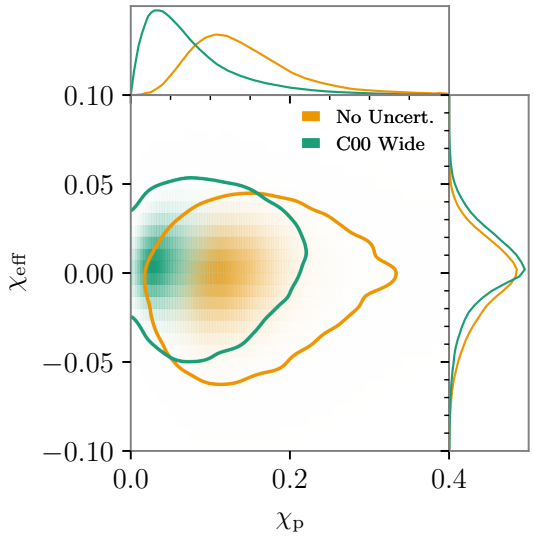

FIG. 5. Inferred effective inspiral spin $\chi_{\rm eff}$ and effective precession spin $\chi_{\rm p}$ for GW250207 under two different calibration prior assumptions. Results are shown for analyses that neglect calibration uncertainty (yellow) and that adopt a wide prior on the Hanford calibration (green). Two-dimensional plots show the joint posterior probability densities with 90% credible contours; one-dimensional plots show the corresponding marginalized posteriors. While the inferred $\chi_{\rm eff}$ is similar in both cases, neglecting to account for calibration uncertainty leads to a shift in the $\chi_{\rm p}$ posterior that erroneously ascribes the effect of miscalibration to spin precession in the signal.

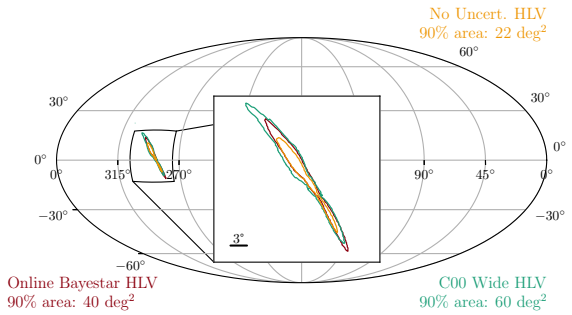

FIG. 6. Sky localization for GW240925 from the low-latency BAYESTAR analysis (which does not account for calibration uncertainty) [87], and Bilby analyses including and excluding calibration uncertainty. All analyses use data from Hanford, Livingston and Virgo. While including calibration uncertainty broadens the localization, failing to marginalize over calibration uncertainty may lead to biased results in the presence of significant miscalibration.

the polarization can be adjusted to an alignment that Virgo is less sensitive to, and this accounts for the disparate SNRs measured in the Virgo and LIGO detectors (the inferred Virgo SNR is $\rho_{\rm V} = 1.9^{+0.2}_{-0.7}$ versus $\rho_{\rm L} = 26.46^{+0.11}_{-0.13}$ in Livingston). Similar to the sky localization of GW170817 [98], this illustrates how multi-detector observations can constrain source properties even when SNR is low in one detector.

We find that systematic differences from the choice of waveform model are small for both signals, as typical Jensen–Shannon divergences (JSDs) between the intrinsic parameter posteriors obtained with different waveforms are $O(0.01)$ nat. The most significant variations appear in parameters that are defined differently between the waveform families, like azimuthal angles and coalescence time. The differences in the posteriors for the calibration parameters are generally smaller than those found for the intrinsic binary parameters; however, larger differences between the calibration-parameter posteriors obtained with the various waveforms are found when using the wide priors than when using the more constraining in-situ envelope priors. Given the properties of the sources, it is expected that the different waveform models should give consistent results [40].

We expect constraints on the calibration to improve with increasing SNR. Low-SNR signals will have posteriors dominated by the prior, while the likelihood becomes more dominant at high SNRs [99, 100]. From Fig. 2 of the main paper, we can see that constraints from GW250207 are often tighter than for GW240925, but that depends on the frequency. Since the frequency contents of the signals differ, they provide different amounts of information about the calibration parameters as a function of frequency. If we examine the constraints around physically comparable frequencies corresponding to the minimum-energy circular orbit $f_{22}^{\rm MECO}$ for the two signals, for GW240925 we infer $\delta\mathcal{A} = -7^{+27}_{-18}\%$ and $\delta\phi = -4^{+19}_{-19}$ deg (90% credible interval) at the spline node at 243.0 Hz, while for GW250207 we infer $\delta\mathcal{A} = -16^{+10}_{-9}\%$ and $\delta\phi = 7^{+8}_{-8}$ deg at the spline node at 56.4 Hz. These constraints are tighter for GW250207. In general, the measurement precision will depend upon both the SNR in the relevant detector, and the network SNR, which governs how well source properties can be determined [5, 67, 101–103].

To assess potential correlations between the inferred source and calibration parameters, we calculate the Pearson correlation coefficient (PCC) [104] between each source parameter and the frequency-dependent amplitude and phase calibration parameters. We find no significant correlations for GW240925 (PCC < 0.3). For the coherent, multi-detector analyses of GW250207, we find that the calibration phase corrections are correlated with both the sky position and polarization, while the calibration amplitude corrections are correlated with the inclination and luminosity distance, with PCC > 0.4 in both cases. For the single-detector, Hanford-only analysis, the calibration amplitude corrections show correlations with the inferred chirp mass and reference time, since the luminosity distance posterior is more prior driven for the single-detector analysis.

We also investigate the cause of the inconsistency between the Livingston–Virgo and three-detector sky localizations. The effect of including Hanford data is illustrated in Fig. 7, which shows the posteriors on the sky location and time delays between detector pairs for both the two- and three-detector analyses for each signal. For GW240925, the two-detector sky localization has multiple modes, each cor-

responding to a different Livingston–Virgo time delay. The low Virgo SNR means that there is posterior support for time delays corresponding to offsets of approximately a GW cycle [105]. Adding Hanford data adds significant extra information to the analysis, resulting in a change in the localization [106, 107]. The three-detector localization is primarily constrained by the Hanford–Livingston time delay, and the three-detector analysis selects one of the secondary modes for the Livingston–Virgo time delay from the two-detector analysis as most probable. The two-detector localization does have some posterior support at the sky location identified in the three-detector analysis, but this is outside of the 90% credible area. For GW250207, the two-detector localization exhibits two separate modes, although again only the primary mode falls within the 90% credible area. The secondary mode corresponds to a much closer inferred distance than the primary mode and is therefore downweighted by the prior in the two-detector analysis. However, adding Hanford data shifts the localization around the ring of constant Livingston–Virgo time delay, to a point consistent with the observed Hanford time; the additional constraint on the timing overcomes the prior suppression of the secondary mode, which then becomes the preferred source location. The precise sky localization obtained with the inclusion of Hanford data for both signals was enabled by the simultaneous inference of the calibration parameters in the parameter-estimation analyses.

Another source of the discrepancy between the low-latency BAYESTAR localization and the offline Bilby results arises from differences in the physical effects included in the waveform models underlying each analysis. The search pipelines that produce the matched-filter SNR time series used by BAYESTAR include only aligned-spin effects and GW emission from the dominant $(2, 2)$ multipole moment, analogous to the IMRPhenomXAS Bilby analyses. They also assume that the data are perfectly calibrated. In Fig. 8, we show how the three-detector sky localization for GW250207 changes when these assumptions are relaxed in full parameter-estimation analyses. The inclusion of higher-order multipole moments and precession effects, which are better measured with three detectors than with two, leads to an improvement in the sky localization relative to analyses using waveforms that only model the dominant $(2, 2)$-multipole emission and aligned spins. Finally, marginalizing over calibration uncertainty broadens the sky localization.

CONSISTENCY TESTS

Analyses that test the consistency of GW signals with general relativity (GR) predictions can reveal potential errors in the underlying assumptions, whether due to deviations from GR, missing physics in waveform models, mismodeling of the noise, or miscalibration of the data [110–112]. Different tests are complementary, since they probe different types of deviations [113]. We perform a set of tests to search for inconsistencies in the signal modeling, DQ and detector calibration.

Residual tests

The residuals test examines the data for excess coherent power remaining in the detector network after subtracting a best-fit waveform [110, 113, 114]. Significant residual power may indicate additional physical effects beyond those captured by binary black hole (BBH) waveform models, alternative physics, unmodeled calibration systematic errors or instrumental noise artifacts.

The residual data are obtained by subtracting the maximum-likelihood waveform, inferred from the parameter estimation, from the original data. If the waveform model adequately captures the GW signal, the resulting residuals should be consistent with stationary Gaussian noise. We analyze the residual data using BayesWave [72, 74, 115] and compute the 90% credible upper limit on the network SNR, denoted ρ_{90}. To assess the significance of the obtained ρ_{90}, we also analyze nearby segments of detector data around the signal (without simulated signals) to estimate the background distribution. The probability of obtaining a ρ_{90} higher than or equal to that calculated from the residual data is reported as the p-value, $P(\rho^n_{90} \geq \rho_{90})$, where ρ^n_{90} is the 90% credible upper limit on the coherent network SNR from the noise-only background segments. A higher p-value indicates that the residual power is likely to originate from instrumental noise. For a single signal consistent with GR, the p-value is expected to follow a uniform distribution on the interval $(0, 1]$ [110]. The goodness-of-fit of the GR-based waveform to the data can also be quantified by calculating the 90% credible lower limit on the fitting factor:

$$\mathrm{FF}_{90} = \rho_{\mathrm{GR}}\left(\rho^2_{\mathrm{GR}} + \rho^2_{90}\right)^{-1/2}, \tag{6}$$

where ρ_{GR} is the optimal network SNR for the maximum-likelihood waveform, and a value of $\mathrm{FF}_{90} = 1$ indicates perfect agreement between the waveform and the data [110].

We compute residuals using the C00 strain data, $r(t) = d_{\mathrm{C00}}(t) - h_{\mathrm{ML}}(t)/\eta$, where $1/\eta$ is the median calibration correction factor inferred from parameter estimation, and $h_{\mathrm{ML}}(t)$ is the maximum-likelihood waveform. In analyses where calibration parameters are not inferred, the residuals are computed in the standard way as $r(t) = d_{\mathrm{C00}}(t) - h_{\mathrm{ML}}(t)$. We show the results in Fig. 9.

For GW240925, when calibration effects are not included in the analysis, $\rho_{90} = 7.2$ with a p-value of 0.26 and $\mathrm{FF}_{90} = 0.97$. When the residuals are computed using the calibration correction factor from the wide prior analysis, $\rho_{90} = 7.2$ with a p-value of 0.30 and $\mathrm{FF}_{90} = 0.97$. These ρ_{90} values fall within the range reported in analyses of previous detections [114]. The top two panels of Fig. 9 show the residuals for these two cases. No excess residual power can be identified, and these results remain consistent with the expectation for Gaussian noise, regardless of the calibration treatment.

For GW250207, $\rho_{90} = 7.2$ with a p-value of 0.28 and $\mathrm{FF}_{90} = 0.99$ when calibration effects are neglected, and $\rho_{90} = 7.6$ with a p-value of 0.19 and $\mathrm{FF}_{90} = 0.99$ when the

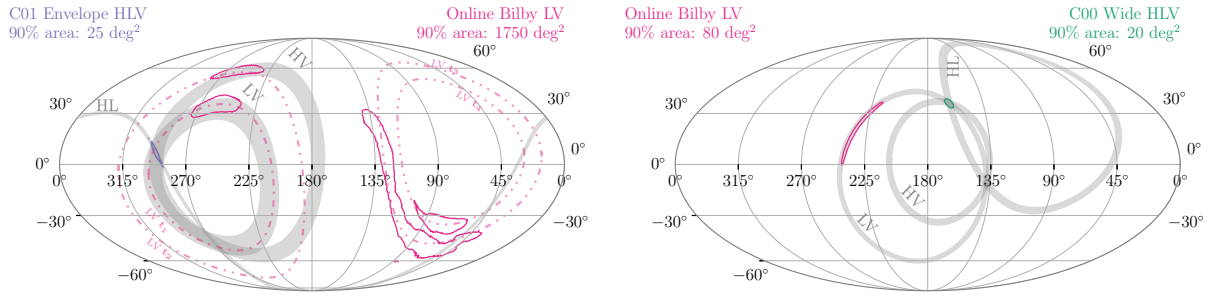

FIG. 7. Posteriors on the sky location together with rings corresponding to the inferred time delays between detector pairs for GW240925 (left) and GW250207 (right). Solid curves enclose the 90% credible areas for the two-detector (Livingston–Virgo; LV) [108, 109] and three-detector (HLV) analyses. The grey bands represent 90% credible intervals for time delays from the three-detector analyses; the three-detector localizations lie at intersections of these bands. For GW240925, the two-detector analysis produces a multimodal posterior distribution for the LV time delay, and the dot–dashed rings illustrate the approximate positions of its peaks. The time delay t_1 is selected by the three-detector analysis through consistency with the Hanford–Livingston (HL) and Hanford–Virgo (HV) time delays. For GW250207, the two- and three-detector analysis lie on a ring of consistent LV time delay. For both signals, inclusion of Hanford data improves the time-delay constraints and shifts the sky-location posterior distributions.

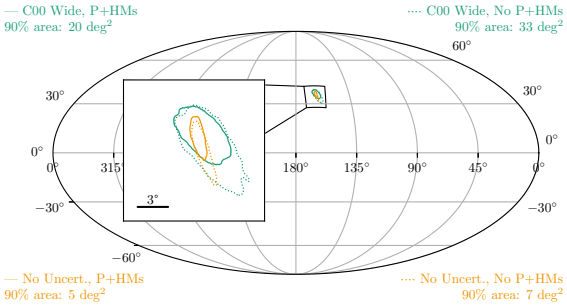

FIG. 8. Sky localization for GW250207 from four different Bilby analyses using data from Hanford, Livingston and Virgo. We show results from analyses with and without (solid and dotted lines, respectively) the inclusion of spin-precession effects and higher-order multipole moments (P+HMs), and that incorporate or neglect calibration uncertainties (green and yellow, respectively).

calibration correction factor from the wide prior analysis is applied. The corresponding residuals are shown in the bottom two panels of Fig. 9. Although the Hanford residuals appear to exhibit visible excess power in both cases, the results are consistent with expectations for Gaussian noise. We confirm that there is no correlation between these peaks and environmental monitors. We further investigate the residuals in Hanford by performing the Anderson–Darling test on a 1 s segment of the residual data centered on the search-analysis trigger time. We then compute a p-value for this statistic against the background segments. Both cases result in an equal Anderson–Darling statistic of 1.07, with a p-value of 0.98. These results indicate that the residuals are consistent with Gaussian noise.

Parametrized tests

The parametrized tests modify the signal waveforms by introducing parametrized deviations away from the GR model [117]. Modifying the waveform in this case is not expected to reproduce a signal in an alternative theory of gravity, but the additional flexibility in the waveform may identify phenomena such as missing physics in the waveform or an issue with DQ [111, 112, 118, 119]. Calibration errors could potentially mimic or mask the parametrized deviations considered here; hence, we expect that biases may result from miscalibration. However, as waveform modifications produce coherent effects across the detector network, while calibration errors affect each detector independently, the impact of miscalibration may not lead to an observable effect [120].

We study the impact of calibration uncertainties on the FTI test [121], which examines deviations in the post-Newtonian (PN) coefficients during the inspiral phase. To assess this, we perform analyses both with and without marginalization over calibration uncertainties for each signal; in the latter case, we assume the data are perfectly calibrated. The analyses adopt the same settings and calibration priors as those used for the inference of source properties. We use the aligned-spin SEOBNRv5HM_ROM waveform model [122] as the GR baseline and allow a single PN deviation parameter to vary at a time.

For both signals, the FTI results are all consistent with their GR values. When calibration uncertainty is not marginalized, the posteriors on the deviation parameters peak further from zero compared to the results obtained using the wide calibration prior. Marginalizing over the wide calibration prior also broadens the posteriors on the deviation parameters. Figure 10 shows the GW250207 results for $\delta\hat{\varphi}_0$ as a representative example. The choice of calibration prior affects GW240925 less than GW250207; an illustrative example for GW240925 is shown in Fig. 10 with the $\delta\hat{\varphi}_7$ posteriors. The resulting

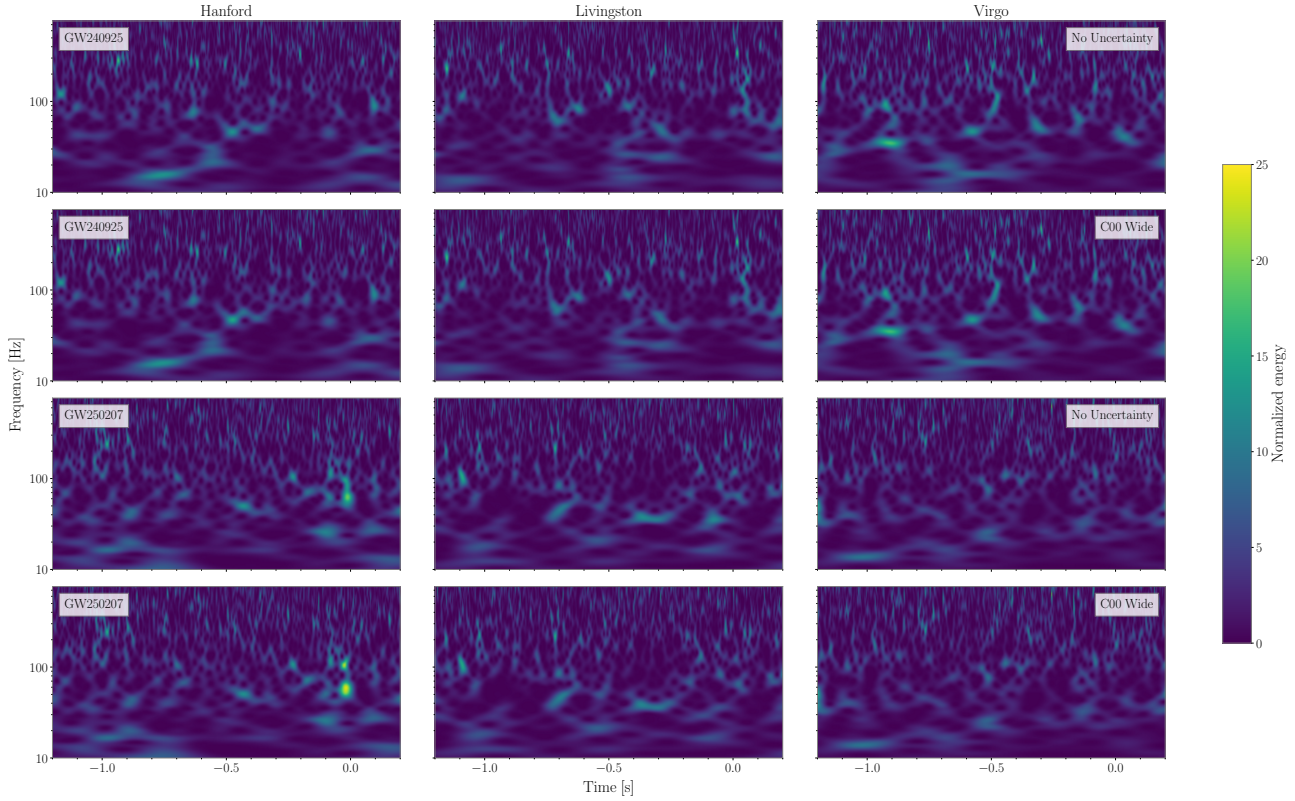

FIG. 9. Time–frequency spectrograms [116] showing residual data from LIGO Hanford (left), LIGO Livingston (middle) and Virgo (right) after waveform subtraction. Data for GW240925 and GW250207 are shown in the two top and bottom panels, respectively; for both signals, the upper panels are for results neglecting calibration uncertainty, and the lower panels are for those using a wide, uninformative prior for Hanford calibration and in-situ measured uncertainty for other detectors. Times are measured relative to the times reported from the search algorithms. The data have been whitened [5], and the scale bar shows the normalized energy (which has a reduced range compared to Fig. 1 of the main paper).

constraints on the PN deviation parameters are summarized as 90% upper bounds in Table III. Despite the wide calibration prior assumed for Hanford, the FTI bounds obtained for GW250207 on the 0PN, 1PN and higher-order PN coefficients are the most stringent achieved from GW observations to date [117, 123, 124]. Such tight constraints would not have been possible if the Hanford data were excluded from the analysis.

We similarly assess the impact of calibration uncertainties on the TIGER test [125], which allows for deviations in both the inspiral PN coefficients and the phenomenological coefficients in the post-inspiral regime. Using the IMRPhenomXPHM waveform model [77, 78] as the GR baseline, we vary one deviation parameter at a time, repeating the analysis with and without calibration uncertainty. As in the FTI case, we find that including calibration uncertainties leads to broader posteriors and improved consistency with the GR values. In the analysis of GW250207, one of the merger–ringdown parameters ($\delta\hat{c}_\ell$, which scales the damping frequency of the ringdown signal) shows that the GR value falls outside the 90% credible interval when calibration uncertainties are neglected, as shown in Fig. 10. The resulting constraints on the PN deviation parameters and the phenomenological post-inspiral parameters are reported as 90% upper bounds in Table III. The TIGER bounds from GW250207 are the most stringent from a single GW observation for the 2PN and higher-order PN coefficients, as well as for the post-inspiral parameters excluding $\delta\hat{c}_\ell$, where GW250114_082203 (GW250114_082203) is the most constraining [117, 123, 124].

To understand why GW250207 is particularly informative, we may consider its source properties. There is significant posterior support for near edge-on inclinations, a configuration that naturally enhances the visibility of higher-order spherical-harmonic moments, in particular the $(4, 4)$ multipole [126], and the component masses are unequal (for comparison, GW250114_082203 does not have support for edge-on inclinations and is consistent with having equal component masses [127]). These characteristics lead to a non-trivial contribution of higher-order multipoles to the observed signal, which are correlated with precise measurements of the properties like chirp mass, mass ratio and component spins, as well as the deviation parameters themselves [125, 128]. Having precise measurements of the source properties is beneficial because deviation parameters are correlated with the underly-

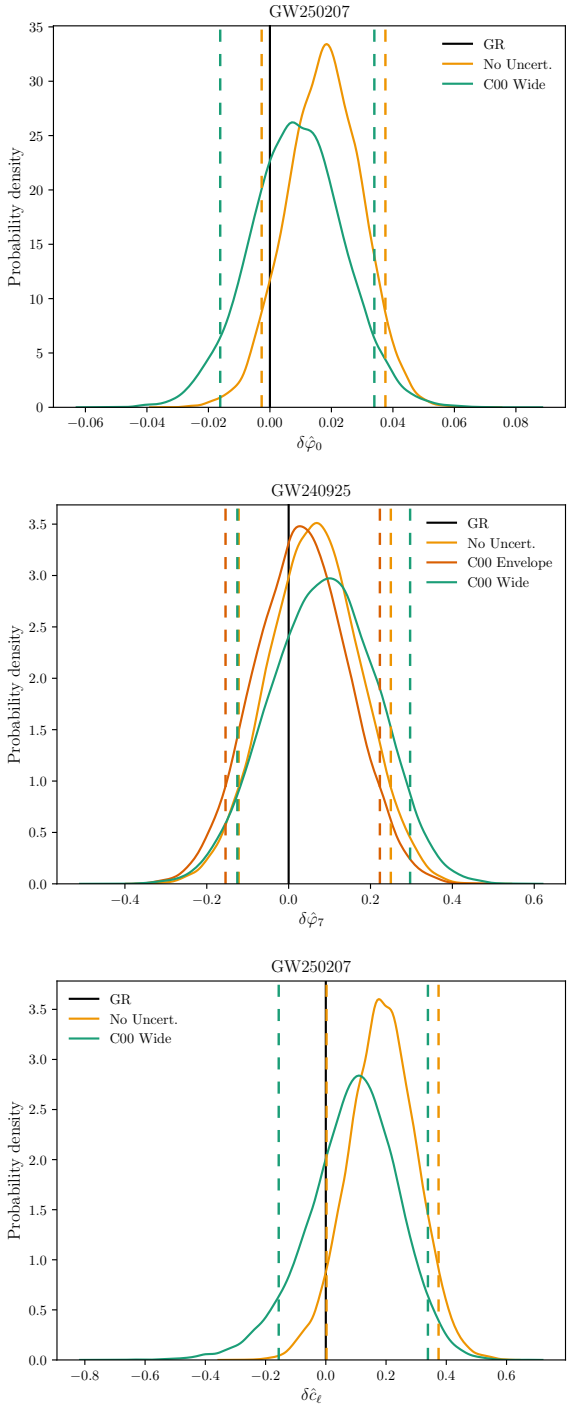

FIG. 10. Illustrative results from the FTI (top, middle) and TIGER (bottom) analyses for GW250207 and GW240925. The dashed lines indicate the 90% credible interval, and the solid black line indicates the GR value. Shown are selected deviation parameters; other parameters exhibit similar behaviors. Results are shown using a wide calibration prior for Hanford (green) and neglecting calibration uncertainty (yellow) for both signals. For GW240925, we additionally show results using the in-situ measured calibration prior for Hanford (orange). Neglecting calibration uncertainty leads to narrower posteriors and reduced agreement with GR.

TABLE III. The 90% upper bounds on the PN deviation parameters obtained with FTI and TIGER, as well as bounds on the post-inspiral deviation parameters from TIGER. Results are from analyses that use the in-situ measured calibration prior for Hanford in the case of GW240925 and a wide calibration prior for Hanford in the case of GW250207. If the GR waveform accurately models the signal and the data are accurately described by our analysis assumptions (well-calibrated data with stationary Gaussian noise), we expect the results to be statistically consistent with deviation values of zero; conversely, a deviation away from zero does not necessarily imply a modification of GR.

		GW240925		GW250207	
PN order	Parameter	FTI	TIGER	FTI	TIGER
-1	$\delta\hat{\varphi}_{-2}$	5.0×10^{-4}	5.8×10^{-4}	2.6×10^{-3}	3.0×10^{-3}
0	$\delta\hat{\varphi}_0$	0.085	0.20	0.034	0.05
0.5	$\delta\hat{\varphi}_1$	0.37	0.22	0.11	0.21
1	$\delta\hat{\varphi}_2$	0.17	0.14	0.056	0.13
1.5	$\delta\hat{\varphi}_3$	0.076	0.10	0.027	0.08
2	$\delta\hat{\varphi}_4$	0.36	0.88	0.14	0.59
2.5 log	$\delta\hat{\varphi}_{5l}$	0.088	0.33	0.037	0.20
3	$\delta\hat{\varphi}_6$	0.11	0.45	0.046	0.31
3 log	$\delta\hat{\varphi}_{6l}$	1.9	2.15	0.87	1.35
3.5	$\delta\hat{\varphi}_7$	0.22	1.15	0.088	0.88
Post-inspiral parameter		TIGER		TIGER	
	$\delta\hat{b}_1$	0.04		0.03	
	$\delta\hat{b}_2$	0.01		0.01	
	$\delta\hat{b}_3$	0.01		0.01	
	$\delta\hat{b}_4$	0.03		0.02	
	$\delta\hat{c}_1$	0.84		0.25	
	$\delta\hat{c}_2$	0.26		0.08	
	$\delta\hat{c}_3$	0.75		0.24	
	$\delta\hat{c}_\ell$	1.30		0.34	

ing GR parameters, and such degeneracies often weaken constraints on the PN terms, even with many inspiral cycles in-band [123, 129]. The signal properties of GW250207 reduce these degeneracies, and allow the analyses to obtain stringent constraints on the higher-PN deviation parameters and the phenomenological post-inspiral parameters.

The TIGER inspiral bounds are generally weaker than those from FTI. In these PN-parametrized tests, the beyond-GR correction is included only in the inspiral regime, while the merger–ringdown sector remains fixed to its GR prediction. The difference in bounds between the two frameworks likely arises from the different choices of upper cutoff frequency for the beyond-GR correction. TIGER terminates these corrections at the $(2,2)$-multipole frequency at approximately the minimum-energy circular orbit $f_{22}^{\rm MECO}$, whereas FTI tapers the waveform at the peak frequency of the $(2,2)$ multipole $f_{22}^{\rm peak}$ [121, 123, 125]. For GW240925 (GW250207), the cutoff frequencies are approximately 259 Hz (64 Hz) for TIGER

and 684 Hz (169 Hz) for FTI. The higher cutoff frequency in the FTI analysis leads to the tighter constraints [117].

We also investigate the impact of detector calibration on the Principal Component Analysis (PCA) test of GR, which looks for correlated deviations across multiple PN orders simultaneously, thereby probing signatures of more complex departures from GR predictions [130–133]. We perform PCA analyses with and without calibration uncertainty for each signal, using the same analysis settings and priors as in the source-parameter estimation. In the TIGER framework, we use the IMRPhenomXPHM waveform model [77, 78], while in the FTI framework, we adopt SEOBNRv5HM_ROM [122] as the baseline GR model and simultaneously vary six fractional PN deformation parameters between 1.5PN and 3.5PN order. We then apply PCA to identify linear combinations of PN deformation parameters that are best constrained by the data: those aligned with the eigenvectors corresponding to the smallest uncertainties define the principal directions of parameter covariance, with the leading component representing the linear combination of PN coefficients that is most tightly constrained. For GW250207, the PCA posteriors in both frameworks are consistent with the GR values. For GW240925, we only report results within the TIGER framework due to the computational cost of analyzing this longer signal; again we find consistency with GR. Incorporating wide, uninformative calibration priors broadens the PCA parameter posteriors and shifts their peaks closer to zero, improving agreement with GR relative to analyses that neglect calibration uncertainty. Figure 11 presents the 90% credible contours for the leading two PCA parameters, $\delta\hat{\varphi}^{(1)}_{\rm PCA}$ and $\delta\hat{\varphi}^{(2)}_{\rm PCA}$.

These results show that not including calibration uncertainty could potentially lead to biases in tests of GR. Using wide calibration priors when the calibration of the detector is unknown can prevent this from happening, at the cost of losing some constraining power.

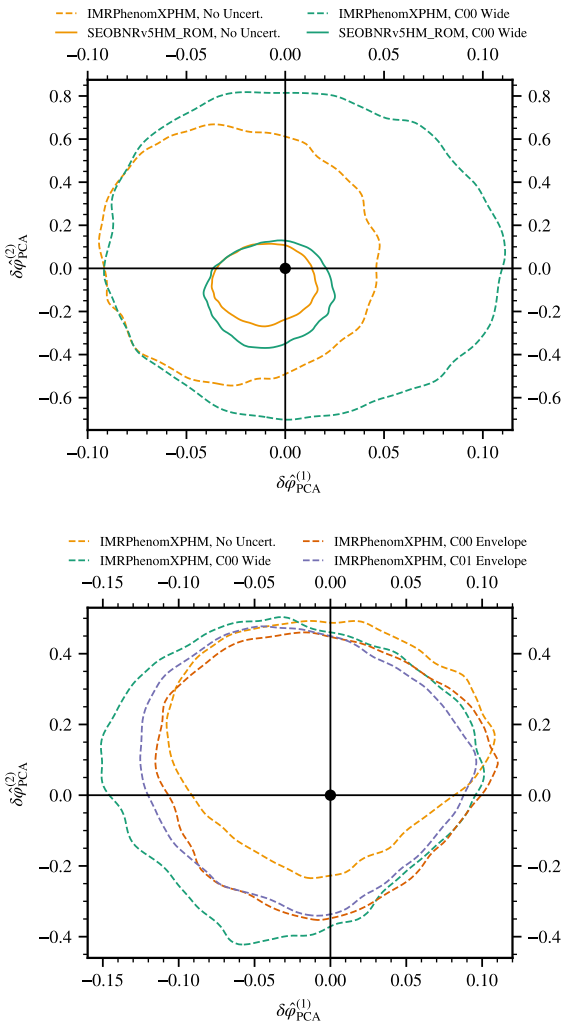

FIG. 11. The 90% credible contours for the leading two PCA parameters for GW250207 (top) and GW240925 (bottom). Results are presented for analyses that neglect calibration uncertainty (yellow) and those that adopt a wide prior on the Hanford calibration (green), for both signals, in the TIGER (dashed) and FTI (solid) frameworks. For GW240925, we additionally show results using the in-situ measured calibration priors applied to the Hanford miscalibrated C00 (orange) and recalibrated C01 (purple) data. Neglecting calibration uncertainty leads to narrower posteriors that peak further from zero, while incorporating a wide calibration prior yields wider posteriors that exhibit improved consistency with GR.

Ringdown tests

We perform a range of quasinormal mode (QNM) analyses using the ringdown [134–136], QNMRF [137–140], and pSEOBNR [141–144] pipelines. For GW240925, the ringdown SNR is low, as expected for a lower-mass BBH, and is therefore not suitable for an in-depth ringdown test. The pSEOBNR analysis requires a SNR of $\rho \geq 8$ in both the inspiral and post-inspiral regimes to break the degeneracy between the QNM frequency deviation and the remnant mass [142], while the ringdown analysis requires a post-peak SNR of $\rho > 10$. Only GW250207 satisfies the SNR thresholds for both pSEOBNR and ringdown analyses. As a check, we perform the QNMRF analysis focusing on the dominant $(2, 2, 0)$ mode for GW240925. The resulting broad posteriors are generally uninformative for both the C00 and C01 data. Therefore, we focus on the results from GW250207, where the higher ringdown SNR enables a more informative test.

Both ringdown and QNMRF pipelines do not implement marginalization over calibration uncertainty, so we perform single-detector analyses of Hanford and Livingston data individually to explore results with different calibration errors. We fit the dominant Kerr $(2, 2, 0)$ mode to the data over a ~ 1 ms range of late starting times, $t_{\rm peak} + \Delta t$, where $\Delta t \in [8, 11] t_{M_{\rm f}}$ with $t_{M_{\rm f}} = G(1+z)M_{\rm f}/c^3$, and $t_{\rm peak}$ is the reference peak time of the strain. The reference peak time $t_{\rm peak}$, remnant mass $M_{\rm f}$, and extrinsic parameters are fixed to those from the maximum-likelihood posterior sample of the NRSur7dq4 IMR analysis using the wide calibration prior. To condition

the data prior to fitting with RINGDOWN, we high-pass filter the time series above a cut-off frequency of 10 Hz, downsample with an anti-aliasing digital filter [136] to a 2048 Hz sampling rate, and crop to a 0.66 s analysis segment beginning at the native 16384 Hz discretized sample closest to $t_{\rm peak} + \Delta t$. In the QNMRF analysis, we analyze the data over the intervals $[\Delta t, \Delta t + 0.2\,\mathrm{s}]$, downsampled at 4096 Hz. A check was also performed over longer 0.66 s segments to verify that differences in configuration between QNMRF and RINGDOWN do not affect the results. We obtain consistent results from both pipelines: for all start times, the remnant black hole properties inferred from the Livingston data are consistent with the maximum-likelihood IMR result, while the measured $(2, 2, 0)$ QNM from the Hanford analysis are biased to lower frequencies and higher damping times, disagreeing with the full-signal inference at the 90% credible level. No evidence for an overtone is found in the Livingston-only analysis within this time window by either pipeline (e.g., with all QNMRF detection statistics falling below threshold [140]), thereby justifying a fit using only the fundamental $(2, 2, 0)$ mode.

The PSEOBNR analysis [141–145] introduces fractional deviations to the frequency and damping time of the fundamental QNMs in the underlying SEOBNRv5PHM waveform model, parametrized as

$$f_{\ell m 0} = f^{\rm GR}_{\ell m 0}(1 + \delta\hat{f}_{\ell m 0}), \quad \tau_{\ell m 0} = \tau^{\rm GR}_{\ell m 0}(1 + \delta\hat{\tau}_{\ell m 0}). \tag{7}$$

Unlike analyses that isolate the ringdown phase, PSEOBNR directly modifies parameters within an IMR waveform model, allowing it to leverage the full signal duration and total SNR, while avoiding ambiguities in the choice of the ringdown start time. The analysis constraints fractional deviations in the $(2, 2, 0)$ QNM in addition to the GR parameters of the waveform model, assuming uniform priors of $\delta\hat{f}_{220} \in [-0.8, 2]$ and $\delta\hat{\tau}_{220} \in [-0.8, 2]$. The high SNR of GW250207 also allows us to constrain fractional deviations in the $(4, 4, 0)$ QNM, assuming uniform priors of $\delta\hat{f}_{440} \in [-0.8, 0.8]$ and $\delta\hat{\tau}_{440} \in [-0.8, 0.8]$. Since this analysis can explicitly incorporate calibration uncertainties, we perform analyses both with and without calibration uncertainties, adopting a wide, uninformative prior for Hanford in the latter case. In both cases, the inferred remnant parameters and QNM deviations remain consistent with the IMR analysis within the 90% credible interval; however, neglecting calibration uncertainty reduces this level of agreement. The $(2, 2, 0)$ QNM is well constrained, with $\delta\hat{f}_{220} = 0.01^{+0.03}_{-0.03}$ and $\delta\hat{\tau}_{220} = -0.02^{+0.17}_{-0.15}$ when calibration uncertainty is incorporated, compared to $\delta\hat{f}_{220} = -0.01^{+0.02}_{-0.02}$ and $\delta\hat{\tau}_{220} = 0.12^{+0.14}_{-0.13}$ when it is neglected, for the analyses with fractional deviations in the $(2, 2, 0)$ QNM in addition to the GR parameters. The frequency of the $(4, 4, 0)$ QNM is constrained to $\delta\hat{f}_{440} = 0.03^{+0.20}_{-0.64}$ and $\delta\hat{f}_{440} = 0.00^{+0.12}_{-0.28}$ with and without incorporating calibration uncertainty, respectively. The damping time of the $(4, 4, 0)$ QNM remains unconstrained. The impact of calibration uncertainties is less pronounced in the PSEOBNR analysis, potentially because it incorporates data from all three detectors and uses a full waveform model, in contrast to single-detector, ringdown-only analyses of QNMRF and RINGDOWN.

REFERENCES

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914, Phys. Rev. D **95**, 062003 (2017), arXiv:1602.03845 [gr-qc].

[2] C. Cahillane *et al.* (LIGO Instrument Science Collaboration), Calibration uncertainty for Advanced LIGO's first and second observing runs, Phys. Rev. D **96**, 102001 (2017), arXiv:1708.03023 [astro-ph.IM].

[3] A. D. Viets *et al.*, Reconstructing the calibrated strain signal in the Advanced LIGO detectors, Classical Quantum Gravity **35**, 095015 (2018), arXiv:1710.09973 [astro-ph.IM].

[4] L. Sun *et al.*, Characterization of systematic error in Advanced LIGO calibration, Classical Quantum Gravity **37**, 225008 (2020), arXiv:2005.02531 [astro-ph.IM].

[5] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, Classical Quantum Gravity **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].

[6] B. Allen, *LIGO calibration accuracy*, Tech. Rep. LIGO-T960189 (LIGO, 1996).

[7] T. Akutsu *et al.* (KAGRA Collaboration), Overview of KAGRA: Calibration, detector characterization, physical environmental monitors, and the geophysics interferometer, Prog. Theor. Exp. Phys. **2021**, 05A102 (2021), arXiv:2009.09305 [gr-qc].

[8] F. Acernese *et al.* (Virgo Collaboration), Calibration of Advanced Virgo and reconstruction of the detector strain $h(t)$ during the Observing Run O3, Classical Quantum Gravity **39**, 045006 (2022), arXiv:2107.03294 [gr-qc].

[9] C. Grimaud, *Virgo interferometer calibration and h(t) strain reconstruction uncertainty during the O4 observing run*, Ph.D. thesis, Université Savoie Mont Blanc (2025).

[10] A. G. Sullivan *et al.*, Timing system of LIGO discoveries, Phys. Rev. D **108**, 022003 (2023), arXiv:2304.01188 [astro-ph.IM].

[11] M. Wade *et al.*, Toward low-latency, high-fidelity calibration of the LIGO detectors with enhanced monitoring tools, Class. Quant. Grav. **42**, 215016 (2025), arXiv:2508.08423 [gr-qc].

[12] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-4.0: Methods for Identifying and Characterizing Gravitational-wave Transients, arXiv e-prints (2025), arXiv:2508.18081 [gr-qc].

[13] S. Karki *et al.*, The Advanced LIGO Photon Calibrators, Rev. Sci. Instrum. **87**, 114503 (2016), arXiv:1608.05055 [astro-ph.IM].

[14] D. Estevez, P. Lagabbe, A. Masserot, L. Rolland, M. Seglar-Arroyo, and D. Verkindt, The Advanced Virgo Photon Calibrators, Classical Quantum Gravity **38**, 075007 (2021), arXiv:2009.08103 [astro-ph.IM].

[15] D. Bhattacharjee, Y. Lecoeuche, S. Karki, J. Betzwieser, V. Bossilkov, S. Kandhasamy, E. Payne, and R. L. Savage, Fiducial displacements with improved accuracy for the global network of gravitational wave detectors, Classical Quantum Gravity **38**, 015009 (2021), arXiv:2006.00130 [astro-ph.IM].

[16] D. Chen *et al.*, Performance of the KAGRA photon calibrators during the fourth joint observing run with LIGO

and Virgo, Classical Quantum Gravity **42**, 185018 (2025), arXiv:2504.12657 [astro-ph.IM].
[17] F. Aubin, E. Dangelser, D. Estevez, A. Masserot, B. Mours, T. Pradier, A. Syx, and P. Van Hove, The Virgo Newtonian calibration system for the O4 observing run, Classical Quantum Gravity **41**, 235003 (2024), arXiv:2406.10028 [gr-qc].
[18] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, *LIGO Virgo KAGRA Calibration Uncertainty (O4)*, Tech. Rep. DCC-T2500288 (LIGO, Virgo, and KAGRA, 2025).
[19] S. Soni *et al.* (LIGO Instrument Science Collaboration), Reducing scattered light in LIGO's third observing run, Classical Quantum Gravity **38**, 025016 (2021), arXiv:2007.14876 [astro-ph.IM].
[20] J. Glanzer, S. Soni, J. Spoon, A. Effler, and G. González, Noise in the LIGO livingston gravitational wave observatory due to trains, Classical Quantum Gravity **40**, 195015 (2023), arXiv:2304.07477 [astro-ph.IM].
[21] S. Soni, J. Glanzer, A. Effler, V. Frolov, G. González, A. Pele, and R. Schofield, Modeling and reduction of high frequency scatter noise at LIGO Livingston, Classical Quantum Gravity **41**, 135015 (2024), arXiv:2311.05730 [astro-ph.IM].
[22] L. K. Nuttall, Characterizing transient noise in the LIGO detectors, Phil. Trans. R. Soc. A **376**, 20170286 (2018), arXiv:1804.07592 [astro-ph.IM].
[23] J. Glanzer *et al.*, Data quality up to the third observing run of Advanced LIGO: Gravity Spy glitch classifications, Classical Quantum Gravity **40**, 065004 (2023), arXiv:2208.12849 [gr-qc].
[24] P. B. Covas *et al.* (LIGO Instrument Science Collaboration), Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of Advanced LIGO, Phys. Rev. D **97**, 082002 (2018), arXiv:1801.07204 [astro-ph.IM].
[25] S. Soni *et al.* (LIGO Instrument Science Collaboration), LIGO Detector Characterization in the first half of the fourth Observing run, Classical Quantum Gravity **42**, 085016 (2025), arXiv:2409.02831 [astro-ph.IM].
[26] D. Davis *et al.* (LIGO Instrument Science Collaboration), LIGO Detector Characterization in the Second and Third Observing Runs, Classical Quantum Gravity **38**, 135014 (2021), arXiv:2101.11673 [astro-ph.IM].
[27] F. Acernese *et al.* (Virgo Collaboration), Virgo detector characterization and data quality: tools, Classical Quantum Gravity **40**, 185005 (2023), arXiv:2210.15634 [gr-qc].
[28] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Machine-learning nonstationary noise out of gravitational-wave detectors, Phys. Rev. D **101**, 042003 (2020), arXiv:1911.09083 [gr-qc].
[29] S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data, Phys. Rev. D **106**, 042006 (2022), arXiv:2205.13580 [gr-qc].
[30] 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), arXiv:2207.03429 [astro-ph.IM].
[31] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85**, 122006 (2012), arXiv:gr-qc/0509116.
[32] H.-Y. Chen, D. E. Holz, J. Miller, M. Evans, S. Vitale, and J. Creighton, Distance measures in gravitational-wave astrophysics and cosmology, Classical Quantum Gravity **38**, 055010 (2021), arXiv:1709.08079 [astro-ph.CO].
[33] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-4.0: An Introduction to Version 4.0 of the Gravitational-Wave Transient Catalog, Astrophys. J. Lett. **995**, L18 (2025), arXiv:2508.18080 [gr-qc].
[34] A. Helmling-Cornell, P. Nguyen, R. Schofield, and R. Frey, Automated evaluation of environmental coupling for Advanced LIGO gravitational wave detections, Classical Quantum Gravity **41**, 145003 (2024), arXiv:2312.00735 [gr-qc].
[35] S. Hourihane and K. Chatziioannou, Glitches far from transient gravitational-wave events do not bias inference, Phys. Rev. D **112**, 084006 (2025), arXiv:2506.21869 [gr-qc].
[36] J. Betzwieser, *Calibration statements on OOOM requests for GW240925 and GW250207*, Tech. Rep. LIGO-T2500292 (LIGO, 2025).
[37] R. Macas and A. Lundgren, Sensitive test of non-Gaussianity in gravitational-wave detector data, Phys. Rev. D **108**, 063016 (2023), arXiv:2306.09019 [gr-qc].
[38] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Low-latency Gravitational-wave Alerts for Multimessenger Astronomy during the Second Advanced LIGO and Virgo Observing Run, Astrophys. J. **875**, 161 (2019), arXiv:1901.03310 [astro-ph.HE].
[39] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run, Phys. Rev. X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].
[40] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run, arXiv e-prints (2025), arXiv:2508.18082 [gr-qc].
[41] S. Klimenko, S. Mohanty, M. Rakhmanov, and G. Mitselmakher, Constraint likelihood analysis for a network of gravitational wave detectors, Phys. Rev. D **72**, 122002 (2005), arXiv:gr-qc/0508068.
[42] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, A coherent method for detection of gravitational wave bursts, Classical Quantum Gravity **25**, 114029 (2008), arXiv:0802.3232 [gr-qc].
[43] S. Klimenko *et al.*, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, Phys. Rev. D **93**, 042004 (2016), arXiv:1511.05999 [gr-qc].
[44] T. Mishra, S. Bhaumik, V. Gayathri, M. J. Szczepańczyk, I. Bartos, and S. Klimenko, Gravitational waves detected by a burst search in LIGO/Virgo's third observing run, Phys. Rev. D **111**, 023054 (2025), arXiv:2410.15191 [astro-ph.HE].
[45] C. Messick *et al.*, Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data, Phys. Rev. D **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].
[46] S. Sachdev *et al.*, The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs, arXiv e-prints (2019), arXiv:1901.08580 [gr-qc].
[47] C. Hanna *et al.*, Fast evaluation of multidetector consistency for real-time gravitational wave searches, Phys. Rev. D **101**, 022003 (2020), arXiv:1901.02227 [gr-qc].
[48] K. Cannon *et al.*, GstLAL: A software framework for grav-

itational wave discovery, SoftwareX **14**, 100680 (2021), arXiv:2010.05082 [astro-ph.IM].

[49] S. Sakon *et al.*, Template bank for compact binary mergers in the fourth observing run of Advanced LIGO, Advanced Virgo, and KAGRA, Phys. Rev. D **109**, 044066 (2024), arXiv:2211.16674 [gr-qc].

[50] B. Ewing *et al.*, Performance of the low-latency GstLAL inspiral search towards LIGO, Virgo, and KAGRA's fourth observing run, Phys. Rev. D **109**, 042008 (2024), arXiv:2305.05625 [gr-qc].

[51] L. Tsukada *et al.*, Improved ranking statistics of the GstLAL inspiral search for compact binary coalescences, Phys. Rev. D **108**, 043004 (2023), arXiv:2305.06286 [astro-ph.IM].

[52] P. Joshi *et al.*, New Methods for Offline GstLAL Analyses, arXiv e-prints (2025), arXiv:2506.06497 [gr-qc].

[53] P. Joshi *et al.*, How Many Times Should We Matched Filter Gravitational Wave Data? A Comparison of GstLAL's Online and Offline Performance, arXiv e-prints (2025), arXiv:2505.23959 [gr-qc].

[54] T. Adams, D. Buskulic, V. Germain, G. M. Guidi, F. Marion, M. Montani, B. Mours, F. Piergiovanni, and G. Wang, Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era, Classical Quantum Gravity **33**, 175012 (2016), arXiv:1512.02864 [gr-qc].

[55] F. Aubin *et al.*, The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run, Classical Quantum Gravity **38**, 095004 (2021), arXiv:2012.11512 [gr-qc].

[56] C. Alléné *et al.*, The MBTA pipeline for detecting compact binary coalescences in the fourth LIGO-Virgo-KAGRA observing run, Classical Quantum Gravity **42**, 105009 (2025), arXiv:2501.04598 [gr-qc].

[57] Q. Chu *et al.*, SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences, Phys. Rev. D **105**, 024023 (2022), arXiv:2011.06787 [gr-qc].

[58] T. Dal Canton, A. H. Nitz, B. Gadre, G. S. Cabourn Davies, V. Villa-Ortega, T. Dent, I. Harry, and L. Xiao, Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live, Astrophys. J. **923**, 254 (2021), arXiv:2008.07494 [astro-ph.HE].

[59] T. Dal Canton *et al.*, Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors, Phys. Rev. D **90**, 082004 (2014), arXiv:1405.6731 [gr-qc].

[60] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[61] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search, Astrophys. J. **849**, 118 (2017), arXiv:1705.01513 [gr-qc].

[62] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), GW150914: First results from the search for binary black hole coalescence with Advanced LIGO, Phys. Rev. D **93**, 122003 (2016), arXiv:1602.03839 [gr-qc].

[63] R. Essick *et al.*, Compact binary coalescence sensitivity estimates with injection campaigns during the LIGO-Virgo-KAGRA Collaborations' fourth observing run, Phys. Rev. D **112**, 102001 (2025), arXiv:2508.10638 [gr-qc].

[64] J. S. Speagle, dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[65] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. Ser. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[66] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue, Mon. Not. R. Astron. Soc. **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].

[67] 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), arXiv:gr-qc/9402014.

[68] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman, Gravitational radiation damping of compact binary systems to second postNewtonian order, Phys. Rev. Lett. **74**, 3515 (1995), arXiv:gr-qc/9501027.

[69] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Binary Black Hole Mergers in the first Advanced LIGO Observing Run, Phys. Rev. X **6**, 041015 (2016), [Erratum: Phys. Rev. X 8, 039903 (2018)], arXiv:1606.04856 [gr-qc].

[70] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Properties of the Binary Black Hole Merger GW150914, Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].

[71] N. Christensen and R. Meyer, Parameter estimation with gravitational waves, Rev. Mod. Phys. **94**, 025001 (2022), arXiv:2204.04449 [gr-qc].

[72] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches, Classical Quantum Gravity **32**, 135012 (2015), arXiv:1410.3835 [gr-qc].

[73] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D **91**, 084034 (2015), arXiv:1410.3852 [gr-qc].

[74] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, Phys. Rev. D **103**, 044006 (2021), arXiv:2011.09494 [gr-qc].

[75] T. Baka *et al.*, *Correcting misspecification of calibration uncertainties in gravitational-wave data analysis with efficient reweighting*, Tech. Rep. LIGO-T2500295 (LIGO, 2025).

[76] W. Farr, B. Farr, and T. Littenberg, *Modelling calibration errors in CBC waveforms*, Tech. Rep. LIGO-T1400682 (LIGO, 2014).

[77] 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), arXiv:2004.06503 [gr-qc].

[78] M. Colleoni, F. A. R. Vidal, C. García-Quirós, S. Akçay, and S. Bera, Fast frequency-domain gravitational waveforms for precessing binaries with a new twist, Phys. Rev. D **111**, 104019 (2025), arXiv:2412.16721 [gr-qc].

[79] E. Hamilton, L. London, J. E. Thompson, E. Fauchon-Jones, M. Hannam, C. Kalaghatgi, S. Khan, F. Pannarale, and A. Vano-Vinuales, Model of gravitational waves from precessing black-hole binaries through merger and ringdown, Phys. Rev. D **104**, 124027 (2021), arXiv:2107.08876 [gr-qc].

[80] E. Hamilton *et al.*, PhenomXPNR: An improved gravitational wave model linking precessing inspirals and NR-calibrated merger-ringdown, arXiv e-prints (2025), arXiv:2507.02604 [gr-qc].

[81] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, Next

generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes, Phys. Rev. D **108**, 124037 (2023), arXiv:2303.18046 [gr-qc].
[82] H. Estellés, A. Buonanno, R. Enficiaud, C. Foo, and L. Pompili, Adding equatorial-asymmetric effects for spin-precessing binaries into the SEOBNRv5PHM waveform model, Phys. Rev. D **113**, 044049 (2026), arXiv:2506.19911 [gr-qc].
[83] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Surrogate model of hybridized numerical relativity binary black hole waveforms, Phys. Rev. D **99**, 064045 (2019), arXiv:1812.07865 [gr-qc].
[84] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes, Phys. Rev. D **102**, 064001 (2020), arXiv:2001.11412 [gr-qc].
[85] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].
[86] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[87] LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration, LIGO/Virgo/KAGRA S240925n: Identification of a GW compact binary merger candidate, GCN **37604** (2024).
[88] A. Krolak and B. F. Schutz, Coalescing binaries — Probe of the universe, Gen. Relativ. Gravit. **19**, 1163 (1987).
[89] S. Nissanke, D. E. Holz, S. A. Hughes, N. Dalal, and J. L. Sievers, Exploring short gamma-ray bursts as gravitational-wave standard sirens, Astrophys. J. **725**, 496 (2010), arXiv:0904.1017 [astro-ph.CO].
[90] T. Narikawa, M. Kaneyama, and H. Tagoshi, Optimal follow-up observations of gravitational wave events with small optical telescopes, Phys. Rev. D **96**, 084067 (2017), arXiv:1705.04008 [gr-qc].
[91] S. A. Usman, J. C. Mills, and S. Fairhurst, Constraining the Inclinations of Binary Mergers from Gravitational-wave Observations, Astrophys. J. **877**, 82 (2019), arXiv:1809.10727 [gr-qc].
[92] B. S. Sathyaprakash and B. F. Schutz, Physics, Astrophysics and Cosmology with Gravitational Waves, Living Rev. Relativity **12**, 2 (2009), arXiv:0903.0338 [gr-qc].
[93] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Effects of waveform model systematics on the interpretation of GW150914, Classical Quantum Gravity **34**, 104002 (2017), arXiv:1611.07531 [gr-qc].
[94] K. S. Thorne, Gravitational radiation, in *Three hundred years of gravitation*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1987) Chap. 9, pp. 330–458.
[95] L. S. Finn and D. F. Chernoff, Observing binary inspiral in gravitational radiation: One interferometer, Phys. Rev. D **47**, 2198 (1993), arXiv:gr-qc/9301003.
[96] B. F. Schutz, Networks of gravitational wave detectors and three figures of merit, Classical Quantum Gravity **28**, 125023 (2011), arXiv:1102.5421 [astro-ph.IM].
[97] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence, Phys. Rev. Lett. **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].
[98] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
[99] R. Essick and D. E. Holz, Calibrating gravitational-wave detectors with GW170817, Classical Quantum Gravity **36**, 125002 (2019), arXiv:1902.08076 [astro-ph.IM].
[100] R. Essick, Calibration uncertainty's impact on gravitational-wave observations, Phys. Rev. D **105**, 082002 (2022), arXiv:2202.00823 [astro-ph.IM].
[101] M. Pürrer and C.-J. Haster, Gravitational waveform accuracy requirements for future ground-based detectors, Phys. Rev. Res. **2**, 023151 (2020), arXiv:1912.10055 [gr-qc].
[102] E. Payne, C. Talbot, P. D. Lasky, E. Thrane, and J. S. Kissel, Gravitational-wave astronomy with a physical calibration model, Phys. Rev. D **102**, 122004 (2020), arXiv:2009.10193 [astro-ph.IM].
[103] S. Vitale, C.-J. Haster, L. Sun, B. Farr, E. Goetz, J. Kissel, and C. Cahillane, Physical approach to the marginalization of LIGO calibration uncertainties, Phys. Rev. D **103**, 063016 (2021), arXiv:2009.10192 [gr-qc].
[104] K. Pearson, Note on regression and inheritance in the case of two parents, Proc. R. Soc. Lond. **58**, 240 (1895).
[105] R. Essick, S. Vitale, E. Katsavounidis, G. Vedovato, and S. Klimenko, Localization of short duration gravitational-wave transients with the early advanced LIGO and Virgo detectors, Astrophys. J. **800**, 81 (2015), arXiv:1409.2435 [astro-ph.HE].
[106] L. P. Singer *et al.*, The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo, Astrophys. J. **795**, 105 (2014), arXiv:1404.5623 [astro-ph.HE].
[107] A. Ouzriat, V. Sordini, and F. Di Renzo, Characterizing Low-Latency Sky Localization in Multi-Detector Gravitational-Wave Networks, arXiv e-prints (2025), arXiv:2510.21930 [astro-ph.HE].
[108] LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration, LIGO/Virgo/KAGRA S240925n: Updated Sky localization and Source Classification, GCN **37607** (2024).
[109] LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration, LIGO/Virgo/KAGRA S250207bg: Updated Sky localization, GCN **39242** (2025).
[110] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), Tests of General Relativity with GWTC-3, Phys. Rev. D **112**, 084080 (2025), arXiv:2112.06861 [gr-qc].
[111] A. Gupta *et al.*, Possible causes of false general relativity violations in gravitational wave observations, SciPost Phys. Comm. Rep. 10.21468/SciPostPhysCommRep.5 (2024), arXiv:2405.02197 [gr-qc].
[112] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GW230814: investigation of a loud gravitational-wave signal observed with a single detector, arXiv e-prints (2025), arXiv:2509.07348 [gr-qc].
[113] N. K. Johnson-McDaniel, A. Ghosh, S. Ghonge, M. Saleem, N. V. Krishnendu, and J. A. Clark, Investigating the relation between gravitational wave tests of general relativity, Phys. Rev. D **105**, 044020 (2022), arXiv:2109.06988 [gr-qc].
[114] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-4.0: Tests of General Relativity. I. Overview and General Tests, arXiv e-prints (2026), arXiv:2603.19019 [gr-qc].
[115] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and

M. Millhouse, Enabling high confidence detections of gravitational-wave bursts, Phys. Rev. D **94**, 044050 (2016), arXiv:1511.08752 [gr-qc].

[116] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts, Classical Quantum Gravity **21**, S1809 (2004), arXiv:gr-qc/0412119.

[117] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GWTC-4.0: Tests of General Relativity. II. Parameterized Tests, arXiv e-prints (2026), arXiv:2603.19020 [gr-qc].

[118] L. Sampson, N. Cornish, and N. Yunes, Gravitational Wave Tests of Strong Field General Relativity with Binary Inspirals: Realistic Injections and Optimal Model Selection, Phys. Rev. D **87**, 102001 (2013), arXiv:1303.1185 [gr-qc].

[119] J. Meidam *et al.*, Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method, Phys. Rev. D **97**, 044033 (2018), arXiv:1712.08772 [gr-qc].

[120] B. Edelman *et al.*, Constraining unmodeled physics with compact binary mergers from GWTC-1, Phys. Rev. D **103**, 042004 (2021), arXiv:2008.06436 [gr-qc].

[121] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, Tests of general relativity with gravitational-wave observations using a flexible theory-independent method, Phys. Rev. D **107**, 044020 (2023), arXiv:2203.13937 [gr-qc].

[122] L. Pompili *et al.*, Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes, Phys. Rev. D **108**, 124035 (2023), arXiv:2303.18039 [gr-qc].

[123] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Tests of General Relativity with GW170817, Phys. Rev. Lett. **123**, 011102 (2019), arXiv:1811.00364 [gr-qc].

[124] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), Black Hole Spectroscopy and Tests of General Relativity with GW250114, Phys. Rev. Lett. **136**, 041403 (2026), arXiv:2509.08099 [gr-qc].

[125] S. Roy, M. Haney, G. Pratten, P. T. H. Pang, and C. Van Den Broeck, Improved parametrized test of general relativity using the IMRPhenomX waveform family: Including higher harmonics and precession, Phys. Rev. D **113**, 024016 (2026), arXiv:2504.21147 [gr-qc].

[126] C. Mills and S. Fairhurst, Measuring gravitational-wave higher-order multipoles, Phys. Rev. D **103**, 024042 (2021), arXiv:2007.04313 [gr-qc].

[127] A. G. Abac *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration), GW250114: Testing Hawking's Area Law and the Kerr Nature of Black Holes, Phys. Rev. Lett. **135**, 111403 (2025), arXiv:2509.08054 [gr-qc].

[128] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Tests of general relativity with binary black holes from the second LIGO–Virgo gravitational-wave transient catalog, Phys. Rev. D **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].

[129] E. M. Sänger *et al.*, Tests of General Relativity with GW230529: a neutron star merging with a lower mass-gap compact object, arXiv e-prints (2024), arXiv:2406.03568 [gr-qc].

[130] A. Pai and K. G. Arun, Singular value decomposition in parametrised tests of post-Newtonian theory, Classical Quantum Gravity **30**, 025011 (2013), arXiv:1207.1943 [gr-qc].

[131] M. Saleem, S. Datta, K. G. Arun, and B. S. Sathyaprakash, Parametrized tests of post-Newtonian theory using principal component analysis, Phys. Rev. D **105**, 084062 (2022), arXiv:2110.10147 [gr-qc].

[132] A. A. Shoom, P. K. Gupta, B. Krishnan, A. B. Nielsen, and C. D. Capano, Testing the post-Newtonian expansion with GW170817, Gen. Rel. Grav. **55**, 55 (2023), arXiv:2105.02191 [gr-qc].

[133] P. Mahapatra *et al.*, Confronting General Relativity with Principal Component Analysis: Simulations and Results from GWTC-3 Events, Phys. Rev. D **112**, 104007 (2025), arXiv:2508.06862 [gr-qc].

[134] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Testing the no-hair theorem with GW150914, Phys. Rev. Lett. **123**, 111102 (2019), arXiv:1905.00869 [gr-qc].

[135] M. Isi and W. M. Farr, Analyzing black-hole ringdowns, arXiv e-prints (2021), arXiv:2107.05609 [gr-qc].

[136] H. Siegel, M. Isi, and W. M. Farr, Analyzing black-hole ringdowns. II. Data conditioning, Phys. Rev. D **111**, 044070 (2025), arXiv:2410.02704 [gr-qc].

[137] S. Ma, K. Mitman, L. Sun, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, W. Throwe, N. L. Vu, and Y. Chen, Quasinormal-mode filters: A new approach to analyze the gravitational-wave ringdown of binary black-hole mergers, Phys. Rev. D **106**, 084036 (2022), arXiv:2207.10870 [gr-qc].

[138] S. Ma, L. Sun, and Y. Chen, Using rational filters to uncover the first ringdown overtone in GW150914, Phys. Rev. D **107**, 084010 (2023), arXiv:2301.06639 [gr-qc].

[139] S. Ma, L. Sun, and Y. Chen, Black Hole Spectroscopy by Mode Cleaning, Phys. Rev. Lett. **130**, 141401 (2023), arXiv:2301.06705 [gr-qc].

[140] N. Lu, S. Ma, O. J. Piccinni, L. Sun, and E. Finch, Statistical identification of ringdown modes with rational filters, Phys. Rev. D **112**, 064047 (2025), arXiv:2505.18560 [gr-qc].

[141] R. Brito, A. Buonanno, and V. Raymond, Black-hole Spectroscopy by Making Full Use of Gravitational-Wave Modeling, Phys. Rev. D **98**, 084038 (2018), arXiv:1805.00293 [gr-qc].

[142] A. Ghosh, R. Brito, and A. Buonanno, Constraints on quasinormal-mode frequencies with LIGO-Virgo binary–black-hole observations, Phys. Rev. D **103**, 124041 (2021), arXiv:2104.01906 [gr-qc].

[143] E. Maggio, H. O. Silva, A. Buonanno, and A. Ghosh, Tests of general relativity in the nonlinear regime: A parametrized plunge-merger-ringdown gravitational waveform model, Phys. Rev. D **108**, 024043 (2023), arXiv:2212.09655 [gr-qc].

[144] L. Pompili, E. Maggio, H. O. Silva, and A. Buonanno, Parametrized spin-precessing inspiral-merger-ringdown waveform model for tests of general relativity, Phys. Rev. D **111**, 124040 (2025), arXiv:2504.10130 [gr-qc].

[145] A. Toubiana, L. Pompili, A. Buonanno, J. R. Gair, and M. L. Katz, Measuring source properties and quasinormal mode frequencies of heavy massive black-hole binaries with LISA, Phys. Rev. D **109**, 104019 (2024), arXiv:2307.15086 [gr-qc].