Investigating the source of inferred parameters' discrepancies between LIGO detectors in GW231123 LIGO SURF 2025 Final report

Krzysztof Król[1,2] and Mentor: Sophie Bini[1]

[1]LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA
[2]Astronomical Observatory, University of Warsaw, Al. Ujazdowskie 4, 00-478 Warszawa, Poland

ABSTRACT

GW231123 is a short-duration gravitational-wave (GW) signal, consistent with a binary black hole merger with a total mass of 190-265 $M_\odot$. It is the most massive binary black hole confidently observed to date. Both components are highly spinning and likely have masses in the mass gap caused by pair-instability supernova processes. The event is challenging to analyze because of its very short duration and the difficulties in modeling such extreme systems. Additionally, significant differences between inferred posteriors for total mass and spins arise when parameter estimation is performed using LIGO Hanford-only and LIGO Livingston-only data, raising concerns about the presence of spurious transient noise overlapping with the GW signal. In this project, we quantify these differences and investigate their causes through simulated signals similar to GW231123. We find that Gaussian noise fluctuations could be the cause of the observed differences, and we deem this a plausible explanation.

1. INTRODUCTION

GW231123 is a transient gravitational-wave (GW) event observed on 23 November 2023 by the two Advanced LIGO detectors (Virgo was offline at the time of detection). The signal is consistent with the coalescence of a binary black hole (BBH) system with total mass of between 190-265 $M_\odot$, which makes it the most massive merger observed to date with high confidence. In addition to being very massive, both component black holes are likely very highly spinning and exhibit significant precession (The LIGO Scientific Collaboration et al. 2025a).

GW231123 might have very important astrophysical implications. Due to pair-instability supernova process (Farmer et al. 2019), it is believed that black holes with masses between 60 $M_\odot$ and 130 $M_\odot$ cannot form as a result of standard stellar collapse. The possible formation channels for black holes with masses in this range include hierarchical merger as well as other stellar and binary evolution processes (The LIGO Scientific Collaboration et al. 2025a; Gerosa & Fishbach 2021; Mapelli et al. 2020; Costa et al. 2021; Spera et al. 2019).

In the case of the previous most-massive merger, GW190521 (Abbott et al. 2020a,b), the primary black hole is believed to lie in the mass gap. For GW231123, the primary is within or above the mass gap produced by the pair-instability supernova process, while the secondary is within the mass gap with a probability of 83%. Furthermore, GW231123 is the second event (after GW190521) observed that produced a remnant that would be classified as an intermediate mass black hole (a black hole with mass between $10^2 M_\odot$ and $10^5 M_\odot$) (The LIGO Scientific Collaboration et al. 2025a).

Inferring the properties of the source of this exceptional event is particularly challenging due to its high mass. The detected signal's frequency is noticeably lower than other events, peaking at close to 50 Hz. The event is also much shorter, with only a few observable cycles before merger and the entire event being shorter than 0.2 seconds.

The data from the LIGO–Virgo–Kagra network is used to infer the source properties. In particular cases in which the GW event overlaps with transient noise, inference using only one detector has been done in the past, such as in Payne et al. (2022). Such analysis has also been performed for GW231123. The LIGO Scientific Collaboration et al. (2025a) noticed differences in the source properties inferred from LIGO Livingston-only data and from LIGO Hanford-only data, especially in inferred spins. In this project, we quantify these differences and investigate their possible sources.

The paper is structured as follows: In Section 2, we list and define the parameters inferred in our analysis, describe the Bayesian framework used for sampling and parameter estimation, and introduce the Jensen–Shannon divergence as a measure of similarity between posterior distributions. In Section 3, we present the inferred properties of GW231123 and quantify the differences between the LIGO Hanford–only and LIGO Livingston–only pa-

k.krol22@student.uw.edu.pl

bini@caltech.edu

rameter estimation runs. In Section 4, we place these results in context by comparing them with two additional gravitational-wave events. In Section 5, we analyze simulated injections to assess the origin and magnitude of similar differences. Finally, in Section 6, we summarize our findings and discuss possible directions for future work.

2. METHODS

2.1. *Binary black hole coalescence parameters*

The main difficulty in parameter estimation from binary black hole GW signal comes from the multidimensionality of the data. There are 15 total parameters that describe a binary black hole waveform (for a quasicircular system), namely:

- masses of the two black holes m_1 and m_2 ,
- spins $\vec{\chi_1}$ and $\vec{\chi_2}$, expressed by the dimensionless spin vector $\vec{\chi_i} = \vec{S_i}c/(Gm_i^2)$ $(0 \leq |\vec{\chi_i}| < 1)$, where $\vec{S_i}$ is the angular moment of the black hole,
- luminosity distance $D_{\rm L}$,
- inclination angle ι (angle between observer's line of sight and orbital plane of the system),
- coalescence time t_0 and phase φ,
- polarization angle ψ,
- source sky position θ and ϕ.

Due to the possibility of more precise measurement and more direct impact on the waveform we also define chirp mass $\mathcal{M}$:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

for $m_1 > m_2$. Due to redshift, the source-frame mass $m_i^{\rm source}$ (as well as total mass and chirp mass) is different than the measured detector-frame mass $m_i = (1+z)m_i^{\rm source}$, where z denotes redshift. We will label the detector-frame masses with no superscript and label the source-frame masses with the superscript $^{\rm source}$ throughout the rest of the report.

In literature, it is common to define the spin parameters $\chi_{\rm eff}$ and $\chi_{\rm p}$, which are the effective aligned spin and effective precessing spin. They are defined as (Ng et al. 2018; Thomas et al. 2021):

$$\chi_{\rm eff} = \frac{\vec{S_1}/m_1 + \vec{S_2}/m_2}{m_1 + m_2} \cdot \vec{L}$$

and

$$\chi_{\rm p} = \frac{\max(A_1|\vec{S_1}|,\ A_2|\vec{S_2}|)}{A_1 m_1^2},$$

where $A_1 = 2 + 3/2q$ and $A_2 = 2 + 3q/2$. $\chi_{\rm eff}$ is bounded by $[-1, 1]$ and $\chi_{\rm p}$ by $[0, 1]$ to avoid naked singularities. Furthermore, $\chi_{\rm eff}$ is a constant of motion up to at least the second post-Newtonian order (Racine 2008; Blanchet 2014) and typically can be measured more precisely than individual spins.

2.2. *Bayesian statistics*

To infer the properties of GW signals, we use Bayesian statistics. Bayesian statistics is based on an interpretation of probability, where, unlike in the frequentist approach, probability expresses a degree of belief. It is achieved by combining prior knowledge with new data to compute new probabilities ("posterior") according to Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

where $P(A)$ is the prior probability, $P(B)$ is the probability of new observations, $P(B|A)$ the likelihood, and $P(A|B)$ the posterior probability. In the context of parameter estimation, Bayes' theorem can be rewritten as

$$p(\theta|d) = \frac{L(d|\theta)\pi(\theta)}{\int L(d|\theta)\pi(\theta)\,{\rm d}\theta},$$

where observations are denoted by d, unknown parameters are denoted by θ, L is the likelihood function and π is the prior probability density function. The denominator serves as a normalization constant (Christensen & Meyer 2022).

2.3. *Data preprocessing and sampling settings*

We perform our parameter estimation (PE) using whitened, bandpassed data. A glitch in LIGO Hanford occurred 1.7-1.1 s before the event in a frequency range between 15-30 Hz. It was modeled along with the BBH waveform using BayesWave (Cornish & Littenberg 2015; Cornish et al. 2021; Chatziioannou et al. 2021). Another glitch occurred in LIGO Livingston 3.0-2.0 s in a frequency range between 10-20 Hz. This glitch, due to low frequency and the extended time between the event and the glitch, was not removed from the LIGO Livingston strain data.

We perform PE using `bilby` (Ashton et al. 2019), a Python library for GW-related Bayesian inference. It allows one to perform parameter estimation for a GW signal using nested sampling. We are using the same power spectral density (PSD), likelihood, and sampler settings as for the GWTC-4 catalog (The LIGO Scientific Collaboration et al. 2025b).

2.4. *Comparing posterior distributions*

To understand the differences between posterior distributions resulting from different PE runs, we will visually examine the corner plots that contain the relevant parameters and try to quantify them. To do the latter, we will use the Jensen–Shannon (JS) divergence between two posterior distributions Q_1 and Q_2, which is defined as (Kullback & Leibler 1951; Menéndez et al. 1997; Endres & Schindelin 2003; Lin 1991):

$$\mathrm{JS}(Q_1||Q_2) = \frac{1}{2}\mathrm{D}_{\mathrm{KL}}(Q_1||M) + \frac{1}{2}\mathrm{D}_{\mathrm{KL}}(Q_2||M),$$

where $M = \frac{1}{2}(Q_1 + Q_2)$ is a mixture distribution of Q_1 and Q_2, and $\mathrm{D}_{\mathrm{KL}}(Q_1||Q_2)$ is the Kullback–Leibler divergence defined as

$$\mathrm{D}_{\mathrm{KL}}(Q_1||Q_2) = \sum_{x\in X} Q_1(x)\log_2\frac{Q_1(x)}{Q_2(x)}.$$

We will use the convention in which the logarithm base in Kullback–Leibler divergence is 2. Then the JS divergence is a real number in the range $[0, 1]$ for discrete probability distributions (as is the case for numerically calculated posterior distributions), where $\mathrm{JS} = 0$ if and only if the two distributions are identical.

Due to different PE seeds affecting different runs, noise and numerical estimations of posterior distributions we expect non-zero JS divergences even for two different PE runs performed on the same data. Typically, two distributions with $\mathrm{JS} < 0.05$ would be considered to be in good agreement (Abbott et al. 2019).

3. SOURCE PROPERTIES OF GW231123

3.1. *Waveform approximant choice and ASD difference quantification*

The LIGO Scientific Collaboration et al. (2025a) found that the NRSur7dq4 (Varma et al. 2019) waveform model is on average more accurate than all other waveform models tested (IMRPhenomXPHM, IMRPhenomXO4a, IMRPhenomTPHM, SEOBNRv5PHM) for events similar to GW231123. We first attempted using NRSur7dq4 to perform PE and for further studies, but quickly discovered that its computational cost along with current LIGO Data Grid's load significantly slowed us down. We therefore decided to use IMRPhenomXPHM (Pratten et al. 2021), which offers close to 5× computational speedup. We offer more insight on waveform systematics in the Appendix in Section A.

We later prepared further visualizations of the data. We plotted the whitened strain along with the maximum likelihood IMRPhenomXPHM waveforms resulting from full, Hanford-only and Livingston-only PE in Figure 1. There is only a slight difference between the LIGO Livingston-only estimated waveform and the estimated waveform using both detectors, which can be expected given that the network inference is driven by LIGO Livingston due to its higher SNR.

3.2. *Comparing PE results*

Table 1. Jensen–Shannon divergences of selected parameters' posteriors between LIGO Hanford-only and LIGO Livingston-only PE runs for GW231123.

Parameter	JS of H1 vs. L1
luminosity distance D_{L}	0.88
viewing angle θ_{JN}	0.72
eff. aligned spin parameter χ_{eff}	0.69
phase φ	0.67
source-frame total mass	0.63
detector-frame primary mass	0.54
detector-frame total mass	0.54
source-frame primary mass	0.53
inclination angle ι	0.51
source-frame secondary mass	0.36
eff. precessing spin parameter χ_{p}	0.34
mass ratio q	0.15
detector-frame secondary mass	0.14
primary spin magnitude χ_1	0.03
secondary spin magnitude χ_2	0.00

We present PE posteriors for selected parameters from LIGO Hanford-only and LIGO Livingston-only data PE runs in Figure 2. We also present JS divergences for selected parameters from these PE runs in Table 1. The estimated parameter values from PE along with 90% confidence intervals are shown in Table 2.

We will be denoting the parameters inferred using LIGO Hanford-only data and LIGO Livingston-only data using superscripts $^{\mathrm{H1}}$ and $^{\mathrm{L1}}$, respectively. The median masses differ noticeably (detector-frame total mass values of $M^{\mathrm{H1}} = 330^{+36}_{-55}\,M_\odot$ vs. $M^{\mathrm{L1}} = 286^{+20}_{-26}\,M_\odot$), but the differences in luminosity distance (as well as inclination angle, coupled with it) and spins are the most striking. LIGO Hanford-only data favors medium values of $\chi^{\mathrm{H1}}_{\mathrm{eff}} = 0.48^{+0.23}_{-0.50}$ and $\chi^{\mathrm{H1}}_{\mathrm{p}} = 0.45^{+0.35}_{-0.26}$, while the LIGO Livingston-only data favor low $\chi^{\mathrm{L1}}_{\mathrm{eff}} = 0.03^{+0.15}_{-0.22}$ and very high precession with $\chi^{\mathrm{H1}}_{\mathrm{p}} = 0.72^{+0.21}_{-0.23}$.

Typically, the luminosity distance and the inclination can be degenerate, but this degeneracy can be lifted in precessing systems (Usman et al. 2019). Us-

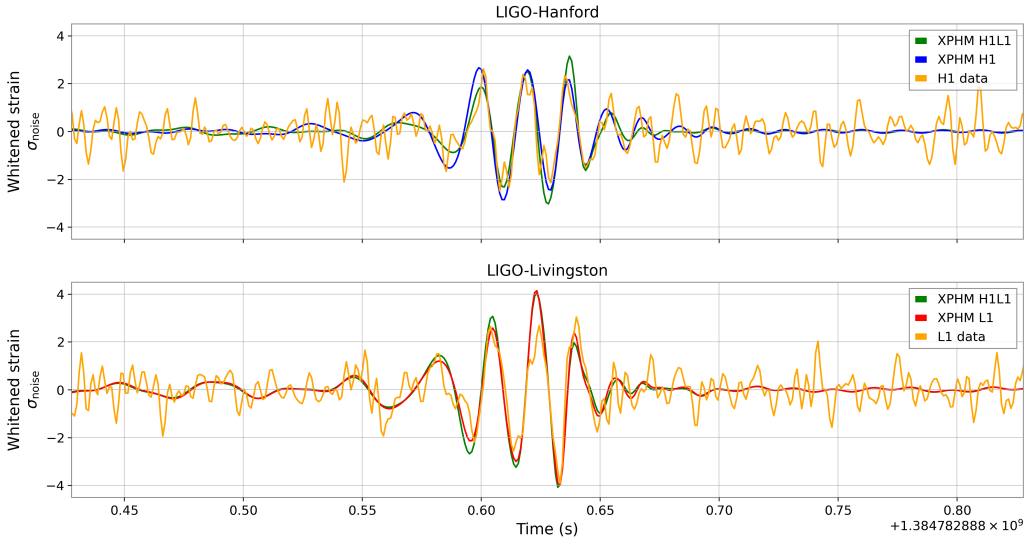

Figure 1. Whitened strain bandpassed to the frequency range [20Hz, 256Hz] (orange) plotted with best fitting IMRPhenomXPHM waveform models from LIGO Hanford+LIGO Livingston (green), LIGO Hanford-only (blue) and LIGO Livingston-only (red) PE. The y-axis is in noise standard deviation units.

Table 2. Posterior values of selected parameters from 3 PE runs for GW231123 with 90% confidence intervals.

Parameter	H1	L1	H1+L1
luminosity distance $D_{\rm L}$ (Mpc)	4064^{+2905}_{-2758}	568^{+390}_{-271}	851^{+349}_{-293}
primary spin magnitude χ_1	$0.79^{+0.18}_{-0.40}$	$0.75^{+0.20}_{-0.26}$	$0.79^{+0.18}_{-0.26}$
secondary spin magnitude χ_2	$0.60^{+0.36}_{-0.53}$	$0.54^{+0.40}_{-0.48}$	$0.68^{+0.28}_{-0.55}$
eff. aligned spin parameter $\chi_{\rm eff}$	$0.48^{+0.23}_{-0.50}$	$0.03^{+0.15}_{-0.22}$	$0.05^{+0.16}_{-0.21}$
eff. precessing spin parameter $\chi_{\rm P}$	$0.45^{+0.35}_{-0.26}$	$0.72^{+0.21}_{-0.23}$	$0.74^{+0.19}_{-0.24}$
inclination angle ι (rad)	$1.53^{+1.35}_{-1.27}$	$1.47^{+0.69}_{-0.55}$	$1.41^{+0.80}_{-0.56}$
log likelihood	$90.2^{+3.0}_{-4.6}$	$149.4^{+3.7}_{-5.0}$	$244.9^{+3.7}_{-5.4}$
optimal SNR	$13.4^{+1.7}_{-1.7}$	$17.5^{+1.7}_{-1.7}$	$22.2^{+1.6}_{-1.7}$
source-frame primary mass ($M_\odot$)	128^{+38}_{-31}	157^{+14}_{-13}	150^{+13}_{-12}
source-frame secondary mass ($M_\odot$)	69^{+31}_{-26}	98^{+22}_{-27}	94^{+19}_{-20}
source-frame total mass ($M_\odot$)	330^{+36}_{-55}	286^{+20}_{-26}	285^{+21}_{-21}
mass ratio q	$0.55^{+0.35}_{-0.28}$	$0.62^{+0.15}_{-0.17}$	$0.63^{+0.12}_{-0.13}$
geocentric time since 1384782888 (s)	$0.6018^{+0.0240}_{-0.0217}$	$0.6020^{+0.0258}_{-0.0166}$	$0.6268^{+0.0026}_{-0.0035}$
phase φ (rad)	$2.89^{+2.57}_{-2.10}$	$5.10^{+0.43}_{-0.50}$	$5.03^{+0.60}_{-0.58}$
viewing angle $\theta_{\rm JN}$ (rad)	$1.56^{+1.33}_{-1.30}$	$1.56^{+0.46}_{-0.40}$	$1.54^{+0.41}_{-0.37}$

ing LIGO Hanford-only data results in $\iota^{\rm H1} = 1.53^{+1.35}_{-1.27}$ and $D_{\rm L}^{\rm H1} = 4064^{+2905}_{-2758}$ Mpc, while LIGO Livingston-only data constrains those much better at $\iota^{\rm L1} = 1.47^{+0.69}_{-0.55}$ $D_{\rm L}^{\rm L1} = 568^{+390}_{-271}$ Mpc.

Table 3. Maximum likelihood LIGO Hanford and LIGO Livingston SNRs of different modes from respective single-detector PE runs.

Mode	LIGO Hanford	LIGO Livingston
$(2,2)$	13.3	17.9
$(2,1)$	0.1	3.6
$(3,3)$	2.0	3.7
$(3,2)$	0.9	3.3
$(4,4)$	0.6	3.5
all modes	14.4	17.3

Another difference we can observe, related to the detectability of precession and breaking of the luminosity distance-inclination angle degeneracy, is the presence of higher order modes. We calculated the maximum likelihood signal-to-noise ratios (SNRs) for LIGO Hanford and LIGO Livingston from the respective single-detector PE runs. We present them in Table 3.

Unfortunately, directly comparing these values is not possible due to the difference in PSDs and locations of the detectors. While we could calculate the expected LIGO Hanford SNR from a LIGO Livingston-only PE run and vice-versa, data from a single detector does not allow one to contrain the sky location well, which could skew the SNRs we calculate.

Nevertheless, clearly the difference in the SNRs of higher order modes is significantly higher than the difference in sensitivity of the detectors. We observe the highest relative differences for the modes $(2,1), (4,4), (3,2)$, respectively. It is also clearly visible for the mode $(3,3)$, so it is noticeable for all the higher order modes that get modeled by IMRPhenomXPHM (Pratten et al. 2021).

We consider these differences significant and we believe their further investigation is warranted.

3.3. *Low frequency part of the spectrum*

We have also ran PE on data from both detectors in a different frequency range of 10-448 Hz instead of the previously used 20-448 Hz. Typically, including data from this low frequency region does not improve the fit noticeably because of much lower sensitivity of the detectors in lower frequencies. However, it could be theorized that including this part of the spectrum is helpful in GW231123's case, as it is a significantly more massive event and therefore is detected at lower frequencies.

After inspecting the results and calculating JS divergences between posteriors from two runs we concluded that changing the lower limit of frequency did not impact the posterior distributions in a way that would be distinguishable from random sampler seed change in different PE runs. We will therefore keep the minimum frequency at 20 Hz.

3.4. *GW231123's most likely parameters*

Throughout this report, as noted previously, we will analyze the results obtained from IMRPhenomXPHM PE for constistency. For the injection study in Section 5, we will use the results of catalog NRSur7dq4 analysis, as that is believed to be the most precise model for high-spin mergers such as GW231123. However, we acknowledge that new work has emerged in the field that proposes a new method to incorporate the precision of the model in the inference of the GW parameters (Hoy et al. 2025). We decide not to incorporate this method because its applicability to an event as extreme as GW231123 would need to be studied first.

4. COMPARISON OF OBSERVED DIFFERENCES TO OTHER EVENTS

Typically, the LVK Collaboration runs the PE using only the entire network of detectors. We have ran PE using only LIGO Hanford data and only LIGO Livingston data for GW230927be, an O4a event with much more typical source parameters than GW231123, as well as for GW190521, the previous most massive BBH merger.

4.1. *GW230927be*

GW230927be was a BBH merger with a total mass of $38.3^{+3.3}_{-2.2} M_\odot$, effective aligned spin parameter consistent with 0 at a luminosity distance of $1.2^{+0.4}_{-0.5}$ Gpc, detected with a network SNR of $19.7^{+0.2}_{-0.2}$ (The LIGO Scientific Collaboration et al. 2025b). There was no significant evidence of precession, with a poorly constrained effective precessing spin parameter. The catalog analysis, due to moderate mass of the event, was performed using IMRPhenomXPHM.

A comparison between posteriors inferred from the two PE runs can be seen in Figure 3. Clearly the two are in good agreement for most of the parameters. LIGO Livingston-only posteriors are generally more constrained, as expected due to the detector's superior sensitivity compared to LIGO Hanford. The only major difference can be noticed for mass ratio, where the LIGO Hanford-only posterior seems to be shifted towards unity.

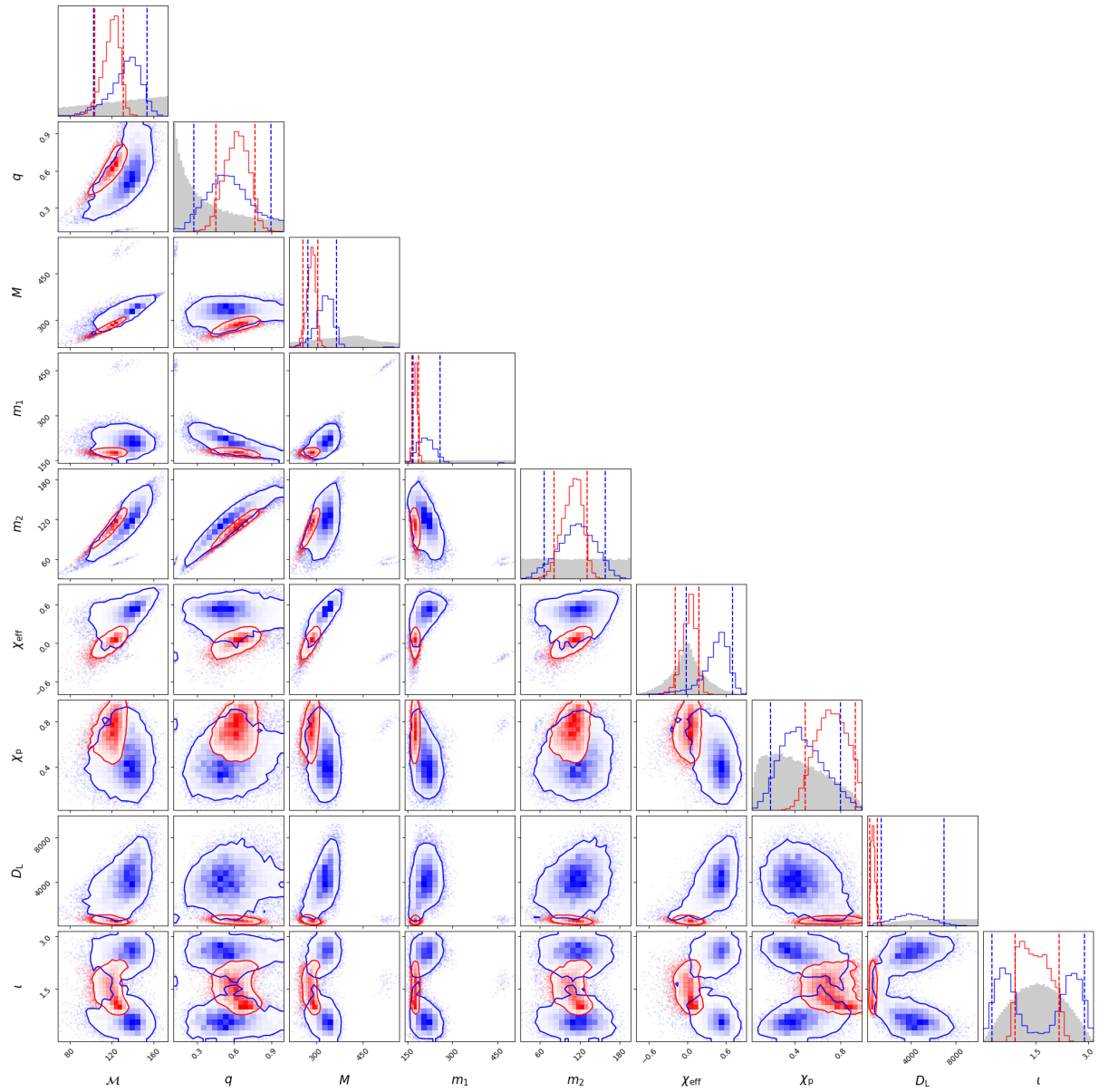

Figure 2. Comparison of LIGO Hanford-only (blue) and LIGO Livingston-only (red) parameter estimation results. Gray on 1D histograms indicates prior. Dashed lines on 1D histograms and contours on 2D histograms mark the 90% confidence levels for respective PE runs.

4.2. *GW190521*

GW190521 was the most massive observed BBH merger up until the detection of GW231123. The source-frame total mass of the system before merger was $150^{+29}_{-17} M_\odot$, the effective inspiral spin parameter was consistent with zero, and the luminosity distance was $5.3^{+2.4}_{-2.6}$ Gpc. There was moderate support for precession, with inferred effective precessing spin parameter of $0.68^{+0.25}_{-0.37}$ and $\log_{10}$ Bayes factor for orbital precession of $1.06^{+0.06}_{-0.06}$ (Abbott et al. 2020).

Figure 4 shows that there are noticeable differences between posteriors from the two PE runs. In particular, we see two distinct peaks in the mass ratio posterior and three distinct peaks in inclination angle posterior. Still, the 90% confidence intervals from the two runs have significant overlap.

4.3. *Quantitative analysis*

To quantify the differences between LIGO Hanford-only PE and LIGO Livingston-only PE in GW230927be and in GW190521 we calculated the JS divergence be-

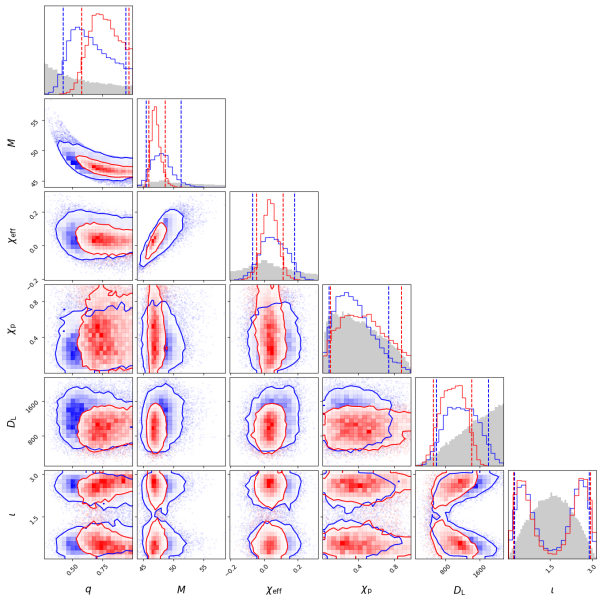

Figure 3. Comparison of LIGO Hanford-only (blue) and LIGO Livingston-only (red) parameter estimation results for GW230927be. Gray line on histograms indicates prior. Dashed lines on 1D histograms and contours on 2D histograms mark the 90% confidence levels for respective PE runs.

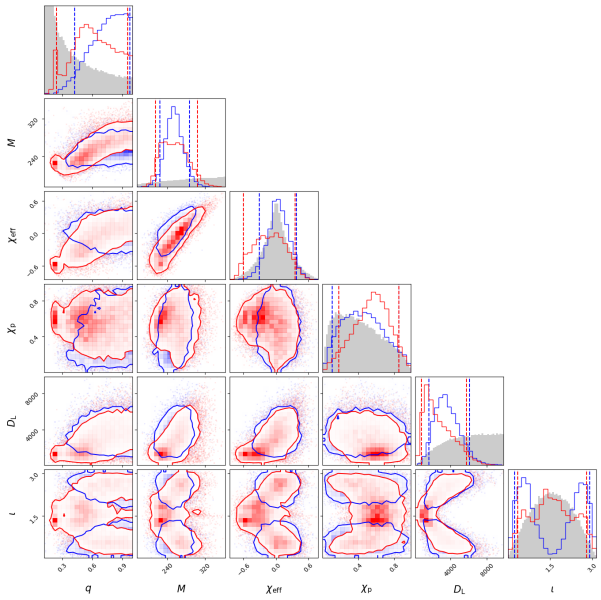

Figure 4. Comparison of LIGO Hanford-only (blue) and LIGO Livingston-only (red) parameter estimation results for GW190521. Gray line on histograms indicates prior. Dashed lines on 1D histograms and contours on 2D histograms mark the 90% confidence levels for respective PE runs.

tween the posteriors. The results are presented in Table 4.

Table 4. Jensen–Shannon divergences of selected parameters' posteriors for GW230927be and GW190521.

Parameter	GW230927be	GW190521
luminosity distance $D_{\rm L}$	0.11	0.11
viewing angle $\theta_{\rm JN}$	0.01	0.02
eff. aligned spin parameter $\chi_{\rm eff}$	0.09	0.11
phase φ	0.15	0.17
source-frame total mass	0.05	0.10
detector-frame primary mass	0.15	0.13
detector-frame total mass	0.18	0.06
source-frame primary mass	0.07	0.17
inclination angle ι	0.00	0.12
source-frame secondary mass	0.21	0.03
eff. precessing spin parameter $\chi_{\rm p}$	0.04	0.04
mass ratio q	0.15	0.08
detector-frame secondary mass	0.21	0.03
primary spin magnitude χ_1	0.03	0.08
secondary spin magnitude χ_2	0.00	0.00

Notably, all of the JS divergence values included in the table are equal to or lower than 0.21. Some of the inferred parameters' posteriors' JS divergences were higher than that, but they were either associated with SNR (such as likelihood) or with the sky position (which we do not expect to be able to reliably measure from a single detector for most events) (such as L1 time), so we decided to omit them. Although some of the JS divergence values are above the commonly adopted threshold of 0.05, they are still significantly lower than the values calculated for GW231123, with the only exception being the detector-frame secondary mass in GW230927be and χ_1 in GW190521.

Based on the JS divergence values, as well as visual inspection of the corner plots, we conclude that the differences in these two events are significantly smaller than for GW231123 and that the differences observed in GW231123 are significantly above the norm for most parameters.

5. INJECTION STUDY

5.1. *Injection setup*

Throughout the project we performed multiple sets of injections. Here we will only report on the results of the set of injections we deemed to be most relevant. Other injections and their results will be briefly discussed in Appendix Sections B and C.

We performed our injections by using the maximum likelihood waveform from the catalog (LIGO Hanford and LIGO Livingston data) NRSur7dq4 PE run, described in The LIGO Scientific Collaboration et al. (2025a). We chose those parameters as the waveform resulting from them is the one that matches the observed waveform most closely and we wanted to ensure our results are applicable to GW231123. We present selected injected parameters' values in Table 5.

Table 5. NRSur7dq4 maximum likelihood injected parameters.

Parameter	Value
luminosity distance $D_{\rm L}$ (Mpc)	962
primary spin magnitude χ_1	0.988
secondary spin magnitude χ_2	0.883
eff. aligned spin parameter $\chi_{\rm eff}$	-0.047
eff. precessing spin parameter $\chi_{\rm p}$	0.986
inclination angle ι (rad)	1.03
source-frame primary mass ($M_\odot$)	131
source-frame secondary mass ($M_\odot$)	119
source-frame total mass ($M_\odot$)	250
detector-frame primary mass ($M_\odot$)	156
detector-frame secondary mass ($M_\odot$)	142
detector-frame total mass ($M_\odot$)	298
mass ratio q	0.91
geocentric time since 1384782888 (s)	0.619
phase φ (rad)	2.35
viewing angle $\theta_{\rm JN}$ (rad)	1.34
LIGO Hanford optimal SNR	13.6
LIGO Livingston optimal SNR	16.0

5.2. *Zero-noise injections*

We performed three zero-noise injections, for LIGO Hanford-only data, LIGO Livingston-only data and using both detectors. We present their results in Figure 5.

As we can see, LIGO Livingston-only posteriors are significantly narrower than LIGO Hanford-only posteriors, which matches our expectations. Generally the posteriors agree, we do not see systematic shift between the two. The only exception to that is luminosity distance (and source masses that are related to it), but that can be linked to poorer constraint on inclination angle leading to samples with higher inferred luminosity distance.

We also see that the posteriors are not in good agreement with the injected values. In particular, the injected values of $m_1, \chi_{\rm p}, \chi_1, q$ are not in the 90% confidence intervals recovered with IMRPhenomXPHM from either PE run. We also see that the injected values are far from the maximum likelihood inferred parameters for $m_1, m_2, \chi_{\rm p}, \chi_1, q, D_{\rm L}$ and geocentric time. That once again is unsurprising, as significant differences between IMRPhenomXPHM and NRSur7dq4 PE results have already been noticed for this high-spin part of the parameter space (The LIGO Scientific Collaboration et al. 2025a).

5.3. *Gaussian noise injections*

We performed 40 injections of NRSur7dq4 maximum likelihood inferred waveform into Gaussian noise – 20 in LIGO Hanford, 20 in LIGO Livingston. We used each detector's respective PSD from the time of detection computed using BayesWave (Cornish & Littenberg 2015; Cornish et al. 2021; Chatziioannou et al. 2021).

Since the Gaussian noise realizations in both detectors are independent, we calculated JS divergence between all combinations of PE results. That resulted in 400 JS values. We present the extreme posteriors for effective aligned spin parameter, effective precessing spin parameter, detector-frame total mass and source-frame total mass in Figure 6 as well as all posteriors for luminosity distance and inclination angle in Figure 7.

Noticeably the extreme posterior distributions differ significantly from each other. In LIGO Hanford-only analysis, the maximum likelihood $\chi_{\rm eff}$ shifts from ~ -0.2 to ~ 0.7; for LIGO Livingston-only analysis it shifts from ~ -0.4 to ~ 0.3. The posteriors for $\chi_{\rm p}$ are generally wider, and the maximum likelihood values range from ~ 0.2 to ~ 0.7 in LIGO Hanford and from ~ 0.5 to ~ 0.8 in LIGO Livingston. Analogously, the maximum likelihood detector-frame total mass values range from $\sim 240M_\odot$ to $\sim 360M_\odot$ for LIGO Hanford and from $\sim 260M_\odot$ to $\sim 330M_\odot$ for LIGO Livingston, while the extreme values for source-frame total mass are shifted by $\sim 25M_\odot$ to $\sim 100M_\odot$.

In many cases, the overlap between extreme posteriors is minimal, as can be seen for LIGO Livingston-only detector-frame total mass posterior distributions. Further, for luminosity distance and inclination angle, we see both posteriors indicating face-on, as well as edge-on, especially in LIGO Hanford-only PE runs.

To quantify the frequency of differences of magnitude equal or higher than in GW231123, we calculated the fraction of runs' combinations with JS divergence equal or higher than the value calculated for GW231123. We present these results in Table 6.

We interpret these results as supporting the hypothesis that Gaussian noise might have caused the discrepancies between posteriors inferred from LIGO Hanford-only and LIGO Livingston-only PE runs. We also be-

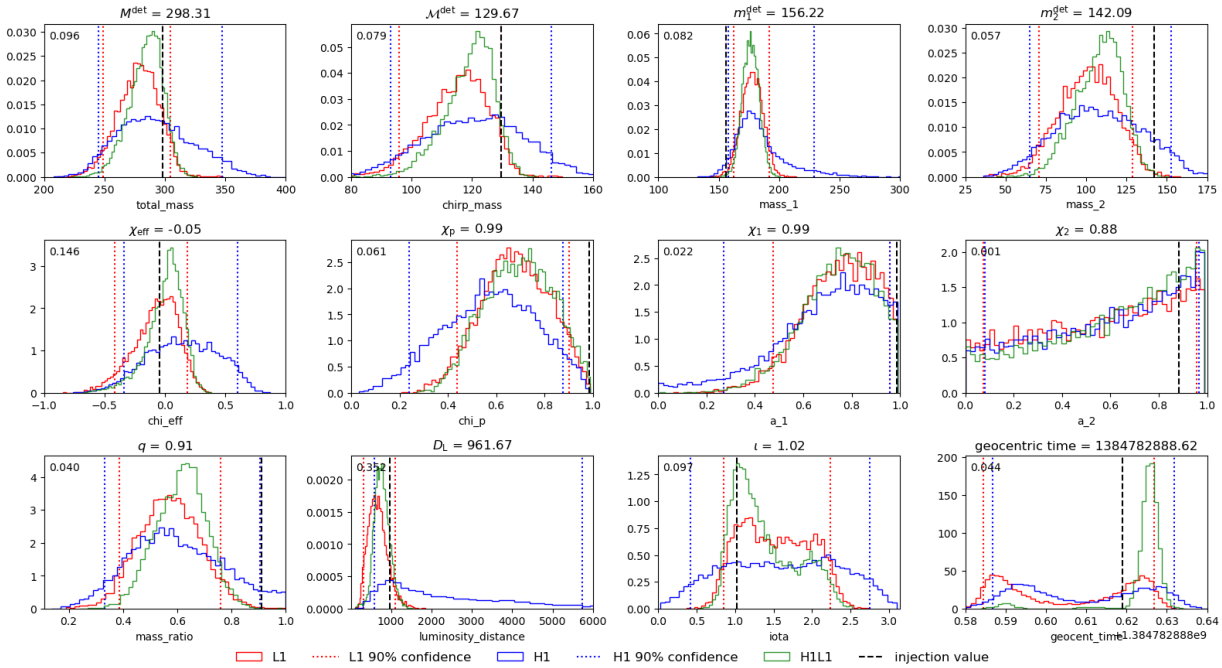

Figure 5. Inferred posteriors from PE performed on three zero noise NRSur7dq4 maximum likelihood injections. Blue indicates LIGO Hanford-only run, red LIGO Livingston-only run and green the network run. Dashed lines in blue and red correspond to 90% confidence intervals. Black dashed line indicates the values of injected parameters.

lieve that JS divergence is not a good way to measure similarity of phase or inclination angle posteriors, since these are angles and both often show some level of symmetry (inclination angle) or periodicity (phase for very similar black holes). We therefore do not find the low fraction for phase alarming and conclude that these fractions make this hypothesis plausible.

Table 6. Fraction of injection seed combinations with JS divergence higher or equal to that of the event.

Parameter	$\mathrm{JS} \geq \mathrm{JS}_{\mathrm{GW231123}}$
primary spin magnitude χ_1	0.81
secondary spin magnitude χ_2	0.73
detector-frame secondary mass	0.68
mass ratio q	0.63
detector-frame total mass	0.30
eff. precessing spin parameter χ_{P}	0.22
eff. aligned spin parameter χ_{eff}	0.20
luminosity distance D_{L}	0.18
detector-frame primary mass	0.18
inclination angle ι	0.13
phase φ	0.06

We also plotted the maximum likelihood waveforms from the injections performed. We present that in Figure 8. We see clear deviations from the injected waveform in both the inspiral, merger and ringdown part of the signal. Apart from a single seed in LIGO Livingston, the waveforms stay synced in phase. The one outlying LIGO Livingston seed displays a significant phase shift in the ringdown phase, at $t \sim 1384782883.3$.

We also performed 10 LIGO Hanford injections with LIGO Livingston's PSD, which will be discussed in Appendix C. They do not show that this change in LIGO Hanford's PSD significantly decreases the differences we see between detectors. We therefore conclude that the difference in PSD between detectors is not what is responsible for the observed high JS divergence values in GW231123.

6. CONCLUSIONS AND FUTURE WORK

GW231123's source is the most massive BBH merger detected using gravitational waves so far. The two component black holes are also likely very highly spinning. Significant differences in inferred parameters' posteriors between LIGO Hanford-only PE and LIGO Livingston-only PE can be noticed. These differences are significantly higher than what we observe for other events, as evidenced by the comparison we performed for GW230927be (The LIGO Scientific Collaboration et al. 2025b) and GW190521 (Abbott et al. 2020).

We performed zero-noise injections that further highlight the waveform systematics observed in The LIGO Scientific Collaboration et al. (2025a), as the performed

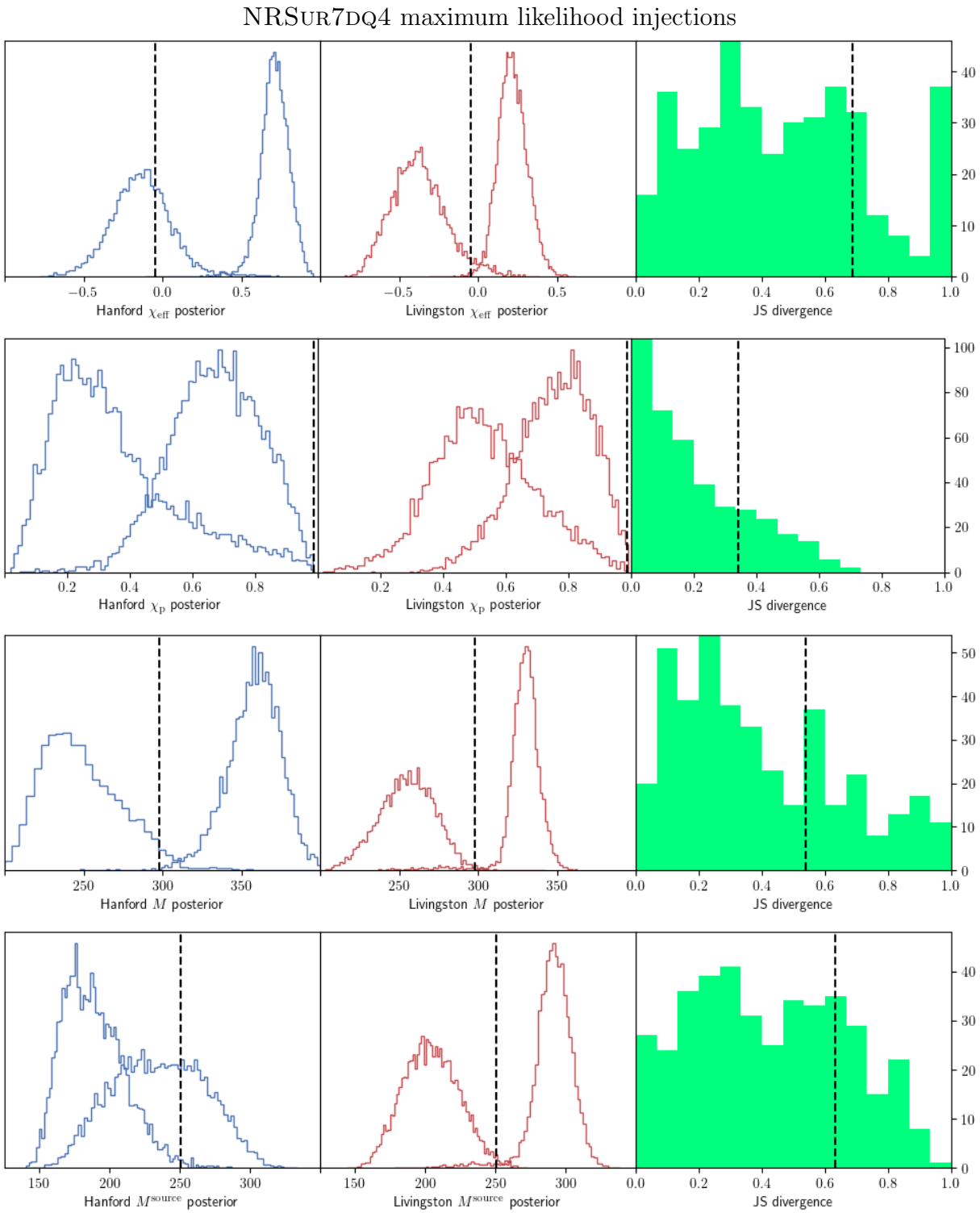

Figure 6. Left and central panels: extreme inferred posteriors for effective aligned spin parameter, effective precessing spin parameter, detector-frame total mass and source-frame total mass from the Gaussian noise maximum likelihood injections for LIGO Hanford-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds. Gray dashed lines on the left two histograms mark the injected NRSur7dq4 parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

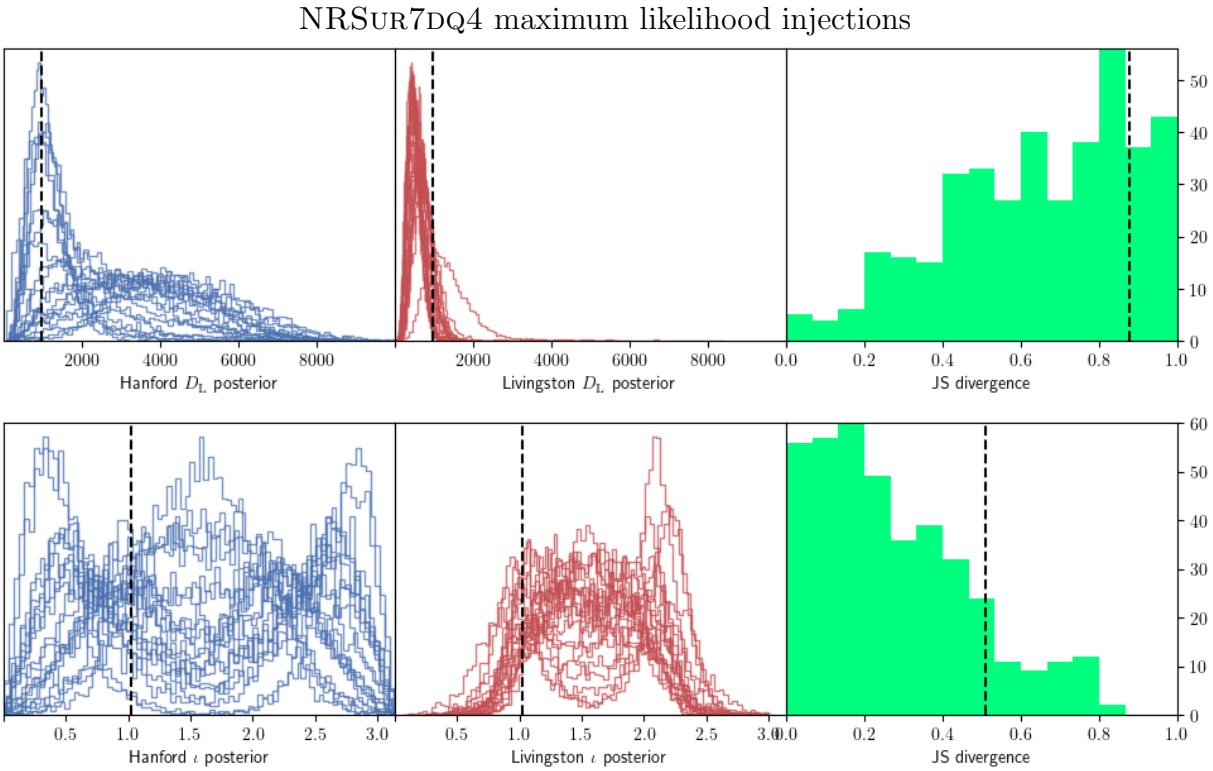

Figure 7. Left and central panels: inferred posteriors for luminosity distance and inclination angle from the Gaussian noise maximum likelihood injections for LIGO Hanford-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds. Gray dashed lines on the left two histograms mark the injected NRSur7dq4 parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

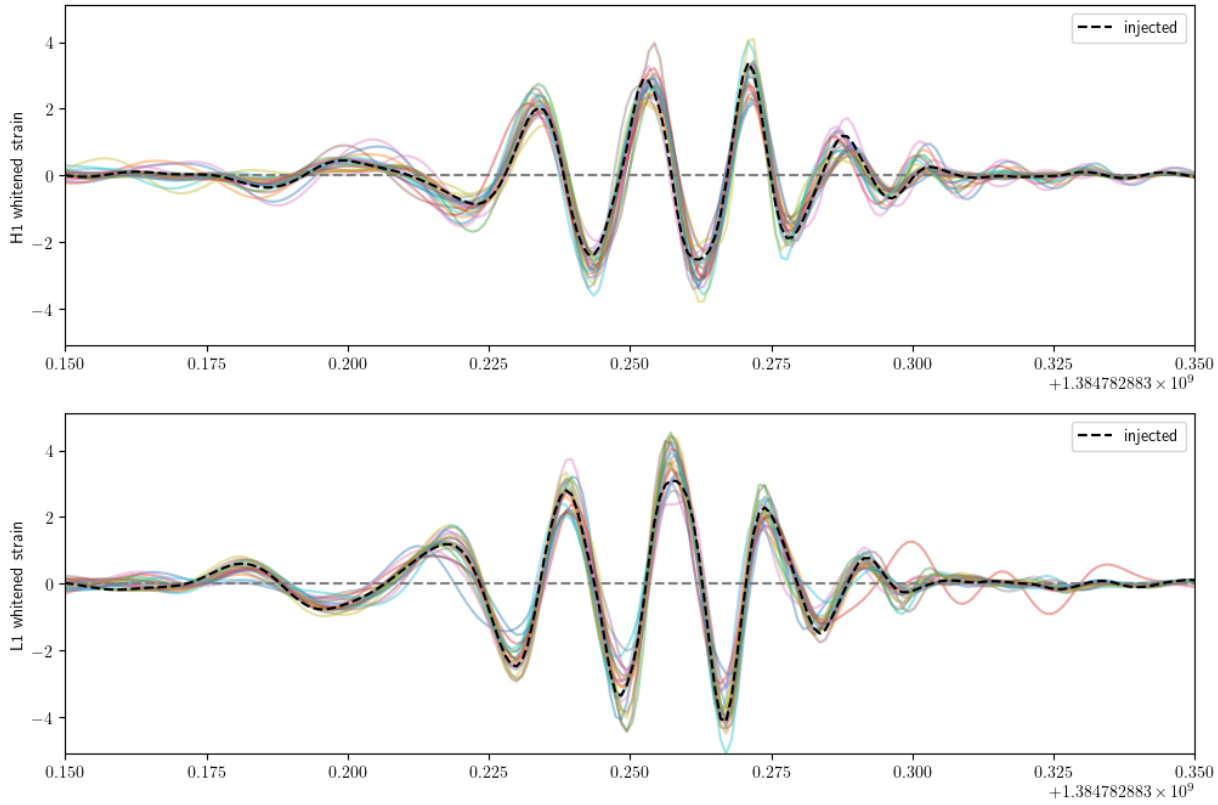

Figure 8. Maximum likelihood whitened waveforms from the 20 Gaussian noise injections in LIGO Hanford and 20 Gaussian noise injections in LIGO Livingston. The injected waveform is plotted with a dashed black line. The y-axis is in noise standard deviation units.

IMRPhenomXPHM injections fail to recover the injected NRSur7dq4 values. We also performed Gaussian noise injections, which show that signals similar to GW231123 are very sensitive to Gaussian noise fluctuations. We find that these can lead to the the same degree of discrepancies between single-detector PE runs as the ones we observe in GW231123. This behavior occurs in $\sim 10\%$ – $\sim 70\%$ of the Gaussian noise realizations for the parameters for which the observed differences were the highest.

In particular, the inferred one-detector-only posteriors for effective aligned spin parameter, effective precessing spin parameter, total mass, luminosity distance and inclination angle can differ very significantly with different Gaussian noise realizations. Spin orientations can vary from low to extremely high effective spin parameters, total mass can differ by tens of solar masses and the inferred inclination angle can change from values implying face-on to values implying edge-on orientation, which is connected with a significant change in luminosity distance. The differences are generally higher in LIGO Hanford, which matches our expectations, as, due to the difference in sensitivity, noise fluctuations are more pronounced compared to the signal there.

Based on our injections we conclude that Gaussian noise fluctuations could explain the differences between detectors we see in GW231123. Based on the fraction of injection seeds' combinations that exhibit differences equal to or higher than GW231123 as measured by Jensen–Shannon divergence, we believe this explanation is plausible. While our study has once again highlighted the issue of waveform systematics, we do not find that they are the driver of the discrepancies we observe between detectors.

In the future, we will repeat a similar analysis with NRSur7dq4 to ensure that this finding is not unique to IMRPhenomXPHM and is an inherent attribute of this part of the parameter space instead.

7. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (NSF), the LIGO Laboratory Summer Undergraduate Research Fellowship program (LIGO SURF), and the California Institute of Technology Student-Faculty Programs.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, Physical Review X, 9, 031040, doi: 10.1103/PhysRevX.9.031040

Abbott, R., et al. 2020a, PhRvL, 125, 101102, doi: 10.1103/PhysRevLett.125.101102

—. 2020b, ApJL, 900, L13, doi: 10.3847/2041-8213/aba493

Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, PhRvL, 125, 101102, doi: 10.1103/PhysRevLett.125.101102

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27, doi: 10.3847/1538-4365/ab06fc

Blanchet, L. 2014, Living Reviews in Relativity, 17, 2, doi: 10.12942/lrr-2014-2

Boyle, M., Hemberger, D., Iozzo, D. A. B., et al. 2019, Classical and Quantum Gravity, 36, 195006, doi: 10.1088/1361-6382/ab34e2

Chatziioannou, K., Cornish, N., Wijngaarden, M., & Littenberg, T. B. 2021, Phys. Rev. D, 103, 044013, doi: 10.1103/PhysRevD.103.044013

Christensen, N., & Meyer, R. 2022, Reviews of Modern Physics, 94, 025001, doi: 10.1103/RevModPhys.94.025001

Cornish, N. J., & Littenberg, T. B. 2015, Class. Quant. Grav., 32, 135012, doi: 10.1088/0264-9381/32/13/135012

Cornish, N. J., Littenberg, T. B., Bécsy, B., et al. 2021, Phys. Rev. D, 103, 044006, doi: 10.1103/PhysRevD.103.044006

Costa, G., Bressan, A., Mapelli, M., et al. 2021, MNRAS, 501, 4514, doi: 10.1093/mnras/staa3916

Endres, D., & Schindelin, J. 2003, IEEE Transactions on Information Theory, 49, 1858, doi: 10.1109/TIT.2003.813506

Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, ApJ, 887, 53, doi: 10.3847/1538-4357/ab518b

Gerosa, D., & Fishbach, M. 2021, Nature Astronomy, 5, 749, doi: 10.1038/s41550-021-01398-w

Hoy, C., Akçay, S., Mac Uilliam, J., & Thompson, J. E. 2025, Nature Astronomy, doi: 10.1038/s41550-025-02579-7

Kullback, S., & Leibler, R. A. 1951, The Annals of Mathematical Statistics, 22, 79 , doi: 10.1214/aoms/1177729694

Lin, J. 1991, IEEE Transactions on Information Theory, 37, 145, doi: 10.1109/18.61115

Mapelli, M., Spera, M., Montanari, E., et al. 2020, ApJ, 888, 76, doi: 10.3847/1538-4357/ab584d

Menéndez, M., Pardo, J., Pardo, L., & Pardo, M. 1997, Journal of the Franklin Institute, 334, 307, doi: https://doi.org/10.1016/S0016-0032(96)00063-4

Miller, S. J., Isi, M., Chatziioannou, K., Varma, V., & Hourihane, S. 2025, arXiv e-prints, arXiv:2505.14573, doi: 10.48550/arXiv.2505.14573

Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. 2018, PhRvD, 98, 083007, doi: 10.1103/PhysRevD.98.083007

Payne, E., Hourihane, S., Golomb, J., et al. 2022, PhRvD, 106, 104017, doi: 10.1103/PhysRevD.106.104017

Pratten, G., García-Quirós, C., Colleoni, M., et al. 2021, PhRvD, 103, 104056, doi: 10.1103/PhysRevD.103.104056

Racine, É. 2008, PhRvD, 78, 044021, doi: 10.1103/PhysRevD.78.044021

Scheel, M. A., Boyle, M., Mitman, K., et al. 2025, arXiv e-prints, arXiv:2505.13378, doi: 10.48550/arXiv.2505.13378

Spera, M., Mapelli, M., Giacobbo, N., et al. 2019, MNRAS, 485, 889, doi: 10.1093/mnras/stz359

The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. 2025a, arXiv e-prints, arXiv:2507.08219, doi: 10.48550/arXiv.2507.08219

The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2025b, arXiv e-prints, arXiv:2508.18082, doi: 10.48550/arXiv.2508.18082

Thomas, L. M., Schmidt, P., & Pratten, G. 2021, Phys. Rev. D, 103, 083022, doi: 10.1103/PhysRevD.103.083022

Usman, S. A., Mills, J. C., & Fairhurst, S. 2019, ApJ, 877, 82, doi: 10.3847/1538-4357/ab0b3e

Varma, V., Field, S. E., Scheel, M. A., et al. 2019, Phys. Rev. Research., 1, 033015, doi: 10.1103/PhysRevResearch.1.033015

Xu, Y., & Hamilton, E. 2023, PhRvD, 107, 103049, doi: 10.1103/PhysRevD.107.103049

APPENDIX

A. WAVEFORM SYSTEMATICS

Significant differences in LIGO Hanford and LIGO Livingston's sensitivities can immediately be noticed, especially in the 20-50 Hz frequency range. In particular, at 20 Hz LIGO Livingston's ASD value is almost two times lower than LIGO Hanford's ASD.

When choosing the waveform approximant to use for our study we inspected results from PE using both NRSur7dq4 and IMRPhenomXPHM and discovered a significant difference in the average modeled waveform's ASD between the two models. We plot this in Figure A1 and Figure A2 for NRSur7dq4 and IMRPhenomXPHM, respectively. That finding is in agreement with the discrepancies noticed between different waveform approximants in The LIGO Scientific Collaboration et al. (2025a). We decided to focus our project on IMRPhenomXPHM. We acknowledge the fact that using the quicker model might influence the results we achieve and note that further investigation using other waveform approximants will be performed by us in the future.

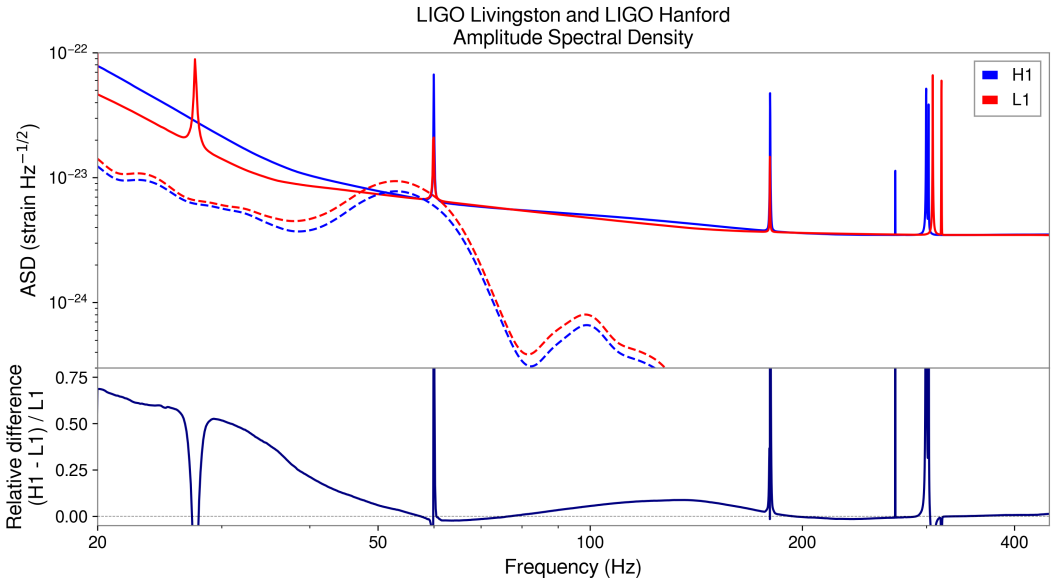

Figure A1. Top plot: ASD of LIGO Livingston (red), LIGO Hanford (blue) and of the average waveform resulting from NRSur7dq4 PE posterior samples for both of the interferometers (average waveform resulting from the network-inferred posteriors). Bottom plot: relative difference in ASD of the two interferometers defined as $(\mathrm{ASD}^{\mathrm{H1}} - \mathrm{ASD}^{\mathrm{L1}})/\mathrm{ASD}^{\mathrm{L1}}$.

B. NUMERICAL RELATIVITY INJECTIONS

B.1. *Numerical relativity*

A numerical relativity (NR) waveform has only 9 intrinsic parameters associated with it: ratio of Christodoulou masses q, dimensionless spins $\vec{\chi_1}, \vec{\chi_2}$ (which also dictate the values of $\chi_{\mathrm{eff}}, \chi_{\mathrm{p}}$), eccentricity e (which will be very close to 0 for all waveforms we might examine), and mean anomaly at reference time ℓ (which is irrelevant for quasi-circular orbits) (Scheel et al. 2025; Boyle et al. 2019). The system's masses can be scaled, and all extrinsic parameters can be varied.

NR waveforms are the ground truth in General Relativity and are used to calibrate the waveform approximants we use in PE. Unlike with waveform approximants, we can be certain of their parameters. This, however, comes at a very large computational cost, hence they cannot be used directly for PE. Additionaly, there are only 3756 NR waveforms in the current Simulating eXtreme Spacetimes (SXS) catalog (Scheel et al. 2025) and a limited number of waveforms in other catalogs. As such, we cannot choose a waveform with arbitrary parameters, but have to limit ourselves to already computed waveforms.

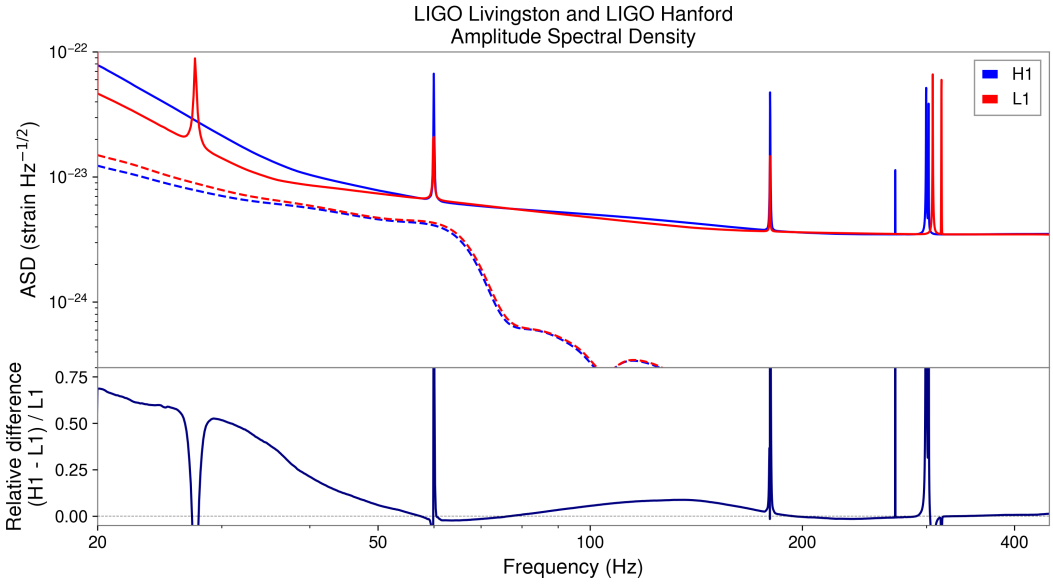

Figure A2. Top plot: ASD of LIGO Livingston (red), LIGO Hanford (blue) and of the average waveform resulting from IMRPhenomXPHM PE posterior samples for both of the interferometers (average waveform resulting from the network-inferred posteriors). Bottom plot: relative difference in ASD of the two interferometers defined as $(\mathrm{ASD}^{\mathrm{H1}} - \mathrm{ASD}^{\mathrm{L1}})/\mathrm{ASD}^{\mathrm{L1}}$.

B.2. *NR injection setup*

We will be examining two NR waveforms from the current SXS catalog. We will describe them shortly below. We injected them into noise generated with the same PSDs that we used for the maximum likelihood injections. We performed 20 injection in LIGO Hanford and 20 injections in LIGO Livingston for each configuration. The precise values of the injected NR waveforms can be found in Table B1. We will describe them shortly below.

The first waveform we examined in two injections was SXS:BBH:4043. It is a simulated BBH merger with two equal-mass, highly spinning black holes. Both components have $\chi = 0.949$ and are oriented in such a way, that $\chi_{\mathrm{eff}} = 0.159, \chi_{\mathrm{p}} = 0.940$. We scaled the mass to detector-frame total mass of $\sim 300 M_{\odot}$. The first injection we performed had $\iota = 0.03, \varphi = 2.00, \psi = 2.4$ and the second one had $\iota = 1.43, \varphi = 2.00$, which represents two very different orientations, face-on and edge-on, and different phases. We expected to see difference in precession measurability and thus presence of higher order modes, which could impact lifting of luminosity distance–inclination angle degeneracy (Miller et al. 2025; Xu & Hamilton 2023).

We performed the third NR injection using SXS:BBH:4030. It was very similar to the previously described injection, with the main difference being spin orientation – they were oriented such that $\chi_{\mathrm{eff}} = 0.810$ and $\chi_{\mathrm{p}} = 0.589$. The total mass was scaled to the same value. We injected it at an orientation and location such that $\iota = 1.58, \varphi = 2.35$, the same setup that was used in The LIGO Scientific Collaboration et al. (2025a).

B.3. *NR Injections in Gaussian noise*

We present the fractions of injection seed combinations from NR injections with JS divergence for parameter posteriors exceeding those observed in GW231123 in Table B2.

We notice very significant differences in these values between different injections. This shows that the impact of Gaussian noise on PE is highly dependent on the signal from the BBH merger, depending both on the source intrinsic parameters, as well as the location and orientation of the source. This difference is the most striking for luminosity distance D_{L} and inclination angle ι, where the number of seed combinations exceeding event's JS divergence for those parameters varies between 0 for both SXS:BBH:4043 injections and $\sim 30\%$ – $\sim 40\%$ of the combinations for SXS:BBH:4043. While, as we argued in Section 5.3, the JS divergence value for the inclinations angle ι might not be most informative, the JS divergence for luminosity distance D_{L} (closely related to ι) is an important result.

Table B1. NR Gaussian noise injections' injected parameters. Column names refer to SXS:BBH:4043 and SXS:BBH:4030 NR waveforms and the inclination angle ι at which they were evaluated.

Parameter	4043, $\iota = 0.03$	4043, $\iota = 1.43$	4030, $\iota = 1.58$
luminosity distance $D_{\rm L}$ (Mpc)	4043	2358	1220
primary spin magnitude χ_1	0.949	0.949	0.949
secondary spin magnitude χ_2	0.949	0.949	0.949
eff. aligned spin parameter $\chi_{\rm eff}$	0.159	0.159	0.810
eff. precessing spin parameter $\chi_{\rm p}$	0.940	0.940	0.589
inclination angle ι (rad)	0.03	1.43	1.58
source-frame primary mass ($M_\odot$)	150	150	150
source-frame secondary mass ($M_\odot$)	150	150	150
source-frame total mass ($M_\odot$)	300	300	300
detector-frame primary mass ($M_\odot$)	90.6	106	121
detector-frame secondary mass ($M_\odot$)	90.6	106	121
detector-frame total mass ($M_\odot$)	181	212	243
mass ratio q	1.00	1.00	1.00
geocentric time since 1384782888 (s)	0.637	0.637	0.619
phase φ (rad)	2.00	2.00	2.35
viewing angle $\theta_{\rm JN}$ (rad)	0.50	1.90	1.51
LIGO Hanford optimal SNR	12.8	13.1	13.8
LIGO Livingston optimal SNR	17.9	18.0	17.0

We present the inferred posteriors from these three NR injections for effective precessing spin parameter $\chi_{\rm eff}$, source-frame total mass, luminosity distance $D_{\rm L}$ and inclination angle ι in Figure B1 for SXS:BBH:4043 with $\iota = 0.03$, in Figure B2 for SXS:BBH:4043 with $\iota = 1.43$ and in Figure B3 for SXS:BBH:4030 with $\iota = 1.58$.

We notice that for both SXS:BBH:4043 injections the resulting posteriors are much more similar to each other than for SXS:BBH:4030 injections, which is corroborated by the calculated JS divergence values. However, the posteriors do not recover the injected values correctly, as we can see a systematic shift of the posteriors of $\chi_{\rm eff}$ in relation to the injected value in SXS:BBH:4043 with $\iota = 0.03$. The differences in luminosity distance and inclination angle are also less pronounced, but, clearly, there are a few seeds where the inclination angle posterior switched between face-on and edge-on orientation.

C. PSD

We investigated whether the differences in PSDs of the two LIGO detectors could explain the differences we see in GW231123 PE runs for individual detectors. To do that, in addition to the 20 maximum likelihood injections in LIGO Hanford and 20 injections in LIGO Livingston, performed 10 maximum likelihood injections in LIGO Hanford with LIGO Livingston's PSD.

We do not see a significant, systematic shift between the values in Table C1, neither do we see clearly different shape of the histograms in Figure C1 or Figure C2. This suggests that the difference in PSD between the two detectors is not responsible for the differences we see in one-detector-only runs for GW231123. Otherwise, we would expect the fractions of seed combinations with JS divergence exceeding that of the event to be noticeably lower than for the initial runs.

We present the resulting extreme posteriors for selected parameters from this run in Figure C1 and all posteriors for the luminosity distance and inclination angle in Figure C2. In Table C1 we present a comparison of the fraction of seed combinations with JS divergence higher than that for the event for previously described runs in LIGO Hanford-with-LIGO-Hanford-PSD and for the newly performed LIGO Hanford-with-LIGO-Livingston-PSD runs.

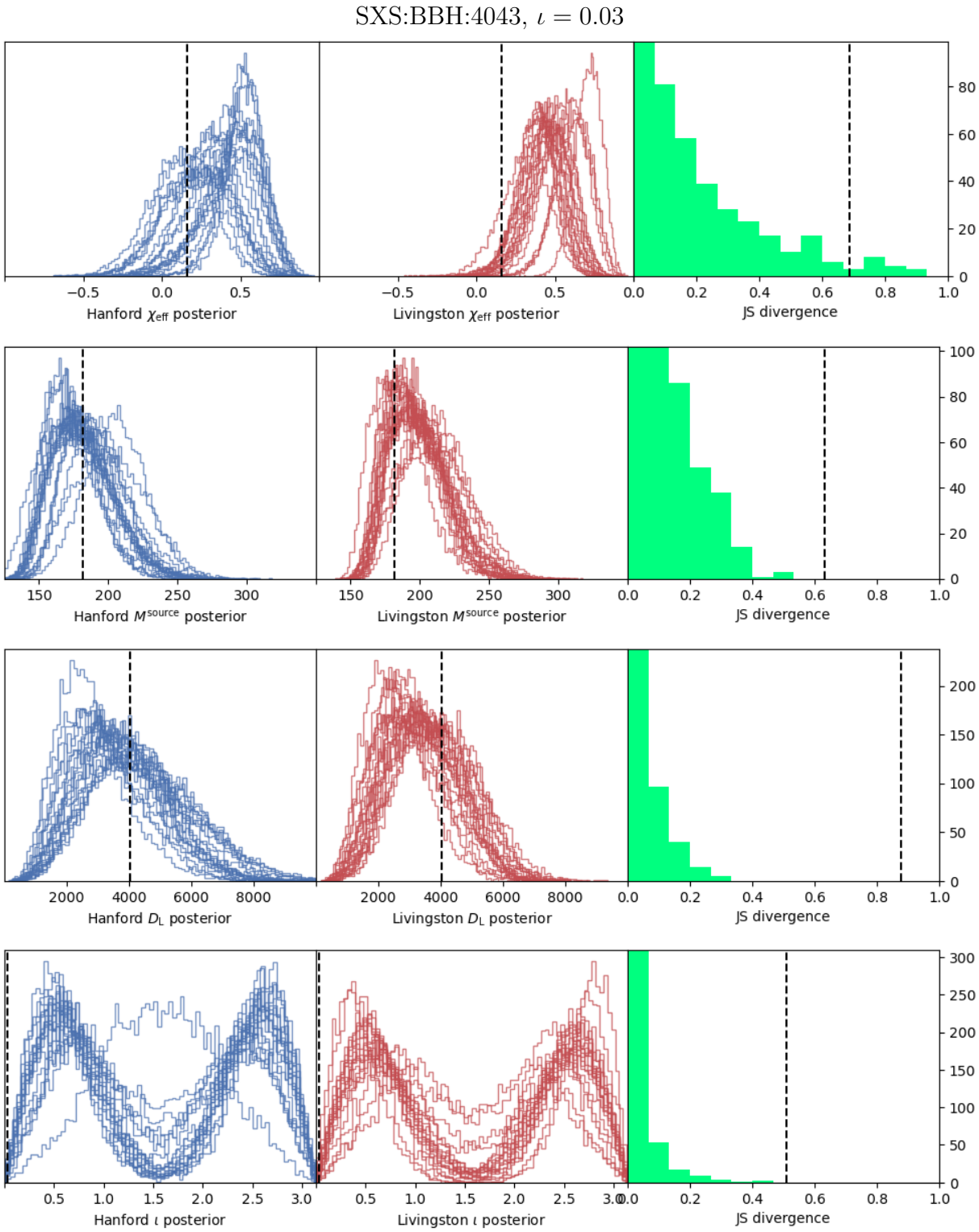

Figure B1. Left and central panels: inferred posteriors for effective precessing spin parameter $\chi_{\rm eff}$, source-frame total mass, luminosity distance $D_{\rm L}$ and for inclination angle ι from the SXS:BBH:4043 injection in Gaussian noise with $\iota = 0.03$ for LIGO Hanford-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds. Gray dashed lines on the left two histograms mark the injected NR parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

Table B2. Fraction of injection seed combinations with JS divergence higher or equal to that of GW231123 for three performed NR injections. Column names refer to maximum likelihood injection from Section 5, SXS:BBH:4043 and SXS:BBH:4030 NR waveforms and the inclination angle ι at which they were evaluated.

Parameter	max. l.	4043, $\iota = 0.03$	4043, $\iota = 1.43$	4030, $\iota = 1.58$
primary spin magnitude χ_1	0.81	0.75	0.26	0.80
secondary spin magnitude χ_2	0.73	0.62	0.27	0.96
detector-frame secondary mass	0.68	0.22	0.27	0.88
mass ratio q	0.63	0.12	0.16	0.80
detector-frame total mass	0.30	0.09	0.07	0.22
eff. precessing spin parameter $\chi_{\rm p}$	0.22	0.01	0.00	0.53
eff. aligned spin parameter $\chi_{\rm eff}$	0.20	0.04	0.00	0.36
luminosity distance $D_{\rm L}$	0.18	0.00	0.00	0.29
detector-frame primary mass	0.18	0.01	0.01	0.52
inclination angle ι	0.13	0.00	0.00	0.43
phase φ	0.06	0.00	0.00	0.10

Table C1. Fraction of injection seed combinations with JS divergence higher or equal to that of GW231123.

Parameter	H1 vs. L1	H1 with L1 PSD vs. L1
primary spin magnitude χ_1	0.81	0.78
secondary spin magnitude χ_2	0.73	0.72
detector-frame secondary mass	0.68	0.64
mass ratio q	0.63	0.62
detector-frame total mass	0.30	0.35
eff. precessing spin parameter $\chi_{\rm p}$	0.22	0.30
eff. aligned spin parameter $\chi_{\rm eff}$	0.20	0.21
luminosity distance $D_{\rm L}$	0.18	0.14
detector-frame primary mass	0.18	0.23
inclination angle ι	0.13	0.12
phase φ	0.06	0.10

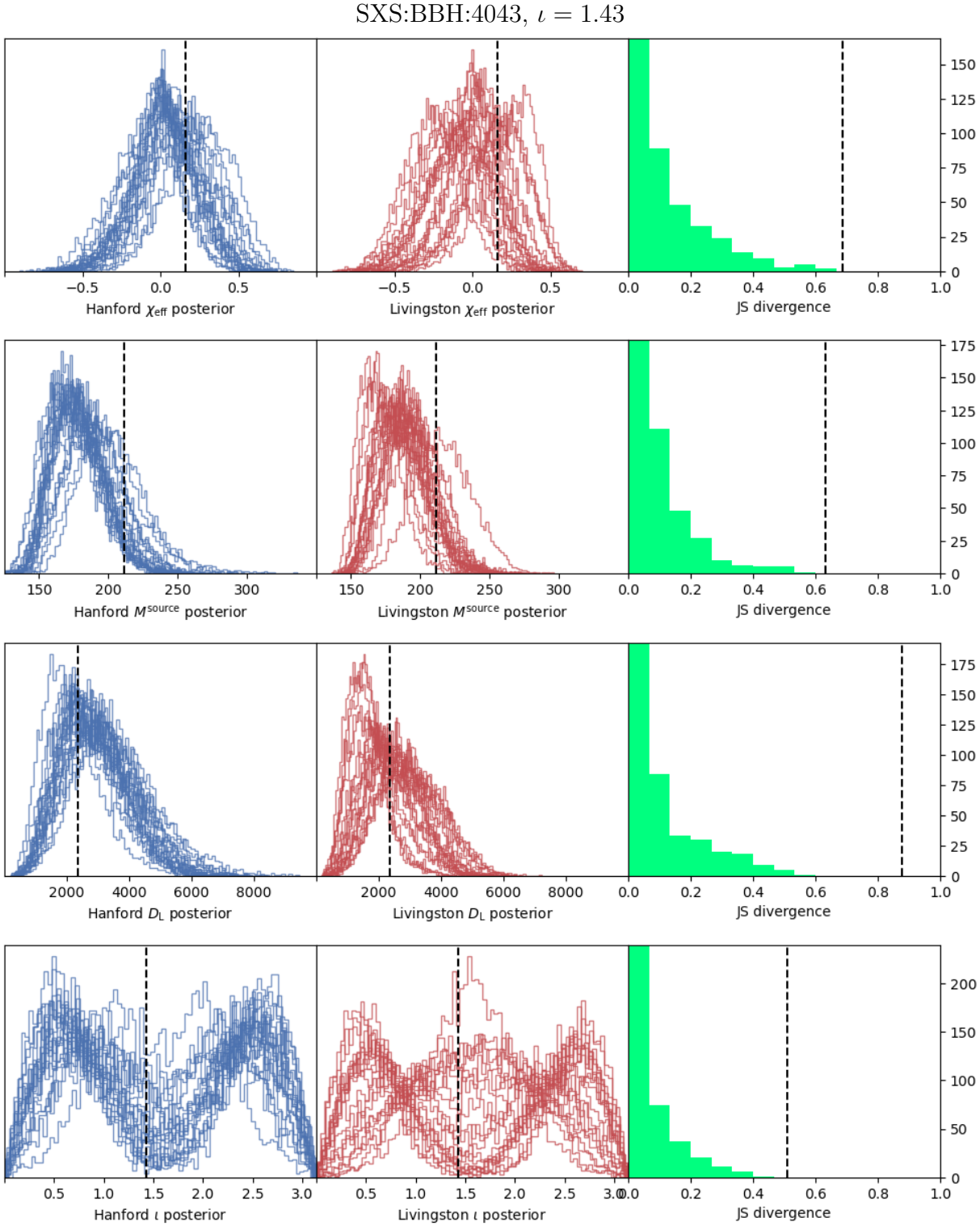

Figure B2. Left and central panels: inferred posteriors for effective precessing spin parameter $\chi_{\rm eff}$, source-frame total mass, luminosity distance $D_{\rm L}$ and for inclination angle ι from the SXS:BBH:4043 injection in Gaussian noise with $\iota = 1.43$ for LIGO Hanford-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds. Gray dashed lines on the left two histograms mark the injected NR parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

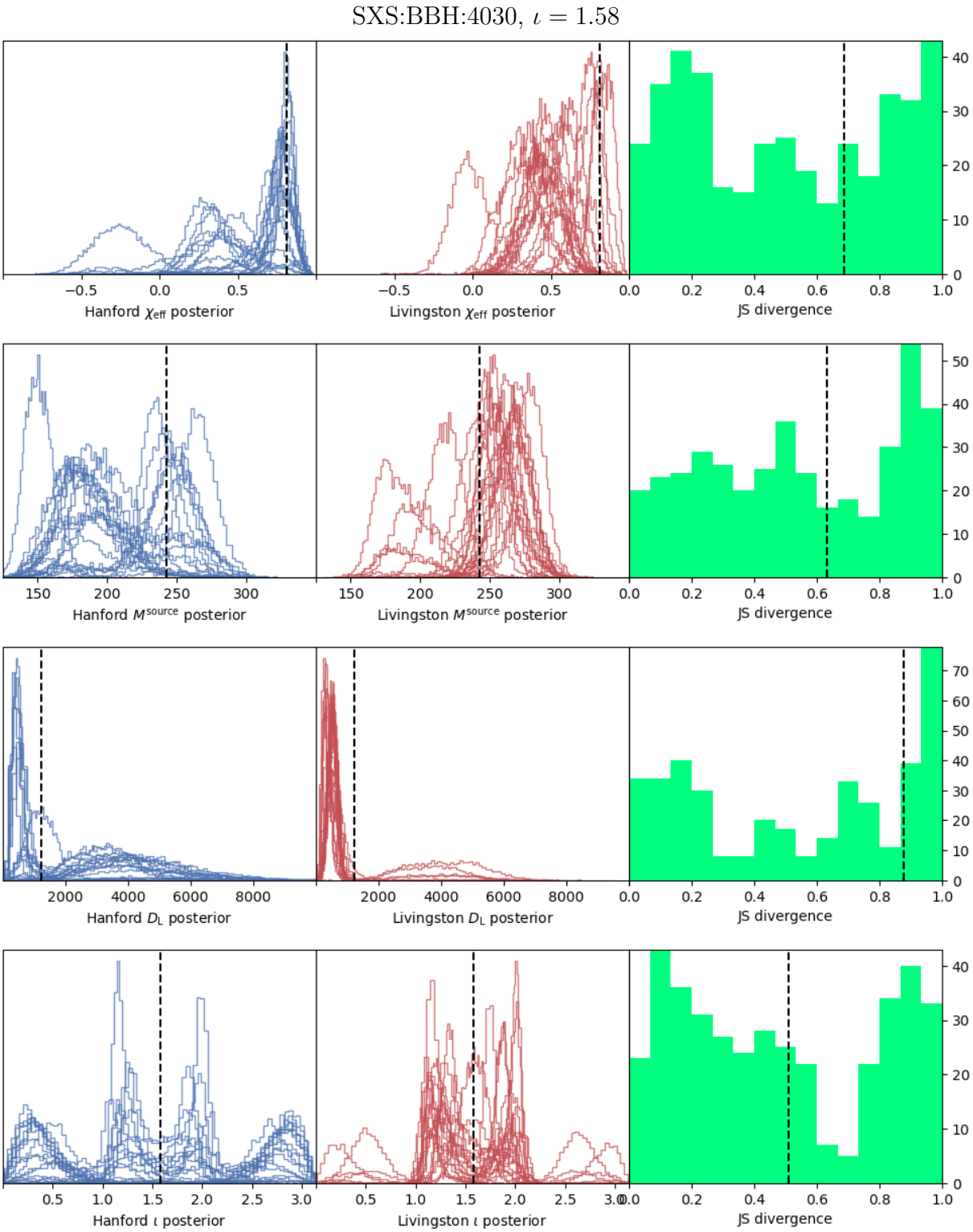

Figure B3. Left and central panels: inferred posteriors for effective precessing spin parameter $\chi_{\rm eff}$, source-frame total mass, luminosity distance $D_{\rm L}$ and for inclination angle ι from the SXS:BBH:4030 injection in Gaussian noise with $\iota = 1.58$ for LIGO Hanford-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds. Gray dashed lines on the left two histograms mark the injected NR parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

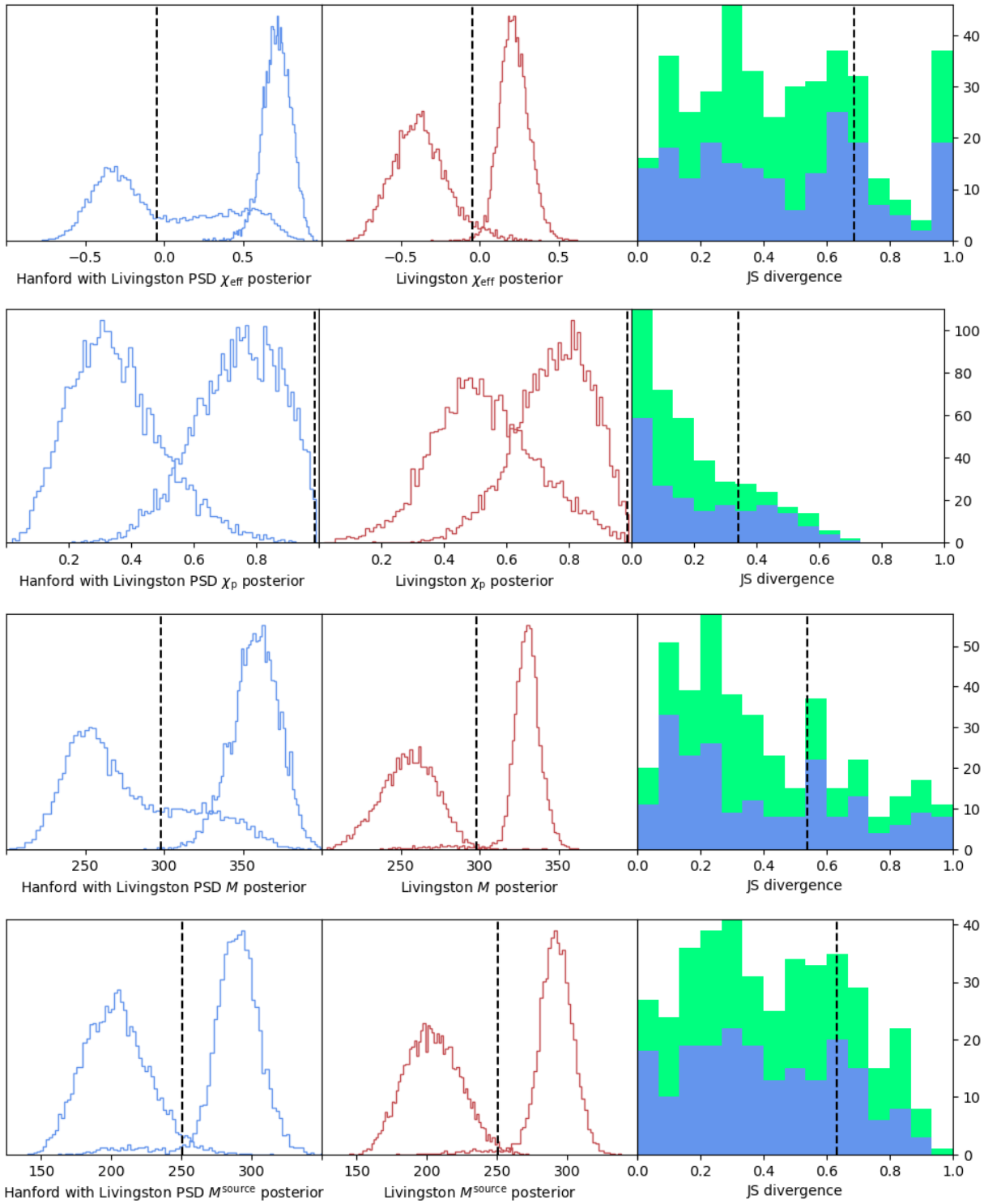

Figure C1. Left and central panels: extreme inferred posteriors for effective aligned spin parameter, effective precessing spin parameter, detector-frame total mass and source-frame total mass from the Gaussian noise maximum likelihood injections for LIGO Hanford-with-LIGO-Livingston-PSD-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds; blue shows JS divergences of combinations of LIGO Hanford-with-LIGO-Livingston-PSD-only and LIGO Livingston-only runs, green shows JS divergences of combinations of LIGO Hanford-only and LIGO Livingston-only runs. Gray dashed lines on the left two histograms mark the injected NRSur7dq4 parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.

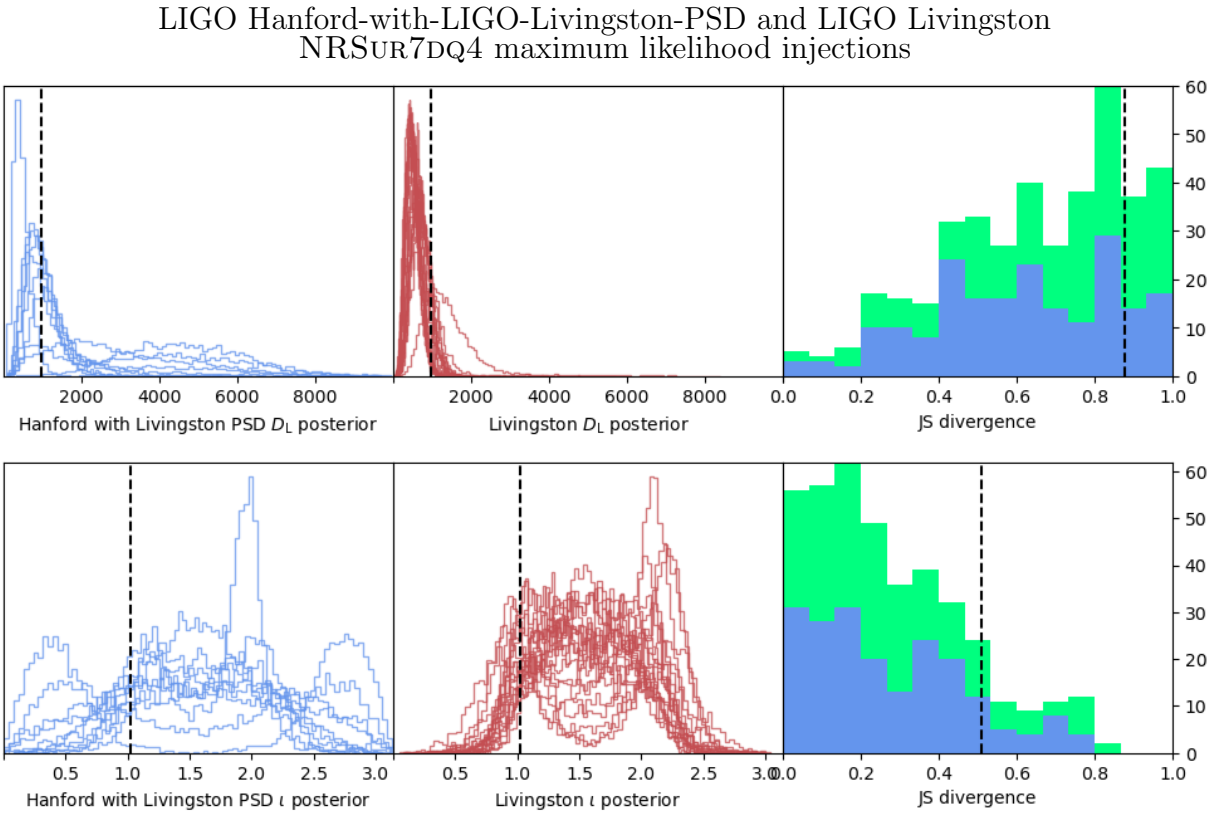

Figure C2. Left and central panels: inferred posteriors for luminosity distance and inclination angle from the Gaussian noise maximum likelihood injections for LIGO Hanford-with-LIGO-Livingston-PSD-only and LIGO Livingston-only runs, respectively. Right: calculated JS divergences for combinations of seeds; blue shows JS divergences of combinations of LIGO Hanford-with-LIGO-Livingston-PSD-only and LIGO Livingston-only runs, green shows JS divergences of combinations of LIGO Hanford-only and LIGO Livingston-only runs. Gray dashed lines on the left two histograms mark the injected NRSur7dq4 parameters. Gray dashed lines on the rightmost histograms mark the JS divergence values for GW231123.