

# The mass gap, the spin gap, and the origin of merging binary black holes

Vishal Baibhav<sup>1,\*</sup>, Davide Gerosa<sup>2</sup>, Emanuele Berti<sup>1</sup>, Kaze W. K. Wong<sup>1</sup>, Thomas Helfer<sup>1</sup>, and Matthew Mould<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA*

<sup>2</sup>*School of Physics and Astronomy & Institute for Gravitational Wave Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom*



(Received 6 April 2020; accepted 26 June 2020; published 6 August 2020)

Two of the dominant channels to produce the black-hole binary mergers observed by LIGO and Virgo are believed to be the isolated evolution of stellar binaries in the field and dynamical formation in star clusters. Their relative efficiency can be characterized by a “mixing fraction.” Pair instabilities prevent stellar collapse from generating black holes more massive than about  $45 M_{\odot}$ . This “mass gap” only applies to the field formation scenario, and it can be filled by repeated mergers in clusters. A similar reasoning applies to the binary’s effective spin. If black holes are born slowly rotating, the high-spin portion of the parameter space (the “spin gap”) can only be populated by black-hole binaries that were assembled dynamically. Using a semianalytical cluster model, we show that future gravitational-wave events in either the mass gap, the spin gap, or both can be leveraged to infer the mixing fraction between the field and cluster formation channels.

DOI: [10.1103/PhysRevD.102.043002](https://doi.org/10.1103/PhysRevD.102.043002)

## I. INTRODUCTION

Gravitational-wave (GW) observations of merging black-hole (BH) binaries are bringing us into a new era where many questions are still unanswered. How, when, and where do these binaries form? What is the core physics that drives them to merge?

The two most popular formation channels are isolated binary evolution in the field and dynamical formation in clusters (see e.g., [1,2] for reviews). For isolated binaries, the most promising mechanism to catalyze mergers is a common-envelope phase in between the formation of the two BHs. Alternatively, dynamical channels predict that binary BHs (BBHs) form and harden through three-body encounters in dense stellar clusters. Other scenarios for the formation and merger of BBHs include chemically homogeneous evolution [3,4], active galactic nuclei (AGN) disks [5–7], secular interactions in triples [8–10], and primordial BHs [11]. Different formation pathways leave different imprints on the properties of the BBH population, including the binary masses, spins, eccentricities, and redshift evolution. Measuring these distributions informs us on the environment in which BBHs form and evolve [12–16].

One of the most promising signatures is the distribution of BH spins: systems formed through dynamical interactions are expected to have isotropic spin orientations, whereas binaries born in the field are more likely to have aligned spins [17–20].

However, if BHs are naturally born with low spins, it becomes harder to differentiate between formation channels using spin alignment. LIGO/Virgo observations indicate that this may be the case: the majority of the events reported so far have “effective spin”<sup>1</sup>  $\chi_{\text{eff}} \simeq 0$ . More specifically, all detections from Ref. [21] but two (GW151226 and GW170729) are compatible with  $\chi_{\text{eff}} = 0$  at 90% confidence, although this is a somewhat prior-dependent statement [22]. For GW151226 and GW170729, the 90% lower limit on the effective spin is as low as  $\chi_{\text{eff}} \sim 0.1$  [21]. A recent study [23] showed that the effective spin distribution of LIGO observations is almost consistent with a Dirac delta centered at  $\chi_{\text{eff}} = 0$ . The additional triggers reported in Refs. [24,25], if astrophysical in nature, might be high-spin outliers with  $\chi_{\text{eff}} \gtrsim 0.5$ , but this is also a prior-dependent statement [26].

Recent stellar-physics simulations also suggest that BHs are born with very low spins. Efficient core-envelope interactions may transfer the angular momentum of the progenitor star away from the collapsing core, resulting in BH spins  $\chi \sim 10^{-2}$  [27].

If stellar-mass BHs do indeed rotate very slowly, we will not be able to differentiate the aligned and isotropic

<sup>1</sup>For a binary with component masses  $m_1 > m_2$ , mass ratio  $q = m_2/m_1$ , and dimensionless spins of magnitude  $\chi_i$  at angles  $\theta_i$  ( $i = 1, 2$ ) with respect to the orbital angular momentum, the effective spin  $\chi_{\text{eff}} \equiv (\chi_1 \cos \theta_1 + q\chi_2 \cos \theta_2)/(1 + q)$  is a mass-weighted combination of the components of the BH spins parallel to the binary’s orbital angular momentum.

\*vbaibha1@jhu.edu

populations, making it difficult (if not impossible) to use spin alignment to disentangle BH mergers formed in the field from dynamically formed binaries.

In this paper we identify specific observational signatures that are enhanced if spins are indeed small. In a nutshell, we exploit specific regions of the parameter space which can plausibly be populated by only one of the two scenarios. These “reserved regions” or “gaps” provide a new handle to infer the mixing fraction between the underlying formation channels.

Suppose, for simplicity, that only two formation channels (“field” and “cluster”) are at play for  $N$  BBH detections:

$$N_{\text{field}} + N_{\text{cluster}} = N. \quad (1)$$

The fraction of observation from the “cluster” scenario is

$$f \equiv \frac{N_{\text{cluster}}}{N}, \quad (2)$$

while the “field” fraction is given by  $1 - f$ . Let us further separate the fraction of the catalog entries that are inside or outside a specific region of the parameter space (“gap”), i.e.,

$$N = N_{\text{no gap}} + N_{\text{gap}}. \quad (3)$$

This gap is a reserved region, in the sense that it can only be populated by one of the models (say the “cluster” model): this implies  $N_{\text{field,gap}} = 0$ , and therefore  $N_{\text{gap}} = N_{\text{cluster,gap}}$ . If the efficiency of the “cluster” model at populating the gap

$$\lambda \equiv \frac{N_{\text{cluster,gap}}}{N_{\text{cluster}}} \quad (4)$$

can be reliably estimated, one immediately obtains an estimate of the number of binaries coming from each population:

$$N_{\text{cluster}} = \frac{N_{\text{gap}}}{\lambda}, \quad N_{\text{field}} = N - \frac{N_{\text{gap}}}{\lambda}, \quad (5)$$

or equivalently of the mixing fraction:

$$f = \frac{N_{\text{gap}}}{\lambda N}. \quad (6)$$

For instance, if  $N \sim 100$  events are detected during LIGO/Virgo’s third observing run O3 and one of them lies in the gap, an efficiency  $\lambda \sim 5\%$  would imply that  $f \sim 20\%$  of the observed BH binaries must have formed in clusters, and the remaining  $1 - f \sim 80\%$  must have formed in the field.

Gaps in the parameter space are naturally populated by hierarchical BH mergers. When two BHs merge in the field, the remnant BH does not interact again with other BHs.

This is not necessarily true for BHs that merge in clusters. If these “second-generation” (henceforth 2g) BHs remain in the cluster, they might continue to interact with other BHs, eventually forming new binaries and merging again [28–31]. These 2g BHs will, on average, be heavier than their ancestors. Moreover, binary formation and hardening tend to occur faster for heavier objects, and thus mergers occur more often.

Both supernova models and LIGO observations [16] indicate the presence of a mass gap (usually referred to as the “upper mass gap,” to distinguish it from the putative gap between BHs and neutron stars). Pair-instability supernova (PISN) and pulsational pair-instability supernova (PPISN) [32] prevent the formation of BHs with masses larger than  $\sim 45 M_{\odot}$  [33–37]. The pair-instability mass gap is our first reserved region: if a merging binary with a component BH heavier than the PISN threshold is found, this would point towards a hierarchical origin.

When two “first-generation” (henceforth 1g) BHs merge, they form a remnant with a unique distribution of spins, which is largely independent of the spins of their progenitors. In particular, remnant spins are strongly peaked at  $\chi \sim 0.7$  [28,29,38]. This is our second reserved region, which we call the “spin gap” (although to be rigorous we should call it the “effective spin gap”): if BHs are indeed born with low spins from stellar collapse, the detection of a highly spinning object would also indicate a hierarchical origin. The mass and spin gaps are illustrated in Fig. 1.

A 2g merger can occur only if (i) the preceding 1g merger happened *in situ* and (ii) the merger remnant remains bound to the cluster. Only BHs that receive kicks smaller than the escape speed of the clusters can be retained and potentially merge again. Conversely, the detection of 2g mergers can be used to constrain the escape speed of clusters [31,39]. Generic BH recoils are  $\mathcal{O}(100 \text{ km/s})$  [31,40], but kick velocities tend to zero for BHs with similar masses and small spins, as indicated by current observations.

While we assume that 2g mergers happen only in dense star clusters, other astrophysical mechanisms (such as gas accretion [41], stellar mergers [36], population III stars [42–44], or gravitational lensing [45,46]) could lead to events that contaminate these gaps and complicate the measurement of the mixing fraction  $f$ . However, these mechanisms are expected to be subdominant. Furthermore, it should still be possible to disentangle the population of dynamically formed 2g mergers from other sources because of the unique relationship between the 1g and 2g populations.

The rest of this paper puts these ideas on more solid footing. In Sec. II we describe a semianalytical cluster model based on simple prescriptions, which, however, can replicate the main features relevant to BH mergers of the more complex and computationally expensive Monte Carlo simulations [47–49] and direct  $N$ -body simulations [50,51].

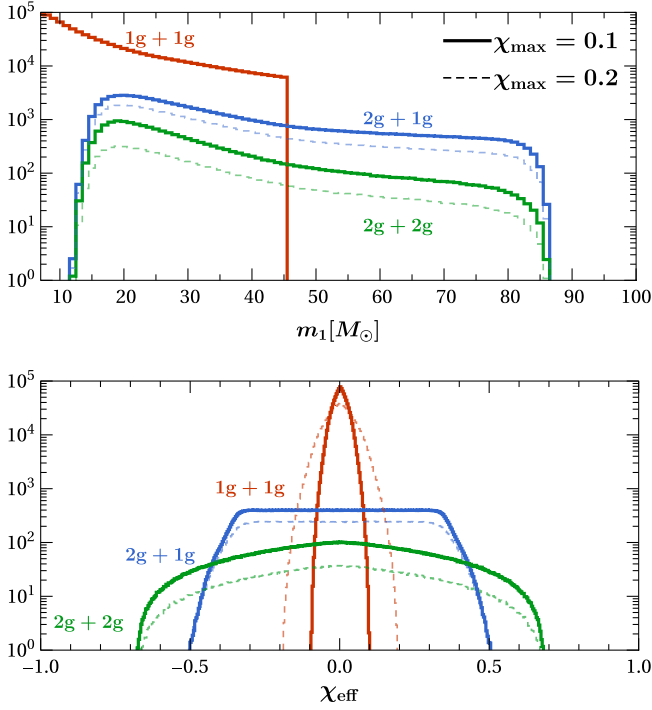


FIG. 1. Illustration of the “mass gap” in the primary mass  $m_1$  (top panel) and of the “spin gap” in the effective spin  $\chi_{\text{eff}}$  (bottom panel). Solid (dashed) lines are computed under the assumption that the maximum individual BH spin at birth is  $\chi_{\text{max}} = 0.1(0.2)$ . Only 2g events can populate the regions of the parameter space with high values of  $m_1$  and/or  $\chi_{\text{eff}}$ .

In Sec. III we use this model to predict the fraction of events populating the mass and spin gaps. In Sec. IV A we use simple analytical approximations for the effective spin probability distribution functions (PDFs) in field binaries and cluster binaries to estimate measurement errors on the “mixing fraction” between field and cluster events using only 1g mergers, and in Sec. IV B we show that using the mass and spin gaps can yield better estimates of the mixing fraction. In Sec. V we summarize our results and discuss directions for future research. The derivation of the spin PDFs is presented in Appendix. Throughout the paper, we use cosmological parameters from Ref. [52].

## II. HIERARCHICAL MERGERS WITH A SEMIANALYTICAL CLUSTER MODEL

In this section we use a semianalytical cluster model which is not meant to replace  $N$ -body simulations but serves our main purpose: relating the bulk properties of clusters to the characteristics of binary mergers which can be observed in GWs.

### A. Binary formation and mergers in clusters

Massive clusters are hotbeds for multiple-generation BBH mergers, but a good understanding of their evolution is elusive because the large number of particles comprising

these systems makes numerical simulations extremely challenging. We evolve binary BHs in clusters following Refs. [53,54]. Reference [53] used a semianalytical approach to predict rates and properties of inspiraling BH binaries forming in nuclear star clusters (NSCs), while Ref. [54] combined a cosmological model of globular cluster (GC) formation with analytical prescriptions from Ref. [53] to study the properties of dense clusters that form merging BH binaries.

We calibrate the half-mass radius  $r_h$  for GCs to fits of late-type galaxies [53,55]:

$$r_h = \begin{cases} 3 \text{ pc} & \text{if } M_{\text{cl}} \leq M_{\text{NSC}}, \\ 2.14 \left( \frac{M_{\text{cl}}}{10^6 M_\odot} \right)^{0.321} \text{ pc} & \text{if } M_{\text{cl}} > M_{\text{NSC}}, \end{cases} \quad (7)$$

where we set  $M_{\text{NSC}} = 2.87 \times 10^6 M_\odot$  (slightly lower than the value  $M_{\text{NSC}} = 5 \times 10^6 M_\odot$  used in Ref. [53]) to ensure continuity between the GC and the NSC regime.

The escape velocity from the cluster is [53]

$$v_{\text{esc}} \simeq 0.1 \sqrt{\frac{M_{\text{cl}} \text{ pc}}{M_\odot r_h}} \text{ km s}^{-1}, \quad (8)$$

and the velocity dispersion is given by  $\sigma_\star = v_{\text{esc}}/(2\sqrt{3})$ , as predicted by Plummer’s model. The number density of stars at the center of the cluster is set to [53]

$$n_\star = 4 \times 10^6 \left( \frac{\sigma_\star}{100 \text{ km s}^{-1}} \right)^2 \text{ pc}^{-3} \quad (9)$$

to match observations [56,57].

### 1. Mass segregation

Clusters containing a subpopulation of BHs with average mass  $\langle m_{\text{BH}} \rangle$  will segregate to the cluster core on a timescale [58]

$$t_{\text{MS}} = \frac{\langle m_\star \rangle}{\langle m_{\text{BH}} \rangle} t_r(r_h), \quad (10)$$

where  $\langle m_\star \rangle$  is the mass of a typical star in the cluster and  $t_r(r_h)$  is the relaxation time at the half-mass radius

$$t_r(r_h) = 4.2 \times 10^9 \left( \frac{15}{\ln \Lambda} \right) \left( \frac{r_h}{4 \text{ pc}} \right)^{3/2} \left( \frac{M_{\text{cl}}}{10^7 M_\odot} \right)^{1/2} \text{ yr}. \quad (11)$$

We set the Coulomb logarithm parameter to  $\Lambda \simeq 0.4 N_\star$  [58], where  $N_\star \simeq M_{\text{cl}}/\langle m_\star \rangle$  is the number of stars in the cluster.

As they fall into the core, BHs lose energy to stars, which become more energetic and migrate outwards. Over time the BHs become confined to an ever smaller core, where

fewer stars are available to carry out the energy. Eventually, BHs decouple from the rest of the cluster population. Assuming that the fraction of the total cluster mass contained in BHs is [54]

$$f_{\text{BH}} = \frac{M_{\text{BH}}}{M_{\text{cl}}} = 0.05 \quad (12)$$

and that BHs are confined in the “BH half-mass radius”

$$r_{\text{BH}} \equiv \frac{M_{\text{BH}}}{M_{\text{cl}}} r_{\text{h}} = f_{\text{BH}} r_{\text{h}}, \quad (13)$$

one can find the number density of BHs as [59]

$$n_{\text{BH}} = n_{\star} \frac{M_{\text{BH}}}{M_{\text{cl}}} \frac{r_{\text{h}}^3}{r_{\text{BH}}^3} = f_{\text{BH}}^{-2} n_{\star}, \quad (14)$$

where  $n_{\star}$  is the number density of stars in the core.

The velocity dispersion of BHs in this dynamically decoupled core is related to the stellar dispersion through the temperature ratio

$$\xi = \frac{\langle m_{\text{BH}} \rangle \sigma_{\text{BH}}^2}{\langle m_{\star} \rangle \sigma_{\star}^2} = 5, \quad (15)$$

where for the latest equality we follow Refs. [54,60,61].

## 2. Formation of BH binaries

In the dense environment of the cluster core, BHs can efficiently form binaries, which will then harden and eventually merge through the following processes.

- (1) *Three-body interactions.* If the density is high enough, a close encounter between three single BHs can lead to the formation of a BH binary, with the third BH carrying away the energy needed to bind the pair. The timescale to form a binary via three-body interactions is [59]

$$t_{3\text{bb}} = 6.45 \times 10^9 \left( \frac{n_{\text{BH}}}{10^6 \text{ pc}^{-3}} \right)^{-2} \left( \frac{\sigma_{\text{BH}}}{10 \text{ km s}^{-1}} \right)^9 \times \left( \frac{m_1}{10 M_{\odot}} \right)^{-5} \text{ yr.} \quad (16)$$

where  $m_1$  is the mass of the heaviest BH in the triple system. Three-body binary formation is highly efficient because of the strong dependence on the velocity dispersion, which is much smaller for BHs compared to stars: cf. Eq. (15).

- (2) *Binary-single interactions.* Clusters also have a population of stellar binaries, which tend to sink towards their cores because they are heavier than single stars. Once inside the core, these binaries undergo binary-single interactions with BHs. Most such encounters end up in exchanges between the

BH and the lighter of the two stars in the binary. If the cluster core contains enough hard stellar binaries, a BH of mass  $m_{\text{BH}}$  can form a binary with a star via exchange interactions on a timescale [62]

$$t_{1-2} = 5 \times 10^9 \left( \frac{f_b}{0.1} \right)^{-1} \left( \frac{n_{\star}}{10^4 \text{ pc}^{-3}} \right)^{-1} \frac{\alpha_{\text{BH}}}{10 \text{ km s}^{-1}} \times \left( \frac{2\langle m_{\star} \rangle + m_{\text{BH}}}{20 M_{\odot}} \right)^{-1} \left( \frac{a_{\text{hard}}}{1 \text{ AU}} \right)^{-1} \text{ yr.} \quad (17)$$

where  $f_b = 0.1$  is the binary fraction in the core [63] and  $a_{\text{hard}}$  is the typical separation of a hard stellar binary. The latter is estimated as [64,65]

$$a_{\text{hard}} = 1.5 \left( \frac{r_{\text{h}}}{3 \text{ pc}} \right) \left( \frac{\langle m_{\star} \rangle / M_{\text{cl}}}{10^{-5}} \right) \text{ AU}, \quad (18)$$

which corresponds to the maximum separation of a hard stellar binary in the core. This BH-star system might form a BH-BH binary following another exchange interaction with a single BH on timescales smaller than  $t_{1-2}$ . Comparing Eqs. (16) and (17) shows that three-body binary formation is likely to dominate the dynamical formation of BH binaries because the binary fraction  $f_b$  is rather small.

- (3) *GW captures.* BHs can also form binaries through single-single GW capture. In this case, two single BHs become bound after a close encounter if sufficient energy is dissipated via GWs. Such BH binaries are predicted to be very eccentric [66] and, consequently, merge almost instantly [67]. The rate of single-single GW capture mergers is comparable to that of binary-single interactions only when the binary fraction is at the percent level [66]. For  $f_b = 0.1$  as assumed here, GW captures can be safely neglected.

To summarize, we define the BBH formation timescale to be  $\min(t_{1-2}, t_{3\text{bb}})$ .

## 3. Hardening and merger

After a “hard” binary (i.e., a binary with binding energy greater than the kinetic energy of cluster particles) is formed, it typically undergoes a series of strong encounters with stars in the core. Because of “Heggie’s law” [64], these repeated encounters tend to make hard binaries harder and soft binaries softer. In a cluster with BH mass density  $\rho_{\text{BH}}$ , a binary will harden at a rate [68]

$$\dot{a}_{\text{dyn}} = -20 \frac{G \rho_{\text{BH}}}{\sigma_{\text{BH}}^2} a^2. \quad (19)$$

If, after an interaction with another BH, the semimajor axis  $a$  of the binary decreases to  $a_{\text{fin}}$ , the binary will recoil with velocity  $\propto (a_{\text{fin}})^{-1/2}$ . This happens because the extra



binding energy is converted to kinetic energy, most of which is carried away by the interloper, while some of it gets transferred to the binary system.

So, while binaries become harder with every encounter, these binaries also receive larger and larger recoils and may eventually be kicked out of the cluster. By equating the recoil speed to the escape velocity of the cluster  $v_{\text{esc}}$ , one can estimate the binary separation at which the binary could be ejected:

$$a_{\text{ej}} = 3.9\eta \left( \frac{v_{\text{esc}}}{30 \text{ km s}^{-1}} \right)^{-2} \frac{m_3}{20 M_{\odot}} \frac{m_3}{m_1 + m_2 + m_3} \text{ AU}, \quad (20)$$

where  $m_1$  and  $m_2$  are the BBH component masses,  $\eta \equiv m_1 m_2 / (m_1 + m_2)^2$  is the symmetric mass ratio of the BBH, and  $m_3 = \langle m_{\text{BH}} \rangle$  is the mass of the BH interloper.

A binary can avoid ejection if gravitational radiation takes over and drives it to merger *before* another dynamical interaction kicks it out of the cluster. The binary separation decays due to gravitational radiation as

$$\dot{a}_{\text{GW}} = -\frac{64 G^3}{5 c^5} \frac{m_1 m_2 (m_1 + m_2)}{a^3} f(e), \quad (21)$$

where  $e$  is the eccentricity and

$$f(e) = \left( 1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right) (1 - e^2)^{-7/2}. \quad (22)$$

We adopt the median value  $e = 1/\sqrt{2}$  expected for a thermal distribution  $p(e) = 2e$ . GWs start dominating the dynamics at the separation  $a_{\text{GW}}$  where dynamical hardening [Eq. (19)] balances GW emission [Eq. (21)]. Setting  $\dot{a}_{\text{dyn}} = \dot{a}_{\text{GW}}$  yields

$$a_{\text{GW}} = 0.05 \left( \frac{m_1 + m_2}{20 M_{\odot}} \right)^{3/5} \times \left[ n \frac{\sigma_{\text{BH}}}{30 \text{ km s}^{-1}} \frac{10^6 M_{\odot} \text{ pc}^{-3}}{\rho_{\text{BH}}} f(e) \right]^{1/5} \text{ AU}. \quad (23)$$

For  $a < a_{\text{GW}}$ , GWs dominate the energy loss from the binary. If  $a_{\text{GW}} < a_{\text{ej}}$ , dynamical encounters eject the BBH from the cluster before GWs can drive the BHs to coalescence. In this case, the ejected BBH (with a separation  $a_{\text{ej}}$ ) can continue to harden *ex situ* via GW emission. On the other hand, if  $a_{\text{GW}} > a_{\text{ej}}$ , GW emission will cause the BBHs to coalesce *in situ* before ejection.

Assuming that each interaction extracts 20% of the binary's binding energy [68], the time to harden to a separation  $a_{\text{crit}} = \max(a_{\text{GW}}, a_{\text{ej}})$ , from an initial separation  $a \gg a_{\text{crit}}$ , is [69]

$$t_{\text{hard}} = \frac{3.16 \times 10^8}{q_3} \left( \frac{\sigma_{\text{BH}}}{30 \text{ km s}^{-1}} \right) \left( \frac{a_{\text{crit}}}{0.05 \text{ AU}} \right)^{-1} \times \left( \frac{m_1 + m_2}{20 M_{\odot}} \right)^{-1} \left( \frac{n_{\text{BH}}}{10^6 \text{ pc}^{-3}} \right)^{-1} \text{ yr}. \quad (24)$$

The binary will continue to interact with other cluster members until it reaches a semimajor axis  $a_{\text{crit}}$ . After reaching  $a_{\text{crit}}$ , the binary's hardening is dominated by GW emission, which drives the system to coalescence on a timescale [67]

$$t_{\text{GW}} = 0.56 \times 10^8 \left[ \frac{m_1 m_2 (m_1 + m_2)}{2 \times 10^3 M_{\odot}^3} \right]^{-1} \times \left( \frac{a_{\text{crit}}}{0.1 \text{ AU}} \right)^4 \left( \frac{1 - e^2}{0.5} \right)^{7/2} \text{ yr}. \quad (25)$$

#### 4. Timescale comparison

The total delay time between the formation of the cluster and a BBH merger is the sum of the timescales for mass segregation [Eq. (10)], BBH formation [ $\min(t_{1-2}, t_{3\text{bb}})$ ], hardening [Eq. (24)], and GW-induced merger [Eq. (25)]:

$$t_{\text{tot}} = t_{\text{MS}} + \min(t_{1-2}, t_{3\text{bb}}) + t_{\text{hard}} + t_{\text{GW}}. \quad (26)$$

Our estimates neglect the lifespans of massive stars that lead to BH formation, which is of order  $\mathcal{O}(1)$  Myr. Figure 2 shows how the different terms in this sum depend on  $M_{\text{cl}}$ :

- (1) *Mass segregation.* BHs sink into the core on a timescale  $t_{\text{MS}} \propto (r_{\text{h}}^3 M_{\text{cl}})^{1/2}$  [Eq. (10)]. In our model GCs have a fixed  $r_{\text{h}}$ , so  $t_{\text{MS}} \propto M_{\text{cl}}^{1/2}$ ; for NSCs the mass segregation timescale is approximately  $t_{\text{MS}} \propto M_{\text{cl}}$  [Eqs. (7) and (11)].
- (2) *BBH formation.* BH binaries form predominantly through three-body interactions, which are very sensitive to the velocity dispersion [Eq. (16)]. Heavier clusters have a larger velocity dispersion, which makes three-body interactions inefficient; therefore, BBH formation timescales increase very steeply with cluster size ( $t_{3\text{bb}} \propto M_{\text{cl}}^{5/2}$ ). As a result, binary formation through three-body interactions is slower in most NSCs compared to GCs, where BBHs could also form through binary-single interactions on a timescale  $t_{1-2} \propto M_{\text{cl}}^{1/6}$  [Eq. (17)].
- (3) *Hardening.* For *ex situ* mergers, binaries are ejected more rapidly for smaller clusters, and therefore they spend less time hardening ( $t_{\text{hard}} \propto M_{\text{cl}}^{1/2}$ ). On the other hand, the hardening timescale for clusters that retain their binaries decreases with cluster mass ( $t_{\text{hard}} \propto M_{\text{cl}}^{-2/5}$ ): larger clusters have higher densities and a larger influx of BHs to the center, which makes dynamical hardening more efficient.
- (4) *GWs.* Small clusters have low escape speeds, and BBHs get ejected at large orbital separations

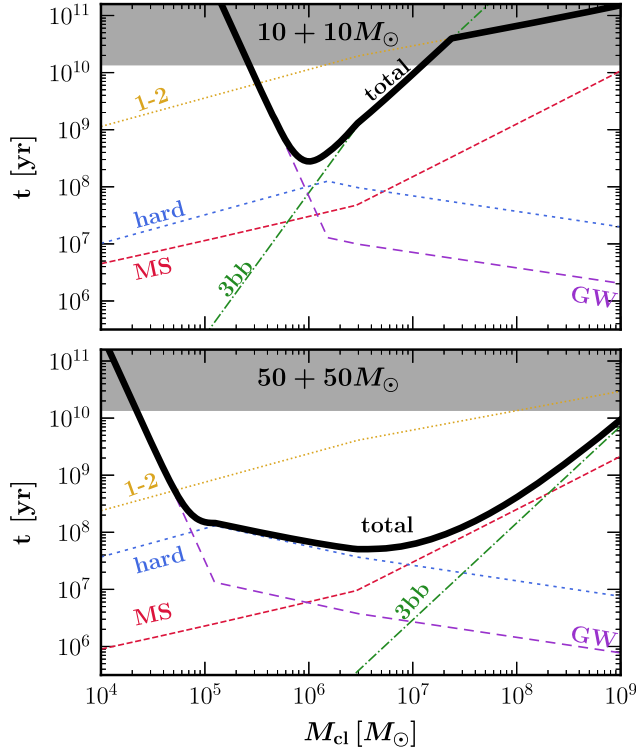


FIG. 2. Timescales involved in the merger of  $(10 + 10) M_\odot$  (top) and  $(50 + 50) M_\odot$  (bottom) BBHs. The timescales related to three-body interactions, binary-single interactions, mass segregation, GW inspiral, and critical hardening are indicated in green, yellow, red, purple, and blue, respectively. The thick black line marks the sum  $t_{\text{tot}}$  of Eq. (26). The gray shaded region marks time delays larger than the age of the Universe.

( $a_{\text{ej}} \propto M_{\text{cl}}^{-1}$ ) due to dynamical interactions. Therefore, the gravitational radiation reaction timescale  $t_{\text{GW}} \propto a_{\text{crit}}^4 \propto M_{\text{cl}}^{-4}$  increases sharply for small clusters. The situation is different for heavier clusters, which retain and dynamically harden BBHs until GW emission takes over: in this case  $a_{\text{GW}} \propto M_{\text{cl}}^{-1/10}$ , and thus  $t_{\text{GW}} \propto M_{\text{cl}}^{-2/5}$ .

The dominant term depends on both the mass of the cluster and the mass of the binary (Fig. 2). For lighter BBHs of  $\sim(10 + 10) M_\odot$ , only the GW timescale matters for clusters with mass  $\lesssim 10^6 M_\odot$ , while the three-body interaction timescale is dominant for large clusters. For larger BBHs of  $(50 + 50) M_\odot$ , hardening time and mass segregation timescales also play an important role, while the GW radiation-reaction timescale and the three-body timescale become important only for  $M_{\text{cl}} \lesssim 10^5 M_\odot$  and  $M_{\text{cl}} \gtrsim 10^8 M_\odot$ , respectively. Figure 2 also confirms our earlier claim that binary-single interactions are not an efficient channel for BBH formation. They only become important for light binaries in very massive clusters, but at that point  $t_{1-2}$  becomes comparable to the age of the Universe.

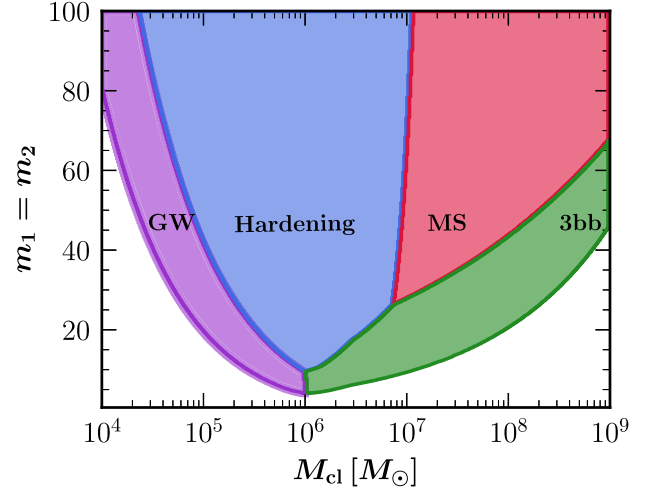


FIG. 3. Dominant timescales in the  $(M_{\text{cl}}, m_1)$  plane for an equal-mass binary ( $m_1 = m_2$ ). Regions where three-body interactions, mass segregation, hardening, and GW inspiral dominate are indicated in green, red, blue, and purple, respectively.

Figure 2 also shows the total delay time for equal-mass binaries as a function of  $M_{\text{cl}}$ . For a binary with fixed component masses  $m_1 = m_2$ , the delay time decreases with  $M_{\text{cl}}$  for clusters of mass  $M_{\text{cl}} \lesssim 10^6 M_\odot$ , where the gravitational radiation or hardening timescales dominate. For  $M_{\text{cl}} \gtrsim 10^6 M_\odot$ , other processes dominate and the delay time increases. The minimum time delay  $t_{\text{tot}}^{\text{min}}$  over all cluster masses is, in general, a function of  $m_1$  and  $m_2$ :

$$t_{\text{tot}}^{\text{min}}(m_1, m_2) = \min[t_{\text{tot}}(m_1, m_2, M_{\text{cl}})], \quad (27)$$

where we take  $M_{\text{cl}} \in [10^5 M_\odot, 10^9 M_\odot]$ . For a given binary of masses  $(m_1, m_2)$ , as long as  $t_0 > t_{\text{tot}}^{\text{min}}$ , there are two values of  $M_{\text{cl}}$ —say  $M_{\text{cl}}^{\text{min}}$  and  $M_{\text{cl}}^{\text{max}}$ —such that

$$\begin{aligned} t_{\text{tot}}(m_1, m_2, M_{\text{cl}}^{\text{min}}) &= t_0, \\ t_{\text{tot}}(m_1, m_2, M_{\text{cl}}^{\text{max}}) &= t_0. \end{aligned} \quad (28)$$

These represent bounds on the range of cluster masses that can produce merging BHs with masses  $m_1$  and  $m_2$  within time  $t_0$ : BBH mergers are possible, on average, when  $M_{\text{cl}}^{\text{min}} \leq M_{\text{cl}} \leq M_{\text{cl}}^{\text{max}}$ . This point will be important later (cf. Sec. II B 1).

Figure 3 illustrates the dominant timescale as a function of cluster mass and BBH mass for an equal-mass binary ( $m_1 = m_2$ ). Most BBHs have small masses (cf. Sec. II B), where  $t_{\text{tot}}$  is dominated by  $t_{\text{GW}}$  and  $t_{3\text{bb}}$  [54]. Both of these timescales decrease sharply with the BH mass ( $t_{\text{GW}} \propto m_1^{-7}$  and  $t_{3\text{bb}} \propto m_1^{-5}$ ), so mass segregation and hardening (which decay as  $m_1^{-1}$  and  $m_1^{-1.6}$ , respectively) can become the dominant timescales only for BBHs with larger masses. The hardening time dominates for  $m_1 \gtrsim 10 M_\odot$ , while the

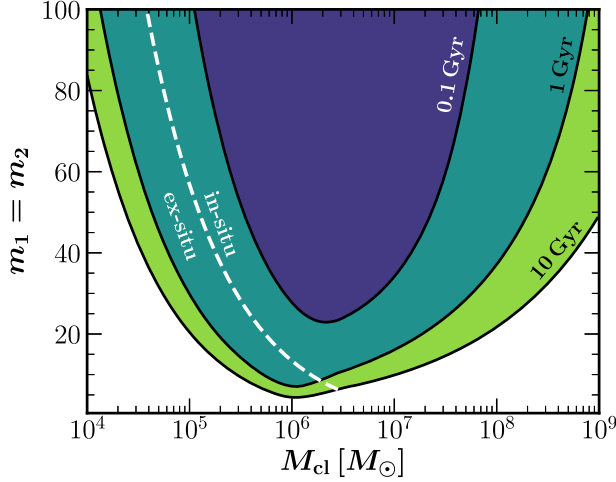


FIG. 4. Contour plot of the total delay time for the merger of an equal-mass BBH system ( $m_1 = m_2$ ) as a function of the cluster mass. The white dashed line is the boundary between the regions where most mergers happen *ex situ* (left) and *in situ* (right).

mass segregation timescale dominates for  $m_1 \gtrsim 25 M_\odot$  in some NSCs with mass  $M_{\text{cl}} \gtrsim 10^7 M_\odot$ .

Figure 4 shows how the time  $t_{\text{tot}}$  varies in the  $(M_{\text{cl}}, m_1)$  plane, assuming again an equal-mass binary ( $m_1 = m_2$ ). In particular, the white dashed line separates the region where mergers occur *in situ* and *ex situ*: small BHs and BBHs hosted in small clusters are ejected because of large recoils and insufficient escape speed, respectively. After they are ejected, these BHs then take a long time to merge under gravitational radiation reaction: as discussed earlier, this phase dominates their entire evolution. More quantitatively, BHs with  $m_1 \lesssim 7 M_\odot$  take  $\gtrsim 10^9$  yrs to merge. Our model does not predict mergers of BHs with  $m_1 \lesssim 4 M_\odot$ , which are always ejected before a binary is formed. We also find that below a minimum primary BBH component mass for *in situ* mergers

$$m_{\text{min, is}} \simeq 6 M_\odot, \quad (29)$$

all mergers happen outside the cluster.

### B. First-generation mergers

Ideally, one should generate the population of merging binaries by convolving a cluster formation model with the delay times discussed in the previous section. There are large uncertainties in this process [70–72], so we choose instead to start by considering the observed population of first-generation BBHs.

We distribute the primary mass of  $1g + 1g$  mergers according to [16]

$$p(m_1) \propto m_1^\alpha, \quad m_1 \in [m_{\text{min}}, m_{\text{max}}], \quad (30)$$

while the secondary mass is drawn from

$$p(m_2|m_1) \propto m_2^\beta, \quad m_2 \in [m_{\text{min}}, m_1]. \quad (31)$$

We fix  $\alpha = -1.6$  and  $\beta = 6.7$ , as estimated from GW observations [16], and  $m_{\text{min}} = 5 M_\odot$ . The parameter  $m_{\text{max}}$  marks the onset of the mass gap, and it is set to  $45 M_\odot$  [35] unless specified otherwise. The spin magnitudes of the component BHs are drawn from a uniform distribution in the range  $[0, \chi_{\text{max}}]$ .

We sample redshifts from the Madau star-formation-rate fit [73]:

$$\text{sfr}(z) = \frac{0.015(1+z)^{2.7}}{1 + [(1+z)/2.9]^{5.6}}. \quad (32)$$

Lighter BHs have longer total delay times  $t_{\text{tot}}$ . Therefore, at any given redshift there is a lower bound on the masses of BBHs that could form and merge within a given time. Assuming that clusters could not have formed earlier than  $t_{\text{max}} = 13.4 \times 10^9$  years ago ( $z_{\text{max}} = 11.34$ ), we discard all binaries that could not have merged at the sampled redshift. In other words, we only keep binaries that satisfy the constraint

$$t_{\text{lookback}}(z) + t_{\text{tot}}^{\text{min}}(m_1, m_2) < t_{\text{max}}, \quad (33)$$

where  $t_{\text{lookback}}$  is the cosmological lookback time [74]. This procedure removes some low-mass and/or high- $z$  binaries, slightly modifying our merging population relative to the sampled distribution.

Starting from this  $1g + 1g$  population, we can now use the semianalytical scheme outlined above to obtain the distribution of  $2g + 1g$  and  $2g + 2g$  binaries.

### 1. Clusters that merge black holes

For each binary we sample the cluster mass from a distribution of the form [75]

$$p(M_{\text{cl}}) \propto M_{\text{cl}}^{-2} \quad (34)$$

in a range of cluster masses  $M_{\text{cl}} \in [M_{\text{cl}}^{\text{min}}, M_{\text{cl}}^{\text{max}}]$  that could support BBH mergers at redshift  $z$ . Here  $M_{\text{cl}}^{\text{min}}$  and  $M_{\text{cl}}^{\text{max}}$  are calculated from Eq. (28), with  $t_0 = t_{\text{max}} - t_{\text{lookback}}(z) > t_{\text{tot}}$ .

The black line in the left panel of Fig. 5 shows the resulting distribution of cluster masses. Smaller clusters are more abundant but relatively inefficient at bringing binaries to merger: BBHs get ejected from the cluster with large orbital separations and therefore have long GW-driven inspiral timescales  $t_{\text{GW}}$ . Most of the merging binaries come from GCs with mass  $\sim 10^6 M_\odot$ : these clusters lie in the “sweet spot” where delay times are the smallest, especially for lighter BBHs, which form the bulk of the population (cf. Figs. 2–4). This behavior is consistent with Refs. [54, 76]. In our model, smaller BHs of component masses  $\sim 5 M_\odot$  can only form in clusters with

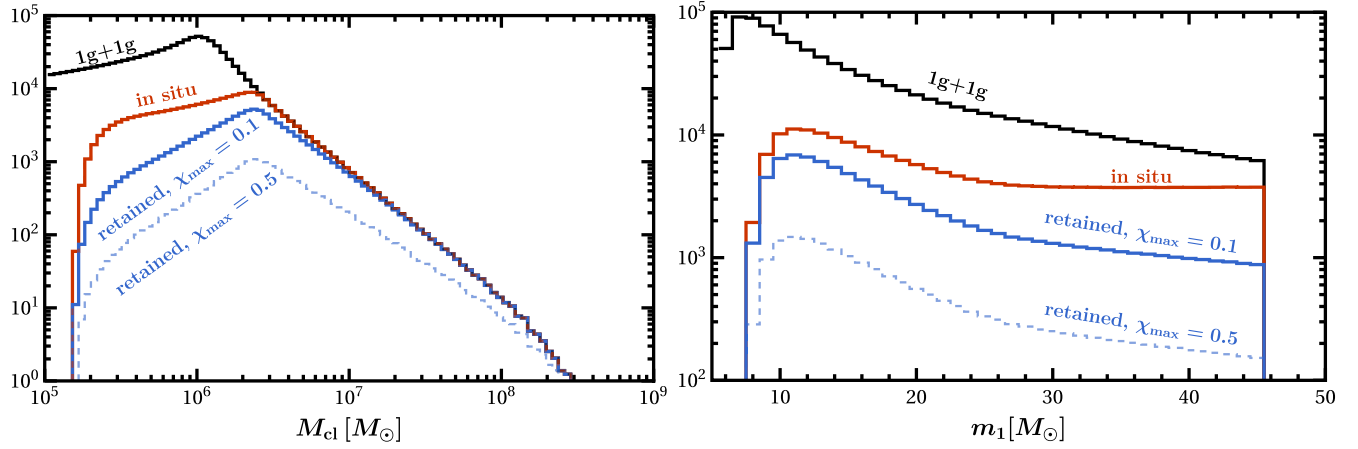


FIG. 5. Distribution of host cluster masses (left) and primary BBH masses (right) for the 1g + 1g populations. Black curves show the full sample of 1g mergers (higher generations are excluded from this plot). Red curves show the fraction of binaries that survive dynamical kicks and are able to merge inside the cluster. Blue curves show systems that further survive GW kicks and remain available to assemble the second generation of BH mergers. We assume  $\chi_{\max} = 0.1$  (solid) and  $\chi_{\max} = 0.5$  (dashed).

$M_{cl} \simeq 10^{5.8} - 10^{6.2} M_{\odot}$ : these are the only systems that can efficiently lead light BBHs to merger. On the other hand, heavier BBHs have smaller delay times, and they can merge more easily within a wider range of cluster masses. Because of the shape of the probability distribution function  $p(M_{cl})$ , most of these massive BBHs come from clusters at the lower end of the mass spectrum.

### C. Hierarchical mergers

Our goal is to look for smoking guns that can be used to identify the 2g population. We must first address a key question: how efficiently do clusters produce 2g mergers?

#### 1. Retention in the cluster

Given a sample of binaries with masses  $(m_1, m_2)$  merging in a cluster of mass  $M_{cl}$ , we assume that all binaries with  $a_{GW} < a_{ej}$  merge inside the cluster. At merger, the remnant receives an additional kick  $v_{kick}$  due to asymmetric GW emission. If  $v_{kick} < v_{esc}$  the remnant is retained in the cluster, where it can merge again and form a 2g binary. The properties of the merger remnant are computed using fits to numerical-relativity simulations for final mass [77], spin [78], and recoils [79–83] as implemented in Ref. [84], assuming isotropic spin orientations.

For illustrative purposes, in Fig. 6 we focus only on the retention power of clusters ignoring the prescriptions of Sec. II B 1, as well as the fact that some clusters might not be able to drive small BHs to merger. We plot both the fraction of *in situ* mergers ( $a_{ej} < a_{GW}$ ) and retained BHs ( $v_{kick} < v_{esc}$ ) for different assumptions on the quantity  $\chi_{\max}$  that marks the “edge” of the spin gap. *In situ* BH mergers are only possible for clusters with mass  $M_{cl} > 1.6 \times 10^5 M_{\odot}$ , where this threshold is mainly set by  $m_{\max}$ . For  $M_{cl} > 4 \times 10^6 M_{\odot}$ —a threshold now set by  $m_{\min}$ —all mergers are found *in situ*. Postmerger kicks increase when

the merging BH spins are large, so the retention fraction decreases steeply with  $\chi_{\max}$ . For larger values of  $\chi_{\max}$ , the population contains more BHs that receive kicks larger than the escape speed of all but the most massive clusters.

In the right panel of Fig. 5, we convolve this retention power and the prescription of Sec. II B 1 to illustrate the final distribution of BBH primary masses  $m_1$ . The 1g + 1g distribution of primary masses follows the injected power law  $m_1^{-1.6}$ . Most small-mass BHs get ejected due to kicks from dynamical interactions, resulting in an almost flat distribution for *in situ* mergers with  $m_1 \gtrsim 30 M_{\odot}$ . In addition, *in situ* mergers do not occur for  $m_1 < m_{\min, is} \simeq 6 M_{\odot}$  (cf. Figs. 4 and 5).

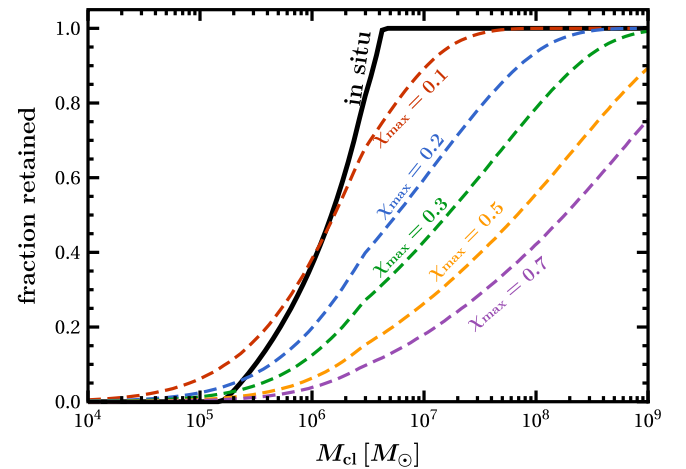


FIG. 6. Fraction of BBHs retained after kicks due to either (i) dynamical interactions before merger (solid line) or (ii) GW recoil at merger (dashed lines). The largest spin of 1g BHs  $\chi_{\max}$  increases from top (red,  $\chi_{\max} = 0.1$ ) to bottom (purple,  $\chi_{\max} = 0.7$ ).



The mass distribution of the retained binaries depends on  $\chi_{\max}$ . Larger  $\chi_{\max}$  leads to large kicks: this reduces the number of retained remnants but also affects the slope of the distribution. This can be understood as follows. Because  $\beta \gg 1$ , most binaries have mass ratio  $q$  close to unity. For  $\chi_{\max} = 0$ , symmetry in the merger process (e.g., [40]) implies  $v_{\text{kick}} = 0$ . In this case, the  $m_1$  distributions of *in situ* and retained mergers should be very similar, with the same plateau at  $m_1 \sim m_{\max}$ . For larger values of  $\chi_{\max}$ , however, most of the small clusters do not retain post-merger remnants, which removes a significant fraction of the heaviest BBHs. These systems can form and merge easily in a wider range of cluster masses, which implies that a significant fraction of them comes from the lighter, more abundant environments. Therefore, large values of  $\chi_{\max}$  lead to an  $m_1$  distribution of retained BBHs which drops more sharply.

The left panel of Fig. 5 shows the mass distribution of clusters that host BBHs. As expected, *in situ* and retained cluster events are more likely in heavier systems with a larger escape speed. In particular, clusters with  $M_{\text{cl}} \gtrsim 3 \times 10^6 M_{\odot}$  are able to produce *in situ* events. The cluster masses that can retain BHs following GW kicks and thus support 2g mergers are sensitive to the maximum spin  $\chi_{\max}$  of 1g BHs. For  $\chi_{\max} \gtrsim 0.5$ , some BHs are ejected even from the most massive NSCs.

## 2. 2g + 1g or 2g + 2g?

If retained, a BBH merger remnant (a “2g BH”) can merge with either another 2g BH (2g + 2g merger) or with a 1g BH (2g + 1g merger). Selective pairing of BH component masses implies that the retained remnants of 1g mergers are more likely to merge with a 2g BH (cf. [31]). However, because heavier BHs merge more quickly, 2g systems have a very small survival time in the cluster compared to 1g BHs. Although 2g BHs would tend to pair with other 2g BHs, their short merger time implies a lower merger probability for 2g + 2g mergers.

In order to take into account some of these complications, we assign a merger probability based on (i) the number density of 1g BHs and (ii) the number density of BHs retained within a given cluster at a given redshift. We divide the 1g population into bins of cluster mass and redshift,  $\mathbf{n}_{1g}(M_{\text{cl}}, z)$ . We do the same for the retained remnants and calculate the number of binaries in each bin,  $\mathbf{n}_{\text{rem}}(M_{\text{cl}}, z)$ . The number of ways in which a 2g + 1g binary can form is proportional to  $\mathbf{n}_{1g} \times \mathbf{n}_{\text{rem}}$ , while for a 2g + 2g binary it is proportional to  $\mathbf{n}_{\text{rem}}(\mathbf{n}_{\text{rem}} - 1)/2$ . So the ratio between 2g + 2g and 2g + 1g mergers in a cluster of mass  $M_{\text{cl}}$  at redshift  $z$  is  $\propto (\mathbf{n}_{\text{rem}} - 1)/(2\mathbf{n}_{1g})$ .

If a 2g BH of mass  $m_1$  merges with another 2g BH, we extract the companion mass  $m_2 \in [m_{\min}, m_1]$  from the same distribution  $p(m_2|m_1) \propto m_2^\beta$  used in Sec. II B. Its spin is estimated by binning and resampling the 2g remnant spin distribution. For the case of 2g + 1g mergers, we extract  $m_2$

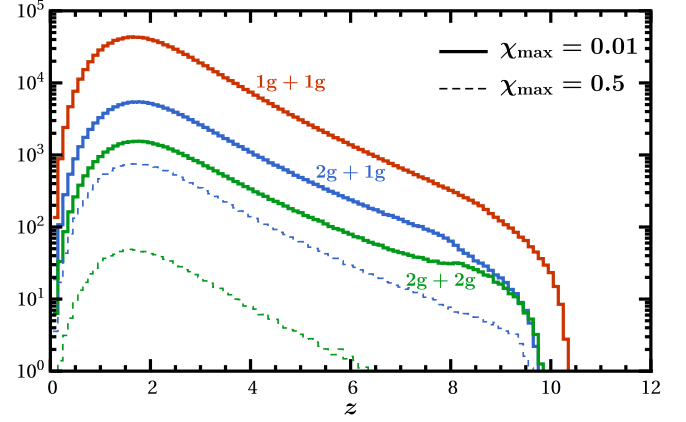


FIG. 7. Distribution of merger redshifts  $z$ . Red, green, and blue curves indicate the 1g + 1g, 2g + 1g, and 2g + 2g populations, respectively. Solid (dashed) histograms are obtained with  $\chi_{\max} = 0.01$  ( $\chi_{\max} = 0.5$ ).

from  $p(m_2|m_1) \propto m_2^\beta$  but now restrict  $m_2 \in [m_{\min}, \min(m_1, m_{\max})]$ . The spin is extracted from the 1g distribution  $p(\chi) = \text{constant}$  with  $\chi \in [0, \chi_{\max}]$ .

We calculate the time delay between a 1g + 1g merger and the next merger involving its 2g remnant using the same  $t_{\text{tot}}$  introduced in Eq. (26) above but replacing the mass segregation timescale  $t_{\text{MS}}$  by  $t_{\text{MS}} \times (v_{\text{kick}}/v_{\text{esc}})^3$  to take into account the time needed by the kicked remnant to sink back into the cluster core. This is obtained from Eq. (10) by assuming that the GW recoil displaces the remnant to  $(v_{\text{kick}}/v_{\text{esc}})^2 r_h$  [53].

The redshift distribution of 1g + 1g mergers was assumed to follow the Madau star formation rate of Eq. (32). We find that 2g BBHs closely follow the same distribution, contrary to the expectation that repeated mergers may suffer further time delays [28]: cf. Fig. 7. This is because merger products are heavier compared to their progenitors and thus merge on very short timescales in our model.

We find that, overall, the number of 2g + 1g events dominates over the 2g + 2g populations with a relative fraction of about 4:1 for  $\chi_{\max} = 0.01$ . The fraction of 2g + 2g mergers decreases and tends to zero as  $\chi_{\max} \rightarrow 1$ .

## D. Caveats

The  $M_{\text{cl}}^{-2}$  scaling of cluster masses [Eq. (34)] is based on present-day observations. A larger number of heavier clusters might have been present (and subsequently disrupted) at larger redshifts. This could increase the fraction of binaries retained in the cluster and hence the number of 2g mergers. Moreover, we fixed  $r_h$  as a function of  $M_{\text{cl}}$  [Eq. (7)] based on fits from Ref. [53]. However, a wide distribution of  $r_h$  is observed for a given  $M_{\text{cl}}$ . This means that even for small  $M_{\text{cl}}$ , one can find smaller  $r_h$  (and hence larger  $v_{\text{esc}}$ ) than assumed here. Therefore, 2g mergers could occur efficiently in many clusters smaller than those

discussed here. We also assumed that all clusters have the same distribution of BHs given by  $\alpha$  and  $m_{\max}$ . More realistically, the mass distribution of BHs should be a function of metallicity, which in turn has a complicated correlation with  $M_{\text{cl}}$ , the mass of the host galaxy, etc. [85]. These complications were ignored in our simple model, and they are beyond the scope of the present work.

### III. FILLING THE MASS AND SPIN GAPS BY HIERARCHICAL MERGERS

We now discuss two key features of the observed merger distribution that can help to identify the origin of BBH mergers: the mass gap and the spin gap.

#### A. The mass gap

Theoretical studies suggest the existence of a gap in the BH mass function of 1g mergers above  $m_{\max} \sim 45M$  [35] due to PISN and PPISN [32]. The distribution of BBH mergers detected in O1 and O2 already hints at the possible existence of an upper bound of  $\sim 40 M_{\odot}$  on the component BH masses [16]. The detection of BH binaries with component masses in the mass gap could be evidence that repeated mergers are at play.

In our model, only 2g BHs can have masses above  $m_{\max}$ . We thus define the number of mergers in the mass gap as

$$N_{\text{Mgap}} \equiv N_{1g+2g}(m_1 > m_{\max}) + N_{2g+2g}(m_1 > m_{\max}). \quad (35)$$

The efficiency of the cluster model at populating the mass gap is given by

$$\lambda_{\text{M}} \equiv \frac{N_{\text{Mgap}}}{N_{\text{cluster}}}. \quad (36)$$

The top panel of Fig. 1 shows the distribution of primary masses for  $\chi_{\max} = 0.1, 0.2$ . In both cases, around 25% of 2g mergers lie in the mass gap.

#### B. The spin gap

While core collapse might leave behind slowly rotating BHs [27], BBH mergers produce remnants with a spin distribution peaked at  $\chi \sim 0.7$  [28,29,38]. This is the second smoking-gun signature of 2g mergers, as shown in the bottom panel of Fig. 1 for  $\chi_{\max} = 0.1$ . The effective spin distributions of the 2g + 1g and 2g + 2g populations is broader compared to the 1g + 1g case, with events leaking in the region where  $|\chi_{\text{eff}}| > \chi_{\max}$ . This region is the spin gap.

Much like evading the PISN/PPISN constraint is a prerogative of repeated mergers, we find that populating the spin gap is also a strong indication of 2g events. The 2g + 1g and 2g + 2g populations are, collectively, quite distinct from the 1g + 1g binaries. In particular, 2g events constitute only  $\sim 1\%$  of mergers with effective spin outside

the gap ( $|\chi_{\text{eff}}| < \chi_{\max}$ ). This assumption is solid as we change  $\chi_{\max}$ : we find that the fraction of 2g events outside the spin gap peaks at 2% for  $\chi_{\max} = 0.24$ .

The number of events in the spin gap is defined as

$$N_{\text{Sgap}} \equiv N_{1g+2g}(|\chi_{\text{eff}}| > \chi_{\max}) + N_{2g+2g}(|\chi_{\text{eff}}| > \chi_{\max}), \quad (37)$$

while

$$\lambda_{\text{S}} \equiv \frac{N_{\text{Sgap}}}{N_{\text{cluster}}} \quad (38)$$

is the cluster efficiency at populating this region. In the bottom panel of Fig. 1, for example, 74% of all 2g mergers lie in the spin gap when  $\chi_{\max} = 0.1$ . This number reduces to 50% for  $\chi_{\max} = 0.2$ .

#### C. Gap efficiencies

The maximum 1g spin  $\chi_{\max}$  is the main parameter that determines the efficiencies  $\lambda_{\text{M}}$  and  $\lambda_{\text{S}}$ , which are shown as thick lines in Fig. 8. For  $\chi_{\max} = 10^{-2}$ , about 5% (14%) of all mergers lie in the mass (spin) gap. These are conservative upper limits: if  $\chi_{\max}$  is increased, merger products receive larger and larger GW kicks, and the number of 2g mergers decreases drastically. For  $\chi_{\max} = 0.5$ , only 0.4% of mergers lie in the mass gap. The effect is even more severe for the spin gap, which remains nearly empty ( $\approx 0.01\%$  of the events).

The spin efficiency is largely independent of both  $m_{\max}$  and  $\alpha$ . The mass efficiency, on the other hand, depends on  $\alpha$

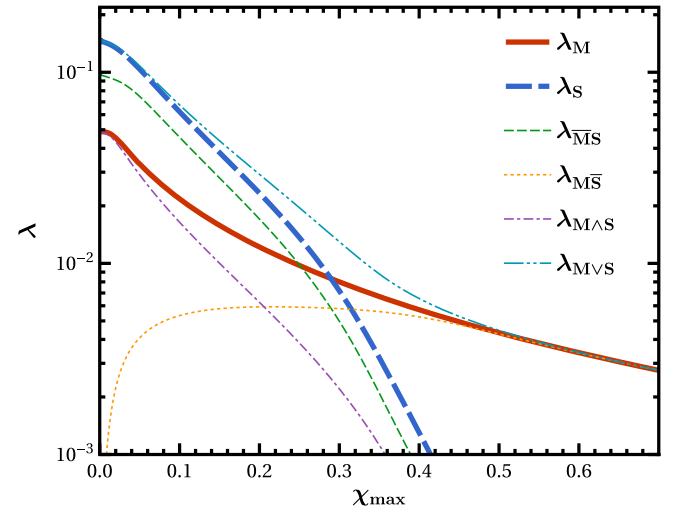


FIG. 8. Fraction of events that lie in one or both gaps. The total contributions to the mass ( $\lambda_{\text{M}}$ ) and spin ( $\lambda_{\text{S}}$ ) gap are given by the thick solid red and thick dashed blue curves, respectively. The other curves indicate contributions from binaries that are in one gap but not in the other one ( $\lambda_{\text{MS}}$  and  $\lambda_{\text{MS}}$ ), in both gaps ( $\lambda_{\text{M} \wedge \text{S}}$ ), or in either of the two gaps ( $\lambda_{\text{M} \vee \text{S}}$ ).

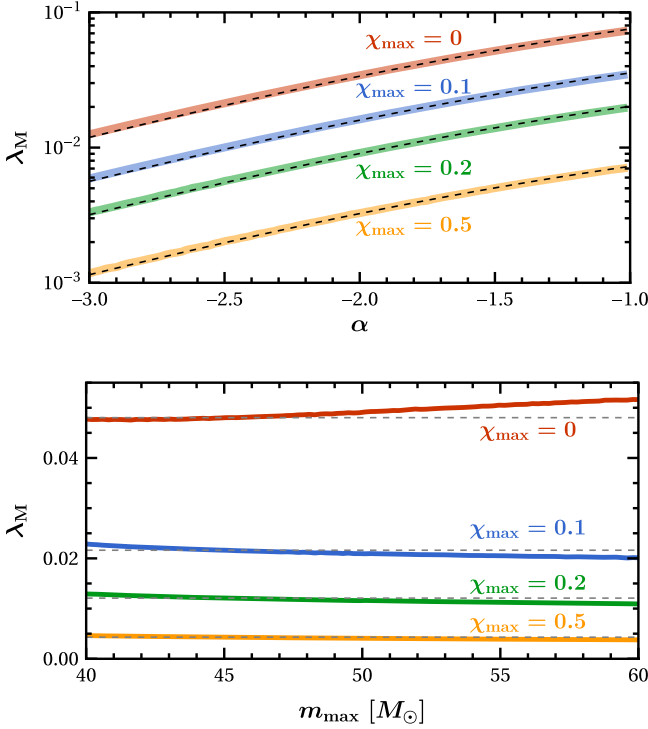


FIG. 9. Mass-gap efficiency  $\lambda_M$  as a function of the mass spectral index  $\alpha$  (top panel) and the gap edge  $m_{\max}$  (bottom panel). The largest 1g spin  $\chi_{\max}$  is varied from 0 to 0.5 (top to bottom in each panel). In the top panel, black dashed lines show the approximate dependence from Eq. (42) with  $\alpha = -1.6$ . In the bottom panel, black dashed lines represent our default value  $m_{\max} = 45M_\odot$ .

and, to a lesser extent, on  $m_{\max}$ . These trends are illustrated in Fig. 9, which can be understood by a simplified model as follows.

Let us assume that the component masses of the population retained after merger follow some power law  $p_{1g}(m) \propto m^{\alpha'}$  for  $m \in [m_{\min,1g}, m_{\max}]$ . Let us also neglect energy dissipation during merger. The primary masses of 2g events will be distributed according to

$$p_{2g}(m) = \frac{1}{2} \frac{1 + \alpha'}{m_{\max}^{1+\alpha'} - m_{\min,1g}^{1+\alpha'}} \left(\frac{m}{2}\right)^{\alpha'}, \quad (39)$$

where  $m \in [2m_{\min,1g}, 2m_{\max}]$ . The probability that an event lies in the mass gap, i.e.,  $m > m_{\max}$ , is given by

$$\Lambda(\alpha') = \int_{m_{\max}}^{2m_{\max}} p_{2g}(m) dm = \frac{2^{-(\alpha'+1)} - 1}{Q^{-(\alpha'+1)} - 1}, \quad (40)$$

where

$$Q \equiv \frac{m_{\max}}{m_{\min,1g}} \sim 7.5. \quad (41)$$

We can approximate the dependence of  $\lambda_M$  on  $\alpha$  by rescaling

$$\lambda_M(\chi_{\max}, \alpha) = \frac{\Lambda(\alpha)}{\Lambda(\alpha')} \lambda_M(\chi_{\max}, \alpha'). \quad (42)$$

The scaled value  $\lambda(\alpha = -1.6)$  is shown in the top panel of Fig. 9 with black dashed lines. Our analytical approximation closely follows the estimate provided by the full cluster model. This agreement is somewhat surprising because the primary components of retained BHs are not distributed according to a power law at low masses (cf. the right panel of Fig. 5).

The bottom panel of Fig. 9 shows that  $\lambda_M$  is very mildly dependent on  $m_{\max}$ . For example, for  $\chi_{\max} = 0.1$ ,  $\lambda_M$  changes from 2.5% at  $m_{\max} = 40 M_\odot$  to 2% at  $m_{\max} = 60 M_\odot$ .

#### D. One or both gaps?

Although we argue that both gaps are smoking-gun signatures of hierarchical mergers, only a subset of binaries will have both large masses and large spins. Other sources will lie in one of the gaps but not the other.

The various contributions are shown in Fig. 8. In particular,

- (i)  $\lambda_S$  and  $\lambda_M$  are the spin- and mass-gap efficiencies introduced above in Eqs. (36) and (38);
- (ii)  $\lambda_{M \wedge S}$  is the fraction of events that lie in *both* the mass gap and the spin gap;
- (iii)  $\lambda_{M \vee S}$  is the fraction of events that lie in *either* the mass gap or the spin gap;
- (iv)  $\lambda_{M \bar{S}}$  is the fraction of events that lie in the mass gap but not in the spin gap; and
- (v)  $\lambda_{\bar{M} S}$  is the fraction of events that lie in the spin gap but not in the mass gap.

For  $\chi_{\max} \sim 0$ , the spin gap occupies a large portion of the parameter space, which implies that all binaries in the mass gap must also be in the spin gap:  $\lambda_M \rightarrow \lambda_{M \wedge S}$ . The opposite is true for  $\chi_{\max} \gtrsim 0.7$ : the spin gap shrinks and, consequently,  $\lambda_M \rightarrow \lambda_{M \bar{S}} \sim \lambda_{M \vee S}$ . A future event with large mass but small effective spin ( $\bar{M}\bar{S}$ ) can be explained by our model only if  $\chi_{\max}$  is sufficiently large. If BHs are indeed born with negligible spins  $\chi_{\max} \sim 10^{-2}$ , we find that *a mass-gap event should also be in the spin gap*. This is an important feature of our model, which can potentially allow us to disentangle the contribution to the mass gap provided by hierarchical mergers (as considered here) from other mechanisms.

We can similarly define the probability of producing events in one of the two gaps but not in the other:

$$p(\bar{S}|M) = \frac{\lambda_{M \bar{S}}}{\lambda_M} = 1 - \frac{\lambda_{M \wedge S}}{\lambda_M}, \quad (43)$$

$$p(\bar{M}|S) = \frac{\lambda_{\bar{M} S}}{\lambda_S} = 1 - \frac{\lambda_{M \wedge S}}{\lambda_S}. \quad (44)$$

These are shown in Fig. 10. Events in the spin gap have a  $\sim 67\%$  probability of not being in the mass gap (blue curve)

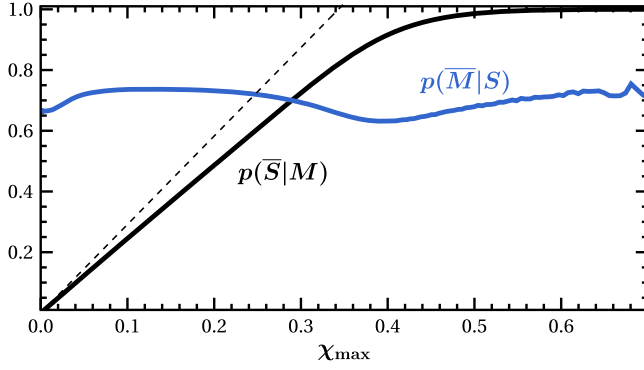


FIG. 10. Probability of an event being in only one of the two gaps. The blue (black) curve shows the probability that a binary lies in the spin (mass) gap but not in the mass (spin) gap. The dashed black line corresponds to the approximation  $\chi_{\max}/0.34$  (see text).

at  $\alpha = -1.6$ . Notably, this probability is almost independent of  $\chi_{\max}$ . On the other hand, the probability of observing a mass-gap event outside the spin gap grows from 0 to 1 as  $\chi_{\max}$  increases.

More specifically, we find a linear behavior  $p(\bar{S}|M) \propto \chi_{\max}$  for  $\chi_{\max} \rightarrow 0$ . As we show in Appendix, the  $\chi_{\text{eff}}$  distribution of 2g + 1g mergers (which form the bulk of 2g mergers) is roughly uniform for  $\chi_{\text{eff}} \in [-\chi_f/2, \chi_f/2]$ , where  $\chi_f \simeq 0.68$  is the most probable remnant spin (cf. Fig. 1). To a first approximation and for small  $\chi_{\max}$ , the probability of a mass-gap event not lying in the spin gap for 2g + 1g mergers can be approximated by  $\chi_{\max}/(\chi_f/2) \simeq \chi_{\max}/0.34$  (black dashed line in Fig. 10).

In this work we have ignored the relatively rare possibility of 3g mergers, which happens when the remnant of a 2g merger is also retained in the cluster. Since 2g BHs have large spins  $\sim 0.7$ , they receive large merger recoils. Only clusters with very high escape speed can successfully retain a meaningful fraction of 3g mergers [31]. For example, at  $\chi_{\max} \approx 0$ , when 2g events account for  $>10\%$  of all cluster events, only  $\sim \mathcal{O}(0.01\%)$  events are 3g.

#### IV. INFERENCE WITH MASS AND SPIN GAPS

The effective spin has long been proposed as a tool to infer the fraction  $f$  of BHs formed in clusters. The orientations of BHs formed in clusters should be isotropically distributed, leading to a symmetric  $\chi_{\text{eff}}$  distribution centered at  $\chi_{\text{eff}} = 0$ , while field binaries should be preferentially aligned, leading to a distribution skewed towards positive values of  $\chi_{\text{eff}}$  [18–20,86,87]. However, this argument fails if 1g BHs are all born with small spins because in that case the effective spin  $\chi_{\text{eff}} \sim 0$  irrespective of the spin orientations. Indeed, as we show below, if we only focus on 1g + 1g mergers, the error on the mixing fraction  $f$  scales as  $\delta f \propto 1/\chi_{\max}$ .

In this paper we argue that the mass and the spin gap can provide a powerful alternative which, crucially, remains viable also for small BH spins at birth.

#### A. Measuring the mixing fraction and $\chi_{\max}$ with 1g mergers

As we argued earlier, the efficiencies  $\lambda_M$  and  $\lambda_S$  depend on  $\chi_{\max}$ . We first need to estimate the accuracy with which GW detectors can measure  $\chi_{\max}$  using only 1g BHs.

To simplify the notation, let us introduce a “normalized effective spin”

$$\hat{\chi} \equiv \frac{\chi_{\text{eff}}}{\chi_{\max}} = \frac{\chi_1 \cos \theta_1 + q\chi_2 \cos \theta_2}{\chi_{\max}(1+q)} \in [-1, 1]. \quad (45)$$

We distribute the spin angles  $\theta_1$  and  $\theta_2$  uniformly in the cosine between 0 and  $\theta_{\max}$ . For the cluster model, we set  $\theta_{\max} = \pi$ , such that the distribution is isotropic. For the field binaries, perfect alignment would imply  $\theta_{\max} = 0$ , but nonvanishing values are predicted in more realistic models (e.g., [17,20]). Hereafter,  $\theta_{\max}$  refers to the largest spin tilt of the field binaries.

If  $\hat{\chi}$  has PDF  $\hat{p}(\hat{\chi})$ , the PDF of  $\chi_{\text{eff}}$  can be recovered easily as

$$p(\chi_{\text{eff}}|\chi_{\max}) = \frac{\hat{p}(\hat{\chi})}{\chi_{\max}}. \quad (46)$$

The derivation of analytical approximations for  $\hat{p}(\hat{\chi})$  for cluster and field binaries is presented in Appendix. The main result consists of the PDFs for the cluster model  $\hat{p}_{\text{cluster}}(\hat{\chi})$  [Eq. (A14)] and that of the field formation channel  $\hat{p}_{\text{field}}(\hat{\chi})$  [Eq. (A9)], shown here in Fig. 11.

If we denote by  $f$  the mixing fraction between the two channels, the total PDF is given by

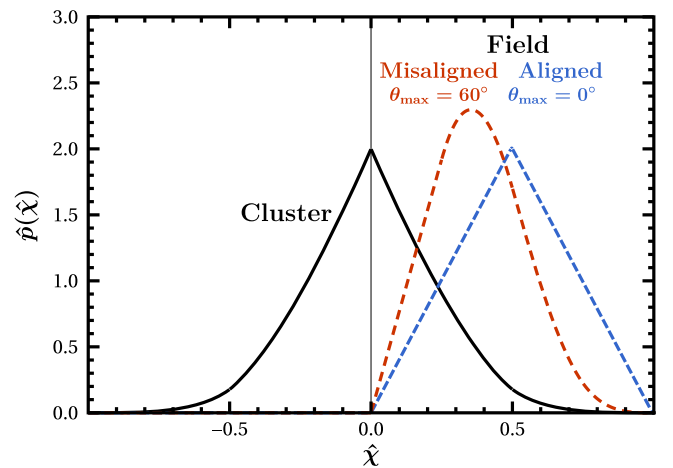


FIG. 11. PDF  $\hat{p}_{\text{cluster}}(\hat{\chi})$  for cluster binaries from Eq. (A14) (solid black line) and field binaries  $\hat{p}_{\text{field}}(\hat{\chi})$  from Eq. (A9) (dashed lines). For the field binaries, we assume either  $\theta_{\max} = 0^\circ$  (blue) or  $\theta_{\max} = 60^\circ$  (red).



$$\hat{p}(\hat{\chi}) = f\hat{p}_{\text{cluster}}(\hat{\chi}) + (1-f)\hat{p}_{\text{field}}(\hat{\chi}). \quad (47)$$

Suppose we have  $N_{\text{lg}}$  effective spin measurements  $\{\chi_{\text{eff}}^i\}$ . The log-likelihood of this sample can be written as

$$\mathcal{L} \equiv \mathcal{L}(f, \chi_{\text{max}} | \{\chi_{\text{eff}}^i\}) = \sum_i^{N_{\text{lg}}} \log \frac{\hat{p}(\hat{\chi})}{\chi_{\text{max}}}. \quad (48)$$

The quantities  $f$  and  $\chi_{\text{max}}$  can be estimated from the observed  $\{\chi_{\text{eff}}^i\}$  by maximizing the likelihood. The variance of the estimator quantifies the associated uncertainties.

### 1. Mixing fraction errors

LIGO measures  $\chi_{\text{eff}}$  better than any other spin parameter but still with some errors. The uncertainties  $\delta\chi_{\text{eff}}^i$  in the  $i$ th event affect the estimate of  $f$  via

$$(\delta f)^2 = \sum_i^{N_{\text{lg}}} \left( \frac{\partial f}{\partial \chi_{\text{eff}}^i} \delta\chi_{\text{eff}}^i \right)^2, \quad (49)$$

where

$$\frac{\partial f}{\partial \chi_{\text{eff}}^i} = -\frac{\partial^2 \mathcal{L} / \partial \chi_{\text{eff}}^i \partial f}{\partial^2 \mathcal{L} / \partial f^2} = -\frac{\mathcal{L}_{f\chi_{\text{eff}}^i}}{\mathcal{L}_{ff}} \quad (50)$$

and

$$\begin{aligned} \mathcal{L}_{f\chi_{\text{eff}}^i} &= \sum_i \frac{\hat{\chi}_i [\hat{p}_{\text{cluster}}(\hat{\chi}_i) \hat{p}'_{\text{field}}(\hat{\chi}_i) - \hat{p}_{\text{field}}(\hat{\chi}_i) \hat{p}'_{\text{cluster}}(\hat{\chi}_i)]}{\chi_{\text{max}} \hat{p}(\hat{\chi}_i)^2}, \\ \mathcal{L}_{ff} &= -\sum_i \left[ \frac{\hat{p}_{\text{cluster}}(\hat{\chi}_i) - \hat{p}_{\text{field}}(\hat{\chi}_i)}{\hat{p}(\hat{\chi}_i)} \right]^2. \end{aligned} \quad (51)$$

Instead of evaluating these quantities as Monte Carlo sums, we consider the integrated expectation value

$$\langle g(\hat{\chi}) \rangle \equiv \frac{1}{N} \sum_i g(\hat{\chi}_i) \rightarrow \int_{-1}^1 g(\hat{\chi}) \hat{p}(\hat{\chi}) d\hat{\chi}. \quad (52)$$

We also express  $\delta\chi_{\text{eff}}^i$  as

$$\delta\chi_{\text{eff}}^i = \frac{\sigma_{\text{LIGO}}}{\rho_i}, \quad (53)$$

where the signal-to-noise ratio (SNR)  $\rho_i$  is drawn from the distribution  $p(\rho) \propto \rho^{-4}$  [88,89] in the range  $[8, \infty)$ , and  $\sigma_{\text{LIGO}} = \langle \rho \delta\chi_{\text{eff}} \rangle \simeq 1.4$  is the median error  $\delta\chi_{\text{eff}}$  from LIGO/Virgo observations scaled by the SNR [21]. Ignoring the relatively weak SNR dependence on the effective spins [20,87,90] and marginalizing over  $\delta\chi_{\text{eff}}$ , a Monte Carlo sum like that in Eq. (49) can be approximated as

$$\begin{aligned} \frac{1}{N} \sum_i^N (g(\hat{\chi}_i) \delta\chi_{\text{eff}}^i)^2 &= \langle g(\hat{\chi})^2 \rangle \int_{\rho_{\text{thr}}}^{\infty} \frac{\langle \rho \delta\chi_{\text{eff}} \rangle^2}{\rho^2} p(\rho) d\rho \\ &= \langle g(\hat{\chi})^2 \rangle (\overline{\delta\chi_{\text{eff}}})^2, \end{aligned} \quad (54)$$

where

$$\overline{\delta\chi_{\text{eff}}} = \sqrt{\frac{3}{5}} \frac{\sigma_{\text{LIGO}}}{\rho_{\text{thr}}} \simeq 0.136. \quad (55)$$

By combining Eqs. (49)–(54) one gets

$$\delta f = \frac{\mathcal{F}_f(f, \theta_{\text{max}})}{\chi_{\text{max}} \sqrt{N_{\text{lg}}}}, \quad (56)$$

where

$$\mathcal{F}_f(f, \theta_{\text{max}})^2 = (\overline{\delta\chi_{\text{eff}}})^2 \frac{\left\langle \left( \hat{\chi} \frac{\hat{p}_{\text{cluster}} \hat{p}'_{\text{field}} - \hat{p}_{\text{field}} \hat{p}'_{\text{cluster}}}{\hat{p}^2} \right)^2 \right\rangle}{\left\langle \left( \frac{\hat{p}_{\text{cluster}} - \hat{p}_{\text{field}}}{\hat{p}} \right)^2 \right\rangle^2} \quad (57)$$

and we omitted the arguments of  $\hat{p}(\hat{\chi})$ ,  $\hat{p}_{\text{cluster}}(\hat{\chi})$ , and  $\hat{p}_{\text{field}}(\hat{\chi})$ , for clarity. The function  $\mathcal{F}_f(f, \theta_{\text{max}})$  varies only mildly with  $f$  and increases slowly with the largest misalignment angle of field binaries  $\theta_{\text{max}}$ . For  $\theta_{\text{max}} < 60^\circ$ , one has  $\mathcal{F}_f(f, \theta_{\text{max}}) \sim \mathcal{O}(0.1)$ . Equation (56) returns  $\delta f \propto \chi_{\text{max}}^{-1}$ , as expected (see the solid black lines in Fig. 12): one cannot rely on the spin orientations to measure the mixing fraction if BH spins are too low. If  $\chi_{\text{max}} \sim 0.01$  [27], one would need  $\gtrsim \mathcal{O}(10^5)$  1g detections to achieve errors on the mixing fraction  $\delta f \sim 0.1$ . For larger values of  $\chi_{\text{max}}$ , we can use 1g events to measure  $f$  quite accurately. For example, if  $\chi_{\text{max}} = 0.5$  we can achieve an error  $\delta f \lesssim 0.1$  with only  $\sim 100$  events.

### 2. Errors on $\chi_{\text{max}}$

We estimate  $\delta\chi_{\text{max}}$  due to errors on the individual  $\chi_{\text{eff}}^i$  measurements by error propagation:

$$(\delta\chi_{\text{max}})^2 = \sum_i^{N_{\text{lg}}} \left( \frac{\partial \chi_{\text{max}}}{\partial \chi_{\text{eff}}^i} \delta\chi_{\text{eff}}^i \right)^2, \quad (58)$$

where using the same notation as above we get

$$\frac{\partial \chi_{\text{max}}}{\partial \chi_{\text{eff}}^i} = -\frac{\partial^2 \mathcal{L} / \partial \chi_{\text{eff}}^i \partial \chi_{\text{max}}}{\partial^2 \mathcal{L} / \partial \chi_{\text{max}}^2} = -\frac{\mathcal{L}_{\chi_{\text{max}} \chi_{\text{eff}}^i}}{\mathcal{L}_{\chi_{\text{max}} \chi_{\text{max}}}}, \quad (59)$$

with

$$\begin{aligned} \mathcal{L}_{\chi_{\text{max}} \chi_{\text{eff}}^i} &= -\frac{1}{\chi_{\text{max}}^2} \sum_i [\hat{\chi}_i \log'' \hat{p}(\hat{\chi}_i) + \log' \hat{p}(\hat{\chi}_i)], \\ \mathcal{L}_{\chi_{\text{max}} \chi_{\text{max}}} &= \frac{1}{\chi_{\text{max}}^2} \sum_i [1 + 2\hat{\chi}_i \log' \hat{p}(\hat{\chi}_i) + \hat{\chi}_i^2 \log'' \hat{p}(\hat{\chi}_i)]. \end{aligned} \quad (60)$$

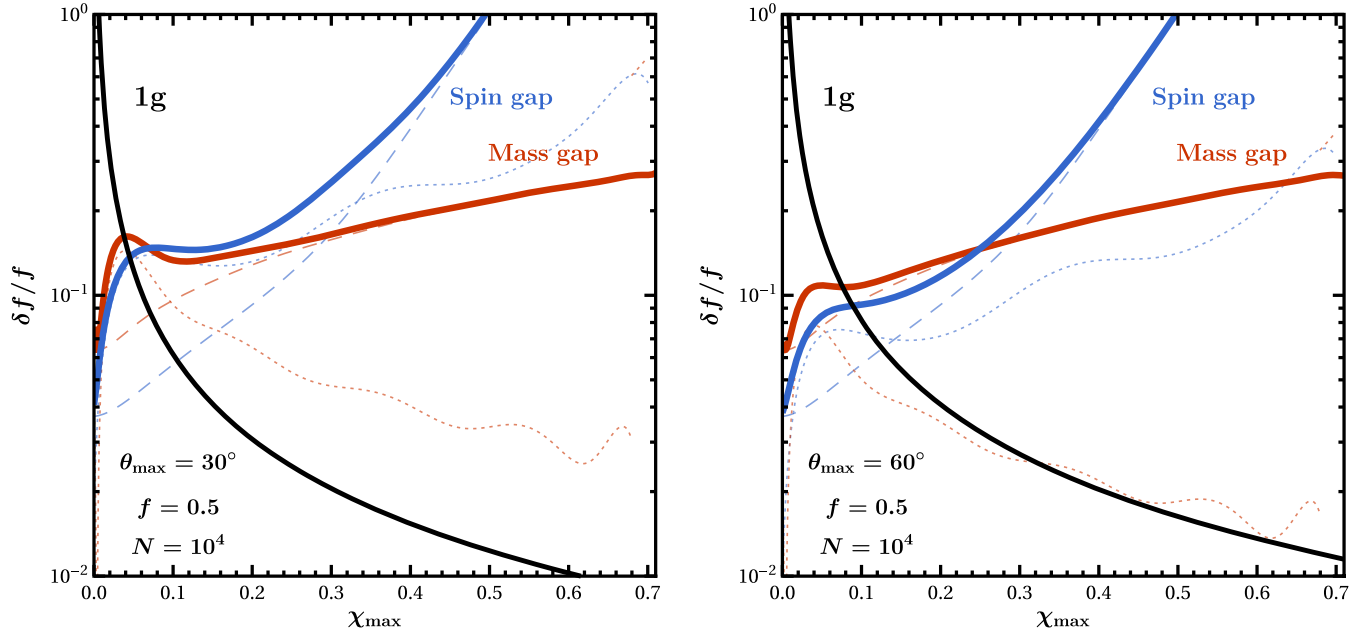


FIG. 12. Relative error  $\delta f/f$  on the fraction of dynamical mergers as a function of  $\chi_{\max}$  considering either only 1g mergers (black), the mass gap (red), or the spin gap (blue). We assume a catalog of  $N = 10^4$  observations, a mixing fraction  $f = 0.5$ , and the largest misalignment angles for field binaries  $\theta_{\max} = 30^\circ$  (left) and  $60^\circ$  (right). The contributions due to Poisson counting errors and efficiency uncertainties are marked with dashed and dotted lines, respectively.

Combining these results and replacing the sum by an integral as in Eq. (52), Eq. (58) yields

$$\delta\chi_{\max} = \frac{\mathcal{F}_{\chi_{\max}}(f, \theta_{\max})}{\sqrt{N_{1g}}}, \quad (61)$$

where

$$\mathcal{F}_{\chi_{\max}}(f, \theta_{\max})^2 = (\overline{\delta\chi_{\text{eff}}})^2 \frac{\langle (\hat{\chi} \log'' \hat{p} + \log' \hat{p})^2 \rangle}{\langle 1 + 2\hat{\chi} \log' \hat{p} + \hat{\chi}^2 \log'' \hat{p} \rangle^2}. \quad (62)$$

Here, again, we have suppressed the argument of  $\hat{p}(\hat{\chi})$ .

The function  $\mathcal{F}_{\chi_{\max}}(f, \theta_{\max})$  presents a mild dependence on  $f$  for nonzero misalignment angles, but it diverges as  $\theta_{\max} \rightarrow 0$ . Mathematically, the reason for the divergence is that, in the limit  $\theta_{\max} \rightarrow 0$ ,  $\log'' \hat{p}(\hat{\chi})$  diverges as  $\hat{\chi} \rightarrow 1$  in our simple analytical model. The divergence may not occur in a more accurate Bayesian inference analysis. In the limit  $f \rightarrow 1$ , when all events come from clusters [Eq. (47)] and  $\theta_{\max}$  is irrelevant, we have  $\mathcal{F}_{\chi_{\max}}(f, \theta_{\max}) \simeq 1$ .

Our estimate considers the errors on  $\chi_{\max}$  and  $f$  due only to measurement errors on the  $\chi_{\text{eff}}^i$ . Another source of error is the variance of the maximum-likelihood estimator itself (given by the inverse of the Fisher information matrix), which would be present even if all of the  $\chi_{\text{eff}}^i$  were measured perfectly. However, these errors are always subdominant with respect to the measurement errors estimated above because  $\overline{\delta\chi_{\text{eff}}}$  [Eq. (55)] is relatively large.

We have also assumed that the maximum likelihood estimator  $\chi_{\max}^{\text{mle}}$  coincides with the true value  $\chi_{\max}$ . This approximation must break down for a finite sample of size  $N$ , and then  $\hat{\chi}^{\text{mle}} = \chi_{\text{eff}}/\chi_{\max}^{\text{mle}}$  will not follow the distribution  $\hat{p}(\hat{\chi})$ . A more careful error analysis is an interesting topic for future research.

## B. Measuring the mixing fraction with the gaps

We finally address the crucial point of this paper: can the mass and spin gaps be used to constrain the mixing fraction between different formation channels?

Suppose we are given a catalog of  $N$  observations, which include 1g and 2g events. As discussed in the Introduction, the number of identifiable 2g mergers because of the mass and spin gaps is

$$\begin{aligned} \frac{N_{\text{Mgap}}}{N} &= f\lambda_{\text{M}}(\chi_{\max}), \\ \frac{N_{\text{Sgap}}}{N} &= f\lambda_{\text{S}}(\chi_{\max}), \end{aligned} \quad (63)$$

where  $f$  is the fraction of all detections that were produced in clusters. The number of 1g events available to infer  $\chi_{\max}$ , as described in Sec. IV A, is

$$N_{1g} = N[1 - f\lambda_{\text{S}}(\chi_{\max})]. \quad (64)$$

For either the mass gap ( $\lambda_{\text{gap}} = \lambda_{\text{M}}$ ) or the spin gap ( $\lambda_{\text{gap}} = \lambda_{\text{S}}$ ), the uncertainty in measuring  $f$  is given by

$$\left(\frac{\delta f}{f}\right)_{\text{gap}}^2 = \left(\frac{\partial \lambda_{\text{gap}} / \partial \chi_{\text{max}}}{\lambda_{\text{gap}}} \delta \chi_{\text{max}}\right)^2 + \left(\frac{\delta N_{\text{gap}}}{N_{\text{gap}}}\right)^2. \quad (65)$$

The first term on the right-hand side represents the uncertainty in measuring the efficiency propagated from  $\chi_{\text{max}}$ . As we argued in Sec. III C, the dependencies on the other population parameters (such as  $\alpha$  and  $m_{\text{max}}$ ) are mild and can be neglected. The second term in Eq. (65) is the Poisson counting error associated with the number of mass- or spin-gap events,

$$\frac{\delta N_{\text{gap}}}{N_{\text{gap}}} = \frac{1}{\sqrt{\lambda_{\text{gap}} f N}}. \quad (66)$$

In Fig. 12 we plot the relative errors in the mixing fraction  $\delta f/f$  for both the mass gap (red) and the spin gap (blue) as a function of  $\chi_{\text{max}}$ . Figure 12 also shows the individual contributions due to efficiency errors and Poisson errors. For illustration we consider  $N = 10^4$  observations with a mixing fraction of  $f = 0.5$ , and we select two values of the largest misalignment angle for field binaries:  $\theta_{\text{max}} = 30^\circ, 60^\circ$ .

Spin-gap events yield a more accurate measurement of  $f$  at small values of  $\chi_{\text{max}}$ . This is because (i) the number of spin-gap events is higher (i.e.,  $\lambda_S$  is large) and thus Poisson errors are low, and (ii) the derivative  $\partial \lambda_S / \partial \chi_{\text{max}}$  vanishes for  $\chi_{\text{max}} \rightarrow 0$ , leading to small efficiency errors. The mass gap is more informative for larger values of  $\chi_{\text{max}}$  because  $\lambda_S < \lambda_M$  for  $\chi_{\text{max}} \gtrsim 0.3$ , while  $(\partial \lambda_S / \partial \chi_{\text{max}}) / \lambda_S > (\partial \lambda_M / \partial \chi_{\text{max}}) / \lambda_M$  for  $\chi_{\text{max}} \gtrsim 0.05$ .

The “critical” value of  $\chi_{\text{max}}$  at which the mass gap is preferred over the spin gap depends on both  $f$  and  $\theta_{\text{max}}$ . In particular, the threshold is  $\chi_{\text{max}} \simeq 0.07(0.25)$  for  $f = 0.5$  and  $\theta_{\text{max}} = 30^\circ(60^\circ)$ . This is one of the most important findings of this paper: *If BHs are born slowly rotating, spin-gap events are more effective than mass-gap events at pinning down the mixing fraction between formation channels.*

In general, we find that counting errors dominate for large values of  $\chi_{\text{max}}$ , simply because there are not enough gap events: cf. Eq. (66). From Fig. 12, at  $f = 0.5$  efficiency errors are important only when  $\chi_{\text{max}} < 0.1(0.02)$  for mass-gap events and when  $\chi_{\text{max}} < 0.33(0.14)$  for spin-gap events at  $\theta_{\text{max}} = 30^\circ(60^\circ)$ . The parameter  $\theta_{\text{max}}$  enters  $\delta f/f$  in Eq. (65) only through the efficiency error [cf. Eq. (58)]. As a result, only the small- $\chi_{\text{max}}$  behavior of the  $\delta f/f$  errors is affected by  $\theta_{\text{max}}$ : for example, compare the left and right panels of Fig. 12, where  $\delta f/f$  is the same for  $\theta_{\text{max}} = 30^\circ$  and  $\theta_{\text{max}} = 60^\circ$  at large  $\chi_{\text{max}}$ .

Besides  $\chi_{\text{max}}$  and  $f$ , the two parameters determining the mass spectrum ( $\alpha$  and  $m_{\text{max}}$ ) mildly affect the number of events and thus the Poisson errors. For instance,  $\lambda_M$  increases for  $|\alpha| < 1.6$ , which implies that Poisson errors become less relevant. As expected, we find that spin-gap estimates are more robust against changes in  $m_{\text{max}}$  and  $\alpha$  compared to mass-gap estimates. This is simply because the mass spectrum parameters have a direct impact on the 1g mass distribution.

Figure 13 shows contours of  $\delta f/f \sqrt{N}$  obtained from the mass gap (left panel) and the spin gap (right panel) in the

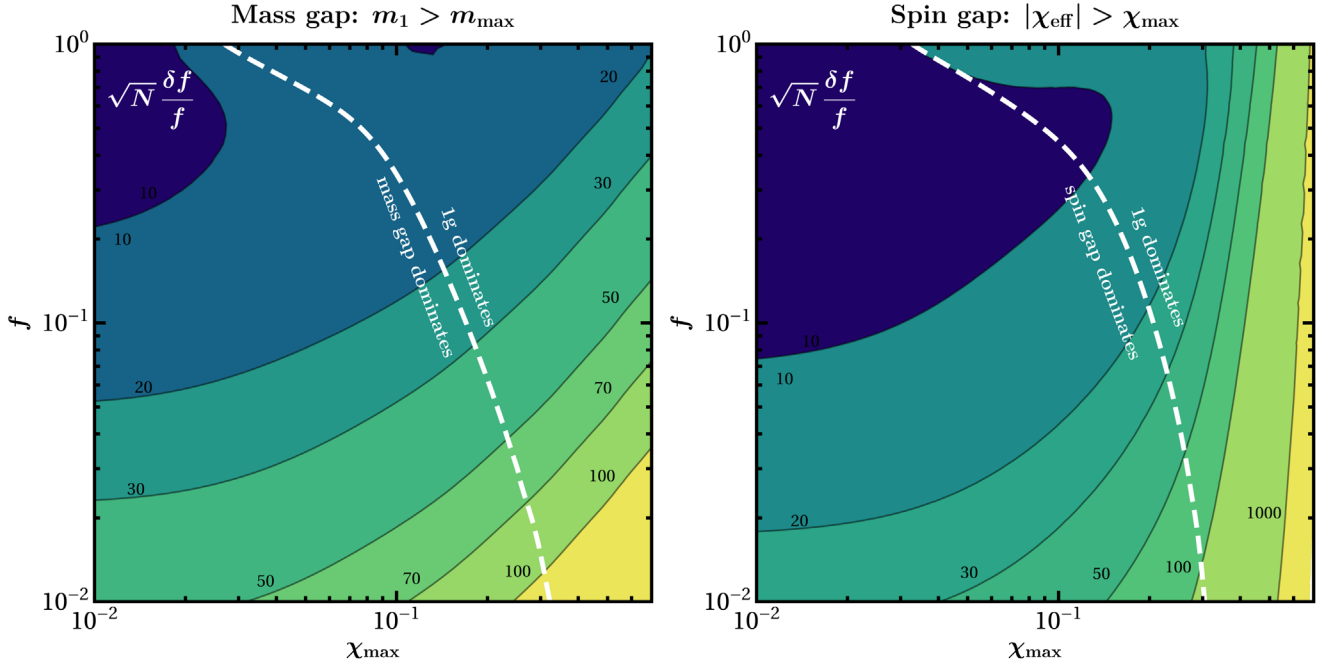


FIG. 13. Fractional error on the mixing fraction  $\sqrt{N} \delta f/f$  obtained using the mass gap (left) and spin gap (right) as a function of  $\chi_{\text{max}}$  and  $f$  assuming  $\theta_{\text{max}} = 60^\circ$ . The white dashed line marks the location where the error  $\delta f$  from 1g events equals the one obtained with the gaps. In particular, gap (1g) events dominate to the left (right) of the white dashed lines.

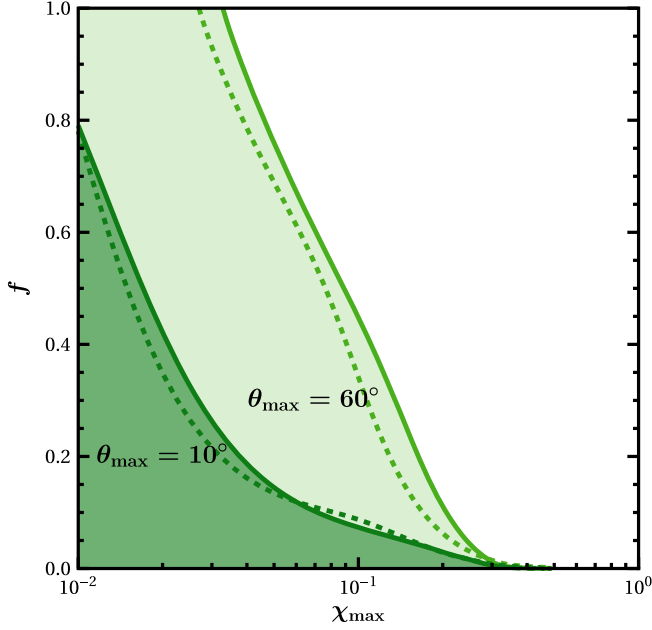


FIG. 14. The shaded areas mark values of  $f$  and  $\chi_{\max}$  where gaps provide a more accurate measurement of the mixing fraction  $f$  compared to 1g events. Results for the spin (mass) gap are indicated with solid (dashed) curves. Darker (lighter) regions show results for  $\theta_{\max} = 10^\circ$  ( $60^\circ$ ).

$(\chi_{\max}, f)$  plane, assuming, for concreteness,  $\theta_{\max} = 60^\circ$ . We also plot contours (dashed white lines) where either of the gap measurements and 1g detections achieve the same  $\delta f$  accuracy: gaps are better than 1g observations at constraining  $f$  to the left of these lines and worse to the right of these lines. As a rough rule of thumb, the mixing fraction is better constrained through gap measurements when  $\chi_{\max} \lesssim 0.1$ , while the 1g population is more constraining if  $\chi_{\max} \gtrsim 0.1$ . This is another central result of our work.

Suppose for concreteness that  $\chi_{\max} = 10^{-2}$ , as predicted by Ref. [27]. Then spin-gap events would allow us to measure the mixing fraction  $f$  with an accuracy of  $\delta f = 0.1$  with a catalog of  $N \sim 150$  events if  $f = 0.2$  ( $N \sim 3000$  events if  $f = 1$ ), corresponding to 5 (400) events in the gap. Achieving the same  $\delta f = 0.1$  accuracy with mass-gap events would require  $N \sim 450$  events if  $f = 0.2$  ( $N \sim 4400$  events if  $f = 1$ ), corresponding to 4 (200) events with  $m_1 > m_{\max}$ . This should be compared with the  $\gtrsim 10^5$  events needed for measuring  $f$  at  $\chi_{\max} = 10^{-2}$  using 1g events only.

To make the previous rule of thumb more precise, in Fig. 14 we show regions in the  $(\chi_{\max}, f)$  plane where gap events would lead to smaller errors compared to 1g events. This plot confirms that looking at events in the gaps is crucial if BHs are born slowly rotating (i.e., at small  $\chi_{\max}$ ) and cluster formation is inefficient (i.e., at small values of  $f$ ). The region where the gaps are important increases with  $\theta_{\max}$  because (as we discussed in Sec. IV A)  $\delta\chi_{\max}$  diverges for small values of  $\chi_{\max}$ .

We summarize our main points as follows:

- (i) Spin-gap events measure  $f$  more accurately than mass-gap events at small  $\chi_{\max}$ .
- (ii) The error  $\delta f$  at  $\chi_{\max} \gtrsim 0$  is dominated by the efficiency error; this is governed by the maximum misalignment angle  $\theta_{\max}$  in field formation scenarios, and it is only mildly affected by the mass spectrum parameters  $\alpha$  and  $m_{\max}$ .
- (iii) Poisson errors dominate for large  $\chi_{\max}$ . For spin-gap events the Poisson errors depend only on  $\chi_{\max}$ , while for mass-gap events they are also governed by the mass spectrum parameters  $\alpha$  and  $m_{\max}$ .
- (iv) Gap events measure  $f$  better than 1g events at small  $\chi_{\max}$ , small  $f$ , and large  $\theta_{\max}$ .

## V. CONCLUSIONS

Each event in the growing LIGO/Virgo BBH catalog yields three main “intrinsic” observable quantities: the binary component masses and the effective spin. The observed events are likely to come from at least two, and possibly more, formation channels. As the number of observations grows, reserved regions in the intrinsic parameter space (the mass and spin gaps) could allow us to measure the relative contributions of different channels. In this paper we quantified this statement, showing that the mass and spin gaps allow us to efficiently separate the contribution of field binaries from the contribution of binaries formed dynamically in star clusters.

Supernova instabilities [32] and efficient core-envelope interactions [27] imply that massive stars should form BHs with an upper mass ( $m \lesssim 45 M_\odot$ ) and spin ( $\chi \lesssim 0.1$ ) limit. Therefore, our current understanding of field binary evolution hints at the existence of *both* a “mass gap” and a “spin gap” within this formation scenario. Repeated mergers in clusters provide a natural way to evade these constraints (see e.g., [30,31,91]). Assuming to a first approximation that only these two scenarios are at play, observations of BBHs in either the mass gap, the spin gap, or both would not only imply that those events were formed dynamically, but it would also improve our understanding of the origin of the whole observed population.

The key theoretical input to perform this analysis is the efficiency with which dynamical environments like GCs and NSCs can populate the gap(s). We used a semianalytical model specifically designed to predict the occurrence of repeated (2g) mergers in dense star clusters. The main prediction of our model is that the gap efficiencies are of the order of a few % and that they are mostly sensitive to  $\chi_{\max}$ , the largest spin magnitude of individual BHs at birth.

We propose the following broad observational strategy. We can assume that the bulk of the population consists of 1g BHs, which are outside the gaps and can be used to measure  $\chi_{\max}$  (Sec. IV). As shown in Sec. III, a measurement of  $\chi_{\max}$  can be converted into a solid estimate of the cluster efficiencies at populating the mass gap ( $\lambda_M$ ) and the



spin gap ( $\lambda_S$ ). Combined with the measured distribution of the effective spins  $\{\chi_{\text{eff}}^i\}$ , this results in an estimate of the mixing fraction  $f$  between cluster and field binaries (Sec. IV B).

Many studies in the literature (see e.g., [18–20,92–94]) devised strategies to infer the mixing fraction  $f$  between field and cluster formation channels from the observed distribution of effective spins  $\{\chi_{\text{eff}}^i\}$ . The underlying idea behind most of these studies is that cluster formation predicts a distribution of  $\chi_{\text{eff}}$  which is symmetric about zero, while binaries in the field should show a preference for  $\chi_{\text{eff}} > 0$ . All strategies that rely on a measurement of the spin orientations, however, are bound to fail if the spin magnitudes are small. Indeed, in Sec. IV we show that, within this approach, the accuracy in determining the mixing fraction scales like  $\delta f \propto 1/\chi_{\text{max}}$ .

On the other hand, by exploiting the gaps we can estimate the mixing fraction  $f$  between different channels even if BH spins at birth are zero. If anything, the spin gap is larger if the natal spins are small, and outliers can be more easily identified. Indeed, we find that both the mass and spin gaps allow for a better measurement of the mixing fraction (compared to the standard “ $\chi_{\text{eff}}$  distribution test”) as long as  $\chi_{\text{max}} \lesssim 0.1$ . Both observations [16,21] and theoretical modeling [27,39] suggest that this is indeed the case, making our new observational strategy timely and relevant.

We assumed that the mass and spin gaps can only be populated by repeated mergers in clusters. This is an important caveat of our study. While 2g mergers constitute a well-motivated scenario (see e.g., [30,31,91]) which is now being implemented in LIGO/Virgo parameter estimation pipelines [95], other astrophysical mechanisms could “pollute” the gaps and deteriorate the measurement of  $f$ . These include gas accretion [41], stellar mergers [36], population III stars [42–44], or gravitational lensing<sup>2</sup> of 1g events [45,46].

An independent reanalysis of data from the first and second LIGO/Virgo observing runs identified at least one BBH event (GW170817A, not to be confused with the famous binary neutron star merger that occurred on the same day) which may have *both* high mass and high spin

[25]. Based on these properties of the binary, some authors [99] suggested that GW170817A might have formed in an AGN disk [6,7]. However, as pointed out in the context of the first candidate mass-gap event GW170729 [100], it is dangerous to evaluate individual outlier events without reference to the entire population. We postpone a more complete study, including an outlier analysis along the lines of Ref. [100], to future work.

Two predictions of our model are particularly noteworthy because they could be verified or disproved in the near future:

- (1) Future events with large mass but small effective spin ( $M\bar{S}$  in our notation) can be explained only if  $\chi_{\text{max}}$  is sufficiently large. In other words, we find that *a mass-gap event should also be in the spin gap*: cf. Fig. 8, and note that Ref. [39] recently proposed a similar argument. This is an important feature of our model, which can potentially allow us to disentangle the hierarchical merger contribution to the mass gap considered in this work from other astrophysical mechanisms.
- (2) If BHs are born slowly rotating, high-spin events are more effective than high-mass events to pin down the mixing fraction between formation channels: the spin gap (which was largely neglected in the literature so far) is actually more discriminating than the mass gap if spins are small, as suggested by astrophysical theory and LIGO/Virgo observations so far (Fig. 14).

In this exploratory study we have made simplifying assumptions that should be relaxed in the future. First of all, to keep the analysis general, we focused on the fraction of the total number of observations that end up in the mass or spin gaps. In practice, this fraction will be detector dependent: third-generation detectors such as the Einstein Telescope or Cosmic Explorer will be more sensitive to low-mass binaries, while current detectors introduce a selection bias that favors large masses.

Second, our ability to distinguish between different formation scenarios could improve if we considered not only the number of events in the gap but also their *distribution*. In our strategy we proposed estimating  $\chi_{\text{max}}$  from the 1g spin distribution and to simply count events with  $|\chi_{\text{eff}}| > \chi_{\text{max}}$ . However, 2g + 2g mergers can result in effective spins  $\chi_{\text{eff}} \gtrsim 0.34$ , while this is not possible for 2g + 1g events (cf. Fig. 1). Third-generation detectors could allow us to infer the spin distribution of spin-gap events and possibly to measure the relative number of 2g + 2g and 2g + 1g events. Similarly, the mass distribution of mass-gap events contains useful information. For large  $\chi_{\text{max}}$ , only very large NSCs could retain postmerger remnants, leading to a steeper mass distribution (cf. Sec. II C). This dependence is weak, but mass-gap events should typically have large SNRs; therefore, their mass distribution is easier to measure.

The upcoming release of LIGO/Virgo data from the third observing run O3 will bring us closer to the large-statistics

<sup>2</sup>A gravitational lens with magnification  $\mu$  increases the amplitude of a GW event by a factor of  $\sqrt{\mu}$  [45,46]. This magnification reduces the inferred luminosity distance, increasing the apparent source-frame mass and producing “fake” mass-gap events. However, the lensing probability is rather small ( $\sim 10^{-3}$  according to Ref. [45]). The distribution of magnification factors depends on the lens model, but it roughly scales like  $p(\mu) \sim \mu^{-3}$  for  $\mu > 2$  [96–98]. In order to contribute significantly to the mass gap, lensed events must be located at high redshifts, have large magnification, and have source-frame masses close enough to the mass gap. Taking into account the lensing probability, the shape of the mass function, and the magnification function, we estimate that the probability of observing a mass-gap event due to lensing is  $\sim 10^{-5}$ , so it can safely be neglected in the present context.

regime of GW astronomy. As we enter this new era, the observational strategy outlined in this paper could lead us to a better estimate of the relative contribution of different formation channels.

### ACKNOWLEDGMENTS

We thank Alberto Vecchio, Cyril Creque-Sarbinowski, Giacomo Fragione, and Mohammadtaher Safarzadeh for discussions. V. B., E. B., T. H., and K. W. K. W. are supported by NSF Grants No. PHY-1912550 and No. AST-1841358, NASA ATP Grants No. 17-ATP17-0225 and No. 19-ATP19-0051, and NSF-XSEDE Grant No. PHY-090003. E. B. also acknowledges support from the Amaldi Research Center funded by the MIUR program “Dipartimento di Eccellenza” (CUP: B8I18001170001). D. G. is supported by Leverhulme Trust Grant No. RPG-2019-350. The authors would like to acknowledge networking support by the GWverse COST Action CA16104, “Black holes, gravitational waves and fundamental physics.” Computational work was performed on the University of Birmingham BlueBEAR cluster, the Athena cluster at HPC Midlands+ funded by EPSRC Grant No. EP/P020232/1, and the Maryland Advanced Research Computing Center (MARCC).

### APPENDIX: ANALYTICAL APPROXIMATIONS OF $\chi_{\text{eff}}$ PROBABILITY DISTRIBUTIONS

The goal of Sec. IV is to compute errors on the maximum effective spin  $\chi_{\text{max}}$  and on the mixing fraction  $f$  by error propagation, which requires the evaluation of first and second derivatives of the PDFs. In principle, this could be done by sampling the distributions and numerically interpolating the results, which, however, would result in large errors on the derivatives. To overcome this problem, in this Appendix we find analytical expressions for the PDF of  $\hat{\chi}$  in the two scenarios of interest: cluster and field formation.

Our starting point is the rescaled effective spin  $\hat{\chi}$  of Eq. (45). Since in Eq. (31) we set  $\beta = 6.7 \gg 1$ , we can assume  $q = 1$  for the vast majority of our sources, so that

$$\hat{\chi} \simeq \frac{\hat{\chi}_1 \cos \theta_1 + \hat{\chi}_2 \cos \theta_2}{2}, \quad (\text{A1})$$

where  $\hat{\chi}_i = \chi_i / \chi_{\text{max}}$ . We have verified that setting  $q = 1$  leads to negligible deviations with respect to the PDFs found by using generic values of  $q$ .

We will repeatedly use the following standard identities from probability theory. If  $X$  and  $Y$  are two independent, continuous random variables with PDFs  $f_X$  and  $f_Y$ , the PDF of their product  $XY$  and of their sum  $X + Y$  are

$$f_{XY}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z/x) \frac{1}{|x|} dx, \quad (\text{A2})$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx. \quad (\text{A3})$$

The PDF of a generic bijective function  $g(X)$  is

$$f_{g(X)}(z) = f_X(g^{-1}(z)) \left| \frac{dg^{-1}(x)}{dx} \right|_{x=z}. \quad (\text{A4})$$

#### 1. Field binaries

For field binaries the spins will be nearly aligned, so we draw  $\cos \theta_i$  ( $i = 1, 2$ ) uniformly in the range  $[1 - \delta, 1]$ , where  $\delta$  is related to the maximum misalignment angle  $\theta_{\text{max}}$  of each spin by  $\theta_{\text{max}} = \arccos(1 - \delta)$ .

First we find the distribution of

$$z_i = \hat{\chi}_i \cos \theta_i, \quad (\text{A5})$$

which is a product of the two uniform distributions

$$\begin{aligned} p(\cos \theta_i) &= 1/2\delta, \quad \cos \theta_i \in [1 - \delta, 1], \\ p(\hat{\chi}_i) &= 1, \quad \hat{\chi}_i \in [0, 1]. \end{aligned} \quad (\text{A6})$$

From Eq. (A2) one gets

$$p(z_i) = \begin{cases} -\log(1 - \delta)/\delta & \text{for } 0 < z_i \leq 1 - \delta, \\ -\log(z_i)/\delta & \text{for } 1 - \delta \leq z_i \leq 1. \end{cases} \quad (\text{A7})$$

For  $q = 1$ , the distribution of

$$\hat{\chi} = \frac{z_1 + z_2}{2} \quad (\text{A8})$$

follows directly from Eqs. (A3) and (A4).

$$\hat{p}_{\text{field}}(\hat{\chi}) = \frac{4}{\delta^2} \times \begin{cases} \xi^2 \hat{\chi} & \text{for } 0 \leq \hat{\chi} \leq (1 - \delta)/2, \\ -\xi(\delta + (\xi + 2)\hat{\chi} - 2\hat{\chi} \log(2\hat{\chi}) - 1) & \text{for } (1 - \delta)/2 \leq \hat{\chi} \leq 1/2, \\ -\xi(\delta + \xi\hat{\chi}) & \text{for } 1/2 \leq \hat{\chi} \leq 1 - \delta, \\ \hat{\chi} \left[ \text{Li}_2\left(\frac{1-\delta}{2\hat{\chi}}\right) - \text{Li}_2\left(\frac{2\hat{\chi}+\delta-1}{2\hat{\chi}}\right) \right] - \xi[\hat{\chi} \log(2(\delta + 2\hat{\chi} - 1)) + \hat{\chi} \log \hat{\chi} & \\ - 2\hat{\chi} + 1] + 2(\delta + \hat{\chi} - 1) + (-\delta - 2\hat{\chi} + \hat{\chi} \log(2\hat{\chi}) + 1) \log(\delta + 2\hat{\chi} - 1) & \text{for } 1 - \delta \leq \hat{\chi} \leq 1 - \delta/2, \\ \hat{\chi} \text{Li}_2\left(1 - \frac{1}{2\hat{\chi}}\right) - \hat{\chi} \text{Li}_2\left(\frac{1}{2\hat{\chi}}\right) - 2\hat{\chi} - (-2\hat{\chi} + \hat{\chi} \log(2\hat{\chi}) + 1) \log(2\hat{\chi} - 1) + 2 & \text{for } 1 - \delta/2 \leq \hat{\chi} \leq 1, \end{cases} \quad (\text{A9})$$

$$\hat{p}_{\text{field}}(\hat{\chi}) = \frac{4}{\delta^2} \times \begin{cases} \xi^2 \hat{\chi} & \text{for } 0 \leq \hat{\chi} \leq (1-\delta)/2, \\ -\xi(\delta + (\xi+2)\hat{\chi} - 2\hat{\chi} \log(2\hat{\chi}) - 1) & \text{for } (1-\delta)/2 \leq \hat{\chi} \leq (1-\delta), \\ \hat{\chi} [\text{Li}_2(\frac{1-\delta}{2\hat{\chi}}) - \text{Li}_2(\frac{2\hat{\chi}+\delta-1}{2\hat{\chi}})] + 2(\delta + \hat{\chi} - 1) & \text{for } (1-\delta) \leq \hat{\chi} \leq 1/2, \\ -(\delta - \hat{\chi} \log(\frac{2\hat{\chi}}{1-\delta}) + 2\hat{\chi} - 1) \log(\delta + 2\hat{\chi} - 1) + \xi \hat{\chi} \log(2\hat{\chi}) & \text{for } 1/2 \leq \hat{\chi} \leq 1 - \delta/2, \\ \hat{\chi} [\text{Li}_2(\frac{1-\delta}{2\hat{\chi}}) - \text{Li}_2(\frac{2\hat{\chi}+\delta-1}{2\hat{\chi}})] + 2(\delta + \hat{\chi} - 1) & \text{for } 1 - \delta/2 \leq \hat{\chi} \leq 1, \\ -(\delta - \hat{\chi} \log(\frac{2\hat{\chi}}{1-\delta}) + 2\hat{\chi} - 1) \log(\delta + 2\hat{\chi} - 1) - \xi(\hat{\chi} \log(2\hat{\chi}) - 2\hat{\chi} + 1) & \text{for } 1/2 \leq \hat{\chi} \leq 1 - \delta/2, \\ \hat{\chi} \text{Li}_2(1 - \frac{1}{2\hat{\chi}}) - \hat{\chi} \text{Li}_2(\frac{1}{2\hat{\chi}}) - 2\hat{\chi} - (-2\hat{\chi} + \hat{\chi} \log(2\hat{\chi}) + 1) \log(2\hat{\chi} - 1) + 2 & \text{for } 1 - \delta/2 \leq \hat{\chi} \leq 1, \end{cases} \quad (\text{A10})$$

where Eqs. (A9) hold for  $\delta \leq 1/2$  ( $\theta_{\text{max}} = 60^\circ$ ), Eqs. (A10) hold for  $1/2 < \delta \leq 1$ , we defined  $\xi = \log(1-\delta)$ , and

$$\text{Li}_2(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^2} = \int_z^0 \frac{\ln(1-t)}{t} dt \quad (\text{A11})$$

is the dilogarithm.

## 2. Cluster binaries

### a. First generation

For cluster BBHs we assume an isotropic distribution; i.e., we draw  $\cos \theta_i$  from a uniform distribution in  $[-1, 1]$ .

First we find the distribution of  $z_i = \hat{\chi}_i \cos \theta_i$ , which is a product of two uniform distributions:

$$\begin{aligned} p(\cos \theta_i) &= 1/2, & \cos \theta_i &\in [-1, 1], \\ p(\hat{\chi}_i) &= 1, & \hat{\chi}_i &\in [0, 1]. \end{aligned} \quad (\text{A12})$$

Using Eq. (A2) one has

$$p(z_i) = -\frac{1}{2} \log |z_i|, \quad \text{for } z_i \in [-1, 1]. \quad (\text{A13})$$

Finally, the distribution of  $\hat{\chi}$  from Eq. (A8) can be calculated using Eqs. (A3) and (A4), with the result

$$\hat{p}_{\text{cluster}}(\hat{\chi}) = \begin{cases} -2|\hat{\chi}| \text{Li}_2(\frac{2|\hat{\chi}|}{2|\hat{\chi}|-1}) - \frac{1}{2}(4 + \pi^2)|\hat{\chi}| - \log(1-2|\hat{\chi}|)(-2|\hat{\chi}| + |\hat{\chi}| \log(1-2|\hat{\chi}|) + 1) + 2 & \text{for } 0 < |\hat{\chi}| \leq 1/2, \\ |\hat{\chi}| \text{Li}_2(1 - \frac{1}{2|\hat{\chi}|}) - |\hat{\chi}| \text{Li}_2(\frac{1}{2|\hat{\chi}|}) - 2\hat{\chi} - (-2|\hat{\chi}| + |\hat{\chi}| \log(2|\hat{\chi}|) + 1) \log(2|\hat{\chi}| - 1) + 2 & \text{for } 1/2 < |\hat{\chi}| \leq 1. \end{cases} \quad (\text{A14})$$

### b. 2g + 1g and 2g + 2g mergers

Here we provide some approximations to the PDFs of 2g mergers. These are not used explicitly in the main body of the paper, but they are useful to understand some of the trends observed in our model.

Let us start from 2g + 1g events. In the small- $\chi_{\text{max}}$  limit, the spin of 1g BHs is negligible, while 2g remnants will have spins  $\chi_f \simeq 0.68$ , which yields

$$\chi_{\text{eff}} \simeq \frac{\chi_f \cos \theta_1}{1+q}. \quad (\text{A15})$$

Because  $\beta \gg 1$ , a 2g BH with  $m_1 < m_{\text{max}}$  will pair with a 1g BH of similar mass (i.e.,  $q = 1$ ), resulting in  $\chi_{\text{eff}} \simeq 0.34 \cos \theta_1$ . Since  $\cos \theta_1$  is distributed uniformly in  $[-1, 1]$ , the resulting distribution of  $\chi_{\text{eff}}$  is also uniform in  $[-0.34, 0.34]$ :

$$p_{2g+1g}(\chi_{\text{eff}}) \simeq \frac{1}{\chi_f} \quad \text{for } |\chi_{\text{eff}}| \leq \frac{\chi_f}{2}, \quad (\text{A16})$$

as seen in the bottom panel of Fig. 1. The equal-mass assumption breaks down for 2g + 1g mergers in the mass gap, which leads to events with  $|\chi_{\text{eff}}| > 0.34$ , causing the “tail” in the  $\chi_{\text{eff}}$  distribution observed in Fig. 1.

For 2g + 2g mergers, the equal-mass approximation remains appropriate. One has

$$\chi_{\text{eff}} \simeq \frac{\chi_f}{2} (\cos \theta_1 + \cos \theta_2). \quad (\text{A17})$$

Both  $\cos \theta_i$  are distributed uniformly in  $[-1, 1]$ . Equation (A3) returns a PDF,

$$p_{2g+2g}(\chi_{\text{eff}}) \simeq \frac{1}{\chi_f} \left(1 - \frac{|\chi_{\text{eff}}|}{\chi_f}\right) \quad \text{for } |\chi_{\text{eff}}| \leq \chi_f, \quad (\text{A18})$$

in good agreement with Fig. 1.

With the above distributions, we can also provide an analytical approximation for the probability that an event lies in the spin gap:

$$p_{2g+1g}(|\chi_{\text{eff}}| > \chi_{\text{max}}) = 1 - \frac{2\chi_{\text{max}}}{\chi_f}, \quad (\text{A19})$$

$$p_{2g+2g}(|\chi_{\text{eff}}| > \chi_{\text{max}}) = \left(1 - \frac{\chi_{\text{max}}}{\chi_f}\right)^2. \quad (\text{A20})$$

Suppose that a fraction  $f_{2g+2g}$  of all 2g events are assigned to 2g + 2g (cf. Sec. II C 2), so that

$$p_{2g}(\chi_{\text{eff}}) = (1 - f_{2g+2g})p_{2g+1g}(\chi_{\text{eff}}) + f_{2g+2g}p_{2g+2g}(\chi_{\text{eff}}). \quad (\text{A21})$$

The probability of having a spin-gap event is given by

$$p_{2g}(|\chi_{\text{eff}}| > \chi_{\text{max}}) = 1 - \frac{2\chi_{\text{max}}}{\chi_f} + f_{2g+2g} \left(\frac{\chi_{\text{max}}}{\chi_f}\right)^2. \quad (\text{A22})$$

The term of order  $\mathcal{O}(\chi_{\text{max}}^2)$  can be neglected for small values of  $\chi_{\text{max}}$ . Furthermore, let us note that  $f_{2g+2g}$  is a monotonically decreasing function of  $\chi_{\text{max}}$ , with maximum  $\simeq 0.25$ . We can thus approximate the spin-gap efficiency as being independent of  $f_{2g+2g}$ :

$$\lambda_S(\chi_{\text{max}}) \simeq \left(1 - \frac{2\chi_{\text{max}}}{\chi_f}\right) \lambda_{2g}(\chi_{\text{max}}), \quad (\text{A23})$$

where  $\lambda_{2g}(\chi_{\text{max}})$  is the efficiency of producing 2g events, i.e., the fraction of all events that are either 2g + 2g or 1g + 2g.

- 
- [1] I. Mandel and A. Farmer, [arXiv:1806.05820](#).
  - [2] M. Mapelli, [arXiv:1809.09130](#).
  - [3] P. Marchant, N. Langer, P. Podsiadlowski, T. M. Tauris, and T. J. Moriya, *Astron. Astrophys.* **588**, A50 (2016).
  - [4] S. E. de Mink and I. Mandel, *Mon. Not. R. Astron. Soc.* **460**, 3545 (2016).
  - [5] N. W. C. Leigh *et al.*, *Mon. Not. R. Astron. Soc.* **474**, 5672 (2018).
  - [6] N. C. Stone, B. D. Metzger, and Z. Haiman, *Mon. Not. R. Astron. Soc.* **464**, 946 (2017).
  - [7] I. Bartos, B. Kocsis, Z. Haiman, and S. Márka, *Astrophys. J.* **835**, 165 (2017).
  - [8] K. Silsbee and S. Tremaine, *Astrophys. J.* **836**, 39 (2017).
  - [9] B.-M. Hoang, S. Naoz, B. Kocsis, F. A. Rasio, and F. Dosopoulou, *Astrophys. J.* **856**, 140 (2018).
  - [10] G. Fragione, E. Grishin, N. W. C. Leigh, H. Perets, and R. Perna, *Mon. Not. R. Astron. Soc.* **488**, 47 (2019).
  - [11] S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, *Phys. Rev. Lett.* **116**, 201301 (2016).
  - [12] M. Zevin, C. Pankow, C. L. Rodriguez, L. Sampson, E. Chase, V. Kalogera, and F. A. Rasio, *Astrophys. J.* **846**, 82 (2017).
  - [13] S. R. Taylor and D. Gerosa, *Phys. Rev. D* **98**, 083017 (2018).
  - [14] D. Wysocki, J. Lange, and R. O'Shaughnessy, *Phys. Rev. D* **100**, 043012 (2019).
  - [15] J. Roulet and M. Zaldarriaga, *Mon. Not. R. Astron. Soc.* **484**, 4216 (2019).
  - [16] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), *Astrophys. J.* **882**, L24 (2019).
  - [17] D. Gerosa, M. Kesden, E. Berti, R. O'Shaughnessy, and U. Sperhake, *Phys. Rev. D* **87**, 104028 (2013).
  - [18] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, *Astrophys. J.* **832**, L2 (2016).
  - [19] B. Farr, D. E. Holz, and W. M. Farr, *Astrophys. J.* **854**, L9 (2018).
  - [20] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gładysz, *Phys. Rev. D* **98**, 084036 (2018).
  - [21] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
  - [22] S. Vitale, D. Gerosa, C.-J. Haster, K. Chatziioannou, and A. Zimmerman, *Phys. Rev. Lett.* **119**, 251103 (2017).
  - [23] S. Miller, T. A. Callister, and W. Farr, *Astrophys. J.* **895**, 128 (2020).
  - [24] B. Zackay, T. Venumadhav, L. Dai, J. Roulet, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023007 (2019).
  - [25] B. Zackay, L. Dai, T. Venumadhav, J. Roulet, and M. Zaldarriaga, *Phys. Rev. D* **101**, 083030 (2020).
  - [26] Y. Huang, C.-J. Haster, S. Vitale, A. Zimmerman, J. Roulet, T. Venumadhav, B. Zackay, L. Dai, and M. Zaldarriaga, [arXiv:2003.04513](#).
  - [27] J. Fuller and L. Ma, *Astrophys. J.* **881**, L1 (2019).
  - [28] D. Gerosa and E. Berti, *Phys. Rev. D* **95**, 124046 (2017).
  - [29] M. Fishbach, D. E. Holz, and B. Farr, *Astrophys. J.* **840**, L24 (2017).
  - [30] C. L. Rodriguez, M. Zevin, P. Amaro-Seoane, S. Chatterjee, K. Kremer, F. A. Rasio, and C. S. Ye, *Phys. Rev. D* **100**, 043027 (2019).
  - [31] D. Gerosa and E. Berti, *Phys. Rev. D* **100**, 041301 (2019).
  - [32] S. E. Woosley, *Astrophys. J.* **836**, 244 (2017).
  - [33] K. Belczynski *et al.*, *Astron. Astrophys.* **594**, A97 (2016).



- [34] M. Renzo, R. J. Farmer, S. Justham, S. E. de Mink, Y. Göteborg, and P. Marchant, *Mon. Not. R. Astron. Soc.* **493**, 4333 (2020).
- [35] R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham, *Astrophys. J.* **887**, 53 (2019).
- [36] U. N. Di Carlo, M. Mapelli, Y. Bouffanais, N. Giacobbo, S. Bressan, M. Spera, and F. Haardt, [arXiv:1911.01434](#).
- [37] S. Stevenson, M. Sampson, J. Powell, A. Vigna-Gómez, C. J. Neijssel, D. Szécsi, and I. Mandel, *Astrophys. J.* **882**, 121 (2019).
- [38] E. Berti and M. Volonteri, *Astrophys. J.* **684**, 822 (2008).
- [39] K. Belczynski and S. Banerjee, [arXiv:2002.08050](#).
- [40] D. Gerosa, F. Hébert, and L. C. Stein, *Phys. Rev. D* **97**, 104049 (2018).
- [41] Z. Roupas and D. Kazanas, *Astron. Astrophys.* **632**, L8 (2019).
- [42] T. Kinugawa, A. Miyamoto, N. Kanda, and T. Nakamura, *Mon. Not. R. Astron. Soc.* **456**, 1093 (2016).
- [43] T. Hartwig, M. Volonteri, V. Bromm, R. S. Klessen, E. Barausse, M. Magg, and A. Stacy, *Mon. Not. R. Astron. Soc.* **460**, L74 (2016).
- [44] K. Belczynski, T. Ryu, R. Perna, E. Berti, T. L. Tanaka, and T. Bulik, *Mon. Not. R. Astron. Soc.* **471**, 4702 (2017).
- [45] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, *Phys. Rev. D* **97**, 023012 (2018).
- [46] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, *Astrophys. J.* **874**, L2 (2019).
- [47] K. Kremer, C. S. Ye, N. Z. Rui, N. C. Weatherford, S. Chatterjee, G. Fragione, C. L. Rodriguez, M. Spera, and F. A. Rasio, *Astrophys. J. Suppl. Ser.* **247**, 48 (2020).
- [48] A. Askar, M. Szkudlarek, D. Gondek-Rosińska, M. Giersz, and T. Bulik, *Mon. Not. R. Astron. Soc.* **464**, L36 (2017).
- [49] C. L. Rodriguez, S. Chatterjee, and F. A. Rasio, *Phys. Rev. D* **93**, 084029 (2016).
- [50] S. Banerjee, *Mon. Not. R. Astron. Soc.* **467**, 524 (2017).
- [51] U. N. Di Carlo, N. Giacobbo, M. Mapelli, M. Pasquato, M. Spera, L. Wang, and F. Haardt, *Mon. Not. R. Astron. Soc.* **487**, 2947 (2019).
- [52] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016).
- [53] F. Antonini and F. A. Rasio, *Astrophys. J.* **831**, 187 (2016).
- [54] N. Choksi, M. Volonteri, M. Colpi, O. Y. Gnedin, and H. Li, *Astrophys. J.* **873**, 100 (2019).
- [55] I. Y. Georgiev, T. Böker, N. Leigh, N. Lützgendorf, and N. Neumayer, *Mon. Not. R. Astron. Soc.* **457**, 2122 (2016).
- [56] D. Merritt, *Astrophys. J.* **718**, 739 (2010).
- [57] W. E. Harris, *Astron. J.* **112**, 1487 (1996).
- [58] J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton University Press, Princeton, NJ, 1987).
- [59] H. M. Lee, *Mon. Not. R. Astron. Soc.* **272**, 605 (1995).
- [60] M. Atakan Gurkan, M. Freitag, and F. A. Rasio, *Astrophys. J.* **604**, 632 (2004).
- [61] M. Morscher, B. Pattabiraman, C. Rodriguez, F. A. Rasio, and S. Umbreit, *Astrophys. J.* **800**, 9 (2015).
- [62] M. C. Miller and V. M. Lauburg, *Astrophys. J.* **692**, 917 (2009).
- [63] N. Ivanova, K. Belczynski, J. M. Fregeau, and F. A. Rasio, *Mon. Not. R. Astron. Soc.* **358**, 572 (2005).
- [64] D. C. Heggie, *Mon. Not. R. Astron. Soc.* **173**, 729 (1975).
- [65] G. D. Quinlan and S. L. Shapiro, *Astrophys. J.* **343**, 725 (1989).
- [66] J. Samsing, D. J. D’Orazio, K. Kremer, C. L. Rodriguez, and A. Askar, *Phys. Rev. D* **101**, 123010 (2020).
- [67] P. C. Peters and J. Mathews, *Phys. Rev.* **131**, 435 (1963).
- [68] G. D. Quinlan, *New Astron.* **1**, 35 (1996).
- [69] M. C. Miller and D. P. Hamilton, *Mon. Not. R. Astron. Soc.* **330**, 232 (2002).
- [70] K. El-Badry, E. Quataert, D. R. Weisz, N. Choksi, and M. Boylan-Kolchin, *Mon. Not. R. Astron. Soc.* **482**, 4528 (2019).
- [71] G. Fragione and B. Kocsis, *Phys. Rev. Lett.* **121**, 161103 (2018).
- [72] D. A. Forbes, N. Bastian, M. Gieles, R. A. Crain, J. M. D. Kruijssen, S. S. Larsen, S. Ploekinger, O. Agertz, M. Trenti, A. M. N. Ferguson, J. Pfeffer, and O. Y. Gnedin, *Proc. R. Soc. A* **474**, 20170616 (2018).
- [73] P. Madau and M. Dickinson, *Annu. Rev. Astron. Astrophys.* **52**, 415 (2014).
- [74] D. W. Hogg, [arXiv:astro-ph/9905116](#).
- [75] W. E. Harris and R. E. Pudritz, *Astrophys. J.* **429**, 177 (1994).
- [76] F. Antonini and M. Gieles, *Mon. Not. R. Astron. Soc.* **492**, 2936 (2020).
- [77] E. Barausse, V. Morozova, and L. Rezzolla, *Astrophys. J.* **758**, 63 (2012); **786**, 76(E) (2014).
- [78] F. Hofmann, E. Barausse, and L. Rezzolla, *Astrophys. J.* **825**, L19 (2016).
- [79] M. Campanelli, C. O. Lousto, Y. Zlochower, and D. Merritt, *Astrophys. J.* **659**, L5 (2007).
- [80] J. A. Gonzalez, U. Sperhake, B. Bruegmann, M. Hannam, and S. Husa, *Phys. Rev. Lett.* **98**, 091101 (2007).
- [81] C. O. Lousto and Y. Zlochower, *Phys. Rev. D* **77**, 044028 (2008).
- [82] C. O. Lousto, Y. Zlochower, M. Dotti, and M. Volonteri, *Phys. Rev. D* **85**, 084015 (2012).
- [83] C. O. Lousto and Y. Zlochower, *Phys. Rev. D* **87**, 084027 (2013).
- [84] D. Gerosa and M. Kesden, *Phys. Rev. D* **93**, 124066 (2016).
- [85] N. Choksi, O. Y. Gnedin, and H. Li, *Mon. Not. R. Astron. Soc.* **480**, 2343 (2018).
- [86] D. Wysocki, D. Gerosa, R. O’Shaughnessy, K. Belczynski, W. Gladysz, E. Berti, M. Kesden, and D. E. Holz, *Phys. Rev. D* **97**, 043014 (2018).
- [87] K. K. Y. Ng, S. Vitale, A. Zimmerman, K. Chatziioannou, D. Gerosa, and C.-J. Haster, *Phys. Rev. D* **98**, 083007 (2018).
- [88] B. F. Schutz, *Classical Quantum Gravity* **28**, 125023 (2011).
- [89] H.-Y. Chen and D. E. Holz, [arXiv:1409.0522](#).
- [90] M. Dominik, E. Berti, R. O’Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale, *Astrophys. J.* **806**, 263 (2015).
- [91] P. Christian, P. Mocz, and A. Loeb, *Astrophys. J.* **858**, L8 (2018).

- [92] M. Arca Sedda and M. Benacquista, *Mon. Not. R. Astron. Soc.* **482**, 2991 (2019).
- [93] M. Safarzadeh, *Astrophys. J. Lett.* **892**, L8 (2020).
- [94] M. A. Sedda, M. Mapelli, M. Spera, M. Benacquista, and N. Giacobbo, *Astrophys. J.* **894**, 133 (2020).
- [95] Z. Doctor, D. Wysocki, R. O’Shaughnessy, D. E. Holz, and B. Farr, *Astrophys. J.* **893**, 35 (2020).
- [96] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (Springer-Verlag, Berlin, Heidelberg, 1992).
- [97] S. Hilbert, S. D. M. White, J. Hartlap, and P. Schneider, *Mon. Not. R. Astron. Soc.* **382**, 121 (2007).
- [98] M. Lima, B. Jain, and M. Devlin, *Mon. Not. R. Astron. Soc.* **406**, 2352 (2010).
- [99] V. Gayathri, I. Bartos, Z. Haiman, S. Klimentko, B. Kocsis, S. Marka, and Y. Yang, *Astrophys. J.* **890**, L20 (2020).
- [100] M. Fishbach, W. M. Farr, and D. E. Holz, *Astrophys. J.* **891**, L31 (2020).