




# Hierarchical binary black hole mergers in globular clusters: Mass function and evolution with redshift<sup>★</sup>

Stefano Torniamenti<sup>1,2,3,4</sup>, Michela Mapelli<sup>1,2,3,4</sup>, Carole P erigois<sup>2,3</sup>, Manuel Arca Sedda<sup>2,5,6,7</sup>,  
Maria Celeste Artale<sup>2,3,8</sup>, Marco Dall’Amico<sup>2,3</sup>, and Maria Paola Vaccaro<sup>1,2</sup>

<sup>1</sup> Institut f ur Theoretische Astrophysik, ZAH, Universit at Heidelberg, Albert-Ueberle-Stra e 2, 69120 Heidelberg, Germany  
e-mail: [stefano.torniamenti@unipd.it](mailto:stefano.torniamenti@unipd.it), [mapelli@uni-heidelberg.de](mailto:mapelli@uni-heidelberg.de)

<sup>2</sup> Physics and Astronomy Department Galileo Galilei, University of Padova, Vicolo dell’Osservatorio 3, 35122 Padova, Italy

<sup>3</sup> INFN – Padova, Via Marzolo 8, 35131 Padova, Italy

<sup>4</sup> INAF – Osservatorio Astronomico di Padova, Vicolo dell’Osservatorio 5, 35122 Padova, Italy

<sup>5</sup> Gran Sasso Science Institute (GSSI), 67100 L’Aquila, Italy

<sup>6</sup> INFN, Laboratori Nazionali del Gran Sasso, 67100 Assergi, Italy

<sup>7</sup> INAF – Osservatorio Astronomico d’Abruzzo, Teramo, Italy

<sup>8</sup> Instituto de Astrof sica, Facultad de Ciencias Exactas, Universidad Andres Bello, Fernandez Concha 700, Santiago, Chile

Received 18 January 2024 / Accepted 22 April 2024

## ABSTRACT

Hierarchical black hole (BH) mergers are one of the most straightforward mechanisms producing BHs inside and above the pair-instability mass gap. We investigated the impact of globular cluster (GC) evolution on hierarchical mergers, accounting for the uncertainties related to BH mass pairing functions on the predicted primary BH mass, mass ratio, and spin distribution. We find that the evolution of the host GC quenches the hierarchical BH assembly at the third generation, mainly due to cluster expansion powered by a central BH subsystem. Hierarchical mergers match the primary BH mass distribution from GW events for  $m_1 > 50 M_\odot$  regardless of the assumed BH pairing function. At lower masses, however, different pairing functions lead to dramatically different predictions on the primary BH mass merger-rate density. We find that the primary BH mass distribution evolves with redshift, with a larger contribution from mergers with  $m_1 \geq 30 M_\odot$  for  $z \geq 2$ . Finally, we calculate the mixing fraction of binary black holes (BBHs) from GCs and isolated binary systems. Our predictions are very sensitive to the spins, which favor a large fraction ( $>0.6$ ) of BBHs born in GCs in order to reproduce misaligned spin observations.

**Key words.** black hole physics – gravitational waves – stars: black holes – stars: kinematics and dynamics – galaxies: star clusters: general

## 1. Introduction

The first three observing runs of the Advanced LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) interferometers have delivered a gravitational-wave (GW) transient catalog (GWTC-3) of approximately 90 merger candidates (Abbott et al. 2019, 2021a, 2023a, 2024). Most of these events are associated with the merger of binary black holes (BBHs), with masses spanning from the so-called lower mass-gap (e.g., the secondary component of GW190814; Abbott et al. 2020a) to the intermediate-mass range (e.g., GW190521; Abbott et al. 2020b,c). The number of detected black holes (BHs) has now become sufficiently large to allow exploration of the underlying BH mass distribution (Abbott et al. 2021b, 2023b), which encodes information about the astrophysical processes governing their formation and merger (Fishbach et al. 2017, 2021; Vitale et al. 2017; Callister et al. 2021a; Callister & Farr 2024; Farah et al. 2023).

One of the most puzzling features of the primary BH mass distribution from GWTC-3 is the absence of a high-mass cutoff above  $\sim 60 M_\odot$  (Callister & Farr 2024), as we would

expect from (pulsational) pair instability (Belczynski et al. 2017; Woosley 2017; Spera & Mapelli 2017; Marchant et al. 2019; Stevenson et al. 2019). The formation of BHs in this mass range is still a matter of debate, because of the large uncertainties that affect both the lower and upper boundary of the claimed pair-instability mass gap (Farmer et al. 2019, 2020; Mapelli et al. 2020; Belczynski et al. 2020; van Son et al. 2020; Marchant & Moriya 2020; Farrell et al. 2021; Vink et al. 2021; Woosley & Heger 2021; Costa et al. 2021; Briel et al. 2023; Hendriks et al. 2023).

In principle, dynamical interactions in star clusters can lead to the formation of BHs in the pair-instability mass gap through direct star–star collisions (Di Carlo et al. 2019; Renzo et al. 2020; Kremer et al. 2020; Costa et al. 2022; Ballone et al. 2023) or hierarchical (i.e., repeated) BH mergers (Quinlan & Shapiro 1990; Miller & Hamilton 2002; O’Leary et al. 2006; Antonini & Rasio 2016; Fishbach et al. 2017; Gerosa & Berti 2017; Gerosa & Fishbach 2021). Direct  $N$ -body simulations show that star–star collisions are not rare in young and open clusters (Di Carlo et al. 2020; Torniamenti et al. 2022), where the short relaxation time-scales (Portegies Zwart et al. 2010) lead to mass segregation within less than 1 Myr, enhancing the possibility of direct star collisions. In more massive globular clusters (GCs) and nuclear star clusters, BHs born from the merger of other BHs

<sup>★</sup> The data underlying this article are available on Zenodo at <https://doi.org/10.5281/zenodo.11367590> (Torniamenti et al. 2024).

can efficiently be retained and pair up again to produce second- and higher-generation mergers. Such hierarchical assembly can repeat several times and lead to significant BH mass growth (Antonini & Rasio 2016; Mapelli et al. 2021, 2022; Chattopadhyay et al. 2023), thus populating the pair-instability mass gap and intermediate-mass regime.

One of the main challenges in studying hierarchical BH mergers in such massive and long-lived clusters is the computational cost. Accurate but computationally expensive Monte Carlo and/or  $N$ -body simulations permit the exploration of only a limited portion of the parameter space for both the BBH population (e.g., first-generation masses, spins) and the host star cluster (total mass, central density, initial binary fraction). For this reason, fast and flexible semi-analytic models are usually adopted to probe hierarchical BH assembly (Antonini & Rasio 2016; Antonini et al. 2019, 2023; Mapelli et al. 2021, 2022; Kritos et al. 2022, 2023; Arca Sedda et al. 2023a; Fragione & Rasio 2023). In particular, our semi-analytic code FASTCLUSTER<sup>1</sup> (Mapelli et al. 2021, 2022; Vaccaro et al. 2024) integrates the evolution of the BBH orbital properties (semi-major axis and eccentricity) driven by GW emission and dynamical encounters while also providing the possibility to explore different assumptions for the BBH masses, spins, orbital eccentricities, and for different star cluster masses, densities, and initial binary fractions.

In the present work, we account for the temporal evolution of the globular cluster, including stellar mass loss, two-body relaxation, and tidal stripping by the host galaxy. Our main goal is to explore the impact of GC evolution and the BBH pairing function on the population of hierarchical BBH mergers. We investigated how the internal GC evolution and tidal stripping from the host galaxy quench the BH assembly. Also, we accounted for uncertainties related to first-generation mass distribution by relying on different BBH pairing functions, that is, the criterion to pair up the primary and secondary BHs in GCs. We investigated the peculiar imprint that hierarchical mergers have on the mass, mass ratio, and spin distributions, and we find it can affect the detected parameters from GWs. Finally, we compared the mass, mass ratio, and spin distributions of BBH mergers in GCs to those found in isolation in order to obtain insights into their relative contributions to GW detections.

This work is organized as follows. In Sect. 2 we introduce the details of the code implementation. Section 3 describes our main results, which we discuss in Sect. 4. Finally, Sect. 5 summarizes our conclusions.

## 2. Methods

Here, we summarize how we model the evolution of the cluster. We assume the formalism presented by Antonini & Gieles (2020). In particular, we adopt a two-component formalism (Breen & Heggie 2013) in which the cluster consists of two populations: stars, with mass  $m_*$  and total mass  $M_*$ , and BHs, with mass  $m_{\text{BH}}$  and total mass  $M_{\text{BH}}$ . The resulting total star cluster mass is  $M_{\text{tot}} = M_* + M_{\text{BH}}$ .

The total mass of the stellar population decreases because of stellar evolution (stellar winds, supernova explosions), and tidal stripping from the host galaxy. We model the stellar mass loss as in Antonini & Gieles (2020), assuming that  $M_*$  evolves as a power of time. As a consequence, we assume that the cluster undergoes an adiabatic expansion. Also, we implement

tidal stripping from the host galaxy following the formalism by Gieles et al. (2011), who describe the evolution of GCs in circular orbits and in different tidally filling states.

We model GC expansion due to a central BH subsystem, and relate it to BH ejection from the cluster core (Antonini & Gieles 2020). As the GC evolves toward a state of energy equipartition (Bianchini et al. 2016; Spera et al. 2016), BHs segregate to the cluster core, giving birth to a central BH subsystem (Spitzer 1987). The temperature difference between BHs and the bulk of the cluster (Spitzer 1969) generates a heat transfer from the core toward the outer parts (Breen & Heggie 2013). The energy flux is produced by BBH hardening, until dynamical recoils eject most BBHs (Heggie 1975; Goodman 1984). The GC eventually achieves a state of balanced evolution, where the energy flux from the core is regulated by the energy demand of the cluster as a whole (Hénon 1961).

We model the BH mass loss during the phase of balanced evolution, when it can be related to the GC energy transfer (Breen & Heggie 2013). We consider the onset of balanced evolution as the time of core collapse of the host GC. Then, we implement the decrease in the total BH mass and the resulting expansion rate following Antonini & Gieles (2020). More details can be found in Appendix A.

### 2.1. Global properties of GCs

Following Mapelli et al. (2022), we drew the initial cluster total mass from a log-normal distribution with a mean of  $\langle \log_{10} M_{\text{tot}}/M_{\odot} \rangle = 5.6$  (Harris 1996), and assumed a fiducial standard deviation of  $\sigma_M = 0.4$ . We drew the GC density at the half-mass radius from a log-normal distribution with a mean of  $\langle \log_{10} \rho/(M_{\odot} \text{pc}^{-3}) \rangle = 3.7$ , and assumed a fiducial standard deviation of  $\sigma_{\rho} = 0.4$ . In Appendix B, we discuss how different initial GC masses and densities affect our results.

We set the core density to  $\rho_c = 20\rho$  (Mapelli et al. 2021). This choice yields a good approximation of the ratio of the core to the half-mass density in GCs with the considered BH mass fraction, as shown in Breen & Heggie (2013). We calculate the cluster escape velocity as (Georgiev et al. 2009):

$$v_{\text{esc}} = 40 \text{ km s}^{-1} \left( \frac{M_{\text{tot}}}{10^5 M_{\odot}} \right)^{1/3} \left( \frac{\rho}{10^5 M_{\odot} \text{pc}^{-3}} \right)^{1/6}. \quad (1)$$

### 2.2. Hierarchical BBH mergers

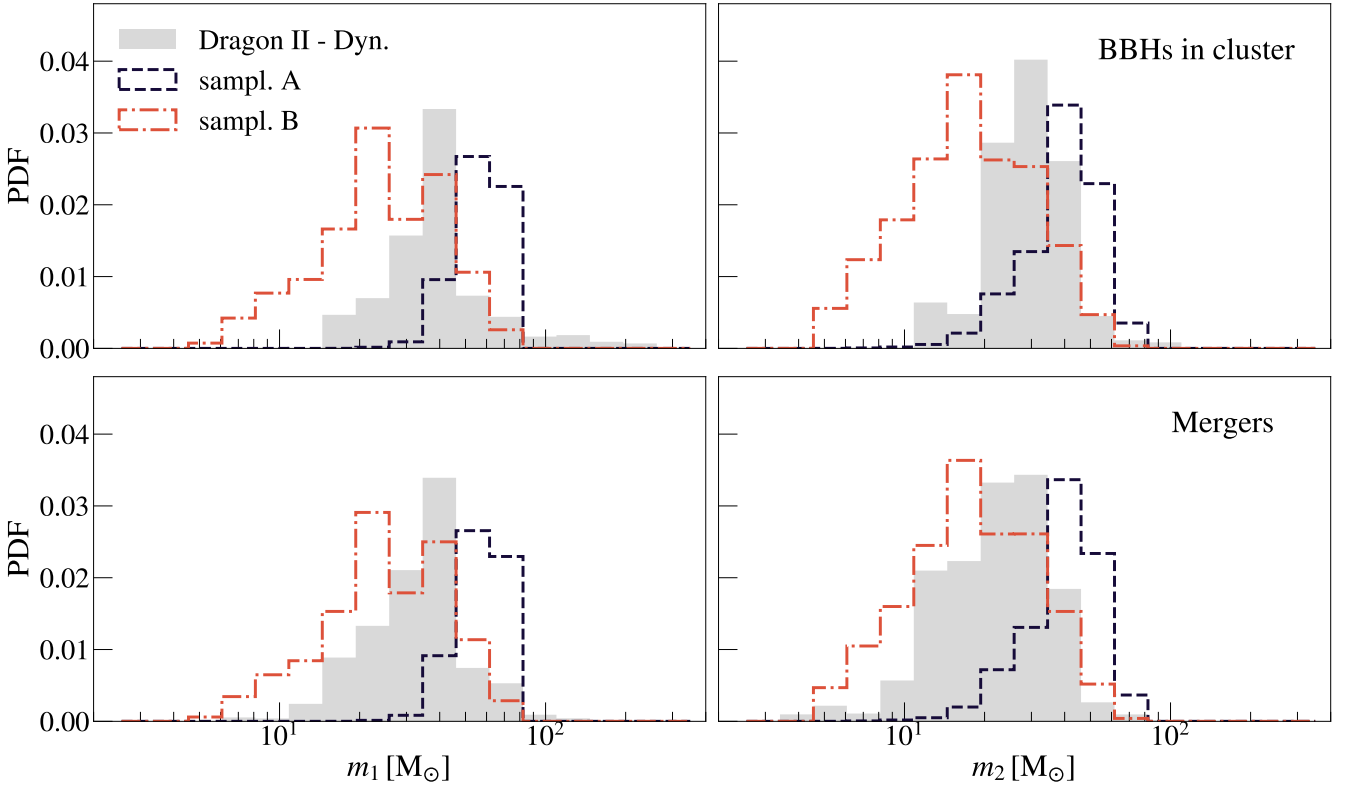
In the following, we present our model for evolving dynamically formed BBHs in GCs. As the details of the implementation have already been presented by Mapelli et al. (2021), here we mainly focus on the upgrades we introduced.

#### 2.2.1. Pairing function and other properties of first-generation (1g) BHs

We generate the masses of first generation (1g) BHs with the population synthesis code MOBSE<sup>2</sup> (Mapelli et al. 2017; Giacobbo et al. 2018). In this particular case, we adopted the rapid model by Fryer et al. (2012) for core-collapse supernovae. For (pulsational) pair-instability supernovae, we used the equations reported in the appendix of Mapelli et al. (2020). This yields a minimum BH mass of  $\approx 5 M_{\odot}$  and a maximum BH mass

<sup>1</sup> FASTCLUSTER is an open-source code available at [https://gitlab.com/micmap/fastcluster\\_open](https://gitlab.com/micmap/fastcluster_open)

<sup>2</sup> MOBSE (Mapelli et al. 2017; Giacobbo et al. 2018) is a custom and upgraded version of BSE (Hurley et al. 2002), and is publicly available at [https://gitlab.com/micmap/mobse\\_open](https://gitlab.com/micmap/mobse_open)



**Fig. 1.** Distribution of first-generation primary (left) and secondary (right) masses of dynamical BBHs that are retained within the cluster (upper panels) and merge (lower panels). We show the distributions from the Dragon-II direct  $N$ -body simulations (Arca Sedda et al. 2024a, gray shaded area) and in FASTCLUSTER, produced by sampling A (blue dashed line) and B (red dot-dashed line).

that depends on metallicity, and can be as high as  $\approx 70 M_{\odot}$  for metal-poor stars at  $Z = 0.0002$ , which approximately corresponds to  $0.01 Z_{\odot}$ .

We assume that our 1g BHs pair up with other BHs via three-body encounters at the time of core collapse ( $t_{cc}$ , Eq. (A.3)), that is when dynamical interactions are expected to become effective within the BH subsystem core (Heggie & Hut 2003).

We paired the BH masses following the mass sampling criterion from Antonini et al. (2023). This pairing function is conceived to describe the dynamical BBH formation in clusters from three-body encounters, and favors the pairing of the most massive objects. The BH sampling is based on the formation rate of hard binaries per unit of volume and energy (Heggie 1975), under the assumption that energy equipartition holds in the BH subsystem:

$$\Gamma_{3b}(m_1, m_2, m_3) \propto \frac{n_1 n_2 n_3 m_1^4 m_2^4 m_3^{5/2}}{(m_1 + m_2 + m_3)^{1/2} (m_1 + m_2)^{1/2}} \beta^{9/2}, \quad (2)$$

where  $n_i = n(m_i)$  is the number density of BHs with mass  $m_i$  (with  $i = 1, 2$ , and  $3$ ), while  $\beta$  is the inverse of their mean internal energy assuming that they are in energy equipartition.

As Eq. (2) is symmetric with respect to  $m_1$  and  $m_2$ , the probability density function of sampling one of the two is:

$$p_{1,2}(m_{1,2}) = \int_{m_{low}}^{m_{up}} \int_{m_{low}}^{m_{up}} dm_3 dm_{2,1} \Gamma_{3b}(m_1, m_2, m_3), \quad (3)$$

where  $m_{low}$  and  $m_{up}$  are the lower and upper limit of the mass distribution. We randomly sampled two BH masses per time from the MOBSE catalog following the probability distribution function in Eq. (3). We assigned the primary (secondary) mass  $m_1$

( $m_2$ ) to the larger (smaller) sampled mass. Hereafter, we refer to this mass sampling criterion as model A.

To encompass the uncertainties, we also considered the pairing function introduced by Mapelli et al. (2021), who uniformly sample the primary BH mass from the BH mass function (as in the MOBSE catalog), and drew the secondary BH mass with the probability distribution function defined by O’Leary et al. (2016):

$$p(m_2) \propto (m_1 + m_2)^4, \quad (4)$$

in the interval  $[m_{min}, m_1)$ , where  $m_{min} = 5 M_{\odot}$ . Hereafter, we refer to this mass sampling criterion as sampling B. In Sect. 2.2.2 and Fig. 1, we compare the BBH masses generated by the aforementioned pairing functions with those obtained from direct  $N$ -body models.

For both sampling criteria, we applied a further correction to the BH masses based on the supernova kick, calculated as (Mapelli et al. 2021):

$$v_{SN} = v_{H05} \frac{\langle m_{NS} \rangle}{m_{BH}}, \quad (5)$$

where  $m_{BH}$  is the BH mass,  $\langle m_{NS} \rangle = 1.33 M_{\odot}$  is the average neutron star mass (Özel & Freire 2016), and  $v_{H05}$  is randomly drawn from a Maxwellian distribution with a root-mean square of  $265 \text{ km s}^{-1}$ . If one member of the BBH received a kick of larger than  $v_{esc}$  (Eq. (1)), we assumed that was ejected from the cluster, and did not integrate its dynamics any further.

We sampled the dimensionless spin magnitudes ( $\chi_1$  and  $\chi_2$ ) of the BHs from a Maxwellian distribution with root-mean square  $\sigma_{\chi} = 0.1$ . We chose this distribution as a toy model, because there is no consensus on the spin distribution from

first principles (see, e.g., Fuller & Ma 2019; Belczynski et al. 2020; Bavera et al. 2020; Périgois et al. 2023). Our choice qualitatively matches the spin distribution inferred by the LIGO–Virgo–KAGRA (LVK) Collaboration (Abbott et al. 2023b). We drew spin directions isotropic over the sphere, because dynamical interactions reset any spin alignments with the orbital angular momentum of the binary system.

Finally, we sampled the initial orbital eccentricity from a thermal probability distribution (Heggie 1975):

$$p(e) = 2e \quad e \in [0, 1), \quad (6)$$

and the semi-major axis from:

$$p(a) \propto a^{-1} \quad a \in [1, 10^3] R_{\odot}. \quad (7)$$

As soft binaries tend to be disrupted by dynamical interactions (Heggie 1975), we first checked whether the newly generated binary was hard. If it was not, we did not integrate it any further, assuming it was disrupted by further interactions (Mapelli et al. 2021).

### 2.2.2. Testing the BH pairing functions with $N$ -body models

In the following, we compare the distributions of BBHs given by the BBH pairing functions introduced in Sect. 2.2.1 with those from the  $N$ -body models by Arca Sedda et al. (2023b, 2024a,b), whose stellar and binary evolution prescriptions are also based on the formulae of (Hurley et al. 2000, 2002; see Kamlah et al. 2022 for more details). Specifically, we ran  $10^6$  models at the same metallicity as the  $N$ -body simulations ( $Z = 0.0005$ ), and compared the distributions of first-generation BBHs that are retained within the cluster and that merge.

As shown in Fig. 1, the BBH distributions from  $N$ -body simulations lie somewhere in between the two considered pairing functions. The sampling A captures the expected dynamical trend, which favors the coupling of the most massive BHs. The position of the dynamical peak, however, is very sensitive to the underlying stellar evolutionary models. For this reason, different prescriptions can result in discrepancies at the high-mass end of the distribution. The  $N$ -body models also display a long high-mass tail due to gas accretion from companion stars. The sampling B displays a milder increase at higher masses, without the presence of a clear peak, as in the  $N$ -body models. However, this case can reproduce the trend at low masses, despite the different BH mass distribution at first generation.

Our comparison suggests that there are many uncertainties related to the BH pairing function that are due to the impact of primordial binaries (which are not included in our formalism), deviations with respect to energy equipartition (Spera et al. 2016), stellar and binary evolution models, BH–star interactions, and possibly stochastic fluctuations. For example, Arca Sedda et al. (2024a) assume that BHs can accrete up to 50% of the stellar mass during BH–star collisions. This results in the high-mass tail above the lower limit of the pair-instability mass gap, which is not present in our models. At the same time, the prescriptions of these latter authors for (pulsational) pair-instability supernovae produce a peak that is  $\sim 30\%$  lower than in our models at  $Z = 0.0005$ . These differences affect the resulting primary mass distribution, which is expected to peak toward the 1g BH mass upper limit. At lower masses, that is, at  $\lesssim 30 M_{\odot}$ , different schemes for core collapse supernovae affect the BH mass distribution. In this case, the delayed model used by Arca Sedda et al. (2024a) does not predict any mass gap below

$5 M_{\odot}$ , in contrast to the rapid model used for our runs. These differences mainly affect the secondary masses, which can extend to the lower limit of the 1g BH mass distribution.

A detailed exploration of all these factors goes beyond the scope of this work and merits a dedicated study. In the following, we show the results of sampling both A and B to bracket these uncertainties.

### 2.2.3. BBH evolution

We evolve the semi-major axis  $a$  and orbital eccentricity  $e$  of hard BBHs through the formalism described in Eq. (16) of Mapelli et al. (2021), that is, by integrating a system of ordinary differential equations for  $a$  and  $e$  that takes into account dynamical hardening and GW decay or circularization. When a BBH merges within a Hubble time (13.6 Gyr), we estimate the mass and spin of the merger remnant from the fitting formulas provided by Jiménez-Forteza et al. (2017). Also, we calculate the relativistic kick magnitude  $v_K$  (Lousto et al. 2012). The merger remnant is retained inside its parent cluster if the relativistic kick magnitude  $v_K < v_{\text{esc}}$ , with  $v_{\text{esc}}$  given by Eq. (1). In this case, we model the dynamical formation of an  $N$ th generation BBH and its evolution, as described in Mapelli et al. (2022).

We iterate the hierarchical merger process until we reach one of the following conditions: the merger remnant is ejected from the cluster, the cluster evaporates, no BHs remain within the cluster, or we reach the Hubble time.

### 2.2.4. $N$ th-generation secondary mass sampling

We sampled the  $N$ th-generation (ng) secondary BHs following the criteria introduced in Sect. 2.2.1 for the 1g BBHs. For the sampling A, we drew  $m_2$  from Eq. (3) by setting  $m_1$  to the mass of the ng primary. As this model relies on the definition of a probability density function, which cannot be constrained in a direct way for ng BHs; we only consider ng1g BBHs.

For the sampling B, we drew the secondary mass from Eq. (4). We considered two different cases, hereafter ngng and ng1g. In the ngng case, we assumed that the secondary BH mass  $m_2$  ranges within  $[m_{\text{min}}, m_{1,\text{ng}}]$ , where  $m_{\text{min}}$  is the minimum BH mass and  $m_{1,\text{ng}}$  is the mass of the primary BH belonging to the  $N$ th generation. This yields the possibility that the secondary BH comes from previous mergers, as done by Mapelli et al. (2021).

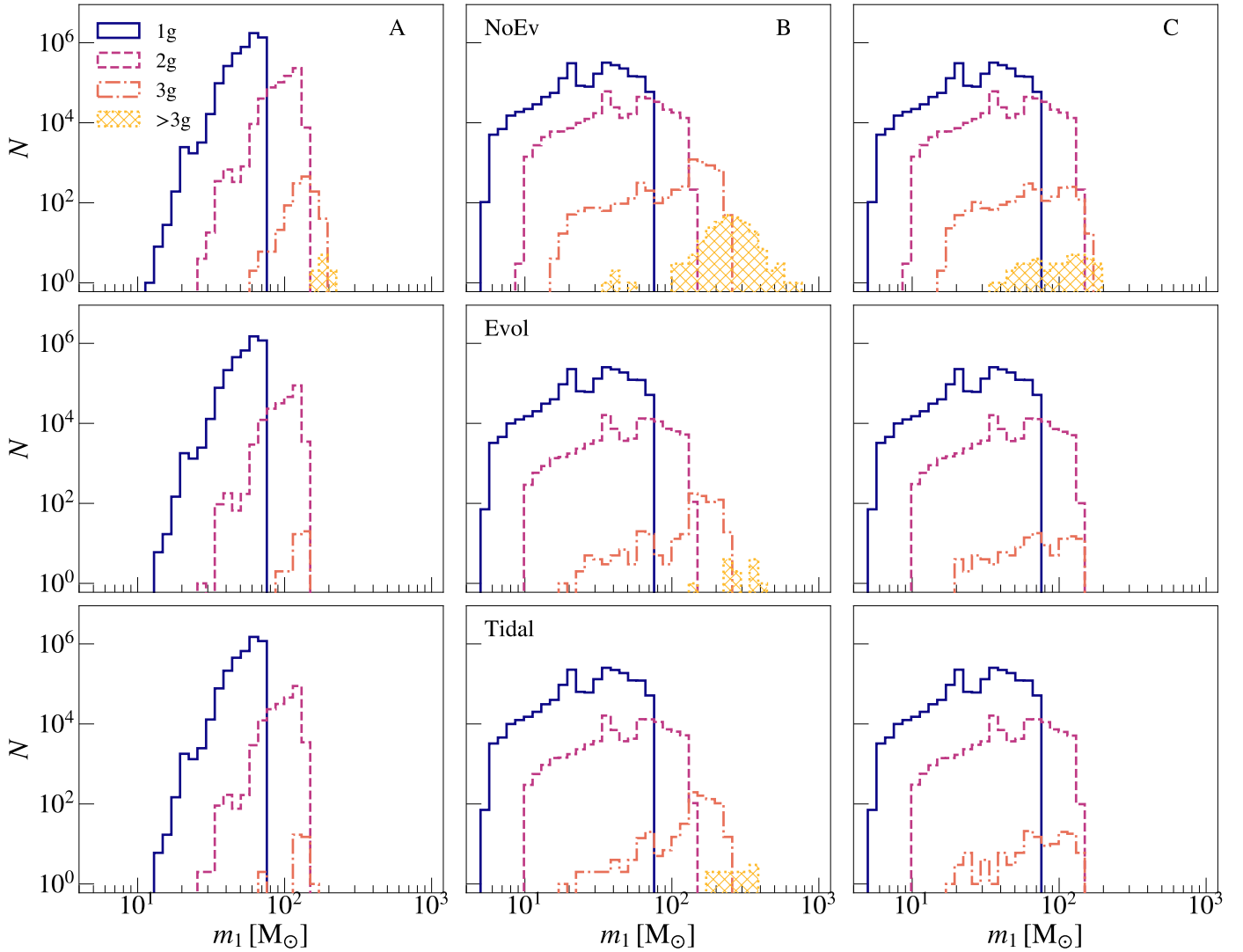
In the ng1g case, we assumed that the secondary is always a 1g BH, and  $m_2 \in [m_{\text{min}}, \max(m_{1,\text{1g}})]$ , where  $\max(m_{1,\text{1g}})$  is the maximum mass of a primary first-generation BH. This assumption leads to smaller mass ratios at  $\text{ng} > 1$ , because our secondary sampling criterion (Eq. (4)) favors larger masses but BBHs with  $q = 1$  can no longer form due to the increasing primary mass. In turn, this can result in larger relativistic recoil during a merger, the magnitude of which decreases in more symmetric configurations (González et al. 2007; Schnittman & Buonanno 2007; Herrmann et al. 2007).

## 2.3. Merger rate

We evaluated the merger rate density of dynamical BBHs in GCs (Mapelli et al. 2022) as:

$$\mathcal{R}(z) = \int_{z_{\text{max}}}^z \psi(z') \frac{dt(z')}{dz'} \left[ \int_{Z_{\text{min}}}^{Z_{\text{max}}} \eta(Z) \mathcal{F}(z', z, Z) dZ \right] dz', \quad (8)$$

where  $t(z)$  is the look-back time at redshift  $z$ ,  $\psi(z')$  is the GC formation rate density at redshift  $z'$ , and  $Z_{\text{min}}(z')$  ( $Z_{\text{max}}(z')$ ) is the



**Fig. 2.** Primary BH mass distribution of BBH mergers for models A (left), B (center), and C (right) for the three evolutionary cases considered: NoEv (upper panels), Evol (middle panels), and Tidal (lower panels). Different colors correspond to different BH generations: 1g (blue), 2g (purple), 3g (salmon), and >3g (yellow) BHs. Here, we show our models with  $Z = 0.0002$  because they are representative of metal-poor GCs in the Galactic halo (Muratov & Gnedin 2010).

minimum (maximum) metallicity of stars formed at redshift  $z'$ . The merger efficiency  $\eta$  is defined as the total number of BBHs of a given population that merge within a Hubble time divided by the total initial stellar mass of that population (Giacobbo et al. 2018; Klencki et al. 2018). Finally,  $\mathcal{F}(z', z, Z)$  is the merger rate of BBHs that form at redshift  $z'$  from stars with metallicity  $Z$  and merge at redshift  $z$ , normalized to all BBHs that form from stars with metallicity  $Z$ . We calculated the cosmological parameters ( $H_0$ ,  $\Omega_M$  and  $\Omega_\Lambda$ ) from Planck Collaboration XIII (2016).

We assumed a Gaussian distribution to model the GC formation rate as a function of redshift:

$$\psi(z) = \mathcal{B}_{\text{GC}} \exp\left[-\frac{(z - z_{\text{GC}})^2}{(2\sigma_{\text{GC}}^2)}\right], \quad (9)$$

with its maximum at  $z_{\text{GC}} = 3.2$  and a spread of  $\sigma_{\text{GC}} = 1.5$  (Mapelli et al. 2022). The fiducial normalization constant is  $\mathcal{B}_{\text{GC}} = 2 \times 10^{-4} M_\odot \text{Mpc}^{-3} \text{yr}^{-1}$ , consistently with El-Badry et al. (2019) and Reina-Campos et al. (2019).

We calculated the merger efficiency  $\eta(Z)$  as:

$$\eta(Z) = \frac{\mathcal{N}_{\text{merg, sim}}(Z)}{\mathcal{N}_{\text{sim}}(Z)} \frac{\mathcal{N}_{\text{BH}}(Z)}{M_\star(Z)}, \quad (10)$$

where, for a given  $Z$ ,  $\mathcal{N}_{\text{merg, sim}}(Z)$  is the number of simulated merging BHs,  $\mathcal{N}_{\text{sim}}(Z)$  is the number of simulated BHs,  $\mathcal{N}_{\text{BH}}$  is the total number of BHs associated with a given metallicity, and  $M_\star(Z)$  is the total initial stellar mass for a given metallicity  $Z$ . We estimated  $M_\star(Z)$  as the sum of the initial total mass of all simulated star clusters with a given  $Z$ , and  $\mathcal{N}_{\text{BH}}(Z)$  as the number of BHs that we expect from a stellar population following a Kroupa mass function between 0.1 and  $150 M_\odot$  (Kroupa 2001), assuming that all stars with a zero-age main sequence mass of  $\geq 20 M_\odot$  are BH progenitors (Heger et al. 2003).

We calculated

$$\mathcal{F}(z', z, Z) = \frac{d\mathcal{N}(z', z, Z)/dt}{\mathcal{N}_{\text{TOT}}(Z)} p(z', Z), \quad (11)$$

where  $d\mathcal{N}(z', z, Z)/dt$  is the total number of BBHs that form at redshift  $z'$  with metallicity  $Z$  and merge at redshift  $z$  per unit time, and  $\mathcal{N}_{\text{TOT}}(Z)$  is the total number of BBH mergers with progenitor's metallicity  $Z$ . We evaluated the dispersion around the mass-weighted metallicity by assuming a log-normal

distribution for the metallicity:

$$p(z', Z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left\{-\frac{[\log(Z(z')/Z_\odot) - \langle \log Z(z')/Z_\odot \rangle]^2}{2\sigma_Z^2}\right\}, \quad (12)$$

where

$$\langle \log Z(z')/Z_\odot \rangle = \log \langle Z(z')/Z_\odot \rangle - \frac{\ln(10)\sigma_Z^2}{2}. \quad (13)$$

We took the average metallicity  $\langle Z(z)/Z_\odot \rangle$  from [Madau & Fragos \(2017\)](#) and assumed a standard deviation of  $\sigma_Z = 0.3$  ([Mapelli et al. 2022](#)). We refer the interested reader to [Mapelli et al. \(2022\)](#) for all the details of the calculation and the underlying assumptions.

#### 2.4. Description of the runs

For our GC models, we considered seven metallicities:  $Z = 0.0002, 0.0008, 0.002, 0.004, 0.006, 0.012,$  and  $0.02$ . We performed a denser sampling between  $0.002$  and  $0.02$ , where the metallicity variation has a stronger impact on the  $1g$  BH distribution.

For each metallicity, we ran three sets of  $N = 5 \times 10^6$  BBHs, corresponding to the mass sampling criteria introduced in Sects. 2.2.1 and 2.2.4. In the following, we refer to these three models as follows:

- Model A: we draw the primary and secondary mass of  $1g$  BBHs according to sampling A (Sect. 2.2.1), and we sample the secondary mass of hierarchical BBHs according to the  $ng1g$  criterion (Sect. 2.2.4);
- Model B: we draw the primary and secondary mass of  $1g$  BBHs according to sampling B (Sect. 2.2.1), and we sample the secondary mass of hierarchical BBHs according to the  $ngng$  criterion (Sect. 2.2.4); that is, we assume that the secondary can be a  $ng$  BH;
- Model C: we draw the primary and secondary mass of  $1g$  BBHs according to sampling B (Sect. 2.2.1), and we sample the secondary mass of hierarchical BBHs according to the  $ng1g$  criterion (Sect. 2.2.4); that is, we assume that the secondary is always a  $1g$  BH.

We considered three evolutionary cases for the host GCs:

- NoEv: the absence of star cluster evolution.
- Evol: star cluster evolution in isolation.
- Tidal: star cluster evolution in the presence of a tidal field.

To highlight the impact of GC evolution, for each evolutionary case we set the same initial seed of our random generations. In total, we ran 189 different models, corresponding to  $9.45 \times 10^8$   $1g$  BBHs. Table 1 reports the details of our models.

## 3. Results

### 3.1. Impact of GC evolution

Figure 2 shows the impact of GC evolution on the primary BH mass distribution for different BBH generations. Independently of the selected mass sampling, GC evolution quenches hierarchical mergers at the third generation. This is at least one generation earlier than what is predicted by nonevolving GC models. Only in model B can GCs still produce  $4g$  mergers, but their efficiency

**Table 1.** Summary of the runs.

ID	1g	ng
A	A	ng1g
B	B	ngng
C	B	ng1g

**Notes.** Column 1: model name. Column 2: first-generation ( $1g$ ) mass sampling criterion. Column 3:  $N$ th-generation ( $ng$ ) secondary mass sampling criterion. See Sects. 2.2.1 and 2.2.4 for details.

is reduced by a factor of  $\sim 50$  with respect to models where GC evolution is not taken into account. The maximum primary BH mass is  $\sim 150 M_\odot$  if only  $ng1g$  mergers are considered, while for model B it can be as high as  $\sim 300 M_\odot$ .

The process that mainly quenches the hierarchical chain is the internal GC relaxation. Cluster expansion, powered by the central BH subsystem, reduces the GC central density and, in turn, the dynamical hardening rate. GC expansion and mass loss also lower the cluster escape velocity, favoring the ejection by relativistic recoil upon merger. For  $1g$  BBH mergers, the percentage of ejected BH remnants increases from 83% to 92% when GC evolution is implemented. Only in GCs with an initial  $v_{\text{esc}} \gtrsim 80 \text{ km s}^{-1}$  does this percentage fall below 80%.

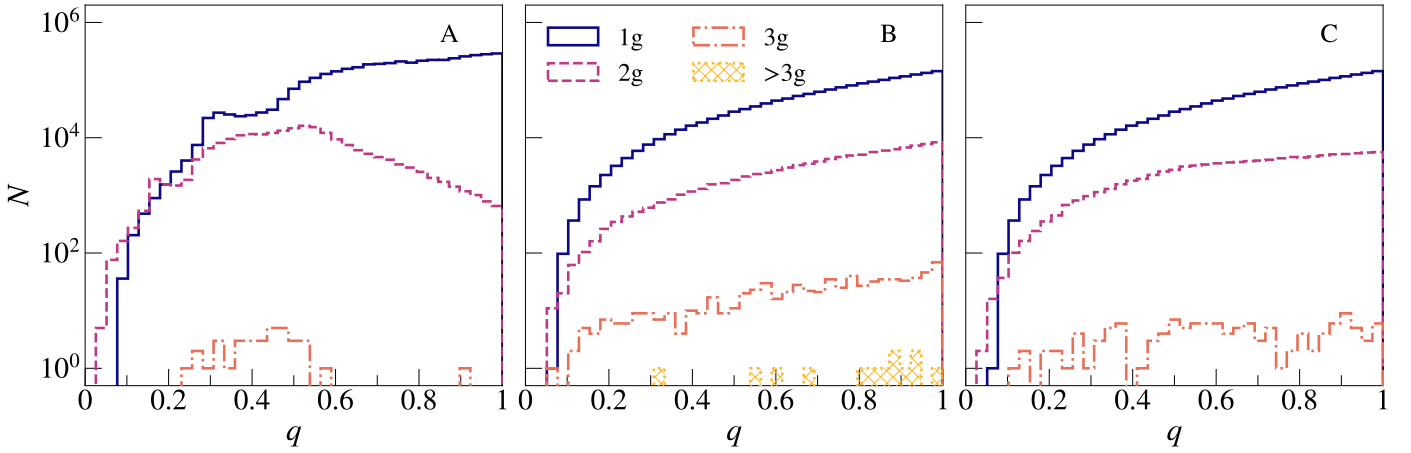
As shown in Fig. 2, the external tidal field does not affect the hierarchical assembly significantly. However, we point out that our implementation is relatively conservative: it does not account for disk and bulge shocking, and for possible encounters with molecular clouds, which can significantly enhance star cluster dissolution ([Gieles et al. 2006](#)). Furthermore, here we only consider GCs at a galactocentric distance of 8 kpc from the center of a Milky-Way like galaxy.

### 3.2. Impact of the pairing function

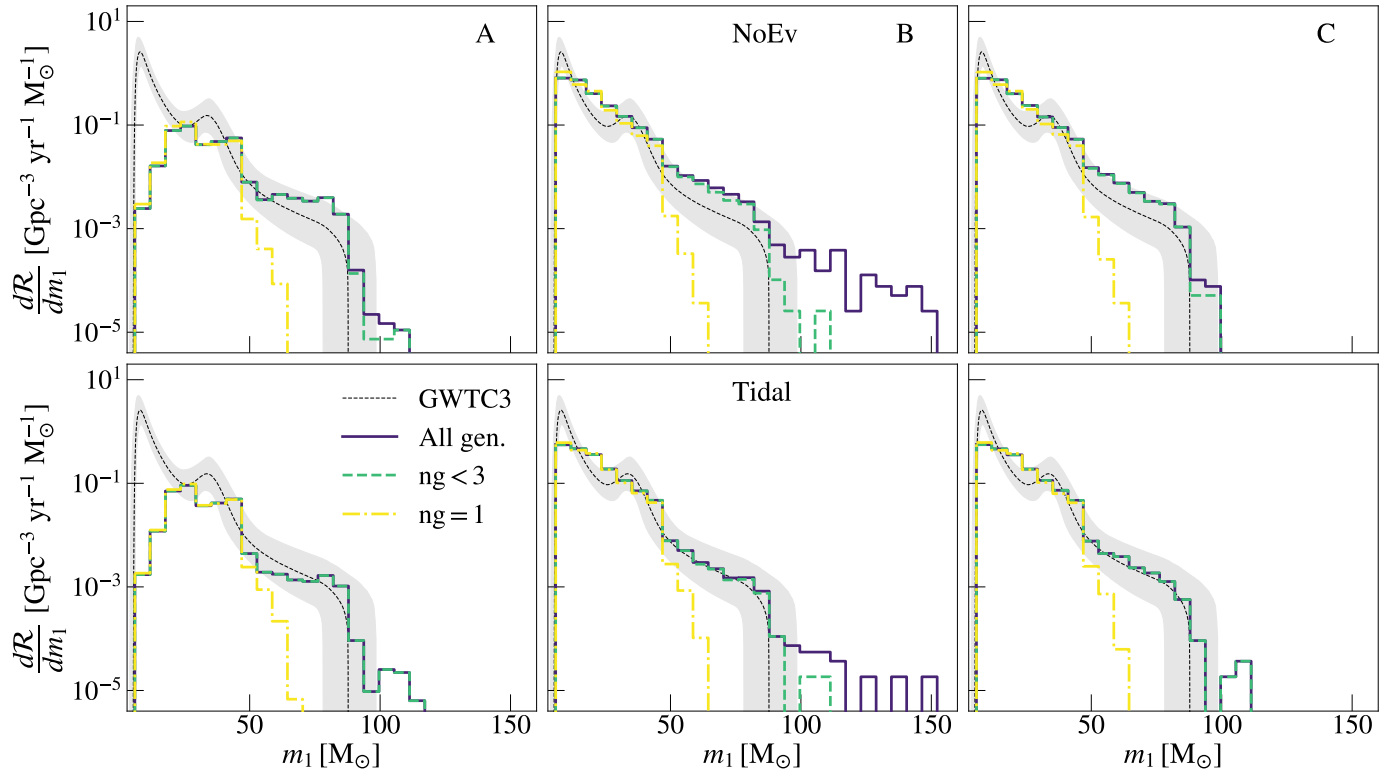
The BBH mass pairing criterion deeply affects the mass distribution of BBH mergers (Fig. 2). In our model A,  $m_1$  is skewed towards the maximum  $1g$  BH mass, and there are almost no mergers with  $m_1 \lesssim 15 M_\odot$ . In contrast, models B and C display a smoother primary mass distribution, extending down to the minimum BH mass in our models ( $5 M_\odot$ ).

Figure 3 compares the mass ratio distributions at different generations. At  $1g$ , all the mass samplings result in similar distributions, with a monotonic trend and a strong preference for  $q = 1$ . For higher generations, the mass ratio distribution displays significant differences. In model A, the  $ng$  secondary mass is likely to be  $m_2 \sim \max(m_{1,1g})$ , because this sampling favors the most massive BHs. As  $m_{1,2g} \sim 2 \max(m_{1,1g})$ , the  $2g$  mass ratio distribution peaks at  $\sim 0.5$ , and the  $3g$  distribution spreads around  $1/3$ . In model B, the mass ratios display similar distributions at all generations, because  $m_2$  is determined from Eq. (4). In model C, the distribution shows a flatter trend after the second generation, as  $m_{1,3g} \gtrsim \max(m_{1,1g})$  and mass ratios close to 1 are no longer possible.

We evaluate the impact of  $ng1g$  mergers on BH retention by comparing the fraction of ejected BH remnants for models B and C, where the BBH masses only differ for the secondary BH mass sampling at  $ng > 1$ . For model C, the fraction of BHs ejected by the GW relativistic recoil increases by 1% at  $2g$  and by 4% at  $3g$ , indicating that the  $ng1g$  assumption plays only a minor role in their retention.



**Fig. 3.** Mass ratio distributions of BBH mergers for models A (left), B (center), and C (right) in the Tidal evolutionary case. Different colors correspond to different BH generations: 1g (blue), 2g (purple), 3g (salmon), and >3g (yellow) BHs. Here, we show models with  $Z = 0.0002$ .



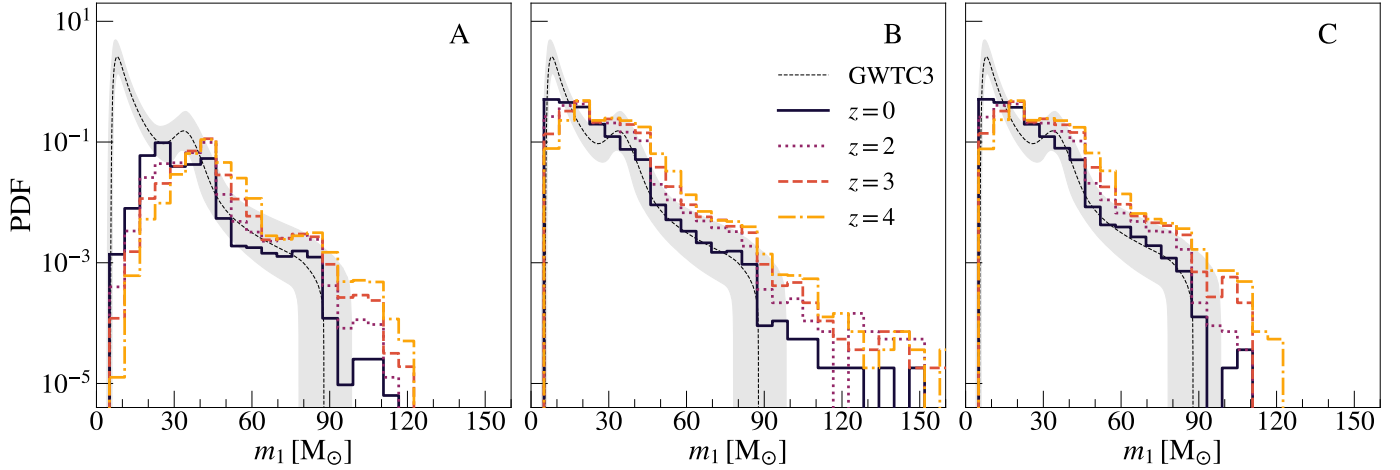
**Fig. 4.** Primary BH mass differential merger rate density in the Local Volume for models A (left), B (center), and C (right) for the NoEv (upper panels) and Tidal (lower panels) cases. We show the mass distribution resulting from 1g mergers (yellow dash-dotted line), 1g and 2g mergers (green dashed line), and from all BBH generations together (purple solid line). The black dashed line is the median value of the POWER LAW + PEAK model, as inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals. In this and the following figures, the black line and shaded gray areas are shown only for a qualitative comparison, because our simulations do not assume the parametric models adopted by the LVK.

### 3.3. Primary BH mass merger rate density in the Local Volume

Figure 4 shows the primary mass-differential-merger-rate density in the Local Volume. To obtain this quantity, we integrated the merger-rate density as described in Sect. 2.3, taking into account different progenitor metallicities. Overall, despite the large differences for low primary BH masses ( $m_1 \leq 20 M_\odot$ ), all of our models (A, B, C) reproduce both the peak at  $m_1 \approx 30\text{--}40 M_\odot$  and the high-mass tail ( $>50 M_\odot$ ) reported by the LVK Collaboration (Abbott et al. 2016).

In all of our models, the high-mass tail of the primary BH mass distribution ( $m_1 > 40 M_\odot$ ) is almost uniquely populated by hierarchical mergers. The host GC evolution quenches the contribution of hierarchical mergers to the merger-rate density, because it results in a less efficient hardening rate at late times.

The BH sampling criterion plays a key role in determining the primary mass distribution. All the models are able to account for the high-mass tail inferred by the LVK Collaboration (Abbott et al. 2023a,b), independently of the sampling criteria. At lower masses, instead, the considered sampling criteria differ dramatically: models B and C predict a larger BBH population



**Fig. 5.** Probability distribution function of primary BH masses at redshift  $z = 0$  (black solid line),  $z = 1$  (purple dotted line),  $z = 2$  (salmon dashed line), and  $z = 3$  (yellow dash-dotted line) for models A (left), B (center), and C (right) in the Tidal evolutionary case. In model B, the distribution extends to higher masses, but the plot is truncated at  $150 M_{\odot}$  to improve the readability of the figure. The black dashed line is the median value of the POWER LAW + PEAK model, as inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals.

at lower masses, with a main peak at  $\approx 10 M_{\odot}$  and a monotonically decreasing trend. Also, they approximately match the second peak at  $\approx 35 M_{\odot}$ .

In contrast, model A does not produce any peak at  $\approx 10 M_{\odot}$ . This model results in a main peak extending from  $\approx 25 M_{\odot}$  to  $\approx 50 M_{\odot}$ , which roughly encloses the  $35 M_{\odot}$  peak inferred from GW observations. Here, the peak is mostly due to 1g BBH mergers. A secondary peak due to 2g BBH mergers is present at  $\approx 70\text{--}80 M_{\odot}$ . These results are consistent with those from Antonini et al. (2023), who considered the same mass sampling criterion, but a power-law BH mass function.

This demonstrates the importance of quantifying the uncertainties related to the BH pairing function: according to models B and C, BBH mergers from GCs can account for a larger fraction of the LVK population, whereas model A suggests that GCs can produce only a subpopulation with a primary BH mass of  $\gtrsim 30 M_{\odot}$ .

### 3.4. Evolution of the primary BH mass with redshift

Figure 5 shows the dependence of the primary BH mass distribution on redshift. In all our models, the primary mass distribution evolves with redshift: BH mergers with  $m_1 < 20 M_{\odot}$  ( $m_1 > 30 M_{\odot}$ ) are progressively quenched (become more common) as redshift increases. The reason for this trend is the relative contribution of different progenitor-star metallicities at different redshifts (see Appendix C).

In model A, the main peak of the distribution shifts from  $\sim 35 M_{\odot}$  at  $z = 0$  to  $\sim 50 M_{\odot}$  at  $z = 3$ . As this mass sampling is skewed toward the maximum mass of the 1g BH spectrum, the larger contribution of subsolar metallicities for  $z > 0$  produces different peaks, depending on the dominant metallicity at the considered redshift. In models B and C, the primary BH mass distribution for  $z > 0$  is no longer monotonically decreasing, but shows a depletion at  $10 M_{\odot}$ , and a main peak at  $20 M_{\odot}$ .

Previous works suggest the existence of a weak trend (Fishbach et al. 2021), or even a strong correlation (Rinaldi et al. 2024) between primary BH mass and redshift. Our result indicates that dynamically assembled BBH mergers in GCs may account for such a trend. Finding a more top-heavy BH mass function at higher redshift ( $z \gtrsim 1$ ) would therefore strongly sug-

gest a dynamical origin of these BBH mergers. Next-generation ground-based GW detectors will be able to address this point (Reitze et al. 2019; Maggiore et al. 2020; Evans et al. 2021; Branchesi et al. 2023).

## 4. Discussion

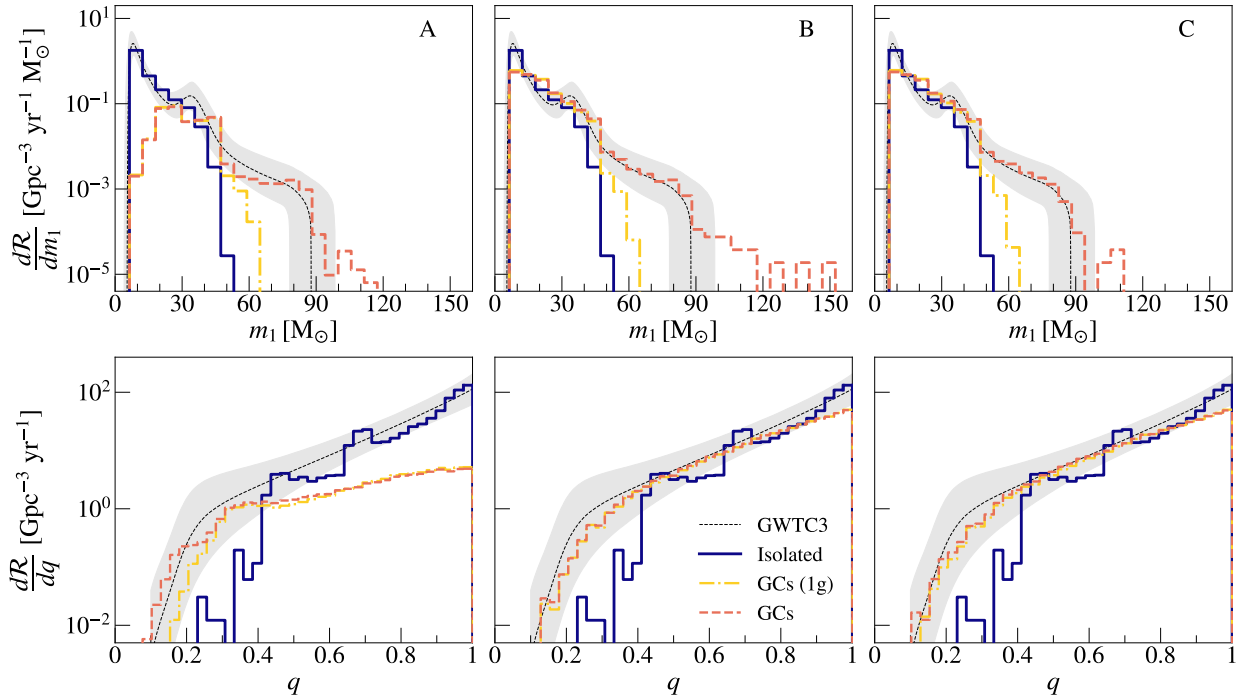
### 4.1. Comparison with isolated mergers

Here, we compare the primary mass, mass ratio, and spin distribution of BBHs in GCs to those predicted from isolated BBH mergers evolved with the same stellar evolution prescriptions as those used in our dynamical runs.

We take the isolated BBH mergers from a simulation run with MOBSE (Mapelli et al. 2017; Giacobbo et al. 2018). This simulation consists of  $1.8 \times 10^8$  binary systems comprising nine metallicities ( $Z = 0.0002, 0.0008, 0.002, 0.004, 0.006, 0.008, 0.012, 0.016, 0.02$ ). We drew the zero-age main sequence mass of the primary stars from a Kroupa (Kroupa 2001) initial mass function between  $5$  and  $150 M_{\odot}$ . We sampled the initial mass ratios, orbital eccentricity, and semi-major axis from the distribution by Sana et al. (2012), which are fits to observations of massive binary stars. We included core-collapse and pair-instability supernovae as described in Sect. 2.2.1 and we integrated each binary system accounting for wind mass transfer, Roche-lobe overflow, common envelope (assuming efficiency  $\alpha = 3$ ), tides, and GW decay as described in Giacobbo et al. (2018). In order to calculate the merger rate density of isolated BBHs, we used Eq. (8), assuming the cosmic star formation rate density  $\psi(z)$  and metallicity evolution from Madau & Fragos (2017). We refer to Santoliquido et al. (2020, 2021) for details.

#### 4.1.1. Primary mass and mass ratios

Figure 6 shows the primary mass and mass-ratio distribution of BBH mergers from GCs and isolated binaries. As expected, primary BH masses from isolated binaries peak at  $10 M_{\odot}$ , matching the main peak of the LVK population (Abbott et al. 2023b). Below  $\approx 40 M_{\odot}$ , the main difference between isolated BBHs and models B and C is the relative importance of the  $10 M_{\odot}$  peak,



**Fig. 6.** Primary BH mass (upper panels) and mass ratio (lower panels) differential merger rate density in the Local Volume for models A (left), B (center), and C (right) in the Tidal evolutionary case. We show the mass distribution resulting from isolated BBH mergers (blue solid line), 1g BBH mergers in GCs (yellow dash-dotted line), and all BBH mergers in GCs (including hierarchical mergers, salmon dashed line). The black dashed line is the median value of the POWER LAW + PEAK model inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals: these are shown only for a qualitative comparison, because our simulations do not assume the parametric models adopted by the LVK. We refer to Appendix D for a quantitative analysis.

which is a factor of about 3 higher in isolated binaries. In contrast, model A is dramatically different from isolated BBHs even at low primary BH masses.

For large primary masses ( $\geq 40 M_{\odot}$ ), all of our GC models differ from the isolated BBH model, as the latter cannot account for the high-mass tail, while the former produce high-mass BHs via hierarchical mergers. The distribution of mass ratios (lower panel of Fig. 6) shows that unequal-mass BBHs are more common in our dynamical BBHs with respect to isolated BBHs. The mass ratios from dynamical BBHs are qualitatively consistent with the lower end of the distribution inferred from the parametric model adopted by the LVK Collaboration (Abbott et al. 2023b). Dynamical mergers dominate the distribution at low  $q$  values. In contrast, the distribution from the isolated channel cannot explain the LVK rate for  $q < 0.5$ . In this case, the steep decrease in the mass ratio distribution at low  $q$  stems from common-envelope evolution, which is responsible for BBH hardening in our models and favors the formation of equal-mass binaries.

At high values of  $q$ , model A shows a flatter trend than both models B and C, and, for  $q \gtrsim 0.8$ , is about one order of magnitude below the median value inferred by the LVK Collaboration for the adopted parametric model (Abbott et al. 2023b). In contrast, models B and C display an increasing trend throughout the entire range in  $q$  and are always within the 90% credible interval of the LVK parametric model shown in the figure (Abbott et al. 2023b).

#### 4.1.2. Spins

For each BBH merger, we estimated the effective spin ( $\chi_{\text{eff}}$ ) and the precessing spin ( $\chi_p$ ):

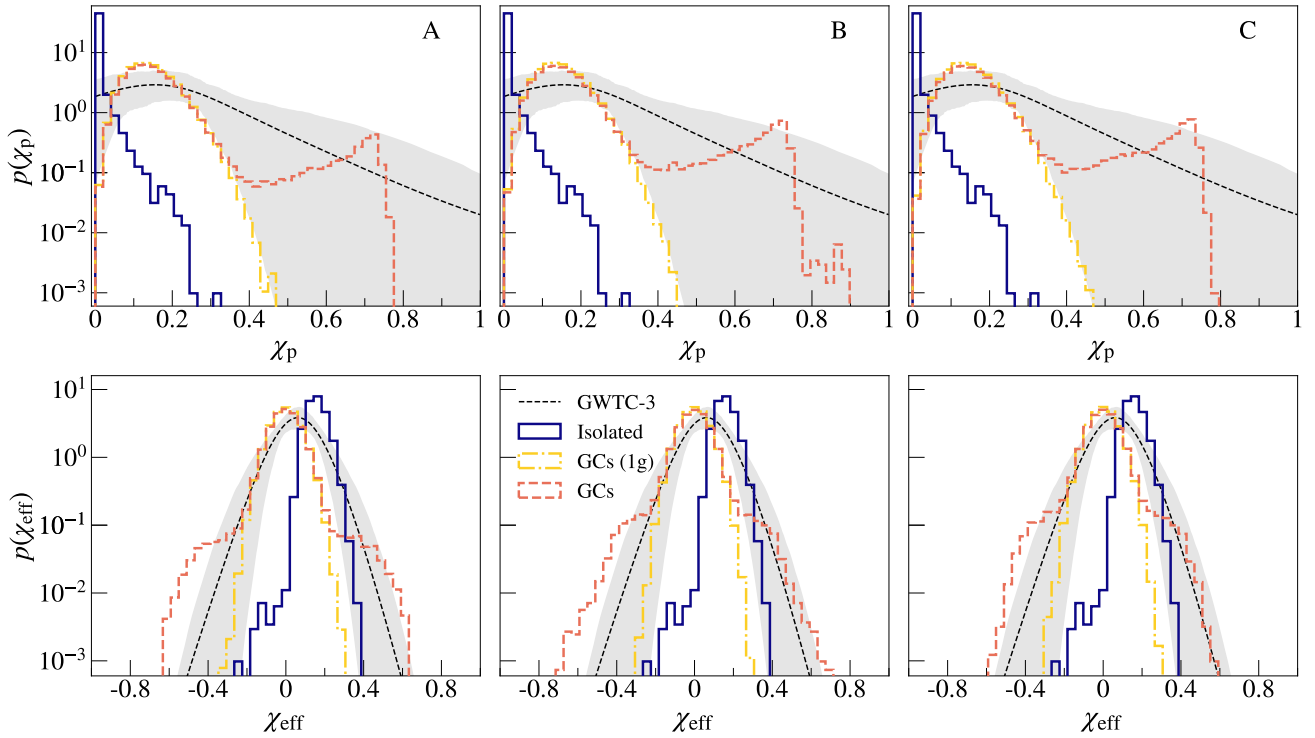
$$\chi_{\text{eff}} = \frac{(m_1 \chi_1 + m_2 \chi_2)}{m_1 + m_2} \cdot \frac{\mathbf{L}}{L},$$

$$\chi_p = \frac{c}{B_1 G m_1^2} \max(B_1 S_{1\perp}, B_2 S_{2\perp}), \quad (14)$$

where  $\mathbf{L}$  is the orbital angular momentum of the BBH,  $B_1 \equiv 2 + 3q/2$ , and  $B_2 \equiv 2 + 3/(2q)$ .  $S_{1\perp}$  and  $S_{2\perp}$  are the components of the spin vectors ( $\mathbf{S}_1$  and  $\mathbf{S}_2$ ) perpendicular to the orbital angular momentum.  $\chi_{\text{eff}}$  characterizes a mass-averaged spin angular momentum in the direction parallel to the binaries orbital angular momentum, while  $\chi_p$  approximately corresponds to the degree of in-plane spin, and traces the rate of relativistic precession of the orbital plane.

Figure 7 shows the distributions of  $\chi_{\text{eff}}$  and  $\chi_p$  from our models in the Tidal evolutionary case compared to the inferred distributions through the Gaussian spin model presented by Abbott et al. (2023b). The isolated channel mainly leads to the formation of low  $\chi_p$  and positive  $\chi_{\text{eff}}$ , because of angular momentum alignment driven by mass transfer and tidal effects. First-generation BBHs in GCs show a distribution of  $\chi_{\text{eff}}$  that is symmetric around zero due to the assumption of isotropic spins, and a peak at  $\chi_p \approx 0.15$ . The spread of the distributions of  $\chi_{\text{eff}}$  and  $\chi_p$  depend on the initial value of the parameter  $\sigma_{\chi}$  of the initial Maxwellian distribution, which, in our models is set to 0.1 (Sect. 2.2.1).

Black holes from previous mergers typically display larger spins, resulting from the sum of their parent BBH orbital angular momentum and BH spins. This produces a spin distribution peaked at  $\chi \sim 0.7$  (Gerosa & Berti 2017; Fishbach et al. 2017), and, in turn, a peak at  $\chi_p = 0.7$ . In model B, where the contribution of 3g mergers is not negligible, a tail at  $\chi_p \approx 0.8$  is present, which is due to the further spin increase in successive



**Fig. 7.** Precessing (upper panels) and effective (lower panels) spin distribution of BBH mergers at redshift  $z = 0$  for models A (left), B (center), and C (right) in the Tidal evolutionary case. We show the mass distribution resulting from isolated BBH mergers (blue solid line), 1g BBH mergers in GCs (yellow dash-dotted line), and BBH mergers in GCs, including hierarchical mergers (salmon dashed line). The black dashed line is the median value of the Gaussian spin model inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals.

generations. Also, the high spins produced by hierarchical mergers can extend the distribution of  $\chi_{\text{eff}}$  to much lower and higher values.

The population inferred from GWs is mostly consistent with slowly rotating BHs (Callister et al. 2022), but the decrease at high spins is milder than what is expected from 1g mergers only, according to our models. At the same time, high absolute values of  $\chi_{\text{eff}}$  from hierarchical mergers are not ruled out by current GW detections. Given the current uncertainties on spin parameters, additional GW detections will be crucial to constrain a possible secondary peak in the distribution of  $\chi_p$  due to hierarchical mergers.

#### 4.2. Mixing fractions

As a simple but instructive exercise, we calculated the mixing fraction between two channels: BBHs born in GCs and BBHs formed in isolated binary systems. For the latter, we used the models described in Sect. 4.1. Here, we neglected the other dynamical channels for simplicity, because a two-channel analysis allows a more straightforward interpretation of the impact of the main model features. We refer to Mapelli et al. (2022) for a more sophisticated but somewhat less transparent four-channel analysis.

We calculated the probability of recovering the parameters  $\theta = \{\mathcal{M}_{\text{chirp}}, q, z, \chi_{\text{eff}}, \chi_p\}$  of the observed LVK events<sup>3</sup> (Abbott et al. 2023b) given the mixing fraction  $f_{\text{GC}}$  and the other hyperparameters of our models  $\lambda$ :

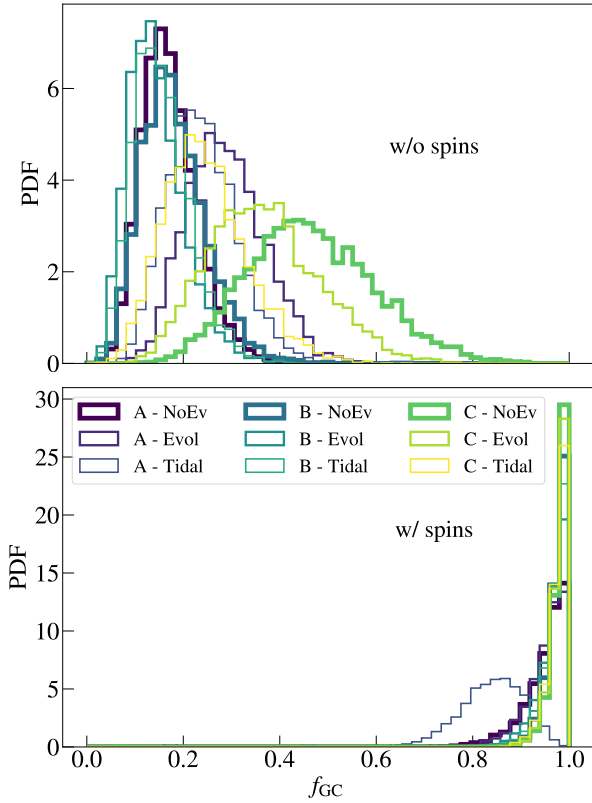
$$p(\theta|f_{\text{GC}}, \lambda) = f_{\text{GC}} p(\theta|\text{GC}, \lambda) + (1 - f_{\text{GC}}) p(\theta|\text{iso}, \lambda). \quad (15)$$

<sup>3</sup> In our formalism,  $\mathcal{M}_{\text{chirp}} = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$  is the BBH chirp mass.

According to this formalism, a mixing fraction  $f_{\text{GC}} = 0$  ( $f_{\text{GC}} = 1$ ) means that all BBHs are born from isolated binary systems (GCs). In our analysis, we considered 58 BBH events from Abbott et al. (2023a) with a probability of having an astrophysical origin of  $p_{\text{astro}} > 0.9$  and a false-alarm rate (FAR) of  $< 0.25 \text{ yr}^{-1}$ . We describe our hierarchical Bayesian framework in Appendix D. With this approach, we also take into account observational biases, associating a detection probability to our simulated catalogs. This allows us to make a fair comparison between our models and LVK data.

Figure 8 shows the probability distribution function (PDF) of the GC mixing fraction  $f_{\text{GC}}$  we obtain if we marginalize (upper panel) or do not marginalize over  $\chi_{\text{eff}}$  and  $\chi_p$  (lower panel). This figure shows a clear predominance of the GC channel if we include the spin parameters in the calculation of the mixing fraction, almost regardless of the GC model we assume. The main reason for this is that our isolated binary model tends to produce spins that are overly aligned with the orbital angular momentum in order to account for the majority of the LVK data (see Fig. 7 and Appendix D for details). This result confirms that the evolution of spin orientation during binary evolution is a key process (e.g., Stegmann & Antonini 2021; Callister et al. 2021b). As shown by Mapelli et al. (2022), this result is degenerate with the initial assumption on spin magnitudes. In particular, the choice of lower 1g spin magnitudes (e.g.,  $\sigma_\chi = 0.01$ ) can lead to a higher contribution of the isolated BBHs to the mixing fraction.

If we do not include  $\chi_{\text{eff}}$  and  $\chi_p$  in our analysis, that is if we marginalize over them, the GC mixing fraction  $f_{\text{GC}}$  decreases significantly, implying an important contribution of the isolated BBH channel. The contribution of the isolated BBHs when we neglect the spins mostly stems from the impact of  $m_1$ . In fact,



**Fig. 8.** Mixing fraction ( $f_{GC}$ ) probability distribution function. A mixing fraction  $f_{GC} = 0$  ( $f_{GC} = 1$ ) means that all BBHs are born from isolated binary systems (GCs). In the upper and lower panels, we do and do not marginalize over the spin parameters  $\chi_{\text{eff}}$  and  $\chi_p$  in our analysis, respectively. All the models presented in this paper are shown with different colors, as described in the legend.

our isolated binary models have a preference for  $m_1 \sim 8\text{--}10 M_\odot$ , which corresponds to the main peak in the LVK population (Fig. 6). In contrast, this peak is suppressed in all of our dynamical models, especially model A.

Model C – NoEv results in the largest GC mixing fraction among dynamical models if we neglect the spins (median value of  $f_{GC} \approx 0.4$  for model C – NoEv), whereas models A and B cluster at lower values of  $f_{GC}$ . This difference is more prone to stochastic fluctuations in our simulations. However, we note that model C – NoEv better reproduces the events with both a large value of chirp mass  $M_{\text{chirp}}$  and a low value of  $q$  in the LVK data (Appendix D). Overall, the differences between models Evol, NoEv, and Tidal are negligible in the analysis of the mixing fractions.

## 5. Summary

We investigated the process of hierarchical mergers in GCs, accounting for the main physical processes that drive GC evolution, namely mass loss by stellar evolution, expansion powered by a central BH subsystem, and tidal stripping. We also bracket the uncertainties related to BH pairing functions on the predicted primary BH mass, mass ratio, and spin distribution.

We summarize our main results as follows.

- Globular cluster expansion quenches the hierarchical BH assembly at the third generation. Our models predict a maximum BH mass of  $\sim 140 M_\odot$  if the mergers only involve first-

generation secondary BHs, and  $\sim 300 M_\odot$  if we include the possibility of ng secondary BHs.

- In all of our models, the hierarchical mergers dominate the primary BH mass distribution for  $m_1 > 50 M_\odot$ , and can reproduce the high-mass tail inferred from GW events.
- The adopted mass pairing function deeply affects the low-mass end of the primary BH mass distribution. Model A suppresses the formation of low-mass BBHs, resulting in a main peak around  $m_1 = 35 M_\odot$ . In contrast, models B and C produce a peak in the primary BH mass function at  $m_1 \sim 10 M_\odot$  and then a monotonic decrease for larger masses.
- The primary BH mass distribution evolves with redshift in all of our GC models, with a larger contribution from mergers with  $m_1 \geq 30 M_\odot$  at  $z \geq 2$ , while low-mass BBH mergers are quenched at high redshift. This effect stems from the interplay between GC dynamics and stellar metallicity evolution.
- Overall, despite the large differences for low primary BH masses ( $m_1 \leq 20 M_\odot$ ), all of our models (A, B, and C) match both the peak at  $m_1 \approx 30\text{--}40 M_\odot$  and the high-mass tail ( $>50 M_\odot$ ) reported by the LVK Collaboration (Abbott et al. 2023b). This confirms the importance of star-cluster dynamics and hierarchical mergers in producing the most massive BBH mergers observed with LVK.
- We compared the primary mass, mass ratio, and spin distribution of dynamical BBHs in GCs to those derived from isolated BBH mergers. Isolated binaries result in a three times greater contribution than models B and C at  $10 M_\odot$ , matching the main peak from LVK data. In contrast, dynamical mergers in GCs are necessary to reproduce the events with  $m_1 > 50 M_\odot$  and  $q < 0.5$ . Also, hierarchical mergers imprint a secondary peak in the  $\chi_p$  distribution and contribute to the positive and negative wings of the  $\chi_{\text{eff}}$  distribution.
- We calculated the mixing fraction between dynamical BBHs in GCs and isolated BBHs. Our predictions are very sensitive to the spin parameters, which favor a large fraction of dynamical BBH mergers in GCs ( $f_{GC} > 0.6$ ).

Our analysis confirms the importance of hierarchical mergers in matching the high-value tails of the primary BH mass ( $m_1 \geq 50 M_\odot$ ) and precession spin distributions ( $\chi_p \sim 0.7$ ). We predict that the mass of BBH mergers in GCs increases with redshift, regardless of the uncertainties related to the BBH pairing function. The ongoing fourth observing run of the LVK and next-generation GW detectors will be an important test-bed for our key predictions.

*Acknowledgements.* We thank the anonymous referee for their comments, which helped us to improve this manuscript. We thank Mark Gieles and Fabio Antonini for sharing CLUSTERBH and for productive discussions. M.M., S.T., C.P., M.D. and M.P.V. acknowledge financial support from the European Research Council for the ERC Consolidator grant DEMOBLACK, under contract no. 770017. S.T., M.M. and M.P.V. also acknowledge financial support from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181 – 390900948) STRUCTURES. M.D. also acknowledges financial support from Cariparo foundation under grant 55440. M.A.S. acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101025436 (project GRACE-BH). M.C.A. acknowledges financial support from the Seal of Excellence @UNIPD 2020 program under the ACROGAL project.

## References

- Aasi, J., Abbott, B. P., Abbott, R., et al. 2015, *Class. Quant. Grav.*, **32**, 074001  
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *ApJ*, **818**, L22  
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, *Phys. Rev. X*, **9**, 031040  
 Abbott, R., Abbott, T. D., Abraham, S., et al. 2020a, *ApJ*, **896**, L44

- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020b, *Phys. Rev. Lett.*, **125**, 101102
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020c, *ApJ*, **900**, L13
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, *Phys. Rev. X*, **11**, 021053
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, *ApJ*, **913**, L7
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, *Phys. Rev. X*, **13**, 041039
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023b, *Phys. Rev. X*, **13**, 011048
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2024, *Phys. Rev. D*, **109**, 022001
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *Class. Quant. Grav.*, **32**, 024001
- Antonini, F., & Gieles, M. 2020, *MNRAS*, **492**, 2936
- Antonini, F., & Rasio, F. A. 2016, *ApJ*, **831**, 187
- Antonini, F., Gieles, M., & Gualandris, A. 2019, *MNRAS*, **486**, 5008
- Antonini, F., Gieles, M., Dosopoulou, F., & Chattopadhyay, D. 2023, *MNRAS*, **522**, 466
- Arca Sedda, M., Mapelli, M., Benacquista, M., & Spera, M. 2023a, *MNRAS*, **520**, 5259
- Arca Sedda, M., Kamlah, A. W. H., Spurzem, R., et al. 2023b, *MNRAS*, **526**, 429
- Arca Sedda, M., Kamlah, A. W. H., Spurzem, R., et al. 2024a, *MNRAS*, **528**, 5119
- Arca Sedda, M., Kamlah, A. W. H., Spurzem, R., et al. 2024b, *MNRAS*, **528**, 5140
- Ballone, A., Costa, G., Mapelli, M., et al. 2023, *MNRAS*, **519**, 5191
- Bavera, S. S., Fragos, T., Qin, Y., et al. 2020, *A&A*, **635**, A97
- Belczynski, K., Ryu, T., Perna, R., et al. 2017, *MNRAS*, **471**, 4702
- Belczynski, K., Klencki, J., Fields, C. E., et al. 2020, *A&A*, **636**, A104
- Bianchini, P., van de Ven, G., Norris, M. A., Schinnerer, E., & Varri, A. L. 2016, *MNRAS*, **458**, 3644
- Bouffanais, Y., Mapelli, M., Gerosa, D., et al. 2019, *ApJ*, **886**, 25
- Bouffanais, Y., Mapelli, M., Santoliquido, F., et al. 2021, *MNRAS*, **507**, 5224
- Branchesi, M., Maggiore, M., Alonso, D., et al. 2023, *JCAP*, **2023**, 068
- Breen, P. G., & Hogg, D. C. 2013, *MNRAS*, **432**, 2779
- Briel, M. M., Stevance, H. F., & Eldridge, J. J. 2023, *MNRAS*, **520**, 5724
- Callister, T. A., & Farr, W. M. 2024, *Phys. Rev. X*, **14**, 021005
- Callister, T. A., Haster, C.-J., Ng, K. K. Y., Vitale, S., & Farr, W. M. 2021a, *ApJ*, **922**, L5
- Callister, T. A., Farr, W. M., & Renzo, M. 2021b, *ApJ*, **920**, 157
- Callister, T. A., Miller, S. J., Chatzioannou, K., & Farr, W. M. 2022, *ApJ*, **937**, L13
- Chattopadhyay, D., Stegmann, J., Antonini, F., Barber, J., & Romero-Shaw, I. M. 2023, *MNRAS*, **526**, 4908
- Costa, G., Bressan, A., Mapelli, M., et al. 2021, *MNRAS*, **501**, 4514
- Costa, G., Ballone, A., Mapelli, M., & Bressan, A. 2022, *MNRAS*, **516**, 1072
- Dehnen, W. 1993, *MNRAS*, **265**, 250
- Di Carlo, U. N., Giacobbo, N., Mapelli, M., et al. 2019, *MNRAS*, **487**, 2947
- Di Carlo, U. N., Mapelli, M., Bouffanais, Y., et al. 2020, *MNRAS*, **497**, 1043
- El-Badry, K., Quataert, E., Weisz, D. R., Choksi, N., & Boylan-Kolchin, M. 2019, *MNRAS*, **482**, 4528
- Evans, M., Adhikari, R. X., Afle, C., et al. 2021, arXiv e-prints [arXiv:2109.09882]
- Farah, A. M., Edelman, B., Zevin, M., et al. 2023, *ApJ*, **955**, 107
- Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, *ApJ*, **887**, 53
- Farmer, R., Renzo, M., de Mink, S. E., Fishbach, M., & Justham, S. 2020, *ApJ*, **902**, L36
- Farrell, E., Groh, J. H., Hirschi, R., et al. 2021, *MNRAS*, **502**, L40
- Fishbach, M., Holz, D. E., & Farr, B. 2017, *ApJ*, **840**, L24
- Fishbach, M., Holz, D. E., & Farr, W. M. 2018, *ApJ*, **863**, L41
- Fishbach, M., Doctor, Z., Callister, T., et al. 2021, *ApJ*, **912**, 98
- Fragione, G., & Rasio, F. A. 2023, *ApJ*, **951**, 129
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, **749**, 91
- Fuller, J., & Ma, L. 2019, *ApJ*, **881**, L1
- Georgiev, I. Y., Puzia, T. H., Hilker, M., & Goudfrooij, P. 2009, *MNRAS*, **392**, 879
- Gerosa, D., & Berti, E. 2017, *Phys. Rev. D*, **95**, 124046
- Gerosa, D., & Fishbach, M. 2021, *Nat. Astron.*, **5**, 749
- Giacobbo, N., Mapelli, M., & Spera, M. 2018, *MNRAS*, **474**, 2959
- Gieles, M., & Baumgardt, H. 2008, *MNRAS*, **389**, L28
- Gieles, M., Portegies Zwart, S. F., Baumgardt, H., et al. 2006, *MNRAS*, **371**, 793
- Gieles, M., Hogg, D. C., & Zhao, H. 2011, *MNRAS*, **413**, 2509
- González, J. A., Spherake, U., Brüggemann, B., Hannam, M., & Husa, S. 2007, *Phys. Rev. Lett.*, **98**, 091101
- Goodman, J. 1984, *ApJ*, **280**, 298
- Harris, W. E. 1996, *AJ*, **112**, 1487
- Harris, W. E., Harris, G. L. H., & Alessi, M. 2013, *ApJ*, **772**, 82
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, *ApJ*, **591**, 288
- Hogg, D. C. 1975, *MNRAS*, **173**, 729
- Hogg, D., & Hut, P. 2003, *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics* (Cambridge: Cambridge University Press)
- Hendriks, D. D., van Son, L. A. C., Renzo, M., Izzard, R. G., & Farmer, R. 2023, *MNRAS*, **526**, 4130
- Hénon, M. 1961, *Ann. Astrophys.*, **24**, 369
- Herrmann, F., Hinder, I., Shoemaker, D., & Laguna, P. 2007, *Class. Quant. Grav.*, **24**, S33
- Hurley, J. R., Pols, O. R., & Tout, C. A. 2000, *MNRAS*, **315**, 543
- Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, *MNRAS*, **329**, 897
- Jiménez-Forteza, X., Keitel, D., Husa, S., et al. 2017, *Phys. Rev. D*, **95**, 064024
- Kamlah, A. W. H., Leveque, A., Spurzem, R., et al. 2022, *MNRAS*, **511**, 4060
- Klencki, J., Moe, M., Gladysz, W., et al. 2018, *A&A*, **619**, A77
- Kremer, K., Spera, M., Becker, D., et al. 2020, *ApJ*, **903**, 45
- Kritos, K., Stokov, V., Baibhav, V., & Berti, E. 2022, arXiv e-prints [arXiv:2210.10055]
- Kritos, K., Berti, E., & Silk, J. 2023, *Phys. Rev. D*, **108**, 083012
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Loredo, T. J. 2004, *AIP Conf. Proc.*, **735**, 195
- Lousto, C. O., Zlochower, Y., Dotti, M., & Volonteri, M. 2012, *Phys. Rev. D*, **85**, 084015
- Madau, P., & Fragos, T. 2017, *ApJ*, **840**, 39
- Maggiore, M., Van Den Broeck, C., Bartolo, N., et al. 2020, *JCAP*, **2020**, 050
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, **486**, 1086
- Mapelli, M., Giacobbo, N., Ripamonti, E., & Spera, M. 2017, *MNRAS*, **472**, 2422
- Mapelli, M., Spera, M., Montanari, E., et al. 2020, *ApJ*, **888**, 76
- Mapelli, M., Dall'Amico, M., Bouffanais, Y., et al. 2021, *MNRAS*, **505**, 339
- Mapelli, M., Bouffanais, Y., Santoliquido, F., Arca Sedda, M., & Artale, M. C. 2022, *MNRAS*, **511**, 5797
- Marchant, P., & Moriya, T. J. 2020, *A&A*, **640**, L18
- Marchant, P., Renzo, M., Farmer, R., et al. 2019, *ApJ*, **882**, 36
- Miller, M. C., & Hamilton, D. P. 2002, *MNRAS*, **330**, 232
- Muratov, A. L., & Gnedin, O. Y. 2010, *ApJ*, **718**, 1266
- O'Leary, R. M., Rasio, F. A., Fregeau, J. M., Ivanova, N., & O'Shaughnessy, R. 2006, *ApJ*, **637**, 937
- O'Leary, R. M., Meiron, Y., & Kocsis, B. 2016, *ApJ*, **824**, L12
- Özel, F., & Freire, P. 2016, *ARA&A*, **54**, 401
- Pérgois, C., Mapelli, M., Santoliquido, F., Bouffanais, Y., & Rufolo, R. 2023, *Universe*, **9**, 507
- Planck Collaboration XIII. 2016, *A&A*, **594**, A13
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, **48**, 431
- Quinlan, G. D., & Shapiro, S. L. 1990, *ApJ*, **356**, 483
- Reina-Campos, M., Kruijssen, J. M. D., Pfeffer, J. L., Bastian, N., & Crain, R. A. 2019, *MNRAS*, **486**, 5838
- Reitze, D., Adhikari, R. X., Ballmer, S., et al. 2019, *BAAS*, **51**, 35
- Renzo, M., Cantiello, M., Metzger, B. D., & Jiang, Y. F. 2020, *ApJ*, **904**, L13
- Rinaldi, S., Del Pozzo, W., Mapelli, M., Lorenzo-Medina, A., & Dent, T. 2024, *A&A*, **684**, A204
- Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Science*, **337**, 444
- Santoliquido, F., Mapelli, M., Bouffanais, Y., et al. 2020, *ApJ*, **898**, 152
- Santoliquido, F., Mapelli, M., Giacobbo, N., Bouffanais, Y., & Artale, M. C. 2021, *MNRAS*, **502**, 4877
- Schnittman, J. D., & Buonanno, A. 2007, *ApJ*, **662**, L63
- Spera, M., & Mapelli, M. 2017, *MNRAS*, **470**, 4739
- Spera, M., Mapelli, M., & Jeffries, R. D. 2016, *MNRAS*, **460**, 317
- Spitzer, L. 1969, *ApJ*, **158**, L139
- Spitzer, L. 1987, *Dynamical Evolution of Globular Clusters* (Princeton: Princeton University Press)
- Spitzer, L., & Hart, M. H. 1971, *ApJ*, **164**, 399
- Stegmann, J., & Antonini, F. 2021, *Phys. Rev. D*, **103**, 063007
- Stevenson, S., Sampson, M., Powell, J., et al. 2019, *ApJ*, **882**, 121
- Torniamenti, S., Rastello, S., Mapelli, M., et al. 2022, *MNRAS*, **517**, 2953
- Torniamenti, S., Mapelli, M., Pérgois, C., et al. 2024, <https://doi.org/10.5281/zenodo.11367590>
- Vaccaro, M. P., Mapelli, M., Pérgois, C., et al. 2024, *A&A*, **685**, A51
- VandenBerg, D. A., Brogaard, K., Leaman, R., & Casagrande, L. 2013, *ApJ*, **775**, 134
- van Son, L. A. C., De Mink, S. E., Broekgaarden, F. S., et al. 2020, *ApJ*, **897**, 100
- Vesperini, E. 1998, *MNRAS*, **299**, 1019
- Vink, J. S., Higgins, E. R., Sander, A. A. C., & Sabhahit, G. N. 2021, *MNRAS*, **504**, 146
- Vitale, S., Lynch, R., Raymond, V., et al. 2017, *Phys. Rev. D*, **95**, 064053
- Woosley, S. E. 2017, *ApJ*, **836**, 244
- Woosley, S. E., & Heger, A. 2021, *ApJ*, **912**, L31

## Appendix A: Star cluster evolution

In the following, we summarize the description of GC evolution we included in our treatment. We use the model presented by Antonini & Gieles (2020) and the code CLUSTERBH<sup>4</sup>.

### A.1. Stellar mass loss

Stellar mass loss by stellar winds and supernova explosions, and the subsequent GC expansion are modeled as

$$\dot{M}_{\star, \text{sev}} = \begin{cases} 0 & t < t_{\text{sev}}, \\ -\nu \frac{M_{\star}}{t} & t \geq t_{\text{sev}}, \end{cases} \quad (\text{A.1})$$

$$\dot{r}_{\text{h, sev}} = -\frac{\dot{M}_{\star, \text{sev}}}{M_{\text{tot}}} r_{\text{h}}, \quad (\text{A.2})$$

where  $M_{\star}$  is the total stellar mass,  $t_{\text{sev}}$  is the timescale for stellar evolution,  $r_{\text{h}}$  is the half-mass radius, and  $\nu = 0.07$  (Antonini & Gieles 2020). We calibrate  $t_{\text{sev}}$  to match the trend from direct  $N$ -body simulations, as described in Sect. A.4.

### A.2. Two-body relaxation

We consider the onset of balanced evolution as the moment of the GC core collapse (Antonini & Gieles 2020):

$$t_{\text{cc}} = 3.2 t_{\text{rh}, \psi}, \quad (\text{A.3})$$

where  $t_{\text{rh}, \psi}$  is the half-mass relaxation timescale (Spitzer & Hart 1971):

$$t_{\text{rh}, \psi} = 0.138 \sqrt{\frac{M_{\text{tot}} r_{\text{h}}^3}{G \langle m \rangle \psi \ln \Lambda}}. \quad (\text{A.4})$$

Here,  $G$  is the gravitational constant,  $M_{\text{tot}}$  is the total cluster mass (both stars and black holes),  $\langle m \rangle$  is the mean mass within the cluster (including both stars and black holes), and  $\ln \Lambda = 10$  is the Coulomb logarithm. The parameter  $\psi$  is the heat conduction efficiency due to the mass function.

Following Spitzer & Hart (1971), for a two-component model (the two components being black holes and stars):

$$\psi = \left( m_{\star}^{3/2} M_{\star} + m_{\text{BH}}^{3/2} M_{\text{BH}} \right) N_{\text{tot}}^{3/2} / M_{\text{tot}}^{5/2}, \quad (\text{A.5})$$

where  $N_{\text{tot}}$  is the total number of stars within the cluster. In our runs, the mean stellar mass  $m_{\star} = 0.6 M_{\odot}$ , following a Kroupa (2001) initial mass function, while  $m_{\text{BH}}$  is the mean mass of first-generation BHs (see Sect. 2.2) and thus depends on the metallicity. We set the total BH mass,  $M_{\text{BH}}$ , based on the BH mass fraction from direct  $N$ -body models, as described in Sect. A.4.

When the GC reaches a state of balanced evolution, the expansion rate is given by (Antonini & Gieles 2020):

$$\dot{r}_{\text{h, rlx}} = \zeta \frac{r_{\text{h}}}{t_{\text{rh}, \psi}} + 2 \frac{\dot{M}_{\text{tot}}}{M_{\text{tot}}} r_{\text{h}}. \quad (\text{A.6})$$

Here,  $\dot{M}_{\text{tot}} = \dot{M}_{\star} + \dot{M}_{\text{BH}}$ , and  $\dot{M}_{\text{BH}}$  is the BH mass-loss rate due to dynamical ejections within the core (Breen & Heggie 2013):

$$\dot{M}_{\text{BH}} = \begin{cases} 0, & t < t_{\text{cc}} \text{ or } M_{\text{BH}} = 0 \\ -\beta \frac{M_{\text{tot}}}{t_{\text{rh}, \psi}} & t \geq t_{\text{cc}} \text{ and } M_{\text{BH}} > 0, \end{cases} \quad (\text{A.7})$$

where  $\beta = 3 \times 10^{-3}$  (Antonini & Gieles 2020).

<sup>4</sup> CLUSTERBH is available at <https://github.com/mgieles/clusterbh>

### A.3. Tidal mass loss

Tidal stripping from the host galaxy brings about additional stellar-mass loss to the cluster. Following Gieles et al. (2011), we model the tidal mass-loss rate from a point-mass galaxy under the assumption of circular orbits as:

$$\dot{M}_{\star, \text{tid}} = -\xi_{\text{J}} \frac{M_{\text{tot}}}{t_{\text{rh}, \psi=1}}. \quad (\text{A.8})$$

Here, we set the tidal mass loss timescale to the relaxation time for  $\psi = 1$  (Gieles et al. 2011), because this process takes place at the tidal boundary, where almost no BHs are present. The parameter  $\xi_{\text{J}}$  embodies the impact of different tidally filling regimes on the cluster mass loss:

$$\xi_{\text{J}} = \frac{3}{5} \zeta \left( \frac{r_{\text{h}}/r_{\text{J}}}{[r_{\text{h}}/r_{\text{J}}]_1} \right)^{3/2}. \quad (\text{A.9})$$

Here,  $[r_{\text{h}}/r_{\text{J}}]_1$  is defined as in Gieles et al. (2011) and  $r_{\text{J}}$  is the Jacobi (tidal) radius (Gieles & Baumgardt 2008):

$$r_{\text{J}} = \left( \frac{G}{3\omega^2} \right)^{1/3} M_{\text{tot}}^{1/3}, \quad (\text{A.10})$$

where  $\omega = V_{\text{C}} R_{\text{G}}^{-1}$  depends on the cluster galactocentric distance  $R_{\text{G}}$  and its circular velocity  $V_{\text{C}}$  (Gieles & Baumgardt 2008). For our runs, we consider  $R_{\text{G}} = 8 \text{ kpc}$  and  $V_{\text{C}} = 220 \text{ km s}^{-1}$ , which correspond to a solar-neighborhood-like static tidal field.

The combined stellar mass loss, including stellar evolution and tidal stripping, is:

$$\dot{M}_{\star} = \begin{cases} \dot{M}_{\star, \text{tid}} & t < t_{\text{sev}}, \\ \dot{M}_{\star, \text{tid}} + \dot{M}_{\star, \text{sev}} & t \geq t_{\text{sev}}. \end{cases} \quad (\text{A.11})$$

We also introduced the possibility to model the galactic rotational curve at different galactocentric distances, based on a Dehnen (1993) profile:

$$V_{\text{C}} = G M_{\text{g}} \frac{R_{\text{G}}^{2-\gamma}}{(R_{\text{G}} + r_{\text{s}})^{3-\gamma}}, \quad (\text{A.12})$$

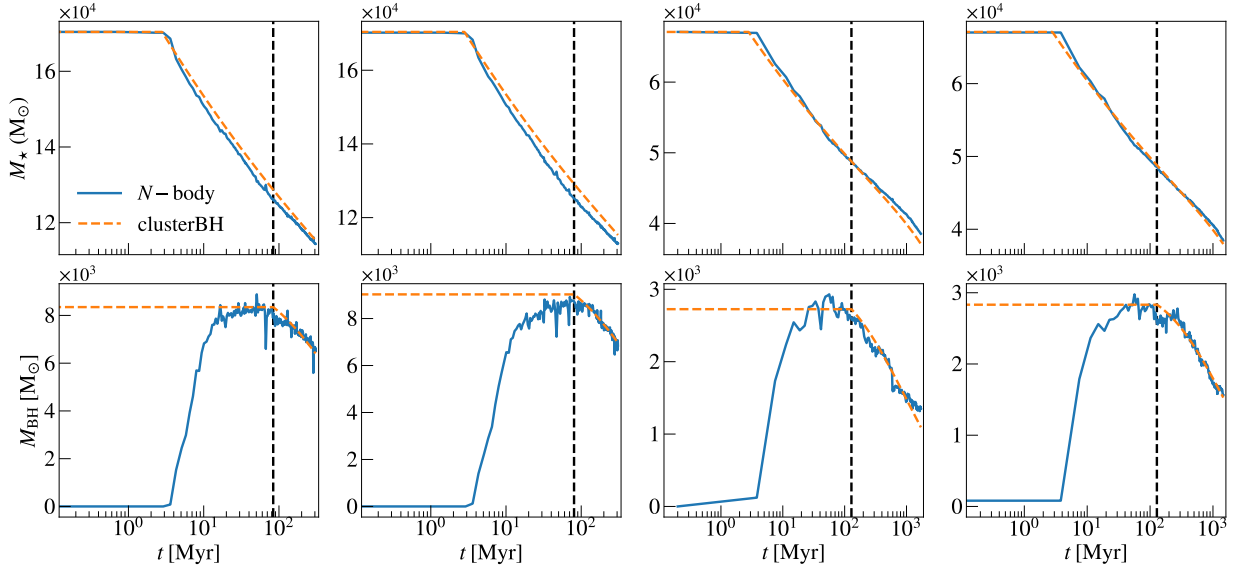
where  $M_{\text{g}}$  is the galaxy total mass and  $r_{\text{s}}$  is the galaxy length scale. Although this feature was not considered for this work, it will be a useful tool for investigating the evolution of GC populations in galaxies.

### A.4. Calibration of $t_{\text{sev}}$ and $M_{\text{BH}}$

The stellar evolution timescale,  $t_{\text{sev}}$ , and the BH mass within the cluster,  $M_{\text{BH}}$ , depend on the details of stellar evolution and BH ejection by supernova kicks, respectively. Here, we tuned these parameters to reproduce the evolution of the stellar and BH mass of the direct  $N$ -body simulations by Arca Sedda et al. (2024a,b, 2023b), following a similar approach to Antonini & Gieles (2020).

As shown in Figure A.1, our semi-analytic implementation can trace the trend of stellar mass loss. By setting  $t_{\text{sev}} = 3 \text{ Myr}$ , we capture the early change in the stellar mass slope due to the onset of supernova explosions. For this reason, in the following we consider  $t_{\text{sev}} = 3 \text{ Myr}$ .

In the  $N$ -body models, the total BH mass within the half-mass radius initially increases because the most massive stars progressively give birth to BHs, and peaks at  $t_{\text{cc}}$ . Subsequently, the total BH mass decreases because of dynamical ejections within the BH subsystem. For this comparison, we set the initial  $M_{\text{BH}}$  to the value given by the  $N$ -body simulations at  $t_{\text{cc}}$ , which ranges from 0.042 to 0.05. In our runs, we assume an initial BH mass fraction of 0.045.



**Fig. A.1.** Evolution of  $M_*$  (top) and  $M_{\text{BH}}$  (bottom) in the  $N$ -body simulations (blue, solid line) presented in [Arca Sedda et al. \(2024a\)](#) and in this work (orange dashed line). Different columns show the evolution of some representative  $N$ -body simulations. The two left-hand panels show the evolution of a GC with  $N = 3 \times 10^5$  stars, while the two right-hand panels show the evolution of GCs with  $N = 1.2 \times 10^5$  stars. The black vertical dashed line is the time of the GC core collapse.

#### A.5. Coupling between CLUSTERBH and FASTCLUSTER

In this section, we describe how we interface the two semi-analytic codes CLUSTERBH ([Antonini & Gieles 2020](#)) and FASTCLUSTER ([Mapelli et al. 2021](#)). First, we generate the GC masses and densities and calculate the time of core collapse (eq. A.3). We draw the 1g BBH masses, spins and orbital properties as described in Sect. 2.2.1. For each metallicity, we set in both FASTCLUSTER and CLUSTERBH the value of  $m_{\text{BH}}$  to the mean BH mass in the MOBSE catalog.

We evolve the host GC until the time of core collapse, if BHs are not ejected by their natal kick. To ensure efficiency along with accuracy, we implemented a Runge-Kutta fourth order integrator and an adaptive time-step in CLUSTERBH. Specifically, we increase (reduce) the time-step by a factor of 2 (10) if the percentage change of the integrated quantities ( $M_*$ ,  $M_{\text{BH}}$  or  $r_{\text{hm}}$ ) between two time-steps falls below 0.1% (exceeds 1%). After BBH formation, we integrate the GC mass, radius, density and velocity dispersion, along with the BBH eccentricity and semi-major axis (see eq. 16 in [Mapelli et al. 2021](#)). Since the evolution timescales of the BBH and the host GC generally differ by several orders of magnitude, we implement two separate adaptive time-step schemes for integrating the BBH hardening and the GC evolution, respectively.

We stop GC evolution if the BH is ejected from the cluster (e.g. as a consequence of three-body interactions or relativistic recoil) or when the integration reaches the Hubble time. If the GC evaporates, or no BHs are retained within the cluster, we continue the BBH evolution as an isolated binary, as described in [Mapelli et al. \(2021\)](#). Our implementation delivers an efficient integration of both the GC and the BBH properties, and enables

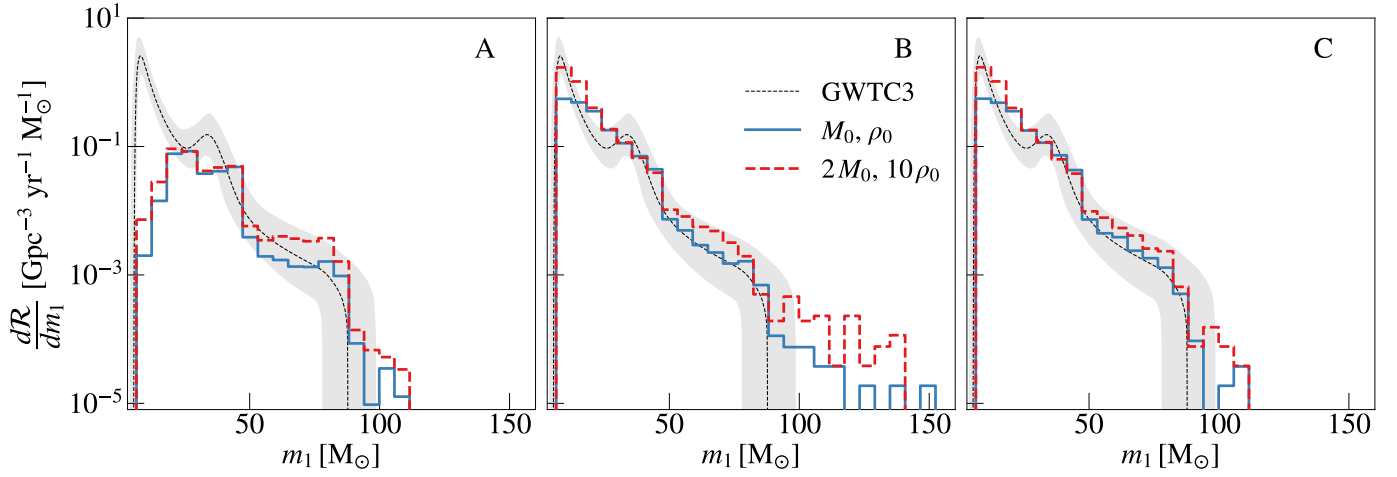
us to evolve  $\sim 3 \times 10^5$  BBHs in GCs per day with a single core.

## Appendix B: Impact of the initial GC mass and density on our results

The initial GC mass and density distributions we assumed in the main text (Sect. 2.1) match the observed properties of current Milky Way GCs ([Harris et al. 2013](#); [VandenBerg et al. 2013](#)). However, GCs might have been more massive and denser at their formation time (e.g., [Vesperini 1998](#)). In the following, we estimate the impact of GC initial conditions on the BBH merger rate density by considering larger initial GC masses and densities. In particular, we assume that the initial median value of the GC mass (density) distribution is two (ten) times as large as the values considered in Sect. 2.1.

Figure B.1 shows that larger GC masses and densities have a mild impact on the primary BH mass. As expected, denser, and more massive clusters produce a higher BBH merger rate density, because of a higher BH retention fraction. In models B and C, the merger rate of low-mass BHs is enhanced, because supernova kicks are less efficient in ejecting them from more massive GCs. In contrast, the mass sampling presented in model A is only marginally affected by this effect, because only massive BHs tend to pair-up efficiently with this mass sampling.

The longer cluster lifetime and the shorter evolution timescales also enhance the production of hierarchical mergers. However, also in this case we find a minor contribution from  $n_{\text{g}} \geq 3$  mergers at redshift  $z = 0$ , especially for the  $n_{\text{g}}1_{\text{g}}$  cases (models A and C).



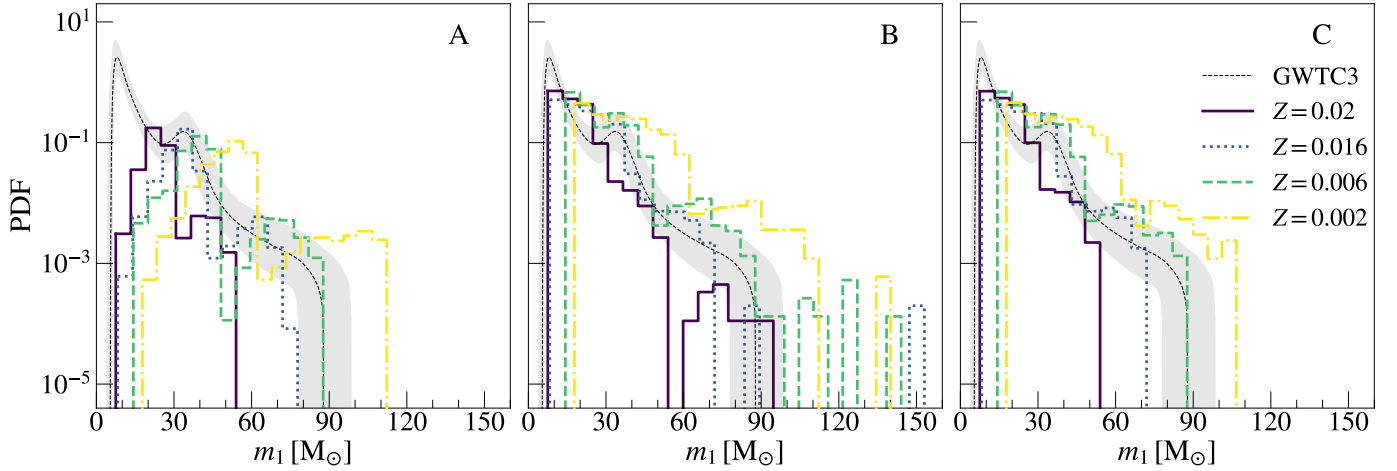
**Fig. B.1.** Primary BH mass merger rate density for models A (left), B (center), and C (right) in the Tidal evolutionary case. The GC masses and densities are initialized as described in Sect. 2.1, ( $M_0, \rho_0$ , blue solid line), and with mass (density) distributions that are shifted to larger values by a factor 2 (10, red dashed line). The black dashed line is the median value of the POWER LAW + PEAK model inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals.

### Appendix C: Hierarchical BH mass function as a function of metallicity

Figure C.1 shows the primary BH mass distribution produced by BBH mergers at different metallicities. In model A, the main peak shifts with the maximum mass of first-generation BHs. Also, for each metallicity, this pairing function predicts a secondary peak due to hierarchical mergers at  $\sim 2 \max(m_{1,1g})$ .

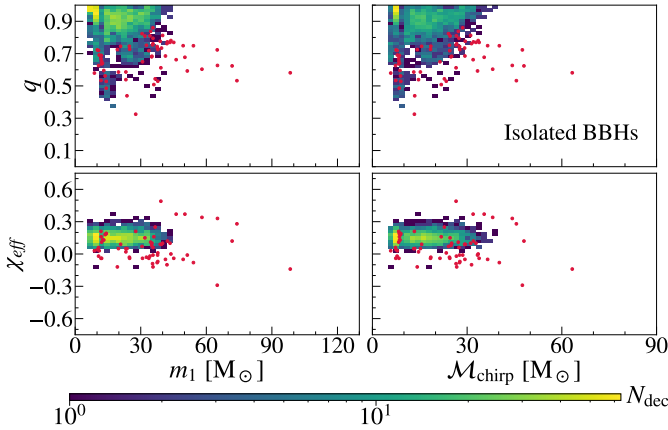
In models B and C, the main peak at  $10 M_{\odot}$  is mainly due to BBH mergers at solar ( $Z = 0.02$ ) and slightly subsolar metallic-

ity ( $Z = 0.016$ ), while lower metallicities extend the mass distribution toward larger masses. As discussed in Sect. 3.4, the relative contribution of different metallicities results in a variation of the BH primary mass distribution with redshift. In models B and C, primary BH masses  $m_1 < 20 M_{\odot}$  become less and less common at  $z \geq 2$ , because metal-rich progenitors ( $Z = 0.02, 0.016$ ) also become less frequent at high redshift. In model A, the main peak of the primary BH mass shifts from  $\sim 35 M_{\odot}$  to  $\sim 50 M_{\odot}$  as a function of redshift, because metal-poor stars become more and more common at high redshift.

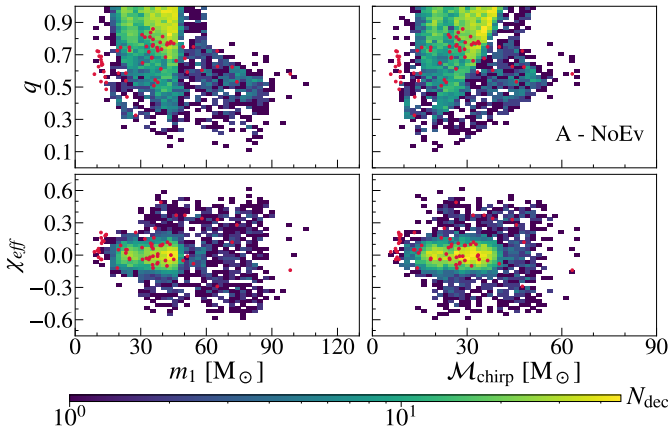


**Fig. C.1.** Probability distribution function of primary BH masses in GCs, for models A (left), B (center), and C (right) in the Tidal evolutionary case. We consider different progenitor metallicities:  $Z = 0.02$  (purple solid line),  $Z = 0.016$  (blue dotted line),  $Z = 0.006$  (green dashed line), and  $Z = 0.002$  (yellow dash-dotted line). The black dashed line is the median value of the POWER LAW + PEAK model inferred from GWTC-3 (Abbott et al. 2023b). The gray shaded areas are the corresponding 90% credible intervals.

## Appendix D: Bayesian hierarchical analysis



**Fig. D.1.** Two dimensional histograms of the detectable simulated BBHs in the isolated binary evolution model. The four panels show, from left to right and from top to bottom,  $q - m_1$ ,  $q - \mathcal{M}_{\text{chirp}}$ ,  $\chi_{\text{eff}} - m_1$ , and  $\chi_{\text{eff}} - \mathcal{M}_{\text{chirp}}$ . The red points are the data from [Abbott et al. \(2023b\)](#). We do not add error bars in order to avoid overcrowding the plot.



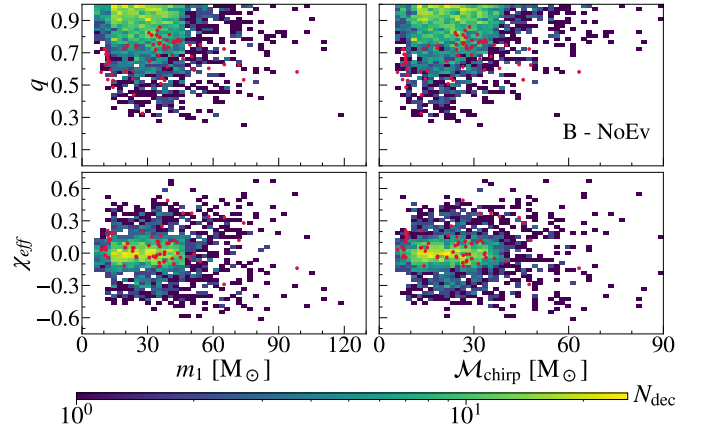
**Fig. D.2.** Same as Figure D.1 but for model A NoEv.

To compare our models against GW events in the third-GW transient catalog, we use a hierarchical Bayesian approach, as we already described in previous work ([Bouffanais et al. 2019, 2021; Mapelli et al. 2022](#)). Given a number  $N_{\text{obs}}$  of GW observations,  $\mathcal{H} = \{h^k\}_{k=1}^{N_{\text{obs}}}$ , described by an ensemble of parameters  $\theta$ , the posterior distribution of the hyper-parameters  $\lambda$  associated with the models is described as an inhomogeneous Poisson distribution ([Loredo 2004; Mandel et al. 2019](#)):

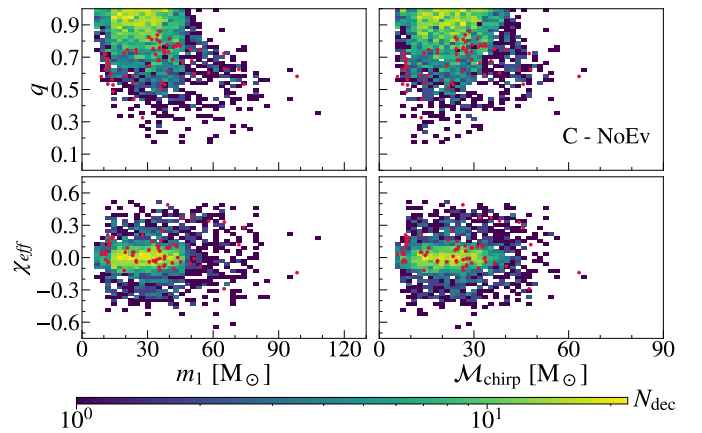
$$p(\lambda, N_\lambda | \mathcal{H}) = e^{-\mu_\lambda} \pi(\lambda, N_\lambda) \prod_{k=1}^{N_{\text{obs}}} N_\lambda \int_{\theta} \mathcal{L}^k(h^k | \theta) p(\theta | \lambda) d\theta, \quad (\text{D.1})$$

where  $\theta$  are the GW parameters,  $N_\lambda$  is the number of events predicted by the astrophysical model,  $\mu_\lambda$  is the predicted number of detections associated with the model and the GW detector,  $\pi(\lambda, N_\lambda)$  is the prior distribution on  $\lambda$  and  $N_\lambda$ , and  $\mathcal{L}^k(\{h\}^k | \theta)$  is the likelihood of the  $k$ th detection. The predicted number of detections is given by  $\mu(\lambda) = N_\lambda \beta(\lambda)$ , where

$$\beta(\lambda) = \int_{\theta} p(\theta | \lambda) p_{\text{det}}(\theta) d\theta \quad (\text{D.2})$$



**Fig. D.3.** Same as Figure D.1 but for model B NoEv.



**Fig. D.4.** Same as Figure D.1 but for model C NoEv.

is the detection efficiency of the model. In eq. D.2,  $p_{\text{det}}(\theta)$  is the probability of detecting a source with parameters  $\theta$  and can be inferred by computing the optimal signal-to-noise ratio and comparing it to a detection threshold, as described, e.g., in [Bouffanais et al. \(2021\)](#).

The values for the event's log-likelihood are derived from the posterior and prior samples released by the LVC, such that the integral in eq. D.1 is approximated with a Monte Carlo approach as

$$\mathcal{I}^k = \int_{\theta} \mathcal{L}^k(h^k | \theta) p(\theta | \lambda) d\theta \sim \frac{1}{N_s^k} \sum_{i=1}^{N_s^k} \frac{p(\theta_i^k | \lambda)}{\pi^k(\theta_i^k)}, \quad (\text{D.3})$$

where  $\theta_i^k$  is the  $i$ th posterior sample for the  $k$ th detection and  $N_s^k$  is the total number of posterior samples for the  $k$ th detection. Both the model and prior distributions are estimated with Gaussian kernel density estimation. We further marginalise eq. D.1 over  $N_\lambda$  using a prior  $\pi(N_\lambda) \sim 1/N_\lambda$  ([Fishbach et al. 2018](#)), which yields the following expression:

$$p(\lambda | \mathcal{H}) \sim \pi(\lambda) \prod_{k=1}^{N_{\text{obs}}} \frac{\mathcal{I}^k}{\beta(\lambda)}. \quad (\text{D.4})$$

Based on the above calculations, we can estimate the mixing fraction as described in eq. 15. Figures D.1, D.2, D.3, and D.4 show the behavior of four selected models (isolated binaries, and models A, B, and C in the NoEv case, respectively) in the planes

defined by the four parameters  $q$ ,  $\chi_{\text{eff}}$ ,  $m_1$ , and  $\mathcal{M}_{\text{chirp}}$ . In the histograms, we do not show all possible simulated BBHs, but rather only the detectable ones (i.e., with a  $S/N^5$  of above 8 assuming O3b sensitivity). From these figures, we see that BBHs from isolated binary systems tend to be associated with nearly aligned spins, while our dynamical models have a broader distribution

of  $\chi_{\text{eff}}$ , which is favored by the LVK data. This explains the large values of the mixing fraction  $f_{\text{GC}}$  we obtain when we include the impact of  $\chi_{\text{eff}}$  into our hierarchical Bayesian analysis. From these figures, we also see that isolated BBHs favor a strong peak at  $m_1 \sim 8 - 10 M_{\odot}$ , similar to the main peak reported by the LVK Collaboration.

---

<sup>5</sup>  $S/N^2 = 4 \int_{f_{\text{min}}}^{f_{\text{max}}} \frac{\tilde{h}^2(f)}{S_n(f)} df$ , where  $\tilde{h}$  is the waveform and  $S_n$  the power spectral noise density.