

Identifying type II strongly lensed gravitational-wave images in third-generation gravitational-wave detectors

Yijun Wang¹,* Rico K. L. Lo¹, Alvin K. Y. Li, and Yanbei Chen
California Institute of Technology, Pasadena, California 91125, USA

 (Received 27 January 2021; accepted 16 April 2021; published 24 May 2021; corrected 23 June 2021)

Strong gravitational lensing is a gravitational wave (GW) propagation effect that influences the inferred GW source parameters and the cosmological environment. Identifying strongly lensed GW images is challenging as waveform amplitude magnification is degenerate with a shift in the source intrinsic mass and redshift. However, even in the geometric-optics limit, type II strongly lensed images cannot be fully matched by type I (or unlensed) waveform templates, especially with large binary mass ratios and orbital inclination angles. We propose to use this mismatch to distinguish individual type II images. Using planned noise spectra of Cosmic Explorer, Einstein Telescope and LIGO Voyager, we show that a significant fraction of type II images can be distinguished from unlensed sources, given sufficient SNR (~ 30). Incorporating models on GW source population and lens population, we predict that the yearly detection rate of lensed GW sources with detectable type II images is 172.2, 118.2 and 27.4 for CE, ET and LIGO Voyager, respectively. Among these detectable events, 33.1%, 7.3% and 0.22% will be distinguishable via their type II images with a log Bayes factor larger than 10. We conclude that such distinguishable events are likely to appear in the third-generation detector catalog; our strategy will significantly supplement existing strong lensing search strategies.

DOI: [10.1103/PhysRevD.103.104055](https://doi.org/10.1103/PhysRevD.103.104055)

I. INTRODUCTION

Successful detection of gravitational wave (GW) signals from compact binary mergers by the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) and Virgo collaboration has greatly enriched our understanding of gravity and many aspects of astrophysics [see, e.g., [1,2]]. To extract physical information from detector data, proper signal interpretation is crucial. For this purpose, it is important to study changes in the waveform as it propagates through the universe, since, if unaccounted for, propagation effects can be confused with intrinsic GW features and introduce bias in subsequent analysis. On the other hand, results of propagation effects depend on properties both of the GW and the objects along its path that it interacts with [see, e.g., [3–8]]. Therefore, identifying such signatures also maximizes the scientific output of GW detection.

One GW propagation effect is strong gravitational lensing, in which the rays of a GW are bent strongly enough by a gravitational potential and form multiple images with different magnifications. Gravitational lensing of gravitational waves has attracted enormous interest. It has been estimated that third-generation detectors can detect up to hundreds of strongly lensed events [9–12];

such events can then be used to study cosmological structures [10,13–16] and fundamental physics [17–20].

However, identifying strongly lensed images is challenging since the predominant effect of strong lensing, namely the amplitude magnification by $\sqrt{\mu}$, is degenerate with scaling down the luminosity distance, D_L , by $\sqrt{\mu}$ and keeping the redshifted mass, $M_*(1+z_s)$, constant [see, e.g., [21]], where M_* is the total mass of the binary and z_s is the redshift of this GW source. This degeneracy stems from the fact that general relativity is a scale-free geometric theory, and that GW frequency evolution is unaffected by strong lensing [21].

Current search strategies typically look for multiple events in a catalog that are consistent in intrinsic properties and sky locations, and have orbital phase related in characteristic ways [see, e.g., [22,23]]. For example, this has been used to study the series of events GW170104, GW170814 and a subthreshold trigger, GWC170620, as potential candidates for lensed images [23]. In Ref. [24], the event GW170814 and GW170809 are analyzed as potential strongly lensed companion images using a similar consistency test.

It is also proposed that a sharp transition in the inferred source intrinsic mass distribution at high mass values could single out strongly lensed images [16,25]. This mass distribution anomaly argument, however, must be made in reference to an expected GW source distribution. Currently, such source population models are subject to considerable uncertainties.

* yijunw@caltech.edu

The above strategies share two other drawbacks: (1) without prior knowledge of the lensed source parameters, all pairs of cataloged events must be searched over to find strongly lensed candidates. As detector sensitivity improves and next-generation detectors start observing, the computational cost of such analysis will surge with the increased number of detected events; (2) it is also required that more than one lensed images are detected. If all but one of the images are missed, the methods described above cannot ascertain if a GW image is strongly lensed.

For the above reasons, an intrinsic waveform distortion in a lensed image can be both a more definitive and efficient indicator of strong lensing. If such a lensed image is found, its estimated parameters help narrowing down the search space in the more general pairwise search method mentioned above. An example is the frequency-specific GW diffraction in weak lensing [26,27]. Diffraction signature was searched for in current detected events, but it has yet to be found [22]. It is also predicted that, for GWs within the frequency range of LIGO, diffraction becomes important when the lens mass ranges from 1–100 M_\odot [4,28,29]. Strong lensing by such small lenses requires a small impact parameter, which places stringent requirement on the alignment of the GW source, the lens and the observer [see, e.g., [30]]. Consequently, we expect such events to be rare.

Though diffraction is negligible for strong lensing (within a similar frequency range as LIGO), waveform distortion does occur in the geometric-optics limit when an image originates from a saddle-point solution to the lens equation [see, e.g., [30]]. Such images are called type II images, their waveforms are the Hilbert transforms of the corresponding unlensed waveforms. By contrast, waveforms of type I and type III images are identical to the unlensed waveform, up to a rescaling—and, for type III images, a sign flip. In [15], it is pointed out that type II images are degenerate with type I images with an azimuthal angle shift of $\pi/4$, if only the dominant ($2, \pm 2$) modes are considered. For highly eccentric orbits, this degeneracy is partially lifted. For quasicircular binaries, the degeneracy is also broken if higher multipoles are considered. Recently, Ref. [31] has systematically examined the type II images of a wide range of GW sources, including the effects of spin precession and orbital eccentricity. It was found that the type I/II waveform difference is still small upon tuning the azimuthal angle, the polarization angle and relative phases between GW modes. In this paper, we build upon the work in Ref. [31] by exploring whether such type II images can be distinguished from regular images despite the small waveform mismatch.

For third-generation detectors, such as the LIGO Voyager [32], the Einstein Telescope¹ [33] (ET) and LIGO Cosmic Explorer² [34] (CE) with current models, we expect to be able

to detect a non-trivial number of such distinguishable type II events thanks to the expected high signal-to-noise ratio (SNR).

This paper is organized as follows. In Sec. II we review the geometric optics theory for GW lensing. In Sec. III, we calculate the best-match overlap between type II and type I waveforms over a range of detector-frame binary mass, mass ratio and orbital inclinations. We briefly discuss the implication of waveform mismatch for detection triggering in the current LIGO pipeline framework. In Sec. IV we discuss the distinguishability of type II images in the high-SNR regime by comparing the log likelihoods under type I and type II image hypothesis. Based on this, we compute the fraction of distinguishable type II images. In Sec. V, we incorporate population models on GW sources and lensing galaxies, and predict the expected number of events with distinguishable type II images for LIGO Voyager, ET and CE. We then discuss the results and draw the conclusion.

Throughout this work, we assume a Λ CDM universe with $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$ and a Hubble Constant of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

II. LENS THEORY AND IMAGE TYPE

The geometric optics treatment of gravitational lensing is thoroughly investigated and well established by many authors [see, e.g., [26,30,35]]. In this section, we summarize and discuss scenarios where type II images are distinctive from type I counterparts. We closely follow the discussion in [35] and keep mostly consistent notations.

A. Thin gravitational lens: Geometric-optics limit

We adopt the thin lens model, in which the line-of-sight lens dimension is much smaller than separations between the GW source, the lens and the observer. The source plane and the lens plane are defined by the GW source and the lens center, and both planes are perpendicular to the optical axis connecting the lens center and the observer. All the lens mass is projected onto the lens plane. Lensing deflection to GW paths occurs only on the lens plane.

On each plane, the origin is established as its intersection with the optical axis. The source position has the dimensionless coordinate $\vec{y} = \vec{\eta} D_d / (r_* D_s)$ and the GW path intersects the lens plane at $\vec{x} = \vec{\xi} / r_*$. $\vec{\eta}, \vec{\xi}$ are coordinates with physical units of length, r_* is the lens' Einstein radius, while D_d, D_s are the observer's angular diameter distance to the lens and the source.

The amplitude of the observed image is then expressed as a Kirchhoff integral over the lens plane [see also, e.g., [27,30]],

$$F(\omega, \vec{y}) = \frac{\omega}{2\pi i} \int d^2\vec{x} e^{i\omega t(\vec{x}, \vec{y})}, \quad (1)$$

where ω is the source-frame GW frequency and $t(\vec{x}, \vec{y})$ is the GW travel time difference between lensed paths and the unlensed path,

¹<http://www.et-gw.eu/>.

²<https://cosmicexplorer.org/>.

$$t(\vec{x}, \vec{y}) \approx \frac{1}{2} |\vec{x} - \vec{y}|^2 + t_\Phi, \quad (2)$$

where the first term accounts for the geometrical extra path length in the small deflection limit and the second term, t_Φ , is the Shapiro time delay inside the lens' gravitational potential. In the geometric optics limit, only paths very close to the stationary points of t contribute to the integral, and we may Taylor-expand the time delay around the j th stationary point,

$$t(\vec{x}, \vec{y}) = t(\vec{x}_j, \vec{y}) + \frac{1}{2} dx^a dx^b T_{,ab}(\vec{x}_j, \vec{y}) + \mathcal{O}(|d\vec{x}|^3), \quad (3)$$

where dx^a is a component of the two-dimensional vector $d\vec{x} \equiv \vec{x} - \vec{x}_j$ on the lens plane, and $|,|$ denotes partial derivatives and repeated upper and lower indices imply summation. The integral in Eq. (1) then reduces to two Gaussian integrals after diagonalizing the time delay Jacobian, $T_{,ab}$.

When $\det(T_{,ab}) > 0$, phase shifts from both the ω/i prefactor and the two Gaussian integrals depend on the sign of ω . When $\text{Tr}(T_{,ab}) > 0$, the phase factor is 1, giving type I images. When $\text{Tr}(T_{,ab}) < 0$, the phase shift is $-\text{sgn}(\omega)\pi$, where the function sgn returns the sign of its argument. This phase shift gives type III images, which differ from type I by an overall phase of π . (Note that $\pm\pi$ phases are equivalent.)

When $\det(T_{,ab}) < 0$, the two Gaussian integrals give opposite phase shifts regardless the sign of ω , and no longer contribute to the overall phase of $F(\omega, \vec{y})$. The overall phase shift is then $-\text{sgn}(\omega)\pi/2$, giving type II images which are equivalent to a Hilbert transform of type I images.

B. Gravitational waves from circular, nonspinning binaries

For compact binaries, the complex GW strain at infinity can be written as

$$h = h_+ - ih_\times = \sum_{l,m} {}_{-2}Y_{lm}(\iota, \phi) h_{lm}, \quad (4)$$

where the subscripts $+, \times$ denote plus and cross polarizations, and ${}_{-2}Y_{lm}(\iota, \phi)$ is the $s = -2$ spin-weighted spherical harmonics. For non-spinning binaries with quasicircular orbits, we choose the coordinate system such that the orbital angular momentum is along the z axis. In this way, arguments ι and ϕ of the spin-weighted spherical harmonic also corresponds to the orbital inclination angle and the azimuthal angle, respectively.

Let us start out by considering $m \neq 0$ modes. The contribution from modes with $m = \pm m_0$, where m_0 is a positive integer, is

$$\tilde{h}_{I,m_0} = \sum_l \sum_{m=\pm m_0} {}_{-2}Y_{lm}(\iota, \phi) \tilde{h}_{I,lm}, \quad (5)$$

where the subscript I denotes the regular type I waveforms. The quantity $\tilde{h}_{I,lm}$ is the Fourier transform of h_{lm} in Eq. (4) via

$$\tilde{h}_{I,lm}(f) = \int_{-\infty}^{\infty} h_{lm}(t) e^{-2\pi i f t} dt. \quad (6)$$

We note that ϕ appears only in the factor of $\exp(im\phi)$ in ${}_{-2}Y_{lm}(\iota, \phi)$. Furthermore, for nonspinning, circular binaries, with orbital angular momentum along the z axis, in frequency domain, $m > 0$ modes only have negative frequency components and the inverse is true for $m < 0$ modes. Therefore, the Hilbert transform of \tilde{h}_{I,m_0} , \tilde{h}_{II,m_0} is written as

$$\begin{aligned} \tilde{h}_{II,m_0}(\iota, \phi) &= -i \text{sgn}(f) \tilde{h}_{I,m_0}(\iota, \phi) \\ &= \tilde{h}_{I,m_0}\left(\iota, \phi + \frac{\pi}{2m_0}\right). \end{aligned} \quad (7)$$

Therefore, for each subset of GW modes with $m = \pm m_0$, the Hilbert transform is degenerate with an additional orbital azimuthal angle $\Delta\phi = \pi/(2m_0)$. For example, the required angle change is $\Delta\phi = \pi/4$, provided that only the $(l, \pm 2)$ GW modes are considered. Modes with different $|m|$ require different angle changes to compensate for the Hilbert transform (e.g., the $(l, \pm 3)$ modes require $\Delta\phi = \pi/6$). This difference in the compensation requirements breaks the degeneracy between Hilbert-transformed signals and orbital azimuthal angle change.

Physically, $|m| \neq 2$ modes can be significant when the orbit is significantly eccentric [15]. For binaries with significant mass ratios and inclination angles, the $(3, \pm 3)$ modes become significant, breaking degeneracy. Figure 1 is analogous to Fig. 2 in [15] and plots example type I/II waveforms from a binary with a detector-frame mass $\tilde{M} = 150 M_\odot$, a mass ratio $q = 2.2$ and an orbital inclination angle $\iota = 80$ deg. The binary is nonspinning in a quasicircular orbit, and all multipoles with $l \leq 4$ are included. The top two panels show that the type II image is not degenerate with the type I image with an additional time shift. The bottom panels show that, when we include only the $m = \pm m_0$ modes, the Hilbert transform is degenerate with the original waveform with $\Delta\phi = \pi/(2m_0)$.

For $m = 0$ modes, h is independent from ϕ , and one cannot recover its Hilbert transform via shifting ϕ . This in principle further breaks the degeneracy, although $m = 0$ modes are generally weak for nonspinning binaries in circular orbits. However, note that these are where the GW memory effects take place [36–38].

In this paper, we systematically explore GW sources which are nonspinning binary black holes in quasicircular

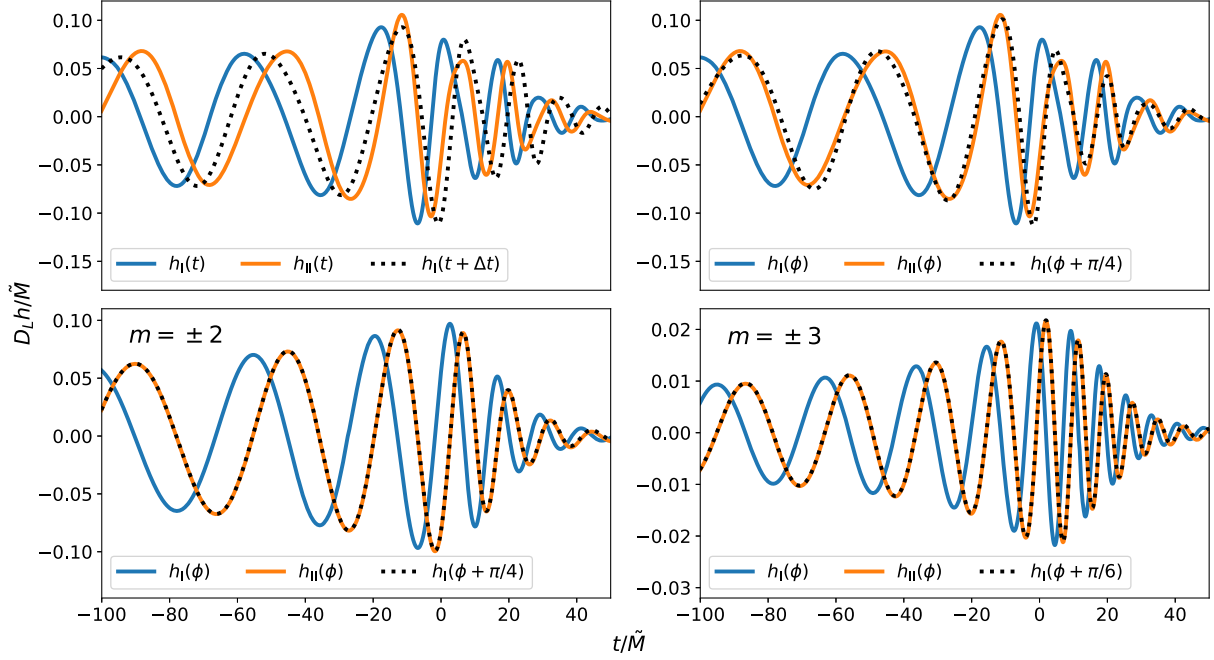


FIG. 1. Type I/II NRsur7dq4 surrogate model waveforms from a binary with $\tilde{M} = 150 M_\odot$, $q = 2.2$, $\iota = 80^\circ$. The binary is nonspinning in a quasicircular orbit. The black dotted line shows the type I waveform with a $\pi/4$ shift in the orbital azimuthal angle, and the shifted waveform is completely degenerate with the type II image. The orange dotted line shows the type II waveform, such that its peak overlaps with that of the type I waveform. We observe that the type I/II waveform offset cannot be compensated by a time shift.

orbits. The distinguishable signature of type II images will be due to higher order GW modes, which is related to binary mass ratio, q , and orbital inclination, ι .

III. WAVEFORM MISMATCH

In this section, we quantify the mismatch between type I/II waveforms for nonspinning binaries, in preparation for discussion on their distinguishability in the next section. We also discuss the implication of this mismatch for the GW signal veto process, namely, whether the mismatch leads to type II signal rejection in the current LIGO data analysis pipeline.

A. Best-match overlap

In this section, we describe the procedure to compute the type I/II waveform difference over a large parameter space. We model only nonspinning binaries in quasicircular orbits. Highly spinning binaries or those with highly eccentric orbits are expected to be fewer than the population we consider [see, e.g., [39,40]]. Since, the optical depth for type II images is also small, on the order of 10^{-3} – 10^{-4} [10,11,21], we exclude these less frequent sources from our analysis.

For this source population, frequency-domain GW strain is given by the Fourier transform of Eq. (4),

$$\tilde{h}_1(f) = \sum_{l,m} {}_{-2}Y_{lm}(\iota, \phi) \frac{\tilde{H}_{1,lm}(\tilde{M}, q, f)}{D_L} e^{-2\pi i f t_0 - i\Phi}, \quad (8)$$

where $\tilde{H}_{1,lm}(\tilde{M}, q, f)/D_L$ is equal to $\tilde{h}_{1,lm}(f)$ in Eq. (6), with the dependence on D_L explicitly shown. The waveform is a function of the detector-frame mass (or equivalently, the redshifted mass), $\tilde{M} = (1+z)M$ (where M is the intrinsic mass), the mass ratio, $q \equiv \tilde{M}_1/\tilde{M}_2 \geq 1$ ($\tilde{M}_1 + \tilde{M}_2 = \tilde{M}$), and the luminosity distance, D_L . The polarization angle, Φ , and signal time-of-arrival, t_0 , add additional phase shifts to the signal.

For any two waveforms, \tilde{h}_1, \tilde{h}_2 , we define the overlap by

$$\text{overlap} = \frac{\Re(\langle \tilde{h}_1 | \tilde{h}_2 \rangle)}{\sqrt{\langle \tilde{h}_1 | \tilde{h}_1 \rangle \langle \tilde{h}_2 | \tilde{h}_2 \rangle}} = 1 - \epsilon, \quad (9)$$

where ϵ is the mismatch and $\langle \cdot \rangle$ denotes inner product given by

$$\langle a | b \rangle = \int_{-\infty}^{\infty} \frac{a^*(f)b(f)}{S_n(f)} df, \quad (10)$$

where $S_n(f)$ is the two-sided noise power spectral density. By applying the optimal matched filter, the SNR of \tilde{h} , ρ , is given by $\sqrt{\langle \tilde{h} | \tilde{h} \rangle}$.

Throughout this paper, we use Roman numeral subscripts to denote the image types and Arabic numeral subscripts to represent any individual waveform. We also adopt the simplifying assumption that both GW polarizations can be independently detected, i.e., the time-domain

waveform is taken to be complex, as in Eq. (4). In Sec. VI, we discuss in more detail the validity of this assumption.

To obtain highly accurate models for \tilde{h} , we adopt the time-domain numerical relativity surrogate waveform model, NRSur7dq4, [41] extracted through the PYTHON package, GWSURROGATE [42]. This surrogate model provides all $l \leq 4$ mode waveforms, $h_{l,m}(t)$, through the inspiral, merger and ringdown phases.

To avoid spurious edge effects due to the finite-length of surrogate waveforms, we apply a time-domain kaiser window function from `numpy.kaiser` [43] with $\beta = 4$. The window is centered at the waveform amplitude peak to maximally preserve waveform features. The signal is zero-padded prior to the Fourier transform to ensure sufficiently smooth transformed waveform.

To maximize the overlap, Ref. [31] separately tunes the azimuthal angle, polarization angle, and the relative phases between each of the GW multipoles. We adopt a different approach by tuning the intrinsic parameters of the GW sources. We adopt a nested search method. We first generate a type II signal template, $\tilde{h}_{\text{II},0}$, with $(\tilde{M}_0, q_0, t_0, \phi_0)$. Since the waveform amplitude scaling does not contribute to the overlap, we fix $D_L = 3$ Gpc for all waveforms. We make a (\tilde{M}, q) grid, with the mass range centered on \tilde{M}_0 and mass ratio between 1 and 4 (the range of q used to train the surrogate model). At each grid point, we construct the type I template and use the PYTHON module `scipy.optimize.dual_annealing` [44] to find the (ι, ϕ) that maximize the overlap between the type I template and the type II target. The spin-weighted spherical harmonics are computed using the PYTHON package `spherical_functions`³ and `quaternion`.⁴ To implicitly maximize over t_0 and Φ , we take the Fourier transform of the integrand in Eq. (9) and pick the element with the largest absolute value [see, e.g., [45]]:

$$(1 - \epsilon)_{\max} = \max_{t_0} \left| \frac{1}{a} \int_{-\infty}^{\infty} df \frac{\tilde{h}_I^*(f) \tilde{h}_{\text{II}}(f)}{S_n(f)} e^{-2\pi i f t_0} \right|, \quad (11)$$

$$a = \sqrt{\langle \tilde{h}_I | \tilde{h}_I \rangle \langle \tilde{h}_{\text{II}} | \tilde{h}_{\text{II}} \rangle}.$$

Figure 2 shows an example maximization result contour plot for a type II signal with $\tilde{M} = 150 M_\odot$, $q = 1.7$ and $\iota = 70$ deg with CE noise curve. Due to the waveform mismatch, the best-match template has different parameter values from those of the true signal, with a maximal overlap of 99.06%.

We calculate the overlap for sample points on the grid $X(\tilde{M}) \otimes Y(q) \otimes Z(\iota)$, with

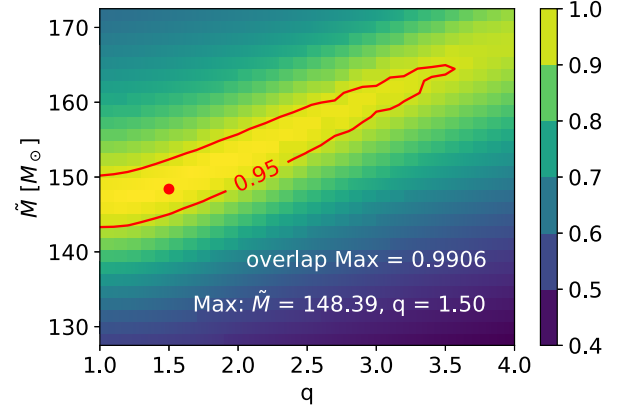


FIG. 2. Contour plot for maximized overlap for a type II waveform with $\tilde{M} = 150 M_\odot$, $q = 1.7$ and $\iota = 70$ deg. Grid point with the maximum overlap is shown with the red dot at $\tilde{M} = 148.39 M_\odot$, $q = 1.50$. 95% overlap contour is shown in red.

$$X(\tilde{M}) = \{60, 80, 100, 150, 200, 230, 260, \\ 300, 400, 500, 600, 700, 800\} [M_\odot],$$

$$Y(q) = \{1.2, 1.7, 2.2, 2.7, 3.2\},$$

$$Z(\iota) = \{15, 30, 40, 50, 60, 70, 80\} [\text{deg}].$$

We then interpolate between the samples using the `scipy.interpolate` module [44] to construct a function $\epsilon(\tilde{M}, q, \iota)$. For nonspinning binaries, the interpolated function ensures that $90 \text{ deg} < \iota < 180 \text{ deg}$ is symmetric to $0 \text{ deg} < \iota < 90 \text{ deg}$. We perform the same analyses for CE, ET, and LIGO Voyager with their respective noise power spectral density (PSD) [32–34,46].

Figure 3 shows the amplitude spectral density of CE, ET, and LIGO Voyager [32–34,46], as well as the waveform of a

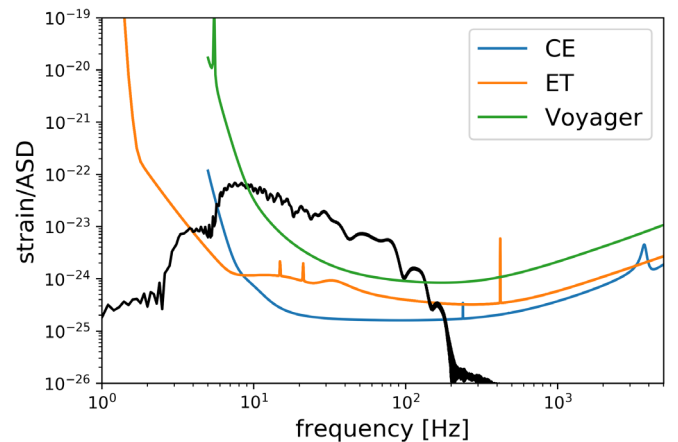


FIG. 3. Positive frequency band waveform for a binary with $\tilde{M} = 200 M_\odot$, $q = 2.2$, $\iota = 80$ deg and $D_L = 1$ Gpc, plotted in black. The amplitude spectral densities (ASDs) for CE, ET and LIGO Voyager are plotted with colored traces. Note that ASDs for CE and LIGO Voyager are available starting from 3 Hz.

³https://github.com/moble/spherical_functions.

⁴<https://github.com/moble/quaternion>.

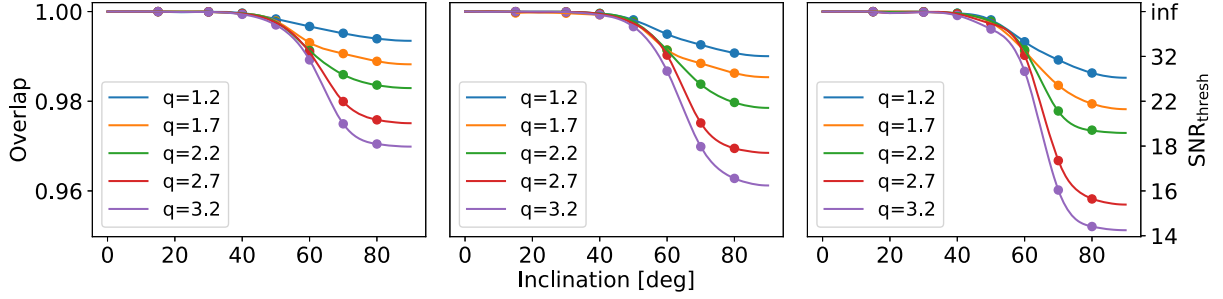


FIG. 4. Overlap between type I and type II waveforms for $\tilde{M} = 150 M_\odot$ at selected mass ratio values. The axis on the right shows the threshold ρ to distinguish such type II images from type I counterparts by a log Bayes factor of 10, for the corresponding waveform overlap value. See Section IV for details. Panels from left to right are overlaps for CE, ET, and LIGO Voyager respectively. In all panels, data points are shown with dots, and the interpolated overlap functions are shown in smooth curves.

binary with $\tilde{M} = 200 M_\odot$, $q = 2.2$, $\iota = 80$ deg at $D_L = 1$ Gpc as an example. The low-frequency amplitude loss of the surrogate waveform is due to the finite length of the NRSur7dq4 waveforms. For less massive binaries, this effect results in significant loss of ρ , especially in the case of CE, where the low-frequency sensitivity degrades slower.

To estimate how this ρ loss affects the overlap values, we compute the maximum overlap for a ($\tilde{M}_0 = 60 M_\odot$, $q = 3$, $\iota = 80$ deg) binary with the CE PSD, filtering all frequency components below 30 Hz, where the loss of ρ becomes significant. Compared with the unfiltered case ($(1 - \epsilon)_{\max} = 0.981$), the overlap decreases only by 3.3×10^{-3} . Since $\tilde{M} = 60 M_\odot$, $q = 3$ and $\iota = 80$ deg are roughly the smallest redshifted mass, largest mass ratio and inclination we consider, other binaries within our parameter space should have a smaller loss of the overlap. Considering the small size of the difference, we do not filter signals in subsequent analysis.

Figure 4 shows the best-match overlap for GW waveforms with a redshifted mass of $150 M_\odot$ for the three GW detectors at selected mass ratio values. Maximization data points are shown with solid dots, and the interpolation functions are shown as smooth curves. The right axis shows the required ρ to distinguish type I/II waveforms with a log Bayes factor of 10 at the corresponding overlap values on the left axis. See discussion in Sec. IV. Consistent with intuition, the best-match overlap is the lowest for high mass-ratio signals at large inclinations. Over our parameter space, the mismatch value for such signals is typically on the order of 2%. We note that the same type II waveforms have the largest mismatch with type I waveforms in LIGO Voyager, as the LIGO Voyager PSD emphasizes high-frequency waveform components, where the Hilbert transform effect is more pronounced.

B. Signal veto

An ensuing concern from the mismatch is whether the difference in waveforms could lead to type II signals vetoed or assigned a lower significance value during observing runs. For the current GW data analysis pipelines,

once a threshold ρ is reached, the data typically go through a χ^2 veto test to screen out spurious signals. In this section, we calculate the noncentral parameter in the χ^2 statistic distribution from using type I templates to match type II signals.

The χ^2 veto was described in detail in [47]. This test characterizes the distribution of ρ over frequency bins and vetoes detector “glitches,” or loud bursts of non-Gaussian noise that might have a high ρ , but have a frequency distribution very different from that of a genuine GW signal.

Suppose the best-match template to the signal, $\tilde{n} + \tilde{h}_0$ is \tilde{h}_T , where \tilde{n} is noise and \tilde{h}_0 is the embedded waveform. We divide the detector sensitive frequency range into p disjoint sub-bands, Δf_j , such that the template ρ in each bin is $1/p$ of its total ρ ,

$$\rho_{T,j} = \int_{-\Delta f_j, \Delta f_j} \frac{|\tilde{h}_T|^2}{S_n(f)} df = \frac{1}{p} \int_{-\infty}^{\infty} \frac{|\tilde{h}_T|^2}{S_n(f)} df. \quad (12)$$

We then calculate the signal ρ in each frequency bin as:

$$s_j \equiv \frac{1}{\sqrt{\langle \tilde{h}_T | \tilde{h}_T \rangle}} \int_{-\Delta f_j, \Delta f_j} \frac{\tilde{h}_T^*(\tilde{n} + \tilde{h}_0)}{S_n(f)} df. \quad (13)$$

We then define the χ^2 statistic as

$$\chi^2 \equiv p \sum_{i=1}^p (s_i - s/p)^2, \quad s \equiv \sum_{j=1}^p s_j. \quad (14)$$

In the case where the best-match template in the template bank does not exactly match the embedded waveform, the distribution of χ^2 over many Gaussian noise realizations is a classical χ^2 distribution with a noncentral parameter,

$$\langle \chi^2 \rangle = p - 1 + \kappa \langle s \rangle^2, \quad (15)$$

where $\langle \cdot \rangle$ denotes the average over noise realizations. The factor, κ , in the noncentral parameter is bound by

$$0 < \kappa < \frac{1}{(1 - \epsilon)^2} - 1 \approx 2\epsilon, \quad (16)$$

where ϵ is the minimized mismatch between the template and the underlying waveform, as is defined in Eq. (11). The approximate equality is satisfied when $\epsilon \ll 1$. This bound is agnostic of the specific waveform of the signal and templates. Consequently, the noncentral parameter introduced by using type I templates on type II signals is smaller than $0.12\langle s \rangle^2$ in most cases, if we take the largest mismatch to be 6%. If such a noncentral parameter lies within the χ^2 threshold during detection, type II images are unlikely to be vetoed.

C. Type II signal recovery

There have been ongoing efforts to look for possible weaker (sub-threshold) strongly lensed counterparts of confirmed GW detections, assuming the latter being strongly lensed signals themselves [48,49]. One method is to simulate lensed injections of a super-threshold GW event, then use a generic template bank to search for these injections through an injection run, and produce a targeted template bank for searching possible lensed counterparts of the target event by retaining only templates that can find the injections.

However, only type I lensed images have been considered for current searches. The question we would like to investigate is: Should type II lensed images be present in the data, would a type I template bank be able to find them? The answer to this question may be a crucial step for us to identify possible lensed GWs that we might have already detected but still not being discovered.

As a preliminary test to this question, we apply the search method detailed in [48] to the high-mass-ratio compact binary coalescence event GW190814 [50]. Using the waveform approximant IMRPhenomXPHM [51], we generate a set of simulated lensed injections for GW190814. They are then injected into real LIGO-Virgo data in two ways: (1) by treating them as type I images, and (2) by treating them as type II images, i.e., applying Hilbert transform to the waveform in the frequency domain as discussed previously. Through the GW CBC search pipeline GstLAL [52], we apply the previously generated type I image target bank to search for these injections in both tests, and finally we compare the number of missed injections to roughly estimate the effectiveness of a type I image bank to look for type II images.

As discussed in [48,52], each GW candidate found in the GstLAL search will be assigned a log likelihood ratio statistic $\ln \mathcal{L}$ to measure its significance. The false-alarm-rate (FAR) can be calculated accordingly, which corresponds to how often noise will produce a trigger with ranking statistic $\ln \mathcal{L}$ larger or equal to the ranking statistic $\ln \mathcal{L}^*$ of the trigger we are considering. In the search, an injection is said to be found if its FAR passes the usual

threshold 1 in 30 days, as usual for a generic gravitational-wave search [53].

In both tests, we have injected a total of 8036 simulated lensed injections. We assume that the injected events are registered by both detectors in the aLIGO network and the Virgo detector. In test **A**, we apply a type I image bank to look for injected type II images. For test **B**, we use the same image bank and look for the type I counterpart of the injections in test **A**. In test **A**, 638 injections are missed, whereas in test **B** the missed count is 536. We observe that the number of missed injections increases when the injections were treated as type II images, indicating that the current search method for subthreshold lensed GWs may be missing possible type II lensed signals.

However, it is important to remark that our current results are inconclusive since: (1) we have only been testing on one particular GW event, and (2) the exact reason for the extra number of injections to be missed are yet to be investigated. Nevertheless, our results indicate there could be improvements to the current search method for subthreshold lensed GW signals, and further investigation will be done as future work.

IV. DISTINGUISHING TYPE II EVENTS

While we have systematically examined the type I/II waveform mismatch, whether it enables us to distinguish type II images in actual GW experiments deserves further discussion. In this section, we use the waveform overlap and quantify the fraction of strongly lensed GW sources that have distinguishable type II images.

A. Bayes factor

Using a Bayesian model (or equivalently hypothesis) selection framework, we quantify the distinguishability between a type I image and a type II image by computing the Bayes factor \mathcal{B} , which is the ratio of the probability of observing the data \vec{d} under the hypothesis that the signal is of type II over that under the hypothesis that the signal is of type I, namely

$$\begin{aligned} \mathcal{B} &= \frac{p(\vec{d}|\text{type II image})}{p(\vec{d}|\text{type I image})} \\ &= \frac{\int d\vec{\theta} \mathcal{L}(\vec{\theta}|\text{type II image}) \pi(\vec{\theta}|\text{type II image})}{\int d\vec{\theta} \mathcal{L}(\vec{\theta}|\text{type I image}) \pi(\vec{\theta}|\text{type I image})}, \quad (17) \end{aligned}$$

where $\mathcal{L}(\vec{\theta})$ is the (Whittle) likelihood as a function of the waveform parameters $\vec{\theta}$, and $\pi(\vec{\theta})$ is the prior distribution, which is different under the two hypotheses. The log likelihood function, up to a normalization constant, is given by

$$\begin{aligned} \ln \mathcal{L}_i(\vec{\theta}) &\propto -\frac{1}{2} \langle d - h_i(\vec{\theta}) | d - h_i(\vec{\theta}) \rangle \\ &\propto -\frac{1}{2} \langle d | d \rangle + \langle d | h_i \rangle - \frac{1}{2} \langle h_i | h_i \rangle, \end{aligned} \quad (18)$$

where the subscript $i = \text{I, II}$ denotes the assumed image type. In an actual inference analysis, we do not know *a priori* the “true” waveform parameters. Therefore, we usually evaluate the integrals in Eq. (17) using a sampling algorithm that explores the parameter space spanned by $\vec{\theta}$ stochastically.

Still, we can give an analytical approximate of the Bayes factor for distinguishing a type II image from a type I image using only the SNR ρ and the mismatch ϵ we calculated in Sec. III. Following the treatment in Refs. [54–56], with the Laplace approximation we can write the log Bayes factor as

$$\ln \mathcal{B} \approx \ln \left[\frac{\mathcal{L}_{\text{II}}(\vec{\theta}_{\text{MLE}})}{\mathcal{L}_{\text{I}}(\vec{\theta}_{\text{MLE}})} \right] + \ln \left(\frac{\sigma_{\text{II}}^{\text{posterior}}}{\sigma_{\text{I}}^{\text{posterior}}} \right), \quad (19)$$

where $\sigma_i^{\text{posterior}}$ is the posterior (uncertainty) volume assuming that the lensed GW is of type- i . The log likelihood ratio in Eq. (19) can be shown [56], in the high SNR limit, to scale as

$$\ln \left[\frac{\mathcal{L}_{\text{II}}(\vec{\theta}_{\text{MLE}})}{\mathcal{L}_{\text{I}}(\vec{\theta}_{\text{MLE}})} \right] \approx \epsilon \rho^2, \quad (20)$$

when the (minimized) mismatch $\epsilon \ll 1$. If we ignore the correlation between the parameters, we can estimate the posterior volume $\sigma_i^{\text{posterior}}$ roughly as

$$\sigma_i^{\text{posterior}} \approx \prod_{j=1}^N \sqrt{2\pi} \Delta \theta_i^{j,\text{posterior}}, \quad (21)$$

with j loops over the N -dimensional vector $\vec{\theta}$ and $\Delta \theta_i^{j,\text{posterior}}$ is the uncertainty of the 1D marginal posterior distribution for θ_j assuming that the image is of type- i . Note that here we assumed that identical prior was used when calculating the Bayes factor, except for the image type. The posterior volume ratio also scales with the mismatch, actually. Since $\langle h_{\text{I}} | h_{\text{I}} \rangle \approx (1 - \epsilon)^2 \langle h_{\text{II}} | h_{\text{II}} \rangle$ and that $\Delta \theta_i^{j,\text{posterior}} \propto 1 / \sqrt{\langle h_{\text{I}} | h_{\text{I}} \rangle}$, therefore we have

$$\ln \left(\frac{\sigma_{\text{II}}^{\text{posterior}}}{\sigma_{\text{I}}^{\text{posterior}}} \right) \approx -N\epsilon. \quad (22)$$

Indeed, in the high SNR limit, the first term in Eq. (19) is much larger than the second term as $N \sim 10$ and $\epsilon \ll 1$. Hence, we will ignore the contribution from the log

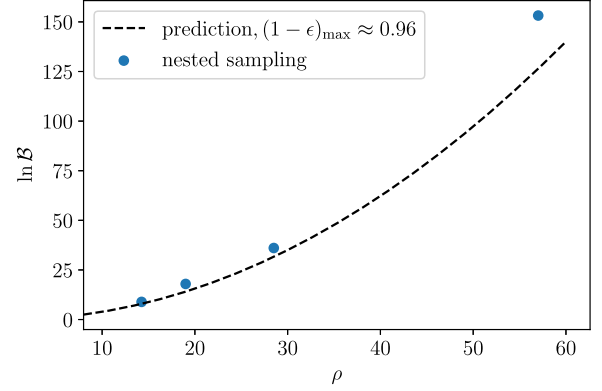


FIG. 5. The log Bayes factor $\ln \mathcal{B}$ as a function of the SNR ρ of the injections with different luminosity distances and fixed mismatch ϵ , computed using Eq. (17) with nested sampling and Eq. (23). We see that the simulation results roughly follow the expected quadratic scaling with the SNR.

posterior volume ratio in this paper. Therefore, we can estimate the log Bayes factor simply as⁵

$$\ln \mathcal{B} \approx \epsilon \rho^2. \quad (23)$$

Figure 5 shows the log Bayes factor as a function of the SNR ρ using nested sampling with the help of the library *bilby* [57] and *dynesty* [58] as in Eq. (17), as well as its approximate using only the optimal SNR and the mismatch using Eq. (23). Here we use the IMRPhenomXHM waveform model [59] for both the simulated signals and the inference. All simulated signals have a redshifted total mass of $\tilde{M} = 150 M_{\odot}$, $q = 3.2$ viewing at an inclination angle of $\iota = 80$ deg with different luminosity distances to adjust the optimal SNR. We see that the simulation results roughly follow the expected quadratic scaling with the optimal SNR. Indeed, by performing a least-squares fit we found that the exponent is 2.02 ± 0.07 .

Since $\ln \mathcal{B}$ scales as SNR^2 , even a small type I/II mismatch could lead to significant $\ln \mathcal{B}$ in the high-SNR regime. For instance, for a mismatch of 3%, an SNR of 20 would yield a log Bayes factor larger than 10, favoring the type II waveform hypothesis, thereby identifying this event as a strongly lensed image regardless whether other images are detected. The right axis in Fig. 4 shows the required SNR to produce $\ln \mathcal{B} = 10$ for the corresponding type I/II overlap values. While such SNR is high for the current aLIGO, for third-generation detectors, it occurs frequently. For example, an equal-mass binary with a detector frame total mass of $100 M_{\odot}$ at $D_L = 8$ Gpc has an $\rho = 30$ for LIGO Voyager. The same source with $D_L = 17$ Gpc has $\rho = 131$ for CE.

⁵Note that posterior volume also depends on dependences of h_{I} and h_{II} on θ_j .

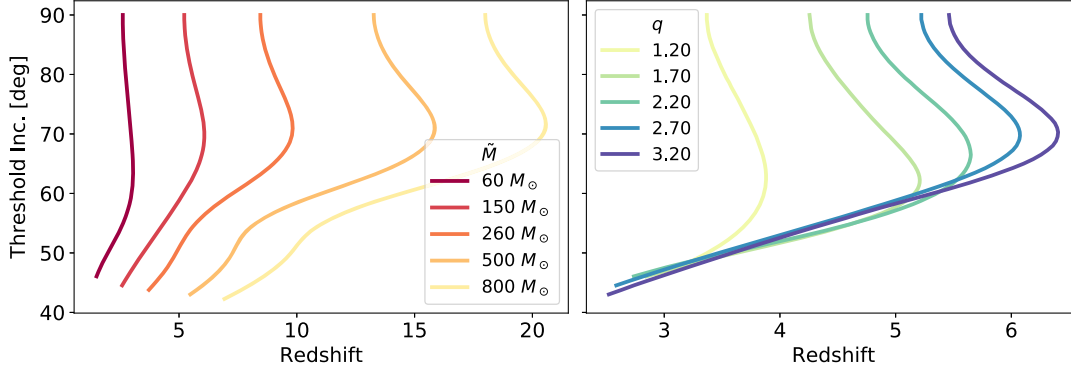


FIG. 6. Inclination threshold curves for distinguishable type II sources as a function of redshift assuming CE sensitivity. Left: inclination threshold curves for fixed mass ratio $q = 2.67$ at selected redshifted mass values. Right: inclination threshold curves for binaries with redshifted mass $\tilde{M} = 150 M_\odot$ with selected mass ratio values. The curve-crossing at low inclination values are due to systematic errors; see text for discussion.

B. Threshold inclination

In this section, we find the range of parameters, (\tilde{M}, q, z_s) , where type II images can be distinguished via the log Bayes factor test. We choose $\ln \mathcal{B}_{\text{thresh}} = 10$ as the criterion for distinguishability.

We begin by computing the distinguishable threshold inclination, ι , for sources with certain redshifted mass, mass ratio and redshift. Since both ρ and ϵ in Eq. (23) depend on ι , it is more straightforward to first fix \tilde{M} , q , and ι to obtain ϵ , and then scale ρ via D_L to achieve the $\ln \mathcal{B}_{\text{thresh}}$ condition. Inverting $D_{L,\text{thresh}}(\tilde{M}, q, \iota)$ yields $\iota_{\text{thresh}}(\tilde{M}, q, D_L(z_s))$, where z_s is the GW source redshift.

To calculate ρ , we assume both GW polarizations can be detected, and the total amplitude is $\sqrt{h_+^2 + h_\times^2}$. In the Discussion section, we further discuss the justifications for this assumption in the context of third-generation GW detectors. However, the finite length of the surrogate model waveform can lead to significant loss in ρ , even though the effect on waveform overlap is negligible, as demonstrated in Sec. III. For a binary with $\tilde{M} = 60 M_\odot$, $q = 3$, approximately 15% of ρ is lost in the case of CE. For LIGO Voyager, the noise increase starts earlier and steeper toward lower frequencies; consequently, the ρ loss for the same binary is only $\sim 5\%$. To accurately estimate ρ , we supplement the surrogate model waveform with analytical inspiral stage waveform, whose amplitude scales as $f^{-7/6}$ [60]. The inspiral amplitude is matched to the surrogate waveform amplitude at $0.5f_{\text{ISCO}}$, where f_{ISCO} is the innermost stable circular orbit frequency, approximated as [see, e.g., [61]]

$$f_{\text{ISCO}} = \frac{c^3}{6^{3/2} \pi G \tilde{M}}. \quad (24)$$

We note that, by compensating for the lost ρ , our result is optimistic in estimating the distinguishability; while the early inspiral phase contributes significantly to ρ , the type II waveform mismatch is less pronounced. For the same ρ ,

the compensated inspiral waveform does not offer as much information as the higher frequency GW phases for distinguishing type II images. Nonetheless, this overestimate is significant only for systems toward the low mass limit, where the expected detectable number of events is low due to the small ρ .

The mismatch ϵ is available from the interpolation function in Sec. III. We do not consider binaries with best-match overlap larger than 0.999, i.e., we consider such mismatch a result of systematic errors and does not reflect actual waveform difference. As discussed in Sec. III, the truncated surrogate waveform leads to errors in the best-match overlap, though for high mass systems, the error will be much smaller than 3.3×10^{-3} for the $\tilde{M} = 60 M_\odot$, $q = 3$, $\iota = 80$ deg example binary. For computational cost concerns, we also limit the grid density in the nested maximization process. If the actual best-match binary is not on the grid points, the maximization result will deviate from the true value, and the size of the deviation depends on the distance between the true best-match and its closest grid point. Aside from systematic errors in the waveform and overlap optimization process, interpolation for ϵ also introduces errors. In particular, the cubic spline fit may introduce spurious trace curves to guarantee smoothness when connecting the limited number of samples. Especially in the case of CE, ρ can be very large, thus exaggerating the physical significance of such a small mismatch. The exact value of this threshold is tuned to exclude spurious interpolation function results. In the next subsection, we discuss our choice of the upper limit value for the best-match overlap and assess the impact of this mismatch resolution in the next subsection.

Figure 6 shows the threshold inclination as a function of source redshift assuming CE sensitivity. The left panel shows threshold inclination with fixed redshifted mass $\tilde{M} = 150 M_\odot$. We observe that the mass ratio becomes an increasingly important factor at high inclinations. At low redshift, the threshold inclination is constrained primarily

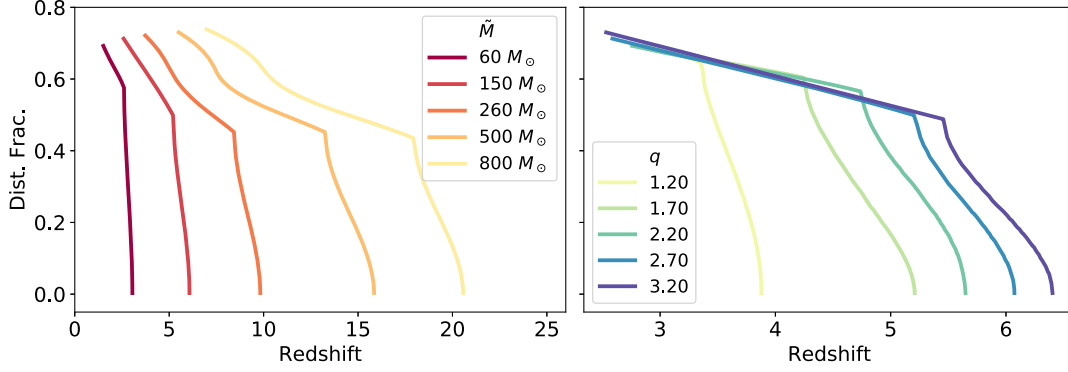


FIG. 7. The fraction of distinguishable type II images as a function of redshift for CE sensitivity. Right: distinguishable fraction, $\text{fr}(\tilde{M}, q, z_s)$, for constant $\tilde{M} = 150 M_\odot$. Left: distinguishable fraction for constant $q = 2.67$. The cusps in the fraction correspond to the exclusion of high-inclination binaries with sub-threshold ρ . The fraction curves directly correspond to the threshold inclination curves in Fig. 6.

by the mismatch ϵ ; at higher redshift (e.g., $z_s \sim 4.2$ for $q = 1.73$), the high inclination regions start to be excluded despite the large mismatch value, as ρ becomes too small. Beyond a certain redshift (e.g., $z_s \sim 5.2$ for $q = 1.73$), no combination of ρ and ϵ meets the $\ln \mathcal{B}_{\text{threshold}}$ condition, and no more type II images can be distinguishable. The right panel shows similar threshold curves fixing the mass ratio to be 2.67. We observe a similar curve shape, although lighter binaries have smaller ρ and consequently a larger threshold inclination.

C. Distinguishable image fraction

From the threshold inclination, we can further calculate the fraction of GW sources with distinguishable type II images, $\text{fr}(\tilde{M}, q, z_s)$. For simplicity, we assume that GW

sources and type II images are isotropically distributed, therefore the fraction of distinguishable type II images scales as the area of the celestial sphere within the ι threshold limits. The differential fraction is then proportional to $\sin \iota$. Figure 7 shows the distinguishable fraction of type II sources for the same binaries as in Fig. 6. The cusps mark the redshift when high inclination regions start to be excluded due to smaller ρ . We observe that, for CE, large fractions of sources with type II images can be distinguished via the log Bayes factor test out to high redshift. Similar plots for ET and LIGO Voyager are shown in Fig. 8 as dashed lines.

We have so far considered only the redshifted mass (detector-frame mass), \tilde{M} , as it is the direct input to the surrogate model, which assumes an asymptotically flat and

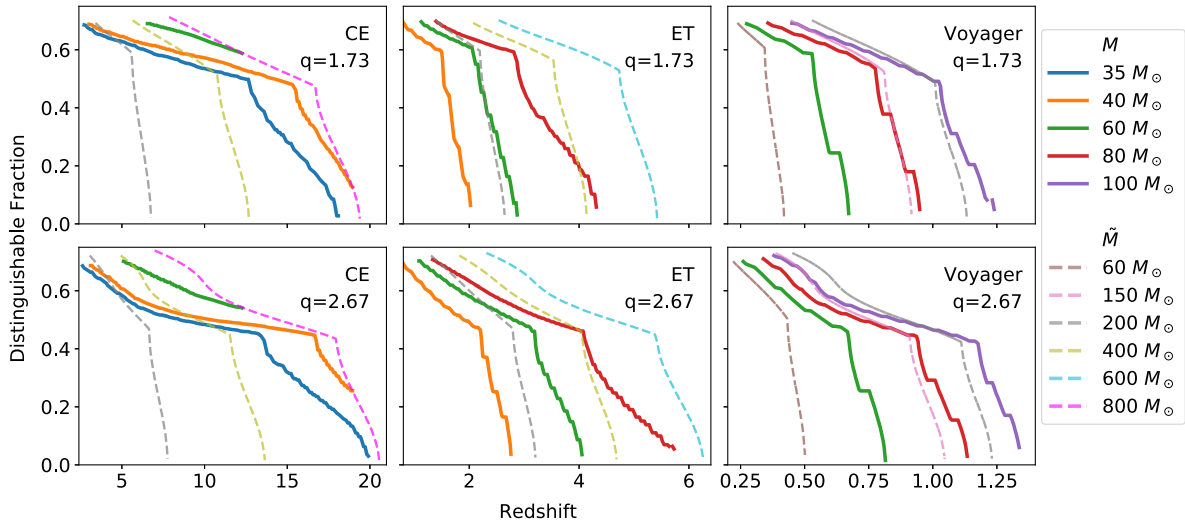


FIG. 8. The fraction of distinguishable type II images as a function of redshift. Top row: GW sources have constant mass ratio of 1.73. Bottom row: GW sources have constant mass ratio of 2.67. The Left, Middle, and Right columns show distinguishable fractions assuming CE, ET, and LIGO Voyager sensitivity, respectively. The fractions for fixed redshifted mass, $\text{fr}(\tilde{M}, q, z_s)$, are shown in dashed lines, and those for fixed apparent mass, $\text{fr}_{\text{app}}(M, q, z_s)$, are shown in solid lines. The mass values are shown in the legend. Note that the fraction curves are jagged due to interpolation errors and limited data density.

stationary universe. The “apparent” total mass of the binary, M , is related to the redshifted mass by $M = \tilde{M}/(1 + z_s)$. Due to lensing magnification, this inferred apparent total mass could be larger or smaller than the actual GW source total mass. We discuss magnification effects in Sec. V. Therefore, the fraction of distinguishable type II sources with apparent mass M , mass ratio q at redshift z_s is given by,

$$\text{fr}_{\text{app}}(M, q, z_s) = \text{fr}(\tilde{M}(1 + z_s), q, z_s). \quad (25)$$

Figure 8 shows $\text{fr}_{\text{app}}(M, q, z_s)$ for selected apparent mass values in solid traces. The top row shows the fractions with a fixed mass ratio of 1.73, and the bottom row shows similar plots with mass ratio fixed at 2.67. The left, middle, and right columns show results for CE, ET, and LIGO Voyager, respectively. The distinguishable fractions for fixed redshifted mass, $\text{fr}(\tilde{M}, q, z_s)$, are plotted for reference in dashed lines. The exact mass values are specified in the legend.

As expected, the resulting traces show similar trends and features as in Fig. 7: at lower redshifts, the distinguishable fraction decreases with ρ . It then undergoes a cusp where the high inclination regions start to be excluded before continuing to decrease. The trace is jagged due to the finite spacing of the interpolation data points, rather than any physical jumps in the fraction.

In most cases, there is a significant fraction of GW sources with distinguishable type II images via the log Bayes factor test. As Fig. 4 suggests, the mismatch value is not drastically different across the three detectors with different noise curve shapes. The redshift reach is rather primarily determined by ρ , related to the overall sensitivity level of different detectors. For example, for type II images with apparent mass $\tilde{M} = 60 M_\odot$ and mass ratio $q = 1.76$, 60% can be distinguished in CE out to $z_s \sim 12.5$. For ET, 60% of the same population can be distinguished out to $z_s \sim 2$. Due to the lower sensitivity of LIGO Voyager, a similar fraction of such type II images can be identified only out to $z_s \sim 0.5$. However, for type II images with a higher apparent mass of $100 M_\odot$, 50% can still be registered out to $z_s \sim 1$.

Finally, we assess the impact of mismatch resolution. Throughout this paper, we adopt a minimum mismatch value of $\epsilon = 0.001$. Figure 9 shows the changes in the distinguishable fraction of GW sources with type II images for CE. The mass ratio is fixed to be $q = 1.73$, and the solid lines from left to right represent $\tilde{M} = 100 M_\odot, 200 M_\odot, 260 M_\odot, 400 M_\odot, 600 M_\odot$, and $800 M_\odot$. The dashed horizontal traces show the largest distinguishable fraction as a function of the redshifted mass.

As the waveform mismatch resolution becomes coarser, the distinguishable fraction decreases significantly. For instance, 70% of all sources with type II images with redshifted mass $m = 100 M_\odot$ and $q = 1.73$ have

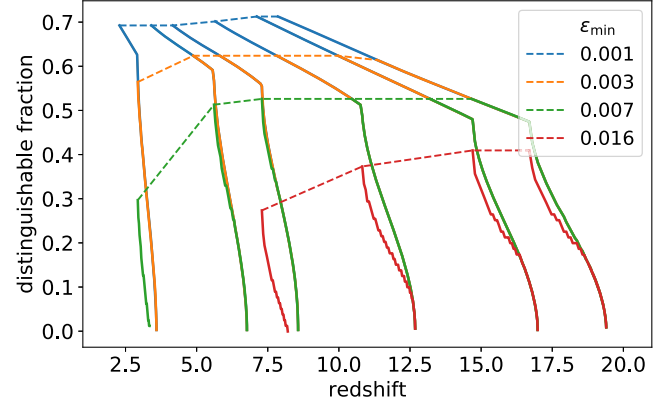


FIG. 9. Distinguishable fractions of sources with type II images assuming a minimum resolvable mismatch of 0.001, 0.003, 0.007 and 0.016 with CE sensitivity. The solid lines show the distinguishable fractions for selected redshifted mass. The mass ratio is fixed to be 1.73. From left to right, the traces correspond to a redshifted mass of $100 M_\odot, 200 M_\odot, 260 M_\odot, 400 M_\odot, 600 M_\odot$, and $800 M_\odot$. Dashed lines show the maximum fraction for this range of redshifted mass. For a mismatch resolution of 0.016 (or 0.984 overlap), fraction traces for type II images with redshifted mass $\tilde{M} = 100, 200 M_\odot$ (the leftmost two traces) are absent, since the resolution is larger than the maximum possible waveform mismatch for such GW sources.

distinguishable type II images out to redshift $z_s \sim 2.5$ if a mismatch of 0.001 is resolvable, but the fraction drops to 30% if the mismatch resolution is 0.007. With a mismatch resolution of 0.016, no such type II images are distinguishable. This critical role of the minimum resolvable mismatch suggests that the distinguishability of type II images does not solely depend on the SNR. In the era of third-generation GW detectors, the possible scientific output from GW detection events is not solely determined by the noise level. As is discussed, the waveform template bank density in the matched filtering search limits the waveform difference resolution. In addition, the detector calibration must also be sufficiently accurate such that we can be confident that the mismatch from the data reflects a real signal difference, rather than an instrument systematic error. Otherwise, we cannot take full advantage of the large SNR offered by exquisite detector sensitivity. Since these factors during the third-generation GW detector era are still subject to much uncertainty, to our knowledge, we have chosen $\epsilon = 0.001$ as the fiducial value. The analysis should be refined as such information becomes available.

V. DETECTABLE POPULATION

The distinguishable fraction calculations depend only on the waveform mismatch, and do not assume astrophysical estimates on GW source and lens distributions. In this section, we describe how results of image distinguishability can be combined with astrophysical models to give a more

detailed prediction of the detectable GW events with type II images and those with distinguishable type II images. In the following subsections, we first offer an overview of the procedures for calculating the lensed population, followed by more detailed discussions on each ingredient.

A. Overview

We define a GW event as a particular binary black hole (BBH) merger with possibly multiple images due to strong lensing. For our calculation, a GW image is detectable if its single-detector $\rho \geq 8$; we defer the detector network scenario to future studies. A GW event has a distinguishable type II image if this image satisfy the log Bayes factor threshold. The differential detectable and distinguishable merger rate is given by

$$\begin{aligned} \frac{\partial^3 \dot{N}_{\text{II, det}}}{\partial M_* \partial q \partial z_s} &= \tau_{\text{II}}(z_s) \frac{\partial^3 \dot{N}}{\partial M_* \partial q \partial z_s} \int d \log_{10} \mu \\ &\times \left[\frac{\partial P_{\text{II}}(\mu, z_s)}{\partial \log_{10} \mu} \int_0^{\pi/2} d\iota \sin \iota \Theta(\sqrt{\mu} \rho(M_*, q, z_s, \iota) - 8) \right], \end{aligned} \quad (26)$$

$$\begin{aligned} \frac{\partial^3 \dot{N}_{\text{II, dis}}}{\partial M_* \partial q \partial z_s} &= \tau_{\text{II}}(z_s) \frac{\partial^3 \dot{N}}{\partial M_* \partial q \partial z_s} \\ &\times \int d \log_{10} \mu \frac{\partial P_{\text{II}}(\mu, z_s)}{\partial \log_{10} \mu} \text{fr}(\tilde{M}, q, \tilde{z}_s), \end{aligned} \quad (27)$$

where $\partial^3 \dot{N} / \partial M_* \partial q \partial z_s$ is GW event rate per intrinsic binary mass, M_* , mass ratio, q , and GW source redshift, z_s , measured in the observer frame. The weighting factor $\sin \iota$ comes from the assumption that BBH mergers are distributed evenly on the sky. Θ is the Heaviside function. Multiplying with the optical depth $\tau_{\text{II}}(z_s)$, we obtain the rate of events with at least one type II images. The quantity $\partial P_{\text{II}}(\mu, z_s) / \partial \log_{10} \mu$ describes the distribution of magnification μ for type II images for sources at z_s , normalized such that

$$\int d \log_{10} \mu \frac{\partial P_{\text{II}}(\mu, z_s)}{\partial \log_{10} \mu} = 1. \quad (28)$$

Due to magnification, the source appears to have the same redshifted mass, but the inferred luminosity distance is different. Therefore,

$$\begin{aligned} \tilde{M} &= M_*(1 + z_s) \\ D_L(\tilde{z}_s) &= D_L(z_s) / \sqrt{\mu}. \end{aligned} \quad (29)$$

The differential merger number per observer time is calculated as [see also [21]]

$$\frac{\partial^3 \dot{N}}{\partial M_* \partial q \partial z_s} = R_{\text{mrg}}(M_*, q, z_s) \frac{1}{1 + z_s} \frac{dV_c}{dz_s}, \quad (30)$$

where dV_c/dz_s is the differential comoving volume. The $1/(1 + z_s)$ factor accounts for the cosmological redshift and converts the source-frame merger rates into detector-frame merger rates. In Fig. 12, we plot this “modified” differential comoving volume and the total merger rates for reference. Since we have adopted the same population models, Fig. 12 replicates Fig. 1 in [10].

For fast calculation of $\rho(M_*, q, z_s, \iota)$, we use the phenomenological model IMRPhenomHM [62], called from the PYTHON package `pycbc.waveform` [63].

We note that for a type I/II waveform mismatch of 6%, the required ρ to be distinguishable is approximately 13, larger than the threshold SNR of 8, and none of the GW sources we consider have a larger waveform mismatch. Therefore, we may assume that the *distinguishable* images are all *detectable*, leading to the omission of the Heaviside function in Eq. (27). In addition, only 0.2% of all sample lens systems have a brighter type II image than the type I image. Considering errors from the lens-equation solution algorithm and the small number of events with distinguishable type II images, we may assume that the events with detectable or distinguishable type II images will most certainly have a detectable type I companion image.

In the following subsections, we compute the type II image optical depth and the magnification distribution. We then summarize procedures to calculate the total BBH merger rates. Detailed steps and adopted parameter values are presented in Appendix. We then make concrete detection population for CE, ET, and LIGO Voyager and discuss results.

B. Optical depth and magnification

To obtain $\tau_{\text{II}}(z_s)$ and $\partial P_{\text{II}}(\mu, z_s) / \partial \log_{10} \mu$, we perform a Monte Carlo simulation. We consider elliptical galaxies as lenses, as they are expected to be the predominant lensing objects [11]. While the lens geometry and properties are expected to be more complex and varied in nature, studies showed that a simple lens model, such as the singular isothermal sphere model, is sufficient to capture most of the results from more sophisticated hydrodynamic simulations of the Universe [64]. In this study, we adopt the slightly more generalized singular isothermal ellipsoid model following the examples of Refs. [10, 11]. References [9, 13] adopt the singular isothermal sphere lens model and predict the detectable strongly lensed events for ET. For one of the BBH evolutionary scenarios they investigate, it is predicted that 57.2 strongly lensed events can be detected out of the 2.08×10^5 total detectable BBH events per year, roughly a factor of three smaller than our prediction (see Table I). In the future, we can adapt our analysis using different lens models and systematically study the uncertainty in the strong-lensing population predictions.

TABLE I. Predicted yearly detection rates. The columns show the detectable BBH merger rates, the rates for strongly lensed (SL) BBH mergers, the rates for BBH mergers with detectable type II images and the rates for BBH mergers with distinguishable type II images. The distinguishable event rates are given with $\ln \mathcal{B} \geq 10, 5, 2$. The fraction of events with detectable type II images that are also distinguishable is shown in the parenthesis.

		CE	ET	Voyager
Det.		2.17×10^5	1.96×10^5	7.59×10^4
Det. SL		184.7	157.1	38.4
Det. type II		172.2	118.2	27.4
Dist. type II	$\ln \mathcal{B} \geq 10$	56.9 (33.1%)	8.6 (7.3%)	0.06 (0.22%)
	$\ln \mathcal{B} \geq 5$	65.6 (38.1%)	14.9 (12.6%)	0.22 (0.81%)
	$\ln \mathcal{B} \geq 2$	73.7 (42.8%)	25.3 (21.4%)	0.80 (2.93%)

We restrict the GW source redshift to $0.05 \leq z_s \leq 7$. In the low-redshift limit, GW sources in our local universe ($z_s \ll 0.05$) is unlikely to be strongly lensed, since lensing rates are expected to be low, and there are not sufficiently many massive galaxies in between to compensate. We set the upper limit of the galaxy redshift to $z_s = 7$, since such galaxies are faint and robust observational data is relatively scarce for developing a reliable phenomenological model of the mass function [65].

At each redshift, we generate samples of lenses, parametrized by surface velocity dispersion, σ_v , ellipticity, e , lens redshift, z_l , and the lens-plane angular coordinates of the lens, $\vec{\theta} = (x, y)$. For the number of lenses per unit σ_v per comoving volume, $\Psi(\sigma_v, z_l)$, we first adopt the modified Schechter function [66], which is calibrated to observation on galaxies in the solar neighborhood,

$$\Psi(\sigma_v, 0) = \phi_* \left(\frac{\sigma_v}{\sigma_*} \right)^\alpha \exp \left[- \left(\frac{\sigma_v}{\sigma_*} \right)^\beta \right] \frac{\beta}{\sigma_v \Gamma(\alpha/\beta)}, \quad (31)$$

where $\phi_* = 8.0 \times 10^{-3} h^3 \text{ Mpc}^{-3}$, $\sigma_* = 161 \text{ km/s}$, $\alpha = 2.32$ and $\beta = 2.67$. h is the Hubble parameter.

To account for the redshift dependence, we follow the prescription in [11], in which

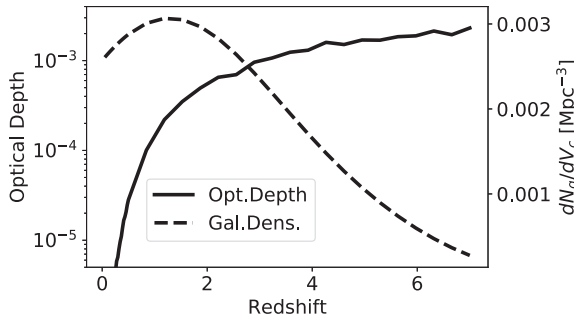


FIG. 10. Left axis: optical depths, $\tau_{\text{II}}(z_s)$, for GW sources with at least one type II image as a function of source redshift. Optical depths lower than $\sim 10^{-5}$ are omitted, as they are too low to predict an observable GW source at such redshifts with type II images in future detectors. Right axis: comoving number density of all galaxies modeled as lenses.

$$\Psi(\sigma_v, z_l) = \Psi(\sigma_v, 0) \frac{\Psi_{\text{hyd}}(\sigma_v, z_l)}{\Psi_{\text{hyd}}(\sigma_v, 0)}, \quad (32)$$

where $\Psi_{\text{hyd}}(\sigma_v, z_l)$ is the velocity dispersion function derived from hydrodynamical simulation in [67]. The redshift dependence of the galaxy comoving number density is shown on the right axis of Fig. 10. We truncate σ_v at 50 km/s and 400 km/s to include the major part of the distribution in Eq. (31). In general, the galaxy number density peaks around $z \sim 1, 2$ and decreases toward higher redshift as they have less time to form.

For galaxy ellipticity, we adopt the same Gaussian distribution as in [10], where the mean and standard deviation are 0.7 and 0.16, truncated at $e = 0.2$ and $e = 1$. The lens redshift is uniformly sampled from $[0, z_s]$. Since strong lensing occurs only when the angular separation between the lens and the source is small, we uniformly sample the lens positions within a square region centered at the source with the side length equal to four times the Einstein radius of the lens, given by [see [10]]

$$\theta_E = 4\pi \left(\frac{\sigma_v}{c} \right)^2 \frac{D_{ls}}{D_s}, \quad (33)$$

where D_{ls} and D_s are the lens-source and observer-source separations, respectively.

For each sampled lens parameter set $(\sigma_v, e, z_l, \vec{\theta})$, we solve the lens equation with the PYTHON package `lenstronomy`⁶[68] and obtain the number of images, image types and magnifications. Since our interest in distinguishable type II images is to identify strongly lensed GW sources, we compute the “source-based” optical depth, the fraction of GW sources with type II images, rather than the fraction of all images that are type II. Each sample with at least one type II image contributes to τ_{II} , while depending on the solution for μ , it contributes to $\partial P_{\text{II}} / \partial \log_{10} \mu$ accordingly. To account for the lens population, each sample is weighted by the expected count of such a galaxy within the defined lens position range.

⁶<https://github.com/sibirrer/lenstronomy>.

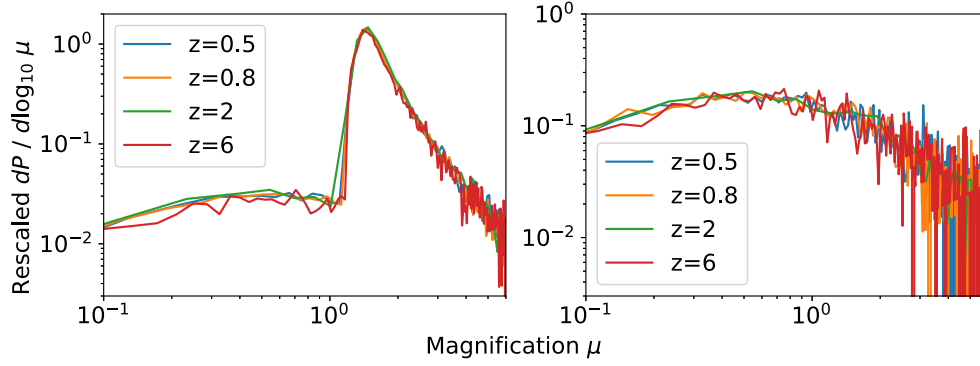


FIG. 11. Rescaled magnification distribution at redshift $z = 0.5, 0.8, 2, 6$. The left panel shows $\partial P(\mu, z_s)/\partial \log_{10} \mu$, including all images from the Monte Carlo samples. The right panel shows the rescaled distribution of only type II images, $\partial P_{\text{II}}(\mu, z_s)/\partial \log_{10} \mu$. The traces are rescaled such that the largest image count is normalized to 1.

Figure 10 shows the type II image optical depth at various source redshifts on the left axis. Optical depths smaller than $\sim 10^{-5}$ are truncated, as they are too low to produce a possible lensed source. We observe that the optical depth is on the order of 10^{-3} – 10^{-4} , consistent with results from ray-tracing studies using N-body simulations [see, e.g., [69,70]].

In the generated sample, the probability of a strongly lensed GW event (i.e., with multiple images) to have no type II images is smaller than 0.01% and therefore negligible. We conclude that the type II image optical depth is effectively identical to the strong lensing optical depth. We calculate that roughly 91.5% of all sources with multiple images have a type II image as the second “brightest” image, which suggests that if multiple images were to be detected, it is likely that at least one of the images may be a candidate for type II image distinction via the log Bayes factor test.

For larger redshifts, we do not extrapolate optical depth due to the lack of information on extremely high redshift galaxy velocity dispersion function from hydrodynamical simulations. Instead, we take the conservative limit and assume the optical depth to be constant beyond $z_s = 7$.

For the magnification distribution, we extract the type II images from the Monte Carlo simulation samples for each redshift. Figure 11 shows the rescaled image magnification distribution per $\log_{10} \mu$ at selected redshifts $z_s = 0.5, 0.8, 2, 6$. The *left* panel includes all images with the peak dominated by the slightly magnified type I images. The *right* panel contains only type II images, which constitute the demagnified image population. The rescaling normalizes the highest image count in each case to 1. Since the magnification for all images peaks around 1, we ignore it when calculating the detectable strongly lensed GW events; instead, the detection rate of all strongly lensed events can be estimated by multiplying the detectable BBH merger rate under the no-lensing hypothesis by the strong lensing optical depth, which, as the Monte Carlo samples show, is effectively identical to the type II image optical depth.

We note that $\partial P_{\text{II}}(\mu, z_s)/\partial \log_{10} \mu$ is independent from source redshift z_s by construction, as Fig. 11 confirms. To explain this feature, we first note that the lens equation solution depends only on $\vec{\theta}/\theta_E$, where $\vec{\theta} = (x, y)$ and θ_E is defined in Eq. (33). Since the range of the possible lens angular positions, $\vec{\theta}$, is directly determined by θ_E , the image solution (image count, magnification, etc.) and its distribution remain constant under the scaling. The only remaining redshift-dependent quantity is the galaxy velocity dispersion function. However, Eq. (32) shows that only the overall magnitude of $\Psi(\sigma_v, z_l)$ changes with redshift. Consequently, we expect a universal normalized magnification distribution for all redshifts.

Finally, we fit $dP_{\text{II}}(\mu)/d \log_{10} \mu$ by a log normal distribution with a mean of -0.35 and standard deviation of 0.57 , truncated at $\log_{10} \mu = -2, 1$.

C. GW source population

We adopt GW source population models provided by [10,71]. In summary, we assume the merger rate of the primary black hole in a binary to be proportional to the formation rate of black holes and their progenitor stars. The merger rate is then calibrated to the observed BBH merger density in the local universe. We follow the prescription and the chosen astrophysical models in [10,71], and we provide more details in Appendix for reference. The BBH merger population is calibrated to a local merger rate of $103 \text{ Gpc}^{-3} \text{ yr}^{-1}$ based on LIGO detection data up until GW170104 [72]. With the second LIGO-Virgo Gravitational-Wave Transient Catalog (GWTC-2), the local BBH merger rate is more tightly constrained to be $23.9^{+14.3}_{-8.6} \text{ Gpc}^{-3} \text{ yr}^{-1}$ [72]. The data slightly favor that the merger rate increases with redshift, but remain statistically consistent with a nonevolving merger rate hypothesis [72]. This updated value suggests that our merger rate model could be an overestimate. However, the local rate difference is less than an order of magnitude, and the high-redshift merger rates are not constrained by LIGO data. As the local merger

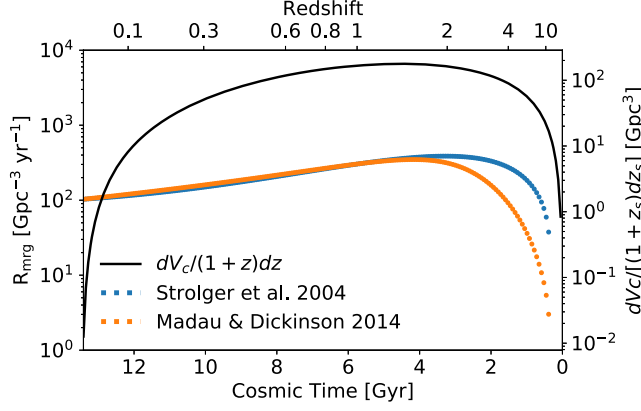


FIG. 12. BBH merger rate density, assuming two SFR models in [73,74]. The blue overarching trace plots the modified differential comoving volume and corresponds to values on the right axis. The merger rate density is directly analogous to Fig. 1 in [10].

rate only acts as an overall scaling in the population model, our result can be easily scaled to reflect any differences. Consequently, our predictions serve as an adequate reference and can be easily adapted in light of new data and more accurate BBH population models.

Figure 12 replicates Fig. 1 in [10] and shows the predicted BBH merger rate density under the two star formation rate (SFR) models in [73] and [74], respectively. Due to the intrinsic uncertainty in these analytical SFR models, we choose one, the more optimistic SFR in [74], for the following population estimates.

The total BBH merger rate per source redshift and the detectable merger rate are plotted in Fig. 13. Since the strong lensing optical depth is in general smaller than 0.1% at the redshift with the most GW sources, we will neglect the magnification effect when calculating the total detectable GW events. To keep the detectable population estimate general, we assume the detectors to always be online. To incorporate the detector duty cycle, the detectable and

distinguishable populations simply scale proportionally with the fraction of detector online time, since the type II waveforms can be identified by their own waveforms, and the duty cycle does not disproportionately affect particular GW image types. We estimate a total of 2.17×10^5 BBH mergers per year up to $z_s = 23$. The detectable total merger number is 2.17×10^5 for CE (99.96%), 1.96×10^5 for ET (90.3%), and 7.59×10^4 for LIGO Voyager (35.0%). We note that the detection rate is not only affected by the detector sensitivity, but also by the redshift distribution of BBH mergers and the comoving volume. Even though ET has lower sensitivity than CE overall, it already covers the redshift range with peak GW source count ($z_s \sim 2$). At large redshift with $z_s \gg 7$, BBH mergers happen far less frequently due to a lack of black hole formation and the decreasing comoving volume per redshift. Consequently, the detection rate of ET is only slightly lower than CE. In the case of LIGO Voyager, the lower sensitivity excludes many sources from $z_s \sim 2$, leading to a larger loss in the detectable source fraction.

D. Type II image rate

In this section, we combine lensing statistics and GW source population models to study the rate of detectable and distinguishable type II images in third generation GW detectors.

Figure 13 shows the differential event rate as a function of redshift for three detectors. In each panel, four different populations are shown. The total rate of BBH mergers are plotted as solid black curves. The dashed curves show the rate of detectable GW events. The dotted curves show the rate of events with a detectable type II image as in Eq. (26). The dot-dash curves show the rate of GW events with at least one distinguishable type II image.

As expected, the rate of BBH mergers in all three categories decreases with the detector sensitivity, especially at high redshifts. For LIGO Voyager, in particular, the rate of expected GW sources with distinguishable type II

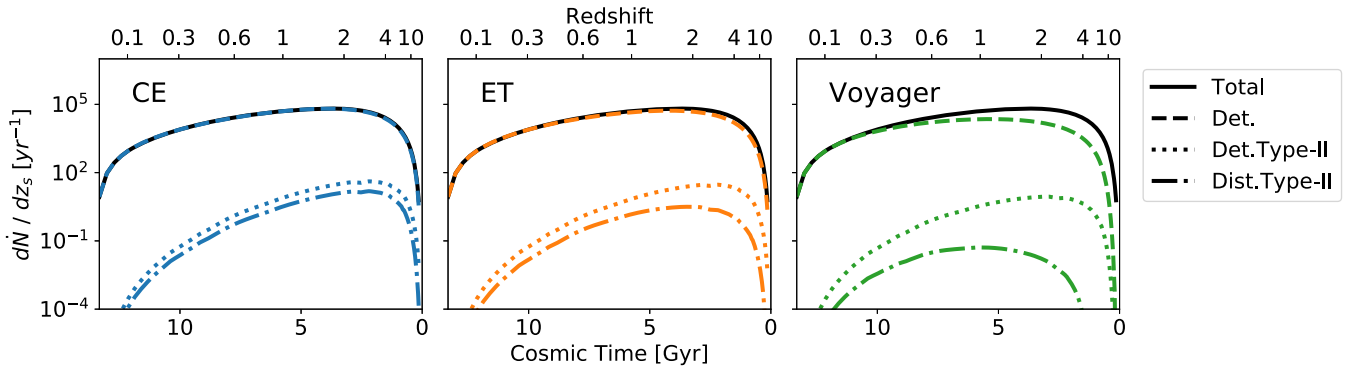


FIG. 13. Yearly detected population per unit redshift prediction as a function of redshift. The panels from left to right show the detection population for CE, ET and LIGO Voyager. In all panels, the solid black line denotes the total BBH merger rate. The dashed curves show the rate of detectable GW sources (i.e., $\rho > 8$) when unlensed. The dotted curve shows the rate of GW sources with a detectable type II image. The dot-dash curve shows the event population with a distinguishable type II image. See text for total detection rates.

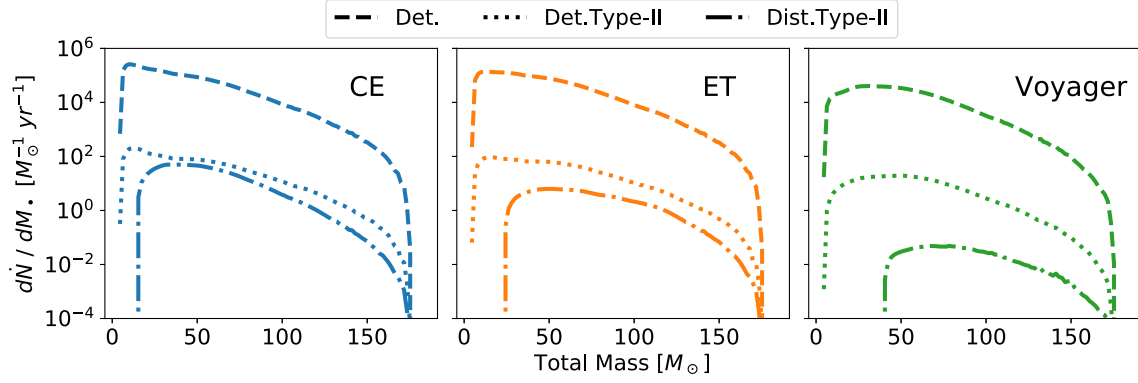


FIG. 14. Detection rate as a function of BBH intrinsic mass. The panels from left to right shows the detection prediction for CE, ET, and LIGO Voyager. The line styles are consistent with those in Fig. 13.

images drops quickly with redshift, consistent with the trend of the distinguishable fraction in Fig. 8.

Figure 14 plots the same population prediction binned by the total mass of the BBH, with consistent line styles as in Fig. 13. For all detectors, the detection rate decreases with increasing total mass, consistent with the underlying initial mass function. The detection rate of events with distinguishable type II images shows a cutoff at small total mass, which is primarily due to two factors. When the waveform mismatch for low-mass BBHs is smaller than the imposed mismatch resolution (i.e., $\epsilon < 0.001$), their type II images are considered indistinguishable from type I images. When the mismatch has just exceeded the resolution threshold, distinguishability requires very large SNRs, which may not be achievable depending on the detector sensitivity. Consequently, we observe a mass cutoff in all three detectors, which shifts to higher masses as the detector sensitivity decreases.

Overall, we predict that CE will detect roughly 184.7 strongly lensed GW events per year, among which 172.2 have at least one detectable type II image. Among these strongly lensed GW sources, 56.9 per year have a type II image distinguishable via the log Bayes factor test. ET will be able to detect 157.1 strongly lensed events per year, and 118.2 of these have detectable type II images. However, due to reduced sensitivity, the number of sources with a distinguishable type II image drops to 8.6 per year. For LIGO Voyager, the yearly detection rate of GW events with detectable type II images is 27.4 per year out of the 38.4 strongly lensed events. The distinguishable type II image rate is 0.06 per year, which suggests that the possibility of observing a GW source with distinguishable type II images with LIGO Voyager is relatively slim. The detection rates are summarized in Table I.

E. Discussion

In this section, we discuss the implication of the predicted detection rates for GW sources with distinguishable type II images. We re-examine assumptions in our

analysis and explore how relaxing these assumptions lead to more an optimistic detection prediction.

As Fig. 14 shows, the yearly detection rates of GW sources with distinguishable type II images are 56.9, 8.6, and 0.06 for CE, ET and LIGO Voyager, respectively. In particular, in the case of CE, more than 30% of all detectable strongly lensed sources will have distinguishable type II images. For such sources, detection of the type II image alone can confirm the existence of strongly lensed images, without pair-wise GW event inference on the strong lensing hypothesis. Once such images are identified, the inferred source parameter values can act as a prior during the subsequent and more elaborate catalog search for the other images.

For ET and LIGO Voyager, the expected detection rate is smaller, thus the distinguishable type II images will not be as powerful for confirming the strong lensing hypothesis as in the case of CE. However, we emphasize that if several of our conservative constraints can be relaxed, distinguishable type II images can still contribute to the identification of strong lensing.

The first condition we revisit is the waveform mismatch resolution. Throughout the analysis, we consistently adopt

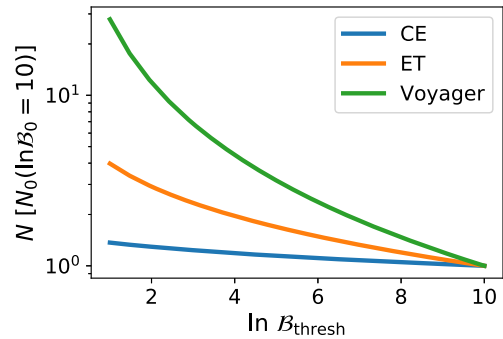


FIG. 15. The number of GW sources with distinguishable type II images for different log Bayes factor threshold values, expressed as a fraction of the distinguishable number with the threshold value $\ln B_{\text{thresh}} = 10$.

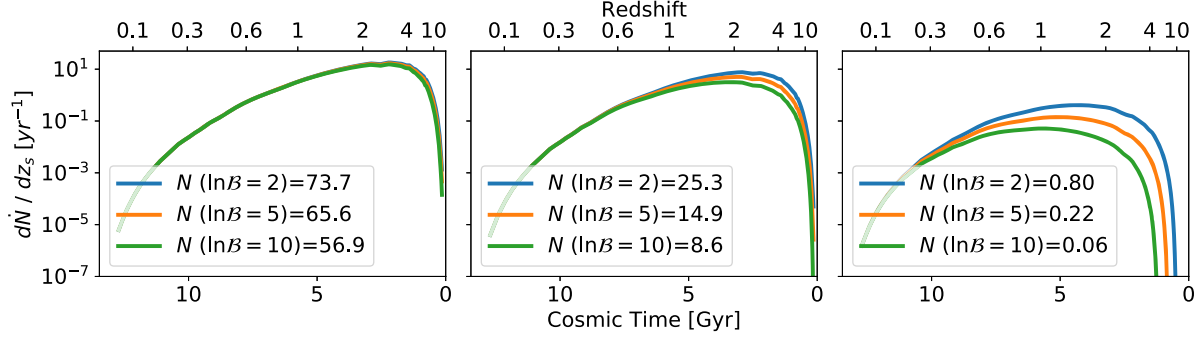


FIG. 16. The expected number of GW sources with distinguishable type II images with $\ln \mathcal{B}_{\text{thresh}} = 2, 5$, and 10 . The expected yearly detection count for each threshold value is shown in the legend. The panels from left to right correspond to CE, ET and LIGO Voyager. As expected, the detection number increase is most significant at large redshift for all three detectors.

$\epsilon_{\min} = 0.001$, which excludes the binaries at small inclinations, and the distinguishable fraction is “saturated” at roughly 70% (see, e.g., Fig. 7 and Fig. 9). As Fig. 9 suggests, the waveform mismatch resolution significantly affects the fraction of distinguishable type II images. If we can expect a better waveform resolution from third-generation GW detectors, the distinguishable fraction should increase considerably; as Fig. 9 shows, the distinguishable fraction roughly doubles as the mismatch resolution improves from $\mathcal{O}(1\%)$ to $\mathcal{O}(0.1\%)$ assuming CE sensitivity. For CE and ET, this increase results in many more detectable sources at small redshifts ($z_s \sim 1, 2$), where the BBH population also peaks. This requirement has two implications for third-generation GW detector performance and data analysis process. As is discussed briefly in Sec. III, the error in detector calibration should be much smaller, such that the waveform mismatch is not obscured by systematic uncertainties. In terms of the data analysis process, the density of the matching template bank should be such that the waveform difference is large compared with the template spacing. If such conditions are not satisfied, the high SNR detection offered by the third-generation detectors cannot be taken full advantage of to maximize the scientific output.

We have also taken a conservative estimate by setting the threshold log Bayes factor to be 10. Even for $\ln \mathcal{B}_{\text{thresh}} = 5$, the type II image hypothesis is more than 100 times more likely than the type I image hypothesis, and an even smaller threshold value may be sufficient for realistic data analysis. Figure 15 shows the increase in the number of events with distinguishable type II images with a lower $\ln \mathcal{B}_{\text{thresh}}$, normalized to the number when $\ln \mathcal{B}_{\text{thresh}} = 10$. We observe that the increase is the most dramatic for LIGO Voyager, as a lower threshold extends the sensitive range to higher redshift ($z_s \sim 2$), where the GW source population peaks. For CE and ET, the increase is more modest, as they already detect most sources at $z_s \sim 2$ with high SNR. The extended range is then expected to add relatively fewer GW sources in comparison. Figure 16 shows the redshift distribution of

the GW sources with distinguishable type II images with $\ln \mathcal{B}_{\text{thresh}} = 2, 5$, and 10 . As expected, the distinguishable rate increase is more significant at high redshift, and the effect is the strongest for LIGO Voyager; at $\ln \mathcal{B}_{\text{thresh}} = 2$, 42.8% of all strongly lensed GW sources in CE are accompanied by at least one distinguishable type II image and 21.4% for ET. For LIGO Voyager, the distinguishable number is still small, but at $\sim 1/\text{yr}$, it is more promising that such an event will appear in the LIGO Voyager catalog with a few years of observing run. The predicted detection rates are summarized in Table I.

In addition, we have so far considered the single-detector scenario, and we estimate the advantage of a detector network via the simplifying assumption that both GW polarizations can be independently detected, i.e., the time-domain waveform for calculating the overlap and ρ is complex. If ET implements a triangular design, the detector itself is sufficient to capture the polarization content [33]. For LIGO Voyager and CE, the polarization content can be obtained if a concurrent detector network exists. In the upcoming decades, more GW observatories across the globe will start to observe, such as the expansion of the LIGO network to include IndiGO⁷ [75]. This global network offers increased detector-networks ρ and an increased detection spatial resolution. On the other hand, the uncertainty in the polarization content from a realistic detector-network model may be partially degenerate with the Hilbert transform signal, thus subtracting away from the type I/II waveform difference and their distinguishability. A thorough investigation on realistic detector network effect is deferred to future studies.

VI. CONCLUSIONS

In this paper, we study an intrinsic waveform signature of type II images of strongly lensed GW sources. For CE, ET and LIGO Voyager, we compute the best-match overlap between type I/II waveforms. We then calculate the

⁷<http://www.gw-indigo.org/tiki-index.php>.

required threshold orbit inclination to establish the type II waveform hypothesis by a favoring log Bayes factor of 10. The fraction of GW sources with distinguishable type II images is computed from the threshold inclination accordingly. For all three detectors, we find that significant fractions of type II images (e.g., 50–70%) of sufficiently high SNR GW events can be identified. In other words, if such a type II image is detected with reasonable SNR, it can likely be distinguished from regular type I images and used as the tell-tale evidence of strongly lensed events.

We also assess the effects of the type II signature in the context of the current LIGO data analysis process. We apply the targeted subthreshold search method described in [48] on an example high-mass-ratio compact binary coalescence event GW190814 [50]. We generate a reduced template bank based on injection run results using simulated type I lensed injections of the target event. The resulting reduced bank is used, then, to search for the same set of simulated lensed injections in two different searches, in which they are injected as type I images (original waveforms) and type II images (Hilbert transform of the same waveforms) respectively.

Our preliminary result shows that there is a slight increase in the number of injections missed when they are treated as type II images. This hints at the possibility that the current search scheme may suffer from sensitivity loss without considering type II images. However, we remark that the current results in this study are only preliminary and will require further studies.

We then incorporate GW source population model and lensing probabilities to predict the expected number of GW sources with distinguishable type II images in CE, ET, and LIGO Voyager respectively. For these three detectors, we predict the yearly detection rates are 56.9, 8.6, and 0.06 with a conservative threshold at $\ln \mathcal{B}_{\text{thresh}} = 10$. A relaxed log Bayes factor threshold boosts the expected detection rates, especially for LIGO Voyager; at $\ln \mathcal{B}_{\text{thresh}} = 2$, the yearly detection rate for LIGO Voyager approaches 1/yr.

Such distinguishable type II images are “shortcuts” for identifying strongly lensed events, as they guarantee the existence of at least one other lensed image. They also improve the computational efficiency of searching for the companion images, as the estimated parameters, such as the redshifted mass, mass ratio and sky location, can inform a more comprehensive catalog search. As illustrated, this method will be most powerful with the unprecedented sensitivity offered by third-generation GW detectors.

Our work can be extended and refined in several directions. We can relax the constraints on GW source range by including spin and orbit eccentricity, as is studied in Ref. [31]. On one hand, the Hilbert transform of GWs from such sources may have a larger mismatch from the original waveform, favoring type II image distinguishability. On the other hand, the Hilbert transform may be

partially degenerate with a parameter bias with the additional degrees of freedom. The effect of these competing factors warrants careful treatment.

We may consider realistic detector networks instead of assuming complete knowledge on both GW polarizations, which adds to the underlying waveform uncertainties. Similar to the hypothesized effect of binary spin and orbital eccentricity, uncertainty in the polarization may be partially degenerate with the Hilbert transform signature. However, a detector network yields larger signal SNR, which should promote the distinguishability of type II images.

It is also important to refine the lens modeling. While theoretical works on various lens types and their respective image characteristics abound, to our knowledge, the effect of the model choice on predictions for realistic detection has yet to be systematically investigated. Therefore, such a follow-up study is essential for understanding the uncertainty and robustness of this strongly lensing detection forecast.

In conclusion, this study shows that the intrinsic waveform characteristics of type II images can be a powerful supplemental tool for hunting strongly lensed events in the catalog of third-generation GW detectors, when tens of such events may be identified.

ACKNOWLEDGMENTS

We thank the anonymous referee for helpful suggestions that improve this manuscript. Y. W. would like to thank the David and Ellen Lee Distinguished Fellowship for support during this research. Research of Y. W. and Y. C. are supported by the Simons Foundation (Grant No. 568762), and the National Science Foundation, through Grants No. PHY-2011961, No. PHY-2011968, and No. PHY-1836809. R. K. L. L. would also like to thank the Croucher Foundation for support during this research. The computations presented here were conducted on the Caltech High Performance Cluster partially supported by a grant from the Gordon and Betty Moore Foundation. This paper carries LIGO Document No. LIGO-P2100002.

APPENDIX: BINARY BLACK HOLE MERGER RATE

In this Appendix, we elaborate on the astrophysical models adopted to calculate the merger rate of binary black holes. In summary, we compute the BBH merger rate from population models on black hole progenitor stars and calibrate to the observed rate in the local universe.

Adapted from Eq. (B1) and (B2) in [10], the birth rate of individual black holes with mass m_* at redshift z_s , $R_{\text{birth}}(m_*, z_s)$, is given by

$$R_{\text{mrg}}(m_*, z_s) = \int dm_* dt_d dZ \phi(m_*) \dot{\psi}(t(z_s) - t_d) P(t_d) P(Z, t(z_s) - t_d) \delta[m_* - g^{-1}(m_*, Z)], \quad (\text{A1})$$

where m_* is the mass of the progenitor star, Z is stellar metallicity and $t(z_s)$ is the cosmic time as a function of redshift. $g^{-1}(m_*, Z)$ gives the stellar mass m_* with metallicity Z that leaves a black hole remnant with mass m_* . The expression of remnant black hole mass as a function of stellar mass and metallicity is given in [76], with $20 M_\odot < m_* < 105 M_\odot$ and $-5 < \log_{10} Z < -1.7$. $P(Z, t_z)$ is the redshift-dependent distribution of metallicity. The mean log metallicity at any redshift is given in [77]. At each redshift, the metallicity follows a log normal distribution [77].

t_d is the time delay between black hole formation and its merger with another black hole. $P(t_d)$ is the distribution of time delay, and we adopt the form $P(t_d) \propto t_d^{-1}$, truncated at $t_d = 50$ Myr and the Hubble time [10]. Note that we ignore the time delay between the formation of a star and the formation of its remnant. Since stellar evolution is on the

order of Myr, which is negligibly small compare to the evolution time scale of galaxies and hence that of black holes, we can neglect it for model simplicity without incurring large errors.

The quantity $\phi(m_*)$ is the initial mass function that describes the stellar mass distribution, which we assume to remain constant across redshift. Specifically, we adopt the Chabrier initial mass function [78] for $m_* > 1 M_\odot$, where $\phi(m_*) \propto m_*^{-2.3}$. The quantity $\dot{\psi}(t)$ is the star formation rate (SFR) including all m_* at cosmic time t . We adopt the analytic SFR expression in [74]. We calibrate the merger rate at $z = 0$ to be $103 \text{ Gpc}^{-3} \text{ yr}^{-1}$, which is the expected local black hole merger rate given LIGO detection data up until GW170104 [71,72]. See text for the effects of an updated local merger rate based on GWTC-2.

We note that $R_{\text{mrg}}(m_*, z_s)$ is the rates for black hole binary at z_s whose *primary* black hole, i.e., the heavier one, is m_* . We then assign a mass ratio value according to the distribution $P(q) \propto q$, with q truncated at 1.2 and 3.2. We can then directly convert the rates into $R_{\text{mrg}}(M_*, q, z_s)$ where M_* is the total binary mass.

-
- [1] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **896**, L44 (2020).
 - [2] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen *et al.* (The LIGO Scientific and the Virgo Collaborations), *Astrophys. J. Lett.* **900**, L13 (2020).
 - [3] K.-H. Lai, O. A. Hannuksela, A. Herrera-Martín, J. M. Diego, T. Broadhurst, and T. G. F. Li, *Phys. Rev. D* **98**, 083005 (2018).
 - [4] A. K. Meena and J. S. Bagla, *Mon. Not. R. Astron. Soc.* **492**, 1127 (2020).
 - [5] K. Pardo, M. Fishbach, D. E. Holz, and D. N. Spergel, *J. Cosmol. Astropart. Phys.* **07** (2018) 048.
 - [6] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **100**, 104036 (2019).
 - [7] S. Perkins and N. Yunes, *Classical Quant. Grav.* **36**, 055013 (2019).
 - [8] A. Vijaykumar, M. V. S. Saketh, S. Kumar, P. Ajith, and T. R. Choudhury, *arXiv:2005.01111*.
 - [9] M. Biesiada, X. Ding, A. Piórkowska, and Z.-H. Zhu, *J. Cosmol. Astropart. Phys.* **10** (2014) 080.
 - [10] S.-S. Li, S. Mao, Y. Zhao, and Y. Lu, *Mon. Not. R. Astron. Soc.* **476**, 2220 (2018).
 - [11] M. Oguri, *Mon. Not. R. Astron. Soc.* **480**, 3842 (2018).
 - [12] L. Yang, X. Ding, M. Biesiada, K. Liao, and Z.-H. Zhu, *Astrophys. J.* **874**, 139 (2019).
 - [13] X. Ding, M. Biesiada, and Z.-H. Zhu, *J. Cosmol. Astropart. Phys.* **12** (2015) 006.
 - [14] K. Liao, X.-L. Fan, X. Ding, M. Biesiada, and Z.-H. Zhu, *Nat. Commun.* **8**, 1148 (2017).
 - [15] L. Dai and T. Venumadhav, *arXiv:1702.04724*.
 - [16] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, *Phys. Rev. D* **97**, 023012 (2018).
 - [17] X.-L. Fan, K. Liao, M. Biesiada, A. Piórkowska-Kurpas, and Z.-H. Zhu, *Phys. Rev. Lett.* **118**, 091102 (2017).
 - [18] T. E. Collett and D. Bacon, *Phys. Rev. Lett.* **118**, 091101 (2017).
 - [19] S. Mukherjee, B. D. Wandelt, and J. Silk, *Phys. Rev. D* **101**, 103509 (2020).
 - [20] S. Mukherjee, B. D. Wandelt, and J. Silk, *Mon. Not. R. Astron. Soc.* **494**, 1956 (2020).
 - [21] L. Dai, T. Venumadhav, and K. Sigurdson, *Phys. Rev. D* **95**, 044011 (2017).
 - [22] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, *Astrophys. J. Lett.* **874**, L2 (2019).
 - [23] L. Dai, B. Zackay, T. Venumadhav, J. Roulet, and M. Zaldarriaga, *arXiv:2007.12709* [Phys. Rev. D (to be published)].
 - [24] T. Broadhurst, J. M. Diego, and G. F. Smoot III, *arXiv:1901.03190*.

- [25] T. Broadhurst, J.M. Diego, and G. Smoot III, [arXiv:1802.05273](#).
- [26] T. T. Nakamura, *Phys. Rev. Lett.* **80**, 1138 (1998).
- [27] R. Takahashi and T. Nakamura, *Astrophys. J.* **595**, 1039 (2003).
- [28] J. M. Diego, O. A. Hannuksela, P. L. Kelly, G. Pagano, T. Broadhurst, K. Kim, T. G. F. Li, and G. F. Smoot, *Astron. Astrophys.* **627**, A130 (2019).
- [29] M. H. Y. Cheung, J. Gais, O. A. Hannuksela, and T. G. F. Li, *Mon. Not. R. Astron. Soc.* **503**, 3326 (2021).
- [30] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (Springer-Verlag, Berlin, Heidelberg, 1992).
- [31] J. María Ezquiaga, D. E. Holz, W. Hu, M. Lagos, and R. M. Wald, *Phys. Rev. D* **103**, 064047 (2021).
- [32] R. X. Adhikari, K. Arai, A. F. Brooks, C. Wipf, O. Aguiar, P. Altin, B. Barr, L. Barsotti, R. Bassiri, A. Bell *et al.*, *Classical Quant. Grav.* **37**, 165003 (2020).
- [33] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge, S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik *et al.*, [arXiv:1108.1423](#).
- [34] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen *et al.*, in *Bulletin of the American Astronomical Society* (American Astronomical Society, Washington, DC, 2019), Vol. 51, p. 35.
- [35] T. T. Nakamura and S. Deguchi, *Prog. Theor. Phys. Suppl.* **133**, 137 (1999).
- [36] M. Favata, *Classical Quant. Grav.* **27**, 084036 (2010).
- [37] M. Favata, *Phys. Rev. D* **80**, 024002 (2009).
- [38] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, *Phys. Rev. D* **101**, 023011 (2020).
- [39] M. Zaldarriaga, D. Kushnir, and J. A. Kollmeier, *Mon. Not. R. Astron. Soc.* **473**, 4174 (2018).
- [40] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, G. Allen, A. Allocca *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **100**, 064064 (2019).
- [41] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. Research* **1**, 033015 (2019).
- [42] S. Field, C. Galley, and J. Blackman, in *APS April Meeting Abstracts*, APS Meeting Abstracts (American Physical Society, College Park, MD, 2018), Vol. 2018, p. G14.005.
- [43] S. van der Walt, S. C. Colbert, and G. Varoquaux, *Comput. Sci. Eng.* **13**, 22 (2011).
- [44] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt *et al.*, *Nat. Methods* **17**, 261 (2020).
- [45] B. Moore, T. Robson, N. Loutrel, and N. Yunes, *Classical Quant. Grav.* **35**, 235006 (2018).
- [46] M. Evans, R. Sturani, S. Vitale, and E. Hall (LIGO Scientific Collaboration), Unofficial sensitivity curves (ASD) for aLIGO, Kagra, Virgo, Voyager, Cosmic Explorer and ET, Technical Report No. LIGO-T1500293-v11, 2018.
- [47] B. Allen, *Phys. Rev. D* **71**, 062001 (2005).
- [48] A. K. Y. Li, R. K. L. Lo, S. Sachdev, C. L. Chan, E. T. Lin, T. G. F. Li, and A. J. Weinstein, [arXiv:1904.06020](#).
- [49] C. McIsaac, D. Keitel, T. Collett, I. Harry, S. Mozzon, O. Edy, and D. Bacon, *Phys. Rev. D* **102**, 084031 (2020).
- [50] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **896**, L44 (2020).
- [51] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, [arXiv:2004.06503](#) [Phys. Rev. D (to be published)].
- [52] C. Messick *et al.*, *Phys. Rev. D* **95**, 042001 (2017).
- [53] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
- [54] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius, *Phys. Rev. D* **84**, 062003 (2011).
- [55] M. Vallisneri, *Phys. Rev. D* **86**, 082001 (2012).
- [56] W. Del Pozzo, K. Grover, I. Mandel, and A. Vecchio, *Classical Quant. Grav.* **31**, 205006 (2014).
- [57] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [58] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [59] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, *Phys. Rev. D* **102**, 064002 (2020).
- [60] C. Cutler and É. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
- [61] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, *Astrophys. J.* **611**, 623 (2004).
- [62] C. Kalaghatgi, M. Hannam, and V. Raymond, *Phys. Rev. D* **101**, 103004 (2020).
- [63] A. Nitz *et al.*, gwastro/pycbc: PyCBC release v1.16.11 (2020).
- [64] A. Robertson, G. P. Smith, R. Massey, V. Eke, M. Jauzac, M. Bianconi, and D. Rychanowski, *Mon. Not. R. Astron. Soc.* **495**, 3727 (2020).
- [65] S. L. Finkelstein, *Pub. Astron. Soc. Aust.* **33**, e037 (2016).
- [66] Y.-Y. Choi, C. Park, and M. S. Vogeley, *Astrophys. J.* **658**, 884 (2007).
- [67] P. Torrey, S. Wellons, F. Machado, B. Griffen, D. Nelson, V. Rodríguez-Gomez, R. McKinnon, A. Pillepich, C.-P. Ma, M. Vogelsberger, V. Springel, and L. Hernquist, *Mon. Not. R. Astron. Soc.* **454**, 2770 (2015).
- [68] S. Birrer and A. Amara, *Phys. Dark Universe* **22**, 189 (2018).
- [69] S. Hilbert, S. D. M. White, J. Hartlap, and P. Schneider, *Mon. Not. R. Astron. Soc.* **382**, 121 (2007).
- [70] S. Hilbert, S. D. M. White, J. Hartlap, and P. Schneider, *Mon. Not. R. Astron. Soc.* **386**, 1845 (2008).
- [71] L. Cao, Y. Lu, and Y. Zhao, *Mon. Not. R. Astron. Soc.* **474**, 4997 (2018).
- [72] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **118**, 221101 (2017).
- [73] P. Madau and M. Dickinson, *Annu. Rev. Astron. Astrophys.* **52**, 415 (2014).
- [74] L.-G. Strolger, A. G. Riess, T. Dahlen, M. Livio, N. Panagia, P. Challis, J. L. Tonry, A. V. Filippenko, R. Chornock, H. Ferguson, A. Koekemoer, B. Mobasher, M. Dickinson *et al.*, *Astrophys. J.* **613**, 200 (2004).

- [75] 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), [Public], Technical Report No. LIGO-M1100296-v2, 2011.
- [76] M. Spera, M. Mapelli, and A. Bressan, *Mon. Not. R. Astron. Soc.* **451**, 4086 (2015).
- [77] K. Belczynski, D. E. Holz, T. Bulik, and R. O’Shaughnessy, *Nature (London)* **534**, 512 (2016).
- [78] G. Chabrier, *Publ. Astron. Soc. Pac.* **115**, 763 (2003).
- Correction:* The original article ID number in Ref. [14] was incorrect and has been fixed.