The most powerful astrophysical events: gravitational-wave peak luminosity of binary black holes as predicted by numerical relativity

David Keitel,[1,2,*] Xisco Jiménez Forteza,[1,†] Sascha Husa,[1,‡] Lionel London,[3,§] Alessandro Nagar,[4] Sebastiano Bernuzzi,[5] Enno Harms,[6] Mark Hannam,[3] Sebastian Khan,[3,7] Michael Pürrer,[8] Geraint Pratten,[1] and Vivek Chaurasia[9,6]

[1]*Universitat de les Illes Balears, IAC3—IEEC, 07122 Palma de Mallorca, Spain*
[2]*School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom*
[3]*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, United Kingdom*
[4]*Institut des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France*
[5]*DiFeST, University of Parma, and INFN, I-43124 Parma, Italy*
[6]*Theoretical Physics Institute, University of Jena, 07743 Jena, Germany*
[7]*Albert-Einstein-Institut, Callinstraße 38, 30167 Hannover, Germany*
[8]*Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany*
[9]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560012, India*

(Dated: 30 December 2016)
(LIGO document number: LIGO-P1600279-v5)

For a brief moment, a binary black hole (BBH) merger can be the most powerful astrophysical event in the visible universe. Here we present a model fit for this gravitational-wave *peak luminosity* of nonprecessing quasicircular BBH systems as a function of the masses and spins of the component black holes, based on numerical relativity (NR) simulations and the hierarchical fitting approach introduced in [1]. This fit improves over previous results in accuracy and parameter-space coverage and can be used to infer posterior distributions for the peak luminosity of future astrophysical signals like GW150914 and GW151226. The model is calibrated to the $\ell \leq 6$ modes of 378 nonprecessing NR simulations up to mass ratios of 18 and dimensionless spin magnitudes up to 0.995, and includes unequal-spin effects. We also constrain the fit to perturbative numerical results for large mass ratios. Studies of key contributions to the uncertainty in NR peak luminosities, such as (i) mode selection, (ii) finite resolution, (iii) finite extraction radius, and (iv) different methods for converting NR waveforms to luminosity, allow us to use NR simulations from four different codes as a homogeneous calibration set. This study of systematic fits to combined NR and large-mass-ratio data, including higher modes, also paves the way for improved inspiral-merger-ringdown waveform models.

I. INTRODUCTION

With Advanced LIGO's [2, 3] first detections [4–6], BBH coalescences have become objects of observational astronomy. The peak rate at which BBHs radiate gravitational wave (GW) energy makes them the most luminous known phenomena in the universe. The source of the first GW detection GW150914 has been inferred to be consistent with two black holes of $29^{+4}_{-4}\,M_\odot$ and $36^{+5}_{-4}\,M_\odot$ inspiralling, merging and ringing down as described by General Relativity (GR) . In such a proces GW energy is emitted at a peak rate of $3.6^{+0.5}_{-0.4} \times 10^{56}$ erg/s, equivalent to $200^{+30}_{-20}\,M_\odot\,c^2/\mathrm{s}$ [4, 7, 8]. Though this *peak luminosity*, L_{peak}, is not electromagnetic, but gravitational, we can compare its numerical value to the photon luminosity of other astrophysical sources to illustrate its scale: GW150914 at its peak emitted as much power as $\sim 10^{23}$ suns, $\gtrsim 10^{11}$ times more than all stars in the Milky Way, and still 60–90 times more than the ultra-luminous gamma-ray burst GRB 110918A [9].[1]

The peak luminosities for LIGO's first BBH events were inferred using a fit [10] to data from numerical relativity (NR) simulations, which we will improve upon in this paper through an enhanced fitting method and a significantly larger calibration data set. Source parameters of GW events are determined through Bayesian inference [6, 7, 11–13], comparing LIGO data with waveform models, which are approximate maps between the masses and spins of the binary components and the GW signal. As of the aLIGO O1 run, the state-of-the-art NR-calibrated BBH waveform models were PhenomPv2 [14–16] (a precessing model based on the aligned-spin PhenomD [17, 18]) and SEOBNRv2/3 [13, 19–22]. A detailed recent study has found these models to be at least sufficiently accurate in the parameter region corresponding to the first detection [23].

The primary products of this inference are multi-dimensional sample chains that approximate posterior probability density functions (PDFs) for the intrinsic and extrinsic BBH parameters. Subsequently, such a sampled PDF can be used to infer other quantities, typically obtained through fitting formulae calibrated to NR. Examples include final-state properties: the final spin and final mass of the merger remnant, a single Kerr black hole (BH) , which also yield the total radiated energy (e.g. [1, 17, 24–26]). In fact, full inspiral-merger-ringdown waveform models include such final-state NR fits to describe the ringdown phase, but due to practical implementation details and for greater flexibility in using updated fits, the values reported in [4–6, 13] come from standalone fitting formulae evaluated on posterior PDFs.

The same approach is used for inference of the GW peak luminosity. However, a robust L_{peak} model for general BBH configurations was not available in the literature prior to O1. An early phenomenological formula [27] is limited to non-spinning BBHs, and thus a new fit [10] had to be developed. To accurately capture the luminosities of the NR calibration set, it also took into account contributions from subdominant harmonics not included in most current waveform models.

In this paper, we present an improved version of that model

*Electronic address: david.keitel@ligo.org
†Electronic address: francisco.forteza@ligo.org
‡Electronic address: sascha.husa@ligo.org
§Electronic address: lionel.london@ligo.org

[1] Assuming $L_\odot = 3.8 \times 10^{33}$erg/s, $L_{\mathrm{MW}} = 2 \times 10^{11} L_\odot$ and the GRB's estimated peak isotropic equivalent luminosity of $(4.7\pm0.2)\times10^{54}$ erg/s [9].

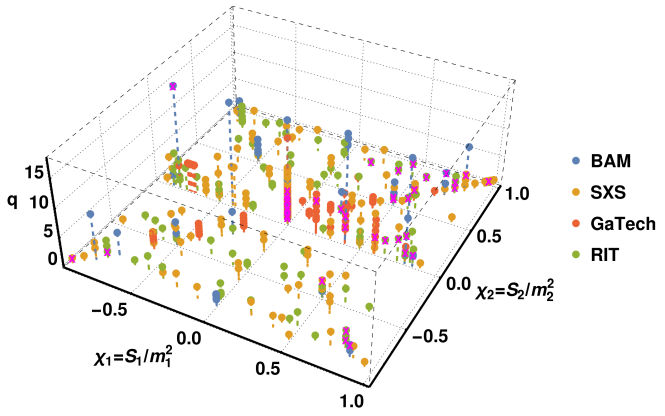

FIG. 1: Parameter-space coverage of the combined NR data set from BAM, SXS, GaTech, RIT, shown against the individual BH spins and the mass ratio $q = m_1/m_2$ of the system. Simulations not used in the fit (see Tbl. VI) are marked with magenta crosses.

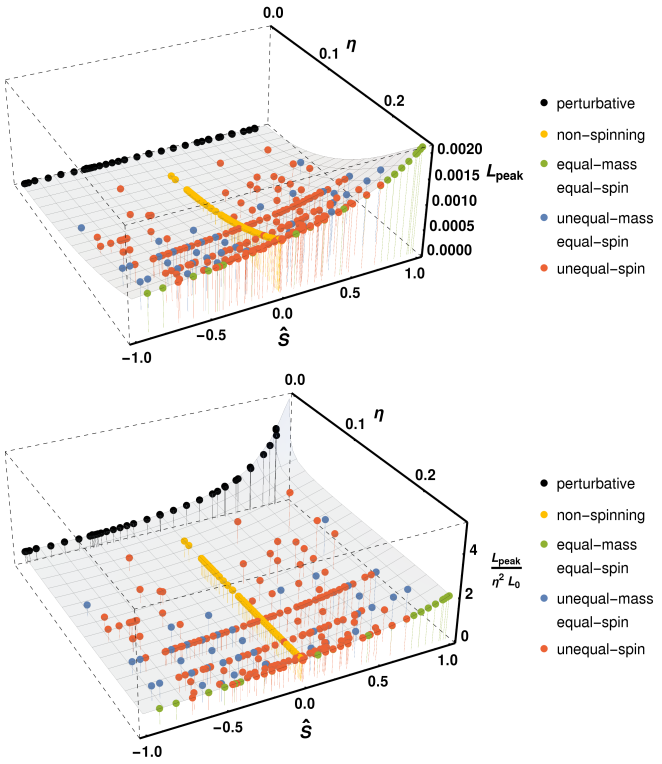

FIG. 2: Combined data set over the two-dimensional space spanned by symmetric mass ratio η and effective spin $\widehat{S}$, defined in Eq. (3). Top panel: peak luminosity $L_{\rm peak}$, lower panel: rescaled as $L_{\rm peak}/\eta^2 L_0$. Subsets used in the various steps of Sec. III are highlighted by colors. The shaded surface is added here to guide the eye, but is in fact the 2D projection of the new fit developed in this paper.

fit for the GW peak luminosity from the merger of more general BBH systems, including spins on both binary components. Still, we concentrate on cases where the spin of each BH is aligned with the system's total angular momentum. We use the hierarchical data-driven approach introduced in [1] to develop a three-dimensional ansatz and fit it to a total of 378 simulations from four separate NR codes, including more subdominant modes than before, and to independent numerical results for large mass ratios obtained with the perturbative scheme of [28–30]. This addition is essential in producing a well-constrained fit at very unequal masses where NR coverage is sparse or nonexistent.

Besides the interest in peak luminosity as an astrophysical quantity, which is not currently a direct observable but only accessible through NR-calibrated fits, another motivation for this improved fit is its role as a test case of the general fitting method from [1] for quantities that require accurate treatment of the higher modes from NR simulations, and of combining NR and perturbative large-mass-ratio results. In these aspects, the present study prepares for the development of improved inspiral-merger-ringdown waveform models.

Notably, the GW peak luminosity $L_{\rm peak}$ does not depend on the total mass of a BBH system: Luminosity generally scales with emitted energy over emission timescale, $L \sim E_{\rm rad}/\Delta t$. But for a BBH, both the total radiated energy $E_{\rm rad}$ and the characteristic merger timescale Δt are proportional to the total mass, so that $L_{\rm peak}$ is independent of it. Hence, the GW peak luminosities even of supermassive BBHs, observable by eLISA-like missions [31–33] or by pulsar timing arrays (PTAs, [34–36]), are similar to those of stellar-mass BBHs. The results of this paper will be applicable to such systems as well. Furthermore, we use geometric units with $G = 1$, $c = 1$ and unit total mass M, so that luminosity is a dimensionless quantity, corresponding to $1M$ of energy radiated per $1M$ of time. The conversion factor to Watt is c^5/G, and another factor 10^7 for the usual astronomical unit of erg/s.

We will first review the catalog of NR simulations of BBH mergers and perturbative large-mass-ratio data used to construct our model in Sec. II. We discuss the challenges in combining NR data from different simulation codes, the steps taken to process the different sets into a single, effectively homogeneous data set, and how this set is augmented with independent results for large-mass-ratio systems. Given this data set, we discuss the construction and validation of our model fit for peak luminosity in Secs. III and IV. We also compare our new fit with the previous result [10], calibrated to a smaller NR data set, that was used during O1 [4–7, 13]; and to another independent, recently-published fit [26]. The appendices include more details on investigations of the NR data.

II. INPUT DATA

A. Numerical Relativity Datasets

We begin by considering 419 non-precessing NR simulations from four sources, with their coverage of the three-dimensional BBH parameter space illustrated in Figs. 1 and 2:

- 47 simulations performed by the authors using the BAM [37, 38] code, including those first used in [1, 17];
- 160 simulations from the public SXS catalog [39, 40] performed with the Spectral Einstein Code [41];
- 105 simulations from the public GaTech catalog [42, 43], performed with the MAYA code [44–47];
- 107 simulations with the LAZEV code [48] with results published in [24, 26], labeled as "RIT" data in the following.

BAM, MAYA and LAZEV are finite difference codes to solve the Baumgarte-Shapiro-Shibata-Nakamura formulation of the GR initial value problem [49] with a singularity-avoiding slicing

condition following the moving puncture approach [50, 51], whereas the simulations of the SXS collaboration have been performed with the pseudospectral SpEC code [41] which employs the generalized harmonic gauge (GHG) combined with black-hole excision.

We use mass and spin parameters of the component BHs after equilibration and the initial burst of 'junk' radiation. To compute the luminosity for BAM, SXS and GaTech simulations, we begin with the Weyl scalar ψ_4 decomposed into its spin-two spherical harmonic multipoles,

$$\psi_{\ell m}(t) \;=\; \frac{1}{r}\int_\Omega {}_{-2}\overline{Y}_{\ell m}(\theta,\phi)\,\psi_4(t,\theta,\phi)\,\mathrm{d}\Omega\,. \qquad (1)$$

For SXS ψ_4 data, we have applied corrections for center-of-mass drifts [52–54], which remove some unphysical oscillations in higher modes. From these spherical harmonic multipoles, we calculate the GW strain-rate multipoles $\dot{h}_{\ell m}(t)$ via the fixed-frequency-integration (FFI) method described in [55]. We then compute the peak luminosity according to

$$L_{\rm peak} = \max_t \lim_{r\to\infty} \frac{r^2}{16\,\pi}\sum_{\ell=2}^{\ell_{\rm max}}\sum_{m=-\ell}^{+\ell}\left|\,\dot{h}_{\ell m}(t)\,\right|^2\,, \qquad (2)$$

truncating the sum over ℓ at $\ell_{\rm max}=6$. For RIT simulations, we use directly the peak luminosity values as given in [26], which again include all modes up to $\ell_{\rm max}=6$.

NR simulation results always have finite accuracy, and the post-processing from the initial $\psi_4(t,\theta,\phi)$ to the final product $L_{\rm peak}$ can lead to additional sources of inaccuracies that could substantially affect the accuracy of individual data points and of any NR-calibrated fits. As we aim to fit small subdominant effects, such as those of unequal spins, possible error sources must be carefully analyzed. Thus, we have considered the impact of the following effects on $L_{\rm peak}$:

- Conversion between strain $h(t)$, $\psi_4(t)$ and luminosity $L(t)$: The FFI approach is known to be accurate at the 1% level or better [55]. We have also tested its validity by developing an alternative technique to compute $h(t)$ and $L(t)$ from the time-domain $\psi_4(t)$.

- Extrapolation effects: Gauge invariability of the waveforms is only well defined for an observer at null infinity. We have extrapolated $L_{\rm peak}$, as computed from waveforms available at a range of finite extraction radii, to infinity and estimated the extrapolation uncertainty.

- Peak accuracy and discreteness: The peak in luminosity can be quite steep, but we have verified that with a sampling of 0.1M in time the difference between two points next to the peak is only on the order of 0.01–0.2%.

- Mode selection: While the $\ell=|m|=2$ mode has the largest peak amplitude for all mass ratios, the importance of other spherical harmonics monotonically increases toward the extreme-mass-ratio limit. We can thus bound higher-mode contributions by comparison with the large-mass-ratio results, where neglecting modes with $\ell>8$ incurs an error below 1% even for mass ratios of 10^5 (see below in Sec. II B). For the available NR waveforms, we determine that it is generally sufficient to consider modes up to $\ell_{\rm max}=6$, with contributions of the $\ell=7,8$ modes rising to only 1% for mass-ratio 18 non-spinning cases.

Details on these aspects are given in appendix A. We remove 41 cases from the initial catalog for reasons as discussed in appendix A, e.g. because they are inconsistent with equivalent or nearby configurations from the same or other codes. Thus, we perform our fit with a final set of 378 NR results.

B. Perturbative large-mass-ratio data

As the computational cost of NR simulations increases rapidly for unequal mass ratios, no data for BBH systems with mass ratios $q>18$ are currently available from any of the simulation codes discussed above. ($q=m_1/m_2$ with the convention $m_1>m_2$.) However, constraining fits to some known behavior in or close to the extreme-mass-ratio limit is essential in ensuring a sane extrapolation towards that limit, and also to reduce uncertainty in the intermediate-mass-ratio region where there is some NR coverage, but it is still very sparse.

For the final-state quantities studied in [1], we used analytical expressions from [56] for the limiting case of a test particle orbiting a Kerr black hole. For the peak luminosity, it is known [57, 58] that the leading-order term as $\eta\to 0$ must be $L_{\rm peak}\propto\eta^2$, with the symmetric mass ratio $\eta=(m_1m_2)/(m_1+m_2)^2=q/(1+q)^2$. However, no fully analytical results for the spin dependence in the extreme-mass-ratio limit exist. Instead, here we constrain our fit by numerical results for finite, but very large mass ratios.

The simulation method for BBH mergers in the test-mass (large-mass-ratio) limit developed in [28, 30, 59] combines a semi-analytical description of the dynamics with a time-domain numerical approach for computing the multipolar waveform based on BH perturbation theory. The small BH's dynamics are prescribed using the effective-one-body (EOB) test-mass dynamics, i.e. the conservative (geodesic) motion augmented with a linear-in-η radiation reaction expression [28, 60]. The latter is built from the factorized and resummed circularized waveform introduced in [61] (and [62] for spin) and uses Post-Newtonian (PN) information up to 5.5PN (see also Refs. [63, 64] for extensions up to 22PN). Waveforms are calculated by solving either the Regge-Wheeler-Zerilli (RWZ) 1+1 equations (nonspinning case) or the Teukolsky 2+1 equation (spinning case). Those equations are solved in the time-domain using hyperboloidal coordinates to extract the radiation unambiguously at scri [29, 30, 65].

The method has been extensively applied for developing and assessing the quality of EOB waveforms [29, 30], informing the EOB model in the test-mass limit [66], QNM excitation [30, 59], and computing gravitational recoils [67, 68].

The large-mass-ratio waveforms employed here were produced in [29] and [30] (RWZ and Teukolsky data respectively). These waveforms are approximate because $O(\eta)$ effects are not taken into account in the conservative dynamics. However, the consistency of the method was proven in the nonspinning case by showing that, for $\eta\to 0$, the analytical mechanical flux assumed for the small BH's motion agrees with the numerical GW fluxes to few percents up to the last stable orbits [29]. The same check has been performed for the spinning case where, instead, significant deviations were found where the spin of the central BH is large ($\chi_1\gtrsim 0.7$) and aligned with the orbital angular momentum [30]. The discrepancy originates from poor performance of the straight, 5PN-accurate, EOB-resummed analytical multipolar wave-

form, from which the radiation reaction force is built, for large spins [62]. An iterative method to produce consistent $O(\eta)$ spinning waveforms has been proposed [30]; and two such waveforms at $\chi_1 = \pm 0.9$ are available for consistency tests. The method proposed in [30] is very expensive, since several iterations are needed to find good consistency between the radiation reaction that is used to drive the dynamics and the waveform. Ref. [64] proposed an additional factorization and resummation of the residual waveform amplitudes of [62] that delivers a more accurate analytical waveform amplitude up to the last stable orbit (or even the light-ring) when the BH dimensionless spin tends to 1. The additional resummation discussed in Ref. [64] (or minimal variations of it), once incorporated in the radiation reaction, is expected to strongly improve the self-consistency of the Teukolsky waveforms computed as in Ref. [30] without need for the iteration procedure.

We use only the Teukolsky results at $q = 10^3$ (31 data points) and the RWZ results at $q = 10^4$ and $q = 10^5$ (7 data points each), as the RWZ at $q = 10^3$ are expected to be less accurate, and indeed their luminosities deviate at negative spins. In Fig. 2 we compare the qualitative behavior of peak luminosities from NR and perturbative data, and the spin dependence is analyzed in Sec. III C.

III. CONSTRUCTING THE PHENOMENOLOGICAL FIT

We apply the hierarchical modeling scheme for the three-dimensional non-precessing BBH parameter space that was introduced in [1] and is summarized in Fig. 3. The general idea is to construct a fit ansatz that matches the structure actually seen in the data set, and to model effects in order of their importance: first fit well-constrained subspaces as functions of the dominant parameters, then add sub-dominant effects only to the degree that they are supported by the data.

The parameter-space dimensionality is the same for peak luminosity as for final spin or radiated energy: just as the final (dimensionless) spin, the peak luminosity is manifestly independent of the total mass of the system, while for radiated energy the total mass is still only an overall scale factor. Hence, for non-precessing quasi-circular BBHs, this leaves a three-dimensional parameter space: mass ratio and two spin parameters χ_1 and χ_2.

As expected, and obvious visually in Fig. 2, one principal direction of curvature in the data set is given by the mass ratio, equivalently expressed as q or the symmetric η. We can then perform a 3D hierarchical fit by changing the spin parametrization from the two component spins $\chi_i = S_i/m_i^2$ to a dominant symmetric component

$$\widehat{S} = (m_1^2\chi_1 + m_2^2\chi_2)/(m_1^2 + m_2^2) \tag{3}$$

and a subdominant antisymmetric component $\Delta\chi = \chi_1 - \chi_2$. This is the same choice of effective spin parameter as in [1, 17]. Tests like in appendix C of [1] show that the fitting method is robust under changing to different parametrizations like the $\chi_{\rm eff}$ parameter previously used in [10].

We perform our fits not on the peak luminosity $L_{\rm peak}$ itself, but on the rescaled quantity $L'_{\rm peak} = L_{\rm peak}/\left(\eta^2 L_0\right)$, which removes the expected dominant η^2 dependence (known analytically for large mass ratios [57, 58] and found as the dominant term in [10]) to make small subdominant corrections easier to identify, and where we have also scaled out the equal-mass, zero-spin value $L_0 = L_{\rm peak}\,(\eta = 0.25, \chi_1 = \chi_2 = 0)\,/\,0.25^2$ to get typical values of order unity. We use $L_0 \approx 0.0164$, the average of the SXS, GaTech and RIT results at this configuration. (These three simulations agree within 0.2%.)

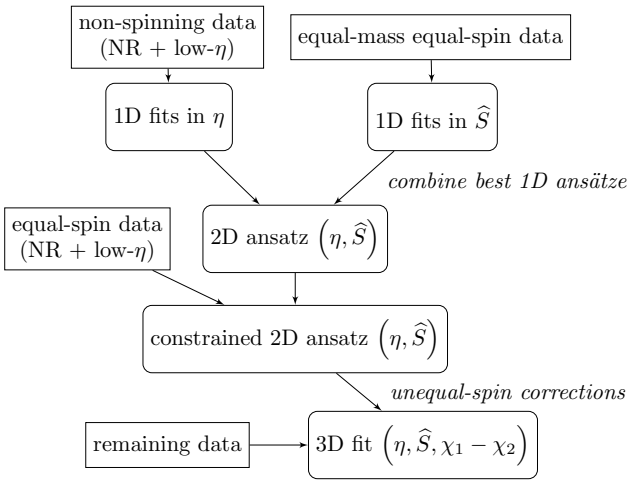

FIG. 3: Flow chart of the hierarchical step-by-step construction leading to the final 3D ansatz, as introduced in [1], but adjusted for the different handling of large-mass-ratio (low-η) information.

All fits are performed with Mathematica's `NonlinearModelFit` function. To avoid overfitting, our model selection is guided by the Akaike and Bayesian information criteria (AICc, BIC, [69, 70]), which help to choose between fits based not only on the overall goodness of fit, as measured e.g. by the root-mean-square error (RMSE) , but also penalize excessively high numbers of free coefficients. The AICc is defined as

$$\mathrm{AICc} = -2\ln\mathcal{L}_{\rm max} + 2N_{\rm coeffs} + \frac{2N_{\rm coeffs}(N_{\rm coeffs}+1)}{N_{\rm data} - N_{\rm coeffs} - 1}\,, \tag{4}$$

with the maximum log-likelihood $\mathcal{L}_{\rm max}$ (assumed Gaussian). This is the standard AIC proposed by Akaike [69] plus a correction for low $N_{\rm data}$. Schwarz' alternative criterion [70]

$$\mathrm{BIC} = -2\ln\mathcal{L}_{\rm max} + N_{\rm coeffs}\ln(N_{\rm data})\,, \tag{5}$$

despite its name, generally cannot be understood directly as a Bayesian evidence. For specific advantages and disadvantages of these two criteria, their mathematical and philosophical basics and other alternatives see e.g. [71] and references therein. For both, the model with the *lowest* value is preferred. The BIC tends to impose a slightly stronger penalty on extra parameters than the AICc, and we use it as a default ranking of fits, though in practice we do not find disagreements between the two criteria.

A. One-dimensional non-spinning fit

First, we analyse 84 non-spinning cases, including 81 NR simulations as well as the non-spinning large-mass-ratio data points. As in [1], we consider several ansatz choices for the one-dimensional function $L'_{\rm peak}(\eta)$: polynomials up to seventh order, denoted as P(m), as well as rational functions, denoted as R(m, k) for polynomial orders m and k in the numerator and

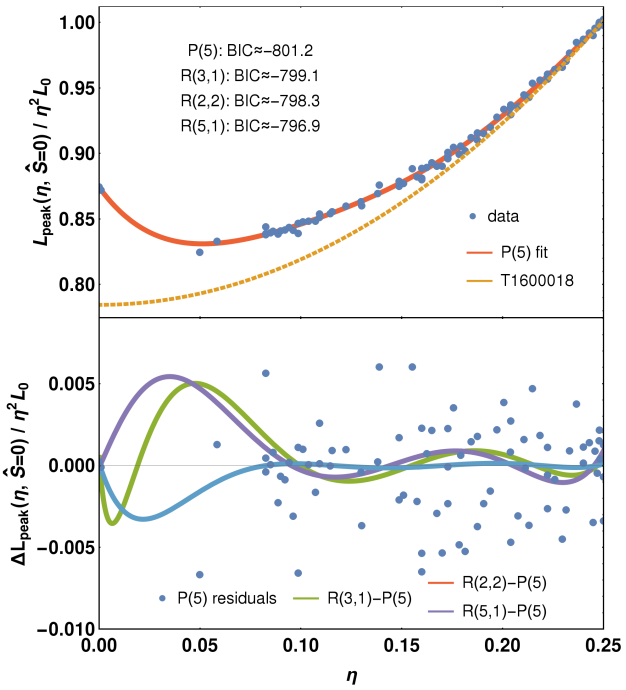

FIG. 4: One-dimensional fits of the rescaled non-spinning peak luminosity $L'_{\rm peak}\left(\eta, \widehat{S}=0\right)$. Top panel: the preferred fifth-order polynomial, see Eq. (6), and comparison with the previous fit from Ref. [10]. Lower panel: residuals of this fit (points) and differences from the three next-highest-ranking fits in terms of BIC (lines).

	estimate	std.err.	rel.err.[%]
a_0	0.874	0.00104	0.1
a_1	−2.11	0.282	13.3
a_2	35.2	7.01	19.9
a_3	−245.	63.6	26.0
a_4	877.	249.	28.3
a_5	−1170.	354.	30.2

TABLE I: Fit coefficients for the one-dimensional non-spinning $L'_{\rm peak}\left(\eta, \widehat{S}=0\right)$ fit over 84 data points, along with their uncertainties (standard errors) and relative errors (std.err./estimate).

denominator, respectively. We construct the latter as Padé approximants from an initial polynomial fit to simplify the handling of initial values in the fitting algorithm.

With the dominant η^2-dependence already scaled out, fitting the higher-order corrections allows us to achieve subpercent accuracy, though the additional fit coefficients are not very tightly constrained. The top-ranked fit both by BIC and AICc (with marginally significant differences) is a fifth-order polynomial

$$L'_{\rm peak}\left(\eta, \widehat{S}=0\right) = a_5\eta^5 + a_4\eta^4 + a_3\eta^3 + a_2\eta^2 + a_1\eta + a_0 \quad (6)$$

with its fit coefficients and their uncertainties given in Tbl. I. Fig. 4 shows this fit, its residuals and comparisons with both the previous fit from [10] ("T1600018") and the next-highest-ranked alternatives. These next-best alternatives are all rational functions, with the next-simpler polynomial P(4) disfavored by 7 in BIC and 20% in RMSE and the next-higher-order P(6) marginally disfavored by 4 in BIC with almost identical RMSE. We find a clear upwards correction over the T1600018 result at low η, and differences between highly-ranking fits that are much smaller than this correction or the

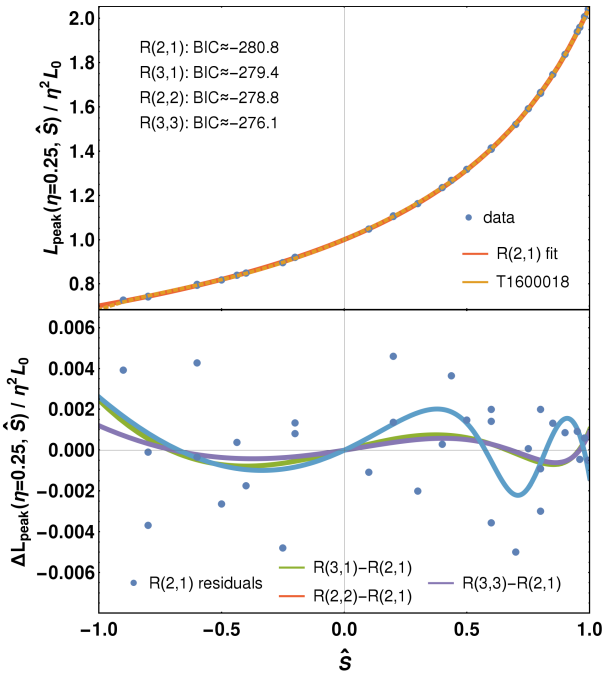

FIG. 5: One-dimensional fits of the rescaled equal-mass-equal-spin peak luminosity $L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right)$. Top panel: best fit in terms of BIC, a rational function R(2,1), see Eq. (7), and the almost indistinguishable P(5) from [10]. Lower panel: residuals of this fit (points) and differences from three next-best-ranked fits by BIC (lines).

	estimate	std.err.	rel.err.[%]
b_1	0.98	0.00232	0.2
b_2	−0.178	0.0276	15.5
b_4	1.79	0.0104	0.6

TABLE II: Fit coefficients for the one-dimensional equal-mass-equal-spin $L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right)$ fit over 32 data points.

typical residuals. In the data-less region between the lowest-η NR case ($q = 18$) and the perturbative results, differences between the highest-ranking fits are larger, but still at most at the same level as the typical fit residuals at higher η, corresponding to relative errors below 0.6%. As another comparison, refitting the simple $L_{\rm peak}(\eta) = a_2\eta^2 + a_4\eta^4$ ansatz that we used in [10] (which in $L'_{\rm peak}$ corresponds to just const. $+\,\eta^2$) is disfavored by over 280 in BIC over this data set, and has a four times higher RMSE.

All highly-ranked fits agree that the NR data cannot be connected to the large-mass-ratio regime with a simple monotonic function. This behavior might seem surprising, but can be explained by studying the individual modes: the observed behavior of the total peak luminosity results from competing trends of modes that either fall or rise towards $\eta \to 0$. (See Sec. A 5 and Fig. 20 for details.) Also we recall that the full $L_{\rm peak}\left(\eta, \widehat{S}=0\right)$ is of course monotonic after the dominant η^2 term has been factored back in.

B. One-dimensional equal-mass-equal-spin fit

Next, we consider the 32 equal-mass and equal-spin NR simulations, i.e. configurations with $\eta = 0.25$ and $\chi_1 = \chi_2$, fitting the one-dimensional function $L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right)$. We

use a similar set of polynomial and rational ansätze, with the intercept fixed by requiring consistency with the η fit in the non-spinning case, $L'_{\rm peak}\left(\eta = 0.25, \widehat{S} = 0\right)$. The curvature of this spin dependence at equal masses is relatively mild and can be best fit by a three-coefficient rational function ansatz

$$L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right) = \frac{0.107 b_2 \widehat{S}^2 + 0.465 b_1 \widehat{S}}{1 - 0.328 b_4 \widehat{S}} + 1.00095\,, \quad (7)$$

with the numerical prefactors due to constructing the ansatz as a Padé approximant to simplify handling of initial values in the fitting code. This fit is marginally top-ranked by both AICc and BIC; it is shown in Fig. 5 and the coefficients b_i are given in Tbl. II. Low-order rational functions are clearly preferred over polynomials, with the P(5) we used in [10] disfavored by +14 in BIC and having 12% higher RMSE, and the simple R(2,1) ansatz is fully sufficient to describe the data to similar sub-percent accuracy as the non-spinning set. Adding another term in either the numerator or denominator is possible, but does not improve the statistics; while adding too many terms tends to induce unconstrained coefficients or singularities within the fitting region.

C. Spin dependence at large mass ratios

Analyzing the perturbative data at large mass ratios, we verify that the mass-ratio dependence is completely dominated by the leading-order η^2 scaling in this regime, with the rescaled $L'_{\rm peak}$ equal up to 0.2% for the three non-spinning data points at the mass ratios $q \in \{10^3, 10^4, 10^5\}$.

We treat the single-spin perturbative data as equivalent to equal-spin results, which is easily accurate enough as $\widehat{S}(q = 10^3, \chi_1 = 1.0, \chi_2 = 1.0) - \widehat{S}(10^3, 1.0, 0.0) \approx 10^{-6}$ and the spin-difference terms (to be fitted below in Sec. III E) are expected to be suppressed at least by η^2.

We find that the spin dependence at each of these mass-ratio steps is much steeper than for equal masses, requiring a higher-order spin ansatz. A spin term at least as complex as R(4,1) is clearly preferred over any lower-order alternatives, with the $q = 10^3$ data yielding ΔBIC ≈ 44 and a factor > 2 in RMSE in favor of R(4,1) against the R(2,1) found at equal masses, and similar preference even for the two highest mass ratios for which we have only 7 data points available each.

Since this analysis of the large-mass-ratio data alone is only used to guide the ansatz choice in the next section, but not used directly as a constraint, we do not list the detailed results of these fits here. Instead, the final 3D fit using NR and perturbative data will be compared with the high-q data in Sec. IV B.

D. Two-dimensional fits

In proceeding with the hierarchical modeling approach, we can now make a two-dimensional equal-spin ansatz informed and constrained by the previous 1D steps and the large-mass-ratio information. In [1], we constructed 2D final-state ansätze by first simply adding the two one-dimensional fits and then generalizing each spin coefficient by a polynomial in η. This time, we find that we need to introduce additional η-dependent higher-order terms in $\widehat{S}$, as the curvature

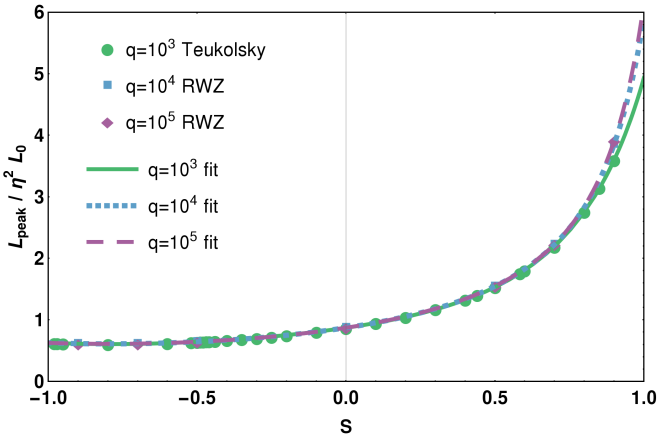

FIG. 6: Numerical data from two perturbative codes and at three mass ratios, together with fits of the form const. + R(4, 1).

of $L'_{\rm peak}$ along the spin dimension increases from equal masses towards the largest mass ratios.

We thus consider a 2D ansatz of the general form

$$L'_{\rm peak}\left(\eta, \widehat{S}\right) = L'_{\rm peak}\left(\eta, \widehat{S} = 0\right) + \mathrm{R}(m,k)\left(\eta, \widehat{S}\right) \quad (8)$$

with the η fit from Eq. (6) and the rational function $\mathrm{R}(m,k)$ in $\widehat{S}$ inheriting the coefficients b_i from Tbl. II and filled up with $b_i = 1.0$ for orders not present in $L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right)$ from Eq. (7). We then introduce the required freedom to change the curvature along the η dimension through the substitution

$$b_i \to b_i \sum_{j=0}^{j=J} f_{ij}\,\eta^j\,, \quad (9)$$

with a maximum expansion order J.

On the other hand, the number of free coefficients is reduced again by consistency constraints with the 1D fits:

$$f_{i2} = 16 - 16 f_{i0} - 4 f_{i1} \quad \text{for } b_i \text{ from } \eta = 0.25 \text{ fit}\,, \quad (10a)$$
$$f_{i2} = -16 f_{i0} - 4 f_{i1} \quad \text{for other } b_i\,. \quad (10b)$$

In practice, we use R(4,2) to match the $q = 10^3$ result, thus introducing one extra power of $\widehat{S}$ in both numerator and denominator compared to $L'_{\rm peak}\left(\eta = 0.25, \widehat{S}\right)$ in Eq. (7).

With 92 equal-spin data points not yet used in the two one-dimensional subspace fits (including 50 NR simulations and the single-spin large-mass-ratio results, which as discussed above can be considered as effectively equal-spin), we can easily expand the polynomials in η from Eq. (9) to order $J = 2$, $b_i \to b_i\left(f_{i0} + f_{i1}\eta + f_{i2}\eta^2\right)$, and still obtain a well-constrained fit. The only further constraint is that we set the remaining highest-order coefficient in the denominator, f_{71}, to zero to avoid a singularity within the physical $\left(\eta, \widehat{S}\right)$ region, leaving 11 free coefficients.

The resulting fit and its residuals over the equal-spin data set are plotted in Fig. 7. We again find sub-percent relative errors over most of the calibration set, with an RMSE of ≈ 0.0057 and only two cases over 1% relative error (both $q = 8$ from `BAM`). There is no apparent curvature or oscillatory feature except for the large-mass-ratio region, where the $L'_{\rm peak}$ quantity plotted in Fig. 7 over-emphasizes any remaining features and the corresponding relative errors are below

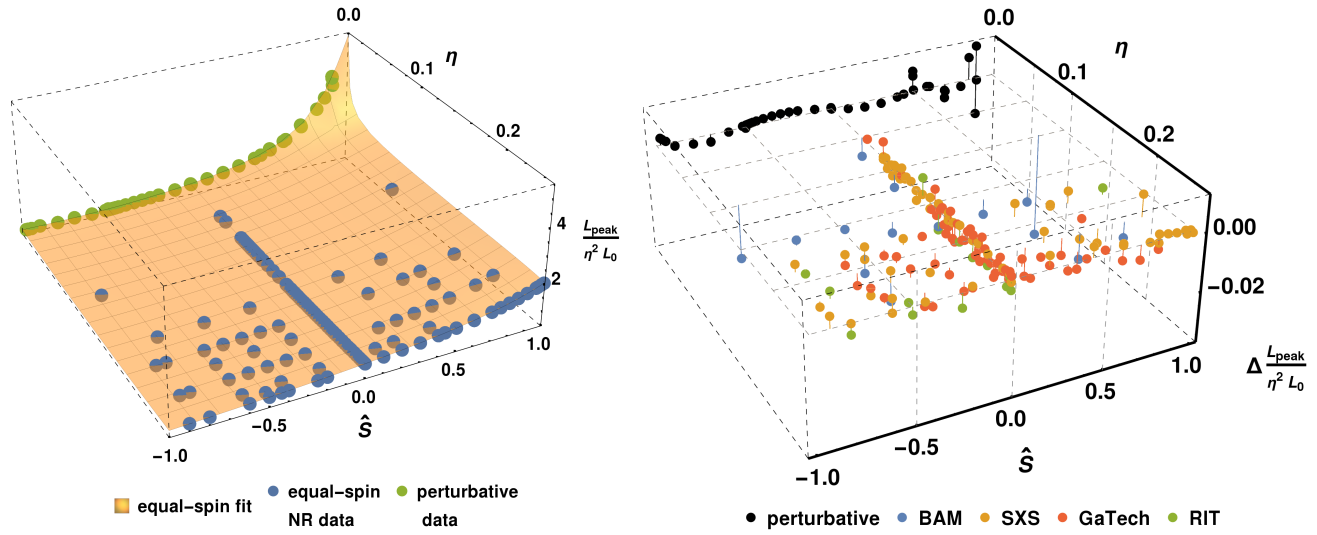

FIG. 7: Results of the two-dimensional equal-spin $L'_{\rm peak}\left(\eta, \widehat{S}\right)$ fit. First panel: comparison of the smooth fit surface with the equal-spin NR data and perturbative results; second panel: residuals over the parameter space, color-coded by data provenance.

0.5%. This accuracy is similar to that of the large-mass-ratio-only fits in Sec. III C, thus proving that the combined two-dimensional fit successfully captures both the shallow spin slope at similar masses and the steep slope in the perturbative regime. As discussed in appendix Sec. A 6, several outliers have been removed before the fit; the 2D fit still matches all equal-spin outliers to below 4% relative error.

As this equal-spin part of the full $L'_{\rm peak}\left(\eta, \widehat{S}, \Delta\chi\right)$ will be re-fitted, together with unequal-spin corrections, in the next and final step of the hierarchical procedure, we do not tabulate its best-fit coefficients at this point.

E. Unequal-spin contributions and 3D fit

Simply extending the 2D fit to the full 3D parameter space either by evaluating fit errors of the equal-spin-only calibrated fit over the whole data set, or by refitting the 2D ansatz from Eq. (8)), more than doubles the RMSE and induces oscillations at high $\left|\widehat{S}\right|$. But even for such a naive approach, relative errors are still limited to below 10%, so that the effects of unequal spins can evidently be treated as subdominant corrections. We follow here the same approach as in [1] to model spin-difference effects, constructing a 3D ansatz as

$$L'_{\rm peak}\left(\eta, \widehat{S}, \Delta\chi\right) = L'_{\rm peak}\left(\eta, \widehat{S}\right) + \Delta L'_{\rm peak}(\eta, \widehat{S}, \Delta\chi) \qquad (11)$$

where we choose the correction terms $\Delta L'_{\rm peak}$ with guidance from (i) PN analytical results and (ii) an analysis of the residuals of unequal-spin NR simulations under the 2D equal-spin fit.

Though PN cannot be expected to be quantitatively accurate for the late-inspiral and merger stages of BBH coalescence – where the peak luminosity emanates – it can still give some intuition on the qualitative shape of spin and spin-difference effects. The PN spin-orbit flux terms as given in Eq. (3.13) of Ref. [72] and Eq. (4.9) of Ref. [73] include linear terms in $\Delta\chi$ with an η-dependent prefactor that can be expressed as $\mathrm{P}(\eta)\sqrt{1-4\eta}$ with a polynomial $\mathrm{P}(\eta)$. The next-to-leading-order contributions would be quadratic in $\Delta\chi$ and a mixed term proportional to $\widehat{S}\,\Delta\chi$.

At equal masses ($\eta = 0.25$) BBH configurations are symmetric under relabeling of the component BHs, so that terms linear in $\Delta\chi$ must vanish; this is ensured by the $\sqrt{1-4\eta}$ factor, which we therefore expect both in the linear and the mixture term, but not in the quadratic term. Hence, we make the general spin-difference ansatz

$$\Delta L'_{\rm peak}\left(\eta, \widehat{S}, \Delta\chi\right) = A_1(\eta)\,\Delta\chi + A_2(\eta)\,\Delta\chi^2 + A_3(\eta)\,\widehat{S}\,\Delta\chi \qquad (12)$$

with a simple polynomial for $A_2(\eta)$ and $A_1(\eta)$, $A_3(\eta)$ both being a polynomial multiplied by the symmetry factor.

To check that these up to three terms accurately describe our available set of 215 unequal-spin NR cases, and to get a handle on the functions $A_i(\eta)$, we visually inspect the data in steps of fixed mass ratio with sufficient numbers of data points. Examples for $q = 1$ and $q = 3$ are shown in Fig. 8. For each of these steps, $q = \{1, 1.33, 1.5, 1.75, 2, 3, 4, 5, 6, 8\}$, we compute the residuals under the non-spinning fit from Eq. (8), then perform four fits in $\Delta\chi$: linear, linear+quadratic, linear+mixed, or the sum of all three terms. Fits of the collected coefficients, as functions of η, give estimates of the functions $A_i(\eta)$, as displayed with the "per-mass-ratio data" points and "per-mass-ratio fit" lines in Fig. 9.

This procedure yields clear evidence for a linear spin-difference term and some preference for both second-order terms, though the data is too noisy to constrain their η-dependence very well. For example, there is an apparent sign switch in the linear term at mass ratio $q = 4$ ($\eta = 0.16$), which is most likely due to a combination of the 2D fit being relatively weakly constrained in this region and non-negligible errors in some of the unequal-spin data points, which however cannot easily be discarded as outliers. The overall fits in η are reasonably robust against such problems, but a full model selection of $A_i(\eta)$ is clearly not the right thing to do at this point without a much more detailed understanding of the point-by-point data quality. Hence, we make very simple choices for the $A_i(\eta)$ with just one power of η each:

$$A_1(\eta) = d_{10}\sqrt{1-4\eta}\eta^3 \qquad (13a)$$

$$A_2(\eta) = d_{20}\eta^3 \qquad (13b)$$

$$A_3(\eta) = d_{30}\sqrt{1-4\eta}\eta^3\,, \qquad (13c)$$

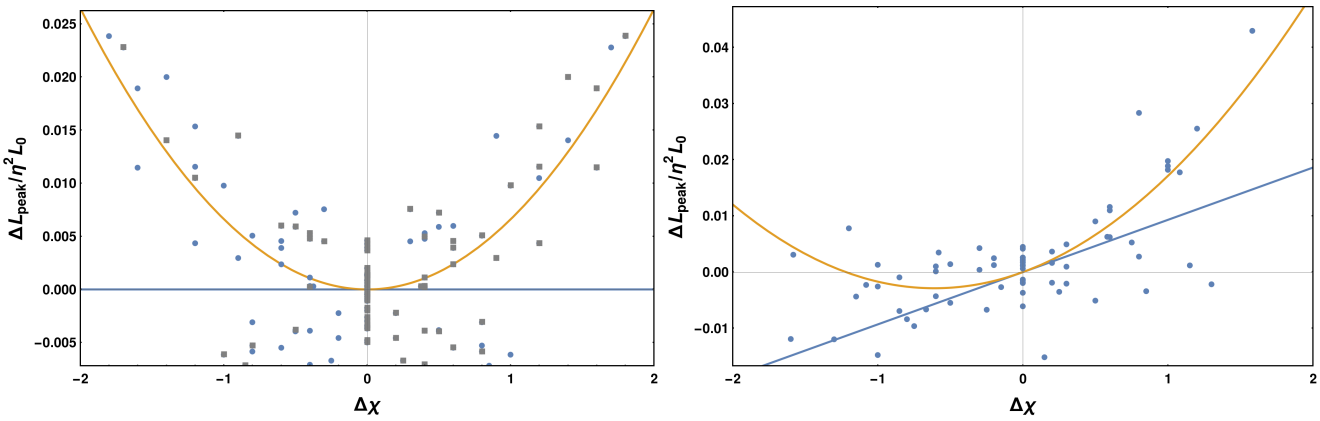

FIG. 8: Examples of spin-difference behavior at fixed mass ratios, for scaled NR data $L'_{\rm peak}$ after subtraction of the 2D $\left(\eta, \widehat{S}\right)$ fit. First panel: $q = 1$ with linear term vanishing due to symmetry and mainly quadratic dependence; points in gray are mirror duplicates exploiting the $\chi_1 \leftrightarrow \chi_2$ symmetry at equal masses. Second panel: $q = 3$ where the linear term dominates and the apparent quadratic dependence likely is noise-dominated.

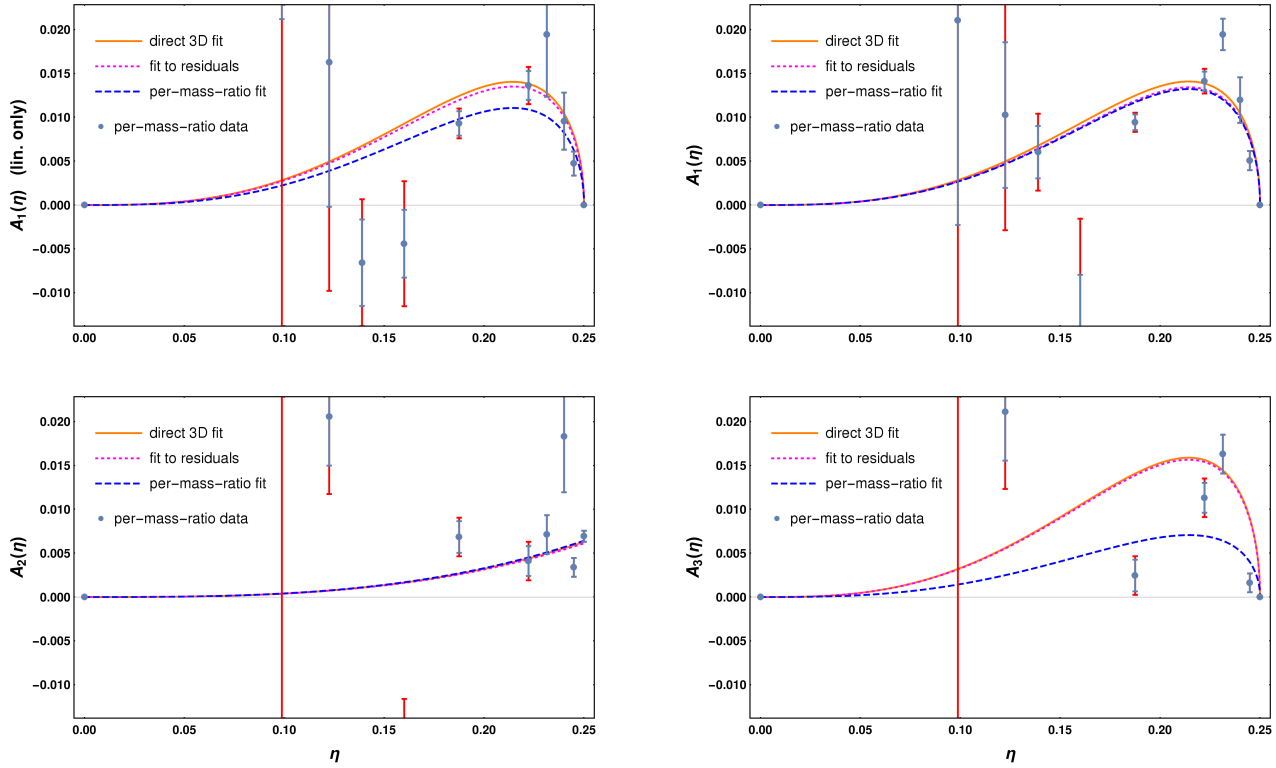

FIG. 9: Spin-difference behavior of the scaled NR luminosities $L'_{\rm peak}$ after subtraction of the 2D $\left(\eta, \widehat{S}\right)$ fit over mass ratio η, showing the results of fits as in Fig. 8 at η steps corresponding to $q = \{1, 1.33, 1.5, 1.75, 2, 3, 4, 5, 6, 8\}$ and three estimates for the ansatz functions $A_i(\eta)$ from Eq. (13): (i) unequal-spin part of the final 3D fit from Eq. (14) ("direct 3D fit"), (ii) fit of the unequal-spin terms from Eq. (13) ("fit to residuals") to the residuals of the 2D fit from Eq. (8) over all mass ratios, (iii) fits of Eq. (13) to the per-mass-ratio results. Top-left panel: linear term A_1 only. The remaining panels are for the combined linear+quadratic+mixture fit, in clock-wise order: linear term A_1, quadratic term A_2 and mixture term A_3. $A_1(\eta)$ from the combined ansatz is very similar to the linear-only fit, demonstrating its robustness. Error bars for the per-mass-ratio points include components from the fit uncertainty at that ratio (blue) and the average data weight of the contributing NR cases (red). At the lowest η, points lie outside the plot range, but are so uncertain that they do not contribute significantly to the total fit. The direct-3D and residuals-only results are consistent, while the per-mass-ratio analysis only matches them qualitatively, which is however sufficient since it was only used to investigate the possible shapes of $A_i(\eta)$.

and investigate how much improvement this can yield over the 2D fit. We now use the full data set except for the 1D subspaces (307 data points, including 265 NR simulations) to fit the full 3D ansatz from Eq. (11), with the equal-spin and spin-difference contributions from Eqs. (8) and (12)+(13), respectively. The sets of coefficients a_i, b_i and f_{i2} are already fixed from the 1D fits and consistency constraints (see Tbl. I, Tbl. II and Eq. (10)), leaving between 11 and 14 free coefficients in this final 3D stage. When including all three spin-difference terms, the full ansatz (with the constraints from Eq. (10) for the f_{i2} still to be applied) is:

$$L'_{\rm peak}\left(\eta,\widehat{S},\Delta\chi\right) = a_5\eta^5 + a_4\eta^4 + a_3\eta^3 + a_2\eta^2 + a_1\eta + a_0 \tag{14}$$
$$+\frac{0.107b_2\widehat{S}^2\left(f_{22}\eta^2+f_{21}\eta+f_{20}\right)+0.465b_1\widehat{S}\left(f_{12}\eta^2+f_{11}\eta+f_{10}\right)+\widehat{S}^4\left(f_{42}\eta^2+f_{41}\eta+f_{40}\right)+\widehat{S}^3\left(f_{32}\eta^2+f_{31}\eta+f_{30}\right)}{-0.328b_4\widehat{S}\left(f_{62}\eta^2+f_{61}\eta+f_{60}\right)+\widehat{S}^2\left(f_{72}\eta^2+f_{70}\right)+1.0}$$
$$+d_{20}\,\eta^3\,(\chi_1-\chi_2)^2 + d_{10}\,\sqrt{1-4\eta}\,\eta^3\,(\chi_1-\chi_2) + d_{30}\,\sqrt{1-4\eta}\,\eta^3\,\widehat{S}\,(\chi_1-\chi_2)\ .$$

We consider residuals and information criteria, summarized in Tbl. III, to check which spin-difference terms are actually supported by the data. These rankings depend on the specific choice of $A_i(\eta)$, but with the current parameter-space coverage and understanding of NR data quality, the main goal is to find general evidence for spin-difference effects and a general idea of their shape, not to exactly characterize them. With the choices made in Eq. (13), we find a 14-coefficient fit with linear+quadratic+mixture corrections that has well-constrained coefficients (see Tbl. IV), is evidently preferred in terms of AICc and BIC, and reduces overall residuals by about 20% in RMSE. Different choices for the powers of η in Eq. (13) yield compatible results, while polynomials in η with several free coefficients tend to produce under-constrained fits.

	$N_{\rm data}$	$N_{\rm coeff}$	RMSE	AICc	BIC
1D η	84	6	2.81×10^{-3}	−817.101	−801.176
1D $\hat{S}$	32	3	2.42×10^{-3}	−285.789	−280.783
2D ($\chi_1=\chi_2$)	92	11	5.65×10^{-3}	−751.73	−724.77
2D all	307	11	1.67×10^{-2}	−1914.19	−1870.36
3D lin	307	12	1.51×10^{-2}	−2007.95	−1960.56
3D lin+quad	307	13	1.39×10^{-2}	−2134.23	−2083.30
3D lin+mix	307	13	1.41×10^{-2}	−2082.59	−2031.66
3D lin+quad+mix	307	14	1.36×10^{-2}	−2157.80	−2103.34

TABLE III: Summary statistics for the various steps of the hierarchical fit. Note that it is not meaningful to compare AICc and BIC between datasets of different sizes. There is preference for the 3D fit including all three linear+mixture+quadratic terms, although many different choices of the $A_i(\eta)$ ansatz functions yield similar results with just ± a few percent in RMSE and ± a few in AICc/BIC, so that the shape of these terms is not yet strongly constrained.

	estimate	std.err.	rel.err.[%]
d_{10}	3.79	0.282	7.5
d_{20}	0.402	0.0439	10.9
d_{30}	4.27	0.84	19.7
f_{10}	1.63	0.0116	0.7
f_{11}	−3.63	0.229	6.3
f_{20}	31.7	1.32	4.2
f_{21}	−274.	28.6	10.4
f_{30}	−0.235	0.0109	4.7
f_{31}	6.96	0.438	6.3
f_{40}	0.211	0.0224	10.6
f_{41}	1.53	0.452	29.6
f_{60}	3.09	0.0439	1.4
f_{61}	−16.7	1.67	10.0
f_{70}	0.836	0.0231	2.8

TABLE IV: Fit coefficients for the final 3D fit stage, cf. Eq. (14).

IV. FIT ASSESSMENT

In this section, we assess in some detail the properties and statistical quality of the new three-dimensional peak luminosity fit, with the actual non-rescaled luminosity (in geometric units of $G = c = M = 1$) obtained as $\eta^2\,L_0\,L'_{\rm peak}\left(\eta,\widehat{S},\Delta\chi\right)$.

We compare with our previous fit [10] used for LIGO parameter estimation during O1 [4–7, 13], which used a much smaller calibration set of 89 BAM and SXS simulations, only modes up to $\ell_{\rm max}=4$ and no extreme-mass-ratio constraints; and with the recent Healy&Lousto fit [26] based on 107 RIT simulations, using modes up to $\ell_{\rm max}=6$. We attempt to present a fair comparison by analyzing NR and perturbative large-mass-ratio results separately, and also consider the improvement from refitting the unmodified ansätze of [10, 26] to the present NR data set.

A. Residuals and information criteria

In Fig. 10 we show the distribution of residuals for the 3D fit in $L'_{\rm peak}$ projected to the $\left(\eta,\widehat{S}\right)$ parameter space, so that it can be compared to the 2D results in Fig. 7. The strongest visible outliers in this scaling are at low η and correspond to mild actual deviations; of at most a 7% relative error at $q=18$, with 417 of the 423 data points below 3% relative error.

For a comparison with the two previous fits, we first concentrate on the 378 NR simulations only and revisit large mass ratios in Sec. IV B. In Fig. 11 we show histograms of the residuals in $L_{\rm peak}$ for the three fits over this data set, demonstrating that the new fit achieves a narrower distribution. As listed in Tbl. V, the standard deviation of residuals is only half of that for our previous fit and three times lower than for the RIT fit. With a mean offset by only a ninth of a standard deviation, there is no evidence for bias, though that notion is notoriously ambiguous for a data set that samples the parameter space non-uniformly.

The same table contains AICc and BIC values evaluated over the same NR-only data set, which both find a very significant preference for the new fit. Note that, being computed over a different data selection and for $L_{\rm peak}$ instead of $L'_{\rm peak}$, these values are not directly comparable with the previous Tbl. III. Since we have removed 41 NR cases from the full available data set (see appendix Sec. A), it is advisable to check that the statistical preference still holds when including these in the evaluation set. Indeed, the reduction in standard deviations of residuals is then less against the T1600018 and RIT fits, but still roughly 20% and 30%, and there is still a preference of several hundreds in both information criteria.

We also show results for re-fitting the T1600018 and RIT ansätze to the present NR+perturbative data set, with the

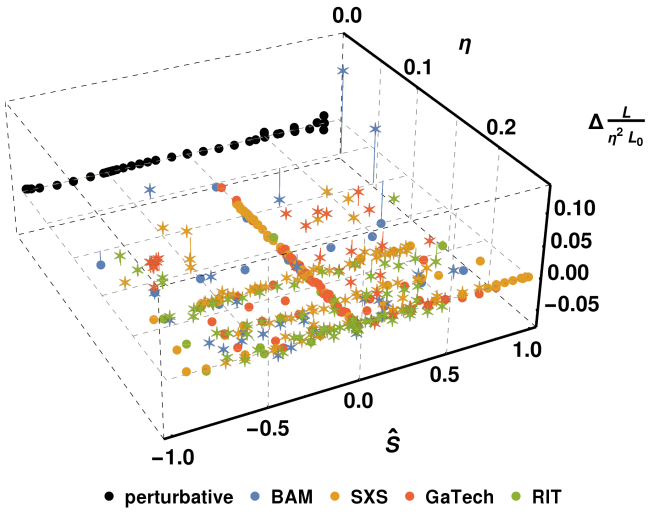

FIG. 10: Residuals of the final 3D peak luminosity fit, for NR and perturbative large-mass-ratio data, projected to the 2D parameter space of η and $\widehat{S}$. The data sets are distinguished by colors, and unequal-spin points highlighted with stars.

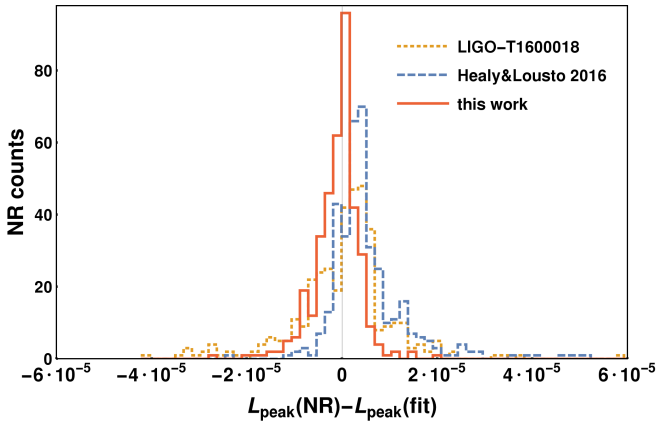

FIG. 11: Fit residuals of the final 3D peak luminosity fit compared with previous fits [10, 26], evaluated over the set of 378 NR simulations shown in Fig. 1. 6 outliers for Healy&Lousto with |NR − fit| > 0.00006 are outside of the plot range.

	$N_{\rm coef}$	mean	stdev	AICc	BIC
LIGO-T1600018	11	3.0×10^{-7}	1.0×10^{-5}	−7732.1	−7685.6
(refit)	11	-1.8×10^{-6}	4.0×10^{-5}	−6706.0	−6659.5
HealyLousto 2016	19	6.9×10^{-6}	1.5×10^{-5}	−7225.5	−7148.9
(refit)	19	-4.9×10^{-7}	1.0×10^{-5}	−7708.3	−7631.7
this work	23	-9.8×10^{-7}	4.8×10^{-6}	−8298.1	−8206.7
(refit)	23	-5.5×10^{-7}	4.8×10^{-6}	−8323.6	−8232.3

TABLE V: Summary statistics for the final 3D peak luminosity fit compared with previous fits [10, 26], evaluated over the 378 NR simulations shown in Fig. 1. The new fit has a total of 23 free coefficients, corresponding to tables I, II and IV. We also show results for re-fitting the three ansätze to the full NR+large-mass-ratio data set, again evaluating the statistics over NR only.

statistics then again evaluated over NR data only. Our old ansatz with only 11 coefficients is not well suited to matching the large-mass-ratio region and the large unequal-spin population in the NR data set, and its refitted version performs worse than the original. On the other hand, the RIT ansatz with 19 coefficients was only weakly constrained in the original [26] version fitted to 107 simulations, with large errors on several

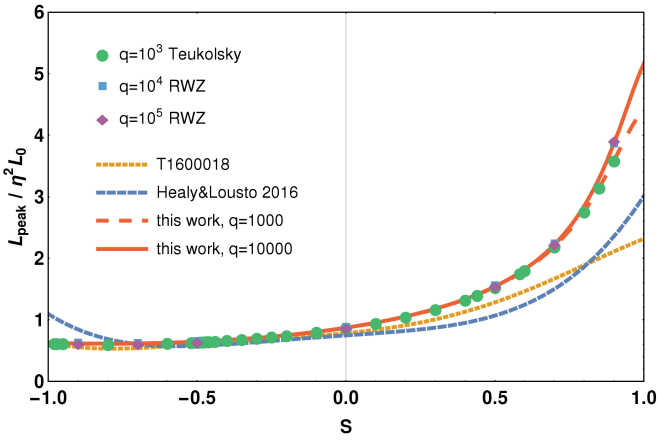

FIG. 12: Comparison between perturbative large-mass-ratio data and the three fits from Refs. [10, 26] and this work, reproducing the same data as in Fig. 6 but with projections of the full NR-calibrated fits. The T160018 and RIT fits are essentially converged at $q = 10^3$ (e.g. only 0.4% change at $S = 1.0$ for the RIT fit going to $q = 10^4$), and the visually identical lines for higher q not shown.

fit coefficients, but improves now significantly through the refit. Yet, it does not achieve the same level of accuracy as the new ansatz and fit developed in this paper.

As a test of robustness, we also perform a similar refit of our final hierarchically-obtained ansatz directly using the full data set, instead of using the constraints from the 1D subsets. This produces somewhat better summary statistics, but it also allows uncertainties from the less well-controlled unequal-spin set to influence the non-spinning part of the fit. The more conservative approach is to have the non-spinning part of the fit calibrated only to the corresponding data subset, as done in Sec. III A; so that we recommend the stepwise fit, with coefficients as reported in Tables I, II and IV, for further applications.

B. Large-mass-ratio and extremal-spin limits

In Fig. 12, we compare our full 3D fit with the perturbative large-mass-ratio data and find that it correctly reproduces the behavior it is meant to be constrained to. The T1600018 fit did not predict the steep rise for positive spins, and while at negative spins it matches the shape roughly, it is still off by about 10% in that region. The RIT fit disagrees with the perturbative data at high spin magnitudes, either negative or positive, and does not reproduce the increasing steepness for even higher mass ratios.

The clearest difference between this fit and the previous ones in the NR-dominated region is for high aligned spins, which is shown in Fig. 13 for the extremal spin limit, $\chi_1 = \chi_2 = \widehat{S} = 1$. The RIT fit estimates a lower luminosity at equal masses, but higher values at $\eta < 0.16$ before approaching the $\eta \to 0$ limit rather flatly, as discussed before. Our older fit and the new one roughly agree at similar masses, but in the lower panel with the rescaled $L'_{\rm peak}$ it is obvious that the previous fit did not anticipate the steep $\eta \to 0$ limit that we are now implementing through fitting the perturbative data.

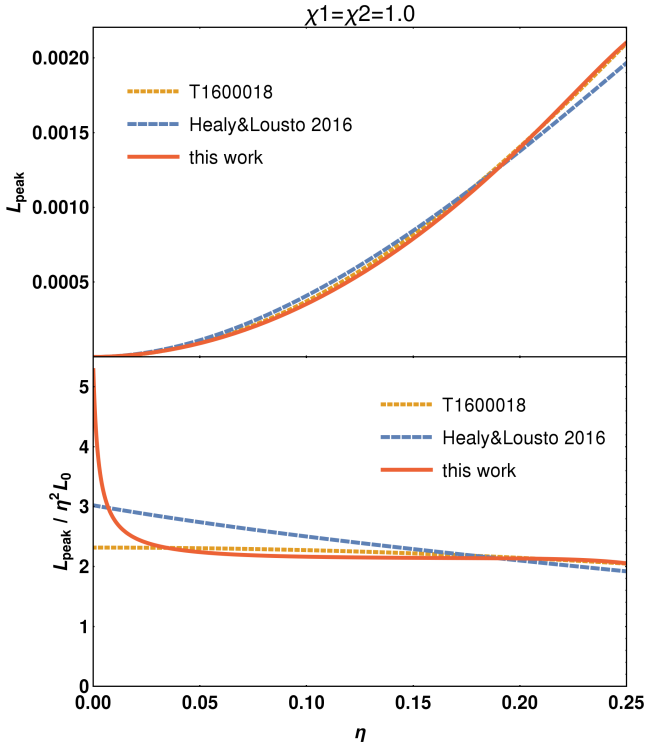

FIG. 13: Behavior of the full 3D fit (11) in the extremal-spin limit, $\chi_1 = \chi_2 = \widehat{S} = 1$, as a function of mass ratio η; again compared with the fits from Refs. [10, 26]. Top panel: in terms of physical peak luminosity $L_{\rm peak}$, lower panel: in terms of rescaled $L'_{\rm peak} = L_{\rm peak}/\eta^2 L_0$.

V. CONCLUSIONS

Using the hierarchical analysis approach to the three-dimensional parameter space of nonprecessing quasicircular binary black hole (BBH) coalescences introduced in [1], we have developed a new model for the peak of the gravitational-wave luminosity of BBH coalescence events, $L_{\rm peak}$. This model fit is based on the largest-yet combined set of Numerical Relativity results from four independent simulation codes, as well as on perturbative numerical data for the large-mass-ratio regime that cannot currently be probed by NR.

The result that BBHs are, for a brief moment during their merger, the most powerful astrophysical events is already clear from simple dimensional analysis, as is the rough scaling of this peak luminosity with the initial binary parameters. Yet, only detailed NR-calibrated fits allow for a precise understanding of the parameter-space dependence of $L_{\rm peak}$. Our new fit significantly reduces the residuals for most available NR cases in comparison with a previous version of this fitting procedure [10] used in LIGO O1 data analysis and an alternative fit [26], both calibrated to much smaller data sets.

We also characterize the quality of the luminosity data set considering various sources of NR inaccuracies and the compatibility between different simulation codes, finding that the peak luminosity's subdominant parameter dependencies are of a similar or even smaller order than typical discrepancies between simulations. This limits the level of detail to which we can model spin-difference effects, though we can still improve over an equal-spin-only fit and find that the spin-difference dependence qualitatively matches expectations. These statistical improvements, wider parameter-space coverage and systematic understanding of sources of uncertainty can make the new fit a useful ingredient for future parameter estimation studies of BBH events.

The final fit ansatz is given in Eq. (14), with coefficient estimates listed in Tables I, II and IV. Example implementations of this fit for Mathematica and python are available as supplementary material for this paper, along with an ASCII table of the data set. The python implementation is equivalent to that included in the free software LALInference [74] package.

As more GW detections are made, there will be more opportunities to infer the luminosities of stellar-mass BBH systems. In particular, the aLIGO observatories in the USA [2, 3], Advanced Virgo [75] in Italy, and forthcoming observatories in Japan [76, 77] and India [78] are poised to become pivotal tools for earth-based GW astronomy, eventually enabling daily BBH detections [79] over a wide parameter range.

The accuracy of the fit presented in this paper should be sufficient for the expected sensitivity during the second aLIGO observing run and for 'vanilla' BBH events (similar masses, low spins, no strong precession), with sampling uncertainties in mass ratio and spins still dominating over fit errors. Still, a continued expansion of the NR calibration set and an improved understanding of higher-mode contributions, precession and the transition from similar-mass to extreme-mass-ratio regimes will be important to improve the understanding of BBH peak luminosities and waveforms.

Meanwhile, this project of fitting peak luminosities is an important step in extending the 'Phenom' waveform family [14–18], as our analysis of higher-mode contributions and the demonstration of joint calibration to NR and perturbative large-mass-ratio data can form the basis for improved modeling of full inspiral-merger-ringdown waveforms.

Acknowledgments

DK, XJ and SH were supported by the Spanish Ministry of Economy and Competitiveness grants FPA2010-16495, CSD2009-00064 and FPA2013-41042-P, European Union FEDER funds, the Conselleria d'Economia i Competitivitat del Govern de les Illes Balears and the Fons Social Europeu; and DK recently also by the EU H2020-MSCA-IF-2015 grant 704094 GRANITE. LL was supported by Science and Technology Facilities Council (ST) grants ST/I001085/1 and ST/H008438/1. MH was supported by ST/I001085/1 and ST/H008438/1 and European Research Council Consolidator Grant 647839, and MP by ST/I001085/1. VC thanks ICTS for support during the S.N. Bhatt Memorial Excellence Fellowship Program 2014. `BAM` simulations were carried out at Advanced Research Computing (ARCCA) at Cardiff, as part of the European PRACE petascale computing initiative on the clusters Hermit, Curie and SuperMUC, on the UK DiRAC Datacentric cluster and on the BSC MareNostrum computer under PRACE and RES (Red Española de Supercomputación) allocations. We thank the CBC working group of the LIGO Scientific Collaboration, and especially Ilya Mandel, Nathan Johnson-McDaniel, Alex Nielsen, Christopher Berry, Ofek Birnholtz and Juan Calderon Bustillo for discussions of the general fitting method, the previous results described in [10], as well as the current results and manuscript. This paper has been assigned document number LIGO-P1600279-v5.

[1] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer (2016), 1611.00332.
[2] J. Aasi et al. (LIGO Scientific Collaboration), Class. Quant. Grav. **32**, 074001 (2015), 1411.4547.
[3] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 131103 (2016), 1602.03838.
[4] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016), 1602.03837.
[5] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016), 1606.04855.
[6] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. **X6**, 041015 (2016), 1606.04856.
[7] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 241102 (2016), 1602.03840.
[8] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **116**, 221101 (2016), 1602.03841.
[9] D. D. Frederiks et al., Astrophys. J. **779**, 151 (2013), 1311.5734.
[10] X. Jiménez Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, L. London, and M. Pürrer, Tech. Rep. LIGO-T1600018-v4 (2016), URL https://dcc.ligo.org/LIGO-T1600018-v4/public.
[11] J. Aasi et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. **D88**, 062001 (2013), 1304.1775.
[12] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, et al., Phys. Rev. **D91**, 042003 (2015), 1409.7215.
[13] B. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. **X6**, 041014 (2016), 1606.01210.
[14] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014), 1308.3271.
[15] A. Bohé, M. Hannam, S. Husa, F. Ohme, M. Pürrer, and P. Schmidt, Tech. Rep. LIGO-T1500602, LIGO Scientific Collaboration and Virgo Collaboration (2016), URL https://dcc.ligo.org/T1500602.
[16] A. Bohé, M. Pürrer, M. Hannam, S. Husa, F. Ohme, P. Schmidt, et al., in preparation.
[17] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, Phys. Rev. **D93**, 044006 (2016), 1508.07250.
[18] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, Phys. Rev. **D93**, 044007 (2016), 1508.07253.
[19] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. **D89**, 084006 (2014), 1307.6232.
[20] M. Pürrer, Class. Quant. Grav. **31**, 195010 (2014), 1402.4146.
[21] M. Pürrer, Phys. Rev. **D93**, 064041 (2016), 1512.02248.
[22] S. Babak, A. Taracchini, and A. Buonanno (2016), 1607.05661.
[23] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration) (2016), 1611.07531.
[24] J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. **D90**, 104004 (2014), 1406.7295.
[25] F. Hofmann, E. Barausse, and L. Rezzolla, Astrophys. J. **825**, L19 (2016), 1605.01938.
[26] J. Healy and C. O. Lousto (2016), 1610.09713.
[27] J. G. Baker, W. D. Boggs, J. Centrella, B. J. Kelly, S. T. McWilliams, and J. R. van Meter, Phys. Rev. **D78**, 044046 (2008), 0805.1428.
[28] A. Nagar, T. Damour, and A. Tartaglia, Class. Quant. Grav. **24**, S109 (2007), gr-qc/0612096.
[29] S. Bernuzzi, A. Nagar, and A. Zenginoğlu, Phys. Rev. **D84**, 084026 (2011), 1107.5402.
[30] E. Harms, S. Bernuzzi, A. Nagar, and A. Zenginoğlu, Class. Quant. Grav. **31**, 245004 (2014), 1406.5983.
[31] P. Amaro Seoane et al., GW Notes **6**, 4 (2013), 1201.3621.
[32] P. Amaro Seoane et al. (eLISA Consortium) (2013), 1305.5720.
[33] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme, et al., Phys. Rev. **D93**, 024003 (2016), 1511.05581.
[34] M. Rajagopal and R. W. Romani, Astrophys. J. **446**, 543 (1995), astro-ph/9412038.
[35] A. Sesana, A. Vecchio, and M. Volonteri, Mon. Not. Roy. Astron. Soc. **394**, 2255 (2009), 0809.3412.
[36] A. Sesana and A. Vecchio, Phys. Rev. **D81**, 104008 (2010), 1003.0677.
[37] B. Bruegmann, J. A. Gonzalez, M. Hannam, S. Husa, U. Sperhake, and W. Tichy, Phys. Rev. **D77**, 024027 (2008), gr-qc/0610128.
[38] S. Husa, J. A. Gonzalez, M. Hannam, B. Bruegmann, and U. Sperhake, Class. Quant. Grav. **25**, 105006 (2008), 0706.0740.
[39] A. H. Mroué, M. A. Scheel, B. Szilágyi, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, S. Ossokine, N. W. Taylor, et al., Phys. Rev. Lett. **111**, 241104 (2013), 1304.6077.
[40] SXS Collaboration, *SXS Gravitational Waveform Database* (2016), URL http://www.black-holes.org/waveforms/.
[41] SXS Collaboration, *SpEC: Spectral Einstein Code* (2016), URL http://www.black-holes.org/SpEC.html.
[42] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, Class. Quant. Grav. **33**, 204001 (2016), 1605.03204.
[43] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, *Georgia Tech catalog of binary black hole simulations* (2016), URL http://www.einstein.gatech.edu/catalog/.
[44] F. Herrmann, I. Hinder, D. Shoemaker, and P. Laguna, Class. Quant. Grav. **24**, S33 (2007), gr-qc/0601026.
[45] B. Vaishnav, I. Hinder, F. Herrmann, and D. Shoemaker, Phys. Rev. **D76**, 084020 (2007), 0705.3829.
[46] J. Healy, J. Levin, and D. Shoemaker, Phys. Rev. Lett. **103**, 131101 (2009), 0907.0671.
[47] L. Pekowsky, R. O'Shaughnessy, J. Healy, and D. Shoemaker, Phys. Rev. **D88**, 024040 (2013), 1304.3176.
[48] Y. Zlochower, J. G. Baker, M. Campanelli, and C. O. Lousto, Phys. Rev. **D72**, 024021 (2005), gr-qc/0505055.
[49] T. W. Baumgarte and S. L. Shapiro, Phys. Rev. **D59**, 024007 (1998), gr-qc/9810065.
[50] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006), gr-qc/0511048.
[51] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. Lett. **96**, 111102 (2006), gr-qc/0511103.
[52] M. Boyle, Phys. Rev. **D87**, 104006 (2013), 1302.2919.
[53] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer (2014), 1409.4431.
[54] M. Boyle, Phys. Rev. **D93**, 084031 (2016), 1509.00862.
[55] C. Reisswig and D. Pollney, Class. Quant. Grav. **28**, 195015 (2011), 1006.1632.
[56] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, Astrophys. J. **178**, 347 (1972).
[57] S. A. Teukolsky and W. H. Press, Astrophys. J. **193**, 443 (1974).
[58] R. Fujita, PTEP **2015**, 033E01 (2015), 1412.5689.
[59] S. Bernuzzi and A. Nagar, Phys. Rev. **D81**, 084056 (2010), 1003.0597.
[60] T. Damour and A. Nagar, Phys. Rev. **D76**, 064028 (2007),

0705.2519.
[61] T. Damour, B. R. Iyer, and A. Nagar, Phys. Rev. **D79**, 064004 (2009), 0811.2069.
[62] Y. Pan, A. Buonanno, R. Fujita, E. Racine, and H. Tagoshi, Phys. Rev. **D83**, 064003 (2011), [Erratum: Phys. Rev.D87,no.10,109901(2013)], 1006.0431.
[63] R. Fujita, Prog. Theor. Phys. **128**, 971 (2012), 1211.5535.
[64] A. Nagar and A. Shah, Phys. Rev. **D94**, 104017 (2016), 1606.00207.
[65] S. Bernuzzi, A. Nagar, and A. Zenginoğlu, Phys. Rev. **D83**, 064010 (2011), 1012.2456.
[66] T. Damour, A. Nagar, and S. Bernuzzi, Phys. Rev. **D87**, 084035 (2013), 1212.4357.
[67] A. Nagar, Phys. Rev. **D88**, 121501 (2013), 1306.6299.
[68] A. Nagar, E. Harms, S. Bernuzzi, and A. Zenginoğlu, Phys. Rev. **D90**, 124086 (2014), 1407.5033.
[69] H. Akaike, IEEE Transactions on Automatic Control **19**, 716 (1974).
[70] G. E. Schwarz, Annals of Statistics **6**, 461 (1978).
[71] A. R. Liddle, Mon. Not. Roy. Astron. Soc. **377**, L74 (2007), astro-ph/0701113.
[72] A. Bohé, S. Marsat, and L. Blanchet, Class. Quant. Grav. **30**, 135009 (2013), 1303.7412.
[73] S. Marsat, A. Bohé, L. Blanchet, and A. Buonanno, Class. Quant. Grav. **31**, 025023 (2014), 1307.6793.
[74] LIGO Scientific Collaboration, *LSC Algorithm Library - LALSuite*, free software, URL https://wiki.ligo.org/DASWG/LALSuite.
[75] F. Acernese et al. (VIRGO Collaboration), Class. Quant. Grav. **32**, 024001 (2015), 1408.3978.
[76] K. Somiya (KAGRA Collaboration), Class. Quant. Grav. **29**, 124007 (2012), 1111.7185.
[77] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA Collaboration), Phys. Rev. **D88**, 043007 (2013), 1306.6747.
[78] C. S. Unnikrishnan, Int. J. Mod. Phys. **D22**, 1341010 (2013), 1510.06059.
[79] J. Aasi et al. (LIGO Scientific Collaboration and Virgo Collaboration), Living Rev. Rel. **19**, 1 (2016), 1304.0670.
[80] M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky, Phys. Rev. **D76**, 124038 (2007), 0710.0158.
[81] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[82] F. J. Zerilli, Phys. Rev. Lett. **24**, 737 (1970).
[83] O. Sarbach and M. Tiglio, Phys. Rev. **D64**, 084016 (2001), gr-qc/0104061.
[84] O. Rinne, L. T. Buchman, M. A. Scheel, and H. P. Pfeiffer, Class. Quant. Grav. **26**, 075009 (2009), 0811.3593.
[85] H. Nakano, J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. **D91**, 104022 (2015), 1503.00718.

Appendix A: NR data investigations

As a first estimate of the *overall* accuracy of the peak luminosity data set, we study the differences between results from different codes for equal initial parameters. We then give additional details on the possible error sources listed in Sec. II A and on the properties of higher modes, and discuss the 41 cases not used in the calibration set.

1. Comparison between different codes

To analyze typical deviations between results from different NR codes, we identify simulations with initial BH parameters

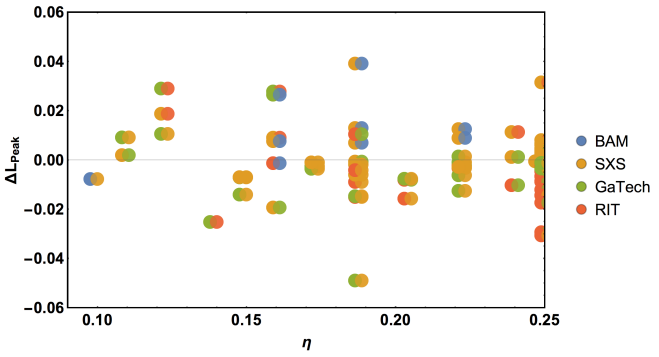

FIG. 14: Relative differences in the peak luminosity for equal-parameter configurations from different NR codes, shown against symmetric mass ratio η. Pairs of simulations are shown with a small horizontal offset for ease of visual identification.

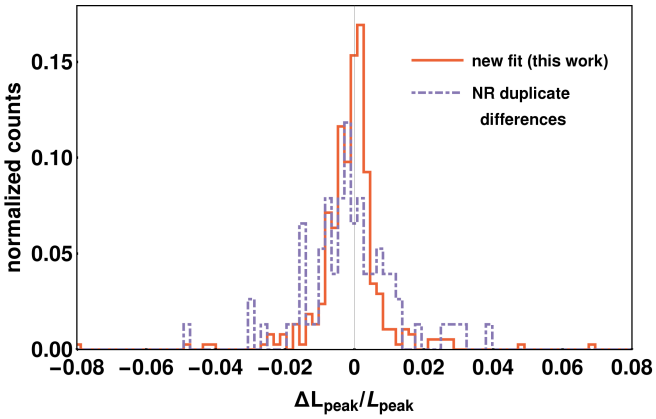

FIG. 15: Comparison of the distribution of relative fit errors (NR only, same set as in Fig. 11) and of differences between NR codes for equivalent parameters.

equal to within numerical accuracy, with a tolerance criterion

$$\left|\lambda_i - \lambda_j\right| \leq \epsilon = 0.0002 \text{ for } \lambda_i = \{\eta_i, \chi_{1i}, \chi_{2i}\} \,. \tag{A1}$$

This threshold was found in appendix A of [1] to be strict enough to reliably identify equivalent initial configurations, and is also tolerant enough to accomodate the minor relaxation of parameters after the initial 'junk' radiation which may be different between codes. In Fig. 14 we show the relative difference in $L_{\rm peak}$ between such matching cases, including the non-spinning $q = 4$ case where we have results from all four codes and a few triple coincidences. The set of these tuples is too sparse for clear conclusions on the parameter-space dependence of discrepancies between codes, though there might be some indication of increasing differences at large positive spins, which are particularly challenging to simulate due to increased resolution requirements for capturing the larger metric gradients in the near-horizon zone. We find many pairs with differences below 1%, but also several up to a few % even at not particularly challenging configurations.

This study gives a useful overall estimate of the possible error magnitude on the NR data set: while certainly many simulations are accurate to more than the few-% level, in general for any given simulation that does not have a paired case from another code, or at least nearby neighbors in parameter space, we cannot confidently assume that the errors will be low. This affects in particular the unequal-spin cases, where

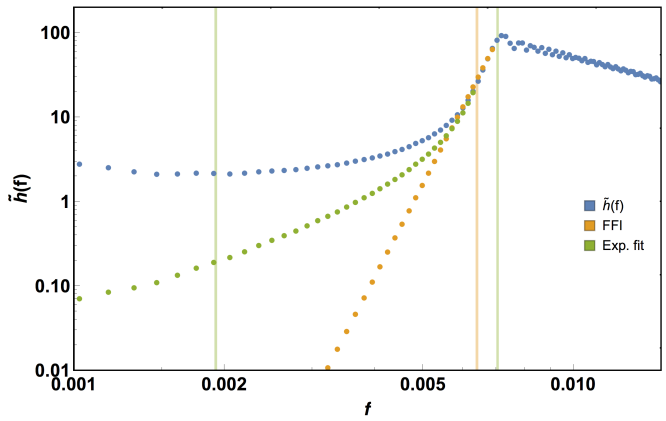

FIG. 16: Comparison of the FFI algorithm and the exponential-fit method for the low-frequency regime of $\widetilde{h}(f)$, for the example of the $q = 10$ non-spinning SXS waveform "SXS:BBH:0185". The green vertical lines indicate the local maximum and minimum of $\widetilde{h}(f)$ and the yellow line is at the tuned value f_0 for the FFI.

due to the much larger 3D parameter space very few duplicates exist. On the other hand, for equal spins – and particularly for the densely covered non-spinning or equal-mass subsets – we can use the duplicates analysis to make a very strict selection of calibration points, allowing the sub-percent calibration demonstrated in Secs. III A and III B. The specific decisions are detailed below in Sec. A 6.

As shown in the histograms of Fig. 15, the overall distribution of (relative) differences between equivalent configurations is of a similar width than that of the fit residuals. This demonstrates that we are indeed not overfitting the data, but also that one would need to characterize the accuracy of all NR cases to a significantly lower level to extract more information on subdominant effects.

2. Luminosity computation from ψ_4

Equations (1) and (2) describe the general computation of peak luminosities from the Weyl curvature component $\psi_4(t, \vec{r})$. This conversion is normally performed by either integrating $\psi_4(t, \vec{r})$ twice in time or by first applying a Fourier transform to the data, both to finally obtain the strain $h(t)$. However, both strategies for computing $h(t)$ carry the same technical issue: non-linear drifts in the final strain as a consequence of the characteristic low-frequency noise present when operating on finite segments of data.

This problem was already solved in [55] by means of the FFI algorithm, which we briefly describe here. One takes the Fourier-domain strain as

$$\widetilde{h}(f) = \begin{cases} -\frac{\widetilde{\psi_4}(f)}{f^2}, & \text{if} \quad f \geq f_{\rm max}\,, \\ -\frac{\widetilde{\psi_4}(f)}{f_0^2}, & \text{if} \quad f < f_{\rm max}\,. \end{cases} \tag{A2}$$

All physical frequency content must be contained in $[f_0, f_{\rm QNM}]$ where f_0 must be tuned close to the lowest physical frequency for a given mode and $f_{\rm QNM}$ is the quasi-normal mode frequency of the same mode. Thus, a proper selection of f_0 down-weights contributions from the low-frequency regime, driving these effects to zero – see Fig. 16.

As a consistency check, we have developed an alternative technique to convert from $\psi_4(t)$ to $h(t)$ which avoids the step of tuning f_0. In Fig. 16 we show an example of the Fourier transform of the dominant-mode strain $\widetilde{h}_{22}(f)$. In general, both local maxima and minima are located in the range $[f_{\rm min} \sim 0.5 f_0, f_{\rm max} \sim 1.2 f_0]$. The low-frequency behavior shown in the figure occurs for any $\widetilde{h}_{22}(f)$ independently of the system's physical parameters, being a consequence of the finiteness and discreteness of the time-domain waveforms. Empirically we found that the data in $[f_{\rm min}, f_{\rm max}]$ can be well fit with an exponential ansatz, which is then extended to all data in $[0, f_{\rm max}]$ and combined with the original data above $f_{\rm max}$:

$$\widetilde{h}_{\ell m}(f) = \begin{cases} -\frac{\widetilde{\psi}_{4,\ell m}(f)}{f^2}, & \text{if } f \geq f_{\rm max}\,, \\ a \exp^{-(b-f)} f^c e^{i\phi_{\ell m}}, & \text{if } f < f_{\rm max}\,, \end{cases} \tag{A3}$$

where $\phi_{\ell m}$ is the original ψ_4 phase. The split in fit coefficients a (amplitude) and b (peak position) is introduced here so that good starting values for the fit function can be picked more easily. With this approach, we smoothly drive the low-frequency noise to zero, eliminating non-physical artifacts in the Fourier-domain data.

We find that the difference between peak luminosities from the two different algorithms, when f_0 is optimally selected, is generally negligible, e.g. it is about $\sim 0.05\%$ in the example of the $q = 10$ non-spinning SXS waveform, and no significantly larger discrepancies have been found over the data set. So this effect is negligible for our analysis in comparison with other sources of uncertainty.

3. Extrapolation

The NR waveforms used in this paper are extracted at finite radii, which implies ambiguities, in particular due to gauge effects. We therefore extrapolate all waveforms to null infinity, where unambiguous waveforms can be defined. This allows us to assemble a consistent set of peak luminosity values for different codes, and to estimate the errors due to finite radius effects.

However, the extraction properties of the codes are not equal, and thus we have extrapolated the available waveforms following the following prescriptions:

- `BAM`: We have calculated $L_{\rm peak}$ at each finite radius and then performed a linear-in-$1/R$ extrapolation using only the well-resolved extraction radius. The maximum used for any case is $R \leq 180M$, but for some cases significantly fewer radii can be used for a robust extrapolation, depending on simulation grid resolutions.

- GaTech: $L_{\rm peak}$ is again calculated at finite radii and then extrapolated with a fit quadratic in $1/R$, only using up to $R \leq 100M$ because the slope generally changes for higher radii; this choice of extrapolation order and radius cut yields the most consistent results with other codes in the analysis of equivalent configurations.

- SXS: These waveforms are already provided at 2nd, 3rd and 4th order polynomial extrapolation, and we compute $L_{\rm peak}$ from these data products, after a correction [52–54] for center-of-mass drift, using the 2nd order extrapolation as the preferred value following [80]. We use waveforms based on the Weyl-scalar ψ_4, but

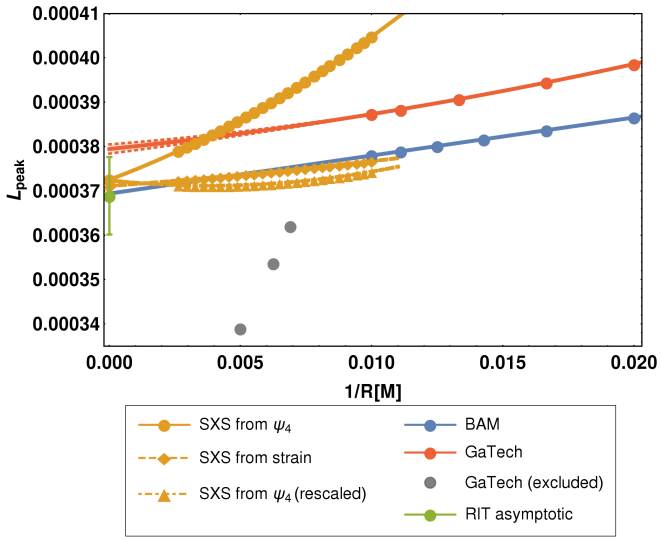

FIG. 17: $q = 4$ non-spinning example of extrapolation from finite radii for BAM, SXS and GaTech, with second-order fits for SXS and GaTech and linear for BAM; as well as the RIT value extrapolated with the method of Ref. [85]. In this case we find consistent values from BAM, SXS and RIT, with the GaTech case an outlier. The $R > 100$M GaTech data would make the trend more inconsistent, and are excluded from extrapolation. SXS luminosities computed from strain, or from ψ_4 but with the α^4 rescaling discussed in the text, show a flatter finite-R behavior more similar to the other codes, and extrapolated values consistent with the luminosity from ψ_4.

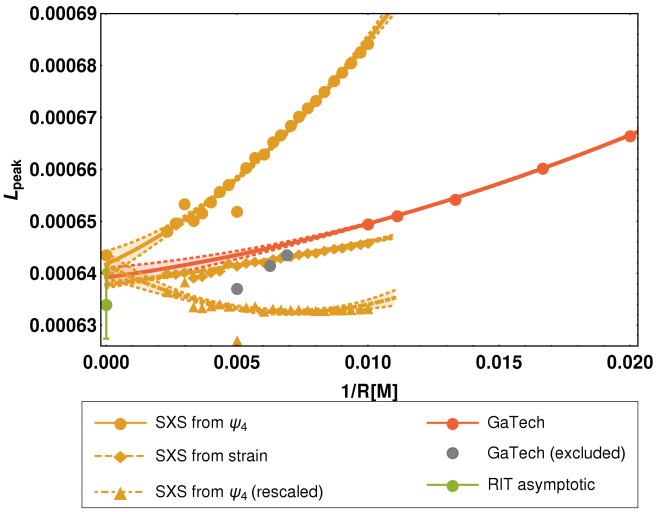

FIG. 18: $q = 2.5$ non-spinning example of extrapolation behavior for SXS and GaTech, compared with the extrapolated RIT value. In this case we find consistent GaTech and RIT values, with the asymptotic SXS computed from ψ_4 a marginal outlier. Extrapolating the finite-radius peak luminosities computed from ψ_4 with rescaling, or from strain, seems to improve consistency in this example, which however needs further study before applying it to the whole data set.

also compare with waveforms based on a computation of the strain. The SXS ψ_4 data use a definition of null-tetrad which is different from their Regge-Wheeler-Zerilli strain data [81–84], and from the definition used in the other codes. For the luminosity this difference corresponds to an overall scaling factor of the lapse function to the fourth power as a consequence of the difference between Eqs. (30-33) in [37] and Eqs. (11-12) in [80]. A rough correction for the different tetrad scaling used to compute the Weyl-scalar ψ_4 is to multiply it by α^4, with $\alpha = 1 - 2M/R$, where M is the final mass and R is an approximation to the luminosity distance using the standard relation between luminosity distance and isotropic radial coordinate for the Schwarzschild spacetime (compare also with the analysis in [85]). A comparison of luminosities computed from SXS ψ_4, strain, and heuristically rescaled ψ_4 with data from other codes is included in Figs. 17 and 18.

- RIT: The luminosity data provided in [26] uses the extrapolation method of [85].

In Fig. 17 we show the only configuration, the non-spinning $q = 4$ case, for which we have data from all four codes. This includes peak luminosities computed from the finite-radius strain data available as additional data products from SXS to cross-check the pre-extrapolated value. We see that extrapolation for $R \to \infty$ reduces discrepancies in $L_{\rm peak}$ between the different codes, but cannot completely alleviate it in this case. Another similar example is shown in Fig. 18 for a $q = 2.5$ non-spinning configuration where we have three simulations from SXS, GaTech and RIT, with the GaTech and RIT values more consistent in this case.

As extrapolations for BAM, SXS and GaTech are based on fits, their uncertainty can be estimated by the standard deviation on the intersection parameter (equivalent to the confidence interval on the extrapolation to $1/R = 0$). For the plotted non-spinning $q = 4$ case, these are much smaller than the remaining differences between the results, while for the $q = 2.5$ the uncertainties are almost wide enough to make the results marginally consistent. For some other cases, these uncertainties can reach up to a few %, especially when we want to be conservative and take the maximum of (i) the statistical uncertainty for the standard extrapolation-order choice and (ii) the difference between this and the closest alternative order. In general, such an uncertainty estimate cannot provide information about any systematics present in the data from different codes, and indeed for example we find that for BAM the purely statistical extrapolation uncertainties are much smaller in some high-q cases than for low-q cases which are generally considered more reliable.

Hence, a study of the extrapolation uncertainties over the whole parameter space is useful in gaining an understanding of the properties of the different codes, but cannot directly be used as a measure of total NR uncertainties.

4. Peak accuracy

Since we are dealing with discrete numerical data sets, the peak finding might be also a problem if the sampling is not fine enough; particularly for high mass-ratio cases where the higher modes become more relevant and it is important to sample each mode accurately so that the overall peak profile is not washed out. We have estimated this contribution to NR uncertainties by applying two different time samplings to the data: for the actual $L_{\rm peak}$ values used in this paper, we use $\Delta t = 0.1M$, while here we compare also with a coarser $\Delta t = 1M$ to illustrate the possible loss of accuracy.

In Fig. 19 we show, for an SXS mass-ratio 10 non-spinning

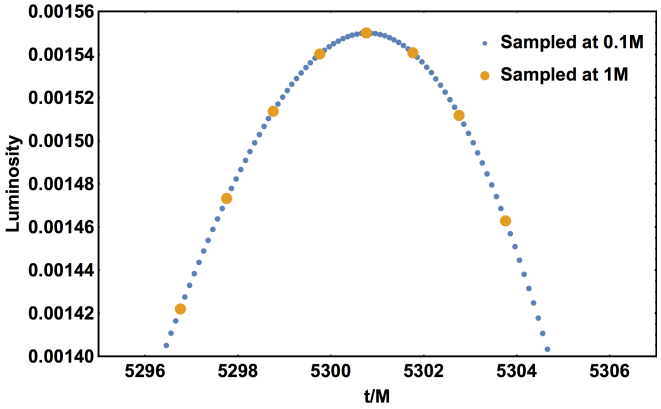

FIG. 19: Luminosity computed with different sampling rates in time, for the non-spinning $q = 10$ SXS waveform "SXS:BBH:0185" or "BBH_CFMS_d11d2_q10_sA_0_0_0_sB_0_0_0".

case, that the uncertainty contribution, measured as the difference between two points bracketing the peak, would be about 1% of the total peak luminosity with the coarser sampling, but is only about 0.05% for the finer sampling that we actually use. As a worst case, we found 0.2% for the non-spinning $q = 18$ BAM result.

5. Mode selection

As introduced in Eq. (2), we compute NR peak luminosities for BAM, SXS and GaTech waveforms as sums over all modes up to $\ell_{\rm max} = 6$. The RIT luminosities from [24, 26] use the same cut-off. For the perturbative data from [28–30] at large mass ratios, we use $\ell_{\rm max} = 8$. These choices are based on studying the individual contribution of each mode to the total luminosity, finding that $\ell > 6$ contributions are sufficiently small to be discarded for the NR data in comparison with other sources of uncertainty.

As an illustrative example, we show in the top panel of Fig. 20 the cumulative peak luminosity when adding modes ℓ by ℓ (including all $|m| \leq \ell$ at each step) for the $q = 10$ non-spinning SXS waveform, and the per-ℓ contributions in the lower panel. The fall-off of the higher-ℓ contributions to the global peak is expected to be exponential, which is indeed found in this case.

To quantify and extrapolate the loss generally expected for non-spinning configurations, we have estimated the relative loss in $L_{\rm peak}$ from not including the $\ell = 7, 8$ modes for non-spinning SXS waveforms up to mass-ratio $q = 10$ (maximum loss of 0.6%) and the non-spinning BAM simulation at $q = 18$ (loss of 1%), and fit a quadratic function in η:

$$\frac{\Delta L_{\rm peak}}{L_{\rm peak}} = 0.017611 - 0.153760\eta + 0.334803\eta^2 \,. \qquad \text{(A4)}$$

This result is illustrated in Fig. 21, together with a marginally consistent fit when including the $q = 10^3$ Teukolsky result (loss of 2%). The $\ell > 6$ contributions are smaller for negative spins and larger for positive spins, as illustrated in the same figure with $\chi_1 = \pm 0.8$ results at $q = 10^3$ and from BAM at $q = 18$. The largest loss for any NR case investigated is $\lesssim 2\%$ for the $q = 18$, $\chi_1 = +0.8$ BAM case, which is a significant contribution to the overall error budget but still on the

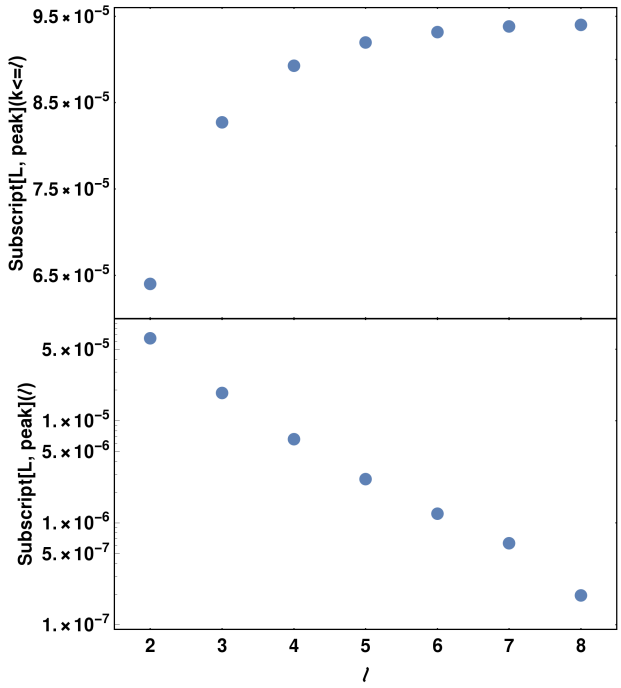

FIG. 20: Per-mode contributions to the total peak luminosity for the same SXS case as in Fig. 19. Top panel: cumulative sum up to ℓ. Lower panel: Natural logarithm of the luminosity contribution per ℓ. Each point contains all m for the given ℓ. Similar behavior for large mass ratios was found in [65].

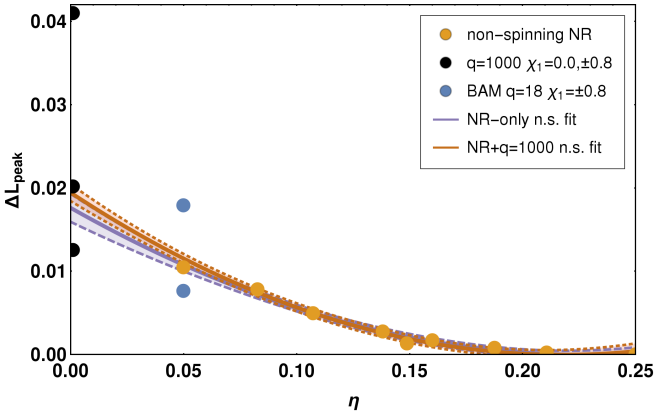

FIG. 21: Relative loss in the peak luminosity including modes up to $\ell_{\rm max} = 6$ against $\ell_{\rm max} = 8$, for non-spinning SXS cases up to $q = 10$, a non-spinning BAM case with $q = 18$ and the $q = 10^3$ Teukolsky result. Also shown are $q = 18$ and $q = 10^3$ results for $\chi_1 = +0.8$ (above the non-spinning line) and for $\chi_1 = -0.8$ (below), as well as the quadratic non-spinning fit from Eq. (A4) to NR data points only and a fit of the same order including the $q = 10^3$ point, with 90% confidence intervals for both fits.

level of other error sources. For the perturbative large-mass-ratio results, with a worst-case $\ell > 6$ of $\approx 5\%$, we use $\ell_{\rm max} = 8$ instead, so that the loss from $\ell > 8$ is limited to $< 1\%$.

Another useful investigation is to consider the η dependence, and especially the $\eta \to 0$ behavior, for individual modes. Fitting $L'_{\rm peak}\left(\eta, \widehat{S} = 0\right)$ in Sec. III A we found, as illustrated in Fig. 4, that the peak luminosity of all modes summed up to $\ell_{\rm max} = 6$, after scaling out the dominant η^2 dependence, is not a monotonic function towards low η. Repeating the same comparison of rescaled non-spinning peak lumi-

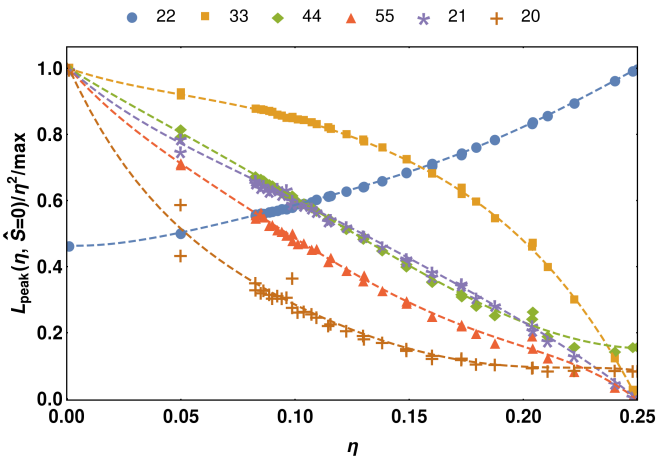

FIG. 22: Comparison of rescaled peak luminosities for non-spinning NR and perturbative large-mass-ratio data, for a small selection of modes. The points for each mode have been scaled by the maximum for that mode, which is at $\eta = 0.25$ for the 22 mode and at $\eta \to 0$ for the other modes. The connecting lines are fifth-order polynomial fits, which were not statistically optimized and just added to guide the reader's eye. This can be compared with the sum over modes in Fig. 4. As a guide to the overall strength of the individual modes, we list the non-rescaled maxima $\max_\eta L^{\ell m}_{\rm peak}(\widehat{S} = 0)$ of each of the displayed modes $(\ell m) = \{22, 33, 44, 55, 21, 20\}$: $\{1.0\times10^{-3}, 5.9\times10^{-5}, 1.5\times10^{-5}, 5.3\times10^{-6}, 9.8\times10^{-6}, 6.3\times10^{-7}\}$.

nosities between NR and perturbative large-mass-ratio data, but for individual modes, we find – as shown in Fig. 22 for a subset of modes – that these are all montonic as $\eta \to 0$; however, the slopes are very different, with the dominant 22 mode falling off faster than η^2 and the subdominant and higher modes falling off much slower, consistent with the general expectation of stronger contributions at low η. This finding of monotonicity in each mode increases our trust in the combination of NR and perturbative results, and the non-monotonicity of the rescaled peak luminosities after summing the modes can thus be explained as a superposition of these counteracting trends in the individual modes.

6. Cases not used in fit calibration

Of the full catalog of 419 NR simulations from four codes, we have only used 378 to calibrate our new fit. 22 of the 41 removed cases are non-spinning or equal-spin configurations. Of these, 17 belong to one of the pairs or groups of equivalent initial parameters identified in Sec. A 1, with differences between the paired results inconsistent at a level higher than the fit residuals we can otherwise achieve in the corresponding subspace fit; or are individual points inconsistent with an otherwise consistent set of direct neighbors. In these cases we removed from each tuple the case most discrepant with the others and with the global trend. This includes for example the GaTech $q = 4$ and SXS $q = 2.5$ non-spinning cases shown in the extrapolation comparisons of Figs. 17 and 18, or the SXS $\left(q = 1, \widehat{S} = 0.97\right)$ point whose luminosity seems inconsistent with other $q = 1$, high-spin SXS results.

We emphasize that in the one-dimensional fits for non-spinning and equal-mass-equal-spin BBHs we calibrate the fits to sub-percent accuracies, so that this is a very strict criterion for removing cases, which mainly serves to guarantee a very clean calibration of the well-covered subspaces and dominant effects so that in the later steps we have a better chance of isolating and extracting subdominant effects from the general, more noisy data set. In terms of total absolute or relative errors compared with the whole NR data set, several of these cases are not overly inaccurate, and we do not imply that necessarily there are data quality issues with the waveforms from which the luminosities are calculated.

The remaining cases were identified as strong outliers outside of the main distribution in visual inspection of the two-dimensional equal-spin fit (Sec. III D) and the per-mass-ratio analysis of residuals of unequal-spin cases against the 2D fit (Sec. III E). For these simulations, there are no equivalent or nearby comparison cases, so that it cannot be said with certainty if they would still be outliers in a more densely covered future data set; and at the same time a small residual for any given point is no guarantee for its absolute accuracy when there are no equivalent comparison points. Hence, we have made much less strict exclusions in the sparsely covered unequal-spin range, which limits the accuracy to which we can extract the subdominant spin-difference effects (which are of a similar scale as the remaining scatter in the data set), but also reduces the risk of overfitting to spurious trends in a more strongly-trimmed data set.

	q	χ_1	χ_2	$L_{\rm peak}$	$\Delta L_{\rm peak}$	$\Delta L_{\rm peak}/L_{\rm peak}$	tag	code
1	1.00	0.20	0.80	0.00133540	−0.00001456	−0.011	Q1.00_0.20_0.80	RIT
2	1.00	0.25	0.25	0.00114910	−0.00001078	−0.009	Q1.0000_0.2500_0.2500	RIT
3	1.00	0.40	0.80	0.00143030	−0.00001617	−0.011	Q1.00_0.40_0.80	RIT
4	1.00	0.50	0.50	0.00132610	−0.00002155	−0.016	Q1.0000_0.5000_0.5000	RIT
5	1.00	0.80	0.80	0.00165190	−0.00005163	−0.031	Q1.0000_0.8000_0.8000	RIT
6	1.00	0.97	0.97	0.00185963	−0.00017055	−0.092	d15_q1_sA_0_0_0.97_sB_0_0_0.97_ecc6e-4	SXS
7	1.00	−0.80	−0.80	0.00075683	−0.00000522	−0.007	d15_q1_sA_0_0_-0.8_sB_0_0_-0.8	SXS
8	1.00	−0.95	−0.95	0.00071785	−0.00001083	−0.015	d15_q1_sA_0_0_-0.95_sB_0_0_-0.95	SXS
9	1.10	0.00	0.00	0.00102562	0.00000646	0.006	735261.71914351848	GaT
10	1.33	0.50	0.50	0.00127610	−0.00001496	−0.012	Q0.7500_0.5000_0.5000	RIT
11	1.33	0.80	−0.80	0.00113510	0.00001336	0.012	Q0.7500_-0.8000_0.8000	RIT
12	1.33	0.60	0.80	0.00144390	−0.00002266	−0.016	Q1.33_0.80_0.60	RIT
13	1.50	0.00	0.00	0.00092086	−0.00000929	−0.010	Q0.6667_0.0000_0.0000	RIT
14	1.67	0.00	0.00	0.00089059	0.00001118	0.013	Q0.6000_0.0000_0.0000	RIT
15	2.00	0.85	−0.85	0.00104805	−0.00005372	−0.051	q2_-85_85_0.2833_it2_T_96_468	BAM
16	2.00	0.60	0.60	0.00113005	−0.00001154	−0.010	735103.51365740737	GaT
17	2.00	0.85	0.00	0.00119969	−0.00004465	−0.037	q2_0_85_0.566667_T_80_360	BAM
18	2.00	0.80	0.80	0.00133220	−0.00004371	−0.033	Q2.00_0.80_0.80	RIT
19	2.00	0.60	0.50	0.00109870	−0.00002568	−0.023	Q0.5000_0.5000_0.6000	RIT
20	2.00	0.80	0.00	0.00115110	−0.00004828	−0.042	Q0.5000_0.0000_0.8000	RIT
21	2.50	0.00	0.00	0.00064369	0.00000637	0.010	BBH_CFMS_d16.9_q2.50_sA_0_0_0_sB_0_0_0	SXS
22	3.00	0.50	−0.50	0.00067168	−0.00002270	−0.034	q3_-50_50_0.25_T_80_400	BAM
23	3.00	0.00	0.00	0.00051866	−0.00000761	−0.015	735125.35489583330	GaT
24	3.00	0.40	0.00	0.00065030	−0.00001591	−0.024	735132.32116898149	GaT
25	3.00	0.50	0.80	0.00074376	−0.00001267	−0.017	Q0.3333_0.8000_0.5000	RIT
26	3.00	0.60	0.00	0.00074392	−0.00003003	−0.040	735137.68403935188	GaT
27	3.00	0.67	0.00	0.00078909	−0.00002904	−0.037	Q3.00_0.00_0.67	RIT
28	3.00	0.80	−0.80	0.00084159	−0.00002278	−0.027	Q3.00_-0.80_0.80	RIT
29	3.00	0.85	0.85	0.00107685	0.00003335	0.031	BBH_SKS_d13.9_q3_sA_0_0_0.850_sB_0_0_0.850	SXS
30	4.00	0.75	0.75	0.00069840	0.00001188	0.017	q4a075_T_112_448	BAM
31	4.00	0.75	0.00	0.00063280	−0.00002841	−0.045	Q4.00_0.00_0.75	RIT
32	4.00	0.00	0.00	0.00037948	0.00000782	0.021	735088.53305555554	GaT
33	4.30	0.00	0.00	0.00034217	0.00000421	0.012	735383.79907407402	GaT
34	4.50	0.00	0.00	0.00031462	−0.00000329	−0.010	735250.44696759258	GaT
35	5.00	0.80	0.00	0.00052483	−0.00000926	−0.018	Q5.00_0.00_0.80	RIT
36	5.00	0.00	0.00	0.00026999	−0.00000480	−0.018	735094.40543981478	GaT
37	5.00	0.40	0.00	0.00034792	−0.00001784	−0.051	735171.32571759261	GaT
38	6.00	0.00	0.00	0.00020707	−0.00000395	−0.019	Q0.1667_0.0000_0.0000	RIT
39	6.00	0.00	0.00	0.00021325	0.00000234	0.011	735192.40052083333	GaT
40	6.00	0.20	0.00	0.00023419	−0.00000829	−0.035	735198.69027777773	GaT
41	18.00	−0.80	0.00	0.00006179	0.00003868	0.626	q18a0aM08c025_96_fine	BAM

TABLE VI: NR cases from the source catalogs not included in the fit calibration, for reasons detailed below.