Stepping stone toward detecting gravitational wave memory: A cumulative analysis with the full $(\ell = 2, m = 0)$ spherical harmonic using events from GWTC-4.0 and GWTC-5.0

Maria Rosselló-Sastre,[1,*] Sascha Husa,[2,1] Yumeng Xu,[1] Jorge Valencia,[1] Joan Llobera-Querol,[1] and Antoni Ramos-Buades[1]

[1]*Departament de Física, Universitat de les Illes Balears, IAC3—IEEC, Carretera de Valldemossa km 7.5, E-07122 Palma, Spain*

[2]*Institut de Ciències de l'Espai (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193 Cerdanyola del Vallès, Spain*

(Received 20 July 2026; accepted 18 September 2026; published 13 October 2026)

We perform Bayesian model selection to test for the presence of the $(\ell = 2, m = 0)$ spherical harmonic mode in gravitational wave events that have previously been identified as binary black hole mergers. As our signal model we use the quasicircular, nonprecessing `IMRPhenomTHM_20` waveform model, which includes the oscillatory and displacement memory contributions. Including the oscillatory component of the (2, 0) mode increases the signal-to-noise ratio and evidence for this mode, compared to testing only for the presence of gravitational wave memory. Our analysis thus constitutes a natural stepping stone toward detecting gravitational wave memory. We perform our analysis for the binary black hole signals identified in the GWTC-4.0 catalog, and for selected GWTC-5.0 events. In our Bayesian model comparison we find a cumulative $\log_{10}\mathcal{B} = 1.38 \pm 0.79$ in favor of the presence of the (2, 0) mode for the GWTC-4.0 catalog. We also stack the signal-to-noise ratio of the full (2, 0) mode and of its individual contributions, obtaining results consistent with previous studies and reaching $\mathrm{SNR}_{\mathrm{memory}} = 0.89^{+0.29}_{-0.11}$ after approximately 7.5 months of O4a observations. In addition, we study the precessing candidate GW241127_061008, and find no additional evidence for the (2, 0) mode when precession is included in `IMRPhenomTPHM_20`. Overall, our results provide an assessment of the observational support for the (2, 0) mode in current gravitational wave data and allow us to discuss prospects for its future detection. We find that decisive statistical evidence will likely require a larger catalog, with an optimistic estimated number of events of $N_{\mathrm{events}} = 166^{+82}_{-55}$, based on the specific assumptions adopted in this work. We also expect that decisive evidence will require a more extensive waveform systematics study.

DOI: 10.1103/brrl-jsb2

I. INTRODUCTION

Accurate inference of source parameters in gravitational wave (GW) data analysis require waveform models that capture as many measurable signal contributions as possible in the detector network. The emission of GWs is commonly described in terms of a decomposition into a sum of spin-weighted spherical harmonic modes (ℓ, m), which provide a natural basis for describing the angular structure of the radiation [1]. Here we are interested in compact binary coalescence, in particular of orbiting binary black hole (BBH) systems. In this case, the dominant harmonic is the $(\ell = 2, m = \pm 2)$ mode when the component spins are aligned with the orbital angular momentum. For misaligned binaries this statement more naturally holds in a coprecessing frame [2]. However, subdominant modes can play a significant role for asymmetric or highly inclined systems [3], where they help break degeneracies among intrinsic and extrinsic parameters and thereby enhance measurement accuracy [4–8]. Waveform models incorporating higher-order modes have been used systematically in the analysis of GW data since the third observing run (O3). To date, LIGO-Virgo-KAGRA (LVK) [9–11] observations have provided evidence for higher-order modes in the asymmetric events GW190412 and GW190814 [12–14]. These measurements are dominated by the $(\ell = 3, m = \pm 3)$ multipole, as analyses of both events found this to be the only measurable subdominant mode. In the GWTC-2.1 sample, no event exhibited measurable higher-order-mode content beyond $\ell = 3$, although a few additional events showed only marginal indications of the $(3, \pm 3)$ mode.

Here we are interested in the $(\ell = 2, m = 0)$ mode, which contains the main contribution from the nonoscillatory displacement memory [15–17], but also an oscillatory ringdown component. The memory arises from the cumulative emission of gravitational radiation throughout the

*Contact author: maria.rossello@uib.es

 © 2026 American Physical Society

evolution of the binary and manifests as a secular, permanent change between the initial and final asymptotic states of the spacetime. Unlike the oscillatory modes ($m \neq 0$), which encode the transient quasinormal relaxation of the remnant black hole, the displacement memory instead is sourced by the net flux of energy and momentum carried away by the GW emission. As such, it provides complementary information about the nonlinear dynamics of the coalescence and offers a direct signature of the strong-field, radiative regime of general relativity (GR) [18–20]. Moreover, the distinct dependence of the (2, 0) mode on the inclination angle, opposite to that of the dominant $(2, \pm 2)$ modes, helps break the distance–inclination degeneracy [21–23]. Including the oscillatory component of the (2, 0) mode increases the signal-to-noise ratio (SNR) for this mode (compare e.g., Ref. [24]), which will facilitate its detection and increases its relevance for parameter estimation (PE) of very high SNR events. Investigating the evidence for this mode thus constitutes a natural stepping stone for the detection of gravitational wave memory.

Motivated by these considerations and by the high number of GW events already observed by the LVK we perform a systematic reanalysis of BBH events detected during the first two parts of the fourth observing run (O4a and O4b), and published as the GWTC-4.0 [25–29] and GWTC-5.0 [30–34] editions of the LVK GW transient catalog. Building on the PE framework documented in our first analysis of the GWTC-4.0 catalog [35], we use the phenomenological waveform models to analyze the set of 84 BBH detections reported in GWTC-4.0 and a selection of 6 high significance BBH events reported in GWTC-5.0. They are chosen based on high network SNR values, or high effective spin parameter, $|\chi_{\rm eff}|$ [32,36,37], which enhances the (2, 0) mode, especially the amplitude of the memory contribution in the case of high positive spin components.

For our analysis we are using the `IMRPhenomT` family of time-domain phenomenological waveform models, restricted to the quasicircular sector [38–40]. The `IMRPhenomT` family has been extended to include the $(\ell = 2, m = 0)$ spherical harmonic mode with the oscillatory ringdown and the displacement memory contributions `IMRPhenomTHM_20` [24] for nonprecessing, quasicircular BBHs; and further to spin precession `IMRPhenomTPHM_20` [41].

We quantify the evidence for the presence of the (2, 0) mode in individual events via Bayesian model comparison, assessing whether including the mode yields a statistically significant improved description of the observed data, and evaluate its cumulative detectability across the population by stacking event level statistics and computing a population-level SNR. In contrast to approaches that isolate only the displacement memory contribution, we perform a per event comparison between complete inspiral–merger–ringdown phenomenological models with and without the full (2, 0) mode, using O4 data and incorporating measured power spectral densities (PSDs) and calibration uncertainties. The resulting Bayes factors are then combined across events.

The paper is organized as follows. We start with a review of detectability prospects for GW memory in Sec. II. In Sec. III, we briefly describe the waveform model used in this work. In Sec. IV, we outline the treatment of the data, and the methods used in the analysis. In Sec. V, we present the results of the reanalysis of the dataset using the `IMRPhenomTHM_20` model to include the $(\ell = 2, m = 0)$ mode. Finally, we summarize our findings and conclude in Sec. VI.

Throughout this paper, component masses are denoted by m_i. We define the mass ratios $q = m_2/m_1 \leq 1$ and $Q = 1/q$, and the total mass $M = m_1 + m_2$. Masses are reported in the detector frame. The components of the dimensionless spin vectors in the direction of the orbital momentum (our z-axis orthogonal to the orbital plane) are denoted by $\chi_i = S_i^z/m_i^2$.

II. PROSPECTS FOR GW MEMORY DETECTABILITY

The detectability of GW memory has been the subject of extensive investigation, spanning analytical estimates, population-level forecasts, and Bayesian inference studies. Early observational analyses searched for memory signatures in existing GW catalogs, including GWTC-1 [42,43], GWTC-2 [44,45], and GWTC-3 [46], but did not yield conclusive detections. Foundational work established that detection with second-generation detectors is most naturally achieved through stacking multiple events. In particular, [47] showed that hundreds of GW150914-like signals at Advanced LIGO design sensitivity would be required for a confident detection using frequentist SNR Subsequent population-based forecasts, incorporating updated BBH distributions and detector sensitivities, predicted cumulative SNRs of order ∼3 after several years of observation, reinforcing the conclusion that detection is more likely through population stacking than via individual events [48]. More studies, including the effects of spin memory and next generation detector networks, have further refined this picture; while a second generation network operating at O4/O5 sensitivities may achieve population level detection of displacement memory, third generation observatories such as Cosmic Explorer [49] are expected to enable detections in individual events. In contrast, the detection of spin memory is likely to still rely on stacking even in these advanced scenarios [50]. In our previous work [51], we studied the parameter biases obtained when neglecting the presence of the full (2, 0) mode, in future ground-based detectors, finding important biases at sufficiently high SNRs.

In a recent work [52], the authors use hierarchical Bayesian inference on the GWTC-5.0 catalog of BBH observations to constrain the memory enhancement factor across the population, finding a value consistent with the

GR prediction and forecasting that roughly 2000 detections would be needed to measure a nonzero memory signal at the 1σ level. Complementary analyses have explored the detectability of memory at the level of individual events with current detectors. In particular, [53] performed Bayesian model selection studies for signals observed with O4 sensitivity and third generation detectors. For an optimally oriented (edge-on) injection with a network SNR of 44.7 at O4 sensitivity, repeated injection-recovery analyses across multiple noise realizations yielded Bayes factors clustered around zero, remaining below detection thresholds and providing no statistically significant preference for either the memory or no memory hypothesis.

In the space-based context, early theoretical studies also highlighted the strong potential of LISA [54] for per event memory detection in massive black hole binaries (MBHBs) [15,16]. This expectation has been confirmed and substantially refined by recent work. Using time domain modeling of the LISA response, [55] demonstrated that the memory signal induces a resolvable, nonoscillatory component in the strain, with SNRs sufficient for detection in a subset of MBHB mergers over the mission lifetime. Building on this, [56] carried out a full Bayesian parameter estimation analysis, showing that the memory amplitude can be constrained with meaningful precision and that waveform models including memory are robustly supported by the data. Finally, [57] developed a rigorous framework for memory detection, introducing parametrized descriptions of the memory growth and establishing hypothesis-testing strategies based on model selection, thereby defining the conditions for confident detection. In contrast, searches for bursts with memory in pulsar timing array data have thus far yielded only upper limits. For example, analyses of the NANOGrav 12.5-year dataset report no statistically significant evidence for memory signals, with Bayes factors of ~ 2.8 for models including memory in addition to a stochastic GW background [58,59].

III. WAVEFORM MODEL

The `IMRPhenomTHM` family of time domain phenomenological models [24,38–41,60] offers an alternative to the computationally efficient frequency domain models. The flexibility of modeling effects in the time domain facilitates the inclusion of complicated new physical effects such as spin-precession [40], orbital eccentricity [60], or GW memory [24]. For this analysis, we employ the quasicircular, nonprecessing-spin `IMRPhenomTHM` model [38,39], which has been calibrated to numerical relativity (NR) and test-particle waveforms and it includes $(\ell, m) = \{(2, \pm 2), (2, \pm 1), (3, \pm 3), (4, \pm 4), (5, \pm 5)\}$ multipoles; and the quasicircular precessing-spin `IMRPhenomTPHM` model [40], which is based on the "twisting-up" approximation [61] and has the same mode content in the coprecessing frame as the `IMRPhenomTHM` model.

To incorporate the memory effect in this analysis, we employ the phenomenological waveform model of the $(\ell = 2, m = 0)$ spherical harmonic mode, which captures both the dominant displacement memory contribution and the quasinormal mode (QNM) oscillations of the ringdown in this mode [24]. This model is constructed in the time domain by fitting phenomenological *Ansätze* to NR simulations and memory waveforms computed using the Bondi-Metzner-Sachs (BMS) balance laws [62,63]. It is valid for BBH systems in quasicircular orbits, covering mass ratios up to $Q = 10$ and spin components aligned with the orbital angular momentum, $\chi_1, \chi_2 \in [-1, 1]$. In order to extend our description of the (2, 0) mode to include spin-precession, we apply the "twisting-up" approximation to the aligned-spin model of the (2, 0) mode and we include the displacement memory contribution in all the $\ell = 2$ modes [41]. Both the aligned-spin and the precessing models are implemented within `IMRPhenomTHM` and `IMRPhenomTPHM`, respectively, in the Python package PHENOMXPY [64].

IV. METHODS

The strain data are taken from the Gravitational Wave Open Science Center (GWOSC) [65] using the GWpy package [66] with sample rate 4096 Hz. The data conditioning procedures, including the duration of the analyzed segments, the treatment of instrumental glitches, and the choice of the starting frequency used in the likelihood evaluation, follow the methodology described in the original analysis presented in [35]. For completeness, we refer the reader in particular to Sec. III B of this reference, as the inference framework and analysis configuration adopted in this work are identical to those employed in that study. The prior distributions used in the analysis are chosen to reflect standard astrophysical assumptions. Uniform priors are imposed on the redshifted component masses, the merger time, and the coalescence phase. The orientation of the binary and its sky location are assigned isotropic priors. For the luminosity distance, we adopt a prior corresponding to a constant merger rate density in comoving volume and source-frame time. For the spin components, we use Bilby's [67] `AlignedSpin` with spin magnitudes uniformly distributed between 0 and 0.99 and spins constrained to be aligned with the orbital angular momentum.

Note that time domain waveforms which contain the displacement memory contribution are nonperiodic signals. We need to take this into account when converting the waveforms to frequency domain. For this, we use the Fourier transform method `gw-foutstep` [68], which decomposes the waveform into a known sigmoidlike component (capturing the persistent offset and transformable analytically) plus a residual transient part that can be safely Fourier transformed numerically without artifacts. This method avoids the spurious artifacts introduced by standard Fast Fourier Transform techniques applied to nonperiodic signals.

Because the present study requires the execution of multiple PE runs, minimizing manual interaction is essential in order to reduce the possibility of human error and

ensure reproducibility. To this end, we employ the automated workflow tool PEAutomator [69], which manages job submission, monitoring, and postprocessing of the results.

A. Setup and hypotheses

Our goal is to assess the presence of the $(\ell = 2, m = 0)$ mode in the observed signals. This mode contains two physically distinct contributions: an oscillatory component associated with the QNM ringdown of the remnant black hole, and the displacement memory contribution. To quantify the evidence for this mode, we perform a paired Bayesian model comparison analysis for each event [70]. We consider two hypotheses:

(i) $\mathcal{H}_{20}$: `IMRPhenomTHM_20` model with the (2, 0) mode (oscillatory ringdown + displacement memory); and
(ii) $\mathcal{H}_{\varnothing}$: `IMRPhenomTHM` model without the (2, 0) mode.

For a given event i, we quantify the relative evidence for the presence of the mode through the Bayes factor,

$$\mathcal{B}_i \equiv \frac{p(d_i|\mathcal{H}_{20})}{p(d_i|\mathcal{H}_{\varnothing})},$$
$$\log_{10}\mathcal{B}_i = \log_{10} Z_i(\mathcal{H}_{20}) - \log_{10} Z_i(\mathcal{H}_{\varnothing}), \quad (4.1)$$

where $Z_i(\mathcal{H})$ denotes the Bayesian evidence for the hypotheses $\mathcal{H}$ computed from the data d_i of the corresponding event. The uncertainty on the log-10 Bayes factor is propagated from the uncertainties in the evidence via

$$\sigma[\log_{10}\mathcal{B}_i] = \sqrt{\sigma^2[\log_{10} Z_i(\mathcal{H}_{20})] + \sigma^2[\log_{10} Z_i(\mathcal{H}_{\varnothing})]}. \quad (4.2)$$

The evidence is computed using the DYNESTY nested-sampling algorithm as implemented in Bilby [67]. To assess the cumulative evidence for the (2, 0) mode across the population of detected events, we combine the individual Bayes factors assuming statistical independence between events. Under this assumption, the stacked Bayes factor is given by the sum of the Bayes factor logarithms,

$$\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} = \sum_{i=1}^{N} \log_{10}\mathcal{B}_i, \quad (4.3)$$

where N denotes the number of analyzed events. The corresponding uncertainty is obtained by adding the individual uncertainties in quadrature,

$$\sigma[\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}] = \sqrt{\sum_{i=1}^{N} \sigma^2[\log_{10}\mathcal{B}_i]}. \quad (4.4)$$

To interpret the Bayes factor results, we adopt the Jeffreys' scale [71], which provides a qualitative classification of the strength of evidence in favor of a given hypothesis. According to this scale, Bayes factors with $|\log_{10}\mathcal{B}| < 0.5$ are generally considered inconclusive, corresponding to evidence that is not worth more than a bare mention. Values in the range $0.5 \leq |\log_{10}\mathcal{B}| < 1$ are interpreted as substantial evidence, while $1 \leq |\log_{10}\mathcal{B}| < 2$ indicate strong evidence in favor of one model over the other. Finally, $|\log_{10}\mathcal{B}| \geq 2$ is typically regarded as decisive evidence. This classification applies symmetrically to negative values of $\log_{10}\mathcal{B}$, in which case the same levels of evidence favor the alternative hypothesis. Note that here we directly work with the Bayes factor instead of an odds ratio, since the parameter space that corresponds to the two models we are using for model selection is the same, and due to the smallness of the contribution of the (2, 0) mode to the signal no significant effect on detection is expected, and we conclude from our results that there is also no significant effect on the estimated parameters.

B. Stacking signal-to-noise ratio

As an additional diagnostic, we compute the SNR associated with the (2, 0) mode and use it as a proxy for the strength of the evidence obtained from the Bayesian model comparison. This approach allows direct comparison with the methodology adopted in [48,50]. In particular, we evaluate the SNR for three cases: the complete (2, 0) mode including both the oscillatory and displacement memory contributions; the displacement memory contribution alone; and the oscillatory component alone. For a set of independent events, the combined SNR is obtained by summing the individual SNRs in quadrature,

$$\rho^2 = \sum_i \rho_i^2, \quad (4.5)$$

where the SNR for a signal $h(t)$ is computed in frequency domain as

$$\rho = \sqrt{\langle h|h\rangle} = \sqrt{4\int_0^{\infty} \frac{|\tilde{h}(f)|^2}{S_n(f)} df}, \quad (4.6)$$

with $S_n(f)$ the detector's PSD.

We restrict the (2, 0) mode SNR stacking analysis to the O4a event sample, excluding events selected from O4b in order to avoid introducing selection biases, since the O4b subset was not chosen in an unbiased or complete manner but rather according to specific selection criteria. Results are reported in Sec. V B, where we find consistency with the predictions in [48,50].

C. Bayes factor's scaling with the SNR

In order to compare the strength of the evidence across different PE analyses in a consistent way, it is useful to estimate how the Bayes factor changes as a function of the SNR. In the high-SNR approximation [72], and assuming that the two analyses are performed with the same priors and the same waveform mismatch, the Bayes factor is expected to scale as

$$\log_{10}\mathcal{B} \propto \frac{\rho^2}{2}(1-\mathrm{FF}^2), \tag{4.7}$$

where ρ is the SNR and FF is the fitting factor between the two waveforms, which quantifies the proximity of the target signal to a template manifold by measuring the reduction in recovered SNR when the signal is filtered with the best matching template from that manifold [73]. This allows us to rescale the Bayes factor measured for one run to estimate its expected value, and uncertainty, for another run with a different SNR. This scaling provides a simple way to extrapolate the significance of the results obtained for the louder events. Computing the ratio between two runs, assuming the same value for FF and that we use the same priors in both injections, we can compute the Bayes factor and its associated error as

$$\log_{10}\mathcal{B} = \left(\frac{\rho}{\rho_0}\right)^2 \log_{10}\mathcal{B}_0, \qquad \sigma = \left(\frac{\rho}{\rho_0}\right)^2 \sigma_0. \tag{4.8}$$

With this expression we can easily check the scaling of the Bayes factor with the SNR. This is presented in Sec. IV E using a set of PE, zero-noise model injections, for which we obtain results consistent with the expected scaling of these quantities.

D. Sampler settings

As previously mentioned, we employ the DYNESTY [74] sampler as implemented in Bilby [67]. We use the acceptance-walk method for the Markov Chain Monte Carlo evolution with `naccept=60` and `nlive=1000` with two independent seeds for each run. The stopping criterion of the sampler is controlled by the parameter $\Delta\ln Z$, which sets the threshold for the estimated remaining contribution of the unexplored prior volume to the total evidence. The sampler terminates once this residual contribution falls below the chosen tolerance. Therefore, reducing $\Delta\ln Z$ enforces a stricter convergence requirement, drives the nested sampling run further into the high likelihood region before stopping, and decreases the residual uncertainty in the final estimate of $\ln Z$. The default value in Bilby is set to $\Delta\ln Z = 0.10$, however here we adopt $\Delta\ln Z = 0.05$ in order to obtain better-converged evidence estimates. Since the obtained Bayes factors are close to zero and, in most cases, compatible with zero within the statistical uncertainty, the evidence difference between the competing hypotheses is small and a more accurate determination of the residual error is desirable. A tighter stopping criterion yields smaller numerical uncertainties and more trustworthy Bayes factors at the expense of increased computational cost. However, the total wall time is not deterministic, since the early stages of the algorithm involve stochastic exploration of the parameter space and the number of iterations required to reach convergence can vary between runs, even for identical settings. Consequently, the observed runtime reflects both the systematic effect of the stopping criterion and the intrinsic variability of the sampling process, which can mask any strictly monotonic dependence on $\Delta\ln Z$. In Appendix A, we provide a more detailed discussion of the runtimes of all the events analyzed with both values of $\Delta\log Z$ and each model.

E. Parameter estimation injection studies

To validate our setup we perform a small set of injection and recovery studies.

1. Numerical relativity injection

First we perform a zero-noise NR injection using a quasicircular, nonprecessing, waveform from the Simulating eXtreme Spacetimes (SXS) public Catalog [75–77], SXS:BBH:2497 ($Q = 2, \chi_1 = \chi_2 = 0, N_{\mathrm{orbits}} = 20$). The simulations in the latest version of this catalog include the memory contribution, which is introduced in the extrapolated waveforms using the BMS balance laws, as explained in [77]. We choose a nonequal mass simulation so that all the modes in the model have a nonzero contribution to the waveform. In the injected waveform, we include the modes $(\ell, m) = \{(2,\pm2),(2,\pm1),(2,0),(3,\pm3),(4,\pm4),(5,\pm5)\}$ to match the mode content in `IMRPhenomTHM`. We recover the injected parameters with the model with and without the (2, 0) mode to be able to compute the Bayes factor between the two recoveries. We choose the same mode content in order to be able to see the biases that are introduced in the parameters only due to the omission of the (2, 0) mode. We select an inclination of $\pi/2$, as this is the optimal configuration for the (2, 0) mode and the network (H1 and L1) SNR of the injected signal is 37.6, with a (2, 0) mode SNR of 0.98. We pick the values in this range to keep the analysis realistic and comparable to the SNR values observed in the O4a and O4b observing runs. The injected parameters for this injection are listed in Table I. The posterior distributions are shown in Appendix B. At this SNR and for this set of source parameters we observe no appreciable bias in the recovered intrinsic or extrinsic parameters when the (2, 0) mode is omitted from the recovery model. The Bayes factor between the two hypotheses in this case is $\log_{10}\mathcal{B} = 0.48 \pm 0.12$, indicating only weak support for the model including the mode. These findings are consistent with our results for the O4a and O4b data; in both

cases, omitting the (2, 0) mode does not lead to noticeable biases in the posterior distributions, and the Bayes factors remain of comparable magnitude.

2. Model injections

Furthermore, to test the consistency of the Bayes factor with the expected SNR scaling in the PE runs that have been performed in this work, we perform two zero-noise model injections: one at considerably high SNR and one at lower SNR, to check the scaling of the Bayes factor as a function of the SNR introduced in Sec. IV C. We choose model rather than NR injections for this purpose, in order to avoid waveform systematics between the injected waveform and the model used for recovery. First, we inject a signal at a high SNR of 435.3 and we compute the Bayes factor between the recovery with and without the (2, 0) mode in the model, obtaining the result: $\log_{10}\mathcal{B} = 26.06 \pm 0.24$. Then, we perform the same injection at a much lower SNR of 15.3 and we compare the obtained Bayes factor, $\log_{10}\mathcal{B} = 0.08 \pm 0.16$ with the analytical prediction based on Eq. (4.7). Using the expression in Eq. (4.8), with $\rho = 15.3, \rho_0 = 435.3, \log_{10}\mathcal{B}_0 = 26.06, \sigma_0 = 0.24$, we obtain $\log_{10}\mathcal{B} \pm \sigma = 0.03209 \pm 0.00030$. For these injection runs, however, we use the standard Bilby stopping criterion of $\Delta \ln Z = 0.10$, rather than the more stringent $\Delta \ln Z = 0.05$ used for the PE analyses presented in the rest of this work. Therefore, we conservatively take $\Delta \ln Z = 0.10$ as the minimum uncertainty associated with the sampler, yielding a final result of $\log_{10}\mathcal{B} \pm \sigma = 0.03 \pm 0.10$ which is consistent with the result of the actual PE injection; $\log_{10}\mathcal{B} = 0.08 \pm 0.16$. The parameters used for these injections are shown in Table I and the results are summarized in Table II.

In summary, at the low SNR considered here, there is no robust evidence for the presence of the (2, 0) mode. We emphasize that these zero-noise model injections only quantify modeling and sampling behavior. In real detector data, noise realizations add an additional stochastic contribution to the evidence. For Bayes factors of order unity or smaller such noise fluctuations can change the sign of $\log_{10}\mathcal{B}$ or produce values that are consistent with zero within the uncertainty, so small measured Bayes factors should be interpreted cautiously.

The purpose of these zero-noise model injections is to validate the consistency of the waveform models and sampling setup, in particular the expected scaling of the Bayes factor with SNR, rather than to provide an additional ingredient for the population level forecasts presented in Sec. V D. The results therefore serve as a validation of the PE methodology used in this work and are not used directly in the subsequent detectability estimates.

TABLE I. Parameters of the NR injection (middle column) and model injections (right column). The only difference between the high and low SNR model injections is the luminosity distance, which is quoted in brackets for the low SNR injection. In all cases, we set the minimum frequency of the analysis to $f_{\rm min} = f_{\rm ref} = 20$ Hz.

Parameter	NR injection	Model injection high (low) SNR
q	2.00	1.03
$M[M_\odot]$	50	100
d_L [Mpc]	400	140 (4000)
χ_1	0.0	0.2
χ_2	0.0	0.2
ι [rad]	$\pi/2$	$\pi/2$
ϕ [rad]	1.20	5.43
ψ [rad]	1.26	2.70
α [rad]	2.60	2.77
δ [rad]	0.53	−0.50

TABLE II. Bayes factor scaling for two model injections at different SNR compared with the analytical estimate with Eq. (4.7).

	SNR	$\log_{10}\mathcal{B} \pm 1\sigma$
High SNR injection	435.3	26.06 ± 0.24
Low SNR injection	15.3	0.08 ± 0.16
Analytical value for low SNR injection	⋯	0.03 ± 0.10

V. RESULTS

We perform an analysis of the BBH events in GWTC-4.0 and the set of six events from GWTC-5.0 [32] listed here (quoting the results obtained by the LVK analysis):

(i) GW240920_124024 is a high SNR $= 37.0^{+0.1}_{-0.1}$ event, with some support for spin precession, $\chi_{\rm p} = 0.47^{+0.44}_{-0.30}$.
(ii) GW240925_005809 [78] is a high SNR $= 31.9^{+0.1}_{-0.1}$ event, the posterior however shows a bimodality in the inferred redshift.
(iii) GW241011_233834 [37] is another high SNR $= 35.8^{+0.1}_{-0.1}$ event. It shows unequal masses, high spins $\chi_{\rm eff} = 0.50^{+0.06}_{-0.04}$, and is the closest event among the candidates reported in GWTC-5.0, $d_L = 210^{+40}_{-40}$ Mpc.
(iv) GW241110_124123 [37] is a low SNR $= 9.9^{+0.3}_{-0.5}$ event, with unequal masses and high spins $\chi_{\rm eff} = -0.31^{+0.23}_{-0.18}$.
(v) GW241127_061008 has high SNR $= 31.2^{+0.1}_{-0.2}$, unequal masses, and support for spin precession $\chi_{\rm p} = 0.57^{+0.14}_{-0.21}$.
(vi) GW250114_082203 [36] has the highest SNR among observed GW events to date, SNR $= 76.9^{+0.0}_{-0.1}$, providing a particularly interesting case for searching for subdominant features in the waveforms.

We use the `IMRPhenomTHM_20` and `IMRPhenomTHM` waveform models in order to investigate the presence of the $(\ell = 2, m = 0)$ mode in the data. For the event GW241127_061008, we also compare

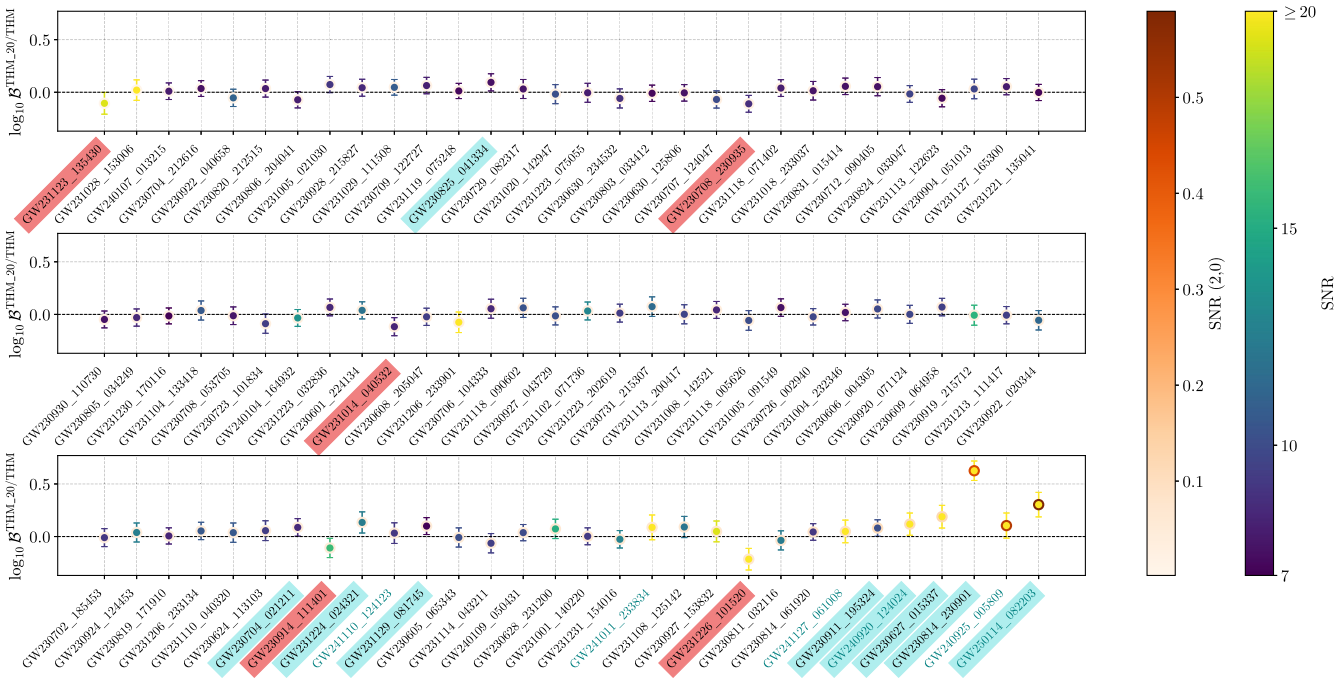


FIG. 1. Log-10 Bayes factors for each individual event in increasing SNR of the (2, 0) mode order. Each point corresponds to the value of $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}}$ with 1σ sampling uncertainty shown as vertical error bars. The color filling the markers represents the network SNR of the full signal for each event, as indicated by the rightmost colorbar; while the color of the border of the marker represents the network SNR of only the (2, 0) mode, as indicated in the leftmost colorbar. The horizontal dashed line marks the reference value $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} = 0$, corresponding to no preference between the models. The events with $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} - 1\sigma > 10^{-3}$ are highlighted in blue and the ones with $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} + 1\sigma < -10^{-3}$, in red. The events from GWTC-5.0 are marked with the color of the font label in blue.

`IMRPhenomTPHM_20` and `IMRPhenomTPHM` since its χ_{p} value is significantly different from zero. We do not use precessing waveform models for the full event set because they are more computationally expensive than the aligned spin versions, and the majority of events show no decisive evidence for precession when analyzed with the waveform models included in GWTC-4.0 and GWTC-5.0.

A. Bayes factors

Using Eqs. (4.1) and (4.2), we compute the Bayes factors and their associated uncertainties for all tested hypotheses across the full set of considered events. Figure 1 shows the individual values of $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}}$ for each BBH event in GWTC-4.0 and for the selected events from GWTC-5.0 (marked by the blue font color). The events are ordered by increasing SNR of the (2, 0) mode, which is encoded by the color of the marker edges, while the marker fill color represents the SNR of the full signal. Overall, the results indicate that for the majority of events the data do not provide strong evidence in favor of including the (2, 0) mode. Most measurements cluster around zero and their associated 1σ uncertainties typically overlap the null value, implying that the two waveform models are statistically indistinguishable. A mild tendency toward positive values is observed at increasing SNR, suggesting that the contribution of the (2, 0) mode may become marginally more relevant in louder signals, although the inferred Bayes factors remain small and correspond to weak evidence.

The labels highlighted in blue (red) correspond to events with $\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} - 1\sigma > 10^{-3}$ ($\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} + 1\sigma < -10^{-3}$), i.e., those showing the most robust, although still weak, preference for the inclusion (exclusion) of the (2, 0) mode; their values are reported in Table III. We adopt 10^{-3} rather than 0 to introduce a margin above the null evidence threshold $\mathcal{B} = 1$, thereby avoiding the classification of marginal cases as positive due to numerical precision limits or fluctuations induced by the sampling procedure. The highest evidence for the presence of the (2, 0) mode is found for GW230814_230901 [79], with substantial evidence according to Jeffreys' scale ($\log_{10} \mathcal{B}^{\mathrm{THM_20/THM}} = 0.63 \pm 0.09$), although, as it will be explained later in the text, this result can be affected by waveform systematics. Consistently, the posterior distributions presented in Appendix C show that, for all events, the inferred parameters are largely unaffected by the inclusion of the (2, 0) mode. The detectability of this mode is not solely driven by signal strength but also depends on additional source properties, such as the binary orientation and intrinsic parameters. To assess the agreement between the two waveform models and the data, in Appendix C we also present the

TABLE III. List of events with log-10 Bayes factors above the selection threshold $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} - 1\sigma > 10^{-3}$ and below $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} + 1\sigma < -10^{-3}$. We include the SNR obtained with the `IMRPhenomTHM_20` model for the full signal and for the (2, 0) mode, and the median value of the inclination angle. The quoted uncertainties represent one standard deviation. In Fig. 9 we show the posteriors for some of the parameters of the nine events with positive log-10 Bayes factors.

Event	SNR	$\mathrm{SNR}_{(2,0)}$	θ_{JN}[rad]	$\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}$
GW230814_230901	$41.6^{+1.0}_{-1.0}$	$0.44^{+0.39}_{-0.28}$	$1.38^{+1.16}_{-0.79}$	0.63 ± 0.09
GW250114_082203	$75.5^{+1.0}_{-1.0}$	$0.59^{+0.26}_{-0.23}$	$0.91^{+1.48}_{-0.21}$	0.30 ± 0.12
GW230627_015337	$28.1^{+1.1}_{-1.1}$	$0.49^{+0.30}_{-0.45}$	$1.27^{+1.30}_{-0.76}$	0.19 ± 0.11
GW231224_024321	$12.4^{+1.1}_{-1.0}$	$0.05^{+0.15}_{-0.04}$	$1.53^{+1.10}_{-1.03}$	0.13 ± 0.10
GW240920_124024	$36.8^{+1.0}_{-1.0}$	$0.10^{+0.15}_{-0.07}$	$0.96^{+1.62}_{-0.44}$	0.12 ± 0.11
GW231129_081745	$7.1^{+1.1}_{-1.1}$	$0.05^{+0.07}_{-0.04}$	$1.57^{+0.81}_{-0.80}$	0.10 ± 0.08
GW230825_041334	$7.6^{+1.0}_{-1.1}$	$0.02^{+0.05}_{-0.02}$	$1.68^{+0.86}_{-1.06}$	0.09 ± 0.08
GW230704_021211	$8.6^{+1.0}_{-1.1}$	$0.05^{+0.10}_{-0.04}$	$1.38^{+1.09}_{-0.79}$	0.09 ± 0.08
GW230911_195324	$10.2^{+1.0}_{-1.1}$	$0.10^{+0.07}_{-0.08}$	$1.55^{+0.78}_{-0.75}$	0.08 ± 0.08
GW231123_135430	$19.3^{+1.0}_{-1.0}$	$0.00^{+0.0}_{-0.01}$	$0.41^{+2.31}_{-0.14}$	-0.11 ± 0.10
GW230914_111401	$15.8^{+1.0}_{-1.0}$	$0.05^{+0.08}_{-0.04}$	$1.36^{+1.15}_{-0.75}$	-0.11 ± 0.09
GW230708_230935	$8.7^{+1.0}_{-1.0}$	$0.03^{+0.07}_{-0.03}$	$2.08^{+0.56}_{-1.22}$	-0.11 ± 0.08
GW231014_040532	$8.2^{+1.0}_{-1.0}$	$0.03^{+0.09}_{-0.03}$	$1.47^{+1.14}_{-0.96}$	-0.12 ± 0.09
GW231226_101520	$33.3^{+1.1}_{-1.1}$	$0.07^{+0.15}_{-0.06}$	$0.89^{+1.76}_{-0.47}$	-0.22 ± 0.10

whitened strain for the two highest SNR events in the dataset; GW230814_230901 and GW250114_082203.

In our previous reanalysis of GWTC-4.0 [35], we identified systematic differences between `IMRPhenomTPHM` and `IMRPhenomXPNR` [80] for the event GW230814_230901, which arise from their different mode content. In particular, these discrepancies can be traced to the absence of the $(3, \pm 2)$ mode in the time domain phenomenological model. See Fig. 5 of Ref. [35] for the systematics in the mass ratio-inclination parameters across `IMRPhenomXPNR`, `IMRPhenomTPHM`, and `SEOBNRv5PHM` [81–83], which also includes the $(3, \pm 2)$ mode. This suggests that an important improvement to these waveform models would be the inclusion of the $(3, \pm 2)$ multipole contribution. These differences in the posteriors can also be seen in the top left panel of Fig. 9 in Appendix C, where the mixed posterior samples from GWTC-4.0 are plotted together with the results from the `IMRPhenomTHM` and `IMRPhenomTHM_20` runs. To assess whether `IMRPhenomTHM_20` remains preferred over `IMRPhenomXHM` [aligned-spin, frequency domain model which includes the $(3, \pm 2)$ mode] under a consistent comparison, we perform a PE analysis using identical settings with the latter model. We obtain $\log_{10}\mathcal{B}^{\mathrm{THM_20/XHM}} = 0.18 \pm 0.10$. While the evidence is 3.5 times lower than that obtained in the comparison against `IMRPhenomTHM`, the result still provides mild support for the presence of the (2, 0) mode in the data.

As expected, these results indicate that the (2, 0) mode is too weak to yield statistically significant evidence for its presence in individual events at the current sensitivity. We therefore perform a combined analysis by stacking the Bayes factors over all GWTC-4.0 BBH events. We restrict this calculation to the GWTC-4.0 sample to avoid selection effects associated with including only the events selected for analysis from GWTC-5.0. This approach enhances the overall sensitivity to a common signal component by coherently combining the individual evidence, allowing us to assess whether the data collectively support the presence of this mode beyond statistical fluctuations. The cumulative evidence trajectory is displayed in Fig. 2. The events are sorted in growing network SNR order, with the corresponding value represented in the color bar. The shaded region shows the 1σ uncertainty accumulated over all the events.

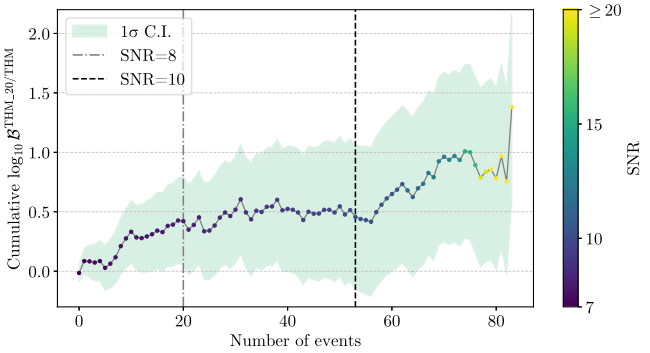


FIG. 2. Cumulative sum of log-10 Bayes factors as a function of the number of events, in increasing SNR order for the BBH events in GWTC-4.0. The shaded region shows the propagated 1σ uncertainty and the color scale indicates the SNR of each event. The vertical lines indicate where the SNR thresholds of 8 and 10 are located, respectively.

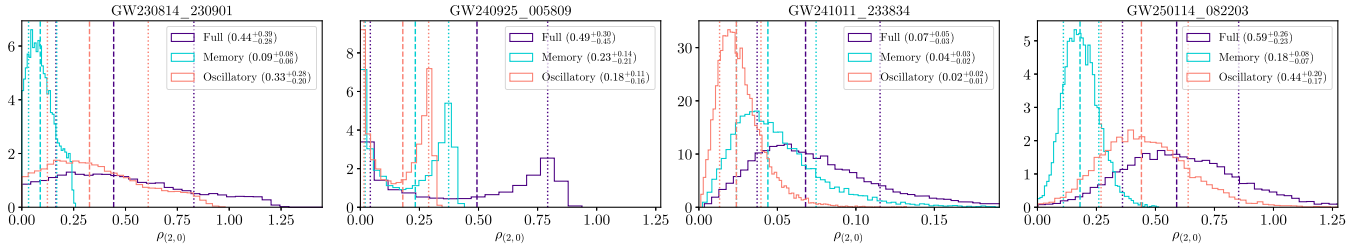


FIG. 3. Posterior distributions of the SNR of the (2, 0) mode for four selected events. Each curve represents the distribution obtained with the different components in this mode. The numbers in the legend show the median, 16% and 84% percentiles of the distributions, denoted by the vertical lines in the plots (dashed-median, dotted-percentiles).

The cumulative trajectory reaches a final value of $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} = 1.38 \pm 0.79$. Accounting for the uncertainty, the result does not provide significant evidence for the presence of the (2, 0) mode in the GWTC-4.0 sample. The high SNR events have the major contribution to the cumulative evidence. Although low SNR events contribute less to the cumulative Bayes factor than higher SNR events, the cumulative evidence continues to grow as additional events are included, indicating that the signal is not driven solely by a small number of high-significance detections. Most detected events are compatible with face-on orientations of the binaries, where the dominant quadrupolar modes are strongest and the (2, 0) mode effectively vanishes. Overall, these results are consistent with previous studies under current detector sensitivity, as the (2, 0) mode represents a subdominant contribution relative to the primary signal, producing a weak per event imprint whose sign and magnitude depend on the SNR, and especially on parameters such as the inclination, the mass ratio and the spin components.

B. Stacking signal-to-noise ratio

To compute the SNR of the (2, 0) mode for each event, we take the posterior samples of the parameters, we generate the corresponding waveform for each sample and compute the SNR of the mode, so that we obtain a posterior distribution of this quantity. In Fig. 3, we show the posteriors for four events; GW230814_230901, GW240925_005809, GW241011_233834 and GW250114_082203. We do the same for the three options in the (2, 0) mode, so including only the displacement memory component, only the oscillatory component and both contributions. For most cases in the whole catalog, the median SNR of only the displacement memory is the lowest among the three contributions. This is not the case for GW240925_005809 and GW241011_233834. GW240925_005809 shows a bimodal distribution for the three posteriors. This bimodality is consistent with bimodality in the inferred redshift reported in GWTC-5.0 [32,78]. As expected, the median for the full mode is the highest in all the events. For the rest of events not shown here, the SNR distributions rail against zero. However, for these four high SNR events, the posteriors deviate from zero, especially for GW250114_082203, for which the median values of the three distributions are larger than zero, meaning that for this event the (2, 0) mode SNR is different from zero, but still below 1.

Once we have the estimation of this quantity for the three contributions and all events, we perform a stacking of the SNR using Eq. (4.5). The curves obtained for the three contributions and all the events in GWTC-4.0 are shown in Fig. 4 as a function of time (in years). For reference, we add a horizontal line at the threshold SNR = 3, which is the threshold for memory used in [48,50] and that we also adopt for comparison. This value is surpassed for the full (2, 0) mode, but not reached when we consider the individual components of this mode. The final values obtained from this calculation are summarized in Table IV, where we also include the final SNR reached when considering the selected events from GWTC-5.0.

We now compare the light blue curve in Fig. 4 [SNR of the memory in the (2, 0)] with the results presented in:

(i) Figure 4 from [48]; the curves show the total memory SNR with a detector network composed by Advanced LIGO and Advanced Virgo at design sensitivity.

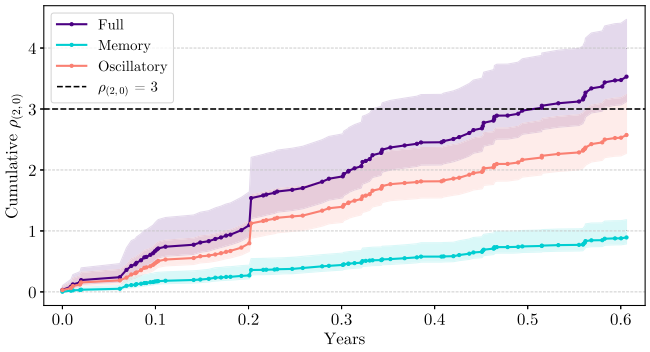


FIG. 4. Cumulative sum of the SNR of the (2, 0) mode for the different contributions: the full mode, only the displacement memory and only the oscillatory components as a function of the time in years including only the events from GWTC-4.0. The horizontal black, dashed line shows the threshold of SNR = 3.

TABLE IV. Final values of the cumulative SNR of the (2, 0) mode for each of the contributions in this mode and the full signal. Top row corresponds to the case where we only consider the events in GWTC-4.0 (see Fig. 4), and the bottom row shows the results when adding the selected events from GWTC-5.0.

Final value of (2, 0) SNR stacking	Full	Oscillatory	Memory
GWTC-4.0 events	$3.53^{+0.94}_{-0.41}$	$2.57^{+0.66}_{-0.30}$	$0.89^{+0.29}_{-0.11}$
GWTC-4.0 + 6 GWTC-5.0 events	$4.93^{+1.04}_{-0.66}$	$3.43^{+0.73}_{-0.39}$	$1.40^{+0.34}_{-0.25}$

(ii) Figure 5 from [50]; the blue curve corresponds to the displacement memory SNR estimate for the O4 network of detectors (LIGO Hanford, LIGO Livingston, Virgo, KAGRA and LIGO India) at design sensitivity.

The obtained results in this work are consistent with the prospects in these previous works, reaching a value which is near SNR $= 1$ after ∼7.5 months of observations during O4a.

As expected, the SNR of the full (2, 0) mode accumulates more rapidly than either contribution considered separately, reaching a median value greater than 3 for the GWTC-4.0 event set. Since the GWTC-5.0 sample was selected to contain high SNR events, its inclusion in the final GWTC-4.0 results leads to a substantial increase in the corresponding values.

C. GW241127_061008: Analysis with spin-precessing models

Since GW241127_061008 shows clear evidence of spin precession [32], with a recovered value of $\chi_{\rm p} = 0.57^{+0.14}_{-0.21}$ that significantly deviates from zero, we analyze this event using the precessing version of the time domain waveform models. In particular, we compare `IMRPhenomTPHM`, against the extended model, `IMRPhenomTPHM_20`, which includes the (2, 0) mode in the coprecessing frame and the displacement memory contribution in all the $\ell = 2$ modes. This comparison allows us to assess whether these additional physical effects provide a measurable improvement in the description of the data. For the aligned-spin analysis, we obtain $\log_{10} \mathcal{B}^{\rm THM_20/THM} = 0.05 \pm 0.12$, while for the precessing analysis we find $\log_{10} \mathcal{B}^{\rm TPHM_20/TPHM} = -0.01 \pm 0.13$. These values indicate that, within the quoted uncertainties, the evidence in favor of including the additional (2, 0) mode and memory contributions is not statistically significant.

In Fig. 5, we show the posterior distributions for a set of parameters obtained with `IMRPhenomTHM_20`, `IMRPhenomTHM`, `IMRPhenomTPHM_20`, and `IMRPhenomTPHM`. For comparison, we also include the posterior samples from `IMRPhenomXPNR` reported in GWTC-5.0. Overall, we find that including the (2, 0) mode in both the aligned-spin and precessing versions of the model does not produce any noticeable shift in the recovered parameters. Since `IMRPhenomTPHM` does not model mode asymmetries, isolating a subdominant mode is more challenging, as its imprint can be partially masked by these unmodeled asymmetries. Some differences arise between `IMRPhenomTHM` and `IMRPhenomTPHM`; the precessing models recover smaller values of q and a $\chi_{\rm eff}$ distribution peaked near zero, whereas the aligned-spin models favor more unequal mass ratios and moderately negative values of $\chi_{\rm eff}$, consistent with the fact that the q–$\chi_{\rm eff}$ degeneracy is broken differently when precession is allowed. The precessing versions of the models also recover distributions shifted toward larger luminosity distances and chirp masses. Relative to `IMRPhenomXPNR`, the results of the aligned-spin models are broadly consistent across all parameters, whereas `IMRPhenomTPHM` and `IMRPhenomTPHM_20` show larger discrepancies, yielding a larger chirp mass, a slightly more equal mass ratio, and a $\chi_{\rm eff}$ closer to zero than `IMRPhenomXPNR`. This is not entirely unexpected, since `IMRPhenomXPNR` includes mode asymmetries and additional precession effects beyond those modeled in `IMRPhenomTPHM`. The inclination angle, $\theta_{\rm JN}$, remains bimodal for all models.

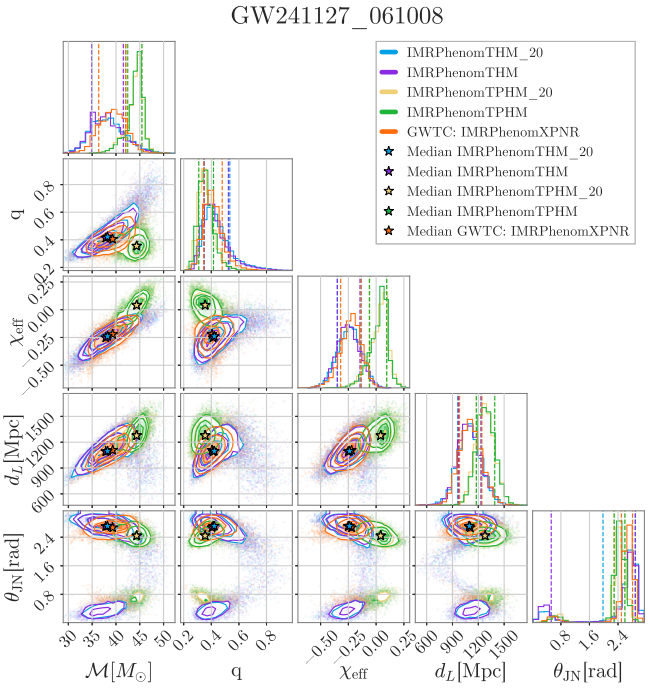


FIG. 5. Posterior distributions for the event GW241127_061008. The figure shows the parameters (chirp mass $\mathcal{M}$, mass ratio q, effective spin parameter $\chi_{\rm eff}$, luminosity distance d_L and inclination angle $\theta_{\rm JN}$), recovered with `IMRPhenomTHM_20` (blue), `IMRPhenomTHM` (purple), `IMRPhenomTPHM_20` (yellow), `IMRPhenomTPHM` (green), and the samples from GWTC-5.0 obtained with `IMRPhenomXPNR` (orange).

D. Detection prospects

Using state-of-the-art phenomenological waveform models calibrated to numerical relativity simulations this work has set constraints on the detectability of the (2, 0) mode in individual events, which remains a challenge in ground-based detectors. In the following we use the present analysis to make projections for the number of observations required to obtain clear evidence for this mode at the current detector sensitivity.

To model the empirical distributions of $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}$ and $\sigma[\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}]$ inferred from the GWTC-4.0 BBH events, we fit parametric probability distributions to the observed samples. We adopt a student's t distribution for $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}$, as the data exhibit heavier tails than a Gaussian, and a skew-normal distribution for the associated uncertainties, which display clear asymmetry. We further verify that these choices provide the best description of the data among the tested families, based on the Akaike and Bayesian information criteria (AIC and BIC). The distribution parameters (including the degrees of freedom and shape parameter) are obtained via maximum-likelihood estimation. In particular, the fitted student's t distribution yields 2.85 degrees of freedom, reflecting the pronounced non-Gaussianity and heavy tailed nature of the Bayes factor distribution, while the skew-normal fit to the uncertainty distribution gives a shape parameter of 8.33, capturing its positive skewness.

Random samples are then drawn from these distributions and accumulated until the cumulative sum reaches a fixed threshold, set to $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} - 1\sigma = 2$. This threshold is motivated by the Jeffreys scale, according to which values above 2 correspond to decisive statistical evidence. The procedure is repeated 500 times, yielding 500 independent realizations and, consequently, a distribution for the number of events required to reach the threshold. The corresponding results are shown in Fig. 6. We find that decisive evidence for the (2, 0) mode at the current detector sensitivity is expected to require a substantial catalog of detections, with an estimated number of events of $N_{\mathrm{events}} = 166^{+82}_{-55}$. This estimate should be regarded as optimistic, as it is limited to the data considered in this work and to the assumptions adopted in the analysis. A limitation of our work is that we do not account for waveform systematics, and the waveform model employed does not include the $(3, \pm 2)$ mode. This omission may affect the Bayes factor estimates, as illustrated in particular by GW230814_230901, for which the observed systematics appear to be driven primarily by the absence of this mode. Moreover, the estimate is based solely on the set of GWTC-4.0 events, which may be too small a sample from which to infer robust distributional fits. Nonetheless, under these assumptions, our results suggest that decisive evidence may be achievable by the end of O4, using the current BBH candidates reported from O4a and O4b (84 and 104 for a false alarm rate $>1\ \mathrm{yr}^{-1}$, respectively), and future candidates from O4c.

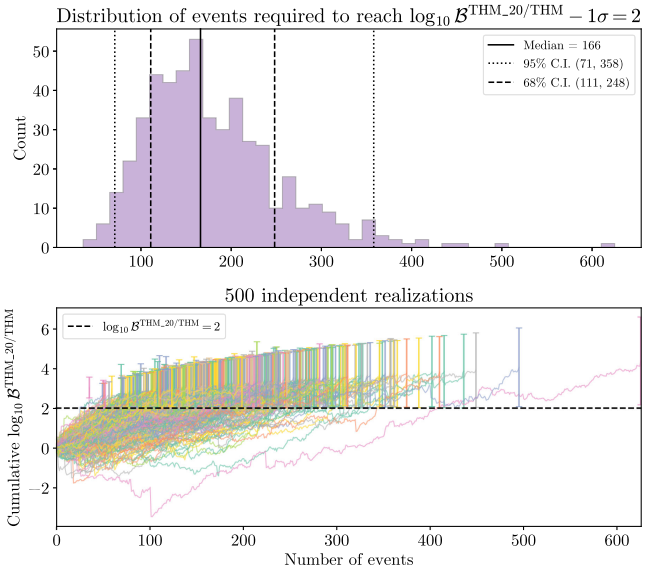


FIG. 6. Distribution of the number of events required for the cumulative Bayes factor, $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}$, to reach the threshold value of 2 accounting for the accumulated uncertainty, obtained from 500 independent realizations. The upper panel shows the histogram of threshold crossing event counts; the solid, dashed, and dotted vertical lines indicate the median, 68% credible interval, and 95% credible interval, respectively. The lower panel shows the cumulative evolution of $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}}$ as a function of the number of events for all realizations, with the horizontal dashed line marking the adopted threshold.

VI. CONCLUSIONS

In this work, we have carried out a systematic analysis of the BBH events in GWTC-4.0, together with a selected subset of six high significance GWTC-5.0 events, using the `IMRPhenomTHM` and `IMRPhenomTHM_20` waveform models. The latter includes the full $(\ell = 2, m = 0)$ spherical harmonic mode, incorporating both the oscillatory ringdown contribution and the displacement memory component. Our analysis has combined event-by-event Bayesian model comparison with cumulative evidence accumulation and SNR stacking in order to assess the detectability of the (2, 0) mode in current ground-based GW data.

At the level of individual events, the data do not provide statistically significant evidence for the inclusion of the (2, 0) mode in the majority of cases. The Bayes factors remain close to zero, with uncertainties that typically overlap the null value, indicating that the waveform models with and without this mode are generally indistinguishable at the current detector sensitivity. A mild trend toward larger Bayes factors is observed for louder events, and the strongest preference for the presence of the mode is

obtained for GW230814_230901 (bearing in mind that this result may be affected by waveform systematics). Importantly, the inclusion of the (2, 0) mode does not produce appreciable changes in the inferred source parameters for the events studied, which indicates that the omission of this mode is not yet a dominant source of systematic bias in current PE analyses.

When the evidence is combined across the full O4a sample, we obtain a cumulative $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} = 1.38 \pm 0.79$, which, with its large uncertainty, constitutes only moderate support for the presence of the (2, 0) mode and falls short of decisive evidence according to the Jeffreys' scale. The stacking of the (2, 0) mode SNR further indicates that the full mode accumulates more rapidly than its separate oscillatory and memory contributions, reaching a median value above the commonly adopted threshold of 3 for GWTC-4.0, while the memory contribution alone remains below this level. For the precessing event GW241127_061008, the comparison between the precessing waveform models similarly yields no statistically significant preference for including the additional (2, 0) contribution.

These results suggest that the (2, 0) mode is beginning to contribute measurably to the loudest events in the current catalog, but that robust event level detection remains out of reach for present ground-based sensitivities. Our prospects analysis indicates that decisive evidence for the mode may only require a moderately larger catalog, with an estimated number of events of $N_{\mathrm{events}} = 166^{+82}_{-55}$. The main shortcoming of this first cumulative analysis with the full (2, 0) mode is that we have not studied effects of waveform systematics. While an analysis such as the one presented here can reveal very small effects such as the contributions of the (2, 0) mode, it can also be sensitive to systematics. Sufficient support for a robust claim of detection will require such systematics studies and more extended injection studies beyond the zero noise simplification.

Beyond the observational implications for the (2, 0) mode, this work also constitutes a step toward the detection of GW displacement memory. Given the additional SNR that is obtained from the oscillatory component of the (2, 0) mode, we foresee that decisive evidence for the entire mode will be obtained before a clear detection of the separate memory component. We expect that controlling the presence of the full (2, 0) mode will support memory detection also in terms of controlling waveform systematics, in particular to exclude contamination from systematics effects in the part of waveform that is not the memory.

ACKNOWLEDGMENTS

The authors would like to thank Shubhanshu Tiwari for the LSC Publication and Presentation Committee review of this manuscript. We thank Héctor Estellés for useful discussions and suggestions. This paper has document number LIGO-P2600341. We thankfully acknowledge the computer resources (MN5 Supercomputer), technical expertise and assistance provided by Barcelona Supercomputing Center (BSC) through funding from the Red Española de Supercomputación (RES) (AECT-2024-3-0027); and the computer resources (Picasso Supercomputer), technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga (AECT-2025-1-0035). This research has made use of data or software obtained from the Gravitational Wave Open Science Center ([84]), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. M. R.-S. and J. V. were supported by the Spanish Ministry of Universities, Grants No. FPU21/05009 and No. FPU22/02211, respectively. Y. X. was supported by the INVESTIGA@UIB programme of the Universitat de les Illes Balears (UIB), cofunded by the 2023 Sustainable Tourism Promotion Plan (ITS2023-086—Research Promotion Programme). J. L.-Q. was supported by the Conselleria d'Educació i Universitats del Govern de les Illes Balears via FPI-CAIB doctoral Grant No. FPI_093_2022. A. R.-B. is supported by the Veni research programme which is (partly) financed by the Dutch Research Council (NWO) under Grant No. VI.Veni.222.396; acknowledges support from the Spanish Agencia Estatal de Investigación Grant No. PID2024-157460NA-I00; and the Spanish Ministerio de Ciencia, Innovación y Universidades (Beatriz Galindo, BG23/00056), cofinanced by UIB. This work was supported by the Universitat de les Illes Balears (UIB) with funds from the Programa de Foment de la Recerca i la Innovació de la UIB 2024–2026 (supported by the yearly plan of the Tourist Stay Tax ITS2023-086); the Spanish Agencia Estatal de Investigación Grants No. PID2022-138626NB-I00, No. RED2024-153978-E, and No. RED2024-153735-E, funded by MICIU/AEI/10.13039/501100011033 and the ERDF/EU; and the Comunitat Autònoma de les Illes Balears through an Ajut per a projectes de recerca científica i tecnològica (PRD2025_099) and through the Conselleria d'Educació i Universitats with funds from the European Union—European Regional Development Fund (ERDF) (SINCO2022/18146—Plataforma HiTech-IAC3-BIO).

DATA AVAILABILITY

We plan to make our results publicly available in the near future via the Zenodo repository [28].

APPENDIX A: COMPUTATIONAL COST OF THE RUNS WHEN VARYING $\Delta \log Z$

In Fig. 7, we show the relative runtime difference, in percent, between analyses performed with the default setting, $\Delta \log Z = 0.10$, and with half that value, $\Delta \log Z = 0.05$, for each GW event and waveform model version. Markers above (below) 0 indicate that the runtime for the reduced threshold is higher (lower) than for the default setting. The color bar shows the absolute runtime, in hours, obtained for the default setting, providing a reference for the computational cost associated with each event. The largest relative differences are found for short duration events, indicating that lowering $\Delta \log Z$ does not necessarily lead to a substantial increase in the computational cost of the analysis. All runs were performed on the Picasso supercomputer using 128 CPU cores. The reported runtimes therefore correspond to this computing configuration and may vary for different hardware resources.

APPENDIX B: NUMERICAL RELATIVITY INJECTION

We present in Fig. 8 the zero-noise NR injection with the simulation SXS:BBH:2497 ($Q = 2, \chi_1 = \chi_2 = 0$, $N_{\rm orbits} = 20$) recovering with `IMRPhenomTHM` and `IMRPhenomTHM_20`. As discussed in Sec. IV E, this injection study is performed as a consistency check to validate the behavior observed in the analysis of real data. We find that the parameter biases introduced by neglecting the (2, 0) mode are generally negligible, with only marginal effects observed across most of the recovered source parameters.

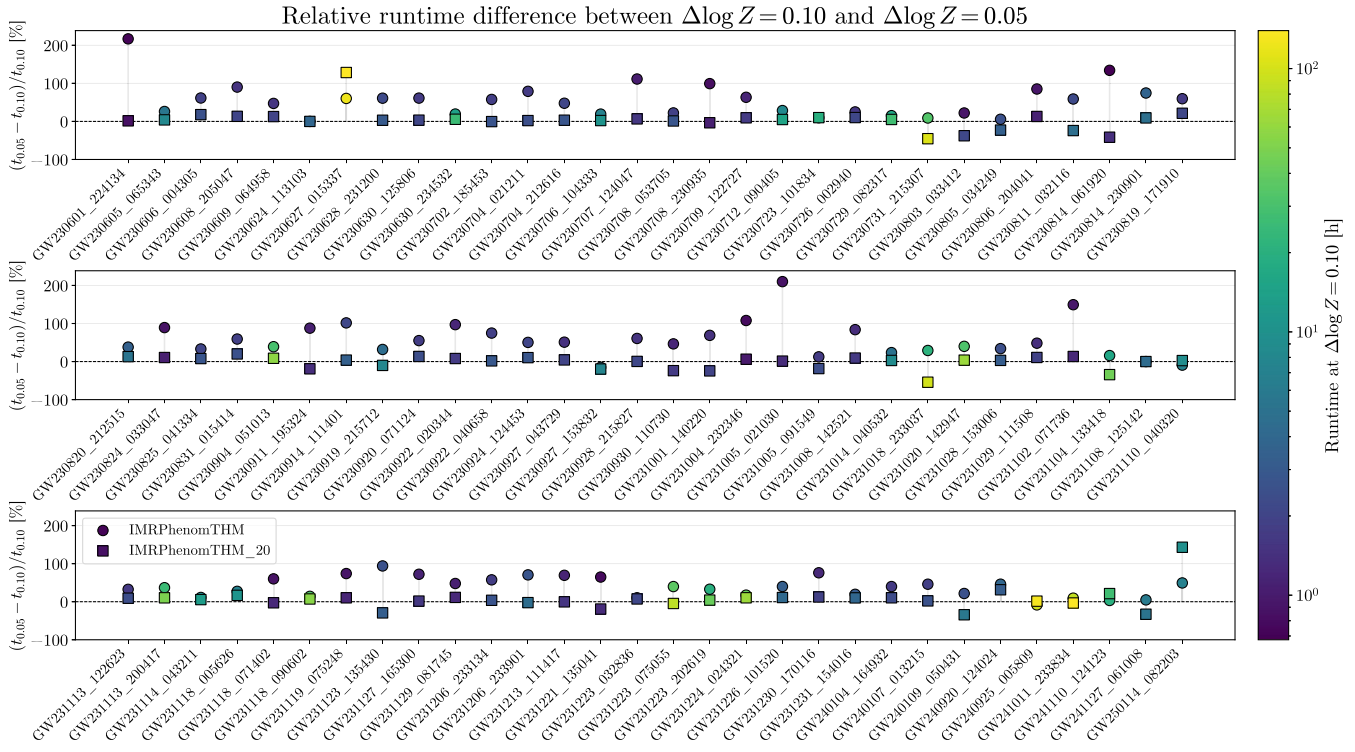


FIG. 7. Relative runtime difference between the runs performed with $\Delta \log Z = 0.10$ and $\Delta \log Z = 0.05$ for all analyzed events. The vertical axis shows the relative difference in runtime, defined as $(t_{0.05} - t_{0.10})/t_{0.10}$ in percent, where positive values indicate longer runtimes for the more stringent stopping criterion. Circles and squares correspond to the `IMRPhenomTHM` and `IMRPhenomTHM_20` models, respectively. The marker color indicates the absolute runtime obtained with $\Delta \log Z = 0.10$ (in hours), as shown by the color bar on the right. The events are chronologically ordered.

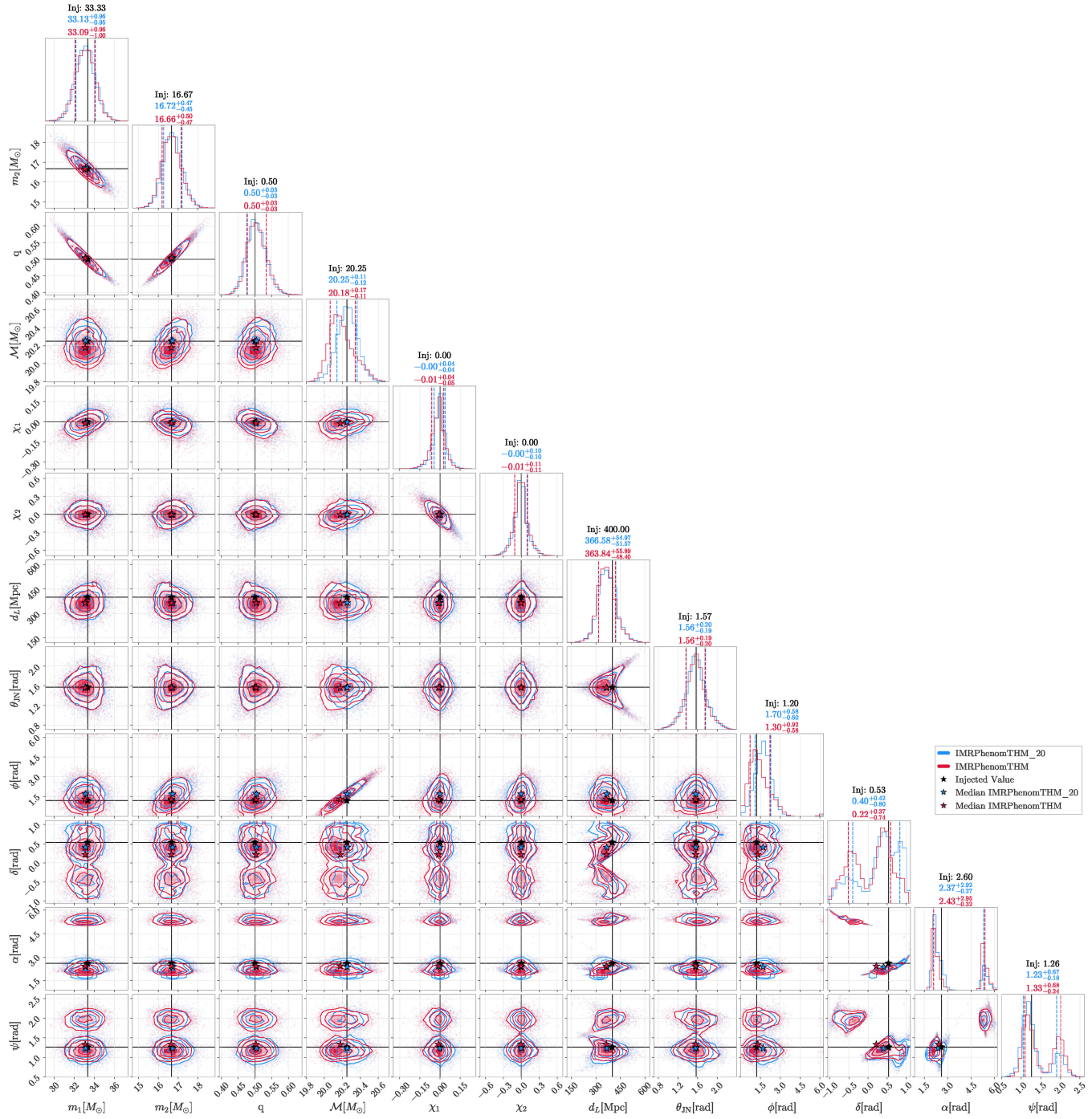


FIG. 8. Corner plot for the NR injection with SXS:BBH:2497, recovering with `IMRPhenomTHM_20` (blue) and `IMRPhenomTHM` (red), showing the 2D and 1D posteriors of the individual masses m_1 and m_2, mass ratio q, chirp mass $\mathcal{M}$, individual spin components aligned with the orbital angular momentum χ_1 and χ_2, luminosity distance d_L, inclination $\theta_{\rm JN}$, phase ϕ, declination δ, right ascension α and polarization angle ψ. The injected values are shown in black.

APPENDIX C: FURTHER RESULTS FOR THE EVENTS WITH HIGHEST $\log_{10}\mathcal{B}^{\rm THM_20/THM}$

Figure 9 shows the posterior distributions of selected source parameters for the nine events exhibiting the highest statistical evidence for the presence of the (2, 0) mode. For each event, we compare the parameter recovery obtained with waveform models including and excluding this mode in order to assess potential systematic biases arising from its omission. In all cases, the recovered posterior distributions are found to be largely consistent between the two models, indicating that neglecting the (2, 0) mode produces only negligible effects on the inferred source parameters.

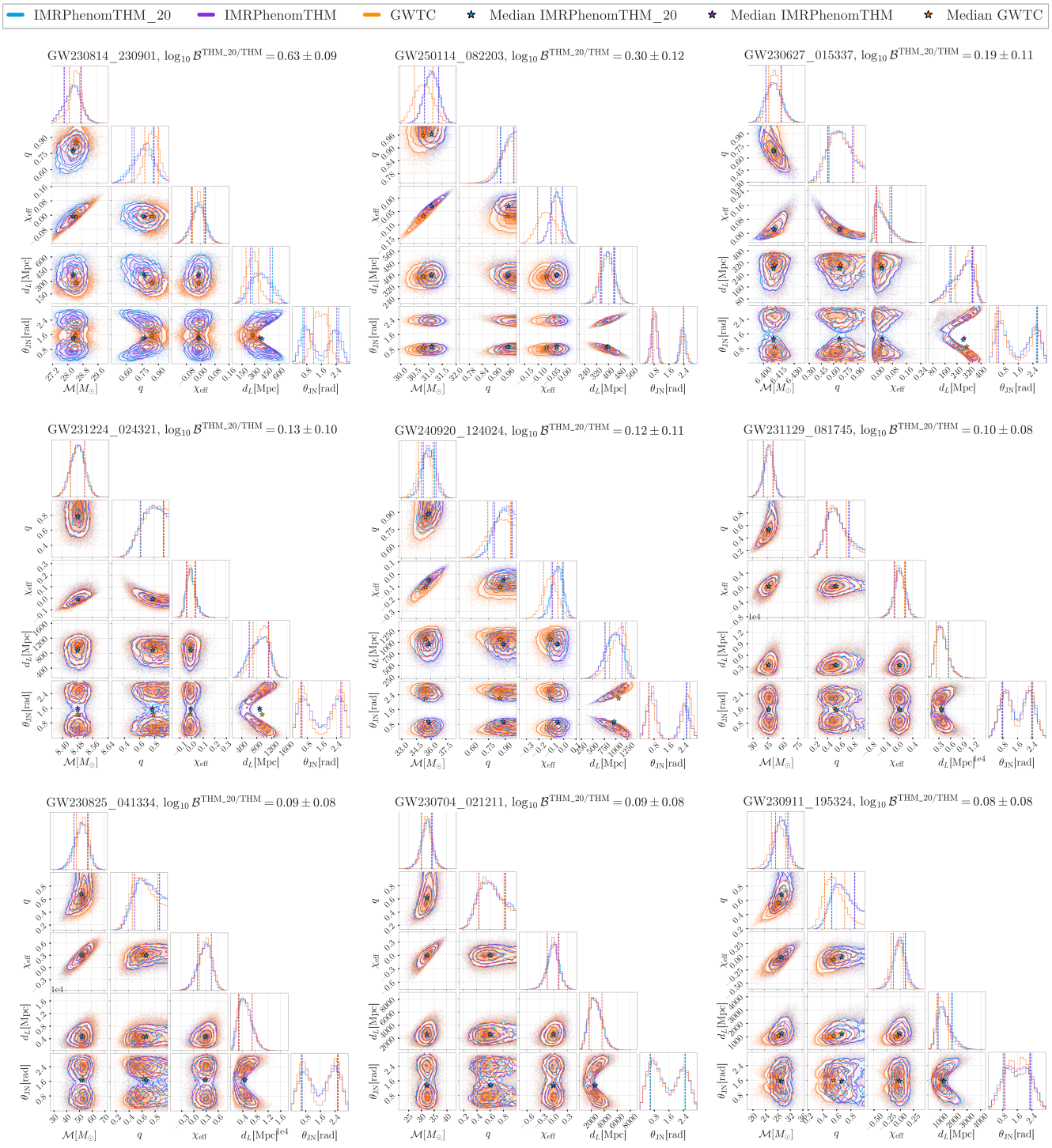


FIG. 9. Posterior distributions for the events listed in Table III satisfying $\log_{10}\mathcal{B}^{\mathrm{THM_20/THM}} - 1\sigma > 10^{-3}$. The figure shows the parameters (chirp mass $\mathcal{M}$, mass ratio q, effective spin parameter χ_{eff}, luminosity distance d_L and inclination θ_{JN}), recovered with `IMRPhenomTHM_20` (blue), `IMRPhenomTHM` (purple), and the samples from GWTC (orange).

For reference, we also include the posteriors reported by the LVK in the corresponding GWTCs.

In Fig. 10, we show the whitened waveforms of GW230814_230901 (single detector event, L1 data shown in the top panel) and GW250114_082203 (H1 data in the middle panel and L1 data in the bottom panel). In these events, the signal morphology is clearly distinguishable from the detector noise, enabling a direct visual comparison between models and data. In both cases, the maximum-likelihood waveforms obtained from the two models

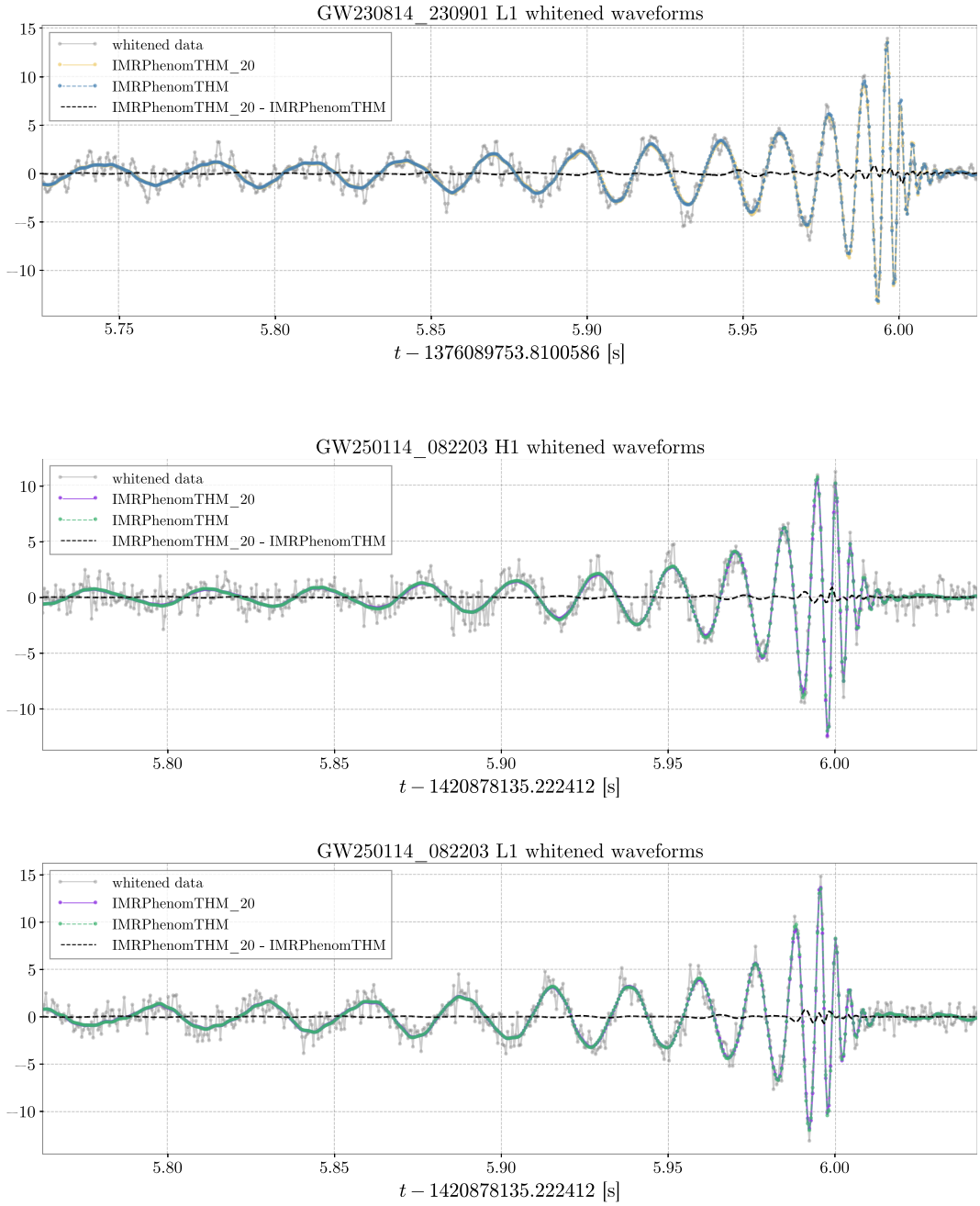


FIG. 10. Whitened waveforms for GW230814_230901 (only detected in L1) and GW250114_082203 (detected in both L1 and H1). The gray lines show the whitened data, while the colored lines show the maximum likelihood waveforms for each version of the model. In black we include the difference between the two waveform models.

provide a good fit to the data. The differences between the two waveform models with and without the (2, 0) mode are small and are shown by the dashed black curves. These discrepancies become more pronounced in the late inspiral and merger phase, consistent with the expectation that the signal contribution of the (2, 0) mode is predominantly concentrated in this regime. The absence of any apparent memory induced offset in the whitened waveforms arises as a direct consequence of the whitening procedure. By dividing the strain by the noise PSD, whitening effectively acts as a high pass filter that suppresses the low frequency content of the signal. As a result, the low frequency components associated with the memory are strongly attenuated, removing any constant strain offset and leaving only the oscillatory structure of the waveform.

[1] J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. (N.Y.) **8**, 2155 (1967).
[2] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Phys. Rev. D **84**, 024046 (2011).
[3] C. Mills and S. Fairhurst, Phys. Rev. D **103**, 024042 (2021).
[4] J. Calderón Bustillo, S. Husa, A. M. Sintes, and M. Pürrer, Phys. Rev. D **93**, 084019 (2016).
[5] F. H. Shaik, J. Lange, S. E. Field, R. O'Shaughnessy, V. Varma, L. E. Kidder, H. P. Pfeiffer, and D. Wysocki, Phys. Rev. D **101**, 124054 (2020).
[6] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, Phys. Rev. D **92**, 022002 (2015).
[7] J. Calderón Bustillo, S. H. W. Leong, T. Dietrich, and P. D. Lasky, Astrophys. J. Lett. **912**, L10 (2021).
[8] J. Calderón Bustillo, S. H. W. Leong, and K. Chandra, Nat. Astron. **9**, 1530 (2025).
[9] LIGO Scientific Collaboration, Classical Quantum Gravity **32**, 074001 (2015).
[10] Virgo Collaboration, Classical Quantum Gravity **32**, 024001 (2014).
[11] KAGRA Collaboration, Prog. Theor. Exp. Phys. **2021**, 05A101 (2020).
[12] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **102**, 043015 (2020).
[13] C. Hoy, C. Mills, and S. Fairhurst, Phys. Rev. D **106**, 023019 (2022).
[14] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **896**, L44 (2020).
[15] M. Favata, Astrophys. J. Lett. **696**, L159 (2009).
[16] M. Favata, Classical Quantum Gravity **27**, 084036 (2010).
[17] K. Mitman, J. Moxon, M. A. Scheel, S. A. Teukolsky, M. Boyle, N. Deppe, L. E. Kidder, and W. Throwe, Phys. Rev. D **102**, 104007 (2020).
[18] Y. B. Zel'dovich and A. G. Polnarev, Astron. Zh. **51**, 30 (1974), https://inspirehep.net/literature/1760701.
[19] L. Blanchet and T. Damour, Phys. Rev. D **46**, 4304 (1992).
[20] D. Christodoulou, Phys. Rev. Lett. **67**, 1486 (1991).
[21] Y. Xu, M. Rosselló-Sastre, S. Tiwari, M. Ebersold, E. Z. Hamilton, C. García-Quirós, H. Estellés, and S. Husa, Phys. Rev. D **109**, 123034 (2024).
[22] S. Gasparotto, R. Vicente, D. Blas, A. C. Jenkins, and E. Barausse, Phys. Rev. D **107**, 124033 (2023).
[23] J. Calderón Bustillo, N. Sanchis-Gual, A. Torres-Forné, J. A. Font, A. Vajpeyi, R. Smith, C. Herdeiro, E. Radu, and S. H. W. Leong, Phys. Rev. Lett. **126**, 081101 (2021).
[24] M. Rosselló-Sastre, S. Husa, and S. Bera, Phys. Rev. D **110**, 084074 (2024).
[25] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **995**, L18 (2025).
[26] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **1004**, L21 (2025).
[27] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **1004**, L22 (2025).
[28] LIGO Scientific, Virgo, and KAGRA Collaborations, GWTC-4.0: Parameter estimation data release (Version 2), 10.5281/zenodo.17014085 (2025).
[29] S. Soni *et al.* (LIGO Collaboration), Classical Quantum Gravity **42**, 085016 (2025).
[30] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), arXiv:2605.27223.
[31] LIGO Scientific, Virgo, and KAGRA Collaborations, arXiv:2605.27224.
[32] LIGO Scientific, Virgo, and KAGRA Collaborations, arXiv:2605.27225.
[33] LIGO Scientific, Virgo, and KAGRA Collaborations, arXiv:2605.27227.
[34] LIGO Scientific, Virgo, and KAGRA Collaborations, arXiv:2605.27226.
[35] Y. Xu *et al.*, Phys. Rev. D **113**, 083001 (2026).
[36] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. Lett. **135**, 111403 (2025).
[37] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **993**, L21 (2025).
[38] H. Estelles, A. Ramos-Buades, S. Husa, C. Garcia-Quiros, M. Colleoni, L. Haegel, and R. Jaume, Phys. Rev. D **103**, 124060 (2021).
[39] H. Estelles, S. Husa, M. Colleoni, D. Keitel, M. Mateu-Lucena, C. Garcia-Quiros, A. Ramos-Buades, and A. Borchers, Phys. Rev. D **105**, 084039 (2022).
[40] H. Estellés, M. Colleoni, C. García-Quirós, S. Husa, D. Keitel, M. Mateu-Lucena, M. d. L. Planas, and A. Ramos-Buades, Phys. Rev. D **105**, 084040 (2022).
[41] M. Rosselló-Sastre and S. Husa, Phys. Rev. D **112**, 124059 (2025).
[42] N. Khera, B. Krishnan, A. Ashtekar, and T. De Lorenzo, Phys. Rev. D **103**, 044012 (2021).
[43] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, Phys. Rev. D **101**, 023011 (2020).
[44] Z.-C. Zhao, X. Liu, Z. Cao, and X. He, Phys. Rev. D **104**, 064056 (2021).
[45] M. Hübner, P. Lasky, and E. Thrane, Phys. Rev. D **104**, 023004 (2021).
[46] S. Y. Cheung, P. D. Lasky, and E. Thrane, Classical Quantum Gravity **41**, 115010 (2024).
[47] P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen, Phys. Rev. Lett. **117**, 061102 (2016).
[48] O. M. Boersma, D. A. Nichols, and P. Schmidt, Phys. Rev. D **101**, 083026 (2020).
[49] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, 035 (2019), https://inspirehep.net/literature/1743201.
[50] A. M. Grant and D. A. Nichols, Phys. Rev. D **107**, 064056 (2023); **108**, 029901(E) (2023).
[51] M. Rosselló-Sastre, S. Husa, S. Bera, and Y. Xu, Phys. Rev. D **112**, 084021 (2025).
[52] K. Mitman, M. Isi, and W. M. Farr, Phys. Rev. D **114**, 064016 (2026).
[53] Y. Chen *et al.*, Phys. Rev. D **110**, 064049 (2024).
[54] M. Colpi *et al.* (LISA Collaboration), arXiv:2402.07571.
[55] H. Inchauspé, S. Gasparotto, D. Blas, L. Heisenberg, J. Zosso, and S. Tiwari, Phys. Rev. D **111**, 044044 (2025).
[56] A. Cogez, S. Gasparotto, J. Zosso, H. Inchauspé, C. Pitte, L. Magaña Zertuche, A. Petiteau, and M. Besancon, Phys. Rev. D **113**, 104034 (2026).
[57] J. Zosso, L. Magaña Zertuche, S. Gasparotto, A. Cogez, H. Inchauspé, and M. Jacobs, Phys. Rev. D **113**, 104033 (2026).

[58] G. Agazie *et al.* (NANOGrav Collaboration), Astrophys. J. **963**, 61 (2024).
[59] S. M. Tomson *et al.*, Astrophys. J. Lett. **996**, L9 (2026).
[60] M. d. L. Planas, A. Ramos-Buades, C. Garcia-Quiros, H. Estelles, S. Husa, and M. Haney, Phys. Rev. D **113**, 024006 (2026).
[61] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015).
[62] A. Ashtekar, T. De Lorenzo, and N. Khera, Gen. Relativ. Gravit. **52**, 107 (2020).
[63] G. Compère, R. Oliveri, and A. Seraj, J. High Energy Phys. 10 (2020) 116; 06 (2024) 045(E).
[64] C. García-Quirós, S. Tiwari, and S. Babak, Phys. Rev. D **112**, 064017 (2025).
[65] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. **1004**, 232 (2026).
[66] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban, SoftwareX **13**, 100657 (2021).
[67] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[68] J. Valencia, R. Tenorio, M. Rosselló-Sastre, and S. Husa, Phys. Rev. D **110**, 124026 (2024).
[69] Y. Xu, M. Rossello-Sastre, J. Valencia, A. Ramos Buades, J. Llobera Querol, and F. A. Ramis Vidal, Pe automator: V0.20.3 (Version v0.20.3), Zenodo, 10.5281/zenodo.17629330 (2025).
[70] E. Thrane and C. Talbot, Publ. Astron. Soc. Aust. **36**, e010 (2019); **37**, e036(E) (2020).
[71] H. Jeffreys, *The Theory of Probability* (Oxford University Press, Oxford, 1998).
[72] W. Del Pozzo, K. Grover, I. Mandel, and A. Vecchio, Classical Quantum Gravity **31**, 205006 (2014).
[73] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).
[74] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).
[75] Simulating eXtreme Spacetimes (SXS) Collaboration, SpEC: Spectral Einstein Code, https://www.black-holes.org/code/SpEC.html.
[76] M. Boyle *et al.*, Classical Quantum Gravity **36**, 195006 (2019).
[77] M. A. Scheel *et al.*, Classical Quantum Gravity **42**, 195017 (2025).
[78] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. Lett. **137**, 071401 (2026).
[79] LIGO Scientific, Virgo, and KAGRA Collaborations, Astrophys. J. Lett. **1004**, L23 (2026).
[80] E. Hamilton, M. Colleoni, J. E. Thompson, C. Hoy, A. Heffernan, M. Kinnear, J. Valencia, F. A. Ramis Vidal, C. García-Quirós, S. Ghosh, L. London, M. Hannam, and S. Husa, Phys. Rev. D **113**, 084055 (2026).
[81] L. Pompili *et al.*, Phys. Rev. D **108**, 124035 (2023).
[82] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, Phys. Rev. D **108**, 124037 (2023).
[83] H. Estellés, A. Buonanno, R. Enficiaud, C. Foo, and L. Pompili, Phys. Rev. D **113**, 044049 (2026).
[84] https://gwosc.org/.