PHYSICAL REVIEW D **93,** 024003 (2016)

Science with the space-based interferometer eLISA: Supermassive black hole binaries

Antoine Klein,[1] Enrico Barausse,[2,3] Alberto Sesana,[4] Antoine Petiteau,[5] Emanuele Berti,[1,6] Stanislav Babak,[7] Jonathan Gair,[8,9] Sofiane Aoudia,[10] Ian Hinder,[7] Frank Ohme,[11] and Barry Wardell[12,13]

[1]*Department of Physics and Astronomy, The University of Mississippi, University, Mississippi 38677, USA*
[2]*Sorbonne Universités, UPMC Université Paris 6, UMR 7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014 Paris, France*
[3]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014 Paris, France*
[4]*School of Physics and Astronomy, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*
[5]*APC, Université Paris Diderot, Observatoire de Paris, Sorbonne Paris Cité, 10 rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France*
[6]*CENTRA, Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Avenida Rovisco Pais 1, 1049 Lisboa, Portugal*
[7]*Max Planck Institute for Gravitational Physics, Albert Einstein Institute, Am Mühlenberg 1, 14476 Golm, Germany*
[8]*Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, United Kingdom*
[9]*School of Mathematics, University of Edinburgh, The King's Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom*
[10]*Laboratoire de Physique Théorique, Faculté des Sciences Exactes Université de Bejaia, 06000 Bejaia, Algeria*
[11]*School of Physics and Astronomy, Cardiff University, Queens Building, CF24 3AA Cardiff, United Kingdom*
[12]*Department of Astronomy, Cornell University, Ithaca, New York 14853, USA*
[13]*School of Mathematical Sciences and Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland*

(Received 24 November 2015; published 6 January 2016)

We compare the science capabilities of different eLISA mission designs, including four-link (two-arm) and six-link (three-arm) configurations with different arm lengths, low-frequency noise sensitivities and mission durations. For each of these configurations we consider a few representative massive black hole formation scenarios. These scenarios are chosen to explore two physical mechanisms that greatly affect eLISA rates, namely (i) black hole seeding, and (ii) the delays between the merger of two galaxies and the merger of the black holes hosted by those galaxies. We assess the eLISA parameter estimation accuracy using a Fisher matrix analysis with spin-precessing, inspiral-only waveforms. We quantify the information present in the merger and ringdown by rescaling the inspiral-only Fisher matrix estimates using the signal-to-noise ratio from nonprecessing inspiral-merger-ringdown phenomenological waveforms, and from a reduced set of precessing numerical relativity/post-Newtonian hybrid waveforms. We find that all of the eLISA configurations considered in our study should detect some massive black hole binaries. However, configurations with six links and better low-frequency noise will provide much more information on the origin of black holes at high redshifts and on their accretion history, and they may allow the identification of electromagnetic counterparts to massive black hole mergers.

DOI: 10.1103/PhysRevD.93.024003

I. INTRODUCTION

Gravitational waves (GWs) are a generic prediction of general relativity (GR) [1] and of other relativistic gravitational theories [2–4]. Indirect evidence for the existence of GWs comes from observations of binary systems involving at least one pulsar [5], which allow us to track the orbital period very accurately over long time scales and to observe small secular changes due to the emission of GWs. The observed damping is in sub-percent-level agreement with the predictions of GR's quadrupole formula for GW emission [6,7], and overall the combined measurements of secular changes are consistent with the predictions of GR within 0.05% [8].

A worldwide experimental effort towards a direct detection of GWs is also under way. Ground-based, kilometer-scale laser interferometers target the GW emission from a variety of sources, including the late inspiral of neutron-star binaries; the inspiral, merger and ringdown of systems comprised of two stellar-mass black holes, or a neutron-star and a stellar-mass black hole; supernova explosions; and isolated pulsars. These ground-based interferometers work

as a network of "second-generation" detectors, as opposed to the first generation of ground-based interferometers (i.e., the initial LIGO and Virgo experiments), which were active from 2002 to 2010. They include the two Advanced LIGO [9] interferometers in the U.S. (which are currently taking data in science mode) and the French-Italian detector Advanced Virgo [10] (which will undergo commissioning in 2016). Within the next few years the Japanese interferometer KAGRA [11] and a LIGO-type detector in India [12] will join this network.

At the same time, pulsar-timing arrays [13] are targeting the GW signal from binaries of massive black holes (MBHs) with masses $\sim 10^8 - 10^{10}\ M_\odot$ at separations of hundreds to thousands of gravitational radii. Low-frequency GWs emitted by these systems passing between a pulsar and Earth leave a characteristic imprint in the time of arrival of the radio pulses. This signal can be disentangled from other sources of noise and unambiguously identified by timing an ensemble of ultrastable millisecond pulsars (i.e., a pulsar-timing array) [14]. The European Pulsar Timing Array (EPTA) [15], the Parkes Pulsar Timing Array (PPTA) [16] and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [17], joining together in the International Pulsar Timing Array (IPTA) [18,19], are constantly improving their sensitivity in the frequency range $\sim 10^{-9} - 10^{-6}$ Hz, and have already significantly constrained the stochastic GW signal from the massive black hole binary (MBHB) population [20–22].

A common feature of current ground-based interferometers and pulsar-timing arrays is that they primarily probe the low-redshift Universe. For instance, pulsar-timing arrays are mostly sensitive to MBHBs at $0.2 \lesssim z \lesssim 1.5$, while the range of ground-based detectors depends on the particular family of sources being observed, but never exceeds $z \sim 2$ [23–25]. The two classes of experiments explore hugely different frequency ranges for the GW signal ($f \sim 10^{-9} - 10^{-6}$ Hz for pulsar-timing array, and $f \sim 10 - 10^3$ Hz for ground-based interferometers), and the GW frequency band in between these widely separated bands is difficult to probe from the ground. Pulsar-timing arrays are intrinsically limited at the high-frequency end by the Nyquist frequency, set by the typical interval between subsequent pulsar observation sessions (typically a few weeks, corresponding to roughly 10^{-6} Hz). Second-generation ground-based detectors are intrinsically limited at frequencies lower than $f \sim 10$ Hz by "seismic noise" (the main noise source is actually given not by geological activity, but by vibrations due to surface events: human activity, waves, wind, etc.). While third-generation ground-based interferometers (such as the Einstein Telescope [26]), which have been proposed for construction in the next decade, might abate seismic noise by going underground, they will not be able to probe frequencies lower than $f \sim 0.1$ Hz [27].

However the frequency window between 10^{-7} Hz and ~ 1 Hz is expected to be populated by a rich variety of astrophysical sources. For instance, our current understanding of galaxy formation and an increasing body of observational evidence support the idea that MBHs (with masses from $10^5\ M_\odot$—or even lower—up to $10^9 - 10^{10}\ M_\odot$) are hosted in the centers of almost all galaxies [28]. When galaxies merge to form bigger systems [as predicted by the Λ cold dark-matter (ΛCDM) model and supported by observations], these MBHs are expected to form tight binaries [29], inspiral and finally merge into a perturbed black hole, which sheds away these perturbations in the ringdown phase [30,31]. The inspiral, merger and ringdown phases are expected to be the main GW source in the $10^{-8} - 1$ Hz spectrum. Their detection will provide precious information on the coevolution between MBHs and their host galaxies (and henceforth on the hierarchical formation of structures in the ΛCDM model), on the very origin of MBHs at high redshift, on the dynamics of gas and accretion onto MBHs, and on the strong-field, highly relativistic dynamics of GR (see e.g. [32–35], and [36] for a review of the science achievable with GW observations of MBHBs). Moreover, if these sources could also be identified in the electromagnetic band, they may be used as "standard sirens" to probe the expansion history of the Universe [37,38].

Other sources that may populate this unprobed frequency band include binary systems formed of a compact object (i.e., a stellar-mass black hole, a neutron star or a white dwarf) orbiting a MBH [39]. The observation of the GW signal from these systems, known as extreme mass ratio inspirals (EMRIs), would provide a way to test GR and the geometry of MBHs with unprecedented accuracy, as well as a way to understand the dynamics of stellar objects in galactic nuclei. Also present in this band will be countless almost monochromatic sources from our own Galaxy, e.g. neutron-star or white-dwarf binary systems at wide separations. White-dwarf binary systems will be particularly numerous, especially at $f \sim 10^{-4} - 10^{-5}$ Hz [40], and may give rise to a background GW signal only partially resolvable as individual sources. Finally, more exotic sources may also be present, for instance stochastic background signals of cosmological origin, arising from new physics at the TeV energy scale or beyond, such as cosmic strings or a first-order electroweak phase transition [41].

Given its tremendous potential for fundamental physics and astrophysics, the European Space Agency (ESA) has selected the observation of the Universe at GW frequencies around one mHz as one of the three main science themes of the "Cosmic Vision Program" [42]. Indeed, a call for mission proposals for the "Gravitational Universe" science theme is expected for late 2016, and the L3 launch slot in 2034 has been reserved for the selected mission. The main candidate mission for this call (for which a decision will be made by 2018–19, so as to allow sufficient time for industrial production before the nominal 2034 launch date) is the evolving Laser Interferometer Space Antenna (eLISA) [43], named after the "classic LISA" concept of

the late 1990s and early 2000s [44]. The eLISA mission concept consists of a constellation of three spacecraft, trailing Earth around the Sun at a distance of about 15°. Each spacecraft will contain one or two test masses in almost perfect free fall, and laser transponders which will allow measurements of the relative proper distances of the test masses in different spacecraft via laser interferometry. This will allow the detection of the effect of possible GW signals (which would change the distance between the test masses). The most technically challenging aspect of the mission will be to maintain the test masses in almost perfect free fall. For this reason, a scaled-down version of one of eLISA's laser links will be tested by the "LISA Pathfinder" mission. Pathfinder was launched by ESA in December 2015, and it will provide crucial tests of how well eLISA's low-frequency acceleration noise can be suppressed.

There are, however, other aspects to the eLISA mission that are yet to be evaluated and decided upon by ESA, within the constraints imposed by the allocated budget for the Gravitational Universe science theme. A "Gravitational Observatory Advisory Team" (GOAT) [45] has been established by ESA to advise on the scientific and technological issues pertaining to an eLISA-like mission. Variables that affect the cost of the mission include (i) the already mentioned low-frequency acceleration noise; (ii) the mission lifetime, which is expected to range between one and several years, with longer durations involving higher costs because each component has to be thoroughly tested for the minimum duration of the mission, and may also require higher fuel consumption, since the orbital stability of the triangular constellation sets an upper limit on the mission duration and therefore achieving a longer mission may require the constellation to be further from Earth; (iii) the length L of the constellation arms, which may range from one to several million kilometers, with longer arms involving higher costs to put the constellation into place and to maintain a stable orbit and slowly varying distances between the spacecraft; (iv) the number of laser links between the spacecraft, i.e., the number of "arms" of the interferometer (with four links corresponding to two arms, i.e., only one interferometer, and six links to three arms, i.e., two independent interferometers at low frequencies [46]): giving up the third arm would cut costs (mainly laser power, industrial production costs), while possibly hurting science capabilities (especially source localization) and allowing for no redundancy in the case of technical faults in one of the laser links.

This paper is the first in a series that will evaluate the impact of these four key design choices on the scientific performance of eLISA. Here we focus on the main scientific target of eLISA, namely the inspiral, merger and ringdown of MBHBs. We assess how the number of observed sources, their distance and the accuracy with which their parameters can be extracted from the data change under different design choices. In subsequent papers in this series we will repeat this exercise for other sources/science capabilities of eLISA, namely (a) EMRIs and their science impact; (b) the measurement of the expansion history of the Universe ("cosmography"); (c) Galactic white-dwarf binaries and their science impact; (d) the stochastic background from a first-order phase transition in the early Universe.

The plan of the paper is as follows. In Sec. II we review the different mission designs used in our analysis, and how the corresponding noise curves are produced. In Sec. III we describe our MBH evolution models. In Sec. IV we review the gravitational waveforms, instrument response model and data analysis tools used in the present work. In Sec. V we present the results of the parameter estimation study in detail. Some conclusions and a discussion of possible directions for future work are given in Sec. VI. Appendix A describes the construction of the precessing-binary hybrid waveforms that we used to extrapolate our inspiral-only error estimates on sky location and distance determination. In Appendix B we give a simple analytical prescription to estimate the error on the remnant spin from the ringdown radiation. Unless otherwise specified, throughout this paper we adopt geometrical units ($G = c = 1$).

II. eLISA MISSION DESIGNS

We investigated six detector noise curves by varying two key parameters characterizing the noise, namely (1) the arm length L, chosen to be either 1, 2 or 5×10^6 km (A1, A2 and A5, respectively); (2) the low-frequency acceleration, which is either projected from the expected performance of LISA Pathfinder (N2) or ten times worse, assuming a very pessimistic outcome of LISA Pathfinder (N1). The laser power and the diameter of the telescope have been adjusted based on the interferometer arm length to get similar sensitivities at high frequencies. We considered a laser power of 0.7 W for configuration A1 and 2 W for configurations A2 and A5; the telescope mirror size has been chosen to be 25 cm for A1, 28 cm for A2, 40 cm for A5. Note that fixing a 2 W laser and 40 cm telescope improves the high-frequency performance of some configurations, but this only affects the high-frequency noise, and it has a very mild impact on the study presented here.

Analytic fits to the six sky-averaged sensitivity curves obtained in this way are of the form

$$S_n(f) = \frac{20}{3}\frac{4S_{n,\mathrm{acc}}(f) + S_{n,\mathrm{sn}}(f) + S_{n,\mathrm{omn}}(f)}{L^2} \times \left[1 + \left(\frac{f}{0.41\,\frac{c}{2L}}\right)^2\right]. \quad (1)$$

Here L is the arm length in meters, and $S_{n,\mathrm{acc}}(f)$, $S_{n,\mathrm{sn}}(f)$ and $S_{n,\mathrm{omn}}$ denote the noise components due to low-frequency acceleration, shot noise and other measurement noise, respectively. We use the following values:

$$S_{n,\mathrm{acc}}(f)=\begin{cases}9\times10^{-28}\frac{1}{(2\pi f)^4}\left(1+\frac{10^{-4}\,\mathrm{Hz}}{f}\right)\mathrm{m^2\,Hz^{-1}} & \text{for N1},\\ 9\times10^{-30}\frac{1}{(2\pi f)^4}\left(1+\frac{10^{-4}\,\mathrm{Hz}}{f}\right)\mathrm{m^2\,Hz^{-1}} & \text{for N2},\end{cases}$$

$$S_{n,\mathrm{sn}}(f)=\begin{cases}1.98\times10^{-23}\ \mathrm{m^2\,Hz^{-1}} & \text{for A1},\\ 2.22\times10^{-23}\ \mathrm{m^2\,Hz^{-1}} & \text{for A2},\\ 2.96\times10^{-23}\ \mathrm{m^2\,Hz^{-1}} & \text{for A5},\end{cases}$$

$$S_{n,\mathrm{omn}}(f)=2.65\times10^{-23}\,\mathrm{m^2\,Hz^{-1}}\quad\text{for all configurations.}$$

Here $S_{n,\mathrm{omn}}$—for "other measurement noise," as estimated in the eLISA study of [43]—might vary across configurations, however we keep it fixed according to the most conservative choice. The noise curves obtained in this way are shown in Fig. 1.

Each analytic curve was compared to its numerical counterpart generated by the LISACode simulator. We found very good agreement in all cases, the only difference being that the analytic fit does not reproduce the high-frequency oscillatory behavior. This is because the analytic fit assumes the long-wavelength approximation, which breaks down at $f=c/(2\pi L)\approx0.05(1\mathrm{Gm}/L)\mathrm{Hz}$. Still, the analytic approximation is sufficient for our purposes, since most of the relevant sources emit radiation at $f\lesssim0.05$ Hz.

In addition to the instrumental noise, we expect an astrophysical foreground coming from the compact white dwarf (CWD) binaries in our Galaxy. Millions of galactic binaries emit almost monochromatic GW signals which superpose with random phase, creating an unresolved stochastic foreground above a few millihertz. Sufficiently loud signals (standing above the background) and all individual signals at high frequencies can be identified and removed from the data [47–49]. Based on the population synthesis model of [50], we can estimate the unresolvable (stochastic) part of the GW signal generated by the population of galactic white dwarf binaries and produce a piecewise analytic fit of this signal, which is given below for each configuration[1]:

$$S_{\mathrm{gal,N2A1}}(f)=\begin{cases}f^{-2.1}\times1.55206\times10^{-43} & 10^{-5}\le f<5.3\times10^{-4},\\ f^{-3.235}\times2.9714\times10^{-47} & 5.3\times10^{-4}\le f<2.2\times10^{-3},\\ f^{-4.85}\times1.517\times10^{-51} & 2.2\times10^{-3}\le f<4\times10^{-3},\\ f^{-7.5}\times6.706\times10^{-58} & 4\times10^{-3}\le f<5.3\times10^{-3},\\ f^{-20.0}\times2.39835\times10^{-86} & 5.3\times10^{-3}\le f\le10^{-2}\end{cases}$$

$$S_{\mathrm{gal,N2A2}}(f)=\begin{cases}f^{-2.1}\times1.3516\times10^{-43} & 10^{-5}\le f<5.01\times10^{-4},\\ f^{-3.3}\times1.4813\times10^{-47} & 5.01\times10^{-4}\le f<2.07\times10^{-3},\\ f^{-5.2}\times1.17757\times10^{-52} & 2.07\times10^{-3}\le f<3.4\times10^{-3},\\ f^{-9.1}\times2.7781\times10^{-62} & 3.4\times10^{-3}\le f<5.2\times10^{-3},\\ f^{-20.0}\times3.5333\times10^{-87} & 5.2\times10^{-3}\le f\le10^{-2}\end{cases}$$

$$S_{\mathrm{gal,N2A5}}(f)=\frac{20}{3}\begin{cases}f^{-2.3}\times10^{-44.62} & 10^{-5}\le f<10^{-3},\\ f^{-4.4}\times10^{-50.92} & 10^{-3}\le f<10^{-2.7},\\ f^{-8.8}\times10^{-62.8} & 10^{-2.7}\le f<10^{-2.4},\\ f^{-20.0}\times10^{-89.68} & 10^{-2.4}\le f\le10^{-2}.\end{cases}\tag{2}$$

The fit for the LISA-like configuration $S_{\mathrm{gal,A5}}(f)$ is taken from [51]. This astrophysical stochastic GW foreground is added to the instrumental noise in quadrature,[2] and the resulting curves are also shown in Fig. 1. Note that the GW foreground is below the instrumental nose for N1 configurations. For each acceleration noise (N1/N2) we consider either four or six laser links (L4/L6), and we assume a mission lifetime of either two or five years (M2/M5). This amounts to a total of 24 mission configurations, labeled as N*i*A*j*M*k*L*l* (where $i=1,2$, $j=1,2,5$, $k=2,5$, $l=4,6$).

The configuration N2A5M5L6 corresponds to Classic LISA. Configuration N2A1M2L4 corresponds approximately to the New Gravitational Observatory (NGO) [52] concept, which was proposed to ESA during the selection process for the L1 large satellite mission. and which was also used to illustrate the science case in the

[1]We provide here fits for the N2 configurations only, since the CWD background was found to give negligible contribution to the noise budget in all N1 baselines.

[2]The CWD unresolved background depends on the mission duration: as the observation time increases, more individual CWDs can be identified and subtracted. The fits of Eq. (2) were derived for a two-year mission lifetime (M2). We expect the background to be slightly lower for a five-year mission (M5), but for simplicity we omit this effect and use Eq. (2) for both M2 and M5 configurations.

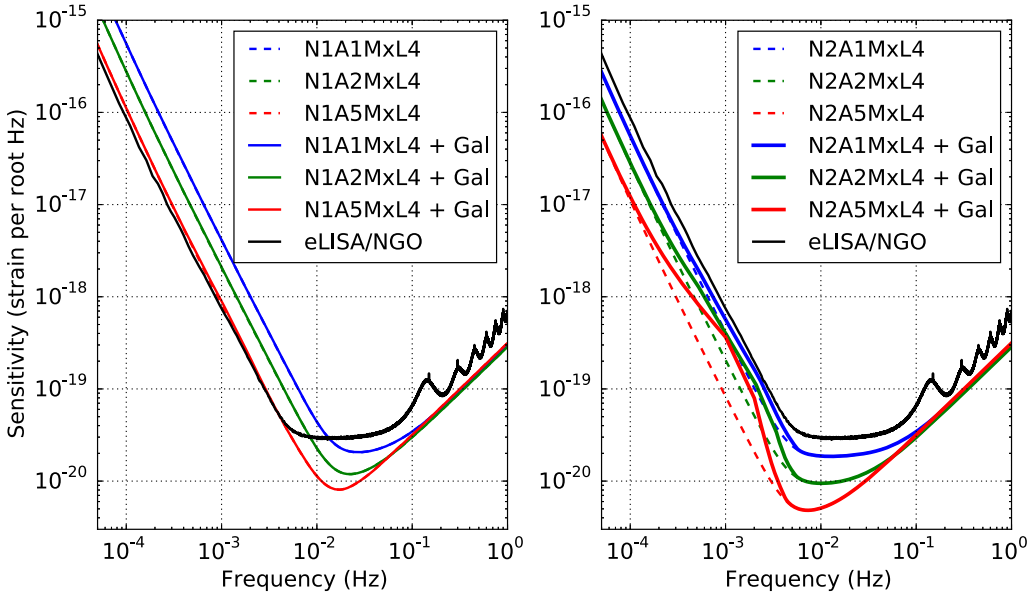

FIG. 1. Analytic fits to the sensitivity curves for different configurations investigated in this paper. In both panels the curves running from the top down are for A1 (blue), A2 (green) and A5 (red) configurations, and all curves have the same high-frequency noise by design. The (black) curve with wiggles at high frequency is the numerical sensitivity for eLISA/NGO, which is shown in both panels for reference. Dashed curves include only instrumental noise, while solid lines represent the total noise (including the contribution from CWD confusion noise). The left panel shows all configurations with pessimistic (N1) acceleration noise levels; the right panel assumes optimistic (N2) acceleration noise levels.

Gravitational Universe [43]. Figure 1 shows that configuration N2A1M2L4 differs from NGO only in a multiplicative factor $1/0.65$, which was included in the NGO design as a safety margin. In Sec. V we will therefore use N2A1MxL4 as a proxy for NGO, and "normalize" our science performance results to this configuration.

III. SUPERMASSIVE BLACK HOLE EVOLUTION

Sound observational evidence [53] as well as theoretical considerations (see e.g. [54,55]) suggest that the evolution of MBHs on cosmological time scales is inextricably coupled to the evolution of their host galaxies. Methods to follow this MBH-galaxy coevolution include Eulerian and smoothed-particle hydrodynamics simulations, as well as semianalytical galaxy-formation models (see [54,55] for recent reviews of the theory of galaxy formation with an overview of these techniques). While hydrodynamical simulations have a better handle of the small-scale physics, subgrid dissipative phenomena such as star formation and feedback are not yet treatable self-consistently by simulations, and it is unlikely that they will be in the near future. The same is true for the scale of the horizon of MBHs, on which key phenomena for the cosmological evolution of these objects, such as mergers and accretion, take place.

Clearly, the same drawback applies to semianalytical models, in which not only the subgrid physics, but also the scales that would be within reach of hydrodynamical simulations are treated with simplified prescriptions depending on a (limited) number of free parameters that are calibrated against observations. For our purposes, semianalytical models have the advantage of being computationally more convenient, which allows us to explore the parameter space of galaxy formation and MBH evolution with large statistics.

Here we adopt the semianalytical model of [56], which was later improved in [57] (where the prescriptions for the MBH spin evolution and the star formation were improved) and [58,59] (which implemented the cosmological evolution of nuclear star clusters, and accounted for the delay between a galaxy merger and that of the MBHs hosted by the two galaxies). Our model follows the evolution of baryonic structures along a dark-matter merger tree produced by an extended Press-Schechter formalism, suitably modified to reproduce the results of N-body simulations following [60]. The model evolves the hot unprocessed intergalactic medium; the cold, metal-enriched interstellar medium (in both its galactic disk and bulge components); the stellar galactic disk and the stellar spheroid; the nuclear gas and the nuclear star cluster; and, of course, the MBHs. These components are linked by a number of gravitational and nongravitational interactions, which are summarized graphically in Fig. 2. We refer to [56,57,59] for detailed descriptions of our implementation of the various processes represented in this diagram. Highlighted in red are the key assumptions that we will vary in this paper, and that we discuss below: black hole seeding and delays.

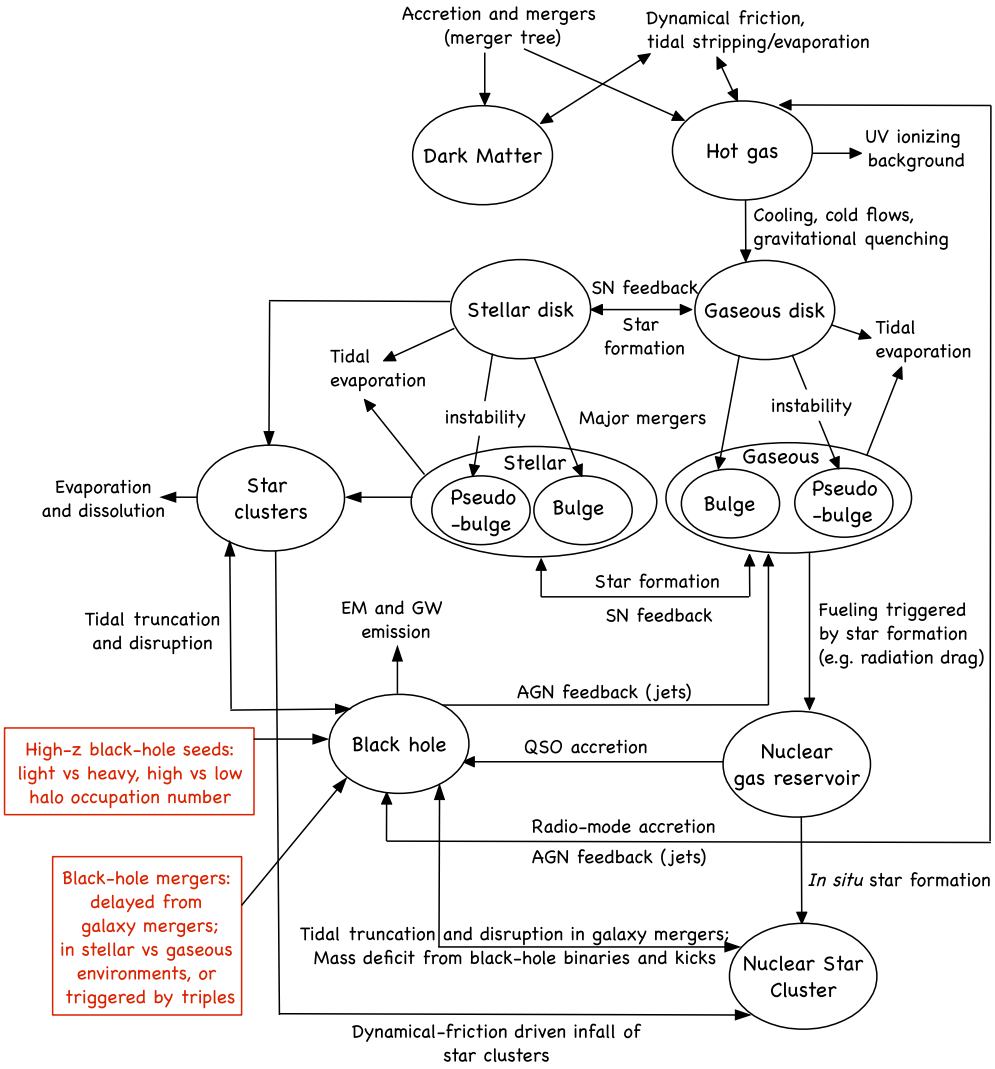

FIG. 2. Schematic summary of the model of [56], with the improvements of [57–59]. Red boxes on the left highlight the elements that heavily affect rates (black hole seeding and delays), for which we consider multiple options in this paper.

A. Black hole seeding

One of the crucial uncertainties in rate predictions is the birth mechanism of MBHs, which are thought to grow from high-redshift "seeds" whose exact nature is still debated.

The "light-seed" scenario assumes that these seeds may be the remnants of population III (popIII) stars forming in the low-metallicity environments characterizing the Universe at $z \approx 15-20$ [61]. While the mass of these first stars (and therefore that of their remnants) is uncertain, it can be of the order of a few hundred $M_\odot$ [62] (although recent simulations favor more fragmentation and lower masses, see e.g. [63]). In this light-seed scenario, we draw the popIII star mass from a log-normal distribution centered at 300 $M_\odot$ with a rms of 0.2 dex and an exclusion region between 140 and 260 $M_\odot$: in this mass range popIII stars explode as supernovae due to the electron-positron pair instability, without forming a black hole [62]. Outside this mass range, a black hole will generally form at the end of the popIII star's evolution. We assume that the mass of this black hole is about 2/3 the mass of the initial star [62]. Also, because active star formation at $z \approx 15-20$ is only expected in the deepest potential wells, we place a black-hole seed only in the rare massive halos collapsing from the 3.5σ peaks of the primordial density field [61,64], between $z = 15$ and $z = 20$ (the latter being the initial redshift of our merger trees). Because in light-seed scenarios it is difficult to reproduce the active galactic nuclei (AGN) luminosity function at high redshifts unless super-Eddington accretion is allowed [65], we assume the maximum MBH accretion rate to be $\dot{M} = A_{\rm Edd}\dot{M}_{\rm Edd}$, with $A_{\rm Edd}$ being a free parameter that we set to $\approx$2.2 as in [57–59].

In the "heavy-seed" scenario, MBHs already have masses $\sim 10^5\ M_\odot$ at high redshifts $z \sim 15-20$. These seeds may arise from the collapse (due e.g. to bar instabilities) of protogalactic disks. This would funnel large quantities of cold gas to the nuclear region, where

a black hole seed might then form. Several flavors of this scenario have been proposed (see e.g. [66–69]). In this paper we adopt a particular model, namely that of [69]. Like in the light-seed model, we start our evolutions at $z = 20$ and stop seed formation at $z = 15$, where we assume that the Universe has been metal enriched by the first generation of stars, which results in quenching of the seed formation due to the enhanced radiative cooling. The model of [69] has a free parameter, the critical Toomre parameter Q_c at which the protogalactic disks are assumed to become unstable. By changing Q_c one varies the probability that a halo hosts a black hole seed (i.e., the halo occupation number). Plausible values of Q_c range from 1.5 to 3, with larger values corresponding to larger halo occupation numbers, but one must have $Q_c \gtrsim 2$ to ensure that a significant fraction of massive galaxies host a MBH at $z = 0$ [69].

B. Delays

Another ingredient that is particularly important for calculating MBH merger rates is the delay between MBH mergers and galaxy mergers. When two dark-matter halos coalesce, the galaxies that they host initially maintain their identity, because they are smaller and more compact than the halos. The galaxies are then brought together by dynamical friction on typical time scales of a few gigayears. During this time, environmental effects such as tidal stripping and tidal evaporation remove mass from the smaller galaxy, which in turn affects the dynamical evolution of the system: see [56] for more details about our treatment of dynamical friction, tidal stripping and evaporation.

After the two galaxies have merged, the MBHs they host are slowly brought to the center of the newly formed galaxy by dynamical friction against the stellar background. As a result, the MBHs eventually form a bound system (a "hard" binary), i.e., one such that their relative velocity exceeds the velocity dispersion of the stellar background. From this moment on, the MBHB will further harden by three-body interactions with stars. It is unclear if this mechanism alone can bring the binary to the small separations ($\lesssim 10^{-3}$ pc) where GW emission can drive the system to merger within a Hubble time. This is known as the "last parsec problem" [70]. Recently, however, it has been suggested that triaxiality of the galaxy potential (resulting e.g. from a recent galaxy merger) would allow three-body stellar interactions to harden the binary to the GW-dominated regime on a typical time scale of a few gigayears [71–75]. Galaxy rotation has also been suggested as a possible mechanism helping the binary reach GW-dominated separations [76]. Moreover, if the nuclear region contains a significant amount of gas in a disk geometry, planetlike migration might drive the binary to merger on significantly shorter time scales, typically of $\sim 10^7 - 10^8$ yr [77,78] (but see e.g. [79] for possible complications arising in this scenario). Finally, if a MBHB stalls and in the meantime another galaxy merger happens, a third MBH may be added to the system. Triple interactions are expected to trigger the merger of the two most massive MBHs and ejection of the lightest one on time scales $\sim 10^8$ yr [80]. Nevertheless, this process is likely to be effective at inducing coalescence of the inner binary only for systems with masses $\gtrsim 10^6 - 10^7\ M_\odot$. Below that threshold, the lightest MBH may be ejected before the triple interactions trigger the merger of the inner binary, especially if its mass is much lower than that of the inner binary [59,80].

We refer to [59] for a detailed description of our implementation of these delays. To highlight their impact on our results for MBH merger rates, we consider both models where the delays are included, and models where no delays are present (i.e., the MBHs merge at the same time as their host galaxies).

C. Population models used in our study

We have generated possible realizations of the MBH population based on this semianalytical galaxy-formation model. We present results for variants of the model which are representative of the possible combinations that can be obtained by varying the prescriptions for seeding and delays presented above. More specifically, we focus on three models:

(1) Model popIII: This model assumes light MBH seeds from popIII stars, while accounting for the delays between MBH and galaxy mergers (cf. Sec. III B). The inclusion of these delays makes the model more realistic, and certainly more conservative. We verified that typical eLISA event rates change by less than a factor of 2 when setting the delays to zero, hence we decided to omit the variant of this light-seed model in which delays are not present.

(2) Model Q3-d: This assumes heavy MBH seeds from the collapse of protogalactic disks, while accounting for the delays between MBH and galaxy mergers. The halo occupation fraction of the seeds at high redshifts is determined by the critical Toomre parameter for disk instability, which we set to $Q_c = 3$ (cf. Sec. III A). Note however that setting $Q_c = 2$ only decreases the merger rates by a factor ~ 2. As in the case of the popIII model, the inclusion of delays makes this model more "realistic" (and conservative).

(3) Model Q3-nod: This is the same as model Q3-d, but without accounting for the delays between galaxy and MBH mergers. For this reason, this model should be considered an "optimistic" (upper bound) scenario for eLISA event rates.

For each of the three models above we simulate about 1300 galaxies/galaxy clusters, with dark-matter masses ranging from $10^{10}\ M_\odot$ to $10^{16}\ M_\odot$. By tracking self-consistently the mass and spin evolution of MBHs and their interaction (e.g. via feedback and accretion) with the

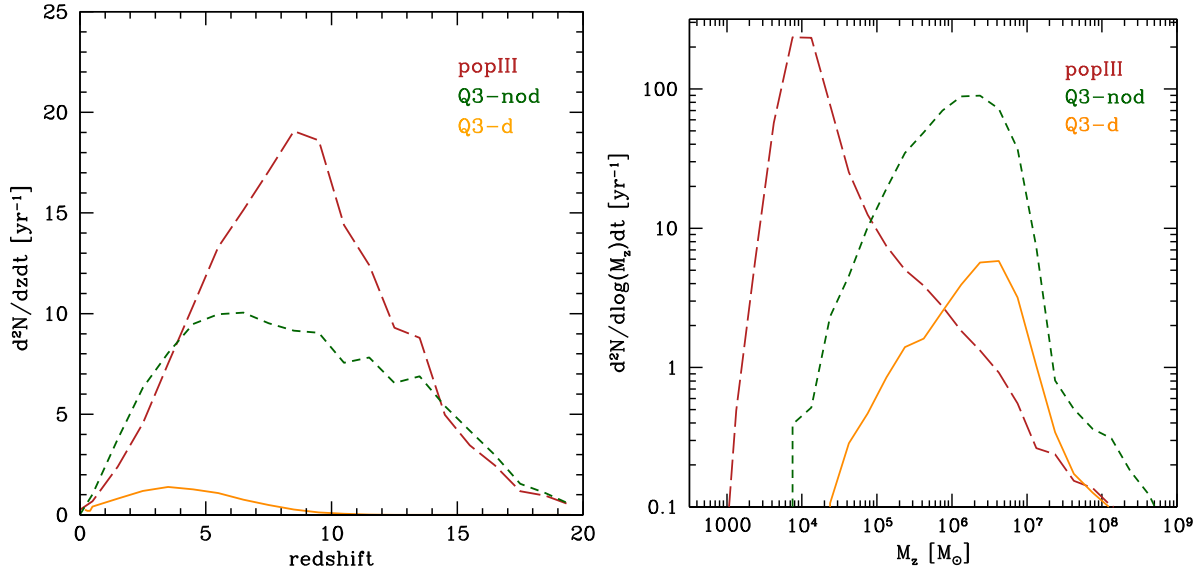

FIG. 3. Predicted merger rates per unit redshift (left panel) and per unit total redshifted mass $M_z = (m_1 + m_2)(1 + z)$ (right panel) for the three models described in the text.

galactic host, our model allows us to predict the masses, spin magnitudes and spin orientations of the MBHs when they form a GW-driven binary system. As such, while MBHBs often present partially aligned, high spins in our simulations, systems with low and/or misaligned spins are also possible in the three models listed above.

The merger rates for popIII models are rather insensitive to the inclusion of delays, but this is not true for the heavy-seed models. This is illustrated in Fig. 3, which shows the predicted MBH merger rates as a function of mass and redshift in the three models considered above. Note that while the popIII and Q3-nod models predict a high merger rate up to $z > 15$, very few or no events at $z > 10$ are expected in the Q3-d model.

The difference in merger rates among the various seed models is due to two factors. First, different models have different mass functions and occupation numbers at high z. Second, whether a MBHB stalls or merges depends on the details of its interactions with the stars, nuclear gas and other MBHs, which depend critically on the MBH masses (cf. [59] for more details on our assumptions regarding these interactions). The different MBH mass functions at high redshifts in the light- and heavy-seed models imply that many more binaries can "stall" at high redshifts in the heavy-seed scenarios. Nevertheless, since our treatment of the delays is quite simplified (in particular when it comes to modeling triple MBH systems and the interaction with nuclear gas), models Q3-d and Q3-nod can be thought of as bracketing the possible range of merger rates.

D. Population completeness

Because our models follow the coevolution of MBHs with galaxies including both their dark-matter and baryonic constituents (cf. Fig. 2), at fixed resolution for the dark-matter merger trees our simulations become computationally expensive for high redshifts and very massive galaxies. For galaxies with dark-matter halo masses $M_H > 10^{13}\ M_\odot$ at $z = 0$, we find that it becomes difficult to resolve the halos where MBHs form at high redshifts within acceptable computational times. Therefore it is possible that we may "miss" merging binaries at high redshift, when simulating the most massive halos at $z = 0$.

We quantify this effect in Fig. 4, which shows the number of MBH mergers as a function of halo mass. The linear trend seen at $M_H < 10^{13}\ M_\odot$ is easily explained. Suppose that seeds form in halos of mass M_S. A halo of mass M_0 has formed from roughly $N = M_0/M_S$ halos of mass M_S. This implies a number of seeds proportional to M_0. Suppose for simplicity that we start off with 2^n seeds. If we consider a perfect hierarchy in which two remnants of a previous round of mergers keep merging with each other until there is only one MBH left, the number of mergers is $\sum_{i=0}^{n-1} 2^i = 2^n - 1$, which approximately matches the number of seeds (i.e., 2^n), and which is therefore proportional to M_0. Although simplistic, this argument highlights the reason why the trend shown in Fig. 4 may hint at a lack of resolution in our simulations for $M_H > 10^{13}\ M_\odot$.

To assess the impact of this issue on our results, we computed the MBH merger rate per unit (dark-matter) mass in the low-mass halos, and used it to correct the merger rates at larger halo masses (cf. the thick lines in Fig. 4). The results of this exercise (reported in Table I) suggest that this lack of resolution may lead to our simulations missing up to a factor of 2 in terms of merger events. In this sense, the event rates in our study are therefore conservative. To further confirm this finding, we also ran a few test simulations with increased halo resolution. These

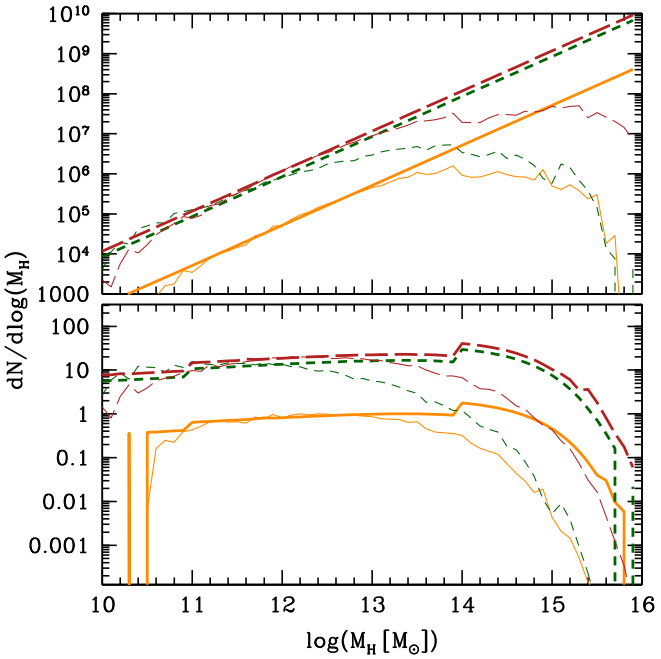

FIG. 4. Contribution of each halo mass to the total merger rate for models popIII (long-dashed brown lines), Q3-nod (short-dashed green lines) and Q3-d (solid orange lines). Lines in the top panel are proportional to the number of mergers per halo (i.e., removing the Press and Schechter weights), whereas lines in the bottom panel represent the halo contribution to the cosmic merger rate $dN/d\log M_H$. In both panels, thin lines are the results of our MBH population models; thick lines are extrapolations assuming a linear relation between the number of mergers and the halo mass.

higher-resolution runs show that the number of mergers is essentially resolution independent at $M_H < 10^{13}\ M_\odot$, but that MBH mergers are much more numerous for $M_H > 10^{13}\ M_\odot$, in line with the expected linear trend with M_H. We also confirmed the expectation that the missing events are mostly seed-mass, high-z MBHBs. This does not make a large difference for eLISA rates in the popIII model, but it may increase more significantly the number of detections in the Q3-nod and Q3-d scenarios, where the MBH seeds are massive enough to be within the instrument's sensitivity range.

TABLE I. MBHB merger rates (total number of mergers per year) along the cosmic history predicted by our population models. The second column reports the rates found in the simulations, the third column reports the rate corrected as described in the text and in Fig. 4, and the fourth column is the ratio of the two (i.e., the expected level of incompleteness of the populations adopted in this study).

Model	Total rate	Extrapolated rate	Ratio
popIII	175.36	332.65	1.89
Q3-d	8.18	14.06	1.72
Q3-nod	121.80	240.96	1.98

IV. GRAVITATIONAL WAVEFORMS AND DATA ANALYSIS

An accurate waveform model, encapsulating the complexity produced by a potentially precessing spinning MBHB, is required in order to make a realistic assessment of eLISA's capabilities. Since many MBHBs merge within the detector band, including merger and ringdown in the computation is also crucial. Unfortunately, inspiral-merger-ringdown (IMR) waveform models for precessing-spinning binaries suitable for efficient parameter estimation are still under active development [81–83]. We therefore employ the following strategy:

(a) our core analysis is based on a generic precessing inspiral-only gravitational waveform [the shifted uniform asymptotics (SUA) waveform described in [84], see next section], with the detector response modeled as in [46] (see also [85–87] for similar studies in the context of LISA);
(b) since the signal-to-noise ratio (SNR) ρ of the inspiral phase depends only mildly on spins, we compared our precessing waveforms to the restricted second-order post-Newtonian (2PN) waveform described in [88,89], ensuring that they yield comparable SNR distributions and detection rates;
(c) results are then rescaled with the aid of the spin-aligned IMR waveform family of [90] (commonly referred to as "PhenomC") in combination with a restricted set of dedicated precessing IMR hybrid waveforms, which are constructed from numerical relativity (NR) simulations stitched to a post-Newtonian (PN) model of the early inspiral.

In the following, we first describe our core inspiral-precessing waveform model (Sec. IV A). Then we briefly summarize the basics of the adopted Fisher matrix analysis (Sec. IV B) and our IMR rescaling (Sec. IV C).

A. Inspiral-precessing waveform model

The spacecraft in all eLISA configurations considered in this study share the same orbits, modulo a rescaling proportional to the detector arm length. If we choose an orthogonal reference system $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$ in the orbital plane tied to the detector arms, in a fixed Solar System frame $(\hat{\boldsymbol{x}}',\hat{\boldsymbol{y}}',\hat{\boldsymbol{z}}')$ tied to the ecliptic (with $\hat{\boldsymbol{z}}'$ perpendicular to the ecliptic) we can write [46,52]

$$\hat{\boldsymbol{x}} = \left(\frac{3}{4}-\frac{1}{4}\cos 2\Phi\right)\hat{\boldsymbol{x}}' - \frac{1}{4}\sin 2\Phi\hat{\boldsymbol{y}}' + \frac{\sqrt{3}}{2}\cos\Phi\hat{\boldsymbol{z}}', \quad (3)$$

$$\hat{\boldsymbol{y}} = -\frac{1}{4}\sin 2\Phi\hat{\boldsymbol{x}}' + \left(\frac{3}{4}+\frac{1}{4}\cos 2\Phi\right)\hat{\boldsymbol{y}}' + \frac{\sqrt{3}}{2}\sin\Phi\hat{\boldsymbol{z}}'. \quad (4)$$

A third vector $\hat{\boldsymbol{z}} = \hat{\boldsymbol{x}}\times\hat{\boldsymbol{y}}$ completes the three-dimensional orthogonal reference system. The constellation drifts away from Earth at the rate of 7.5° per year, so that $\dot{\Phi} = 2\pi(352.5/360)/\mathrm{yr}$.

Consider a binary with orbital angular momentum direction $\hat{\boldsymbol{L}}$, located in a direction specified by the unit vector $\hat{\boldsymbol{N}}$ in the Solar System frame. The response of a single (four-link) detector to the GWs emitted by such a binary can be described as [46]

$$h = \frac{\sqrt{3}}{2}(F_+ h_+ + F_\times h_\times), \tag{5}$$

where

$$F_+(\theta,\phi,\psi) = \frac{1}{2}(1+\cos^2\theta)\cos 2\phi\cos 2\psi - \cos\theta \sin 2\phi \sin 2\psi, \tag{6}$$

$$F_\times(\theta,\phi,\psi) = F_+(\theta,\phi,\psi-\pi/4), \tag{7}$$

$$\cos\theta = \hat{\boldsymbol{N}}\cdot\hat{\boldsymbol{z}}, \tag{8}$$

$$\tan\phi = \frac{\hat{\boldsymbol{N}}\cdot\hat{\boldsymbol{y}}}{\hat{\boldsymbol{N}}\cdot\hat{\boldsymbol{x}}}, \tag{9}$$

$$\tan\psi = \frac{[\hat{\boldsymbol{L}} - (\hat{\boldsymbol{L}}\cdot\hat{\boldsymbol{N}})\hat{\boldsymbol{N}}]\cdot\hat{\boldsymbol{z}}}{(\hat{\boldsymbol{N}}\times\hat{\boldsymbol{L}})\cdot\hat{\boldsymbol{z}}}. \tag{10}$$

In the six-link case, we model the response of a second independent detector as

$$h^{(\mathrm{II})} = \frac{\sqrt{3}}{2}(F_+^{(\mathrm{II})} h_+ + F_\times^{(\mathrm{II})} h_\times), \tag{11}$$

$$F_{+,\times}^{(\mathrm{II})}(\theta,\phi,\psi) = F_{+,\times}(\theta,\phi-\pi/4,\psi). \tag{12}$$

In addition, we have to take into account the fact that the barycenter of eLISA is traveling around the Sun. Instead of modeling this using a Doppler phase as in [46], we prefer to time shift the waveform accordingly:

$$h(t) = F_+(t)h_+(t-t_D) + F_\times(t)h_\times(t-t_D), \tag{13}$$

$$t_D = \frac{R}{c}\sin\theta'\cos(\Phi-\phi'), \tag{14}$$

where $R = 1$ AU, and (θ',ϕ') are the spherical angles of $\hat{\boldsymbol{N}}$ in the Solar System frame. The two descriptions are equivalent in the limit where the orbital frequency varies slowly with respect to the light-travel time across the orbit of the constellation: $\dot{\omega}/\omega \ll c/R$. While this condition is satisfied in the early inspiral, it breaks down near merger.

We decompose the time-domain waveform $h(t)$ into a sum of orbital harmonics as

$$h_{+,\times} = \sum_n A_{+,\times}^{(n)}(\iota) e^{in\varphi}, \tag{15}$$

$$\cos\iota = -\hat{\boldsymbol{L}}\cdot\hat{\boldsymbol{N}}, \tag{16}$$

$$\varphi = \phi_C + \phi_T, \tag{17}$$

$$\phi_C = \phi_{\mathrm{orb}} - 3v^3(2-v^2)\log v, \tag{18}$$

where ι is the inclination angle (defined with a minus sign to agree with the common convention in the literature), $A_{+,\times}^{(n)}(\iota)$ can be found at 2.5PN in [91] and at 3PN in [92,93], ϕ_{orb} is the orbital phase of the binary, $v = (M\omega)^{1/3}$ (with M being the total binary mass) is a post-Newtonian parameter, ϕ_C is the carrier phase and ϕ_T is the Thomas phase, taking into account the fact that the orbital plane is precessing [94] and satisfying

$$\dot{\phi}_T = -\frac{\cos\iota}{1-\cos^2\iota}(\hat{\boldsymbol{L}}\times\hat{\boldsymbol{N}})\cdot\dot{\hat{\boldsymbol{L}}}. \tag{19}$$

We use a signal $h(t)$ in the so-called TaylorT4-form at 3.5PN order, i.e., we integrate the following equations of motion:

$$M\dot{\phi}_{\mathrm{orb}} = v^3, \tag{20}$$

$$M\dot{v} = v^9\sum_{n=0}^{7} a_n v^n, \tag{21}$$

together with the equations of precession at 3.5PN spin-orbit [95] and 2PN spin-spin orders [96]

$$M\dot{\hat{\boldsymbol{L}}} = -v^6(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2), \tag{22}$$

$$M\dot{\boldsymbol{s}}_A = \mu_B v^5 \boldsymbol{\Omega}_A, \tag{23}$$

$$\boldsymbol{\Omega}_A = (C_{A,0} + C_{A,2}v^2 + C_{A,4}v^4 + Dv)\hat{\boldsymbol{L}}\times\boldsymbol{s}_A + \frac{1}{2}v\boldsymbol{s}_B\times\boldsymbol{s}_A. \tag{24}$$

Here $\boldsymbol{s}_A = \boldsymbol{S}_A/m_A M$ are the dimensionless reduced spins, $\mu_A = m_A/M$ are the dimensionless individual masses, and the couplings are

$$C_{A,0} = 2\mu_A + \frac{3}{2}\mu_B, \tag{25}$$

$$C_{A,2} = 3\mu_A^3 + \frac{35}{6}\mu_A^2\mu_B + 4\mu_A\mu_B^2 + \frac{9}{8}\mu_B^3, \tag{26}$$

$$C_{A,4} = \frac{27}{4}\mu_A^5 + \frac{31}{2}\mu_A^4\mu_B + \frac{137}{12}\mu_A^3\mu_B^2 + \frac{19}{4}\mu_A^2\mu_B^3 + \frac{15}{4}\mu_A\mu_B^4 + \frac{27}{16}\mu_B^5, \tag{27}$$

$$D = -\frac{3}{2}\hat{\boldsymbol{L}} \cdot (\boldsymbol{s}_1 + \boldsymbol{s}_2). \tag{28}$$

To compute the Fourier transform of the waveform, we use a SUA transformation [84]. We first separate each harmonic of the waveform into parts varying on the orbital time scale and parts varying on the precession time scale:

$$h^{(n)}(t) = h^{(n)}_{\rm prec}(t) h^{(n)}_{\rm orb}(t), \tag{29}$$

$$h^{(n)}_{\rm prec}(t) = \{F_+(t) A^{(n)}_+(t - t_D) + F_\times(t) A^{(n)}_\times(t - t_D)\} e^{in\phi_T(t-t_D)}, \tag{30}$$

$$h^{(n)}_{\rm orb}(t) = e^{in\phi_C(t-t_D)}. \tag{31}$$

The Fourier transform of the signal is then given by

$$\tilde{h}(f) = \sum_n \tilde{h}^{(n)}(f), \tag{32}$$

$$\tilde{h}^{(n)}(f) = \sqrt{2\pi} T e^{i[2\pi f t_0 - n\phi_C(t_0 - t_D) - \pi/4]} \times \sum_{k=0}^{k_{\rm max}} \frac{a_k}{2} [h^{(n)}_{\rm prec}(t_0 + kT) + h^{(n)}_{\rm prec}(t_0 - kT)], \tag{33}$$

where a_k are constants satisfying the system

$$\frac{(-i)^p}{2^p p!} = \sum_{k=0}^{k_{\rm max}} a_k \frac{k^{2p}}{(2p)!} \tag{34}$$

for $p \in \{0, \ldots, k_{\rm max}\}$, and t_0 and T are defined through

$$2\pi f = n\dot{\phi}_{\rm orb}(t_0 - t_D), \tag{35}$$

$$T = \sqrt{\frac{1}{n\ddot{\phi}_{\rm orb}(t_0 - t_D)}}. \tag{36}$$

In our simulations we used the value $k_{\rm max} = 3$ as a good compromise between computational efficiency and waveform accuracy [84].

B. Fisher matrix analysis

A careful estimate of the likely errors in eLISA measurements of MBHB parameters will ultimately require numerical simulations and full evaluations of the multidimensional posterior probability distributions (see e.g. [97]), but these techniques are computationally expensive. The simple Fisher matrix analysis described in this section allows us to efficiently estimate errors on ensembles of thousands of systems in different MBHB population scenarios, and it is expected to be sufficiently accurate in the high SNR regime.

We first define the detector-dependent inner product

$$(a|b) = 4{\rm Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df, \tag{37}$$

where a tilde denotes the Fourier transform, a star denotes complex conjugation and $S_n(f)$ is the one-sided noise power spectral density of the detector, equal to 3/20 times the sky-averaged sensitivity for each configuration given in Eq. (1) [30]. The SNR of the signal h is given by

$$\rho^2 = (h|h). \tag{38}$$

The Fisher information matrix Γ for the signal h has elements

$$\Gamma_{ij} = \left(\frac{\partial h}{\partial \theta^i}\middle|\frac{\partial h}{\partial \theta^j}\right), \tag{39}$$

where $\boldsymbol{\theta}$ is the vector of source parameters. The combined Fisher matrix for several independent detectors is the sum of the single-detector Fisher matrices, $\Gamma = \sum_i \Gamma^{(i)}$, and the combined squared SNR is the sum of the individual squared SNR ρ_i: $\rho^2 = \sum_i \rho_i^2$.

The correlation matrix Σ is the inverse of the Fisher matrix, $\Sigma = \Gamma^{-1}$. The estimated expectation of the statistical error on a parameter $\Delta\theta^i$ is given by the corresponding diagonal element of the correlation matrix, $(\Delta\theta^i)^2 = \Sigma^{ii}$. The estimated error in some function of the parameters is obtained by a linear propagation of errors:

$$(\Delta\alpha)^2 = \sum_{i,j} \frac{\partial\alpha}{\partial\theta^i}\frac{\partial\alpha}{\partial\theta^j}\Sigma^{ij}. \tag{40}$$

The signal from a MBHB in a quasicircular orbit is described by 15 parameters: the sky location of the source in ecliptic coordinates (colatitude, θ, and longitude, ϕ), the luminosity distance, D_l, the time at coalescence, t_c, the total redshifted mass, $M_z = m_{1z} + m_{2z}$, the symmetric mass ratio, $\eta \equiv m_{1z}m_{2z}/M_z^2$, the initial phase, ϕ_0, the dimensionless spin parameters, χ_1 and χ_2, the direction of the spins (two polar angles, θ_{χ_1} and θ_{χ_2}, and two azimuthal angles, ϕ_{χ_1} and ϕ_{χ_2}), the inclination, ι, of the orbital angular momentum with respect to the line of sight and the polarization angle, Ψ. Because the system is precessing, the latter six parameters must be specified at some reference time t_0.

In this study we focus in particular on

(a) the errors in the two redshifted masses (Δm_{1z}, Δm_{2z});
(b) the error in the sky location, related to the errors on the θ and ϕ angles via $\Delta\Omega = 2\pi \sin\theta \sqrt{\Delta\theta\Delta\phi - (\Sigma^{\theta\phi})^2}$;
(c) the error in the luminosity distance, ΔD_l;
(d) the errors in the magnitudes of the two individual spins ($\Delta\chi_1$, $\Delta\chi_2$) and the errors on their misalignment angles

relative to the orbital angular momentum at the innermost stable circular orbit ($\Delta\theta_{\chi_1}$, $\Delta\theta_{\chi_2}$).

Additionally, we use simple analytical expressions (described in Appendix B) to estimate the accuracy in measuring the spin of the MBH remnant $\Delta\chi_r$ from the radiation emitted in the ringdown phase.

We compute the Fisher matrices using the SUA waveform model, but we additionally model the effect of merger and ringdown by rescaling the errors as described in the next section. The rescaling cannot take into account the fact that the eigenvectors of the Fisher matrix are different in the merger-ringdown part and in the inspiral part, so we expect the results of this calculation to be generally conservative.

C. Inspiral-merger-ringdown rescaling

There is a very limited literature trying to estimate the effect of merger and ringdown on parameter estimation for space-based detectors [98–100]. These works focused on specific choices for the intrinsic parameters of the binary, which makes it hard to use their conclusions in population studies. Babak *et al.* [98] first claimed that adding merger and ringdown can provide roughly an order-of-magnitude improvement in angular resolution. McWilliams *et al.* [99] studied the improvement in the estimation of various parameters as a function of time as one transitions from the innermost stable circular orbit to the postmerger phase. They found negligible improvement in the determination of the system's mass, but their Fig. 1 shows that angular accuracy improves by a factor of ∼4 for a four-link configuration, and by a factor of ∼5 to 6 for a six-link configuration, hinting that the inclusion of merger and ringdown may have a greater impact on angular resolution for six-link configurations. Their Fig. 5 claims "2 to 3 orders of magnitude improvement in a mix of the five angular parameters and in $\ln(D_l)$" when the SNR improves by a factor of ∼10 due to the inclusion of merger and ringdown.

Here we estimate the impact of merger and ringdown by extrapolating our results using spin-aligned PhenomC IMR waveforms [90] in combination with a restricted set of dedicated precessing IMR hybrid waveforms, constructed from NR simulations stitched to a PN model of the early inspiral.

We first use the IMR PhenomC waveform model to rescale the SUA waveform SNR as follows. For each MBHB in our catalog, we construct the PhenomC waveform corresponding to the given MBHB parameters. Since the PhenomC model is valid for systems with spins (anti-) aligned with the orbital angular momentum, this requires computing an "effective spin" obtained by projecting the two individual spins along the orbital angular momentum. Note that the SNR produced by PhenomC is a very good proxy (i) when spins are partially (anti-)aligned with the orbital momentum and (ii) for systems with low spin magnitude. The population of MBHBs considered here usually have nearly aligned spins, especially in the small-seed scenario. From the PhenomC waveform, we compute (for each detector configuration) the SNR produced by the inspiral portion of the waveform alone ($\rho_{\mathrm{PhenC,I}}$) and by considering also merger and ringdown ($\rho_{\mathrm{PhenC,IMR}}$). The ratio of the two, $\mathcal{R}(\rho)=\rho_{\mathrm{PhenC,IMR}}/\rho_{\mathrm{PhenC,I}}$, defines the gain due to the inclusion of merger and ringdown. This quantity is shown in Fig. 5 for the Q3-nod model and the N2A1M2L6 configuration. Note that $\mathcal{R}(\rho)\to 1$ for total redshifted masses $M_z < 3\times 10^5\ M_\odot$, but the gain becomes much larger (ranging between ∼4 and ∼20) for $M_z > 3\times 10^6 M_\odot$: at these redshifted masses the inspiral is out of band, and the merger-ringdown contribution to the signal is dominant. The plot also shows $\mathcal{R}(\rho)$ for a family of nonspinning PhenomC waveforms. Although the trend is the same, the average gain is smaller in this case, because highly spinning MBHBs are louder GW sources in the merger-ringdown phase.

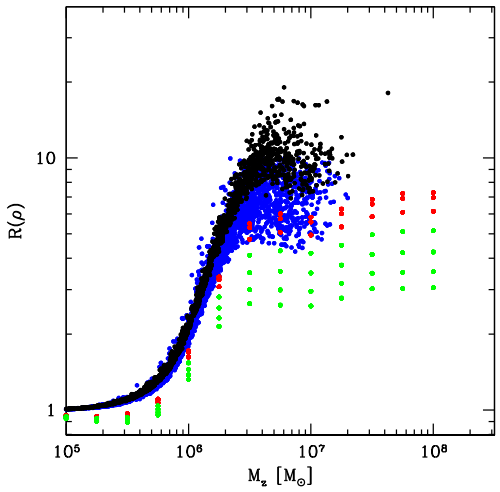

FIG. 5. SNR gain $\mathcal{R}(\rho)$ as a function of redshifted total mass M_z. PhenomC waveforms applied to one realization of the Q3-nod population model are represented by black ($0.5 < m_{2z}/m_{1z} < 1$) and blue ($0 < m_{2z}/m_{1z} < 0.5$) dots. The red and green dots are computed using nonspinning PhenomC waveforms at a fixed M_z for decreasing values of m_{2z}/m_{1z} (from top to bottom); red dots are for $m_{2z}/m_{1z} > 0.5$ and green dots are for $m_{2z}/m_{1z} < 0.5$. This calculation refers to the detector configuration that we labeled N2A1M2L6.

Next, we need to check how the parameter estimation accuracy scales with $\mathcal{R}(\rho)$. To this end we constructed a set of six analytic IMR precessing waveforms by "stitching" NR simulations to PN approximations of the early inspiral phase (see Appendix A for details). Each waveform is constructed for fixed values of the symmetric mass ratio η, of the initial phase ϕ_0 and of the six parameters defining the spin magnitudes and orientations. We are therefore left with seven free parameters determining the response of the detector to a given waveform:

$$\mathcal{C} = \{\theta, \phi, M_z, t_c, D_l, \iota, \Psi\}. \tag{41}$$

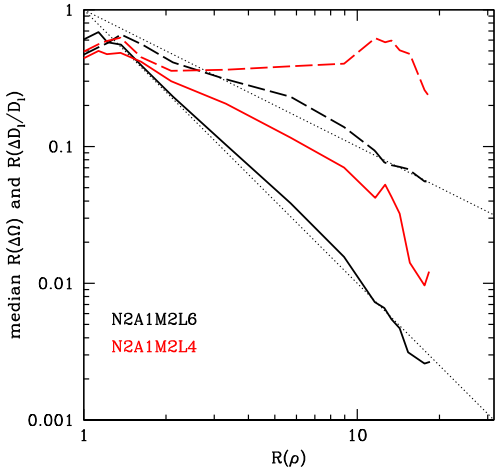

FIG. 6. Median improvement in sky localization $\mathcal{R}(\Delta\Omega)$ (solid lines) and luminosity distance errors $\mathcal{R}(\Delta D_l/D_l)$ (dashed lines) as a function of the SNR gain $\mathcal{R}(\rho)$. Black and red lines are for a six-link and four-link detector configuration, respectively, as indicated in figure. Dotted lines represent linear and quadratic scalings to guide the eye.

We perform 10^4 Monte Carlo drawings of $\mathcal{C}$ by assuming isotropic distributions in all angles, a flat distribution in t_c between one week and two years, a flat-in-log distribution in M_z between 10^5 and 7×10^7 $M_\odot$, and a flat distribution in D_l between 1 and 250 Gpc. We then perform an error analysis using the submatrix of the "complete" Fisher matrix that corresponds to these parameters. For each event, we first compute parameter errors for the inspiral portion of the waveform ($\Delta\mathcal{C}_{\rm I}$), and then for the full hybrid precessing waveform ($\Delta\mathcal{C}_{\rm IMR}$). The ratio $\mathcal{R}(\Delta\mathcal{C})=\Delta\mathcal{C}_{\rm IMR}/\Delta\mathcal{C}_{\rm I}$ is then compared to the ratio $\mathcal{R}(\rho)=\rho_{\rm IMR}/\rho_{\rm I}$. For a fixed $\mathcal{R}(\rho)$ we find a fairly wide range $\mathcal{R}(\Delta\mathcal{C})$, depending on the parameters of the system (i.e., sky location, inclination, etc.), and we consider the median of the distribution of $\mathcal{R}(\Delta\mathcal{C})$ as a function of $\mathcal{R}(\rho)$. The results for the median $\Delta\Omega$ and $\Delta D_l/D_l$ are shown in Fig. 6 for the detector configurations N2A1M2L4 and N2A1M2L6, and for the waveform Q_2 in Table VII. The figure indicates that (i) for a six-link detector, $\mathcal{R}(\Delta\Omega)\propto[\mathcal{R}(\rho)]^{-2}$ and $\mathcal{R}(\Delta D_l/D_l)\propto[\mathcal{R}(\rho)]^{-1}$, as one would expect from analytical scalings; (ii) for a four-link detector, $\mathcal{R}(\Delta\Omega)\propto[\mathcal{R}(\rho)]^{-1}$ and $\mathcal{R}(\Delta D_l/D_l)\approx0.5$. The latter result is indicative of parameter degeneracy preventing an optimal scaling. Since mass ratios and spins have been fixed, we cannot use this model to scale errors on these parameters. However, we notice that in the merger-ringdown phase, the waveform is characterized by the mass and spin of the MBH remnant (and not of the individual progenitors). It is therefore unlikely that a full error analysis on IMR waveforms would lead to significant improvements in the errors Δm_{1z}, Δm_{2z}, $\Delta\chi_1$, $\Delta\chi_2$.

Based on these scaling estimates, we tentatively extrapolate the results obtained from the precessing inspiral-only waveforms as follows:

(1) For each MBHB we compute the ratio $\mathcal{R}(\rho)=\rho_{\rm PhenC,IMR}/\rho_{\rm PhenC,I}$ using a PhenomC waveform.
(2) We rescale the SNR computed using SUA waveforms by the factor $\mathcal{R}(\rho)$.
(3) We finally rescale the errors $\Delta\Omega$ and $\Delta D_l/D_l$ as described above (but we do not apply any correction to the mass and spin determination errors), to get what we refer to as an "SUA IMR" estimate.

We caution that the scaling is based on the analysis of a seven-parameter Fisher submatrix using a restricted number of selected waveforms, and we could not check whether it holds when the full set of 15 parameters is considered. As such, the IMR results presented below should only be taken as roughly indicative of the effect of adding merger and ringdown. A more rigorous and comprehensive study taking into account the impact of systematic errors is a topic for future work, and it will be crucial to assess the accuracy of our rough estimates and to improve upon them.

V. RESULTS

The models presented in Sec. III were used to generate Monte Carlo catalogs of the population of coalescing MBHBs, for a total observation time of 50 years (i.e., in terms of plausible eLISA lifetimes, ten realizations of five-year catalogs, or 25 realizations of two-year catalogs). The MBH masses, redshifts, spin orientations and magnitudes were chosen according to the output of our semianalytical galaxy formation model, by using appropriate smoothing kernels. The other "extrinsic" parameters (sky location, inclination, polarization angles, time and GW phase at merger), which are not provided by our model, are randomized by assuming either uniform distributions or isotropic angular distributions.

The gravitational waveforms used to model the signal of each merger event in the catalogs were described in Sec. IV. Our parameter estimation analysis is based on the SUA model introduced in Sec. IV A, which includes precession and higher harmonics, taking advantage of the information on the spin magnitudes and orientations provided by our galaxy-formation model. The implementation of precession makes it necessary to taper the waveform toward merger by introducing a window function. This causes a partial damping of the SUA waveform amplitude close to merger. To quantify how this affects the detection rates, we compared it with a 2PN model [88,89] that has no spins and no higher harmonics. The 2PN waveform does not carry any information related to spin precession, and it was not used in our parameter estimation calculations. The impact of merger and ringdown was quantified by the extended SUA IMR model, constructed as detailed in Sec. IV C. Recall that the SUA IMR model is only used to rescale errors in the sky location and luminosity distance.

In the following we compare the performance of all 12 eLISA baselines described in Sec. II for mission durations of two and five years. As already mentioned, a longer

integration time allows the resolution and subtraction of more individual CWDs. This effect is expected to have an impact on the recovery of MBHB signals, but we neglect it here and use Eq. (2) for the CWD noise for both two- and five-year mission lifetimes. Note that even with this approximation, the performance of the instrument does not always scale trivially with the mission lifetime. This is because signals are long lasting in the detector band (especially for massive nearby sources), and longer observations lead to a slightly better-than-linear improvement in the detector performance. This is especially true for the most sensitive baselines, for which several sources generate detectable GWs in the detector band for more than two years. As already mentioned, in the discussion of the results we average over ten independent realizations of the eLISA MBHB data stream in the case of a five-year mission, and over 25 realizations in the case of a two-year mission. Therefore all of our results should be understood as accurate within some Poissonian noise, which we omit in all figures and tables to improve readability.

A. Detection rates

The number of events that would be observed (with a threshold $\rho = 8$ on the SNR) by different eLISA configurations in a five-year mission (averaged over the ten catalog realizations), as well as the number of events with $z > 7$ (roughly corresponding to the farthest observed quasar at the moment of writing [101]), are presented in Table II. These numbers are calculated by using three models: the inspiral-only restricted 2PN waveform model, the (inspiral-only) SUA model and the SUA model with the merger-ringdown correction (SUA IMR). Detection rates scale linearly with the mission duration, well within the Poissonian error due to the stochastic nature of our cosmological models. Therefore, to a very good approximation, a two-year mission would observe a number of events that can be obtained by multiplying the values in Table II by a factor of 0.4.

In the popIII scenario, Table II shows that both the overall detection rates and the detection rates at $z > 7$ are remarkably similar for all waveform models. This is because detectable mergers in the popIII scenario are dominated by low-mass systems (cf. Fig. 3), and therefore neither the tapering of the SUA waveform caused by the window function nor the addition of the merger and ringdown makes a significant difference in the SNR, compared to a simple 2PN waveform.

In the heavy-seed scenarios (Q3-d and Q3-nod) the number of detections is instead waveform dependent, especially as the detector becomes less sensitive. This is also to be expected, because most detectable events have total MBHB redshifted mass $10^5 < M_z < 10^6$. These binaries merge well inside the eLISA band, so the SNR is very sensitive to the final portion of the inspiral (and to whether we include merger or not). For SUA waveforms the tapering at the end of the inspiral tends to suppress the SNR, resulting in fewer detections than with restricted 2PN waveforms, which have a hard cutoff at the innermost stable circular orbit. On the contrary, adding merger to the SUA waveforms significantly boosts the SNR, resulting in more detections than with restricted 2PN waveforms. The inclusion of the merger is especially important for events with $z > 7$ and less sensitive detector configurations—particularly those with high low-frequency noise (N1) and/or short arm length (A1, A2), for which the SNR is dominated by the high-frequency part of the waveform. This consideration highlights the

TABLE II. Number of detected MBH mergers for the three MBH population models discussed in the text, using the two different waveform models (SUA and restricted 2PN) discussed in the text. Numbers in parentheses (IMR) are for SUA waveforms, with the SNR rescaled to account for the contribution of merger and ringdown as described in Sec. IV. For each model we report both the overall number of detections and only those at $z > 7$, assuming a five-year mission lifetime and a detection threshold $\rho = 8$ on the SNR. Approximate rates for a two-year mission can be obtained by multiplying by 0.4.

	SUA (IMR)						Restricted 2PN					
	popIII		Q3-nod		Q3-d		popIII		Q3-nod		Q3-d	
Configuration ID	All	$z>7$	All	$z>7$	All	$z>7$	All	$z>7$	All	$z>7$	All	$z>7$
N2A5M5L6	659.7(660.4)	401.1(401.1)	595.6(611.8)	342.6(358.0)	40.4(40.8)	3.6(3.6)	665.8	402.7	610.2	357.0	40.4	3.6
N2A5M5L4	510.7(511.8)	277.5(277.5)	555.6(608.7)	306.4(355.0)	40.2(40.8)	3.4(3.6)	507.6	278.5	602.4	349.8	40.4	3.6
N2A2M5L6	356.8(357.9)	160.1(160.1)	558.8(609.4)	307.6(355.9)	40.2(40.8)	3.6(3.6)	359.3	162.6	593.8	341.8	40.4	3.6
N2A2M5L4	233.1(235.0)	78.8(78.8)	495.9(598.1)	253.2(346.1)	39.8(40.8)	3.4(3.6)	223.4	76.8	557.5	309.6	39.9	3.6
N2A1M5L6	157.6(159.5)	34.9(34.9)	498.1(602.9)	251.6(350.0)	39.1(40.8)	3.1(3.6)	152.4	34.6	570.5	320.0	40.4	3.6
N2A1M5L4	97.2(99.9)	16.4(16.4)	417.9(574.1)	186.8(327.5)	37.9(40.6)	2.8(3.4)	96.3	14.9	519.1	278.2	39.1	3.3
N1A5M5L6	246.6(249.3)	86.8(86.8)	416.2(598.3)	177.5(345.5)	37.5(40.8)	2.5(3.6)	245.9	87.0	533.0	283.9	39.9	3.6
N1A5M5L4	153.9(158.7)	36.1(36.1)	342.9(565.4)	125.6(317.7)	33.7(40.7)	2.0(3.5)	149.1	35.6	470.8	231.6	38.7	3.4
N1A2M5L6	118.7(122.1)	22.5(22.5)	255.7(554.2)	66.5(305.0)	27.8(40.8)	1.1(3.6)	120.3	21.9	398.2	167.5	36.8	2.4
N1A2M5L4	70.6(78.0)	8.0(8.1)	189.7(484.1)	37.3(249.0)	22.4(40.6)	0.7(3.4)	69.5	7.8	316.7	113.4	31.1	1.8
N1A1M5L6	48.8(58.6)	3.9(4.1)	142.1(456.4)	17.0(223.0)	16.8(40.1)	0.5(3.4)	56.1	4.1	262.0	69.6	29.2	1.1
N1A1M5L4	28.4(38.2)	1.3(1.5)	95.3(371.4)	6.1(161.5)	11.7(38.5)	0.3(2.9)	35.4	1.4	193.5	39.3	24.0	0.7

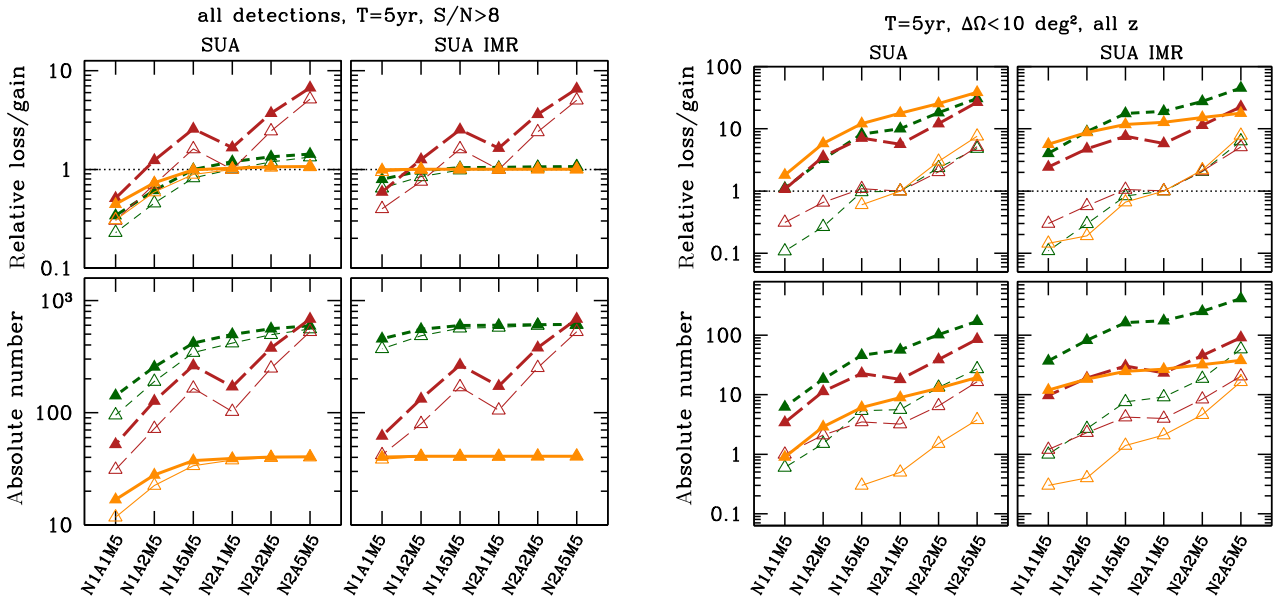

FIG. 7. Total number of detections (i.e., sources with $\rho > 8$, left plot) and total number of detections with $\Delta\Omega < 10$ deg^2 (right plot) assuming a five-year mission (M5). In each plot, the left and right panels are for inspiral and IMR-rescaled waveforms, respectively. The bottom panels represent the Absolute number of detections for different eLISA configurations, while the top panels represent the gain/loss of a given configuration with respect to the standard NGO design, i.e., the ratio [number of sources for (NiAjMkLℓ)]/[number of sources with (N2A1MkL4)]. Long-dashed brown lines are for model popIII, solid orange lines for model Q3-d, and short-dashed green lines for model Q3-nod. Thick lines with filled triangles are for six links (L6), while thin lines with open triangles are for four links (L4).

importance of having accurate IMR waveform models *even for detection*, and not just for parameter estimation.

A comparative view of the performance of the different designs is given in the left panel of Fig. 7. In this figure (and in the following ones) thick lines with filled triangles refer to six-link configurations (L6), while thin lines with open triangles refer to four-link configurations (L4). Long-dashed brown lines refer to model popIII, solid orange lines to model Q3-d and short-dashed green lines to model Q3-nod. The bottom panels represent the absolute number of detections as a function of the eLISA configuration, while the top panels represent the gain/loss of a given configuration with respect to the standard NGO design [52], i.e., the ratio [number of sources for (NiAjMkLm)]/[number of sources for (N2A1MkL4)]. The figure shows that in terms of event rates alone, four- or six-link configurations yield relatively similar results: roughly speaking, the SNR of an event only increases by a factor of $\sqrt{2}$ as we move from a four-link (single-detector) configuration to a six-link (two-detector) configuration. However, the arm length (A1, A2 or A5) and the level of the low-frequency noise (either N1 or N2) are of key importance. Either of these factors can modify the event rates by more than a factor of 10, depending on the MBH population scenario. For instance, the N1A1 configurations are likely to see just a few tens of MBHBs in a five-year mission in the "conservative/realistic" popIII and Q3-d models. Even more dramatically, as can be seen from Table II, these same configurations are likely to see at most a handful of binaries at $z > 7$ in the popIII and Q3-d models. This could severely jeopardize the mission's potential to investigate the origin of MBH seeds at high redshifts.

The right panel of Fig. 7 shows an example of the potential advantages of a six-link configuration in terms of science return. We compare the number of sources that can be localized in the sky within 10 deg^2, a figure of merit indicative of how many detections can be used for electromagnetic follow-up observations (a 10 deg^2 error box is comparable to the SKA and LSST fields of view). On average, six-link configurations perform about ten times better than their four-link counterparts. The difference is even larger when the SUA IMR scaling is adopted, because the improvement in parameter estimation is more prominent for six links (cf. Sec. IV C). Note that any six-link configuration performs better than NGO for all the considered MBHB population models, highlighting the importance of adopting this feature in the mission design. Including merger somewhat mitigates the difference across designs for six-link configurations, but a factor of ∼10 difference still persists between the best and the worst configuration (see e.g. the top-right plot in the right panel of Fig. 7).

B. Parameter estimation

We assess the accuracy with which various eLISA configurations can estimate MBHB parameters using the Fisher matrix approach described in Sec. IV B,

either with inspiral-only SUA waveforms or including a merger-ringdown correction as described in Sec. IV C. As a sanity check, we verified that qualitatively similar trends for the parameter estimation errors are found with an independent Fisher matrix code employing restricted 2PN, nonspinning waveforms [88] (although the absolute errors are typically larger for the 2PN models, which omit spin precession information).

Our main goal is to assess the scientific return of the mission, so we report mostly the number of systems for which selected parameters can be measured within a certain error, rather than the average (or median) absolute errors on those parameters. This representation is more directly linked to the mission's science goal of testing the formation and evolution of the MBH population, which requires parameters to be measured with reasonable precision for a large sample of the astrophysical MBH population. For other mission goals, it might be more appropriate to quote the absolute errors: for instance, in order to test the black hole no-hair theorem of GR using MBH mergers, a single MBHB with very well determined parameters (remnant spin and dominant quasinormal mode frequencies [30,102]) might be enough. We will keep this in mind below (e.g. when we report the absolute error with which the final remnant spin can be measured), but we defer a more complete analysis of absolute errors and tests of GR to future work.

Our "success metrics" to assess the science capabilities of various mission designs are the expected number of observed binaries that meet one or more of the following conditions:

(i) Both redshifted masses (m_{1z} and m_{2z}) are measured with a relative statistical error of 1% or better: $\Delta m_{1z}/m_{1z} < 0.01$ and $\Delta m_{2z}/m_{2z} < 0.01$. This metric is useful to gauge the mission's capability to probe MBH growth across cosmic history.

(ii) Spin magnitudes and directions are measured accurately. For the spin magnitude, we require that either the spin parameter χ_1 of the more massive black hole (the "primary") be measured with absolute statistical error of 0.01 or better, or that the spin parameter χ_2 of the less massive black hole (the "secondary") be measured with absolute statistical error of 0.1 or better. Note that we use different thresholds because the secondary's spin is typically harder to measure. As for the spin direction, we require that the angles of both spins with respect to the orbital angular momentum of the system (θ_{χ_1}, θ_{χ_2}) be determined to within an error of 10° or less. Spin magnitudes and directions are related to the global accretion history and to the local dynamics of the accretion flow [56,57,103], as well as to the interaction between the MBHs and the gas in the nuclear region via the Bardeen-Petterson effect [104–106]. Therefore this metric is useful to gauge the mission's capability to probe the nature of MBH feeding by discriminating the role of various growth mechanisms (coherent vs chaotic gas accretion, mergers, etc.).

(iii) The remnant spin parameter χ_r is measured with (statistical) absolute precision of 0.1 or better. This is particularly useful for tests of GR and tests of the black hole no-hair theorem.

(iv) The statistical sky position error is $\Delta\Omega < 10\,\mathrm{deg}^2$, the statistical relative error on the luminosity distance is of 10% or better ($\Delta D_l/D_l < 0.1$) and $z < 5$. These systems have sky localization error comparable to (or smaller than) the SKA and LSST fields of view, and they are close enough that a possible transient electromagnetic counterpart might be identified, thus allowing us to measure their redshift and potentially test the cosmological $D_l(z)$ relation (and therefore the Hubble parameter, the composition of the Universe, and the equation of state of dark energy).

(v) The statistical error on the luminosity distance is $\Delta D_l/D_l < 0.3$ and $z > 7$. These are the systems that will provide the most information on the formation of MBH seeds at high z. Clearly, in order to test competing scenarios for MBH seeds we should be able to detect MBHBs at high redshift, but we should also be able to ascertain that they *are* indeed at high redshift (hence the requirement on $\Delta D_l/D_l$, which can be translated into a requirement on the redshift error by assuming a standard ΛCDM cosmology).

The results for these figures of merit are reported for all configurations in Tables III and V (for a two-year mission) and in Tables IV and VI (for a five-year mission). The same results are also represented graphically in Figs. 7–12, and are briefly described below.

(i) All configurations allow the precise measurement of MBHB masses for at least a few systems (see Fig. 8). Note, however, that these numbers get dangerously close to one for the worst performing mission designs. The number of links does not have a strong impact on this metric. However it is essential to accumulate cycles in band, either with a long arm length (A5) or by preserving the target low-frequency noise (N2).

(ii) Similar considerations apply to spin measurements. In order to determine the spin magnitude (Fig. 9), it is essential to achieve the target low-frequency noise (N2), since we typically predict less than a single good spin measurement (especially of the secondary χ_2) for some population models and baselines with suboptimal low-frequency noise (N1). Spin directions (Fig. 10) are also much better determined using N2A5M*k*L6 baselines, for which we have more than ten satisfactory measurements for all MBHB population models considered in this study. The errors are generally a

TABLE III. Number of MBHBs detected with specific values of the mass and spin errors. $\Delta m_{1z,2z}/m_{1z,2z}$ is the relative error on each of the MBH masses, $\Delta\chi_1$ and $\Delta\chi_2$ are the absolute errors on the individual spin magnitudes, $\Delta\theta_{\chi_{1,2}}$ is the absolute error on each of the spin misalignment angles with respect to the orbital angular momentum at the innermost stable circular orbit, and $\Delta\chi_r$ is the error on the magnitude on the remnant MBH spin. Numbers are for a two-year mission lifetime.

	$\Delta m_{1z,2z}/m_{1z,2z}<0.01$			$\Delta\chi_1 < 0.01$			$\Delta\chi_2 < 0.1$			$\Delta\theta_{\chi_{1,2}} < 10$ deg			$\Delta\chi_r < 0.1$		
Configuration ID	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d
N2A5M2L6	146.6	141.8	13.3	45.3	76.8	2.6	41.8	44.7	3.9	21.0	40.9	9.4	3.5	31.4	10.9
N2A5M2L4	94.6	108.5	11.3	32.4	60.5	2.1	21.2	27.2	2.5	11.5	19.1	4.8	3.0	18.5	10.7
N2A2M2L6	71.4	99.6	10.9	28.3	54.4	2.0	17.1	22.2	2.1	11.7	18.9	5.1	3.3	27.0	10.5
N2A2M2L4	40.7	69.1	8.4	19.6	40.8	1.5	8.2	11.1	1.1	6.0	7.7	2.3	2.9	17.0	10.2
N2A1M2L6	30.4	66.4	8.5	18.7	39.3	1.5	7.4	10.8	1.0	6.1	9.2	2.9	3.1	21.3	9.5
N2A1M2L4	15.3	41.2	6.3	13.4	27.6	1.0	3.8	4.9	0.6	3.1	3.0	1.0	2.9	12.3	9.3
N1A5M2L6	40.7	49.3	7.0	20.5	29.8	0.9	7.3	8.0	0.6	5.7	6.8	1.9	3.0	22.1	10.5
N1A5M2L4	18.7	29.8	4.7	14.6	20.3	0.6	3.6	3.7	0.4	2.5	2.2	0.6	2.7	16.5	10.3
N1A2M2L6	11.6	20.4	3.2	12.6	12.6	0.2	2.2	2.4	0.2	1.8	2.2	0.6	2.7	15.0	9.2
N1A2M2L4	4.4	10.1	2.3	7.5	8.2	0.1	1.1	1.0	0.1	0.8	0.6	0.2	2.6	12.1	9.2
N1A1M2L6	3.3	8.7	2.4	4.8	5.7	0.1	0.6	0.6	0.1	0.7	0.6	0.2	2.2	9.1	6.6
N1A1M2L4	1.6	3.8	1.0	2.4	3.3	0.0	0.3	0.4	0.1	0.3	0.2	0.1	2.1	7.8	6.4

TABLE IV. Same as Table III, but for a five-year mission lifetime.

	$\Delta m_{1z,2z}/m_{1z,2z} < 0.01$			$\Delta\chi_1 < 0.01$			$\Delta\chi_2 < 0.1$			$\Delta\theta_{\chi_{1,2}} < 10$ deg			$\Delta\chi_r < 0.1$		
Configuration ID	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d
N2A5M5L6	510.5	406.6	33.5	114.4	199.5	6.9	153.1	130.2	10.4	63.7	111.6	24.2	8.8	78.3	27.2
N2A5M5L4	366.8	328.5	28.7	89.1	160.0	5.4	81.2	82.7	7.4	35.6	56.8	12.1	7.4	46.8	26.1
N2A2M5L6	255.6	300.0	27.4	73.6	140.5	4.7	61.4	66.2	6.0	34.7	53.6	13.2	8.2	67.4	26.2
N2A2M5L4	157.0	219.6	21.0	52.0	106.3	3.7	30.3	34.8	3.6	16.3	22.1	5.8	7.4	40.4	25.4
N2A1M5L6	101.4	214.0	20.7	46.1	101.3	3.3	24.5	32.8	3.4	16.9	24.7	7.1	7.8	52.9	24.0
N2A1M5L4	53.3	142.4	16.0	32.8	69.9	2.3	11.1	16.5	1.9	7.4	8.1	2.6	7.2	30.8	23.1
N1A5M5L6	148.7	164.6	15.5	52.1	73.8	2.2	25.3	23.3	1.9	15.3	17.7	4.7	7.5	55.0	26.3
N1A5M5L4	79.0	104.9	10.7	36.0	53.2	1.5	10.1	11.6	1.0	6.3	5.8	1.4	6.9	37.3	24.7
N1A2M5L6	52.9	75.8	8.4	31.1	33.4	0.6	6.0	5.7	0.6	4.8	4.5	1.7	6.9	38.2	23.3
N1A2M5L4	25.5	43.9	4.9	22.5	20.9	0.5	2.4	3.2	0.3	1.8	1.7	0.3	6.5	26.3	19.1
N1A1M5L6	14.3	34.4	4.0	15.1	13.0	0.4	1.6	2.3	0.2	1.8	1.1	0.4	5.5	23.1	16.6
N1A1M5L4	7.7	16.7	1.9	6.5	8.0	0.2	0.4	1.3	0.1	0.5	0.4	0.0	5.5	14.2	13.9

TABLE V. Number of MBHBs detected within specific values of the sky location and luminosity distance errors, as reported in the table headers. A mission lifetime of two years is assumed.

	$\Delta\Omega < 10\,\mathrm{deg}^2$ & $\Delta D_l/D_l < 0.1$ & $z < 5$						$z > 7$ & $\Delta D_l/D_l < 0.3$					
	SUA			SUA IMR			SUA			SUA IMR		
Configuration ID	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d
N2A5M2L6	14.5	34.8	6.0	16.1	47.4	10.1	71.6	117.2	1.2	71.6	141.1	1.4
N2A5M2L4	3.2	8.7	1.1	4.8	16.0	4.9	10.2	54.4	0.6	30.4	96.8	1.0
N2A2M2L6	6.8	23.2	3.8	9.2	35.2	9.5	20.8	82.6	0.9	20.8	134.4	1.4
N2A2M2L4	1.6	4.2	0.4	2.6	5.8	1.6	2.8	18.0	0.2	10.1	54.0	0.7
N2A1M2L6	3.4	14.9	2.5	5.7	26.4	7.8	3.9	50.9	0.6	3.9	120.1	1.3
N2A1M2L4	0.6	1.7	0.1	1.0	2.6	0.5	0.5	0.8	0.0	2.6	41.8	0.2
N1A5M2L6	4.0	13.7	1.9	7.0	27.3	7.5	9.8	30.5	0.4	9.9	111.9	1.2
N1A5M2L4	0.7	1.6	0.0	1.2	2.6	0.2	1.3	2.2	0.0	5.2	9.0	0.2
N1A2M2L6	1.9	5.1	0.8	4.4	18.0	5.5	2.3	6.6	0.2	2.4	77.7	1.0
N1A2M2L4	0.4	0.5	0.0	0.6	1.0	0.1	0.2	0.4	0.0	1.0	2.0	0.0
N1A1M2L6	0.7	1.5	0.2	2.7	9.8	3.9	0.2	0.1	0.0	0.5	0.4	0.6
N1A1M2L4	0.2	0.2	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0

TABLE VI. Same as Table V, but for a five-year mission lifetime.

	$\Delta\Omega < 10$ deg^2 & $\Delta D_l/D_l < 0.1$ & $z < 5$						$z > 7$ & $\Delta D_l/D_l < 0.3$					
	SUA			SUA IMR			SUA			SUA IMR		
Configuration ID	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d	popIII	Q3-nod	Q3-d
N2A5M5L6	41.0	90.6	14.8	45.0	119.6	26.1	207.1	299.4	3.4	207.1	352.4	3.6
N2A5M5L4	10.5	23.9	3.5	15.7	43.9	13.4	35.3	147.6	1.6	100.6	258.8	2.7
N2A2M5L6	21.0	62.9	9.3	26.4	94.2	23.1	60.6	210.0	2.3	60.6	338.4	3.6
N2A2M5L4	3.9	11.0	1.4	6.4	16.4	3.7	9.7	53.1	0.9	31.4	147.4	1.7
N2A1M5L6	10.7	37.5	6.0	15.2	68.4	19.2	12.1	134.1	1.6	12.1	306.0	3.4
N2A1M5L4	1.9	4.6	0.4	3.0	7.8	1.4	1.9	13.4	0.1	6.3	64.6	0.9
N1A5M5L6	12.3	34.3	4.4	18.9	72.2	18.0	26.9	79.1	1.3	26.9	286.7	3.4
N1A5M5L4	1.9	4.5	0.3	3.4	6.4	1.0	4.2	5.8	0.1	14.4	26.8	0.3
N1A2M5L6	5.5	14.3	2.4	12.0	45.8	13.5	6.1	17.2	0.5	6.3	197.7	2.4
N1A2M5L4	0.8	1.2	0.0	1.2	2.0	0.1	1.3	0.4	0.0	2.7	4.9	0.1
N1A1M5L6	2.0	4.6	0.9	7.9	24.9	9.0	1.0	2.1	0.1	1.3	110.8	1.7
N1A1M5L4	0.2	0.5	0.0	0.7	0.8	0.0	0.1	0.0	0.0	0.6	0.6	0.0

factor of ≈2 worse for four links (L4), and quickly deteriorate by shortening the arm length, if suboptimal low-frequency noise (N1) is assumed.

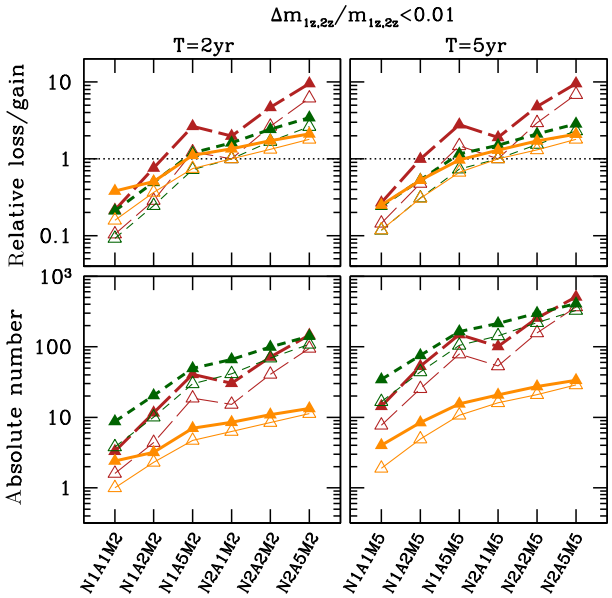

FIG. 8. Number of detections with fractional errors of less than 0.01 for both redshifted masses. The left and right panels are for a mission lifetime of two years and five years, respectively. Inspiral-only waveforms have been used in all cases. The bottom panels represent the number of sources as a function of the eLISA configuration, while the top panels represent the gain/loss of a given configuration with respect to NGO, i.e., the ratio [number of sources for (NiAjMkLℓ)]/[number of sources with (N2A1MkL4)]. Long-dashed brown lines are for model popIII, solid orange lines for model Q3-d and short-dashed green lines for model Q3-nod. Thick lines with filled triangles are for six links (L6), while thin lines with open triangles are for four links (L4).

(iii) Measurements of the remnant spin χ_r are quite insensitive to the details of the detector, as shown in Fig. 11. This is because the individual spins before merger are measured from their imprint on the adiabatic inspiral, and those measurements are sensitive to the detector's arm length and low-frequency noise. Conversely, the postcoalescence MBH mass and spin is measured from the quasinormal modes of the remnant, which lie at higher frequencies, where differences between the various configurations are smaller (cf. Fig. 1).

(iv) The identification of systems suitable for electromagnetic follow-up is heavily dependent on the number of links, as clearly shown in the left panel of Fig. 12. Notice that any six-link configuration (even with pessimistic low-frequency noise N1) performs better than NGO. We also notice that adding merger and ringdown (SUA IMR) mitigates the gap across designs *in the six-link case only*. In the four-link case, the availability of a more complete waveform model does not significantly improve the performance of the modestly performing baselines with respect to this particular metric.

(v) Similar arguments apply to the identification of high-redshift systems (right panel of Fig. 12). As before, adding merger and ringdown (SUA IMR) partially improves the capability of the worst performing six-link designs, but it does not have as much of an impact on the performance of the four-link designs.

From these results we can draw several conclusions about the benefits/drawbacks of specific design choices with respect to concrete science goals. Configurations N1A1M2L4 and N1A1M2L6 will only detect a few systems with mass estimates that are precise at the 1% level or better. Their performance is even worse for the MBH spins. The number of systems with precise measurements of the individual MBH redshifted masses

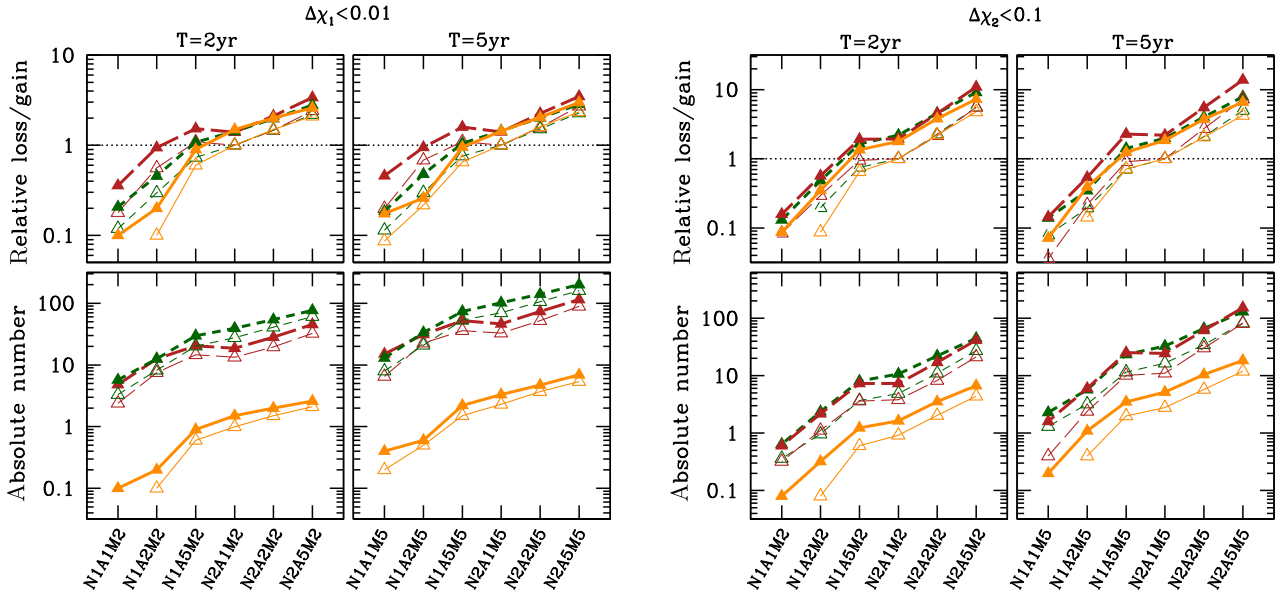

FIG. 9. Number of detections with Absolute error on the primary MBH spin smaller than 0.01 (left plot), and with Absolute error on the secondary MBH spin smaller than 0.1. In each plot, left and right panels are for a mission lifetime of two years and five years, respectively; inspiral-only waveforms have been used in all cases. The bottom panels represent the number of sources as a function of the eLISA configuration, while the top panels represent the gain/loss of a given configuration with respect to NGO, i.e., the ratio [number of sources for (NiAjMkLℓ)]/[number of sources with (N2A1MkL4)]. Long-dashed brown lines are for model popIII, solid orange lines for model Q3-d and short-dashed green lines for model Q3-nod. Thick lines with filled triangles are for six links (L6), while thin lines with open triangles are for four links (L4).

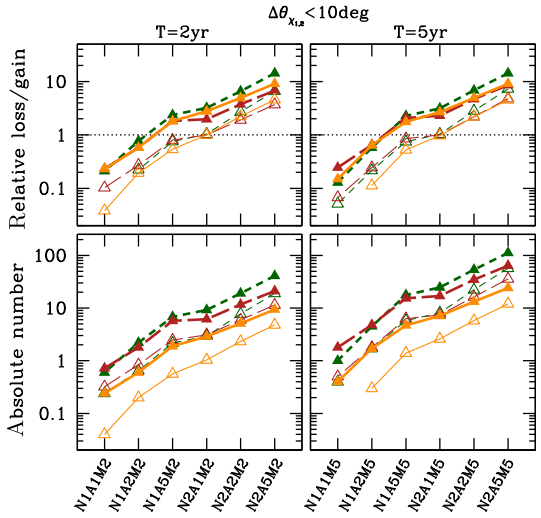

FIG. 10. Number of detections such that the Absolute error in the measurement of both misalignment angles, θ_{χ_1} and θ_{χ_2}, at the innermost stable circular orbit is less than 10°. The left and right panels are for a mission lifetime of two years and five years, respectively. Inspiral-only waveforms have been used in all cases. The bottom panels represent the number of sources as a function of the eLISA configuration, while the top panels represent the gain/loss of a given configuration with respect to NGO, i.e., the ratio [number of sources for (NiAjMkLℓ)]/[number of sources with (N2A1MkL4)]. Long-dashed brown lines are for model popIII, solid orange lines for model Q3-d and short-dashed green lines for model Q3-nod. Thick lines with filled triangles are for six links (L6), while thin lines with open triangles are for four links (L4).

(m_{1z}, m_{2z}), spin magnitudes (χ_1, χ_2) and spin orientation angles at the innermost stable circular orbit $(\theta_{\chi_1}, \theta_{\chi_1})$ varies by a factor of $\sim 30-100$ across different configurations. However, this is not true for the remnant spin χ_r, for the reasons explained above. This means that our ability to probe fundamental physics (e.g. black hole horizons, no-hair theorem, etc) by measuring ringdown modes of merging MBHBs is more intimately related to

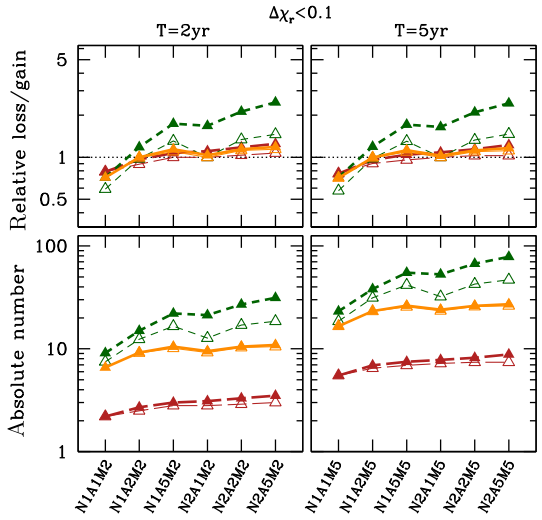

FIG. 11. Same as Fig. 9, but for the remnant spin χ_r.

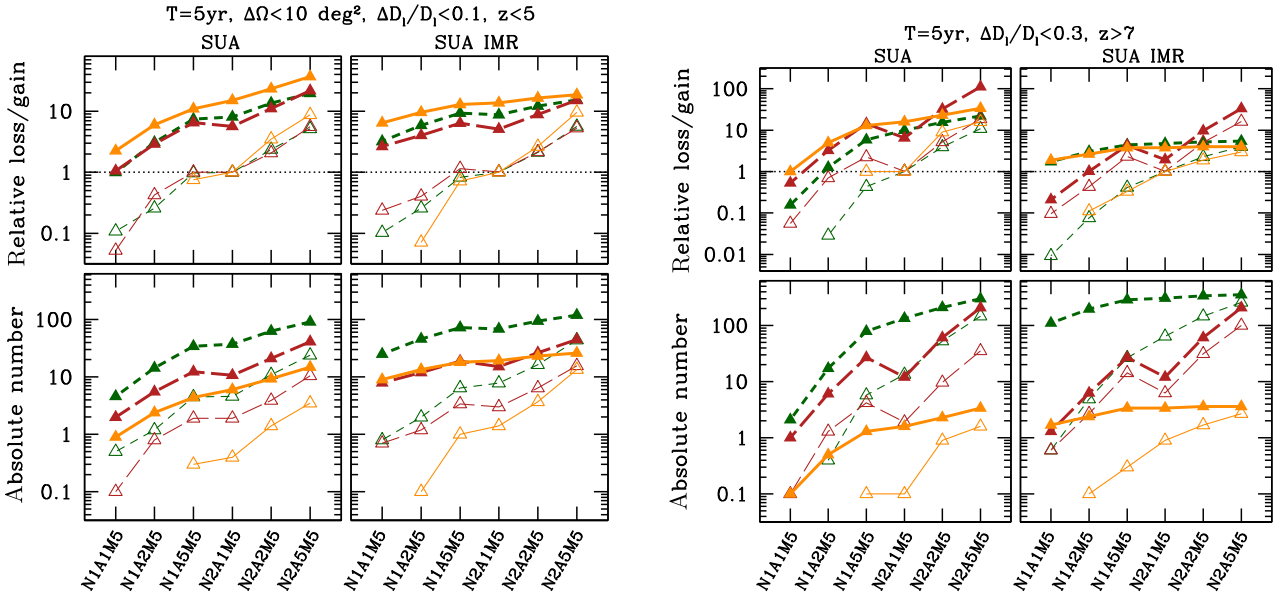

FIG. 12. (Left plot) Total number of detections at $z < 5$ with $\Delta\Omega < 10$ deg^2 and $\Delta D_l/D_l < 0.1$. (Right plot) Total number of detections at $z > 7$ with $\Delta D_l/D_l < 0.3$. In each plot, left and right panels are for inspiral and IMR rescaled waveform, respectively; five years of observations are assumed. The bottom panels represent the number of sources as a function of the eLISA configuration, while the top panels represent the gain/loss of a given configuration with respect to NGO, i.e., the ratio [number of sources for (NiAjMkLℓ)]/[number of sources with (N2A1MkL4)]. Long-dashed brown lines are for model popIII, solid orange lines for model Q3-d and short-dashed green lines for model Q3-nod. Thick lines with filled triangles are for six links (L6), while thin lines with open triangles are for four links (L4).

an adequate knowledge of the waveform and to the intrinsic rate of MBHB mergers in the Universe. The specific detector baseline (at least in the range considered in this study) will affect the results by less than a factor of 4. Only some of the six-link configurations ensure localization of a sufficient number of sources to allow for electromagnetic follow-ups and hence either multimessenger astronomy or (potentially) studies of the dark-energy equation of state, for which $\gtrsim$10 sources would be required. Finally, six-link configurations with good low-frequency noise (N2) are necessary to detect high-redshift systems with a relatively small distance error, thus probing the early epoch of MBH formation. In general, configurations with worse low-frequency noise (N1) and short arm length are expected to provide very few detections at high redshifts. The ability to detect high-redshift MBHs and to localize systems for systematic electromagnetic follow-up are of crucial importance for the "traditional" astronomy community. Our results suggest that six links are a firm requirement to achieve these goals. On the other hand, precise mass and spin measurements are not so sensitive to the number of links, and part of the scientific potential of the mission is preserved even in a four-link scenario.

A different way of comparing instrument performance for different design choices is presented in Fig. 13. Here we have quantified the impact of each single baseline element on selected figures of merit. We generated this figure as follows. For each of the three MBHB population models we considered all pairs of eLISA configurations that differ only by a specific element (number of links, arm length, low-frequency noise) and compared the results of the respective analyses, focusing for simplicity on five-year missions (M5). For example, in assessing the impact of four vs six links (upper-right panel in Fig. 13), we compared the results of our Fisher matrix analysis for each of the three MBHB models (popIII, Q2-nod, Q2d) using the six detector pairs NiAjM5L4 and NiAjM5L6 (i.e., N1A1M5L4 vs N1A1M5L6, N1A2M5L4 vs N1A2M5L6, etc.), for a total of 18 comparisons. This procedure yielded 18 comparisons of L4 vs L6 and N1 vs N2, and 12 comparisons of A2 vs A5 and A1 vs A5. For each metric we plotted a histogram of the ratio of the number of sources satisfying that particular metric across the comparisons. Note that we always take the ratio of the worse over the better configuration (L4/L6, N1/N2, A2/A5, A1/A5), i.e., we quantify the "science loss" related to a specific descoping option. The two top panels show that either dropping the third arm (L6 → L4, top-left panel) or not meeting the target low-frequency sensitivity (N2 → N1, top-right panel) can seriously jeopardize the mission potential. Indeed, the average number (blue triangle) of high-redshift detections and potential electromagnetic follow-up targets might drop by a factor of 10, harming the astrophysical impact of the mission. On the other hand, shortening the arm length to 2 Gm (A5 → A2, bottom-left panel) seems to preserve

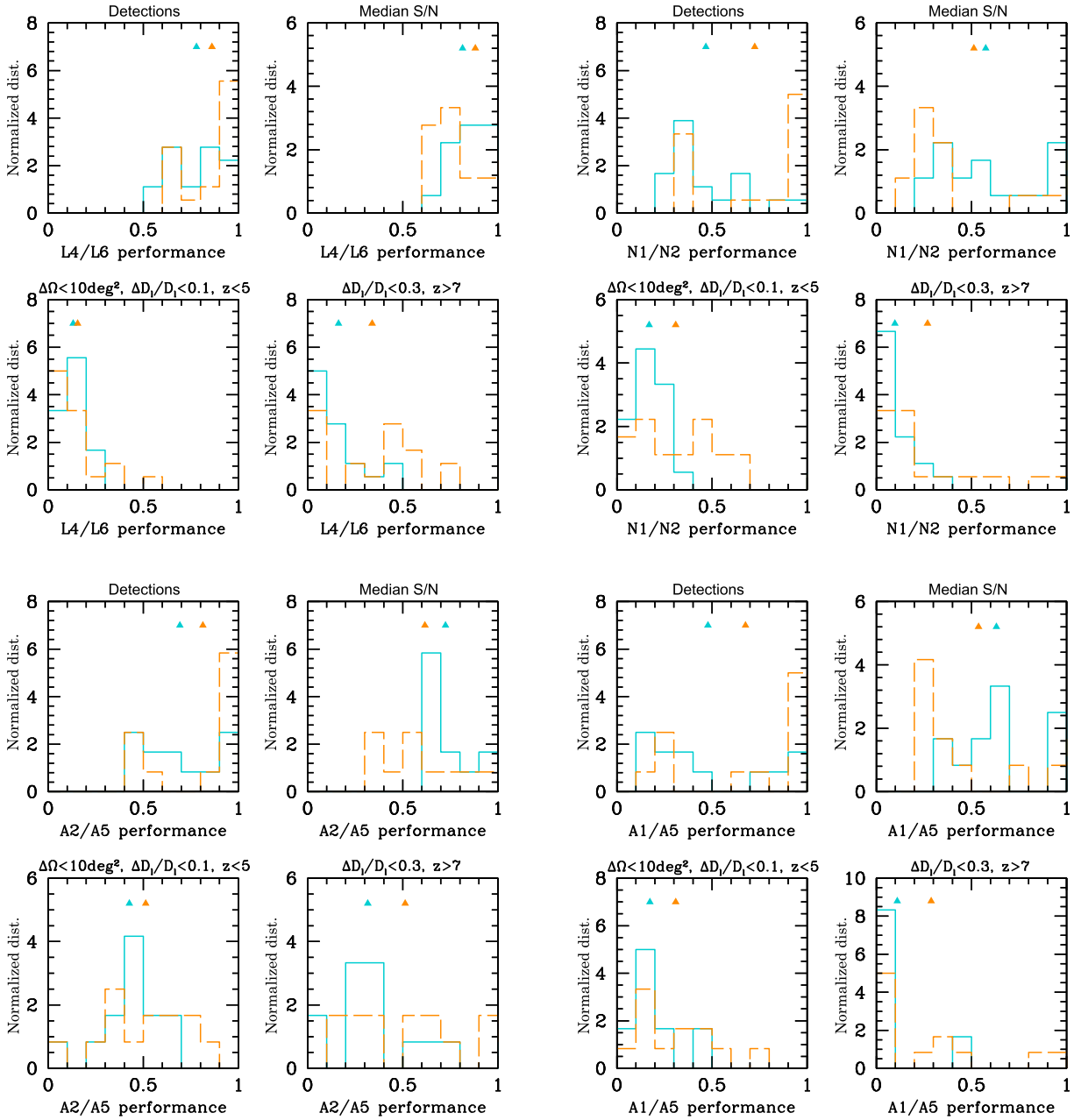

FIG. 13. Science loss as a function of specific design choices. In each plot, we compare the results of simulations that only differ by a specific design element (see the description in the main text). Those are number of links L4/L6 (18 simulations, upper left); low-frequency noise N1/N2 (18 simulations, upper right); arm length A2/A5 (12 simulations, lower left); arm length A1/A5 (12 simulations, lower right). Each of the four plots visualizes the science loss for specific design choices according to four different indicators: total number of Detections (upper left quadrant), Median SNR of detected sources (upper right quadrant), number of low-z sources with good enough sky localization for counterpart searches and good distance determination (lower left quadrant), and number of high-z sources with good distance measurement (lower right quadrant). In each panel we plot the histograms of the indicators over all the simulations, and the mean value (triangles). Blue is for inspiral waveforms only, orange is for IMR-rescaled waveforms.

most of the performance metrics within a factor of 2. Further shortening the arm length to 1 Gm (A2 → A1, bottom-right panel), however, is potentially damaging. Note that these average figures of merit are somewhat waveform dependent. As mentioned above, adding merger and ringdown tends to reduce the gap between different designs, mitigating the average science loss for several of the figures of merit.

VI. CONCLUSIONS

We have investigated the relative performance of different eLISA designs for the study of the formation and growth of MBHs throughout the Universe. We have explored the complex aspects of this problem to the best of our current capabilities, which are necessarily limited by our incomplete knowledge of MBH astrophysics and waveform models. For example, work by Littenberg *et al.* [100] using effective-one-body models to improve on the parameter estimation findings of [98,99] found that systematic errors on NR waveforms were too large to draw any conclusions. It is therefore urgent and important to revisit our study using state-of-the-art waveforms and models for detector design when these are available. In particular, our assessment of the impact of merger and ringdown should be seen as a preliminary estimate.

For the reader's convenience, we summarize here the main science questions that we have addressed (and the figures and tables that summarize our findings):

(i) Can we measure MBH masses and probe the growth of MBHs across cosmic history? (See Fig. 8 and Tables III and IV.)
(ii) Can we measure the premerger spins and probe the nature of MBH feeding? (See Figs. 9 and 10 and Tables III and IV.)
(iii) Can we measure the remnant spin (which allows further tests on MBH feeding, as well as potential tests of the black hole no-hair theorem)? (See Fig. 11 and Tables III and IV.)
(iv) Is the sky localization error small enough for close sources that possible transient electromagnetic counterparts might be identified, thus allowing us to measure the source redshift and potentially test the cosmological $D_l(z)$ relation? (See the left panel of Fig. 12 and Tables V and VI.)
(v) Can we test seed formation scenarios by detecting MBHBs at high redshift *and* ascertain that they are indeed at high redshift? (See the right panel of Fig. 12 and Tables V and VI.)

Our study is somewhat similar in spirit (but different in many details, ranging from instrumental design to MBH modeling and waveforms) to a similar investigation that was carried out in 2012 in the U.S. by the NASA Physics of the Cosmos (PCOS) Gravitational-Wave Mission Concept Study, whose final report is available online [107]. The PCOS report considered several mission concepts, including "SGO High" (essentially the LISA concept modified to include all known cost savings, but with the same science performance); "SGO Mid," where the scalable parameters—the arm length, distance from Earth, telescope diameter, laser power and duration of science operations—were all reduced for near maximum cost savings; and "SGO Low," which eliminates one of the measurement arms, giving a similar performance to ESA's NGO concept.

Here we have adopted the NGO concept as a baseline (N2A1M2L4) and investigated how different "science metrics" vary as we tune different variables in the mission design. We broadly agree with some of the main conclusions of the PCOS study, namely that (i) scientifically compelling mission concepts exist that have worse sensitivity than the classic LISA design; (ii) scaling down the three-arm LISA architecture by shortening the measurement baseline and the mission lifetime (SGO Mid in the PCOS report terminology) preserves compelling science—provided the low-frequency target sensitivity can be achieved (N2 in our notation)—and does not increase risk; (iii) eliminating a measurement arm reduces the science return, as well as increasing mission risk.

In Fig. 13 we have also quantified the impact of each single baseline element on selected figures of merit. Our studies suggest that the cost-saving intervention that preserves the most science is shortening the arm length from 5 to 2 Gm (A5 → A2). In this case, the detector performance in each specific figure of merit is degraded by at most a factor of 2. On the other hand, either dropping the third arm (L6 → L4) or not meeting the target low-frequency sensitivity (N2 → N1) can seriously jeopardize the mission potential. High-redshift detections and potential electromagnetic follow-up targets might drop by a factor of $\sim 5-10$, correspondingly reducing the likelihood of coincident observations with traditional astronomical instruments. Further shortening the arm length to 1 Gm (A2 → A1) is also potentially damaging. Our results indicate that compromising on arm length might be the best way to save on mission costs while preserving most of the original LISA MBHB science.

Our study also suggests that in order to achieve the mission's science goals while cutting cost, a significant effort must be put into modeling the merger and ringdown of the MBHB waveforms: a complete knowledge of the MBHB IMR waveforms can compensate, at least partially, for cost reductions in detector design, but only for six-link configurations. Our study is incomplete in this respect, since we have included the merger/ringdown by simply rescaling the angular resolution and distance determination errors by appropriate powers of the SNR, so further investigations in this direction will be particularly valuable.

ACKNOWLEDGMENTS

E. Barausse acknowledges support from the European Union's Seventh Framework Programme (FP7/PEOPLE-2011-CIG) through Marie Curie Career Integration Grant No. GALFORMBHS PCIG11-GA-2012-321608, and from H2020-MSCA-RISE-2015 Grant No. StronGrHEP-690904. A. K. and E. Berti are supported by NSF CAREER Grant No. PHY-1055103. E. Berti is supported by FCT Contract No. IF/00797/2014/CP1214/CT0012 under the IF2014 Programme. F. O. is supported by European Research Council Consolidator Grant No. 647839, A. S.

is supported by a University Research Fellowship of the Royal Society. B. W. is supported by the Irish Research Council, which is funded under the National Development Plan for Ireland. Computations were performed on the GPC supercomputer at the SciNet HPC Consortium, on the Datura cluster at the AEI, on the XSEDE network (allocation TG-MCA02N014) and on the "Projet Horizon Cluster" at the Institut d'astrophysique de Paris. This work was supported in part by the DFG under Grant No. SFB/Transregio 7 "Gravitational-Wave Astronomy."

APPENDIX A: PRECESSING BLACK HOLE BINARY HYBRID WAVEFORMS

Our study makes use of a set of complete inspiral-merger-ringdown waveforms, each constituting a "hybrid" of an analytical PN prediction and the result of a fully relativistic NR simulation. In this appendix we shall briefly outline their construction.

Our waveforms are computed from numerical solutions of the full Einstein equations in which binary black hole initial data is evolved using a $3+1$ approach through inspiral, merger and ringdown. The initial data is of the Bowen-York type [108,109], and the initial parameters are chosen (i) for minimal eccentricity ($e \lesssim 10^{-4}$), using a variation of the method presented in [110], and (ii) for maximal precession, with the spin directions chosen to maximize the angle between the orbital and total angular momentum, using an iterative effective-one-body method. See Table VII for a summary of the physical parameters of the configurations studied. The initial data is evolved using the Baumgarte-Shapiro-Shibata-Nakamura (BSSN) formulation of the Einstein equations with eighth-order finite-differencing using components of the EINSTEIN TOOLKIT [111–118] in combination with the LLAMA multipatch code [119]. In comparison to purely Cartesian numerical grids, the use of constant angular resolution grids leads to high accuracy in the wave zone for not only the dominant $\ell = 2$, $m = \pm 2$ modes, but also the higher modes which are important for precessing systems.

The waveforms, consisting of 10–16 GW cycles, are constructed in terms of the standard complex Newman-Penrose scalar

$$\Psi_4(t) = \frac{\partial^2}{\partial t^2}[h_+(t) - ih_\times(t)] = |\Psi_4(t)|e^{i\phi(t)}, \qquad \text{(A1)}$$

where h_+ and $h_\times$ are the GW polarizations in the source frame, which is determined by our numerical simulations. The waveforms are measured on coordinate spheres at finite radius, and extrapolated to future null infinity using standard methods. The simulations were all performed at several numerical resolutions to measure the effect of numerical truncation error.

The stitching of a numerically obtained signal $\Psi_4^{\rm NR}$ to inspiral data of a particular PN approximant is a well studied procedure commonly applied to nonprecessing signals [120–128]. However, combining multiple harmonic modes of precessing signals presents additional challenges, and rapid progress in using precessing hybrids as well as comparing analytical and numerical waveforms has been reported recently [81–83,129–133].

Here, we follow an approach that is close in spirit to the treatment of nonprecessing binaries [128], with the additional complication that in the presence of precession we have to track more than just the evolution of the binary's orbital frequency. The black hole spins and the orbital plane constantly change direction, which leads to an extended set of PN equations that have to be integrated. We choose to employ the adiabatic TaylorT4 approximant (see Sec. IV A for details).

Apart from the number of equations to integrate, there is the further difficulty of finding appropriate PN initial data. The spins and the orbital plane constantly change their orientation, and it is not clear *a priori* which PN initial conditions evolve to the same setup as assumed by the respective NR simulation. One could of course approach the problem the other way around, namely start with some PN initial data, evolve the system up to a smaller separation and let the NR code "take over" by feeding in the appropriate quantities from the end of the PN evolution. This idea was explored already by Campanelli *et al.* [130], who found that although the results from PN and NR agree reasonably in the early inspiral, they quickly differ considerably with progressing simulation time. The cause of this disagreement is manifold. Apart from the fact that a truncated PN series will always deteriorate close to the

TABLE VII. Summary of the configurations studied. All quantities are measured at the point where the NR waveform begins and are given in units where they have been adimensionalized by M, the sum of the initial irreducible masses of the black holes. D/M is the separation, $m_i^{\rm h}$ is the irreducible mass and $[S_i^x, S_i^y, S_i^z]$ are the spin vectors.

Configuration	D/M	$m_1^{\rm h}$	$[S_1^x, S_1^y, S_1^z]$	$m_2^{\rm h}$	$[S_2^x, S_2^y, S_2^z]$
Q1	9	0.5	$[-0.02, -0.01, 0.15]$	0.5	$[-0.01, -0.01, 0.14]$
Q2	9	0.67	$[-0.02, -0.01, 0.27]$	0.33	$[-0.00, 0.01, 0.06]$
Q4	9	0.8	$[-0.02, 0.02, 0.38]$	0.2	$[0.00, 0.00, 0.02]$
Q2a	9	0.67	$[0, 0, 0.27]$	0.33	$[0, 0, 0.06]$
Q2HP	9	0.67	$[0.07, -0.18, 0.11]$	0.33	$[-0.01, 0.02, -0.05]$
Q4HP	9	0.8	$[-0.24, -0.30, -0.01]$	0.2	$[0.00, -0.00, -0.00]$

merger, the disagreement potentially stems from the different frameworks used in PN and NR to define physical quantities, and in particular from the transition from Bowen-York initial data [108,109] to the actually modeled system in the NR simulation.

We overcome this issue here by reading the "initial" values $\boldsymbol{S}_i(t_{\rm ini})$, $\boldsymbol{L}(t_{\rm ini})$ and $\omega_{\rm orb}(t_{\rm ini})$ off the NR simulation, at a time $t_{\rm ini}$ when the junk radiation has left the system and we observe a reasonably clean evolution of the numerical solution. Together with the time-independent mass ratio, these quantities complete the set of parameters we need to specify in order to integrate the PN equations both forward and backward in time. Note that both mass and spin measures in NR typically employ the formalism of quasi-local horizons [134], and determining the spin direction is a coordinate-dependent process [135,136]. When combining PN and NR descriptions for precessing binaries, however, we are forced anyway to relate different coordinate systems with each other. If the black holes are still far enough separated in the simulation, we can hope to sensibly identify the NR measurements with PN parameters, in due consideration of the appropriate spin supplementary condition in PN that ensures constant spin magnitudes [137,138]. The orbital frequency as well as the direction of the Newtonian orbital angular momentum are estimated simply through the coordinate motion of the punctures and the Euclidean vector product of the separation and the relative velocity. Again, this is a coordinate-dependent measure, but we merely extract from the NR simulation that the modeled system is (approximately) characterized at some instant by the specified values. We also tested the idea of optimizing the initial orbital frequency by a least-squares fit, and found slightly better agreement between the PN and NR evolutions (of all relevant quantities) when we set $\omega_{\rm orb}(t_{\rm ini}) = \hat{\omega}_{\rm orb}$, with $\hat{\omega}_{\rm orb}$ determined in turn by minimizing the PN-NR difference of, e.g., the spin of the heavier black hole over a few hundred M of evolution time.

Having calculated the PN evolution of $\omega_{\rm orb}(t)$, $\boldsymbol{L}(t)$ and $\boldsymbol{S}_i(t)$, we obtain the GW strain $h = h_+ - ih_\times$ by applying the explicit expressions provided in the appendix of [139]. The Newman-Penrose scalar $\Psi_4^{\rm PN}$ follows through two numerical time derivatives. However, even assuming that we have modeled the same system analytically and numerically, we cannot immediately combine the two waveform parts due to additional subtleties. First, there is another initial parameter, the initial phase, which does not enter the waveform simply as $e^{i\phi_0}$ (this is just the lowest order effect); there are higher-order amplitude corrections that depend on ϕ_0 [139]. Knowing them analytically, however, we can still fit for an optimal ϕ_0 between the PN and NR parts of the waveform. Second, although finding the initial parameters also relates the time between the PN and NR evolution, there is the problem that the physical quantities affect the PN waveform *immediately*, whereas if we consider a waveform extracted at some finite radius in NR, there is a time lag between the spin evolution and the observed GW signal. As already discussed in [130], this time lag approximately corresponds to the travel time between source and observer, but gauge effects will spoil this relation, and we shall determine both ϕ_0 and t_0 by an additional least-squares fit of the GW phase, just as in the nonprecessing case.

In short, we (1) make sure to simulate the same physical system numerically and analytically by reading the PN parameters off the NR simulation, and then (2) combine the NR and PN parts of the waveform by minimizing the phase difference in Ψ_4 over a certain length of evolution time. The choice of this interval would ideally be based on considerations concerning the hybridization accuracy (see e.g. [123,125]). Our goal, however, is not to produce highly accurate template waveforms to be eventually used in GW searches; we merely want to complete the PN description in a reasonably well motivated and robust way. Hence, we simply overlay the PN and NR waveforms in a region of approximately $250M$ length, as early as the NR simulation permits. Within this interval, the GW frequency $\omega(t) = d\phi(t)/dt$ evolves from $M\omega \approx 0.065$ to $M\omega \approx 0.08$, which is more than the minimal frequency evolution suggested in [125] to ensure an unambiguous matching.

Once we have matched the dominant spherical harmonic $\ell = 2$, $m = 2$ mode, all other modes are aligned as well. There is no additional freedom left to apply any time or phase shift to individual modes. We can only check to see that the agreement is similar to the dominant mode, and indeed, we find that the phase difference between PN and NR in the matching region is comparable (< 0.1 rad) for all spherical harmonic modes. The accuracy of the PN amplitude, however, degrades towards higher modes (higher spherical harmonic modes enter at different PN orders and are thus determined to lower relative expansion order [139,140]), which effectively limits our matching procedure to modes with $\ell \leq 4$ and $m = \pm\ell$.

The final hybrid waveform is now constructed mode by mode as a smooth connection of the PN and NR parts of the signal,

$$|\Psi_4^{\rm hyb}(t)| = |\Psi_4^{\rm PN}(t)|[1 - \mathcal{T}(t)] + |\Psi_4^{\rm NR}(t)|\mathcal{T}(t). \quad \text{(A2)}$$

An equivalent transition is also used separately for the phase $\phi(t)$, and $\mathcal{T}$ is a blending function. We employ a form of the Planck taper function,

$$\mathcal{T}(t) = \begin{cases} 0, & t \leq t_1 \\ \left[\exp\left(\frac{t_2-t_1}{t-t_1} + \frac{t_2-t_1}{t-t_2}\right) + 1\right]^{-1}, & t_1 < t < t_2 \\ 1, & t > t_2, \end{cases} \quad \text{(A3)}$$

as suggested by [141]. The parameters t_1 and t_2 used for constructing the phase are defined by the matching interval

that determined the optimal time and phase shift between the PN and NR parts of the waveform. However, we find that a slightly larger value of t_2 results in a smoother amplitude transition, which in turn avoids artifacts in the transformation from $\Psi_4^{\rm hyb}$ to $h^{\rm hyb}$. Finally, we obtain $h^{\rm hyb}$ by two time integrations in the Fourier domain, as suggested by [123,142]. A comparison with the original PN waveform $h^{\rm PN}$ ensures that the transformation is accurate, and we generally find

$$\left|\frac{h^{\rm PN}(t)-h^{\rm hyb}(t)}{h^{\rm hyb}(t)}\right| < 1\% \tag{A4}$$

for the dominant mode. This is not merely a statement about the amplitude accuracy, it also confirms that the phases agree very accurately over thousands of M in evolution time. We further confirmed the robustness of our hybrids by producing NR data at three different resolutions for each configuration. The hybrids we obtain are consistent among all NR resolutions, and we used the highest resolution for the results presented here.

APPENDIX B: ESTIMATE OF THE ERROR ON THE REMNANT SPIN

To estimate the errors on the remnant mass and spin we follow [30]. The postmerger waveform is dominated by quasinormal ringing. The dominant oscillation modes have a large quality factor, so one can use an approximation where each mode is replaced by a δ function at the appropriate oscillation frequency and compute the SNR as in Eq. (3.16) of [30], where we include redshift factors and substitute the Euclidean distance r by the luminosity distance D_l as appropriate:

$$\rho_{\rm FH} = \left(\frac{2}{5}\right)^{1/2}\left(\frac{1}{\pi\mathcal{F}_{lmn}D_l}\right)\left(\frac{\epsilon_{\rm rd}}{S_h^{\rm NSA}(f_{lmn})}\right)^{1/2} \times [M(1+z)]^{3/2}\frac{2Q_{lmn}}{\sqrt{1+4Q_{lmn}^2}}. \tag{B1}$$

Here M is the remnant mass in the source frame, $\mathcal{F}_{lmn} = M\omega_{lmn}$ is the dimensionless oscillation frequency, and Q_{lmn} is the quality factor of a quasinormal mode with angular indices (l,m) and overtone number n. Note that this expression involves the *non-sky-averaged* noise curve; see [30] for details. For the fundamental mode with $l=m=2$, the frequency and damping time of the oscillations are well fitted by (cf. Table VIII of [30])

$$\mathcal{F}_{lmn} = f_1 + f_2(1-\chi_r)^{f_3}, \tag{B2}$$

$$Q_{lmn} = q_1 + q_2(1-\chi_r)^{q_3}, \tag{B3}$$

where $f_1=1.5251$, $f_2=-1.1568$, $f_3=0.1292$, $q_1=0.7000$, $q_2=1.4187$, $q_3=-0.4990$ and χ_r is the dimensionless spin of the final black hole. Then the errors on mass and spin, in the Flanagan-Hughes convention [143], can be estimated as in Eqs. (4.12a) and (4.12b) of [30]:

$$\Delta\chi_r = \frac{1}{\rho_{\rm FH}}\left|2\frac{Q_{lmn}}{Q'_{lmn}}\left(1+\frac{1+4\beta}{16Q_{lmn}^2}\right)\right|, \tag{B4a}$$

$$\frac{\Delta M}{M} = \frac{1}{\rho_{\rm FH}}\left|2\frac{Q_{lmn}f'_{lmn}}{f_{lmn}Q'_{lmn}}\left(1+\frac{1+4\beta}{16Q_{lmn}^2}\right)\right|, \tag{B4b}$$

where a prime denotes a derivative with respect to χ_r. In general we would have

$$\beta = \sin^2\psi\cos 2\phi^{\times}_{lmn} - \cos^2\psi\cos 2\phi^{+}_{lmn} \tag{B5}$$

with $\cos\psi \equiv (1+N_\times^2)^{-1/2}$, $\sin\psi \equiv N_\times(1+N_\times^2)^{-1/2}$. The parameter $N_\times$ is the ratio between plus and cross polarization amplitudes. Following Flanagan and Hughes [143] we set $N_\times = 1$, $\phi^+ = \phi^\times = 0$ and therefore $\beta = 0$, which simplifies the expressions even further.

As an estimate of the ringdown efficiency $\epsilon_{\rm rd}$ we use the "matched-filtering based" estimate of Eq. (4.17) in [144]:

$$\epsilon_{\rm rd} = 0.44\frac{q^2}{(1+q)^4}, \tag{B6}$$

where $q = m_1/m_2 > 1$ is the mass ratio of the binary (see also [145–147] for similar scalings). This estimate is conservative, in the sense that it is appropriate for nonspinning binary mergers. Spin corrections should modify the efficiency by an amount which is roughly proportional to the sum of the components of the binary spins along the orbital angular momentum [148]. These spin-dependent corrections will change $\epsilon_{\rm rd}$ by at most a factor of ≈ 2, which is within the scope of our order-of-magnitude calculation. We plan to improve the accuracy of these estimates in future work.

[1] A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. **1**, 154 (1918).
[2] J. R. Gair, M. Vallisneri, S. L. Larson, and J. G. Baker, Living Rev. Relativity **16**, 7 (2013).
[3] N. Yunes and X. Siemens, Living Rev. Relativity **16**, 9 (2013).
[4] E. Berti *et al.*, Classical Quantum Gravity **32**, 243001 (2015).
[5] R. A. Hulse and J. H. Taylor, Astrophys. J. **195**, L51 (1975).
[6] T. Damour and J. H. Taylor, Phys. Rev. D **45**, 1840 (1992).
[7] T. Damour and J. H. Taylor, Astrophys. J. **366**, 501 (1991).
[8] M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo, Science **314**, 97 (2006).
[9] LIGO, http://www.advancedligo.mit.edu.
[10] VIRGO, http://www.virgo.infn.it.
[11] KAGRA, http://gwcenter.icrr.u-tokyo.ac.jp/en/.
[12] IndIGO, http://www.gw-indigo.org.
[13] R. S. Foster and D. C. Backer, Astrophys. J. **361**, 300 (1990).
[14] R. W. Hellings and G. S. Downs, Astrophys. J. **265**, L39 (1983).
[15] M. Kramer and D. J. Champion, Classical Quantum Gravity **30**, 224009 (2013).
[16] G. Hobbs, Classical Quantum Gravity **30**, 224007 (2013).
[17] M. A. McLaughlin, Classical Quantum Gravity **30**, 224008 (2013).
[18] G. Hobbs *et al.*, Classical Quantum Gravity **27**, 084013 (2010).
[19] R. N. Manchester (for the IPTA Collaboration), Classical Quantum Gravity **30**, 224010 (2013).
[20] L. Lentati *et al.*, Mon. Not. R. Astron. Soc. **453**, 2577 (2015).
[21] Z. Arzoumanian *et al.*, arXiv:1508.03024.
[22] R. M. Shannon *et al.*, Science **349**, 1522 (2015).
[23] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. Holz, T. Bulik, and F. Pannarale, Astrophys. J. **806**, 263 (2015).
[24] K. Belczynski, A. Buonanno, M. Cantiello, C. L. Fryer, D. E. Holz, I. Mandel, M. C. Miller, and M. Walczak, Astrophys. J. **789**, 120 (2014).
[25] K. Belczynski, S. Repetto, D. Holz, R. O'Shaughnessy, T. Bulik, E. Berti, C. Fryer, and M. Dominik, arXiv:1510.04615.
[26] Einstein Telescope, http://www.et-gw.eu.
[27] J. Harms, B. J. J. Slagmolen, R. X. Adhikari, M. C. Miller, M. Evans, Y. Chen, H. Müller, and M. Ando, Phys. Rev. D **88**, 122003 (2013).
[28] J. Kormendy and D. Richstone, Annu. Rev. Astron. Astrophys. **33**, 581 (1995).
[29] M. C. Begelman, R. D. Blandford, and M. J. Rees, Nature (London) **287**, 307 (1980).
[30] E. Berti, V. Cardoso, and C. M. Will, Phys. Rev. D **73**, 064030 (2006).
[31] E. Berti, V. Cardoso, and A. O. Starinets, Classical Quantum Gravity **26**, 163001 (2009).
[32] J. E. Plowman, D. C. Jacobs, R. W. Hellings, S. L. Larson, and S. Tsuruta, Mon. Not. R. Astron. Soc. **401**, 2706 (2010).
[33] J. E. Plowman, R. W. Hellings, and S. Tsuruta, Mon. Not. R. Astron. Soc. **415**, 333 (2011).
[34] J. R. Gair, A. Sesana, E. Berti, and M. Volonteri, Classical Quantum Gravity **28**, 094018 (2011).
[35] A. Sesana, J. Gair, E. Berti, and M. Volonteri, Phys. Rev. D **83**, 044036 (2011).
[36] E. Barausse, J. Bellovary, E. Berti, K. Holley-Bockelmann, B. Farris, B. Sathyaprakash, and A. Sesana, J. Phys. Conf. Ser. **610**, 012001 (2015).
[37] B. F. Schutz, Nature (London) **323**, 310 (1986).
[38] D. E. Holz and S. A. Hughes, Astrophys. J. **629**, 15 (2005).
[39] P. Amaro-Seoane, J. R. Gair, M. Freitag, M. Coleman Miller, I. Mandel, C. J. Cutler, and S. Babak, Classical Quantum Gravity **24**, R113 (2007).
[40] F. Verbunt and G. Nelemans, Classical Quantum Gravity **18**, 4005 (2001).
[41] C. Caprini, R. Durrer, T. Konstandin, and G. Servant, Phys. Rev. D **79**, 083519 (2009).
[42] ESA Cosmic Vision Program, http://sci.esa.int/cosmic-vision.
[43] P. Amaro-Seoane *et al.* (eLISA Collaboration), arXiv:1305.5720.
[44] P. L. Bender, D. Hils, and R. T. Stebbins, in *Proceedings of the 18th Texas Symposium on Relativistic Astrophysics and Cosmology, Chicago, 1996*, edited by A. V. Olinto, D. N. Schramm, and J. A. Frieman (World Scientific, Singapore, 1998).
[45] http://www.cosmos.esa.int/web/goat.
[46] C. Cutler, Phys. Rev. D **57**, 7089 (1998).
[47] A. Blaut, S. Babak, and A. Krolak, Phys. Rev. D **81**, 063008 (2010).
[48] T. B. Littenberg, Phys. Rev. D **84**, 063009 (2011).
[49] J. Crowder and N. J. Cornish, Classical Quantum Gravity **24**, S575 (2007).
[50] G. Nelemans, L. R. Yungelson, and S. F. Portegies Zwart, Mon. Not. R. Astron. Soc. **349**, 181 (2004).
[51] S. E. Timpano, L. J. Rubbo, and N. J. Cornish, Phys. Rev. D **73**, 122001 (2006).
[52] P. Amaro-Seoane *et al.*, GW Notes **6**, 4 (2013).
[53] L. Ferrarese and D. Merritt, Astrophys. J. **539**, L9 (2000).
[54] A. J. Benson, Phys. Rep. **495**, 33 (2010).
[55] J. Silk and G. A. Mamon, Res. Astron. Astrophys. **12**, 917 (2012).
[56] E. Barausse, Mon. Not. R. Astron. Soc. **423**, 2533 (2012).
[57] A. Sesana, E. Barausse, M. Dotti, and E. Rossi, Astrophys. J. **794**, 104 (2014).
[58] F. Antonini, E. Barausse, and J. Silk, Astrophys. J. **806**, L8 (2015).
[59] F. Antonini, E. Barausse, and J. Silk, Astrophys. J. **812**, 72 (2015).
[60] S. Cole, J. Helly, C. S. Frenk, and H. Parkinson, Mon. Not. R. Astron. Soc. **383**, 546 (2008).
[61] P. Madau and M. J. Rees, Astrophys. J. **551**, L27 (2001).
[62] A. Heger and S. E. Woosley, Astrophys. J. **567**, 532 (2002).

[63] T. H. Greif, V. Springel, S. D. M. White, S. C. O. Glover, P. C. Clark, R. J. Smith, R. S. Klessen, and V. Bromm, Astrophys. J. **737**, 75 (2011).
[64] M. Volonteri, F. Haardt, and P. Madau, Astrophys. J. **582**, 559 (2003).
[65] P. Madau, F. Haardt, and M. Dotti, Astrophys. J. Lett. **784**, L38 (2014).
[66] S. M. Koushiappas, J. S. Bullock, and A. Dekel, Mon. Not. R. Astron. Soc. **354**, 292 (2004).
[67] M. C. Begelman, M. Volonteri, and M. J. Rees, Mon. Not. R. Astron. Soc. **370**, 289 (2006).
[68] G. Lodato and P. Natarajan, Mon. Not. R. Astron. Soc. **371**, 1813 (2006).
[69] M. Volonteri, G. Lodato, and P. Natarajan, Mon. Not. R. Astron. Soc. **383**, 1079 (2008).
[70] M. C. Begelman, R. D. Blandford, and M. J. Rees, Nature (London) **287**, 307 (1980).
[71] Q. Yu, Mon. Not. R. Astron. Soc. **331**, 935 (2002).
[72] F. M. Khan, A. Just, and D. Merritt, Astrophys. J. **732**, 89 (2011).
[73] E. Vasiliev, Classical Quantum Gravity **31**, 244002 (2014).
[74] E. Vasiliev, F. Antonini, and D. Merritt, Astrophys. J. **785**, 163 (2014).
[75] E. Vasiliev, F. Antonini, and D. Merritt, Astrophys. J. **810**, 49 (2015).
[76] K. Holley-Bockelmann and F. M. Khan, Astrophys. J. **810**, 139 (2015).
[77] Z. Haiman, B. Kocsis, and K. Menou, Astrophys. J. **700**, 1952 (2009).
[78] M. Colpi, Space Sci. Rev. **183**, 189 (2014).
[79] G. Lodato, S. Nayakshin, A. R. King, and J. E. Pringle, Mon. Not. R. Astron. Soc. **398**, 1392 (2009).
[80] L. Hoffman and A. Loeb, Mon. Not. R. Astron. Soc. **377**, 957 (2007).
[81] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2014).
[82] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).
[83] S. Ossokine, M. Boyle, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **92**, 104028 (2015).
[84] A. Klein, N. Cornish, and N. Yunes, Phys. Rev. D **90**, 124029 (2014).
[85] R. N. Lang and S. A. Hughes, Phys. Rev. D **74**, 122001 (2006); **75**, 089902(E) (2007); **77**, 109901(E) (2008).
[86] R. N. Lang and S. A. Hughes, Astrophys. J. **677**, 1184 (2008).
[87] R. N. Lang, S. A. Hughes, and N. J. Cornish, Phys. Rev. D **84**, 022002 (2011).
[88] E. Berti, A. Buonanno, and C. M. Will, Phys. Rev. D **71**, 084025 (2005).
[89] E. Berti, A. Buonanno, and C. M. Will, Classical Quantum Gravity **22**, S943 (2005).
[90] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishan, Phys. Rev. D **82**, 064016 (2010).
[91] K. G. Arun, L. Blanchet, B. R. Iyer, and M. S. S. Qusailah, Classical Quantum Gravity **21**, 3771 (2004).
[92] L. Blanchet, G. Faye, B. R. Iyer, and S. Sinha, Classical Quantum Gravity **25**, 165003 (2008).
[93] L. Blanchet, Living Rev. Relativity **17**, 2 (2014).
[94] T. A. Apostolatos, C. Cutler, G. S. Sussman, and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994).
[95] A. Bohé, S. Marsat, G. Faye, and L. Blanchet, Classical Quantum Gravity **30**, 075017 (2013).
[96] E. Racine, Phys. Rev. D **78**, 044021 (2008).
[97] E. K. Porter, Phys. Rev. D **92**, 064001 (2015).
[98] S. Babak, M. Hannam, S. Husa, and B. F. Schutz, arXiv:0806.1591.
[99] S. T. McWilliams, R. N. Lang, J. G. Baker, and J. I. Thorpe, Phys. Rev. D **84**, 064003 (2011).
[100] T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly, Phys. Rev. D **87**, 104003 (2013).
[101] D. J. Mortlock, S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, R. G. McMahon, C. Simpson, T. Theuns, E. A. Gonzáles-Solares, A. Adamson, S. Dye, N. C. Hambly, P. Hirst, M. J. Irwin, E. Kuiper, A. Lawrence, and H. J. A. Röttgering, Nature (London) **474**, 616 (2011).
[102] E. Berti, J. Cardoso, V. Cardoso, and M. Cavaglia, Phys. Rev. D **76**, 104044 (2007).
[103] E. Berti and M. Volonteri, Astrophys. J. **684**, 822 (2008).
[104] J. M. Bardeen and J. A. Petterson, Astrophys. J. Lett. **195**, L65 (1975).
[105] T. Bogdanović, C. S. Reynolds, and M. C. Miller, Astrophys. J. Lett. **661**, L147 (2007).
[106] D. Gerosa, B. Veronesi, G. Lodato, and G. Rosotti, Mon. Not. R. Astron. Soc. **451**, 3941 (2015).
[107] http://pcos.gsfc.nasa.gov/physpag/GW_Study_Rev3_Aug2012-Final.pdf.
[108] J. M. Bowen and J. W. York, Jr., Phys. Rev. D **21**, 2047 (1980).
[109] S. Brandt and B. Bruegmann, Phys. Rev. Lett. **78**, 3606 (1997).
[110] M. Purrer, S. Husa, and M. Hannam, Phys. Rev. D **85**, 124051 (2012).
[111] T. Goodale, G. Allen, G. Lanfermann, J. Massó, T. Radke, E. Seidel, and J. Shalf, in *Vector and Parallel Processing—VECPAR 2002, 5th International Conference*, Lecture Notes in Computer Science (Springer, Berlin, 2003).
[112] CACTUS developers, CACTUS COMPUTATIONAL TOOLKIT, http://cactuscode.org/.
[113] E. Schnetter, S. H. Hawley, and I. Hawke, Classical Quantum Gravity **21**, 1465 (2004).
[114] E. Schnetter, P. Diener, E. N. Dorband, and M. Tiglio, Classical Quantum Gravity **23**, S553 (2006).
[115] CARPET, adaptive mesh refinement for the CACTUS framework, http://www.carpetcode.org/.
[116] M. Ansorg, B. Brügmann, and W. Tichy, Phys. Rev. D **70**, 064011 (2004).
[117] F. Löffler, J. Faber, E. Bentivegna, T. Bode, P. Diener, R. Haas, I. Hinder, B. C. Mundim, C. D. Ott, E. Schnetter, G. Allen, M. Campanelli, and P. Laguna, Classical Quantum Gravity **29**, 115001 (2012).
[118] EINSTEIN TOOLKIT, open software for relativistic astrophysics, http://einsteintoolkit.org/.
[119] D. Pollney, C. Reisswig, E. Schnetter, N. Dorband, and P. Diener, Phys. Rev. D **83**, 044045 (2011).

[120] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan *et al.*, Phys. Rev. D **77**, 104017 (2008).
[121] M. Boyle, A. Buonanno, L. E. Kidder, A. H. Mroue, Y. Pan, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **78**, 104020 (2008).
[122] M. Boyle, D. A. Brown, and L. Pekowsky, Classical Quantum Gravity **26**, 114006 (2009).
[123] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband *et al.*, Phys. Rev. D **82**, 064016 (2010).
[124] M. Hannam, S. Husa, F. Ohme, and P. Ajith, Phys. Rev. D **82**, 124052 (2010).
[125] I. MacDonald, S. Nissanke, and H. P. Pfeiffer, Classical Quantum Gravity **28**, 134002 (2011).
[126] M. Boyle, Phys. Rev. D **84**, 064013 (2011).
[127] F. Ohme, M. Hannam, and S. Husa, Phys. Rev. D **84**, 064029 (2011).
[128] P. Ajith, M. Boyle, D. A. Brown, B. Brugmann, L. T. Buchman *et al.*, Classical Quantum Gravity **29**, 124001 (2012).
[129] L. Gualtieri, E. Berti, V. Cardoso, and U. Sperhake, Phys. Rev. D **78**, 044024 (2008).
[130] M. Campanelli, C. O. Lousto, H. Nakano, and Y. Zlochower, Phys. Rev. D **79**, 084010 (2009).
[131] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Bruegmann *et al.*, Phys. Rev. Lett. **106**, 241101 (2011).
[132] P. Schmidt, M. Hannam, and S. Husa, Phys. Rev. D **86**, 104063 (2012).
[133] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, arXiv:1409.4431.
[134] A. Ashtekar and B. Krishnan, Living Rev. Relativity **7**, 10 (2004).
[135] M. Jasiulek, Classical Quantum Gravity **26**, 245008 (2009).
[136] M. Campanelli, C. O. Lousto, Y. Zlochower, B. Krishnan, and D. Merritt, Phys. Rev. D **75**, 064030 (2007).
[137] G. Faye, L. Blanchet, and A. Buonanno, Phys. Rev. D **74**, 104033 (2006).
[138] L. E. Kidder, Phys. Rev. D **52**, 821 (1995).
[139] K. Arun, A. Buonanno, G. Faye, and E. Ochsner, Phys. Rev. D **79**, 104023 (2009).
[140] L. Blanchet, G. Faye, B. R. Iyer, and S. Sinha, Classical Quantum Gravity **25**, 165003 (2008).
[141] D. McKechan, C. Robinson, and B. Sathyaprakash, Classical Quantum Gravity **27**, 084020 (2010).
[142] C. Reisswig and D. Pollney, Classical Quantum Gravity **28**, 195015 (2011).
[143] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4535 (1998).
[144] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam , S. Husa, and B. Brügmann, Phys. Rev. D **76**, 064034 (2007).
[145] A. Nagar, T. Damour, and A. Tartaglia, Classical Quantum Gravity **24**, S109 (2007).
[146] S. Bernuzzi and A. Nagar, Phys. Rev. D **81**, 084056 (2010).
[147] E. Barausse, V. Morozova, and L. Rezzolla, Astrophys. J. **758**, 63 (2012); **786**, 76(E) (2014).
[148] I. Kamaretsos, M. Hannam, and B. Sathyaprakash, Phys. Rev. Lett. **109**, 141102 (2012).