

# Astrophysical gravitational-wave echoes from galactic nuclei

László Gondán <sup>1</sup>★ and Bence Kocsis <sup>2</sup>

<sup>1</sup>ELTE – Eötvös Loránd University, Pázmány Péter sétány 1/A, Budapest 1117, Hungary

<sup>2</sup>Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

Accepted 2022 July 11. Received 2022 June 21; in original form 2021 October 18

## ABSTRACT

Galactic nuclei (GNs) are dense stellar environments abundant in gravitational-wave (GW) sources for the Laser Interferometer Gravitational-Wave Observatory (LIGO), Virgo, and Kamioka Gravitational Wave Detector (KAGRA). The GWs may be generated by stellar-mass black hole (BH) or neutron star mergers following gravitational bremsstrahlung, dynamical scattering encounters, Kozai–Lidov-type oscillations driven by the central supermassive black hole (SMBH), or gas-assisted mergers if present. In this paper, we examine a smoking gun signature to identify sources in GNs: the GWs scattered by the central SMBH. This produces a secondary signal, an astrophysical GW echo, which has a very similar time–frequency evolution as the primary signal but arrives after a time delay. We determine the amplitude and time-delay distribution of the GW echo as a function of source distance from the SMBH. Between  $\sim 10$  per cent and 90 per cent of the detectable echoes arrive within  $\sim (1\text{--}100)M_6$  s after the primary GW for sources between 10 and  $10^4$  Schwarzschild radius, where  $M_6 = M_{\text{SMBH},z}/(10^6 M_\odot)$ , and  $M_{\text{SMBH},z}$  is the observer-frame SMBH mass. The echo arrival times are systematically longer for high signal-to-noise ratio (SNR) primary GWs, where the GW echo rays are scattered at large deflection angles. In particular,  $\sim 10$  per cent–90 per cent of the distribution is shifted to  $\sim (5\text{--}1800)M_6$  s for sources, where the lower limit of echo detection is 0.02 of the primary signal amplitude. We find that  $\sim 5$  per cent–30 per cent ( $\sim 1$  per cent–7 per cent) of GW sources have an echo amplitude larger than 0.2–0.05 times the amplitude of the primary signal if the source distance from the SMBH is 50 (200) Schwarzschild radius. Non-detections can rule out that a GW source is near an SMBH.

**Key words:** black hole physics – gravitational lensing: strong – gravitational waves – galaxies: active – galaxies: nuclei.

## 1 INTRODUCTION

The detection of gravitational waves (GWs) from stellar-mass black hole (BH) and neutron star mergers by the Advanced Laser Interferometer Gravitational-Wave Observatory<sup>1</sup> (aLIGO; Aasi et al. 2015), Advanced Virgo<sup>2</sup> (AdV; Acernese et al. 2015), and Kamioka Gravitational Wave Detector (KAGRA;<sup>3</sup> Kagra Collaboration et al. 2019) has opened the field of GW astronomy (Abbott et al. 2016, 2017, 2019, 2021a). In the past years, a significant effort has been invested in explaining how these compact objects may form binaries and merge as frequently as observed to constrain the possible astrophysical origin of the detected mergers (e.g. Barack et al. 2019; Abbott et al. 2020, 2021b, and references therein).

GWs of compact binaries encode masses, mass ratios, spins (e.g. Cutler & Flanagan 1994; Poisson & Will 1995), and, in some cases, orbital eccentricity (e.g. Gondán et al. 2018a; Lower et al. 2018; Gondán & Kocsis 2019; Romero-Shaw, Lasky & Thrane 2019). The GW signal can also carry information on the formation environment through external effects affecting the GW signal. Gas dynamics affect

the waveform by slowing down or accelerating the inspiral (e.g. Barausse & Rezzolla 2008; Kocsis, Yunes & Loeb 2011; Yunes et al. 2011; Barausse, Cardoso & Pani 2014; D’Orazio & Loeb 2018; Derdzinski et al. 2019, 2021; Cardoso & Maselli 2020; Toubiana et al. 2021). Furthermore, there are multiple situations in which the merger environment can affect the GW signal. An intervening mass distribution (e.g. Kocsis et al. 2006; Bonvin et al. 2017; Randall & Xianyu 2019) or the relative motion of the GW source due to peculiar velocities of GW sources and the orbital motion around a companion (e.g. Inayoshi et al. 2017; Meiron, Kocsis & Loeb 2017; Robson et al. 2018; Chamberlain et al. 2019; Wong, Baibhav & Berti 2019; Tamanini et al. 2020; Torres-Orjuela, Chen & Amaro-Seoane 2020) can result in Doppler boosting, gravitational lensing of the rest-frame GWs, time-dependent gravitational redshift, GW phase shift, or the Shapiro delay.

High-frequency GWs emitted by stellar/intermediate-mass compact binaries are scattered by the gravitational field of a supermassive black hole (SMBH; e.g. Futterman, Handler & Matzner 1988; Dolan 2008) producing second and higher order scattered GWs analogous to the relativistic images of electromagnetic (EM) waves (e.g. Dolan & Ottewill 2011; Zenginoğlu & Galley 2012). These second and higher order scattered GW signals have the same waveform as the unperturbed signal in the geometrical optics limit, including the inspiral, merger, and ringdown waveforms. They only differ in amplitude, arrival time, overall phase, and polarization

\* E-mail: laszlo.gondan@ttk.elte.hu

<sup>1</sup><http://www.ligo.caltech.edu/>

<sup>2</sup><http://www.ego-gw.it/>

<sup>3</sup><https://gwcenter.icrr.u-tokyo.ac.jp/en/>

angle (Dai & Venumadhav 2017; Ezquiaga et al. 2021).<sup>4</sup> Similar to EM waves, the amplitudes of the scattered GW signals decay asymptotically exponentially with the order of the scattered waves (e.g. Zenginoğlu & Galley 2012; Kocsis 2013), indicating that the detections of scattered secondary GWs are typically the most prominent besides the primary GWs. Note that the scattering of GWs and second and higher order scattered GWs are referred to as lensing and multiple images, respectively, in the GW lensing community.<sup>5</sup>

In this study, we examine GW lensing and multiple images due to a single SMBH in the vicinity of the stellar-mass compact binary sources generating a secondary GW signal (hereafter referring to it as an *astrophysical GW echo*). We investigate the main characteristics of the GW echo signal, including its amplitude and time delay relative to the primary GW signal. The detection of an astrophysical GW echo will be a smoking gun signature of GW sources in close proximity to SMBHs in galactic nucleus (GN) host environments. Note that sequences of distinct GW pulses also referred to as GW echoes are expected from exotic compact objects (e.g. Cardoso & Pani 2017; Maggio, Pani & Ferrari 2017; Mark et al. 2017; Bueno et al. 2018) and when going beyond the description of BHs by classical general relativity (e.g. Abedi, Dykaar & Afshordi 2017; Barceló, Carballo-Rubio & Garay 2017; Dong & Stojkovic 2021) as well.

Several source populations are expected to emit GWs in the vicinity of SMBHs in the inner regions of GNs for ground-based GW detectors.

(i) Because of the extremely high stellar number densities in GNs reaching up to  $10^{10} \text{ pc}^{-3}$ , the probability of very close dynamical encounters among single objects is non-negligible, where the binary forms and merges due to GW emission (O’Leary, Kocsis & Loeb 2009). The heavy BHs sink to the densest central regions due to mass segregation, where the velocity dispersion is so high that only very close encounters result in captures. This population of mergers represents the highest eccentricity GW sources in the Universe, which typically form in the aLIGO/AdV/KAGRA frequency band (Kocsis & Levin 2012; Gondán et al. 2018b; Gondán & Kocsis 2021).

(ii) The long-term gravitational perturbations due to the central SMBH drive variations in the binary’s orbital eccentricity, the so-called Kozai–Lidov (KL) effect (Kozai 1962; Lidov 1962; Lidov & Ziglin 1976). In some cases, this effect may lead to close encounters between the binary components, after which the binary merges due to GW emission (Antonini & Perets 2012; Petrovich & Antonini 2017; Hamers et al. 2018; Hoang et al. 2018; Randall & Xianyu 2018a,b; Fragione et al. 2019; Liu & Lai 2020).

(iii) In active galactic nuclei (AGNs), the gaseous accretion disc around the SMBH may facilitate binary formation and mergers of stellar-mass compact objects (McKernan et al. 2014; Bellovary et al. 2016; Bartos et al. 2017; Stone, Metzger & Haiman 2017; Tagawa, Haiman & Kocsis 2020a). In this scenario, BHs form *in situ* in the vicinity of a GN and sink to the inner region due to mass segregation or they are delivered to these regions by infalling globular clusters

(Morris 1993; Miralda-Escudé & Gould 2000; Freitag, Amaro-Seoane & Kalogera 2006; Hopman & Alexander 2006; O’Leary et al. 2009; Antonini 2014), then get captured in the disc by hydrodynamic drag as they cross the disc (e.g. Goldreich, Lithwick & Sari 2002; Bartos et al. 2017; Yang et al. 2019b; Tagawa et al. 2020a). Alternatively, some BHs may have formed in the disc itself (Levin 2007; Stone et al. 2017). Once in the disc, BHs get transported to the inner regions by exchanging angular momentum with the surrounding gas (Goldreich & Tremaine 1979). In certain regions, the BHs open an annular gap in the accretion disc and accumulate in a narrow range of radii, the so-called migration traps (Bellovary et al. 2016; Secunda et al. 2019, 2020, 2021). Bellovary et al. (2016) argue that migration traps may be expected to be close to the SMBH from  $\sim 20$  to  $\sim 300$  Schwarzschild radii ( $r_s = 2GM_{\text{SMBH}}/c^2$ ) from the central SMBH of mass  $M_{\text{SMBH}}$ .<sup>6</sup> However, they may exist in slim discs near the innermost stable circular orbit (ISCO; Peng & Chen 2021) or near the boundary of a gap region if a gap opens due to a heavy stellar-mass BH or an intermediate-mass BH (McKernan et al. 2014). Dynamical encounters frequently happen in migration traps leading to the formation and subsequent merger of binary black holes (BBHs) on short time-scales (Secunda et al. 2019, 2020; Yang et al. 2019a), where the binary separation is efficiently reduced by gas dynamical friction (Escala et al. 2004; Kim & Kim 2007; Baruteau, Cuadra & Lin 2011) to the point where GW emission drives the binaries together. Alternatively, BBHs may also form and merge in the disc outside migration traps (Tagawa et al. 2020a). Because of the deep potential barrier of the SMBH, the merger remnant BH remains near the migration trap and may undergo subsequent mergers with additional BHs, which leads to high BH masses, characteristic spin properties, and possibly non-zero eccentricity identifiable via GW observations (Yang et al. 2019a; Secunda et al. 2020; Tagawa et al. 2020b, 2021a,b; Samsing et al. 2022).

Astrophysical GW echoes are produced in GNs due to strong gravitational lensing by the SMBH. The lensing of GWs has been investigated in the literature in the wave optics and the geometrical optics regimes (e.g. Takahashi & Nakamura 2003) on very different scales, including lensing by the large-scale structure of the Universe (Hilbert et al. 2007, 2008; Robertson et al. 2020), galaxy clusters and galaxies (e.g. Wang, Stebbins & Turner 1996; Takahashi & Nakamura 2003; Dai & Venumadhav 2017; Takahashi 2017; Broadhurst, Diego & Smoot 2018; Li et al. 2018; Smith et al. 2018), intermediate-mass BH in globular clusters (e.g. Kains et al. 2016; Lai et al. 2018; Tatekawa & Okamura 2020), and populations of stellar-mass microlenses in galaxies (e.g. Wang et al. 1996; Cheung et al. 2021; Mishra et al. 2021; Meena et al. 2022). In addition, lensing by SMBHs has been investigated in the geometrical optics limit for the GWs of BBHs detectable by aLIGO/AdV/KAGRA (Kocsis 2013) and the *Laser Interferometer Space Antenna* (LISA;<sup>7</sup> Amaro-Seoane et al. 2017; D’Orazio & Loeb 2020; Yu & Chen 2021). Several searches have been carried out for signatures of strongly lensed and multiply imaged GW signals in the LIGO and Virgo data. So far, some potential candidates have been proposed that may be interpreted as signals strongly lensed by an intervening galaxy or a galaxy cluster (Broadhurst, Diego & Smoot 2019; Dai et al. 2020; Diego, Broadhurst & Smoot 2021; Liu, Magaña Hernandez & Creighton 2021). However, no compelling evidence has been found

<sup>4</sup>The so-called saddle point images may experience non-trivial waveform changes if the GWs exhibit modes higher than (2, 2) modes or precession (Dai & Venumadhav 2017; Ezquiaga et al. 2021). In addition, the polarization of the GW signal may also change by a small amount during the scattering by a Kerr BH with a small impact parameter (Dolan 2008) and due to the intrinsic anisotropy of the emission pattern, Doppler effect, and de Sitter precession (Torres-Orjuela et al. 2019; Gong, Cao & Chen 2021; Yu & Chen 2021). Note that we do not take into account these effects in this paper.

<sup>5</sup>Multiple images are produced when the stationary phase approximation holds, and the diffraction integral can be solved in terms of its stationary points (e.g. Takahashi & Nakamura 2003).

<sup>6</sup>Note that Pan & Yang (2021) have recently found that migration traps cannot exist in standard AGN discs and BHs do not open gaps as long as their mass is limited to  $10 M_{\odot}$ ; see however Kocsis et al. (2011).

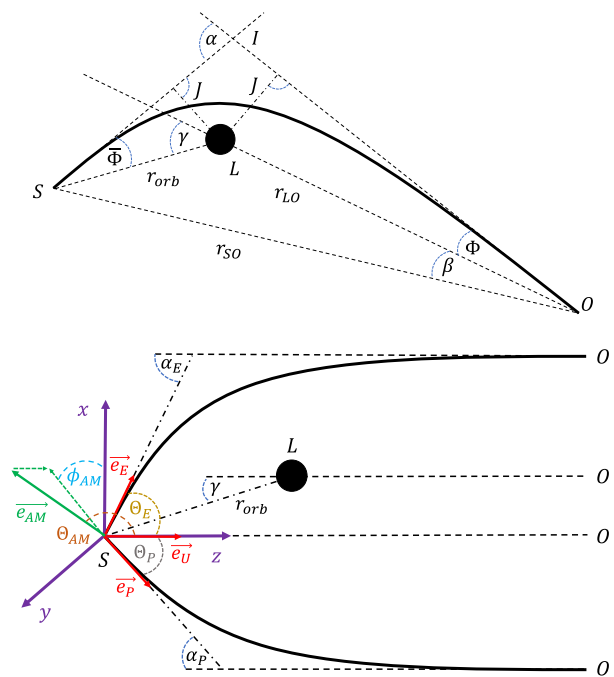
<sup>7</sup><https://www.elisascience.org/>

for lensing yet (Hannuksela et al. 2019; McIsaac et al. 2020; Pang et al. 2020; Abbott et al. 2021c; Kim et al. 2022).

If the SMBH lens happens to lie close to the line of sight behind the GW source, such that the deflection angle  $\alpha$  is close to the Einstein radius, the primary GW signals and astrophysical GW echoes may both be strongly magnified relative to the unlensed GW signal (Einstein 1936). However, unlike for EM waves,<sup>8</sup> the GW echo may be remarkably significant for GW sources in the proximity of the SMBH not just around the Einstein radius but also for larger deflection angles. For large deflection angles, the GW amplitude of the echo relative to the unlensed GW signal scales as  $(r/M_{\text{SMBH}})^{-1}\alpha^{-1}$  (e.g. Kocsis 2013), and the anisotropic emission of the GW source profile significantly broadens the distribution of the relative echo amplitude (see Fig. 2 below). This indicates that GW echoes with large deflection angles may be typically detected for GW sources that are close to the SMBH, although this range has not yet been determined quantitatively in the literature to our knowledge.

In this paper, we examine the possibility of observing astrophysical GW echoes with aLIGO/AdV/KAGRA from the mergers of stellar-mass BBHs in the vicinity of SMBHs as a function of the distance from the SMBH. We assume a generic lensing configuration for calculating the magnification and time delay, i.e. *not* limited to the case of small deflection angles, but neglect possible interference effects between the scattered GW signals. We resort to the geometrical optics limit (i.e. the short-wavelength limit) and assume that the BBH is fixed with respect to the SMBH (i.e. BBHs are in the stationary-lens regime). We account for the anisotropy of the GW emission pattern but neglect the deflection angle dependence of the Doppler effect. We find that these are typically valid approximations to leading order for stellar-mass BBH mergers around SMBHs in LIGO–Virgo–KAGRA searches for SMBHs with masses above  $10^5 M_\odot$  and due to the short time interval the GW signal spends in the aLIGO/AdV/KAGRA band. Since lensing does not alter the waveform in the geometrical optics limit, it is scaled with a constant amplitude and shifted with a time delay (hence the name astrophysical GW echo). This allows us to make population model and detector-independent predictions on the detected signal-to-noise ratio (SNR) distribution and detectability of the GW echo compared to that of the primary GW. We determine the probability of detecting GWs that undergo an arbitrary large-angle deflection due to the SMBH. To this end, we generate random Monte Carlo (MC) samples of BBH mergers in the vicinity of SMBHs to determine the magnification distribution of GW echoes relative to the primary GW signals and the relative amplitude of the GW echo as a function of the radial distance of merging BBHs from the SMBH and the inclination of the plane of merging BBHs. The inclination dependence may be relevant for BBH mergers in AGN discs (Bartos et al. 2017; Corley et al. 2019; Tagawa et al. 2020a) or BH discs (Szölgény & Kocsis 2018; Gruzinov, Levin & Zhu 2020; Máthé, Szölgény & Kocsis 2022) in GNs. We also calculate the fraction of detectable GW echoes in mock binary catalogues by accounting for observational bias and the corresponding time-delay distribution as a function of SMBH mass and binary distance from the SMBH.

In Section 2, we derive the deflection and magnification of primary GWs and astrophysical GW echoes for a generic lensing configuration for an arbitrary binary position relative to the SMBH and the line of sight to the observer. Next, we describe the methodologies with which we carry out our investigations and present the main results of the paper in Section 3. Finally, we summarize the results of the paper and conclude in Section 4. Several details of our analysis are included



**Figure 1.** Top panel: generic lensing geometry by a static spherically symmetric SMBH lens (L). The source is in S, the observer in O, and the image as seen by the observer in I. Illustrated quantities are defined in Section 2.1, and the bending of primary GWs and GW echoes around the SMBH lens are described in Section 2.3. Bottom panel: schematic representation of the emission of unlensed GW (U), primary GW (P), and GW echo (E) with respect to the merging binary (i.e. the source) assuming a distant observer. See definitions in Section 2.1.

in Appendices A–D. First, we discuss the properties of magnification distribution for primary GWs and GW echoes in Appendix A, then estimate the time duration the GW signals of merging BBHs around SMBHs spend in the aLIGO/AdV/KAGRA band in Appendix B. We estimate the impact of additional effects caused by the SMBH on the properties of the GW source in addition to the deflection of GW rays in Appendix C. Finally, we investigate the validity of our assumptions on the geometrical optics limit and the stationary-lens regime for BBH mergers in the vicinity of SMBHs in Appendix D.

We use geometric units ( $G = 1 = c$ ) throughout the paper, where mass  $M$  and distance  $r$  have units of time:  $GM/c^3$  and  $r/c$ .

## 2 DEFLECTION AND MAGNIFICATION OF GRAVITATIONAL WAVES

We introduce the geometric conventions we use to describe the deflection of primary GWs and astrophysical GW echoes around an SMBH in Section 2.1. Section 2.2 introduces the geometric conventions with which GW sources are generated in our models in discs and spherical nuclear star clusters around SMBHs. We determine the deflection angle and time delay in Section 2.3, and the amplitude magnification of a GW echo relative to the primary GW is given in Section 2.4.

### 2.1 Lensing geometry

The top panel of Fig. 1 illustrates the geometric configuration. We assume that the SMBH lens is a Schwarzschild BH for which the bending of rays happens in a plane defined by the lens (L), the observer (O), and the source (S) (e.g. Weinberg 1972). We define

<sup>8</sup>Since intensity scales with the square of the amplitude, it is more suppressed.



all distances and angles of the GW trajectory in the asymptotic flat (Minkowski) space-time.

The line joining the observer and the lens is the optical axis,  $\overline{OL}$ .  $r_{LO}$  and  $r_{SO}$  represent the radial lens–observer and source–observer distances, respectively, while the radial lens–source distance is  $r_{orb}$ . We introduce  $D_{SL}$  and  $D_{SO}$  such that they measure the component of the source–lens and source–observer distances along the optical axis, and the corresponding lens–observer distance is  $r_{LO}$ . The angular position of the source in the observer’s sky with respect to the position of the SMBH is  $\beta = \text{SOL}\angle$ . Furthermore,  $\Phi$  and  $\bar{\Phi}$  are respectively the angles subtended by the tangent of the ray at the observer and the source relative to  $\overline{OL}$  and  $\overline{SL}$ .  $\gamma$  is the angle between the source–lens direction ( $\overline{SL}$ ) and the optical axis. The impact parameters of incoming and outgoing rays are the same  $J$  in the asymptotic approximation. In this lensing configuration, the primary GW signal has  $-\pi \leq \gamma < 0$  where the GW source and its scattered GW are on the same side of the lens, and the GW echo has  $0 < \gamma \leq \pi$ . The high-magnification strong-lensing scenario, where  $\theta$  is at the Einstein radius, corresponds to  $\gamma \sim 0$  (Virbhadra & Ellis 2000), and  $\gamma \sim \pm\pi$  describes retrolensing scenarios (e.g. Holz & Wheeler 2002; Eiroa & Torres 2004).

Given that  $r_{LO} \gg r_{orb}$ , generally  $\beta \ll 1$  and  $\Phi \ll 1$ , while  $\{\alpha, \gamma, \bar{\Phi}\}$  may be of order unity. Furthermore, we will use the following simple approximate geometrical relations:

$$D_{SL} = r_{orb} \cos \gamma, \quad D_{SO} = r_{LO}, \quad r_{SO} = r_{LO}, \quad (1)$$

$$\Phi = \frac{J}{r_{LO}}, \quad \bar{\Phi} = \arcsin\left(\frac{J}{r_{orb}}\right), \quad \beta = \frac{r_{orb}}{r_{LO}} \sin \gamma. \quad (2)$$

The amplitude of GWs emitted by a merging binary is anisotropic as they are described by spin-2-weighted spherical harmonics, whose energy flux changes with the  $\Psi$  polar angle with respect to the binary orbital axis (i.e. the angular momentum vector) as

$$Y(\Psi) \propto \sin^8\left(\frac{\Psi}{2}\right) + \cos^8\left(\frac{\Psi}{2}\right). \quad (3)$$

Thus, the GW energy flux emitted in the direction perpendicular to binary plane is a factor of  $Y(0)/Y(\pi/2) = 8$  times higher than in the plane of the binary and a factor of  $5/2$  higher than average. This shows that a  $90^\circ$  scattering may change the GW energy flux by a factor between  $1/8$  and  $8$  due to the anisotropic GW emission pattern. We take into account the anisotropic emission of GWs of the merging binaries according to equation (3). We describe the emission of the unlensed GWs, primary GWs, and GW echoes next; see the bottom panel of Fig. 1 for an illustration.

## 2.2 Geometric conventions

To account for the anisotropic GW emission pattern, we adopt a Cartesian coordinate system centred at the binary (i.e. the source) as shown in Fig. 1. We define orthonormal basis vectors  $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  such that (i) the binary, the lens, and the observer lie in the  $x$ – $z$  plane, (ii)  $\mathbf{z}$  points from the binary to the observer, (iii) and the SMBH (lens) lies in the half-space with positive  $x$ . We also define the tangent unit vectors along with the rays of the GWs along with the primary signal and the echo at the source respectively with  $\mathbf{e}_P$ ,  $\mathbf{e}_E$ , and let  $\mathbf{e}_U = \mathbf{z}$  label the tangent vector along with rays to the observer in case the lens is neglected. The angular momentum unit vector of the binary is denoted by  $\mathbf{e}_{AM}$  and defined using spherical coordinates as

$$\mathbf{e}_{AM} = (\sin \Theta_{AM} \cos \phi_{AM}, \sin \Theta_{AM} \sin \phi_{AM}, \cos \Theta_{AM}), \quad (4)$$

where the azimuthal angle  $\phi_{AM}$  ranges  $[0, 2\pi]$ , while the polar angle  $\Theta_{AM}$  takes values in the range  $[0, \pi]$ . Since the rays of the GW echo

lie in the  $x$ – $z$  plane,  $\mathbf{e}_E$  is given as

$$\mathbf{e}_E = (\sin \Theta_E, 0, \cos \Theta_E), \quad (5)$$

where for distant observers the polar angle  $\Theta_E$  equals the deflection angle of the GW echo  $\alpha_E$ .<sup>9</sup> Similarly,  $\mathbf{e}_P$  can be given as

$$\mathbf{e}_P = (-\sin \Theta_P, 0, \cos \Theta_P), \quad (6)$$

where the polar angle  $\Theta_P$  equals  $\alpha_P$ . Finally, the  $\Psi$  polar angle with respect to  $\mathbf{e}_{AM}$  can be given separately for the unlensed GW, the primary GW, and the GW echo as

$$\cos \Psi_U = \mathbf{e}_{AM} \cdot \mathbf{e}_U, \quad \cos \Psi_P = \mathbf{e}_{AM} \cdot \mathbf{e}_P, \quad \cos \Psi_E = \mathbf{e}_{AM} \cdot \mathbf{e}_E, \quad (7)$$

where  $\cdot$  denotes the scalar product.

To describe GW sources in AGN/BH discs, we adopt another Cartesian coordinate system centred at the SMBH (lens) with a set of orthonormal basis vectors  $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$  aligned with the AGN/BH disc. In particular, the basis vectors are oriented such that  $\mathbf{X}$  and  $\mathbf{Y}$  lie in the disc plane and  $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$  is the rotation axis of the disc. The unit vector  $\mathbf{e}_O$  points from the SMBH (lens) to the observer<sup>10</sup> where  $\mathbf{e}_O = (\cos \varphi_O \cos i, \sin \varphi_O \cos i, \sin i)$ . Here, the inclination angle of the disc  $i$  is defined as the latitude angle, and  $\varphi_O$  denotes the longitude angle of the observer in the disc.

Here,  $i$  ranges  $[-\pi/2, \pi/2]$ , where  $i = \{-\pi/2, \pi/2\}$  and  $i = 0$  cases describe face-on and edge-on discs, respectively, as viewed by the observer. Furthermore,  $\varphi_O$  ranges between  $[0, 2\pi]$ . In order to describe GW sources (e.g. merging BBHs) at radius  $r_{orb}$  from the SMBH lens, we introduce the source position vector  $\mathbf{r}_{LS}$ , which is defined in polar coordinates in the plane of the disc as  $\mathbf{r}_{LS} = r_{orb}(\cos \varphi_S, \sin \varphi_S, 0)$ . Finally,  $\gamma$  is calculated as

$$\gamma = \arccos(\mathbf{e}_O \cdot \mathbf{r}_{LS}/r_{orb}) = \arccos[\cos(\varphi_S - \varphi_O) \cos i]. \quad (8)$$

Without loss of generality, we restrict the range of  $i$  to  $[0, \pi/2]$  due to symmetry for  $i \rightarrow -i$ .

We decompose the source population into uniform rings around the SMBH. Note that for an inclined uniform ring, the prior distribution of  $\gamma$  for fixed  $r_{orb}$ ,  $i$ , and  $\varphi_O$  follows from  $dN/d\varphi_S = \text{const}$  and  $\varphi_S = \varphi_O + \arccos[(\cos \gamma)/(\cos i)]$  as

$$\frac{dN}{d\gamma} = \frac{dN}{d\varphi_S} \frac{d\varphi_S}{d\gamma} \propto \frac{\sin \gamma}{\sqrt{\cos^2 i - \cos^2 \gamma}} = \frac{\sin \gamma}{\sqrt{\sin^2 \gamma - \sin^2 i}} \quad (9)$$

for  $|\sin \gamma| > \sin i$ , and  $dN/d\gamma = 0$  otherwise if  $|\sin \gamma| \leq \sin i$ . It is uniformly distributed in the range  $[0, \pi]$  for  $i = 0$  (edge-on disc) and it shrinks on to the single value  $\gamma = \pi/2$  for  $i = \pi/2$  (face-on disc). For  $0 < i < \pi/2$ , the  $\gamma$  range  $[\gamma_{\min}, \gamma_{\max}] = [i, \pi - i]$  shrinks gradually from  $[0, \pi]$  to the point  $\{\pi/2\}$  with increasing  $i$ . Note that the  $\gamma$  distribution is symmetric around  $\gamma = \pi/2$  with peaks at the boundaries  $\gamma_{\min} = i$  and  $\gamma_{\max} = \pi - i$ .

We use the same  $(X, Y, Z)$  coordinates to describe an isotropic population of GW sources around a single SMBH (e.g. merging BBHs in either an isotropic GN or an orientation-averaged AGN/BH disc) at radius  $r_{orb}$  but in this case  $\mathbf{r}_{LS} = r_{orb}(\cos \varphi_r \cos l, \sin \varphi_r \cos l, \sin l)$ , where the latitude angle  $l$  plays the role of  $i$  in equations (8) and (9), but while  $i$  was set to a single value for a single disc,  $l$  ranges  $[-\pi/2, \pi/2]$ . This implies a uniform distribution of  $-1 \leq \cos \gamma \leq 1$ . Alternatively, this is equivalent to a distribution of source discs with different  $\sin i$  sampled uniformly in the range  $-1 \leq \sin i \leq 1$ .<sup>11</sup>

<sup>9</sup>For clarity, we note that  $\alpha_E$  refers to the angle corresponding to the GW echo, which is generally *not* the Einstein angle.

<sup>10</sup>Note that since the observer is at infinity,  $\mathbf{e}_O = \mathbf{e}_U = \mathbf{z}$ .

<sup>11</sup>For an isotropic distribution, the polar angle  $p$ , defined by  $p = \pi/2 - i$ , has a uniform distribution in the range  $-1 \leq \cos p \leq 1$ .

### 2.3 Deflection of primary GWs and GW echoes and the time delay of echoes

We use the Ohanian lens equation in the asymptotic flat region of space-time (Ohanian 1987),

$$\gamma = \alpha - \Phi - \bar{\Phi}, \quad (10)$$

to determine the large-angle bending of GW rays around a Schwarzschild BH. This gives accurate results for the bending of rays compared to the exact general relativistic treatment even when  $r_{\text{orb}}$  is relatively small ( $r_{\text{orb}} \sim 10r_s$ ), and it provides the most accurate estimate for light bending among other approximate lens equations (Bozza 2008). Here, for a Schwarzschild BH,  $\alpha$  is given as a function of the closest distance of approach  $r_0$  as<sup>12</sup>

$$\alpha(r_0) + \pi = 2 \int_{r_0}^{\infty} \frac{dr}{r \sqrt{\left(\frac{r}{r_0}\right)^2 \left(1 - \frac{r_s}{r_0}\right) - \left(1 - \frac{r_s}{r}\right)}} = \frac{4F(\vartheta, k)}{\sqrt{u_1 u_3}}, \quad (11)$$

where  $F(\vartheta, k)$  is the elliptic integral of the first kind,<sup>13</sup> and we introduce the following functions of  $r_0$ :

$$u_0 = \frac{r_s}{r_0}, \quad u_1 = \sqrt{1 - u_0}, \quad u_3 = \sqrt{1 - 3u_0}, \quad (13)$$

$$\vartheta = \frac{1}{2} \cos^{-1} \left( \frac{-3u_0}{1+u_1 u_3} \right), \quad k^2 = 2u_0 \frac{(u_1)^{-1} + 3(u_3)^{-1}}{u_1 + u_3}. \quad (14)$$

The impact parameter  $J$  for a given  $r_0$  can be expressed as

$$J = r_0 \left( 1 - \frac{r_s}{r_0} \right)^{-\frac{1}{2}} \quad (15)$$

(Weinberg 1972). As seen, equations (11) and (15) relate  $J$  to  $\alpha$ . Since we are only interested in distant sources,  $\Phi$  can be safely dropped from equation (10).

Next, we express equation (10) as a function of the polar coordinates of the source position ( $r_{\text{orb}}, \gamma$ ) as

$$\gamma = \alpha(r_0) - \arcsin \left( \frac{J(r_0)}{r_{\text{orb}}} \right). \quad (16)$$

This equation leads to a collection of solutions for both  $-\pi \leq \gamma < 0$  and  $0 < \gamma \leq \pi$  as GWs can go around the lens multiple times, where a higher multiplicity corresponds to a lower  $r_0$ .<sup>14</sup> For a given source position  $-\pi \leq \gamma < 0$  with respect to the observer, the largest  $r_0$  corresponds to the primary signal, while the other solutions describe higher odd-order scattered GWs. Similarly, the largest  $r_0$  corresponds to the echo and further  $r_0$  solutions describe higher even-order scattered GWs for a given  $0 < \gamma \leq \pi$ . Here and throughout the paper ‘primary’ signal and ‘echo’ refer to the solutions of equation (16). Finally, substituting the resulting  $r_0$  back

<sup>12</sup>While the integral can be evaluated by integrating between  $r_0 \leq r \leq r_{\text{orb}}$  and  $r_0 \leq r \leq r_{\text{obs}}$  (e.g. equation 6 in Bozza 2008), where  $r_{\text{obs}}$  is the radial distance of the source to the observer, but here we resort to the approximation of  $r_{\text{orb}} \rightarrow \infty$  and  $r_{\text{obs}} \rightarrow \infty$ . This is justified as long as  $r_0 \ll r_{\text{orb}}$ , so that  $\alpha$  is not very close to 0 and  $J(r_0) \leq r_{\text{orb}}$ .

<sup>13</sup>We use the trigonometric form defined by

$$F(\vartheta, k) = \int_0^{\vartheta} \frac{d\chi}{\sqrt{1 - k^2 \sin^2 \chi}}. \quad (12)$$

<sup>14</sup>Note that GWs can go around a lens if  $\pi \leq \alpha < \infty$ , which translates to  $1.5r_s < r_0 \lesssim 1.76r_s$  in case of a Schwarzschild BH due to the one-to-one correspondence between  $\alpha$  and  $r_0$  given by equation (11). Such condition can be given numerically in terms of  $(\gamma, r_{\text{orb}})$  as well by substituting the appropriate  $(\alpha, r_0)$  pairs back into equation (16) and investigating which  $(\gamma, r_{\text{orb}})$  pairs satisfy this equation.

into equation (11) shows that the primary GWs and GW echoes satisfy  $0 < \alpha < \alpha_{\text{Einstein}}$  and  $\alpha_{\text{Einstein}} < \alpha \leq \pi$ , respectively, where  $\alpha_{\text{Einstein}} = 2(2r_{\text{orb}}/r_s)^{-1/2}$  is the deflection angle at the Einstein radius.

In Schwarzschild coordinates  $(t, r, \varphi, \theta)$  in the equatorial plane ( $\theta = \pi/2$ ), GWs follow null geodesics that satisfy

$$\frac{dr}{d\varphi} = \pm r \sqrt{\frac{r^2}{J^2} - \left(1 - \frac{r_s}{r}\right)}, \quad (17)$$

$$\frac{d\varphi}{dt} = \frac{J}{r^2} \left(1 - \frac{r_s}{r}\right). \quad (18)$$

Here,  $J$  is a constant of motion, the impact parameter, which is expressed in terms of the closest distance of approach  $r_0$  in equation (15). Solving equations (17) and (18) for the coordinate time  $t$  between two radial coordinates  $r_A$  and  $r_B$ , we get the following expressions:

$$t(r_B, r_A, J) = \frac{1}{J} \int_{r_A}^{r_B} \frac{r dr}{\left(1 - \frac{r_s}{r}\right) \sqrt{\frac{r^2}{J^2} - \left(1 - \frac{r_s}{r}\right)}}. \quad (19)$$

For GWs or EM waves passing near an SMBH with  $r_0$ , the total change in the coordinate time from the source  $r_{\text{orb}}$  to the observer  $r_{\text{obs}}$  can be given as

$$\Delta t(r_{\text{obs}}, r_{\text{orb}}, J) = t(r_{\text{orb}}, r_0, J) + t(r_{\text{obs}}, r_0, J). \quad (20)$$

We define the time delay  $\tau_d$  between the primary GW and its GW echo as

$$\tau_d = \lim_{r_{\text{obs}} \rightarrow \infty} \left[ \Delta t(r_{\text{obs}}, r_{\text{orb}}, J_E) - \Delta t(r_{\text{obs}}, r_{\text{orb}}, J_P) \right], \quad (21)$$

where  $J_E$  and  $J_P$  are calculated respectively for the primary GW and the GW echo emitted by a merging BBH at a relative position  $(r_{\text{orb}}, \gamma)$  to the SMBH and the line of sight. We consider BBH mergers with a maximum radial distance from the SMBH out to  $r_{\text{orb}} = 10^4 r_s$  throughout the paper for which setting  $r_{\text{obs}} = 10^8 r_s$  yields an excellent approximation for practical purposes so that

$$\tau_d \approx \Delta t(10^8 r_s, r_{\text{orb}}, J_E) - \Delta t(10^8 r_s, r_{\text{orb}}, J_P). \quad (22)$$

Note that  $\tau_d \propto r_s$  (equation 19), and it applies to SMBHs in the local Universe. For SMBHs at cosmological redshift  $z$ , the cosmological redshift effect has to be considered by multiplying  $\tau_d$  by a factor of  $1 + z$  or equivalently by assuming redshifted SMBH mass  $M_{\text{SMBH}, z} = (1 + z)M_{\text{SMBH}}$ . Finally, we note that the additional observation direction-dependent variation in the time delay caused by the Doppler shift is much smaller in comparison, which is neglected here (Appendix C).

### 2.4 Magnification of primary GWs and GW echoes

For an isotropic source with an arbitrary position relative to the SMBH and the line of sight to the observer, the magnification of the energy flux relative to the unlensed signal can be calculated in the short-wavelength limit as

$$\mu_{\text{iso}} = \left( \frac{r_{\text{SO}}}{r_{\text{orb}}} \right)^2 \frac{\sin \Phi}{\sin \gamma} \left| \frac{d\Phi}{d\gamma} \right| \quad (23)$$

(Bozza & Mancini 2004). Using equations (1) and (2), this may be expressed as

$$\mu_{\text{iso}} = \frac{1}{r_{\text{orb}}} \frac{J}{|\sin \gamma|} \left| \frac{dJ}{d\gamma} \right| = \frac{1}{2r_{\text{orb}}} \left| \frac{d(J^2)/dr_0}{d(\cos \gamma)/dr_0} \right|, \quad (24)$$

where  $J(r_0)$  and  $\gamma(r_0)$  are given analytically in equations (11), (15), and (16). Here equation (23) describes the magnification of

the energy flux for point sources. Finally, we note that hereafter equations (23) and (24) refer to the primary GW signal ( $\mu_{\text{iso}, \text{P}}$ ) and GW echo ( $\mu_{\text{iso}, \text{E}}$ ) if  $-\pi \leq \gamma < 0$  and  $0 < \gamma \leq \pi$ , respectively, or equivalently when  $0 < \alpha < \alpha_{\text{Einstein}}$  and  $\alpha_{\text{Einstein}} < \alpha \lesssim \pi$ .

Equation (24) gives the amplitude magnification perceived by an observer relative to the unlensed isotropic source. To take into account the effect of anisotropic GW emission pattern for a merging binary with given  $e_{\text{AM}}$  at a relative position ( $r_{\text{orb}}, \gamma$ ) with respect to the SMBH and the line of sight, we multiply  $\mu_{\text{iso}}$  by the energy flux of lensed GW signal relative to that of the unlensed GW signal  $Y(\Psi)/Y(\Psi_{\text{U}})$ ; see equation (3). Accordingly, the observed magnification of the energy flux of the primary GW and the GW echo relative to the unlensed GW is given as

$$\mu_{\text{P}} = \frac{Y(\Psi_{\text{P}})}{Y(\Psi_{\text{U}})} \mu_{\text{iso}, \text{P}}, \quad \mu_{\text{E}} = \frac{Y(\Psi_{\text{E}})}{Y(\Psi_{\text{U}})} \mu_{\text{iso}, \text{E}}, \quad (25)$$

respectively, where  $\Psi_{\text{P}}$ ,  $\Psi_{\text{U}}$ , and  $\Psi_{\text{E}}$  are determined by equation (7).

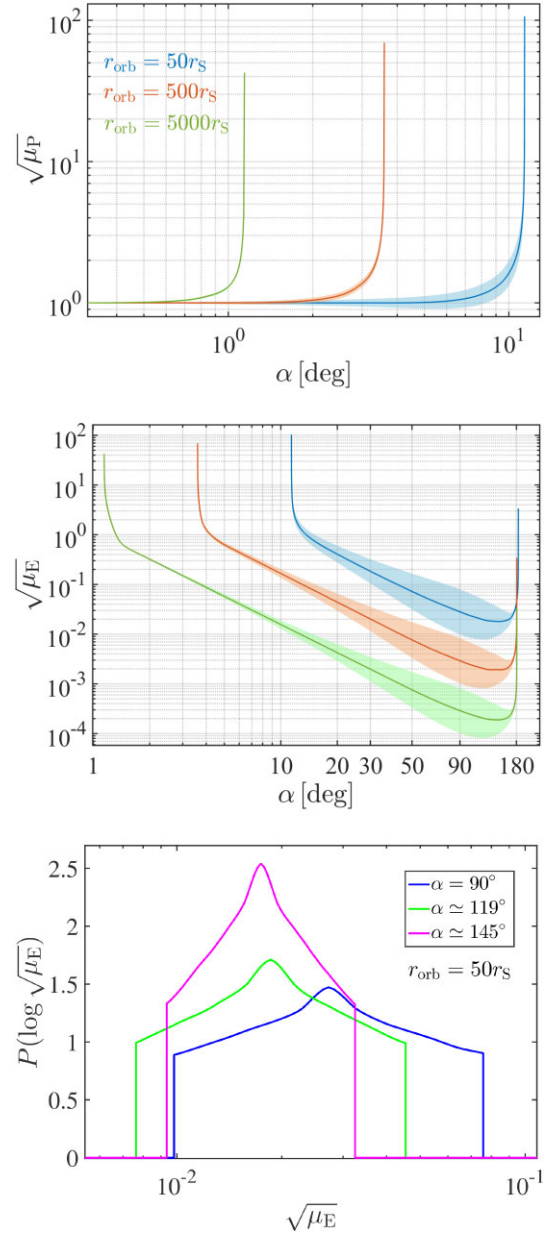
Since the GW amplitude scales with the square root of the energy flux (e.g. Wang et al. 1996; Takahashi & Nakamura 2003),  $\sqrt{\mu}$  represents the *amplification*, i.e. the magnification of the GW amplitude relative to the unlensed GW due to lensing. Accordingly,  $\sqrt{\mu_{\text{P}}}$  and  $\sqrt{\mu_{\text{E}}}$  are the amplification of the primary GW and the GW echo, respectively. While the amplification is always larger than unity for the primary GW, it may be larger or lower than unity for the GW echo. In particular, for large-angle scattering, the echo amplitude is typically fainter than the unperturbed signal, as shown in Fig. 2. Note that additional effects such as gravitational redshift and Doppler beaming are less important and are neglected in this study; see Appendix C for details.

Next, assuming an isotropic distribution of  $e_{\text{AM}}$ ,<sup>15</sup> we investigate the main characteristics of the distribution of amplitude magnification  $P(\sqrt{\mu})$  as a function of  $r_{\text{orb}}$  and  $\alpha$  jointly for primary GWs and GW echoes. Accordingly,  $P(\sqrt{\mu})$  is given as

$$P(\sqrt{\mu}) = \begin{cases} P(\sqrt{\mu_{\text{P}}}) & \text{if } 0 < \alpha < \alpha_{\text{Einstein}}, \\ P(\sqrt{\mu_{\text{E}}}) & \text{if } \alpha_{\text{Einstein}} < \alpha \lesssim \pi \end{cases} \quad (26)$$

at any fixed  $r_{\text{orb}}$ . We find that  $P(\sqrt{\mu})$  peaks at  $\sqrt{\mu_{\text{iso}}}$  at any fixed  $\alpha$  and  $r_{\text{orb}}$ , which is in addition to the median of the distributions. We also find that the width of the distributions is largest at  $\alpha = \pi/2$ , and in this case  $\max(\sqrt{\mu})/\text{med}(\sqrt{\mu}) = \text{med}(\sqrt{\mu})/\min(\sqrt{\mu}) \sim \sqrt{8}$  as expected from equation (3) since a face-on binary is  $Y(0)/Y(\pi/2) = 8$  times brighter in energy flux than an edge-on binary. We list some more properties of  $P(\sqrt{\mu})$  in Appendix A.

We show examples for  $\sqrt{\mu_{\text{P}}}$  and  $\sqrt{\mu_{\text{E}}}$  as a function of  $\alpha$  for various  $r_{\text{orb}}$  separately as incoming GWs are measured independently. We also depict examples for the probability density function (PDF) of the logarithmic GW amplification for GW echoes for some specific  $\alpha$  values at a fixed  $r_{\text{orb}}$ .<sup>16</sup> The top panel of Fig. 2 shows  $\sqrt{\mu_{\text{E}}}$  as a function of  $\alpha$  for various  $r_{\text{orb}}$ , where  $0 < \alpha < \alpha_{\text{Einstein}}$  (Section 2.3).<sup>17</sup> Solid lines show the median of the  $P(\sqrt{\mu_{\text{P}}})$  distribution as a function of  $\alpha$ , which



**Figure 2.** Top and middle panels: the GW amplification of the emitted primary GWs (top,  $\sqrt{\mu_{\text{P}}}$ ) and astrophysical GW echoes (middle,  $\sqrt{\mu_{\text{E}}}$ ) of a merging binary due to an SMBH lens as a function of deflection angle  $\alpha$  for different source distances  $r_{\text{orb}}$  from the SMBH. We consider binaries with isotropic angular momentum distributions that are stationary with respect to the SMBH lens, relevant for aLIGO/AdV/KAGRA sources. Solid lines show the median of the amplification distribution, and shaded areas show the range of possible amplifications resulting from the anisotropic GW emission pattern of merging binaries. The strong amplification peak in the top and middle panels at small  $\alpha$  corresponds to strong lensing at the Einstein radius,  $\alpha_{\text{Einstein}}$ , while the second peak in the middle panel at  $\sim \pi$  corresponds to the glory phenomenon in geometrical optics. Primary GWs and GW echoes generally have  $0 < \alpha < \alpha_{\text{Einstein}}$  and  $\alpha_{\text{Einstein}} < \alpha \lesssim \pi$ , respectively. Bottom panel: the probability density function (PDF) of the logarithmic GW amplification for GW echoes,  $P(\log \sqrt{\mu_{\text{E}}})$ , at fixed  $r_{\text{orb}} = 50r_s$  for three specific  $\alpha$  values as labelled. We show the case with the broadest distribution among sources with fixed deflection angle ( $\alpha = 90^\circ$ ), the case with the minimum  $\sqrt{\mu_{\text{E}}}$  ( $\alpha \sim 119^\circ$ , the minimum of the blue shaded area in the middle panel), and the case where the median of the  $\sqrt{\mu_{\text{E}}}$  distribution has a minimum ( $\alpha \sim 145^\circ$ , the minimum of the solid blue line in the middle panel).

<sup>15</sup>Here and throughout the paper, we assume an isotropic distribution of the angular momentum unit vector  $e_{\text{AM}}$  for merger events in isotropic GNs since there is no preferred direction for  $e_{\text{AM}}$  in this environment. We make the same assumption for merger events in AGN/BH discs, as binary–single interactions are expected to randomize the direction of the binary orbital plane (e.g. Tagawa et al. 2020a, 2021a; Samsing et al. 2022).

<sup>16</sup>Note that similar distributions characterize the logarithmic GW amplification for primary GWs at any fixed  $\alpha$  and  $r_{\text{orb}}$  values.

<sup>17</sup>For each  $r_{\text{orb}}$ , the presented logarithmic  $\alpha$  range was divided into 8000 equal parts, and for each, we sampled  $10^6$  unit vectors  $e_{\text{AM}}$  from an isotropic distribution to determine the lower and upper limits of  $P(\sqrt{\mu_{\text{P}}})$ .



corresponds to the amplitude magnification of an isotropic-equivalent GW source. Furthermore, the shaded areas show the allowed range of amplification  $\sqrt{\mu_P}$  broadened by the anisotropic GW emission pattern of merging binaries. The peak corresponds to strong lensing at the Einstein radius  $\alpha_{\text{Einstein}}$ . Similarly, the middle panel of Fig. 2 shows  $\sqrt{\mu_E}$  for the same  $r_{\text{orb}}$  values as in the top panel. In this case,  $\alpha_{\text{Einstein}} < \alpha \lesssim \pi$  (Section 2.3). The peak at small  $\alpha$  corresponds to strong lensing at  $\alpha_{\text{Einstein}}$ , and the second peak at  $\sim \pi$  corresponds to the glory phenomenon in geometrical optics (Futterman et al. 1988). The bottom panel of Fig. 2 shows the  $\log \sqrt{\mu_E}$  distribution,  $P(\log \sqrt{\mu_E}) = \frac{1}{N} \frac{dN}{d \log \sqrt{\mu_E}}$ , where  $N$  is the number of sources, and we also use a similar logarithmic representation for other parameters throughout the paper when plotting PDFs. The figure shows three cases with large deflection angles for fixed  $r_{\text{orb}} = 50 r_s$ : (i) when the support of  $P(\sqrt{\mu_E})$  is the widest ( $\alpha = 90^\circ$ ); (ii) when  $\sqrt{\mu_E}$  has a minimum ( $\alpha \sim 119^\circ$ ); and (iii) when the median of  $P(\sqrt{\mu_E})$  has a minimum ( $\alpha \sim 145^\circ$ ). Clearly,  $P(\sqrt{\mu_E})$  is approximately symmetric on a logarithmic  $\sqrt{\mu_E}$  scale, and so it peaks near the median.

We define the fraction of the GW echo amplitude relative to the primary GW as

$$\sqrt{\mu_{\text{EP}}} = \frac{\sqrt{\mu_E}}{\sqrt{\mu_P}}. \quad (27)$$

The echo amplitude is typically smaller than the primary amplitude, i.e.  $\sqrt{\mu_{\text{EP}}} < 1$ . However,  $\sqrt{\mu_{\text{EP}}} \gtrsim 1$  is also possible for (i)  $\gamma \sim \pi$  because in this case  $\sqrt{\mu_P} \sim 1$  and  $\sqrt{\mu_E}$  peaks due to the retrolensing glory phenomenon, and for (ii)  $\gamma \sim 0$  because in this case  $\sqrt{\mu_P} \simeq \sqrt{\mu_E}$ ; see Fig. 2 for examples, and for (iii) cases where the energy flux at the source in the direction along the echo rays is stronger than along the primary signal rays.

Importantly, for any  $r_{\text{orb}}$ , the minimum of both  $\sqrt{\mu_P}$  and  $\sqrt{\mu_E}$  are finite non-zero. Accordingly, we determine the minimum value of the relative echo amplitude  $\sqrt{\mu_{\text{EP}}}$  as a function of  $r_{\text{orb}}$  in Appendix A, which sets a strict lower bound on the allowed range of  $r_{\text{orb}}$  for high SNR sources, where an astrophysical GW echo is not detected. In these investigations, we assume that the SNR of both the primary GW and the GW echo are above the SNR detection threshold.

### 3 RESULTS

First, we study the characteristics of the  $\sqrt{\mu_{\text{EP}}}$  distribution,  $P(\sqrt{\mu_{\text{EP}}})$ , for single AGN and BH discs with a fixed inclination and isotropic distributions of BBH mergers around SMBHs in Sections 3.1. Next, in Section 3.2, we determine the  $\sqrt{\mu_{\text{EP}}}$  distribution in a mock SNR-limited survey of the primary GWs accounting for observational bias, i.e. a larger detection volume for BBH mergers with larger primary GW amplitudes. Then, in Section 3.3, we determine the distribution of angular position relative to the SMBH and the line of sight, the distribution of deflection angle for the echo, and the time-delay distribution between the primary GW and the GW echo for BBH mergers in isotropic populations around SMBHs in a mock SNR-limited survey of both GW echoes and primary GWs.

#### 3.1 GW echoes from single AGN/BH discs and isotropic nuclear star clusters around SMBHs

Let us first investigate how the distribution of GW echo amplitude relative to the primary GW signal,  $P_i(\sqrt{\mu_{\text{EP}}})$ , and its mean,  $\langle \sqrt{\mu_{\text{EP}}} \rangle$ , depends on the inclination of the GW source's position plane around the SMBH (i.e. the AGN/BH disc plane) relative to the line of sight and the SMBH-binary distance. Second, we determine  $P(\sqrt{\mu_{\text{EP}}})$ ,

the relative GW amplitude for an isotropic distribution of BBH mergers around a single SMBH or around SMBHs in a mock complete volume-limited survey (e.g. either isotropic GNs or an ensemble of AGN/BH discs in both cases) in which the detection of primary GWs is complete within some maximum detection distance.

To determine  $P_i(\sqrt{\mu_{\text{EP}}})$  for BBH mergers in an AGN/BH disc with fixed inclination  $i$  and fixed SMBH-binary distance  $r_{\text{orb}}$ , we first sample the longitude angles of both the source  $\varphi_S$  and observer  $\varphi_O$  in the disc plane uniformly in the range  $[0, 2\pi]$  and compute the corresponding  $\gamma$  (equation 8). Next, we sample angular momentum unit vectors  $\mathbf{e}_{\text{AM}}$  from an isotropic distribution at each  $\gamma$ , i.e. we sample  $\phi_{\text{AM}}$  uniformly in the range  $[0, 2\pi]$  and  $\cos \Theta_{\text{AM}}$  uniformly in the range  $[-1, 1]$  (Section 2.1), then compute  $\sqrt{\mu_{\text{EP}}}$  for each  $(\gamma, \phi_{\text{AM}}, \Theta_{\text{AM}})$  in the MC sample as discussed in Section 2.4. Finally, using the resulting MC sample, we generate the PDF of the relative echo amplitude, i.e.  $P_i(\sqrt{\mu_{\text{EP}}})$ .<sup>18</sup> Throughout this paper, we verify the convergence of generated distributions as a function of sample size by evaluating the Kolmogorov-Smirnov test with respect to the final distributions.

The top left-hand panel of Fig. 3 displays  $P_i(\log \sqrt{\mu_{\text{EP}}})$  for various  $i$  values at fixed  $r_{\text{orb}} = 50 r_s$ . The obtained distributions are bimodal, where the peak at low  $\sqrt{\mu_{\text{EP}}}$  corresponds to BBH mergers with  $\gamma \sim \pi/2$ , while the second peak is produced by mergers close to either  $\gamma \sim \gamma_{\text{min}} = i$  or  $\gamma_{\text{max}} = \pi - i$  (Section 2.2).  $P_i(\sqrt{\mu_{\text{EP}}})$  is bounded between lower and upper relative amplitudes corresponding to BBH mergers respectively near  $\gamma \sim \gamma_{\text{magn, min}} \sim 113^\circ$  and near either  $\gamma \sim \gamma_{\text{min}}$  or  $\gamma_{\text{max}}$ .<sup>19</sup> Furthermore, the range of  $\sqrt{\mu_{\text{EP}}}$  for fixed  $r_{\text{orb}}$  approaches  $\sqrt{\mu_{\text{EP}}}(\gamma = \pi/2)$  with decreasing  $i$  (Section 2.2) where  $P_i(\sqrt{\mu_{\text{EP}}})$  forms a sharp peak in the  $i \rightarrow \pi/2$  limit. Note that  $P_i(\sqrt{\mu_{\text{EP}}})$  shifts toward lower values as  $r_{\text{orb}}$  increases because  $\sqrt{\mu_E}$  decreases systematically with increasing  $r_{\text{orb}}$ , while  $\sqrt{\mu_P}$  remains  $\sqrt{\mu_P} \sim 1$  for the vast majority of mergers as only a small portion of them are located near  $\alpha_{\text{Einstein}}$ .

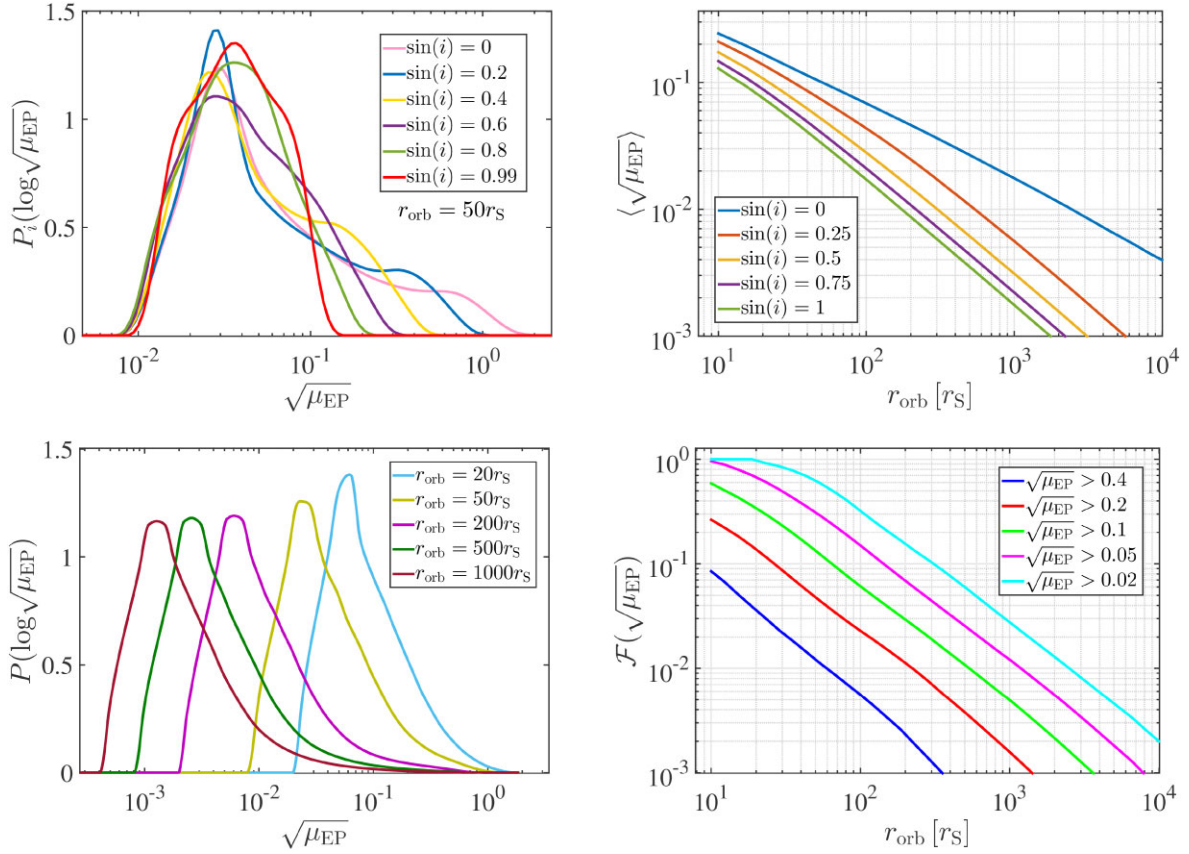
The top right-hand panel of Fig. 3 shows the mean relative echo amplitude  $\langle \sqrt{\mu_{\text{EP}}} \rangle$  as a function of  $r_{\text{orb}}$  for different  $i$  as labelled. Clearly,  $\langle \sqrt{\mu_{\text{EP}}} \rangle$  increases monotonically with  $i$  from the face-on to the edge-on disc configuration, and it decreases monotonically for larger  $r_{\text{orb}}$ . The former trend arises because a systematically higher fraction of  $\gamma$  values occur closer to either  $\gamma \sim 0$  or  $\gamma \sim \pi$  for lower  $i$ , where  $\sqrt{\mu_{\text{EP}}}$  is the highest. The latter trend follows from the  $r_{\text{orb}}$  dependence of  $P_i(\sqrt{\mu_{\text{EP}}})$ . In particular,  $\langle \sqrt{\mu_{\text{EP}}} \rangle$  is  $\sim 2.03 (\log(r_{\text{orb}}/10 r_s))^{2.1} + 1.92$  times higher for edge-on discs ( $\sin i = 0$  blue curve) than for face-on discs ( $\sin i = 1$  green curve), e.g. a factor of  $\sim \{2, 4, 11\}$  times higher for  $r_{\text{orb}} \sim \{10 r_s, 100 r_s, 1000 r_s\}$ .<sup>20</sup> Therefore, astrophysical GW echoes with the highest amplitudes are expected to be typically detected from nearly edge-on AGN/BH discs in the inner regions close to the SMBH.

Similar to the top left-hand panel of Fig. 3, the bottom left-hand panel shows the distribution of the logarithmic relative GW echo amplitude  $P(\log \sqrt{\mu_{\text{EP}}})$  for isotropically distributed BBH mergers around a single SMBH or around SMBHs in a mock complete

<sup>18</sup>We used an MC sample of  $10^6$  mergers to generate  $P_i(\sqrt{\mu_{\text{EP}}})$ .

<sup>19</sup>Here,  $\gamma_{\text{magn, min}} \sim 113^\circ$  is the angle between the SMBH-binary direction and the optical axis where  $\sqrt{\mu_{\text{EP}}}$  attains its minimum (Appendix A). Note that the upper bound of  $\sqrt{\mu_{\text{EP}}}$  decreases systematically with  $i$ , while the lower bound is the smallest at  $i \sim 67^\circ$  ( $\sin i \sim 0.92$ ).

<sup>20</sup>Throughout the paper, we use the non-linear least-squares method to fit the coefficients of functions to be fitted, and resulting values are given at 95 per cent confidence level.



**Figure 3.** Top left-hand panel: the distribution of the logarithmic GW echo amplitude relative to the primary GW signal for BBH mergers in a single AGN or a BH disc with fixed inclination  $i$ ,  $P_i(\log \sqrt{\mu_{\text{EP}}})$ . Examples are shown for various inclination angles as labelled in the legend, and the merger distance from the SMBH is fixed to  $r_{\text{orb}} = 50 r_S$ . Top right-hand panel: the mean relative amplitude of the GW echo compared to the primary GW signal,  $\langle \sqrt{\mu_{\text{EP}}} \rangle$ , as a function of  $r_{\text{orb}}$  for BBH mergers in an AGN/BH disc with a fixed inclination as shown in the legend. Bottom left-hand panel: same as the top left-hand panel, but for an isotropic distribution of BBH mergers around a single SMBH or around SMBHs in a mock complete volume-limited survey (e.g. either isotropic GNs or an ensemble of AGN/BH discs in both cases),  $P(\log \sqrt{\mu_{\text{EP}}})$ , for various  $r_{\text{orb}}$  as labelled in the legend. Bottom right-hand panel: the cumulative fraction of BBH mergers among an isotropically distributed sample of mergers at a distance  $r_{\text{orb}}$  from a single SMBH or in a population of GNs in a mock complete volume-limited survey  $\mathcal{F}(\sqrt{\mu_{\text{EP}}})$  for which  $\sqrt{\mu_{\text{EP}}}$  is higher than the values given in the legend. For example, the bottom curve shows the fraction of mergers in a single GN where the GW echo amplitude relative to the primary GW is larger than  $\sqrt{\mu_{\text{EP}}} = 0.4$ , which may be detectable even for weak marginally detected primary GWs. For very bright primary GWs, the fraction of mergers with a detectable GW echo with a relative amplitude of  $\sqrt{\mu_{\text{EP}}} > 0.02$  reaches 100 per cent for  $r_{\text{orb}} \lesssim 20 r_S$  (see Fig. A1 for the critical radius for a guaranteed GW echo detection for arbitrary  $\sqrt{\mu_{\text{EP}}}$ ).

volume-limited survey.  $P(\sqrt{\mu_{\text{EP}}})$  is determined by marginalizing  $P_i(\sqrt{\mu_{\text{EP}}})$  over  $i$ .<sup>21</sup> The figure shows that  $P(\sqrt{\mu_{\text{EP}}})$  has one peak close to  $\sqrt{\mu_{\text{EP}}} \sim 0.121 (r_{\text{orb}}/10 r_S)^{-1}$ , which corresponds to BBH mergers at  $\gamma \sim \pi/2$ . The lower bound, where  $P(\sqrt{\mu_{\text{EP}}})$  is non-zero, corresponds to equation (A3), while the upper bound corresponds to mergers at either  $\gamma \sim 0$  or  $\gamma \sim \pi$ . Furthermore,  $P(\sqrt{\mu_{\text{EP}}})$  shifts systematically toward lower values with increasing  $r_{\text{orb}}$  roughly as  $r_{\text{orb}}^{-1}$  because  $P_i(\sqrt{\mu_{\text{EP}}})$  also shifts to lower values with increasing  $r_{\text{orb}}$  for all  $i$ .

Finally, the bottom right-hand panel of Fig. 3 shows the cumulative fraction of BBH mergers located at a radial distance of  $r_{\text{orb}}$  to the SMBHs, for which the relative GW echo amplitude is larger than the values shown in the legend,  $\mathcal{F}(\sqrt{\mu_{\text{EP}}})$ . Here we assume an isotropic merger population around a single SMBH or around SMBHs in a mock complete volume-limited survey. We find that  $\mathcal{F}(\sqrt{\mu_{\text{EP}}})$  decreases monotonically with increasing  $r_{\text{orb}}$ , which arises

because  $P(\sqrt{\mu_{\text{EP}}})$  shifts systematically toward lower  $\sqrt{\mu_{\text{EP}}}$  values for larger  $r_{\text{orb}}$ . The figure shows that  $\mathcal{F}(\sqrt{\mu_{\text{EP}}}) = 1$  for  $r_{\text{orb}} \lesssim 20 r_S$  for  $\sqrt{\mu_{\text{EP}}} > 0.02$ , meaning that the GW echo amplitude is at least 0.02 of the primary amplitude for all sources with  $r_{\text{orb}} \lesssim 20 r_S$  (see equation A3). Furthermore, at least 1 per cent of GW echoes have amplitudes larger than  $\{0.4, 0.2, 0.1, 0.05, 0.02\}$  times the primary GW amplitude if  $r_{\text{orb}} \lesssim \{60 r_S, 210 r_S, 500 r_S, 1100 r_S, 2400 r_S\}$ , respectively, and more than 10 per cent of GW echoes have amplitudes larger than  $\{0.2, 0.1, 0.05, 0.02\}$  times the primary amplitude, respectively, for BBH mergers closer to the SMBH than  $\sim \{25 r_S, 60 r_S, 135 r_S, 300 r_S\}$ . This indicates that for the most typical weak primary GW signals in a complete volume-limited survey, there is at least a 1 per cent probability of detecting a GW echo with 0.4 of the primary's amplitude for BBH mergers within  $\sim 60 r_S$  of the SMBH, but for very strong primary signals for which an echo is detectable at 0.02 of the primary's amplitude, there is close to 100 per cent chance to detect a GW echo for sources within  $20 r_S$ , and still a non-zero (1 per cent  $\gtrsim$ ) chance to detect an echo from distances up to  $\sim 2400 r_S$ .

<sup>21</sup>  $P(\sqrt{\mu_{\text{EP}}})$  was generated from an MC sample of  $5 \times 10^6$  elements.



### 3.2 GW echoes from AGN/BH discs and isotropic nuclear star clusters in an SNR-limited survey

In this section, we investigate the impact of observational bias on the previous results presented for single AGN/BH discs and isotropic distributions of BBH mergers around SMBHs in Section 3.1. Here we examine the distributions in a mock SNR-limited GW observational survey of the primary GWs, where BBH mergers with larger primary GW amplitudes are detectable from a larger volume.

As we have seen in Section 3.1, primary GWs typically have higher amplitudes compared to GW echoes, and by definition, arrive earlier at the detector due to their lower light-traveltime around the SMBH (e.g. Kocsis 2013). Thus, assuming that the primary GW signal has been detected above a specific SNR detection limit, we now investigate what fraction of sources may have a detectable GW echo. Here we account for the amplitude dependence of the detection volume of the primary GW signal, which leads to an *observational bias* skewing the echo amplitude distribution of single SMBHs presented above in Section 3.1. The number of detectable mergers scales with the detection volume. Hence, the relative fraction of primary GWs with  $\sqrt{\mu_P}$  in an MC sample scale as  $N_P \propto V_P \propto \mu_P^{3/2}$ . Thus, we assign the detection-volume-based weight  $w_P = \mu_P^{3/2}$  to each element in the MC sample to define the detection-volume-weighted PDF of any given variable  $x$ ,  $P_{\text{det},P}(x)$ .<sup>22</sup> Note that here and throughout the paper we assume a 100 per cent detector duty cycle and a polarization-averaged waveform. Presented results may be affected at some level for a sparse GW detector network when the line of nodes of the antenna beam pattern of the individual instruments is considered.

In practice, we generate the PDF as follows. We first assign a weight  $w_P$  to each element in an MC sample of Section 3.1 that was generated for sources near an individual SMBH or near a population of SMBHs in a mock complete volume-limited survey. Such an MC sample is a table of GW source parameters ( $r_{\text{orb}}$ ,  $\gamma$ ,  $\phi_{\text{AM}}$ ,  $\Theta_{\text{AM}}$ ). Next,  $\sqrt{\mu_P}$ ,  $\sqrt{\mu_E}$ , and  $\sqrt{\mu_{\text{EP}}}$  are calculated using equations (24) and (25) from the GW source parameters. Then we determine the cumulative distribution function (CDF) of the detection-volume-weighted elements,<sup>23</sup> and finally differentiate the CDF with respect to  $x$  to get  $P_{\text{det},P}(x)$  PDF.<sup>24</sup>

The top left-hand panel of Fig. 3 shows the distribution of the GW echo amplitude relative to the primary amplitude for GW detections of BBH mergers in all AGN/BH discs with an inclination  $i$  in a mock SNR-limited survey,  $P_{i,\text{det},P}(\sqrt{\mu_{\text{EP}}})$ , for various fixed inclination angle  $i$  at fixed  $r_{\text{orb}} = 50 r_S$ . As seen, observational bias significantly changes the corresponding  $P_i(\sqrt{\mu_{\text{EP}}})$  distribution for nearly edge-on discs. This is due to the high-magnification strongly lensed mergers in the sample for which  $\sqrt{\mu_{\text{EP}}} \sim 1$ . Observational bias has a gradually less impact for increasing inclination when  $\sin i$

$\gtrsim 0.1$ , or equivalently  $i \gtrsim 6^\circ$ . The reasons are that for BBH mergers with lower  $i$ , a systematically higher fraction of highly magnified mergers ( $\sqrt{\mu_{\text{EP}}} \sim 1$ ) lie closer to the optical axis (Section 3.1), and the distribution is biased towards  $\sqrt{\mu_{\text{EP}}} \sim 1$  for small  $i$ . Note that BBH mergers behind the SMBH close to the optical axis have  $\sqrt{\mu_{\text{EP}}} \lesssim 1$ , which contribute to the peak occurring in  $P_{i,\text{det},P}(\sqrt{\mu_{\text{EP}}})$  at  $\sqrt{\mu_{\text{EP}}} \sim 1$ . Mergers in front of the SMBH close to the optical axis, i.e.  $\alpha \sim \pi$ , contribute typically to the  $\sqrt{\mu_{\text{EP}}} \gtrsim 1$  part of the peak. Finally, the impact of observational bias on  $P_i(\sqrt{\mu_{\text{EP}}})$  decreases with increasing  $r_{\text{orb}}$  because a systematically lower fraction of mergers has  $\sqrt{\mu_{\text{EP}}} \sim 1$  for larger  $r_{\text{orb}}$  (Section 3.1). The important conclusion is that in an SNR-limited GW observational survey, the amplitudes of GW echoes are comparable to primary amplitudes for nearly edge-on discs or, more generally, for sources along the line of sight, but such configurations are expected to be uncommon.

Next, we determine the detection-volume-weighted mean of the GW echo amplitude relative to the primary GW signal  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P}$  as a function of  $r_{\text{orb}}$  for the same inclination angles as in the top right-hand panel of Fig. 3. In comparison,  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P}$  is significantly higher than  $\langle \sqrt{\mu_{\text{EP}}} \rangle$  for nearly edge-on disc configurations with small  $i$ , while it is only marginally higher than  $\langle \sqrt{\mu_{\text{EP}}} \rangle$  if  $\sin i \gtrsim 0.1$ . For nearly edge-on discs,  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P} \gtrsim \{0.91, 0.45\}$  if  $r_{\text{orb}} \lesssim \{150 r_S, 10^4 r_S\}$ , respectively, in a mock SNR-limited survey, while  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P}$  is above  $\sim \{0.046, 0.004\}$  for the same  $r_{\text{orb}}$  limits for single nearly edge-on discs. Furthermore,  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P}$  decreases much less steeply than in a single AGN/BH disc with increasing  $r_{\text{orb}}$  for  $\sin i \sim 0$  but decreases similarly as in a single disc for  $\sin i \gtrsim 0.1$ . Note that similar to the case of single AGN/BH discs, nearly edge-on discs produce on average GW echoes with the highest relative amplitude in the inner regions close to the SMBH in an SNR-limited survey, and the relative echo amplitude decreases systematically with both  $i$  and  $r_{\text{orb}}$ . In comparison to the case of a single AGN/BH disc (Section 3.1), we find that  $\langle \sqrt{\mu_{\text{EP}}} \rangle_{\text{det},P}$  is  $\sim 21.18 (\log(r_{\text{orb}}/10 r_S))^{4.38} + 26.65$  times higher for edge-on discs than for face-on discs, e.g. this enhancement is  $\sim \{27, 48, 470\}$  for  $r_{\text{orb}} \sim \{10 r_S, 100 r_S, 1000 r_S\}$ , respectively.

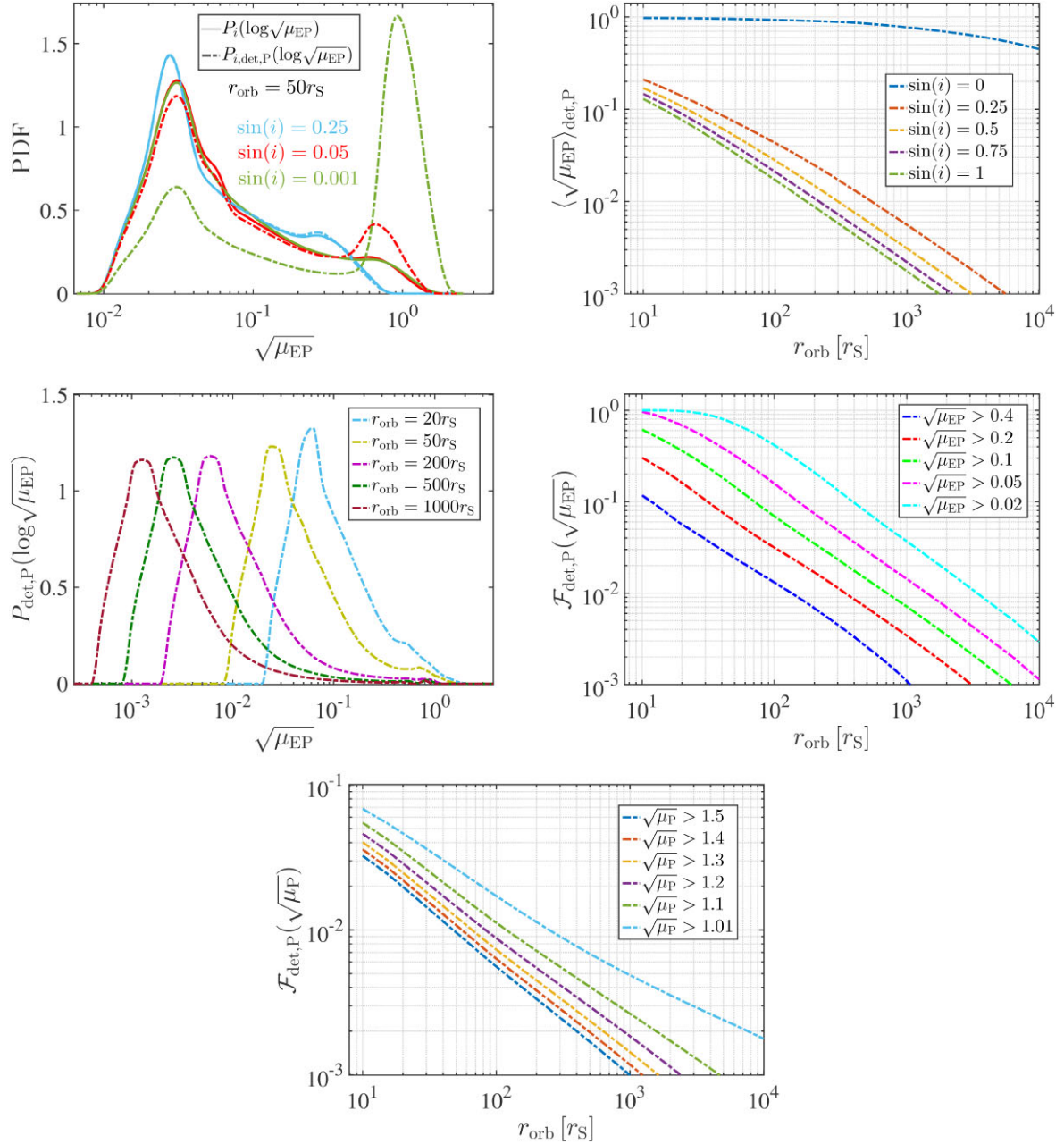
We repeat the analysis of Section 3.1 for an isotropic distribution of BBH mergers around an SMBH by marginalizing  $P_{i,\text{det},P}(\sqrt{\mu_{\text{EP}}})$  over the inclination angle to obtain the distribution function of the relative GW echo amplitude  $P_{\text{det},P}(\sqrt{\mu_{\text{EP}}})$  in an SNR-limited survey. Examples are displayed in the middle left-hand panel of Fig. 4 (cf. Fig. 3). The fraction of mergers with  $\sqrt{\mu_{\text{EP}}} \sim 1$  is increased due to observational bias predominantly for  $r_{\text{orb}} \lesssim 100 r_S$ . The reason is that the Einstein angle is smaller for higher  $r_{\text{orb}}$ , hence the probability for GW echoes with  $\sqrt{\mu_{\text{EP}}} \sim 1$  is suppressed for isotropic merger populations. However, we find that the fraction of mergers with  $\sqrt{\mu_{\text{EP}}} \gtrsim 1$  are typically smaller than a per cent. This is not surprising since  $\sqrt{\mu_E}$  drops at a much faster rate at  $\gamma \sim \pi$  ( $\alpha \sim \pi$ ) than at  $\gamma \sim 0$  ( $\alpha \sim \alpha_{\text{Einstein}}$ ), while (i) the fraction of mergers in the  $\gamma$  distribution are the same at either  $\gamma \sim 0$  or  $\gamma \sim \pi$  as it is symmetric (Section 2.2) and (ii) the corresponding weights are comparable as  $\sqrt{\mu_{\text{EP}}} \sim 1$  at either  $\gamma \sim 0$  or  $\gamma \sim \pi$  (Section 2.4). Therefore, GW echoes with amplitudes higher than that of primary GWs are expected to constitute a small fraction of the detected BBH mergers (e.g.  $\sqrt{\mu_{\text{EP}}} \gtrsim 1$  for  $\sim 0.65$  per cent of mergers at fixed  $r_{\text{orb}} = 20 r_S$ ).

The middle right-hand panel of Fig. 4 shows the cumulative fraction of isotropically distributed BBH mergers around SMBHs in a mock SNR-limited survey  $\mathcal{F}_{\text{det},P}(\sqrt{\mu_{\text{EP}}})$  that have relative GW echo amplitudes higher than the given values shown in the legend as a function of  $r_{\text{orb}}$  (cf. Fig. 3). Clearly, observational bias increases  $\mathcal{F}(\sqrt{\mu_{\text{EP}}})$ , although the degree of increase is significant only for

<sup>22</sup>This method assumes a homogeneous density of BBH mergers in an Euclidean geometry, which gives an adequate first estimate for the observable distribution despite its simplicity of neglecting the inhomogeneous clustering of mergers on small scales and cosmological effects (see also Abadie et al. 2010; O’Shaughnessy, Kalogera & Belczynski 2010; Dominik et al. 2015; Belczynski et al. 2016; Tagawa et al. 2020b, for a similar estimate for various other source populations).

<sup>23</sup>That is, the CDF of the variable  $x$  is given as  $\text{CDF}_{\text{det},P}(x_k) = \sum_k w_k / \sum_g w_g$ , where  $k$  indices those elements in the MC sample for which the corresponding  $x$ s are lower than  $x_k$ , and  $g$  runs over the sample. Here,  $x$  may represent  $\sqrt{\mu_{\text{EP}}}$ , for example, and  $g$  may represent all sources with some fixed  $r_{\text{orb}}$  and  $i$  or cases with fixed  $r_{\text{orb}}$  but with arbitrary  $i$ .

<sup>24</sup>An MC sample of  $5 \times 10^7$  elements was used to generate  $P_{\text{det},P}(x)$ .



**Figure 4.** The first and second rows are similar to Fig. 3, showing the properties of the relative GW echo amplitude distribution over the primary GWs, but for a mock SNR-limited GW observational survey accounting for observational bias by scaling the probabilities with the detection volume of the primary GW signal (Section 3.2). The top left-hand panel shows the distributions of  $\log \sqrt{\mu_{\text{EP}}}$  with and without observational bias using dashed and solid lines, respectively. The inclination angle of the AGN/BH discs is fixed as labelled. The top right-hand panel shows that the mean relative echo amplitude is strongly increased in a mock SNR-limited survey by observational bias for edge-on discs ( $\sin i = 0$ ) due to strong lensing. The middle row left-hand panel shows that the distributions of  $\sqrt{\mu_{\text{EP}}}$  for isotropic BBH merger populations around SMBHs (or an ensemble of AGN/BH discs) acquire a long tail to high echo amplitudes due to strongly lensed sources, but this affects only a tiny fraction of sources at any fixed  $r_{\text{orb}}$ . The right-hand panel in the middle row shows the fraction of sources above a given  $\sqrt{\mu_{\text{EP}}}$  in a mock SNR-limited survey, as labelled, which is very similar to that of Fig. 3 as observational bias affects only a tiny fraction of sources. Bottom panel: the fraction of BBH mergers for which the amplification of primary GWs relative to the unlensed GWs  $\sqrt{\mu_{\text{P}}}$  is higher than selected values as labelled as a function of  $r_{\text{orb}}$ . Results are given for isotropic distributions of BBH mergers around SMBHs, including the effects of observational bias.

relatively high  $\sqrt{\mu_{\text{EP}}}$  limits ( $\sqrt{\mu_{\text{EP}}} \gtrsim 0.1$ ) since observational bias preferentially enhances the fraction of BBH mergers with  $\sqrt{\mu_{\text{EP}}} \sim 1$  in  $P(\sqrt{\mu_{\text{EP}}})$ . For instance, the fraction of mergers with  $\sqrt{\mu_{\text{EP}}} > 0.2$  is  $\mathcal{F}_{\text{det,P}}(\sqrt{\mu_{\text{EP}}}) \sim \{30 \text{ per cent}, 3 \text{ per cent}, 0.33 \text{ per cent}\}$  for  $r_{\text{orb}} \sim \{10 r_S, 100 r_S, 1000 r_S\}$ , respectively, in a mock SNR-limited survey, while these fractions drop to  $\mathcal{F}(\sqrt{\mu_{\text{EP}}}) \sim$

$\{25 \text{ per cent}, 2 \text{ per cent}, 0.15 \text{ per cent}\}$  in a mock complete volume-limited survey for the same radii. Furthermore, a non-negligible fraction of mergers produce GW echoes with amplitudes comparable to those of primary signals. For instance,  $\sim 60 \text{ per cent}$  ( $\sim 10 \text{ per cent}$ ) of mergers have  $\sqrt{\mu_{\text{EP}}} > 0.1$  ( $0.4$ ) at  $r_{\text{orb}} \sim 10 r_S$  in a mock SNR-limited survey, which drops to  $\sim 1 \text{ per cent}$  ( $\sim 0.1 \text{ per cent}$ )

at  $r_{\text{orb}} \sim 1000 r_S$ . Note that similar results stand for a mock complete volume-limited survey. In comparison to the case of a mock complete volume-limited survey (Section 3.1), we find that at least 1 per cent of all GW echoes have amplitudes larger than  $\{0.4, 0.2, 0.1, 0.05, 0.02\}$  times the primary GW amplitude if  $r_{\text{orb}} \lesssim \{135 r_S, 330 r_S, 690 r_S, 1350 r_S, 3300 r_S\}$ , respectively, and more than 10 per cent of echoes have amplitudes larger than  $\{0.2, 0.1, 0.05, 0.02\}$  times the primary GW amplitude if BBHs merge closer to the SMBH than  $\sim \{30 r_S, 65 r_S, 140 r_S, 370 r_S\}$ , respectively. Thus, in an SNR-limited survey, there is at least a 1 per cent probability of detecting GW echoes for BBH mergers within  $\sim 135 r_S$  for the most typical weak primary GW signals, and there is a non-zero (1 per cent  $\lesssim$ ) chance to detect echoes from distances up to  $\sim 3300 r_S$  for very strong primary signals. These limits are respectively  $\sim 60 r_S$  and  $\sim 2400 r_S$  for a complete volume-limited survey (Section 3.1). Finally, we note that the fraction of detectable echoes  $\mathcal{F}_{\text{det},P}(\sqrt{\mu_{\text{EP}}})$  decreases with  $r_{\text{orb}}$  as expected from the  $r_{\text{orb}}$  dependence of the relative echo amplification distribution  $P_{\text{det},P}(\sqrt{\mu_{\text{EP}}})$ .

For further astrophysical applications, we also investigate how the SMBH may amplify the primary GW signal relative to the unlensed GW signal for isotropically distributed BBH mergers around SMBHs in a mock SNR-limited survey irrespective of the GW echo. For this purpose, we assume detection for the primary GW signal and generate the  $P_{\text{det},P}(\sqrt{\mu_P})$  distribution as before. Similarly, we also determine the fraction of mergers  $\mathcal{F}_{\text{det},P}(\sqrt{\mu_P})$  for which the amplification of primary GWs relative to the unlensed GWs is higher than a given limit as a function of  $r_{\text{orb}}$ . Results are displayed in the bottom panel of Fig. 4.  $\mathcal{F}_{\text{det},P}(\sqrt{\mu_P})$  decreases with  $r_{\text{orb}}$  because the fraction of strongly lensed mergers by the SMBH, which are located near  $\alpha_{\text{Einstein}}$ , decrease with  $r_{\text{orb}}$  as well (Section 2.4). We find that only a small fraction of BBH mergers have a prominent  $\sqrt{\mu_P}$ , even in the best-case scenario, when binaries merge closest to the SMBH. For instance,  $\sim \{7 \text{ per cent}, 4 \text{ per cent}, 3 \text{ per cent}\}$  of mergers have  $\sqrt{\mu_P} > \{1.01, 1.3, 1.5\}$ , respectively, for  $r_{\text{orb}} = 10 r_S$  and these merger fractions drop to  $\sim \{1.7 \text{ per cent}, 0.7 \text{ per cent}, 0.5 \text{ per cent}\}$  and  $\sim \{0.5 \text{ per cent}, 0.14 \text{ per cent}, 0.1 \text{ per cent}\}$  for  $r_{\text{orb}} = 100 r_S$  and  $1000 r_S$ , respectively. This is caused mainly by the quick drop of  $\sqrt{\mu_P}$  close to  $\alpha_{\text{Einstein}}$  (Section 2.4) and partly because a small fraction of mergers are located close to the optical axis whose image is near  $\alpha_{\text{Einstein}}$ . Therefore, we conclude that strongly lensed GW signals are not expected to be common in an SNR-limited survey of isotropically distributed BBH mergers around SMBHs. In particular, one in 30 (200) BBH mergers at a distance of  $10 r_S$  ( $100 r_S$ ) of an SMBH may be amplified by 50 per cent.

Note that since highly magnified primary GWs are located near the Einstein radius most such mergers originate in near edge-on disc configurations ( $\sin i \sim 0$ ) (Section 2.2). Furthermore, this subset produces GW echoes with amplitudes comparable to that of primary GWs  $\sqrt{\mu_{\text{EP}}} \sim 1$ . This explains why nearly edge-on AGN/BH disc may produce GW echoes with  $\sqrt{\mu_{\text{EP}}} \sim 1$  in an SNR-limited survey, yet the overall fraction of highly amplified primary GWs is low for an isotropic population as discs with  $\sin i \sim 0$  constitute a small portion of possible configurations in the  $[-1, 1]$  range.

### 3.3 Time delay between the primary GW and the GW echo in an SNR-limited survey

In this section, we determine the time-delay distribution between the primary GW and the GW echo for isotropic distributions of BBH mergers around SMBHs in a mock SNR-limited GW observational survey as a function of SMBH-binary distance. We examine detec-

tions for both GW echoes and primary GWs since the detection volume is different for the two cases, which will skew the time-delay distribution in an observational sample. First, we determine the distribution of time delay assuming that the GW echo was detected irrespective of the primary GW signal. Second, we examine the time-delay distribution in cases where the primary GW signal has a high SNR, such that the echo may be fainter than the primary by a factor of either  $\sqrt{\mu_{\text{EP}}} > 0.02$  or  $> 0.2$ .

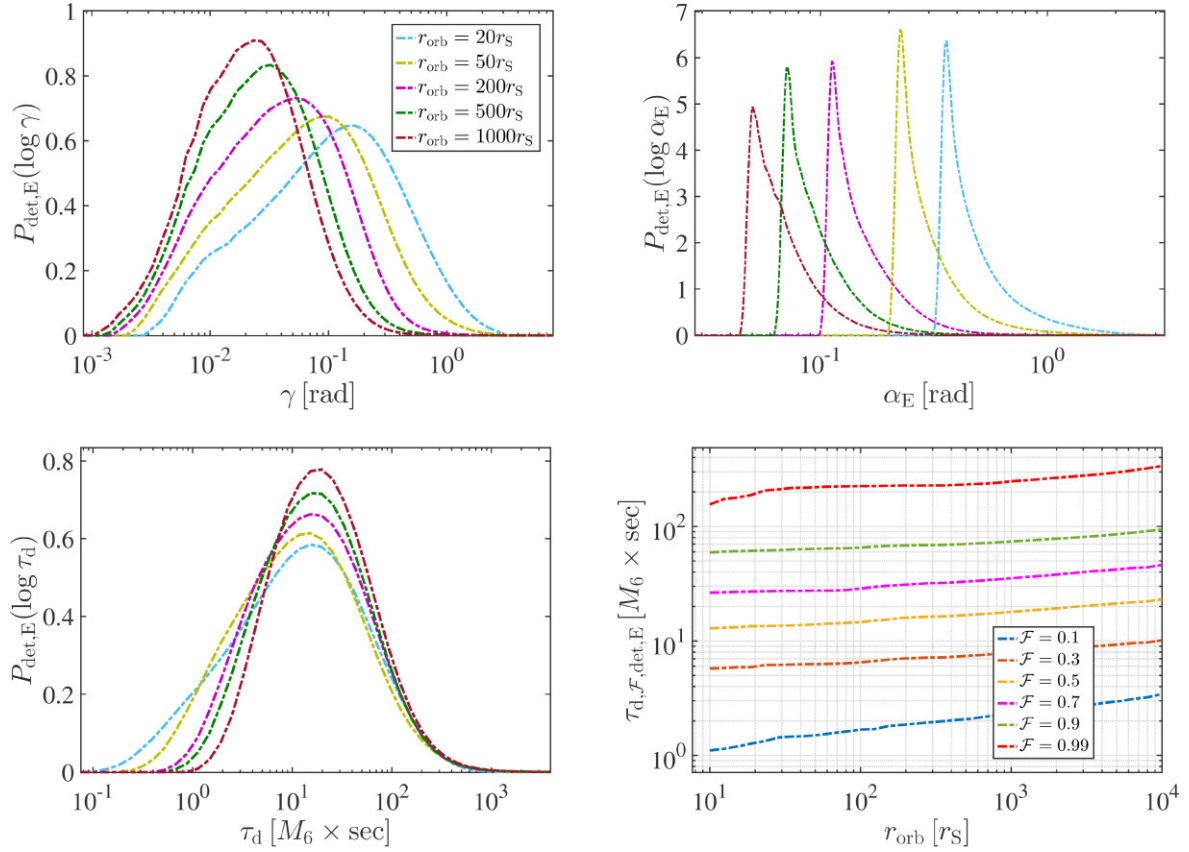
We start by generating a merger sample. For fixed SMBH-binary distance  $r_{\text{orb}}$ , we first generate an isotropic prior MC sample around the SMBH with an isotropic binary orientation ( $\gamma, \phi_{\text{AM}}, \Theta_{\text{AM}}$ ) such that the prior distribution of  $\cos \gamma$  is uniform in the range  $[-1, 1]$  (Section 2.2). We also determine the corresponding deflection angle  $\alpha_E$  for each  $\gamma$  at the considered  $r_{\text{orb}}$  as discussed in Section 2.3. Then, we apply the methodology introduced in Section 3.2 to obtain the detection-volume-weighted distribution  $P_{\text{det},E}(\gamma)$ , the angular position distribution of merging BBHs relative to the SMBH and the line of sight in a mock SNR-limited survey, and the corresponding distribution of deflection angle for the GW echo  $P_{\text{det},E}(\alpha_E)$  by assuming detection for the GW echo. Here we compute the detection-volume weight according to the echo amplitude as  $w_E = \mu_E^{3/2}$ . This assumption also means a detection for the primary signal as typically  $\sqrt{\mu_{\text{EP}}} \lesssim 1$  (Section 3.2). Note that in this type of survey, the vast majority of mergers with observable GW echoes are expected to be in the high-magnification strong-lensing regime, and hence the primary GWs and GW echoes have similar amplitudes. This implies that both the primary GWs and the GW echoes are typically near the SNR detection limit in an observational sample.

Examples for  $P_{\text{det},E}(\log \gamma)$  and  $P_{\text{det},E}(\log \alpha_E)$  are displayed in the top left- and top right-hand panels of Fig. 5, respectively, for various  $r_{\text{orb}}$  as labelled.<sup>25</sup> We find that (i) the vast majority of BBH mergers with detectable GW echoes are strongly lensed (i.e.  $\sqrt{\mu_{\text{EP}}} \sim 1$ ) as most of them have  $\alpha_E \gtrsim \alpha_{\text{Einstein}}$  and accordingly are located behind the SMBH relatively close to the optical axis, (ii) and a marginal fraction of mergers are located in front of the SMBH ( $\lesssim 1$  per cent retrolensing among all events at  $r_{\text{orb}} \gtrsim 10 r_S$ ). Note that  $P_{\text{det},E}(\gamma)$  systematically shifts toward lower  $\gamma$  with increasing  $r_{\text{orb}}$  because mergers with the largest detectable volumes preferentially form near  $\alpha_{\text{Einstein}}$  that systematically shifts toward lower values for larger SMBH-binary distances, while the corresponding  $P_{\text{det},E}(\alpha_E)$  shifts toward lower  $r_{\text{orb}}$  since  $\alpha_{\text{Einstein}} \propto 1/\sqrt{r_{\text{orb}}}$ . For further investigations, we define  $\gamma_{\mathcal{F},\text{det},E}$  using  $P_{\text{det},E}(\gamma)$  as the half-opening angle of the cone in which a cumulative  $\mathcal{F}$  fraction of BBH mergers are located at a distance  $r_{\text{orb}}$  from the SMBH. We find that (i)  $\gamma_{\mathcal{F},\text{det},E}$  decreases with increasing  $r_{\text{orb}}$ , and (ii) the mergers that take place behind the SMBH are located close to the optical axis only for relatively large  $r_{\text{orb}}$  (not shown). For instance, 50 per cent and 90 per cent of detected echoes correspond to mergers located within a half-opening angle of  $\gamma_{\mathcal{F},\text{det},E} \sim \{0.15, 0.05, 0.02\}$  rad and  $\sim \{1.4, 0.6, 0.2\}$  rad, respectively, for  $r_{\text{orb}} = \{10 r_S, 100 r_S, 1000 r_S\}$ . We define similarly  $\alpha_{E,\mathcal{F},\text{det},E}$  and find that the vast majority of mergers have deflection angles close to  $\alpha_{\text{Einstein}}$ . In particular,  $\{30 \text{ per cent}, 50 \text{ per cent}, 70 \text{ per cent}, 90 \text{ per cent}\}$  of mergers are located typically within  $\sim \{1.1, 1.3, 1.6, 2.3\} \times \alpha_{\text{Einstein}}$ .

Next, we determine the distribution of time delay  $P_{\text{det},E}(\tau_d)$  between primary GW signals and GW echoes for BBH mergers with detectable GW echoes in isotropic populations around SMBHs

<sup>25</sup>We use an MC sample of  $2.5 \times 10^7$  elements to generate  $P_{\text{det},E}(\gamma)$ .





**Figure 5.** Left-hand panels: distribution of the relative angular position  $\gamma$  (top left) and the deflection angle of the GW echo  $\alpha_E$  (top right) on a logarithmic scale and the distribution of corresponding time delay  $\log \tau_d$  (bottom left) between primary GWs and GW echoes for isotropic distributions of BBH mergers around SMBHs at a fixed distance  $r_{\text{orb}}$  shown in the legend in a mock SNR-limited GW observational survey of detectable GW echoes. The bottom right-hand panel shows the corresponding contours of fixed cumulative distribution levels  $\mathcal{F}$  of  $\tau_d$  shown in the legend as a function of  $r_{\text{orb}}$ . In all panels, observational bias was calculated using the detection volume of the GW echo. Note that  $\tau_d$  is shown for a redshifted SMBH mass of  $M_{\text{SMBH},z} = (1+z)M_{\text{SMBH}} = 10^6 M_\odot$ . For other values of  $M_{\text{SMBH},z}$  the units must be scaled by the factor  $M_6 = M_{\text{SMBH},z}/10^6 M_\odot$  (Section 3.3).

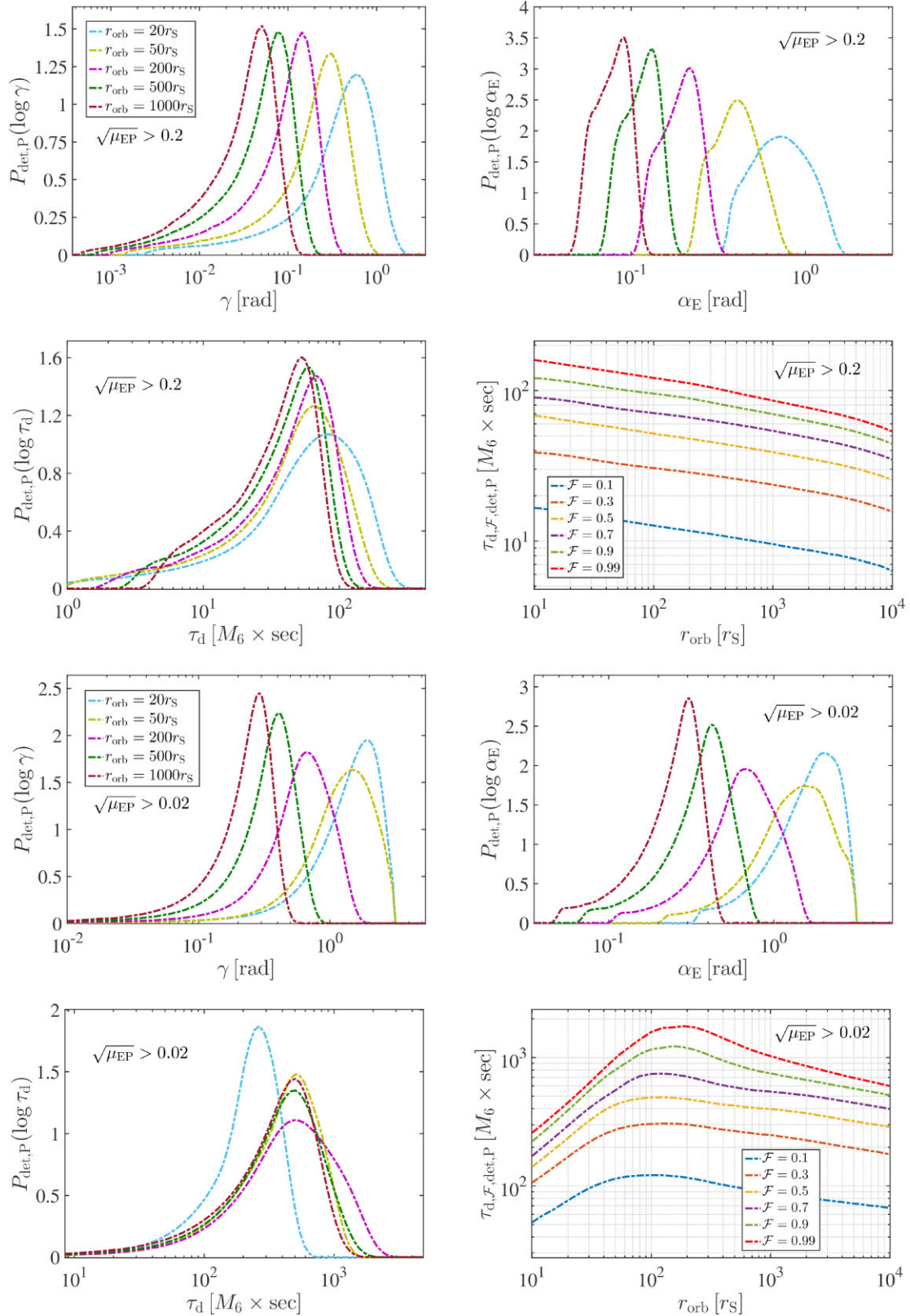
in a mock SNR-limited survey. To do so, we first assign a time delay  $\tau_d$  to a given relative position ( $r_{\text{orb}}, \gamma$ ) and redshifted SMBH mass  $M_{\text{SMBH},z}$  as introduced in Section 2.3, then proceed as prescribed in Section 3.2 to generate  $P_{\text{det},E}(\tau_d)$ . The distributions corresponding to  $P_{\text{det},E}(\log \gamma)$  in the bottom left-hand panel of Fig. 5 are displayed in the bottom right-hand panel for  $M_{\text{SMBH},z} = 10^6 M_\odot$ . Results for other  $M_{\text{SMBH},z}$  can be given by simply scaling the obtained results with the factor  $M_6 = M_{\text{SMBH},z}/10^6 M_\odot$  as  $\tau_d \propto r_s$  (Section 2.3). As seen, the peak of the  $P_{\text{det},E}(\log \tau_d)$  distribution is approximately independent of  $r_{\text{orb}}$ , but the distribution shifts slowly to higher  $\tau_d$  with increasing  $r_{\text{orb}}$ . This feature is due to the fact that  $\tau_d$  increases at a faster rate than how  $P_{\text{det},E}(\gamma)$  shifts toward smaller angles for increasing  $r_{\text{orb}}$ .

Finally, the bottom right-hand panel of Fig. 5 shows the time delay at which a cumulative  $\mathcal{F}$  fraction of isotropically distributed BBH mergers with detectable echoes are located in the vicinity of SMBHs in a mock SNR-limited survey, which we denote by  $\tau_{d,\mathcal{F},\text{det},E}$ . We find that  $\tau_{d,\mathcal{F},\text{det},E}$  increases slowly with  $r_{\text{orb}}$ . For instance, 99 per cent of the systems have a delay time  $\tau_{d,0.99,\text{det},E}$ , which increases slowly between  $20 r_s$  and  $1000 r_s$  from  $\sim 200$  s up to  $\sim 250$  s. Since a non-negligible fraction of GW echoes with amplitudes comparable to the primary amplitudes may be produced by merging BBHs out to a few hundred  $r_s$  (Section 3.2), the maximum time delay between the primary GW and its GW echo may be roughly estimated as  $\max(\tau_d) \sim 250 \text{ s} \times M_6$ . The median time delay between the primary

signal and the echo  $\tau_{d,0.5,\text{det},E}$  represents the typical time delay.<sup>26</sup> As seen, the median also increases slowly with  $r_{\text{orb}}$ , e.g.  $\tau_{d,0.5,\text{det},E}$  is  $\sim 13$  s and  $\sim 18$  s for  $r_{\text{orb}} = 20 r_s$  and  $1000 r_s$ , respectively. Similar to the estimate of  $\max(\tau_d)$ , the typical time delay between the primary signal and its echo is set to be  $\sim 18 \text{ s} \times M_6$ .

Up to this point, the distributions describing GW echoes were obtained by marginalizing over the properties of primary GWs. Let us now consider the distributions for a subset of GW echoes, which arrive after a moderately strong or a very strong primary GW signal that has been detected such that the detection of the GW echo is made possible for e.g.  $\sqrt{\mu_{\text{EP}}} > 0.2$  or  $\sqrt{\mu_{\text{EP}}} > 0.02$ , respectively (cf. Figs 3 and 4). We generate such distribution by applying the detection-volume weights according to the primary GW signal and by also discarding the samples that do not satisfy the above-mentioned conditions on  $\sqrt{\mu_{\text{EP}}}$ . We denote these distributions with the subscript  $\text{det}, \text{P}$  and additionally specify  $\sqrt{\mu_{\text{EP}}} > 0.2$  or  $\sqrt{\mu_{\text{EP}}} > 0.02$  in the figures. The results are shown in Fig. 6. The top four panels show the case of  $\sqrt{\mu_{\text{EP}}} > 0.2$  for primary GWs with moderately high SNRs, and the bottom four panels assume  $\sqrt{\mu_{\text{EP}}} > 0.02$  for primary GWs with very high SNRs. The figure shows a systematic shift of the source angular positions and the deflection angles to higher values

<sup>26</sup>Note that the median of  $P_{\text{det},E}(\tau_d)$  estimates to within  $\sim 10$  per cent the peak location in  $P_{\text{det},E}(\log \tau_d)$ .

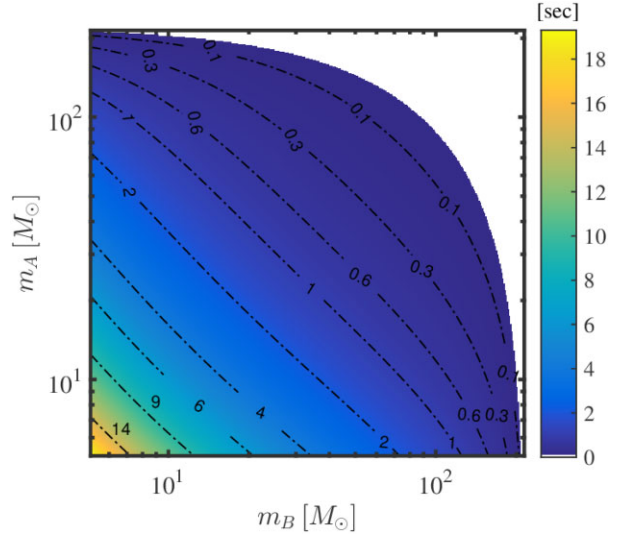


**Figure 6.** Similar to Fig. 5, showing the distribution of the relative angular position  $\gamma$ , the deflection angle of echo  $\alpha_E$ , and the corresponding time delay  $\tau_d$  between the primary GW and its GW echo for isotropic distributions of BBH mergers around SMBHs in a mock SNR-limited survey for fixed  $r_{\text{orb}}$  as shown in the legend. The figure also shows the corresponding lines of constant cumulative distribution levels  $\mathcal{F}$  of  $\tau_d$  labelled in the legend as a function of  $r_{\text{orb}}$ . Observational bias is calculated using the detection volume of the primary GW (cf. Fig. 5, which used the detection volume of the GW echo). The top four and bottom four panels show the distributions assuming that the primary GW has a moderately high or very high SNR, respectively, such that the GW echo may be detected if its amplitude is at least 0.2 or 0.02 times the primary amplitude, respectively; see also Figs 3 and 4.

(cf. Fig. 5). The  $\alpha_E$  distribution is no longer strongly peaked near the Einstein angle, instead it is shifted to higher values by a factor of  $\sim 2$ – $3$  for  $\sqrt{\mu_{EP}} > 0.2$  and by a factor of  $6$ – $7$  for  $\sqrt{\mu_{EP}} > 0.02$  over the radial distance range of  $r_{\text{orb}} = 20 r_S$ – $1000 r_S$ . Here, higher factors correspond to larger  $r_{\text{orb}}$ . In comparison to GW echo detections, we find that {30 per cent, 50 per cent, 70 per cent, 90 per cent} of mergers are located typically within  $\sim \{1.5, 1.9, 2.2, 2.6\} \times \alpha_{\text{Einstein}}$  and within  $\sim \{4.5, 5.8, 7.2, 9.1\} \times \alpha_{\text{Einstein}}$  for echoes with  $\sqrt{\mu_{EP}} > 0.2$  and  $\sqrt{\mu_{EP}} > 0.02$ , respectively, while the corresponding limits for GW echo detections range between  $\sim 1.1$  and  $1.6$ . These differences are not surprising since if the primary has a very high (moderately high) SNR such that the detection of GW echo requires e.g.  $\sqrt{\mu_{EP}} > 0.02$  (e.g.  $> 0.2$ ), the source does not need to lie close to the optical axis, but also could be detected even for scattering configurations with large deflection angles.

Furthermore, Fig. 6 shows that the corresponding time-delay distributions of GW echoes for primary GWs with moderately high and very high SNRs are systematically shifted to higher time-delay values compared to those distributions presented in Fig. 5 for GW echo detections marginalized over all primary GWs. Again, this is due to the fact that the latter case is dominated by primary GWs with low SNRs for which both the primary and the echo are strongly lensed, while for primary GWs with moderately high and very high SNRs, even GW echoes with high deflection angles are detectable. Therefore, among primary GWs with moderately high and very high SNRs, the time delays between primary signals preceding echoes may be much larger, in many cases peaking near  $r_{\text{orb}}/c$  for small  $r_{\text{orb}}$ . This leads to a systematic linear increase of  $\tau_d$  with  $r_{\text{orb}}$  for small  $r_{\text{orb}}$ . For large  $r_{\text{orb}}$ , the delay time decreases weakly with the SMBH-binary distance. Note that these trends are different from the GW echo distribution marginalized over all primary GWs (Fig. 5), which showed a nearly constant, weakly increasing trend with  $r_{\text{orb}}$ . Similar to our considerations on the characteristic time delays for GW echo detections, we set the typical time delay to be the median of  $P_{\text{det}, p}(\tau_d)$  and the maximum time delay to be the 99 per cent quantile of  $P_{\text{det}, p}(\tau_d)$  in the mock observational sample. The typical time delay for primary GWs with moderately high SNRs ( $\sqrt{\mu_{EP}} > 0.2$  in examples) is  $\tau_d \sim (40\text{--}60 \text{ s}) \times M_6$  and it is  $\sim (260\text{--}500 \text{ s}) \times M_6$  for primary GWs with very high SNRs (e.g. when  $\sqrt{\mu_{EP}} > 0.02$ ) over the radial distance range of  $r_{\text{orb}} = 20\text{--}1000 r_S$ . Similarly, the maximum time delay is  $\max(\tau_d) \sim (85\text{--}144 \text{ s}) \times M_6$  for primary GWs with  $\sqrt{\mu_{EP}} > 0.2$  and it is  $\sim (500\text{--}1750 \text{ s}) \times M_6$  for  $\sqrt{\mu_{EP}} > 0.02$  over the same radial distance range.

To put these time-delay values in context, we determine the typical time duration of signals in the sensitive region of the aLIGO/AdV/KAGRA frequency band in Appendix B. Fig. 7 shows the result for quasi-circular BBHs; the signals spend less than 10 s above 20 Hz for BBHs with component masses above  $10 M_\odot$ . The signal duration is longer in the band for highly eccentric BBHs, which may reach a few minutes for the lightest binaries (i.e.  $5\text{--}5 M_\odot$  BBHs) for also a lower frequency cut-off of 20 Hz, but higher mass binaries typically spend much less time in the band. These time-scales are comparable to the typical time delays for SMBHs with masses  $\lesssim 10^6 M_\odot$  for most detectable GW echoes (i.e. those which correspond to high-magnification events). However, they are generally lower than the typical time-delay values for echoes with moderately high or very high SNR primary GWs. Thus, we conclude that the GW echo may often overlap with the primary signal for low-mass BBHs merging in the vicinity of low-mass SMBHs, particularly for GW echo detections marginalized over the primary GWs (Fig. 5 where both the echoes and the primary signal amplitudes are typically near the SNR detection limit). Note that interference effects between



**Figure 7.** The duration of time that quasi-circular BBH mergers' GW signals spend in the advanced GW detectors' sensitive frequency band above 20 Hz as a function of component masses using the 1.5 post-Newtonian (PN) calculation; see Appendix B for details.

the primary GW signal and its GW echo may be important in these scenarios. However, the primary and echo signals are well separated in time and are distinct for higher mass BBHs or massive SMBHs. They are also typically distinct in the case when the primary GW has at least a moderately high SNR.

#### 4 SUMMARY AND CONCLUSIONS

In this paper, we studied the expected properties of astrophysical GW echoes produced by merging BBHs around SMBHs in the sensitive frequency band of aLIGO/AdV/KAGRA. The astrophysical GW echo is generated by scattering of the GWs by the SMBH. It follows a very similar time–frequency evolution and is observed from the same direction as the primary GW signal but arrives with a time delay. The time–frequency evolution of the GW echo may differ from the primary GW for a large deflection angle due to the Doppler shift, which is different if looking at the source from different directions. In addition, the GW polarization of the primary GW and the GW echo is generally different as the emitted signal is intrinsically anisotropic.

We investigated the expected properties of GW echoes for general configurations not limited to highly magnified small deflection angle cases, accounting for the intrinsic anisotropy of the emitted GWs. We carried out an MC-based study to explore the characteristics of the expected magnification distribution and hence the relative amplitude of the GW echo compared to the primary GW signal in mock SNR-limited GW observational surveys and volume-limited surveys focusing on AGN/BH discs and isotropic populations of merging BBHs around SMBHs. We also studied the impact of disc inclination on the resulting distributions. Finally, we explored the time-delay distribution between the primary GW signal and its astrophysical GW echo in mock SNR-limited surveys and the corresponding distribution of the angular position of the merging BBH relative to the SMBH and the line of sight.

Our conclusions are summarized as follows.

- (i) We find that the inclination of the AGN/BH disc  $i$  has a significant impact on the amplitude of the GW echo relative to the primary GW, denoted by  $\sqrt{\mu_{EP}}$ . Echoes with the most prominent



$\sqrt{\mu_{\text{EP}}}$  are expected from the inner regions of edge-on discs with  $i \lesssim 6^\circ$ , and  $\sqrt{\mu_{\text{EP}}}$  systematically decreases with both  $i$  and  $r_{\text{orb}}$ . This holds for BBH mergers in single AGN/BH discs (Section 3.1) and BBH mergers with detectable primary GWs in AGN/BH discs in SNR-limited surveys (Section 3.2).<sup>27</sup> Nearly edge-on discs may produce  $\sim 27$ – $470$  times higher  $\sqrt{\mu_{\text{EP}}}$  on average than face-on discs for  $10r_s \lesssim r_{\text{orb}} \lesssim 1000r_s$  in SNR-limited surveys, where higher fractions correspond to larger SMBH-binary distances (Fig. 4). This enhancement factor reduces to  $\sim 2$ – $11$  for the same  $r_{\text{orb}}$  range in the case of single discs (Fig. 3).

(ii) The GW echo amplitude is comparable to the primary amplitude for strongly lensed signals, which represents a significant fraction of sources in SNR-limited surveys for edge-on discs, but  $\sqrt{\mu_{\text{EP}}}$  is typically much less than unity for discs with moderate inclinations (Fig. 4). We find that the mean  $\sqrt{\mu_{\text{EP}}}$  is expected to be  $\gtrsim \{0.91, 0.45\}$  respectively for  $r_{\text{orb}} \lesssim \{200r_s, 10^4r_s\}$  for nearly edge-on discs in an SNR-limited survey for the primary GW (Fig. 4), while it is  $\gtrsim \{0.046, 0.004\}$  for the same  $r_{\text{orb}}$  range for nearly edge-on discs (Fig. 3). Note that nearly edge-on disc configurations are expected to be uncommon.

(iii) A non-negligible fraction of BBH mergers with detectable primary GWs produce GW echoes with amplitudes comparable to that of primary signals in isotropic populations in close proximity to SMBHs (e.g. either isotropic GNs or an ensemble of AGN/BH discs) both in complete volume-limited surveys<sup>28</sup> (Fig. 3) and in SNR-limited surveys (Fig. 4). For instance, we find that  $\sim 60$  per cent ( $\sim 10$  per cent) of mergers have  $\sqrt{\mu_{\text{EP}}} > 0.1$  ( $0.4$ ) at  $r_{\text{orb}} \sim 10r_s$  in mock SNR-limited surveys, which drops to  $\sim 1$  per cent ( $\sim 0.1$  per cent) at  $r_{\text{orb}} \sim 1000r_s$ .

(iv) There is a non-zero ( $> 1$  per cent) chance to detect GW echoes for BBH mergers within  $\sim 135r_s$  ( $\sim 60r_s$ ) of the SMBH for the most typical weak primary GW signals, and there is a non-zero ( $1$  per cent  $\lesssim$ ) chance to detect echoes from distances up to  $\sim 3300r_s$  ( $\sim 2400r_s$ ) for very strong primary signals in SNR-limited surveys (complete volume-limited surveys) (Sections 3.1 and 3.2).

(v) Strongly lensed GWs by the SMBH are expected to be rare for BBH mergers with detectable primary GWs in isotropic populations around SMBHs in SNR-limited surveys (Section 3.2). We find that the primary GWs have amplitudes larger than  $\{1.01, 1.3, 1.5\}$  times the amplitude of unlensed GWs for  $\sim \{7$  per cent,  $4$  per cent,  $3$  per cent $\}$  of mergers for fixed  $r_{\text{orb}} = 10r_s$  and these fractions drop significantly for higher  $r_{\text{orb}}$  (Fig. 4). However, these percentages increase by a factor of  $8$  for the combined SNR of the primary signal and its echo because the GW echo is also strongly lensed.

(vi) In an SNR-limited survey of BBH mergers in isotropic populations around SMBHs, the vast majority of mergers with observable GW echoes are expected to be in the high-magnification strong-lensing regime (Fig. 5), where primary GWs and GW echoes have similar amplitudes. Although these mergers, which take place behind the SMBH close to the optical axis, are intrinsically rare, their GW echoes can be detected from a much larger volume than high deflection angle weak echoes. However, for primary GWs with high SNR, the GW echo may be much weaker in comparison and still be detected, implying that the GW echo may be in the large

deflection angle regime, especially if the source is relatively close to the SMBH, within a few hundred  $r_s$  (Fig. 6).

(vii) In an SNR-limited survey of GW echoes irrespective of the primary GWs for isotropically distributed BBH mergers around SMBHs, the typical time delay between the primary GW and the GW echo can be roughly estimated as  $\sim 18s \times M_6$ , where  $M_6 = (1+z)M_{\text{SMBH}}/10^6 M_\odot$  (Fig. 5). In practice, primary GWs and GW echoes typically have comparable amplitudes and are near the SNR detection limit. For moderately high or very high SNR primary GWs in an SNR-limited survey of primary GWs, much weaker GW echoes may be detected, which generally arrive with a much longer time delay after the primary, up to  $r_{\text{orb}}/c$  in many cases (Fig. 6). For instance, the typical time delay is  $\sim (40\text{--}60s) \times M_6$  and  $\sim (260\text{--}500s) \times M_6$  respectively, if the amplitude of detected echo is at least  $0.2$  and  $0.02$  times the primary amplitude. These time-scales characterize BBH mergers out to few hundred  $r_s$  since a non-negligible fraction of GW echoes have amplitudes comparable to primary amplitudes in this distance range (Section 3.3).

(viii) The signal duration above  $20\text{Hz}$  for quasi-circular BBHs with component masses  $\gtrsim 10 M_\odot$  and highly eccentric BBHs with relatively high component masses (Appendix B) are comparable to the typical time delays for SMBHs with masses  $\lesssim 10^6 M_\odot$  for most detectable GW echoes (Section 3.3). However, they are generally lower than the typical time-delay values for echoes with moderately high or very high SNR primary GWs. Accordingly, the primary signal and its echo may interfere for very low mass BH and NS mergers around low-mass SMBHs with  $M_{\text{SMBH}} \lesssim 10^6 M_\odot$ , particularly for GW echo detections marginalized over the primary GWs.

(ix) We find that the distribution of GW amplitude magnification  $P(\sqrt{\mu})$  at each deflection angle  $\alpha$  and SMBH-binary distance  $r_{\text{orb}}$  is broadened due to the anisotropic emission of the GW source profile for merging binaries. However, this broadening is found to be negligible in the strong-lensing scenario ( $\alpha \sim \alpha_{\text{Einstein}}$ ) for relatively distant mergers from the SMBH (Appendix A). For instance, the fraction of the upper or the lower bound of  $P(\sqrt{\mu})$  relative to its median is  $\sim \{10$  per cent,  $5$  per cent,  $1$  per cent $\}$  for mergers at  $r_{\text{orb}} \sim \{300r_s, 10^3r_s, 2.2 \times 10^4r_s\}$ .

(x) We identified a strong lower bound for the GW echo amplitude relative to the primary GW  $\sqrt{\mu_{\text{EP}}}$  emitted by merging BBHs in the vicinity of an SMBH as a function of distance from the SMBH  $r_{\text{orb}}$ ,

$$\sqrt{\mu_{\text{EP}}} \gtrsim 0.0411 \left( \frac{r_{\text{orb}}}{10r_s} \right)^{-1} - 0.0021 \left( \frac{r_{\text{orb}}}{10r_s} \right)^{-2} \quad (28)$$

(Appendix A). Its inverse gives the minimum distance of a merging binary from an SMBH upon detection of a primary GW together with the corresponding GW echo with  $\sqrt{\mu_{\text{EP}}}$ ,

$$r_{\text{orb}} \lesssim 4.11r_s \left( \frac{\sqrt{\mu_{\text{EP}}}}{0.1} \right)^{-1} - 0.516r_s. \quad (29)$$

This theoretical bound may be used to rule out the possibility that a BBH merged closer to an SMBH than the estimated minimum radial distance (Fig. A1). In addition, this limit may also offer a conclusive way to rule out that a given BBH merger could have originated near an SMBH.

Chen, Li & Cao (2019) suggested that the high-mass BBH mergers observed by LIGO–Virgo are in reality have lower masses because Doppler and gravitational redshift could increase the ‘apparent’ BH mass and distance by up to a factor of  $\sim 1.9$ – $3.4$  if the mergers lie within  $r_{\text{orb}} = 10r_s$  near an SMBH. However, in this case, the minimum GW echo amplitude is larger than  $0.04$  of the primary GW (equation 28). Thus, if targeted searches rule out the detection of a

<sup>27</sup>Here we assume detection for the primary GW signal above a certain SNR detection threshold and multiply the single GN probability by the detection volume  $\propto \mu_{\text{EP}}^{3/2}$ .

<sup>28</sup>The detection of primary GW signals is complete within some maximum detection distance.

GW echo from the same direction at this level<sup>29</sup> within  $4 M_6$  min after the primary GW signal (i.e. within  $\sim 25$  s–29 d for SMBH masses between  $10^5$  and  $10^{10} M_\odot$ ), this argues against the possibility that the source mass and distance have been systematically overestimated for these sources.

If detected with a consistent value of magnification and time delay, the astrophysical GW echo serves as a smoking gun signature to identify the origin of a GW source from GNs hosting an SMBH. Furthermore, it may be used to test theories describing the astrophysical formation scenarios of BBH mergers in the central regions of AGN/BH discs and GNs with an SMBH in their centre. Moreover, the primary signal and the echo may have a different amplitude and polarization but the same frequency when accounting for the time delay, given that the corresponding detected waves are originally emitted by the source in different directions and at different times. The relative amplitude and time delay between the primary signal and the echo may be used to constrain the geometry of the system and the distance to the SMBH. Note that the recently performed LIGO–Virgo searches can already constrain multiple images by looking for events with overlapping parameters without including the time delay (which was fixed to the scale of galaxy lensing; Abbott et al. 2021c). For very low mass SMBHs and binary mergers, the time-dependent GW diffraction pattern between the primary GW signal and the GW echo may help to constrain the SMBH mass and spin (Kocsis 2013). Detections of astrophysical GW echoes may also be useful to confirm the prediction of general relativity for the large-angle bending of GWs around BHs or to offer new possibilities for testing strong gravity.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for constructive comments that helped improve the quality of this paper. We are grateful to Mária Kolozsvári for help with logistics and administration related to the research. LG is supported by the ÚNKP-18-3 and ÚNKP-21-4 New National Excellence Programmes of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. This work received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Framework Programme ERC-2014-STG under grant agreement no. 638435 (GalNUC).

## DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

## REFERENCES

Aasi J. et al., 2015, *Classical Quantum Gravity*, 32, 115012  
 Abadie J. et al., 2010, *Classical Quantum Gravity*, 27, 173001  
 Abbott B. P. et al., 2016, *Phys. Rev. Lett.*, 116, 061102  
 Abbott B. P. et al., 2017, *Phys. Rev. Lett.*, 119, 161101  
 Abbott B. P. et al., 2019, *Phys. Rev. X*, 9, 031040  
 Abbott R. et al., 2020, *ApJ*, 900, L13  
 Abbott R. et al., 2021a, *Phys. Rev. X*, 11, 021053  
 Abbott R. et al., 2021b, *ApJ*, 913, L7  
 Abbott R. et al., 2021c, *ApJ*, 923, 14

Abedi J., Dykaar H., Afshordi N., 2017, *Phys. Rev. D*, 96, 082004  
 Acernese F. et al., 2015, *Classical Quantum Gravity*, 32, 024001  
 Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)  
 Antonini F., 2014, *ApJ*, 794, 106  
 Antonini F., Perets H. B., 2012, *ApJ*, 757, 27  
 Barack L. et al., 2019, *Classical Quantum Gravity*, 36, 143001  
 Barausse E., Rezzolla L., 2008, *Phys. Rev. D*, 77, 104027  
 Barausse E., Cardoso V., Pani P., 2014, *Phys. Rev. D*, 89, 104059  
 Barceló C., Carballo-Rubio R., Garay L. J., 2017, *J. High Energy Phys.*, 2017, 54  
 Bartos I., Kocsis B., Haiman Z., Márka S., 2017, *ApJ*, 835, 165  
 Baruteau C., Cuadra J., Lin D. N. C., 2011, *ApJ*, 726, 28  
 Belczynski K., Repetto S., Holz D. E., O’Shaughnessy R., Bulik T., Berti E., Fryer C., Dominik M., 2016, *ApJ*, 819, 108  
 Bellovary J. M., Mac Low M.-M., McKernan B., Ford K. E. S., 2016, *ApJ*, 819, L17  
 Blandford R., Teukolsky S. A., 1976, *ApJ*, 205, 580  
 Bonvin C., Caprini C., Sturani R., Tamanini N., 2017, *Phys. Rev. D*, 95, 044029  
 Bozza V., 2008, *Phys. Rev. D*, 78, 103005  
 Bozza V., Mancini L., 2004, *ApJ*, 611, 1045  
 Broadhurst T., Diego J. M., Smoot G., III, 2018, preprint (arXiv:1802.05273)  
 Broadhurst T., Diego J. M., Smoot G. F., III, 2019, preprint (arXiv:1901.03190)  
 Bueno P., Cano P. A., Goelen F., Hertog T., Verenocke B., 2018, *Phys. Rev. D*, 97, 024040  
 Cardoso V., Maselli A., 2020, *A&A*, 644, A147  
 Cardoso V., Pani P., 2017, *Nat. Astron.*, 1, 586  
 Chamberlain K., Moore C. J., Gerosa D., Yunes N., 2019, *Phys. Rev. D*, 99, 024025  
 Chen X., 2021, in Bambi C., Katsanevas S., Kokkotas K. D., eds, *Handbook of Gravitational Wave Astronomy*. Springer, Singapore, p. 1  
 Chen X., Li S., Cao Z., 2019, *MNRAS*, 485, L141  
 Cheung M. H. Y., Gais J., Hannuksela O. A., Li T. G. F., 2021, *MNRAS*, 503, 3326  
 Corley K. R. et al., 2019, *MNRAS*, 488, 4459  
 Cutler C., Flanagan É. E., 1994, *Phys. Rev. D*, 49, 2658  
 Cutler C., Kennefick D., Poisson E., 1994, *Phys. Rev. D*, 50, 3816  
 Dai L., Venumadhav T., 2017, preprint (arXiv:1702.04724)  
 Dai L., Zackay B., Venumadhav T., Roulet J., Zaldarriaga M., 2020, preprint (arXiv:2007.12709)  
 Derdzinski A. M., D’Orazio D., Duffell P., Haiman Z., MacFadyen A., 2019, *MNRAS*, 486, 2754  
 Derdzinski A., D’Orazio D., Duffell P., Haiman Z., MacFadyen A., 2021, *MNRAS*, 501, 3540  
 Diego J. M., Broadhurst T., Smoot G. F., 2021, *Phys. Rev. D*, 104, 103529  
 Dolan S. R., 2008, *Classical Quantum Gravity*, 25, 235002  
 Dolan S. R., Ottewill A. C., 2011, *Phys. Rev. D*, 84, 104002  
 Dominik M. et al., 2015, *ApJ*, 806, 263  
 Dong R., Stojkovic D., 2021, *Phys. Rev. D*, 103, 024058  
 D’Orazio D. J., Di Stefano R., 2018, *MNRAS*, 474, 2975  
 D’Orazio D. J., Loeb A., 2018, *Phys. Rev. D*, 97, 083008  
 D’Orazio D. J., Loeb A., 2020, *Phys. Rev. D*, 101, 083031  
 Einstein A., 1936, *Science*, 84, 506  
 Eiroa E. F., Torres D. F., 2004, *Phys. Rev. D*, 69, 063004  
 Escala A., Larson R. B., Coppi P. S., Mardones D., 2004, *ApJ*, 607, 765  
 Ezquiaga J. M., Holz D. E., Hu W., Lagos M., Wald R. M., 2021, *Phys. Rev. D*, 103, 064047  
 Fragione G., Grishin E., Leigh N. W. C., Perets H. B., Perna R., 2019, *MNRAS*, 488, 47  
 Freitag M., Amaro-Seoane P., Kalogera V., 2006, *ApJ*, 649, 91  
 Futerma J. A. H., Handler F. A., Matzner R. A., 1988, *Scattering from Black Holes*. Cambridge Univ. Press, Cambridge  
 Goldreich P., Tremaine S., 1979, *ApJ*, 233, 857  
 Goldreich P., Lithwick Y., Sari R., 2002, *Nature*, 420, 643  
 Gondán L., Kocsis B., 2019, *ApJ*, 871, 178  
 Gondán L., Kocsis B., 2021, *MNRAS*, 506, 1665  
 Gondán L., Kocsis B., Raffai P., Frei Z., 2018a, *ApJ*, 855, 34

- Gondán L., Kocsis B., Raffai P., Frei Z., 2018b, *ApJ*, 860, 5
- Gong Y., Cao Z., Chen X., 2021, *Phys. Rev. D*, 103, 124044
- Greene J. E., Ho L. C., 2007, *ApJ*, 667, 131
- Gruzinov A., Levin Y., Zhu J., 2020, *ApJ*, 905, 11
- Hamers A. S., Bar-Or B., Petrovich C., Antonini F., 2018, *ApJ*, 865, 2
- Hannuksela O. A., Haris K., Ng K. K. Y., Kumar S., Mehta A. K., Keitel D., Li T. G. F., Ajith P., 2019, *ApJ*, 874, L2
- Hilbert S., White S. D. M., Hartlap J., Schneider P., 2007, *MNRAS*, 382, 121
- Hilbert S., White S. D. M., Hartlap J., Schneider P., 2008, *MNRAS*, 386, 1845
- Hoang B.-M., Naoz S., Kocsis B., Rasio F. A., Dosopoulou F., 2018, *ApJ*, 856, 140
- Holz D. E., Wheeler J. A., 2002, *ApJ*, 578, 330
- Hopman C., Alexander T., 2006, *ApJ*, 645, L133
- Inayoshi K., Tamanini N., Caprini C., Haiman Z., 2017, *Phys. Rev. D*, 96, 063014
- Isaacson R. A., 1968, *Phys. Rev.*, 166, 1263
- Kagra Collaboration et al., 2019, *Nat. Astron.*, 3, 35
- Kains N., Bramich D. M., Sahu K. C., Calamida A., 2016, *MNRAS*, 460, 2025
- Kim H., Kim W.-T., 2007, *ApJ*, 665, 432
- Kim K., Lee J., Hannuksela O. A., Li T. G. F., 2022, preprint ([arXiv:2206.08234](https://arxiv.org/abs/2206.08234))
- Kocsis B., 2013, *ApJ*, 763, 122
- Kocsis B., Levin J., 2012, *Phys. Rev. D*, 85, 123005
- Kocsis B., Frei Z., Haiman Z., Menou K., 2006, *ApJ*, 637, 27
- Kocsis B., Yunes N., Loeb A., 2011, *Phys. Rev. D*, 84, 024032
- Kozai Y., 1962, *AJ*, 67, 591
- Lai D., Rafikov R. R., 2005, *ApJ*, 621, L41
- Lai K.-H., Hannuksela O. A., Herrera-Martín A., Diego J. M., Broadhurst T., Li T. G. F., 2018, *Phys. Rev. D*, 98, 083005
- Levin Y., 2007, *MNRAS*, 374, 515
- Li S.-S., Mao S., Zhao Y., Lu Y., 2018, *MNRAS*, 476, 2220
- Lidov M. L., 1962, *Planet. Space Sci.*, 9, 719
- Lidov M. L., Ziglin S. L., 1976, *Celest. Mech.*, 13, 471
- Liu B., Lai D., 2020, *Phys. Rev. D*, 102, 023020
- Liu X., Magaña Hernandez I., Creighton J., 2021, *ApJ*, 908, 97
- Lower M. E., Thrane E., Lasky P. D., Smith R., 2018, *Phys. Rev. D*, 98, 083028
- McIsaac C., Keitel D., Collett T., Harry I., Mozzon S., Edy O., Bacon D., 2020, *Phys. Rev. D*, 102, 084031
- McKernan B., Ford K. E. S., Kocsis B., Lyra W., Winter L. M., 2014, *MNRAS*, 441, 900
- Maggio E., Pani P., Ferrari V., 2017, *Phys. Rev. D*, 96, 104047
- Mark Z., Zimmerman A., Du S. M., Chen Y., 2017, *Phys. Rev. D*, 96, 084002
- Máthé G., Szölgvény Á., Kocsis B., 2022, preprint ([arXiv:2202.07665](https://arxiv.org/abs/2202.07665))
- Meena A. K., Mishra A., More A., Bose S., Singh Bagla J., 2022, preprint ([arXiv:2205.05409](https://arxiv.org/abs/2205.05409))
- Meiron Y., Kocsis B., Loeb A., 2017, *ApJ*, 834, 200
- Miralda-Escudé J., Gould A., 2000, *ApJ*, 545, 847
- Mishra A., Meena A. K., More A., Bose S., Bagla J. S., 2021, *MNRAS*, 508, 4869
- Morris M., 1993, *ApJ*, 408, 496
- Ohanian H. C., 1987, *Am. J. Phys.*, 55, 428
- O’Leary R. M., Kocsis B., Loeb A., 2009, *MNRAS*, 395, 2127
- O’Shaughnessy R., Kalogera V., Belczynski K., 2010, *ApJ*, 716, 615
- Pan Z., Yang H., 2021, *Phys. Rev. D*, 103, 103018
- Pang P. T. H., Hannuksela O. A., Dietrich T., Pagano G., Harry I. W., 2020, *MNRAS*, 495, 3740
- Peng P., Chen X., 2021, *MNRAS*, 505, 1324
- Petrovich C., Antonini F., 2017, *ApJ*, 846, 146
- Poisson E., Will C. M., 1995, *Phys. Rev. D*, 52, 848
- Randall L., Xianyu Z.-Z., 2018a, *ApJ*, 853, 93
- Randall L., Xianyu Z.-Z., 2018b, *ApJ*, 864, 134
- Randall L., Xianyu Z.-Z., 2019, *ApJ*, 878, 75
- Robertson A., Smith G. P., Massey R., Eke V., Jauzac M., Bianconi M., Ryczanowski D., 2020, *MNRAS*, 495, 3727
- Robson T., Cornish N. J., Tamanini N., Toonen S., 2018, *Phys. Rev. D*, 98, 064012
- Romero-Shaw I. M., Lasky P. D., Thrane E., 2019, *MNRAS*, 490, 5210
- Samsing J. et al., 2022, *Nature*, 603, 237
- Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses*. Springer-Verlag, Berlin
- Secunda A., Bellovary J., Mac Low M.-M., Ford K. E. S., McKernan B., Leigh N. W. C., Lyra W., Sándor Z., 2019, *ApJ*, 878, 85
- Secunda A. et al., 2020, *ApJ*, 903, 133
- Secunda A., Hernandez B., Goodman J., Leigh N. W. C., McKernan B., Ford K. E. S., Adorno J. I., 2021, *ApJ*, 908, L27
- Shapiro I. I., 1964, *Phys. Rev. Lett.*, 13, 789
- Smith G. P. et al., 2018, in González G., Hynes R., eds, *Proc. IAU Symp. Vol. 338, Gravitational Wave Astrophysics: Early Results from Gravitational Wave Searches and Electromagnetic Counterparts*. Cambridge Univ. Press, Cambridge, p. 98
- Stone N. C., Metzger B. D., Haiman Z., 2017, *MNRAS*, 464, 946
- Szölgvény Á., Kocsis B., 2018, *Phys. Rev. Lett.*, 121, 101101
- Tagawa H., Haiman Z., Kocsis B., 2020a, *ApJ*, 898, 25
- Tagawa H., Haiman Z., Bartos I., Kocsis B., 2020b, *ApJ*, 899, 26
- Tagawa H., Kocsis B., Haiman Z., Bartos I., Omukai K., Samsing J., 2021a, *ApJ*, 907, L20
- Tagawa H., Kocsis B., Haiman Z., Bartos I., Omukai K., Samsing J., 2021b, *ApJ*, 908, 194
- Takahashi R., 2017, *ApJ*, 835, 103
- Takahashi R., Nakamura T., 2003, *ApJ*, 595, 1039
- Tamanini N., Klein A., Bonvin C., Barausse E., Caprini C., 2020, *Phys. Rev. D*, 101, 063002
- Tatekawa T., Okamura Y., 2020, preprint ([arXiv:2012.14703](https://arxiv.org/abs/2012.14703))
- Torres-Orjuela A., Chen X., Cao Z., Amaro-Seoane P., Peng P., 2019, *Phys. Rev. D*, 100, 063012
- Torres-Orjuela A., Chen X., Amaro-Seoane P., 2020, *Phys. Rev. D*, 101, 083028
- Toubiana A. et al., 2021, *Phys. Rev. Lett.*, 126, 101105
- Virbhadra K. S., Ellis G. F. R., 2000, *Phys. Rev. D*, 62, 084003
- Wang Y., Stebbins A., Turner E. L., 1996, *Phys. Rev. Lett.*, 77, 2875
- Weinberg S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley-VCH, Weinheim, Germany
- Wen L., 2003, *ApJ*, 598, 419
- Wong K. W. K., Baibhav V., Berti E., 2019, *MNRAS*, 488, 5665
- Yang Y. et al., 2019a, *Phys. Rev. Lett.*, 123, 181101
- Yang Y., Bartos I., Haiman Z., Kocsis B., Márka Z., Stone N. C., Márka S., 2019b, *ApJ*, 876, 122
- Yu H., Chen Y., 2021, *Phys. Rev. Lett.*, 126, 021101
- Yunes N., Kocsis B., Loeb A., Haiman Z., 2011, *Phys. Rev. Lett.*, 107, 171103
- Zenginoğlu A., Galley C. R., 2012, *Phys. Rev. D*, 86, 064030
- Zwick L., Capelo P. R., Bortolas E., Mayer L., Amaro-Seoane P., 2020, *MNRAS*, 495, 2321
- Zwick L., Capelo P. R., Bortolas E., Vázquez-Aceves V., Mayer L., Amaro-Seoane P., 2021, *MNRAS*, 506, 1007

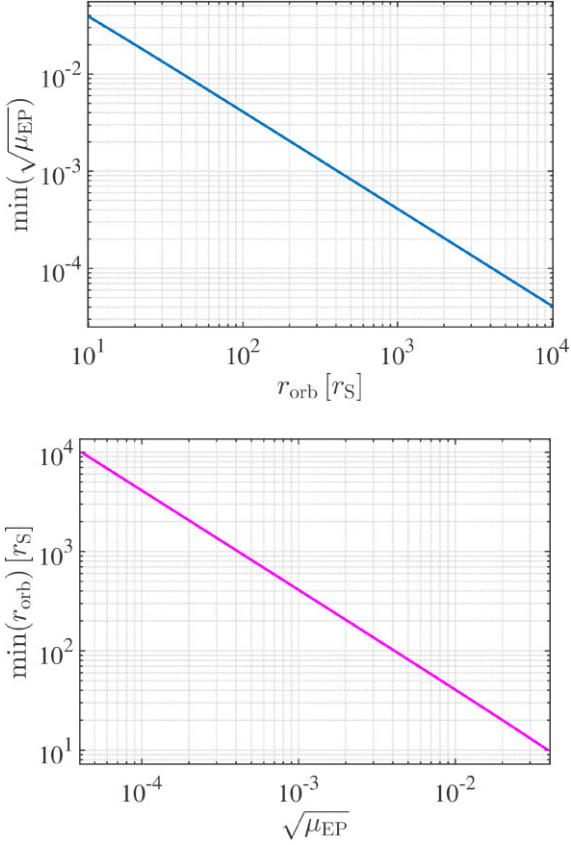
## APPENDIX A: PROPERTIES OF THE AMPLIFICATION DISTRIBUTION

In this section, we carry out investigations regarding  $P(\sqrt{\mu})$  and the minimum value of the relative echo amplitude  $\min \sqrt{\mu_{\text{EP}}}$ . We also note that we work under the assumption of detectable primary GW and GW echo pairs.

First, we note that  $\max(\sqrt{\mu})/\text{med}(\sqrt{\mu}) = \text{med}(\sqrt{\mu})/\min(\sqrt{\mu})$  is due to the isotropy of  $e_{\text{AM}}$ . Furthermore,  $P(\sqrt{\mu})$  forms a sharp peak as a function of deflection angle  $\alpha$  at  $\sqrt{\mu_{\text{IS}}}$  in the  $\alpha \rightarrow 0$  and  $\alpha \rightarrow \pi$  limits, respectively. Finally,  $\max(\sqrt{\mu})/\text{med}(\sqrt{\mu})$  as a function of  $\alpha$  can be best fitted numerically with the quartic polynomial:

$$\frac{\max(\sqrt{\mu})}{\text{med}(\sqrt{\mu})} \simeq 0.165\alpha^4 - 1.035\alpha^3 + 1.295\alpha^2 + 1.036\alpha + 1, \quad (\text{A1})$$





**Figure A1.** Top panel: the minimum echo amplitude relative to the amplitude of the primary GW signal  $\min(\sqrt{\mu_{\text{EP}}})$  assuming that the merging binary is within a radial distance of  $r_{\text{orb}}$  from the SMBH. Bottom panel: the minimum distance of the merging binary from an SMBH for a given GW echo detection amplitude  $\sqrt{\mu_{\text{EP}}}$  relative to the primary GW.

which approximates the fraction in question to within 4 per cent for any  $r_{\text{orb}} \gtrsim 10 r_S$ .  $\max(\sqrt{\mu})/\text{med}(\sqrt{\mu})$  is symmetric about  $\alpha = \pi/2$  and gradually decreases with  $\alpha$  to unity in the  $\alpha \rightarrow 0$  and  $\alpha \rightarrow \pi$  limits. Finally, based on equation (A1), the broadening of  $\sqrt{\mu}$  in the strong-lensing scenario ( $\alpha \sim \alpha_{\text{Einstein}}$ ) due to the anisotropic GW emission pattern of merging binaries at a fixed  $r_{\text{orb}}$  can be given as

$$\frac{\max(\sqrt{\mu})}{\text{med}(\sqrt{\mu})} \Big|_{\alpha_{\text{Einstein}}} \simeq 0.66 \left( \frac{r_{\text{orb}}}{r_S} \right)^{-2} - 2.93 \left( \frac{r_{\text{orb}}}{r_S} \right)^{-3/2} + 2.59 \left( \frac{r_{\text{orb}}}{r_S} \right)^{-1} + 1.47 \left( \frac{r_{\text{orb}}}{r_S} \right)^{-1/2} + 1. \quad (\text{A2})$$

Clearly, the broadening of  $\sqrt{\mu}$  is negligible at  $\sim \alpha_{\text{Einstein}}$  for relatively distant merger events from the SMBH. For instance, the broadening is  $\sim \{10 \text{ per cent}, 5 \text{ per cent}, 1 \text{ per cent}\}$  for mergers at  $\sim \{300 r_S, 10^3 r_S, 2.2 \times 10^4 r_S\}$ .

We also note that  $\sqrt{\mu_E}$  and  $\sqrt{\mu_{\text{EP}}}$  have minima at a certain  $\alpha_{\text{magn, min}}$  (or equivalently  $\gamma_{\text{magn, min}}$ ) for fixed  $r_{\text{orb}}$ .<sup>30</sup> Both minima may be best fitted by

$$\min(\sqrt{\mu_{\text{EP}}}) \approx 0.0411 \left( \frac{r_{\text{orb}}}{10 r_S} \right)^{-1} - 0.0021 \left( \frac{r_{\text{orb}}}{10 r_S} \right)^{-2}, \quad (\text{A3})$$

<sup>30</sup>Note that both  $\alpha_{\text{magn, min}}$  and  $\gamma_{\text{magn, min}}$  do not depend significantly on  $r_{\text{orb}}$ . For instance,  $\{\alpha_{\text{magn, min}}, \gamma_{\text{magn, min}}\} \sim \{117^\circ, 104^\circ\}$  for  $r_{\text{orb}} = 10 r_S$ , while these angles shift to  $\sim \{121^\circ, 121^\circ\}$  for  $r_{\text{orb}} = 10^4 r_S$ .

$$\min(\sqrt{\mu_E}) \approx 0.0398 \left( \frac{r_{\text{orb}}}{10 r_S} \right)^{-1} - 0.0078 \left( \frac{r_{\text{orb}}}{10 r_S} \right)^{-2}, \quad (\text{A4})$$

which estimate  $\min(\sqrt{\mu_E})$  and  $\min(\sqrt{\mu_{\text{EP}}})$  to within  $\sim 3$  per cent and  $\sim 5$  per cent, respectively. The top panel of Fig. A1 shows  $\min(\sqrt{\mu_{\text{EP}}})$  as a function of  $r_{\text{orb}}$ . Upon detection of a GW echo with  $\sqrt{\mu_{\text{EP}}}$ , this defines the minimum radial distance  $\min(r_{\text{orb}})$  of the merging binary from the SMBH lens,<sup>31</sup>

$$\min(r_{\text{orb}}) \approx 4.11 r_S \left( \frac{\sqrt{\mu_{\text{EP}}}}{0.1} \right)^{-1} - 0.516 r_S. \quad (\text{A5})$$

The resulting bound is displayed in the bottom panel of Fig. A1. This theoretical bound may be used to rule out the possibility that any binary has been merged near an SMBH at a smaller radial distance from the SMBH than  $\min(r_{\text{orb}})$  for a measured relative echo amplitude of  $\sqrt{\mu_{\text{EP}}}$ .

Finally, similar to the  $\gamma$  distribution (Section 2.2), the support of the corresponding  $\alpha$ ,  $\sqrt{\mu_E}$ , and  $\sqrt{\mu_{\text{EP}}}$  distributions also shrinks as  $i$  increases from 0 to  $\pi/2$ . More precisely for the  $\sqrt{\mu_E}$  and  $\sqrt{\mu_{\text{EP}}}$  distributions, the lower bound increases with  $i$  only if  $\gamma_{\text{max}} < \gamma_{\text{magn, min}}$  or equivalently if  $i \gtrsim \pi - \gamma_{\text{magn, min}}$  and decreases with  $i$  otherwise as both  $\sqrt{\mu_E}$  and  $\sqrt{\mu_{\text{EP}}}$  attain minima at  $\gamma_{\text{magn, min}}$ .

## APPENDIX B: SIGNAL TIME DURATION IN THE FREQUENCY BAND OF ADVANCED GW DETECTORS

We estimate the time  $\tau_{\text{band}}$  that the GW signals may spend in the advanced GW detectors' sensitive frequency band for BBH mergers originating in AGN/BH discs and isotropic distributions around SMBHs. We consider two scenarios regarding the eccentricity with which inspiraling BBHs may enter the band: (i) quasi-circular BBHs that are expected to form in AGN discs and KL-induced BBH mergers in GNs (e.g. Antonini & Perets 2012; Hammers et al. 2018; Randall & Xianyu 2018b; Fragione et al. 2019; Tagawa et al. 2021a; Samsing et al. 2022); and (ii) eccentric BBHs that are expected to originate mainly from the single-single GW capture channel in GNs (O'Leary et al. 2009; Gondán et al. 2018b; Gondán & Kocsis 2021) and also in part from AGN discs and KL-induced BBH mergers in GNs.

The time duration that the GW signals of isolated quasi-circular BBHs at cosmological redshift  $z$  spend in the aLIGO/AdV/KAGRA band may be estimated using the leading-order estimate as

$$\tau_{\text{band}} = 5(8\pi f_L)^{-8/3} \mathcal{M}_z^{-5/3} \left[ 1 - \left( \frac{f_L}{f_{\text{ISCO}}} \right)^{8/3} \right] \approx 18.9 \text{ s} \left( \frac{M_{\text{tot}, z}}{10 M_\odot} \right)^{-5/3} \left( \frac{\eta}{0.25} \right)^{-1} \left( \frac{f_L}{20 \text{ Hz}} \right)^{-8/3}, \quad (\text{B1})$$

where  $\mathcal{M}_z = M_{\text{tot}, z} \eta^{3/5}$  is the redshifted chirp mass. Here,  $M_{\text{tot}}$  and  $M_{\text{tot}, z} = (1 + z)M_{\text{tot}}$  are the intrinsic and redshifted total binary mass, respectively, and  $\eta = m_1 m_2 / (m_1 + m_2)^2$  is the symmetric mass ratio. We set the lower bound of the detection band to be  $f_L = 20 \text{ Hz}$  based on recent analyses in LIGO–Virgo–KAGRA searches (Abbott et al. 2021a), while the upper bound is the frequency at the ISCO  $f_{\text{ISCO}} = (6^{3/2} \pi M_{\text{tot}, z})^{-1}$ . Note that higher post-Newtonian (PN) corrections beyond leading order affect  $\tau_{\text{band}}$  by more than

<sup>31</sup>We expressed  $r_{\text{orb}}$  as a function of  $\sqrt{\mu_{\text{EP}}}$  from equation (A3), which may be best fitted by a reciprocal power function. We also used the non-linear least-squares method to fit the coefficients, and resulting values are given at 95 per cent confidence level.

$\sim 10$  per cent only in cases where  $\tau_{\text{band}} \lesssim 3.4$  s. Fig. 7 shows  $\tau_{\text{band}}$  as a function of component masses using the 1.5 PN calculation.<sup>32</sup> The white region in the figure corresponds to binaries with  $f_{\text{ISCO}} < 20$  Hz. Clearly, quasi-circular BBHs typically spend a few seconds in the aLIGO/AdV/KAGRA band above  $f_L = 20$  Hz.

In the case of highly eccentric BBHs, the GW signal may spend up to hours in the band in the repeated burst phase.<sup>33</sup> However, the vast majority of the cumulative SNR of the inspiral signal arrives typically within the final few minutes (e.g.  $\sim 50$  per cent within the final minute) of the merger (Kocsis & Levin 2012; Gondán et al. 2018b), even for  $f_L = 10$  Hz. Generally,  $\tau_{\text{band}}$  may reach a few minutes for the lightest BBHs for  $f_L = 20$  Hz based on Gondán et al. (2018b).

Note that additional effects such as gravitational and Doppler shift can be neglected when estimating  $\tau_{\text{band}}$ ; see Appendix C for details.

### APPENDIX C: ADDITIONAL EFFECTS

In addition to the deflection of GW rays considered in detail in this paper, the SMBH may also influence the properties of the GW source in other ways. These include the gravitational redshift and Doppler shift (e.g. Chen et al. 2019; Chen 2021) and time-delay effects (the Shapiro and the Roemer delay, the geometric delay).<sup>34</sup> Following Chen (2021), we estimate the impact of these effects on the physical parameters of GWs investigated in this paper, such as the observed GW amplitude,  $\tau_{\text{band}}$ , and  $\tau_d$ . We denote the leading order estimate of the observed GW amplitude of a merging binary without these additional effects but accounting for cosmological redshift  $z_{\text{cos}}$  with  $h_{\text{cos}} \sim (\mathcal{M}_{\text{cos}}/D_{\text{L,cos}})(\pi f_{\text{cos}}\mathcal{M}_{\text{cos}})^{2/3}$ . For clarity, we denote the cosmological redshift by  $z_{\text{cos}}$  in this section. Furthermore,  $f_{\text{cos}} = f/(1 + z_{\text{cos}})$  is the cosmological redshifted (observed) GW frequency,  $D_{\text{L,cos}} = (1 + z_{\text{cos}})D_C$  is the luminosity distance to the binary, and finally  $f$  and  $D_C$  are the intrinsic (emitted) GW frequency and the comoving distance, respectively.

Close to an SMBH, gravitational redshift also distorts a GW signal causing additional time dilation by the same factor for the primary GW and the GW echo. The redshifted (observed) frequency including both the cosmological and gravitational redshift is  $f_z = f_{\text{cos}}/(1 + z_{\text{grav}})$ . Similarly, the redshifted chirp mass is  $\mathcal{M}_z = (1 + z_{\text{grav}})\mathcal{M}_{\text{cos}}$  (e.g. Chen et al. 2019), where the gravitational redshift is

$$1 + z_{\text{grav}} = \left(1 - \frac{r_S}{r_{\text{orb}}}\right)^{-1/2}. \quad (\text{C1})$$

This implies that the observed GW amplitude and the signal time duration in the band are further distorted as  $h = (1 + z_{\text{grav}})h_{\text{cos}}$  and

<sup>32</sup>We estimate  $\tau_{\text{band}}$  more accurately for quasi-circular BBHs by integrating the frequency evolution equation including PN terms up to 1.5 PN order neglecting spin effects (e.g. Cutler, Kennefick & Poisson 1994). Generally,  $f_{\text{ISCO}}$  sets an upper limit for  $M_{\text{tot}}$  by the condition  $f_{\text{ISCO}} > f_L$  (i.e.  $\tau_{\text{band}} > 0$ ), i.e.  $M_{\text{tot}} \lesssim 220 M_{\odot}$  for  $f_L = 20$  Hz.

<sup>33</sup>To estimate  $\tau_{\text{band}}$ , we first determine the dimensionless pericentre distance  $\rho_{p,20\text{Hz}}$  with which the GW frequency ( $f_{\text{GW}}$ ; Wen 2003) enters the 20 Hz frequency band, then determine  $\tau_{\text{band}}$  by calculating the merger time between  $\rho_{p,20\text{Hz}}$  and the dimensionless pericentre distance characterizing the last stable orbit by applying Peters' formula improved by Zwick et al. (2020), which accounts for the eccentricity evolution and 1 PN order effects. We neglect higher PN order terms because the merger time-scale is weakly affected by these terms (Zwick et al. 2021).

<sup>34</sup>Other effects such as peculiar acceleration of binaries around SMBHs and the influence of gas affect more seriously the low-mass binaries in the LISA band (e.g. Chen 2021, and references therein).

$(1 + z_{\text{grav}})\tau_{\text{band}}$ , respectively; see Section 2.3 for details. As seen, gravitational redshift has a small impact on  $h$  that equals  $h_{\text{cos}}$  to within  $\lesssim 5.5$  per cent for  $r_{\text{orb}} \geq 10 r_S$ , which is marginal compared to the factor of  $\sqrt{3}$  variation caused by the intrinsic anisotropy of the GW emission pattern (Section 2.1). Consequently, we neglected this effect in our investigation. Similarly, the effect of gravitational redshift on  $\tau_{\text{band}}$  is also negligible.

The Doppler shift may cause an additional observation direction-dependent time delay and amplification (Doppler beaming) if the source is moving with respect to the lens and the observer. The effect of Doppler shift can be characterized by the relativistic Doppler factor  $1 + z_{\text{dop}} = \gamma(1 + \mathbf{v} \cdot \mathbf{n})$ , where  $\gamma = 1/\sqrt{1 - v^2}$  is the Lorentz factor, and  $\mathbf{v}$  is the velocity of the source. Finally,  $\mathbf{n}$  is a unit vector tangent to GW rays at the source, i.e.  $\mathbf{e}_p$  and  $\mathbf{e}_E$  in Fig. 1 for the primary signal and its echo, respectively. Thus, the observed GW amplitude and the signal time duration in the band are further distorted as  $(1 + z_{\text{dop}})h_{\text{cos}}$ <sup>35</sup> and  $(1 + z_{\text{dop}})\tau_{\text{band}}$ , respectively.<sup>36</sup> To estimate the maximum effect of Doppler shift on both quantities, we note that to date BBH mergers are expected to merge closest to an SMBH in migration traps (Bellovary et al. 2016; Secunda et al. 2019, 2020, 2021), which may exist in slim discs even near the ISCO (Peng & Chen 2021). Accordingly, the motion of BBHs merging closest to an SMBH happen approximately on a circular orbit ( $v = \sqrt{r_S/2r_{\text{orb}}}$ ) implying that to leading order  $1 + z_{\text{dop}}$  can be approximated as

$$1 + z_{\text{dop}} \approx \left(1 + \cos \zeta \sqrt{\frac{r_S}{2r_{\text{orb}}}}\right) \left(1 + \frac{r_S}{4r_{\text{orb}}}\right), \quad (\text{C2})$$

where  $\cos \zeta$  is the cosine of the angle between  $\mathbf{v}$  and  $\mathbf{n}$ . We find that Doppler shift causes a maximum of  $\sim 25$  per cent variation in  $h_{\text{cos}}$  at  $r_{\text{orb}} = 10 r_S$ , and it drops relatively quickly with increasing  $r_{\text{orb}}$  ( $\sim \{12 \text{ per cent}, 7 \text{ per cent}\}$  for  $r_{\text{orb}} = \{40 r_S, 100 r_S\}$ ). Note that this effect is still negligible compared to the  $\sqrt{3}$  factor of variation due to the anisotropic GW emission pattern. Furthermore, the same results apply for both  $\tau_{\text{band}}$  and  $\tau_d$ . From these, we conclude that the effect of the Doppler shift can also be neglected in our investigations.

Time-delay effects such as the Shapiro and the Roemer delay together with the geometric delay occur in the weak-field regime in the first PN order when determining the coordinate time  $t$  between two spatial coordinates in a curved background geometry, e.g. near a Schwarzschild BH (e.g. Shapiro 1964; Blandford & Teukolsky 1976; Lai & Rafikov 2005). Since we used a fully general relativistic treatment to determine  $t$  and thereby  $\tau_d$  in Section 2.3, these effects are already incorporated in our investigations.

### APPENDIX D: REGIMES FOR BBHS MERGING AROUND SMBHS

In this section, we check the validity of our assumptions on the geometric optics limit and the stationary-lens regime for BBH mergers around SMBHs in Sections D1 and D2, respectively.

#### D1 Geometric optics versus wave optics regime

GWs or EM waves passing near a BH are deflected or lensed (e.g. Schneider, Ehlers & Falco 1992; Takahashi & Nakamura 2003). The geometric optics and wave optics regimes are delineated by the

<sup>35</sup>See e.g. D'Orazio & Loeb (2020) for a detailed derivation.

<sup>36</sup>See Gong et al. (2021) and Yu & Chen (2021) for more detailed estimates for sources close to the horizon.

parameter

$$\chi = 8\pi M_{\text{SMBH},z} f_{\text{GW}} = 2.4 \times 10^3 \frac{M_{\text{SMBH},z}}{10^6 M_\odot} \frac{f_{\text{GW}}}{20 \text{ Hz}} \quad (\text{D1})$$

(Takahashi & Nakamura 2003), where  $f_{\text{GW}} = f_{\text{GW, source}}/(1+z)$  is the observer-frame frequency, and the normalization factor 20 Hz accounts for the cut-off frequency upon detection of a binary merger in recent exploratory analyses in LIGO–Virgo–KAGRA searches (Abbott et al. 2021a). The wave optics treatment asymptotes to the case of geometrical optics in the  $\chi \gg 1$  limit (i.e. in the short-wavelength limit), in which case GWs travel on null geodesics of the curved background geometry (i.e. near an SMBH) in exact analogy to light waves (Isaacson 1968).

SMBH masses range from  $\sim 10^5 M_\odot$  to  $\sim 10^{10} M_\odot$ . Accordingly, equation (D1) implies that GW sources observed by aLIGO/AdV/KAGRA have  $\chi \gtrsim 10^2$  if lensed by SMBHs. Specifically for AGNs, the vast majority of SMBHs have masses between  $\sim 10^6 M_\odot$  and  $\sim 10^8 M_\odot$  (Greene & Ho 2007), indicating that  $\chi \gtrsim 10^3$ . From these results, we conclude that these sources are typically in the short-wavelength limit. Thereby, GWs propagate in approximately the same way near SMBHs as light rays.

Note that interference effects may become significant only in special configurations for sources that happen to be close to the line of sight behind or in front of the SMBH (Kocsis 2013). We conservatively neglected this effect in our investigations.

## D2 Lensing regimes

Three different lensing regimes can be delineated based upon the motion of the GW source with respect to the SMBH lens over its observable lifetime  $\tau_{\text{band}}$  for circular orbits around the SMBH (D’Orazio & Loeb 2020):

- (i) the repeating-lens regime:  $1/f_{\text{orb}} \lesssim \tau_{\text{band}}$ ;
- (ii) the slowly moving lens regime:  $\tau_{\text{band}} \gtrsim \tau_{\text{lens}}$ ,  $1/f_{\text{orb}} \gtrsim \tau_{\text{band}}$ ;
- (iii) the stationary-lens regime:  $1/f_{\text{orb}} \gtrsim \tau_{\text{band}}$  and  $\tau_{\text{band}} \lesssim \tau_{\text{lens}}$ .

Here,  $f_{\text{orb}}$  is the orbital period of the GW source around the SMBH in the observer’s frame. Furthermore,  $\tau_{\text{lens}}$  is the time for the source to cross the Einstein radius of the lens  $\tau_{\text{lens}} = r_{\text{Eins}}^{\text{max}}/v_{\text{orb}}$ , where  $r_{\text{Eins}}^{\text{max}} = \sqrt{4M_{\text{SMBH},z} r_{\text{orb}} \sin i}$  is the Einstein radius when the source is directly behind the lens,  $v_{\text{orb}}$  is the orbital velocity of the source around the lens, and  $i$  is the inclination of the binary plane around the SMBH to the line of sight (D’Orazio & Di Stefano 2018). Note that redshift effects cancel out for the ratio of time-scales. Since both  $1/f_{\text{orb}}$  and  $\tau_{\text{lens}}$  increase for larger distances, while  $\tau_{\text{band}}$  is fixed, merger events far from the SMBH are well in the stationary-lens regime in LIGO–Virgo–KAGRA searches. Here, we examine sources in close proximity to the SMBHs to determine where the stationary-lens approximation breaks down. We estimate the fraction of merging BBHs in the stationary-lens regime for two separate cases due to the

distinct  $M_{\text{SMBH}}$  and  $r_{\text{orb}}$  ranges, one in which BBHs merge in AGN discs and one in which they merge in GNs.

First, the vast majority of SMBH masses of AGN discs lie between  $\sim 10^6 M_\odot$  and  $\sim 10^8 M_\odot$  (Greene & Ho 2007). Here we focus on the hypothetical migration traps in slim discs as BBH mergers may merge closest to the SMBH in this channel (even near the ISCO; Peng & Chen 2021) among other merger channels in AGNs (see the Introduction). In this paper, we examine  $r_{\text{orb}} \geq 10 r_s$  and assume circular orbits around the SMBH. As the eccentricity of the binary thermalizes due to binary–single interactions in AGN discs (especially in slim discs; Tagawa et al. 2021a; Samsing et al. 2022), we assume  $\tau_{\text{band}} \lesssim 2 \text{ min}$  as an upper limit, which corresponds to the lightest BBH mergers on highly eccentric orbits, and  $\tau_{\text{band}}$  is significantly lower for BBH mergers with either lower eccentricities or higher masses (Appendix B, and references therein). Most BBH mergers in migration traps may be higher mass BHs (Secunda et al. 2020). The developed MC routine implements the following steps to estimate the fraction of BBH mergers in the stationary-lens regime. (i) We first draw  $10^6 \leq M_{\text{SMBH}} \leq 10^8 M_\odot$  and  $10 r_s \leq r_{\text{orb}} \leq 10^4 r_s$  and  $\sin i$  values from a uniform distribution in the range  $[-1, 1]$  (Section 2.2). (ii) Next, we calculate  $1/f_{\text{orb}}$ , and calculate the corresponding  $\tau_{\text{lens}}$  value to each  $i$  in the MC sample. (iii) Finally, we determine the fraction of BBHs in the stationary-lens regime using the relations introduced between the time-scales  $\{1/f_{\text{orb}}, \tau_{\text{band}}, \tau_{\text{lens}}\}$ . We find that  $\sim 90$  per cent of mergers lies in the stationary-lens regime even for the lowest  $r_{\text{orb}}$  and  $M_{\text{SMBH}}$  values and for  $\tau_{\text{band}} = 2 \text{ min}$ , and this fraction increases steeply for higher  $M_{\text{SMBH}}$  and  $r_{\text{orb}}$  and for lower  $\tau_{\text{band}}$ . Here we note that the mergers outside of the stationary-lens regime are located at  $\gamma \sim \{0, \pi\}$ . We conclude that the stationary-lens approximation is typically applicable for this study.

Now we consider GN hosts with  $10^5 \leq M_{\text{SMBH}} \leq 10^{10} M_\odot$ . To estimate the lower bound of the fraction of sources in the stationary-lens regime, we focus on the single–single GW capture channel in which BBHs form closest to the SMBH, typically outside of a few hundred  $r_s$  for standard models of mass segregation (e.g. O’Leary et al. 2009; Gondán et al. 2018a; Gondán & Kocsis 2021). Note that BBH mergers in the single–single capture channel may form closer to the SMBH than a few hundred  $r_s$  for strong mass segregation. However, in this case, high-mass BBHs dominate the merger rates, and among them, the most massive ones are located closest to the SMBH, for which  $\tau_{\text{band}}$  is much lower. Repeating the above analysis, we find that more than 90 per cent of binaries are in the stationary-lens regime for  $\{M_{\text{SMBH}}, r_{\text{orb}}, \tau_{\text{band}}\} = \{10^5 M_\odot, 100 r_s, 2 \text{ min}\}$ , and this fraction becomes much higher for higher  $M_{\text{SMBH}}$  or  $r_{\text{orb}}$  or for lower  $\tau_{\text{band}}$ . We conclude that the stationary-lens approximation is typically justified for BBH mergers in GNs.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.