Exceptionality of exceptional gravitational-wave events

Rodrigo Tenorio[*] and Davide Gerosa[†]

Dipartimento di Fisica "G. Occhialini", Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
and INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

(Received 6 January 2026; accepted 9 July 2026; published 31 July 2026)

In gravitational-wave astronomy, as in other scientific disciplines, "exceptional" sources attract considerable interest because they challenge our current understanding of the underlying (astro)physical processes. Crucially, "exceptionality" is defined only relative to the rest of the detected population. For instance, among all gravitational-wave events detected so far, GW231123 is the binary black hole with the largest total mass, while GW241110 is the binary black hole with the most strongly misaligned spin relative to the orbital angular momentum. Mandel [Astrophys. J. Lett. **996**, L4 (2026)] argued that apparent "exceptionality" may reflect measurement error rather than an extreme true value, and suggested that the total mass of GW231123 may be significantly overestimated. Here we present a quantitative analysis that supports this conceptual point. We find that claims of "exceptionality" obtained under population-agnostic priors should be critically questioned whenever measurement uncertainties are comparable to the width of the underlying population. Specifically, we find that the total mass of GW231123 is unlikely to be meaningfully affected by this effect while the spin of GW241110 is far less likely to be antialigned than initially claimed: about 70% of realizations that appear to yield an "exceptionally antialigned" spin are in fact consistent with either nonspinning or aligned configurations.

DOI: 10.1103/ptvx-bkfb

I. INTRODUCTION

Gravitational-wave (GW) detection of black-hole (BH) binaries is now becoming routine. In this context, it is natural to single out some detected events as "exceptional" because of their informative properties, distinguishing them from the bulk of the population (and, in practice for researchers, motivating dedicated "single-event papers"). Most often, such exceptional events correspond to extreme values in some of the binary parameters (masses, spins, redshift) compared to the other events observed so far. For example, GW231123 [1] is the most massive GW event detected at the time of writing, with a total mass of $M_{\rm tot} = 238^{+28}_{-49}, M_\odot$ (90% credibility), and has sparked considerable interest because of its implications for, e.g., pair-instability processes in supernovae and hierarchical BH mergers. The previous record holder for the largest total mass from the preceding data-taking period, GW190521 [2], sparked similar interest at the time. Events GW241011 and GW241110 [3] present the highest spin component aligned/antialigned with the binary orbital angular momentum, respectively, $\chi_{1z} = 0.66^{+0.08}_{-0.09}$ and $\chi_{1z} = -0.39^{+0.34}_{-0.37}$ (90% credibility), with important consequences for the rate of binary BHs in dynamical formation channels.

These measurements are obtained by conducting Bayesian inference on the observed data d using a waveform model parametrized by parameters θ. Specifically, the quantiles quoted above are summary statistics of the posterior distribution $p(\theta|d,\mathcal{I}) \propto p(d|\theta,\mathcal{I})p(\theta|\mathcal{I})$, where the likelihood $p(d|\theta,\mathcal{I})$ encapsulates the generative process of the data (including noise model, detector properties, signal waveform model, etc.) and the prior $p(\theta|\mathcal{I})$ encapsulates our initial belief on the observed signal parameters. Note that inference is conditioned on the broader context $\mathcal{I}$, which is typically omitted from the notation and collects any unstated but relevant assumptions included in the inference process [4].

Mandel [5] recently highlighted that exceptional events are so because their *inferred* parameters are exceptional compared to a population. The emphasis on the word *inferred* is important here: such inference (also known as parameter estimation) is typically performed using population-agnostic assumptions. Let us denote this context by $\mathcal{I}_{\rm ag}$; GW practitioners commonly refer to this as "the PE prior." Crucially, $\mathcal{I}_{\rm ag}$ aims to simplify the analysis of individual events at the expense of population-level

[*]Contact author: rodrigo.tenorio@unimib.it
[†]Contact author: davide.gerosa@unimib.it

information (e.g., uniform prior on component masses). We therefore find ourselves in a conundrum: an event is deemed "exceptional" relative to a population, yet the "exceptionality" itself ignores the population. Such statements of exceptionality are therefore ill posed because the information required demands a different set of assumptions, $\mathcal{I}_{\rm pop}$ (through, e.g., population-informed priors [6,7]).

In this paper, we examine the implications of this line of thought for some GW events that are currently quoted as "exceptional," providing a quantitative analysis that complements Ref. [5].

II. EXCEPTIONAL EVENTS AND MEASUREMENT ERRORS

Let us denote by θ a parameter of interest of a GW event (mass, spin, …). This parameter can be understood as a draw from a detected astrophysical population $\theta \sim p(\theta|\,\mathrm{det}, \{d\})$, where $\{d\}$ denotes data from multiple events, and det refers to the fact that those events passed a detectability threshold. For each event, all we know about the parameters θ is encapsulated in the posterior distribution $p(\theta|d, \mathcal{I})$ from which one often extracts summary statistics $\hat{\theta}$ (such as mean, median, maximum *a posteriori*, etc.). In general, the summary statistic will be shifted with respect to the true value, i.e., $\hat{\theta} \neq \theta$.

Given a collection of $j = 1, \ldots, N_{\rm obs}$ events (i.e., a catalog), an exceptional event, marked with $\star$, is defined as that which extremizes the summary statistic, i.e.,

$$\hat{\theta}^\star = \max_j \hat{\theta}^j \quad \text{or} \quad \hat{\theta}^\star = \min_j \hat{\theta}^j. \tag{1}$$

The crucial realization is that, depending on the typical spread of $p(\theta|\,\mathrm{det}, \{d\})$ and measurement errors, the "exceptionality" of an event may stem not only from the population, but also from the measurement errors.

To illustrate this, suppose there is a single-parameter population following an exponential distribution

$$p(\theta|\,\mathrm{det}, \{d\}) = \begin{cases} e^{-\theta} & \text{if } \theta \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and measurement errors that are Gaussian with a standard deviation σ. The population has unit expected value and unit standard deviation. The condition $\sigma > 1$ ($\sigma < 1$) corresponds to the case where the spread induced by measurement errors is larger (smaller) than the typical spread of the population. We generate 1000 realizations of a catalog with $N_{\rm obs} = 100$ and collect the highest mean $\hat{\theta}^\star$. The results from such a distribution of exceptional events are summarized in Fig. 1.

The main result is that, for exceptional events, the summary statistics tends to significantly overestimate the true parameter whenever the spread of measurement errors becomes compatible with that of the population. For

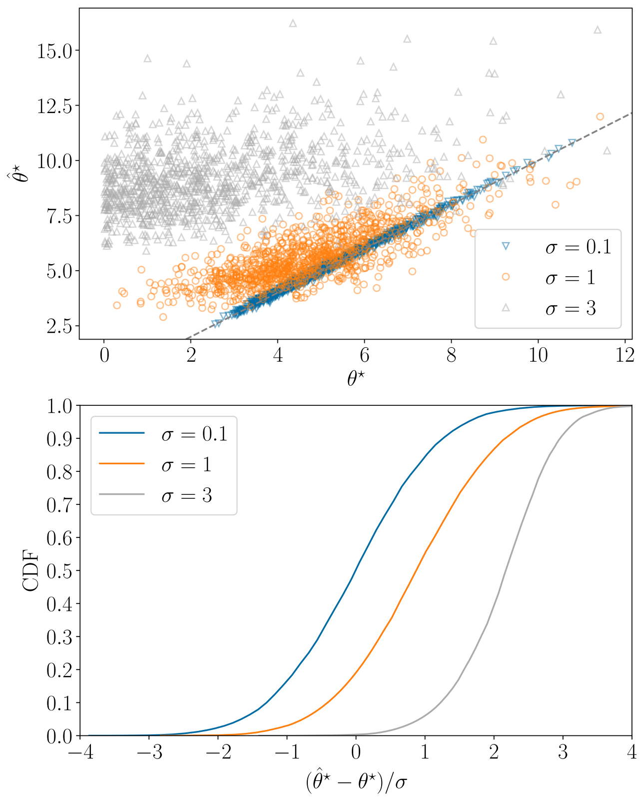

FIG. 1. Distribution of exceptional events in a catalog of $N_{\rm obs} = 100$ using the toy population of Eq. (2) and Gaussian measurement errors. The parameter σ controls the importance of measurement errors compared to the width of the population. The top panel compares the measured parameter $\hat{\theta}^\star$ to the true value $\theta^\star$; the black dashed line corresponds to $\hat{\theta}^\star = \theta^\star$. The bottom panel shows the distribution of measurement errors.

example, for $\sigma = 1$, we find that about 50% of the measured parameters overshoot the true value by more than 1σ. For $\sigma = 3$, more than 90% of the measured values are more than 1σ away, and more than 60% are 2σ away. The latter corresponds to more than a factor of 2 of overestimation.

III. GW231123

We now apply this argument to the current set of GW events. We simulate catalogs by sampling from the default binary BH population of GWTC-4.0 [8] and imposing a detectability threshold corresponding to a network signal-to-noise (SNR) ratio of 8. We verified that our results are not affected by this specific choice. The SNR is computed using the power-spectral density and response of the detectors in the event's observing run, which is sampled proportionally to the duration of the O1–O4 observing runs completed so far. As a cross-check, we show in Fig. 2 the events with largest total mass $M^\star_{\rm tot} = \max_j M^j_{\rm tot}$. For a catalog size comparable to GWTC-4.0 ($N_{\rm obs} = 153$), the resulting distribution peaks at about the total mass of GW231123.

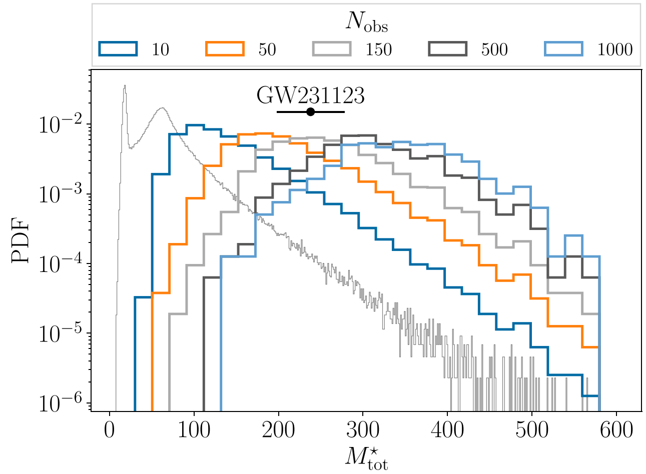

FIG. 2. Exceptional events' total mass distribution using the GWTC-4.0 population as described in the main text. Catalogs of different sizes N_{obs} are shown as colored histograms. The thin gray distribution shows the detected population $p(M_{\mathrm{tot}}|\,\mathrm{det}, \{d\})$. The scatter point indicates the total mass of GW231123 obtained under population-agnostic priors $\mathcal{I}_{\mathrm{ag}}$ from Ref. [1].

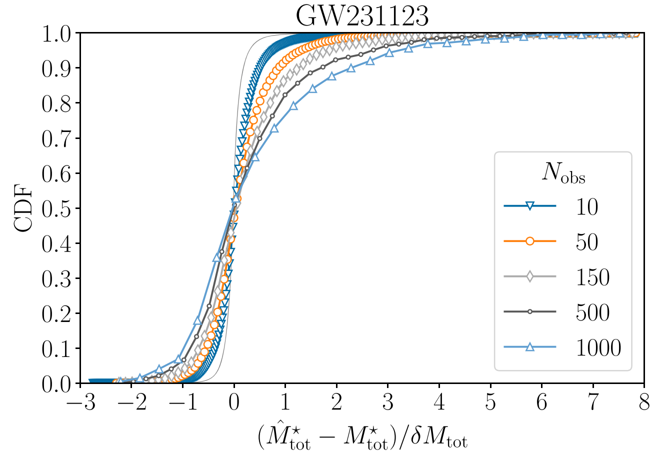

FIG. 4. Deviation of the maximum measured total mass $\hat{M}^{\star}_{\mathrm{tot}}$ with respect to the true parameter $M^{\star}_{\mathrm{tot}}$ in units of $\delta M_{\mathrm{tot}} = 40 M_{\odot}$. Results related to exceptional events from catalogs of sizes N_{obs} are shown with curves and markers. The thin gray curve indicates the corresponding distribution for the entire population $p(M_{\mathrm{tot}}|\,\mathrm{det}, \{d\})$ (i.e., without maximizing). This analysis illustrates the putative "exceptionality" of GW231123's total mass.

To construct the distribution of measurement errors, we standardize the population-agnostic ($\mathcal{I}_{\mathrm{ag}}$) posterior distribution on M_{tot} for all 153 binary BH events in GWTC-4.0 by subtracting and dividing by their posterior mean $\langle M_{\mathrm{tot}}\rangle$, as shown in Fig. 3. As a result, this procedure provides the distribution of relative deviations $M_{\mathrm{tot}}/\langle M_{\mathrm{tot}}\rangle - 1$ for each event in the catalog. This quantity appears to be rather consistent: for all events, the relative deviation is bounded between about -25% and $+30\%$ at 90% credibility. This procedure is appropriate because shifts due to population reweighing are found to be typically contained within the posterior width (see, e.g., Fig. 2 of Ref. [9] as well as Refs. [7,10]). This paper exposes some of the issues associated with maximizing over events to identify exceptionality, for which all that is needed is a reasonable distribution of measurement errors.

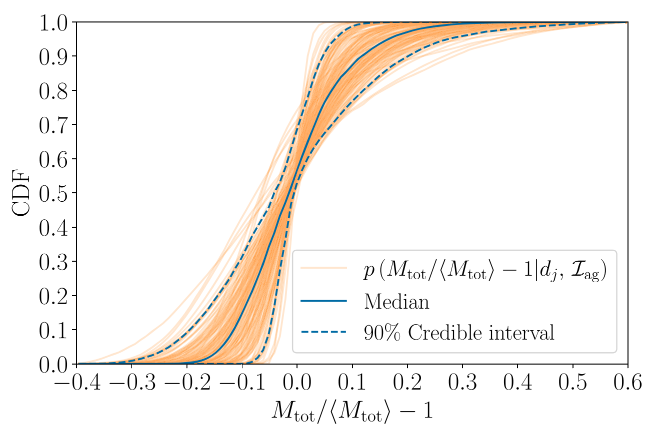

FIG. 3. Distribution of relative errors in total mass with respect to the posterior mean for the 153 events considered from GWTC-4.0 using population-agnostic priors. The solid orange curves correspond to different events. The solid blue curve shows the median of the distributions, while the dashed blue curves enclose the symmetric 90% credible interval.

Given a simulated event M^{j}_{tot}, the corresponding measurement error is obtained by sampling a value from one of the 153 standardized posterior distributions selected at random with equal probability. The resulting distribution of exceptional events is shown in Fig. 4, where we expressed the results in units of $\delta M_{\mathrm{tot}} = 40 M_{\odot}$—a value indicative of GW231123's uncertainty on the total mass. Note that the curves in Fig. 4 are not symmetric: the summary statistic $\hat{M}^{\star}_{\mathrm{tot}}$ is more likely to overestimate than underestimate the true value $M^{\star}_{\mathrm{tot}}$, and, as expected, this asymmetry increases for larger catalogs. The effect in this case is overall rather mild because mass measurement errors are smaller than the size of the targeted population.

One of the conclusions of Ref. [5] is that the true mass of GW231123 could be below $170 M_{\odot}$ as opposed to the value of $M_{\mathrm{tot}} = 238^{+28}_{-49} M_{\odot}$ obtained using population-agnostic priors, which corresponds to a shift of about $78 M_{\odot} = 1.95 \delta M_{\mathrm{tot}}$. We find that less than 5% of our simulated catalogs with $N_{\mathrm{obs}} = 150$ support this conclusion. While the typical spread of exceptional events is broader than that of the individual events, the relatively small size of δM_{tot} with respect to the typical range of M_{tot} suppresses the "exceptionality inflation."

IV. GW241011 AND GW241110

Similarly, we now consider the component χ_{1z} of the primary BH spin aligned with the binary orbital angular momentum for the events GW241011 and GW241110 [3].

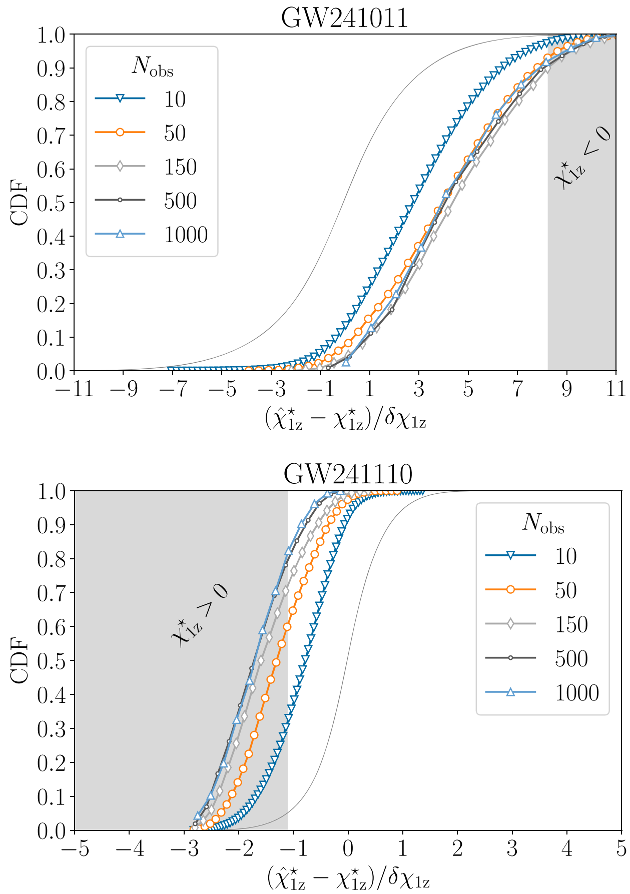

FIG. 5. Deviation of the largest aligned/antialigned spin components $\hat{\chi}^\star_{1z}$ with respect to the true parameter $\chi^\star_{1z}$. In the top (bottom) panel, we maximize (minimize) $\hat{\chi}_{1z}$, which is indicative of GW241011 (GW241110) and use units of $\delta\chi_{1z} = 0.08$ ($\delta\chi_{1z} = 0.35$); see main text. Results related to exceptional events from catalogs of sizes $N_{\rm obs}$ are shown with curves and markers. The thin gray curves indicate the corresponding distribution for the entire population $p(\chi_{1z}|\,\mathrm{det}, \{d\})$ (i.e., without extremizing). Gray areas correspond to cases where BH binaries with aligned spins are recovered with antialigned spins, and vice versa.

Since dimensionless BH spin components are bounded to the $(-1, 1)$ interval, we expect measurement errors to strongly affect events with exceptional spin properties. We simulate GW catalogs as described above. In this case, however, posterior distributions are only shifted (and not rescaled) by subtracting the posterior mean, so that measurement errors correspond to (signed) absolute deviations, and measured values are clipped to the $(-1, 1)$ range after adding measurement errors. The results are shown in Fig. 5.

GW241011 corresponds to the highest observed aligned spin $\hat{\chi}^\star_{1z} = \max_j \hat{\chi}^j_{1z} = 0.66$ with an uncertainty of about $\delta\chi_{1z} = 0.08$ (estimated from the reported 90% credibility interval [3], divided by 2). About 90% of our simulated catalogs with $N_{\rm obs} \geq 50$ report a deviation of more than $1\delta\chi_{1z}$ with respect to the true value; 70% of said catalogs return a deviation beyond $3\delta\chi_{1z}$. Only 10% of the catalogs return a deviation beyond $\approx 8.25\delta\chi_{1z}$, which would imply GW241011 is nonspinning or even antialigned. We conclude that GW241011's high-spin properties are unlikely to be driven by measurement errors. Note that this statement is conservative because the uncertainty associated to GW241011's χ_{1z} is significantly smaller than that of a typical event in the catalog, which is used in the construction of our model.

GW241110 corresponds to the highest observed antialigned spin $\hat{\chi}^\star_{1z} = \min_j \hat{\chi}^j_{1z} = -0.39$. The uncertainty in this case is about $\delta\chi_{1z} = 0.35$ [3]. For $N_{\rm obs} = 150$, about 80% of the exceptional events are more than $1\delta\chi_{1z}$ away from their true parameter values, and 70% of the exceptional events are compatible with BH spins that are preferentially aligned with the binary orbital angular momentum.

V. DISCUSSION

BHs with masses of $\sim 30 \mathrm{M}_\odot$ were considered exceptional ten years ago, following the detection of GW150914 [11], because such values were much higher than those inferred from electromagnetic observations at the time. Today, BHs of $\sim 30 M_\odot$ are considered vanilla [8]. This evolution is bound to continue. As the catalog of detected events grows, the distribution of what is deemed "exceptional" inevitably shifts toward more extreme values. The same argument also applies to measurement errors: as we collect more and more data, it becomes increasingly likely to observe an extreme measurement error. This follows fundamentally from the use of population-agnostic priors to identify exceptional events, which is incompatible with the fact that events are deemed exceptional only with respect to a population [5–7].

In this work, we have presented a simplified model of the interplay between extreme events in a population and extreme measurement errors. We conclude that, for current catalog sizes $N_{\rm obs} \lesssim 10^3$, measurement errors drive apparent exceptional events only when their magnitude is comparable to the span of the population. This is the case for spins with uncertainties of $\delta\chi_{1z} \approx 0.35$, such as GW241110, but it is unlikely to be a concern for GW241011 or for the most massive event observed to date, GW231123.

On the latter, our results are at odds with those of Ref. [5], which suggest that the total mass of GW231123 is likely to be significantly overestimated due to the use of inconsistent priors. While we agree with their overall argument, our quantitative analysis indicates that the typical uncertainty on the total mass renders this effect smaller than initially suggested. Specifically, we find that the 90% credible interval reported under $\mathcal{I}_{\rm ag}$ [1] is still likely to contain the true mass, despite the use of information that is inconsistent with respect to the population.

On the other hand, our analysis suggests that GW241110, the most confidently antialigned spin measured with GWs so far under $\mathcal{I}_{\rm ag}$ ($\chi_{1z} < 0$ at 97.7% credibility [3]), may actually not be so. Specifically, 70% of our simulated catalogs return "exceptional" antialigned events which are in fact nonspinning/aligned. This has major astrophysical implications, as spin is often quoted as a key smoking gun of BH binary formation channels.

Echoing the key message of Ref. [5], our results suggest caution when interpreting astrophysically interesting events in a population-agnostic manner, as the resulting conclusions may not be robust. In addition to Ref. [5], we stress this caveat is important whenever measurement errors are comparable in size to the width of the underlying source population.

ACKNOWLEDGMENTS

This work was partly motivated by discussions related to a bet signed by Ilya Mandel, Thomas Callister, Daniel Holz, Davide Gerosa, and Jonathan Gair at the *"Gravitational-wave snowballs, populations, and models"* workshop held in Sexten, Italy, in January 2025. We thank Ssohrab Borhanian, Federico De Santi, Giulia Fumagalli, Ana Lorenzo-Medina, Matthew Mould, and Alexandre Toubiana for discussions. R. T. and D. G. are supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, Italian-French University (UIF/UFI) Grant No. 2025-C3-386, MUR Grant "Progetto Dipartimenti di Eccellenza 2023-2027" (BiCoQ), and the ICSC National Research Centre funded by NextGenerationEU. D. G. is supported by MSCA Fellowship No. 101149270–ProtoBH and MUR Young Researchers Grant No. SOE2024-0000125. Computational work was performed at CINECA with allocations through INFN, the University of Milano-Bicocca, and ISCRA Type-C project HP10C29O69.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

[1] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **993**, L25 (2025).
[2] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **125**, 101102 (2020).
[3] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **993**, L21 (2025).
[4] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, England, 2003).
[5] I. Mandel, Astrophys. J. Lett. **996**, L4 (2026).
[6] M. Fishbach, W. M. Farr, and D. E. Holz, Astrophys. J. Lett. **891**, L31 (2020).
[7] C. J. Moore and D. Gerosa, Phys. Rev. D **104**, 083008 (2021).
[8] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), arXiv:2508.18083.
[9] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Astrophys. J. Lett. **1004**, L22 (2026).
[10] M. Mancarella and D. Gerosa, Phys. Rev. D **111**, 103012 (2025).
[11] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).