

Measuring gravitational wave speed and Lorentz violation with the first three gravitational-wave catalogs

Anarya Ray ,^{1,*} Pinchen Fan ,² Vincent F. He ,² Malachy Bloom ,² Suyu Michael Yang,²
Jay D. Tasson ,^{2,†} and Jolien D. E. Creighton ,¹

¹*Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA*

²*Department of Physics and Astronomy, Carleton College, Northfield, MN 55057, USA*

The speed of gravitational waves v_g can be measured with the time delay between gravitational-wave detectors. Our study provides a more precise measurement of v_g using gravitational-wave signals only, compared with previous studies. We select 52 gravitational-wave events that were detected with high confidence by at least two detectors in the first three observing runs (O1, O2, and O3) of Advanced LIGO and Advanced Virgo. We use Markov chain Monte Carlo and nested sampling to estimate the v_g posterior distribution for each of those events. We then combine their posterior distributions to find the 90% credible interval of the combined v_g distribution for which we obtain $0.99^{+0.02}_{-0.02}c$ without the use of more accurate sky localization from the electromagnetic signal associated with GW170817. Restricting attention to the 50 binary black hole events generates the same result, while the use of the electromagnetic sky localization for GW170817 gives a tighter constraint of $0.99^{+0.01}_{-0.02}c$. The abundance of gravitational wave events allows us to apply hierarchical Bayesian inference on the posterior samples to simultaneously constrain all nine coefficients for Lorentz violation in the nondispersive, nonbirefringent limit of the gravitational sector of the Standard-Model Extension test framework. We compare the hierarchical Bayesian inference method with other methods of combining limits on Lorentz violation in the gravity sector that are found in the literature.

I. INTRODUCTION

The third observing run (O3) of Advanced LIGO [1] and Advanced Virgo [2] was the first complete run in which all three detectors were used [3, 4]. In total, O3 adds 79 gravitational-wave (GW) candidates, more than seven times the 11 GW candidates from the first (O1) and second (O2) observing runs combined [5]. With the availability of many more GW events, it becomes possible to measure the speed of GWs v_g more precisely than previous works that used similar methods [6, 7]. Furthermore, it allows a direct and comprehensive exploration of the isotropy of v_g for the first time.

General relativity (GR) predicts that the speed of GWs is the same as the vacuum speed of light c . The GWs detected by Advanced LIGO and Advanced Virgo can be used to make statistical inferences about v_g , thereby testing the theory of GR. The first measurement of v_g using the time delay between the GW detectors was performed by Ref. [6]. By applying Bayesian inference, the 90% credible interval of v_g distribution was constrained to be $(0.55c, 1.42c)$ [6]. Reference [7] further constrained the 90% credible interval to $(0.97c, 1.05c)$, by applying similar methods to 11 events from O1 and O2. With a total of 52 high-confidence multidetector GW events accrued through the end of O3, we are able to perform a similar analysis using more events, more robustly testing the theory of GR. The method used here remains much less sensitive than the multimessenger astronomy

approach used in Ref. [8], which placed the constraint $+7 \times 10^{-16} \leq \frac{v_g - c}{c} \leq -3 \times 10^{-15}$. Rather an improved precision, the present approach provides confirmation of the basic conclusion of $v_g = c$ via an alternative approach. More significantly, the many events available that come from different sky directions permit the exploration of the isotropy of v_g .

The large body of events now available, which arrive from a multitude of sky directions, allows for a complete exploration of the isotropy of v_g in the context of the Lorentz invariance test framework provided by the gravitational Standard-Model Extension (SME).¹ Reference [7] simultaneously constrained the first four of nine coefficients for Lorentz violation in the nondispersive, nonbirefringent limit of the gravity sector using four GW events from O1 and O2. Other recent works [13–15] have sought the effects of birefringence and dispersion using the SME. Still others have sought the dependence of GW speed on the motion of the source [16]. In this paper, we use 24 of 52 high-significance multidetector GW events to simultaneously constrain all nine coefficients in the nondispersive, nonbirefringent limit of the SME. While our constraints are much weaker than previous works such as Ref. [17], which have constrained the coefficients for Lorentz violation in the gravity sector down to the order of 10^{-15} to 10^{-14} via multimessenger astronomy, these constraints were obtained using models with only one parameter each. Therefore, our work is the first to

* anarya@uwm.edu

† jtasson@carleton.edu

¹ For an annually updated review of observational and experimental results, see Ref. [9]. For early foundational work on the SME, see Ref. [10]. For foundational gravity-sector work, see Ref. [11, 12].

provide direct limits from GW observations on all nine coefficients simultaneously.

The remainder of this paper is organized as follows. In Sec. II, we discuss the methods used to extract v_g estimates for each event and present the results. Section III presents and compares a number of methods for extracting simultaneous limits on the nine coefficients for Lorentz violation before presenting our final estimate of these coefficients from the O1–O3 data.

II. SPEED OF GRAVITATIONAL WAVES

A. Bayesian inference methods

Here, we briefly describe our method for obtaining the speed of GWs. Interested readers are invited to refer to Ref. [7] for full details.

When a GW passes through Earth, if two or more detectors detect the signal, we can use the relative locations of the detectors and the difference in detection times from those detectors to simultaneously estimate the sky location of the GW event and v_g . With only one detector, we cannot find any v_g information, as there is no difference in detection times in this case. Therefore, we select those events that are detected by at least two GW detectors.

Furthermore, we only consider those events whose median signal-to-noise ratios (SNR) are no smaller than 10.0, as reported in the GWTC-2 and GWTC-3 catalog papers [3, 4]. In total, 41 O3 events meet our selection criteria and are listed in Tables I and II. All O1 and O2 events meet these two selection criteria, so we include their posterior distributions used in Ref. [7] in our analysis. Note that the SNR values used to select the O1 and O2 events (which are the same as what are used in Ref. [7]) correspond to the network SNR with which the events were found by the GstLAL search pipeline as reported in Ref. [5].

The standard parameter estimation using GW data from multiple detectors imposes the constraint that GWs travel at the speed of light [18]. In this work, we remove this constraint such that v_g becomes a parameter to be estimated with all other signal parameters. This causes wider distributions for certain parameter estimations. For example, the calculated sky area is often larger because a defined v_g aids sky localization.

Gravitational wave data d , can be decomposed into a pure GW signal $h(t)$ plus random noise $n(t)$,

$$d(t) = h(t) + n(t). \quad (1)$$

Within the framework of Bayesian inference, the posterior distribution of the parameters $\vec{\theta}$ characterizing a GW signal is computed from the likelihood of obtaining GW data given particular values of said parameters and the *a priori* knowledge of what we expect those values to be. The likelihood function is constructed by assuming the noise $n(t)$ to be stationary and Gaussian distributed.

For details regarding the exact forms of the likelihood see Ref. [7]. Once obtained, the joint posterior distribution of the signal parameters can be used to compute the marginalized posterior distribution of v_g as in

$$p(v_g|d) = \int p(\vec{\theta}|d)d\vec{\theta}, \quad (2)$$

where $\vec{\theta}'$ is the set of parameters in $\vec{\theta}$ except for v_g [7].

To carry out parameter estimation for each event that passes our selection criteria, we use public data [19, 20] from GWTC-1 through GWTC-3. We use `LALINFERENCEMCMC` [18, 21–23], which implements Markov chain Monte Carlo (MCMC) with the Metropolis-Hastings algorithm and `LALINFERENCENEST`, which implements nested sampling to run the Bayesian parameter estimation [18, 24, 25]. For our purposes of extracting v_g distributions, these two algorithms generate comparable results. We use the publicly available power spectral densities and calibration envelopes from the LIGO Scientific, Virgo and KAGRA (LVK) Collaboration in our analysis. In this paper, we use a uniform prior in v_g between 0.1c and 10c. When the v_g posterior rails against the prior, we increase the upper limit of the prior by another 10c. The broadest prior we use is from 0.1c to 30c, which we only use for one event, GW190929_012149. For parameters such as binary masses and spins, we use the same uniform and isotropic priors as those used by the LVK [3–5]. We choose a distance prior that is proportional to luminosity distance squared, similar to Ref. [5]. We do not use the more complicated cosmological priors used by Refs. [3, 4]. For O1 and O2 events, we use the posterior samples from Ref. [7], which used the `IMRPHENOMPv2` [26–28] waveform for all events except for the binary neutron star (BNS) event GW170817, which was analyzed with the `TAYLORF2` waveform [29–34]. For most O3 events, we use the `IMRPHENOMD` waveform [26, 27], which is an aligned spin waveform model for black-hole binaries. We do not use the more sophisticated `IMRPHENOMPv2` model for these events since in the context of our study, we do not expect any significant change in v_g measurements to result from the additional intricacies of the more sophisticated model. We have verified this lack of change for a subset of these events and hence chosen to stick to the `IMRPHENOMD` model consistently for all O3 events except for GW190521. For additional discussion of this point, see Appendix A. For the extremely high-mass binary black hole (BBH) event GW190521, we use the `NRSUR7dq4` waveform [35], which is one of the waveform models used by Ref. [36] for inferring this event's source properties. We note that `IMRPHENOMPv2`, `IMRPHENOMD`, and `NRSUR7dq4` are all waveform models with inspiral, merger, and ringdown.

We can achieve a more precise measurement of v_g by combining data from multiple GW events. By interpreting each observation as an independent experiment, we can multiply the marginalized likelihood as a function of v_g corresponding to each event and obtain the joint pos-

terior distribution of v_g given data from multiple events. For a uniform prior on v_g , the joint posterior can be expressed as a product of individual event posteriors.

Suppose the GW detectors observe n independent GW events with data d_1, d_2, \dots, d_n . For a uniform prior distribution of v_g , the combined posterior distribution of v_g is

$$p(v_g|d_1, d_2, \dots, d_n) \propto p(v_g|d_1)p(v_g|d_2) \cdots p(v_g|d_n). \quad (3)$$

The single event posterior distributions $p(v_g|d_i)$ are obtained as a numerical function of v_g from its parameter estimation samples by means of Gaussian kernel density estimation (KDE) [37, 38]. We use the package SCIPY’s implementation of Gaussian KDE to obtain the posteriors [39]. The joint posterior distribution is then obtained through Eq. (3).

Then, for individual and combined posteriors, we calculate Bayes factors K , via the Savage-Dickey density ratio

$$K = \frac{p(v_g = c | d_1, d_2, \dots)}{p(v_g = c)}, \quad (4)$$

where $p(v_g = c | d_1, d_2, \dots)$ is the posterior probability of $v_g = c$ and $p(v_g = c)$ is the prior probability of $v_g = c$ [40]. Higher Bayes factors suggest stronger evidence for $v_g = c$.

B. Results

In Tables I and II, we show the v_g estimates with 90% credible intervals, network SNRs, sky areas at 90% credible level, and Bayes factors for the selected 41 O3 events. Also shown are the analogous quantities obtained from their combined posteriors. Out of the 41 selected O3 events, 40 events are BBH candidate events. GW200115_042309 is a neutron star–black hole (NSBH) event, with masses of $5.9^{+2.0}_{-2.5} M_\odot$ and $1.44^{+0.85}_{-0.29} M_\odot$ at 90% credible interval [4]. Here, by combining the 41 selected O3 events, we constrain the 90% credible interval of v_g to be $0.99^{+0.02}_{-0.03} c$, with a Bayes factor of 205.9.

We combine the O3 results with the O1 and O2 results discussed in Ref. [7]. The 11 O1 and O2 events are run with `LALINFERENC_MCMC`, which shows results that are consistent with `LALINFERENC_NEST` used for O3a runs [3, 5, 7]. In Table III, we show the v_g estimates with 90% credible intervals, network SNRs, sky areas at 90% credible level, and Bayes factors for the 11 O1 and O2 events and their combined posteriors. We use the same posterior samples as used by Ref. [7], but Table III shows slightly different 90% credible intervals from those in Ref. [7] because we use Gaussian KDE smoothing in this study to extract the credible intervals while Ref. [7] directly used the posterior samples without KDE smoothing [7]. These 11 events were detected by at least two detectors and had median GSTLAL network SNR values greater than 10.0 [5]. GWTC-2.1 [41] shows network SNR values for

O1 and O2 events based on `LALINFERENC` parameter estimations, but we choose GSTLAL SNR values to be consistent with Ref. [7] from which we obtain the v_g posterior samples. GW170817 is a BNS event that was also detected in the electromagnetic spectrum [17, 42]. The “fixed” label means that the result uses the sky localization from the electromagnetic detections, which is much more precise than the localization generated by GW detection pipelines.

Combining the 41 O3 events and 11 O1 and O2 events without fixing GW170817’s sky localization at the detected electromagnetic (EM) signal, we obtain the 90% credible interval of v_g to be $0.99^{+0.02}_{-0.02} c$, with a Bayes factor of 291.9. With GW170817 sky localization fixed, v_g is $0.99^{+0.01}_{-0.02} c$, with a Bayes factor of 249.0. For a total of 49 BBH events, i.e. excluding GW170817, GW190924, and GW200115, v_g is $0.99^{+0.02}_{-0.02} c$, with a Bayes factor of 221.2. Figure 1 shows the combined posterior of v_g .

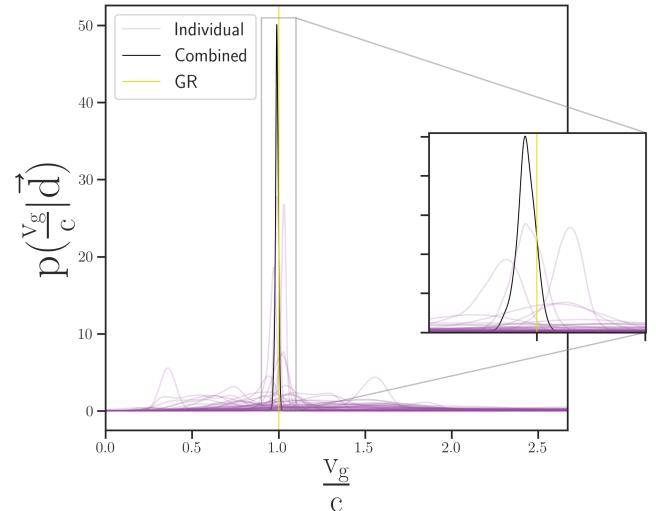


FIG. 1. Posterior distributions of v_g inferred jointly from all 41 events. The black solid line represents the joint posterior, the thin purple lines represent individual event posteriors and GR ($v_g = c$) is marked by the yellow vertical line.

C. Discussion

In Ref. [7], with 11 O1 and O2 events and GW170817’s sky localization unfixed, the combined posterior distribution of v_g was measured to be $1.01^{+0.04}_{-0.05} c$, while here we measure v_g to be $0.99^{+0.02}_{-0.02} c$ with the 52 selected events. With GW170817’s localization fixed, in Ref. [7], the combined posterior distribution of v_g was measured to be $0.99^{+0.02}_{-0.02} c$ for 11 events, while here we find $0.99^{+0.02}_{-0.01} c$ for the 52 events. Given that $1c$ is the GR prediction for v_g , the combined posterior distributions of v_g measured using 52 selected events show no evidence for a violation of GR. All of these combined results have Bayes factors

O3a Event	$v_g(c)$	SNR	$\Omega(\text{deg}^2)$	Bayes factor
*GW190408_181802	$1.66^{+0.72}_{-0.89}$	15.3	1216	3.5
*GW190412	$1.49^{+0.47}_{-0.53}$	18.9	594	2.8
GW190421_213856	$1.15^{+0.46}_{-0.57}$	10.7	2837	11.9
*GW190503_185404	$0.55^{+0.26}_{-0.24}$	12.4	1237	0.5
*GW190512_180714	$1.42^{+0.74}_{-0.98}$	12.2	1637	4.9
*GW190513_205428	$1.26^{+1.51}_{-0.65}$	12.9	1075	7.6
GW190517_055101	$0.88^{+0.67}_{-0.38}$	10.7	2125	13.8
GW190519_153544	$2.04^{+1.75}_{-1.21}$	15.6	2070	3.7
GW190521	$1.82^{+5.01}_{-1.02}$	14.2	2279	3.4
GW190521_074359	$1.20^{+0.64}_{-0.44}$	25.8	2318	11.8
*GW190602_175927	$0.98^{+0.17}_{-0.65}$	12.8	1567	27.1
GW190630_185205	$6.07^{+3.52}_{-4.90}$	15.6	3983	0.5
*GW190701_203306	$0.85^{+0.21}_{-0.43}$	11.3	203	21.4
GW190706_222641	$6.67^{+2.91}_{-4.88}$	12.6	3726	0.3
GW190707_093326	$2.77^{+5.01}_{-1.69}$	13.3	5708	0.9
*GW190720_000836	$1.54^{+0.13}_{-0.37}$	11.0	388	1.3
*GW190727_060333	$3.42^{+2.71}_{-2.05}$	11.9	1521	0.6
*GW190728_064510	$0.97^{+0.82}_{-0.53}$	13.0	1873	14.9
*GW190814	$1.33^{+0.43}_{-0.28}$	24.9	334	4.1
GW190828_063405	$6.11^{+3.41}_{-4.90}$	16.2	2498	0.5
GW190828_065509	$1.64^{+4.77}_{-0.99}$	10.0	2917	4.0
*GW190915_235702	$0.40^{+0.71}_{-0.11}$	13.6	904	5.1
*GW190924_021846	$0.92^{+0.32}_{-0.32}$	11.5	918	21.1
GW190929_012149	$6.08^{+3.53}_{-4.90}$	10.1	5761	0.5

TABLE I. The 90% credible intervals of v_g from individual O3a events posteriors. Median network SNR values are reported from GWTC-2 [3]. The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors [3]. The Bayes factor K indicates how strong the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. III.

on the order of 10^2 , providing strong evidence for $v_g = c$.

Here, our measured distribution of v_g is much narrower than that measured with 11 O1 and O2 events in Ref. [7] using GW signals alone. This is reasonable, given the larger sample size of events included in this study. When we assume that the v_g distributions of individual events are independent and identically distributed, we expect that the measurement errors would decrease by $1/\sqrt{n}$. In our calculations, we find that the combined v_g distribution roughly follows such a pattern as more events are added. For example, with 11 O1 and O2 events, the

O3b Event	$v_g(c)$	SNR	$\Omega(\text{deg}^2)$	Bayes factor
GW191109_010717	$1.80^{+1.20}_{-0.93}$	17.3	4033	2.4
GW191129_134029	$1.69^{+4.11}_{-1.18}$	13.1	4891	4.2
GW191204_171526	$1.18^{+1.26}_{-0.99}$	17.5	3009	5.5
GW191215_223052	$1.44^{+2.02}_{-0.82}$	11.2	3280	5.9
GW191216_213338	$1.31^{+0.59}_{-0.37}$	18.6	2076	11.0
GW191222_033537	$5.11^{+4.19}_{-3.66}$	12.5	3206	0.4
GW191230_180458	$1.42^{+1.07}_{-0.91}$	10.4	2538	5.9
GW200115_042309	$3.02^{+5.76}_{-2.21}$	11.3	3271	1.2
GW200128_022011	$5.57^{+3.79}_{-3.90}$	10.6	8988	0.3
*GW200129_065458	$0.99^{+0.08}_{-0.38}$	26.8	149	53.7
*GW200202_154313	$0.69^{+0.20}_{-0.33}$	10.8	1551	0.7
*GW200208_130117	$1.39^{+0.40}_{-0.60}$	10.8	706	5.6
GW200219_094415	$1.92^{+2.20}_{-1.59}$	10.7	3114	3.2
*GW200224_222234	$1.03^{+0.02}_{-0.04}$	20.0	94	48.1
GW200225_060421	$1.24^{+0.66}_{-0.75}$	12.5	3583	8.0
*GW200311_115853	$0.96^{+0.03}_{-0.05}$	17.8	102	32.1
*GW200316_215756	$3.76^{+5.26}_{-2.76}$	10.3	1881	0.8
O3 combined (BBHs)	$0.99^{+0.02}_{-0.02}$			203.3
O3 combined	$0.99^{+0.02}_{-0.03}$			205.9

TABLE II. The 90% credible intervals of v_g from individual O3b events posteriors and combined posteriors using all O3 events. Median network SNR values are reported from GWTC-3 [4]. The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors [4]. The Bayes factor K indicates how strong the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. III.

combined v_g posterior had an error bar of $0.09c$. With 52 events in total, the combined posterior had an error bar of $0.04c$, which follows $0.09c/\sqrt{52/11} \approx 0.04c$. With GW170817's sky localization unfixed, we find that Bayes factor more than doubles from the value of 149.0 obtained from 11 O1 and O2 events to the value of 291.9 obtained with all 52 events. This, in conjunction with the error bar being reduced by half, implies that our measurement with 52 GW events in total has provided approximately twice stronger evidence for $v_g = c$.

Interestingly, we find that the combined 90% v_g credible interval using the 41 O3 events is approximately the same as the 90% v_g credible interval obtained by only considering GW170817 with the fixed sky localization. GW170817 had an SNR of 33.0, while only four of the 41 O3 events had SNRs above 20.0, with the

Event	$v_g(c)$	SNR	$\Omega(\text{deg}^2)$	Bayes factor
GW150914	$0.68^{+0.50}_{-0.29}$	24.4	2385	7.0
GW151012	$6.01^{+3.45}_{-4.41}$	10.0	6607	0.4
GW151226	$4.04^{+5.17}_{-3.28}$	13.1	6515	1.3
GW170104	$1.61^{+1.88}_{-1.12}$	13.0	5313	4.3
*GW170608	$1.14^{+0.22}_{-0.26}$	14.9	1269	16.9
*GW170729	$3.70^{+1.89}_{-2.12}$	10.8	1287	0.3
GW170809	$0.65^{+0.36}_{-0.34}$	12.4	2252	8.6
*GW170814	$1.00^{+0.09}_{-0.28}$	15.9	250	68.0
GW170817 (unfixed)	$1.01^{+0.04}_{-0.06}$	33.0	53	111.9
*GW170817(fixed)	$0.99^{+0.03}_{-0.02}$	33.0	0	228.5
*GW170818	$0.94^{+0.24}_{-0.38}$	11.3	168	20.5
GW170823	$3.90^{+4.95}_{-3.00}$	11.5	6412	0.9
Combined (All BBHs)	$0.99^{+0.02}_{-0.02}$		221.2	
Combined (All, fixed)	$0.99^{+0.01}_{-0.02}$		249.0	
Combined (All, unfixed)	$0.99^{+0.02}_{-0.02}$		291.9	
Combined (O1/2, fixed)	$0.99^{+0.02}_{-0.02}$		249.0	
Combined (O1/2, unfixed)	$1.01^{+0.04}_{-0.05}$		149.0	

TABLE III. The 90% credible intervals of v_g from individual O1 and O2 events posteriors. Combined posteriors are presented for all selected events (O1, O2, and O3) as well as for O1 and O2 events. Network SNR values are reported from the GSTLAL search pipeline in GWTC-1 [5, 7]. The 90% credible regions of the sky localization (Ω) without fixing v_g at c are calculated from the individual posteriors [5, 7]. The Bayes factor indicates how strongly the posterior distributions support $v_g = c$. The asterisks (*) in front of GW event names represent the GW events chosen for obtaining constraints on all nine coefficients for Lorentz violation in Sec. III. The “fixed” and “unfixed” labels represent whether we fixed the sky localization of GW170817 at the source of its EM counterpart.

highest being 26.8 for GW200129_065458. The similarity between the v_g posterior of GW170817 alone and the 41 O3 events combined suggests that some combination of higher SNRs and better sky localization do help put tighter constraints on v_g . This shows that our decision to exclude events with SNRs lower than 10.0 should not have a high impact on the v_g estimates.

Looking to the future, additional two- and three-detector BBH events with SNRs typical of those above will lead to a slow improvement in v_g measurements as improvements proceed as $1/\sqrt{n}$. However, as GW detectors become more sensitive and the network of detectors expands, we expect more high-SNR, multidetector GW events that would likely lead to a more rapid pace of progress in v_g estimations via the methods used here. Meanwhile, future multimesenger detections can provide

more precise sky localizations, which will likely improve the error bars on the 90% v_g credible interval further.

III. SIMULTANEOUS SME LIMITS

A. Basics

In the nonbirefringent, nondispersive limit of the SME (mass dimension $d = 4$), using natural units and assuming that the nongravitational sectors, including the photon sector, are Lorentz invariant, the difference between the group velocities of gravity and light takes the form [43]

$$\Delta v = - \sum_{lm} Y_{lm}(\hat{n}) \frac{1}{2} (-1)^{1+l} \bar{s}_{lm}, \quad (5)$$

where the Y_{lm} ’s are the spherical harmonics with $l \leq 2$. Here the nine Lorentz-violating degrees of freedom are characterized by the spherical coefficients for Lorentz violation \bar{s}_{lm} , and \hat{n} is the sky location of the source of the GWs. We can expand Eq.(5) over positive m to get its equivalent expression:

$$\Delta v = \sum_l (-1)^l \left(\frac{1}{2} \bar{s}_{l0} Y_{l0} + \sum_{m>0} [\text{Re } \bar{s}_{lm} \text{Re } Y_{lm} - \text{Im } \bar{s}_{lm} \text{Im } Y_{lm}] \right). \quad (6)$$

The SME is a broad and general test framework for testing Lorentz invariance. Unlike models that attempt to describe specific effects with a small number of parameters, test frameworks, because of their generality, have a large number of undetermined coefficients to be explored in experimental data. While a number of studies have proceeded under a simplified approach, sometimes referred to as a maximum reach analysis [44], in which only one coefficient at a time is considered, it is also common to study a family of coefficients together in what is sometimes referred to as a coefficient separation approach [44]. In the context of the maximum reach approach, many coefficients can sometimes be constrained one at a time using a single measurement, while a number of measurements that is greater than or equal to the number of coefficients considered is typically required to simultaneously measure the entire family.

A number of approaches to simultaneously estimate multiple coefficients exist in the literature [7, 13, 45, 46]. One approach involves directly fitting a single data stream to a model involving all of the coefficients in the family.² This approach is well suited to experiments that take data as the laboratory is boosted and rotated.

² For a recent example of this approach in the gravity sector, see Ref. [45].

In the context of astrophysical observations, each individual event provides a measurement of a linear combination of coefficients for Lorentz violation. A system of these inequalities must then be solved, or otherwise disentangled, for estimates of the coefficients for Lorentz violation. Several methods of addressing this issue exist in the literature. In this section, we will compare the implications of several of these approaches in the context of the speed of GW data, as well as introduce new methods based on hierarchical Bayesian inference. Our goal is to consolidate information about these methods and help illuminate their relative merits. We achieve that goal by performing a mock data challenge (MDC), where we generate synthetic data corresponding to a chosen set of “true” values of the SME coefficients and test the efficacy of each method in recovering the true values from the synthetic data.

Given their performance in the MDC, along with other considerations, we choose one of these methods whose merits outweigh that of the others and use it to analyze the real speed of gravity data from a subset of the events analyzed in Sec. II to generate the final results of our SME analysis. Because well-localized events are most informative for the SME analysis, we choose the 24-event subset of those considered in Sec. II with 90% credible posterior sky areas under 2000 deg^2 as obtained from our parameter estimation with v_g as a free parameter.

B. Linear programming method

A number of past studies that have performed an analysis using limits from astrophysical events have taken a linear programming approach. See, for example, Refs. [46–48]. The basic idea translated to the speed of GWs problem proceeds as follows.

From a given event we have an upper and lower bound on Δv . If we suppose that we know an exact sky location, as is effectively the case for GW170817 when the electromagnetic signal’s localization is used, then Eq. (6) can be understood as generating a pair of hyperplanes in \bar{s}_{lm} space that are the boundaries of the parameter space excluded by the event. A subsequent event at a different sky location will generate a distinct pair of hyperplanes. Once a set of n events is collected at distinct sky locations, where n is greater than or equal to the dimensionality of the coefficient space, then a finite maximum and minimum allowed value for each coefficient can be identified via a linear programming scheme such as the simplex method.

In the applications of Refs. [46–48], the sky localizations were sufficiently well known that analysis could proceed directly via the above prescription. In the current problem, for all events except GW170817, the sky localization is comparatively poorly known. This makes the slopes of the hyperplanes bounding the allowed region poorly known.

To address our uncertainty in sky positions, the lin-

ear programming scheme can be adapted as follows. The linear programming process can be applied with all possible hyperplanes generated by samples from our inference that fall within the 68% credible sky localization bands. The worst-case limits generated by the set of linear programming analysis can then be taken as bounds. As might be expected, this method generates very conservative bounds relative to the methods to follow. Testing this approach using four test events and a sky map resolution of $N_{\text{side}} = 64$, which corresponds to $12 \times 64^2 = 49152$ pixels on the celestial sphere [49, 50], we generate bounds that are about an order of magnitude greater than the 1σ credible intervals found via the application of the random draw method that we present in the next subsection. Hence, we do not consider this approach further as a method of extracting SME limits from the speed of GWs data at this time.

C. Random draw method

In Ref. [7], the random draw method for extracting simultaneous limits on coefficients for Lorentz violation was first used. In that work, simultaneous limits were achieved for the set of four \bar{s}_{lm} coefficients with $l \leq 1$ by using the four high-confidence, well-localized events available at the time. In this section, we review this method and discuss ways of extending it to cases in which the number of events exceeds the number of coefficients to be estimated.

The result of the inference discussed in Sec. II A is a set of samples with each sample consisting of values for each of the parameters including the speed of GWs and the sky localization. Hence distributions for each of the sampled parameters are generated. If one randomly draws one sample associated with each event, one can then solve for the coefficients for Lorentz violation that are consistent with that set of samples using Eq. (6). The process of randomly drawing one sample from each event and solving for the coefficients can be iterated to build up a set of samples for the \bar{s}_{lm} coefficients. In other words, a set of points in \bar{s}_{lm} space is built up.

The process described above is straightforward when the number of events observed is equal to the number of coefficients for Lorentz violation to be estimated. Furthermore, in such a scenario, quantile ranges of \bar{s}_{lm} computed from the set of samples of \bar{s}_{lm} , accurately represent the uncertainty in our measurement of the SME coefficients. This is because using one posterior sample of $(\Delta v, \theta, \phi)$ from each event and exactly calculating \bar{s}_{lm} from them by solving a set of nondegenerate linear equations, is equivalent to computing and multiplying the posterior distributions of \bar{s}_{lm} for each event and then drawing one sample from that joint posterior. However, in the case where the number of GW observations exceeds the number of SME coefficients, the linear equations become degenerate and hence no longer exactly solvable. While one can be tempted to cherry-pick the top nine events

with the highest SNRs and lowest sky areas from the set of observations and perform random draw on those, such an analysis will not be maximally informative given the data we have. We can do better using Bayesian hierarchical inference techniques, which can combine information from a large number of events, producing much more informative bounds on the SME coefficients with accurate estimation of measurement uncertainties.

Before discussing our robust Bayesian methods we show how the random draw method can be extended to the case in which the number of observations exceeds the number of coefficients for Lorentz violation by means of singular value decomposition (SVD). However, this extension of the random draw method is susceptible to the limitations of the approximation used to perform the SVD and hence cannot produce reliable uncertainty estimates for the measured Lorentz violation parameters. We elaborate more on this near the end of this section while informing the reader beforehand that this SVD-assisted random draw generalization is useful in the present context only as a consistency check and an optimization tool for the hierarchical Bayesian methods on which we rely for our final results.

With n_{SME} coefficients for Lorentz violation and N_E events, with $n_{\text{SME}} < N_E$, for each random draw, we need to solve the degenerate system of linear equations:

$$\mathbf{A}[\bar{\mathbf{s}}_{lm}] = [\Delta \mathbf{v}]. \quad (7)$$

Here \mathbf{A} is an $N_E \times n_{\text{SME}}$ matrix in which each row corresponds to one of the N_E events under consideration. The entries in each of the n_{SME} columns moving across a given row consist of the coefficients of $\bar{\mathbf{s}}_{lm}$ in Eq. (6), computed for a random sample of θ, ϕ drawn from the event corresponding to that row. The n_{SME} SME coefficients to be computed are organized into a column vector denoted $[\bar{\mathbf{s}}_{lm}]$, while $[\Delta \mathbf{v}]$ denotes a column vector of the randomly drawn Δv corresponding to the samples used in constructing the rows of \mathbf{A} . Before factorizing the nonsquare matrix, we scale both sides of each line of Eq. (7) by the standard deviation of the Δv samples corresponding to that event. We define $[\sigma_{\Delta \mathbf{v}}]$ to be a column vector in which each element corresponds to the standard deviation of the Δv samples from that particular event; then, we can write the scaled version of Eq. (7) as

$$\mathbf{A}'[\bar{\mathbf{s}}_{lm}] = [\Delta \mathbf{v}'], \quad (8)$$

where

$$\mathbf{A}'_{ij} = \frac{\mathbf{A}_{ij}}{[\sigma_{\Delta \mathbf{v}}]_i} \quad (9)$$

$$[\Delta \mathbf{v}']_i = \frac{[\Delta \mathbf{v}]_i}{[\sigma_{\Delta \mathbf{v}}]_i}. \quad (10)$$

The SVD factorizes the nonsquare matrix \mathbf{A}' into two orthogonal square matrices \mathbf{U} and \mathbf{V} , that are $N_E \times N_E$ and $n_{\text{SME}} \times n_{\text{SME}}$ respectively, and a diagonal $N_E \times n_{\text{SME}}$ matrix Σ with nonnegative entries:

$$\mathbf{A}' = \mathbf{U} \Sigma \mathbf{V}^T, \quad (11)$$

where Σ has the form

$$\Sigma = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (12)$$

with

$$\mathbf{S} = \text{diagonal}\{\sigma_1, \dots, \sigma_{n_{\text{SME}}}\}. \quad (13)$$

The nonnegative values $\sigma_1 > \sigma_2 > \dots > \sigma_{n_{\text{SME}}}$ are known as singular values and are estimated along with \mathbf{U} and \mathbf{V} by a linear least squares algorithm [51]. The scaling with the standard deviation of Δv essentially transforms a least-square minimized SVD on \mathbf{A} into a Chi-square minimized SVD on \mathbf{A}' . This allows us to properly account for the fact that some events in our list are less significant than others. Proceeding without this scaling biases the SVD. Once computed, the singular values can be used to solve for $\bar{\mathbf{s}}_{lm}$ in Eq. (7) :

$$[\bar{\mathbf{s}}_{lm}]_i = \frac{1}{\sigma_i} \sum_{k=1}^{n_{\text{SME}}} \mathbf{V}_{ik} [\mathbf{U}^T \Delta \mathbf{v}']_k \quad (14)$$

for each draw. We can then estimate the densities of the SME parameters from all draws and produce constraints on them.

We note that despite being a computationally cheap method for computing constraints on the SME coefficients from multiple GW events, the SVD-assisted random draw method has certain inadequacies. There is ambiguity in the exact meaning and interpretation of the uncertainty estimates produced by this method. In the case where the number of events is larger than the number of SME coefficients, this implementation of the random draw method boils down to randomly choosing a posterior sample of $(\Delta v, \theta, \phi)$ from each event and doing a least chi-square fit for the SME parameters. This procedure is then repeated a large number of times, producing a least chi-square fit of the SME coefficients for each draw. However, this is not equivalent to the multiplication of posterior probabilities of the SME coefficients, over all events, and drawing samples from that joint posterior. Thus, the quantile ranges of the set of chi-square fitted SME coefficients do not hold the same meaning as Bayesian credible intervals. While the Bayesian intervals represent regions of the SME parameter space wherein their true values lie with a particular posterior probability given the data, the SVD-based random draw constraints can be expected to have a different meaning, the exact nature of which remains ambiguous.

Because of these considerations, we conclude that the weighted SVD-assisted random draw method produces constraints that are unreliable and are likely to be underestimates of the true uncertainties in the measurement of SME coefficients. We verify this claim by testing this method against its Bayesian counterparts in a MDC that we describe later in this work. The results of the MDC show that the samples of SME parameters produced by this method are concentrated in a narrow

region around the true values of the parameters, which also coincide with the peaks of the posterior distributions inferred by the Bayesian methods. Therein lies the merit of this method in the present context and its potential to serve as a rapid consistency check for the Bayesian methods. Furthermore, this method is extremely fast and computationally cheap and hence can be used to quickly find the narrow region in the parameter space inside which the peak of the posteriors lies. The stochastic MCMC sampling employed by our Bayesian methods is expected to converge much faster if the MCMC chains are initialized near the maxima of the posterior being sampled. Thus the SVD-assisted random draw method can be used to optimize the MCMC sampling used in our Bayesian methods with significant speed-up gains for narrowly peaked SME posteriors. Given the large number of events expected to be observed in O4 and the width of the Bayesian intervals we compute using our current set of events, the posterior distributions of the SME coefficients can be expected to be very narrow post O4, and hence lead to a drastic increase in the computational cost and latency of the Bayesian methods being applied to such a dataset. This will likely make the optimization of the Bayesian methods as offered by the SVD-assisted random draw method a necessary tool in the near future.

D. Hierarchical Bayesian inference

Since the SME coefficients are properties that are expected to be the same for all events, one can perform Bayesian hierarchical inference on them from the GW data of multiple events. To do so, we can construct the marginalized likelihood of GW data given a particular value of the SME coefficients, jointly from multiple events

$$L(\bar{s}_{lm}) = \prod_{i \in \{\text{events}\}} \int L(d_i | \Delta v', \theta, \phi) \Pi(\Delta v', \theta, \phi | \bar{s}_{lm}) d\Delta v' d\theta d\phi, \quad (15)$$

where the SME sensitive part of the prior imposes the relationship (6) on $\Delta v, \theta, \phi$ for a given value of the SME coefficients :

$$\Pi(\Delta v', \theta, \phi | \bar{s}_{lm}) = \delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) \pi(\theta) \pi(\phi). \quad (16)$$

Here, $\Delta v(\bar{s}_{lm}, \theta, \phi)$ is the right-hand side of Eq. (6). Note that we have chosen to represent the deviation of the speed of gravity from the speed of light by the dummy variable $\Delta v'$ whenever a probabilistic quantity (such as likelihood, posterior, prior, or detection fraction) is expressed as a function of it, so as to distinguish it from the quantity $\Delta v(\bar{s}_{lm}, \theta, \phi)$. The presence of the delta function in Eq. (16) is due to the deterministic nature of Eq. (6).

By Bayes' theorem, for a uniform prior on \bar{s}_{lm} , the likelihood $L(\bar{s}_{lm})$ is proportional to the posterior of these

parameters given GW data. We can now sample this posterior using MCMC to produce joint SME constraints from multiple GW observations. However, this procedure involves a very large number of evaluations of the likelihoods $L(d_i | \Delta v', \theta, \phi)$, which is so computationally expensive that it is practically infeasible.

To get around this problem, one can again use Bayes' theorem to write the likelihood $L(d_i | \Delta v', \theta, \phi)$ as proportional to the ratio of the posterior $p(\Delta v', \theta, \phi | d_i)$ to the prior:

$$L(d_i | \Delta v', \theta, \phi) \propto \frac{p(\Delta v', \theta, \phi | d_i)}{\pi(\Delta v') \pi(\theta) \pi(\phi)} \quad (17)$$

Substituting this into Eq. (15) gives us

$$L(\bar{s}_{lm}) \propto \prod_{i \in \{\text{events}\}} \int p(\Delta v', \theta, \phi | d_i) \delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) d\Delta v' d\theta d\phi. \quad (18)$$

We can now use the samples drawn from the posterior $p(\Delta v', \theta, \phi | d_i)$ obtained using the parameter estimation run described above to evaluate the integral in Eq. (18). Note that we have ignored a factor of $1/\pi(\Delta v')$ in Eq. (18) which is constant since we choose $\pi(\Delta v')$ to be uniform in our parameter estimation runs. However, the presence of the Dirac delta makes it slightly complicated to evaluate this integral directly as a sum over posterior samples. We describe shortly two approximation schemes that can be used to smooth out the discrete sum of Dirac deltas over posterior samples that would entail the evaluation of the integral in Eq. (18) and hence constrain the SME coefficients jointly from multiple GW observations. Before that, we first describe why Bayesian inference of this form is subject to selection biases and how we account for them.

Bayesian hierarchical inference from a set of GW events selected based on a particular criterion introduces selection biases into the inferred posterior distribution of hyperparameters [52, 53]. Since we are selecting events based on whether they were found with a SNR greater than some threshold in at least three detectors, and since each detector has an antenna pattern that makes it more sensitive to certain sky directions than others at the time of detection [54], our analysis might be biased toward some values s_{lm} against others. Particularly, the fact that GW search pipelines such as GstLAL only report multidetector coincidences based on whether or not the time delays between the detectors being triggered are smaller than the light travel time between detectors plus a 5 ms window, has the potential to bias our results greatly [55]. Furthermore, noncoincident events are down-ranked in significance [55], making events even less likely to be detectable for certain cases. Other pipelines such as PyCBC use similar methods for identifying multidetector coincidences albeit with a different value for the timing error window (which is 2 ms for PyCBC [56]).

The existence of this restriction for coincidence formation in search pipelines implies that we are more likely to discover a multidector event if the speed of GWs is greater than or equal to c , as compared to if it were lower than c . Thus, our speed of GW measurements may be biased toward measuring $\Delta v \geq 0$ against $\Delta v < 0$ along any particular sky position.

To account for this bias, we must normalize our hierarchical likelihood over the true rate of events as opposed to the detected rate, with the latter being different from the former, due to selection biases. The constant of normalization is the fraction of events that are detectable given a particular value of the hyperparameters and the detection criteria,

$$L(\bar{s}_{lm}) \propto \frac{1}{\beta_{\text{det}}^N(\bar{s}_{lm})} \times \prod_{i \in \{\text{events}\}}^N \int p(\Delta v', \theta, \phi | d_i) \delta(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) d\Delta v' d\theta d\phi, \quad (19)$$

where $\beta_{\text{det}}(\bar{s}_{lm}) = \frac{R_{\text{det}}(\bar{s}_{lm})}{R_{\text{true}}}$, the fraction of detectable events, is the ratio of the detectable rate of events to the true rate of events [57]. To calculate the fraction accurately, we must simulate a large number of events whose parameters are drawn from broad enough distributions, inject them into the detector noise realizations, and see what fraction of them are recovered given our selection criteria. To do that, we must first quantify our selection criteria in terms of the parameters that characterize the GW signal. Accurate modeling would require us to recalculate the search pipeline's ranking statistic of a simulated event while allowing for nonzero Δv and to find the corresponding false alarm rate (FAR) of that trigger from said ranking statistics. One can then apply a threshold on the combined FAR of the event to classify them as detectable or nondetectable. However, such a calculation would require a pipeline-specific analysis, which is beyond the scope of this work. Instead, we use an approximated selection criteria: for the i th event to be detectable, its recovered parameters must satisfy:

$$\begin{aligned} \text{det} \implies & \{\rho_H \geq \rho_{\text{th}}, \rho_L \geq \rho_{\text{th}}, \rho_V \geq \rho_{\text{th}}, \rho_{\text{net}} \geq \rho_{\text{net,th}}, \\ & \Delta t_{HL}(\Delta v) \leq \Delta t_{HL}(0) + 5 \text{ ms}, t_{HV}(\Delta v) \leq \\ & \Delta t_{HV}(0) + 5 \text{ ms}, t_{VL}(\Delta v) \leq \Delta t_{VL}(0) + 5 \text{ ms}\}, \end{aligned} \quad (20)$$

where ρ_A is the SNR in detector A , ρ_{net} is the network SNR, $\Delta t_{AB}(\Delta v)$ is the timedelay of signal arrival between detectors A and B as a function of Δv , and ρ_{th} is the SNR threshold used for selecting events. Even though we do not select events depending on which search pipeline found them, we use GSTLAL's timing error window to quantify our selection criteria, instead of, say, PyCBC's, due to the following reason. Among the events that survive our three detector SNR thresholds, most are found by both GSTLAL and PyCBC except for GW170818, GW190701, and GW190814, which are found only by GSTLAL. Hence, it is sufficient to model

the selection biases that might have appeared in this particular study based on GSTLAL's value of the timing error window. This would not have been possible if there were events found by PyCBC and not GSTLAL with SNR greater than 10 in three detectors during O3. In such a scenario, a more generalized treatment of selection biases would have been necessary, one that accounts for the difference in timing errors allowed by GSTLAL and PyCBC.

Now that we have a quantifiable detection criterion, we can carry out our simulations. Once the simulated events are injected into detector noise realizations and classified as detectable or nondetectable depending on their recovered parameters, it is possible to compute the fraction of events detectable given a choice of CBC parameters:

$$f_{\text{det}}(\Delta v', \theta, \phi, \vec{\gamma}) \propto \frac{p(\Delta v', \theta, \phi, \vec{\gamma} | \text{det})}{\Pi_{\text{sim}}(\Delta v', \theta, \phi, \vec{\gamma})}. \quad (21)$$

Here, $\vec{\gamma}$ are additional CBC parameters such as masses, spins, etc. that characterize the waveform, $p(\Delta v', \theta, \phi, \vec{\gamma} | \text{det})$ is the probability of detection, which can be calculated from the set of simulated events that are detectable, and Π_{sim} is the prior from which the simulations are drawn, which has to be broad enough so that we have enough events in both the detectable and nondetectable parts of the parameter space. We can marginalize Eq. (21) over suitable priors to get

$$\beta_{\text{det}}(\bar{s}_{lm}) = \int f_{\text{det}}(\Delta v', \theta, \phi, \vec{\gamma}) \Pi(\Delta v', \theta, \phi | \bar{s}_{lm}) \times \Pi(\vec{\gamma}) d\Delta v' d\theta d\phi d\vec{\gamma}. \quad (22)$$

If we choose $\Pi_{\text{sim}}(\Delta v', \theta, \phi, \vec{\gamma}) = \pi(\Delta v')\pi(\theta)\pi(\phi)\Pi(\vec{\gamma})$, where $\pi(\Delta v')$, $\pi(\theta)$, $\pi(\phi)$ are the same as the ones defined in Eqs. (16) and (17), then priors in the denominator and numerator of the integrand in (21) cancel out and we can define the marginalized fraction of detectable events (up to the factors that cancel out later):

$$f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi) \propto \int p(\Delta v', \theta, \phi, \vec{\gamma} | \text{det}) d\vec{\gamma}. \quad (23)$$

As in the case of Eq. (18), we have ignored a factor of $1/\pi(\Delta v')$ in Eq. (23) for the same reason mentioned before. In terms of this marginalized fraction, β_{det} becomes

$$\beta_{\text{det}}(\bar{s}_{lm}) \propto \int f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi) \delta(\Delta v' - \Delta v(\theta, \phi, \bar{s}_{lm})) d\Delta v' d\theta d\phi. \quad (24)$$

To estimate $p(\Delta v', \theta, \phi, \vec{\gamma} | \text{det})$ and hence $f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi)$, we simulate a large number of events whose parameters are drawn from a broad distribution. We then inject the corresponding signals into detector noise realizations and record their SNRs and arrival times. We then apply our selection criteria to find which of these simulated events are detectable given our

criteria and estimate $p(\Delta v', \theta, \phi, \vec{\gamma} | \text{det})$. The estimation schemes will depend on which of the two approximations referred to before are used to smooth out the delta function integral and are hence described in more detail in the corresponding subsections below.

The priors we use to draw the simulated events are truncated power law in the primary mass and mass ratio, uniform in spin, sky position, orientation, comoving volume, geocentric time, and speed of GWs. Particularly, for each observing run, the mass distributions are chosen to be consistent with corresponding population analyses performed by the LVC such that the distributions used have support in regions of the mass space where the events being analyzed are found. For O2, we choose $p(m_1) \propto m_1^{-1.6}$, $m_1 \in (7.9M_{\odot}, 42M_{\odot})$, and $p(q) \propto q^{6.7}$, where $q = \frac{m_2}{m_1}$, which is consistent with Ref. [58] and is identical to the mass distributions used for similar selection function computations [54]. For O3, we choose $p(m_1) \propto m_1^{-1.6}$, $m_1 \in (7M_{\odot}, 80M_{\odot})$ and $p(q) \propto q^{6.7}$, which is broad enough for the O3 events as evident from Ref. [59]. In the next two subsections, we describe the details of our smoothing approximations and the computation β_{det} in each approximation scheme.

1. Narrow Gaussian method

The approach introduced here involves estimating the delta function in Eq. (18) as a narrow Gaussian distribution. For each sample with measured speed difference $\Delta v'$ and sky location θ and ϕ , we construct a Gaussian distribution for the random variable $\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)$ with mean zero and standard deviation σ . Thus, Eq. (19) becomes

$$L(\bar{s}_{lm}) = \frac{1}{\beta_{\text{det}}^N(\bar{s}_{lm})} \prod_{i \in \{\text{events}\}}^N \int p(\Delta v', \theta, \phi | d_i) \times \mathcal{N}(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) d\Delta v' d\theta d\phi, \quad (25)$$

where \mathcal{N} represents Gaussian distributions. Similarly, we can also apply the narrow Gaussian approximation to the computation of β_{det} in Eq. (24):

$$\beta_{\text{det}}(\bar{s}_{lm}) = \int f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi) \mathcal{N}(\Delta v' - \Delta v(\bar{s}_{lm}, \theta, \phi)) \times d\Delta v' d\theta d\phi. \quad (26)$$

Since \mathcal{N} is a smooth function of its arguments we can evaluate the two integrals in Eqs. (25) and (26) as a Monte Carlo sum over samples drawn from $p(\Delta v', \theta, \phi | d_i)$ and $f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi)$, respectively. Since we already have posterior samples drawn from $p(\Delta v', \theta, \phi | d_i)$ for each event during the v_g inference described in Sec. II, and since the samples drawn from $f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi)$ are the parameters of simulated events that survive our selection

criteria, we can compute the log-likelihood of \bar{s}_{lm} ,

$$\ln L(\bar{s}_{lm}) = \sum_{i \in \{\text{events}\}} \ln \frac{\sum_{\{j\}} \mathcal{N}(\Delta v'_j - \Delta v(\bar{s}_{lm}, \theta_j, \phi_j))}{\sum_k \mathcal{N}(\Delta v'_k - \Delta v(\bar{s}_{lm}, \theta_k, \phi_k))}, \quad (27)$$

where the sum in the numerator is over posterior samples corresponding to the i th event while the one in the denominator is over detectable samples. After choosing a width σ for our Gaussian \mathcal{N} appropriately, we can thus use Eq. (27) for fast evaluation of the log-likelihood $\ln L(\bar{s}_{lm})$ as a numerical function of the SME coefficients. Hence, we can use MCMC algorithms to draw samples from $\ln L(\bar{s}_{lm})$ and interpret the quantile ranges of said samples as Bayesian credible intervals of the SME coefficients given GW data.

To determine the appropriate width of our Gaussian distribution σ , we consider the effect of varying its size. Because the Gaussian distribution is an estimation of the delta distribution, theoretically, as the size of σ decreases, the approximation should be more accurate. However, because we sample the log-likelihood with a MCMC algorithm, we encounter numerical difficulties when the σ is too small. Thus our choice of σ has to be tuned in accordance with how the MCMC is implemented numerically.

In the MCMC process, the walkers only make use of local information at each step. Thus, it is possible for walkers to be trapped inside islands of high likelihood. This is what happens when σ is set too small. Since most samples have a high likelihood around zero, walkers can explore freely the region near zero. However, at more peripheral locations in the parameter space, the peaks are usually scattered. Thus, when σ is too small, these peripheral samples form isolated islands of high likelihood. In this case, the walkers will not be able to explore these isolated islands, resulting in false small constraints. On the other hand, as σ gets larger, our approximation becomes less accurate and distributions are artificially broadened. Therefore, we aim to find a σ such that it is big enough for the walkers to explore the sample space fully and small enough such that it gives us useful results.

One way to determine the appropriate value of σ is by applying both the random draw method and the narrow Gaussian method on the same set of data and comparing the results. We divide our list of events into subsets of nine events and apply both methods to each subset using various sizes of σ . The solid lines in Fig. 2 represent the average uncertainty of the resultant \bar{s}_{lm} 's against σ for the nine O3 events from this paper with the smallest sky areas. The average uncertainty is calculated by taking the average of the absolute value of the upper and lower one-standard-deviation value for each \bar{s}_{lm} . For the same sets of events, the random draw method produces uncertainties on the order of 10^0 – 10^1 , which corresponds to dashed lines in Fig. 2. To select a suitable σ , we use superimposed plots such as Fig. 2 to select a σ

such that each uncertainty produced by the narrow Gaussian method is marginally larger than its counterpart from the random draw method. In the example shown, $\sigma = 0.0005$ is a good choice because every solid line lies marginally above the corresponding dashed line in the same color, which means that the choice σ does not artificially tighten the constraints. We perform such analysis for every subset of events and produce a “good σ .” We pick the largest of such “good σ ’s” as the final choice. Even though this final σ produces wider constraints for each subset of events as compared to the random draw method, because the narrow Gaussian method incorporates information from more than nine events, ideally we could still produce tighter constraints than the random draw method.

Another procedure for determining σ is to use information directly from the distribution of $\Delta v'$ samples for each event. One specific procedure is to sort the list of $\Delta v'$ samples and compute the average difference between adjacent values. Choosing σ as half this value produces results that align well with the prior method for the specific distributions tested. That is, the value of σ for a given event is half of the average of the differences between adjacent $\Delta v'$ samples for that event. This procedure is used in the MDC shown in Fig. 3, and has the advantage of not needing to construct plots to determine σ . However, there still is subjectivity in choosing what fraction of the average to use. Moreover, the distribution and quantity of samples will impact the value of σ . Both considerations will affect the final credible intervals for \bar{s}_{lm} .

This method performs much better than the SVD-assisted random draw in the MDC performed in Sec. III F. However, we note that both processes for choosing σ involve a significant amount of user-controlled fine-tuning and can potentially lead to an over-/underestimate of measurement uncertainties of the SME coefficients. For these reasons, we do not choose this method for our final results. We instead choose a different smoothing approximation to the Bayesian method by means of KDE, which can be shown to produce either equally or more accurate results, while requiring almost no user-controlled fine-tuning.

2. KDE methods

In this section we outline a different approach from the one in the previous section, to perform Bayesian hierarchical inference of the SME coefficients from GW data. In this method, instead of smearing out the delta function in $\Pi(\Delta v', \theta, \phi | \bar{s}_{lm})$ with a Gaussian, we approximate the marginalized posterior of $\Delta v, \theta, \phi$ given GW data, for each event, as a fast evaluating function of these quantities, from their single event parameter estimation samples via Gaussian KDE.

The KDE approximation of the posterior is constructed by fitting a multivariate Gaussian around each

posterior sample and then writing the estimate of the posterior as a sum of these individual Gaussians. The covariance matrix of each of the Gaussians is approximated from the sample covariance matrix of the posterior samples themselves up to a constant of proportionality. The constant of proportionality is known as the bandwidth of the estimator and is computed, under reasonable assumptions regarding the true distribution being estimated (see Ref. [38]). We use SciPy’s Gaussian KDE algorithm to obtain our estimate of the marginalized posterior as a fast evaluating function $p_{\text{KDE}, i}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_{lm}), \theta, \phi)$ of the relevant parameters [39]. We then perform the $\Delta v'$ integral of Eq. (18) analytically using the delta function and compute the remaining two integrals (over θ, ϕ) numerically using the trapezoidal rule. We loop over multiple events by multiplying the value of the integral obtained using the KDE corresponding to each event, to evaluate $L(\bar{s}_{lm})$:

$$L(\bar{s}_{lm}) \approx \prod_{i \in \{\text{events}\}} \int p_{\text{KDE}, i}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_{lm}), \theta, \phi) d\theta d\phi. \quad (28)$$

We then sample from it using the same MCMC method described in the previous section to constrain the SME coefficients. To incorporate selection effects in the KDE method, we estimate $f_{\text{det}}^{\text{marg}}(\Delta v', \theta, \phi | \text{det})$ by performing a KDE on the samples of $(\Delta v', \theta, \phi)$ for which the simulated events are detectable given our detection criteria. By restricting our KDE to only these parameters and ignoring other parameters that characterize a simulated event, we effectively marginalize over those other parameters, thus implicitly performing the integral in Eq. (23),

$$f_{\text{det}}^{\text{marg}} \approx p_{\text{KDE, det}}(\Delta v', \theta, \phi), \quad (29)$$

where the subscript det in $p_{\text{KDE, det}}$ represents the fact that this KDE was performed on only those samples for which the simulated events are detectable given our detection criteria. Substituting into Eq. (24), we get

$$\beta_{\text{det}}(\bar{s}_{lm}) \approx \int p_{\text{KDE, det}}(\Delta v' = \Delta v(\theta, \phi, \bar{s}_{lm}), \theta, \phi) d\theta d\phi. \quad (30)$$

We note that the KDE’s bandwidth acts like a control parameter with potential room for user-controlled fine-tuning in the computation of its value, somewhat analogous to the σ of the narrow Gaussian method. However, unlike the narrow Gaussian method where σ can in principle be chosen to be anything, the bandwidth of the KDE is computed directly from the properties of the samples (such as the number of samples and dimensionality of the parameter space) under reasonable assumptions regarding the true density. Thus, the user’s choice is restricted to a number of discrete such assumptions (for example, Scott’s rule [38], Silverman’s rule [37], etc.).

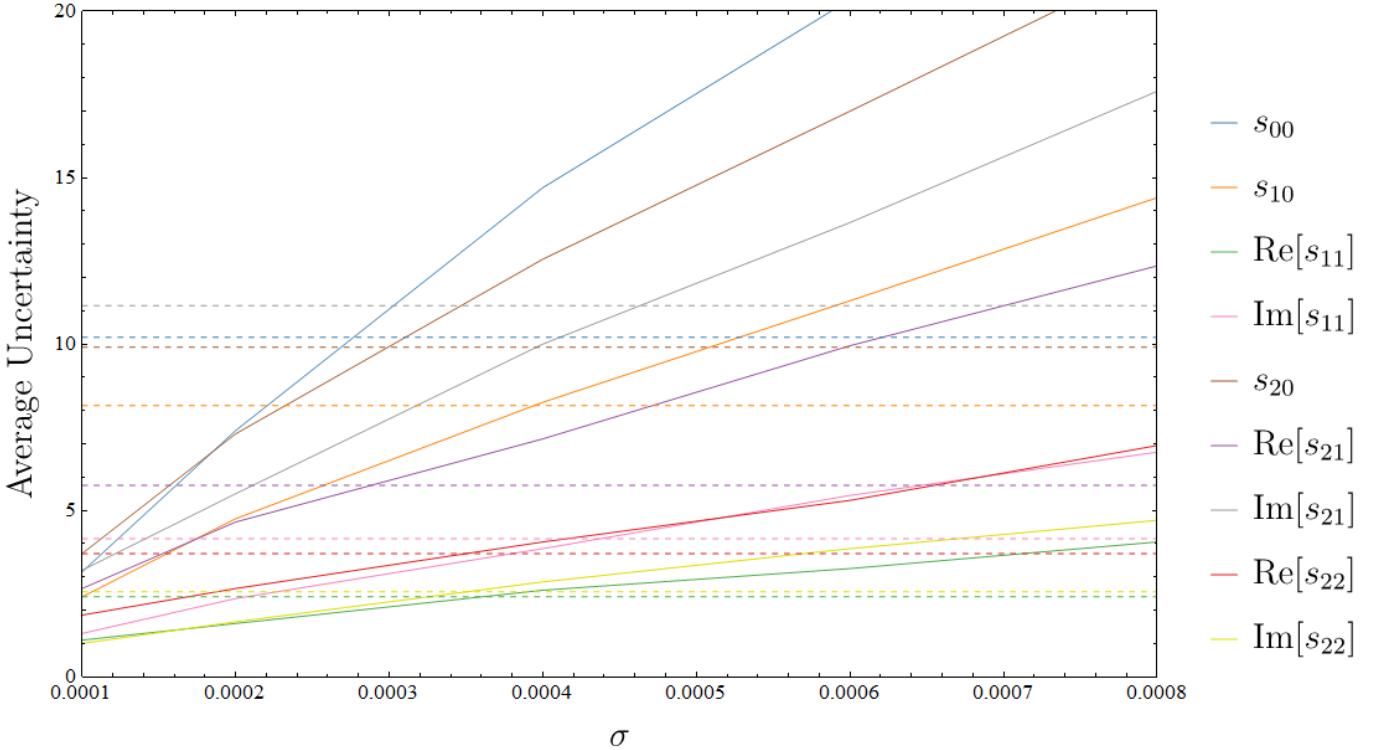


FIG. 2. A portion of the superimposed plot of average uncertainties produced by the random draw method (dashed lines) and the Bayesian method (solid lines). From this plot, we can see $\sigma = 0.0005$ is a potential choice for σ because every solid line is marginally above the corresponding dashed line of the same color.

Furthermore, the effects of choosing a bandwidth on the estimated density (and hence the remainder of the inference) is limited in the sense that the covariance matrix of the Gaussians is determined from the samples themselves with the bandwidth only acting as a scaling parameter that is usually of order unity. This additionally restricts the effects of user-controlled fine-tuning on the inference as compared to the narrow Gaussian method wherein the width of the Gaussian that approximates the delta function is completely determined by the user's choice. A more detailed discussion of this comparison between the two methods in the context of the MDC can be found in Sec. III F. For our chosen bandwidth approximation scheme (Scott's rule) the KDE method can be seen to perform extremely well in the MDC. For these reasons, we choose the KDE method for our final results on the SME constraints.

We present the results of the KDE method upon its use in the analysis of the events marked * listed in tables I through III (except for GW170817 for which we do not use the fixed posteriors due to the inability of the KDE to estimate very narrow densities) in Fig. 4 as our final result.

E. Discussion

The random draw method, which was originally presented in Ref. [7], is very time efficient. However, the number of events that can be used in the analysis is limited by the number of \bar{s}_{lm} to be estimated. Hence, it is only useful in the scenario wherein we have exactly the same number of events available to be used in the analysis as the number of \bar{s}_{lm} coefficients being simultaneously measured. On the other hand, unlike any of the following methods, this method does not involve any approximation of the delta distribution. Thus, this method can be used as a quick feasibility test. For example, we used this method to estimate the size of σ for the narrow Gaussian approach.

With the use of SVD, we were able to take data from a larger set of events. However, the presence of one “bad” event, i.e., a low-significance event with biased posteriors, can disturb the entire analysis since all events are given equal weights. With the SVD chi-square method, this problem is solved by weighting each event with its uncertainty in the v_g measurement. However, this leads to artificially narrower bounds with ambiguity in the meaning of those bounds.

The Bayesian methods are free of all aforementioned pathologies that plague the other methods. They can efficiently handle a large number of events and is unaffected

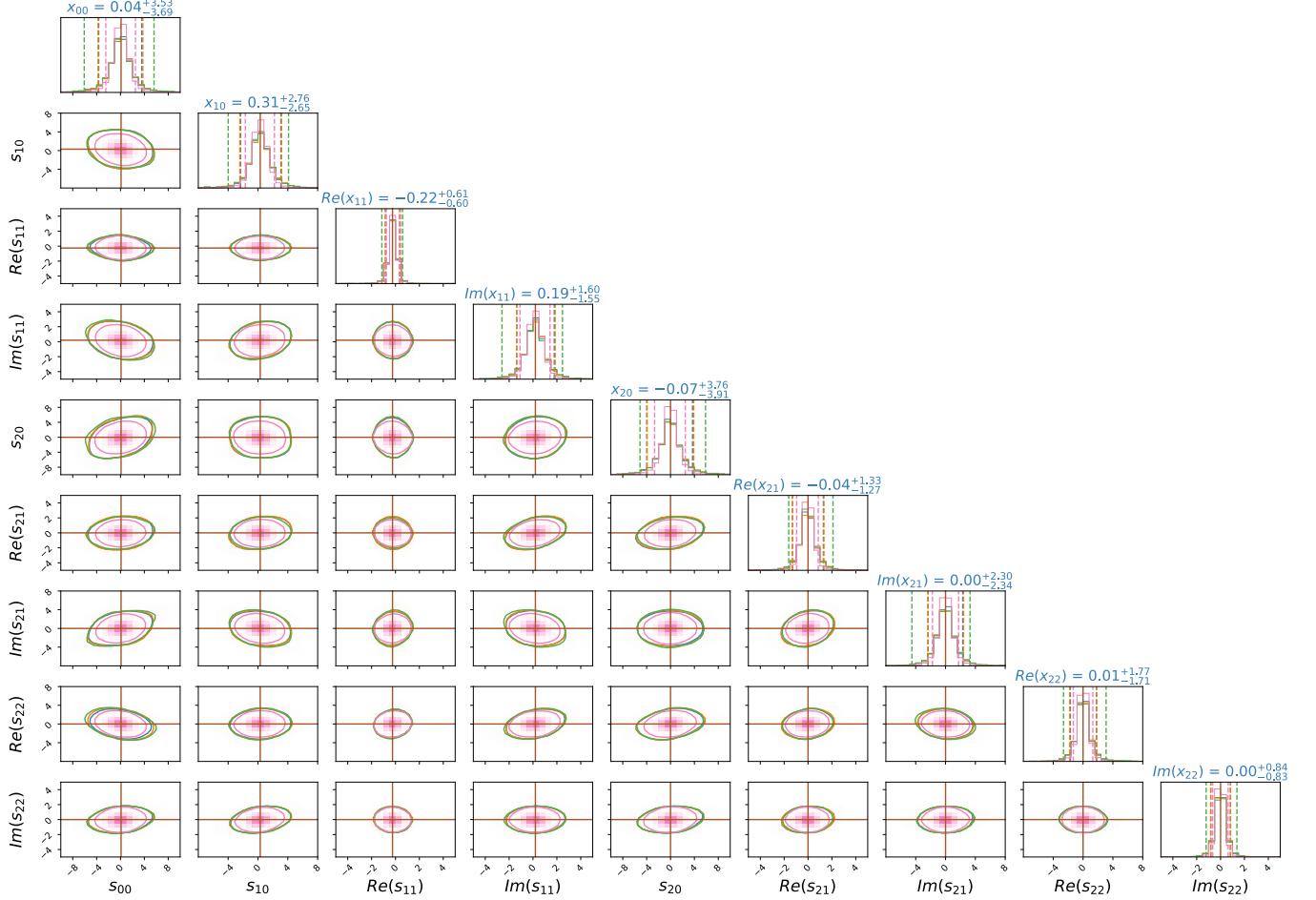


FIG. 3. Distribution of all nine \bar{s}_{lm} for the mock data. The exact Bayesian, KDE, narrow Gaussian, and SVD chi-square methods are color coded blue, orange, green, and magenta respectively. The mock true values of \bar{s}_{lm} , which were chosen to be either equal to or very close to zero, are marked by orange lines. Numbers above the plots show median values with 90% equal tail credible intervals. The displayed constraints were obtained from the inferred \bar{s}_{lm} samples using the PYTHON package CORNER [60]. Note that these mock events are “zero noise” in the sense that the mean values of the Gaussian in Eq. (31) are chosen to be equal to the true values and not perturbed by another random draw.

by a small number of bad events if any. Furthermore, the Bayesian credible intervals have the clear and unambiguous meaning of being regions of the parameter space that contain the true value of said parameters with a certain posterior probability given data.

Among the two approximate Bayesian techniques described in this paper, the narrow Gaussian method has the following issue. The process to find σ is cumbersome and somewhat subjective as it involves partitioning the set of events into subsets and estimating σ from plots, or developing an algorithm that needs to be compared to an independent method. On the other hand, the KDE method has less user-controlled fine-tuning than the narrow Gaussian method as it estimates its control parameter, i.e. the bandwidth, quantitatively from properties of the posterior samples themselves, with its variation having a much more restrictive effect on the estimated density. Hence, we claim that the Bayesian analysis im-

plemented by the KDE method produces the most trustworthy measurements of the SME coefficients.

F. Mock data challenge and comparison of methods

In this section, we describe the MDC that was set up to compare the different methods of SME measurements from GW data in order to verify our claims regarding them that were made in the previous section. To construct the MDC, we choose a fiducial value of the SME coefficients as their true values, say $\bar{s}_{lm,tr}$ and generate data for 15 mock events. The true sky positions of the mock events are chosen to be the mean values of the (θ, ϕ) samples of the real events.

We choose 15 of the “best” real events, i.e. the ones with the most precise sky localizations and v_g measurements, to be represented by our mock events in the MDC.

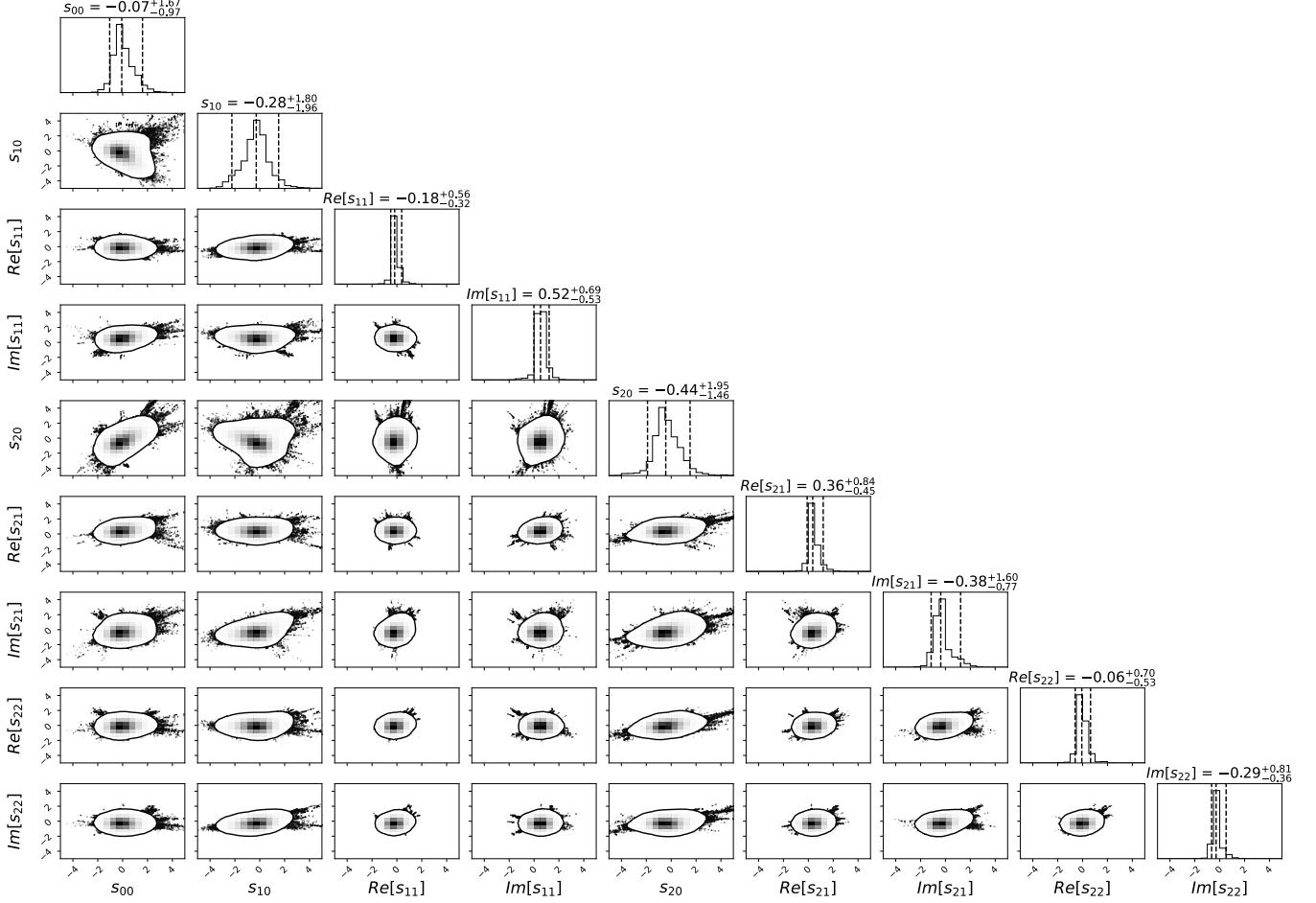


FIG. 4. Distribution of all nine \bar{s}_{lm} using the KDE method from the 24 chosen GW events in Tables I, II, and III. The numbers above the plots show median values with 90% equal tail credible intervals. Each contour represents the two-dimensional 90% credible interval marginalized over the other seven parameters while the discrete points represent posterior samples that lie outside of the 90% contour. The grayscale inside the contour represents the two-dimensional marginalized posterior density in the form of a histogram. The displayed constraints were obtained from the inferred \bar{s}_{lm} samples using the PYTHON package CORNER [60].

We then calculate the true value of Δv for each mock event from the true values of their sky positions and those of the SME coefficients. We then generate mock posterior samples of $(\theta, \phi, \Delta v)$ by adding uncorrelated Gaussian fluctuations to the true sky positions and true Δv values. The width of the fluctuations for each mock event is chosen to be the standard deviations of the posterior samples of the corresponding real event.

This allows us to create a controlled numerical experiment wherein we know the true answer. For Gaussian distributions of $\theta, \phi, \Delta v$ about known true values, one can write down the exact functional form of the likelihood of these parameters given mock data,

$$L_{\text{mdc}}(d_i|\Delta v', \theta, \phi) = \frac{1}{(2\pi)^{3/2}\sigma_{\theta,i}\sigma_{\phi,i}\sigma_{\Delta v,i}} \exp -\frac{1}{2} \left\{ \frac{(\theta - \theta_{tr,i})^2}{\sigma_{\theta,i}^2} + \frac{(\phi - \phi_{tr,i})^2}{\sigma_{\phi,i}^2} + \frac{(\Delta v' - \Delta v_{tr,i})^2}{\sigma_{\Delta v,i}^2} \right\}, \quad (31)$$

where $(\theta_{tr,i}, \phi_{tr,i})$ are the true values of the sky po-

sitions of the i th mock event; $\sigma_{\theta,i}, \sigma_{\phi,i}, \sigma_{\Delta v,i}$ are the widths of the Gaussian fluctuations used to generate the mock posterior samples of the i th mock event; and $\Delta v_{tr,i} = \Delta v(\bar{s}_{lm,tr}, \theta_{tr,i}, \phi_{tr,i})$. Knowledge of these quantities allows us to exactly write down and evaluate Eq. (31) as a function of θ, ϕ without any smoothing approximations.

We can then substitute $L_{\text{mdc}}(d_i|\Delta v', \theta, \phi)$ in place of $L(d_i|\Delta v', \theta, \phi)$ in Eq. (15) and carry out the integral numerically to obtain the “exact Bayesian” likelihood of our mock data given the SME coefficients:

$$L_{\text{mdc}}(\bar{s}_{lm}) = \prod_i \frac{1}{(2\pi)^{3/2}\sigma_{\theta,i}\sigma_{\phi,i}\sigma_{\Delta v,i}} \int \exp -\frac{1}{2} \left\{ \frac{(\theta - \theta_{tr,i})^2}{\sigma_{\theta,i}^2} + \frac{(\phi - \phi_{tr,i})^2}{\sigma_{\phi,i}^2} + \frac{(\Delta v(\bar{s}_{lm}, \theta, \phi) - \Delta v(\bar{s}_{lm,tr}, \theta_{tr,i}, \phi_{tr,i}))^2}{\sigma_{\Delta v,i}^2} \right\} d\theta d\phi. \quad (32)$$

We can then sample the likelihood in Eq. (32), after applying suitable priors on \bar{s}_{lm} , using the MCMC techniques described above and obtain what can be thought of as the true posterior distribution of the SME coefficients given the mock data. We can then compute the constraints obtained from the approximate methods being applied to the mock posterior samples and compare those results with the true posterior.

The results of this comparison are displayed in Fig. 3. We can see that the KDE method agrees remarkably well with the exact Bayesian method, while the narrow Gaussian method deviates from it slightly for some coefficients. We note that a different choice of σ for the narrow Gaussian method leading to better agreement with the exact Bayesian result is possible. However, we conclude that the KDE method's agreement with the exact Bayesian method, independent of any external fine-tuning, justifies its use on the real data for producing our final result. We also note that our claim regarding the SVD chi-square method producing artificially narrower bounds is also verified by this comparison.

G. Final results

With the 24 chosen GW events in Tables I, II, and III, we are able to constrain all nine \bar{s}_{lm} coefficients. We obtain the results shown in Fig. 4.

Note that the measurements of \bar{s}_{lm} shown in Fig. 4 are consistent with zero. Given that zero lies within the 90% credible intervals for all coefficients, we consider these results to be consistent with existing constraints on \bar{s}_{lm} [9].

These limits are considerably weaker than some found in the literature. However, they are valuable as independent tests. Moreover, this is also the first attempt to simultaneously constrain all \bar{s}_{lm} using GW measurements, thus putting direct limits on the full potential anisotropy of the speed of GWs. Additionally, this method can theoretically incorporate as many events as available and thus improve in precision as additional high-quality events become available.

IV. CONCLUSION

In our study, we selected 52 high-SNR gravitational-wave events that were detected by at least two detectors from the first three observing runs of Advanced LIGO and Advanced Virgo. We used `LALINFERENCENEST` and `LALINFERENCEMCMC` to construct posterior distributions of the speed of GWs for each event. We found the 90% credible interval of the combined v_g posterior distribution to be $0.99^{+0.01}_{-0.02}c$. This interval is narrower than the similar one constructed with O1 and O2 events in previous studies, suggesting a more precise measurement of v_g [7]. However, even with the inclusion of a high-SNR BNS event GW170817 with its pinpoint sky local-

ization, we were only able to narrow the 90% credible interval to $0.99^{+0.01}_{-0.02}c$. We then explored multiple methods of extracting SME constraints from v_g -like data. Based on the conclusions of that investigation, we used hierarchical Bayesian inference implemented with KDE methods to simultaneously constrain all nine coefficients for Lorentz violation in the SME framework. The resultant constraints did not exhibit evidence for Lorentz violation. We are optimistic about the possibility of further improvements in speed of gravity and associated Lorentz violation measurements in the future. The search is likely to be aided by the combination of additional detectors at additional locations around the Earth and by the possibility of combining results achieved by the methods presented here with those from additional multimessenger events.

Appendix A: Robustness against choice of waveform approximants

In this section, we elaborate on the insensitivity of our results to the choice of waveform approximants. All measurements presented in this work rely on the posterior samples of v_g, θ, ϕ that are obtained from single-event parameter estimation runs. Samples of these parameters are implicitly computed from those of direct observables such as the time delay between the arrival of signals at different detectors ($\Delta t_{HL}, \Delta t_{HV}, \Delta t_{LV}$), the effective distance to the source as measured by each detector ($D_{eff,H}, D_{eff,L}, D_{eff,V}$), and the coalescence phase (ϕ_0). In Fig. 5, we show that parameter estimation results for such observables are fully consistent between different choices of the waveform approximant for a randomly chosen event in O3 (GW200115). The samples are obtained from LVK's public data release. We note that fixing $v_g = c$ (as was done for the publicly released samples) or allowing it to vary does not affect these observables since v_g is a derived quantity while the observables are directly measurable.

ACKNOWLEDGMENTS

This work was supported by NSF Grants No. PHY-1806990, No. PHY-1912649, and No. PHY-2207728. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The analysis in this work used the LALSuite software library [18]. We thank LIGO and Virgo Collaboration for providing the data from the first, second, and third observing runs. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. We thank Nathan Johnson-McDaniel for helpful comments during a careful review of the manuscript.

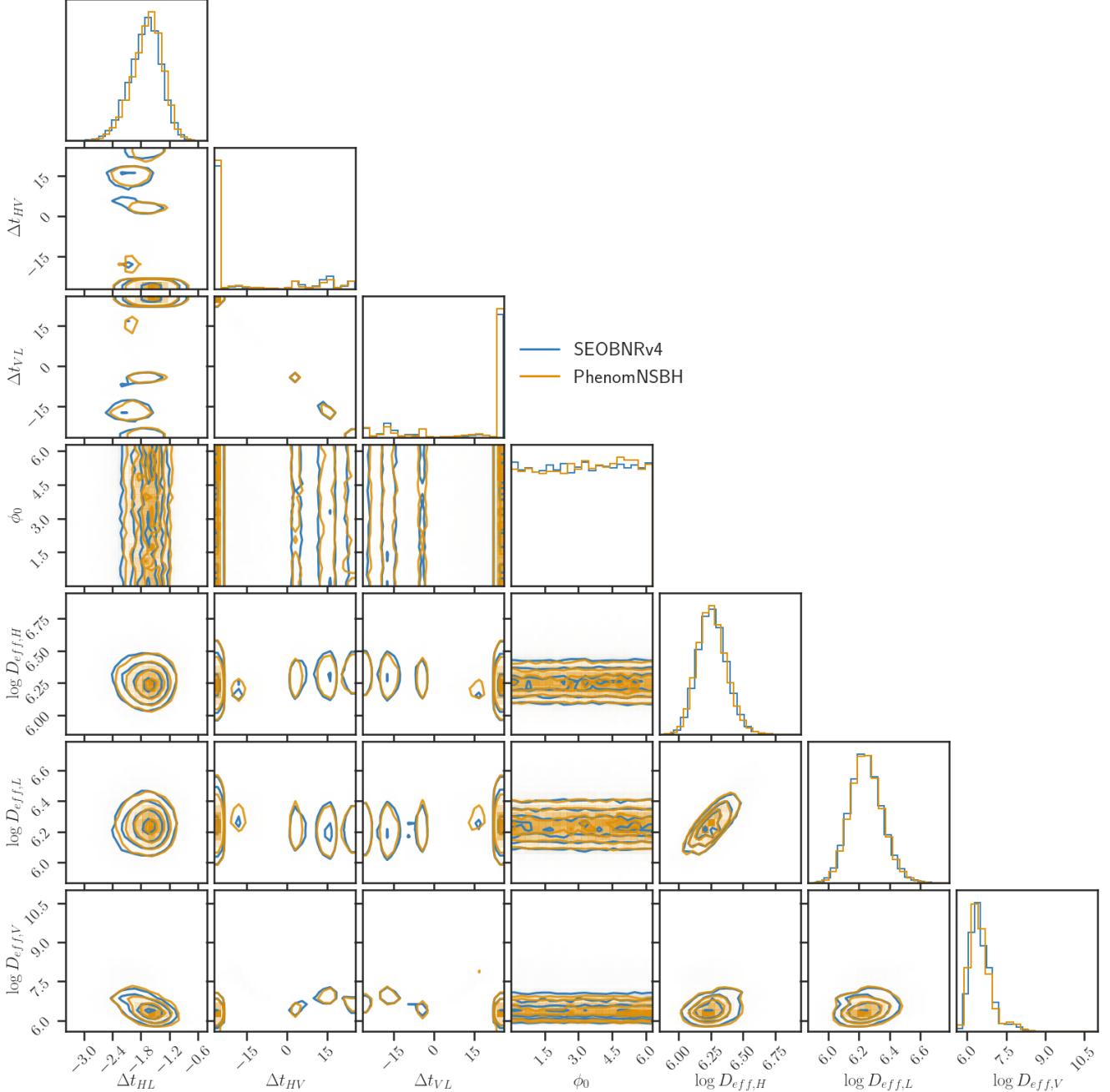


FIG. 5. Insensitivity of observables to the choice of waveform approximants for a randomly selected O3b event (GW200115).

This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Ad-

vanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Min-

istry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

[1] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, *Class. Quantum Grav.* **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[2] F. Acernese *et al.*, Advanced Virgo: a second-generation interferometric gravitational wave detector, *Class. Quantum Grav.* **32**, 024001 (2014), arXiv:1408.3978 [gr-qc].

[3] R. Abbott *et al.* (LIGO–Virgo Collaboration), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. X* **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].

[4] R. Abbott *et al.* (LIGO–Virgo–KAGRA Collaboration), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run (2021), arXiv:2111.03606 [gr-qc].

[5] B. P. Abbott *et al.* (LIGO–Virgo Collaboration), GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, *Phys. Rev. X* **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[6] N. Cornish, D. Blas, and G. Nardini, Bounding the speed of gravity with gravitational wave observations, *Phys. Rev. Lett.* **119**, 161102 (2017), arXiv:1707.06101 [gr-qc].

[7] X. Liu, V. F. He, T. M. Mikulski, D. Palenova, C. E. Williams, J. Creighton, and J. D. Tasson, Measuring the speed of gravitational waves from the first and second observing run of Advanced LIGO and Advanced Virgo, *Phys. Rev. D* **102**, 024028 (2020), arXiv:2005.03121 [gr-qc].

[8] B. P. Abbott *et al.* (LIGO–Virgo, Fermi–GBM, and INTEGRAL Collaborations), Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, *Astrophys. J. Lett.* **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].

[9] V. A. Kostelecký and N. Russell, “Data tables for Lorentz and CPT violation”, 2024 Edition, arXiv:0801.0287 [hep-ph].

[10] D. Colladay and V. A. Kostelecký, Lorentz-violating extension of the Standard Model, *Phys. Rev. D* **58**, 116002 (1998), arXiv:hep-ph/9809521 [hep-ph].

[11] V. A. Kostelecký, Gravity, Lorentz violation, and the Standard Model, *Phys. Rev. D* **69**, 105009 (2004), arXiv:hep-ph/0312310 [hep-ph].

[12] Q. G. Bailey and V. A. Kostelecký, Signals for Lorentz violation in post-Newtonian gravity, *Phys. Rev. D* **74**, 045001 (2006), arXiv:1905.00409 [gr-qc].

[13] L. Haegel, K. O’Neal-Ault, Q. G. Bailey, J. D. Tasson, M. Bloom, and L. Shao, Search for anisotropic, birefringent spacetime-symmetry breaking in gravitational wave propagation from GWTC-3, *Phys. Rev. D* **107**, 064031 (2023), arXiv:2210.04481 [gr-qc].

[14] R. Niu, T. Zhu, and W. Zhao, Testing Lorentz invariance of gravity in the Standard-Model Extension with GWTC-3, *J. Cosmol. Astropart. Phys.* **2022** (12), 011, arXiv:2202.05092 [gr-qc].

[15] C. Gong, T. Zhu, R. Niu, Q. Wu, J.-L. Cui, X. Zhang, W. Zhao, and A. Wang, Gravitational wave constraints on nonbirefringent dispersions of gravitational waves due to Lorentz violations with GWTC-3 events, *Phys. Rev. D* **107**, 124015 (2023), arXiv:2302.05077 [gr-qc].

[16] R. Ghosh, S. Nair, L. Pathak, S. Sarkar, and A. S. Sengupta, Test of the Second Postulate of Relativity from Gravitational Wave Observations (2023), arXiv:2304.14820 [gr-qc].

[17] B. P. Abbott *et al.* (LIGO–Virgo Collaboration), Multi-messenger observations of a binary neutron star merger, *Astrophys. J. Lett.* **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].

[18] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, *LVK Algorithm Library - LALSuite*, Freesoftware(GPL) (2018).

[19] R. Abbott *et al.* (LIGO–Virgo–KAGRA Collaboration), Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo, *SoftwareX* **13**, 100658 (2021), arXiv:1912.11716 [gr-qc].

[20] R. Abbott *et al.* (LIGO–Virgo–KAGRA Collaboration), Open data from the third observing run of LIGO, Virgo, KAGRA and GEO (2023), arXiv:2302.03676 [gr-qc].

[21] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).

[22] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97 (1970).

[23] J. Veitch *et al.*, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library, *Phys. Rev. D* **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[24] J. Veitch and A. Vecchio, Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network, *Phys. Rev. D* **81**, 062003 (2010), arXiv:0911.3820 [astro-ph.CO].

[25] J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Anal.* **1**, 833 (2006).

[26] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, *Phys. Rev. D* **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[27] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, *Phys. Rev. D* **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].

[28] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple model of complete precessing black-hole-binary gravitational waveforms, *Phys. Rev. Lett.* **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].

[29] K. G. Arun, A. Buonanno, G. Faye, and E. Ochsner,

Higher-order spin effects in the amplitude and phase of gravitational waveforms emitted by inspiraling compact binaries: Ready-to-use gravitational waveforms, *Phys. Rev. D* **79**, 104023 (2009), arXiv:0810.5336 [gr-qc].

[30] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, Comparison of post-newtonian templates for compact binary inspiral signals in gravitational-wave detectors, *Phys. Rev. D* **80**, 084043 (2009), arXiv:0907.0700 [gr-qc].

[31] B. Mikóczki, M. Vasúth, and L. A. Gergely, Self-interaction spin effects in inspiralling compact binaries, *Phys. Rev. D* **71**, 124043 (2005), arXiv:astro-ph/0504538 [astro-ph].

[32] J. Vines, E. E. Flanagan, and T. Hinderer, Post-1-newtonian tidal effects in the gravitational waveform from binary inspirals, *Phys. Rev. D* **83**, 084051 (2011), arXiv:1101.1673 [gr-qc].

[33] A. Bohé, S. Marsat, and L. Blanchet, Next-to-next-to-leading order spin-orbit effects in the gravitational wave flux and orbital phasing of compact binaries, *Class. Quantum Grav.* **30**, 135009 (2013), arXiv:1303.7412 [gr-qc].

[34] A. Bohé, G. Faye, S. Marsat, and E. K. Porter, Quadratic-in-spin effects in the orbital dynamics and gravitational-wave energy flux of compact binaries at the 3pn order, *Class. Quantum Grav.* **32**, 195010 (2015), arXiv:1501.01529 [gr-qc].

[35] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, *Phys. Rev. Res.* **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].

[36] R. Abbott *et al.* (LIGO–Virgo Collaboration), GW190521: A binary black hole merger with a total mass of $150 M_{\odot}$, *Phys. Rev. Lett.* **125**, 101102 (2020).

[37] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall/CRC, 1986).

[38] D. W. Scott, *Multivariate density estimation* (Wiley, 1992).

[39] P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nature Methods* **17**, 261 (2020), arXiv:1907.10121 [cs.MS].

[40] E.-J. Wagenmakers, T. Lodewyckx, H. Kuriyal, and R. Grasman, Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method, *Cogn. Psychol.* **60**, 158 (2010).

[41] R. Abbott *et al.* (LIGO–Virgo Collaboration), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run (2021), arXiv:2108.01045 [gr-qc].

[42] B. P. Abbott *et al.* (LIGO–Virgo Collaboration), GW170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[43] V. A. Kostelecký and M. Mewes, Testing local Lorentz invariance with gravitational waves, *Phys. Lett. B* **757**, 510 (2016), arXiv:1602.04782 [gr-qc].

[44] N. A. Flowers, C. Goode, and J. D. Tasson, Superconducting-gravimeter tests of local Lorentz invariance, *Phys. Rev. Lett.* **119**, 201101 (2017), arXiv:1612.08495 [gr-qc].

[45] H. Pihan-le Bars *et al.*, New Test of Lorentz invariance using the MICROSCOPE space mission, *Phys. Rev. Lett.* **123**, 231102 (2019), arXiv:1912.03030 [physics.space-ph].

[46] V. A. Kostelecký and J. D. Tasson, Constraints on Lorentz violation from gravitational Čerenkov radiation, *Phys. Lett. B* **749**, 551 (2015), arXiv:1508.07007 [gr-qc].

[47] J. S. Díaz, V. A. Kostelecký, and M. Mewes, Testing relativity with high-energy astrophysical neutrinos, *Phys. Rev. D* **89**, 043005 (2014), arXiv:1308.6344 [astro-ph.HE].

[48] K. N. Lau and M. D. Seifert, Direct-coupling lensing by antisymmetric tensor monopoles, *Phys. Rev. D* **95**, 025023 (2017), arXiv:1309.2241 [hep-th].

[49] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Górski, Healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python, *J. Open Source Softw.* **4**, 1298 (2019).

[50] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, *Astrophys. J.* **622**, 759 (2005), arXiv:astro-ph/0409513 [astro-ph].

[51] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, *Numer. Math.* **14**, 403 (1970).

[52] I. Mandel, W. M. Farr, and J. R. Gair, Extracting distribution parameters from multiple uncertain observations with selection biases, *Mon. Not. R. Astron. Soc.* **486**, 1086 (2019), arXiv:1809.02063 [physics.data-an].

[53] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, Inferring the properties of a population of compact binaries in presence of selection effects, in *Handbook of Gravitational Wave Astronomy* (Springer Singapore, Singapore, 2020) pp. 1–60, arXiv:2007.05579 [astro-ph.IM].

[54] E. Payne, S. Banagiri, P. D. Lasky, and E. Thrane, Searching for anisotropy in the distribution of binary black hole mergers, *Phys. Rev. D* **102**, 102004 (2020), arXiv:2006.11957 [astro-ph.CO].

[55] C. Messick *et al.*, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, *Phys. Rev. D* **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].

[56] G. S. Davies, T. Dent, M. Tápai, I. Harry, C. McIsaac, and A. H. Nitz, Extending the PyCBC search for gravitational waves from compact binary mergers to a global network, *Phys. Rev. D* **102**, 022004 (2020), arXiv:2002.08291 [astro-ph.HE].

[57] W. M. Farr, Accuracy requirements for empirically measured selection functions, *Res. Notes Am. Astron. Soc.* **3**, 66 (2019), arXiv:1904.10879 [astro-ph.IM].

[58] B. P. Abbott *et al.* (LIGO–Virgo Collaboration), Binary black hole population properties inferred from the first and second observing runs of Advanced LIGO and Advanced Virgo, *Astrophys. J.* **882**, L24 (2019), arXiv:1811.12940 [astro-ph.HE].

[59] R. Abbott *et al.* (LIGO–Virgo–KAGRA Collaboration), Population of merging compact binaries inferred using gravitational waves through GWTC-3, *Phys. Rev. X* **13**, 011048 (2023), arXiv:2111.03634 [astro-ph.HE].

[60] D. Foreman-Mackey, corner.py: Scatterplot matrices in python, *J. Open Source Softw.* **1**, 24 (2016).