Measuring binary black hole orbital-plane spin orientations

Vijay Varma,[1,2,3,*] Maximiliano Isi,[4,5,†] Sylvia Biscoveanu,[4,5] Will M. Farr,[6,7] and Salvatore Vitale[4,5]

[1] *Department of Physics, Cornell University, Ithaca, New York 14853, USA*
[2] *Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[3] *Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany*
[4] *LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[5] *Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA*
[6] *Department of Physics and Astronomy, Stony Brook University, Stony Brook NY 11794, USA*
[7] *Center for Computational Astrophysics, Flatiron Institute, New York NY 10010, USA*

(Dated: January 20, 2022)

Binary black hole spins are among the key observables for gravitational wave astronomy. Among the spin parameters, their orientations within the orbital plane, ϕ_1, ϕ_2 and $\Delta\phi = \phi_1 - \phi_2$, are critical for understanding the prevalence of the spin-orbit resonances and merger recoils in binary black holes. Unfortunately, these angles are particularly hard to measure using current detectors, LIGO and Virgo. Because the spin directions are not constant for precessing binaries, the traditional approach is to measure the spin components at some reference stage in the waveform evolution, typically the point at which the frequency of the detected signal reaches 20 Hz. However, we find that this is a poor choice for the orbital-plane spin angle measurements. Instead, we propose measuring the spins at a fixed *dimensionless* time or frequency near the merger. This leads to significantly improved measurements for ϕ_1 and ϕ_2 for several gravitational wave events. Furthermore, using numerical relativity injections, we demonstrate that $\Delta\phi$ will also be better measured near the merger for louder signals expected in the future. Finally, we show that numerical relativity surrogate models are key for reliably measuring the orbital-plane spin orientations, even at moderate signal-to-noise ratios like $\sim 30 - 45$.

I. INTRODUCTION.

Binary black hole (BH) spins leave characteristic imprints on the gravitational-wave (GW) signals observed by LIGO [1] and Virgo [2]. Measuring the spin parameters (illustrated in Fig. 1) from these signals will allow us to identify which astrophysical processes play a role in the binary evolution. For example, if the spins are tilted with respect to the orbital angular momentum, spin-orbit and spin-spin coupling cause both the spins and the orbital plane to precess [3, 4]. On the other hand, the orbital-plane spin orientations, ϕ_1, ϕ_2 and $\Delta\phi = \phi_1 - \phi_2$, can be used to identify spin-orbit resonances [5] and infer merger kick velocities [6].

Unfortunately, measuring the individual spin degrees of freedom from GW events is challenging at current detector sensitivities. This is particularly true for the orbital-plane spin angles ϕ_1, ϕ_2 and $\Delta\phi$ [7–9] (although see Refs. [10–12]). For instance, these measurements are typically not shown in LIGO-Virgo Collaboration (LVC) publications (e.g. Ref. [13]) as they are very poorly constrained. In this paper, we show that this can be significantly improved by a simple change in the reference point at which the spins are measured.

Because the spin directions are not constant for precessing binaries, spin measurements are inherently tied to a specific moment in the binary's evolution. In practice, the spins are measured by varying the spin parameters of a GW model at a given reference point in the inspiral and matching the predicted signal to the observed data (cf. Sec. II). The traditional approach is to measure the spins at the point where the frequency of the GW signal at the detector reaches a prespecified reference value, typically $f_{\rm ref} = 20\,{\rm Hz}$ [13]. This is mainly motivated by the fact that the sensitivity band of current detectors begins near this value [1, 2].

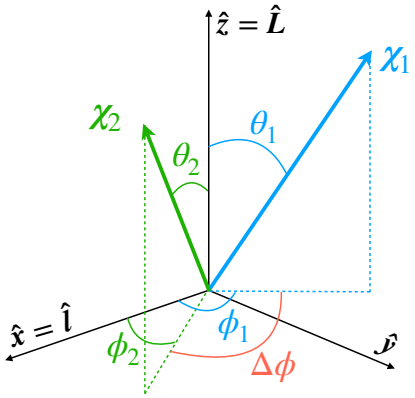

Figure 1. The binary BH spin parameters. The spins are represented by 3-vectors $\boldsymbol{\chi}_1$ and $\boldsymbol{\chi}_2$, with index 1 (2) denoting the heavier (lighter) BH. It is convenient to parameterize the spins by their dimensionless magnitudes $\chi_1, \chi_2 \leq 1$, tilts θ_1, θ_2 w.r.t the orbital angular momentum $\boldsymbol{L}$ [14], and orbital-plane spin angles ϕ_1, ϕ_2 w.r.t the line of separation $\boldsymbol{l}$ from the lighter to the heavier BH. Finally, $\Delta\phi = \phi_1 - \phi_2$.

* vijay.varma@aei.mpg.de; Klarman fellow; Marie Curie fellow
† NHFP Einstein fellow

GW events with $M \gtrsim 60M_\odot$				
GW150914	GW170729	GW170809	GW170818	GW170823
GW190413_052954	GW190413_134308	GW190421_213856	GW190424_180648	GW190503_185404
GW190513_205428	GW190514_065416	GW190517_055101	GW190519_153544	GW190521
GW190521_074359	GW190527_092055	GW190602_175927	GW190620_030421	GW190630_185205
GW190701_203306	GW190706_222641	GW190719_215514	GW190727_060333	GW190731_140936
GW190803_022701	GW190828_063405	GW190909_114149	GW190910_112807	GW190915_235702
GW190929_012149				

Table I. The 31 GW events for which we use the `NRSur7dq4` model.

We propose a different approach: instead of measuring the spins at a given signal frequency, we can measure them at a reference point near the merger. This can be achieved for all binaries by measuring spins at either (i) the point where the GW frequency reaches a fixed *dimensionless frequency*, $Mf_{\rm ref} = Mf_{\rm ISCO} = 6^{-3/2}/\pi$, or (ii) a fixed *dimensionless time*, $t_{\rm ref}/M = -100$ before the GW amplitude reaches its peak, as defined in Eq. (5) of Ref. [15]. Throughout this paper, we use geometric units with $G = c = 1$, and masses refer to the redshifted, detector-frame values. Also, here $M = m_1 + m_2$ is the total mass of the binary with component masses $m_1 \geq m_2$, and ISCO stands for the Schwarzschild innermost stable circular orbit [16]. Although the ISCO is only well-defined in the point-particle limit, we follow previous literature (e.g., Ref.[17]) in *defining* the binary's ISCO frequency to be that of an isolated Schwarzschild BH with mass equal to the total binary mass M. This reference point typically occurs within $\sim 1-4$ GW cycles before the peak amplitude. Similarly, the reference point of $t_{\rm ref}/M = -100$ typically falls within $\sim 2-4$ GW cycles before the peak amplitude. Therefore, independent of the binary parameters, both of these choices allow us to measure the spins near the merger.

Because the spins evolve deterministically, all choices of reference point lead to the same waveform prediction if correctly specified (i.e., the GW model is also evaluated using spins evolved to that reference point). Nevertheless, not all reference points are equivalent for spin measurements in practice. There are two considerations to take into account when choosing a reference point: (1) if the reference point falls well outside the sensitive window of the detector, parameter inference can become inefficient [18]; additionally, (2) the waveform itself can be more (less) sensitive to variations in the spin parameters at some reference point, leading to more (less) precise constraints on the spins at those reference points. In particular, the key finding of this paper is that choosing a reference point near the merger leads to improved constraints on the orbital-plane spin angles. We show that this is due to the waveform being more sensitive to parameterspace variations in these angles near the merger (cf. Sec. III B).

While the traditional choice of $f_{\rm ref} = 20\,{\rm Hz}$ accounts for consideration (1) above, it is not optimal when it comes to consideration (2)—we find that this makes it a poor choice for measuring the orbital-plane spin angles. On the other hand, the reference points that we propose, $t_{\rm ref}/M = -100$ and $Mf_{\rm ref} = 6^{-3/2}/\pi$, satisfy both criteria, and so they provide improved spin measurements. Furthermore, since binary BHs observed by LIGO-Virgo are expected to always merge within the instruments' sensitivity band, $t_{\rm ref}/M = -100$ and $Mf_{\rm ref} = 6^{-3/2}/\pi$ tend to fall within the detector bandwidth and are just as straightforward to interpret as spins measured at $f_{\rm ref} = 20\,{\rm Hz}$.

An executive summary of this paper is as follows. We find that measuring spins near the merger (at either $Mf_{\rm ref} = 6^{-3/2}/\pi$ or $t_{\rm ref}/M = -100$) leads to a marked improvement in the constraints of ϕ_1 and ϕ_2 (but not $\Delta\phi$) for several GW signals in the latest GWTC-2 catalog of events [13, 19–22] released by the LVC. Furthermore, we use numerical relativity (NR) injections to demonstrate that all three angles, including $\Delta\phi$, will be better measured near the merger for louder signals expected in the future. Finally, we study how well different waveform models are able to recover the orbital-plane spin angles from NR injections, and show that NR surrogate models alone are accurate enough to reliably measure these angles, even for moderate signal-to-noise ratios (SNRs) like $\sim 30-45$.

The improvement in the constraints obtained by measuring the spins near merger does not reflect a gain of new information about the source, but rather that the waveform is more sensitive to variations in the orbital-plane spin angles at this point. For instance, we find that evolving the spins measured at $f_{\rm ref} = 20\,{\rm Hz}$ to the reference point $t_{\rm ref}/M = -100$ ($Mf_{\rm ref} = 6^{-3/2}/\pi$) gives results consistent with measuring the spins directly at $t_{\rm ref}/M = -100$ ($Mf_{\rm ref} = 6^{-3/2}/\pi$). This also means that our method can be applied entirely in post-processing. Even though no new information is extracted from the data, our improved constraints can have important implications, providing a better representation of the measurement. For example, in a companion paper Varma *et al.* [23], we use the GWTC-2 spin measurements form this work to constrain the astrophysical distributions of the orbital-plane spin angles at $t_{\rm ref}/M = -100$. Some of the features found in Ref. [23], such as an unexpected peak in the ϕ_1 distribution, are only resolvable when the spins are measured near the merger.

The rest of the paper is organized as follows. We describe our parameter estimation setup in Sec. II. In Sec. III, we discuss the orbital-plane spin angle measurements for GWTC-2 events. In Sec. IV, we describe our NR injection study for louder signals as well as comparison of different waveform models. Finally, in Sec. V, we provide concluding remarks.

II. PARAMETER ESTIMATION SETUP

We obtain measurements of binary parameters from GW signals using Bayes' theorem (see Ref. [24] for a review):

$$p(\boldsymbol{\lambda}|d) \propto \mathcal{L}(d|\boldsymbol{\lambda})\,\pi(\boldsymbol{\lambda}), \tag{1}$$

where $p(\boldsymbol{\lambda}|d)$ is the *posterior* probability distribution of the binary parameters $\boldsymbol{\lambda}$ given the observed data d, $\mathcal{L}(d|\boldsymbol{\lambda})$ is the *likelihood* of the data given $\boldsymbol{\lambda}$, and $\pi(\boldsymbol{\lambda})$ is the *prior* probability distribution for $\boldsymbol{\lambda}$. For quasicircular binary BHs, the full set of binary parameters $\boldsymbol{\lambda}$ is 15 dimensional [13], and includes the masses and spins of the component BHs as well as extrinsic properties such as the distance and sky location. Under the assumption of Gaussian detector noise, the likelihood $\mathcal{L}(d|\boldsymbol{\lambda})$ can be evaluated for any $\boldsymbol{\lambda}$ using a gravitational waveform model and the observed data stream d [24]. A stochastic sampling algorithm is then used to draw *posterior samples* for $\boldsymbol{\lambda}$ from $p(\boldsymbol{\lambda}|d)$.

Our main results are obtained using the time-domain NR surrogate waveform model `NRSur7dq4` [15]. This model accurately reproduces precessing NR simulations and is currently the most accurate model in its regime of validity [15]. `NRSur7dq4` is trained on generically precessing NR simulations with mass ratios $q \leq 4$ and spin magnitudes $\chi_1, \chi_2 \leq 0.8$, but can be extrapolated to $q = 6$ and $\chi_1, \chi_2 \leq 1$ [15]. Wherever comparison with NR is possible in the extrapolated region, `NRSur7dq4` performs better than alternate models [15, 25].

We use the `Parallel Bilby` [26] parameter estimation package with the `dynesty` [27] sampler. Following Ref. [13], we choose a prior that is uniform in spin magnitudes (with $0 \leq \chi_1, \chi_2 \leq 0.99$) and component masses, and isotropic in spin orientations, sky location and binary orientation. Our distance prior is flat-in-comoving-volume [13, 28]. In addition, we place the following constraints: $12 \leq \mathcal{M} \leq 400$, $q \leq 6$, and $60 \leq M \leq 400$, where $\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$ is the chirp mass, and $q = m_1/m_2 \geq 1$ is the mass ratio. These choices are motivated by the regime of validity of `NRSur7dq4`.

In addition to predicting the waveform, `NRSur7dq4` also predicts the spin and orbital dynamics by numerically solving a set of ordinary differential equations (ODEs) [15]. The ODE integration can be initialized at any reference point in the inspiral. The model then evolves the component spins (and orbital dynamics) both forwards and backwards in time, and uses the evolved spins for its internal fits. During the inspiral, the ODE is informed by NR spins and dynamics. However, once the two BHs merge, the individual BH spins are no longer available in NR [29]. Therefore, starting at a time $t/M = -100$ before the peak amplitude, `NRSur7dq4` switches to post-Newtonian-inspired equations to evolve the individual BH spins past the merger-ringdown stage [15, 30]. Here, the choice of $t/M = -100$ is arbitrary, but once again, designed to be near the merger. The spins extended past $t/M = -100$ are not meant to be physical, but rather a convenient parameterization for the `NRSur7dq4` internal fits in the merger-ringdown [15]. For this reason, we choose to measure the spins at $t_{\rm ref}/M = -100$, the closest point to the merger where the spins are still guaranteed to be physical.

Besides $t_{\rm ref}/M = -100$, we measure the spins at $Mf_{\rm ref} = Mf_{\rm ISCO} = 6^{-3/2}/\pi$ and $f_{\rm ref} = 20\,\text{Hz}$ for comparison. Measuring the spins at $Mf_{\rm ref} = 6^{-3/2}/\pi$ also has the same benefits as $t_{\rm ref}/M = -100$, but $Mf_{\rm ref} = 6^{-3/2}/\pi$ is more convenient for frequency-domain models. While `NRSur7dq4` also allows this, we find that for some GW events, the ISCO is reached at a time after $t_{\rm ref}/M = -100$, which can result in unphysical spins. Therefore, while we provide some results at $Mf_{\rm ref} = 6^{-3/2}/\pi$ to demonstrate its efficacy, we will use spin measurements at $t_{\rm ref}/M = -100$ for our main results.

III. ORBITAL-PLANE SPIN ANGLE MEASUREMENTS

The GWTC-2 catalog [13, 19] includes a total of 46 binary BH events. However, because `NRSur7dq4` only includes $\sim$20 orbits before merger, it can only be applied to events with $M \gtrsim 60\,M_\odot$ [15] (for a detector start frequency of 20Hz). This reduces the set of events to 31; these are listed in Tab. I. All results in this section are obtained using `NRSur7dq4` for these events, which we will refer to as the "`NRSur7dq4` events" for convenience. We provide some results for all 46 events using the `IMRPhenomTPHM` model [31] in App. B.

A. Spin measurements for GW events

We first consider GW170818 [19], the event for which we see the greatest improvement when the spins are measured near the merger. Figure 2 shows the ϕ_1, ϕ_2 and $\Delta\phi$ measurements for GW170818 at the three different reference points, $t_{\rm ref}/M = -100$, $Mf_{\rm ref} = 6^{-3/2}/\pi$ and $f_{\rm ref} = 20\,\text{Hz}$. Here, for the joint 2D posteriors, we show 70% contours instead of the more commonly used 90% and 50% contours [13], as we find the 70% contours represent the bulk of the probability mass while being more instructive to discuss the correlations below.

First considering the marginalized 1D distributions in Fig. 2, we find that ϕ_1 and ϕ_2 measured at $t_{\rm ref}/M = -100$ and $Mf_{\rm ref} = 6^{-3/2}/\pi$ are significantly better constrained

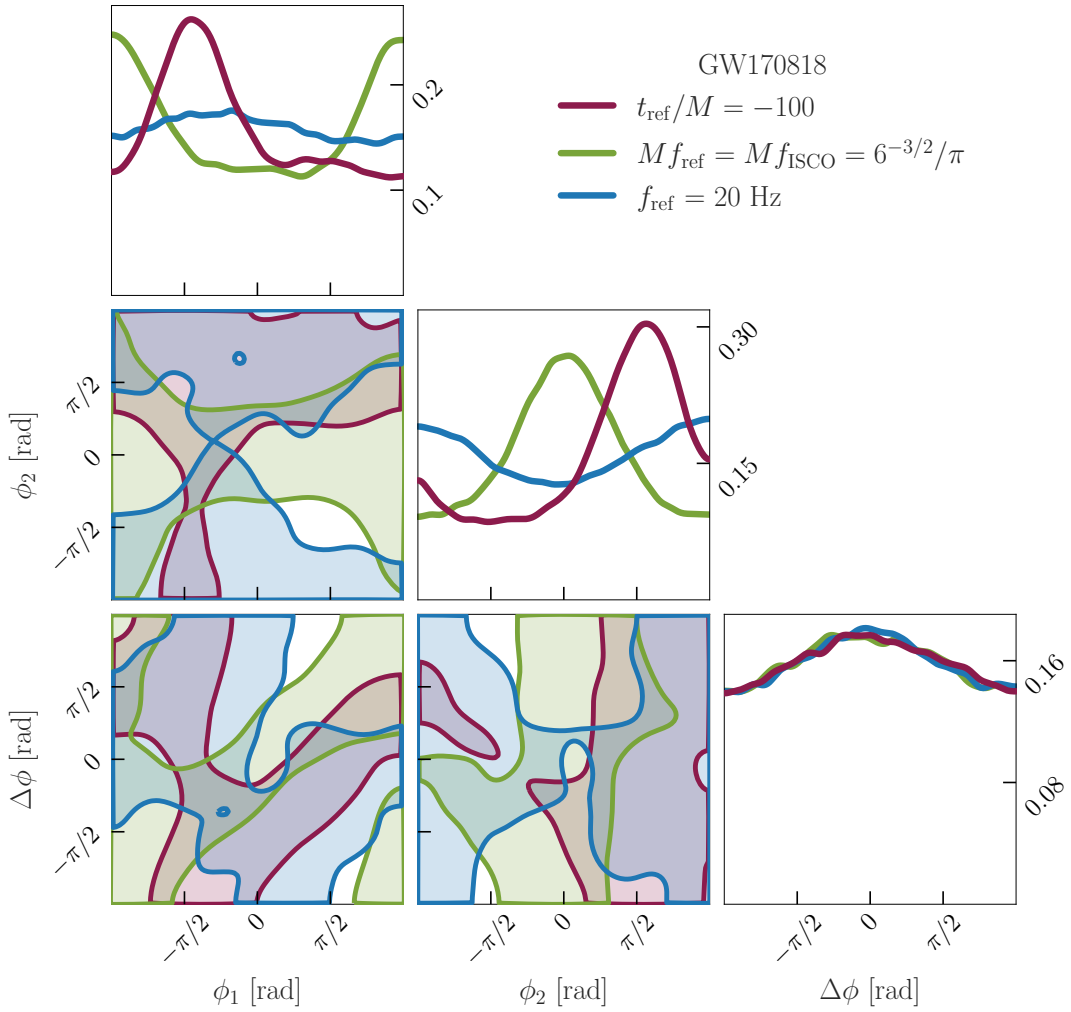

Figure 2. Spin angles ϕ_1, ϕ_2 and $\Delta\phi$ for GW170818 using `NRSur7dq4` at the three different reference points. The lower-triangle subplots show central 70% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors. The 1D ϕ_1 and ϕ_2 posteriors are more sharply peaked when measured at $t_{\rm ref}/M = -100$ or $Mf_{\rm ref} = 6^{-3/2}/\pi$. Similarly, the 2D posteriors are generally better constrained at $t_{\rm ref}/M = -100$ or $Mf_{\rm ref} = 6^{-3/2}/\pi$, compared to $f_{\rm ref} = 20\,$Hz.

than those at $f_{\rm ref} = 20\,$Hz. Note that ϕ_1 and ϕ_2 peak in different regions for the three different reference points, while $\Delta\phi$ is consistent for each of them. This is because ϕ_1 and ϕ_2 change on the orbital timescale, as they are defined with respect to the line-of-separation (cf. Fig. 1). On the other hand, $\Delta\phi$ changes only on the precession timescale, which is longer. Furthermore, while the peaks of ϕ_1 and ϕ_2 are approximately π apart, this does not result in a $\Delta\phi$ peak near $\pm\pi$. Instead, the $\Delta\phi$ posterior is much broader, with a mild peak near 0. This suggests that even for spins measured near the merger ($t_{\rm ref}/M = -100$ or $Mf_{\rm ref} = 6^{-3/2}/\pi$), the data is only informative about ϕ_1 or ϕ_2, but not necessarily both at the same time. In fact, we do not see any significant improvement in the 1D $\Delta\phi$ posterior near the merger for this event.

However, examining the 2D distributions in Fig. 2, we find that the posteriors for all three combinations of ϕ_1, ϕ_2 and $\Delta\phi$ are generally better constrained at $t_{\rm ref}/M = -100$ or $Mf_{\rm ref} = 6^{-3/2}/\pi$ compared to $f_{\rm ref} = 20\,$Hz. The 2D posteriors also show a significant amount of correlation between the three angles. The main feature here is that ϕ_1 and ϕ_2 are better measured than $\Delta\phi$, as noted above. Therefore, the correlations are along vertical and horizontal directions in the $\phi_1 - \phi_2$ posterior, while they are along diagonal directions for the $\phi_1 - \Delta\phi$ posterior (and to a lesser extent for the $\phi_2 - \Delta\phi$ posterior). We note that similar correlations are absent for the higher SNR injections shown in Sec. IV. This suggests that the degeneracies we see in the 2D posteriors of Fig. 2 are a function of the SNR, and can be broken for louder signals.

In the rest of this section, we will focus on 1D marginalized posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz for simplicity. While GW170818 shows the biggest improvement in ϕ_1 and ϕ_2 when measured at $t_{\rm ref}/M = -100$, we find that several other events in GWTC-2 also show significant improvements. Figure 3 compares marginalized

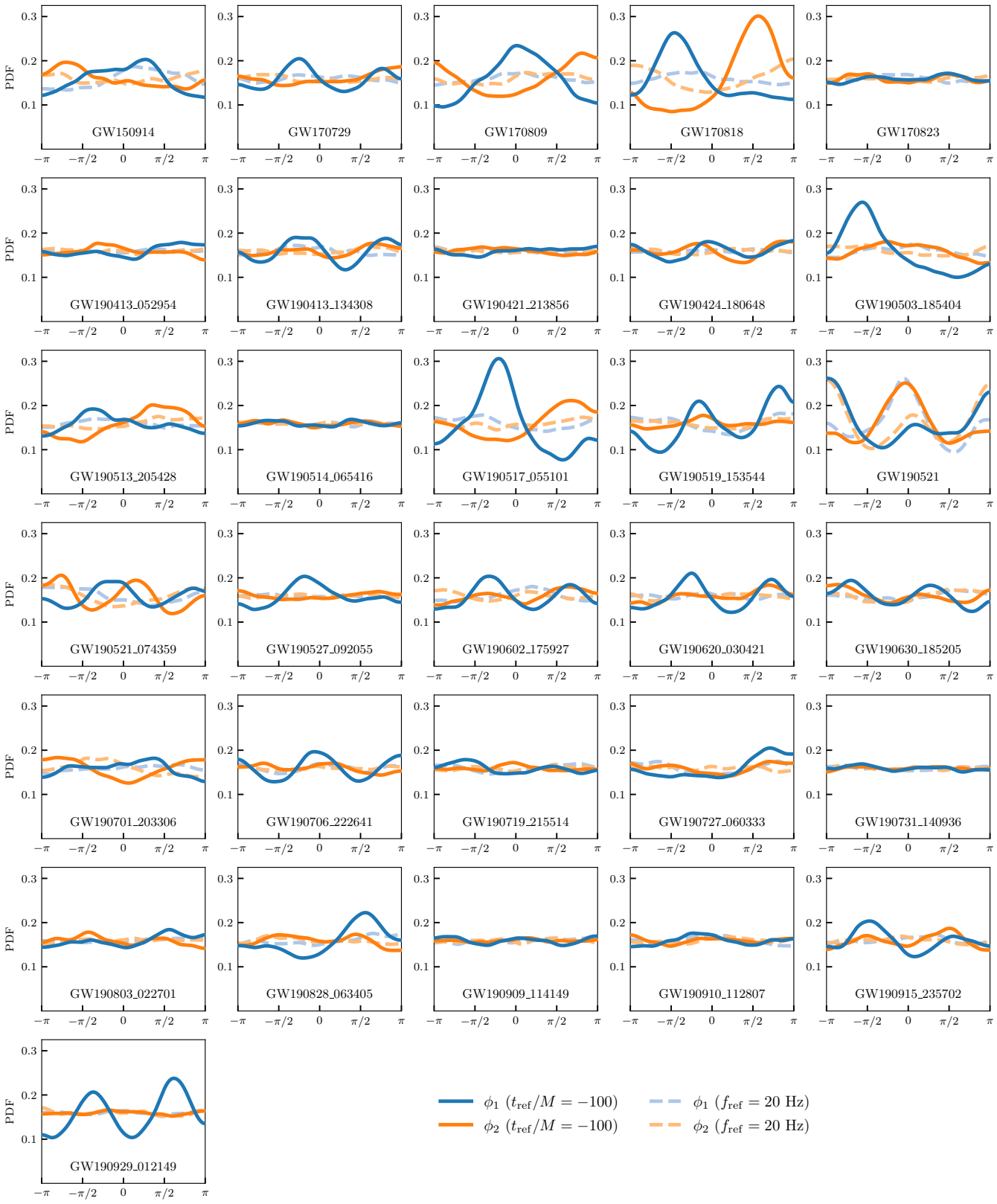

Figure 3. ϕ_1 and ϕ_2 posteriors at $t_{\rm ref}/M = -100$ (solid) and $f_{\rm ref} = 20\,$Hz (dashed) for `NRSur7dq4`. The distributions at $f_{\rm ref} = 20\,$Hz are mostly flat (with the exception of GW190521, which is explained in Sec. III A). By contrast, at $t_{\rm ref}/M = -100$, several cases show a clear deviation from a flat distribution.

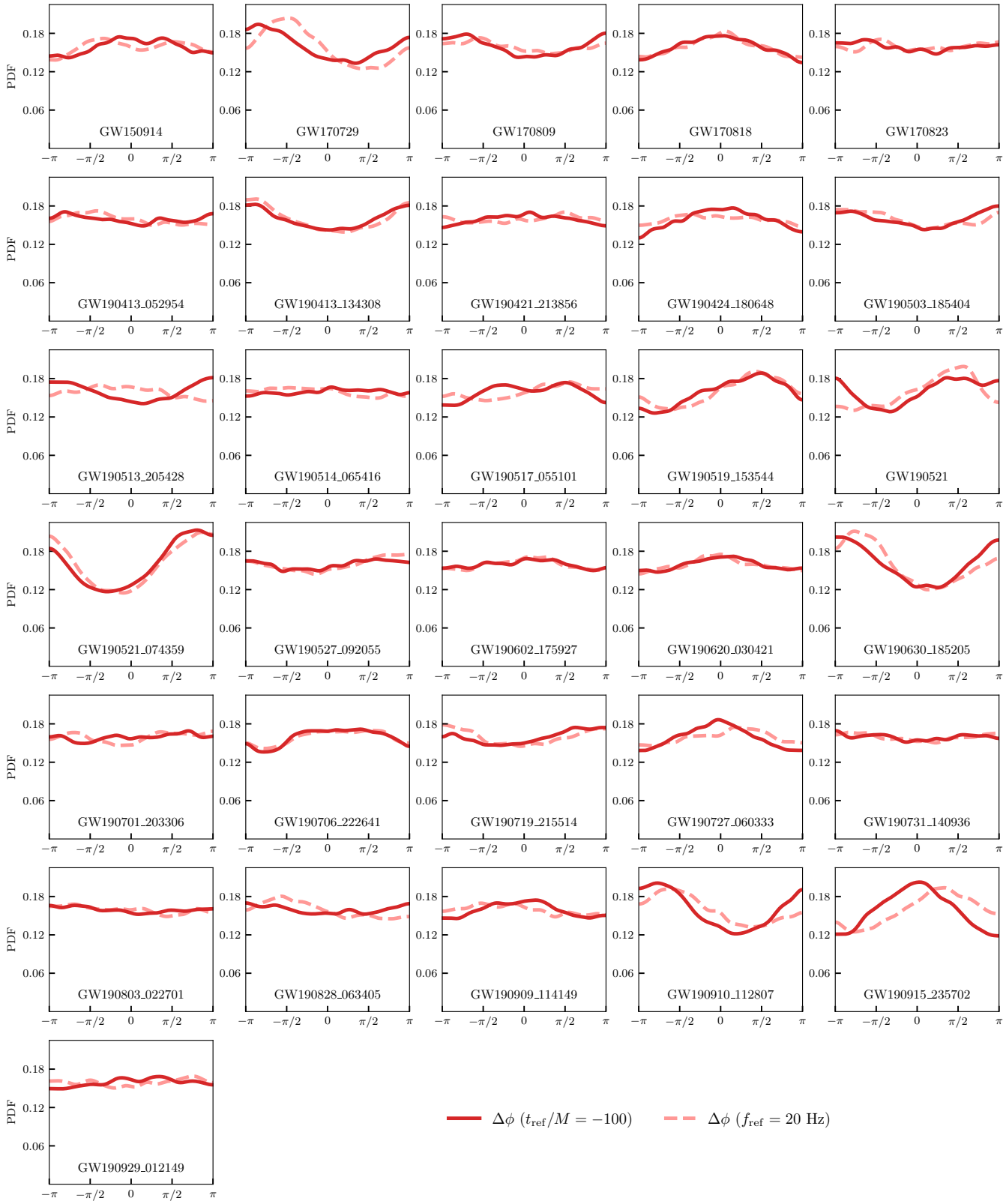

Figure 4. $\Delta\phi$ posteriors at $t_{\rm ref}/M = -100$ (solid) and $f_{\rm ref} = 20\,$Hz (dashed) for `NRSur7dq4`. Even at $t_{\rm ref}/M = -100$, $\Delta\phi$ is less well-measured than ϕ_1 or ϕ_2 (cf. Fig. 3). In fact, $\Delta\phi$ measurements are comparable at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz.

1D posteriors for ϕ_1 and ϕ_2 measured at $f_{\rm ref}=20\,$Hz and $t_{\rm ref}/M=-100$ for all 31 `NRSur7dq4` events. While the measurements at $f_{\rm ref}=20\,$Hz are mostly uninformative and consistent with a uniform distribution, the measurements at $t_{\rm ref}/M=-100$ show clear deviations from uniformity for several events. Interestingly, GW190521 [32] is the only event with a good measurement at $f_{\rm ref}=20\,$Hz. This is explained by the fact that this binary merges at a low frequency due to its high mass ($M\sim 270M_\odot$), therefore its $f_{\rm ISCO}\sim 16$ Hz happens to be close to 20 Hz.

Figure 4 compares 1D $\Delta\phi$ posteriors measured at $f_{\rm ref}=20\,$Hz and $t_{\rm ref}/M=-100$. Similar to Fig. 2, $\Delta\phi$ is less well-measured than ϕ_1 and ϕ_2, and there is no significant improvement when measuring the spins at $t_{\rm ref}/M=-100$. However, as we will show in Sec. IV, we expect this to change with louder signals.

An unambiguous measurement of the orbital-plane spin angles relies on being able to constrain the spin magnitudes away from zero, and the tilt angles to be neither 0 nor π. For the `NRSur7dq4` events, our measurements of the spin magnitudes and tilts are consistent with Refs. [13, 19], and are shown in App. A. Most of these events are consistent with having zero spin magnitudes for both BHs [13], but there is evidence for nonzero spin magnitude in at least some of the events [33]. Secondly, even though there is evidence of precession in the astrophysical binary BH population [34], the individual events are not loud enough to show clear evidence of precession on their own [13]. Finally, the posteriors for the tilt angles do not change significantly between $f_{\rm ref}=20\,$Hz and $t_{\rm ref}/M=-100$ for these events. As a result, even with the improvements at $t_{\rm ref}/M=-100$, the 1D posteriors for ϕ_1, ϕ_2 and $\Delta\phi$ are still relatively broad (cf. Figs. 3 and 4), and do not exclude any of the allowed region between $(-\pi,\pi)$. Nevertheless, these measurements still allow us to place interesting constraints on the astrophysical distributions of the orbital-plane spin angles. This is explored in a companion paper, Ref. [23].

B. Varying the reference point

Next, we systematically study the impact of the reference point at which the spins are measured. Figure 5 shows ϕ_1 measurements at various different reference times for the 31 `NRSur7dq4` events. Rather than repeat the parameter estimation at each $t_{\rm ref}$, we use the `NRSur7dq4` spin dynamics [15] to evolve the spins backwards from $t_{\rm ref}/M=-100$ to the earlier times. As noted in the introduction, we find that this leads to results consistent with measuring spins directly at the new reference time. The earliest reference time we consider is $t_{\rm ref}/M=-4000$, which is near the start of the `NRSur7dq4` waveform's validity [15].

In Fig. 5, as we move the reference point from $t_{\rm ref}/M=-4000$ to $t_{\rm ref}/M=-100$, we see a clear improvement in the ϕ_1 constraint for several events. A possible explanation for this improvement is that the precession (and orbital) timescale decreases as the merger approaches, making the waveform more sensitive to the orbital-plane spin angles near the merger as a result. We provide further justification for this with a Fisher matrix analysis in the following.

1. *Fisher matrix analysis*

As a proxy for the sensitivity of the waveform to the orbital-plane spin angles, we can look to the Fisher information matrix. The Fisher matrix provides a simple way to estimate the statistical uncertainty in measuring binary BH parameters in the high-SNR limit [35, 36]; it is defined as

$$\Gamma_{ij}=\left(\frac{\partial h}{\partial\lambda^i}\middle|\frac{\partial h}{\partial\lambda^j}\right), \tag{2}$$

where $h(t)$ is the gravitational waveform with binary parameters $\boldsymbol{\lambda}=\{\lambda^i\}$ (cf. Sec II), and the inner product $(h|g)$ is defined as

$$(h|g)=4{\rm Re}\int\frac{\tilde h^*(f)\tilde g(f)}{S_n(f)}df, \tag{3}$$

where $\tilde h(f)$ indicates the Fourier transform of $h(t)$, $*$ stands for complex conjugation, and $S_n(f)$ is the one-sided power spectral density for which we use the LIGO design sensitivity noise curve [37]. We use the `NRSur7dq4` waveform model for $h(t)$ and compute the derivatives in Eq. (2) numerically [38]. Then, using the Cramer-Rao inequality [39, 40], the measurement covariance matrix ${\rm Var}(\lambda^i,\lambda^j)$ satisfies

$${\rm Var}(\lambda^i,\lambda^j)\geq\left(\Gamma^{-1}\right)_{ij}. \tag{4}$$

Finally, taking the lower bound of the inequality, the statistical uncertainty in λ^j can be estimated as

$$\delta\lambda^i=\sqrt{(\Gamma^{-1})_{ii}}. \tag{5}$$

and the correlation coefficient between λ^i and λ^j can be estimated as,

$${\rm Corr}(\lambda^i,\lambda^j)=\frac{(\Gamma^{-1})_{ij}}{\sqrt{(\Gamma^{-1})_{ii}(\Gamma^{-1})_{jj}}}. \tag{6}$$

The Fisher matrix method reliably estimates the statistical uncertainty only in the limit of high SNR (see, e.g., Ref. [41] for caveats). Regardless, here we are not interested in the statistical uncertainty itself but in quantifying the sensitivity of the waveform to variations in the binary parameters. The Fisher matrix method is well-suited for this purpose, as we use its bound on statistical uncertainties merely as a proxy for waveform sensitivity: smaller $\delta\lambda^j$ indicates that the waveform is more sensitive to λ^j.

In particular, we are interested in how sensitive the waveform is to changes in the orbital-plane spin angles

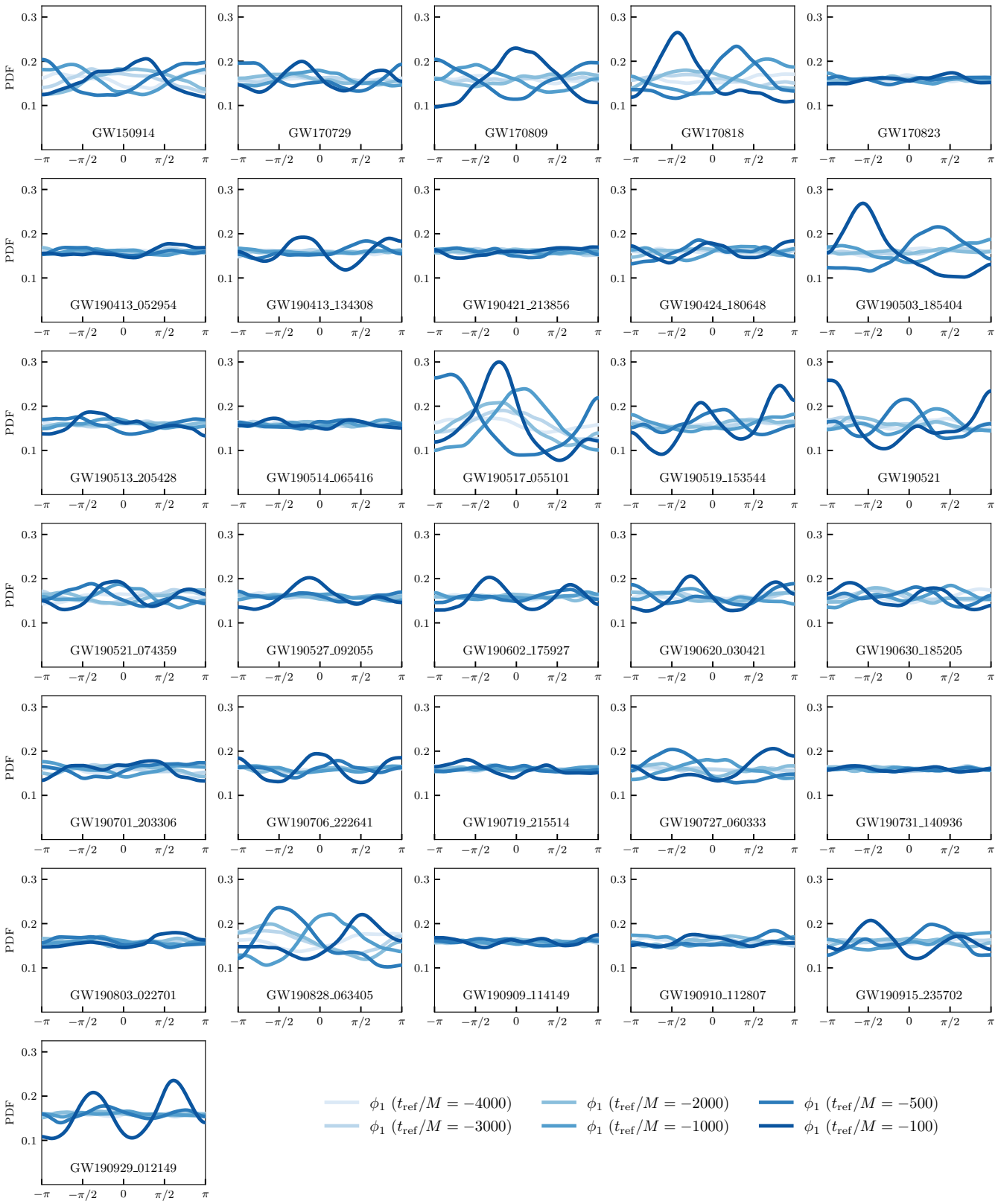

Figure 5. ϕ_1 posteriors when measured at various different $t_{\rm ref}$ for `NRSur7dq4`. Going from $t_{\rm ref}/M = -4000$ to $t_{\rm ref}/M = -100$, we see a clear improvement in the ϕ_1 measurement for several events.

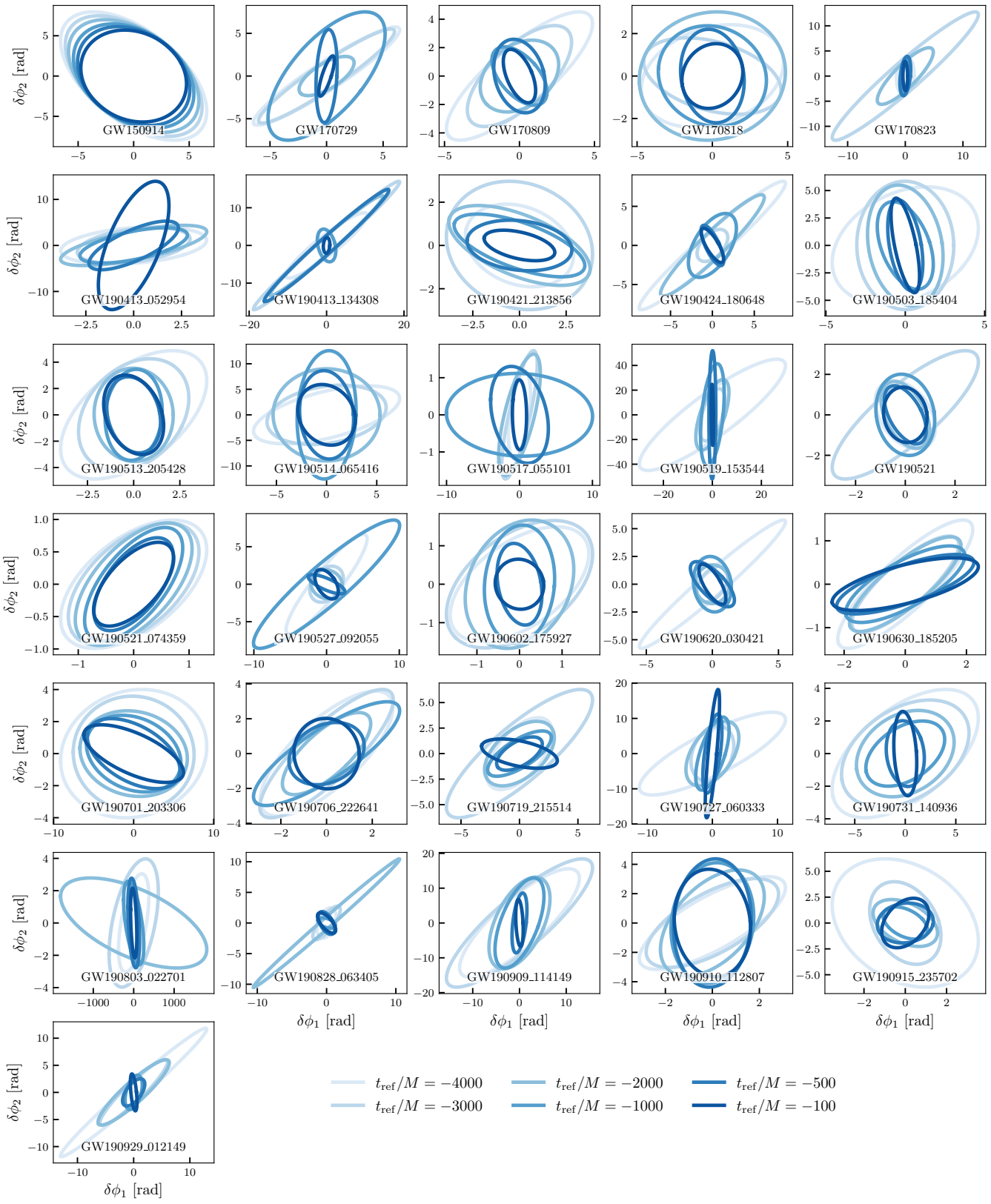

Figure 6. Estimated statistical uncertainty in ϕ_1 and ϕ_2 measurements for `NRSur7dq4` at the maximum likelihood parameters for the `NRSur7dq4` events. The uncertainties are estimated using the Fisher matrix by varying the spins at various different $t_{\rm ref}$. In almost all cases, the expected uncertainty decreases noticeably as one approaches $t_{\rm ref}/M = -100$, indicating that the waveform is more sensitive to changes in ϕ_1 and ϕ_2 near merger. Note that the Fisher matrix method does not place prior bounds on ϕ_1 and ϕ_2 to be within $(-\pi, \pi)$, therefore the statistical biases are not bound to be $\leq 2\pi$.

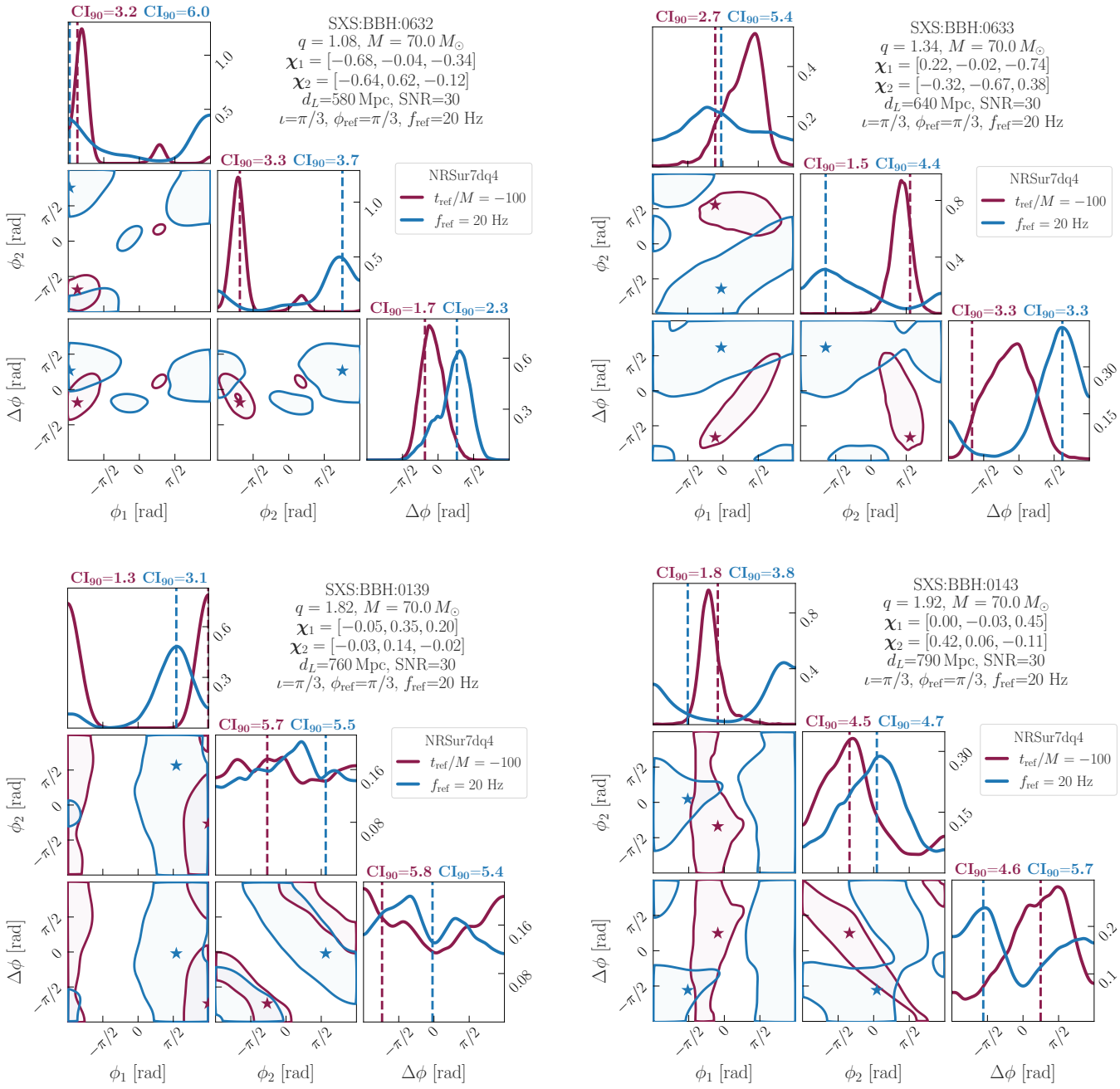

Figure 7. `NRSur7dq4` posteriors for ϕ_1, ϕ_2 and $\Delta\phi$ when measured at $t_{\rm ref}/M=-100$ and $f_{\rm ref}=20\,$Hz for NR injections at SNR=30. In each panel, the inset text shows the binary parameters for the injections, with the spins, inclination angle and orbital phase defined at $f_{\rm ref}=20\,$Hz. The shaded regions in the lower-triangle subplots show the central 90% credible regions for joint 2D posteriors, with the true value indicated by star markers (maroon for $t_{\rm ref}/M=-100$ and blue for $f_{\rm ref}=20\,$Hz). The diagonal subplots show marginalized 1D posteriors, with the true values indicated by vertical dashed lines. The width of the central 90% credible interval ($\rm CI_{90}$) for the 1D distributions are shown in text above the diagonal subplots. All orbital-plane angles, including $\Delta\phi$, are significantly better measured at $t_{\rm ref}/M=-100$.

at various values of $t_{\rm ref}$. For this purpose, we generate the waveform corresponding to the maximum likelihood parameters for each of the 31 `NRSur7dq4` events. Then, we compute $\mathrm{Corr}(\phi_1,\phi_2)$ by varying the spins for the same waveform at different $t_{\rm ref}$. We initially compute the statistical uncertainties at a fixed distance of 100 Mpc, but then rescale them following $\delta\lambda^j \propto 1/\mathrm{SNR}$ to correspond to an SNR (defined as $\sqrt{(h|h)}$) matching that of the observed event.

Figure 6 shows the statistical uncertainties in ϕ_1 and ϕ_2 as we move from $t_{\rm ref}/M=-4000$ to $t_{\rm ref}/M=-100$. In almost all cases, we see that the statistical uncertainty

decreases as we approach the merger, meaning that the waveform is generally more sensitive to variations in ϕ_1 and ϕ_2 near the merger. This explains the improved measurement at $t_{\rm ref}/M = -100$ in Fig. 2 and Fig. 3, as well as the systematic improvement as we approach $t_{\rm ref}/M = -100$ in Fig. 5.

To summarize, since the observed waveform is most sensitive to the orbital-plane spin angles near merger, the data can successfully constrain these angles at that point. However, the precision of that measurement is not preserved as we extrapolate the spins back in time because, even though the dynamics are deterministic, this detail gets smeared out during the inspiral cycles.

Finally, we note that the direction of the $\delta\phi_1 - \delta\phi_2$ correlations in Fig. 6 depend on where in the evolution they are evaluated. This is in agreement with Ref. [38], which found that the inspiral and ringdown regions of the waveform carry complimentary information.

IV. NR INJECTION STUDY

To further investigate the measurability of the orbital-plane spin angles, we consider four NR waveforms, SXS:BBH:0139, SXS:BBH:0143, SXS:BBH:0632, and SXS:BBH:0633, from the public SXS catalog [29, 42, 43]. These waveforms correspond to systems with mass ratios $q < 2$ and substantial orbital-plane spins. Note that none of these waveforms were used to train `NRSur7dq4`. We choose a total mass $M = 70M_\odot$, an inclination angle $\iota = \pi/3$ between $\boldsymbol{L}$ and the line-of-sight, and a reference orbital phase $\phi_{\rm ref} = \pi/3$. Note that ι and $\phi_{\rm ref}$ are defined at $f_{\rm ref} = 20\,$Hz. The luminosity distance is chosen such that the network matched-filter SNR is either 30 or 45. The rest of the binary parameters will be shown in figure insets below. We inject these NR waveforms (in zero-noise) into a simulated LIGO-Virgo network operating at design sensitivity [37], and recover them using different waveform models. The injection and parameter inference are done using the `Parallel Bilby` [26] package.

A. $t_{\rm ref}/M = -100$ vs $f_{\rm ref} = 20\,$Hz for `NRSur7dq4`

In Sec. III, we showed that the constraints on the orbital-plane spin angles become tighter when measured near the merger. It is important to verify that this tighter constraint does not lead to biased estimates for these angles. We verify this in Fig. 7, where we show ϕ_1, ϕ_2 and $\Delta\phi$ measured at $f_{\rm ref} = 20\,$Hz and $t_{\rm ref}/M = -100$, using `NRSur7dq4` against NR injections at SNR=30. We first note that the `NRSur7dq4` model indeed recovers the true values at both $f_{\rm ref} = 20\,$Hz and $t_{\rm ref}/M = -100$. Next, all orbital-plane angles, including $\Delta\phi$, are significantly better measured at $t_{\rm ref}/M = -100$. While this is not always clear from the 1D marginalized distributions for $\Delta\phi$, note that the 2D joint posteriors for all three combinations of ϕ_1, ϕ_2 and $\Delta\phi$ are always better constrained at $t_{\rm ref}/M = -100$.

In Fig. 7, for SXS:BBH:0633 (top-right panel), the true value falls near the edge of the 2D 90% credible region for $t_{\rm ref}/M = -100$. To check whether this is indicative of a systematic bias in `NRSur7dq4` when spins are measured at $t_{\rm ref}/M = -100$, we repeat our injections at at SNR=45 in Fig. 8. As expected, increasing the SNR leads to better constraints on the orbital-plane spins angles for both $f_{\rm ref} = 20\,$Hz and $t_{\rm ref}/M = -100$. For SXS:BBH:0633, the 2D 90% credible regions in Fig. 8 still include the true value at $t_{\rm ref}/M = -100$, suggesting that there are no significant biases. Once again, we find that some 1D posteriors can be more sharply peaked at $f_{\rm ref} = 20\,$Hz (e.g. $\Delta\phi$ in the top-right panel of Fig. 8), but the 2D posteriors are always better constrained at $t_{\rm ref}/M = -100$ for all three combinations of ϕ_1, ϕ_2 and $\Delta\phi$.

In Figs. 3 and 4, we noted that while ϕ_1 and ϕ_2 measurements are improved at $t_{\rm ref}/M = -100$ for current GW events, $\Delta\phi$ measurements are not significantly impacted. Figures 7 and 8 show that as detector sensitivity improves and GW signals are observed at higher SNR, our method will generally lead to improved measurements in $\Delta\phi$ as well. Even when the 1D posterior for $\Delta\phi$ is more sharply peaked at $f_{\rm ref} = 20\,$Hz, the overall constraints on the orbital-plane spin angles are better at $t_{\rm ref}/M = -100$. Measuring the full orbital-plane spin degrees of freedom is necessary to constrain the kick population as done in Ref. [23].

B. Waveform model comparison

In this section, we study the performance of different waveform models in recovering the orbital-plane spins angles. Apart from `NRSur7dq4`, we also consider the phenomenological models `IMRPhenomTPHM` [31] and `IMRPhenomXPHM` [44], as well as the effective-one-body model `SEOBNRv4PHM` [45]. While these models also include some effects of precession, they are not calibrated on precessing NR simulations. Note that `IMRPhenomTPHM` and `SEOBNRv4PHM` are time-domain models, while `IMRPhenomXPHM` is a frequency-domain model. We only consider spin measurements at $f_{\rm ref} = 20\,$Hz for these models as specifying spins at a dimensionless time/frequency would require careful modifications to how these models are implemented. However, because binary BH spin evolution is deterministic, any biases seen at $f_{\rm ref} = 20\,$Hz should translate to biases at $t_{\rm ref}/M = -100$ as well. We repeat our NR injections at SNRs of 30 and 45, but because `SEOBNRv4PHM` is significantly more expensive than the other models, we only apply it to the injections at SNR=45.

Figure 9 shows ϕ_1, ϕ_2 and $\Delta\phi$ posteriors obtained using `NRSur7dq4`, `IMRPhenomTPHM` and `IMRPhenomXPHM` for our NR injections at SNR=30. For two out of the four cases (SXS:BBH:0633 and SXS:BBH:0139), the true value falls on the edge of the 90% credible region of the 2D posteriors for `IMRPhenomXPHM`. The 1D marginalized posteriors are also biased in several cases for `IMRPhenomXPHM`; in partic-

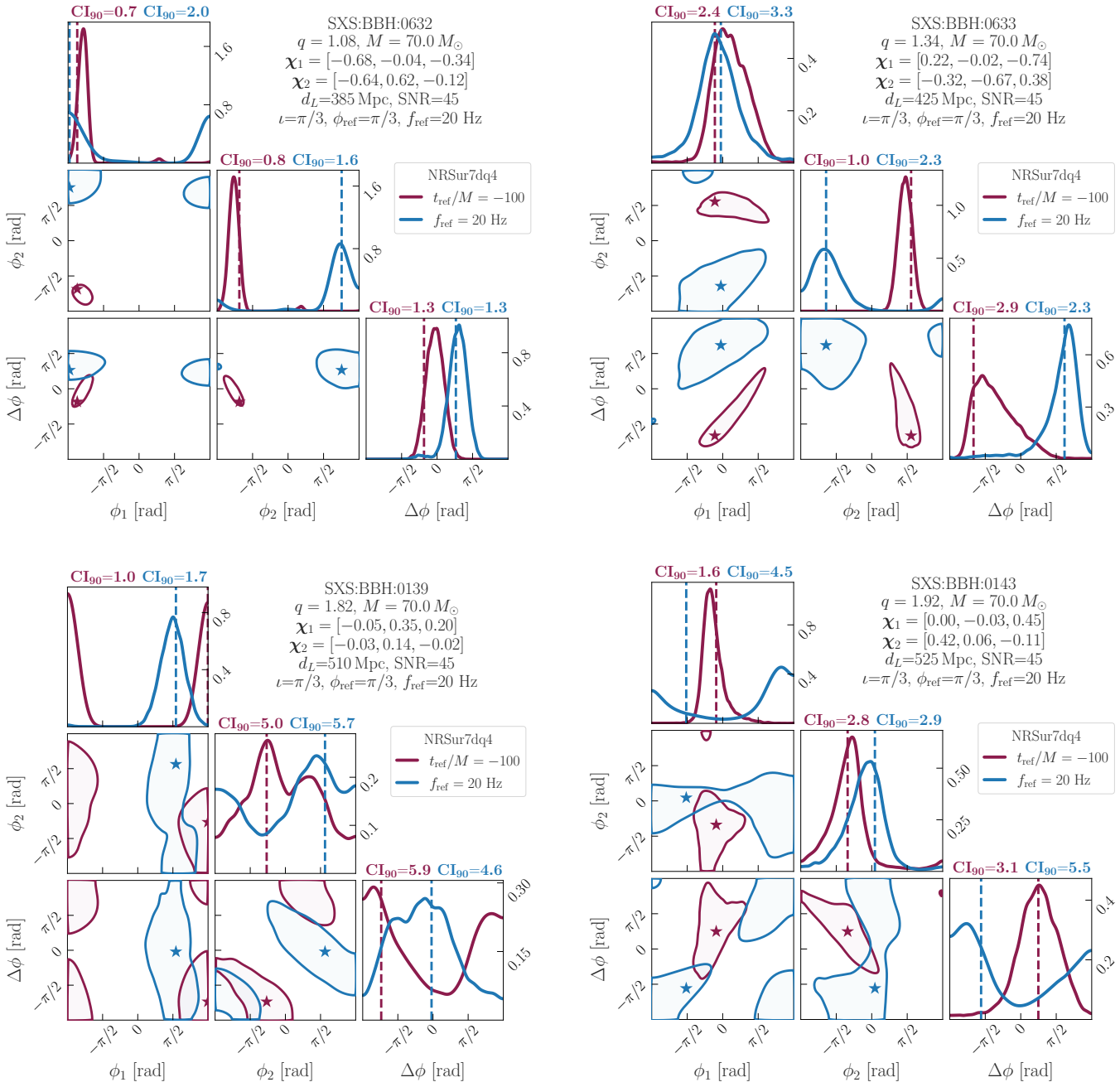

Figure 8. Same as Fig. 7, but now the SNR is increased to 45.

ular, there are cases (e.g. ϕ_1 in top-right panel and $\Delta\phi$ in the bottom-left panel of Fig. 9) where this model has the strongest peaks in the 1D distributions, but prefers the wrong value. By contrast, for both `NRSur7dq4` and `IMRPhenomTPHM`, the true value is always within the 90% credible region of the 2D posteriors.

However, for SXS:BBH:0139 (bottom-left of Fig. 9), the 1D $\Delta\phi$ posterior for `IMRPhenomTPHM` is peaked away from the true value, even though the true value is included in the 90% credible region of the 2D posteriors. For `NRSur7dq4`, the peak in the 1D $\Delta\phi$ posterior is much broader in this case and includes the true value. To check whether this is indicative of a systematic bias in `IMRPhenomTPHM`, we consider the 50% credible region of the 2D posteriors for this case, which is shown only for `IMRPhenomTPHM` for simplicity. The 50% credible region clearly excludes the true value for `IMRPhenomTPHM`, meaning that the bulk of the probability density for `IMRPhenomTPHM` is concentrated in a region away from the true value. This suggests that the deviation in the 1D $\Delta\phi$ posterior for `IMRPhenomTPHM` is indeed due to a systematic bias. This serves as an another example where

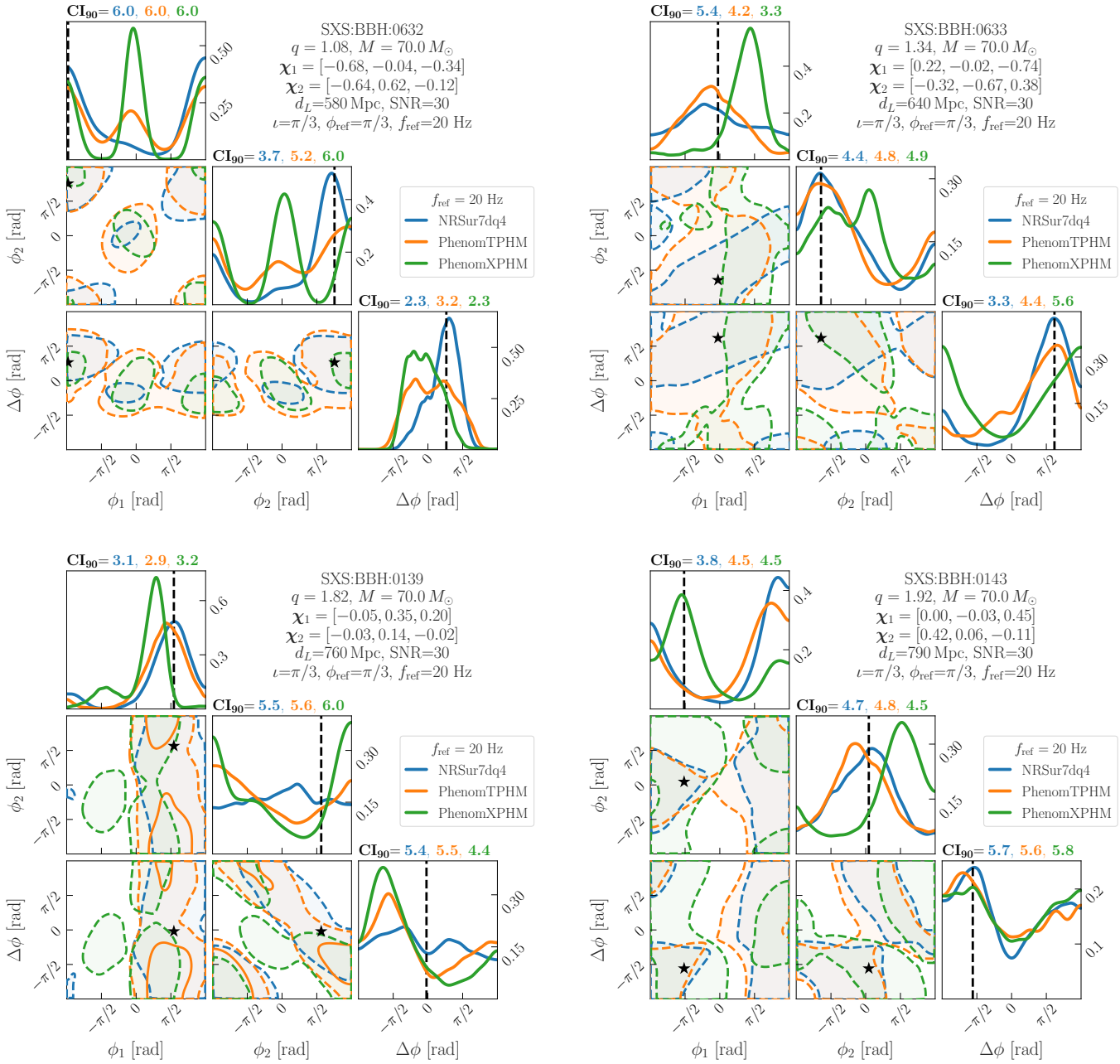

Figure 9. Same NR injections as Fig. 7, but we show posteriors at $f_{\rm ref}=20$ Hz for `IMRPhenomXPHM` and `IMRPhenomTPHM` along with `NRSur7dq4`. The lower-triangle subplots show joint 2D posteriors for ϕ_1, ϕ_2 and $\Delta\phi$, with the true value indicated by a black star. The central 90% credible regions are shown as dashed contours. Only for the bottom-left panel and `IMRPhenomTPHM`, we also show the 50% credible regions as solid contours to demonstrate the systematic bias. The diagonal subplots show marginalized 1D posteriors, with the true value indicated by a black vertical dashed line. All injections are done at an SNR of 30.

a waveform model has a stronger peak than `NRSur7dq4` in the 1D posterior, but is peaked at the wrong value.

Similarly, for SXS:BBH:0143 (bottom-right of Fig. 9), the true value is included in the 2D posteriors but the 1D ϕ_1 posteriors appear biased for both `IMRPhenomTPHM` and `NRSur7dq4`. However, in this case the primary BH has negligible spin in the orbital-plane, therefore ϕ_1 is not a meaningful parameter and hence the offset from the true value is not of concern.

We repeat these NR injections at SNR=45 in Fig. 10, now also including the `SEOBNRv4PHM` model. We now find that the true value is fully excluded from the 90% credible region of the 2D posteriors for `IMRPhenomXPHM` for three out of our four injections. While `IMRPhenomTPHM` still performs better than `IMRPhenomXPHM`, `IMRPhenomTPHM` also excludes the true value from the 90% credible region of the 2D posteriors for SXS:BBH:0143 (bottom-right of Fig. 10). For SXS:BBH:0139 (bottom-left of Fig. 10),

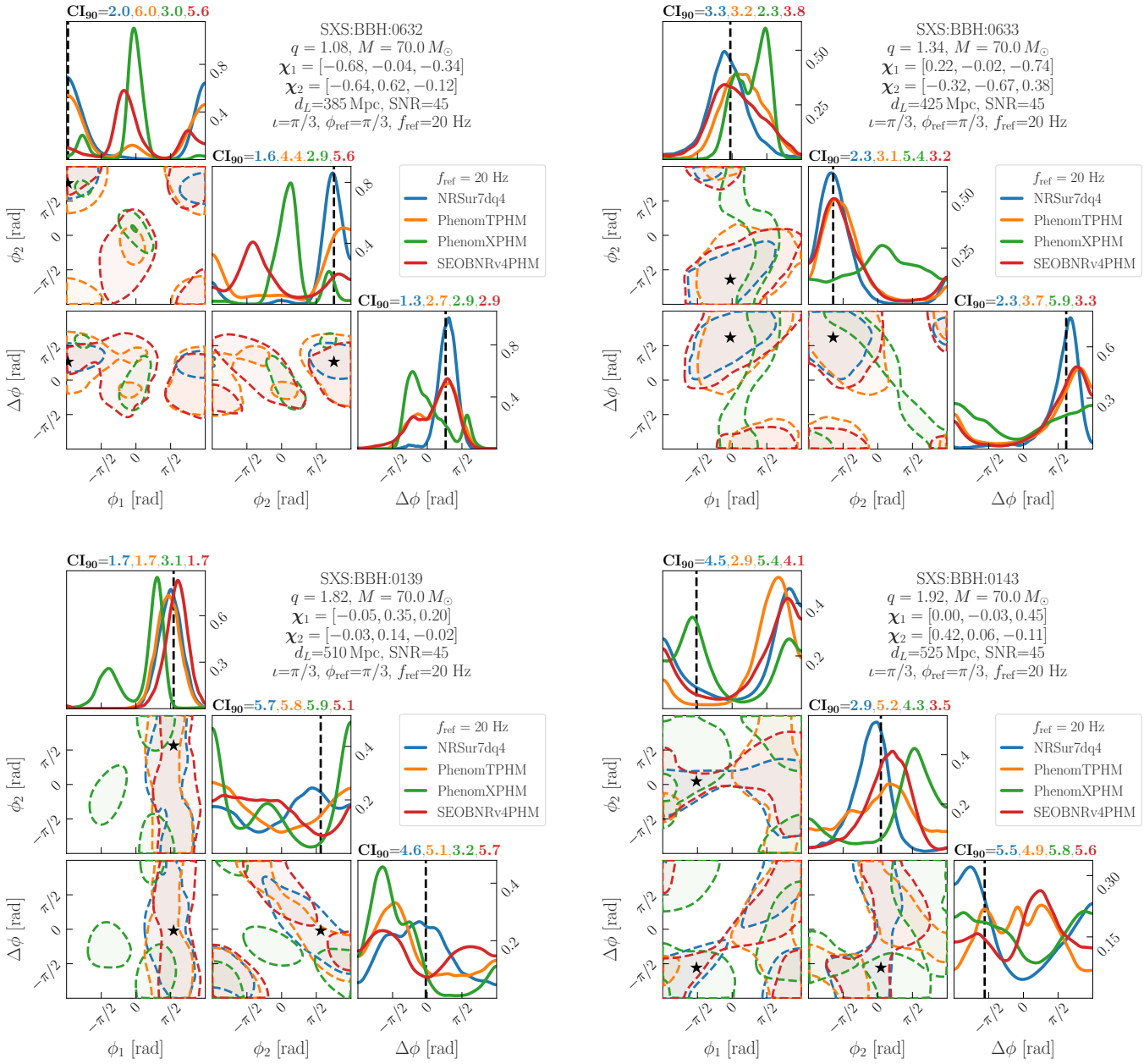

Figure 10. Same as Fig. 9, but now the SNR is increased to 45, and only 90% contours are shown for all joint posteriors.

the 1D $\Delta\phi$ posterior is still biased for `IMRPhenomTPHM`, while `NRSur7dq4` now has a clear peak around the true value. This suggests that `NRSur7dq4` is not prone to the systematic biases present in `IMRPhenomTPHM` (and `IMRPhenomXPHM`) as noted above.

In Fig. 10, `SEOBNRv4PHM` is comparable to `IMRPhenomTPHM`, including the true value in the 90% region of the 2D posteriors for three out of four cases, with the exception being SXS:BBH:0139 (bottom-left of Fig. 10). For this case, `SEOBNRv4PHM` also shows similar biases in the $\Delta\phi$ 1D posterior as `IMRPhenomTPHM` (and `IMRPhenomXPHM`), meaning that this model is also prone to the systematic biases noted above. Furthermore, for SXS:BBH:0632 (top-left of Fig. 10), the 1D posteriors for ϕ_1 and ϕ_2 for `SEOBNRv4PHM` are biased (although they are somewhat included in the smaller secondary modes).

Finally, we note that `NRSur7dq4` generally leads to the best constraints in Figs. 9 and 10, which is expected as this is the only model trained on precessing NR simulations (but not the ones injected here). Instead, the `IMRPhenomTPHM`, `IMRPhenomXPHM` and `SEOBNRv4PHM` models approximate orbital precession by "twisting" [31, 44, 45] a corresponding nonprecessing waveform. While this captures the leading effect of precession, it does not account for effects such as asymmetries between pairs of (ℓ, m) and $(\ell, -m)$ spin-weighted spherical harmonic

waveform modes [15]. Missing physics like this can lead to the systematic biases we see in Figs. 9 and 10. Among the phenomenological models, `IMRPhenomXPHM` is known to have a less accurate precession treatment than `IMRPhenomTPHM` [31], which could be responsible for `IMRPhenomXPHM` having the largest biases in our tests.

V. CONCLUSIONS

We propose that binary BH spins be measured at a reference point close to the merger, at either a dimensionless GW frequency $Mf_{\rm ref} = Mf_{\rm ISCO} = 6^{-3/2}/\pi$ or a dimensionless time $t_{\rm ref}/M = -100$ before the peak of the GW amplitude. We demonstrate that this leads to significant improvements in the measurement of orbital-plane spin orientations ϕ_1 and ϕ_2 for various events in the GWTC-2 catalog, while $\Delta\phi$ is not significantly impacted. However, using NR injections, we show that $\Delta\phi$ will also be better measured near the merger for louder signals expected in the future.

Using the same NR injections, we compare the performance of the waveform models `NRSur7dq4`, `IMRPhenomXPHM`, `IMRPhenomTPHM`, and `SEOBNRv4PHM`, in recovering ϕ_1, ϕ_2 and $\Delta\phi$ at $f_{\rm ref} = 20\,$Hz. As expected, `NRSur7dq4` provides the most accurate constraints for these angles, as this is the only model informed by precessing NR simulations. Among the other models, in general, we find that `IMRPhenomTPHM` and `SEOBNRv4PHM` perform better than `IMRPhenomXPHM`. However, even at moderate SNRs ($\sim 30-45$), we find examples where these models have biased estimates. This highlights the need to train waveform models on precessing NR simulations in order to reliably extract the full spin information from binary BH signals. `IMRPhenomXPHM`, `IMRPhenomTPHM`, and `SEOBNRv4PHM` do not currently allow specifying the spins at $t_{\rm ref}/M = -100$ or $Mf_{\rm ref} = 6^{-3/2}/\pi$, but we expect these biases will persist at those reference points.

In a companion paper, Ref. [23], we use the improved spin measurements obtained here to constrain the astrophysical distribution of the orbital-plane spin angles as well as merger kicks for the binary BH population. Notably, we find a preference for $\Delta\phi \sim \pm\pi$ in the population, which can be a signature of spin-orbit resonances [5].

ACKNOWLEDGMENTS.

We thank Rory Smith and Avi Vajpeyi for support with the `Parallel Bilby` [26] package, and Hector Estelles, Sascha Husa, Geraint Pratten, and Marta Colleoni for support with the phenomenological waveforms. We thank Sizheng Ma for sharing his Fisher matrix code. We thank Davide Gerosa and Katerina Chatziioannou for useful discussions. V.V. was supported by a Klarman Fellowship at Cornell. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 896869. S.B., M.I. and S.V. acknowledge support of the National Science Foundation and the LIGO Laboratory. S.B. is also supported by the NSF Graduate Research Fellowship under Grant No. DGE-1122374. M.I. is supported by NASA through the NASA Hubble Fellowship grant No. HST-HF2-51410.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. This research made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center [46], a service of the LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-1764464. Computations were performed on the Wheeler cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by Caltech; and the High Performance Cluster at Caltech.

Appendix A: Full spin posteriors for `NRSur7dq4`

For completeness, in Figs. 11 – 16 we show the full spin posteriors for the 31 events listed in Tab. I, generated using the `NRSur7dq4` model at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz. We show the spin magnitudes $\chi_{1,2}$, cosines of the tilt angles $\cos\theta_{1,2}$, and orbital-plane spin angles ϕ_1, ϕ_2, and $\Delta\phi$. The priors on all of these parameters are flat in their respective ranges (cf. Sec. II). The spin magnitude and tilt posteriors are consistent between the two reference points, and are consistent with Ref. [13].

In the following, we refer exclusively to the spin measurements at $t_{\rm ref}/M = -100$, and check whether measurements of the orbital-plane spin angles can be tied to a measurement of $\chi_{1,2}$ and $\cos\theta_{1,2}$. As noted in Sec. III A, an unambiguous measurement of the orbital-plane spin angles relies on being able to constrain $\chi_{1,2}$ away from zero, and $\cos\theta_{1,2}$ away from ±1. GW190521 (Fig. 13) is a good example of this: there is a clear preference for large $\chi_{1,2}$ and $\cos\theta_{1,2} \sim 0$, which likely enables the ϕ_1 and ϕ_2 measurement for this event. On the other hand, for GW190517_055101 (Fig. 13), there is a preference for large $\chi_{1,2}$, but with $\cos\theta_{1,2} \sim 1$. However, we still see peaks in the ϕ_1 and ϕ_2 distributions. Finally, for GW170818 (Fig. 11), there is a mild preference for small χ_1 and a similarly mild preference for large χ_2, but this is the event with the best ϕ_1 and ϕ_2 measurement (cf. Fig. 3). We conclude that current constraints on the spin magnitudes and tilts are too broad to look for such correlations with the orbital-plane spin angles: even for the cases where we see a preference for small $\chi_{1,2}$ and/or $\cos\theta_{1,2} \sim \pm1$, there is enough support for large $\chi_{1,2}$ and $\cos\theta_{1,2} \sim 0$, that there can be peaks in the posteriors of the orbital-plane spin angles.

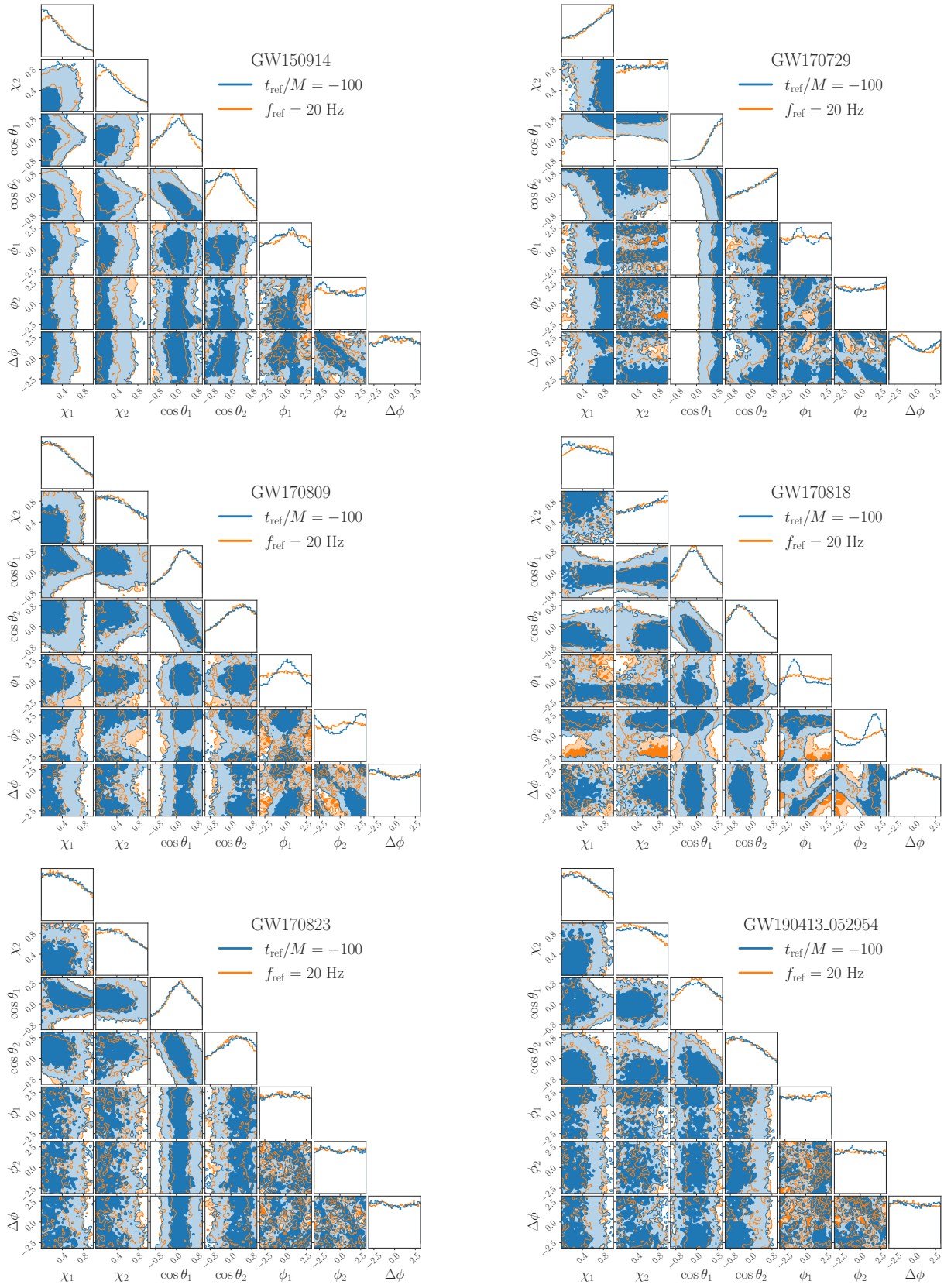

Figure 11. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz using `NRSur7dq4` for the events listed in Tab I. Set 1 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

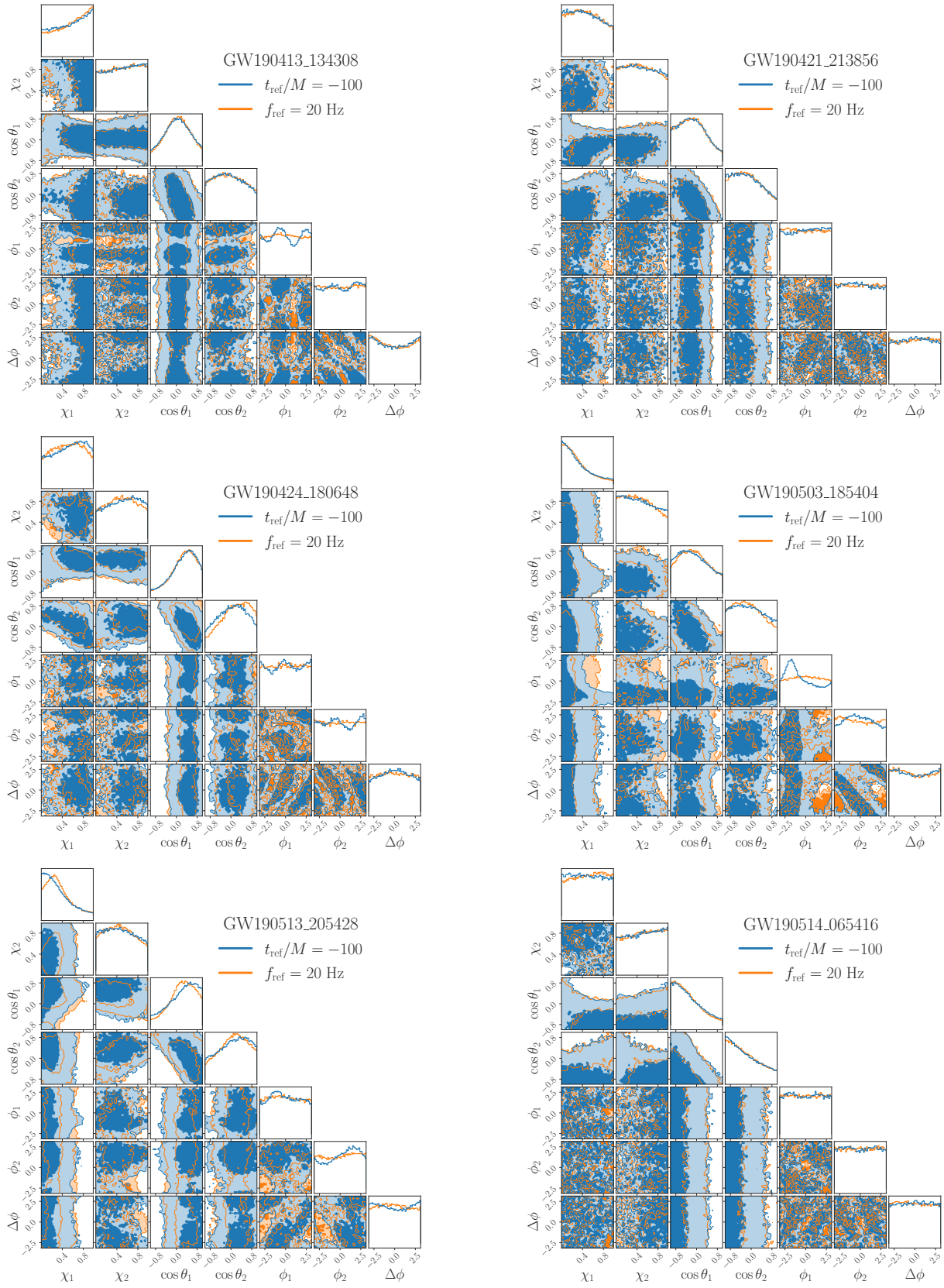

Figure 12. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz using `NRSur7dq4` for the events listed in Tab I. Set 2 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

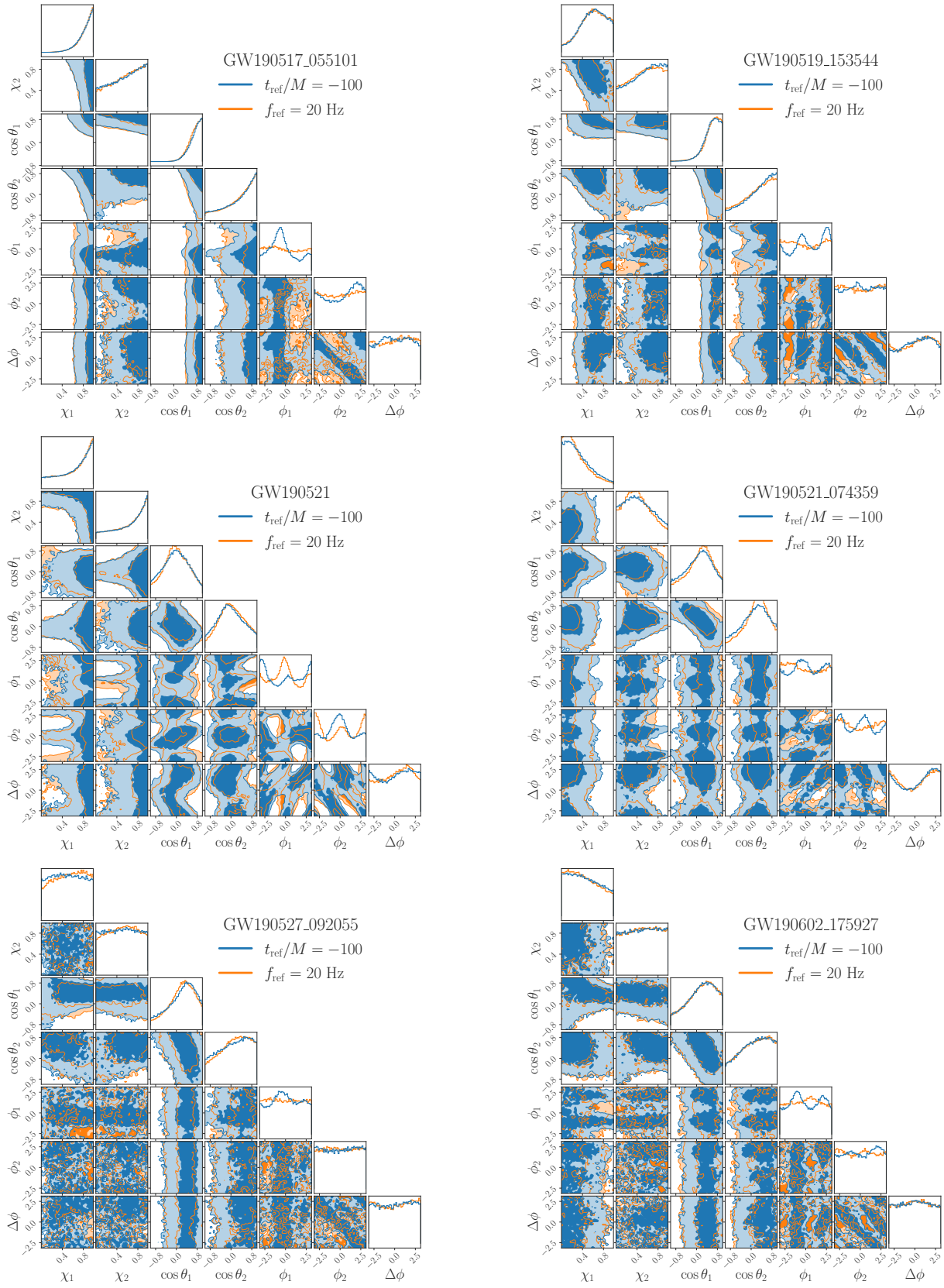

Figure 13. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20$ Hz using `NRSur7dq4` for the events listed in Tab I. Set 3 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

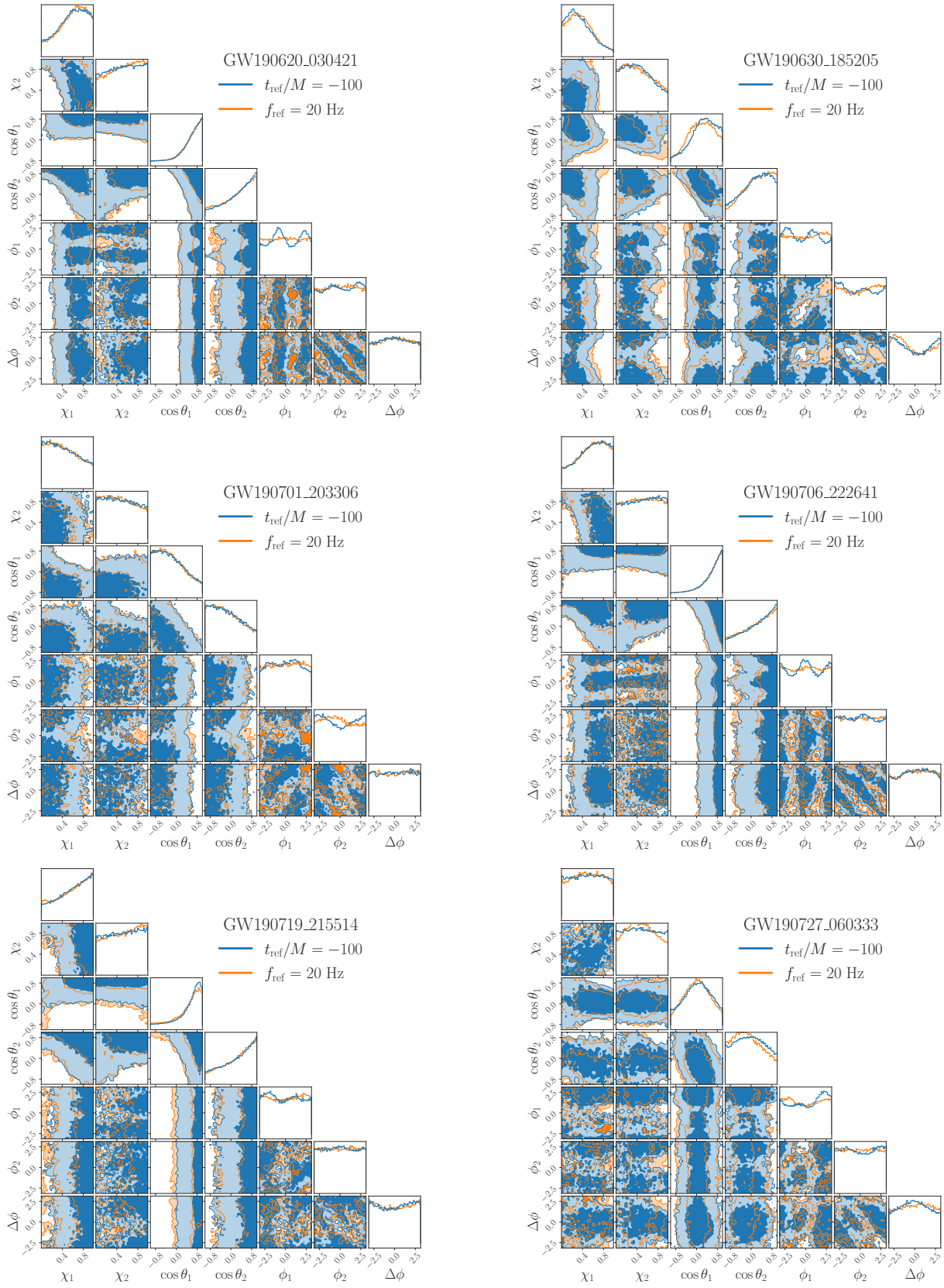

Figure 14. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz using `NRSur7dq4` for the events listed in Tab I. Set 4 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

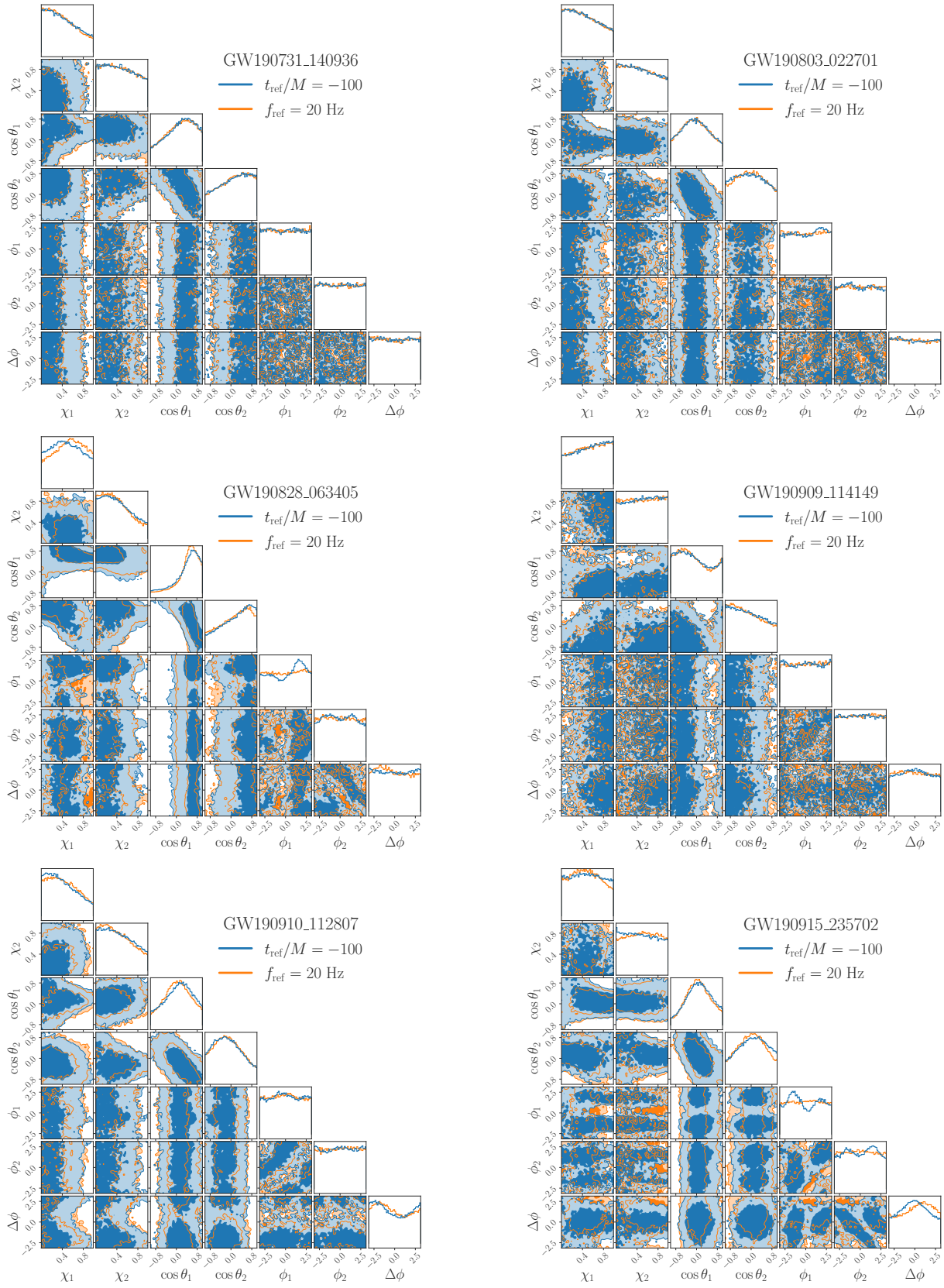

Figure 15. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz using `NRSur7dq4` for the events listed in Tab I. Set 5 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

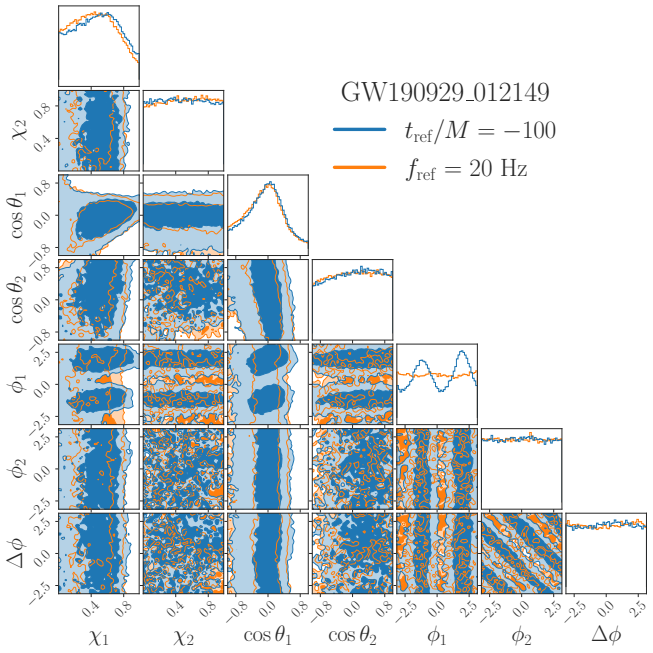

Figure 16. Full spin posteriors at $t_{\rm ref}/M = -100$ and $f_{\rm ref} = 20\,$Hz using `NRSur7dq4` for the events listed in Tab I. Set 6 out of 6. The lower-triangle subplots show central 90% and 50% credible regions of joint 2D posteriors, while the diagonal subplots show marginalized 1D posteriors.

Appendix B: Results for all GWTC-2 events using `IMRPhenomTPHM`

The results in Sec. III were restricted to the 31 GWTC-2 with $M \gtrsim 60M_\odot$ due to the length restrictions of `NRSur7dq4`. The remaining 15 events with $M \lesssim 60M_\odot$ are listed in Tab. II. For completeness, we now analyze all 46 GWTC-2 binary BH events using `IMRPhenomTPHM`. We choose `IMRPhenomTPHM` as this model performed better than `IMRPhenomXPHM` in Sec. IV. Once again, for simplicity, we only consider $f_{\rm ref} = 20\,$Hz for `IMRPhenomTPHM`. For the 15 events with $M \lesssim 60M_\odot$, we relax the prior constraints described in Sec. II to: $5 \leq \mathcal{M} \leq 400$, $q \leq 20$, and $10 \leq M \leq 400$.

GW events with $M \lesssim 60M_\odot$	
GW151012	GW151226
GW170104	GW170608
GW170814	GW190408_181802
GW190412	GW190512_180714
GW190707_093326	GW190708_232457
GW190720_000836	GW190728_064510
GW190828_065509	GW190924_021846
GW190930_133541	

Table II. The remaining 15 binary BH events from GWTC-2 that are not included in Tab. I.

Figure 17 shows $\Delta\phi$ posteriors for `IMRPhenomTPHM` at $f_{\rm ref} = 20\,$Hz for all 46 events. We show the corresponding `NRSur7dq4` posteriors for the applicable events. For most events, there are no strong peaks in $\Delta\phi$ for `IMRPhenomTPHM`, in agreement with Fig. 4. Interestingly, for GW190521, `IMRPhenomTPHM` has a clear peak at $\Delta\phi \sim 0$ which is absent for `NRSur7dq4`. This shows that waveform systematics are already important to consider for current GW events when measuring the orbital-plane spin angles. While further investigation is necessary to understand the nature of this peak, we note once again that `IMRPhenomTPHM` can have biases in 1D $\Delta\phi$ distributions as shown in the bottom-left panels of Fig. 9 and Fig. 10. In these cases, the `IMRPhenomTPHM` $\Delta\phi$ posterior is more sharply peaked (compared to `NRSur7dq4`) but is also biased.

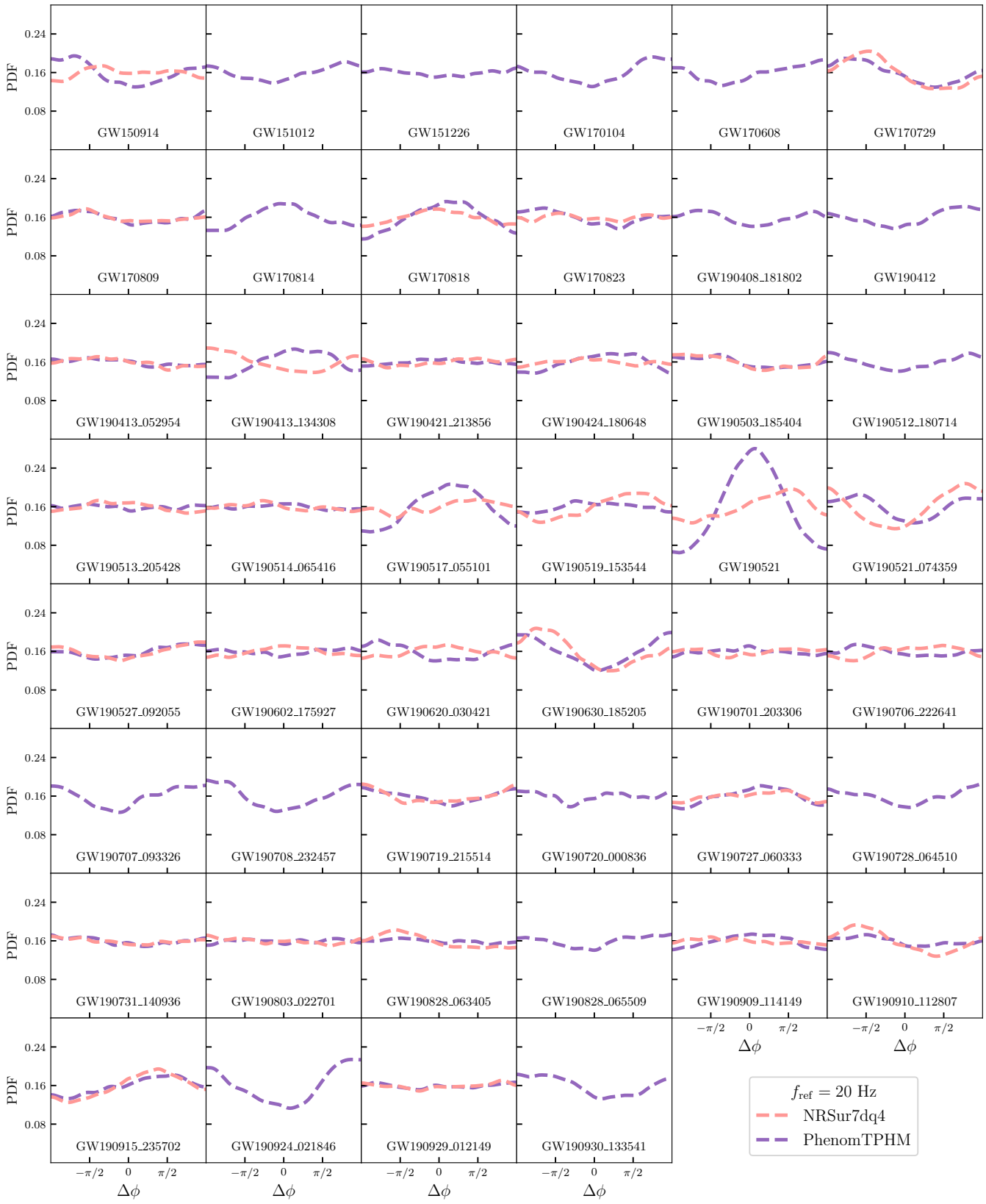

Figure 17. $\Delta\phi$ posteriors at $f_{\rm ref} = 20\,$Hz for all 46 GWTC-2 binary BH events, obtained using the `IMRPhenomTPHM` model. We also show the corresponding `NRSur7dq4` posterior for the 31 `NRSur7dq4` events for comparison.

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

[2] F. Acernese *et al.* (Virgo), "Advanced Virgo: a second-generation interferometric gravitational wave detector," Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[3] Theocharis A. Apostolatos, Curt Cutler, Gerald J. Sussman, and Kip S. Thorne, "Spin-induced orbital precession and its modulation of the gravitational waveforms from merging binaries," Phys. Rev. D **49**, 6274–6297 (1994).

[4] Lawrence E. Kidder, "Coalescing binary systems of compact objects to postNewtonian 5/2 order. 5. Spin effects," Phys. Rev. D **52**, 821–847 (1995), arXiv:gr-qc/9506022.

[5] Jeremy D. Schnittman, "Spin-orbit resonance and the evolution of compact binary systems," Phys. Rev. D **70**, 124020 (2004), arXiv:astro-ph/0409174.

[6] Vijay Varma, Maximiliano Isi, and Sylvia Biscoveanu, "Extracting the Gravitational Recoil from Black Hole Merger Signals," Phys. Rev. Lett. **124**, 101104 (2020), arXiv:2002.00296 [gr-qc].

[7] Salvatore Vitale, Ryan Lynch, John Veitch, Vivien Raymond, and Riccardo Sturani, "Measuring the spin of black holes in binary systems using gravitational waves," Phys. Rev. Lett. **112**, 251101 (2014), arXiv:1403.0129 [gr-qc].

[8] P. Schmidt, F. Ohme, and M. Hannam, "Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter," Phys. Rev. D **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].

[9] Sylvia Biscoveanu, Maximiliano Isi, Vijay Varma, and Salvatore Vitale, "Measuring the spins of heavy binary black holes," Phys. Rev. D **104**, 103018 (2021), arXiv:2106.06492 [gr-qc].

[10] Davide Gerosa, Richard O'Shaughnessy, Michael Kesden, Emanuele Berti, and Ulrich Sperhake, "Distinguishing black-hole spin-orbit resonances by their gravitational-wave signatures," Phys. Rev. D **89**, 124025 (2014), arXiv:1403.7147 [gr-qc].

[11] Daniele Trifirò, Richard O'Shaughnessy, Davide Gerosa, Emanuele Berti, Michael Kesden, Tyson Littenberg, and Ulrich Sperhake, "Distinguishing black-hole spin-orbit resonances by their gravitational wave signatures. II: Full parameter estimation," Phys. Rev. D **93**, 044071 (2016), arXiv:1507.05587 [gr-qc].

[12] Chaitanya Afle *et al.*, "Detection and characterization of spin-orbit resonances in the advanced gravitational wave detectors era," Phys. Rev. D **98**, 083014 (2018), arXiv:1803.07695 [gr-qc].

[13] R. Abbott *et al.* (LIGO Scientific, Virgo), "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].

[14] More precisely, we use the coorbital frame defined in Ref. [15]. In this frame, the z-axis is along the direction that maximises the power in the (2,2) mode, which is taken to be the direction of the orbital angular momentum [47]. The x-axis is along the line of separation from the lighter to the heavier BH, and the y-axis completes the right-handed triad. Note that this frame is defined using the gauge-invariant waveform at future null infinity, rather than the gauge-dependent BH trajectories.

[15] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer, "Surrogate models for precessing binary black hole simulations with unequal masses," Phys. Rev. Research. **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].

[16] S. A. Kaplan, "On Circular orbits in Einsteinian Gravitation theory," ZhETF Pisma Redaktsiiu **19**, 951–952 (1949).

[17] Alessandra Buonanno, Bala Iyer, Evan Ochsner, Yi 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].

[18] Benjamin Farr, Evan Ochsner, Will M. Farr, and Richard O'Shaughnessy, "A more effective coordinate system for parameter estimation of precessing compact binaries from gravitational waves," Phys. Rev. D **90**, 024018 (2014), arXiv:1404.7070 [gr-qc].

[19] B. P. Abbott *et al.* (LIGO Scientific, Virgo), "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. **X9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[20] Rich Abbott *et al.* (LIGO Scientific, Virgo), "Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo," SoftwareX **13**, 100658 (2021), arXiv:1912.11716 [gr-qc].

[21] LIGO Scientific Collaboration and Virgo Collaboration, "GWTC-1," https://doi.org/10.7935/82H3-HH23 (2018).

[22] LIGO Scientific Collaboration and Virgo Collaboration, "GWTC-2," https://doi.org/10.7935/99gf-ax93 (2020).

[23] Vijay Varma, Sylvia Biscoveanu, Maximiliano Isi, Will M. Farr, and Salvatore Vitale, "Hints of spin-orbit resonances in the binary black hole population," Phys. Rev. Lett. **128**, 031101 (2022), arXiv:2107.09693 [astro-ph.HE].

[24] Eric Thrane and Colm Talbot, "An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models," Publications of the Astronomical Society of Australia **36**, e010 (2019), arXiv:1809.02293 [astro-ph.IM].

[25] Marissa Walker, Vijay Varma, and Geoffrey Lovelace, "Extending numerical relativity surrogate models to near extremal spins," (2021), in preparation.

[26] Rory J.E. Smith, Gregory Ashton, Avi Vajpeyi, and Colm Talbot, "Massively parallel Bayesian inference for transient gravitational-wave astronomy," Mon. Not. Roy. Astron. Soc. **498**, 4492–4502 (2020), arXiv:1909.11873 [gr-qc].

[27] Joshua S. Speagle, "DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences," Monthly Notices of the Royal Astronomical Society **493**, 3132–3158 (2020), arXiv:1904.02180 [astro-ph.IM].

[28] I. M. Romero-Shaw *et al.*, "Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue," Mon. Not. Roy. Astron. Soc. **499**, 3295–3319 (2020), arXiv:2006.00714 [astro-ph.IM].

[29] Michael Boyle *et al.*, "The SXS Collaboration catalog of

binary black hole simulations," Class. Quant. Grav. **36**, 195006 (2019), arXiv:1904.04831 [gr-qc].

[30] Jonathan Blackman, Scott E. Field, Mark A. Scheel, Chad R. Galley, Christian D. Ott, Michael Boyle, Lawrence E. Kidder, Harald P. Pfeiffer, and Béla Szilágyi, "Numerical relativity waveform surrogate model for generically precessing binary black hole mergers," Phys. Rev. **D96**, 024058 (2017), arXiv:1705.07089 [gr-qc].

[31] Héctor Estellés, Marta Colleoni, Cecilio García-Quirós, Sascha Husa, David Keitel, Maite Mateu-Lucena, Maria de Lluc Planas, and Antoni Ramos-Buades, "New twists in compact binary waveform modelling: a fast time domain model for precession," (2021), arXiv:2105.05872 [gr-qc].

[32] R. Abbott *et al.* (LIGO Scientific, Virgo), "GW190521: A Binary Black Hole Merger with a Total Mass of 150 $M_\odot$," Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].

[33] Sylvia Biscoveanu, Maximiliano Isi, Salvatore Vitale, and Vijay Varma, "New Spin on LIGO-Virgo Binary Black Holes," Phys. Rev. Lett. **126**, 171103 (2021), arXiv:2007.09156 [astro-ph.HE].

[34] R. Abbott *et al.* (LIGO Scientific, Virgo), "Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog," Astrophys. J. Lett. **913**, L7 (2021), arXiv:2010.14533 [astro-ph.HE].

[35] Lee S. Finn, "Detection, measurement and gravitational radiation," Phys. Rev. D **46**, 5236–5249 (1992), arXiv:gr-qc/9209010.

[36] Curt Cutler and Eanna E. Flanagan, "Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?" Phys. Rev. D **49**, 2658–2697 (1994), arXiv:gr-qc/9402014.

[37] LIGO Scientific Collaboration, *Updated Advanced LIGO sensitivity design curve*, Tech. Rep. (2018) https://dcc.ligo.org/LIGO-T1800044/public.

[38] Sizheng Ma, Matthew Giesler, Vijay Varma, Mark A. Scheel, and Yanbei Chen, "Universal features of gravitational waves emitted by superkick binary black hole systems," Phys. Rev. D **104**, 084003 (2021), arXiv:2107.04890 [gr-qc].

[39] H. Cramér, *Mathematical Methods of Statistics*, Princeton Mathematical Series (Princeton University Press, 1999).

[40] C. Radhakrishna Rao, "Information and the accuracy attainable in the estimation of statistical parameters," Bull. Calcutta Math. Soc. **37**, 81–91 (1945).

[41] Michele Vallisneri, "Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects," Phys. Rev. D **77**, 042001 (2008), arXiv:gr-qc/0703086.

[42] SXS Collaboration, "The SXS collaboration catalog of gravitational waveforms," http://www.black-holes.org/waveforms.

[43] Abdul H. Mroue *et al.*, "Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy," Phys. Rev. Lett. **111**, 241104 (2013), arXiv:1304.6077 [gr-qc].

[44] Geraint Pratten *et al.*, "Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes," Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].

[45] Serguei Ossokine *et al.*, "Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation," Phys. Rev. D **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].

[46] LIGO Scientific Collaboration and Virgo Collaboration, "Gravitational Wave Open Science Center," https://www.gw-openscience.org.

[47] Michael Boyle, Robert Owen, and Harald P. Pfeiffer, "A geometric approach to the precession of compact binaries," Phys. Rev. **D84**, 124011 (2011), arXiv:1110.2965 [gr-qc].