Targeted numerical simulations of binary black holes for GW170104

J. Healy,[1] J. Lange,[1] R. O'Shaughnessy,[1] C. O. Lousto,[1] M. Campanelli,[1] A. R. Williamson,[1] Y. Zlochower,[1] J. Calderón Bustillo,[2] J. A. Clark,[2] C. Evans,[2] D. Ferguson,[2] S. Ghonge,[2] K. Jani,[2] B. Khamesra,[2] P. Laguna,[2] D. M. Shoemaker,[2] M. Boyle,[3] A. García,[4] D. A. Hemberger,[5] L. E. Kidder,[3] P. Kumar,[3,6] G. Lovelace,[4] H. P. Pfeiffer,[7,6] M. A. Scheel,[5] and S. A. Teukolsky[3,5]

[1]*Center for Computational Relativity and Gravitation, School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, New York 14623, USA*
[2]*Center for Relativistic Astrophysics and School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*
[3]*Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[4]*Gravitational Wave Physics and Astronomy Center, California State University Fullerton, Fullerton, California 92834, USA*
[5]*Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, California 91125, USA*
[6]*Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, Ontario M5S 3H8, Canada*
[7]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*

(Received 15 December 2017; published 23 March 2018)

In response to LIGO's observation of GW170104, we performed a series of full numerical simulations of binary black holes, each designed to replicate likely realizations of its dynamics and radiation. These simulations have been performed at multiple resolutions and with two independent techniques to solve Einstein's equations. For the nonprecessing and precessing simulations, we demonstrate the two techniques agree mode by mode, at a precision substantially in excess of statistical uncertainties in current LIGO's observations. Conversely, we demonstrate our full numerical solutions contain information which is not accurately captured with the approximate phenomenological models commonly used to infer compact binary parameters. To quantify the impact of these differences on parameter inference for GW170104 specifically, we compare the predictions of our simulations and these approximate models to LIGO's observations of GW170104.

DOI: 10.1103/PhysRevD.97.064027

I. INTRODUCTION

The LIGO-Virgo Collaboration (LVC) has already reported the confident discovery of five binary black hole (BBH) mergers via gravitational wave (GW) radiation: GW150914 [1] and GW151226 [2] from the first observing run O1 [3], and GW170104 [4], GW170608 [5], and GW170814 [6] from the second observing run. The parameters of these detections were first inferred by the use of nonprecessing (IMRPhenomD, SEOBNRv2, and SEOBNRv4), and approximate precessing models: IMRPhenomPv2 [7–9] and SEOBNRv3 [10–12].

A reanalysis of GW150914 (see [13] for the details of the simulations in Fig. 1 of [1]) implementing full numerical relativity (NR) simulations helped to better constrain the mass ratio of the system [14]. This is due to the fact that NR waveforms include physics omitted by current approximate models, notably higher order modes and accurate precession effects. A full description of this methodology, including detailed tests of systematic errors and parameter estimation improvements, can be found in [15].

This paper is organized as follows. In Sec. II, we describe the two independent techniques we use to solve Einstein's equations numerically for the evolution of binary black hole spacetimes. In Sec. III, we describe the binary's parameters selected for a detailed follow-up, our simulations of these proposed initial conditions, and detailed comparisons between our paired results, for both nonprecessing and precessing simulations. We also contrast our simulations' radiation with the corresponding results derived from the approximate phenomenological models used by LIGO for parameter inference. In Sec. IV, we directly compare our simulations to GW170104. These comparisons provide both a scalar measure of how well each simulation agrees with the data (a marginalized likelihood), as well as the best-fitting reconstructed waveform in each instrument [14,15]. Using our reconstructed waveforms, we graphically demonstrate that our simulations agree with each other and the data, with simulation differences far smaller than the residual noise in each instrument. Using these real observations as a benchmark

for model quality, we then quantify how effectively our simulations reproduce the data, compared to the results of approximate and phenomenological models at the same parameters. Since our simulations parameters were selected using these approximate and phenomenological models, we also have the opportunity to assess how effectively they identified the optimal binary parameters. In Sec. V we discuss the prospects for future targeted simulations in the follow-up of LIGO observations.

II. FULL NUMERICAL EVOLUTIONS

The breakthroughs [16–18] in numerical relativity allowed for detailed predictions for the gravitational waves from the late inspiral, plunge, merger, and ringdown of black hole binary systems. Catalogs of the simulated waveforms are publicly available [19–21] for its use for BBH parameter estimation [15], as well as for determining how the individual masses and spins of the orbiting binary relate to the properties of the final remnant black hole produced after merger. This relationship [22] can be used as a consistency check for the observations of the inspiral and, independently, the merger-ringdown signals as tests of general relativity [3,23,24]. And finally, those full numerically generated waveforms and remnant parameters are used in the fittings of the phenomenological approximate models cited above.

A. Simulations using finite-difference, moving-puncture methods

In order to make systematic studies and build a data bank of full numerical simulations, e.g., [21], it is crucial to develop efficient numerical algorithms, since large computational resources are required. The Rochester Institute of Technology (RIT) group evolved the BBH data sets described below using the LazEv [25] implementation of the moving-puncture approach [17,18] with the conformal function $W = \sqrt{\chi} = \exp(-2\Phi)$ suggested by Ref. [26]. For those runs, they used centered, sixth-order finite differencing in space [27] and a fourth-order Runge-Kutta time integrator (the code does not upwind the advection terms) and a fifth-order Kreiss-Oliger dissipation operator.

The LazEv code uses the EinsteinToolkit [28,29]/Cactus [30]/Carpet [31] infrastructure. The Carpet mesh refinement driver provides a "moving boxes" style of mesh refinement. In this approach, refined grids of fixed size are arranged about the coordinate centers of both holes. The Carpet code then moves these fine grids about the computational domain by following the trajectories of the two BHs.

To compute the initial low eccentricity orbital parameters, RIT used the post-Newtonian techniques described in [32] and then generated the initial data based on these parameters using approach [33] along with the TwoPunctures [34] code implementation.

The LazEv code uses AHFinderDirect [35] to locate apparent horizons, and compute the magnitude and components of the horizon spin using the *isolated horizon* (IH) algorithm detailed in Ref. [36] and as implemented in Ref. [37]. The horizon mass is calculated via the Christodoulou formula $m_H = \sqrt{m_{\rm irr}^2 + S_H^2/(4m_{\rm irr}^2)}$, where $m_{\rm irr} = \sqrt{A/(16\pi)}$, A is the surface area of the horizon, and S_H is the spin angular momentum of the BH (in units of M^2).

The radiated energy, linear momentum, and angular momentum, were measured in terms of the radiative Weyl scalar ψ_4, using the formulas provided in Refs. [38,39], Eqs. (22)–(24), and (27) respectively. However, rather than using the full ψ_4, it was decomposed it into spin weighted -2 spherical harmonic including all ℓ and m modes up to $\ell = 6$. The formulas in Refs. [38,39] are valid at $r = \infty$. To obtain the waveform and radiation quantities at infinity, the perturbative extrapolation described in Ref. [40] was used.

For the RIT simulations, different resolutions are denoted by NXXX where XXX is either 100, 118, or 140 for low, medium, and high resolutions, respectively. This number is directly related to the wave zone resolution in the simulation. For instance, N100 has a resolution of $M/1.0$ in the wave zone (where observer extraction takes place, preliminary to perturbative extrapolation to infinity via [40]), and N140 has $M/1.4$. In each case UID# 1–5, there are 10 levels of refinement in all and the grids followed a pattern close to those described in [41].

Other groups using the moving-puncture [17,18] formalism with finite-difference methods are Georgia Institute of Technology (GT) [20] and those based on bifunctional adaptive mesh (BAM) code [42]. The GT [20] simulations were obtained with the Maya code [43–50], which is also based on the Baumgarte-Shapiro-Shibata-Nakamura (BSSN) formulation with moving punctures. The grid structure for each run consisted of 10 levels of refinement provided by Carpet [31], a mesh refinement package for Cactus [30]. Each successive level's resolution decreased by a factor of 2. Sixth-order spatial finite differencing was used with the BSSN equations implemented with Kranc [51].

B. Simulations using pseudospectral, excision methods

Simulations labeled SXS are carried out using the Spectral Einstein Code (SpEC) [52] used by the Simulating eXtreme Spacetimes Collaboration (SXS). Given initial BBH parameters, a corresponding weighted superposition of two boosted, spinning Kerr-Schild black holes [53] is constructed, and then the constraints are solved [54–56] by a pseudospectral method to yield quasiequilibrium [53,57] initial data. Small adjustments in the initial orbital trajectory are made iteratively to produce initial data with low eccentricity [58–60].

The initial data are evolved using a first-order representation [61] of a generalized harmonic formulation

[62–64] of Einstein's equations, and using damped harmonic gauge [65–67]. The equations are solved pseudospectrally on an adaptively refined [68,69] spatial grid that extends from pure-outflow excision boundaries just inside apparent horizons [67,70–73] to an artificial outer boundary. Adaptive time stepping automatically achieves time steps of approximately the Courant limit.

On the Cal State Fullerton cluster, ORCA, the simulation achieved a typical evolution speed of $O(100M)$/day for the highest resolution (here we measure simulation time in units of M, the total mass of the binary). After the holes merge, all variables are automatically interpolated onto a new grid with a single excision boundary inside the common apparent horizon [70,71], and the evolution is continued. Constraint-preserving boundary conditions [61,74,75] are imposed on the outer boundary, and no boundary conditions are required or imposed on the excision boundaries.

We use a pseudospectral fast-flow algorithm [76] to find apparent horizons, and we compute spins on these apparent horizons using the approximate Killing vector formalism of Cook, Whiting, and Owen [77,78].

Gravitational wave extraction is done by three independent methods: direct extraction of the Newman-Penrose quantity ψ_4 at finite radius [58,70,79], extraction of the strain h by matching to solutions of the Regge-Wheeler-Zerilli-Moncrief equations at finite radius [80,81], and Cauchy-characteristic extraction [82–86]. The latter method directly provides gravitational waveforms at future null infinity, while for the former two methods the waveforms are computed at a series of finite radii and then extrapolated to infinity [87]. Differences between the different methods, and differences in extrapolation algorithms, can be used as error estimates on waveform extraction. These waveform extraction errors are important for the overall error budget of the simulations, and are typically on the order of, or slightly larger than, the numerical truncation error [88,89]. In this paper, the waveforms we compare use Regge-Wheeler-Zerilli-Moncrief extraction and extrapolation to infinity. We have verified that our choice of extrapolation order does not significantly affect our results. We have also checked that corrections to the wave modes [90] to account for a small drift in the coordinates of the center of mass have a negligible effect on our results; hence, we present here the uncorrected results.

III. SIMULATIONS OF GW170104

We extracted the maximum *a posteriori* (MaP) parameters from (preliminary) Bayesian posterior inferences performed by the LIGO Scientific Collaboration and the Virgo Collaboration, using different waveform models [91,92]. As described in Appendix B, this point parameter estimate is one of a few well-motivated and somewhat different choices for follow-up parameters; however, as described in that appendix, we estimate that the specific choice we adopt will not significantly change our principal results. Table I shows parameters simulated with numerical relativity. The first two simulations have been started at a reference frequency of 24 Hz (at the quoted total masses) in order to provide a fast response nonprecessing and precessing simulation to be ready to preliminarily compare with observations within two weeks (for the low resolution runs). The following three simulations have started from 20 Hz to cover the complete nominal low frequency sensitivity band of LIGO.

Spin Conventions: $(\chi_1^x, \chi_1^y, \chi_1^z)$ are specified in a frame where (i) $\hat{L} = (0, 0, 1)$, i.e., the Newtonian orbital angular momentum is along the z axis. (ii) the vector $\hat{n}$ pointing from m_2 to m_1 is the x axis, (1,0,0). Note that the orientation of $\hat{n}$ is essentially undetermined by parameter estimation (PE) methods, so the choice (ii) is meant to break this degeneracy to arrive at concrete parameters. In other words, the spin components given below are those consistent with Eq. (43) of Ref. [93].

The label, UID # 1–5, of the simulation identifies which parameters we are using in the follow-up as given by the initial data from Table I. All simulation initial data are chosen such that, when evolved, the simulation is consistent with the parameters of this table at a reference frequency $f_{\rm ref}$, which denotes the frequency of the (2,2) gravitational wave mode, or equivalently twice the orbital frequency. For aligned spin runs, where spin vectors are preserved, the initial orbital frequency may be smaller than $f_{\rm ref}/2$. For precessing runs, because we target a certain spin configuration at $f_{\rm ref}$, we begin our simulations very close to

TABLE I. Numerical simulations follow-up parameter table (as estimated by the quoted approximant). The two runs started at 24 Hz provided a fast response set of simulations while the following ones, starting at 20 Hz, cover the low frequency sensitivity band. We also report the gravitational wave cycles from those frequencies to merger in the simulations.

Run	$M_{\rm total}/M_\odot$	$f_{\rm ref}$ [Hz]	$q = m_1/m_2$	χ_1	χ_2	$\rm GW_{Cycles}$	Approximant
UID1	58.49	24	0.8514	(0, 0, 0.7343)	(0, 0, −0.8278)	31.1	SEOBNRv4
UID2	58.72	24	0.5246	(0.1607, −0.1023, −0.0529)	(−0.3623, 0.5679, −0.3474)	17.1	SEOBNRv3
UID3	62.13	20	0.4850	(0.0835, −0.4013, −0.3036)	(−0.3813, 0.7479, −0.1021)	24.9	IMRPhenomPv2
UID4	53.46	20	0.7147	(0, 0, 0.2205)	(0, 0, −0.7110)	28.2	SEOBNRv4
UID5	59.11	20	0.4300	(0, 0, −0.3634)	(0, 0, −0.1256)	27.3	IMRPhenomD

TABLE II. Initial data parameters for the quasicircular configurations with a smaller mass black hole (labeled 1), and a larger mass spinning black hole (labeled 2). The punctures are located at $\vec{r}_1 = (x_1, 0, 0)$ and $\vec{r}_2 = (x_2, 0, 0)$, with momenta $P = \pm(P_r, P_t, 0)$, puncture mass parameters m^p/M, horizon (Christodoulou) masses m^H/M, total ADM mass $M_{\rm ADM}$, dimensionless spins $a/m_H = S/m_H^2$, and eccentricity, e.

Run	x_1/M	x_2/M	P_r/M	P_t/M	m_1^p/M	m_2^p/M	m_1^H/M	m_2^H/M	$M_{\rm ADM}/M$	a_1/m_1^H	a_2/m_2^H	e
UID1	−7.9168	6.7407	−2.829e-4	0.074 67	0.3196	0.3056	0.4599	0.5401	0.9928	0.7343	−0.8267	6e-4
UID2	−7.8211	4.1029	−4.837e-4	0.078 37	0.3277	0.4400	0.3441	0.6559	0.9922	0.1977	0.7580	1e-3
UID3	−8.7720	4.2543	−3.316e-4	0.071 60	0.2796	0.3584	0.3266	0.6734	0.9930	0.5101	0.8445	1e-3
UID4	−8.4742	6.0567	−2.918e-4	0.074 21	0.3991	0.4219	0.4168	0.5832	0.9928	0.2205	−0.7110	2e-4
UID5	−9.4395	4.0593	−2.718e-4	0.066 98	0.2753	0.6865	0.3007	0.6993	0.9933	−0.3634	−0.1256	5e-4

this reference frequency. D/M is the initial orbital separation of the NR run in geometric units and the mass ratio and intrinsic spins ($\vec{\chi}_i = \vec{S}_i/m_i^2$) are denoted by $(q, \chi_1^x, \chi_1^y, \ldots,)$. Because of the way NR simulations are set up, the initial parameters can change due to the presence of junk radiation and/or imperfections in setting up initial data; therefore, these quantities should be reported as after-junk masses/spins, at f^r, the relaxed frequency, ideally extracted at the given reference frequency, $r_{\rm ref}$. For precessing runs, in particular, the spin components should be specified at the reference frequency, following the convention $\chi_i^x = \vec{\chi}_i \cdot \hat{n}, \chi_i^z = \vec{\chi}_i \cdot \hat{L}$, with $\hat{n}$ and $\hat{L}$ computed at the reference frequency, too. We also provide e, the orbital eccentricity. For instance, the actual initial data as measured for the RIT's follow-up simulations are described in Table II using the method of Ref. [94] where $Me = r^2\ddot{r}$.

For follow-up UID# 1, the initial spurious burst contains a non-negligible kick which causes a center of mass drift of approximately 0.65 M over the 4000 M of evolution. Because of this, information from the dominant $\ell = 2$, $m = \pm 2$ modes leak into the other modes, particularly the $m =$ odd modes. To reduce this effect, we can recalculate the modes by finding the *average* rest frame of the binary. We calculate the average velocity of the center of mass of the binary (from ψ_4) over the inspiral and then boost the waveform in the opposite direction. This is done using Eqs. (4)–(5) in [95] and Eqs. (7)–(8) in [96]. Note that this does not change the physical waveforms, only how they are spread over modes.

For the RIT simulations, the initial data parameters in Table II for the nonprecessing systems 1, 4, and 5 were determined by choosing the starting frequency just below the reference frequency. This gives the gauge time to settle, and since the spins do not change, this gives us a cleaner waveform once we hit the reference frequency. For the precessing simulations 2 and 3, since the spins will now evolve, we determine the initial data parameters by choosing the initial spins at the specified reference frequency. The initial data used by SXS, being not conformally flat, have less spurious radiation content than the Bowen-York data and hence produce a different set of masses and spins after settling down. See Table III for the specific values of each simulation by the two kinds of initial data families. This process can be iterated to get closer initial parameters for each approach, although it requires some extra evolution time and coordination to reach a fractional agreement below 10^{-3}. This process has been followed for UID # 1, but not for the other cases, in particular, the two precessing ones UID# 2 and UID# 3, and hence the differences, for instance, displayed in Fig. 1 for the precessing case UID# 3. Notwithstanding, we observe a fairly better agreement between the two full numerical approaches than with the SEOBNRv3 model.

The SXS simulations used in this work have been assigned SXS catalog numbers BBH:0626 (UID1), BBH:0627 (UID2), BBH:0628 (UID3), BBH:0625 (UID4), and BBH:0631 (UID5).

Each simulation has an asymptotic frame relative to which we extract $rh_{lm}(t)$. In all cases used here, this axis corresponds to the $\hat{z}$ $(= \hat{L})$ axis of the simulation frame. For all simulations, this axis also agrees with the orbital angular momentum axis $\hat{L}$ at the start of the evolution.

A. Outgoing radiation very similar for different NR methods

Following previous (targeted to GW150914) studies [13], we compare the outgoing radiation mode by mode, using an observationally driven measure: the overlap or *match*. The black and grey lines in Figs. 2 and 3 show the match between the two simulations' (RIT-SXS and RIT-GT, respectively) (2,2) modes, as a function of the minimum frequency used in the match. In this calculation, we use a detector noise power spectrum appropriate to GW170104, and a total mass $M_{\odot=}$ as given in Table I. By increasing the minimum frequency, we increasingly omit the earliest times in the signal, first eliminating transient startup effects associated due to finite duration and eventually comparing principally the merger signals from the two black holes. For comparison, the red, blue, and yellow lines show the corresponding matches between RIT, SXS, and GT simulations, respectively, and effective one body models with identical parameters (faithfulness study). In Fig. 2, which illustrates only nonprecessing simulations, these comparisons are made to the

TABLE III. Values of the individual masses, mass ratio, dimensionless spins, and frequency, given at a time after the gauge settles. Quantities are labeled by the superscript "r" (for relaxed). For the nonprecessing cases (1, 4, and 5), this time is $t^r/M = 200$ for RIT and $t^r/M = 640$ for SXS. For the precessing systems (2 and 3), the values are given such that the relaxed frequency Mf^r is the same between RIT and SXS.

UID#	m_1^r/M	m_2^r/M	q^r	χ_1^r	χ_2^r	Mf^r
UID1-N100	0.459 757	0.539 780	0.851 718	(0,0,0.737 493)	(0, 0, −0.829 064)	0.002 586
UID1-N118	0.459 758	0.539 801	0.851 718	(0,0,0.737 473)	(0, 0, −0.829 040)	0.002 587
UID1-N140	0.459 758	0.539 801	0.851 718	(0,0,0.737 464)	(0, 0, −0.829 030)	0.002 587
UID1-L2	0.459 913	0.540 183	0.851 401	(0,0,0.734 254)	(0, 0, −0.827 442)	0.003 135
UID1-L3	0.459 902	0.540 171	0.851 400	(0,0,0.734 138)	(0, 0, −0.827 657)	0.003 139
UID1-L4	0.459 907	0.540 176	0.851 402	(0,0,0.734 106)	(0, 0, −0.827 690)	0.003 139
UID2-N100	0.344 090	0.655 693	0.524 773	(0.119 494, 0.144 676, −0.063 539)	(−0.675 724, 0.181 910, −0.293 427)	0.004 010
UID2-N118	0.344 091	0.655 693	0.524 774	(0.118 992, 0.144 981, −0.063 537)	(−0.676 089, 0.180 798, −0.293 141)	0.004 010
UID2-N140	0.344 091	0.655 694	0.524 774	(0.118 754, 0.145 124, −0.063 541)	(−0.676 262, 0.180 273, −0.293 005)	0.004 010
UID2-L1	0.344 082	0.655 876	0.524 614	(0.101 171, 0.156 396, −0.066 135)	(−0.693 687, 0.096 203, −0.290 389)	0.004 001
UID2-L2	0.344 069	0.655 966	0.524 523	(0.100 390, 0.156 863, −0.066 508)	(−0.693 803, 0.095 200, −0.289 578)	0.004 013
UID2-L3	0.344 072	0.655 958	0.524 534	(0.100 596, 0.156 736, −0.066 591)	(−0.693 727, 0.095 442, −0.289 439)	0.004 010
UID3-N100	0.326 605	0.672 963	0.485 325	(0.360 515, −0.132 028, −0.338 700)	(−0.685 165, 0.491 281, −0.068 777)	0.003 308
UID3-N118	0.326 606	0.672 963	0.485 326	(0.360 458, −0.131 982, −0.338 637)	(−0.685 154, 0.491 237, −0.068 765)	0.003 308
UID3-N140	0.326 607	0.672 964	0.485 326	(0.360 429, −0.132 025, −0.338 591)	(−0.685 106, 0.491 276, −0.068 763)	0.003 308
UID3-L1	0.326 576	0.673 382	0.484 978	(0.358 221, −0.095 241, −0.350 438)	(−0.749 729, 0.387 741, −0.052 141)	0.003 330
UID3-L2	0.326 598	0.673 451	0.484 962	(0.358 434, −0.094 398, −0.350 478)	(−0.750 204, 0.386 567, −0.051 977)	0.003 339
UID3-L3	0.326 583	0.673 462	0.484 931	(0.352 293, −0.114 550, −0.350 800)	(−0.727 233, 0.427 282, −0.059 549)	0.003 308
UID4-N100	0.416 817	0.583 036	0.714 908	(0,0,0.221 608)	(0, 0, −0.712 189)	0.002 629
UID4-N118	0.416 819	0.583 037	0.714 911	(0,0,0.221 602)	(0, 0, −0.712 184)	0.002 630
UID4-N140	0.416 820	0.583 037	0.714 912	(0,0,0.221 600)	(0, 0, −0.712 182)	0.002 630
UID4-L1	0.416 790	0.583 195	0.714 666	(0,0,0.220 660)	(0, 0, −0.710 930)	0.002 635
UID4-L2	0.416 809	0.583 214	0.714 675	(0,0,0.220 453)	(0, 0, −0.710 967)	0.002 635
UID4-L3	0.416 817	0.583 203	0.714 703	(0,0,0.220 427)	(0, 0, −0.710 930)	0.002 619
UID5-N100	0.300 721	0.699 282	0.430 043	(0, 0, −0.366 067)	(0, 0, −0.125 360)	0.002 932
UID5-N118	0.300 722	0.699 284	0.430 043	(0, 0, −0.366 031)	(0, 0, −0.125 352)	0.002 932
UID5-N140	0.300 723	0.699 285	0.430 043	(0, 0, −0.366 015)	(0, 0, −0.125 348)	0.002 932
UID5-L1	0.300 697	0.699 261	0.430 021	(0, 0, −0.363 371)	(0, 0, −0.125 641)	0.002 900
UID5-L2	0.300 708	0.699 223	0.430 061	(0, 0, −0.363 424)	(0, 0, −0.125 550)	0.002 864
UID5-L3	0.300 722	0.699 178	0.430 107	(0, 0, −0.363 345)	(0, 0, −0.125 616)	0.002 855

nonprecessing model SEOBNRv4 [97]. In Fig. 3, which targets the two precessing UIDs, we instead compare to SEOBNRv3, which approximates some precession effects. For both nonprecessing and precessing simulations, these figures show that the different NR groups' simulations produce similar radiation, with mismatches $\lesssim 10^{-3}$ even at the longest durations considered. By contrast, comparisons with SEOBNRv4 and SEOBNRv3 show that these models do not replicate our simulations' results, particularly for precessing binaries.

To demonstrate good agreement beyond the (2,2) mode for precessing simulations, for multiple resolutions, Tables IV and V systematically compare all modes between RIT and SXS. The match calculations in this table are performed using a strain noise power spectral densities (PSD) characterizing data near GW170104. Following [13], one phase and time shift is computed by maximizing the overlap of the (2,2) mode; this phase and time shift is then applied to all other modes without any further maximization. Table IV shows a resolution test: the match between RIT and SXS simulations, as a function of RIT simulation resolution. As the most challenging precessing case, UID3 is shown by default. Except for the $m = 0$ modes, all the simulations show good agreement mode by mode, for all resolutions.

Based on Figs. 2, 3, and Table IV, we anticipate the lowest production-quality NR resolution (N100 for RIT; L3 for SXS) will usually be sufficient to go well beyond the accuracy of approximate and phenomenological models. To elaborate on this hypothesis, Table V shows the mode-by-mode overlaps between these two lowest NR resolutions. The $\ell = 2$ modes agree without exception. Good agreement also exists for the most significant modes up to $\ell \leq 4$. On the other hand, the last columns of Table IV show that the rough agreement between NR and models for the modes (2,2) displayed in Fig. 3, notably worsens when looking at ones other than the leading modes.

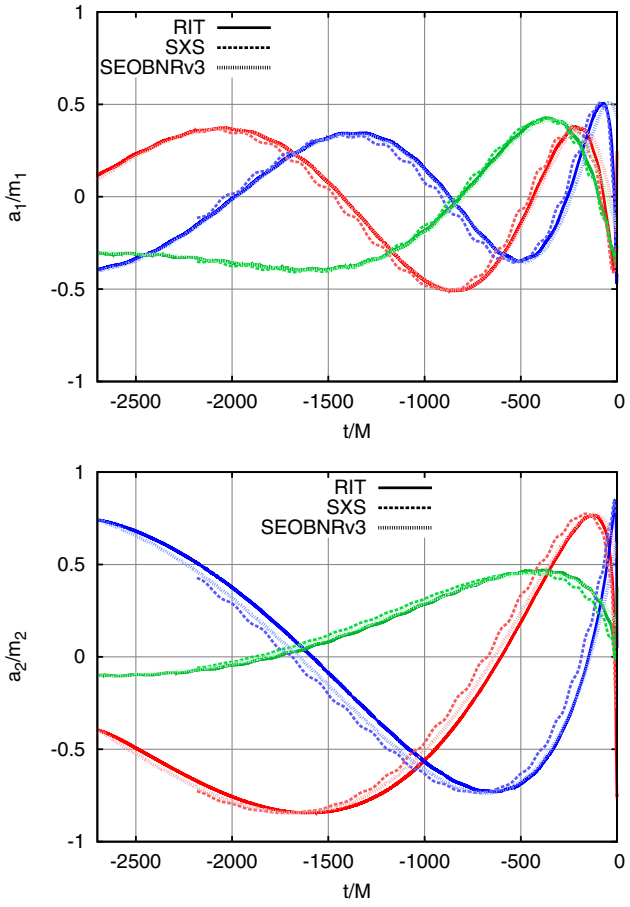

FIG. 1. The small BH (top) and large BH (bottom) spins for follow-up case UID# 3. RIT's simulation has solid lines; SXS's has dashed lines; and spin evolution as predicted by SEOBNRv3 is shown with a wide-dashed line. The spin components, x, y, and z, are red, blue, and green, respectively. A closer agreement between the two full numerical approaches is observed.

In Table VI, we also provide a comparison of the remnant properties, i.e., final mass, spin, and recoil velocity of the final, merged, black hole, as computed by the two NR methods and for a set of three increasing resolutions. We observe good agreement and convergence of their values. In the case of the RIT runs, a nearly fourth-order convergence with a resolution for the recoil velocity of the remnant is displayed. The same convergence properties are shown for the final mass and spin, despite these quantities being over-resolved at these resolutions. Those remnant quantities are also important to model fitting formulas [98,99] to be used to infer the final black hole properties from the binary parameters and thus serve as a test of the general theory of relativity, as in [23]. The phenomenological approximate waveform models [97,100,101] also benefit from information of the final remnant properties as one of their inputs and can thus produce a more accurate precessing and include a higher modes model.

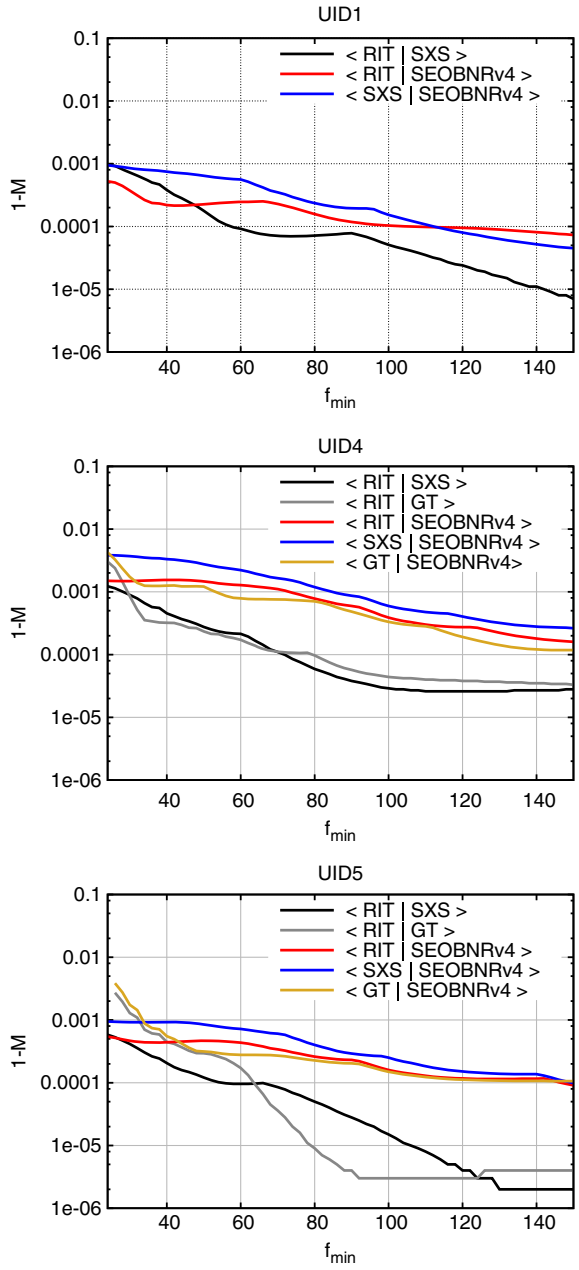

FIG. 2. For the three nonprecessing UIDs# 1,4,5 in Table II, matches between SXS, RIT, and SEOBNRv4 (2,2) modes as a function of $f_{\rm min}$, using the H1 PSD characterizing data near GW170104. We also compare with GT runs for UIDs# 4,5. Compare also to similar plots for GW150914 [13].

In conclusion we see that the typical production NR simulations are well into the convergence regime and produce accurate enough waveforms for all practical applications of the current generation of gravitational wave observations. This includes different NR approaches, modes, and remnant properties. The distinction with the current models is also clear and those still show signs of systematic errors with respect to the most accurate NR waveforms.

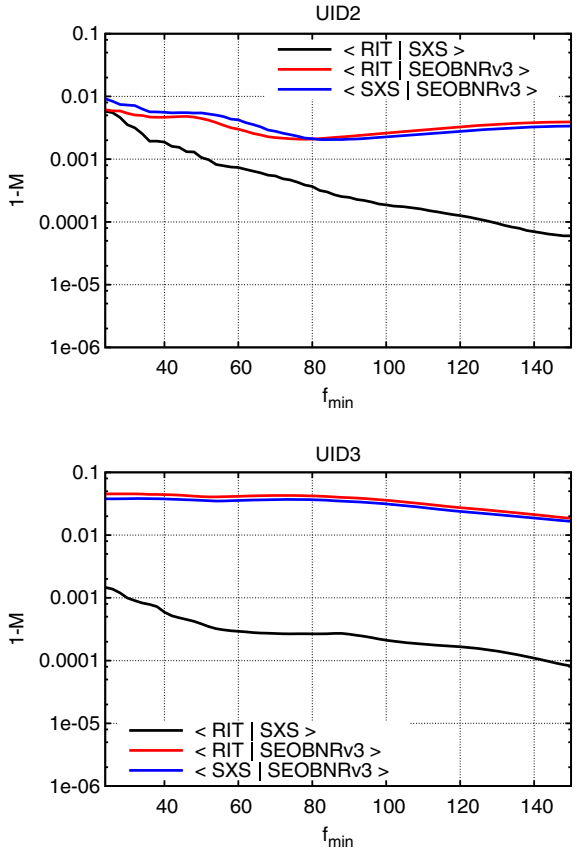

FIG. 3. For the two precessing UIDs# 2,3 in Table II, matches between SXS, RIT, and SEOBNRv3 (2,2) modes as a function of $f_{\rm min}$ as a function of $f_{\rm min}$, using the H1 PSD characterizing data near GW170104. In this comparison, the (2,2) mode of all three simulations and SEOBNRv3 are extracted relative to the $\hat{L}$ axis, identified from their common initial orbital parameters. While these frame identifications are coordinate dependent for precessing binaries—implying our comparisons here could include both intrinsic disagreement and systematic error due to (say) overall misalignment—the good agreement shown in Fig. 1 for the equally coordinate-dependent spins suggests that convention-dependent sources contribute little to the mismatches illustrated here.

IV. COMPARING NR SIMULATIONS WITH OBSERVATIONS OF GW170104

The comparisons above demonstrate that our simulations agree with one another, but differ from approximate and phenomenological models often used to describe precessing mergers. Fortunately, nature has provided us with a natural benchmark with which to assess the efficacy of our calculations and the significance of any discrepancies: observations of BH-BH mergers. We use standard techniques [14,15] to directly compare GW170104 to our simulations. For context, we also compare these observations to the corresponding predictions of approximate and phenomenological models that purport to describe the same event.

TABLE IV. Match between individual spherical harmonic modes (ℓ, m) of the SXS and RIT UID3 waveforms, using the H1 PSD characterizing data near GW170104. Following [13], rather than maximize over time and phase for each independently, our mode-by-mode comparisons fix the event time and overall phase using one mode [here, the (2,2) mode]. The successively higher resolution simulations from RIT, labeled as $N100, N118, N140$ are compared to the L3 (highest) resolution run from SXS. The minimal frequency is taken as $f_{\rm min} = 30m$ Hz for $m \geq 2$ and $f_{\rm min} = 30$ Hz for $m = 0, 1$ for a fiducial total mass of $M = 58.73\ M_\odot$. The column labeled $\mathcal{M}_{v3}$ shows the match between RIT N140 and the corresponding SEOBNRv3 mode. Rows with a "$\cdots$" are not modeled by SEOBNRv3. The column labeled $\mathcal{O}_{N140}$ shows the overlap of N140 with itself, with a minimum frequency of 30 Hz in all cases, to indicate the significance of the mode.

ℓ	m	$N100$	$N118$	$N140$	$\mathcal{M}_{v3}$	$\mathcal{O}_{N140}$
2	−2	0.9989	0.9990	0.9990	0.9347	244.54
2	−1	0.9965	0.9972	0.9968	0.6257	96.12
2	0	0.9972	0.9973	0.9966	0.3091	56.06
2	1	0.9982	0.9983	0.9983	0.5797	102.66
2	2	0.9986	0.9986	0.9986	0.9600	215.48
3	−3	0.9901	0.9902	0.9912	⋯	29.63
3	−2	0.9887	0.9913	0.9902	⋯	17.14
3	−1	0.9785	0.9811	0.9801	⋯	8.98
3	0	0.9803	0.9814	0.9834	⋯	5.57
3	1	0.9848	0.9845	0.9847	⋯	9.17
3	2	0.9867	0.9864	0.9862	⋯	17.01
3	3	0.9899	0.9896	0.9901	⋯	28.87
4	−4	0.9921	0.9927	0.9938	⋯	11.99
4	−3	0.9800	0.9798	0.9814	⋯	6.61
4	−2	0.9830	0.9851	0.9838	⋯	4.17
4	−1	0.9856	0.9871	0.9868	⋯	2.30
4	0	0.9317	0.9341	0.9377	⋯	1.55
4	1	0.9854	0.9862	0.9861	⋯	2.32
4	2	0.9825	0.9845	0.9836	⋯	4.26
4	3	0.9827	0.9825	0.9835	⋯	6.93
4	4	0.9906	0.9911	0.9919	⋯	10.13
5	−5	0.9703	0.9819	0.9848	⋯	3.19
5	−4	0.9646	0.9681	0.9735	⋯	1.72
5	−3	0.9641	0.9674	0.9708	⋯	1.09
5	−2	0.9575	0.9743	0.9765	⋯	0.66
5	−1	0.9657	0.9722	0.9734	⋯	0.36
5	0	0.8730	0.8897	0.9013	⋯	0.25
5	1	0.9636	0.9695	0.9710	⋯	0.37
5	2	0.9541	0.9728	0.9765	⋯	0.67
5	3	0.9688	0.9718	0.9738	⋯	1.13
5	4	0.9643	0.9692	0.9725	⋯	1.71
5	5	0.9657	0.9796	0.9825	⋯	2.73

Figure 4 displays the direct comparison of the nonprecessing simulations by RIT and SXS complementary approaches for the configurations UID# 1, UID# 4, and UID# 5, as given in Table II. They directly compare to the signals as observed by LIGO H1 and L1 and with each other. The lower panel shows the residuals of the signals with

TABLE V. Matches ($\mathcal{M}_i$) between individual spherical harmonic modes (ℓ, m) of the SXS and RIT waveforms, using the H1 PSD characterizing data near GW170104. The lowest resolution simulations from RIT, labeled N100, are compared to the L3 resolution run from SXS. The minimal frequency is taken as $f_{\rm min} = 30m$ Hz for $m \geq 2$ and $f_{\rm min} = 30$ Hz for $m = 0, 1$. The columns labeled $\mathcal{O}_i$ show the overlap of N100 with itself, $\langle h^{N100}_{\ell m} | h^{N100}_{\ell m} \rangle$, to indicate the significance of the mode.

ℓ	m	$\mathcal{M}_1$	$\mathcal{O}_1$	$\mathcal{M}_2$	$\mathcal{O}_2$	$\mathcal{M}_3$	$\mathcal{O}_3$	$\mathcal{M}_4$	$\mathcal{O}_4$	$\mathcal{M}_5$	$\mathcal{O}_5$
2	−2	0.9993	282.02	0.9940	220.13	0.9993	259.04	0.9991	241.44	0.9996	237.10
2	−1	0.9985	28.16	0.9933	127.06	0.9973	96.60	0.9859	24.19	0.9991	19.24
2	0	0.9294	5.81	0.9145	72.81	0.9975	54.88	0.9795	6.72	0.9816	5.14
2	1	0.9986	28.13	0.9946	130.33	0.9985	103.84	0.9859	24.19	0.9991	19.24
2	2	0.9993	282.02	0.9965	201.91	0.9990	226.98	0.9991	241.44	0.9996	237.10
3	−3	0.9463	6.49	0.9870	22.17	0.9904	31.88	0.9363	15.23	0.9978	37.33
3	−2	0.9993	7.80	0.9827	17.40	0.9915	17.89	0.9947	6.06	0.9986	6.15
3	−1	0.8844	1.17	0.9850	9.31	0.9788	9.30	0.7229	1.95	0.9099	0.73
3	0	0.7757	0.78	0.8437	5.46	0.9792	5.73	0.8199	0.69	0.8720	0.28
3	1	0.8879	1.21	0.9766	9.99	0.9838	9.51	0.7228	1.95	0.9099	0.73
3	2	0.9993	7.80	0.9930	18.29	0.9861	17.81	0.9947	6.06	0.9986	6.15
3	3	0.9463	6.49	0.9843	22.66	0.9898	30.67	0.9363	15.23	0.9978	37.33
4	−4	0.9922	12.17	0.9810	9.20	0.9926	12.97	0.9908	10.32	0.9956	13.75
4	−3	0.9930	2.24	0.9824	7.66	0.9794	7.11	0.9869	1.68	0.9967	1.94
4	−2	0.5968	0.69	0.9874	4.90	0.9841	4.37	0.9150	0.53	0.9068	0.49
4	−1	0.6361	0.14	0.9743	2.51	0.9857	2.39	0.7761	0.15	0.6647	0.10
4	0	0.3514	0.49	0.6951	1.45	0.9355	1.60	0.2246	0.52	0.2578	0.28
4	1	0.6533	0.13	0.9606	2.43	0.9848	2.45	0.7761	0.15	0.6647	0.10
4	2	0.5967	0.69	0.9887	4.89	0.9830	4.51	0.9150	0.53	0.9068	0.49
4	3	0.9930	2.24	0.9851	7.75	0.9830	7.37	0.9869	1.68	0.9967	1.94
4	4	0.9922	12.17	0.9814	8.28	0.9912	10.78	0.9908	10.32	0.9956	13.75
5	−5	0.8662	0.64	0.9614	1.85	0.9709	3.29	0.9028	1.38	0.9807	3.83
5	−4	0.9758	0.58	0.9666	1.81	0.9648	1.85	0.9615	0.44	0.9893	0.61
5	−3	0.7287	0.12	0.9775	1.26	0.9631	1.16	0.7838	0.13	0.8748	0.19
5	−2	0.4258	0.22	0.9592	0.74	0.9552	0.71	0.6812	0.12	0.7639	0.13
5	−1	0.1107	0.07	0.9494	0.37	0.9658	0.38	0.6209	0.03	0.4149	0.04
5	0	0.3735	0.10	0.5100	0.25	0.8761	0.27	0.3827	0.08	0.2603	0.06
5	1	0.0553	0.04	0.9128	0.34	0.9639	0.40	0.6208	0.03	0.4152	0.04
5	2	0.4258	0.22	0.9711	0.72	0.9509	0.74	0.6812	0.12	0.7639	0.13
5	3	0.7286	0.12	0.9804	1.30	0.9683	1.22	0.7838	0.13	0.8748	0.19
5	4	0.9758	0.58	0.9764	1.87	0.9661	1.86	0.9615	0.44	0.9893	0.61
5	5	0.8662	0.64	0.9650	1.77	0.9663	2.76	0.9028	1.38	0.9807	3.83

respect to the RIT N118 simulation and compares it with the direct difference of the two approaches and also with the difference of the N118 and N100 resolutions, which measure the finite-difference error of the N118, given the observed near fourth-order convergence seen when included in the N140 run into the analysis. For all three cases we note that the differences in any of the simulations are much smaller than the residuals and hence typical noise of the observations. This shows that the fast response runs performed to simulate BBH (low-medium resolution) are in an acceptable good agreement with the expected higher resolution ones at the required level of errors.

Figure 5 displays the two precessing targeted simulations for GW170104 studied in this paper. We compare them with the L1 and H1 signals in grey and light grey in the plots. Here we also perform a double test of the accuracy of the simulations by considering the two main approaches to solve BBHs by the RIT and SXS groups and by considering the internal consistency of convergence of the waveforms with increasing resolutions. The waveforms again show a good agreement among themselves and their differences, shown in the lower panels waveforms are smaller than the residuals of the signals with respect to the N118 simulations. They are larger than in the aligned cases due to the choice of the initial spin configurations at slightly different reference orbital frequencies. Note also that these comparisons do not align the peak of the waveforms and hence if independently fit to data would show much smaller differential residuals.

A. Residuals versus resolution

For each UID, direct comparison of our simulations to the data selects a fiducial total mass which best fits the

TABLE VI. Remnant results for spinning binaries. We show the remnant mass $m_{\rm rem}$ in units of the total initial mass $M \equiv m_1 + m_2$, the remnant dimensionless spin $\chi^z_{\rm rem} \equiv J^z_{\rm rem}/m^2_{\rm rem}$, and the remnant velocity in the x-y plane $V^{xy}_{\rm rem}$. We show results for different LazEv resolutions (N100, N118, and N140) and different SpEC resolutions (L0, L2, L4, and L6).

UID#	$m_{\rm rem}/M$	$\chi^z_{\rm rem}$	$V^{xy}_{\rm rem}$ (km/s)
UID1-N100	0.955 294	0.619 052	402.78
UID1-N118	0.955 310	0.619 079	404.40
UID1-N140	0.955 311	0.619 100	405.63
UID1-L2	0.955 782	0.618 905	
UID1-L3	0.955 813	0.618 893	
UID1-L4	0.955 829	0.618 899	
UID2-N100	0.963 445	0.581 627	962.55
UID2-N118	0.963 418	0.581 480	996.49
UID2-N140	0.963 405	0.581 392	1016.06
UID2-L1	0.963 768	0.579 124	
UID2-L2	0.964 063	0.579 988	
UID2-L3	0.964 063	0.579 958	
UID3-N100	0.961 903	0.659 634	614.70
UID3-N118	0.961 920	0.659 725	598.96
UID3-N140	0.961 927	0.659 781	587.63
UID3-L1	0.962 123	0.658 707	
UID3-L2	0.962 388	0.657 731	
UID3-L3	0.962 401	0.657 599	
UID4-N100	0.962 020	0.529 128	312.81
UID4-N118	0.962 028	0.529 129	313.65
UID4-N140	0.962 030	0.529 130	314.05
UID4-L1	0.962 114	0.528 897	
UID4-L2	0.962 184	0.529 023	
UID4-L3	0.962 174	0.529 117	
UID5-N100	0.968 160	0.531 761	171.57
UID5-N118	0.968 171	0.531 837	175.35
UID5-N140	0.968 173	0.531 873	177.81
UID5-L1	0.967 872	0.531 920	
UID5-L2	0.968 041	0.531 934	
UID5-L3	0.968 051	0.531 917	

observations, as measured by the marginalized likelihood of the data assuming our simulations. Using the same mass for all simulations performed for that UID (e.g., by all groups and for all resolutions), we can for each simulation select the binary extrinsic parameters like event time and sky location which maximize the likelihood of the data, given our simulation and mass. Then, using these extrinsic parameters, we evaluate the expected detector response in the LIGO Hanford (H1) and Livingston (L1) instruments. This procedure has been used to reconstruct the gravitational wave signal for GW150914 [14] and GW170104 events.

Figure 6 shows an example of these reconstructions for the highest log-likelihood NR waveform (top candidate in Table VII and UID# 1). The top panel of this figure shows the NR predicted response in Hanford (blue-red); the Hanford data (grey); and the Livingston data (in dark grey; shifted by −2.93 ms and sign flipped). The bottom panel shows the residuals, and the difference between the two simulations in green. Note that the difference between waveforms is small compared to the residuals, but enough to make the simulation in blue (top candidate in Table VII) have a slightly higher Likelihood (63.0 versus 62.5) over UID# 1 in red, to match the signals over the whole range of frequencies considered. The same simulation resolution have been considered in both cases.

We have also analyzed the finite-differences errors produced by fast-response, low resolution (yet in the convergence regime) simulations of BBH mergers. The low, medium, and high resolutions runs, N100, N118, and N140, respectively, by the RIT group show a nearly 4th order convergence (there are detailed studies of convergence for similar simulations in Refs. [22,41,98]) that allow us to extrapolate to infinite resolution and evaluate the magnitude of the errors in the waveforms as compared to the residuals for this GW170104 event. We thus can evaluate the error of the N118 simulation that is given by the (N100-N118) difference, while the error of the N100 waveform is twice this difference and that of the N140 waveform is half that difference. This is displayed in the lower half of each panel in Figs. 4 and 5 and provide an alternative evaluation of the errors within a given NR method.

The studies carried out in this paper involving three resolutions for each set of parameters well in the convergence regime of the simulations can be very costly from the resources point of view, totaling over 4 million service units (SUs) in computer clusters, as detailed in Table VIII.

According to the variations in the Table IX that evaluates $\ln \mathcal{L}$ for the different resolutions we may derive as a rule of thumb that the N100 grid provides a good approximation for the nonprecessing binaries, while for the precessing ones, N118 is more appropriate. This leads to a reduction of the SUs needed for these 5 simulations, totaling nearly 1 million SUs, two thirds of which are due to the two precessing cases. The pseudospectral approach used by the SXS Collaboration requires similar total wallclock times than the above finite-differences approach, but spends an order of magnitude less resources. For instance, UID# 1 (SXS:BBH:0626) required 11 kSUs for Lev4, 7.4 kSUs for Lev3, and 4.7 kSUs hours for Lev2.

B. Likelihood of NR and models

For any proposed coalescing binary, characterized by its outgoing radiation as a function of all directions, we can compute a single quantity to assess its potential similarity to GW170104, accounting for all possible ways of orienting the source and placing it in the Universe: the marginalized likelihood ($\ln \mathcal{L}_{\rm marg}$) [15,102,103]. To provide a sense of scale, the distribution of $\ln \mathcal{L}_{\rm marg}$ over the posterior distribution including all intrinsic parameters is roughly universal [103], approximately distributed as

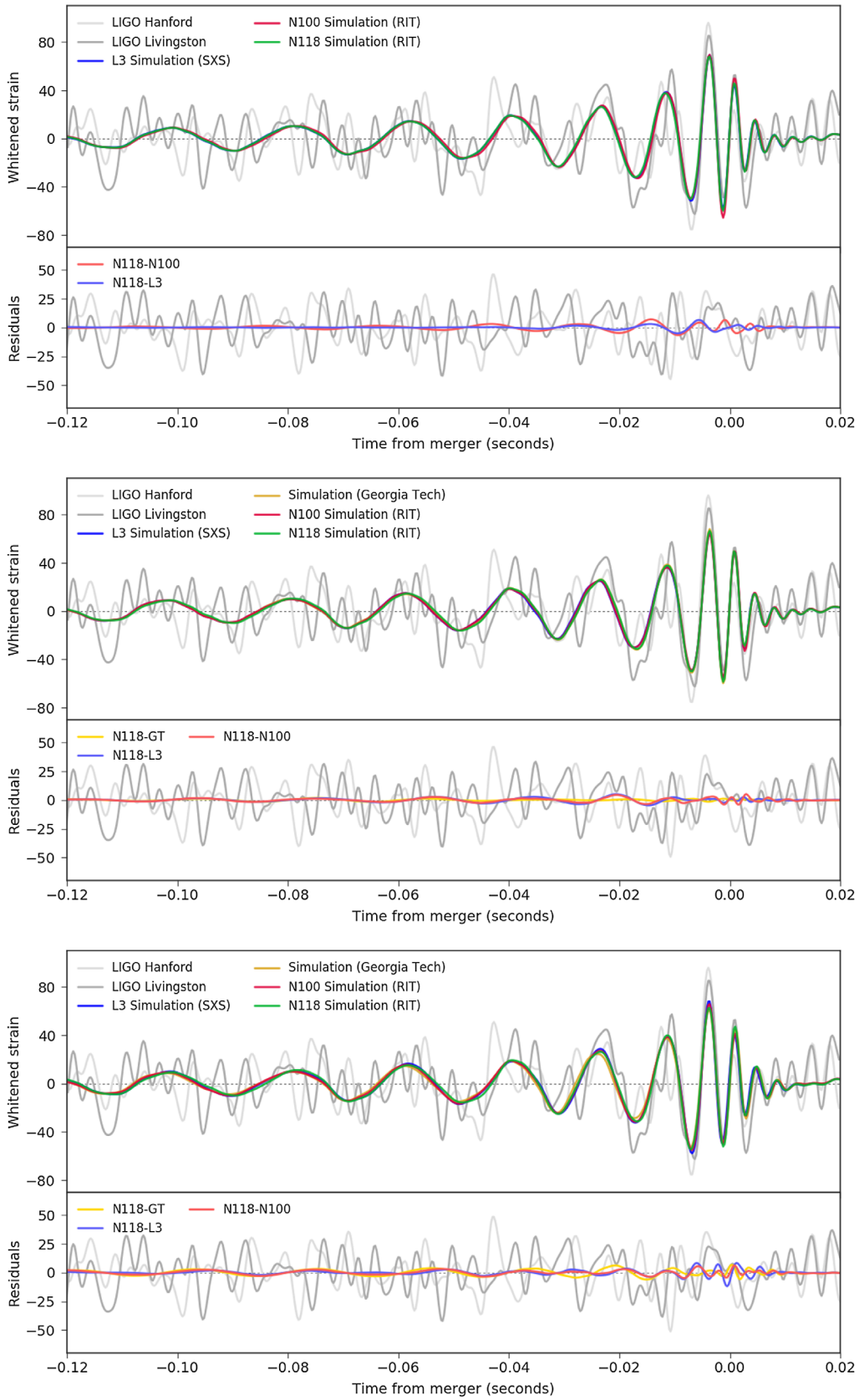

FIG. 4. Comparison of the GW170104 signal seen by LIGO detectors H1 and L1 (in grey and dark grey) with the computer simulations of black hole mergers from SXS, RIT, and GT approaches for the nonprecessing cases labeled as UID# 1, UID# 4, and UID# 5 in Table II.

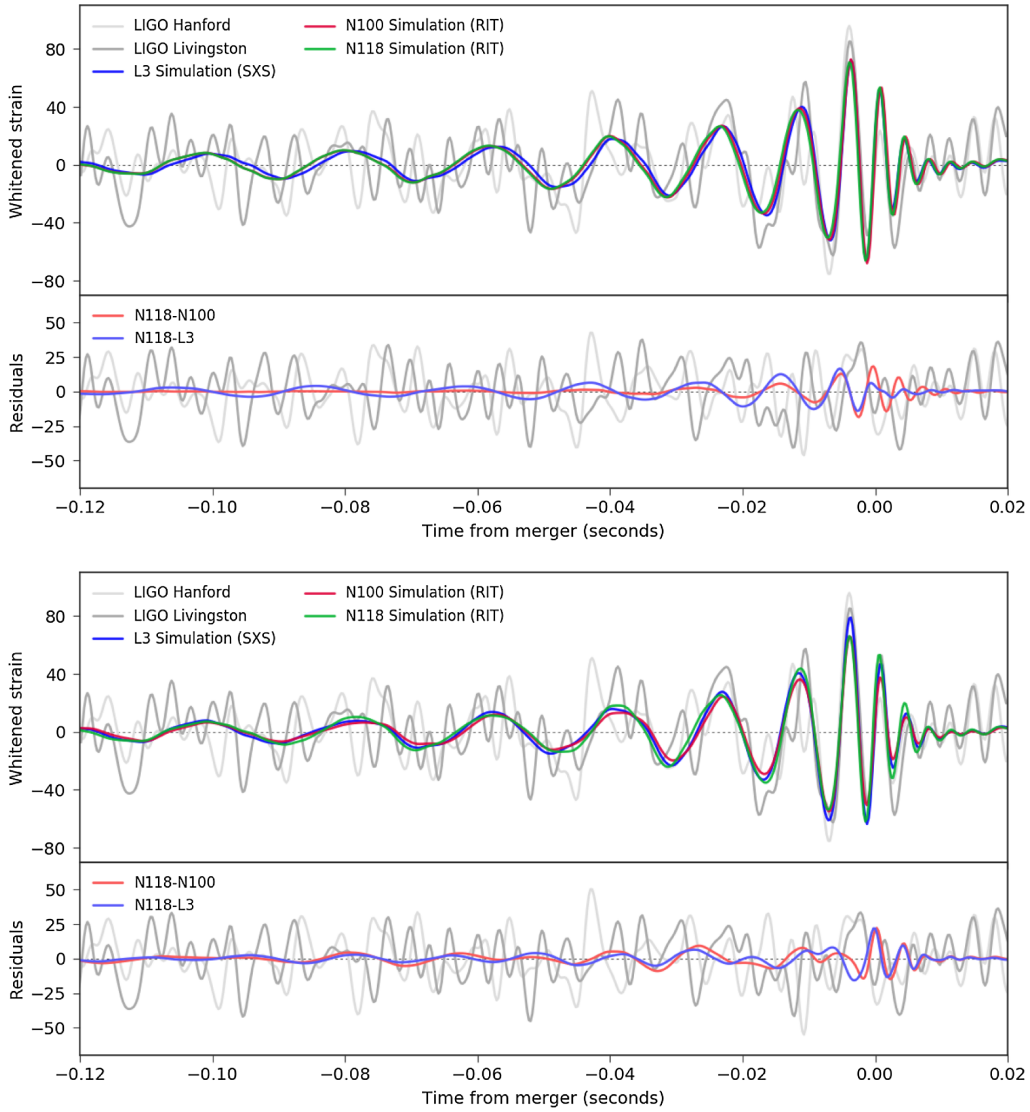

FIG. 5. Comparison of the GW170104 signal seen by LIGO detectors H1 and L1 (in grey and dark grey) with the computer simulations of black hole mergers from SXS and RIT approaches for the precessing cases labeled as UID# 2 and UID# 3 in Table II.

$\ln \mathcal{L}_{\mathrm{marg,max}} - \chi^2/2$ where χ^2 has d degrees of freedom (d.o.f.) (i.e., a mean value of $\ln \mathcal{L}_{\mathrm{marg,max}} - d/2$, and its 90% credible interval is $\ln \mathcal{L}_{\mathrm{marg}} \geq \ln \mathcal{L} - x$, where $x = 3.89$ and $x = 6.68$ for $d = 4$ and $d = 8$, respectively). For each UID and for each proposed total mass M, direct comparison of our simulations to the data allows us to compute a single number measuring the quality of fit: the marginalized likelihood $\mathcal{L}_{\mathrm{marg}}$. The maximum value of this function (here denoted by $\mathcal{L}$) therefore measures the overall quality of fit. Table IX shows $\ln \mathcal{L}$ for the five UIDs simulated here. For comparison, the last column shows $\mathcal{L}$ calculated using an approximate model for the radiation from a coalescing binary. Obviously, if these approximate models and our simulations agree, then we should find the same result for $\ln \mathcal{L}$ at the same parameters. Finally, for context, the peak value of $\ln \mathcal{L}$ computed using SEOBNRv3 with generic parameters is 63.3. If our simulation parameters are well-chosen (and if both our simulations and these models are close to true solutions of Einstein's equations), then this peak value should be in good agreement with the $\ln \mathcal{L}$ evaluated using our simulations.

First and foremost, up to Monte Carlo and fitting error, the marginalized likelihoods calculated with NR agree with each other comparing different resolutions and different approaches to solve the BBH problem, as required given the high degree of similarity between the underlying simulations. Second, the marginalized likelihoods computed at these proposed points are substantially *below* the largest $\mathcal{L}$ found with approximate models like SEOBNRv3, except for UID3. Similar to the explanation described in Appendix B, the exception here is due to the differences between the precessing models ($\ln \mathcal{L}$ was calculated with SEOB but the parameters were suggested with

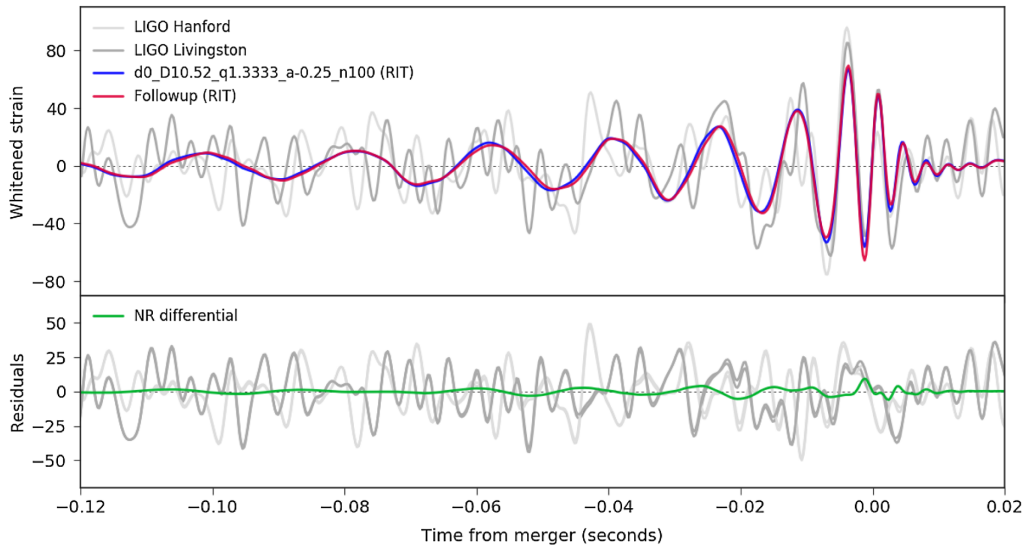

FIG. 6. Comparison of the GW170104 signal seen by LIGO detectors H1 and L1 (in grey and dark grey) with the computer simulations of black hole mergers from RIT at low resolution for the nonprecessing case labeled as UID# 1 in Table II and the highest $\ln \mathcal{L}$ value for an NR simulation given in Table VII (d0_D10.52_q1.3333_a-0.25_n100).

IMRPhenomPv2). Likewise, the binary parameters at which the peak value of $\mathcal{L}$ occurs for SEOBNRv3 are substantially different from any of the proposed parameters explored here. This discrepancy suggests that the model-based procedure that we adopted to target our follow-up simulations was not effective at finding the most likely parameters, as measured with $\ln \mathcal{L}$. The poor performance of our targeted follow-up cannot simply reflect the sampling error; even though the likelihood surface is nearly flat near the peak, small errors are amplified in parameter space. This near flatness also insures that systematic offsets *should* produce a small change in $\ln \mathcal{L}$, *if* the underlying waveform calculations agree; see Appendix B for further discussion. Instead, we suspect the biases in $\mathcal{L}$ arise because the models only approximate the correct solution of Einstein's equations. Third, we confirm our hypothesis in Table VII simply by demonstrating that other simulations (not performed in follow-up) fit the data substantially better than our targeted parameters.

On the one hand, NR follow-up simulations guided by the models (as displayed in Table IX) lead to lower marginalized likelihoods ($\ln \mathcal{L}$). Conversely, other simulations shown in Table VII produce higher $\ln \mathcal{L}$, at points in parameter space where the models predict lower $\ln \mathcal{L}$. This discrepancy suggest the two processes ($\ln \mathcal{L}$ evaluated with NR and with the models) favor different regions of parameter space. In particular, Table VII, which has one of the largest values of $\ln \mathcal{L}$ among all of the (roughly two thousand) simulations available to us, shows that the top precessing simulation is `q50_a0_a8_th_135_ph_30`. This simulation has a mass ratio of 1:2, i.e., $q = 1/2$, where the smaller hole is nonspinning and the larger hole is

TABLE VII. Marginalized likelihood of the data: Selected other simulations: This table shows the results for several other simulations that particularly match the data well and the SEOB model results at those parameter points. These simulations are part of the top 15 simulations in $\ln \mathcal{L}$. When comparing the NR $\ln \mathcal{L}$ values here to the ones in Table IX, one can see these to be generally higher, i.e., better match the data. When comparing the NR $\ln \mathcal{L}$ values to the SEOB at the same points, one sees a consistent lower SEOB $\ln \mathcal{L}$ value. This implies that these points were not picked for NR follow-up due to the lower SEOB $\ln \mathcal{L}$ value.

NR Group	Label	Sim. ID	$q = \frac{m_1}{m_2}$	χ_1	χ_2	$\ln \mathcal{L}$	$\ln \mathcal{L}(SEOB)$ (at NR)	Model
RIT	a	`d0_D10.52_q1.3333_a-0.25_n100`	0.7500	(0, 0, 0)	(0, 0, −0.25)	63.0	62.5	v4
GT	b	`(0.0,1.15)`	0.8696	(0, 0, 0)	(0, 0, 0)	62.2	61.5	v4
RIT	c	`q50_a0_a8_th_135_ph_30`	0.5000	(0, 0, 0)	(0.490, 0.283, −0.566)	62.5	60.7	v3
BAM	d	`BAM150914:24`	0.8912	(−0.278, −0.605, −0.085)	(0.151, 0.396, 0.017)	62.7	61.0	v3
SXS	e	`SXS:BBH:0052`	0.3333	(0.001, 0.008, −0.499)	(0.494, 0.073, 0.001)	62.3	60.4	v3

TABLE VIII. kSUs (1000 core-hours) for each RIT run and resolution.

UID	N100	N118	N140	Total
1	119	184	407	710
2	313	451	557	1321
3	145	217	476	838
4	130	178	565	873
5	67	118	306	491
Total	774	1148	2311	4233

spinning with an intrinsic spin magnitude of 0.8 and pointing initially in a direction downwards with respect to the orbital angular momentum ($\theta = 135°$) and an angle of 30° from the line joining the two black holes ($\phi = 30°$). This simulation belongs to a family of 6 simulations performed in Ref. [104] labeled as NQ50TH135PH [0,30,60,90,120,150]. Those runs, supplemented by two control runs with angles $\phi = 200$, 310 we performed for this paper, are displayed in Fig. 7 versus the $\ln \mathcal{L}$ for this GW170104 event. The lower panels plot all those simulation with respect to their ϕ angle at merger as defined in Ref. [104] and given in Table XXI in that paper. The continuous curve provides a fit (detailed in Table X) for

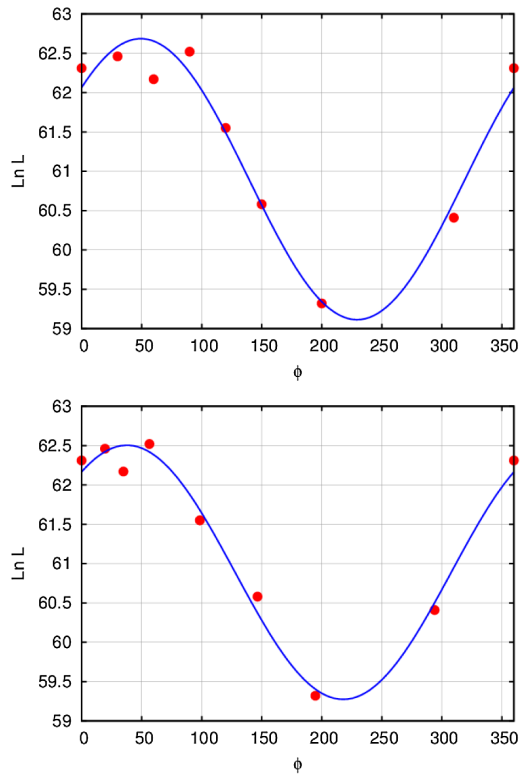

FIG. 7. The log-likelihood of the NQ50TH135 series [104] assuming a period of 2π versus initial angle (top panel) and merger angle (bottom panel.) Data (red) and fits (blue) are given in Table X.

TABLE IX. Marginalized likelihood of the data: This table shows the results for the five simulations when directly compared to the data. For these results, we use the same PSD adopted in all other calculations, with $f_{\rm min} = 30$ Hz (i.e., low-frequency cutoff). The first column is the UID. The second column is the estimated peak log marginalized likelihood $\ln \mathcal{L}$, maximized over binary total mass, for the RIT NR follow-up simulation. The third and fourth columns are the corresponding log marginalized likelihood from SXS and GT, using exactly the same intrinsic parameters (e.g., masses and spins) as maximized by the likelihood in the second column, evaluated using a phenomenological approximate model instead of numerical relativity. The fifth column is the specific model used: either SEOBNRv3 (for precessing simulations) or SEOBNRv4 (for nonprecessing simulations). To see more on this parameter estimation method, see [14,15].

	$\ln \mathcal{L}$ (RIT)			$\ln \mathcal{L}$ (SXS)	$\ln \mathcal{L}$ (GT)	$\ln \mathcal{L}$ (SEOB)	
UID	N100	N118	N140	L3	M120	(at NR)	Model
UID1	60.4	61.0	61.0	60.9	⋯	62.7	v4
UID2	61.0	60.9	60.6	60.9	⋯	61.4	v3
UID3	60.4	60.5	60.7	60.7	⋯	60.4	v3
UID4	60.6	60.7	60.8	60.3	60.4	62.2	v4
UID5	60.0	60.0	60.1	60.0	59.8	61.2	v4

such values as a reference to and an estimate of the maximum value located near the $\phi = 30$ simulation.

The notable results displayed in Fig. 7, where $\ln \mathcal{L}$ seems to be sensitive to the orientation of the spin of the larger hole on the orbital plane, are consistent with broader trends that can be extracted using similar simulations: here, the set of 24 simulations of the family NQ50TH[30,60,90,135]PH [0,30,60,90,120,150] given in Ref. [104] supplemented by the two aligned runs NQ50TH[0,180]PH0 given in Ref. [98] and two runs specifically performed for this paper, NQ50TH135PH[200,310]. These simulations all have $q = 1/2$, a nonspinning smaller BH, and a spinning

TABLE X. The log-likelihood of the NQ50TH135 series [104]. Fittings of the form $\ln \mathcal{L} = A \cos(\pi/180\phi + B) + C$ are also given for both the initial ϕ and $\phi_{\rm merger}$.

ϕ	$\phi_{\rm merger}$	$\ln \mathcal{L}$	$M_z/M_\odot$
0	0	62.3	54.9
30	19.5	62.5	55.2
60	34.8	62.2	54.1
90	56.5	62.5	54.4
120	98.5	61.6	54.1
150	146.5	60.6	54.5
210	194.7	59.3	55.1
310	294.0	60.4	54.6
A	B	C	RMS
1.23 ± 0.21	-0.75 ± 0.15	61.1 ± 0.15	0.38
$A_{\rm merger}$	$B_{\rm merger}$	$C_{\rm merger}$	$\rm RMS_{merger}$
1.08 ± 0.18	-0.47 ± 0.19	61.1 ± 0.15	0.37

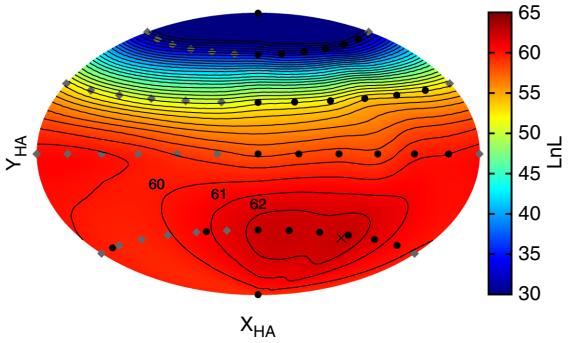

FIG. 8. The log-likelihood of the NQ50THPHI series [104] as a color map with red giving the highest $\ln \mathcal{L}$ and blue the lowest. The black dots (and grey diamonds, obtained by symmetry) represent the NR simulations and we have used Hammer-Aitoff coordinates X_{HA}, Y_{HA}, to represent the map and level curves with the top values of $\ln \mathcal{L} = 60$, 61, 62. The maximum, marked with an X, is located at TH = 137, PH = 87 reaching $\ln \mathcal{L} = 62.6$.

BH with fixed spin magnitude but changing orientation. Figure 8 shows a color map derived from the maximum $\ln \mathcal{L}$ obtained for each of these simulations, using standard (MatLab) plotting tools. The last surface levels indicates the regions of largest likelihood (60,61,62) and a maximum, marked with an X, is located at TH = 137, PH = 87 with $\ln \mathcal{L}$ of 62.6. This results allow us to perform follow-up simulations seeking for this maximum. In the plot, the black points are the NR simulations and the black curves are level sets of the color map. Instead of plotting in the angles theta and phi, we plot in the Hammer-Aitoff coordinates [105], which is a coordinate system where the whole angular space can be viewed as a 2D map. The points at the top left and bottom left are the poles, $\theta = 0$ at the top, and $\theta = \pi$ at the bottom. The line connecting the two is the $\phi = 0$ line. As you move from left to right, ϕ increases from 0 to 150° (the maximum value of ϕ available in these simulations).

C. Reconstructed NR waveforms

The analysis above—a difference in $\ln \mathcal{L}$ for models that should represent the same physical binary which is comparable to the expected range of $\ln \mathcal{L}_{\rm marg}$ over the posterior—suggests modest tension can exist between our NR simulations and the models used to draw inferences about GW170104. To illustrate this tension, in Fig. 9 we display the 90% confidence intervals of the precessing follow-up cases (UID# 2 and UID# 3) computed by the two approximate/phenomenological models comparing them with the full numerical simulations (RIT's with N100 resolutions, note that increasing the numerical resolutions to N118 and N140 reinforces this point). For each simulation, the waveform is generated by first fixing the total mass—selected by maximizing $\mathcal{L}$—and then choosing extrinsic parameters which maximize the likelihood. At

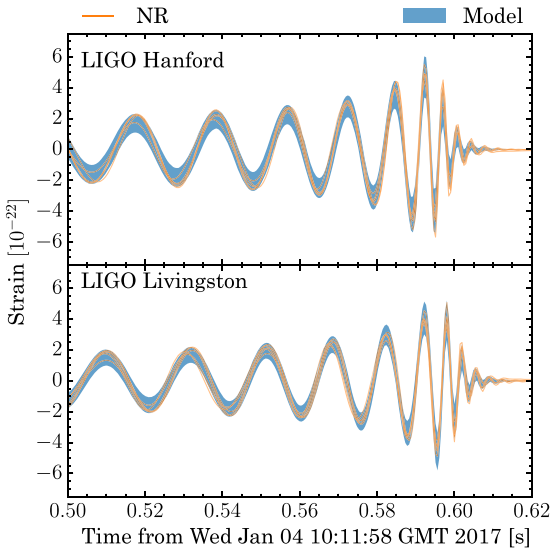

FIG. 9. Comparison of the 90% confidence intervals of GW170104 from the two precessing models with the computer simulations of black hole mergers (in orange) from the best-fitting NR simulations listed in Table VII.

merger, these reconstructed waveforms appear to be in modest tension with the confidence interval reported for $h(t)$; for example, the peaks and troughs of the yellow (NR) curves are consistently at the boundaries of what the 90% credible intervals derived from waveforms allow. This illustration, however, relies on a nonrepresentative metric to assess waveform similarity (i.e, differences in the GW strain as a function of time, without reference to detector sensitivity, assessed by eye).

To remedy this deficiency, Fig. 10 uses the match to compare our reconstructed NR waveforms with reconstructed waveforms drawn from the posterior parameter distribution of GW170104. The top panel uses a violin plot to illustrate the distribution of matches, with a solid bar showing the median value. The bottom panel shows a sample cumulative distribution. The median and maximum of this distribution provides a measure of how consistent the $h(t)$ estimate via NR is with the distribution provided by the model. Using the maximum likelihood waveform from the model and posterior, these distributions should be proportional to a (centrally) χ^2 distributed quantity, with median mismatch $N/2\rho^2$ for N the number of model d.o.f. and ρ the signal to noise ratio (SNR), where the specific choice for ρ depends on the signal and detector/network being studied (e.g., for GW170104, the network SNR was $\simeq$13). By contrast, in several of these overlap distributions, the peak and median values are manifestly offset downward, supporting a significant systematic difference between the radiation predicted from our approximate models and our NR waveforms, each generated from the targeted NR follow-up simulations using parameters drawn from these selfsame model parameter distributions.

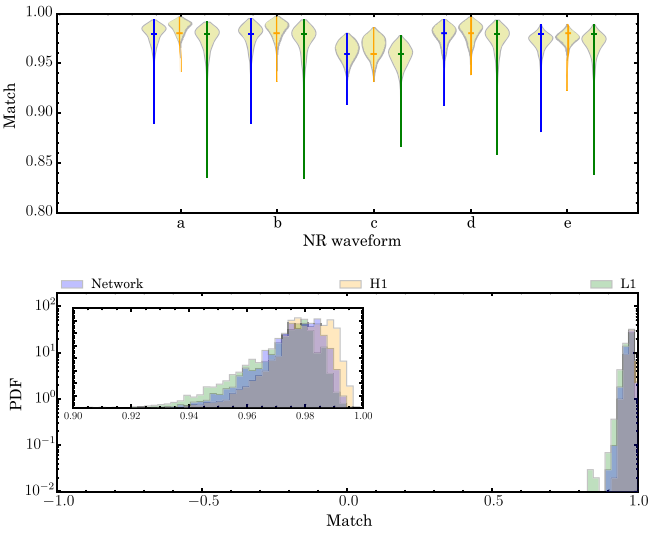

FIG. 10. Distribution of the overall match between each NR waveform (a,b,c,d,e) listed in Table VII and the (distribution of) waveforms produced by model-based parameter inference, as reported in [91]. Matches are produced for the signal in H1, L1, and overall. If a NR signal is perfectly consistent with these models for some parameters, then the distribution of matches will be well approximated by a χ^2 distribution with $d-1$ d.o.f. Many of the best-fitting NR simulations have a distribution of matches that is significantly offset relative to this expected distribution, reflecting the mild tension shown in Fig. 9.

D. Discussion

Using comparisons to data via $\ln \mathcal{L}$ as our guide, we found in Sec. IV B that model-based and NR-based analyses seem to have maxima (in $\ln \mathcal{L}$) in different parts of parameter space; see Appendix B for greater detail. In the region identified as a good fit by model-based analysis, corresponding NR simulations have a low $\ln \mathcal{L}$. Conversely, several NR simulations with distinctly different parameters had a larger $\ln \mathcal{L}$ than the corresponding targeted NR simulations *and* model-based comparisons evaluated at the same parameters. The two functions $\ln \mathcal{L}$, evaluated using models and NR on parameters designed to be similar to and representative of plausible parameters for GW170104, do not agree, implying systematic differences between models and NR (i.e., a change in $\ln \mathcal{L} \simeq 2$). While we have for simplicity adopted one procedure which identifies candidate parameters to select our follow-up simulations, we emphasize that the specific procedure is largely arbitrary, as in this work we simply demonstrate that the two marginalized likelihoods (NR and model-based) disagree somewhere. Changes in the exact location and value of the marginalized likelihood are of less interest than changes in the full posterior distribution; the latter subject is beyond the scope of this study.

The NR follow-up simulations and Bayesian inferences used in this work were performed soon after the identification of GW170104, and as such did not benefit from recent improvements in waveform modeling. Notably, by calibrating to a large suite of numerical relativity simulations, surrogate waveform models have been generated that, in a suitable part of parameter space, are markedly superior to any of the waveform models used for parameter inference to date [106,107]. Parameter inferences performed with these models should be more reliable and (by optimizing $\ln \mathcal{L}$) enable better targets for NR follow-up simulations.

For simplicity and brevity, we have directly compared our nonprecessing and precessing simulations to only one of the two extant families of phenomenological waveform models (SEOBv3/v4). While the two models are in good agreement for nonprecessing binaries (SEOBNRv4/IMRPhenomD), the other precessing model (IMRPhenomPv2) has technical complications that limit its utility for our study. On the one hand, we cannot generate a similar waveform with similar initial conditions, preventing us from performing the straightforward comparisons shown in Fig. 3. [As a frequency domain model, it did not adopt the same time conventions as NR and time-domain models for the precession phase (see, e.g., [108]).] On the other hand, the implementations available do not provide a spin-weighted spherical harmonic decomposition, preventing us from performing the mode-by-mode mismatch calculations in Table IV.

Previous investigations have demonstrated by example that posterior inferences with approximate waveform models can be biased, even for parameters consistent with observed binary black hole [108,109]. For example, a previous large study using simulations consistent with GW150914 found that, despite the brevity and relative simplicity of its signal, the inferred parameters could be biased for certain binary configurations relative to the line of sight [109], and much less so for others (e.g., nonprecessing and comparable-mass binaries). The relevance and frequency of these configurations is not yet determined and depends on the binary black hole population which nature provides.

V. CONCLUSIONS

After the detection of GW170104 [91], we performed several simulations of binary black hole mergers, intending to reproduce LIGO's observations using simulations with similar parameters. The parameters used were selected based on LIGO's reported inferences about GW170104, generated by comparing two approximate models for the binary black hole merger to the GW170104 data. Comparing these targeted simulations of binary black hole mergers, we find good agreement. We have shown that the differences among typical numerical simulations, used as a measure of their error, are much smaller (by over an order of magnitude) than the residuals of observation versus theory. On the other hand, we demonstrate (expected) differences between our numerical solutions to general

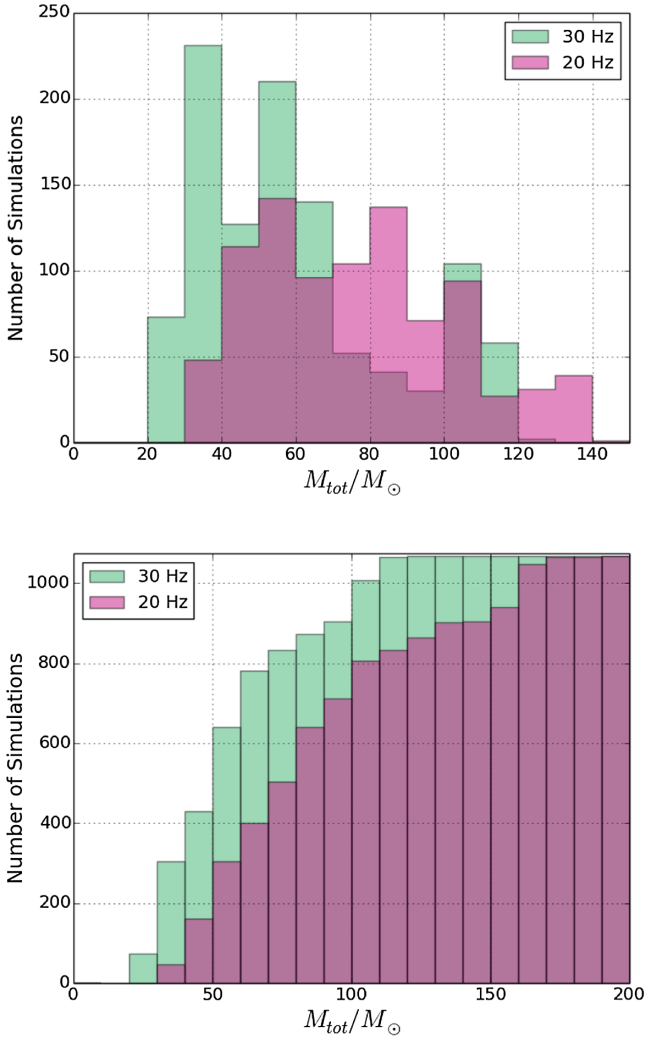

FIG. 11. This plot shows histograms of the number of simulations in a given total mass bin of size 10 $M_\odot$, for starting at 20 Hz and for starting at 30 Hz with the given total mass. This is for all the runs in the public SXS + GT + RIT catalogs [19–21]. The bottom histograms show how many simulations in the catalogs can be used from 20 or 30 Hz from a minimal mass on, i.e., the cumulative of the upper plot.

relativity and the approximate models used to target our simulations. Because we used these models to identify candidate parameters for follow-up, our follow-up simulations were systematically biased away from the best-fitting parameters. These biases are not surprising, as the models used do not fully incorporate all the physics of binary merger, including higher modes and all features of precession, and are known to modestly disagree both with one another and with NR simulations. This does not mean that the models are not recovering the full signal: both models and NR could find similar likelihoods, but for different parameters. These biases can be particularly large for small mass ratios and highly spinning precessing binaries. We demonstrate that other, preexisting simulations with different parameters fit the data substantially better than the configurations targeted by model-based techniques.

We have shown here (and in previous studies [15,110]) that the standard low resolution, fast-response, simulations provide an accurate description of GW signals, and can improve over the parameters determined by the models (See Table VII and Fig. 7) for precessing and nonprecessing cases (note that while SEOBNRv4 improves on the inaccurate [13] SEOBNRv2 [111], it is still not at comparable accuracy to the NR simulations, See Figs. 2–3, for instance). The tension between the models and the full numerical simulations (notwithstanding [112]) may be crucial in determining parameters such as individual spin of the holes and tests of general relativity for the large SNR signals, where the limitations of the models is larger). Both this study, focused on GW170104, and the investigation by [108], carried out on GW151226, point to the limitations of existing models to accurately determine binary parameters in the case of precessing BBH.

Regarding prospects for future follow-ups, Fig. 11 shows the distributions of the minimal total mass of the BBH systems in the NR catalogs [19–21] given a starting gravitational wave frequency of 20 or 30 Hz in the source frame and its cumulative. This provides a coverage for the current events observed by LIGO [redshift effects improve this coverage by a factor of $(1+z)$, where z is the redshift]. Coverage of lower total masses would require longer simulations or hybridization of the current waveforms.

Finally, we demonstrated the power of using purely numerical waveforms to determine parameters of a binary black hole merger as the previous case of GW150914 [103] and similarly in the case of the source GW170104. More work is needed though to systematically and robustly include hybridization of waveforms and the case of generically precessing binaries [15].

ACKNOWLEDGMENTS

The authors thank S. Husa, C. Berry, K. Chatziioannou, and A. Zimmerman for their feedback on this work. The RIT authors gratefully acknowledge the NSF for financial support from Grants No. PHY-1607520, No. PHY-1707946, No. ACI-1550436, No. AST-1516150, No. ACI-1516125, and No. PHY-1726215. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [allocation TG-PHY060027N], which is supported by NSF Grant No. ACI-1548562. Computational resources were also provided by the NewHorizons and BlueSky Clusters at the Rochester Institute of Technology, which were supported by NSF Grants No. PHY-0722703, No. DMS-0820923, No. AST-1028087, and No. PHY-1229173. R. O. S. is supported by NSF Grants No. AST-1412449, No. PHY-1505629, and No. PHY-1607520. The GT authors gratefully acknowledge NSF support through Grants No. PHY-1505824, No. 1505524, and No. XSEDE TG-PHY120016. They also gratefully acknowledge support

from the Cullen-Peck and Dunn Families. The SXS Collaboration authors gratefully acknowledge the Sherman Fairchild Foundation and NSF for financial support from Grants No. PHY-1307489, No. PHY-1606522, No. PHY-1606654, and No. AST-1333129. They also gratefully acknowledge support for this research at CITA from NSERC of Canada, the Ontario Early Researcher Awards Program, the Canada Research Chairs Program, and the Canadian Institute for Advanced Research. Calculations were done on the ORCA computer cluster, supported by NSF Grant No. PHY-1429873, the Research Corporation for Science Advancement, CSU Fullerton, the GPC supercomputer at the SciNet HPC Consortium [113]; SciNet is funded by the Canada Foundation for Innovation (CFI) under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund (ORF)—Research Excellence; and the University of Toronto. Further calculations were performed on the Briarée cluster at Sherbrooke University, managed by Calcul Québec and Compute Canada and with operation funded by the Canada Foundation for Innovation (CFI), Ministére de l'Économie, de l'Innovation et des Exportations du Quebec (MEIE), RMGA and the Fonds de recherche du Québec—Nature et Technologies (FRQ-NT).

APPENDIX A: IMPACT OF DISCRETE POSTERIOR SAMPLING

In the text, we performed several calculations that depend on an inferred posterior distribution: identifying parameters for NR follow-up calculations, maximizing the marginalized likelihood $\ln \mathcal{L}$, and calculating mismatches. These calculations are performed using a finite-size collection of approximately independent, identically distributed samples from a posterior distribution [92]. In this appendix we briefly quantify the (small) effects our finite sample size has on our conclusions and comparisons. For simplicity, we will conservatively standardize our calculations to $N = 3000$ posterior samples; in practice, usually many more were used.

The match and marginalized likelihood distributions are well described by a χ^2 distribution with a suitable number of d.o.f., corresponding to the model dimension of the intrinsic parameter space (i.e., $d = 4$ for calculations which omit precession, and $d = 8$ for calculations which include it). For example if $\ln \mathcal{L}_*$ is the true maximum marginalized likelihood, then the marginalized likelihood distribution over the posterior is well approximated by the distribution of $\ln \mathcal{L} = \ln \mathcal{L}_{\rm max} - x/2$ where x is χ^2 distributed with d d.o.f. If we have N independent draws from the χ^2 distribution, the smallest value of x will be distributed according to $P(>x|d)^N = (1 - P(< x|d))^N$, where $P(<x|d)$ is the cumulative distribution for the χ^2 distribution. At 95% confidence, the maximum value of x over the N samples is therefore $P^{-1}(0.05^{1/N})$. As a result, if we estimate the maximum value of $\ln \mathcal{L}$ with the maximum over our posterior samples, we find an estimate which is smaller than the true maximum value $\ln \mathcal{L}_{\rm max}$ by $0.5P^{-1}(0.05^{1/N})$. Evaluating this expression for $d = 4$ and $d = 8$ in the conservative limit of only $N = 3000$ samples, we find a systematic sampling error of 0.045 (0.31) in $d = 4$ ($d = 8$), respectively, in our estimate of the peak marginalized likelihood. This systematic sampling error in our estimate of the peak marginalized likelihood is smaller than the differences in marginalized likelihoods discussed in the text and figures.

Likewise, given the number of samples, the targeted parameters should be very close to the true maximum *a posteriori* values. Qualitatively speaking, due to finite sample size effects, our estimate of each parameter z has an uncertainty of roughly $\sigma_z/\sqrt{N}$, or roughly 2% of the width of the distribution using our fiducial sample size.

APPENDIX B: MIXED MESSAGES: MAXIMUM LIKELIHOOD, $\ln \mathcal{L}$, AND *A POSTERIORI*

One goal of this work is to demonstrate, by a concrete counterexample, that NR follow-up must be targeted and assessed self-consistently.

One source of inconsistency in our original NR follow-up strategy was the algorithm by which NR simulations were selected from model-based inference. Our NR follow-up simulations were selected by (approximately) maximizing the *a posteriori* probability, proportional to the (15-dimensional) likelihood L; the (seven-dimensional) prior $p(\theta)$ for extrinsic parameters θ; and the (eight-dimensional) prior for intrinsic parameters $p(\lambda)$. This MaP location does not generally correspond to the parameters which maximize the likelihood (maxL). The intrinsic parameters selected by both approaches also do not cause the marginalized likelihood $\mathcal{L}_{\rm marg}(\lambda) = \int d\theta p(\theta) L(\lambda, \theta)$ to take on its largest value. In principle, to avoid introducing artificial inconsistencies simply due to the choice of point estimate, we should have targeted follow-up simulations using $\ln \mathcal{L}$. To assess how much our choice of targeting impacted our estimate of $\ln \mathcal{L}$, we evaluated the marginalized likelihood at our estimates of all three locations. Each location was approximated by our (finite-size) set of posterior samples. For the posterior distribution adopted to generate UID4—a nonprecessing production-quality analysis where SEOBNRv4 was both used to generate the reference posterior used to find the MaP and maxL parameters and to compute a model-based $\ln \mathcal{L}$—we find that the model-based $\ln \mathcal{L}$ values at the MaP and maxL points to be effectively indistinguishable due to Monte Carlo error (61.4 and 61.2, respectively, with an estimated Monte Carlo error of 0.1). This similarity suggests that, when a fully self-consistent analysis is performed, then even if the MaP and maxL parameters differ slightly, they will produce similar values of $\ln \mathcal{L}$, with

differences far smaller than the differences between NR and model-based analysis.

For the reasons described in Sec. IV D, we consistently adopt SEOB-based models to evaluate our model-based $\ln \mathcal{L}$. Because different phenomenological approximants do not quite agree [108], the posterior distributions used to identify the parameters for UID3 and 5 used an IMRPhenomPv2 (precessing) IMRPhenomD (nonprecessing) approximant, respectively, to produce different MaP and maxL parameters. Conversely, to the extent these models somewhat agree, they should estimate model parameters corresponding to the same values of $\ln \mathcal{L}$. In fact, however, when we evaluate $\ln \mathcal{L}$ with SEOBNRv4 on the MaP and maxL parameters of the posterior used to find UID5, we find both values disagree with the values seen for UID4, being lower (60.8) and higher (62), respectively. These differences in $\ln \mathcal{L}$ clearly indicate small differences between the two model-based analyses, comparable to (but smaller than) the differences seen between model-based analyses and NR.

[1] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016).
[2] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016).
[3] B. P. Abbott *et al.* (Virgo Collaboration, LIGO Scientific Collaboration), Phys. Rev. X **6**, 041015 (2016).
[4] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Phys. Rev. Lett. **118**, 221101 (2017).
[5] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Astrophys. J. **851**, L35 (2017).
[6] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. Lett. **119**, 141101 (2017).
[7] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).
[8] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015).
[9] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[10] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu, Phys. Rev. D **89**, 061502 (2014).
[11] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2014).
[12] S. Babak, A. Taracchini, and A. Buonanno, Phys. Rev. D **95**, 024010 (2017).
[13] G. Lovelace, C. O. Lousto, J. Healy, M. A. Scheel, A. García, R. O'Shaughnessy, M. Boyle, M. Campanelli, D. A. Hemberger, L. E. Kidder, H. P. Pfeiffer, B. Szilágyi, S. A. Teukolsky, and Y. Zlochower, Classical Quantum Gravity **33**, 244002 (2016).
[14] B. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. D **94**, 064035 (2016).
[15] J. Lange *et al.*, Phys. Rev. D **96**, 104041 (2017).
[16] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005).
[17] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006).
[18] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. Lett. **96**, 111102 (2006).
[19] A. H. Mroue *et al.*, Phys. Rev. Lett. **111**, 241104 (2013).
[20] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, Classical Quantum Gravity **33**, 204001 (2016).
[21] J. Healy, C. O. Lousto, Y. Zlochower, and M. Campanelli, Classical Quantum Gravity **34**, 224001 (2017).
[22] J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **89**, 104052 (2014); **90**, 104004 (2014).
[23] A. Ghosh *et al.*, Phys. Rev. D **94**, 021101 (2016).
[24] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. Lett. **116**, 221101 (2016).
[25] Y. Zlochower, J. Baker, M. Campanelli, and C. Lousto, Phys. Rev. D **72**, 024021 (2005).
[26] P. Marronetti, W. Tichy, B. Bruegmann, J. Gonzalez, and U. Sperhake, Phys. Rev. D **77**, 064010 (2008).
[27] C. O. Lousto and Y. Zlochower, Phys. Rev. D **77**, 024034 (2008).
[28] F. Löffler, J. Faber, E. Bentivegna, T. Bode, P. Diener, R. Haas, I. Hinder, B. C. Mundim, C. D. Ott, E. Schnetter, G. Allen, M. Campanelli, and P. Laguna, Classical Quantum Gravity **29**, 115001 (2012).
[29] Einstein Toolkit home page, http://einsteintoolkit.org.
[30] Cactus Computational Toolkit, http://cactuscode.org.
[31] E. Schnetter, S. H. Hawley, and I. Hawke, Classical Quantum Gravity **21**, 1465 (2004).
[32] J. Healy, C. O. Lousto, H. Nakano, and Y. Zlochower, Classical Quantum Gravity **34**, 145011 (2017).
[33] S. Brandt and B. Brügmann, Phys. Rev. Lett. **78**, 3606 (1997).
[34] M. Ansorg, B. Bruegmann, and W. Tichy, Phys. Rev. D **70**, 064011 (2004).
[35] J. Thornburg, Classical Quantum Gravity **21**, 743 (2004).
[36] O. Dreyer, B. Krishnan, D. Shoemaker, and E. Schnetter, Phys. Rev. D **67**, 024018 (2003).
[37] M. Campanelli, C. O. Lousto, Y. Zlochower, B. Krishnan, and D. Merritt, Phys. Rev. D **75**, 064030 (2007).
[38] M. Campanelli and C. O. Lousto, Phys. Rev. D **59**, 124022 (1999).
[39] C. O. Lousto and Y. Zlochower, Phys. Rev. D **76**, 041502 (2007).

[40] H. Nakano, J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **91**, 104022 (2015).
[41] J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **96**, 024031 (2017).
[42] B. Brügmann, J. A. González, M. Hannam, S. Husa, U. Sperhake, and W. Tichy, Phys. Rev. D **77**, 024027 (2008).
[43] F. Herrmann, I. Hinder, D. M. Shoemaker, P. Laguna, and R. A. Matzner, Phys. Rev. D **76**, 084032 (2007).
[44] F. Herrmann, I. Hinder, D. Shoemaker, P. Laguna, and R. A. Matzner, Astrophys. J. **661**, 430 (2007).
[45] I. Hinder, B. Vaishnav, F. Herrmann, D. Shoemaker, and P. Laguna, Phys. Rev. D **77**, 081502 (2008).
[46] J. Healy, F. Herrmann, I. Hinder, D. M. Shoemaker, P. Laguna, and R. A. Matzner, Phys. Rev. Lett. **102**, 041101 (2009).
[47] I. Hinder, F. Herrmann, P. Laguna, and D. Shoemaker, Phys. Rev. D **82**, 024033 (2010).
[48] J. Healy, J. Levin, and D. Shoemaker, Phys. Rev. Lett. **103**, 131101 (2009).
[49] J. Healy, P. Laguna, R. A. Matzner, and D. M. Shoemaker, Phys. Rev. D **81**, 081501 (2010).
[50] T. Bode, R. Haas, T. Bogdanovic, P. Laguna, and D. Shoemaker, Astrophys. J. **715**, 1117 (2010).
[51] S. Husa, I. Hinder, and C. Lechner, Comput. Phys. Commun. **174**, 983 (2006).
[52] http://www.black-holes.org/SpEC.html.
[53] G. Lovelace, R. Owen, H. P. Pfeiffer, and T. Chu, Phys. Rev. D **78**, 084017 (2008).
[54] J. W. York, Phys. Rev. Lett. **82**, 1350 (1999).
[55] H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, Comput. Phys. Commun. **152**, 253 (2003).
[56] S. Ossokine, F. Foucart, H. P. Pfeiffer, M. Boyle, and B. Szilágyi, Classical Quantum Gravity **32**, 245010 (2015).
[57] M. Caudill, G. B. Cook, J. D. Grigsby, and H. P. Pfeiffer, Phys. Rev. D **74**, 064011 (2006).
[58] H. P. Pfeiffer, D. A. Brown, L. E. Kidder, L. Lindblom, G. Lovelace, and M. A. Scheel, Classical Quantum Gravity **24**, S59 (2007).
[59] A. Buonanno, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, and A. Taracchini, Phys. Rev. D **83**, 104034 (2011).
[60] A. H. Mroué and H. P. Pfeiffer, arXiv:1210.2958.
[61] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, Classical Quantum Gravity **23**, S447 (2006).
[62] H. Friedrich, Commun. Math. Phys. **100**, 525 (1985).
[63] D. Garfinkle, Phys. Rev. D **65**, 044029 (2002).
[64] F. Pretorius, Classical Quantum Gravity **22**, 425 (2005).
[65] L. Lindblom and B. Szilágyi, Phys. Rev. D **80**, 084019 (2009).
[66] M. W. Choptuik and F. Pretorius, Phys. Rev. Lett. **104**, 111101 (2010).
[67] B. Szilágyi, L. Lindblom, and M. A. Scheel, Phys. Rev. D **80**, 124010 (2009).
[68] G. Lovelace, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **83**, 024010 (2011).
[69] B. Szilágyi, Int. J. Mod. Phys. D **23**, 1430014 (2014).
[70] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009).
[71] D. A. Hemberger, M. A. Scheel, L. E. Kidder, B. Szilágyi, G. Lovelace, N. W. Taylor, and S. A. Teukolsky, Classical Quantum Gravity **30**, 115001 (2013).
[72] S. Ossokine, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. D **88**, 084031 (2013).
[73] M. A. Scheel, M. Giesler, D. A. Hemberger, G. Lovelace, K. Kuper, M. Boyle, B. Szilágyi, and L. E. Kidder, Classical Quantum Gravity **32**, 105009 (2015).
[74] O. Rinne, Classical Quantum Gravity **23**, 6275 (2006).
[75] O. Rinne, L. Lindblom, and M. A. Scheel, Classical Quantum Gravity **24**, 4053 (2007).
[76] C. Gundlach, Phys. Rev. D **57**, 863 (1998).
[77] G. B. Cook and B. F. Whiting, Phys. Rev. D **76**, 041501(R) (2007).
[78] R. Owen, Ph.D. thesis, California Institute of Technology, 2007.
[79] M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky, Phys. Rev. D **76**, 124038 (2007).
[80] L. T. Buchman and O. C. A. Sarbach, Classical Quantum Gravity **24**, S307 (2007).
[81] O. Rinne, L. T. Buchman, M. A. Scheel, and H. P. Pfeiffer, Classical Quantum Gravity **26**, 075009 (2009).
[82] N. T. Bishop, R. Gomez, L. Lehner, M. Maharaj, and J. Winicour, Phys. Rev. D **56**, 6298 (1997).
[83] J. Winicour, Living Rev. Relativity **8** (2005).
[84] R. Gomez, W. Barreto, and S. Frittelli, Phys. Rev. D **76**, 124029 (2007).
[85] C. Reisswig, N. T. Bishop, and D. Pollney, Gen. Relativ. Gravit. **45**, 1069 (2013).
[86] C. J. Handmer and B. Szilágyi, Classical Quantum Gravity **32**, 025008 (2015).
[87] M. Boyle and A. H. Mroué, Phys. Rev. D **80**, 124045 (2009).
[88] N. W. Taylor, M. Boyle, C. Reisswig, M. A. Scheel, T. Chu, L. E. Kidder, and B. Szilágyi, Phys. Rev. D **88**, 124010 (2013).
[89] T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilágyi, Classical Quantum Gravity **33**, 165001 (2016).
[90] M. Boyle, Phys. Rev. D **93**, 084031 (2016).
[91] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. Lett. **118**, 221101 (2017).
[92] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015).
[93] P. Schmidt, I. W. Harry, and H. P. Pfeiffer, arXiv: 1703.01076.
[94] M. Campanelli, C. O. Lousto, H. Nakano, and Y. Zlochower, Phys. Rev. D **79**, 084010 (2009).
[95] M. Boyle, Phys. Rev. D **93**, 084031 (2016).
[96] B. J. Kelly and J. G. Baker, Phys. Rev. D **87**, 084004 (2013).
[97] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **95**, 044028 (2017).
[98] J. Healy and C. O. Lousto, Phys. Rev. D **95**, 024037 (2017).
[99] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, Phys. Rev. D **95**, 064024 (2017).

[100] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilgyi, Phys. Rev. D **96**, 024058 (2017).
[101] L. London, S. Khan, E. Fauchon-Jones, X. J. Forteza, M. Hannam, S. Husa, C. Kalaghatgi, F. Ohme, and F. Pannarale, arXiv:1708.00404.
[102] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. D **92**, 023002 (2015).
[103] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Phys. Rev. D **94**, 064035 (2016).
[104] Y. Zlochower and C. O. Lousto, Phys. Rev. D **92**, 024022 (2015).
[105] E. W. Weinstein, http://mathworld.wolfram.com/Hammer-AitoffEqual-AreaProjection.html.
[106] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Phys. Rev. Lett. **115**, 121102 (2015).
[107] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Phys. Rev. D **95**, 104023 (2017).
[108] A. R. Williamson, J. Lange, R. O'Shaughnessy, J. A. Clark, P. Kumar, J. Caldern Bustillo, and J. Veitch, Phys. Rev. D **96**, 124041 (2017).
[109] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, Classical Quantum Gravity **34**, 104002 (2017).
[110] J. C. Bustillo, A. Boh, S. Husa, A. M. Sintes, M. Hannam *et al.*, arXiv:1501.00918.
[111] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu, Phys. Rev. D **89**, 061502 (R) (2014).
[112] B. P. Abbott *et al.* (LIGO-Virgo Collaboration), Classical Quantum Gravity **34**, 104002 (2017).
[113] C. Loken, D. Gruner, L. Groer, R. Peltier, N. Bunn, M. Craig, T. Henriques, J. Dempsey, C.-H. Yu, J. Chen, L. J. Dursi, J. Chong, S. Northrup, J. Pinto, N. Knecht, and R. V. Zon, J. Phys. Conf. Ser. **256**, 012026 (2010).