Probing fundamental physics with gravitational waves: The next generation

Scott E. Perkins,[1,*] Nicolás Yunes,[1,†] and Emanuele Berti[2,‡]

[1]*Illinois Center for Advanced Study of the Universe (iCASU), Department of Physics,
University of Illinois at Urbana-Champaign, Champaign, Illinois 61820 USA*
[2]*Department of Physics and Astronomy, Johns Hopkins University,
3400 N. Charles Street, Baltimore, Maryland 21218, USA*

(Received 27 October 2020; accepted 5 January 2021; published 12 February 2021; corrected 5 November 2021)

Gravitational wave observations of compact binary mergers are already providing stringent tests of general relativity and constraints on modified gravity. Ground-based interferometric detectors will soon reach design sensitivity, and they will be followed by third-generation upgrades, possibly operating in conjunction with space-based detectors. How will these improvements affect our ability to investigate fundamental physics with gravitational waves? The answer depends on the timeline for the sensitivity upgrades of the instruments, but also on astrophysical compact binary population uncertainties, which determine the number and signal-to-noise ratio of the observed sources. We consider several scenarios for the proposed timeline of detector upgrades and various astrophysical population models. Using a stacked Fisher matrix analysis of binary black hole merger observations, we thoroughly investigate future theory-agnostic bounds on modifications of general relativity as well as bounds on specific theories. For theory-agnostic bounds, we find that ground-based observations of stellar-mass black holes and LISA observations of massive black holes can each lead to improvements of 2–4 orders of magnitude with respect to present gravitational wave constraints, while multiband observations can yield improvements of 1–6 orders of magnitude. We also clarify how the relation between theory-agnostic and theory-specific bounds depends on the source properties.

DOI: 10.1103/PhysRevD.103.044024

I. INTRODUCTION

Einstein's general relativity (GR) has been wildly successful. The agreement with the observed perihelion precession of Mercury and the 1919 eclipse expedition to verify the prediction of relativistic light-bending around the Sun were the beginning of a century of thorough vetting [1]. The theory has passed every experimental test so far, and it was recently validated in the strong-field regime, most notably through the imaging of a black hole (BH) shadow in the electromagnetic spectrum by the Event Horizon Telescope [2] and through the observation of coalescing binary black holes (BBHs) by the LIGO/Virgo Collaboration [3,4].

One century of experimental triumphs did not deter theoretical work on observationally viable extensions of GR for mainly two sets of reasons [5]. The first is observational: some of the most outstanding open questions in physics might be explained by modifying the gravitational sector. For example, one could introduce an additional scalar field to the gravitational action [6,7] or allow the graviton to be massive [8–10] to explain the late-time acceleration of the Universe [11,12] without invoking the cosmological constant or dark energy. The second set of reasons is theoretical: string theory and other ultraviolet completions of the Standard Model usually add higher-order curvature corrections to the Einstein-Hilbert action, implying deviations from GR at high energies and large curvatures [13–15]. Therefore it is important to systematically test the assumptions underlying GR, which are often summarized in terms of Lovelock's theorem [5,16]. More specifically, GR assumes that the gravitational interaction is mediated by the metric tensor alone; the metric tensor is massless; spacetime is four-dimensional; the theory of gravity is position-invariant and Lorentz-invariant; and the gravitational action is parity-invariant. There is no *a priori* reason why these assumptions should be true, and therefore it is reasonable to explore alternatives to GR by systematically questioning each of them [5,17]. Our study is motivated by a combination of these two reasons: we will focus on theories that may address long-standing problems in physics, while questioning the validity of the main assumptions behind GR.

The LIGO-Virgo-KAGRA network of Earth-based detectors just completed their third observing run (O3). A fourth observing run (O4) is planned in 2022, and future observations will combine data from LIGO Hanford [18],

[*] scottep3@illinois.edu
[†] nyunes@illinois.edu
[‡] berti@jhu.edu

LIGO Livingston [18], Virgo [19], KAGRA [20], LIGO India [21], and third-generation (3g) detectors such as Cosmic Explorer (CE) [22] and the Einstein Telescope [23]. The space-based observatory LISA [24], scheduled for launch in 2034, will extend these observations to the low-frequency window. As existing ground-based detectors are improved, new ones are built and space-based detectors are deployed, our ability to test GR will be greatly enhanced, but to what level?

The main goal of this study is to combine the anticipated timeline of technological development for Earth- and space-based gravitational-wave (GW) detectors with astrophysical models of binary merger populations to determine what theories will be potentially ruled out (or validated) over the next three decades. We estimated parameters by running $\sim 10^8$ Fisher matrix calculations using waveform models including the effects of precession [25–27]. Our null hypothesis is that GR correctly describes our Universe, and that all modifications must reduce to GR in some limit for the coupling constants of the modified theory [17]. Under this assumption, we employ the parametrized post-Einsteinian (ppE) framework [28–31] to place upper limits on the magnitudes of any modification, assuming future GW observations to be consistent with GR. As our GW observatories are most sensitive to changes in the GW phase, we ignore modifications to the GW amplitude, an approximation that has been shown to be very good [32].

A. Executive summary

For the reader's convenience, here we provide an executive summary of the main results of this lengthy study.

(i) We use public catalogs of BBH populations observable by LISA and by different combinations of terrestrial networks over the next thirty years, and extract merger rates and detection-weighted source parameter distributions.

While this was not the main goal of this work, we did require astrophysical population models to realistically model GW science over the next three decades. In the pursuit of constructing forecasts of constraints on GR, we developed useful statistics concerning the distribution of intrinsic parameters for detectable merging BBHs for a variety of population models and detectors.

Useful quantities calculated here and related to BBH mergers are the expected detection rates for a large selection of population models and detector networks. These rates are listed in Table VI, and discussed in Secs. III A and III B. Detection rates depend not only on the population model, but also on the detector network. For LISA, we follow the method outlined in Ref. [33] to compute detection rates for multiband and massive black hole (MBH) sources.

We constructed synthetic catalogs by filtering the datasets coming from the full population models based on their signal-to-noise ratio (SNR). This yields a detection-weighted distribution of source parameters (discussed in Sec. IV C) which is useful to understand detection bias and to understand the typical sources accessible by different networks over the next three decades. In Figs. 4 and 5 we show these distributions for a large selection of detection network/population model combinations, considering both stellar-origin black holes (SOBHs) and MBHs.

The main conclusions of this analysis are summarized in Fig. 6, which shows the typical detection rates and SNR distributions for different source models and networks. This plot contains key information on the relative constraining performance of different population model/detector network combinations, which will be important for the following discussion of tests of GR.

(ii) We find that improvements over existing GW constraints on theory-agnostic modifications to GR range from 2 to 4 orders of magnitude for ground-based observations, from 2 to 4 orders of magnitude for LISA observations of MBHs, and from 1 to 6 orders of magnitude for multiband observations, depending on what terrestrial network upgrades will be possible, on LISA's mission lifetime, and on the astrophysical distribution of merging BBHs in the Universe.

The main issue addressed in this work is the scientific return on investment of future detector upgrades in terms of future explorations of strong gravity theories beyond GR. What future detectors and network upgrades are most efficient at constraining beyond-GR physics? Our models use astrophysical populations of SOBHs and MBHs and three reasonable development scenarios for ground-based detectors (ranging from optimistic to pessimistic) to try and answer this question. We first consider generic (theory-agnostic) modifications of GR and then focus on specific classes of theories that test key assumptions underlying Einstein's theory.

Our primary conclusions for generic modifications to GR are summarized in Fig. 7 and in Sec. VI A, where we show bounds on generic deviations from GR at a variety of post-Newtonian (PN) orders, separated by the class of source and marginalized over the detector configurations and population models. A term in the GW phase that is proportional to $(\pi \mathcal{M} f)^{b/3}$, where $\mathcal{M}$ is the chirp mass of the binary and f is the GW frequency, is said to be of $(b+5)/2$ PN order. While the range in constraints between the different models and scenarios is large, we have plotted constraints from current pulsar and GW tests of GR for comparison, where available and competitive. There are several trends present in this figure, most notably:

(1) SOBH multiband sources observed by both LISA and terrestrial networks are the most effective at setting bounds on negative PN effects, outperforming all other classes of sources by at least an order of magnitude. This observation must be tempered, however, because no multiband sources are observed at all in some of the scenarios we have analyzed. The detection rate of multiband sources is an open

question [33,34]. We hope that their importance for tests of GR, outlined here and elsewhere [35–41], will stimulate further work on this class of sources.

(2) The MBH mergers observed by LISA outperform SOBH sources observed only in the terrestrial band for negative PN orders in the more pessimistic ground-based detector scenarios. For most negative PN orders, LISA MBH observations perform at least comparably to the most optimistic terrestrial network scenario and greatly outperform the other two terrestrial scenarios analyzed in this work.

(3) Terrestrially observed SOBH sources are most effective at constraining positive PN effects, outperforming MBHs and multiband sources. Furthermore, for positive PN effects, the difference between the different terrestrial network scenarios closes dramatically. The constraining power between the different terrestrial networks shrinks, spanning a range of 4 orders of magnitude at negative PN orders but showing significant overlap for positive PN orders. This suggests that highly sensitive detectors are less important for constraining deviations that first enter at positive PN order, as opposed to negative PN order.

In terms of what detectors would have the highest return on investment, LISA's contribution to constraints on negative PN effects is quite high. Multiband sources are, by far, the most effective test beds for fundamental physics in the early inspiral of GW signals, but even in the absence of multiband sources (a realistic concern), MBH sources perform as well or better than even the most optimistic terrestrial network scenario we examined. The difference in terrestrial network scenarios is fairly drastic for negative PN effects, and so ground-based detector upgrades would play an important role if LISA were not available. The strongest improvement occurs in our most optimistic scenario (including CE and ET), but there is also a clear separation between the "pessimistic" and "realistic" scenarios.

Terrestrial networks perform the best for positive PN effects, but not by orders of magnitude. Even at positive PN orders, LISA MBH sources are still as effective as the more pessimistic terrestrial network scenarios. Furthermore, while constraining positive PN effects, no single terrestrial network scenario drastically outperforms the others: there is a clear hierarchy between the three scenarios, but with significant overlap.

These conclusions are also summarized in Table I, where we show a concise overview of current constraints on generic ppE parameters coming from observations of pulsars [42] and GWs [3], and we compare them against forecasts from our simulations.

(iii) LISA and future terrestrial network constraints on theory-agnostic modifications to GR follow trends which depend on the PN order, the underlying population of sources, and the detector network.

TABLE I. Summary of the constraints we predict on the theory-agnostic ppE modification parameter β as a function of the PN order parameter b, as defined in Eqs. (25) and (26) below. We compare these constraints against current constraints from pulsar tests [42] and GW observations from the LVC [3], denoted by (*). The LVC analysis used a slightly different formalism, so we mapped their results to the ppE framework for four specific sources (GW150914, GW170104, GW170608, and GW170814), we computed the standard deviation of the Markov Chain Monte Carlo (MCMC) samples and then combined the posteriors assuming a normal distribution to obtain a rough order-of-magnitude estimate of current ppE bounds from the LVC results. The columns list, from left to right: the PN order of each particular modification, the current constraint (if one exists), the best and worst constraints from our simulations, and the class of astrophysical sources those constraints come from. All the constraints are 1σ bounds, and we only show worst-case constraints that still improve on existing bounds. The source class acronyms are as follows: MB stands for multiband observations of SOBHs, T stands for terrestrial-only observations of SOBHs, and MBH stands for space-based detection of MBHs.

PN order	Current	Best (worst)	Best (worst)
(ppE b)	constraint	constraint	source class
$-4(-13)$	$\cdots$	10^{-25} (10^{-14})	MB (T)
$-3.5(-12)$	$\cdots$	10^{-23} (10^{-14})	MB (T)
$-3(-11)$	$\cdots$	10^{-21} (10^{-12})	MB (T)
$-2.5(-10)$	$\cdots$	10^{-19} (10^{-11})	MB (T)
$-2(-9)$	$\cdots$	10^{-17} (10^{-10})	MB (T)
$-1.5(-8)$	$\cdots$	10^{-15} (10^{-9})	MB (T)
$-1(-7)$	2×10^{-11}	10^{-13} (10^{-11})	MB (MBH)
$-0.5(-6)$	1.4×10^{-8}	10^{-11} (10^{-8})	MB (T)
0 $(-5$	1.0×10^{-5}	10^{-7} (10^{-5})	MBH (T)
$.5$ (-4)	$4.4 \times 10^{-3*}$	10^{-7} (10^{-5})	MB (T)
1 (-3)	$2.5 \times 10^{-2*}$	10^{-6} (10^{-4})	MB/T (T)
1.5 (-2)	0.15^*	10^{-5} (10^{-3})	T (MB)
2 (-1)	0.041^*	10^{-4} (10^{-2})	T (MB)

Using suitable approximations, we derive analytical expressions that help to elucidate the reason for the hierarchy of constraining power observed in our simulations. We first examine single observations and show how different source properties influence the constraints. We then attempt to quantify the importance of stacking multiple observations to develop a cumulative constraint from an entire catalog of observations.

In Sec. VI A 1 [Eqs. (32) and (33)] we show that, to leading order, the relative constraining power of one class of sources over another depends on the binary masses and on the initial frequency of observation, raised to a power which depends on the PN order in question. As this power changes sign going from negative to positive PN orders, this scaling explains why multiband and MBH sources are more competitive at negative PN orders, while terrestrial networks are more effective at positive PN orders. This trend is succinctly summarized in Fig. 8.

Besides single-source trends, in Sec. VI A 2 we quantify the effect of stacking observations and the benefit of large catalogs. In Fig. 9 we show that, as the PN order of the modification goes from negative to positive, the number of single observations meaningfully contributing to the cumulative bound from a catalog rises exponentially. This helps to further explain the improvement of terrestrial-only catalogs over LISA catalogs for higher PN orders: the very large catalogs coming from third-generation detectors are effectively leveraged to produce much stronger bounds, but only for positive PN orders. As shown in Fig. 10, this depends on the relation between the three parameters of primary concern (the SNR, the chirp mass, and the constraint), and on how their relation evolves as a function of the PN order.

These considerations help us understand the behavior observed in our simulations. The single-source scaling implies that MBHs and multiband sources should be more efficient at negative PN orders, because of the typical masses and initial frequencies of the observations. At positive PN orders the balance shifts in favor of terrestrial-only catalogs, further enhanced by the fact that large catalogs bear much more weight for positive PN effects.

The considerations made above also explain the significant overlap of different terrestrial detection scenarios at positive PN orders, and their separation at negative PN orders: negative PN effects are well constrained by single, loud events (favoring the most optimistic detector scenarios), while positive PN effects benefit from large catalogs. As detection rates are comparable for all three terrestrial scenarios, they perform comparably for positive PN effects.

(iv) We quantify the expected improvement over current constraints on theory-specific coupling parameters. We derive trends for theory-specific scalings and find that some conclusions following from generic modifications must be *reversed*.

The analysis of generic deviations from GR is a good theory-agnostic diagnostic tool for estimating the efficacy of future efforts to constrain fundamental physics. This is useful to perform null tests of GR, but at the end of the day, tests of GR focused on specific contending candidates provide the most meaningful physical insights [43]. Many of the trends observed for generic modifications remain valid when considering specific theories, but the scaling relations we observe in our simulations can change significantly for some of our target theories.

A bird's eye summary of our conclusions can be found in Table II. There we identify the current bound on theory-specific parameters, our predicted bounds after thirty years, and the class of sources which is most effective at improving the bounds. In this table we only include constraints obtained from actual data with a robust statistical analysis, in an effort to limit our comparisons to reliable experimental limits (as opposed to forecasts, simulations, etcetera). In-depth results by source class and trend derivations are presented in Sec. VI B. We refer the reader to that section for a detailed discussion of individual theories. In broad terms, the process of mapping generic constraints to theory-specific parameters can impose significant modifications to the trends observed in the analysis of generic constraints. These modifications can be significant enough to completely reverse the conclusions derived from generic deviations. This should temper any interpretation of our conclusions from general modifications. We also remark that our analysis for specific theories is far from comprehensive: there is, in principle, a very large number of GR modifications that have different mappings to ppE parameters, and therefore different trends in connection with source distributions.

Our conclusions on the best return of investment from GW detector development from the generic modification analysis *generally* hold also for specific theories. EdGB

TABLE II. Summary of forecasted constraints on specific modifications of GR. The source class acronyms are the same as in Table I. A (*) symbol denotes constraints coming from previous BBH observations, as opposed to other experimental evidence. When necessary, we have mapped all existing constraints to 1σ constraints by assuming the posterior to be normally distributed. We only show worst-case constraints that improve on existing GW bounds. For consistency with previous work, $\dot{M}$ is given in units of $M_\odot/\mathrm{yr}$, while we use geometrical units (so that $\delta\dot{E}$ is dimensionless) for the generic dipole radiation bound. Note that the necessary factor for transforming between the two is $c^3/G = 6.41\times10^{12}\,M_\odot/\mathrm{yr}$. The time derivative of the gravitational constant, $\dot{G}$, is normalized to the current value of G, and it does indeed have units of yr^{-1} in geometrical units (where $G = c = 1$).

			Most (least) stringent	Most (least)
Theory	Parameter	Current bound	forecasted bound	constraining class
Generic dipole	$\delta\dot{E}$	1.1×10^{-3} [44,45]*	10^{-11} (10^{-6})	MB (T)
Einstein-dilaton-Gauss-Bonnet	$\sqrt{\alpha_{\mathrm{EdGB}}}$	1 km [46] 3.4 km [47]*	10^{-3} km (1 km)	T (MBH)
Black hole evaporation	$\dot{M}$	⋯	10^{-8} $M_\odot$/yr (10^{2}) $M_\odot$/yr	MB (T)
Time varying G	$\dot{G}$	10^{-13}–10^{-12} yr^{-1} [48–52]	10^{-9} yr^{-1} (10 yr^{-1})	MB (T)
Massive graviton	m_g	10^{-29} eV [53–56] 10^{-23} eV [3,57]*	10^{-26} eV (10^{-24} eV)	MBH (MB)
dynamic Chern Simons	$\sqrt{\alpha_{\mathrm{dCS}}}$	5.2 km [58]	10^{-2} km (10 km)	T (MB)
Noncommutative gravity	$\sqrt{\Lambda}$	2.1 l_p [59]*	10^{-3} l_p (10^{-1}) l_p	T (MB)

gravity (Sec. VI B 5) and massive graviton theories (Sec. VI B 7) are two notable exceptions: in these cases, the dependence of the theory-agnostic parameters on source mass, spin and distance implies that the generic modifications predictions (at −1PN and 1PN orders, respectively) must be reversed.

The remainder of the paper presents the calculations summarized above in much more detail. The plan of the paper is as follows. In Sec. II we give details on the detector networks implemented in this work. This section includes information about the proposed timelines of detector development, as well as the specific sensitivity curves we have implemented at each stage. In Sec. III we discuss the statistics with which this network is used to filter astrophysical populations, including the calculation of detection probabilities for both terrestrial and space-based detectors. In Sec. IV we describe the population models then discuss the calculation of detection rates and the creation of our synthetic catalog. In Sec. V we outline the statistics of parameter estimation procedures and waveform models, including a brief overview of Fisher analysis and the modified-GR waveforms implemented in this study. In Sec. VI we present the results of our numerical investigation, as well as an analytical analysis to break down certain trends that have appeared in our findings. Finally, in Sec. VII we discuss limitations of this study and directions for future work. To improve readability, some technicalities about Bayesian inference and Fisher matrix calculations, the mapping of the ppE formalism to specific theories and our waveform models are relegated to Appendixes A, B and C, respectively. Throughout this paper we will use geometrical units ($G = c = 1$), and we assume a flat Universe with the cosmological parameters inferred by the Planck Collaboration [60].

II. DETECTOR NETWORKS

The construction and enhancement of GW detectors across the world and in space is expected to proceed steadily over the next thirty years. Tests of GR using GW observations are fundamentally tied to this global timeline of detector development, so it is important to have a realistic range of models for detector networks that spans the inevitable uncertainties intrinsic in planning experiments over such a long time. In this section we describe potential timelines for upgrades and deployment of new detectors, our assumptions on the location of the detectors, and their expected sensitivities.

A. Estimated timeline

Three plausible scenarios for the GW detector roadmap as of the writing of this paper are schematically presented in Fig. 1, with more details in Table III. The timeline starts with the fourth observing run (O4) of the LIGO-Virgo-KAGRA detectors, which are scheduled to take data at their design sensitivities for one year starting in 2022. After this run, the instruments would be taken off-line to be upgraded

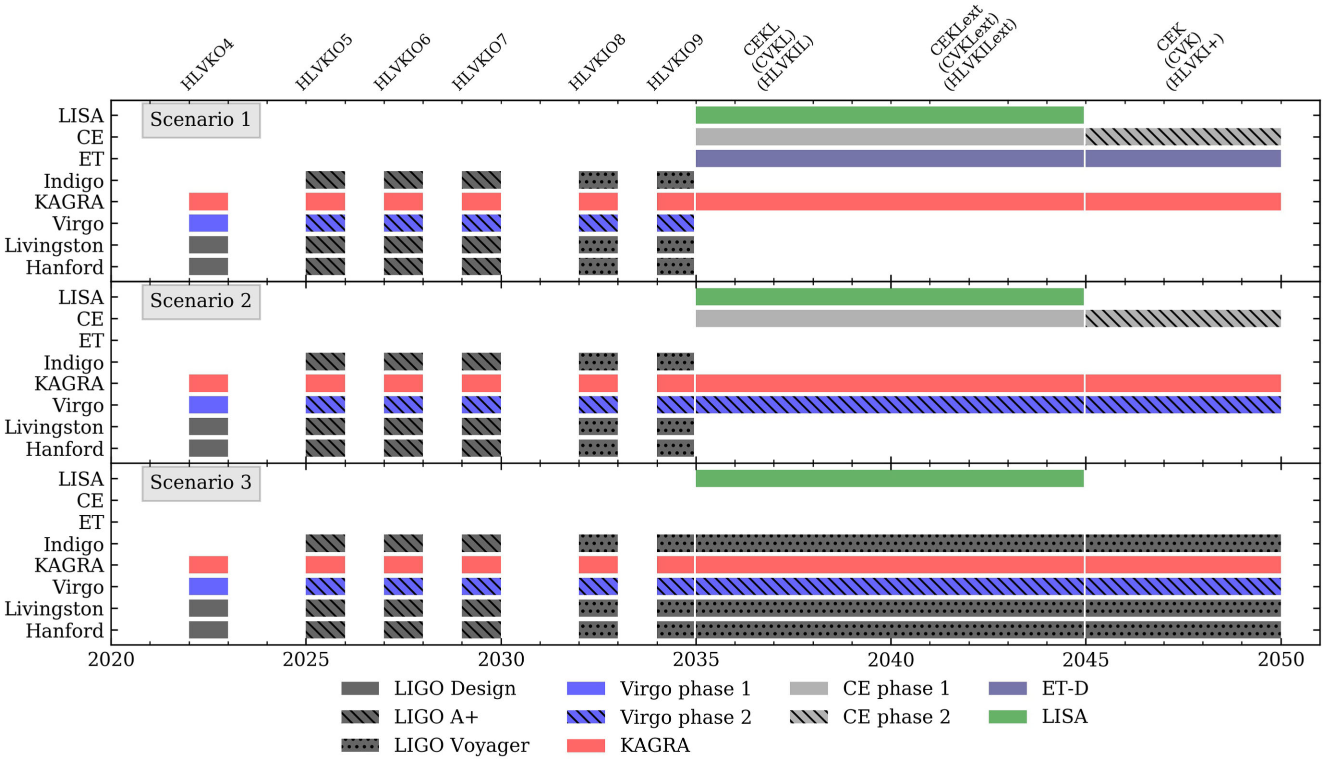

FIG. 1. Graphical representation of Table III. The shaded regions in the figure represent periods of active observation, and the colors/hatching corresponds to the noise curve being implemented, as shown in Fig. 2.

TABLE III. The above timeline tabulates the exact terrestrial detector evolution utilized by this study. There is a single timeline of detectors until 2035, when we model three separate scenarios that could play out in the next three decades: Scenario 1, 2, and 3. A graphical representation is shown in Fig. 1. The various sensitivity curves in column 3 are shown in Fig. 2.

Year	Detectors	Noise curves	Moniker(s)
2022–2023 [62]	LIGO Hanford	Advanced LIGO design [63]	HLVKO4
	LIGO Livingston	Advanced LIGO design	
	Virgo	Advanced Virgo+ phase 1 [63]	
	KAGRA	KAGRA 80 Mpc or 128 Mpc [63]	
2025–2030 [62] (one year observations in alternating years)	LIGO Hanford	Advanced LIGO A+ [63]	
	LIGO Livingston	Advanced LIGO A+	HLVKIO5
	Virgo	Advanced Virgo+ phase 2 high or low [63]	HLVKIO6
	KAGRA	KAGRA 80 Mpc or 128 Mpc	HLVKIO7
	LIGO India	Advanced LIGO A+	
2032–2035 (one year observations in alternating years)	LIGO Hanford	Advanced LIGO Voyager [64]	
	LIGO Livingston	Advanced LIGO Voyager	HLVKIO8
	Virgo	Advanced Virgo+ phase 2 high or low	HLVKIO9
	KAGRA	KAGRA 80 Mpc or 128 Mpc	
	LIGO India	Advanced LIGO Voyager	
	Scenario 1		
2035–2039 [61,65]	Cosmic Explorer	CE phase 1 [66]	CEKL
	Einstein Telescope	ET-D [67]	
	KAGRA	KAGRA 128 Mpc	
	LISA	LISA [68,69]	
2039–2045 [61,65]	Cosmic Explorer	CE phase 1	CEKLext
	Einstein Telescope	ET-D	
	KAGRA	KAGRA 128 Mpc	
	LISA	LISA	
2045–2050 [61,65]	Cosmic Explorer	CE phase 2 [66]	CEK
	Einstein Telescope	ET-D	
	KAGRA	KAGRA 128 Mpc	
	Scenario 2		
2035–2039	Cosmic Explorer	CE phase 1	CVKL
	Virgo	Advanced Virgo+ phase 2 high	
	KAGRA	KAGRA 128 Mpc	
	LISA	LISA	
2039–2045	Cosmic Explorer	CE phase 1	CVKLext
	Virgo	Advanced Virgo+ phase 2 high	
	KAGRA	KAGRA 128 Mpc	
	LISA	LISA	
2045–2050	Cosmic Explorer	CE phase 2	CVK
	Virgo	Advanced Virgo+ phase 2 high	
	KAGRA	KAGRA 128 Mpc	
	Scenario 3		
2035–2039	LIGO Hanford	Advanced LIGO Voyager	HLVKIL
	LIGO Livingston	Advanced LIGO Voyager	
	Virgo	Advanced Virgo+ phase 2 high or low	
	KAGRA	KAGRA 80 Mpc or 128 Mpc	
	LIGO India	Advanced LIGO Voyager	
	LISA	LISA	
2039–2045	LIGO Hanford	Advanced LIGO Voyager	HLVKILext
	LIGO Livingston	Advanced LIGO Voyager	
	Virgo	Advanced Virgo+ phase 2 high or low	
	KAGRA	KAGRA 80 Mpc or 128 Mpc	
	LIGO India	Advanced LIGO Voyager	
	LISA	LISA	

(Table continued)

TABLE III. *(Continued)*

Year	Detectors	Noise curves	Moniker(s)
2045–2050	LIGO Hanford LIGO Livingston Virgo KAGRA LIGO India	Advanced LIGO Voyager Advanced LIGO Voyager Advanced Virgo+ phase 2 high or low KAGRA 80 Mpc or 128 Mpc Advanced LIGO Voyager	HLVKI+

to higher sensitivity, with the next set of one-year-long observing runs starting in 2025. At this point, the network would also be joined by LIGO-India. Subsequent upgrades for the LIGO detectors to LIGO Voyager are planned for the early 2030's. The plans for 3g detectors are understandably more uncertain, with CE and ET potentially joining the network in 2035. After a 5–10 year observing run, CE is expected to be taken off-line for upgrades, with a second set of runs expected in 2045. Meanwhile, LISA is scheduled to fly in 2034, with a minimum mission lifetime of four years and a possible extension by six additional years, for a total of ten years of observation [61].

Given the timeline described above, one can identify several distinct periods of observations in which a different combination of detectors would be simultaneously on-line. During the O4 run, LIGO Hanford (H), LIGO Livingston (L), Virgo (V) and KAGRA (K) are expected to collect data simultaneously, creating the HLVKO4 network. LIGO India is expected to join the data collection effort in the late 2020's for the O5, O6 and O7 observation campaigns, creating the HLVKIO5/O6/O7 networks. In the early 2030's, the LIGO detectors (Hanford, Livingston, and Indigo) will be upgraded to the Voyager design, reflected in the HLVKIO8/09 networks.

The timeline beyond 2035 is quite uncertain, and we cannot model every possible scenario. Therefore, we chose to model three different timelines:

(1) After 2035, an optimistic detector schedule would see the Virgo and LIGO detectors replaced by the Einstein Telescope (E) and CE (C) detectors, respectively. Furthermore, LISA (L) is targeting around 2035 as the beginning of its data collection, with a nominal four-year mission and an additional six-year extension. These assumptions correspond to the CEKL and CEKLext networks, respectively. We follow up the multiband observation campaigns with a final terrestrial-only observation period from 2045–2050 for the CEK network. This timeline is shown as "scenario 1" in Table III.

(2) A less optimistic scenario might see one terrestrial 3g detector receive full funding and come online in the 2030's. We chose to use CE as our one 3g terrestrial detector to create the CVKL, CVKLext, and CVK networks. This is "scenario 2" in Table III.

(3) We also consider a pessimistic scenario where no terrestrial 3g detectors will be observing before the 2050's. The network will remain at its O9 sensitivity, but it will still be joined by LISA in the 2030's. This scenario includes the HLVKIL, HLVKILext, and HLVKI+ networks, and is denoted as "scenario 3" in Table III.

Because these last three observation periods for all three scenarios are less defined and span a wide time range, we assume an 80% duty cycle when estimating terrestrial-only

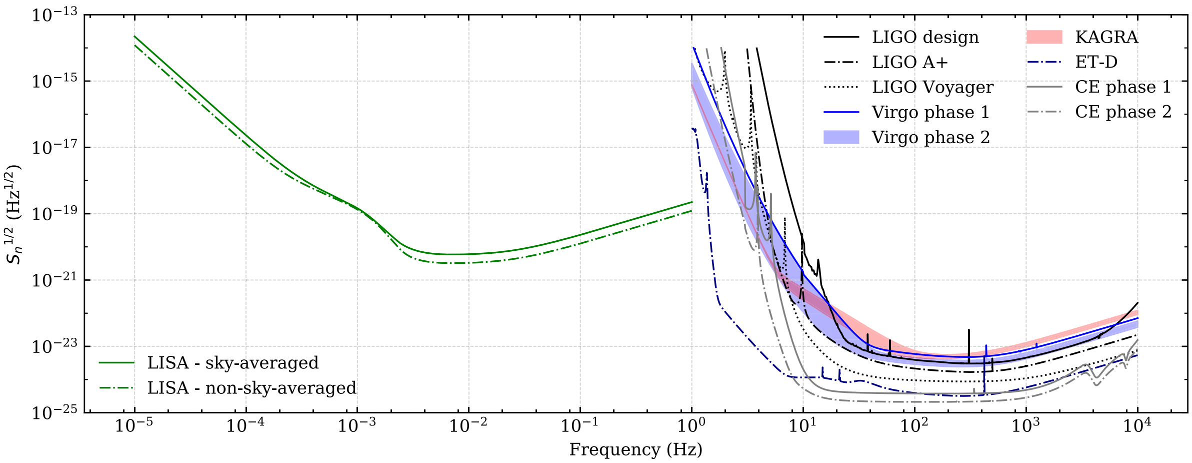

FIG. 2. Noise curves for the various detector configurations studied in this work. The shaded bands observed for the Virgo+ phase 2 and KAGRA sensitivities reflect uncertainties in estimates of their anticipated power spectral densities.

detection rates, but we use the full observation period for calculating multiband rates.

B. Estimated sensitivity

The detector sensitivities can be characterized in terms of their power spectral density S_n, which we present in Fig. 2.

We assume that the LIGO detectors will start operating at design sensitivity (“LIGO design” [63] in Fig. 2) in O4 but will be upgraded to the A+ configuration (“LIGO A+” [63] in Fig. 2) in time for the O5 observing run. In the early 2030’s, the LIGO detectors will be upgraded to the Voyager sensitivity (“LIGO Voyager” [64] in Fig. 2). Virgo observations begin with the Advanced Virgo+ phase 1 noise curve (“Virgo phase 1” [63] in Fig. 2) in O4, and they will subsequently be upgraded to Advanced Virgo+ phase 2 (“Virgo phase 2” [63] in Fig. 2) beginning in O5. To bracket uncertainties, we consider both an optimistic (“high”) configuration and a pessimistic (“low”) configuration for Virgo+ [63]. We model the KAGRA detector using the “128 Mpc” and “80 Mpc” configurations from Ref. [63] for optimistic and pessimistic outlooks, respectively (“KAGRA” in Fig. 2). LIGO India is planned to join the network in O5 with sensitivity well approximated by the A+ noise curve, mirroring the Hanford and Livingston detectors. LIGO India will follow the same development path as its American counterparts and be upgraded to Voyager sensitivity in the early 2030’s.

The US-led 3g detector, CE, may replace the LIGO detectors in 2035 at phase 1 sensitivity (“CE phase 1” in Fig. 2). After upgrades are completed in the early 2040’s, the detector may come back on-line with phase 2 noise sensitivity (“CE phase 2” in Fig. 2) [65].

The European-led 3g counterpart ET could replace the Virgo detector in 2035. ET will be modeled with the ET-D sensitivity in this study (“ET-D” in Fig. 2). In reality, ET is comprised of three individual detectors arranged in an equilateral triangle, and a fully consistent treatment of ET would incorporate the three detectors separately. However, after testing on subsets of our populations, we concluded that modeling ET as three identical copies of one of the constituent detectors minimally impacts our estimates on constraints of modified gravity, because of the small correlations between modified gravity modifications to the phase and the extrinsic parameters of the source, like sky location and orientation. This approximation significantly reduces the computational resources required to perform this study, so we opted to use it when constructing the Fisher matrices themselves (as discussed in Sec. V). When calculating the detection probability, however, we do account for the three detectors separately (cf. Sec. IV B). This is because the different orientations and positions of the detectors affect the rates more than they affect parameter estimation.

TABLE IV. Detector locations used in this paper.

Detector	Latitude (°)	Longitude (°)
LIGO Hanford	46.45	−119.407
LIGO Livingston	30.56	−90.77
Virgo	43.63	10.50
KAGRA	36.41	137.31
LIGO India	14.23	76.43
Cosmic Explorer	40.48	−114.52
Einstein Telescope	43.63	10.50

For networks that include a mixture of 3g and 2g detectors, we will only model the 2g detectors with the most optimistic sensitivity curve, i.e., the “high” configuration for Virgo and the “128 Mpc” configuration for KAGRA. The impact of the different 2g sensitivities is small when implemented alongside a 3g detector, and the shrinking of the parameter space for our models significantly reduces the computational cost of the problem.

For LISA, we model the noise curve using the approximations in Ref. [68]. At different points in this work, we required both sky-averaged and non-sky-averaged response functions to various detectors. For LISA this can be more complicated than terrestrial interferometers, so we plot the sky-averaged noise curve directly from Ref. [68] (“LISA—sky-averaged” in Fig. 2) and the full (non-sky-averaged) sensitivity produced in Ref. [69] (“LISA—non-sky-averaged” in Fig. 2). However, in contrast to Ref. [69], we do include the factor of 2 to account for the second channel, mirroring the approximation we made for ET.

C. Estimated location

The relative locations of the various detectors affects the global response function, and thus it impacts the analysis performed in this paper. For terrestrial detectors, the various geographical locations of each site are shown in Table IV. The sites of detectors currently built or under construction were taken from data contained in LALSuite [70]. Since a site has yet to be decided upon for CE, we chose a reasonable location near the Great Basin desert, in Nevada. For LISA, the detector’s position and orientation as a function of time must be taken into account, so we use the time-dependent response function derived in Refs. [71,72]. Unlike those papers we use the polarization angle defined by the total angular momentum **J**, instead of the orbital angular momentum **L**, because the latter precesses in time, while **J** remains (approximately) constant.

III. STATISTICAL METHODS FOR POPULATION SIMULATIONS

Both terrestrial and space-borne GW detectors have nonuniform sensitivity over the sky. This effect is important when attempting to estimate the expected detection rate and the resulting population catalog.

Terrestrial detector networks can mitigate this selection bias by incorporating more detectors into the network, which can "fill in" low-sensitivity regions in the sky. The incorporation of the most accurate combination of detectors and their locations can be important. This is why in Sec. II C we specified the locations used in this study.

For space-borne detectors, some signals may be detectable for much longer than the observation period, so random sky locations map to random spacetime locations, and the effect of only seeing a portion of the signal must be accounted for.

These issues with terrestrial networks and space detectors, and their associated detection probabilities, are discussed in Secs. III A and Sec. III B, respectively.

We wish to calculate the probability that the GWs emitted by some source will be detected by a terrestrial network of instruments, which we will refer to as the detection probability. We will focus primarily on two classes of sources: SOBH binaries [73] and MBH binaries [74]. We will use publicly available SOBH population synthesis models to produce synthetic catalogs which are mainly of interest for the terrestrial network, but can also be observed as "multiband" events by both the terrestrial network and LISA. We will also use MBH binary simulations to create synthetic catalogs for LISA (these sources are typically well outside the frequency band accessible to terrestrial networks). Intermediate-mass BH binaries could also be of interest [75], but we do not consider them here, mainly because their astrophysical formation models and rates have large uncertainties [36,76,77].

A. Terrestrial detection probability

An accurate calculation of the detection probability for each source requires injections into search pipelines. A simplifying, while still satisfactorily accurate, assumption used in most of the astrophysical literature (see e.g., [78–80]) involves computing the SNR ρ, defined by

$$\rho^2 = 4\mathrm{Re}\left[\int \frac{\tilde{h}\tilde{h}^*}{S_n(f)} df\right], \tag{1}$$

where we recall that $S_n(f)$ is the noise power spectral density of the detector, while $\tilde{h} = \tilde{h}(f)$ is the Fourier transform of the contraction between the GW strain and the detector response function. We can factor out all the detector-dependent quantities from the SNR in the form of the "projection parameter" ω defined as [78,80]

$$\omega^2 = \frac{(1+\cos^2\iota)^2}{4} F_+^2(\theta,\phi,\psi) + \cos^2\iota F_\times^2(\theta,\phi,\psi), \tag{2}$$

where ι is the inclination of the binary relative to the line of sight, θ and ϕ are the spherical angles of the source relative to the vector perpendicular to the plane of the detector, and ψ is the polarization angle. The single-detector antenna pattern functions F_+ and $F_\times$ are given by

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

$$F_\times = \frac{1}{2}(1+\cos^2\theta)\cos 2\phi \sin 2\psi + \cos\theta \sin 2\phi \cos 2\psi. \tag{3}$$

With the projection-parameter approximation, we can approximate the SNR as

$$\rho^2 \approx \omega^2 \rho_{\mathrm{opt}}^2, \tag{4}$$

where ρ_{opt} is the SNR for an optimally oriented binary with $\theta = 0$, $\iota = 0$, and $\psi = 0$. This relation is approximate if the binary is precessing, so that ι is a function of time, but it is exact otherwise.

The calculation of the detection probability can then be rephrased as a search for the extrinsic source parameters that satisfy $\omega \approx \rho/\rho_{\mathrm{opt}} \geq \rho_{\mathrm{thr}}/\rho_{\mathrm{opt}} \equiv \omega_{\mathrm{thr}}$ for some ρ_{thr}. The probability that ω satisfies the above criteria translates into finding the cumulative probability distribution [78],

$$p_{\mathrm{det,terr}}(\vec{\lambda}) = \int \Theta(\omega'(\theta,\phi,\psi,\iota) - \omega_{\mathrm{thr}}) \frac{\sin\theta d\theta d\phi}{4\pi}\frac{d\psi}{\pi}\frac{d\cos\iota}{2}, \tag{5}$$

where $\Theta(\cdot)$ is the Heaviside function, which ultimately describes the selection effects of our terrestrial networks. This cumulative probability clearly depends on the source parameter vector $\vec{\lambda}$, inherited from $\omega_{\mathrm{thr}} = \omega_{\mathrm{thr}}(\vec{\lambda})$.

Equation (5) can be extended to multiple-detector networks by expanding our definition of ω to

$$\omega_{\mathrm{network}}^2 = \sum_i \omega_i^2, \tag{6}$$

where ω_i is the projection parameter for a single detector in the network, and $\omega_{\mathrm{network}} = \rho_{\mathrm{network-thr}}/\rho_{\mathrm{opt}}$ with some threshold network SNR, $\rho_{\mathrm{network-thr}}$, and single-detector optimal SNR, ρ_{opt}. In the case of a multiple-detector network, the locally defined position coordinates θ and ϕ are replaced with the globally defined position coordinates α (the right ascension angle) and δ (the declination angle). The polarization angle ψ is changed to the globally defined polarization angle $\bar{\psi}$, which is defined with respect to an Earth-centered coordinate axis instead of the coordinate system tied to a single detector.

Evaluating Eq. (5) for each network, with the network projection operator defined as Eq. (6), provides a good estimation of the probability we are seeking: a weighting factor for a given binary that incorporates the sensitivity and global geometry of a given detector network, as well as the impact that the intrinsic properties of the source have on its detectability. Importantly, the intrinsic source parameters themselves only enter into Eq. (5) through the calculation of ρ_{opt} in ω_{thr}. Once a threshold SNR ρ_{thr} is set, the detection probability function can be seen as a function of only one number ω_{thr} (for a given network), through its dependence on ρ_{opt}. As Eq. (5) is a

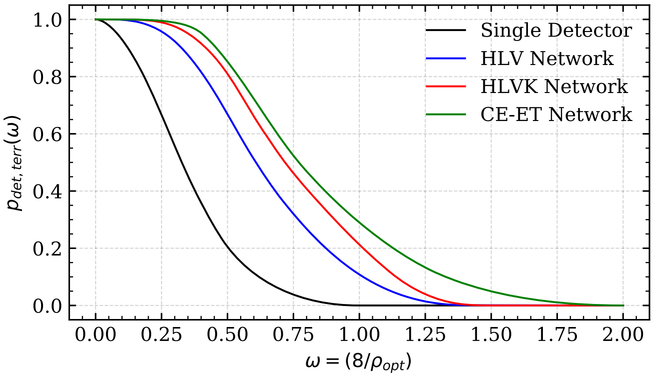

FIG. 3. Detection probability $p_{\rm det}$ for the four networks examined in this paper. The black curve is for a single detector (where global position no longer matters, so this is valid for any single right-angle Michelson interferometer). The blue curve is specifically for the Hanford, Livingston, and Virgo (HLV) network. The red curve is for the Hanford, Livingston, Virgo, and KAGRA (HLVK) network. Finally, the green curve represents a network comprised of CE and ET (which includes all three of the ET detectors as well as the 60° angle between each set of arms).

four-dimensional integral and must be calculated numerically, this detail can significantly save on computational cost if we can approximate the full function $p_{\rm det,terr}(\omega_{\rm thr})$ once for each network. To do this, we form a grid in $\omega_{\rm thr}$ with approximately 100 grid points and evaluate Eq. (5) for each grid point with 10^9 samples uniformly distributed in $\bar{\psi}$, $\cos\iota$, α, and $\sin\delta$. Interpolating across the grid in $\omega_{\rm thr}$ produces an approximation for $p_{\rm det,terr}(\omega_{\rm thr})$. This approximation must be calculated for each specific network, as the quantity ω' in Eq. (5) depends on the number and relative location of the detectors, but it only needs to be evaluated once per network, rather than once per source.

The resulting probability functions for the four terrestrial networks examined in this paper are shown in Fig. 3. Note that the relative location of each detector in a network impacts the form of $p_{\rm det,terr}$, so we label the curves by the detector nodes and not just their number (i.e., the form of $p_{\rm det,terr}$ will be slightly different for a Hanford, Livingston, and Virgo network when compared to a Hanford, Livingson, and KAGRA network). Furthermore, an important assumption in this calculation is that the sensitivity of each detector is identical. This is not a good approximation when jointly considering second- and third-generation detectors, so in these cases we neglect all the 2g detectors in the network. The configurations used at each stage are summarized in Table V.

B. Space detection probability

For space-based detectors, which operate at much lower frequencies, the picture changes quite drastically. The terrestrial detection probability of Sec. III A addresses the issue of random sky location and orientation of the sources, but an important effect for detectors like LISA is the time spent in band. Because signals observable by LISA can be detected for much longer than the observation time $T_{\rm obs}$ of the LISA mission, the time spent in the frequency range accessible to LISA will characterize the detectability of the binary. We characterize this effect as outlined below (we refer the reader to Ref. [33] for a more thorough derivation and further details).

To determine the time the binary spends in the observational frequency band of LISA, we look for the roots of

$$\rho(t_{\rm merger}) - \rho_{\rm thr} = 0, \tag{7}$$

where $t_{\rm merger}$ is the time before merger at which the signal starts, $\rho_{\rm thr}$ is some threshold SNR, and the SNR $\rho(t_{\rm merger})$ is defined as

TABLE V. Configurations used at each stage of our analysis to calculate the probability of detection for a given binary for the terrestrial detector network. Note that networks involving multiple detectors are labeled by the network nodes and not just their number, because the relative position of the detectors impacts the calculation of the detection probability. Our calculation depends on the assumption that all the detectors have approximately the same sensitivity curve, and so the curve used at each stage is given in the last column. Because of this assumption, and the extreme disparity in sensitivity between second- and third-generation detectors, we only use the CE detector to calculate rates when CE is part of the network.

Detection network	Detector locations	Detector sensitivity curve
HLVKO4	Hanford site Livingston site Virgo site	Ad. LIGO design [63]
HLVKIO5-O7	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO A+ [63]
HLVKIO8-O9	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager [64]
CEKL(ext)	Cosmic Explorer site All ET sites	CE phase 1 [66]
CVKL(ext)	Cosmic Explorer site	CE phase 1
HLVKIL(ext)	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager
CEK	Cosmic Explorer site All ET sites	CE phase 2 [66]
CVK	Cosmic Explorer site	CE phase 2
HLVKI+	Hanford site Livingston site Virgo site KAGRA site	Ad. LIGO Voyager

$$\rho(t_{\rm merger}) = 4{\rm Re}\left[\int_{f(t_{\rm merger})}^{\min(f(t_{\rm merger}-T_{\rm obs}),1\ {\rm Hz})} \frac{\tilde{h}\tilde{h}^*}{S_n(f)}df\right]. \tag{8}$$

Note that, at variance with Ref. [33], we use 1 Hz as the upper cutoff for the LISA noise curve.

Once the roots of Eq. (7) (say T_1 and T_2) have been found, we can obtain the probability of mergers for LISA via

$$p^{\rm SOBH}_{\rm det,space}(\vec{\lambda}) = p_{\rm det,terr}(\vec{\lambda}) \times \min\left[\frac{T_1 - T_2}{T_{\rm obs}}, \frac{T_{\rm wait} - T_2}{T_{\rm obs}}\right] \tag{9}$$

for SOBH binaries, and

$$p^{\rm MBH}_{\rm det,space}(\vec{\lambda}) = \min\left[\frac{T_1 - T_2}{T_{\rm obs}}, \frac{T_{\rm wait} - T_2}{T_{\rm obs}}\right] \tag{10}$$

for MBH binaries. The probability $p^{\rm SOBH}_{\rm det,space}$ is weighted by $p_{\rm det,terr}$ because all SOBH binaries we consider for LISA are also candidate multiband events, which must be observed both by LISA and by a terrestrial network to be considered "true" multiband sources. In these expressions, $T_{\rm wait}$ is some maximum waiting time for the binary to merge, which (following Ref. [33]) we choose to be $5 \times T_{\rm obs}$ for each detector network iteration.

C. Waveform model for population estimates

When computing the detection probability of a given source, we need a model for the Fourier transform of the time-domain response function $h = F_+ h_+ + F_\times h_\times$. In the terrestrial case, we implement the full precessing inspiral/merger/ringdown model IMRPhenomPv2 [25–27] with an inclination angle of $\iota = 0°$ to calculate the optimal SNR, $\omega_{\rm opt}$. For the space-based estimates in the next section, we will use the spinning (but nonprecessing) sky-averaged IMRPhenomD waveform model [26,27], with a small modification: since we are interested in LISA rather than terrestrial, right-angle interferometers, we replace the usual factor of 2/5 (that arises from sky-averaging) in favor of the sky-averaged LISA sensitivity curve from [68], which accounts for the second LISA data channel, sky-averaging, and the 60° angle between the detector arms. This waveform model depends on parameters $\vec{\lambda}_D = [\alpha, \delta, \theta_{\rm L}, \phi_{\rm L}, \phi_{\rm ref}, t_{c,\rm ref}, D_L, \mathcal{M}, \eta, \chi_1, \chi_2]$, where α is the right ascension, δ is the declination, $\theta_{\rm L}$ and $\phi_{\rm L}$ are the polar and azimuthal angles of the binary's orbital angular momentum $\mathbf{L}$ in equatorial coordinates at the reference frequency, $\phi_{\rm ref}$ and $t_{c,\rm ref}$ are the orbital phase and the time of coalescence at the reference frequency, D_L is the luminosity distance, $\mathcal{M}$ and η are the redshifted chirp mass and the symmetric mass ratio, and $\chi_i = \hat{\mathbf{L}} \cdot \mathbf{S}_i / m_i^2$ are the dimensionless spin components along $\hat{\mathbf{L}} = \mathbf{L}/|\mathbf{L}|$ with spin angular momentum $\mathbf{S}_i$.

For space-based detectors we must also choose a way to map between time and frequency. The limits of the SNR integral (1) and the antenna patterns (which for LISA are functions of time) depend on this mapping. For multiband SOBH binaries we use the leading-order PN relation [33,72,81],

$$f(t_{\rm merger}) = \frac{5^{3/8}}{8\pi}(\mathcal{M})^{-5/8} t_{\rm merger}^{-3/8}, \tag{11}$$

where again $t_{\rm merger}$ is the time before merger. For massive black hole (MBH) binaries, observed by LISA only through merger, this PN approximation is insufficient, so we use instead [82,83],

$$t_{\rm merger} = \frac{1}{2\pi}\frac{d\phi}{df}, \tag{12}$$

where ϕ is the GW Fourier phase. When calculating detection rates, we will invert these relations numerically as needed.

IV. POPULATION SIMULATIONS

A key ingredient of our work is the use of astrophysically motivated BBH population models (Sec. IV A). Our methodology for computing detection rates and for creating synthetic catalogs from the models is explained in Sec. IV B and in Sec. IV C, respectively.

A. Population models

For ease of comparison with previous work, we use the SPOPS catalogs [73] for SOBH binaries (Sec. IV A 1) and the MBH binary merger catalogs used in Ref. [74] (Sec. IV A 2).

1. Stellar mass simulations

We use the public SPOPS catalog of population synthesis simulations [73] in an effort to accurately capture the full spin orientations of the binaries at merger. The SPOPS catalog uses multiscale solutions of the precessional dynamics [84,85] computed through the public code PRECESSION [86] to quickly evolve the binary's spin orientations in time until the binary is about to merge.

The catalog is parametrized by three different variables: the strength of the BH natal kicks, the BH spin magnitudes at formation, and the efficiency of tidal alignment [73]. In this model, natal kicks are caused by asymmetric mass ejection during core collapse, imparting a torque on one of the constituents of the binary, while the tidal alignment reflects spin-orbital angular momentum coupling through tidal interactions that can realign the spin vectors with the orbital angular momentum vector (see Ref. [73] for further details).

Following Ref. [33], we choose to vary only one parameter of these models while keeping the others fixed. More specifically, we consider a uniform distribution in spin magnitude and the most realistic ("time") prescription for tidal alignment of Ref. [33], while varying the natal kick. To estimate lower and upper constraints on the rates given uncertainties in our population modeling, we use the two most extreme natal kick models, corresponding to $\sigma = 0$ km/s and $\sigma = 265$ km/s, where σ is the one-dimensional dispersion of the Maxwellian distribution the kicks are drawn from. The zero-kick scenario results in a lack of precessional effects and the highest detection rates for all detectors, while the $\sigma = 265$ km/s choice corresponds to a soft upper bound on the size of the kicks, which imparts the largest spin tilts and results in the lowest detection rate. The two chosen values of σ result in optimistic and pessimistic bounds on our projected constraints, and at the same time they provide a useful comparison between highly precessing systems and nonprecessing systems.

2. Massive black hole simulations

To model MBH binary populations, we adopt the semianalytical models of early Universe BH formation [87–89] used in the LISA parameter estimation survey of Ref. [74]. As in that work, we focus on three populations models, characterized by different BH seeding mechanisms and different assumptions on the time delay between BH mergers and the mergers of their host galaxies. These population models are denoted as

(1) PopIII—seeds are produced from the collapse of population III stars in the early Universe (a light-seed scenario);
(2) Q3delays—seeds are produced from the collapse of a protogalactic disk (heavy-seed scenario), and there are delays between galaxy mergers and BH mergers;
(3) Q3nodelays—seeds are produced from the collapse of a protogalactic disk (heavy-seed scenario), and there are no delays between galaxy mergers and BH mergers.

These three models embody two seed formation mechanisms, with two models representing optimistic and pessimistic heavy-seed scenarios. The difference between PopIII simulations with and without delays is less than a factor of 2, so, following Ref. [74], we consider only the more conservative estimate, in which delays are incorporated.

B. Detection rate calculations

With population synthesis simulations at our disposal, we can now estimate expected detection rates for a given detector network. This involves taking a model for our Universe that predicts a certain rate of merging BBHs per comoving volume and filtering the model through the lens of a particular detector configuration and sensitivity. The detection rate r for a given network follows from the following relation [33,90]:

$$r = \iint dz d\vec{\lambda} \mathcal{R}(z) p(\vec{\lambda}) \frac{dV_c(z)}{dz} \frac{1}{1+z} p_{\text{det}}(\vec{\lambda}, z), \qquad (13)$$

where z is the cosmological redshift, $\mathcal{R}$ is the intrinsic merger rate (a function of the redshift), p is the probability of a binary forming and merging given a set of intrinsic source parameters $\vec{\lambda} = \vec{\lambda}_D$ (discussed in Sec. III C), and dV_c/dz is a shell of comoving volume V_c at redshift z.

The quantity p_{det} is the probability of a binary being detected by a given detector network with some threshold SNR, as discussed in Sec. III. The type of detector network affects the quantity p_{det} only, while the other terms in the integral above depend only on information contained in the population simulation. For this study, we have used a threshold SNR of 8 for terrestrial and space detections, while for multiband detections we require the terrestrial SNR and the LISA SNR to both be above 8 independently. Because of the intrinsic difference in the duration of signals observed by space detectors and terrestrial networks, we treat the calculation of p_{det} slightly differently between the two cases, as discussed in Sec. III A for terrestrial detectors, and in Sec. III B for space-based detectors.

For all binaries, we evaluate the integral in Eq. (13) through a large population of binary systems that are evolved to the point of becoming BBHs and are weighted according to the probability that a binary of this type would actually be found in the Universe given some population model. This probability is comprised of factors like the star formation rate (SFR), cosmological evolution of the metallicity, the distribution of masses for these stellar populations, etc.; the continuous equation in Eq. (13) then becomes a discrete sum,

$$r = \sum_i r_i p_{\text{det}}(\vec{\lambda}_i), \qquad (14)$$

where the index i refers to samples in the simulation, r_i is the intrinsic merger rate, which depends on parameters like the SFR and the mass distribution, and $p_{\text{det}}(\vec{\lambda}_i)$ is the detection probability evaluated for the source parameters of the particular sample. This detection probability is $p_{\text{det,terr}}$ when considering a terrestrial network only, $p^{\text{SOBH}}_{\text{det,space}}$ when considering multiband events, or $p^{\text{MBH}}_{\text{det,space}}$ when considering MBH binaries detectable only by LISA.

The intrinsic merger rate r_i varies depending on the catalog used. For the case of the SPOPS simulations, we utilized the original `StarTrack` data at the foundation of each SPOPS catalog (cf. Ref. [90] for details) to construct the intrinsic merger rate in Eq. (13). For MBH catalogs, the intrinsic merger rate r_i becomes [74]

$$r_i = 4\pi W_{\text{PS},i} \left(\frac{D_L(z_i)}{1+z_i} \right)^2, \qquad (15)$$

as outlined in the data release [74,91]. The parameter $W_{\text{PS},i}$ is the weight on the Press-Schechter mass function divided by the number of realizations [87].

C. Synthetic catalog creation

Calculating the BBH detection rate only gets us halfway to our end goal. Once we have the number of mergers we expect to detect for each network and simulated population, we still need to synthesize BBH catalogs to use for the later Fisher analysis in this paper.

To create these synthetic catalogs, we sample directly from the population simulations, using Monte Carlo rejection sampling. The probability of accepting a sample is based on the intrinsic merger rate r_i in Eq. (14), evaluated for a single simulation entry, which comes directly from the simulation data itself. This gives a distribution of sources that reflects the expected BBH distributions for each evolution prescription. With a distribution of "intrinsic" mergers in this realization of the Universe, we assign any remaining parameters according to reasonable distributions. For sky-location and orientation, this distribution is uniform in α, $\sin\delta$, $\cos\theta_L$, and ϕ_L.

For the binary's merger time, we use a uniform distribution in GMST for the terrestrial networks, which impacts the orientation of the terrestrial network at the time of merger. This effect is completely degenerate with the right ascension of the binary, which is also randomly uniform in α. We use a similar prescription for MBH binaries, where the signal duration is typically shorter than the observation period. We employ a uniform distribution in time from 0 to $T_{\rm obs}$, which again translates to a uniform distribution in detector orientation (random position of LISA in its orbit).

Candidates for multiband detection are more nuanced. The signal is typically detectable for much longer than the observation period, and the frequency-time relation is nonlinear because of the familiar chirping behavior of GW signals. For this class of sources, we randomly assign a signal starting time, which has a power-law relation with

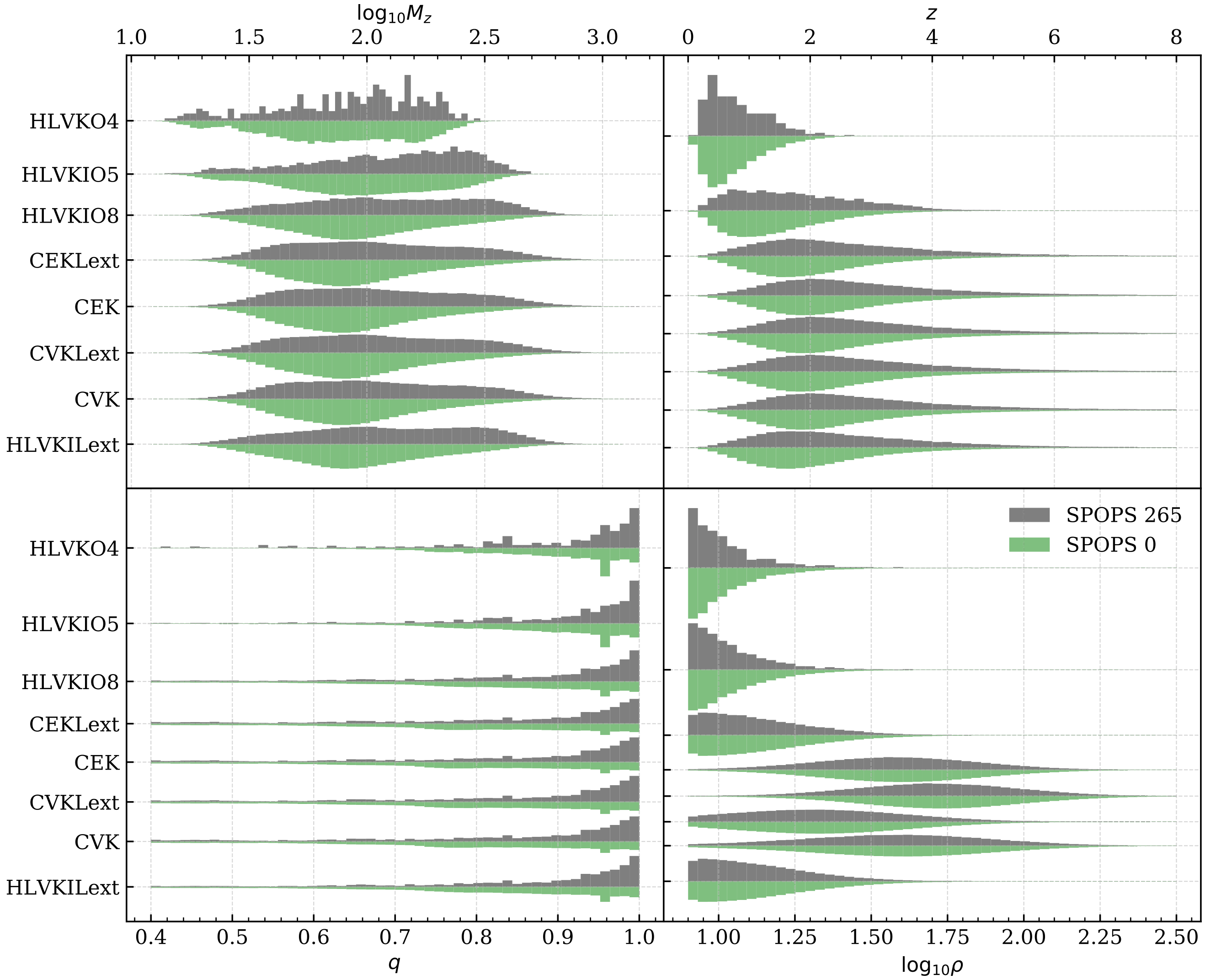

FIG. 4. Distributions of the different source properties detected by each network. For each detector network, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1+z)$, mass ratio $q = m_2/m_1 < 1$, redshift z, and SNR ρ in log-space (base 10). Each plot is split, with the upper (grey) half coming from the $\sigma = 265$ km/s SPOPS simulations, and the lower (green) half coming from the $\sigma = 0$ km/s simulations.

the starting frequency: cf. Eq. (11). In this case, the position of the binary in time not only affects the orientation of LISA, but also the initial and final frequencies of the signal. This assignment of time is important, as assigning a uniformly random initial frequency would create a bias towards seeing sources close to merger.

Once the full parameter vector has been specified, we proceed to calculate the SNR for the source in question. Sources meeting the SNR threshold requirements are retained in the final catalog. This process is repeated as necessary until we have a catalog of sources that matches the number of BBHs predicted by our rate calculations in Sec. IV B.

There are some drawbacks to this scheme. If this process is repeated enough times, sources in the simulation will begin to be reused, as there are a fixed number of possible sources to draw from. For this study, however, these effects are negligible, as the number of the sources in the simulations is larger than any single catalog we construct. Furthermore, the effects will be further mitigated by randomly assigning the rest of the parameter vector not coming from the simulation, which will imbue at least slightly different properties to each source, even if one were reused.

To recap, our process can be broken down into the following steps:

(1) Perform rejection sampling on the simulation entries according to the probability of merging, neglecting detector selection effects.

(2) Keep the "successful" events, and randomly draw the rest of the requisite parameters according to their individual distributions.

(3) Calculate the SNR for the given detector network. If the binary meets the threshold requirements, keep the source in the final catalog.

The source properties of the various *detected* catalogs are shown in Fig. 4 for the SOBH populations, and in Fig. 5 for the MBH populations targeted by LISA. Both figures show the distributions of the redshifted total mass M_z, the mass ratio $q = m_2/m_1 < 1$, the redshift z, and the SNR ρ of the detected populations of sources for different detector configurations and population models. For the SOBH sources shown in Fig. 4, the y-axis labels correspond to different detector combinations, while the upper (grey) and lower (green) histograms correspond to the two different kick magnitudes ($\sigma = 265$ km/s and $\sigma = 0$ km/s) chosen to bracket SOBH population models.

In the LISA SMBH case of Fig. 5, the same properties are plotted for the three populations models and for a four-year and ten-year LISA mission. Note that the y-axis label now corresponds to different population models, and each half of the violin plot corresponds to different mission durations: the upper (grey) half corresponds to the "nominal" four-year LISA mission, and the lower (green) half corresponding to an extended ten-year mission.

The detection rates, cumulative detected sources, and average SNR for each class of sources are shown in Fig. 6, where sources are broken down into four distinct categories:

(i) "SOBH—TERR": SOBH candidates detected only by a terrestrial network;

(ii) "SOBH—MB": SOBH candidates detected by both a terrestrial network and LISA (multiband);

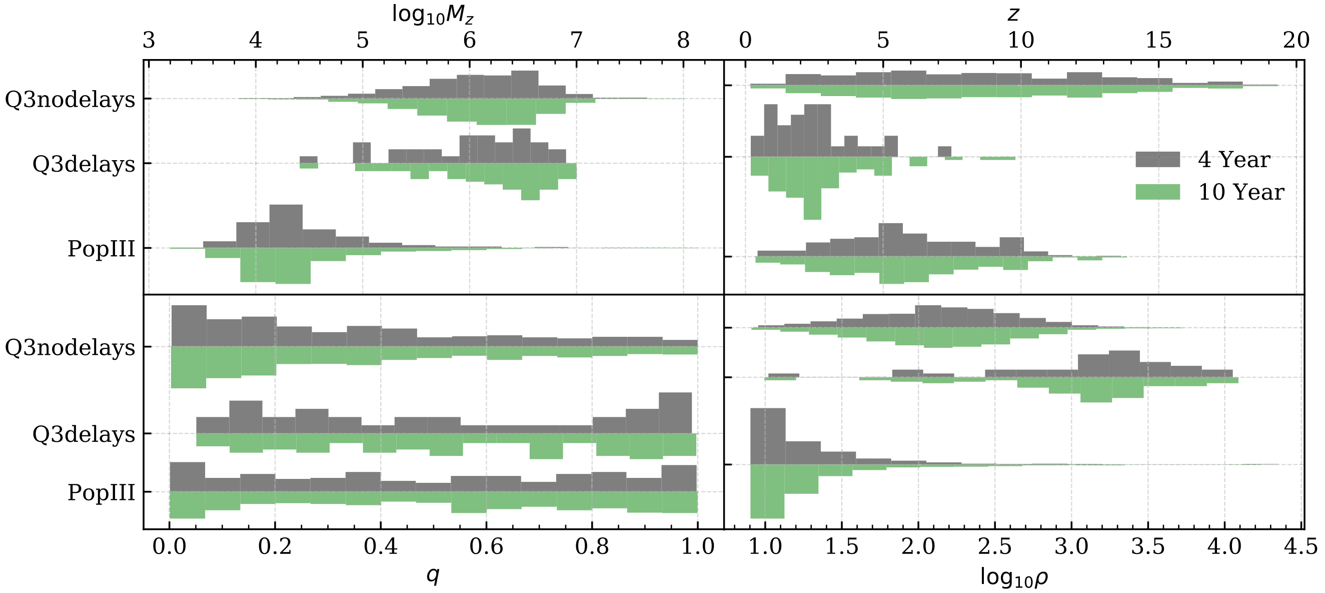

FIG. 5. Distributions of the different MBH binary source properties detected by LISA. For each MBH binary simulations, labeled across the y-axis, we plot the distribution of the total detector-frame mass $M_z = M(1 + z)$, mass ratio $q = m_2/m_1 < 1$, redshift z, and SNR ρ in log-space (base 10). Each plot is split in two, with the upper (grey) half corresponding to a "nominal" four-year LISA mission, and the lower (green) half corresponding to an extended ten-year mission.

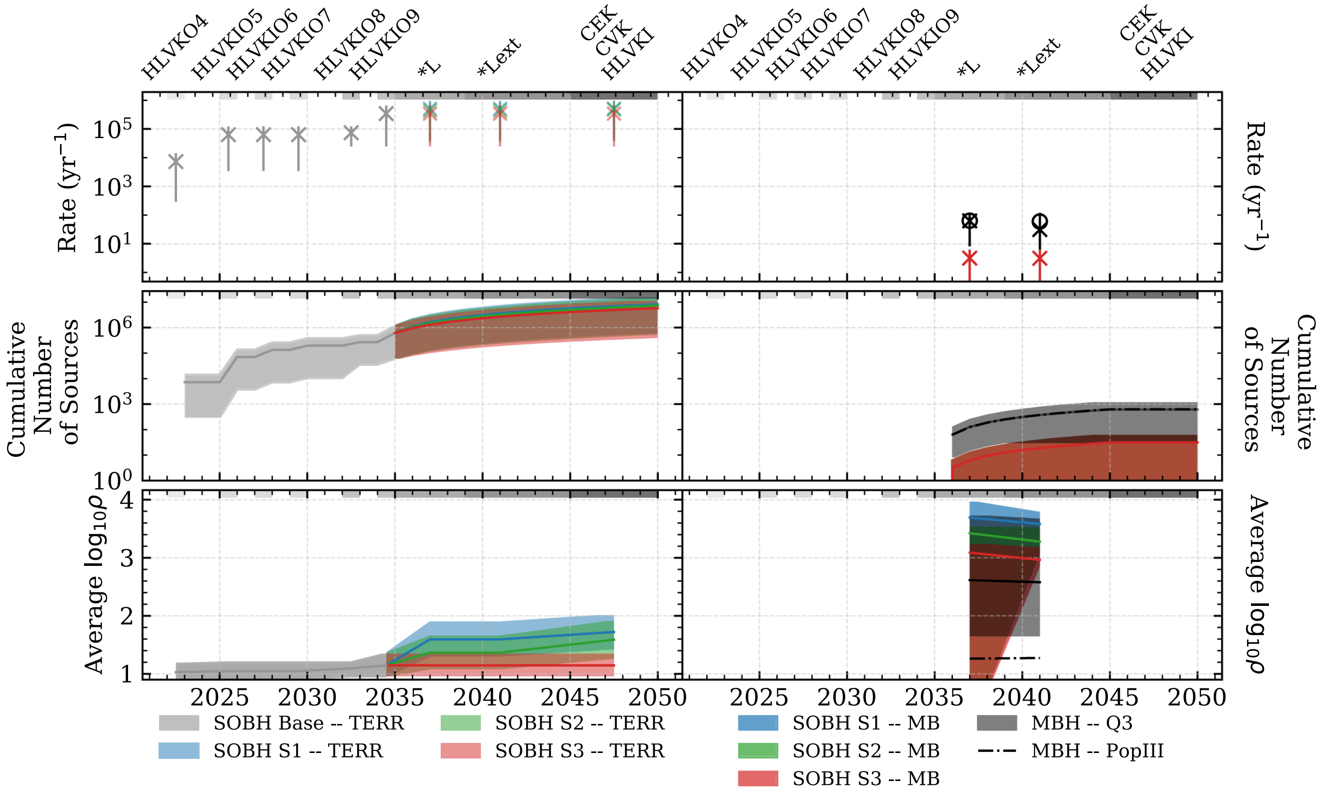

FIG. 6. Properties of detected merger events for various detector networks and population models. The left panels refer to terrestrial-only sources, while MBHs and multiband sources are shown on the right. The points and thick lines show the mean values, while the shaded regions and error bars encompass the optimistic and pessimistic scenarios. The assumed detector network is shown in the top x-axis (using the notation of Table III), while the corresponding years are shown on the bottom x-axis. The top panels show the rates of detected mergers for each class of sources; circles refer to the PopIII MBH population. The middle panels show the cumulative number of observed sources: here the three different multiband scenarios are identical, as the choice of terrestrial network has little impact on the number of multiband sources we can detect [33]. The bottom panels show the average $\log_{10}$SNR. Here the lower (upper) bounds correspond to subtracting (adding) the standard deviation to the mean value of the most pessimistic (optimistic) scenario.

(iii) “MBH—PopIII”: MBH sources from the PopIII model (light seeds);

(iv) “MBH—Q3”: MBH sources from both Q3 (heavy seeds) models, with shaded bands indicating the range of uncertainty on delays between galaxy mergers and BH mergers.

The year is shown across the bottom x-axis, while the detector network timeline is shown across the top x-axis using the acronyms defined in Table III. The solid lines and markers represent the mean values of the different quantities when considering each population model and optimistic/pessimistic detector configurations. The error bars and shaded regions represent the most optimistic and most pessimistic scenarios, except in the case of the SNR in the third panel, where the upper and lower bounds are the optimistic (pessimistic) average plus (minus) the standard deviation of the optimistic (pessimistic) distribution. There is no error for the PopIII model, as we only have one iteration of this model and only one noise curve for LISA. The detection rates for SOBHs and MBHs in the different scenarios are also listed in Table VI.

Roughly speaking, the power of a detector network to reveal new physics comes from a combination of (i) the number of sources the network can detect, and (ii) the typical quality of each signal (as measured by the SNR). Figure 6 attempts to capture the zeroth-order difference between each detector configuration and population model in these two aspects. The punchline is that although LISA will be able, on average, to see events with much larger SNR, these are just a few compared to the abundant number of sources that ground-based detectors will observe (albeit at typically lower SNR). The precision of GR tests scales as ρ^{-1}, and it is approximately proportional to $\sqrt{N}$ for N events [92], therefore it is not immediately obvious which set of observations will be best at testing GR. With our catalogs this question can be answered quantitatively. As we discuss below, ground-based and space-based detectors are complementary to each other.

TABLE VI. Detection rates for the detector networks and population models examined in this study. For SOBH populations, the first number in the parentheses is the detection rate for the terrestrial-only network (neglecting LISA), while the second number is the detection rate for multiband events seen in both the terrestrial network and LISA. For MBH populations, we show the detection rate for LISA for the PopIII, light-seeding scenario, as well as for the Q3, heavy-seeding scenario. In the case of Q3, the first number in parentheses corresponds to delayed mergers (Q3delays) and the second number to the nondelayed version (Q3nodelays).

SOBH rates (yr^{-1})		
Network	SPOPS 0 (T, MB)	SPOPS 265 (T, MB)
HLVKO4	$(1.43\times10^4, 0)$	$(2.90\times10^2, 0)$
HLVKIO5-O7	$(1.22\times10^5, 0)$	$(3.43\times10^3, 0)$
HLVKIO8-O9	$(6.60\times10^5, 0)$	$(2.48\times10^4, 0)$
	Scenario 1	
CEKL	$(9.70\times10^5, 2.58)$	$(3.96\times10^4, 0.0854)$
CEKLext	$(9.70\times10^5, 6.24)$	$(3.96\times10^4, 0.210)$
CEK	$(9.72\times10^5, 0)$	$(3.97\times10^4, 0)$
	Scenario 2	
CVKL	$(8.36\times10^5, 2.58)$	$(3.36\times10^4, 0.0854)$
CVKLext	$(8.36\times10^5, 6.24)$	$(3.36\times10^4, 0.210)$
CVK	$(9.26\times10^5, 0)$	$(3.77\times10^4, 0)$
	Scenario 3	
HLVKIL	$(6.60\times10^5, 2.58)$	$(2.48\times10^4, 0.0854)$
HLVKILext	$(6.60\times10^5, 6.24)$	$(2.48\times10^4, 0.210)$
HLVKI+	$(6.60\times10^5, 0)$	$(2.48\times10^4, 0)$

MBH Rates (yr^{-1})		
Network	PopIII	Q3 (delay, nodelay)
LISA	62.5	(8.11, 119.1)

V. PARAMETER ESTIMATION

In this section we describe the statistical methods we will use to carry out projections on the strength of tests of GR in the future, as well as our waveform model and the numerical implementation.

A. Basics of Fisher analysis

The backbone of this work is built on the estimation of the posterior distributions that might be inferred based on our synthetic signals. Given a loud signal with a large enough SNR, the likelihood of the data, i.e., the probability that one would see a data set d given a model with parameters $\vec{\theta}$, can be expanded about the maximum likelihood (ML) parameters $\vec{\theta}_{\mathrm{ML}}$. This expansion taken out to second order results in the following approximate likelihood function (where we focus on a single detector for the moment) [72,93]:

$$\mathcal{L} \propto \exp\left[-\frac{1}{2}\Gamma_{ij}\Delta\theta^i\Delta\theta^j\right], \qquad (16)$$

where $\Delta\theta^i = \theta^i_{\mathrm{ML}} - \theta^i$ are deviations from the ML values, and Γ_{ij} is the Fisher information matrix,

$$\Gamma_{ij} = (\partial_i h|\partial_j h)|_{\mathrm{ML}}. \qquad (17)$$

As before, h is the template response function, and the noise-weighted inner product is given by

$$(A|B) = 4\mathrm{Re}\left[\int \frac{\tilde{A}\tilde{B}^*}{S_n(f)}df\right], \qquad (18)$$

with $S_n(f)$ the noise power spectral density. By truncating the expansion at second order, we have effectively represented our posterior probability distribution as a multidimensional Gaussian with a covariance matrix given by $\Sigma^{ij} = (\Gamma^{-1})^{ij}$. The variances of individual parameters can then be read off to be $\sigma^i = \sqrt{\Sigma^{ii}}$, where index summation is not implied.

In an attempt to capture the hard boundaries on the spin components (the dimensionless spin magnitudes $|\chi_i|$ and in-plane spin component χ_p in GR should not exceed 1), we incorporate a Gaussian prior on these two parameters with a width of 1. We do so by adding to the Fisher matrix diagonal terms of the form [72,93,94],

$$\Gamma_{ij} \to \Gamma_{ij} + \Gamma^0_{ij}, \qquad (19)$$

where Γ^0_{ii} represents our prior distribution and is given by

$$\Gamma^0_{ij} = \delta_{\chi_1,\chi_1} + \delta_{\chi_2,\chi_2} + \delta_{\chi_p,\chi_p}. \qquad (20)$$

In the case of multiple observations for a single source, we simply generalize the above results through sums. For example, the likelihood for a single event observed with N detectors can be expanded quadratically via

$$\mathcal{L} \propto \exp\left[-\frac{1}{2}\Delta\theta^i\Delta\theta^j\sum_k^N \Gamma_{ij,k}\right], \qquad (21)$$

where the subscript k labels the kth detector, and we have assumed that the parameters $\vec{\theta}$ are globally defined. This gives the final covariance matrix,

$$\Sigma^{ij} = \left(\left(\sum_k^N \Gamma_k + \Gamma^0\right)^{-1}\right)^{ij}. \qquad (22)$$

To improve readability, additional details on the calculation of the Fisher matrix are given in Appendix A.

B. Waveform model for the Fisher analysis

For the Fisher studies carried out in this paper, we model binary merger waveforms using the phenomenological waveform model IMRPhenomPv2 [25–27], which allows us to capture certain spin precessional effects from inspiral until merger. The software used in this work was predominantly written from scratch, but the software library LALSuite [70] was used for comparison and to verify our implementation. For the actual parameter estimation calculation with LISA, we rescale the sensitivity curve to remove the sky-averaging numerical factor, and we account for the geometric factor of $\sqrt{3}/2$ manually in the LISA response function ("LISA—non-sky-averaged" in Fig. 2), following Ref. [69].

To fully specify the waveform produced by the IMRPhenomPv2 template in GR, we need a 13-dimensional vector of parameters,

$$\vec{\lambda}_{\mathrm{Pv2,GR}} = [\alpha, \delta, \theta_{\mathrm{L}}, \phi_{\mathrm{L}}, \phi_{\mathrm{ref}}, t_{c,\mathrm{ref}}, D_L, \mathcal{M}, \eta, \chi_1, \chi_2, \chi_{\mathrm{p}}, \phi_{\mathrm{p}}]. \tag{23}$$

The first 11 parameters are the same as those introduced for the IMRPhenomD model in Sec. III C. The parameters χ_{p} and ϕ_{p} define the magnitude and direction of the in-plane component of the spin, defined as [95]

$$\chi_{\mathrm{p}} = \frac{1}{B_1 m_1^2} \max\left(B_1 S_{1\perp}, B_2 S_{2\perp}\right), \tag{24}$$

where $B_1 = 2 + 3q/2$, $B_2 = 2 + 3/(2q)$, $q = m_2/m_1 < 1$ is the mass ratio, and $S_{i\perp}$ is the projection of the spin of BH i on the plane orthogonal to the orbital angular momentum $\mathbf{L}$.

This IMRPhenomPv2 is then deformed through parametrized post-Einsteinian corrections to model generic, theory-independent modifications to GR [28–31]. We worked with deformations of two types,

$$\tilde{h}_{\mathrm{gen}}(\vec{\lambda}_{\mathrm{Pv2}}, \beta) = \begin{cases} \tilde{h}_{\mathrm{GR}} e^{i\beta(\mathcal{M}\pi f)^{b/3}} & f < 0.018m \\ \tilde{h}_{\mathrm{GR}} & 0.018m < f, \end{cases} \tag{25}$$

$$\tilde{h}_{\mathrm{prop}}(\vec{\lambda}_{\mathrm{Pv2}}, \beta) = \tilde{h}_{\mathrm{GR}} e^{i\beta(\mathcal{M}\pi f)^{b/3}}, \tag{26}$$

where the first waveform h_{gen} represents deviations from GR caused by modified generation mechanisms, and h_{prop} represents deviations from GR caused by modified propagation mechanisms. Details (including the motivation for these implementations, and the disparity of the results between the two types of deviations) are discussed in Appendix C. As outlined there, differences are minor, and therefore from now on we will focus on the propagation mechanism, unless otherwise specified. The parameter β controls the magnitude of the deformation, and b controls the type of deformation considered. The ppE version of the IMRPhenomPv2 model is then controlled by the parameters,

$$\vec{\lambda}_{\mathrm{Pv2,ppE}} = \vec{\lambda}_{\mathrm{Pv2,GR}} \cup \{\beta\}. \tag{27}$$

Recall that, in PN language [96], a term in the phase that is proportional to $(\pi\mathcal{M}f)^{b/3}$ is said to be of $(b+5)/2$ PN order. The waveform model above is identical to the gIMR model coded up in LAL, and used by the LVC when performing parametrized PN tests of GR on GW data.

The main power of the ppE approach is its ability to map the ppE deformations to known theories of gravity. Table VII presents the mapping between (β, b) and the coupling constants in various theories of gravity (see Appendix B for a more detailed review of these mappings).

This table makes it clear then that ppE deformations are not false degrees of freedom, in the language of [43]. Once a constraint is placed on β, one can easily map it to a constraint on the coupling constants of a given theory through Table VII. This reparametrization is typically computationally trivial, and therefore it saves significant resources by reusing generic results, instead of repeating the analysis for every individual theory.

C. Numerical implementation

Common methods for calculating the requisite derivatives for the Fisher matrices typically involve either symbolic manipulation software, such as *Mathematica* [97], or the use of numerical differentiation based on a

TABLE VII. A summary of the theories examined in this work (adapted and updated from [39,45]). The columns (in order) list the theory in question (unless a generic deviation is being examined), the physical interpretation of the modification, the way the modification is introduced into the waveform, the PN order at which the modification is introduced, the equation specifying the ppE-theory mapping, and the b parameter in the ppE framework. The practical ramifications between "generation" vs "propagation" effects relates to how the modification is introduced into the waveform, as explained in Appendix C.

Theory or physical process	Physical modification	G/P	PN order	β	Theory parameter	b
Generic dipole radiation	Dipole radiation	G	−1	(B2)	$\delta\dot{E}$	−7
Einstein-dilaton Gauss-Bonnet	Dipole radiation	G	−1	(B3)	$\sqrt{\alpha_{\mathrm{EdGB}}}$	−7
Black hole evaporation	Extra dimensions	G	−4	(B6)	$\dot{M}$	−13
Time varying G	LPI	G	−4	(B7)	$\dot{G}_z$	−13
Massive graviton	Nonzero graviton mass	P	1	(B11)	m_g	−3
dynamical Chern-Simons	Parity violation	G	2	(B8)	$\sqrt{\alpha_{\mathrm{dCS}}}$	−1
Noncommutative gravity	Lorentz violation	G	2	(B10)	$\sqrt{\Lambda}$	−1

finite difference scheme. The calculation of the derivatives is always followed by some sort of numerical integration, which can be based on a fairly simple method such as Simpson's rule, or some more advanced integration algorithm that might appear prepackaged in *Mathematica*.

All of these methods have their respective benefits: symbolic manipulation and complex integration algorithms provide the most accuracy, while numerical differentiation and simpler integration schemes are typically much faster. All methods also come with their respective drawbacks. The maximally accurate method of adaptive integration and symbolic differentiation in *Mathematica* can be computationally taxing, while the fully numerical approach can be prone to large errors if the stepsizes are not tuned correctly, both for the differentiation with respect to the source parameters $\vec{\theta}$, as well as for the frequency spacing in the Fisher matrix integrals. On top of these aspects, using a program like *Mathematica* can be cumbersome at times, as interfacing with lower-level (or even scripting) languages adds an extra layer of complexity.

A combination of the two extremes implemented in one low-level language would be ideal, and it is the route chosen for this work. While symbolic manipulation is not available in the language that we chose (C++), we instead implemented an automatic differentiation (AD) software package natively written in C/C++: ADOL-C [98]. The basic premise of AD (as implemented in ADOL-C) is to use operator-overloading to perform the chain-rule directly on the program itself. By hard-coding a select number of derivatives on basic mathematical functions and operations (such as trigonometric functions, exponentials, addition, multiplication, etc.) and tracing out all the operations performed on an input parameter as it is transformed into an output parameter, ADOL-C can stitch together the derivative of the original function. This results in derivatives that are exact to numerical precision. As no final, mathematical expression is output, this does not exactly constitute symbolic differentiation, but perfectly fulfills our requirements.

To complete the Fisher calculation, we take our exact derivatives (to floating-point error) and integrate them with a Gaussian quadrature scheme based on Gauss-Legendre polynomials, as in Ref. [72]. To calculate the weighting factors and the evaluation points, we have implemented a modified version of the algorithm found in Ref. [99]. While this typically incurs a high computational cost to calculate the weights and abscissas, we mitigate this fact by doing the calculation only once, and reusing the results for each Fisher matrix. This results in integration errors orders of magnitude lower than a typical "Simpson's rule" scheme, with the same computational speed per data point.

VI. TESTS OF GENERAL RELATIVITY

In this section we summarize the main results of the analysis described above. We begin with the constraints on generic modifications as a function of time for each population and network (Sec. VI A). Next, we translate these into constraints on specific theories (Sec. VI B, and in particular Table VII).

A. Constraints on generic modifications

Let us begin by showing in Fig. 7 the projected strength of constraints on modifications at various PN orders (shown in different panels) as a function of time. Detector scenarios are labeled at the top, and the various astrophysical population classes are separated to facilitate visual comparisons. Recall from Sec. II that we consider three detector scenarios (S1, S2, and S3) bracketing funding uncertainties in the development of the future detector network. The source classes include the following:

(i) SOBH—TERR: SOBH populations as seen by only terrestrial networks;
(ii) SOBH—MB: SOBH events observed by both terrestrial networks and LISA;
(iii) MBHs: heavy-seed (Q3) and light-seed (PopIII) scenarios as seen by LISA.

When relevant, the error estimates shown in the figures below come from the different versions of the population model (i.e., SPOPS 265 vs SPOPS 0 and Q3delays vs Q3nodelays), as well as marginalization over the different estimates of the noise curves (i.e., the "high" and "low" sensitivity curve for Virgo and the "128 Mpc" and "80 Mpc" curves for KAGRA). The uncertainties correspond to the minimum and maximum bounds from all the combinations we studied at that point in the timeline.

Figure 7 is one of the main results of this paper. It allows us to draw many conclusions, itemized below for ease of reading[1]:

(i) Multiband sources yield the best constraints at negative PN orders. This is expected from previous work [35,45]: the long, early (almost monochromatic) inspiral signals coming from LISA observations stringently constrain deviations at low frequencies.
(ii) LISA MBH observations do better than terrestrial SOBH observations at negative PN orders. Constraints coming from the large-SNR MBH populations outperform the terrestrial networks at negative PN order, despite the large number of expected SOBH sources in the terrestrial network.
(iii) Terrestrial SOBH observations can do slightly better than LISA MBH observations at positive PN orders. Positive PN order effects can be constrained better when the merger is in band. The terrestrial networks begins to benefit from the millions of sources in the

[1] Throughout this analysis, the zeroth PN order in the GW phase refers to the first (often called "Newtonian") term in the GR series, which is proportional to $v^{-5} \propto f^{-5/3}$. Consistently, negative (positive) PN orders identify modifications entering in at lower (higher) powers of v, relative to this leading-order term.

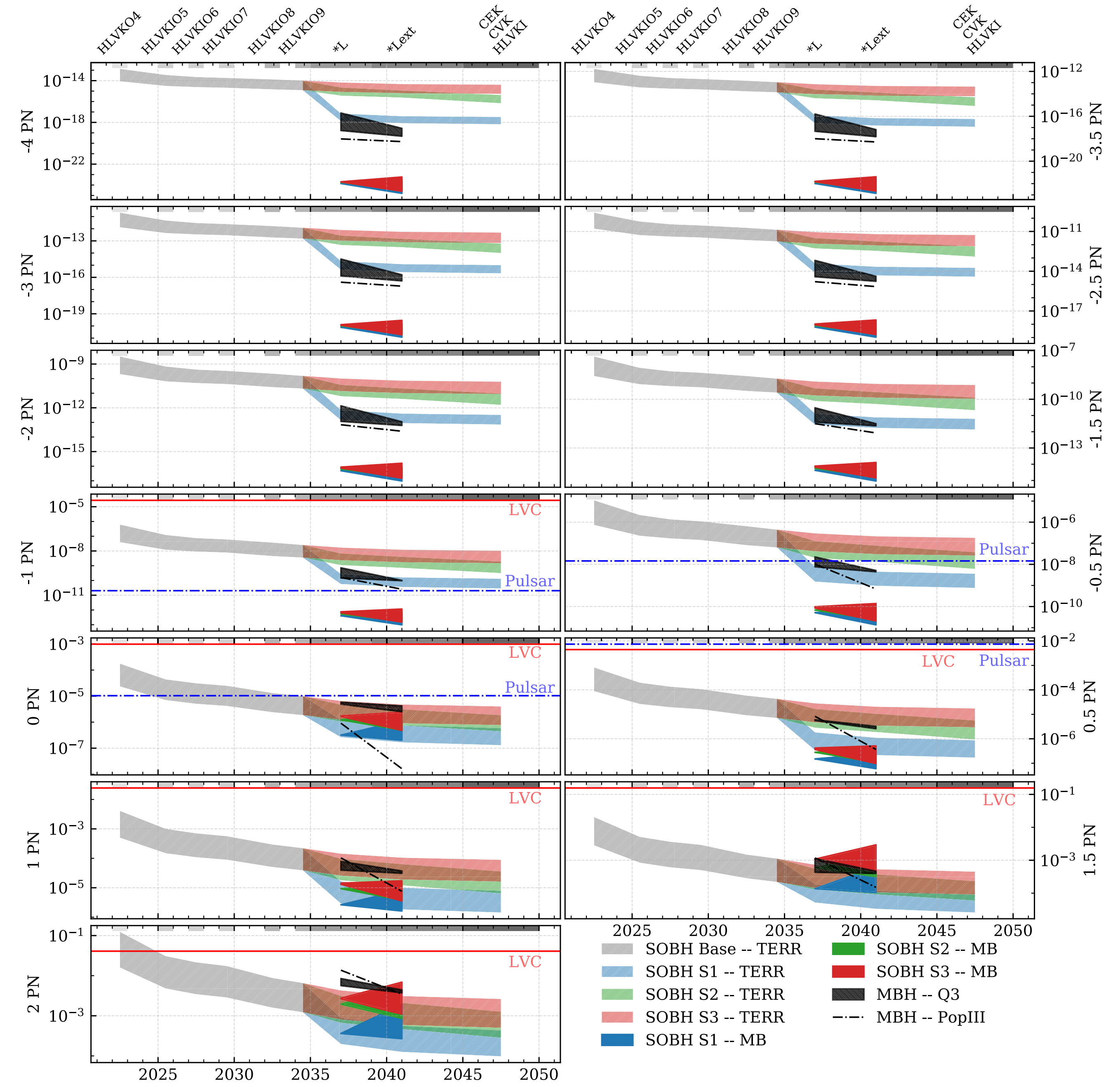

FIG. 7. Constraints on modifications to GR at various PN orders as a function of time. The colors represent different classes of populations (including SOBH terrestrial-only sources, SOBH multiband sources, MBH sources from the Q3 heavy-seed scenario, and MBH sources from the light-seed PopIII scenario). The bands in all of these scenarios—except for PopIII—correspond to astrophysical uncertainties: kick velocities $\sigma = 265$ km/s and $\sigma = 0$ km/s give the upper and lower bounds for SOBHs, while the inclusion of delays affects Q3 scenarios. Greyscale patches at the top of each panel correspond to the observation period for each network, labeled across the top. Multiband sources and MBHs yield strong constraints at negative PN orders. Terrestrial-only SOBH sources begin to contribute substantially at positive PN orders for all detector networks, with the optimistic scenario S1 yielding the best constraints. We overlay as horizontal lines the most stringent current bounds, where available and competitive, from pulsars [42] and LVC observations of GWs [3].

SOBH catalogs, but the extremely high-SNR sources in the MBH catalogs mean that LISA constraints are still competitive with terrestrial constraints.

(iv) Terrestrial network improvements make a big difference at negative PN orders. The different terrestrial network scenarios are widely separated for the negative PN effects, with the most optimistic S1 scenario vastly outperforming the S2 and S3 scenarios. This conclusions is robust with respect to astrophysical uncertainties in the population models.

(v) Network improvements are less relevant at higher PN order. In this case the three different scenarios overlap considerably (but the S1 scenario maintains a clear edge over the other two).

To understand some of these features, it can be illuminating to model the scaling behavior of bounds at different PN orders with respect to various source parameters. Below we consider an analytical approximation that can reproduce most of the observed features. We first model constraints on individual sources and then fold in the enhancement achieved by stacking multiple events.

1. Analytical scaling: Individual sources

A good first approximation is to ignore any covariances between parameters by treating the Fisher matrix as approximately diagonal, so that the bounds on the generic ppE parameter β is roughly

$$\sigma_{\beta\beta} \approx \left(\frac{1}{\Gamma_{\beta\beta}}\right)^{1/2} = \left[4\mathrm{Re}\int_{f_{\mathrm{low}}}^{f_{\mathrm{high}}} \frac{(\pi\mathcal{M}f)^{2b/3}|\tilde{h}|^2}{S_n(f)}df\right]^{-1/2}, \tag{28}$$

where f_{low} and f_{high} are the lower and upper bounds of integration. This expression can be simplified further by assuming white noise, so that $S_n(f) = S_0$ is constant, and by ignoring PN corrections to the amplitude, i.e., $|\tilde{h}| = Af^{-7/6}$, where $A \propto \mathcal{M}^{5/6}/D_L$ is an overall amplitude (see e.g., [100]). This leads to

$$\sigma_{\beta\beta} \approx \left[\frac{6A^2}{S_0}\frac{(f_{\mathrm{low}}^{2(b-2)/3} - f_{\mathrm{high}}^{2(b-2)/3})(\pi\mathcal{M})^{2b/3}}{2-b}\right]^{-1/2}, \tag{29}$$

as long as $b \neq 2$. We can further simplify the expression for $\sigma_{\beta\beta}$ by using the fact that, within the same approximations, the SNR scales like

$$\rho^2 = 4\mathrm{Re}\left[\int_{f_{\mathrm{low}}}^{f_{\mathrm{high}}} \frac{hh^*}{S_n(f)}df\right] \approx \frac{3A^2}{S_0}(f_{\mathrm{low}}^{-4/3} - f_{\mathrm{high}}^{-4/3}), \tag{30}$$

which then leads to

$$\sigma_{\beta\beta} \approx \frac{(\pi\mathcal{M})^{-b/3}}{\rho}\left[\left(1-\frac{b}{2}\right)\frac{f_{\mathrm{low}}^{-4/3} - f_{\mathrm{high}}^{-4/3}}{f_{\mathrm{low}}^{2(b-2)/3} - f_{\mathrm{high}}^{2(b-2)/3}}\right]^{1/2}. \tag{31}$$

Assuming the higher frequency cutoff to be at the Schwarzschild ISCO, so that $f_{\mathrm{high}} = f_{\mathrm{ISCO}} = 6^{-3/2}\eta^{3/5}/(\pi\mathcal{M})$, and expanding to leading order in the small quantity $\pi\mathcal{M}f_{\mathrm{low}} \ll 1$, we finally obtain the approximate scaling,

$$\sigma_{\beta\beta} \approx \left[6^{b-2}\left(\frac{b}{2}-1\right)\right]^{1/2}\frac{(\pi\mathcal{M}f_{\mathrm{low}})^{-2/3}}{\eta^{(b-2)/5}\rho}, \quad b > 2, \tag{32}$$

$$\sigma_{\beta\beta} \approx \left(1-\frac{b}{2}\right)^{1/2}\frac{(\pi\mathcal{M}f_{\mathrm{low}})^{-b/3}}{\rho}, \quad b < 2. \tag{33}$$

The expressions above do not apply to the case $b = 2$, as the integration would lead to a logarithmic scaling. Recall that $b > 2$ corresponds to PN orders higher than 3.5.

As expected, all bounds on generic ppE parameters approximately scale as the inverse of the SNR, regardless of the PN order at which they enter. What is more interesting is that they also scale with the chirp mass as $\mathcal{M}^{-b/3}$ when $b < 2$, or as $\mathcal{M}^{-2/3}$ when $b > 2$. For a single event, we then have the ratio

$$\frac{\sigma_{\beta\beta}^{\mathrm{TERR}}}{\sigma_{\beta\beta}^{\mathrm{MBH}}} \approx \frac{\rho^{\mathrm{MBH}}}{\rho^{\mathrm{TERR}}}\left(\frac{\mathcal{M}^{\mathrm{TERR}}}{\mathcal{M}^{\mathrm{MBH}}}\right)^{-b/3}\left(\frac{f_{\mathrm{low}}^{\mathrm{TERR}}}{f_{\mathrm{low}}^{\mathrm{MBH}}}\right)^{-b/3}, \tag{34}$$

for $b < 2$. Since $\rho^{\mathrm{MBH}}/\rho^{\mathrm{TERR}} \sim 10^2$, $\mathcal{M}^{\mathrm{TERR}}/\mathcal{M}^{\mathrm{MBH}} \sim 10^{-4}$ and $f_{\mathrm{low}}^{\mathrm{TERR}}/f_{\mathrm{low}}^{\mathrm{MBH}} \sim 10^5$, we conclude that the ratio $\sigma_{\beta\beta}^{\mathrm{TERR}}/\sigma_{\beta\beta}^{\mathrm{MBH}} \approx 10^{3-b/3}$. This ratio is large (favoring MBH sources) when b is negative and large, i.e., at highly negative PN orders, and slowly transitions to favor terrestrial, SOBH sources at positive PN orders, explaining the observations in items (ii) and (iii) above. The ratio degrades by approximately 4 orders of magnitude between -4 PN and 2 PN, in favor of the terrestrial network, and in agreement with Fig. 7. This scaling with b holds true regardless of the typical SNRs of the sources, as the ratio of SNRs depends on the ratio of the chirp masses of the sources, but not on the PN order.

Let us now consider the scaling of the bounds with PN order in more detail. Figure 8 shows an averaged ratio $\sigma_{\beta\beta}^{\mathrm{TERR}}/\sigma_{\beta\beta}^{\mathrm{MBH}}$ computed from the full numerical simulations of Fig. 7 (solid blue line), together with the prediction in Eq. (34) that the ratio should scale as $\propto 10^{-b/3}$ (solid black line). The numerical results (blue line, with an "uncertainty" quantified by the shaded blue region) were computed as follows. We first averaged the constraints for each population model at each PN order and for each detector network that concurrently observes with LISA; this allowed us to isolate the effect of the combination of source class and detector, neglecting the sometimes significant contribution from stacking. Ratios of the averaged quantities were then calculated for each combination of SOBH model (SPOPS 0 and SPOPS 265) and heavy-seeding MBH model (Q3delays and Q3nodelays) and for each detector network—the CEKLext, CVKLext, and HLVKILext (optimistic and pessimistic) configurations—resulting in 16 combinations in all at each PN order, assuming an extended ten-year LISA mission duration. The average of these combinations is shown as the solid blue line in Fig. 8, and the region bounded by the minimum and maximum

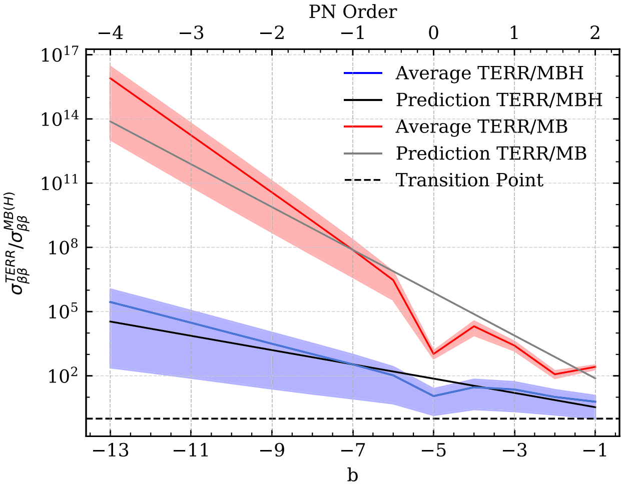

FIG. 8. Scaling relations discussed in Sec. VI A 1. The ratio $\sigma_{\beta\beta}^{\rm TERR}/\sigma_{\beta\beta}^{\rm MBH}$, calculated from the full Fisher simulations including the realistic noise curves shown in Fig. 2 and the IMRPhenomPv2 waveform, is shown in blue. The empirically measured trend is derived from averaging the constraints from each terrestrial network and each population model, then calculating the ratios of every combination of terrestrial network and SOBH model against each MBH heavy-seeding model. The blue line shows the mean ratio, and the blue shaded region is the area bounded by the maximum and minimum ratios. The red line and the red shaded region refer instead to the ratio between the terrestrial-only constraints and the multiband constraints, i.e., $\sigma_{\beta\beta}^{\rm TERR}/\sigma_{\beta\beta}^{\rm MB}$. For this class of sources, we calculate the ratio for each population model and detector network, one at a time. That is, the terrestrial-only constraints from the S1 network derived from the SPOPS 265 model are compared against the multiband constraints from the S1 network and the SPOPS 265 model. The trends predicted analytically in the text are shown in black and grey for MBH and multiband sources, respectively. The trend lines we show for our predictions have been shifted along the y-axis to better compare the with the data.

ratios is shown shaded in blue. Observe that the scaling of Eq. (34) is consistent with the averaged ratio in the entire domain; the small dip at $b = -5$ (or 0PN order) is due to degeneracies with the chirp mass, which the scaling relation does not account for.

The relation $\sigma_{\beta\beta}^{\rm TERR}/\sigma_{\beta\beta}^{\rm MBH}$ can be pushed further by comparing multiband sources against the rest of the SOBH sources detected *only* by the terrestrial network. For these two classes of sources, the masses would be comparable. Let us focus on the impact of the early inspiral observation. The ratio of the SNRs in the LISA band is of $\mathcal{O}(1)$ for typical sources, so we will neglect it for now. Typical initial frequencies, however, are quite different, with multiband sources having initial frequencies of about 10^{-2} Hz for SOBH sources that merge within several decades in the terrestrial band. This makes the ratio $f_{\rm low}^{\rm TERR}/f_{\rm low}^{\rm MB} \sim 10^3$, and thus, the constraining power of multiband sources relative to that of terrestrial-only sources is approximately $\sigma_{\beta\beta}^{\rm TERR}/\sigma_{\beta\beta}^{\rm MB} \sim 10^{-b}$, which explains the scaling observed in item (i) above. In Fig. 8 we show the averaged ratio measured from our full simulations including the noise curves shown in Fig. 2 and the IMRPhenomPv2 waveform (solid red line) as well as the 10^{-b} scaling derived from Eq. (34) (solid gray line). Again, we average the constraints from each population model at each PN order, assuming a ten-year LISA mission duration. However we do not consider every combination of population models and detector networks, but instead compare the multiband constraints from each network and SOBH model against the terrestrial-only constraints from the same combination of terrestrial network and SOBH model. That is, we compare S1 terrestrial-only constraints derived from the SPOPS 265 model against the multiband constraints with the S1 network and from the SPOPS 265 model, repeating the procedure for each terrestrial network and population model. This yields eight different combinations of population models and networks. The red line shows the average ratio for all the combinations considered, and the red-shaded region shows the area bounded by the maximum and minimum ratios. The simple analytical scaling reproduces the numerics quite well at negative PN orders, where the contribution to the constraint on the ppE parameter primarily comes from LISA observations. At positive PN orders the scaling relation breaks down for two main reasons: (i) our scaling relation neglects covariances, and (ii) the dominant source of information is no longer LISA's observation of the early inspiral, but the signal from the merger-ringdown seen by the terrestrial network.

2. Analytical scaling: Multiple sources

Our analysis above helps to elucidate some of the trends observed in our numerical simulations by examining individual sources, but it fails to capture the power of combining observations to enhance constraints on modified theories of gravity. Especially when considering terrestrial networks, this element is critical in predicting future constraints, and it is connected with our observations (iv) and (v) in the previous list.

To fully explore this facet of our predictions, we try to isolate the impact of the total number of sources on the final, cumulative constraint for a given network. As shown in Eq. (A11) of Appendix A, the combined constraint from an ensemble of simulated detections is

$$\sigma_\beta^2 = \left(\sum_i^N \frac{1}{\sigma_{\beta,i}^2}\right)^{-1}, \tag{35}$$

where $\sigma_{\beta,i}$ is the variance on β of the ith source marginalized over the source-specific parameters, including all detectors and priors, and N is the total number of sources in the ensemble. The effect of the population on all the different combinations of detector networks and PN orders can be summarized by the distribution in $\sigma_{\beta,i}$, and we find empirically that they all lie somewhere in the spectrum bounded by the following extreme scenarios:

(a) all the constraints contribute more or less equally,

(b) the total constraint is dominated by a single (or a few) observations.

When the covariances are all approximately equal, the sum above reduces to $\sigma_\beta \approx \sigma_{\beta,i}/\sqrt{N}$, but when one constraint (say $\sigma_{\beta,\mathrm{strongest}}$) dominates the ensemble, the sum reduces to $\sigma_\beta \approx \sigma_{\beta,\mathrm{strongest}}$. Naturally, in the case where all sources are more or less equally important, the power of large catalogs is maximized, and one would expect terrestrial networks observing hundreds of thousands to millions of sources to outperform networks with smaller populations, such as MBHs and multiband sources (everything else being equal). When one observation dominates the cumulative bound because of loud SNR or source parameters that maximize the constraint, then large catalogs are not as important.

In an attempt to quantify this effect, we can ask the following question: what is the minimum number of sources we can retain and still achieve a similar constraint on β? To answer this question, we take all the variances calculated with our Fisher analysis for a given population model and detector network, and order them according to the strength of the constraint from each individual source. With some threshold constraint set, we can work our way down the list, calculating the cumulative bound for the "best" N' sources at a time. We define N_{eff} as the value of N' such that our threshold constraint is achieved. Comparing the values of N_{eff} at each PN order for a single population model and network provides useful insights into how generic constraints benefit from the catalog size.

The upper panel of Fig. 9 shows the values of N_{eff} calculated using the results from our full Fisher analysis, including the noise curves shown in Fig. 2 and the IMRPhenomPv2 waveform, for the CEK network with the SPOPS 0 population model and a threshold constraint of $\log_{10}\sigma_{\beta,\mathrm{thr}} = 0.95\log_{10}\sigma_\beta$. A pronounced trend is evident: positive PN orders require up to $\sim 10^5$ sources to retain a constraint equal to our threshold value, while the most negative PN effects only require a single, highly favorable source to reach the threshold value. The lower panel of Fig. 9 merely shows the value of the full numerical constraint (red × signs) compared with our value of the threshold constraint (blue + signs): by our own definition, the threshold constraint captures most (i.e., 95%) of the full constraint.

Figure 10 shows several different facets of the data relevant to the analysis of Fig. 9. For each PN order, we have plotted three different quantities: (i) a heat map of all the sources in the catalog in the $\mathcal{M}$–ρ plane (shown in blue), which is the same for all PN orders, (ii) the contours showing the strength of the individual constraints from each source for the entire catalog (in black), and (iii) the subset of sources required to meet the threshold constraint $\sigma_{\beta,\mathrm{thr}}$ (in red), where the shade corresponds to the strength of the individual bounds.

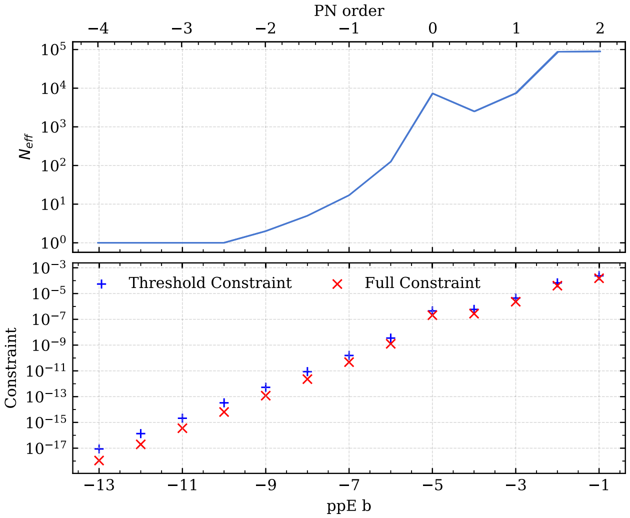

FIG. 9. Empirically determined values of N_{eff} for the CEK (scenario 1) network and the SPOPS 0 catalog, derived from our full Fisher analysis, including the noise curves shown in Fig. 2 and the IMRPhenomPv2 waveform. The parameter N_{eff} is defined as the number of sources needed from the full catalog in order to achieve a threshold constraint $\sigma_{\beta,\mathrm{thr}}$, using the most constraining sources first. Here we choose $\log_{10}\sigma_{\beta,\mathrm{thr}} = 0.95\log_{10}\sigma_\beta$, where σ_β is the cumulative bound from the full Fisher analysis for the entire catalog. The values of the threshold constraint (blue + signs) are shown alongside the full constraint (red × signs) in the lower panel. The number of required sources grows exponentially as a function of PN order: large catalogs benefit positive PN orders, but they are not as important for highly negative PN orders.

Several interesting conclusions can be drawn from this figure. First, the relation between the constraint, the SNR, and the chirp mass changes as a function of PN order. The highly positive PN orders benefit highly from loud sources, with only a slight preference for the lower mass systems (if at all), while highly negative PN effects benefit greatly from low-mass systems, with a slight preference for louder sources. This agrees with our intuition about low-mass systems being most important for negative PN effects: in Eq. (32) the chirp mass is raised to the $-b/3$ power, significantly enhancing the impact of low-mass systems for negative PN effects, while minimizing their impact for positive PN effects (assuming $b < 2$). As these figures are constructed from our fully numerical data, these trends take into account the nonlinear relation between SNR and chirp mass, as these are not independent parameters when considering realistic population models. Reasonably accurate population models are important in studies of this type, as bounds can be significantly altered by changing the distributions of source properties.

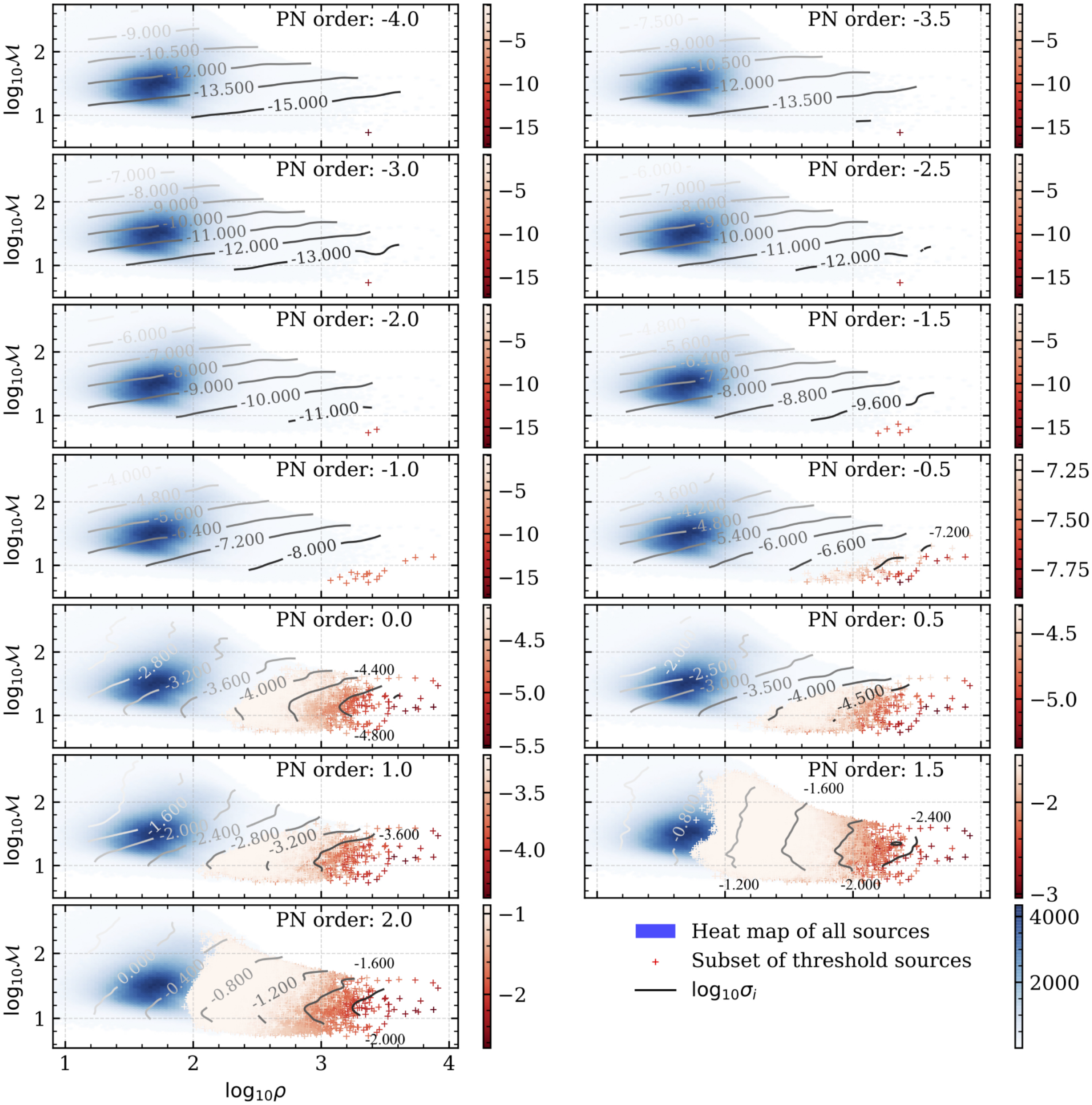

FIG. 10. Three different distributions in the $\mathcal{M}$–ρ plane for the CEK network and the SPOPS 0 population model. The blue heat map shows the distribution of the sources directly in the $\mathcal{M}$–ρ plane, and it is the same for all PN orders. The black contours show the constraints from individual sources. The red scatter plots show the sources needed to obtain a threshold cumulative constraint $\log_{10}\sigma_{\beta,\mathrm{thr}} = 0.95\log_{10}\sigma_\beta$, where the shade of red indicates the strength of the individual bounds (in log base 10). We utilized a 2σ Gaussian filter over the data to smooth out the noise and create more easily interpretable contour plots. In conjunction with Fig. 9, the growing number of scatter points as a function of PN order illustrates the increasing dependence of the cumulative constraint on the size of the source catalog. Furthermore, the relation between chirp mass, SNR, and individual bound can be seen to shift significantly between positive and negative PN orders, agreeing with the commonly held intuition that lower-mass sources are better for constraining negative PN effects. In more detail, the negative PN orders benefit highly from low-mass systems, with slight dependence on SNR, while positive PN order effects depend much more strongly on the SNR and have more minimal dependence on the chirp mass. Finally, the range of individual bounds (∼4 orders of magnitude at negative PN orders and ∼2 orders of magnitude at positive PN orders) helps to explain the different scaling relations between the cumulative bounds and the total number of sources.

A second observation one can draw from Fig. 10 relates to the change in the relation between SNR and individual constraints, which explains why the constraining-power gap between the different terrestrial network scenarios closes at positive PN orders [items (iv) and (v) from above]. The relaxation in the SNR-constraint correlation at high positive PN orders means that the huge boost in SNR from utilizing 3g detectors, as compared to a 2g only network, has only a moderate impact on the cumulative bound, *if* the 2g network is sensitive enough to observe a comparable number of sources to the 3g network. In the case of the Voyager network (HLVKI+), the much lower average SNR (shown in Fig. 4 and Fig. 6) hinders the network's capability greatly at negative PN orders, but only minimally at positive PN orders, as compared with the CEK or CVK networks shown in Fig. 7. This is because the total number of sources observed in each scenario is comparable with scenario 3, only differing by ∼30%, and allowing HLVKI+ to maintain competitive constraining power through comparably sized catalogs.

A third observation that we can make about Fig. 10 is that the range in individual bounds is also clearly PN-order dependent. The most negative PN corrections change by ∼4 orders of magnitude, while the most positive PN corrections only change by ∼2 orders of magnitude. This change in constraint range lends credence to the interpretation outlined above. When constraints are clustered closer together and contribute equally, the cumulative constraint scales strongly with the number of sources. The opposite is true when the clustering is weaker and one constraint dominates over the whole ensemble. The analysis performed here, coupled with that done in Sec. VI A 1, further clarifies the trend observed in items (ii) and (iii). The combination of the individual source scaling favoring LISA at negative PN orders is enhanced by the significant benefit from large catalogs for terrestrial networks for positive PN orders.

B. Specific theories

We can now recast the constraints on generic ppE parameters from Sec. VI A into constraints on relevant quantities in a variety of specific modified gravity theories. We list and categorize these theories in Table VII.

We will utilize the scaling analysis outlined in the previous section, with the additional step,

$$\Gamma_{\text{theory}} = \mathcal{J}^T \cdot \Gamma_{\text{ppE}} \cdot \mathcal{J}, \tag{36}$$

where $\mathcal{J}$ is the Jacobian $\partial\vec{\theta}_{\text{ppE}}/\partial\vec{\theta}_{\text{theory}}$ of the transformation, and $(\cdot)^T$ is the transpose operation. In our case, the Jacobian is diagonal. This is because the off diagonal components are all proportional to the theory-specific modifying parameter; as we inject with GR models, these are always set to zero for any specific beyond-GR theory. We can then write

$$\Gamma_{\alpha_{\text{theory}}\alpha_{\text{theory}}} = \left(\frac{\partial\beta}{\partial\alpha_{\text{theory}}}\right)^2 \Gamma_{\beta\beta}, \tag{37}$$

where β is the generic ppE modification at the corresponding PN order for a given theory, and α_{theory} is the theory-specific modifying parameter. The interested reader can find the mappings $\beta(\alpha_{\text{theory}})$ between each theory and the ppE formalism, and more in-depth explanations of their motivations, in Appendix B.

This mapping between ppE constraints and theory-specific constraints changes the scaling relations between the theory-specific bound and different source parameters, with many of the conclusions made by examining the generic constraints changing quite drastically. This is because the Jacobian typically depends on source parameters, like $\mathcal{M}$, η, χ_1, and χ_2, and this can strongly enhance the constraining power of one population of BBHs over another. No general trend can be ascertained across multiple modified theories since each coupling is different, so we will examine each theory in turn. As we will see, constraints on different theory-specific parameters scale differently with SNR, chirp mass, etcetera, impacting how the cumulative bound improves with stacking and how dependent the bound is on small numbers of loud sources. To examine this in more detail, we will focus on a single detector network (HLVKIO8) with a single population model (SPOPS 0) to try and isolate the pertinent effects for each theory.

1. Generic dipole radiation

Dipole radiation is absent in GR, since in Einstein's theory GWs are sourced by the time variation of the quadrupole moment of the stress-energy tensor. Therefore, any observation of dipole radiation would indicate a departure from GR. Dipole radiation must be sourced by additional channels of energy loss, due to the presence of new (scalar, vector or tensor) propagating degrees of freedom. By the balance law, these new channels of energy loss affect the time variation of the binding energy E, and therefore dipole effects generically enter the GW Fourier phase at −1 PN (to leading order) [45]. While many theories predict specific forms of dipole radiation, we can constrain any process leading to dipole radiation by the time rate of change of the binding energy, $\dot{E}$.

We show in Appendix B, the Jacobian in this specific class of modifications scales as

$$\left(\frac{\partial\beta}{\partial\delta\dot{E}}\right)^2 \propto \eta^{4/5}, \tag{38}$$

where $\delta\dot{E} = \dot{E} - \dot{E}_{\text{GR}}$ is the variation in $\dot{E}$ due to dipole radiation: see Eq. (B1). This implies that the scaling relations found earlier for generic ppE modifications should not change much when we translate them into constraints on dipole radiation.

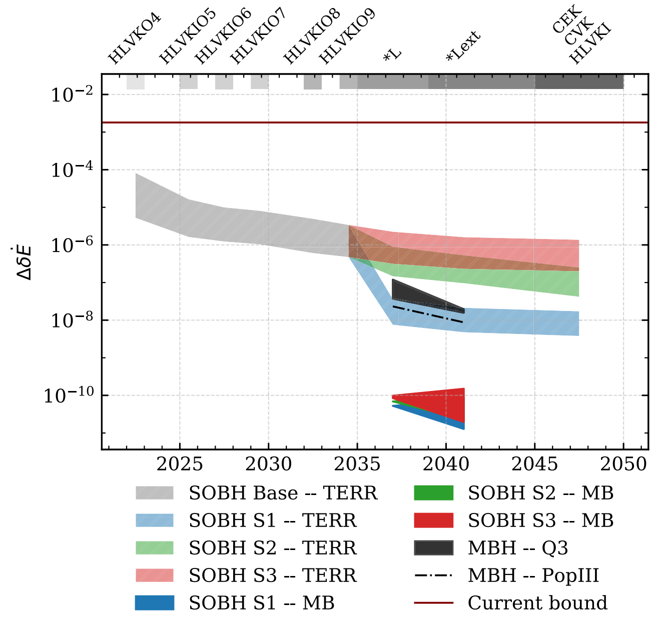

FIG. 11. Projected cumulative constraints on generic dipolar radiation for the detector networks and population models examined in this paper. The multiband sources outperform all other source classes by at least ~2 orders of magnitude, with MBH sources and the most optimistic terrestrial scenario performing comparably.

These constraints are shown in Fig. 11. As dipole radiation is a negative PN effect, multiband sources will contribute significantly, improving bounds by at least 2 orders of magnitude over any other detector network or population class. LISA observations of MBH binaries are still highly competitive, outpacing the terrestrial-only network in all cases except the most optimistic detector schedule. Furthermore, the different terrestrial networks see a wide variation, as the difference between the typical SNRs between the networks are quite large. After thirty years of GW measurements, our models suggest an improvement of 3–9 orders of magnitude over existing constraints, depending on source populations and detector characteristics, but a 9-orders-of-magnitude improvement is only possible with multiband events. All of these trends are consistent with the analysis presented in Sec. VI A 1, with constraints on this negative PN order effect benefitting from the low initial frequency and low chirp masses of LISA multiband sources. This is because dipole radiation approximately scales like a generic ppE modification in terms of SNR and chirp mass, meaning that most of the analysis from above is still valid in this case.

To better understand the numerical results presented in Fig. 11, we can look at our analytical approximation of $\Delta\delta\dot{E}$ using the methods from the previous section. After mapping the bound on the generic β to $\delta\dot{E}$, expanding in $\epsilon = \mathcal{M} f_{\text{low}}$, and setting the upper frequency to the ISCO frequency, we have the approximation,

$$\Delta\delta\dot{E} \approx \frac{112\sqrt{2}}{\eta^{2/5}} \frac{(\mathcal{M}\pi f_{\text{low}})^{7/3}}{\rho}. \tag{39}$$

Results related to this approximation are shown in Fig. 12. The left panel shows a density map of the bounds on $\delta\dot{E}$ versus the SNR of the source, with a numerical fit overlaid showing the SNR scaling trend in black. Our $1/\rho$ scaling prediction, shown in red, matches the numerics very well.

The right panel shows a density plot of the bound on $\delta\dot{E}$ versus chirp mass. To isolate the impact of the chirp mass on the attainable bound on $\delta\dot{E}$, we restrict ourselves to thin slices in different ranges of SNR (the ranges are highlighted in the top panel). This is to insulate our results from the fact that the SNR typically scales with the mass, causing a nonlinear relationship between the mass, SNR, and constraint. To ensure that the scaling does not change for different ranges of SNR, we have separately analyzed three different ranges. For lower mass systems, we see good agreement with the analytically predicted $\mathcal{M}^{7/3}$ scaling relationship, but around $\mathcal{M} \sim 30\, M_\odot$ we see a sharp transition, and our approximations fail.

The impact of these different scaling relations can be seen in the range of constraints and the cumulative constraint shown in Fig. 12. In the left panel, we have plotted the strongest and weakest constraint as solid blue lines, bounding the parameter space of single-source bounds. The cumulative bound for this one network-population combination is shown as a green line, near the bottom of the panel. As is evident in the figure, the improvement of the cumulative bound over the most stringent bound is marginal. This can be explained by the huge range of single-source bounds, covering 5 orders of magnitude, consistent with the analysis performed in Sec. VI A 2.

2. Local position invariance—Variable G theories

If the gravitational constant G were time-dependent, we would observe anomalous acceleration in the inspiral of BBHs [101]. At leading order, this affects the GW Fourier phase at −4 PN. From the transformation in Appendix B, the Jacobian to map from the generic ppE modification to the parameter $\dot{G}$ itself is

$$\left(\frac{\partial\beta}{\partial\dot{G}}\right)^2 \propto \left(\frac{\mathcal{M}}{1+z}\right)^2. \tag{40}$$

The mapping now includes a chirp mass-dependent factor, which can vary by orders of magnitude between source classes. From this scaling with chirp mass, and the fact that this modification enters at a highly negative PN order (−4PN), we expect that the best sources will be those that are seen at the widest separations (like multiband sources) and have the largest chirp mass.

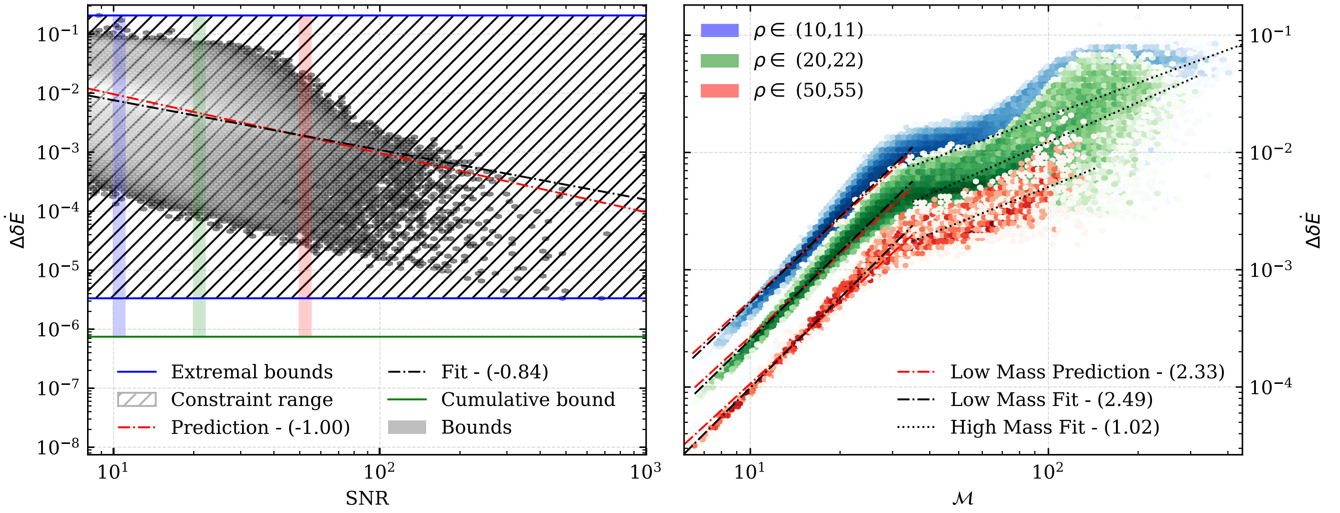

FIG. 12. Result of the scaling analysis outlined in Sec. VI B 1 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The left panel shows a heat map of the constraint on $\delta\dot{E}$ versus the SNR of the source. The solid blue lines correspond to the strongest and weakest single-source constraint, and the area between these two bounds is shown in hatching. The cumulative bound from the entire catalog is shown as the solid green line. The power-law fit to the data in the left panel is shown as the solid black curve, and our prediction for the scaling is shown as the solid red curve. The right panel shows three distinct slices of the catalog, with ranges in SNR from 10 to 11 (blue), 20 to 22 (green), and 50 to 55 (red). These ranges are highlighted in the left panel. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. There is a noticeable transition point in the distribution, so low-mass and high-mass systems were analyzed separately. The powers used in all trend lines are shown in the legend. For trend lines, the (logarithmic) offset for the predicted scaling relations has been adjusted to coincide with the empirically fit offset, to better compare the slopes of the trends. Of particular interest is the strong trend relating the SNR and the bound, as well as the tight correlation between chirp mass and constraint for low-mass systems, which seems to taper off for high-mass systems.

Our predictions for the constraints on $\dot{G}$ can be seen in Fig. 13. Multiband constraints again outperform all other source classes and detector configurations, as expected. However, because the Jacobian is proportional to $\mathcal{M}^2$, MBH sources seen by LISA are not far behind. Comparatively, the terrestrial-only bounds trail significantly behind both of these source classes, by as much as 3 orders of magnitude. There is also a wide separation between the three different terrestrial-only observation scenarios. This suggests that the cumulative bound does not benefit too much from large catalogs, but instead is dominated by a small number of favorable observations.

A variable G modification presents the first departure from our analysis on the scaling of generic results. MBH sources receive a sizeable benefit over the SOBH sources due to the Jacobian factor between parameters. Consequently, constraints on this particular modification benefit greatly from the inclusion of LISA in the GW network, both in the form of multiband and MBH observations.

Even after thirty more years of GW detections with the most ideal networks, our models indicate that the bounds will still fall far short of the current constraints on $\dot{G}$ coming from cosmology. These constraints, however, are qualitatively different from those considered here. Cosmological constraints assume a Newton constant that is linearly dependent on time in the entire cosmological history of the Universe, i.e., that $G \to G(t) \sim G_{\rm BBN} + \dot{G}_{\rm BBN} t$,

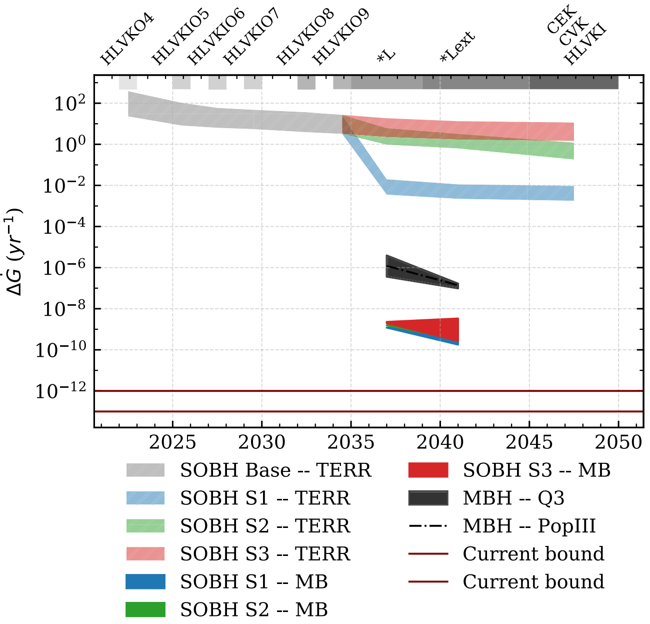

FIG. 13. Projected cumulative constraints on the time derivative of the gravitational constant $\dot{G}$ for the detector networks and population models examined in this paper. Multiband sources outperform all other source classes by ~1–2 orders of magnitude, with MBH sources performing the next best. SOBHs observed by the terrestrial network alone perform the worst, but with scenario 1 outperforming scenarios 2 and 3 due to the high SNR of the observations in the former network.

where t is time from the big bang until today, and where G_{BBN} and $\dot{G}_{\mathrm{BBN}}$ are constants. Our $\dot{G}$ constraints only assume a linear time dependence *near* the BBH merger, i.e., that $G \to G(t) \sim G_{t_c} + \dot{G}_{t_c}(t - t_c)$ for $t < t_c$ where t_c is the time of coalescence, G_{t_c} and $\dot{G}_{t_c}$ are constants, and $G(t)$ relaxes back to G_{t_c} in a few horizon light-crossing times. In our stacking analysis, we are implicitly assuming that $\dot{G}_{t_c}$ is the same for all sources in all catalogs. Therefore, it is not strictly fair to compare cosmological and GW bounds.

We can again repeat the analysis from Sec. VI A to better understand the relationship between the bound on $\dot{G}$ and various source parameters. Making the approximations outlined in Sec. VI A 1, we can approximately rewrite the constraint on $\dot{G}$ as

$$\Delta\dot{G} \approx \frac{32763}{5}\sqrt{\frac{6}{5}}\frac{(\pi\mathcal{M}f_{\mathrm{low}})^{13/3}(1+z)}{\mathcal{M}\rho}, \tag{41}$$

where we obtain the expected extra dependence on the chirp mass from the Jacobian transformation. Results pertinent to this approximation are shown in Fig. 14. The left panel shows a heat map of the $\dot{G}$ constraints against the SNR for the sources in the HLVKIO8 network and the SPOPS 0 model. The right panel shows a heat map of the constraint on $\dot{G}$ against the chirp mass, for different slices in the SNR. Notably, the scaling of the constraint on $\dot{G}$ with respect to the chirp mass matches well with our prediction of $\mathcal{M}^{10/3}$, which differs from the generic constraint by a factor of $\mathcal{M}^{-1}$ due to the Jacobian factor. Again, we see a large spread in the magnitude of the constraint, ranging over ~6 orders of magnitude. This leads to a marginal improvement of the cumulative bound over the strongest bound from a single observation, further hampering the terrestrial-only networks, in agreement with our analysis in Sec. VI A 2. After accounting for the modified scaling due to the Jacobian, the scaling relations and techniques from Sec. VI A generally hold for predicting constraints on variable G theories.

3. Lorentz violation—Noncommutative gravity

If a commutation relation is enforced between momentum and position, as in quantum mechanics, the leading order effect occurs at 2PN. Predictions for the constraints on the scale of the noncommutative relation are shown in Fig. 15. The Jacobian of the transformation found in Appendix B is given by

$$\left(\frac{\partial\beta}{\partial\Lambda^2}\right)^2 \propto \eta^{-4/5}(2\eta - 1). \tag{42}$$

The Jacobian only introduces source-dependent terms of $\mathcal{O}(1)$, and as such, bounds on Λ^2 should generally follow the scaling trends found in Sec. VI A. Given that this modification comes at 2PN, we would expect the terrestrial-only source catalogs to constrain noncommutative gravity the strongest: the power of large catalogs is enhanced, and the effect of LISA observations of the early inspiral is less relevant for positive PN effects.

The bounds predicted by our models are shown in Fig. 15. As expected, the terrestrial networks contribute the most to any future bound on noncommutative gravity. Even when just considering the three terrestrial-only

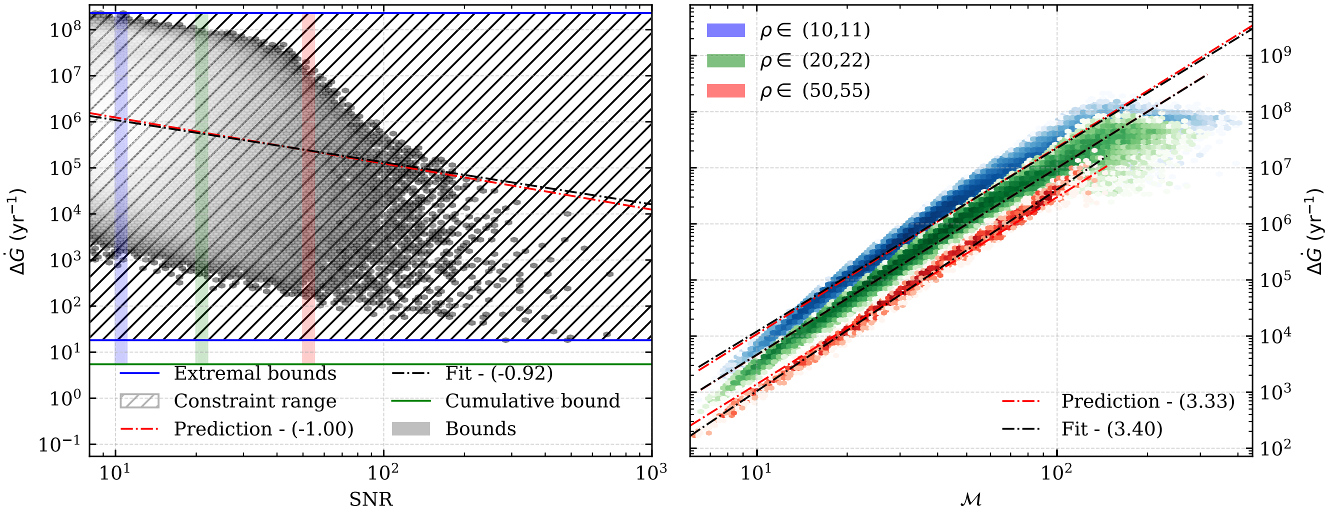

FIG. 14. Result of the scaling analysis outlined in Sec. VI B 2 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on $\dot{G}$ versus the SNR of the source. The right panel shows the density of the constraint versus $\mathcal{M}$, with empirical trends shown in black and predicted trends shown in red. Again, the strong trend relating the SNR and the bound agrees well with the prediction, and there seems to be a tight correlation between $\mathcal{M}$ and constraint, well approximated by our analysis in Sec. VI B 2.

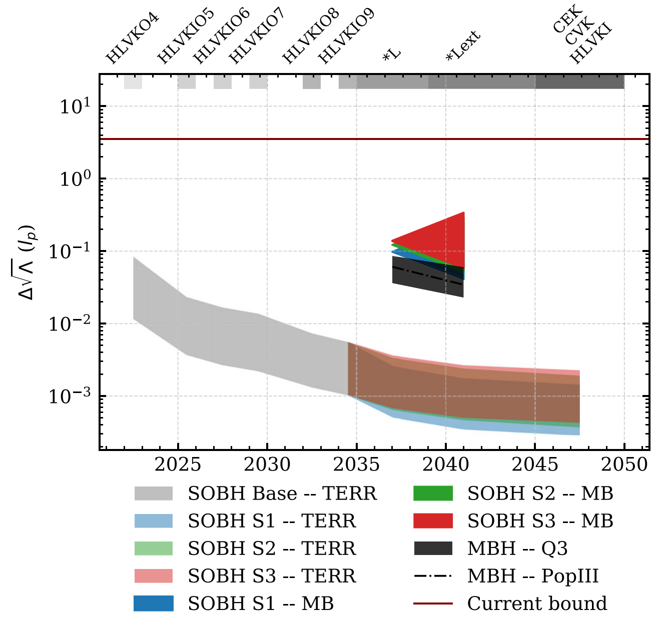

FIG. 15. Projected cumulative constraints on $\sqrt{\Lambda}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to dominate any future constraint on this particular deviation, with an improvement by 1–2 orders of magnitude over any other source classification. This conclusion seems independent of the particular terrestrial scenario we pick, with comparable performance from all three.

scenarios, the differences are minimal. Furthermore, the other source classes (MBH and multiband) perform almost identically. All of these trends further solidify our conclusion that the key to future constraints on this particular modification is large catalogs of observations, as opposed to single, favorable sources. Future constraints from all source classes should improve by 1–3 orders of magnitude over present constraints.

Continuing our analysis to explore the more subtle trends we are seeing, we can repeat the analysis outlined in Sec. VI A. This gives us the following approximation for the variance on $\sqrt{\Lambda}$:

$$\Delta\sqrt{\Lambda} \approx \left(\frac{32768}{1875}\right)^{1/8} \frac{\eta^{1/5}(\pi\mathcal{M}f_{\rm low})^{1/12}}{(1-2\eta)^{1/4}\rho^{1/4}}. \qquad (43)$$

Although the bound on Λ^2 scales as expected from Sec. VI A, approximating our bound on $\sqrt{\Lambda}$ given our constraint on Λ^2 introduces modifications to the trends we would not have expected from a straightforward extrapolation from constraints on generic modifications. Namely, we see that the bound should generically scale with the SNR as $\rho^{-1/4}$, and the constraint should scale with the chirp mass as $\mathcal{M}^{1/12}$.

Pertinent trends related to this approximation are shown in Fig. 16, where the HLVKIO8 network and the SPOPS 0 model were used to do the analysis. The left panel shows a heat map in the constraint-SNR plane, with the extremal, single source bounds shown as solid blue lines. The cumulative bound for only this network-population combination is shown as the solid green line. Our predicted trend for the constraint with respect to the SNR is shown in red, while the empirically determined trend is shown in black. The right panel shows a heat map in the constraint-chirp

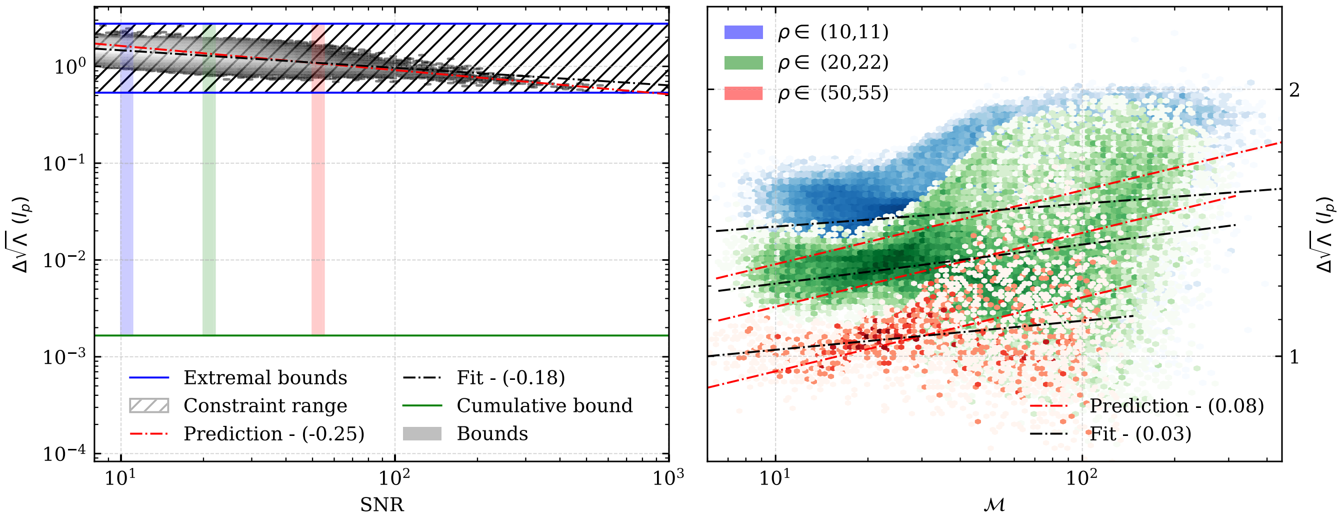

FIG. 16. Result of the scaling analysis outlined in Sec. VI B 3 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on $\sqrt{\Lambda}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The small range of constraints from the catalog lead to considerable enhancements of the cumulative bound when stacking observations, and the weak scaling with chirp mass and moderate scaling with SNR further benefit SOBH sources over other source classes.

mass plane, where we have separately analyzed three different slices of sources with specific SNRs, denoted by the colors red, blue, and green.

In the left panel of Fig. 16, we can see that our approximation for the relation between the constraint and the SNR does fairly well relative to the empirically determined trend. Furthermore, we see that the range of constraints is considerably tighter than even the generic constraints at 2PN. The largest and smallest bound for noncommutative gravity are separated by 1 order of magnitude, leading to a significant improvement of the cumulative bound over the tightest single-observation bound. This feature further explains to some degree the discrepancy between LISA sources and terrestrial-only sources in Fig. 15.

In the right panel of Fig. 16, we see much wider distributions in the constraint-chirp mass plane, as compared to the previously analyzed modifications. Our predicted trends are moderately accurate, although with noticeably lower accuracy. This is consistent with the fact the constraint scales very weakly with chirp mass ($\mathcal{M}^{1/12}$), and other correlations are widening the distribution and complicating the relation.

4. Parity violation—Dynamical Chern Simons

One of the fundamental tenets of GR is the parity invariance of the gravitational action. Dynamical Chern-Simons (dCS) gravity includes a parity-odd, second-order curvature term in the action, known as the Pontryagin density, coupled to a scalar field through a dimensionful parameter α_{dCS}. The fact that the Pontryagin density is parity-odd necessarily restricts the scalar field to also be odd in vacuum, making it an axial field. The leading-order effect in the GW phase sourced by these deviations enters at 2PN order. In Appendix B we recall that the following mapping holds:

$$\left(\frac{\partial\beta}{\partial\alpha_{\mathrm{dCS}}^2}\right)^2 \propto \frac{[\hat{m}_1 s_2^{\mathrm{dCS}} - \hat{m}_2 s_1^{\mathrm{dCS}}]^4\eta^{8/5}}{(1+z)^{-8}\mathcal{M}^8}, \tag{44}$$

where s_i^{dCS} is the BH sensitivity, defined in Eq. (B9), and $\hat{m}_i = m_i/\mathcal{M} = \eta^{-3/5}(1 \pm \sqrt{1-4\eta})/2$ for the larger (+) and smaller (−) mass. Here, we have only shown the Jacobian to leading order in spin, and we have transformed the mass components to explicitly show the chirp mass dependence. As the mass ratio and spin factors are bounded to a magnitude of $\mathcal{O}(1)$, the dependence of the Jacobian on $\mathcal{M}^{-8}$ should have the most significant effect on $\Delta\alpha_{\mathrm{dCS}}$ and *strongly* favor low-mass systems, suggesting that SOBHs would be considerably more effective than MBHs. Furthermore, as this is a positive PN modification, we would expect to see a sizeable benefit from large catalogs, given the analysis in Sec. VI A 2, and the impact of LISA observations of the early inspiral should be considerably less important. All of these factors point to the terrestrial-observation only scenarios outperforming LISA detections of MBH sources and LISA-terrestrial joint detections of multiband sources.

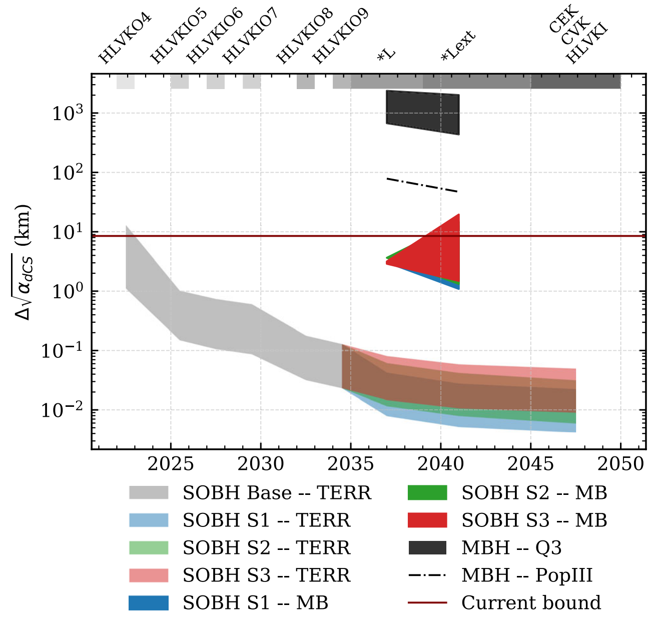

FIG. 17. Projected cumulative constraints on $\sqrt{\alpha_{\mathrm{dCS}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, dominate any future constraint on this particular deviation, with an improvement of 2–5 orders of magnitude over other source classification. This conclusion is independent of the terrestrial scenario we pick, with comparable performance from all three. Multiband sources, with their low chirp masses, seem to perform the next best.

Our predictions for the constraints on the strength of this coupling are shown in Fig. 17. Indeed, terrestrial-only detections perform the best at constraining dCS modifications to GW, with bounds up to ∼2 orders of magnitude tighter than multiband sources and ∼4–5 orders of magnitude better than MBH sources. As expected, MBH sources detected by LISA are severely inhibited by the particular Jacobian for this specific modification. Furthermore, we also see little variation between the three terrestrial scenarios, indicating that a significant weight lies with the size of the catalogs, as opposed to the source properties of a select minority of favorable observations. As the power of constraining this particular modification to GR benefits strongly from large numbers of sources, we can expect to slowly push the current bound down by ∼3 orders of magnitude, with minimal dependence on the actual detector schedule, over the course of the next thirty years.

Further analysis using the techniques in Sec. VI A 1 leads to the following approximate form of the variance:

$$\Delta\sqrt{\alpha_{\mathrm{dCS}}} \approx \left(\frac{3584\sqrt{6}}{5\pi}\right)^{1/4} \frac{(\pi\mathcal{M}f_{\mathrm{low}})^{1/12}\mathcal{M}}{(1+z)\eta^{1/5}\rho^{1/4}} \times (|3015\chi_2^2\hat{m}_1^2 - 5250\chi_1\chi_2\hat{m}_1\hat{m}_2 + 3015\chi_1^2\hat{m}_2^2 - 14(\hat{m}_2 s_1^{\mathrm{dCS}} - \hat{m}_1 s_2^{\mathrm{dCS}})^2|)^{-1/4}. \tag{45}$$

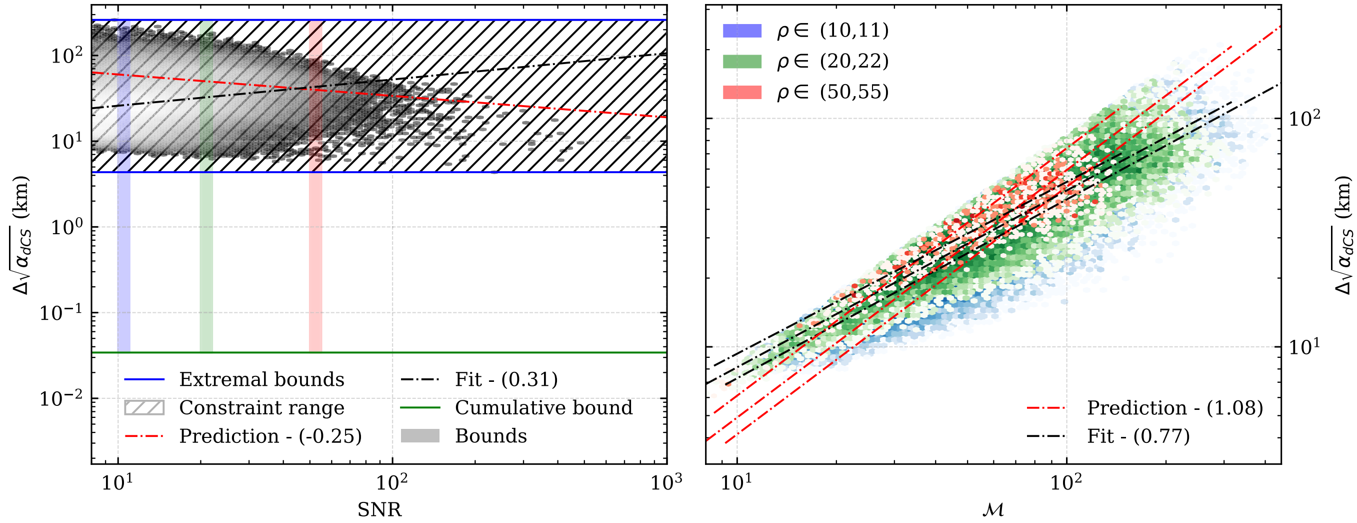

FIG. 18. Result of the scaling analysis outlined in Sec. VI B 4 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{\mathrm{dCS}}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Our prediction for the SNR scaling is considerably less accurate than for previous theories, presumably from covariances with other source parameters and competing scaling trends with the chirp mass. The tight range of constraints and large improvement of the cumulative bound over all other single source constraints, seen in the left panel, indicate strong dependence on the total number of sources in the catalog.

Beyond the additional terms coming from the Jacobian of the parameter transformation, we now see additional deviations from our analysis on generic modifications in Sec. VI A. Raising the bound on α_{dCS}^2 to the one-fourth power to obtain our further approximated bound on $\sqrt{\alpha_{\mathrm{dCS}}}$ has introduced new dependence of the constraint on all the source parameters of interest. Namely, the dependence on ρ has been amended to scale as $\rho^{-1/4}$, and the dependence on the chirp mass is now $\mathcal{M}^{13/12}$.

Results related to this analysis are shown in Fig. 18, derived from data produced with the HLVKIO8 network and the SPOPS 0 model. The left panel shows a heat map of the sources in the catalog in the $\Delta\alpha_{\mathrm{dCS}}$–SNR plane, with the extremal bounds shown in blue, and the cumulative bound (for this single catalog) shown in green. The right panel shows a heat map of the sources in the $\Delta\alpha_{\mathrm{dCS}}$–$\mathcal{M}$ plane for three slices in SNR-range (in red, blue, and green). The trends we have predicted are shown in red, while the empirically determined trends are shown in black, for both panels.

Starting in the left panel, the range in single-observation constraints on $\sqrt{\alpha_{\mathrm{dCS}}}$ is quite small. The tight range of the constraints (just 1–2 orders of magnitude between the strongest and weakest constraints) helps to explain the enhanced effectiveness of the terrestrial networks at constraining this modification, as the constraint scales favorably with large numbers of observations. This is explicitly seen by the sizable improvement of the cumulative constraint over the constraint coming from the strongest single observation.

Furthermore, in the left panel, we see that our prediction for the SNR trend does not accurately reflect what we observe in the synthetic data. This is in stark contrast with noncommutative gravity, where the modification enters at the same PN order and predicts identical scaling with respect to the SNR. Notably, this deviation also occurs in EdGB gravity, detailed below, which has a similarly complicated Jacobian. The primary differences between the modification introduced by dCS and noncommutative gravity are (i) the scaling of the constraint with respect to the chirp mass, and (ii) covariances between the modified gravity coupling constant and all other sources parameters (such as the spins and mass ratio).

For difference (i), we can examine the right panel of Fig. 18, where we see moderate agreement with our predicted scaling trend for the chirp mass and much tighter correlations for dCS than for noncommutative gravity. Not only is the trend more accurately predicted, but the scaling with chirp mass in dCS, as compared with noncommutative gravity, is considerably stronger ($\mathcal{M}^{13/12}$ as opposed to $\mathcal{M}^{1/12}$). Considering there is a negative correlation between the constraint and the SNR, a positive correlation between the constraint and the chirp mass, and a positive correlation between the SNR and chirp mass, a shift in the different trends as significant as that found in dCS may lead to the observed deterioration in our predictions.

For difference (ii), the mild agreement of the chirp mass scaling in the right panel suggests that covariances between parameters are degrading the accuracy of all of our approximations, not just the SNR. To further explore this idea, we can look at the typical range of values that the other source-dependent terms from the Jacobian in Eq. (44) can take. For the final bound from a given source, the magnitude of these additional terms in an absolute sense is important, but in terms of the trends we expect to see, the range of values these terms can take is the quantity of interest. If certain sources with comparable SNR and chirp mass have Jacobian transformations that span several orders of magnitude because of these additional terms, our simple analytical approximations cannot be expected to accurately match the synthetic data. A histogram of the spin- and mass ratio-dependent terms for both dCS and EdGB are shown in Fig. 19, where we do indeed see a non-negligible range of values. Figure 18 shows that the SNR and chirp mass both span approximately 1–2 orders of magnitude for this particular catalog, while the complicated Jacobian factors that we have neglected in our analysis span approximately 4–5 orders of magnitude. A range this large can easily erase any structure we would hope to see with our simple approximations, and helps to explain why our simple analytical approximation fails for dCS (and for EdGB, as we will discuss below).

Between these two factors, our ability to predict scaling trends of the constraint on $\sqrt{\alpha_{\rm dCS}}$ as a function of source parameters has moderate success with regards to the chirp mass but is definitely degraded in general when compared with the same analysis for general modifications. The dCS example provides direct evidence that conclusions derived from generic constraints may be highly misleading when focusing on a particular modified theory.

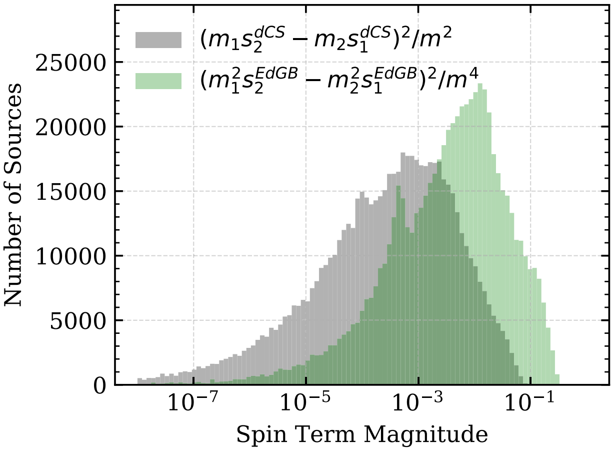

FIG. 19. Histogram of spin-related terms contributing to the relevant Fisher element for dCS and EdGB. The sources were taken from the catalog derived from the HLVKIO8 network and SPOPS 0 population model. For dCS, this only includes the term to first order in spin. The wide range of magnitudes that this term can take (5–6 orders of magnitude) helps to explain the breakdown of our ability to predict trends concerning the constraints on these theories. From Fig. 18 we see that the SNR and chirp mass only span a range of 1 or 2 orders of magnitude, and as such, the trends we would expect to see for these parameters could be completely washed out by this additional spin-dependent term, which we have neglected in our simple analysis.

5. Quadratic gravity—Einstein-dilaton-Gauss-Bonnet

Similar to dCS, Einstein-dilaton-Gauss-Bonnet (EdGB) gravity is also quadratic in curvature at the level of the action. In this case, a scalar field is coupled to the Gauss-Bonnet invariant through a dimensionful coupling constant $\alpha_{\rm EdGB}$. In contrast to dCS, the scalar field in EdGB is parity-even in vacuum (because the Gauss-Bonnet invariant is also parity-even), and the leading order correction to the GW phase comes at -1PN order, because the dominant modification to the generation of GWs is the introduction of dipolar radiation. The Jacobian for this particular theory is

$$\left(\frac{\partial\beta}{\partial\alpha_{\rm EdGB}^2}\right)^2 \propto \frac{[\hat{m}_2^2 s_1^{\rm EdGB} - \hat{m}_1^2 s_2^{\rm EdGB}]^4 \eta^{12/5}}{(1+z)^{-8}\mathcal{M}^8}, \qquad (46)$$

where $s_i^{\rm EdGB}$ is the BH sensitivity defined in Eq. (B4), and we again use the mass parameters $\hat{m}_i = m_i/\mathcal{M} = \eta^{-3/5}(1 \pm \sqrt{1-4\eta})/2$ for the larger $(+)$ and smaller $(-)$ mass. Given the new dependencies on source parameters introduced by the Jacobian, we would expect to see SOBH sources receive a sizeable boost due to the chirp mass scaling. Furthermore, this is a negative PN effect, which already tends to favor small chirp masses (cf. Sec. VI A 1). Both of these considerations imply that multiband and terrestrial networks should outperform LISA MBH sources.

Constraints on $\sqrt{\alpha_{\rm EdGB}}$ are shown in Fig. 20. Indeed, we see SOBH sources of all kinds outperforming MBH sources. Within the SOBH source classes, terrestrial networks outperform multiband sources by 1–2 orders of magnitude. While multiband sources benefit from long early inspiral observations from LISA, which encodes much information for a negative PN effect, the large catalogs of sources in the terrestrial-only catalogs are enhanced by the modified dependence on the SNR, discussed below. As a further consequence of the adjusted SNR dependence, we also see fairly minor variations between the three terrestrial network scenarios. After approximately thirty years of observations, our models indicate that we could see ~2–4 orders of magnitude improvement on previous constraints on $\sqrt{\alpha_{\rm EdGB}}$. This conclusion is fairly robust under variations of the terrestrial network.

Analyzing the constraints on $\sqrt{\alpha_{\rm EdGB}}$ with the machinery of Sec. VI A, we obtain the following approximation on the variance of the coupling parameter:

$$\Delta\sqrt{\alpha_{\rm EdGB}} \approx \left(\frac{903168}{25\pi^6}\right)^{1/8} \frac{(\pi\mathcal{M}f_{\rm low})^{7/12}\mathcal{M}}{(1+z)\eta^{3/10}\rho^{1/4}} \times (\hat{m}_2^2 s_1^{\rm EdGB} - \hat{m}_1^2 s_2^{\rm EdGB})^{-1/2}. \qquad (47)$$

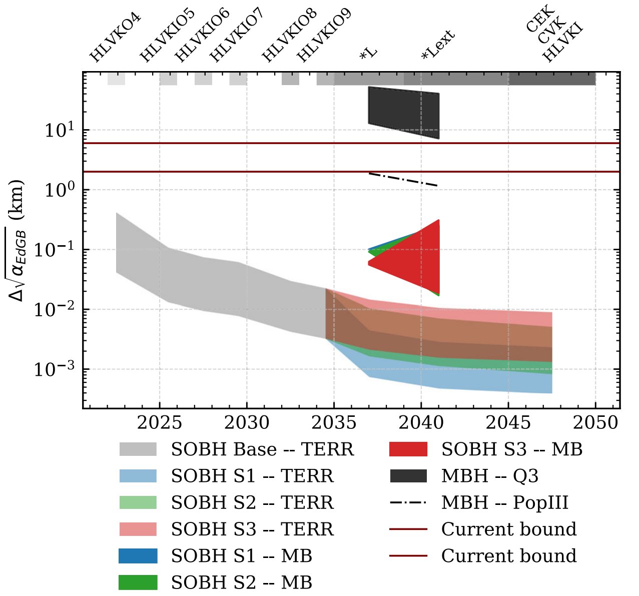

FIG. 20. Projected cumulative constraints on $\sqrt{\alpha_{\mathrm{EdGB}}}$ for the detector networks and population models examined in this paper. Terrestrial-only catalogs, with their populations of millions of sources, seem to most efficiently constraint EdGB, but multiband sources are not far behind. The modified scaling of the constraint with SNR and chirp mass work in favor of terrestrial networks, but the fact that EdGB produces a negative PN modification to leading order benefits multiband sources. MBHs are not effective at constraining EdGB, and will not contribute much to future bounds on this theory.

We now see additional modifications to the dependencies on source parameters, beyond the Jacobian shown above. Just as in the cases of dCS and noncommutative gravity, we must transform from α^2_{EdGB} to $\sqrt{\alpha_{\mathrm{EdGB}}}$, which forces the constraint to scale with $\rho^{-1/4}$ and $\mathcal{M}^{19/12}$.

Trends related to this approximation are shown in Fig. 21, produced from our simulations based on HLVKIO8 and SPOPS 0. The left panel shows a heat map of all the sources in the $\Delta\sqrt{\alpha_{\mathrm{EdGB}}}$-SNR plane, with extremal single-source constraints shown in blue, and the cumulative constraint for this catalog shown in green. The right panel shows a heat map in the $\Delta\sqrt{\alpha_{\mathrm{EdGB}}}$–$\mathcal{M}$ plane, for three different slices of SNR, shown as blue, green, and red.

In the left panel, we again see that our prediction for the SNR scaling is not accurate. Just as in dCS gravity, this discrepancy lies in covariances complicating the relationships beyond the point where our simple approximations are valid. For comparison, we can examine what we found for generic dipole radiation constraints in Sec. VI B 1, where we saw a much better agreement with our predictions for the constraint-SNR relationship. Referring again to the histogram in Fig. 19, we see that the terms related to the BH sensitivity in EdGB span several decades, washing out the trends we would expect to see from the analysis of Sec. VI A. As a by-product, these complications lead to a tight range in single-observation constraints, spanning 1–2 orders of magnitude. This in turn leads to a large enhancement for terrestrial networks: cumulative bounds from

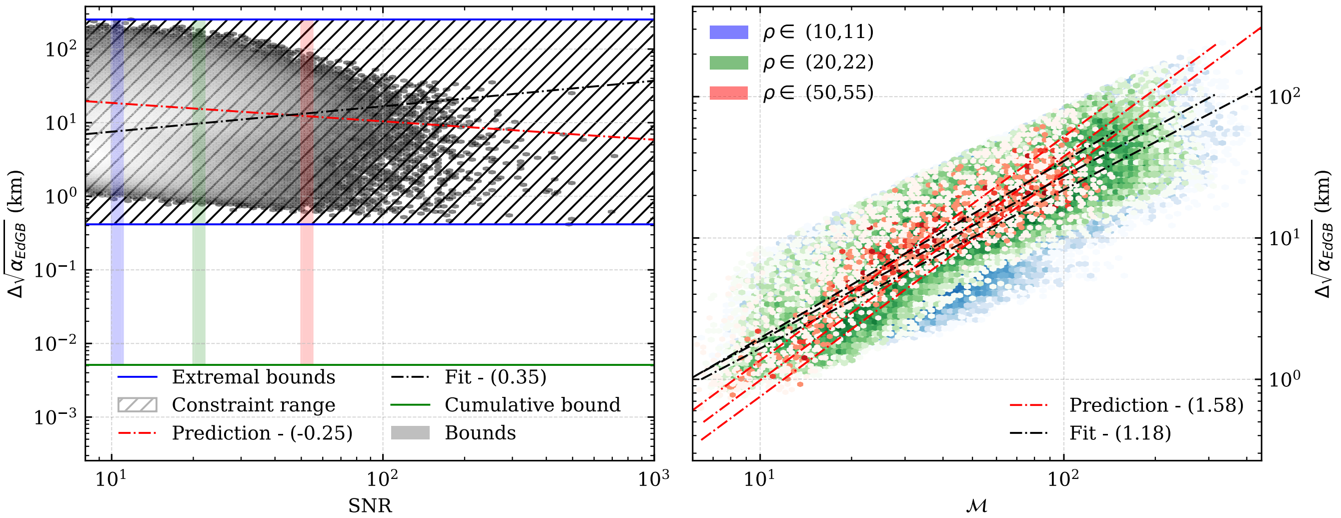

FIG. 21. Result of the scaling analysis outlined in Sec. VI B 5 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on $\sqrt{\alpha_{\mathrm{EdGB}}}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. Because of the small range in single-observation constraints (about 1–2 orders of magnitude), the cumulative bound greatly benefits from large numbers of observations, despite this being a negative PN effect that would typically be dominated by a small cadre of favorable sources. The predicted trend for the constraint-SNR relationship fails, presumably due to covariances introduced through the Jacobian. The predicted trend for the constraint-$\mathcal{M}$ relationships performs fairly well, as the correlation is enhanced through the Jacobian.

tightly grouped populations of constraints benefit from large numbers of sources, which is not typically expected from a modification at −1PN.

In the right panel, we see moderate agreement between our prediction for the $\Delta\sqrt{\alpha_{\mathrm{EdGB}}}$–$\mathcal{M}$ relationship, but again, covariances seem to degrade the quality of simple analytical scaling relationships between the constraint and the source parameters. In contrast, for generic dipole radiation we see a much tighter correlation between the constraint and the chirp mass. The difference between the two trends further confirms our explanation: more complex Jacobians tend to complicate the source parameter-constraint relation we identified in Sec. VI A.

6. *Black hole evaporation*

In the case of BH evaporation, the modification first enters the GW phase at −4PN order. The Jacobian from the ppE parameter to this particular process, as shown in Appendix B, is given by

$$\left(\frac{\partial \beta}{\partial \dot{M}}\right)^2 \propto \left[\frac{3 - 26\eta + 34\eta^2}{\eta^{2/5}(1-2\eta)}\right]^2. \tag{48}$$

As the Jacobian only depends on the system parameters through the symmetric mass ratio [bounded to (0,0.25)], no parameters specific to a given system will induce large changes in the attainable bound. This fact leads us to the conclusion that the driving factors in the constraint magnitude will be the chirp mass (benefitting SOBH sources) and the SNR (benefitting LISA MBH sources and the most sensitive ground-based detector networks). Furthermore, as this modification also enters at a highly negative PN order, multiband sources can also be expected to perform competitively.

Constraints on the rate of BH evaporation are show in Fig. 22. As expected, multiband sources constrain BH evaporation the tightest, with MBH sources from LISA's catalog trailing by 4–6 orders of magnitude. The most sensitive terrestrial network scenario examined in this paper is also competitive with the LISA MBH sources, but the other two scenarios we have considered fall behind by 2–3 orders of magnitude.

By using the machinery of Sec. VI A, we obtain the following approximate form of the bound on $\dot{M}$:

$$\Delta\dot{M} \approx \frac{425984}{5}\sqrt{\frac{6}{5}}\frac{(f_{\mathrm{low}}\pi\mathcal{M})^{13/3}\eta^{2/5}}{\rho}\left|\frac{1-2\eta}{3-26\eta+34\eta^2}\right|. \tag{49}$$

The Jacobian does not depend on the total mass and the phase modification scales linearly with the modifying parameter, so we see a scaling relation as expected from Sec. VI A.

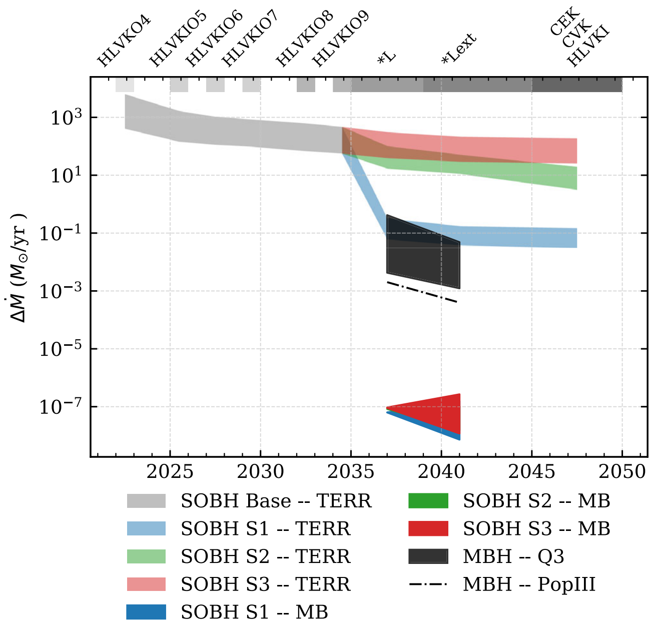

FIG. 22. Projected cumulative constraints on the rate of black hole evaporation $\dot{M}$, for the detector networks and population models examined in this paper. Our models predict multiband sources to perform the best from the three classes of sources examined in this paper, followed next by MBH observations by LISA. Terrestrial-only observations from the most optimistic scenario are competitive with LISA's MBH sources, but the other two scenarios considered in this work trail behind by 2–3 orders of magnitude.

Results related to this approximation are shown in Fig. 23. The left panel depicts a heat map of the sources in the HLVKIO8 network and the SPOPS 0 population model in the $\Delta\dot{M}$–SNR plane. The solid blue lines correspond to the strongest and weakest constraints coming from single observations, while the green line represents the cumulative bound for the entire catalog. The right panel shows a heat map in the $\Delta\dot{M}$–$\mathcal{M}$ plane for different slices of SNR (in red, blue, and green). The empirically determined scaling trends are shown in black, while our predictions for the trends are shown in red.

The left panel of Fig. 23 shows good agreement between the trends predicted by our simple, analytic calculations and the data from our fully numerical treatment. The wide distribution in constraints coming from single sources in the catalog indicates weak scaling with the size of the catalog, giving a relative boost in power to the smaller source populations in the MBH LISA and MB catalogs. This conclusion is supported by the very modest improvement of the cumulative bound for the catalog over the strongest single-source constraint. In the right panel, we see good agreement with our predicted chirp mass scaling relation. The correlation between the chirp mass and the constraint is quite tight for this particular modification, due to the strong scaling and the highly negative PN order (reducing correlations that widen the distribution).

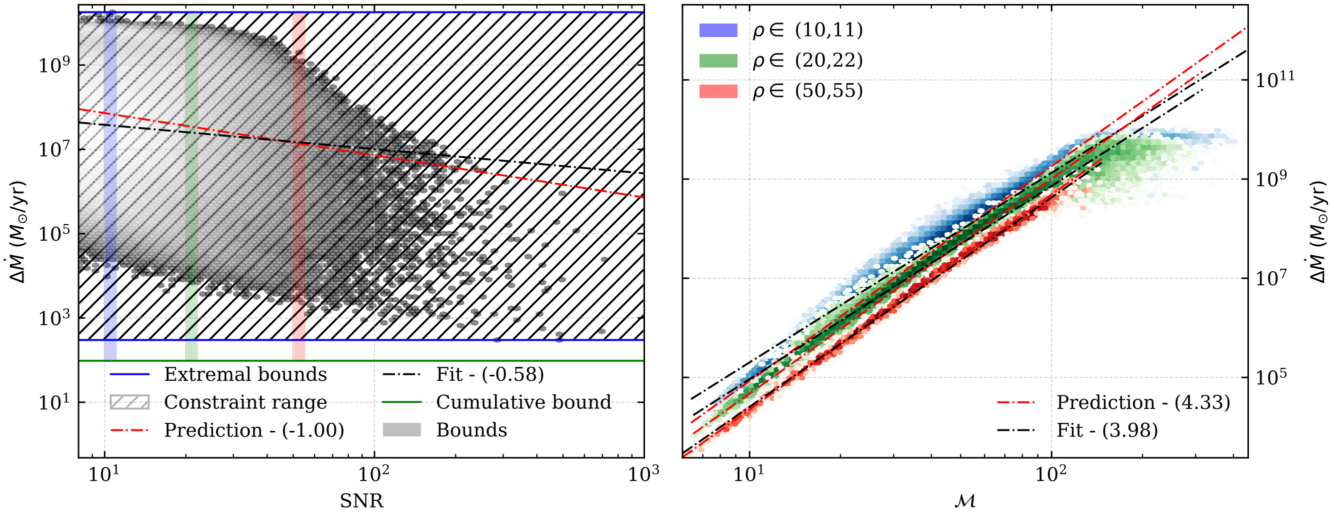

FIG. 23. Result of the scaling analysis outlined in Sec. VI B 6 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on $\dot{M}$ versus the SNR of the source. The right panel shows the density of the constraint versus the chirp mass, with empirical trends shown in black and predicted trends shown in red. The wide distribution of constraints in this catalog indicate that the benefit of large catalogs is minimal, and the total bound is dominated by a select few, highly favorable observations. The distribution of the sources in the $\Delta\dot{M}$–$\mathcal{M}$ plane is to a very good approximation linear, showing a tight correlation between the two quantities. The $\Delta\dot{M}$-SNR relationship also agrees fairly well with our predictions.

7. *Modified dispersion—Massive graviton*

If the graviton were massive, contrary to what is predicted when considering GR as the classical limit of a quantum theory of gravity, the leading order effect would enter the GW phase at 1PN. The Jacobian of the transformation from the ppE framework to this particular modification is

$$\left(\frac{\partial\beta}{\partial m_g^2}\right)^2 \propto \left(\frac{\mathcal{M}D_0}{1+z}\right)^2, \tag{50}$$

where the quantity D_0 is a new cosmological distance defined in Appendix B. We get modified scaling with the chirp mass, and similarly to the variable-G mapping, this Jacobian causes the constraint to inversely scale with the mass. As a result, this new mass factor will benefit MBHs over SOBHs. Furthermore, we now have strong dependence on the distance to the source, D_0, where constraints from farther sources will be enhanced as compared to those sources closer to Earth (see e.g., [72]). These facts benefit LISA MBH sources, which therefore should provide the best constraints.

This is confirmed in Fig. 24. The MBH sources observed by LISA do indeed perform the best, but only marginally. The effectiveness of stacking is seen to still be quite high for this particular modification, as the three terrestrial scenarios all perform comparably. Furthermore, as this is a positive PN effect, terrestrial networks receive a boost from the generic scaling effects discussed in Sec. VI A 1. Multiband sources perform the worst, as they receive little benefit from early inspiral observation, they typically have low mass, and are located at low redshifts. Ultimately, we can expect to improve on the current bound on m_g by 2–3 orders of magnitude over the next thirty years, and this conclusion is robust under variations of the terrestrial detector schedule. This improvement will be insufficient to rule out a massive graviton as a possible explanation of the late-time acceleration of the Universe: in a cosmological context, the graviton would need a mass of the order of the inverse of the Hubble constant, H_0^{-1}, which is of the order of 10^{-30} eV, much smaller than our predicted final constraints.

To explore these relations deeper, we can apply our approximation from Sec. VI A, giving us the following approximation for the constraint on m_g:

$$\Delta m_g \approx \frac{h}{\pi}\left(\frac{5}{2}\right)^{1/4}\sqrt{\frac{(1+z)}{D_0}\frac{\pi f_{\rm low}}{\rho}}. \tag{51}$$

This approximation has produced a notably different scaling relation than what has been seen previously. Namely, the constraint no longer scales with the chirp mass, as the Jacobian factor has canceled the chirp mass dependence from the generic ppE scaling. While this final form of the constraint does not explicitly benefit MBH systems, generic constraints scale with the chirp mass as $\mathcal{M}$. The removal of this chirp mass dependence benefits MBH sources much more than SOBH sources. Also different from previous constraints, we have strong scaling with the distance to the source. For low redshifts, the

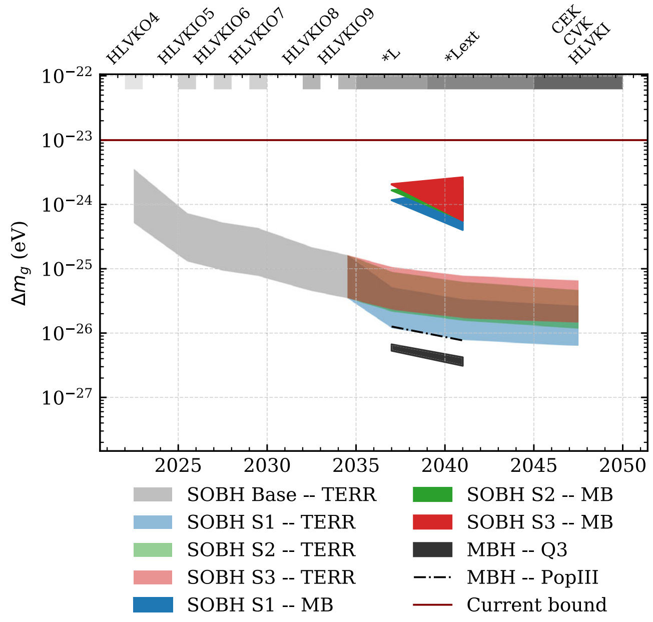

FIG. 24. Projected cumulative constraints on the mass of the graviton, m_g, for the detector networks and population models examined in this paper. Our models show that MBH sources observed by LISA will perform the best at constraining this modification, but only slightly better than the terrestrially observed only sources. Multiband sources perform the worst, as they received no benefits from the Jacobian and already perform only moderately well for positive PN order effects.

distance parameter $D_0 \approx zH_0$ to lowest order in redshift. Extending this expansion to the constraint, the leading-order term should scale as $z^{-1/2}$ for low-redshift sources.

The results related to this approximation are shown in Fig. 25. The left panel shows a heat map of the sources in the catalog created from the HLVKIO8 network and SPOPS 0 population model in the Δm_g-SNR plane, with the solid blue lines denoting the extremal, single observation constraints. The solid green line represents the cumulative bound from this particular catalog. We see good agreement between our predicted scaling for the SNR, after accounting for the Jacobian above. There is a narrow range for the constraints, only spanning one order of magnitude between all sources. This leads to sizeable benefits for large catalogs, also evident from the overlap between the different terrestrial network scenarios.

The right panel shows a heat map of the sources in the Δm_g-redshift plane. We do indeed see a trend in this particular relationship, although the distributions are moderately wide. Our predictions for the scaling relation agrees fairly well with the synthetic data.

C. Effect of precession on the constraints

The differences between the two SOBH population models go beyond the size of the catalogs, which has been our focus so far. An aspect differentiating the SPOPS 0 and SPOPS 265 catalogs, that could have a large impact on our analysis, is the typical magnitude of the in-plane

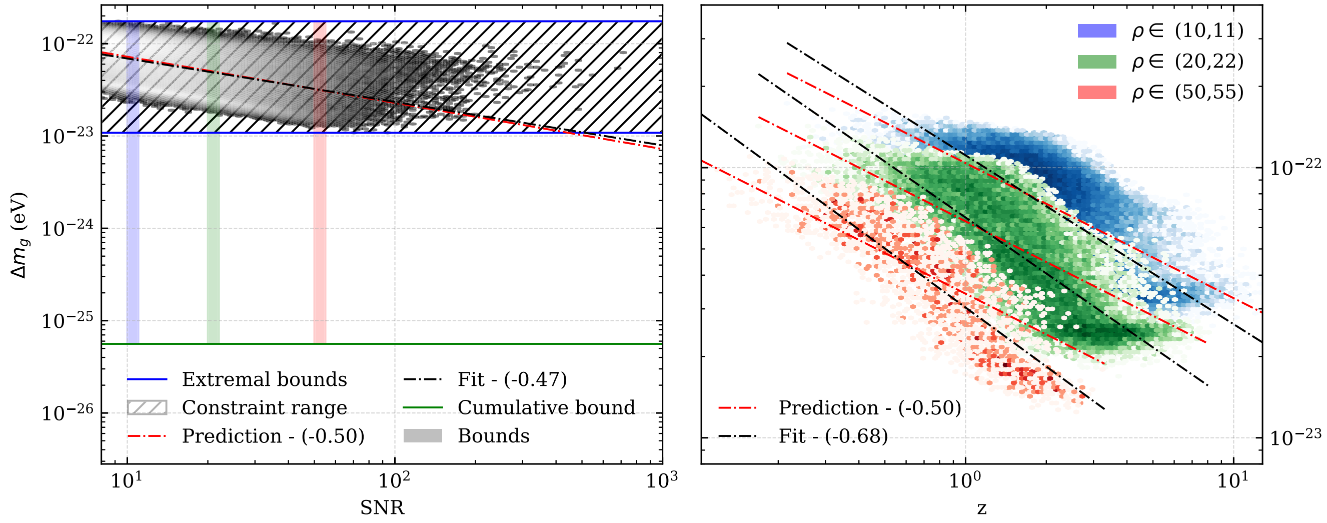

FIG. 25. Result of the scaling analysis outlined in Sec. VI B 7 performed on the data synthesized with the HLVKIO8 network and the SPOPS 0 population. The plotting style is the same as in Fig. 12. The left panel shows a heat map of the constraint on m_g versus the SNR of the source. The right panel shows the density of the constraint versus the redshift z, with empirical trends shown in black and predicted trends shown in red. Because of the narrow range of constraints in the catalog and the large enhancement of the cumulative bound over the strongest single observation, stacking observations is quite efficient for this modification. The right panel shows that there is indeed a trend in the $\Delta m_g - z$ relation (although the distributions are moderately wide) which would favor sources far from Earth, and would primarily benefit MBH sources.

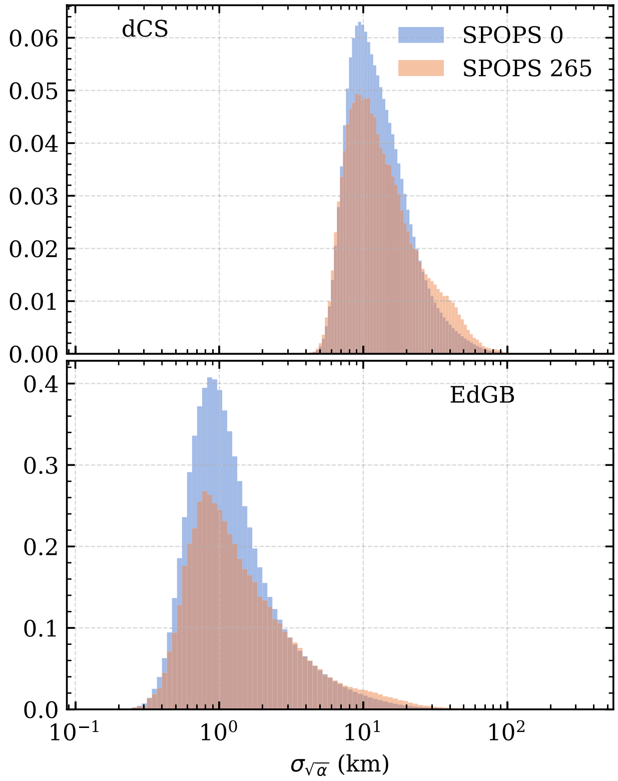

FIG. 26. Distributions of single-source constraints on the GR-modifying parameters $\sqrt{\alpha}_{\rm dCS}$ (top) and $\sqrt{\alpha}_{\rm EdGB}$ (bottom) from the two population models SPOPS 0 (blue) and SPOPS 265 (orange) as detected by the CEK network. The histograms are normalized to provide a comparison of the shapes of the distributions, as opposed to the raw numbers of sources. We see that the distributions only diverge slightly, towards the larger-constraint side of the spectrum. This suggests that the larger precessional effects seen in the SPOPS 265 catalog do not significantly modify the typical constraints attainable by individual sources, or that any effect we may have seen was washed out by the differences in the distributions of other source parameters, such as the total mass and mass ratio. This lack of difference could also be an artifact of our waveform model (IMRPhenomPv2), which is not the most up-to-date waveform available, or of the Fisher approximation, which could be improved upon by a full MCMC analysis.

component of the binary's spins, which is the cause of relativistic precession. The question we now address is whether the stronger constraints coming from the SPOPS 0 catalog over the SPOPS 265 catalog are entirely due to the larger catalog sizes or if the difference in source parameter distributions also impacts the cumulative bounds attainable through GWs.

Previous work has shown that the inclusion of precessional effects can break degeneracies in various source parameters when considering a full MCMC analysis, allowing for significantly tighter constraints on various source properties [102]. To determine if this effect can be seen in our data, in Fig. 26 we show histograms of the individual source constraints on dCS and EdGB, using the two different catalogs (SPOPS 0 and SPOPS 265) and the CEK network. These two theories in particular were chosen because conventional thinking would suggest that they would be the most sensitive to precessional effects, due to the dependence of the ppE parameter on spins.

The figure shows little deviation between the two population models for these theories. The distribution changes slightly on the larger-constraint side of the histogram, but the difference is negligible when considering cumulative constraints. Furthermore, these small deviations in the distributions of constraints cannot be solely attributable to precessional effects, as the parameter distributions shown in Fig. 4 are all modified as well.

To explore the impact of precession on generic modifications in a more controlled environment, we did a direct comparison between systems with zero precession and "maximal" precession (in a sense to be defined shortly), but which are otherwise identical. The results of this analysis are shown in Fig. 27. The methodology we implemented to produce Fig. 27 began with a set grid in the total mass, ranging from 5 $M_\odot$ to 20 $M_\odot$, mass ratio in the range [0.05, 1], and aligned-spin components for each BH ranging from −0.8 to 0.8. With this grid of intrinsic source parameters, we populated the other extrinsic parameters using randomly generated numbers in the conventional ranges. The range on the luminosity distance was chosen such that the SNRs would range from ∼20 to 150. Once a set of full parameter vectors had been created, we calculated one set of Fishers for a fixed detector network with the in-plane component of the spin, χ_p, set to 0. Then, without changing any other parameters, the in-plane spin component was increased to $\chi_p = \sqrt{1-\chi_1^2}$, which is approximately the maximal spin one can achieve while still maintaining a total spin magnitude less than 1. The top panel shows the mean constraint for both configurations as a solid line, with the 1σ interval of the distribution of constraints shown as the shaded region. In the bottom panel we compare the constraints from each configuration (precessing and nonprecessing) for each individual source. The mean of this ratio is then plotted as a solid line, and the 1σ region is shown as the shaded region.

The conclusion from Fig. 27 is that precession seems to have a moderate influence, but one that could be easily washed out by other physical effects. In the most favorable scenario where the binary is maximally precessing, our analysis suggests an improvement of at most a factor of ∼2. Given previous work (see e.g., [102]), one may expect more significant improvements when considering even mild precession. While we do predict improvements from the use of precessing templates, our more restrained conclusions could be the result of two facets of our analysis. Our use of a more rudimentary statistical model, the Fisher

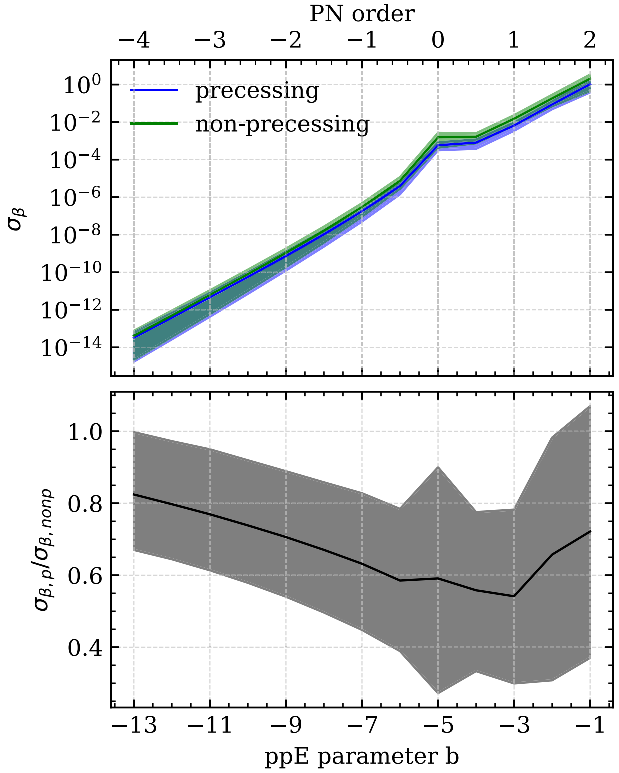

FIG. 27. To create the data involved in this figure, we have created a set grid in parameter space with total mass ranging from 5 $M_\odot$ to 20 $M_\odot$, mass ratio in the range 0.05 to 1, and aligned-spin components for each binary ranging from −0.8 to 0.8. The rest of the parameters were populated with random numbers in the usual ranges, and the luminosity distance was set such that the typical SNRs ranged from ∼20–150. We computed Fisher matrices for each set of parameters, with the in-plane component of the spin set to 0, and then we recomputed them setting the in-plane component of the spin $\chi_p = \sqrt{1-\chi_1^2}$, so that the binary is approximately "maximally" precessing. The top panel shows the distribution of the bounds for the two binary subsets—precessing (blue) and nonprecessing (green)—as a function of PN order. The solid line denotes the average of the synthetic catalog, while the shaded region denotes the 1σ interval. The lower panel shows the ratio $\sigma_{\beta,\mathrm{p}}/\sigma_{\beta,\mathrm{nonp}}$. Each ratio is calculated for a single parameter set, and the mean of these ratios is shown as a solid black line, with the 1σ spread shown by the shading. Even in this more extreme comparison, the improvement in constraint as the result of larger precession effects only amounts to a factor of ∼2. However, more drastic difference may be possible if we performed a full MCMC analysis, or if we used different waveform models.

matrix, does not capture all the more nuanced artifacts in the posterior surface, like a full MCMC analysis would. Furthermore, we here use the IMRPhenomPv2 waveform, which is in some ways more limited in modeling precession with respect to the waveforms used in Ref. [102]. Future studies of precession could focus on these two areas in particular.

VII. CONCLUSIONS

In this work, we have constructed forecasts of what constraints can be placed on a variety of modifications to GR, both generic and theory-specific, using astrophysical models and the most current projections for detector development over the next thirty years. Our analysis spans several topics of interest to the GW community concerned with tests of GR.

We investigate what fundamental physics can be done with a variety of source populations (heavy-seed MBHs, light-seed MBHs, terrestrially observed SOBHs, and multiband SOBHs) and plans for detector development. All of these aspects are connected to what fundamental science is achievable. Ours is the first robust study of this breadth and scope that is capable of quantifying the effects of detector development choices and astrophysical uncertainties.

We identify trends and scaling relationships of constraints for individual GW observations, studying how they evolve with PN order and how they depend on the target source class (MBHs, terrestrially observed SOBHs, and multiband SOBHs). We also quantify the effect of combining constraints from a full, synthetic catalog, appropriately informed by robust population models. We find that the effectiveness of stacking observations is a PN-dependent conclusion. The techniques developed here have important implications for the future of GW-based tests of GR, especially in the era of 3g detectors. The two components of our analysis (individual scaling and studies of the stacking of multiple observations) combine to create a full picture of some of the most important aspects involved in testing GR with GWs. We hope that this information will be valuable in driving design choices for future detector development.

We map our generic constraints to theory-specific constraints, where we analyze specific parameters in viable, interesting theories. Repeating some of the scaling analysis done in previous sections leads, in some cases, to a reversal of the conclusions drawn for generic modifications. This reinforces the need to incorporate theory-specific waveforms in future analyses, when available.

This work opens up several new avenues of research. We focused on BBH systems, neglecting future contributions from neutron star-neutron star and neutron star-BH binaries. These binaries have much longer inspiral signals relative to typical BH mergers observed by the LVC, and they could provide crucial information concerning early inspiral, negative PN effects. Beyond the signal length, neutron stars are sometimes treated on unequal footing in the context of specific theories, such as scalar-tensor gravity, EdGB and dCS. This could provide other insights into specific theories that do not affect BBH mergers.

Because of the scale of the catalogs involved we used simple Fisher matrix forecasts, running $\sim 10^8$ Fisher matrix calculations. A more thorough analysis using MCMC, or other more robust data analysis techniques, could provide more information about some of the trends we have identified. An MCMC population study on the scale of this work is currently intractable, but even an analysis of a subset of sources could be enlightening.

Our work has focused on estimating only a single PN modification at a time, but any modified theory of gravity will correct the waveform at all orders in a PN expansion. Recent work studied how constraints are affected when one attempts to simultaneously constrain ppE deformations that enter at multiple orders [103]. Here we have chosen to limit ourselves to a single parameter at a time for the following reasons. While allowing for multiple parameters to vary in a completely independent way at several PN orders is a more robust and general framework, this treatment is probably overly pessimistic. Past work [30,104] showed that, indeed, varying multiple generic parameters simultaneously drastically lowers our ability to constrain them. However, in the context of a given, physically motivated theory there should be some relation between the different ppE modifications. Any PN expansion should converge in the appropriate domains, ensuring a hierarchy on the size of the modifications. Moreover, the modification at each PN order should at least depend on the coupling parameters of the theory, ensuring that no two PN orders are totally independent from each other. These criteria suggest that the overall bound on a given modification, in the context of a physically motivated theory, should not be significantly weakened by the inclusion of higher-order corrections (except in the most unfortunate of fine-tuning scenarios). Therefore our conclusions should be robust under the inclusion of higher-order PN corrections to the waveform.

Our investigation of the effects of precession on modified GR constraints could be improved in at least three ways. While we did include a full inspiral/merger/ringdown model of precession by implementing IMRPhenomPv2 [25–27], more recent and complex waveform models (such as IMRPhenomPv3 [105], IMRPhenomXPHM [106] or SEOBNRv4PHM [107]) could encode more information in the signal, helping to break degeneracies. A more robust statistical analysis, such as a full MCMC, could explore the posterior space more thoroughly, shedding light on the effects of precession. Last but not least, the astrophysical SOBH models considered here only allow for isolated field formation under restrictive assumptions. Dynamical formation generally predicts a larger fraction of precessing systems [108], and it is important to consider other pathways for producing BBHs with large misaligned spins even within the isolated formation channel [73,109].

ACKNOWLEDGMENTS

We thank Vishal Baibhav, Davide Gerosa, Gabriela González, Bangalore Sathyaprakash, Sashwat Tanay, and Kaze Wong for many useful discussions on various aspects of this work. N. Y. acknowledges support from NSF Grants No. PHY-1759615, No. PHY-1949838, and NASA ATP Grant No. 17-ATP17-0225. E. B. is supported by NSF Grants No. PHY-1912550 and No. AST-2006538, NASA ATP Grants No. 17-ATP17-0225 and No. 19-ATP19-0051, and NSF-XSEDE Grant No. PHY-090003. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 690904. The authors would like to acknowledge networking support by the GWverse COST Action CA16104, "Black holes, gravitational waves and fundamental physics." This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign. It also used computational resources at the Maryland Advanced Research Computing Center (MARCC). The following software libraries were used at various stages in the analysis for this work, in addition to the packages explicitly mentioned above: GSL [110], NumPy, SciPy, FILLTEX [111].

APPENDIX A: BAYESIAN THEORY AND FISHER ANALYSIS DETAILS

Signal analysis in GW science is usually based on Bayes' theorem,

$$p(\vec{\theta}, d) = \frac{p(d, \vec{\theta}) p(\vec{\theta})}{p(d)}, \tag{A1}$$

where $p(\vec{\theta}, d)$ is the posterior probability of the vector of parameters $\vec{\theta}$ given some data set d. The quantity $p(\vec{\theta})$ is the prior information about the source parameters, reflecting any initial beliefs held before the data were taken. The evidence, $p(d)$, is the normalization of the posterior, which also generally holds valuable information about the signal, but will not be the focus of this work. The quantity $p(d, \vec{\theta})$ is the likelihood of the data, and describes the probability that one would see a data set d given some set of parameters $\vec{\theta}$. For GW data analysis, this is given by

$$p(d, \vec{\theta}) \propto \exp\left[-\frac{1}{2} \sum_{i}^{N_{\text{detector}}} (d_i - h_i | d_i - h_i)\right], \tag{A2}$$

for each data series d_i and detector response template h_i from the ith detector, where the noise-weighted inner product is given by

$$(d-h|d-h) = 4\mathrm{Re}\left[\int \frac{(d-h)(d-h)^*}{S_n(f)} df\right]. \tag{A3}$$

To estimate the posterior using real data from LIGO, one would use a Markov Chain Monte Carlo [112,113] to explore the parameter space of the signal. This would yield a set of independent samples from the posterior that quantifies not only the most likely values for the vector $\vec{\theta}$, but also includes information about our confidence in those estimates. This approach is the most reliable and accurate, but it is too computationally expensive for our purposes. Even the most optimized algorithms would take considerable computational resources to analyze the number of sources examined in this paper. We therefore turn to a commonly used approximation of the posterior to estimate the confidence intervals on $\vec{\theta}$ that is much more computationally tractable: the Fisher information matrix.

We calculate the Fisher matrices for each detector and combine them to construct a total Fisher matrix for each source according to Eq. (22). To properly reflect the ability of a terrestrial network to localize sources in the sky, we incorporate a time delay between detectors that is α- and δ-dependent. That is, for each detector besides the reference detector, we append the following factor to the phase:

$$t_{c,i} \to t_{c,\mathrm{ref}} + \delta t_{c,i}(\alpha,\delta), \tag{A4}$$

where $\delta t_{c,i}$ is defined as

$$\delta t_{c,i}(\alpha,\delta) = \frac{\mathbf{x}_{\mathrm{ref}} \cdot \hat{\mathbf{x}}_{\mathrm{source}}(\alpha,\delta) - \mathbf{x}_i \cdot \hat{\mathbf{x}}_{\mathrm{source}}(\alpha,\delta)}{c}. \tag{A5}$$

The detector positions $\mathbf{x}_{\mathrm{ref}}$ and $\mathbf{x}_i$ are in Earth-centered coordinates, the unit vector $\hat{\mathbf{x}}_{\mathrm{source}}$ points to the source in the sky in the same coordinates, and we have reintroduced the speed of light c for clarity. The positions of the detectors in these Earth-centered coordinates were taken from LALSuite [70]. This procedure is neglected when considering LISA, as sky localization comes from the orbital motion of the satellites and long signal durations for space-based detectors.

An additional concern in the context of utilizing Fisher matrices with consistent parameters is the description of the binary's orientation. There are three coordinate systems that naturally arise in the description of terrestrial and space detectors. The natural coordinate system to use for LISA is the ecliptic coordinate system, specifically the parameters θ_j and ϕ_j, as these are the quantities that show up in LISA's response function. For terrestrial detectors, the polarization angle $\bar{\psi}$ and the inclination angle ι naturally arise in the response function, where the polarization angle is naturally defined in the equatorial coordinate system. Finally, the source properties themselves are stipulated in the source frame, aligned with the orbital angular momentum $\mathbf{L}$ and subsequently used to calculate the waveform. Any choice is valid as long as it is consistently enforced, so we chose to use the equatorial coordinates, and we accounted for the coordinate transformation in the calculation of the derivative of the response function. An equally simple solution would be to compute the Fisher matrices in their respective, natural coordinates, then use the Jacobian matrix to transform them as follows:

$$\Gamma_{i'j'} = \frac{\partial x^i}{\partial x^{i'}} \Gamma_{ij} \frac{\partial x^j}{\partial x^{j'}}, \tag{A6}$$

which is exactly how we transform our bounds on generic modifications to theory-specific modifications.

The actual transformation relies on the construction of an explicit rotation matrix between the different frames of reference. Transforming between ecliptic and equatorial coordinates is a trivial rotation by a constant angle, so we will instead just describe the transformation between the source frame and the equatorial system.

The first frame in question is the equatorial frame, which is the frame that defines the parameters θ_{L}, ϕ_{L}, α, and δ. From these quantities, one can construct two vectors: the direction of propagation $\hat{\mathbf{N}}$ (which points from the solar system to the source), and the direction of the orbital angular momentum $\hat{\mathbf{L}}$ at some reference frequency. These two vectors also define the inclination angle of the orbital angular momentum,

$$\cos\iota = -\hat{\mathbf{L}} \cdot \hat{\mathbf{N}}, \tag{A7}$$

which will be needed in the next frame.

The second frame is the source frame, in which the waveform is naturally constructed. This frame is defined by a coordinate system with $\hat{\mathbf{L}} = \hat{\mathbf{z}}$, while the other two Cartesian axes are chosen such that the direction of propagation $-\hat{\mathbf{N}}$ (where $\hat{\mathbf{N}}$ points from the solar system to the source) lies in the $x-z$ plane when the reference phase $\phi_{\mathrm{ref}} = 0$. The vector $\hat{\mathbf{N}}$ is then rotated azimuthally by an angle ϕ_{ref} for nonzero reference phases. The angle between $\hat{\mathbf{L}}$ and $\hat{\mathbf{N}}$ in the source frame is just the inclination defined in Eq. (A7), which fully specifies this vector in the second frame. Using these two vectors, we can construct a third, orthogonal vector as the cross product of these two, which we will call $\hat{\mathbf{K}} = \hat{\mathbf{L}} \times \hat{\mathbf{N}}$.

With three vectors in each frame, we can construct an explicit rotation matrix to transform any quantities from one frame to the other by the set of equations,

$$\begin{aligned}\hat{\mathbf{L}}_{\mathrm{eq}} &= \mathbf{R} \cdot \hat{\mathbf{L}}_{\mathrm{SF}},\\ \hat{\mathbf{N}}_{\mathrm{eq}} &= \mathbf{R} \cdot \hat{\mathbf{N}}_{\mathrm{SF}},\\ \hat{\mathbf{K}}_{\mathrm{eq}} &= \mathbf{R} \cdot \hat{\mathbf{K}}_{\mathrm{SF}},\end{aligned} \tag{A8}$$

where $\mathbf{R}$ is the unspecified rotation matrix and the subscripts "eq" and "SF" correspond to equatorial coordinates and source-frame coordinates, respectively. The system (A8) can be inverted analytically, resulting in analytical expressions for the rotation matrix $\mathbf{R}$.

This rotation matrix allows us to transform any quantity between the two frames. This can be used to calculate the ecliptic angles of the total angular momentum $\hat{\mathbf{J}}$, which is needed for the LISA response function. The vector $\hat{\mathbf{J}}$ is easily constructed in the source frame, as the spins are defined in this frame and the orbital angular momentum already defines the coordinate system. The vector is simply rotated into the equatorial frame and subsequently into the ecliptic frame, to compute the LISA response function.

We also need to specify the polarization angle for the terrestrial network. We simply use the relation [114],

$$\tan\bar{\psi} = \frac{\hat{\mathbf{J}}\cdot\hat{\mathbf{z}} - (\hat{\mathbf{J}}\cdot\hat{\mathbf{N}})(\hat{\mathbf{z}}\cdot\hat{\mathbf{N}})}{\hat{\mathbf{N}}\cdot(\hat{\mathbf{J}}\times\hat{\mathbf{z}})}, \tag{A9}$$

where $\hat{\mathbf{z}}$ is the unit vector of the equatorial coordinate system aligned with the axis of rotation of the Earth, defining a globally consistent polarization angle. These transformations allow us to use the vector of parameters outlined above, where all the quantities are consistently defined.

Once a combined Fisher for each source is calculated, the inversion of each Fisher results in the individual covariance matrices, which effectively acts as marginalization. We extract the variance of the ppE parameter β by taking the diagonal element $\sigma_{\beta\beta}$, which gives us a marginalized posterior on β for a single source. Finally, to combine the sources, we multiply the marginalized posteriors together (because each source is completely independent), which for a series of Gaussians becomes

$$\begin{aligned} p(\beta|\vec{\theta}) &\propto \prod_i^N \exp\left(-\frac{1}{2}\frac{\beta^2}{\sigma_{\beta,i}^2}\right) \\ &\propto \exp\left(-\frac{1}{2}\beta^2\sum_i^N\frac{1}{\sigma_{\beta,i}^2}\right). \end{aligned} \tag{A10}$$

Therefore, our resulting bound on β is simply given by

$$\sigma_\beta^2 = \left(\sum_i^N \frac{1}{\sigma_{\beta,i}^2}\right)^{-1}. \tag{A11}$$

APPENDIX B: MAPPING TO SPECIFIC THEORIES

The main goal of this Appendix is to map parametrized deviations, that do not necessarily have a physical interpretation, to specific parameters appearing in beyond-GR theories.

1. Dipole radiation

In GR, the generation of GWs is sourced from the second time derivative of the mass quadrupole moment, resulting in quadrupolar radiation. This connection to the quadrupole moment is tied to the conservation of the stress energy tensor, rooted in the Bianchi identities (a purely geometrical constraint). If additional fields were added to the gravitational sector that were not subject to such energy conditions, one would generically expect dipolar radiation, providing an additional avenue of energy loss for the system. An additional channel for outgoing power would drive the binary to inspiral faster than what would be predicted by GR, and this faster inspiral would produce a measurable effect on the waveform.

To determine this effect on the waveform, we can write the time derivative of the gravitational binding energy of the system as [45]

$$\dot{E} = \dot{E}_{\mathrm{GR}} + \delta\dot{E}, \tag{B1}$$

where $\dot{E}_{\mathrm{GR}}$ is the GW power output in GR, and $\delta\dot{E}$ is our generic deviation. In terms of these parameters, our modification to the waveform becomes (in the language of ppE parameters) [45]

$$\beta_{\mathrm{dipole}} = \frac{-3}{224}\eta^{2/5}\delta\dot{E}, \tag{B2}$$

where $\eta = m_1 m_2/(m_1+m_2)^2$ is the symmetric mass ratio of the binary system.

Of course, $\delta\dot{E}$ is written generically in Eq. (B1). Once a specific theory has been selected, this term will be a function of the source parameters and of any fundamental constants of the theory in question. For example, in Einstein-dilaton Gauss-Bonnet gravity (EdGB) [115] the waveform modification can be calculated to be [47]

$$\beta_{\mathrm{EdGB}} = -\frac{5}{7168}\frac{\zeta_{\mathrm{EdGB}}}{\eta^{18/5}}\frac{(m_1^2 s_2^{\mathrm{EdGB}} - m_2^2 s_1^{\mathrm{EdGB}})^2}{m^4}, \tag{B3}$$

$$s_i^{\mathrm{EdGB}} = \frac{2[(1-\chi_i^2)^{1/2} - 1 + \chi_i^2]}{\chi_i^2}, \tag{B4}$$

where ζ_{EdGB} is related to the coupling parameter of the theory α_{EdGB} by $\zeta_{\mathrm{EdGB}} = 16\pi\alpha_{\mathrm{EdGB}}^2(1+z)^4/m^4$, and $m = m_1 + m_2$ is the total redshifted mass of the system. The quantities s_i^{EdGB} given in Eq. (B4) are the sensitivities of the BHs, and χ_i are the dimensionless, (anti)aligned spin components of the ith BH.

Because of the approximations used to derive Eq. (B3), this particular formula is only valid when $\sqrt{\alpha_{\mathrm{EdGB}}} \leq m_s/2$, where m_s is the smallest length scale of the system (see e.g., [116]). For this work, the smallest length scale will be the mass of the smaller BH, m_2.

2. Black hole evaporation

High-energy theories that might be candidates for quantum theories of gravity often involve the embedding of our four-dimensional spacetime in a higher-dimensional space, where the extra dimensions are often compactified. For example, Arkani-Hamed, Dimopoulos, and Dvali proposed a model which had implications for the hierarchy problem between the electroweak and Planck scale [117,118]. Another set of models proposed by Randall and Sundrum (RS-I/II) [119,120] postulate a braneworld model where the four-dimensional brane we occupy resides in a five-dimensional anti–de Sitter bulk spacetime. In RS-II, BHs were initially predicted to evaporate much faster as compared with analogous situations in four dimensions, with an evaporation rate given by [121,122]

$$\frac{dm}{dt} = -2.8 \times 10^{-7} \left(\frac{1\, M_{\odot}(1+z)}{m}\right)^2 \left(\frac{l}{10\ \mu m}\right)^2 \frac{M_{\odot}}{\text{yr}}, \tag{B5}$$

where l is the length scale of the extra dimension and m is the detected mass. However, more recent work has shown that black holes in RS-II are actually stable and evaporation does not occur [123,124].

Regardless of the physical origin of the evaporation, it is still interesting to consider its effect on the gravitational waveform. Let us imagine that either the volume or the area of a BH changes with time due to some quantum or classical extension of GR. The volume and the area are common geometric quantities associated with a BH, so it is plausible that if BH solutions become time-dependent, then it is these quantities that acquire the time dependence. Assume then that $dV/dt = c_V \ell^2$ or $dA/dt = c_A \ell$, where $c_{V,A}$ are dimensionless constants and ℓ is a new length that controls the scale at which time dependence kicks in. If so, using that $V = (32\pi/3)m^3$ and $A = 16\pi m^2$ for a Schwarzschild BH, we then have that $dm/dt = [c_V/(32\pi)](\ell/m)^2$ or $dm/dt = [c_A/(32\pi)](\ell/m)$. On general grounds, then, one would expect $dm/dt \sim (\ell/m)^q$ with $q = 1$ of $q = 2$, depending on whether the time dependence acts on the area or the volume of the BH, and a constraint on dm/dt would then imply a constraint on the evaporation scale ℓ.

Regardless of the process that leads to evaporation, the waveform modification has the form [125],

$$\beta_{\text{BHE}} = \frac{25}{851968} \dot{M} \left(\frac{3 - 26\eta + 34\eta^2}{\eta^{2/5}(1-2\eta)}\right), \tag{B6}$$

where $\dot{M} = dM/dt = dm_1/dt + dm_2/dt$ is the anomalous evaporation rate.

3. Local position invariance violation

In the case where Newton's gravitational constant is promoted to a time-dependent quantity, conspicuous additional accelerations could be experienced by binaries inspiraling together. This phenomenon could come about, for example, because the gravitational constant is tied to a background scalar field which evolves on cosmological timescales. This effect can be observed as alterations to the binding energy of the binary, and it has a mapping to the ppE framework [101],

$$\beta_{\dot{G}} = \frac{-25}{65526} \frac{\dot{G}\mathcal{M}}{(1+z)G}, \tag{B7}$$

where $\dot{G} = dG/dt$ is the time derivative of the gravitational constant and $\mathcal{M}$ is the redshifted chirp mass.

4. Parity violation

Many attempts to unify quantum mechanics and gravity involve terms quadratic in curvature at the level of the action in the low-energy limit, as well as additional fields coupled to these higher-order terms. The strength of this coupling is determined by the coupling parameter of the theory, and therefore determines the magnitude of the effect on the waveform. EdGB (discussed above) is an example of this type of modification where the modifying parameter comes at a negative PN order because of dipolar radiation. EdGB, however, preserves parity because the term added to the action is parity-even, introducing a scalar field that is also parity-even. A quadratic theory that does not preserve parity is dynamical Chern-Simons (dCS) gravity [115], which incorporates an additional quadratic curvature term into the action that is parity-odd. In order to keep the action invariant under parity transformations, this odd-parity term must be coupled to an odd-parity scalar field, leading to a variety of implications in different gravitational interactions [115].

This modification affects the waveform as follows [47]:

$$\beta_{\text{dCS}} = -\frac{5}{8192} \frac{\zeta_{\text{dcs}}}{\eta^{14/5}} \frac{(m_1 s_2^{\text{dCS}} - m_2 s_1^{\text{dCS}})^2}{m^2} + \frac{15075}{114688} \frac{\zeta_{\text{dCS}}}{\eta^{14/5}} \frac{(m_2^2\chi_1^2 - \frac{350}{201} m_1 m_2 \chi_1 \chi_2 + m_1^2 \chi_2^2)}{m^2}, \tag{B8}$$

$$s_i^{\text{dCS}} \equiv \frac{2 + 2\chi_i^4 - 2(1-\chi_i^2)^{1/2} - \chi_i^2[3 - 2(1-\chi_i^2)^{1/2}]}{2\chi_i^3}, \tag{B9}$$

where ζ_{dCS} is related to the coupling parameter by $\zeta_{\text{dCS}} = 16\pi\alpha_{\text{dCS}}^2(1+z)^4/m^4$. The quantity s_i^{dCS} given in Eq. (B9) is the sensitivity of the ith BH in dCS.

As the result of the approximations involved in the derivation of the flux, Eq. (B8) is only valid if $\sqrt{\alpha_{\mathrm{dCS}}} \leq m_s/2$, where m_s is the smallest length scale of the system, just as in EdGB. Here we are interested in BBHs, and m_s is the mass of the smaller BH.

5. Lorentz violation

Noncommutative gravity promotes the coordinates in GR to operators with a nontrivial commutation relation defined by $[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu}$, where $\theta^{\mu\nu}$ is a real, constant antisymmetric tensor [59,126]. This tensor plays a role analogous to the role of Planck's constant in quantum mechanics, and it defines a length scale at which there is a fundamental uncertainty between physical parameters.

Defining the quantity $\Lambda^2 = \theta^{0i}\theta_{0i}/(l_p t_p)^2$, where l_p and t_p are the Planck length and time, respectively, one can derive the modification to the waveform as [59,126]

$$\beta_{\mathrm{NC}} = -\frac{75}{256}\eta^{-4/5}(2\eta - 1)\Lambda^2. \tag{B10}$$

In this parameterization, $\sqrt{\Lambda}$ defines the energy scale of noncommutativity, relative to the Planck scale.

6. Modified dispersion

Another assumption made by GR is that gravitons are massless. If this is not assumed, the leading-order correction to the measured GW signal would come about through the propagation of GW [44,127]. The graviton would be ascribed a massive-particle dispersion relation $E^2 = p^2 + m_g^2$, where E is the graviton energy, p is the graviton momentum, and m_g is the graviton mass. With a nonlinear relation between energy and momentum, one would expect that the group velocity would become frequency-dependent. This introduces an additional term in the GW phase [127],

$$\beta_{\mathrm{MG}} = \pi^2 \frac{D_0}{1+z}\frac{\mathcal{M}_z}{\lambda_{\mathrm{MG}}^2}, \tag{B11}$$

$$D_0 \equiv (1+z)\int_0^z \frac{1}{H(z')}\frac{dz'}{(1+z')^2}, \tag{B12}$$

where D_0 is a new cosmological distance similar to the luminosity distance, and λ_g is the Compton wavelength of the graviton, related to the mass by $\lambda_g = h/m_g$. To evaluate the Hubble parameter $H(z)$ we use the cosmological parameters inferred from the Planck Collaboration [60] and software from the Astropy Python package [128,129].

APPENDIX C: INSPIRAL/MERGER/RINGDOWN VS INSPIRAL WAVEFORMS

Concerning the deviations away from GR that we have injected into the waveforms, we examine two families of modifications: those that affect GW propagation and those that modify GW generation. The difference between these two mechanisms arises from our lack of knowledge about the dynamics of BBHs close to merger in modified theories of gravity. To reflect this ignorance, we include the modification due to generation effects in the inspiral portion of the waveform only. Propagation effects are under no such shroud as the mechanism responsible acts in the low-curvature regions between galaxies and should equally affect the waveform across the entire frequency range. We therefore include modifications due to propagation effects in the entire waveform. As we are only ever looking at one effect at a time, these two families of effects are never examined concurrently. To incorporate these modifications, we utilize the ppE methodology [28–31]. In the case of precessing systems, the modifications are treated slightly differently. For generational effects, we append a phase modification to the waveform in the coprecessing frame, where the physics of GW generation are approximately the same as those for a nonprecessing binary. The waveform is then "twisted-up" in the usual fashion for IMRPhenomPv2 waveforms, but with the modified coprecessing waveform. For propagation effects, we append the modification to the waveform at all frequencies, after the waveform has been transformed to the inertial frame. In equations,

$$\tilde{h}_{\mathrm{coprec,gen}} = \begin{cases} \tilde{h}_{\mathrm{coprec,GR}} e^{i\beta(\mathcal{M}\pi f)^{-b/3}} & f < 0.018m \\ \tilde{h}_{\mathrm{coprec,GR}} & 0.018m < f \end{cases} \tag{C1}$$

$$\tilde{h}_{\mathrm{inertial,prop}} = \tilde{h}_{\mathrm{inertial,GR}} e^{i\beta(\mathcal{M}\pi f)^{-b/3}}. \tag{C2}$$

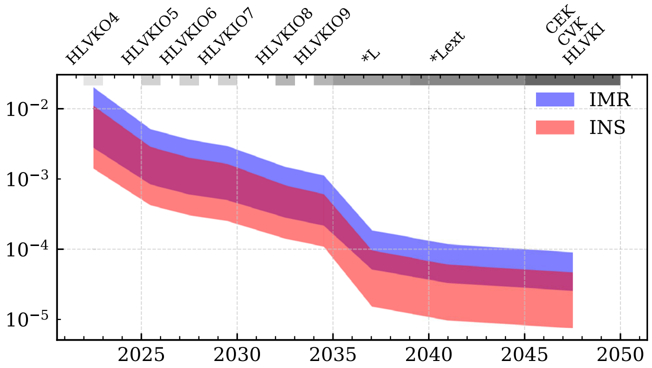

FIG. 28. Comparison between the constraints on β at 1.5PN predicted by using the generation modification (INS), as opposed to the propagation modification (IMR). We used the same catalogs and networks (SOBH base and SOBH S1) in both cases. The difference is negligible when considering the order-of-magnitude constraints of interest in this work.

A comparison between the two methods is shown in Fig. 28, which illustrates that the difference is small. Because of this, we used the full inspiral-merger-ringdown modification in all of this paper.

[1] C. M. Will, Living Rev. Relativity **17**, 4 (2014).
[2] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), Astrophys. J. Lett. **875**, L1 (2019).
[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **100**, 104036 (2019).
[4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 221101 (2016); **121**, 129902 (E) (2018).
[5] E. Berti *et al.*, Classical Quantum Gravity **32**, 243001 (2015).
[6] C. Charmousis, E. J. Copeland, A. Padilla, and P. M. Saffin, Phys. Rev. Lett. **108**, 051101 (2012).
[7] C. Charmousis, E. J. Copeland, A. Padilla, and P. M. Saffin, Phys. Rev. D **85**, 104040 (2012).
[8] C. de Rham, G. Gabadadze, L. Heisenberg, and D. Pirtskhalava, Phys. Rev. D **83**, 103516 (2011).
[9] G. D'Amico, C. de Rham, S. Dubovsky, G. Gabadadze, D. Pirtskhalava, and A. J. Tolley, Phys. Rev. D **84**, 124046 (2011).
[10] C. de Rham, G. Gabadadze, and A. J. Tolley, Phys. Rev. Lett. **106**, 231101 (2011).
[11] A. G. Riess *et al.* (Supernova Search Team Collaboration), Astron. J. **116**, 1009 (1998).
[12] S. Perlmutter *et al.* (Supernova Cosmology Project Collaboration), Astrophys. J. **517**, 565 (1999).
[13] J. Polchinski, *String Theory. Vol. 1: An Introduction to the Bosonic String*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2007).
[14] J. Polchinski, *String Theory. Vol. 2: Superstring Theory and Beyond*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2007).
[15] Y. Fujii and K. Maeda, *The Scalar-Tensor Theory of Gravitation*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2007).
[16] S. H. Alexander and N. Yunes, Phys. Rev. D **97**, 064033 (2018).
[17] N. Yunes and X. Siemens, Living Rev. Relativity **16**, 9 (2013).
[18] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 115012 (2015).
[19] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[20] T. Akutsu *et al.* (KAGRA Collaboration), arXiv: 2005.05574.
[21] B. Iyer, T. Souradeep, C. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta (LIGO Scientific Collaboration), LIGO-India, proposal of the consortium for Indian Initiative in Gravitational-wave Observations (IndIGO), Tech. Report No. LIGO-M1100296-v2, 2020.
[22] S. Dwyer, D. Sigg, S. W. Ballmer, L. Barsotti, N. Mavalvala, and M. Evans, Phys. Rev. D **91**, 082001 (2015).
[23] M. Punturo *et al.*, Classical Quantum Gravity **27**, 084007 (2010).
[24] P. Amaro-Seoane *et al.*, arXiv:1702.00786.
[25] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).
[26] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[27] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).
[28] N. Yunes and F. Pretorius, Phys. Rev. D **80**, 122003 (2009).
[29] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius, Phys. Rev. D **84**, 062003 (2011).
[30] L. Sampson, N. Cornish, and N. Yunes, Phys. Rev. D **87**, 102001 (2013).
[31] K. Chatziioannou, N. Yunes, and N. Cornish, Phys. Rev. D **86**, 022004 (2012); **95**, 129901(E) (2017).
[32] S. Tahura, K. Yagi, and Z. Carson, Phys. Rev. D **100**, 104001 (2019).
[33] D. Gerosa, S. Ma, K. W. K. Wong, E. Berti, R. O'Shaughnessy, Y. Chen, and K. Belczynski, Phys. Rev. D **99**, 103004 (2019).
[34] C. J. Moore, D. Gerosa, and A. Klein, Mon. Not. R. Astron. Soc. **488**, L94 (2019).
[35] E. Barausse, N. Yunes, and K. Chamberlain, Phys. Rev. Lett. **116**, 241104 (2016).
[36] C. Cutler *et al.*, arXiv:1903.04069.
[37] G. Gnocchi, A. Maselli, T. Abdelsalhin, N. Giacobbo, and M. Mapelli, Phys. Rev. D **100**, 064024 (2019).
[38] Z. Carson and K. Yagi, Classical Quantum Gravity **37**, 02LT01 (2020).
[39] Z. Carson and K. Yagi, Phys. Rev. D **101**, 044047 (2020).
[40] A. Toubiana, S. Marsat, S. Babak, E. Barausse, and J. Baker, Phys. Rev. D **101**, 104038 (2020).
[41] C. Liu, L. Shao, J. Zhao, and Y. Gao, Mon. Not. R. Astron. Soc. **496**, 182 (2020).
[42] R. Nair and N. Yunes, Phys. Rev. D **101**, 104011 (2020).
[43] A. J. K. Chua and M. Vallisneri, arXiv:2006.08918.
[44] N. Yunes, K. Yagi, and F. Pretorius, Phys. Rev. D **94**, 084002 (2016).
[45] K. Chamberlain and N. Yunes, Phys. Rev. D **96**, 084039 (2017).
[46] K. Yagi, Phys. Rev. D **86**, 081504 (2012).
[47] R. Nair, S. Perkins, H. O. Silva, and N. Yunes, Phys. Rev. Lett. **123**, 191101 (2019).
[48] C. Bambi, M. Giannotti, and F. L. Villante, Phys. Rev. D **71**, 123524 (2005).
[49] C. J. Copi, A. N. Davis, and L. M. Krauss, Phys. Rev. Lett. **92**, 171301 (2004).
[50] R. N. Manchester, Int. J. Mod. Phys. D **24**, 1530018 (2015).
[51] A. S. Konopliv, S. W. Asmar, W. M. Folkner, Ö. Karatekin, D. C. Nunes, S. E. Smrekar, C. F. Yoder, and M. T. Zuber, Icarus **211**, 401 (2011).
[52] F. Hofmann, J. Müller, and L. Biskupek, Astron. Astrophys. **522**, L5 (2010).
[53] M. G. Hare, Can. J. Phys. **51**, 431 (1973).
[54] A. S. Goldhaber and M. M. Nieto, Phys. Rev. D **9**, 1119 (1974).
[55] C. Talmadge, J. P. Berthias, R. W. Hellings, and E. M. Standish, Phys. Rev. Lett. **61**, 1159 (1988).
[56] S. Desai, Phys. Lett. B **778**, 325 (2018).
[57] R. Brito, V. Cardoso, and P. Pani, Phys. Rev. D **88**, 023514 (2013).
[58] H. O. Silva, A. M. Holgado, A. Cárdenas-Avendaño, and N. Yunes, arXiv:2004.01253.

[59] A. Kobakhidze, C. Lagger, and A. Manning, Phys. Rev. D **94**, 064033 (2016).

[60] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).

[61] J. Baker *et al.*, arXiv:1907.06482.

[62] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), Living Rev. Relativity **21**, 3 (2018).

[63] B. O'Reilly, M. Branchesi, S. Haino, and G. Gemme, LIGO Document, Technical Report No. T2000012-v1 (2020), https://dcc.ligo.org/LIGO-T2000012/public.

[64] D. McClelland, M. Cavaglia, M. Evans, R. Schnabel, B. Lantz, V. Quetschke, and M. Iain (LIGO Scientific Collaboration), The LSC-Virgo white paper on instrument science (2016–2017 edition), Tech. Report No. LIGO-T1600119-v4, 2020.

[65] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, 141 (2019).

[66] Cosmic Explorer, https://cosmicexplorer.org/researchers.html (2020).

[67] S. Hild *et al.*, Classical Quantum Gravity **28**, 094013 (2011).

[68] T. Robson, N. J. Cornish, and C. Liu, Classical Quantum Gravity **36**, 105011 (2019).

[69] S. Tanay, A. Klein, E. Berti, and A. Nishizawa, Phys. Rev. D **100**, 064006 (2019).

[70] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).

[71] C. Cutler, Phys. Rev. D **57**, 7089 (1998).

[72] E. Berti, A. Buonanno, and C. M. Will, Phys. Rev. D **71**, 084025 (2005).

[73] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, Phys. Rev. D **98**, 084036 (2018).

[74] A. Klein *et al.*, Phys. Rev. D **93**, 024003 (2016).

[75] S. Datta, A. Gupta, S. Kastha, K. Arun, and B. Sathyaprakash, Phys. Rev. D **103**, 024036 (2021).

[76] J. R. Gair, I. Mandel, M. C. Miller, and M. Volonteri, Gen. Relativ. Gravit. **43**, 485 (2011).

[77] K. Jani, D. Shoemaker, and C. Cutler, Nat. Astron. **4**, 260 (2020).

[78] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D.E. Holz, T. Bulik, and F. Pannarale, Astrophys. J. **806**, 263 (2015).

[79] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[80] L. S. Finn, Phys. Rev. D **53**, 2878 (1996).

[81] P. C. Peters, Phys. Rev. **136**, B1224 (1964).

[82] K. Chamberlain, C. J. Moore, D. Gerosa, and N. Yunes, Phys. Rev. D **99**, 024025 (2019).

[83] N. J. Cornish and K. Shuman, Phys. Rev. D **101**, 124008 (2020).

[84] M. Kesden, D. Gerosa, R. O'Shaughnessy, E. Berti, and U. Sperhake, Phys. Rev. Lett. **114**, 081103 (2015).

[85] D. Gerosa, M. Kesden, U. Sperhake, E. Berti, and R. O'Shaughnessy, Phys. Rev. D **92**, 064016 (2015).

[86] D. Gerosa and M. Kesden, Phys. Rev. D **93**, 124066 (2016).

[87] E. Barausse, Mon. Not. R. Astron. Soc. **423**, 2533 (2012).

[88] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi, Astrophys. J. **794**, 104 (2014).

[89] F. Antonini, E. Barausse, and J. Silk, Astrophys. J. Lett. **806**, L8 (2015).

[90] K. Belczynski, S. Repetto, D. E. Holz, R. O'Shaughnessy, T. Bulik, E. Berti, C. Fryer, and M. Dominik, Astrophys. J. **819**, 108 (2016).

[91] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme *et al.*, MBH simulation data release, https://people.sissa.it/~barausse/catalogs/.

[92] E. Berti, J. Gair, and A. Sesana, Phys. Rev. D **84**, 101501 (2011).

[93] E. Poisson and C. M. Will, Phys. Rev. D **52**, 848 (1995).

[94] C. Cutler and E.E. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[95] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).

[96] L. Blanchet, Living Rev. Relativity **17**, 2 (2014).

[97] W. R. Inc., Mathematica, Version 12.1.

[98] A. Griewank, D. Juedes, and J. Utke, ACM Trans. Math. Softw. **22**, 131 (1996).

[99] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, USA, 2007).

[100] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Master Series in Physics (Oxford University Press, Oxford, 2007).

[101] N. Yunes, F. Pretorius, and D. Spergel, Phys. Rev. D **81**, 064018 (2010).

[102] K. Chatziioannou, N. Cornish, A. Klein, and N. Yunes, Phys. Rev. D **89**, 104023 (2014).

[103] A. Gupta, S. Datta, S. Kastha, S. Borhanian, K. Arun, and B. Sathyaprakash, Phys. Rev. Lett. **125**, 201101 (2020).

[104] K. Arun, B. R. Iyer, M. Qusailah, and B. Sathyaprakash, Classical Quantum Gravity **23**, L37 (2006).

[105] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Phys. Rev. D **100**, 024059 (2019).

[106] G. Pratten *et al.*, arXiv:2004.06503.

[107] S. Ossokine *et al.*, Phys. Rev. D **102**, 044055 (2020).

[108] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, Astrophys. J. Lett. **832**, L2 (2016).

[109] N. Steinle and M. Kesden, arXiv:2010.00078.

[110] B. Gough, *GNU Scientific Library Reference Manual* (Network Theory Ltd., 2009).

[111] D. Gerosa and M. Vallisneri, J. Open Source Softw. **2**, 222 (2017).

[112] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).

[113] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Classical Quantum Gravity **37**, 055002 (2020).

[114] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Phys. Rev. D **95**, 104004 (2017).

[115] S. Alexander and N. Yunes, Phys. Rep. **480**, 1 (2009).

[116] F.-L. Julié and E. Berti, Phys. Rev. D **100**, 104061 (2019).

[117] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, Phys. Lett. B **429**, 263 (1998).

[118] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, Phys. Rev. D **59**, 086004 (1999).

[119] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).

[120] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999).

[121] R. Emparan, J. Garcia-Bellido, and N. Kaloper, J. High Energy Phys. 01 (2003) 079.

[122] E. Berti, K. Yagi, and N. Yunes, Gen. Relativ. Gravit. **50**, 46 (2018).

[123] P. Figueras and T. Wiseman, Phys. Rev. Lett. **107**, 081101 (2011).

[124] S. Abdolrahimi, C. Cattoen, D. N. Page, and S. Yaghoobpour-Tari, Phys. Lett. B **720**, 405 (2013).

[125] K. Yagi, N. Tanahashi, and T. Tanaka, Phys. Rev. D **83**, 084036 (2011).

[126] S. Tahura and K. Yagi, Phys. Rev. D **98**, 084042 (2018); **101**, 109902(E) (2020).

[127] C. M. Will, Phys. Rev. D **57**, 2061 (1998).

[128] Astropy Collaboration, Astron. Astrophys. **558**, A33 (2013).

[129] Astropy Collaboration and Astropy Contributors, Astron. J. **156**, 123 (2018).

Correction: Errors in Table II and the affiliated caption, and errors in the last sentence in the last paragraph in Sec. III C have been fixed. Replacement figures for errors in axis labels in Figs. 11, 14, 16, 18, 21, 23, and 25 have been rendered.