

When to sweat the small stuff: Identifying the most informative events from ground-based gravitational-wave detectors

Reed Essick^{1,2,3,*} and Daniel E. Holz⁴

¹*Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, Ontario M5S 3H8, Canada*

²*Department of Physics, University of Toronto, 60 St. George Street, Toronto, Ontario M5S 1A7, Canada*

³*David A. Dunlap Department of Astronomy, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada*

⁴*Department of Physics, Department of Astronomy and Astrophysics, Enrico Fermi Institute, and Kavli Institute for Cosmological Physics, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA*



(Received 16 July 2024; accepted 17 October 2024; published 14 November 2024)

We explore scaling relations for the information carried by individual events, and how that information accumulates in catalogs like those from ground-based gravitational-wave detectors. For a variety of situations, the larger number of quiet/distant signals dominates the overall information over the fewer loud/close sources, independent of how many model parameters are considered. We consider implications for a range of astrophysical scenarios, including calibration uncertainty and standard siren cosmology. However, the large number of additional events obtained by lowering the detection threshold can rapidly increase costs. We introduce a simple analysis that balances the costs of analyzing increasingly large numbers of low information events against retaining a higher threshold and running a survey for longer. With the caveat that precise cost estimates are difficult to determine, current economics favor analyzing low signal-to-noise ratio events. However, the higher detection rates expected for next-generation detectors may argue for a higher signal-to-noise ratio threshold for optimal scientific return.

DOI: [10.1103/PhysRevD.110.103018](https://doi.org/10.1103/PhysRevD.110.103018)

I. INTRODUCTION

Common questions that arise when combining information from many separate observations are, “which observations contain the most information?” and “to what extent does this depend on the type of measurement being performed?” Both considerations are important for forecasts in general, but we focus on catalogs of compact binary coalescences (CBCs) observed with current (e.g., LIGO [1], Virgo [2], and KAGRA [3]) and proposed (e.g., Cosmic Explorer [4] and Einstein Telescope [5]) ground-based gravitational-wave (GW) detectors. With large catalogs and finite computational and/or follow-up resources, knowing how information is encoded within the observations can help limit the cost required to reach a target precision.

This manuscript reviews how information from multiple events can be combined and establishes basic rules-of-thumb to guide cost-benefit analyses. In particular, it corrects previous claims in the literature about the impact of the number of parameters constrained simultaneously [6] and provides improved guidelines for which events contribute to a given measurement (Sec. II). It then provides a few examples in Sec. III, concluding with a discussion of

cost optimization with current and proposed detectors in Sec. IV. Large sets of samples used in all figures (both astrophysical and detected distributions) are available in Essick [7].

II. COMBINING OBSERVATIONS

Our analysis of CBCs’ information content is based on the Fisher information matrix [8],

$$\Gamma_{\alpha\beta} \equiv \int dx p(x|\theta) \frac{\partial \ln p(x|\theta)}{\partial \theta_\alpha} \frac{\partial \ln p(x|\theta)}{\partial \theta_\beta}, \quad (1)$$

which provides an estimate of the constraining power of a single observation (with data x) on source parameters (θ) given a likelihood $p(x|\theta)$. Specifically, we investigate how Γ scales with the signal-to-noise ratio (ρ) of individual events and the distribution of ρ within GW catalogs.

For high- ρ signals, one can approximate the likelihood as Gaussian with precision (inverse covariance) matrix Γ [9,10]. As such, a convenient way to quantify the information carried by an event is the determinant of Γ , which serves as a proxy for the (inverse of the) hypervolume in θ -space that is consistent with observations. Larger $\det|\Gamma|$ imply a tighter constraint and more information about θ .

*Contact author: essick@cita.utoronto.ca

The joint likelihood of a catalog is the product of the individual events' likelihoods (assuming independent noise), each of which can be approximated as a Gaussian with Fisher matrix¹ $\Gamma_{\alpha\beta}^{(e)}$. That is,

$$\begin{aligned} \ln p(\{x_e\}|\theta) &= \sum_e^{N_e} \ln p(x_e|\theta) \\ &\sim -\frac{1}{2} \sum_e^{N_e} \theta_\alpha \Gamma_{\alpha\beta}^{(e)} \theta_\beta \\ &\sim -\frac{1}{2} \theta_\alpha \left(\sum_e^{N_e} \Gamma_{\alpha\beta}^{(e)} \right) \theta_\beta, \end{aligned} \quad (2)$$

and the total precision matrix is the sum of the Fisher matrixes from each event: $\Gamma_{\alpha\beta}^{(\text{tot})} = \sum_e \Gamma_{\alpha\beta}^{(e)}$. If one quantifies the total information as the hypervolume to which the parameters are constrained by the joint analysis, one should consider the determinant of the total precision matrix $\Gamma^{(\text{tot})}$,

$$\begin{aligned} \mathcal{I}_{\det \Sigma} &= \det \left| \sum_e^{N_e} \Gamma_{\alpha\beta}^{(e)} \right| \\ &\approx \det \left| N_e \int_{\rho_{\text{thr}}}^{\rho_{\text{max}}} d\rho p(\rho) \Gamma_{\alpha\beta} \right|, \end{aligned} \quad (3)$$

where $p(\rho)$ is the distribution of ρ expected for a catalog of detected events. For a Euclidean universe (or nearby sources) [12,13],

$$p(\rho) = \frac{\Theta(\rho_{\text{thr}} \leq \rho)}{\rho^4} \left(\frac{3\rho_{\text{max}}^3 \rho_{\text{thr}}^3}{\rho_{\text{max}}^3 - \rho_{\text{thr}}^3} \right), \quad (4)$$

where ρ_{thr} is the detection threshold, Θ is the indicator function, and one typically takes $\rho_{\text{max}} \rightarrow \infty$. This distribution also implicitly assumes that we have a signal-dominated catalog (i.e., noise events are rare), which will not be true if ρ_{thr} is lowered arbitrarily. For CBC parameters that primarily affect the GW phase and stationary Gaussian noise, $\Gamma_{\alpha\beta} \sim \rho^2$, which yields

$$\mathcal{I}_{\det \Sigma} \propto \left[N_e \left(\frac{1}{\rho_{\text{thr}}} - \frac{1}{\rho_{\text{max}}} \right) \left(\frac{3\rho_{\text{max}}^3 \rho_{\text{thr}}^3}{\rho_{\text{max}}^3 - \rho_{\text{thr}}^3} \right) \right]^D. \quad (5)$$

¹Importantly, this model assumes the parameters are shared by all events, like the graviton's mass or (approximately) the neutron star radius as opposed to, e.g., the mass of each component in a binary. This caveat also applies to the analysis within, e.g., Haster *et al.* [6]. Expressions for the Fisher matrix for more general hierarchical inferences also exist [11], but they cannot be used within Eq. (2).

where $D = \text{rank}\{\Gamma\}$ is the dimensionality of the problem (rank of the Fisher matrix). It is interesting to note that $\mathcal{I}_{\det \Sigma}$ is not affected if one asks unrelated questions separately or simultaneously (it scales in the same way as the product of separate questions). See Appendix A for more discussion.

Equation (3) can also incorporate (certain types of) prior information trivially. Defining the information as the amount of hypervolume that is viable *a posteriori* and including a Gaussian prior with covariance $P_{\alpha\beta}$, the joint posterior becomes

$$\ln p(\theta|\{x_e\}) \sim -\frac{1}{2} \theta_\alpha \left(P_{\alpha\beta}^{-1} + \sum_e^{N_e} \Gamma_{\alpha\beta}^{(e)} \right) \theta_\beta, \quad (6)$$

in which case the relevant information metric is

$$\mathcal{I} = \det \left| P_{\alpha\beta}^{-1} + \sum_e^{N_e} \Gamma_{\alpha\beta}^{(e)} \right|. \quad (7)$$

If the Fisher matrix elements are small (i.e., weak constraints from individual events), the (inverse) prior can dominate the sum. However, a sufficient number of signals will eventually overwhelm the prior, and the overall constraint will be dominated by the quietest signals in the catalog in the long run.

III. MORE COMPLICATED Γ VS ρ SCALINGS

The quietest events carry a lot of information in aggregate when $\Gamma \sim \rho^2$. That is, $|d\mathcal{I}_{\det \Sigma}/d\rho_{\text{thr}}| \sim \rho_{\text{thr}}^{-(D+1)}$ (taking $N_e \sim \rho_{\text{thr}}^{-3}$ and $\rho_{\text{max}} \rightarrow \infty$), and information is most-rapidly accumulated at small ρ_{thr} due to the large number of signals. However, in general, Γ may have a more complicated relationship with ρ , in which case the total constraint may no longer be dominated by the quietest events in the catalog. The following sections provide a few concrete examples, which are summarized in Fig. 1 for advanced LIGO and Virgo. Appendix B discusses how this behavior can differ between current and proposed GW detectors.

A. Breakdown of the Fisher approximation

The assumption that the single-event precision matrix scales as $\Gamma_{\alpha\beta} \sim \rho^2$ only holds in the high- ρ limit. In truth, the Fisher information matrix only provides a lower-limit on the covariance (upper-limit on the precision matrix), and the actual covariance can be much larger [10]. This happens for small ρ , which means that the events which contribute meaningfully to the sum in Eq. (3) may effectively need to pass a higher threshold than detection.

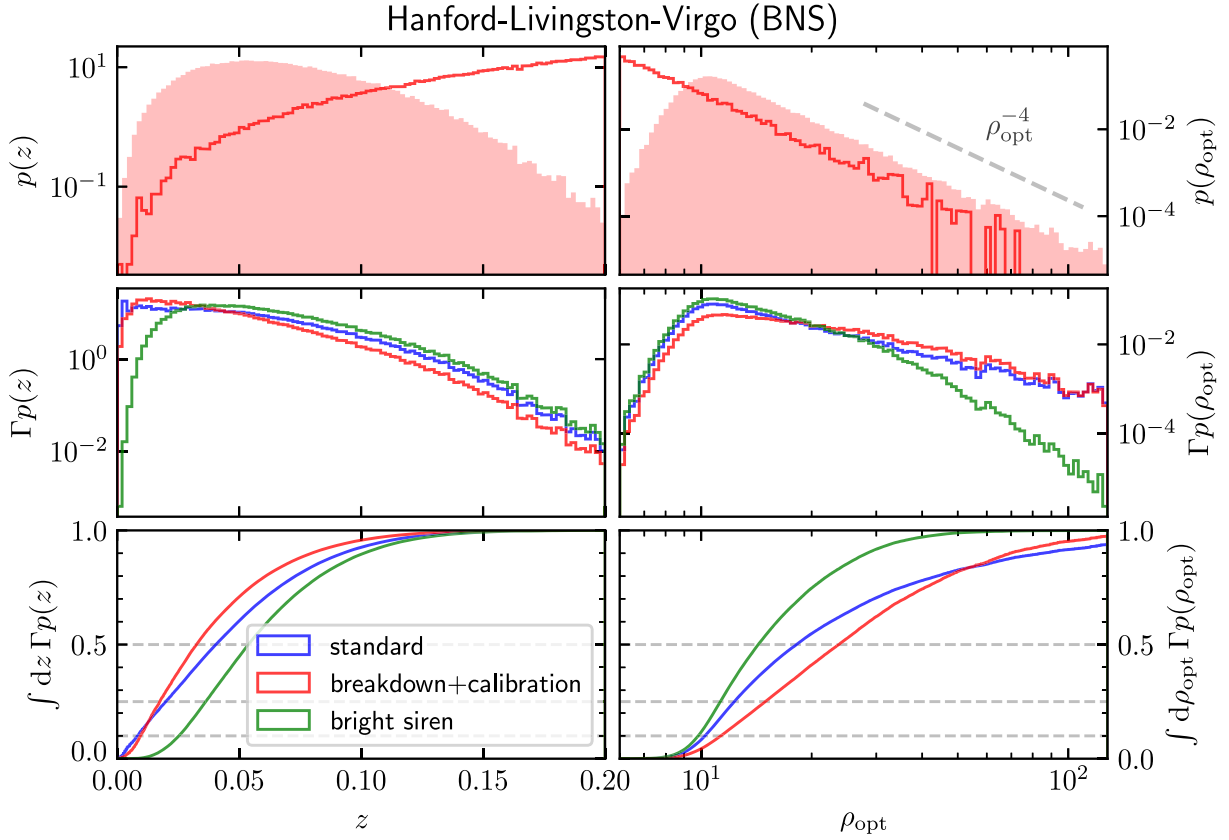


FIG. 1. Distributions of binary neutron star (BNS) signals and their contributions to $\mathcal{I}_{\text{det}\Sigma}$ over the redshift to the source (left, z) and the source's optimal signal-to-noise ratio (right, ρ_{opt}) with a network of advanced LIGO and Virgo detectors at design sensitivity. Top to bottom: distributions of astrophysical (solid) and detected (shaded) signals (with observed signal-to-noise ratio ≥ 10 [14]), the detected-signal distribution weighted by individual events' Fisher matrixes, and the (normalized) integral for $\mathcal{I}_{\text{det}\Sigma}$ showing how much information is accumulated as the upper limit of the integral is increased. The bottom panels show the relative importance of different events assuming $\Gamma \sim \rho^2$ (blue, standard scaling), $\Gamma \sim \rho^4(1 + (\rho/20)^2)^{-1}(1 + (\rho/100)^2)^{-1}$ (red, breakdown of the Fisher approximation at $\rho \lesssim 20$ [15] and pessimistic calibration uncertainty above $\rho \gtrsim 100$ [16]), and $\Gamma \sim [((200 \text{ km/s})/cz)^2 + (10\%(10/\rho))^2]^{-1}$ (green, bright sirens) [17]. Horizontal lines highlight the points at which 10%, 25%, and 50% of the total information is accumulated.

For example, one possible scaling is

$$\Gamma \sim \rho^2 \left(\frac{\rho^2}{\rho_o^2 + \rho^2} \right), \quad (8)$$

for which $\Gamma \sim \rho^4$ at small ρ and $\Gamma \sim \rho^2$ for large ρ . Here, ρ_o sets the scale below which the Fisher approximation begins to break down, which will almost certainly be problem-dependent. Additionally, the breakdown in $\Gamma \sim \rho^2$ may not exactly follow Eq. (8), but Eq. (8) captures the general behavior and makes the associated integrals analytically tractable. Assuming $\rho_{\text{max}} \rightarrow \infty$,

$$\int dN\Gamma \sim N_e \frac{3\rho_{\text{thr}}^3}{\rho_o} \left(\frac{\pi}{2} - \tan^{-1} \left(\frac{\rho_{\text{thr}}}{\rho_o} \right) \right). \quad (9)$$

Since $N_e \sim \rho_{\text{thr}}^{-3}$ for nearby sources, Eq. (5) diverges as ρ_{thr}^{-1} as $\rho_{\text{thr}} \rightarrow 0$. However, Eq. (9) remains finite in the same

limit. Indeed, ρ_o determines which events carry the most information; at least half the total information available is carried by the events with $\rho \geq \rho_o$ (Fig. 1).

Additionally, a breakdown in the Fisher approximation is a likely explanation for the behavior reported in Lackey and Wade [15] (neutron star equation of state constraints are dominated by the loudest events in their catalog, and there was almost no additional information carried by quieter events). Motivated by Lackey and Wade [15], Fig. 1 assumes $\Gamma \sim \rho^2$ breaks down below $\rho_o \sim 20$.

B. Calibration uncertainty

Alternatively, $\Gamma \sim \rho^2$ may break down at high ρ in the presence of calibration or waveform uncertainty. Essick [16] shows that Γ can approach a constant at large ρ (e.g., the fractional uncertainty from calibration), which can be modeled as

$$\Gamma \sim \rho^2 \left(\frac{\rho_o^2}{\rho_o^2 + \rho^2} \right), \quad (10)$$

where ρ_o defines the scale at which calibration uncertainties begin to dominate, in which case²

$$\int dN\Gamma \sim N_e \frac{3\rho_{\text{thr}}^3}{\rho_o} \left(\frac{\rho_o}{\rho_{\text{thr}}} + \tan^{-1} \left(\frac{\rho_{\text{thr}}}{\rho_o} \right) - \frac{\pi}{2} \right). \quad (11)$$

The quietest signals in the catalog dominate the overall information even more than for the standard scaling, although this approaches the standard scaling if $\rho_{\text{thr}} \ll \rho_o$ (i.e., if almost all of the catalog is too quiet to be affected by calibration errors). See Fig. 1, which assumes calibration limits the precision of signals at $\rho_o \sim 100$ (much lower than is likely to be the case for realistic calibration uncertainties [16]).

C. Standard sirens

Finally, single-event constraints may always be a complicated function of ρ , and the approximation $\Gamma \sim \rho^2$ may simply never hold. Standard sirens, where one measures the Hubble parameter (H_0) at the ratio of a redshift (z) from a (candidate) host galaxy and a luminosity distance (D_L) from the GW strain, are one such example.

In the simplest approximation,

$$H_0 = \frac{cz}{D_L}, \quad (12)$$

where c is the speed of light. Standard error propagation suggests that

$$\Gamma_{H_0}^{-1} \sim \sigma_{H_0}^2 \sim H_0^2 \left(\frac{\sigma_z^2}{z^2} + \frac{\sigma_{D_L}^2}{D_L^2} \right), \quad (13)$$

where z is estimated from a set of candidate host galaxies and D_L is estimated from the GW signal.

A straightforward Fisher analysis for the GW data (neglecting the degeneracy with inclination) suggests that

$$\sigma_{D_L} \propto \frac{D_L}{\rho}. \quad (14)$$

Including the degeneracy with inclination, one might expect $\sigma_{D_L} \propto D_L$. Breaking this degeneracy depends on measuring the GW polarization and/or higher-order modes,

²This assumes calibration errors are independent for each event. If calibration errors are not independent, the joint constraint will be more complicated. However, we expect our conclusions about which events carry the most information about astrophysical parameters to still approximately hold because individual GW events carry relatively little information about calibration uncertainties [16,18] (i.e., most signals are too quiet to be affected by calibration errors).

which has been done but often requires relatively large ρ [19,20].

The uncertainty in z from a galaxy catalog is the sum of two components: the (average) measurement uncertainty for the redshift of individual galaxies (which could include their peculiar velocity) and a term that captures the variance of redshifts within the catalog ($\{\mu_g\}$)

$$\sigma_z^2 \sim \frac{1}{N_g} \sum_g \sigma_g^2 + \frac{1}{N_g} \sum_g \left(\mu_g - \frac{1}{N_g} \sum_g \mu_g \right)^2. \quad (15)$$

Assuming most galaxies have similar measurement uncertainty (σ_g), only the second term depends on ρ (through the set of galaxies consistent with the GW localization). Here, we focus on the case of a bright siren, where a single host galaxy is identified ($N_g = 1$ and $\sigma_z^2 = \sigma_g^2$).

As discussed in Chen *et al.* [17], the precision on H_0 for bright sirens is then limited by the distance uncertainty for quiet/distant events while the precision of the loudest/nearest sources is limited by the uncertainty in the redshift measurement for individual galaxies, which does not vanish as $z \rightarrow 0$.³ Therefore, the most informative events in the catalog come from intermediate ρ . Figure 1 assumes $\sigma_g \sim (200 \text{ km/s})/c$ and $\sigma_{D_L}/D_L \sim 10\%$ at $\rho \sim 10$.

IV. DISCUSSION

It is useful to rewrite Eq. (5) in terms of only N_e so that $\mathcal{I}_{\text{det}\Sigma} \sim N_e^{1/3}$ when $\Gamma \sim \rho^2$, which implies $\sigma \sim N_e^{-1/6}$. As such, lowering ρ_{thr} enough to decrease σ by a factor of ~ 2 will increase the catalog size (and associated computational cost) by a factor of $\sim 2^6 = 64$. Figure 2 plots the cumulative information vs the fraction of the catalog for several detector networks and scalings of Γ vs ρ . It is clear that a large fraction of the overall information can come from a relatively small fraction of the catalog.

In contrast, if one fixes ρ_{thr} but runs the experiment longer, a factor of ~ 4 larger catalog size will reduce σ by a factor of ~ 2 . This begs the question of whether it could be more cost effective to run an experiment for longer with a high threshold instead of lowering ρ_{thr} .

Consider the following cost model

$$C = C_F + C_O\tau + C_A(N_e), \quad (16)$$

where C_F , C_O , and $C_A(N_e)$ are the fixed cost of constructing an experiment, the marginal cost of operating the experiment (with duration τ), and the cost of analyzing N_e events, respectively. For current detectors, the number of events scales approximately as $N_e \sim R\tau(\rho_{\text{ref}}/\rho_{\text{thr}})^3$, where

³The second term in Eq. (15) may grow rapidly at low ρ (large N_g), and therefore the precision of (dark) standard sirens may always be limited by σ_z^2 . See Appendix C.

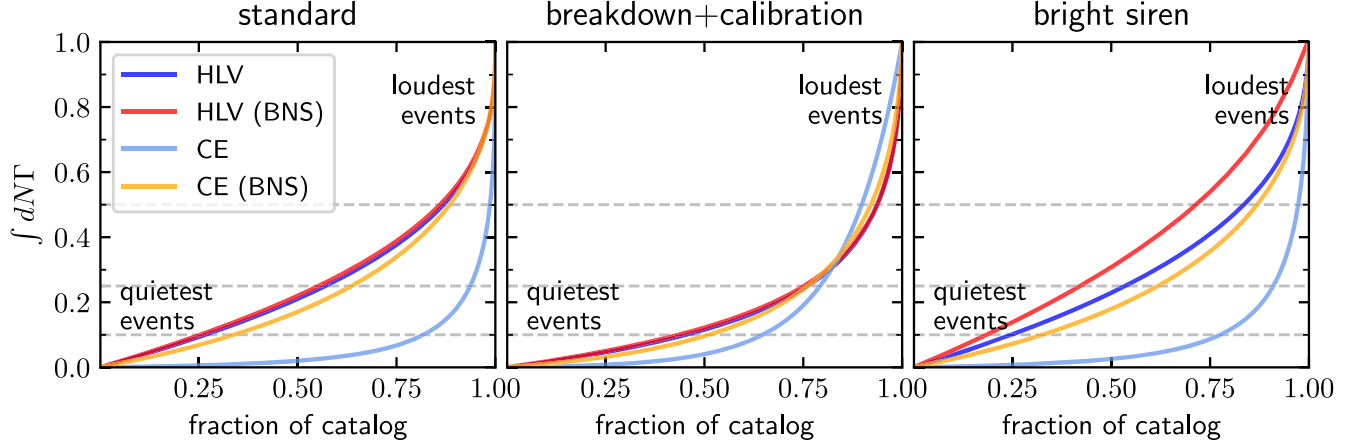


FIG. 2. Integrated information vs the fraction of the catalog considered (including quieter events first) when Γ scales with ρ in the ways shown in Fig. 1. We show the results for low-mass signals observed with advanced LIGOs and Virgo (red) or CE (orange) as well as the full range of CBCs observed with advanced detectors (dark blue) or CE (light blue). Horizontal lines highlight when 10%, 25%, and 50% of the total information has been accumulated.

R is the rate of detections with a threshold ρ_{ref} . The precision of the measurement will scale as $\mathcal{I} \sim N_e/f(\rho_{\text{thr}}) \sim R\tau\rho_{\text{ref}}^3/f(\rho_{\text{thr}})\rho_{\text{thr}}^3$, and, in the simplest approximation [Eq. (3)], $f \sim \rho_{\text{thr}}^{-2}$. Minimizing C with respect to ρ_{thr} at fixed $\mathcal{I}$ yields

$$\left. \frac{dC}{d\rho_{\text{thr}}} \right|_{\mathcal{I}} = C_O \frac{\mathcal{I}}{R\rho_{\text{ref}}^3} \frac{d}{d\rho_{\text{thr}}} (\rho_{\text{thr}}^3 f) + \frac{dC_A}{dN_e} \mathcal{I} \frac{df}{d\rho_{\text{thr}}} = 0 = \frac{C_O}{R} \left(\frac{\rho_{\text{thr}}}{\rho_{\text{ref}}} \right)^3 - 2 \frac{dC_A}{dN_e}, \quad (17)$$

which defines an implicit expression for the optimal ρ_{thr} in terms of $\mathcal{I}$ and marginal costs. Once this is set, the required duration is given by $\tau \sim \mathcal{I}\rho_{\text{thr}}^3/f/\rho_{\text{ref}}^3 R$.

Interestingly, in the simplest model where the marginal analysis cost (dC_A/dN_e) is constant, the optimal ρ_{thr} does not depend on the precision $\mathcal{I}$,⁴ and

$$\rho_{\text{thr}} = \rho_{\text{ref}} \left(\frac{2RdC_A/dN_e}{C_O} \right)^{1/3}. \quad (18)$$

This balances the cost of analyzing additional events (per unit time) and the cost of operating the instrument (per unit time).

Estimating the costs in Eq. (18) precisely is nontrivial, but even rough numbers may provide guidance for current and/or planned detectors. In this spirit, we provide such estimates below. However, we stress that these examples are merely illustrative, and more detailed calculations

⁴Although the general optimization can depend on the amount of information per event and the amount of information per event (at fixed ρ) may change between, e.g., current and next-generation detectors, it is often similar [16].

should be conducted before any policy is set by such considerations.

The current operating cost of advanced LIGO is $\sim 10^7$ USD/yr [21] with an expected rate of detections of $R \sim 1/\text{day}$ at design sensitivity [22]. Estimating the marginal cost of analyzing an event as approximately 10 hrs of an analyst's time ($dC_A/dN_e \sim 300\text{USD/event}$),⁵ Eq. (18) becomes $\rho_{\text{thr}}/\rho_{\text{ref}} \approx 0.28$. As such, the relatively low rate of detections drives the most efficient threshold lower than the nominal value (assuming the Fisher approximation holds at low ρ).

However, the rate of detections may dramatically increase with next-generation detectors, by a factor of $\gtrsim 100$.⁶ Assuming C_O and dC_A/dN_e are roughly the same for future detectors, the optimal ρ_{thr} will be a factor of 1.4 higher than ρ_{ref} .⁷ Again, this also assumes that the Fisher approximation holds for low- ρ events (it almost certainly will not) and that the catalog is flux-limited. Alternatively, if dC_A/dN_e improves even moderately (by as little as a factor of ~ 3), then the optimal ρ_{thr} will still be $\lesssim \rho_{\text{ref}}$. This modest reduction in analysis costs may be readily achievable before next-generation detectors come online. Additionally, the rate of binary neutron star (BNS) detections may continue to be much smaller than the overall detection rate. This would almost guarantee that the optimal ρ_{thr} for questions involving BNS will remain $\lesssim \rho_{\text{ref}}$.

⁵An hourly wage of 30 USD corresponds to 6×10^4 USD/yr, which is likely a low estimate of labor costs including both salary and overhead.

⁶Although the detection horizon may increase by an order of magnitude (naively suggesting an increase of $\gtrsim 10^3$ in detection rate), this extends beyond the peak of star formation.

⁷If operating costs increase proportional to the detector's arms' lengths (i.e., by an order of magnitude), then $\rho_{\text{thr}}/\rho_{\text{ref}} \sim 0.66$.

even with higher detection rates in next-generation catalogs.

All this being said, Fig. 2 shows that, while the simple scaling motivated by Eq. (5) may approximately hold for low-mass CBCs observed with next-generation detectors, it is unlikely to hold for high-mass CBCs. Such surveys will not be flux-limited, and the distribution $p(\rho)$ will be fundamentally different than Eq. (4). Indeed, one will likely lose even less information by considering only the loudest high-mass events observed with CE. See Appendix B for more discussion.

Furthermore, dC_A/dN_e is almost certainly not constant. For example, contamination from noise events will grow rapidly with decreasing ρ_{thr} , and the overall number of events that would require follow-up would be much larger than the number of signals [23]. The confusion between noise and signal is also bound to reduce the information per signal at low ρ . This effectively sets the detection threshold within current catalogs (e.g., GWTC-3 [24]).

Putting such concerns aside, one might additionally optimize over facilities by first optimizing over $(\tau, \rho_{\text{thr}})$ separately, and then optimizing over facilities while taking into account their C_F . This is often done only in the context of the science accessible with different detector designs, neglecting their relative costs (see, e.g., Refs. [25,26]). However, a prerequisite for such an analysis would be an enumeration of science targets, and it is difficult to concisely capture the full range of discovery potential enabled by, e.g., next-generation GW detectors. At the very least, such an exercise is beyond the scope of this paper.

Our results suggest that a more detailed cost-benefit analysis may help determine the optimal balance of survey duration and detection threshold for next-generation detectors. The simplest scaling and cost models suggest that it makes sense to go deep within current catalogs (although our analysis is intentionally only a rough estimate). With reasonable improvements in analysis costs, this may remain true of future catalogs even with much higher detection rates.

ACKNOWLEDGMENTS

We are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. R. E. is supported by the Natural Sciences & Engineering Research Council of Canada (NSERC) through a Discovery Grant (No. RGPIN-2023-03346). D. E. H. is supported by NSF Grants No. AST-2006645 and No. PHY2110507, as well as by the Kavli Institute for Cosmological Physics through an endowment from the Kavli Foundation and its founder Fred Kavli. This study would not have been possible without the following software: NUMPY [27], SCIPY [28], and MATPLOTLIB [29]. The large sample of detected events for both HLV at design sensitivity and CE, including the impact of the

frequency-dependent interferometric response [30], are publicly available in Essick [7]. Samples were generated with monte-carlo-vt [31], gw-distributions [32], and gw-detectors [33].

APPENDIX A: A COMMENT ON HASTER *ET AL.* [6]

Previously, Haster *et al.* [6] argued that the dimensionality of the question(s) being asked plays a role in which events carry the most information. That is, they claim that the set of events which are most informative depends on the number of parameters being constrained: quiet/distant events dominate one-dimensional questions but higher-dimensional questions depend most strongly on loud/nearby events. However, if one considers two properties that are unrelated (i.e., uncorrelated at the single-event level), the amount of information, and which events carry that information, should not depend on whether these properties are estimated separately or simultaneously.

Haster *et al.* [6] correctly identify that each element of Γ often scales with ρ^2 for a single event. As such, $\det |\Gamma|$ should have an overall scaling of ρ^{2D} , where $D = \text{rank}\{\Gamma\}$ is the dimensionality of the problem. They then estimate the total information by summing $\det |\Gamma|$ from separate events. That is, they compute

$$\mathcal{I}_{\Sigma \text{ det}} = \sum_e^{N_e} \det |\Gamma_{\alpha\beta}^{(e)}| \approx N_e \int_{\rho_{\text{thr}}}^{\rho_{\text{max}}} d\rho p(\rho) \det |\Gamma_{\alpha\beta}| \quad (\text{A1})$$

which has very different behavior than Eq. (3). If $\det |\Gamma_{\alpha\beta}|$ scales as ρ^{2D} , then Eq. (A1) becomes

$$\mathcal{I}_{\Sigma \text{ det}} \propto N_e \left(\frac{\rho_{\text{max}}^{2D-3} - \rho_{\text{thr}}^{2D-3}}{2D-3} \right) \left(\frac{3\rho_{\text{max}}^3 \rho_{\text{thr}}^3}{\rho_{\text{max}}^3 - \rho_{\text{thr}}^3} \right) \quad (\text{A2})$$

which diverges as $\rho_{\text{max}} \rightarrow \infty$ if $D \geq 3/2$. This logic suggests that the single loudest event carries almost all the information for high-dimensional questions ($D \gg 1$).

$\mathcal{I}_{\Sigma \text{ det}}$ also suggests that the constraint on each parameter scales worse with the size of the catalog if multiple questions are asked simultaneously ($\mathcal{I}_{\Sigma \text{ det}}^{1/D} \sim N^{1/D}$), even if they are unrelated. That is, even if $\Gamma_{\alpha\beta}^{(e)}$ is always block-diagonal, Haster *et al.* [6] incorrectly predict that asking multiple questions simultaneously affects our ability to answer them.

Fundamentally, then, the issue with the analysis in Haster *et al.* [6] boils down to the fact that

$$\sum_e^{N_e} \det |\Gamma_{\alpha\beta}^{(e)}| \neq \det \left| \sum_e^{N_e} \Gamma_{\alpha\beta}^{(e)} \right| \quad (\text{A3})$$

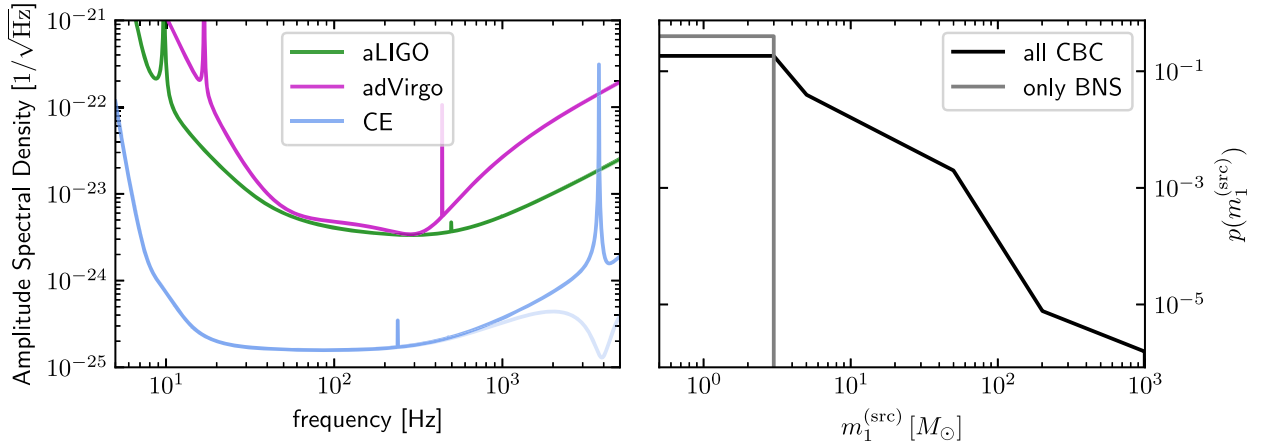


FIG. 3. Left: amplitude spectral densities for advanced LIGO and Virgo at design sensitivity and Cosmic Explorer. For CE, the dark line shows the response to a source that is directly overhead the detector, and light line shows the response without frequency-dependent interferometric effects [30]. Right: source-frame primary mass $[m_1^{(\text{src})}]$ distributions used within this study, intended to mimic the observed distribution of merging binaries (all CBC [36]) or capture the behavior of only low-mass objects (only BNS).

Many of their conclusions still hold, though. Their prediction that the cosmologically distant sources at $z \sim 1$ will dominate the total information from BNS mergers observed with next-generation GW detectors is correct (most of the observed sources will be at $z \sim 1$; see Appendix B). Similarly, they correctly point out that tidal information at high frequencies will be redshifted into the sensitive band of GW detectors for cosmologically distant BNS (see also Refs. [34,35]), and this makes Γ a more complicated function of ρ (other examples are given in Sec. III). In this way, Haster *et al.* [6] correctly point out that the set of events which contain the most information can depend on the behavior of $\Gamma(\rho)$, even though they incorrectly attribute this to the dimensionality of the question. Lastly, their observation that the BNS merger rate may not scale with the comoving volume is also correct.

APPENDIX B: OTHER POPULATIONS AND DETECTOR NETWORKS

This appendix compares the behavior of different populations in different detector networks.

Figure 3 shows the one-sided amplitude spectral densities (ASDs) used for the two LIGO detectors (Hanford and Livingston), Virgo, and CE. It also shows the two source-frame primary-mass distributions considered: one that is based roughly on the current observed population of merging binaries [36] and one restricted to low-masses intended to model BNS systems [$1M_\odot \leq m_2^{(\text{src})} \leq m_1^{(\text{src})} \leq 3M_\odot$].

Figure 4 shows the distribution of inclinations (θ_{jn}) for detected events with both populations shown in Fig. 3 for both the HLV network and for CE. Although the selection against edge-on binaries ($\theta_{jn} = 90^\circ$) is reduced for CE compared to HLV, it is still noticeable. However, unlike

with HLV, it is also clear that the extent of this selection depends on the binary's masses with CE. Indeed, Fig. 4 also shows that there is still a strong selection toward more massive binaries within low-mass populations. Catalogs of low-mass systems from next-generation detectors will still be flux-limited and have strong selections on, e.g., θ_{jn} and the component masses.⁸

Finally, Fig. 5 shows which events are most important with catalogs of sources detected by CE (compare to Fig. 1). The main difference boils down to $p(\rho)$, which, again, is not flux-limited for massive BBH in CE. As such, some of the scalings reported in the main text break for high-mass signals in CE. Indeed, one may increase ρ_{thr} from ~ 10 up to ~ 100 without losing much information if $\Gamma \sim \rho^2$ for high-mass CBCs. However, this would correspond to a reduction of a factor of only ~ 10 in the catalog size (Fig. 2), rather than the expected factor of 10^3 for a flux-limited catalog.

APPENDIX C: STATISTICAL STANDARD SIREN REDSHIFT UNCERTAINTY

The distribution of redshifts from a galaxy catalog can be approximated as

$$p(z) = \sum_g \frac{w_g}{\sqrt{2\pi\sigma_g^2}} e^{-(z-\mu_g)^2/2\sigma_g^2} \quad (\text{C1})$$

assuming Gaussian errors on each galaxy's measured redshift and a weight

⁸The conclusions in, e.g., Vitale [37] are only true for massive binary black holes (BBHs). Vitale [37] focuses on BBHs and acknowledge that their observations may not apply for low-mass systems. See also Vitale [38].

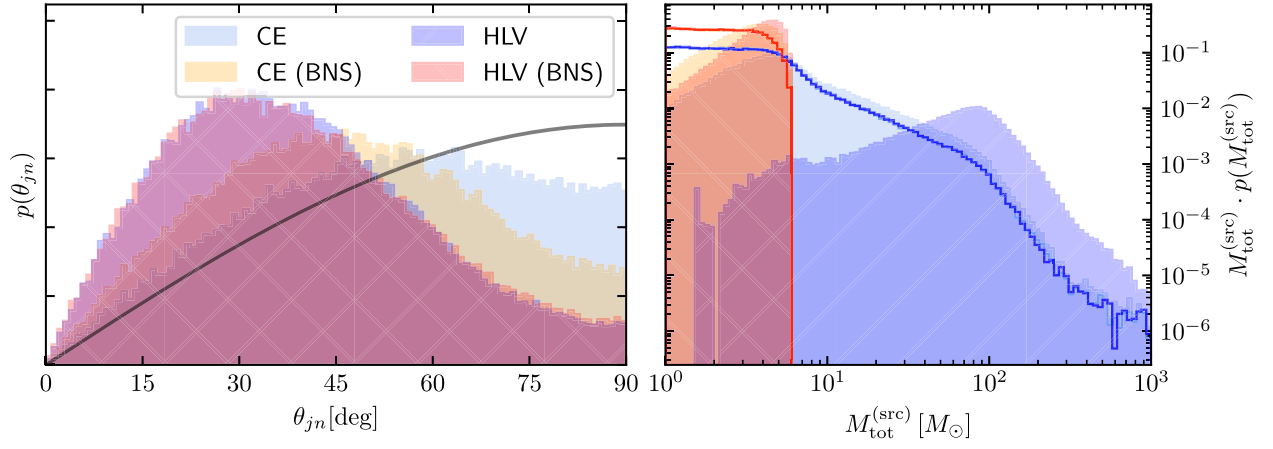


FIG. 4. Left: distribution of orbital inclinations (θ_{jn}) for detected events with both populations from Fig. 3 for HLV at design sensitivity and CE. Because the distribution is symmetric about $\theta_{jn} = 90^\circ$, we only show $0^\circ \leq \theta_{jn} \leq 90^\circ$. An isotropic distribution (black line) is also shown for reference. Right: distributions of total source-frame mass [$M_{\text{tot}}^{(\text{src})}$] for the astrophysical (solid lines) and detected (shaded) populations. While the selection in favor of high-mass systems is reduced at high masses for CE, it remains for $M_{\text{tot}}^{(\text{src})} \gtrsim 5M_\odot$.

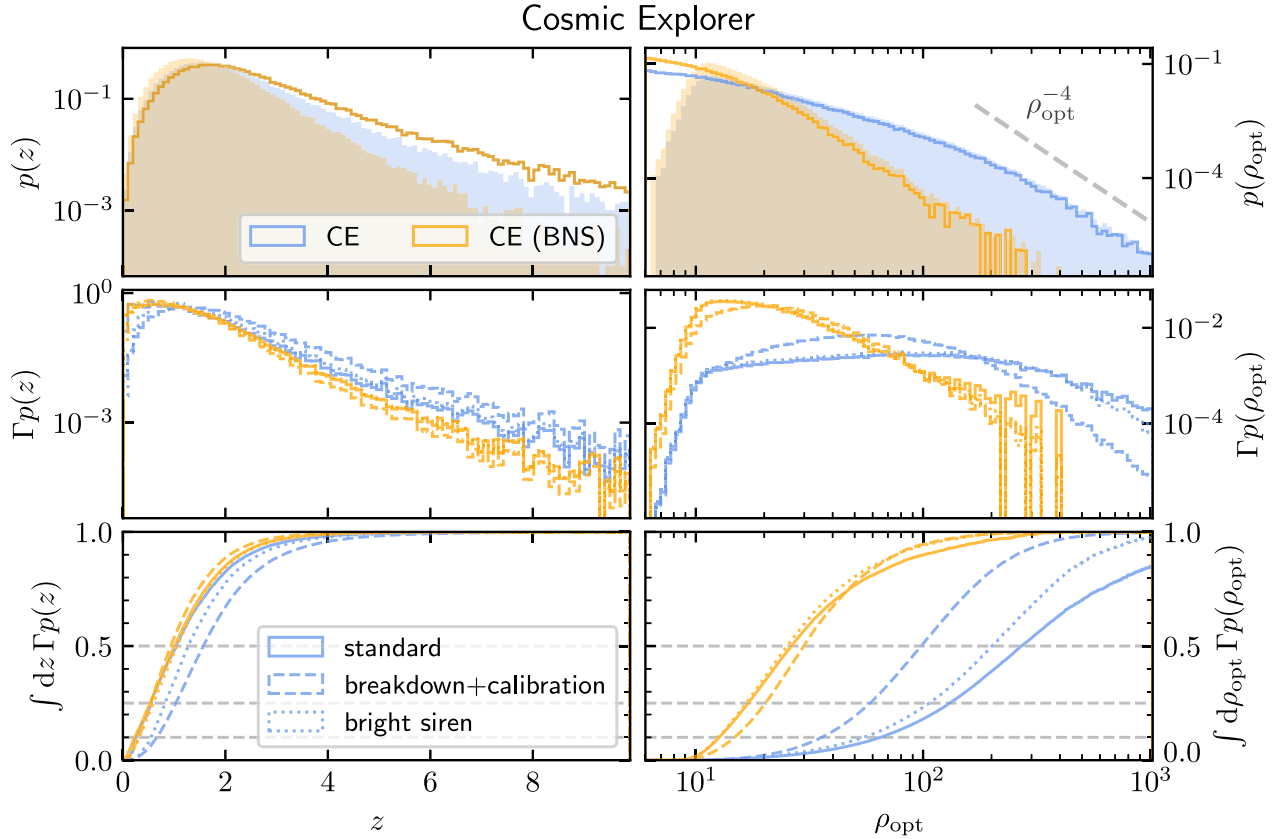


FIG. 5. Analogous to Fig. 1 but for sets of events detected with CE. The cosmological reach of next-generation detectors (to beyond the peak of star formation) means that the distribution of detected ρ falls off much less steeply for high-mass systems. As such, relatively little information is lost by neglecting the low- ρ , high-mass systems. However, catalogs of low-mass systems remain flux-limited, and a significant fraction of the overall information available is carried by the low- ρ , low-mass events.

$$w_g \sim p(\Omega_g | x_{\text{GW}}) \omega_g \quad (\text{C2})$$

for each galaxy, which is the product of the consistency of the Galaxy's position on the sky (Ω_g) with the GW data (x_{GW}) and an additional weight based on the galaxy's properties (ω_g), which captures prior beliefs about the types of galaxies that are more likely to host binary mergers (see, e.g., Hanselman *et al.* [39]). With such a catalog, one can immediately compute

$$\sigma_z^2 \equiv \text{V}[z] = \sum_g^{N_g} w_g \sigma_g^2 + \sum_g^{N_g} w_g \left(\mu_g - \sum_g^{N_g} w_g \mu_g \right)^2 \quad (\text{C3})$$

which is Eq. (15) when $w_g = 1/N_g$.

One can estimate σ_z^2 for a representative catalog by taking the expected value of Eq. (C3) with respect to set of galaxies that are uniformly distributed in comoving volume with a Poisson distribution for the number of galaxies in the catalog, neither of which are perfect models of reality. That is, for simplicity, assume

$$w_g = \frac{1}{N_g} \quad (\text{C4})$$

and

$$p(\mu_g) \propto \frac{dV_c}{dz} \quad (\text{C5})$$

for $z_{\min} < z < z_{\max}$. This yields

$$\text{E}[\sigma_z^2]_{\mu_g, N_g} = \frac{1}{N_g} \sum_g \sigma_g^2 + \left(1 - \frac{1}{N_g} \right) \text{V}[\mu_g] \quad (\text{C6})$$

where

$$\text{V}[\mu_g] = \int_{z_{\min}}^{z_{\max}} dz p(z) z^2 - \left(\int_{z_{\min}}^{z_{\max}} dz p(z) z \right)^2 \quad (\text{C7})$$

For well-localized events, $N_g \rightarrow 1$ and the second term vanishes. For poorly localized events, $N_g \gg 1$ and the second term may dominate.

Taking the expectation value with respect to the number of galaxies, which is Poisson-distributed

$$P(N_g) = \frac{\Lambda^{N_g}}{N_g!} \left(\frac{1}{e^\Lambda - 1} \right) \quad (\text{C8})$$

for $N_g = 1, 2, \dots$, yields

$$\begin{aligned} \text{E}[\sigma_z^2]_{\mu_g, N_g} &= \langle \sigma_g^2 \rangle + \text{V}[\mu_g] \left(1 - \sum_{N_g=1}^{\infty} P(N_g) \frac{1}{N_g} \right) \\ &= \langle \sigma_g^2 \rangle + \text{V}[\mu_g] \left(1 - \frac{1}{e^\Lambda - 1} \int_0^\Lambda d\lambda \frac{e^\lambda - 1}{\lambda} \right) \end{aligned} \quad (\text{C9})$$

Although the integral is difficult to express in closed form, one can still gain insight from recognizing that the second term approaches $\text{V}[\mu_g] \Lambda/4$ as $\Lambda \rightarrow 0$, and that

$$\text{E}[N_g] = \Lambda \propto (\Delta\Omega) \left(\frac{dN_g}{dz} \Delta z \right) \quad (\text{C10})$$

could scale in several ways. $\Delta\Omega$ should scale as ρ^{-1} (ρ^{-2}) for 2-IFO (3 + IFO) localizations, and Δz should scale as $\Delta z = z_{\max} - z_{\min} \sim (D_L/c) \Delta H_0 \sim \rho^{-1}$ where ΔH_0 is the extent of the H_0 prior, although it could also scale as a more complicated function of ρ if one took into account the uncertainty on D_L .

Similarly, if galaxies are uniformly distributed in volume, then $dN_g/dz \sim z^2 \sim \rho^{-2}$ (for nearby sources). As such, one may expect $\text{E}[N_g]$ to scale strongly with ρ , and the uncertainty $(1 - N_g^{-1}) \text{V}[\mu_g]$ may shrink rapidly for loud signals. Additionally, $\text{V}[\mu_g]$ may scale as $(\Delta z)^2 \sim \rho^{-2}$ when $N_g \gg 1$, so that $\text{V}[\mu_g]/z^2 \sim \text{constant}$.

-
- [1] J. Aasi *et al.*, Advanced LIGO, *Classical Quantum Gravity* **32**, 074001 (2015).
 [2] F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, *Classical Quantum Gravity* **32**, 024001 (2014).
 [3] T. Akutsu *et al.*, Overview of KAGRA: Detector design and construction history, *Prog. Theor. Exp. Phys.* **2021**, 05A101 (2020).

- [4] M. Evans *et al.*, A Horizon Study for Cosmic Explorer: Science, observatories, and community, [arXiv:2109.09882](https://arxiv.org/abs/2109.09882).
 [5] M. Branchesi *et al.*, Science with the Einstein Telescope: A comparison of different designs, *J. Cosmol. Astropart. Phys.* **07** (2023) 068.
 [6] C.-J. Haster, K. Chatziioannou, A. Bauswein, and J. A. Clark, Inference of the neutron star equation of state from cosmological distances, *Phys. Rev. Lett.* **125**, 261101 (2020).

- [7] R. Essick, When to sweat the small stuff, [10.5281/zenodo.12546536](https://arxiv.org/abs/10.5281/zenodo.12546536) (2024).
- [8] R. A. Fisher, On the mathematical foundations of theoretical statistics, *Phil. Trans. R. Soc. A* **222**, 309 (1922).
- [9] C. Cutler and E. E. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?, *Phys. Rev. D* **49**, 2658 (1994).
- [10] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, *Phys. Rev. D* **77**, 042001 (2008).
- [11] J. R. Gair, A. Antonelli, and R. Barbieri, A Fisher matrix for gravitational-wave population inference, *Mon. Not. R. Astron. Soc.* **519**, 2736 (2022).
- [12] B. F. Schutz, Networks of gravitational wave detectors and three figures of merit, *Classical Quantum Gravity* **28**, 125023 (2011).
- [13] H.-Y. Chen and D. E. Holz, The loudest gravitational wave events, [arXiv:1409.0522](https://arxiv.org/abs/1409.0522).
- [14] R. Essick, Semianalytic sensitivity estimates for catalogs of gravitational-wave transients, *Phys. Rev. D* **108**, 043011 (2023).
- [15] B. D. Lackey and L. Wade, Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars, *Phys. Rev. D* **91**, 043002 (2015).
- [16] R. Essick, Calibration uncertainty's impact on gravitational-wave observations, *Phys. Rev. D* **105**, 082002 (2022).
- [17] H.-Y. Chen, M. Fishbach, and D. E. Holz, A two per cent Hubble constant measurement from standard sirens within five years, *Nature (London)* **562**, 545 (2018).
- [18] R. Essick and D. E. Holz, Calibrating gravitational-wave detectors with GW170817, *Classical Quantum Gravity* **36**, 125002 (2019).
- [19] R. Abbott *et al.*, GW190412: Observation of a binary-black-hole coalescence with asymmetric masses, *Phys. Rev. D* **102**, 043015 (2020).
- [20] R. Abbott *et al.*, GW190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, *Astrophys. J. Lett.* **896**, L44 (2020).
- [21] Laser interferometer gravitational-wave observatory funding, https://nsf.gov/resources/nsf.gov/about/budget/fy2022/pdf/66i_fy2022.pdf (2022).
- [22] B. P. Abbott *et al.*, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **19**, 1 (2016).
- [23] R. Lynch, M. Coughlin, S. Vitale, C. W. Stubbs, and E. Katsavounidis, Observational implications of lowering the LIGO-Virgo alert threshold, *Astrophys. J. Lett.* **861**, L24 (2018).
- [24] R. Abbott *et al.*, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, *Phys. Rev. X* **13**, 041039 (2023).
- [25] L. S. Finn and D. F. Chernoff, Observing binary inspiral in gravitational radiation: One interferometer, *Phys. Rev. D* **47**, 2198 (1993).
- [26] V. Srivastava, D. Davis, K. Kuns, P. Landry, S. Ballmer, M. Evans, E. D. Hall, J. Read, and B. S. Sathyaprakash, Science-driven tunable design of cosmic explorer detectors, *Astrophys. J.*, **931**, 22 (2022).
- [27] C. R. Harris *et al.*, Array programming with NUMPY, *Nature (London)* **585**, 357 (2020).
- [28] P. Virtanen *et al.*, SCIPY 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [29] J. D. Hunter, MATPLOTLIB: A 2d graphics environment, *Comput. Sci. Eng.* **9**, 90 (2007).
- [30] R. Essick, S. Vitale, and M. Evans, Frequency-dependent responses in third generation gravitational-wave detectors, *Phys. Rev. D* **96**, 084004 (2017).
- [31] R. Essick, Monte-Carlo-Vt, <https://git.ligo.org/reed.essick/monte-carlo-vt> (2023).
- [32] R. Essick *et al.*, gw-distributions, <https://git.ligo.org/reed.essick/gw-distributions> (2023).
- [33] R. Essick, gw-detectors, <https://git.ligo.org/reed.essick/gw-detectors> (2023).
- [34] D. Chatterjee, A. Hegade K. R., G. Holder, D. E. Holz, S. Perkins, K. Yagi, and N. Yunes, Cosmology with Love: Measuring the Hubble constant using neutron star universal relations, *Phys. Rev. D* **104**, 083528 (2021).
- [35] Y. Xie, D. Chatterjee, G. Holder, D. E. Holz, S. Perkins, K. Yagi, and N. Yunes, Breaking bad degeneracies with Love relations: Improving gravitational-wave measurements through universal relations, *Phys. Rev. D* **107**, 043010 (2023).
- [36] R. Abbott *et al.*, Population of merging compact binaries inferred using gravitational waves through GWTC-3, *Phys. Rev. X* **13**, 011048 (2023).
- [37] S. Vitale, Three observational differences for binary black holes detections with second and third generation gravitational-wave detectors, *Phys. Rev. D* **94**, 121501 (2016).
- [38] S. Vitale, D. Brown, and B. Sathyaprakash, On the relative orientation of two cosmic explorers in the us, https://dcc.cosmicexplorer.org/public/0163/T2300002/005/CE_orientations_v5.pdf (2023).
- [39] A. G. Hanselman *et al.*, Gravitational-wave dark siren cosmology systematics from galaxy weighting, [arXiv:2405.14818](https://arxiv.org/abs/2405.14818).