arXiv:2101.08264v1 [gr-qc] 20 Jan 2021

Identifying Type-II Strongly-Lensed Gravitational-Wave Images in Third-Generation Gravitational-Wave Detectors

Yijun Wang,[1,*] Rico K.L. Lo,[1] Alvin K.Y. Li,[1] and Yanbei Chen[1]

[1] *California Institute of Technology, Pasadena, CA 91125, USA*

(Dated: January 22, 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.

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 [9–11]; such events can then be used to study cosmological structures [10, 12–15] and fundamental physics [16, 17].

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, $M_\bullet(1+z_s)$, constant [see, e.g., 18], where $M_\bullet$ 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 [18].

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., 19, 20]. For example, this has been used to study the series of events GW170104, GW170814 and a sub-threshold trigger, GWC170620, as potential candidates for lensed images [20]. In Ref. [21], 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 [15, 22]. 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.

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 pair-wise search method mentioned above. An example is the frequency-specific GW diffraction in weak lensing [23, 24]. Diffraction signature was searched for in current detected events, but it has yet to be found [19]. It is also predicted that, for GWs within the frequency range of LIGO, diffraction becomes important when the lens mass ranges from $1 \sim 100\ M_\odot$ [4, 25]. Strong lensing by such small

* yijunw@caltech.edu

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., 26]. 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., 26]. 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 [14], it is pointed out that Type-II images are degenerate with Type-I images with an orbital phase shift of $\pi/4$, if only the dominant (2,±2) modes are considered. The degeneracy is partially lifted for highly eccentric orbit. It has, however, recently been shown in [27] that the difference between the Type-I/II images is still small if we tune the orbital azimuthal angle, the polarization angle and relative phases between GW modes.

In this paper, we show that Type-II images cannot be completely matched by unlensed GW waveforms by tuning *physical* parameters once we consider higher multipoles of the GWs. For third-generation detectors, such as the LIGO Voyager[28], the Einstein Telescope[1] [29] (ET) and LIGO Cosmic Explorer[2] [30] (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 Section II we review the geometric optics theory for GW lensing. In Section 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 Section 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 Section 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 ΛCDM universe with $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$ and a Hubble Constant of $H_0 = 70\ \mathrm{km\ s^{-1}Mpc^{-1}}$.

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., 23, 26, 31]. In this section, we summarize and discuss scenarios where Type-II images are distinctive from Type-I counterparts. We closely follow the discussion in [31] 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., 24, 26],

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

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,

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

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,

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

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 $\mathrm{Tr}(T_{,ab}) > 0$, the phase factor is 1, giving Type-I images. When $\mathrm{Tr}(T_{,ab}) < 0$, the phase shift is $-\mathrm{sgn}(\omega)\pi$, 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

[1] http://www.et-gw.eu/
[2] https://cosmicexplorer.org/

is then $-\mathrm{sgn}(\omega)\pi/2$, giving Type-II images which are equivalent to a Hilbert transform of Type-I images.

B. Gravitational waves from circular, non-spinning binaries

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

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

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 quasi-circular 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

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

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

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

We note that ϕ appears only in the factor of $\exp(im\phi)$ in ${}_{-2}Y_{lm}(\iota,\phi)$. Furthermore, for non-spinning, 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 $\tilde{h}_{\mathrm{I},m_0}$, $\tilde{h}_{\mathrm{II},m_0}$ is written as

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

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 $(l,\pm2)$ GW modes are considered. Modes with different $|m|$ require different angle changes to compensate for the Hilbert transform (e.g., the $(l,\pm3)$ 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 [14]. For binaries with significant mass ratios and inclination angles, the $(3,\pm3)$ modes become significant, breaking degeneracy. Figure 1 is analogous to Figure 2 in [14] 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 non-spinning in a quasi-circular 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 non-spinning binaries in circular orbits. However, note that these are where the GW memory effects take place [32–34].

In this paper, we systematically explore GW sources which are non-spinning binary black holes in quasi-circular 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 non-spinning 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 non-spinning binaries in quasi-circular orbits. Highly spinning binaries or those with highly eccentric orbits are expected to be fewer than the population we consider [see, e.g., 35, 36]. Since, the optical depth for Type-II images is also small, on the order of $10^{-3}\sim10^{-4}$ [10, 11, 18], 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),

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

where $\tilde{H}_{\mathrm{I},lm}(\tilde{M},q,f)/D_L$ is equal to $\tilde{h}_{\mathrm{I},lm}(f)$ 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, $q\equiv\tilde{M}_1/\tilde{M}_2\geq1$ ($\tilde{M}_1+\tilde{M}_2=\tilde{M}$), 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

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

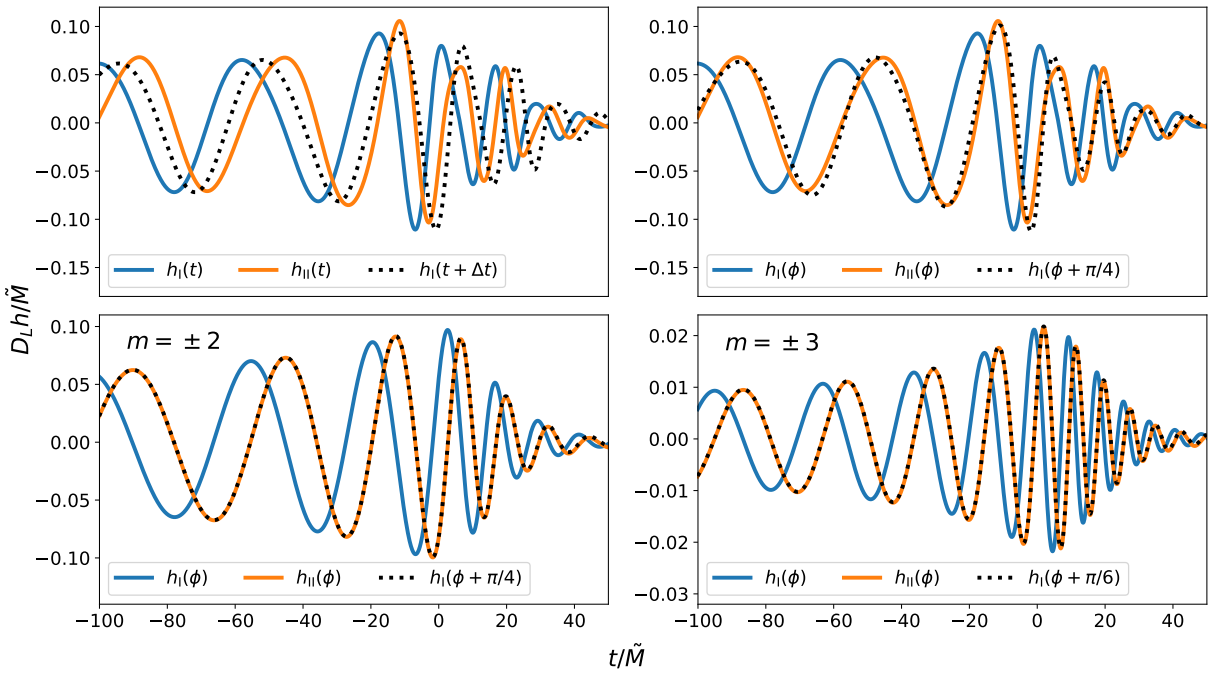

FIG. 1. Type-I/II `NRSur7dq4` surrogate model waveforms from a binary with $\tilde{M} = 150\ M_\odot, q = 2.2, \iota = 0$ deg. The binary is non-spinning in a quasi-circular orbit. The black dotted line shows the Type-I waveform with a $-\pi/4$ shift in the orbital azimuthal angle, and the shifted waveform 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.

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

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

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 Section 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`, [37] extracted through the Python package, `gwsurrogate` [38]. This surrogate model provides all $l \leq 4$ mode waveforms, $h_{{\rm I},lm}(t)$, 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` [39] 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, we adopt a nested search method. We first generate a Type-II signal template, $\tilde{h}_{{\rm II},0}$, with $(\tilde{M}_0, q_0, \iota_0, \phi_0)$. 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` [40] 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., 41]:

$$\begin{aligned}(1-\epsilon)_{\rm max} &= \max_{t_0}\left|\frac{1}{a}\int_{-\infty}^{\infty} df \frac{\tilde{h}_{\rm I}^*(f)\tilde{h}_{\rm II}(f)}{S_n(f)} e^{-2\pi i f t_0}\right|, \\ a &= \sqrt{\langle\tilde{h}_{\rm I}|\tilde{h}_{\rm I}\rangle\langle\tilde{h}_{\rm II}|\tilde{h}_{\rm II}\rangle}\ .\end{aligned} \tag{11}$$

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 val-

[3] `https://github.com/moble/spherical_functions`
[4] `https://github.com/moble/quaternion`

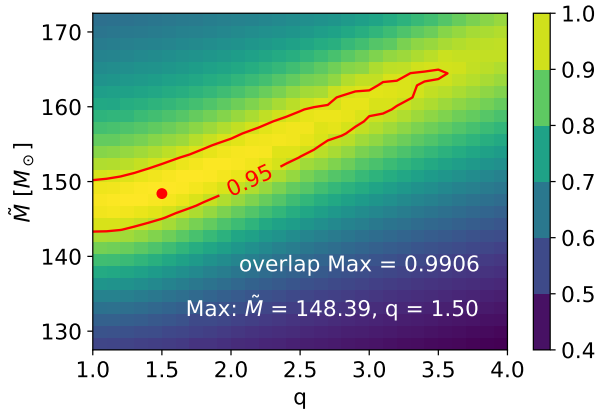

FIG. 2. Contour plot for maximized overlap for a Type-II waveform with $\tilde{M} = 150\ M_\odot$, $q = 1.7$ and $\iota = 70$ deg. Grid point with the maximum overlap is shown with the red dot at $\tilde{M} = 148.39\ M_\odot$, $q = 1.50$. 95% overlap contour is shown in red.

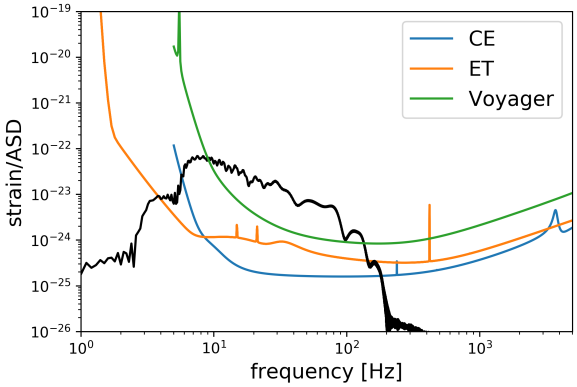

FIG. 3. Positive frequency band waveform for a binary with $\tilde{M} = 200\ M_\odot$, $q = 2.2$, $\iota = 80$ deg and $D_L = 1$ Gpc, plotted in black. The amplitude spectral densities (ASDs) for CE, ET and LIGO Voyager are plotted with colored traces. Note that ASDs for CE and LIGO Voyager are available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

ues from those of the true signal, with a maximal overlap of 99.06%.

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

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

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

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

We then interpolate between the samples using the `scipy.interpolate` module [40] to construct a function $\epsilon(\tilde{M}, q, \iota)$. For non-spinning binaries, the interpolated function ensures that 90 deg $< \iota <$ 180 deg is symmetric to 0 deg $< \iota <$ 90 deg. We perform the same analyses for CE, ET and LIGO Voyager with their respective noise power spectral density (PSD) [42].

Figure 3 shows the amplitude spectral density of CE, ET and LIGO Voyager, as well as the waveform of a 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 Section 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 non-central 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 [43]. 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

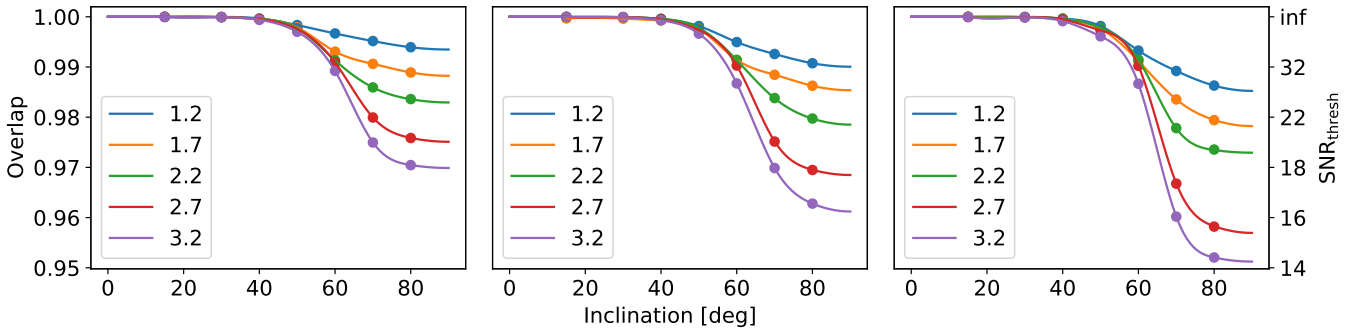

FIG. 4. Overlap between Type-I and Type-II waveforms for $\tilde{M} = 150\ M_\odot$ at selected mass ratio values. The axis on the right shows the threshold ρ to distinguish such Type-II images from Type-I counterparts by a log Bayes factor of 10, for the corresponding waveform overlap value. See Section 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.

its total ρ,

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

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

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

We then define the χ^2statistic as

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

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 non-central parameter,

$$\langle \chi^2 \rangle = p - 1 + \kappa \langle s \rangle^2 \ , \tag{15}$$

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

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

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 non-central 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 non-central 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 [44, 45]. 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 to the high-mass-ratio compact binary coalescence event GW190814 [46]. Using the waveform approximant `IMRPhenomXPHM` [47], 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 [48], we apply the 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.

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 sub-threshold 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 sub-threshold 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

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

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

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

where the subscript $i = \mathrm{I}, \mathrm{II}$ 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. [49–51], with the Laplace approximation we can write the log Bayes factor as

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

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 [51], in the high SNR limit, to scale as

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

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

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

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 the mismatch, actually. Since $\langle h_{\mathrm{I}}|h_{\mathrm{I}}\rangle \approx (1-\epsilon)^2\,\langle h_{\mathrm{II}}|h_{\mathrm{II}}\rangle$ and that $\Delta\theta_i^{j,\text{ posterior}} \propto 1/\sqrt{\langle h_i|h_i\rangle}$, therefore we have

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

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 posterior volume ratio in this paper. Therefore, we can estimate the log Bayes factor simply as [5]

$$\ln \mathcal{B} \approx \epsilon\,\rho^2. \tag{23}$$

Figure 5 shows the log Bayes factor as a function of the SNR ρ using nested sampling with the help of the library `bilby` [52] and `dynesty` [53] 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 [54] for both the simulated signals and the inference. All simulated signals have a redshifted total mass of $\tilde{M} = 150M_\odot, q = 3.2$ viewing at an inclination angle of $\iota = 80\,\mathrm{deg}$ 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 Figure 4 shows the required SNR to produce $\ln \mathcal{B} = 10$

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

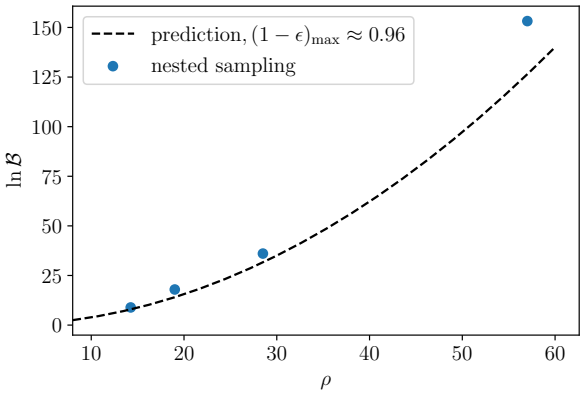

FIG. 5. The log Bayes factor $\ln \mathcal{B}$ as a function of the SNR ρ of the injections with different luminosity distances and fixed mismatch ϵ, computed using Eq. (17) with nested sampling and Eq. (23). We see that the simulation results roughly follow the expected quadratic scaling with the SNR.

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.

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 $\ln \mathcal{B}_{\rm thresh} = 10$ 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 $\ln \mathcal{B}_{\rm thresh}$ condition. Inverting $D_{L,\rm thresh}(\tilde{M}, q, \iota)$ yields $\iota_{\rm thresh}(\tilde{M}, q, D_L(z_s))$, where z_s is the GW source redshift.

To calculate ρ, we assume both GW polarizations can be detected, and the total amplitude is $\sqrt{h_+^2 + h_\times^2}$. 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 Section 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 towards 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 $f^{-7/6}$ [55]. The inspiral amplitude is matched to the surrogate waveform amplitude at 0.5 $f_{\rm ISCO}$, where $f_{\rm ISCO}$ is the Innermost Stable Circular Orbit frequency, approximated as [see, e.g., 56]

$$f_{\rm ISCO} = \frac{c^3}{6^{3/2}\pi G \tilde{M}} \,. \tag{24}$$

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-I/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 towards 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 Section 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 Section 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. We 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 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 $\ln \mathcal{B}_{\rm threshold}$ condition, and no more Type-II images can be distinguishable. The right panel shows similar threshold cures 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, $\mathrm{fr}(\tilde{M}, q, z_s)$. For simplicity, we assume that GW sources and

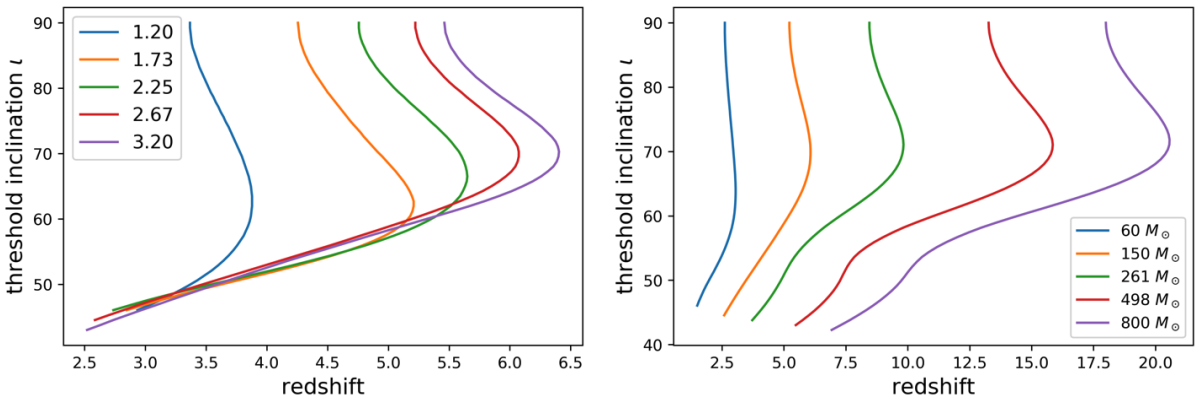

FIG. 6. Inclination threshold curves for distinguishable Type-II sources as a function of redshift assuming CE sensitivity. *Left:* inclination threshold curves for binaries with redshifted mass $\tilde{M} = 150\ M_\odot$ with selected mass ratio values. The curve-crossing at low inclination values are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

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 Figure 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 Figure 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 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 Section V. Therefore, the fraction of distinguishable Type-II sources with apparent mass M, mass ratio q at redshift z_s is given by,

$$\mathrm{fr}_{\mathrm{app}}(M, q, z_s) = \mathrm{fr}(M(1+z_s), q, z_s)\ . \tag{25}$$

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

As expected, the resulting traces show similar trends and features as in Figure 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 Figure 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 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, not only does the possible scientific output from GW detection events depend on the noise level, but also on the detector calibration. Specifically,

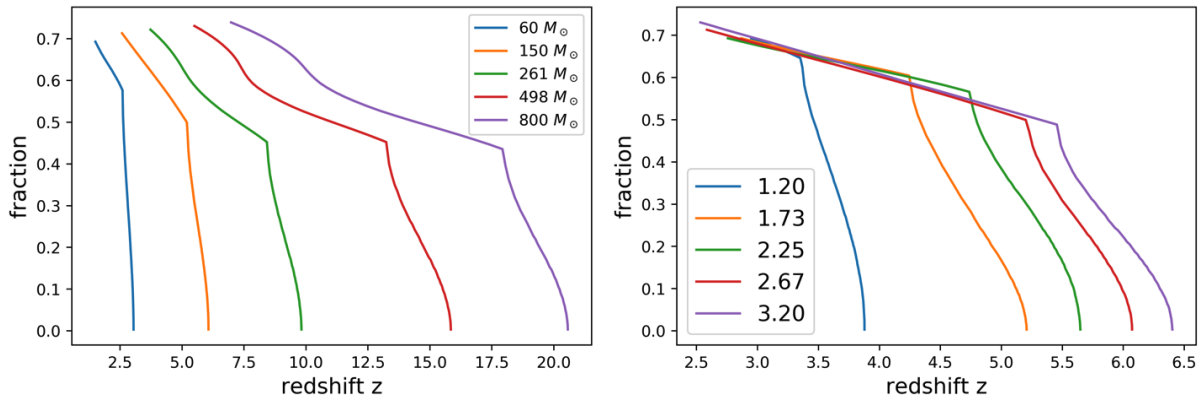

FIG. 7. The fraction of distinguishable Type-II images as a function of redshift for CE sensitivity. *Left:* distinguishable fraction, $\mathrm{fr}(\tilde{M}, q, z_s)$, for constant $\tilde{M} = 150\ M_\odot$. *Right:* distinguishable fraction for constant $q = 2.67$. The cusps in the fraction corresponds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

the calibration must 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.

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. We define a GW event as a particular 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

$$\frac{\partial^3 \dot{N}_{\mathrm{II,det}}}{\partial M_\bullet \partial q \partial z_s} = \tau_{\mathrm{II}}(z_s) \frac{\partial^3 \dot{N}}{\partial M_\bullet \partial q \partial z_s} \int d\log_{10}\mu \left[\frac{\partial P_{\mathrm{II}}(\mu, z_s)}{\partial \log_{10}\mu} \int_0^{\pi/2} d\iota\ \sin\iota\, \Theta\left(\sqrt{\mu}\rho(M_\bullet, q, z_s, \iota) - 8\right) \right] , \tag{26}$$

$$\frac{\partial^3 \dot{N}_{\mathrm{II,dis}}}{\partial M_\bullet \partial q \partial z_s} = \tau_{\mathrm{II}}(z_s) \frac{\partial^3 \dot{N}}{\partial M_\bullet \partial q \partial z_s} \int d\log_{10}\mu\ \frac{\partial P_{\mathrm{II}}(\mu, z_s)}{\partial \log_{10}\mu} \mathrm{fr}\left(\tilde{M}, q, \tilde{z}_s\right) , \tag{27}$$

where $\partial^3 \dot{N}/\partial M_\bullet \partial q \partial z_s$ is GW event rate per intrinsic binary mass, $M_\bullet$, 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 $\tau_{\mathrm{II}}(z_s)$, we obtain the rate of events with at least one Type-II images. The quantity $\partial P_{\mathrm{II}}(\mu, z_s)/\partial \log_{10}\mu$ describes the distribution of magnification μ for Type-II images for sources at z_s, normalized such that

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

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

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

The differential merger number per observer time is calculated as [see also 18]

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

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 Figure 12, we plot this "modified" differential comoving volume and the total merger rates for reference.

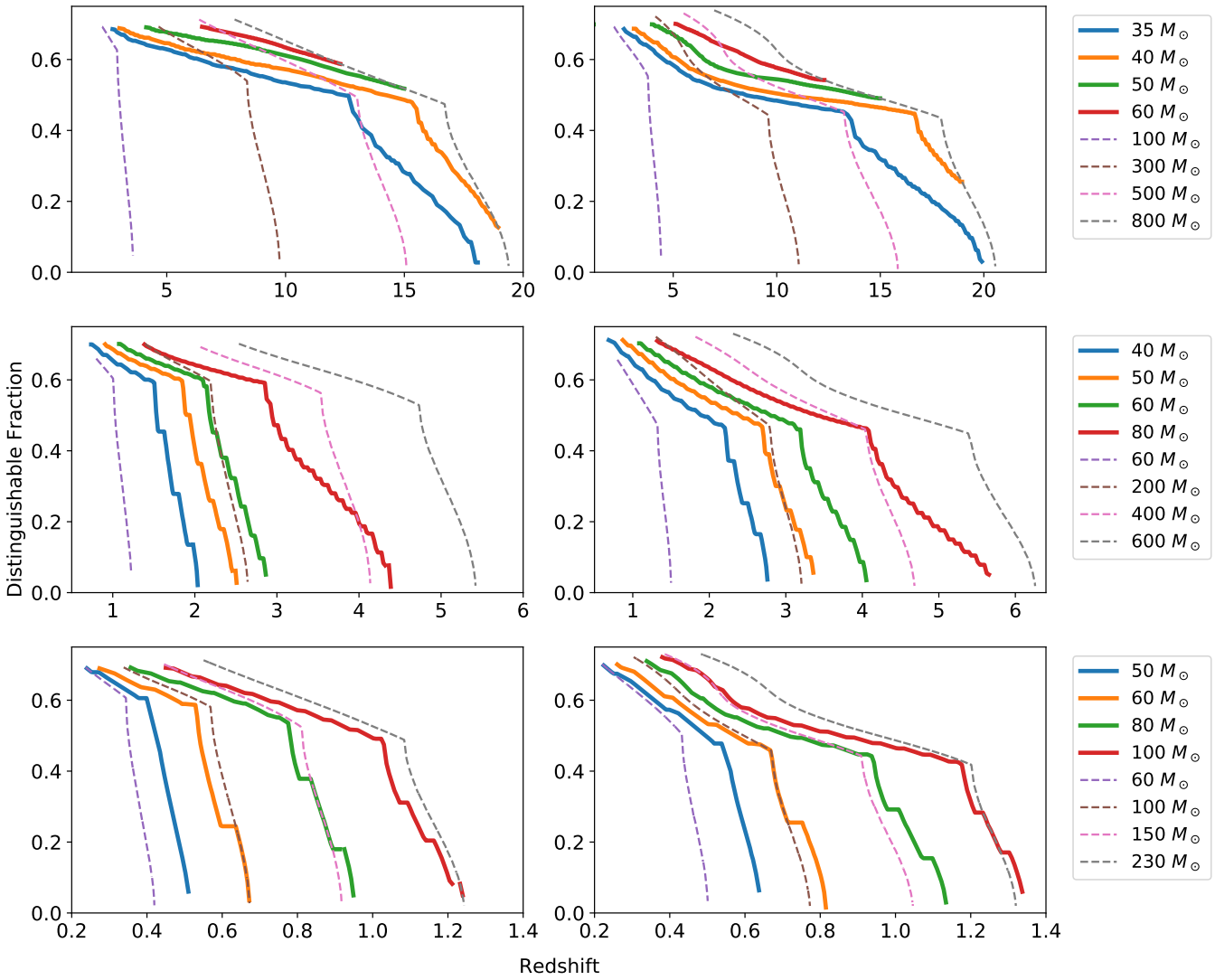

FIG. 8. The fraction of distinguishable Type-II images as a function of redshift. *Left Column:* GW sources have constant mass ratio of 1.73. *Right Column:* GW sources have constant mass ratio of 2.67. The *Top, Middle,* and *Bottom* rows show distinguishable fractions assuming CE, ET and LIGO Voyager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

Since we have adopted the same population models, Figure 12 replicates Figure 1 in [10].

For fast calculation of $\rho(M_\bullet, q, z_s, \iota)$, we use the phenomenological model `IMRPhenomHM` [57], called from the Python package `pycbc.waveform` [58].

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 A. We then make concrete detection population for CE, ET and LIGO Voyager and discuss results.

A. Optical Depth and Magnification

To obtain $\tau_{\mathrm{II}}(z_s)$ and $\partial P_{\mathrm{II}}(\mu, z_s)/\partial \log_{10}\mu$, we perform a Monte Carlo simulation. We consider elliptical galaxies as

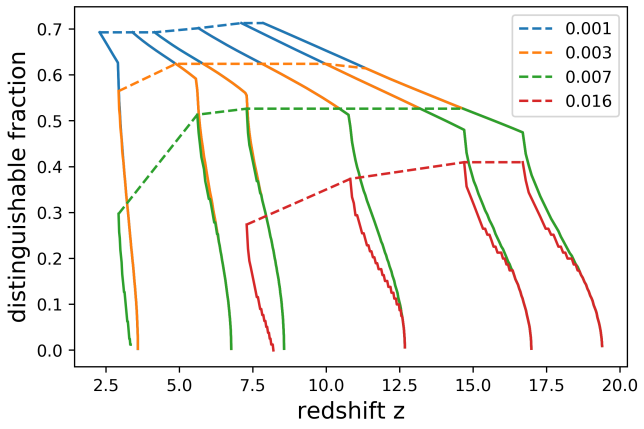

FIG. 9. Distinguishable fractions of sources with Type-II images assuming a minimum resolvable mismatch of 0.001, 0.003, 0.007 and 0.016 with CE sensitivity. The solid lines show the distinguishable fractions for selected redshifted mass. The mass ratio is fixed to be 1.73. From left to right, the traces 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.

lenses, as they are expected to be the predominant lensing objects [59]. Such lenses can be modeled as singular isothermal ellipsoid [see e.g., 10]. 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 [60].

At each redshift, we generate samples of lenses, parameterized 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 [61], which is calibrated to observation on galaxies in the solar neighborhood,

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

where $\phi_* = 8.0 \times 10^{-3} h^3\ \mathrm{Mpc}^{-3}$, $\sigma_* = 161$ 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

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

where $\Psi_{\rm hyd}(\sigma_v, z_l)$ is the velocity dispersion function derived from hydrodynamical simulation in [62].

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]

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

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][63] 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 $\tau_{\rm II}$, while depending on the solution for μ, it contributes to $\partial P_{\rm II}/\partial \log_{10}\mu$ 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.

Figure 10 shows the Type-II image optical depth at various source redshifts. Optical depths smaller than 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 $10^{-3} \sim 10^{-4}$, consistent with results from ray-tracing studies using N-body simulations [see, e.g., 64, 65].

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

[6] `https://github.com/sibirrer/lenstronomy`

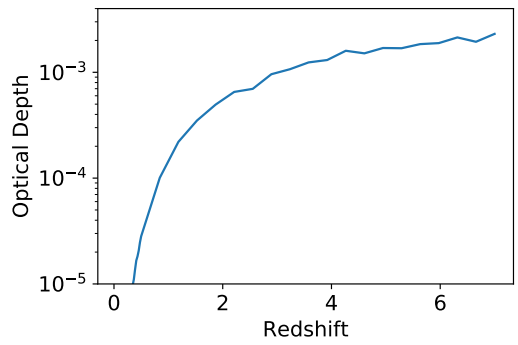

FIG. 10. Optical depths, $\tau_{\rm II}(z_s)$, for GW sources with at least one Type-II image as a function of source redshift. Optical depths lower than 10^{-5} are omitted, as they are too low to predict an observable GW source at such redshifts with Type-II images in future detectors.

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 $\partial P_{\rm II}(\mu, z_s)/\partial \log_{10}\mu$ is independent from source redshift z_s by construction, as Figure 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 $dP_{\rm II}(\mu)/d\log_{10}\mu$ by a log normal distribution with a mean of -0.35 and standard deviation of 0.57, truncated at $\log_{10}\mu = -2, 1$.

B. GW Source Population

We adopt GW source population models provided by [10, 66]. 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, 66], and we provide more details in Appendix A for reference.

Figure 12 replicates Figure 1 in [10] and shows the predicted BBH merger rate density under the two Star Formation Rate (SFR) models in [67] and [68], respectively. Due to the intrinsic uncertainty in these analytical SFR models, we choose one, the more optimistic SFR in [68], for the following population estimates.

The total BBH merger rate per source redshift and the detectable merger rate are plotted in Figure 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. 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.

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

Figure 14 plots the same population prediction binned by the total mass of the BBH, with consistent line styles as in Figure 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

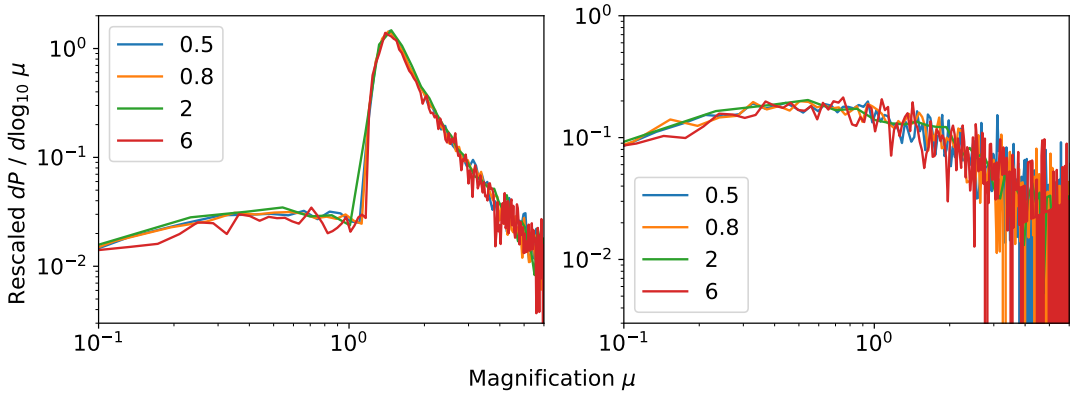

FIG. 11. Rescaled magnification distribution at redshift $z = 0.5, 0.8, 2, 6$. The *left* panel shows $\partial P(\mu, z_s)/\partial \log_{10}\mu$, including all images from the Monte Carlo samples. The *right* panel shows the rescaled distribution of only Type-II images, $\partial P_{\rm II}(\mu, z_s)/\partial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

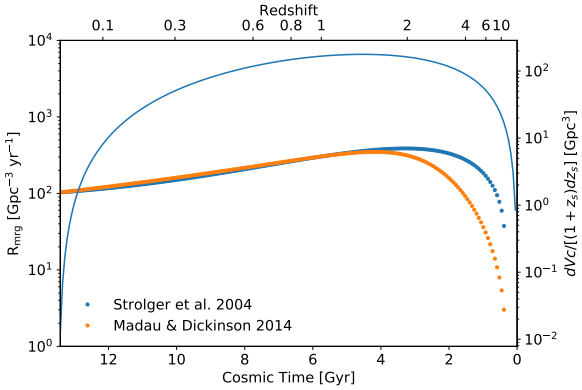

FIG. 12. BBH merger rate density, assuming two SFR models in [67] and [68]. The blue over-arching trace plots the modified differential comoving volume and corresponds to values on the right axis. The merger rate density is directly analogous to Figure 1 in [10].

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 1.

D. 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 Figure 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 $\epsilon_{\rm min} = 0.001$, which excludes the binaries at small inclinations, and the distinguishable fraction is "saturated" at roughly 70% (see, e.g., Figure 7 and Figure 9). As Figure 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 Figure 9 shows, the distinguishable fraction roughly doubles as the mismatch resolution improves from $O(1\%)$ to $O(0.1\%)$ 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

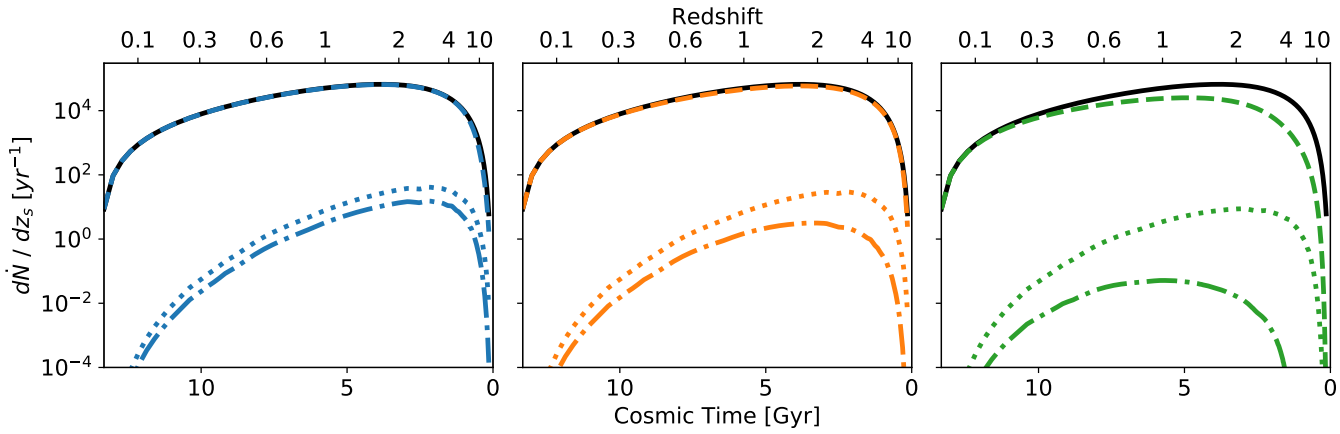

FIG. 13. Yearly detected population per unit reshift prediction as a function of redshift. The panels from left to right show the detection population for CE, ET and LIGO Voyager. In all panels, the solid black line denotes the total BBH merger rate. The dashed curves show the rate of detectable GW sources (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.

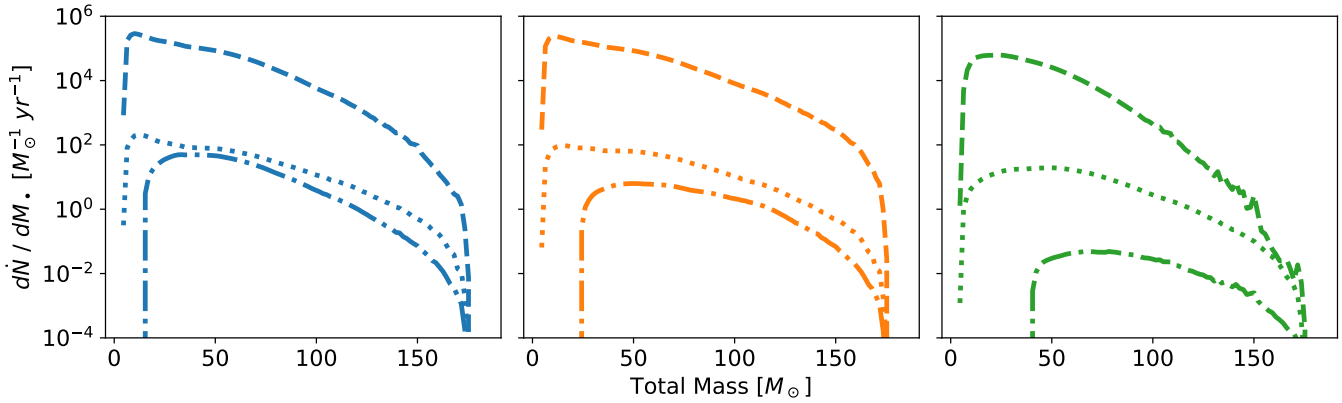

FIG. 14. Detection rate as a function of BBH intrinsic mass. The panels from left to right shows the detection prediction for CE, ET and LIGO Voyager. The line styles are consistent with those in Figure 13.

requirement has two implications for third-generation GW detector performance and data analysis process. As is discussed briefly in Section 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 $\ln \mathcal{B}_{\rm thresh} = 5$, 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 $\ln \mathcal{B}_{\rm thresh}$, normalized to the number when $\ln \mathcal{B}_{\rm thresh} = 10$. 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 $\ln \mathcal{B}_{\rm thresh} = 2$, 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 $\ln \mathcal{B}_{\rm thresh} = 2$, 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 $\sim 1/\mathrm{yr}$, 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 1.

In addition, we have so far considered the single-detector scenario, and we estimate the advantage of a detector network

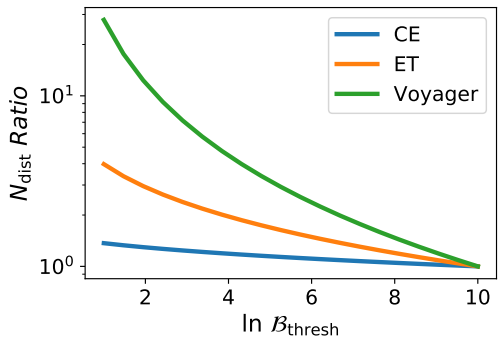

FIG. 15. The increase in the number of GW sources with distinguishable Type-II images for different log Bayes factor threshold values. The detection numbers are normalized to the number with $\ln \mathcal{B}_{\rm thresh} = 10$, for each detector.

		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%)

TABLE 1. 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.

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 [29]. 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 IndiGO[7] [69]. This global network offers increased detector-networks ρ and an increased detection spatial resolution. 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 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 sub-threshold search method described in [44] on an example high-mass-ratio compact binary coalescence event GW190814 [46]. 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 $\ln \mathcal{B}_{\rm thresh} = 10$. A relaxed log Bayes factor threshold boosts the expected detection rates, especially for LIGO Voyager; at $\ln \mathcal{B}_{\rm thresh} = 2$, the yearly detection rate for LIGO Voyager approaches $1/\mathrm{yr}$.

Such distinguishable Type-II images are "short-cuts" 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. 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

[7] http://www.gw-indigo.org/tiki-index.php

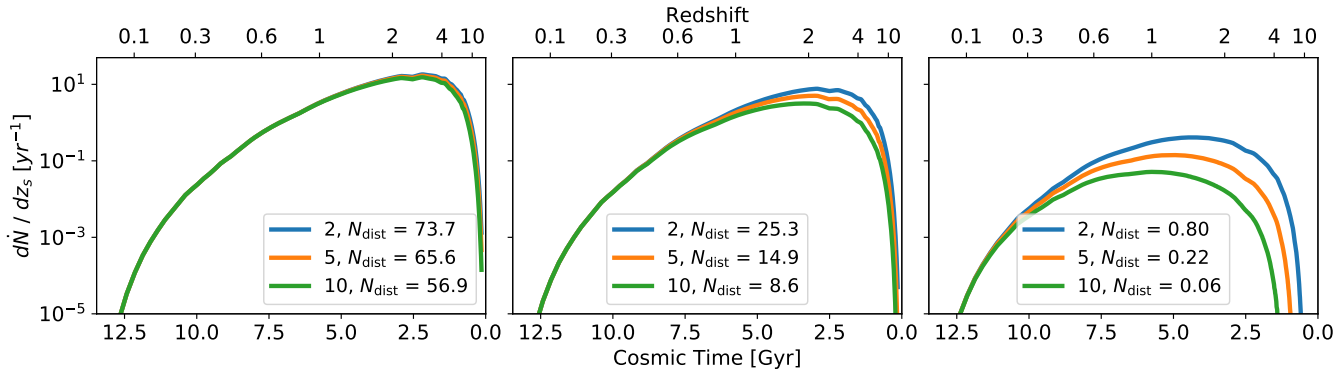

FIG. 16. The expected number of GW sources with distinguishable Type-II images with $\ln \mathcal{B}_{\rm thresh} = 2,\ 5$, and 10. The expected yearly detection count for each threshold value is shown in the legend. The panels from left to right correspond to CE, ET and LIGO Voyager. As expected, the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

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.

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

YW would like to thank the David and Ellen Lee Distinguished Fellowship for support during this research. Research of YW and YC are supported by the Simons Foundation (Award Number 568762), and the National Science Foundation, through Grants PHY-2011961, PHY-2011968, and PHY-1836809. RKLL 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 Number LIGO-P2100002.

[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, others, LIGO Scientific Collaboration, and Virgo Collaboration, ApJ **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].

[2] The LIGO Scientific Collaboration, the Virgo Collaboration, 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.*, arXiv e-prints , arXiv:2009.01190 (2020), arXiv:2009.01190 [astro-ph.HE].

[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), arXiv:1801.07840 [gr-qc].

[4] A. K. Meena and J. S. Bagla, MNRAS **492**, 1127 (2020), arXiv:1903.11809 [astro-ph.CO].

[5] K. Pardo, M. Fishbach, D. E. Holz, and D. N. Spergel, J. Cosmology Astropart. Phys. **2018**, 048 (2018), arXiv:1801.08160 [gr-qc].

[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, others, LIGO Scientific Collaboration, and Virgo Collaboration, Phys. Rev. D **100**, 104036 (2019), arXiv:1903.04467 [gr-qc].

[7] S. Perkins and N. Yunes, Classical and Quantum Gravity **36**, 055013 (2019), arXiv:1811.02533 [gr-qc].

[8] A. Vijaykumar, M. V. S. Saketh, S. Kumar, P. Ajith, and T. R. Choudhury, arXiv e-prints , arXiv:2005.01111 (2020), arXiv:2005.01111 [astro-ph.CO].

[9] M. Biesiada, X. Ding, A. Piórkowska, and Z.-H. Zhu, J. Cosmology Astropart. Phys. **2014**, 080 (2014), arXiv:1409.8360 [astro-ph.HE].

[10] S.-S. Li, S. Mao, Y. Zhao, and Y. Lu, MNRAS **476**, 2220 (2018), arXiv:1802.05089 [astro-ph.CO].

[11] M. Oguri, MNRAS **480**, 3842 (2018), arXiv:1807.02584 [astro-ph.CO].

[12] X. Ding, M. Biesiada, and Z.-H. Zhu, J. Cosmology Astropart. Phys. **2015**, 006 (2015), arXiv:1508.05000 [astro-ph.HE].

[13] K. Liao, X.-L. Fan, X. Ding, M. Biesiada, and Z.-H. Zhu, Nature Communications **8**, 1 (2017).

[14] L. Dai and T. Venumadhav, arXiv e-prints , arXiv:1702.04724 (2017), arXiv:1702.04724 [gr-qc].

[15] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, Phys. Rev. D **97**, 023012 (2018), arXiv:1703.06319 [astro-ph.CO].

[16] X.-L. Fan, K. Liao, M. Biesiada, A. Piórkowska-Kurpas, and Z.-H. Zhu, Physical Review Letters **118**, 091102 (2017).

[17] T. E. Collett and D. Bacon, Phys. Rev. Lett. **118**, 091101 (2017), arXiv:1602.05882 [astro-ph.HE].

[18] L. Dai, T. Venumadhav, and K. Sigurdson, Phys. Rev. D **95**, 044011 (2017), arXiv:1605.09398 [astro-ph.CO].

[19] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, ApJ **874**, L2 (2019), arXiv:1901.02674 [gr-qc].

[20] L. Dai, B. Zackay, T. Venumadhav, J. Roulet, and M. Zaldarriaga, arXiv e-prints , arXiv:2007.12709 (2020), arXiv:2007.12709 [astro-ph.HE].

[21] T. Broadhurst, J. M. Diego, and I. Smoot, George F., arXiv e-prints , arXiv:1901.03190 (2019), arXiv:1901.03190 [astro-ph.CO].

[22] T. Broadhurst, J. M. Diego, and I. Smoot, George, arXiv e-prints , arXiv:1802.05273 (2018), arXiv:1802.05273 [astro-ph.CO].

[23] T. T. Nakamura, Phys. Rev. Lett. **80**, 1138 (1998).

[24] R. Takahashi and T. Nakamura, ApJ **595**, 1039 (2003), arXiv:astro-ph/0305055 [astro-ph].

[25] M. H. Y. Cheung, J. Gais, O. A. Hannuksela, and T. G. F. Li, arXiv e-prints , arXiv:2012.07800 (2020), arXiv:2012.07800 [astro-ph.HE].

[26] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (1992).

[27] J. María Ezquiaga, D. E. Holz, W. Hu, M. Lagos, and R. M. Wald, arXiv e-prints , arXiv:2008.12814 (2020), arXiv:2008.12814 [gr-qc].

[28] 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 and Quantum Gravity **37**, 165003 (2020), arXiv:2001.11173 [astro-ph.IM].

[29] 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 e-prints , arXiv:1108.1423 (2011), arXiv:1108.1423 [gr-qc].

[30] 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*, Vol. 51 (2019) p. 35, arXiv:1907.04833 [astro-ph.IM].

[31] T. T. Nakamura and S. Deguchi, Progress of Theoretical Physics Supplement **133**, 137 (1999).

[32] M. Favata, Classical and Quantum Gravity **27**, 084036 (2010), arXiv:1003.3486 [gr-qc].

[33] M. Favata, Phys. Rev. D **80**, 024002 (2009), arXiv:0812.0069 [gr-qc].

[34] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, Phys. Rev. D **101**, 023011 (2020), arXiv:1911.12496 [astro-ph.HE].

[35] M. Zaldarriaga, D. Kushnir, and J. A. Kollmeier, MNRAS **473**, 4174 (2018), arXiv:1702.00885 [astro-ph.HE].

[36] 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, others, LIGO Scientific Collaboration, and Virgo Collaboration, Phys. Rev. D **100**, 064064 (2019), arXiv:1907.09384 [astro-ph.HE].

[37] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Physical Review Research **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].

[38] S. Field, C. Galley, and J. Blackman, in *APS April Meeting Abstracts*, APS Meeting Abstracts, Vol. 2018 (2018) p. G14.005.

[39] S. van der Walt, S. C. Colbert, and G. Varoquaux, Computing in Science and Engineering **13**, 22 (2011), arXiv:1102.1523 [cs.MS].

[40] 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.*, Nature Methods **17**, 261 (2020).

[41] B. Moore, T. Robson, N. Loutrel, and N. Yunes, Classical and Quantum Gravity **35**, 235006 (2018), arXiv:1807.07163 [gr-qc].

[42] 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*, Tech. Rep. LIGO-T1500293-v11 (2018).

[43] B. Allen, Phys. Rev. D **71**, 062001 (2005), arXiv:gr-qc/0405045 [gr-qc].

[44] 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 e-prints , arXiv:1904.06020 (2019), arXiv:1904.06020 [gr-qc].

[45] C. McIsaac, D. Keitel, T. Collett, I. Harry, S. Mozzon, O. Edy, and D. Bacon, Phys. Rev. D **102**, 084031 (2020), arXiv:1912.05389 [gr-qc].

[46] 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, others, LIGO Scientific Collaboration, and Virgo Collaboration, ApJ **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].

[47] 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 e-prints , arXiv:2004.06503 (2020), arXiv:2004.06503 [gr-qc].

[48] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cariou, S. Caudill, S. J. Chamberlin, J. D. E. Creighton, R. Everett, C. Hanna, D. Keppel, R. N. Lang, T. G. F. Li, D. Meacher, A. Nielsen, C. Pankow, S. Privitera, H. Qi, S. Sachdev, L. Sadeghian, L. Singer, E. G. Thomas, L. Wade, M. Wade, A. Weinstein, and K. Wiesner, Phys. Rev. D **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].

[49] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius, Phys. Rev. D **84**, 062003 (2011), arXiv:1105.2088 [gr-qc].

[50] M. Vallisneri, Phys. Rev. D **86**, 082001 (2012), arXiv:1207.4759 [gr-qc].

[51] W. Del Pozzo, K. Grover, I. Mandel, and A. Vecchio, Class. Quant. Grav. **31**, 205006 (2014), arXiv:1408.2356 [gr-qc].

[52] G. Ashton *et al.*, Astrophys. J. Suppl. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[53] J. S. Speagle, MNRAS **493**, 3132 (2020), arXiv:1904.02180 [astro-ph.IM].

[54] 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), arXiv:2001.10914 [gr-qc].

[55] C. Cutler and É. E. Flanagan, Phys. Rev. D **49**, 2658 (1994), arXiv:gr-qc/9402014 [gr-qc].

[56] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, ApJ **611**, 623 (2004), arXiv:astro-ph/0401543 [astro-ph].

[57] C. Kalaghatgi, M. Hannam, and V. Raymond, Phys. Rev. D **101**, 103004 (2020), arXiv:1909.10010 [gr-qc].

[58] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. Dal Canton, C. Capano, L. Pekowsky, T. Dent, A. R. Williamson, G. S. Davies, S. De, M. Cabero, B. Machenschalk, P. Kumar, S. Reyes, D. Macleod, Dfinstad, F. Pannarale, T. Massinger, S. Kumar, M. Tápai, L. Singer, S. Khan, S. Fairhurst, A. Nielsen, S. Singh, and Shasvath, gwastro/pycbc: PyCBC release v1.16.11 (2020).

[59] M. Oguri and P. J. Marshall, MNRAS **405**, 2579 (2010), arXiv:1001.2037 [astro-ph.CO].

[60] S. L. Finkelstein, PASA **33**, e037 (2016), arXiv:1511.05558 [astro-ph.GA].
[61] Y.-Y. Choi, C. Park, and M. S. Vogeley, ApJ **658**, 884 (2007), arXiv:astro-ph/0611607 [astro-ph].
[62] P. Torrey, S. Wellons, F. Machado, B. Griffen, D. Nelson, V. Rodriguez-Gomez, R. McKinnon, A. Pillepich, C.-P. Ma, M. Vogelsberger, V. Springel, and L. Hernquist, MNRAS **454**, 2770 (2015), arXiv:1507.01942 [astro-ph.GA].
[63] S. Birrer and A. Amara, Physics of the Dark Universe **22**, 189 (2018), arXiv:1803.09746 [astro-ph.CO].
[64] S. Hilbert, S. D. M. White, J. Hartlap, and P. Schneider, MNRAS **382**, 121 (2007), arXiv:astro-ph/0703803 [astro-ph].
[65] S. Hilbert, S. D. M. White, J. Hartlap, and P. Schneider, MNRAS **386**, 1845 (2008), arXiv:0712.1593 [astro-ph].
[66] L. Cao, Y. Lu, and Y. Zhao, MNRAS **474**, 4997 (2018), arXiv:1711.09190 [astro-ph.GA].
[67] P. Madau and M. Dickinson, ARA&A **52**, 415 (2014), arXiv:1403.0007 [astro-ph.CO].
[68] 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.*, ApJ **613**, 200 (2004), arXiv:astro-ph/0406546 [astro-ph].
[69] 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]*, Tech. Rep. LIGO-M1100296-v2 (2011).
[70] M. Spera, M. Mapelli, and A. Bressan, MNRAS **451**, 4086 (2015), arXiv:1505.05201 [astro-ph.SR].
[71] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy, Nature **534**, 512 (2016), arXiv:1602.04531 [astro-ph.HE].
[72] G. Chabrier, PASP **115**, 763 (2003), arXiv:astro-ph/0304382 [astro-ph].

Appendix A: 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 $m_\bullet$ at redshift z_s, $R_{\rm birth}(m_\bullet, z_s)$, is given by

$$R_{\rm mrg}(m_\bullet, z_s) = \int dm_\star\, dt_d\, dZ\, \phi(m_\star)\dot{\psi}\left(t\left(z_s\right) - t_d\right) P(t_d) P\left(Z, t\left(z_s\right) - t_d\right) \delta\left[m_\star - g^{-1}(m_\bullet, Z)\right] , \tag{A1}$$

where $m_\star$ is the mass of the progenitor star, Z is stellar metallicity and $t(z_s)$ is the cosmic time as a function of redshift. $g^{-1}(m_\bullet, Z)$ gives the stellar mass $m_\star$ with metallicity Z that leaves a black hole remnant with mass $m_\bullet$. The expression of remnant black hole mass as a function of stellar mass and metallicity is given in [70], with $20\ M_\odot < m_\star < 105\ M_\odot$ 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 [71]. At each redshift, the metallicity follows a log normal distribution [71].

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 $\phi(m_\star)$ 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 [72] for $m_\star > 1\ M_\odot$, where $\phi(m_\star) \propto m_\star^{-2.3}$. The quantity $\dot{\psi}(t)$ is the Star Formation Rate (SFR) including all $m_\star$ at cosmic time t. We adopt the analytic SFR expression in [68]. We calibrate the merger rate at $z = 0$ to be 103 $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$, which is the expected local black hole merger rate given LIGO detection data [66].

We note that $R_{\rm mrg}(m_\bullet, z_s)$ is the rates for black hole binary at z_s whose *primary* black hole, i.e. the heavier one, is $m_\bullet$. 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 $R_{\rm mrg}(M_\bullet, q, z_s)$ where $M_\bullet$ is the total binary mass.