The mass-ratio distribution of tertiary-induced binary black hole mergers

Su, Yubo; Liu, Bin; Lai, Dong

Published in: Monthly Notices of the Royal Astronomical Society

DOI: 10.1093/mnras/stab1617

Publication date: 2021

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
The mass-ratio distribution of tertiary-induced binary black hole mergers

Yubo Su, Bin Liu and Dong Lai

1Cornell Center for Astrophysics and Planetary Science, Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
2Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
3Tsung-Dao Lee Institute and School of Physics and Astronomy, Shanghai Jiao Tong University, 200240 Shanghai, China

ABSTRACT

Many proposed scenarios for black hole (BH) mergers involve a tertiary companion that induces von Zeipel–Lidov–Kozai (ZLK) eccentricity cycles in the inner binary. An attractive feature of such mechanisms is the enhanced merger probability when the octupole-order effects, also known as the eccentric Kozai mechanism, are important. This can be the case when the tertiary is of comparable mass to the binary components. Since the octupole strength $\mu(1-q)/(1+q)$ increases with decreasing binary mass ratio $q$, such ZLK-induced mergers favour binaries with smaller mass ratios. We use a combination of numerical and analytical approaches to fully characterize the octupole-enhanced binary BH mergers and provide semi-analytical criteria for efficiently calculating the strength of this enhancement. We show that for hierarchical triples with semimajor axial ratio $a_{\text{out}} \gtrsim 0.01–0.02$, the binary merger fraction can increase by a large factor (up to $\sim 20$) as $q$ decreases from unity to 0.2. The resulting mass-ratio distribution for merging binary BHs produced in this scenario is in tension with the observed distribution obtained by the LIGO/VIRGO collaboration, although significant uncertainties remain about the initial distribution of binary BH masses and mass ratios.

Key words: binaries: close – stars: black holes – black hole mergers.

1 INTRODUCTION

The 50 or so black hole (BH) binary mergers detected by the LIGO/VIRGO collaboration to date (Abbott et al. 2020b) continue to motivate theoretical studies of their formation channels. These range from the traditional isolated binary evolution, in which mass transfer and friction in the common-envelope phase cause the binary orbit to decay sufficiently that it subsequently merges via emission of gravitational waves (GWs; e.g. Lipunov, Postnov & Prokhorov 1997; Podsiadlowski, Rappaport & Han 2003; Belczynski et al. 2010, 2016; Dominik et al. 2012, 2013, 2015; Lipunov et al. 2017), to various flavours of dynamical formation channels that involve either strong gravitational scatterings in dense clusters (e.g. Portegies Zwart & McMillan 2000; O’Leary et al. 2006; Miller & Lauberg 2009; Banerjee, Baumgardt & Kroupa 2010; Downing et al. 2010; Ziosi et al. 2014; Rodriguez et al. 2015; Samsing & Ramirez-Ruiz 2017; Gondán et al. 2018; Rodriguez et al. 2018; Samsing & D’Orazio 2018) or mergers in isolated triple and quadruple systems induced by distant companions (e.g. Miller & Hamilton 2002; Wen 2003; Antonini & Perets 2012; Antonini, Toonen & Hamers 2017; Liu & Lai 2017, 2018, 2019, 2020, 2021; Silsbee & Tremaine 2017; Hoang et al. 2018; Randall & Xiang 2018a,b; Fragione & Kocsis 2019; Fragione & Loeb 2019; Liu, Lai & Wang 2019a,b).

Given the large number of merger events to be detected in the coming years, it is important to search for observational signatures to distinguish various BH binary formation channels. The masses of merging BHs obviously carry important information. The recent detection of BH binary systems with component masses in the mass gap (such in GW190521) suggests that some kinds of ‘hierarchical mergers’ may be needed to explain these exceptional events (Abbott et al. 2020a; see Liu & Lai 2021 for examples of such ‘hierarchical mergers’ in stellar multiples). Another possible indicator is merger eccentricity: Previous studies find that dynamical binary-single interactions in dense clusters (e.g. Samsing & Ramirez-Ruiz 2017; Rodríguez et al. 2018; Samsing & D’Orazio 2018; Fragione & Bromberg 2019) or in galactic triples (Antonini et al. 2017; Fragione & Loeb 2019; Liu et al. 2019a; Silsbee & Tremaine 2017) may lead to BH binaries that enter the LIGO band with modest eccentricities. The third possible indicator is the spin-orbit misalignment of the binary. In particular, the mass-weighted projection of the BH spins,

$$x_{\text{eff}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \cdot \hat{L},$$

(1)

can be measured through the binary inspiral waveform (here, $m_{1,2}$ is the BH mass, $x_{1,2} = cS_{1,2}/(Gm_{1,2}^2)$ is the dimensionless BH spin, and $\hat{L}$ is the unit orbital angular momentum vector of the binary). Different formation histories yield different distributions of $x_{\text{eff}}$ (Liu & Lai 2017, 2018; Antonini et al. 2018; Gerosa et al. 2018; Rodríguez et al. 2018; Liu et al. 2019a; Su, Lai & Liu 2020).

The fourth possible indicator of BH binary formation mechanisms is the distribution of masses and mass ratios of merging BHs. In Fig. 1, we show the distribution of the mass ratio $q = m_2/m_1$, where $m_1 \geq m_2$, for all LIGO/VIRGO binaries detected as of the O3a
et al. 2020). On the other hand, dynamical formation channels may produce a larger variety of distributions for the binary mass ratio \( q \) (Liu & Lai 2018). As excitation, and thus enhance the rate of successful binary mergers generally increase the inclination window for extreme eccentricity (Kozinsky & Rasio 2000; Blaes, Lee & Socrates 2002; Lithwick & Naoz 2011; Liu, Munoz & Lai 2015). The strength of the octupole effect depends on the dimensionless parameter \( \epsilon \)

\[
\epsilon = \frac{m_1 - m_2}{m_1 + m_2} \frac{a}{a_{\text{out}}} - 1
\]

where \( a \) and \( a_{\text{out}} \) are the semimajor axes of the inner and outer binaries, respectively. Previous studies have shown that the octupole terms generally increase the inclination window for extreme eccentricity excitation, and thus enhance the rate of successful binary mergers (Liu & Lai 2018). As \( \epsilon_{\text{out}} \propto (1 - q)(1 + q) \) increases with decreasing \( q \), we expect that ZLK-induced BH mergers favour binaries with smaller mass ratios. The main goal of this paper is to quantify the dependence of the merger fraction/probability on \( q \), using a combination of analytical and numerical calculations. We focus on the cases where the tertiary mass is comparable to the binary BH masses. When the tertiary mass \( m_3 \) is much larger than \( m_{12} = m_1 + m_2 \) (as in the case of a supermassive BH tertiary), dynamical stability of the triple requires \( a_{\text{out}}(1 - e_{\text{out}})/(1 + e) \gtrsim 3.7(m_{12}/m)_{\text{12}}^{1/3} \gtrsim 1 \) (Kiseleva et al. 1996), which implies that the octupole effect is negligible.

This paper is organized as follows. In Section 2, we review some analytical results of ZLK oscillations and examine how the octupole terms affect the inclination window and probability for extreme eccentricity excitation. In Section 3, we study tertiary-induced BH mergers using a combination of numerical and analytical approaches. We propose new semi-analytical criteria (Section 3.2) that allow us to determine, without full numerical integration, whether an initial BH binary can undergo a ‘one-shot merger’ or a more gradual merger induced by the octupole effect of an tertiary. In Section 4, we calculate the merger fraction as a function of mass ratio for some representative triple systems. In Section 5, we study the mass-ratio distribution of the initial BH binaries based on the properties of main-sequence (MS) stellar binaries and the MS mass to BH mass mapping. Using the result of Section 4, we illustrate how the final merging BH binary mass distribution may be influenced by the octupole effect for tertiary-induced mergers. We summarize our results and their implications in Section 6.

2 ZLK Oscillations: Analytical Results

Consider two BHs orbiting each other with masses \( m_1 \) and \( m_2 \) on a orbit with semimajor axis \( a \), eccentricity \( e \), and angular momentum \( L \). An external, tertiary BH of mass \( m_3 \) orbits this inner binary with semimajor axis \( a_{\text{out}} \), eccentricity \( e_{\text{out}} \), and angular momentum \( L_{\text{out}} \). The reduced masses of the inner and outer binaries are \( \mu = m_1 m_2/m_{12} \) and \( \mu_{\text{out}} = m_{12} m_{3}/m_{123} \), respectively, where \( m_{12} = m_1 + m_2 \) and \( m_{123} = m_1 + m_2 + m_3 \). These two binary orbits are further described by three angles: the inclinations \( i \) and \( i_{\text{out}} \), the arguments of pericentres \( \omega \) and \( \omega_{\text{out}} \), and the longitudes of the ascending nodes \( \Omega \) and \( \Omega_{\text{out}} \). These angles are defined in a coordinate system where the \( z \)-axis is aligned with the total angular momentum \( J = L + L_{\text{out}} \) (i.e. the invariant plane is perpendicular to \( J \)). The mutual inclination between the two orbits is denoted \( i \equiv i_{\text{out}} \). Note that \( \Omega_{\text{out}} = \Omega + 180^\circ \).

To study the evolution of the inner binary under the influence of the tertiary BH, we use the double-averaged secular equations of motion, including the interactions between the inner binary and the tertiary up to the octupole level of approximation as given by Liu et al. (2015). Throughout this paper, we restrict to hierarchical triple systems where the double-averaged secular equations are valid – systems with relatively small \( a_{\text{out}}/a \) may require solving the single-averaged equations of motion or direct N-body integration (see Antonini & Perets 2012; Antonini, Murray & Mikkola 2014; Luo, Katz & Dong 2016; Lei, Ciri & Orto 2018; Liu & Lai 2019; Liu et al. 2019a; Hamers 2020a). For the remainder of this section, we include general relativistic apsidal precession of the inner binary, a first-order post-Newtonian (1PN) effect, but omit the emission of GWs, a 2.5PN effect – this will be considered in Section 3. We group the results by increasing order of approximation, starting by ignoring the octupole-order effects entirely.

Note that Fig. 1 should not be interpreted as directly reflecting the distribution of merging BH binaries, as there are many selection effects and observational biases, e.g. systems with smaller \( q \) are harder to detect for the same \( M_{\text{chirp}} \) or \( M_{12} \). For a detailed statistical analysis, see Abbott et al. (2020b).

Although we do not study such systems in this paper, we expect that a qualitatively similar dependence of the merger probability on the mass ratio remains, since the strength of the octupole effect in the single-averaged secular equations is also proportional to \( (1 - q)(1 + q) \) (see equation 25 of Liu & Lai 2019).
2.1 Quadrupole order

At the quadrupole order, the tertiary induces eccentricity oscillations in the inner binary on the characteristic time-scale,

\[ t_{ZLK} = \frac{1}{\eta} \left( \frac{a_{\text{out, eff}}}{a} \right)^3, \quad (3) \]

where \( \eta \equiv \sqrt{Gm_2/a^2} \) is the mean motion of the inner binary, and \( a_{\text{out, eff}} \equiv a_{\text{out}} \sqrt{1 - e_{\text{out}}} \). During these oscillations, there are two conserved quantities, the total energy and the total orbital angular momentum. Through some manipulation, the total angular momentum can be written in terms of the conserved quantity \( K \) given by

\[ K = j(e) \cos I - \frac{\eta}{2} e^2. \quad (4) \]

Here, \( j(e) \equiv \sqrt{1 - e^2} \) and \( \eta \) is the ratio of the magnitudes of the angular momenta at zero inner binary eccentricity:

\[ \eta \equiv \left( \frac{L}{E_{\text{out}}} \right)_{e=0} = \frac{\mu}{m_{\text{out}}} \left[ \frac{m_{12} a}{m_{12} a_{\text{out}} (1 - e_{\text{out}})} \right]^{1/2}. \quad (5) \]

Note that when \( \eta = 0 \), \( K \) reduces to the classical ‘Kozai constant’, \( K = j(e)\cos I \).

The maximum eccentricity \( e_{\text{max}} \) attained in these ZLK oscillations can be computed analytically at the quadrupole order. It depends on the ‘competition’ between the IPN apsidal precession rate \( \omega_{\text{GR}} \) and the ZLK rate \( t_{ZLK}^{-1} \). The relevant dimensionless parameter is

\[ \epsilon_{GR} \equiv \left( \frac{\omega_{GR} t_{ZLK}}{e_{\text{out}}} \right)_{e=0} = \frac{3Gm_1 m_2}{a^2} \frac{a_{\text{out, eff}}}{a^2}. \quad (6) \]

It can then be shown that, for an initially circular inner binary, \( e_{\text{max}} \) is related to the initial mutual inclination \( I_0 \) by (Liu et al. 2015; Anderson, Storch & Lai 2016)

\[ \frac{3}{8} \left( j^2(e_{\text{max}}) - j^2(e_{\text{max}}) \right) \left[ 5 \left( \cos I_0 + \frac{\eta}{2} \right)^2 - \left( 3 + 4\eta \cos I_0 + \frac{9}{4} \right) \right]
\]

\[ \times j^2(e_{\text{max}}) + \eta^2 j^4(e_{\text{max}}) + \epsilon_{GR} \left[ 1 - \frac{1}{j(e_{\text{max}})} \right] = 0. \quad (7) \]

In the limit \( \eta \to 0 \) and \( \epsilon_{GR} \to 0 \), we recover the well-known result \( e_{\text{max}} = \sqrt{1 - (5/3)\cos^2 I_0} \). For general \( \eta \), \( e_{\text{max}} \) attains its limiting value \( e_{\text{lim}} \) when \( I_0 = I_{0,\text{lim}} \), where (see also Hamers 2020b)

\[ \cos I_{0,\text{lim}} = \frac{\eta}{2} \left[ \frac{4}{5} j^2(e_{\text{lim}}) - 1 \right]. \quad (8) \]

Note that \( I_{0,\text{lim}} \geq 90^\circ \) with equality only when \( \eta = 0 \). Substituting equation (8) into equation (7), we find that \( e_{\text{lim}} \) satisfies

\[ \frac{3}{8} \left[ j^2(e_{\text{lim}}) - 1 \right] \left[ -3 + \frac{\eta^2}{4} \left( \frac{4}{5} j^2(e_{\text{lim}}) - 1 \right) \right] + \epsilon_{GR} \left[ 1 - \frac{1}{j(e_{\text{lim}})} \right] = 0. \quad (9) \]

On the other hand, eccentricity excitation (\( e_{\text{max}} \geq 0 \)) is only possible when \( (\cos I_0)_c \leq \cos I_0 \leq (\cos I_0)_c \), where

\[ (\cos I_0)_c = \frac{1}{10} \left( -\eta \pm \sqrt{\eta^2 + 60 - \frac{80}{3} \epsilon_{GR}} \right). \quad (10) \]

For \( I_0 \) outside of this range, no eccentricity excitation is possible. This condition reduces to the well-known \( \cos^2 I_0 \leq 3/5 \) when \( \eta = \epsilon_{GR} = 0 \).

2.2 Octupole order: test-particle limit

The relative strength of the octupole-order potential to the quadrupole-order potential is determined by the dimensionless parameter \( \epsilon_{\text{oct}} \) (equation 2). When \( \epsilon_{\text{oct}} \) is non-negligible, \( K \) is no longer conserved, and the system evolution becomes chaotic (Ford et al. 2000; Katz, Dong & Malhotra 2011; Lithwick & Naoz 2011; Li et al. 2014; Liu et al. 2015). As a result, analytical (and semi-analytical) results have only been given for the test-particle limit, where \( m_2 = \eta = 0 \). We briefly review these results below.

Due to the non-conservation of \( K \), \( e_{\text{max}} \) evolves irregularly ZLK cycles, and the orbit may even flip between prograde (\( I < 90^\circ \)) and retrograde (\( I > 90^\circ \)) if \( K \) changes sign (in the test-particle limit, \( K = j(e)\cos I \)). During these orbit flips, the eccentricity maxima reach their largest values but do not exceed \( e_{\text{lim}} \) (Lithwick & Naoz 2011; Li et al. 2015; Anderson et al. 2016). These orbit flips occur on characteristic time-scale \( t_{ZLK,\text{oct}} \), given by (Antognini 2015)

\[ t_{ZLK,\text{oct}} = \frac{t_{ZLK}}{\epsilon_{\text{oct}}} \left( \frac{128\sqrt{10}}{15\pi\sqrt{\epsilon_{\text{oct}}}} \right). \quad (11) \]

The octupole potential tends to widen the inclination range for which the eccentricity can reach \( e_{\text{lim}} \); we refer to this widened range as the octupole-active window. Fig. 2 shows the maximum eccentricity attained by an inner binary orbited by a tertiary companion with inclination \( I_0 \). The octupole-active window is visible as a range of inclinations centred on \( I_0 = 90^\circ \) that attain \( e_{\text{lim}} \) (the red horizontal dashed line in Fig. 2). Katz et al. (2011) show that this window can be approximated using analytical arguments when \( \epsilon_{\text{oct}} \ll 1 \). Muñoz et al. (2016) give a more general numerical fitting formula describing the octupole-active window for arbitrary \( \epsilon_{\text{oct}} \). They find...
In Fig. 2, we see that with the octupole effect included, \( MNRAS \) \( e \) \( \geq \) retrograde inclinations \( I \) attains eccentricity of a system with the same parameters as Fig. 2 except for top panel of Fig. 4, the blue dots show the maximum achieved \( q \) with \( \eta \) window as orbit flip occurs (this follows by inspection of equation 4). In the first panel, \( e_{\text{lim}} \) is denoted by the black dashed line. By comparing the second and third panels, we see that orbit flips occur when \( K \) crosses the dotted line, given by \( K = K_0 \equiv -\eta/2 \).

In Fig. 2, we see that with the octupole effect included, \( e_{\text{lim}} \) indeed attains \( e_{\text{lim}} \) when \( I_0 \) is within the broad octupole-active window given by equation (12) (denoted by the vertical purple lines in Fig. 2).

### 2.3 Octupole order: general masses

For general inner binary masses, when the angular momentum ratio \( \eta \) is non-negligible, the octupole-level ZLK behaviour is less well studied (see Liu et al. 2015). Fig. 3 shows an example of the evolution of a triple system with significant \( \eta \) and \( \epsilon_{\text{oct}} \). Many aspects of the evolution discussed in Section 2.2 are still observed: The ZLK eccentricity maxima and \( K \) evolve over time-scales \( \gg T_{\text{ZLK}} \); the eccentricity never exceeds \( e_{\text{lim}} \); and when \( K \) crosses \( K_e \equiv -\eta/2 \), an orbit flip occurs (this follows by inspection of equation 4).

However, equation (12) no longer describes the octupole-active window as \( \eta \) is non-negligible (see also Rodet et al. 2021). In the top panel of Fig. 4, the blue dots show the maximum achieved eccentricity of a system with the same parameters as Fig. 2 except with \( q = 0.5 \) (so \( \epsilon_{\text{oct}} = 0.007 \) and \( \eta = 0.087 \)). Here, it can be seen that no prograde systems can attain \( e_{\text{lim}} \), and only a small range of retrograde inclinations \( \geq I_{0,\text{lim}} \) (see equation 8) are able to reach \( e_{\text{lim}} \).

In fact, there is even a clear double valued feature around \( I \approx 75^\circ \) in the top panel of Fig. 4 that is not present in Fig. 2. If \( q \) is decreased to 0.3 (Fig. 5) or further to 0.2 (Fig. 6), \( \epsilon_{\text{oct}} \) increases while \( \eta \) decreases. This permits a larger number of prograde systems to reach \( e_{\text{lim}} \).
though a small range of inclinations near $I_0 = 90^\circ$ still do not reach $e_{\text{lim}}$; we call this range of inclinations the 'octupole-inactive gap'. On the other hand, if $q$ is held at 0.5 as in Fig. 4 and $e_{\text{out}}$ is increased to 0.9 while holding $a_{\text{out}}$, eff = 3600 AU constant, both $\epsilon_{\text{oct}}$ and $\eta$ increase; the top panel of Fig. 8 illustrates the behaviour when the inner binary is substantially more compact ($a = 10$ AU): even though $\epsilon_{\text{oct}}$ is larger than it is in any of Figs 4–7, we see that prograde perturbers fail to attain $e_{\text{lim}}$. All of these examples (top panels of Figs 4–8) illustrate importance of $\eta$ in determining the range of inclinations for the system to be able to reach $e_{\text{lim}}$.

In general, we find that a symmetric octupole-active window (as in equation 12) can be realized for sufficiently small $\eta$. Rodet et al. (in preparation) considered some examples of triple systems (consisting of MS stars with planetary companions and tertiaries, for which the short-range forces is dominated by tidal interaction) and found that $\eta \approx 0.054$ in Fig. 6).

Emission of GWs affects the evolution of the inner binary, which can be incorporated into the secular equations of motion for the triple (e.g. Peters 1964; Liu & Lai 2018). The associated orbital and eccentricity decay rates are (Peters 1964)

$$\frac{1}{a} \frac{da}{dt}_{\text{GW}} = -\frac{1}{t_{\text{GW}}}$$

$$= -\frac{64}{5} \frac{G^3 \mu m^2_{12}}{c^5 a^4} \frac{1}{j^2} \left( 1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right).$$

$$\frac{de}{dt}_{\text{GW}} = -\frac{334}{15} \frac{G^3 \mu m^2_{12}}{c^5 a^4} \frac{1}{j^2} \left( 1 + \frac{121}{30} e^2 \right).$$

GW emission can cause the orbit to decay significantly when extreme eccentricities are reached during the ZLK cycles described in the previous section. This allows even wide binaries ($\sim 100$ AU) to merge efficiently within a Hubble time. While various numerical examples of such tertiary-induced mergers have been given before (e.g. Liu & Lai 2018; see also Liu et al. 2019a for ‘population synthesis’), in this section we examine the dynamical process in detail in order to develop an analytical understanding. Our fiducial system parameters are as in Fig. 3: $a_{\text{out}}$, eff = 4500 AU, $e_{\text{out}} = 0.6$, $m_{12} = 50 M_\odot$ (with varying $q$), $m_3 = 30 M_\odot$, and the inner binary has initial $a_0 = 100$ AU and $e_0 = 10^{-3}$.

3 TERTIARY-INDUCED BH MERGERS

Emission of GWs affects the evolution of the inner binary, which can be incorporated into the secular equations of motion for the triple (e.g. Peters 1964; Liu & Lai 2018). The associated orbital and eccentricity decay rates are (Peters 1964)

$$\frac{1}{a} \frac{da}{dt}_{\text{GW}} = -\frac{1}{t_{\text{GW}}}$$

$$= -\frac{64}{5} \frac{G^3 \mu m^2_{12}}{c^5 a^4} \frac{1}{j^2} \left( 1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right),$$

$$\frac{de}{dt}_{\text{GW}} = -\frac{334}{15} \frac{G^3 \mu m^2_{12}}{c^5 a^4} \frac{1}{j^2} \left( 1 + \frac{121}{30} e^2 \right).$$

GW emission can cause the orbit to decay significantly when extreme eccentricities are reached during the ZLK cycles described in the previous section. This allows even wide binaries ($\sim 100$ AU) to merge efficiently within a Hubble time. While various numerical examples of such tertiary-induced mergers have been given before (e.g. Liu & Lai 2018; see also Liu et al. 2019a for ‘population synthesis’), in this section we examine the dynamical process in detail in order to develop an analytical understanding. Our fiducial system parameters are as in Fig. 3: $a_{\text{out}}$, eff = 4500 AU, $e_{\text{out}} = 0.6$, $m_{12} = 50 M_\odot$ (with varying $q$), $m_3 = 30 M_\odot$, and the inner binary has initial $a_0 = 100$ AU and $e_0 = 10^{-3}$.
3.1 Merger windows and probability: numerical results

To understand what initial conditions lead to successful mergers within a Hubble time, we integrate the double-averaged octupole-order ZLK equations including GW radiation. We terminate each integration if either $a = 0.005 a_0$ (a successful merger) or the system age reaches 10 Gyr. We can verify that the inner binary is effectively decoupled from the tertiary for this orbital separation by evaluating $\epsilon_{GR}$ (equation 6):

$$\epsilon_{GR} = 1.8 \times 10^6 \left( \frac{m_3}{50 M_\odot} \right)^2 \left( \frac{a_{\text{out, eff}}}{3600 \text{ AU}} \right)^3 \times \left( \frac{m_3}{30 M_\odot} \right)^{-1} \left( \frac{a}{0.5 \text{ AU}} \right)^{-4}. \quad (15)$$

The middle panel of Fig. 4 shows the merger time $T_m$ as a function of $I_0$ for our fiducial parameters with $q = 0.5$. We note that only retrograde inclinations lead to successful mergers, and almost all successful mergers are rapid, with $T_m \sim t_{\text{ZLK, oct}}$. These are the result of a system merging by emitting a single large burst of GW radiation during an extreme-eccentricity ZLK cycle, which we term a ‘one-shot merger’.

In Fig. 5, $q$ is decreased to 0.3, and some prograde systems are also able to merge successfully. However, these prograde systems exhibit a broad range of merger times, with $T_m \gtrsim t_{\text{ZLK, oct}}$. These occur when a system gradually emits a small amount of GW radiation at every eccentricity maximum – we term this a ‘smooth merger’. Additionally, the octupole-inactive gap near $I_0 = 90^\circ$ is visible in the merger time plot (middle panel of Fig. 5). The middle panels of Figs 6–8 show the behaviour of $T_m$ for the other parameter regimes and also exhibit these two categories of mergers and the octupole-inactive gap.

Due to the chaotic nature of the octupole-order ZLK effect, the initial inclination $I_0$ alone is not sufficient to determine with certainty whether a system can merge within a Hubble time. Instead, for a given $I_0$, we can use numerical integrations with various $\omega, \omega_{\text{out}},$ and $\Omega$ to compute a merger probability, denoted by

$$P_{\text{merger}} (I_0; q, \epsilon_{\text{out}}) = P (T_m < 10 \text{ Gyr}) \quad (16)$$

It is important to note that these ‘one-shot mergers’ are distinct from the ‘fast’ mergers previously discussed in the literature (e.g. Wen 2003; Randall & Xianyu 2018b; Su et al. 2020): The one-shot mergers discussed here occur when the maximum eccentricity attained by the inner binary over an octupole cycle (i.e. within the first $\sim t_{\text{ZLK, oct}}$) is sufficiently large to produce a prompt merger. When the octupole effect is non-negligible, it can drive systems to much more extreme eccentricities than can the quadrupole-order effects alone (compare the blue dots and black dashed line in Fig. 4), and thus our ‘one-shot mergers’ occur for a larger range of $I_0$ than do quadrupole-order ‘fast’ mergers.
where the notation \( P_{\text{merger}}(I_0; q, e_{\text{out}}) \) highlights the dependence of \( P_{\text{merger}} \) on \( q \) and \( e_{\text{out}} \), two of the key factors that determine the strength of the octupole effect (of course \( P_{\text{merger}} \) depends on other system parameters such as \( m_1, m_2, a_{\text{out}}, \) etc.). The bottom panels of Figs 4–8 show our numerical results. In all of these plots, there is a retrograde inclination window for which successful merger is guaranteed. In Fig. 5, it can be seen that a large range of prograde inclinations have a probabilistic outcome. In Fig. 6, while the enhanced octupole strength allows for most of the prograde inclinations to merge with certainty, there is still a region around \( I_0 \approx 80^\circ \) where \( P_{\text{merger}} < 1 \).

### 3.2 Merger probability: semi-analytic criteria

By comparing the top and bottom panels of Figs 4–8, it is clear that their features are correlated: In all five cases, the retrograde merger window occupies the same inclination range as the retrograde octupole-active window, while \( P_{\text{merger}} \) is only non-zero for prograde inclinations where \( e_{\text{max}} \) nearly attains \( e_{\text{out}} \). Here, we further develop this connection and show that the non-dissipative simulations can be used to predict the outcomes of simulations with GW dissipation rather reliably.

In Section 3.1, we identified both one-shot and smooth mergers in our simulations. Towards understanding the one-shot mergers, we first define \( e_{\text{os}} \) to be the \( e_{\text{max}} \) required to dissipate an order-unity fraction of the binary’s orbital energy via GW emission in a single ZLK cycle. Since a binary spends a fraction \( -j(e_{\text{max}}) \) of each ZLK cycle near \( e_{\text{max}} \) (e.g. Anderson et al. 2016), we set

\[
 j(e_{\text{os}}) = -\frac{1}{t_{\text{ZLK}}},
\]

where \( d(\ln a)/d\tau \) is given by equation (13). This yields

\[
 j^\prime(e_{\text{os}}) = \frac{425 t_{\text{ZLK}}}{96 t_{\text{GW}, 0}} \frac{170 G^2 m_1^2}{3 m^3 c^4 n} \left( \frac{a_{\text{out}, \text{eff}}}{a} \right)^3,
\]

where \( t_{\text{GW}, 0} = (t_{\text{GW}})_{e = 0} \) (see equation 13) is given by

\[
 t_{\text{GW}, 0}^{-1} = \frac{64 G^3 m_1^2}{5 c^5 e_{\text{os}}^4}.
\]

We have approximated \( e_{\text{os}} \approx 1 \). Equation (18) is equivalent to

\[
 1 - e_{\text{os}} \approx 3 \times 10^{-6} \left( \frac{m_1}{50 \text{ M}_\odot} \right)^7/6 \left( \frac{q/(1 + q)^2}{1/4} \right)^{1/3} \times \left( \frac{a_{\text{out}, \text{eff}}}{3600 \text{ AU}} \right)^{1/3} \left( \frac{a}{100 \text{ AU}} \right)^{-11/6}.
\]

Then, if a system satisfies \( e_{\text{max}} > e_{\text{os}} \) with \( e_{\text{max}} \) based on non-dissipative integration, it is expected attain a sufficiently large eccentricity to undergo a one-shot merger.

Towards understanding smooth mergers, we seek a characteristic eccentricity that captures GW emission over many ZLK cycles. We define \( e_{\text{eff}} \) as an effective ZLK maximum eccentricity, i.e.

\[
 \left\langle \frac{d \ln a}{d \tau} \right\rangle = -\frac{1}{t_{\text{ZL}K}} \left( \frac{1 + 73 c^2/24 + 37 c^4/96}{j^\prime(e)} \right) \equiv -\frac{425/96}{t_{\text{GW}, 0}} \frac{1}{j^\prime(e_{\text{eff}})},
\]

where the angle brackets denote averaging over many \( t_{\text{ZL}K, \text{osc}} \) in order to capture the characteristic eccentricity behaviour over many octupole cycles. In the second line of equation (21), we have essentially replaced the ZLK-averaged orbital decay rate by \( d(\ln a)/d\tau \) evaluated at \( e_{\text{eff}} \) multiplied by \( j(e_{\text{eff}}) \). In practice (see Figs 4–8), we typically average over 2000\( t_{\text{ZL}K} \) of the non-dissipative simulations to compute \( e_{\text{eff}} \).

With \( e_{\text{eff}} \) computed using equation (21), we can define the critical effective eccentricity \( e_{\text{eff}, \text{c}} \) such that the ZLK-averaged inspiral time is a Hubble time, i.e. \( (d(\ln a)/d\tau) \equiv -(10 \text{ Gyr})^{-1} \). This gives

\[
 j^\prime(e_{\text{eff}, \text{c}}) = \frac{425 \text{ 10 Gyr}}{96 t_{\text{GW}, 0}},
\]

or equivalently

\[
 1 - e_{\text{eff}, \text{c}} \approx 10^{-4} \left( \frac{m_1}{50 \text{ M}_\odot} \right) \left( \frac{q/(1 + q)^2}{1/4} \right)^{1/3} \left( \frac{a}{100 \text{ AU}} \right)^{-4/3}.
\]

Thus, if a system is evolved using the non-dissipative equations of motion and satisfies \( e_{\text{eff}} > e_{\text{eff}, \text{c}} \), then it is expected to successfully undergo a smooth merger within a Hubble time.

Therefore, a system can be predicted to merge successfully if it satisfies either the one-shot or smooth merger criteria. The semi-analytical merger probability (as a function of \( I_0 \) and other parameters) is

\[
 P_{\text{merger}}(I_0; q, e_{\text{out}}) = P \left( e_{\text{eff}} > e_{\text{eff}, \text{c}} \text{ or } e_{\text{max}} > e_{\text{os}} \right).
\]

Although not fully analytical (since numerical integrations of non-dissipative systems are needed to obtain \( e_{\text{eff}} \) and \( e_{\text{max}} \) in general), equation (24) provides efficient computation of the merger probability without full numerical integrations including GW radiation.

The top panels of Figs 4–8 show \( e_{\text{eff}} \) and \( e_{\text{max}} \), and their critical values, \( e_{\text{eff}, \text{c}} \) and \( e_{\text{os}} \). Using these, we compute the semi-analytical merger probability, shown as the thick green lines in the bottom panels of Figs 4–8. We generally observe good agreement with the numerical \( P_{\text{merger}} \). However, \( P_{\text{merger}} \) slightly but systematically underpredicts \( P_{\text{merger}} \) for some configurations, such as the prograde inclinations in Figs 5 and 8. These regions coincide with the inclinations for which the merger outcome is uncertain. This underprediction is due to the restricted integration time of 2000\( t_{\text{ZL}K} \) \( \approx 3 \) Gyr used for the non-dissipative simulations. To illustrate this, we also calculate \( P_{\text{merger}} \), using a shorter integration time of 500\( t_{\text{ZL}K} \) for our non-dissipative simulations. The results are shown as the light green lines in the bottom panels of Figs 4–8, performing visibly worse. A more detailed discussion of this issue can be found in Section 4.4.

A few observations about equation (24) can be made. First, it explains why some prograde systems merge probabilistically (\( 0 < P_{\text{merger}} < 1 \)): For the prograde inclinations in Fig. 5, the \( e_{\text{eff}} \) values scatter widely around \( e_{\text{eff}, \text{c}} \) [or more precisely, \( j(e_{\text{eff}}) \) scatters around \( j(e_{\text{eff}, \text{c}}) \)], even for a given \( I_0 \), so the detailed merger outcome depends on the initial conditions. For the prograde inclinations in Fig. 6, the double-valued feature in the \( e_{\text{max}} \) plot (the top panel) pointed out in Section 2.3 represents a subpopulation of systems that do not satisfy equation (24). Second, \( e_{\text{max}} > e_{\text{os}} \) often ensures \( e_{\text{eff}} > e_{\text{eff}, \text{c}} \) in practice, as the averaging in equation (21) is heavily weighted towards extreme eccentricities. As such, \( e_{\text{eff}} > e_{\text{eff}, \text{c}} \) alone is often a sufficient condition in equation (24).

The one-shot merger criterion \( e_{\text{max}} > e_{\text{os}} \) can also be used to distinguish two different types of system architectures: If \( e_{\text{lim}} > e_{\text{os}} \) for a particular architecture, then all initial conditions leading to orbit flips (i.e. in an octupole-active window) also execute one-shot mergers. For \( e_{\text{lim}} \approx 1 \), equation (9) reduces to

\[
 j(e_{\text{lim}}) \approx \frac{8 e_{\text{os}}}{9} \left( \frac{1 + n^2}{12} \right)^{-1}.
\]
For the system architecture considered in Figs 4–7, this condition is not possible, and when the condition (equation 26) is not satisfied, one-shot mergers give a new and outer binaries, i.e. we draw cos $I_{\text{q}}$ values and compute the merger fraction as a function of the mass ratio.

We first consider the simple case where $e_{\text{out}}$ is fixed at a few specific values and compute the merger fraction as a function of the mass ratio $q$. We consider isotropic mutual orientations between the inner and outer binaries, i.e. we draw $\cos I_0$ from a uniform grid over the range $[-1, 1]$ (recall that $\alpha$, $\alpha_{\text{out}}$, and $\Omega$ are drawn uniformly from the range $[0, 2\pi]$ when computing the merger probability $P_{\text{merger}}$ at a given $I_0$). The merger fraction is then given by

$$f_{\text{merger}}(q, e_{\text{out}}) = \frac{1}{2} \int_{-1}^{1} \cos I_0 \, P_{\text{merger}}(I_0; q, e_{\text{out}}).$$

(27)

This is proportional to the integral of the black lines (weighted by $\sin I_0$) in the bottom panels of Figs 4–7. We can also use semi-analytical criteria introduced in Section 3.2 to predict the outcome and merger fraction. This is computed by using $P_{\text{merger}}$ as the integrand in equation (27), or by evaluating the integral of the thick green lines (weighted by $\sin I_0$) in the bottom panels of Figs 4–7. Fig. 9 shows the resulting $f_{\text{merger}}$ and the analytical estimates for all combinations of $q \in \{0.2, 0.3, 0.4, 0.5, 0.7, 1.0\}$ and $e_{\text{out}} \in \{0.6, 0.8, 0.9\}$. It is clear that the numerical $f_{\text{merger}}$ and the analytical estimate agree well, and that the merger fraction increases steeply for smaller $q$.

To explore the impact of our choice of isotropic mutual orientations between the two binaries, we also consider a wedge-shaped distribution of $\cos I_0$ as was found in the population synthesis studies of Antonini et al. (2017). We still use the same uniform grid of $\cos I_0$ as before, but weight each eccentricity by its probability probability density following the distribution:

$$P(\cos I_0) = \frac{1}{4} + \frac{[\cos I_0]}{2}.$$  

(28)

The resulting $f_{\text{merger}}$ for a tertiary with $\cos I_0$ distributed like equation (28) is shown as the dashed lines in Fig. 9. While the total merger fractions decrease, the strong enhancement of the merger fraction at smaller $q$ is unaffected.

In the right-hand panel of Fig. 9, we see that the merger fractions for the three $e_{\text{out}}$ values overlap for small $e_{\text{out}}$. This implies that $f_{\text{merger}}$ depends only on $e_{\text{out}}$ in this regime, and not on the values of $q$ and $e_{\text{out}}$ independently. From Fig. 4 (which has $e_{\text{out}} = 0.007$), we see that this suggests that the size of the retrograde merger window only depends on $e_{\text{out}}$. Much like what equation (12) shows for the test-particle limit. However, once $e_{\text{out}}$ is increased sufficiently, the three curves in the right-hand panel of Fig. 9 cease to overlap. This can be attributed to their different $\eta$ values: for sufficiently small $e_{\text{out}}$, no prograde initial inclinations successfully merge (e.g. Fig. 4), and the merger fraction is solely determined by the size of the retrograde...
octupole-active window. But once $\epsilon_{\text{out}}$ is sufficiently large, prograde mergers become possible, and the merger fraction is also affected by the size of the octupole-inactive gap, which depends on $\eta$. This again illustrates the importance of the octupole-inactive gap, which we comment on in Appendix A.

Fig. 10 depicts the merger fractions for systems with $a_0 = 50 \, \text{AU}$ (the other parameters are the same as in Fig. 9). According to equation (26), these systems no longer satisfy $\epsilon_{\text{out}} > \epsilon_{\text{in}}$, so the merger fraction is expected to diminish strongly and vary much more weakly with $q$, as one-shot mergers are no longer possible. This is indeed observed, particularly for the $e_{\text{out}}$ = 0.6 curve in Fig. 10. We also remark that the semi-analytical prediction accuracy is poorer in this case than in Fig. 9. This is because the only mergers in this regime are smooth mergers. As can be seen for the prograde $l_0$ in Figs 5 and 8, smooth mergers occur over a wide range of merger times $T_{\text{in}}$, and the specific $T_{\text{in}}$ that a system experiences depends sensitively on its chaotic evolution. Thus, equation (21) is a rather approximate estimate of the amount of GW emission that a real system emits during a smooth merger; indeed, the prograde regions of Figs 5 and 8 show that the merger times for smooth mergers are systematically underpredicted by the semi-analytic merger criterion (see discussion in Section 4.4). The non-monotonicity of the semi-analytic merger fraction for $e_{\text{out}} = 0.6$ from $q = 0.2$ to 0.3 is due to small sample sizes and finite grid spacing in $\cos l_0$.

4.2 Merger fraction for a distribution of tertiary eccentricities

For a distribution of tertiary eccentricities, denoted $P(e_{\text{out}})$, the merger fraction is given by

$$\eta_{\text{merger}}(q) = \int d e_{\text{out}} P(e_{\text{out}}) f_{\text{merger}}(q, e_{\text{out}}),$$

$$= \int d e_{\text{out}} \frac{P(e_{\text{out}})}{} \frac{1}{2} \int_1 ^1 d \cos l_0 \epsilon_{\text{merger}}(l_0; q, e_{\text{out}}).$$

(29)

We consider two possible $P(e_{\text{out}})$ with $e_{\text{out}} \in [0, 0.9]$; (i) a uniform distribution, $P(e_{\text{out}}) = $ constant, and (ii) a thermal distribution, $P(e_{\text{out}}) \propto e_{\text{out}}$.

The top panel of Fig. 11 shows $\eta_{\text{merger}}$ (black dots) for the fiducial triple systems (with the same parameters as in Figs 4–7). For each $q$, the integral in equation (29) is computed using 1000 realizations of random $e_{\text{out}}$, $\cos l_0$, $\omega$, $\omega_{\text{out}}$, and $\Omega$. Not surprisingly, we see $\eta_{\text{merger}}$ increases with decreasing $q$. When $q$ is small, thermal distribution of $e_{\text{out}}$ tends to yield higher $\eta_{\text{merger}}$ than does a uniform distribution. We also compute the merger fraction using the semi-analytical merger probability of equation (24) on a dense grid of initial conditions uniformly sampled in $e_{\text{out}}$ and $\cos l_0$; the result is shown as the blue dotted line in Fig. 11, which is in good agreement with the uniform-$e_{\text{out}}$ simulation result (black).

To characterize the properties of merging binaries, the middle and bottom panels of Fig. 11 show the distributions of merger times and merger eccentricities (at both the LISA and LIGO bands) for different mass ratios. To obtain the LISA and LIGO band eccentricities (with GW frequency equal to 0.1 and 10 Hz, respectively), the inner binaries are evolved from when they reach 0.005$a_0$ (at which point we terminate the integration of the triple system evolution as the inner binary’s evolution is decoupled from the tertiary; see equation (15) to physical merger using equations (13)–(14). While the LIGO band eccentricities are all quite small ($\lesssim 10^{-3}$), the LISA band eccentricities (at 0.1 Hz) are significant, with median $\gtrsim 0.2$ for $q \lesssim 0.5$. We note that these eccentricities are generally smaller than those found in the population studies of Liu et al. (2019a). This is because in this paper we consider only sufficiently hierarchical systems for which double-averaged evolution equations are valid, whereas Liu et al. (2019a) included a wider range of triple hierarchies and had to use $N$-body integrations to evolve some of the systems.

For comparison, Fig. 12 shows the results when $a_{\text{out, eff}} = 5500 \, \text{AU}$ (instead of $a_{\text{out, eff}} = 3600 \, \text{AU}$ for Fig. 11) with all other parameters unchanged. While $\eta_{\text{merger}}$ is lower than it is for $a_{\text{out, eff}} = 3600 \, \text{AU}$, there is still a large increase of $\eta_{\text{merger}}$ with decreasing $q$. Since equation (26) is still satisfied, this is expected.

4.3 $q \ll 1$ limit

For fixed $m_{12}$ (and other parameters), even though the octupole strength $\epsilon_{\text{out}}$ increases as $q$ decreases, the efficiency of GW radiation...
Figure 11. Upper panel: binary BH merger fraction as a function of mass ratio \( q \) for the fiducial triple systems (with parameters the same as in Figs 4–7), assuming random mutual inclinations (uniform in \( \cos I_0 \)), and either uniform (black dots) or thermal distribution (red dots) for the tertiary eccentricity distribution [with \( e_{\text{out}} \in [0, 0.9] \)]. These are obtained numerically using equation (29) by sampling 1000 combinations of \( e_{\text{out}}, \cos I_0, \omega, \omega_{\text{out}}, \Omega_1 \). The blue dotted line is the semi-analytical result obtained by applying equation (24) in equation (29) (evaluated using a dense uniform grid of \( \cos I_0 \) and \( e_{\text{out}} \)). The thick green line is a power-law fit to the analytical \( \eta_{\text{merger}} \) with a power-law index of \(-2.5\). Middle panel: merger times of successful mergers for a uniform \( e_{\text{out}} \) distribution (the median is denoted with the large black dot). Bottom panel: merging binary eccentricities (again, for a uniform \( e_{\text{out}} \) distribution) in the LISA (0.1 Hz; green) and LIGO bands (10 Hz; black), with medians marked with large dots.

Also decreases. It is therefore natural to ask at what \( q \) these competing effects become comparable and the merger fraction is maximized. We show that this does not happen until \( q \) is extremely small.

We see from Figs 4–7 that \( e_{\text{lim}} > e_{\text{os}} \) for our fiducial triple systems. Indeed, from equation (26), we see that even for \( q \) as small as \( 10^{-5} \), the condition \( e_{\text{lim}} > e_{\text{os}} \) is satisfied. This implies that most binaries execute one-shot mergers when undergoing an orbit flip. In addition, recall that the characteristic time for the binary to approach \( e_{\text{lim}} \) can be estimated by equation (11), which, for our fiducial triple systems, is given by

\[
t_{\text{ZLK, oct}} \simeq 10^8 \left( \frac{m_1}{50 \, M_\odot} \right)^{1/2} \left( \frac{a_{\text{out, eff}}}{3600 \, \text{AU}} \right)^{7/2} \left( \frac{a}{100 \, \text{AU}} \right)^{-2} \times \left( \frac{m_3}{30 \, M_\odot} \right)^{-1} \left[ \frac{1-q}{1+q} \sqrt{1-e_{\text{out}}^2} \right]^{-1/2} \text{yr.}
\]  

(30)

4.4 Limitations of semi-analytic calculation

It can be seen in Fig. 9 that the semi-analytical merger fractions are systematically lower than the values obtained from the direct simulations. One reason that this discrepancy arises is because the non-dissipative simulations used to compute \( e_{\text{eff}} \) and \( e_{\text{max}} \) are only run for \( 2000 \, t_{\text{LK}} \approx 3 \, \text{Gyr} \), while the full simulations including GW dissipation are run for 10 Gyr. Owing to the chaotic nature of the octupole-order ZLK effect, this means that, if an initial condition leads to extreme eccentricities only after many Gyr, then \( e_{\text{eff}} \) and \( e_{\text{max}} \) are underpredicted by the non-dissipative simulations. Additionally, there are times when eccentricity vector of the inner binary is librating, during which orbit flips are strongly suppressed (Katz et al. 2011). Since the librating phase can last an unpredictable amount of time, this suggests that the semi-analytical merger criteria can become more complete as the integration time is increased.
where \( f \) is the mass ratio distribution of merging binaries, we would need to know both the mass ratio and the eccentricity distribution of the outer binaries, we have

\[
\frac{dF_{\text{merger}}}{dqdm_{12}} = \int dq_0 dm_{12} da_{\text{out}} \frac{dF}{dqdm_{12}da_0} \times \frac{dF_{\text{out}}}{dm_3da_{\text{out},\text{eff}}} \eta_{\text{merger}}(q; m_{12}, a_0, m_3, a_{\text{out},\text{eff}}),
\]

where \( \eta_{\text{merger}} \) is given by equation (29). Some examples of \( \eta_{\text{merger}} \) are shown in the top panels of Figs 11–12.

Clearly, to properly evaluate equation (32) or (33) would require large population synthesis calculations and in any case would involve significant uncertainties, a task beyond the scope of this paper. For illustrative purposes, we consider the fiducial triple systems as studied in Section 4, and estimate the mass-ratio distribution of BH mergers as

\[
\frac{dF_{\text{merger}}}{dq} \sim \frac{dF}{dq} \eta_{\text{merger}}(q).
\]

We quantify the ‘completeness’ of the semi-analytical merger fraction via the ratio \( f_{\text{merge}}/f_{\text{merger}} \) as a function of non-dissipative integration time. We focus on the fiducial triple systems for demonstrative purposes and compute the completeness for each of the simulations in light grey lines and their mean in the thick black line. We see that the completeness is still increasing even as the non-dissipative simulation time is increased to 2000\( t_{\text{ZLK}} \), so we expect that even longer integration times would give even better agreement with the dissipative simulations.

5 MASS RATIO DISTRIBUTION OF MERGING BH BINARY

In Section 4, we have calculated the binary BH merger fractions \( f_{\text{merger}} \) and \( \eta_{\text{merger}} \) as a function of the mass ratio \( q \) for some representative triple systems. To determine the distribution in \( q \) and \( m_{12} \) (total mass) of the merging binaries, we would need to know both the initial distribution in \( q \) and \( m_{12} \), and \( a_0 \) of the inner BH binaries and the distribution in \( m_3 \), \( a_{\text{out}} \), and \( e_{\text{out}} \) of the outer binaries, denoted by

\[
\frac{dF}{dqdm_{12}da_0}, \quad \frac{dF_{\text{out}}}{dm_3da_{\text{out}}de_{\text{out}}}.
\]

The distribution in \( q \) and \( m_{12} \) of the merging binaries is then

\[
\frac{dF_{\text{merger}}}{dqdm_{12}} = \int dq_0 dm_{12} da_{\text{out}} \frac{dF}{dqdm_{12}da_0} \times \frac{dF_{\text{out}}}{dm_3da_{\text{out}}de_{\text{out}}} \eta_{\text{merger}}(q, e_{\text{out}}, m_{12}, a_0, m_3, a_{\text{out},\text{eff}}, e_{\text{out}}),
\]

where \( \eta_{\text{merger}} \) is given by equation (27) (assuming random mutual inclinations between the inner and outer binaries), and we have spelled out its dependence on various system parameters. Some examples of \( \eta_{\text{merger}} \) are shown in Figs 9–10. If we further specify the eccentricity distribution of the outer binaries, we have

5.1 Initial \( q \)-distribution of BH binaries

The initial mass-ratio distribution of BH binaries, \( dF/dq \), is uncertain. It can be derived from the mass distributions of MS binaries, together with the MS mass (\( m_{\text{ms}} \)) to BH mass (\( m \)) relation.

For the distribution of MS binary masses, we assume that each MS component mass is drawn from a Salpeter-like initial mass function (IMF) independently, with

\[
\frac{dF_{\text{ms}}}{dm_{\text{ms}}} \propto m_{\text{ms}}^{-\alpha},
\]

in the range \( m_{\text{ms}} \leq m_{\text{ms}} \leq m_{\text{ms}} \). Note in this case, the MS binary mass-ratio distribution is (for \( q \leq 1 \))

\[
\frac{dF_{\text{ms}}}{dq} \propto q^{\alpha-1} \left[ 1 - \left( \frac{q}{q_{\text{min}}} \right)^{2-2\alpha} \right],
\]

where \( q_{\text{min}} = m_{\text{min}}/m_{\text{max}} \) is the minimum possible binary mass ratio (this is a generalization of the result of Tout 1991). We consider two representative values of \( \alpha \): (i) \( \alpha = 2.35 \), the canonical Salpeter IMF (Salpeter 1955), and (ii) \( \alpha = 2 \), resulting in a uniform q distribution (for \( q \geq 2q_{\text{min}} \)). The latter case is consistent with observational studies of the mass ratio of high-mass MS binaries (Sana et al. 2012; Duchêne & Kraus 2013; Kobulnicky et al. 2014; Moe & Di Stefano 2017).

To obtain \( dF/dq \), we compute the BH binary mass ratio when each MS mass \( m_{\text{ms}} \) is mapped to its corresponding BH mass \( m \). This mapping is taken from Spera & Mapelli (2017) for the mass range \( 25 \leq m_{\text{ms}} \leq 117 M_{\odot} \). We consider both the case where \( Z = 0.02 \) (‘high Z’) and \( 2.0 \times 10^{-4} \) (‘low Z’), the two limiting metallicities used in Spera & Mapelli (2017). We can then numerically compute \( dF/dq \) by sampling masses for stellar binaries from the IMF, translating these into BH masses, then calculating the resulting BH mass ratios for each binary. The upper panel of Fig. 15 shows the \( dF/dq \) obtained via this procedure for a Salpeter IMF (\( \alpha = 2.35 \)) when sampling \( 10^3 \) MS binaries for each metallicity. In the lower four panels, we also show \( dF/dq \) restricted to particular ranges of \( m_{12} \). Note that the distributions differ significantly among the \( m_{12} \) ranges and also between the two metallicities. Fig. 16 shows the case when \( \alpha = 2 \), which mostly resembles Fig. 15.

5.2 \( q \)-distribution of merging BH binaries

Using the results of Section 5.1, we can also estimate the mass-ratio distribution of merging BHs using equation (34). We consider...
representative triple systems considered in Section 4. For \( \eta_{\text{merger}} \), we use a simple approximation that lies roughly between the two cases shown in Figs 11–12:

\[
\eta_{\text{merger}}(q) \approx 0.2 \times \left[ \max(q, 0.2) \right]^{-2}.
\]  

The results for \( dF_{\text{merger}}/dq \) are displayed as the dotted curves in Figs 15–16 in each panel. Broadly speaking, \( dF_{\text{merger}}/dq \) peaks around \( q \approx 0.3 \) for low-\( Z \) systems, and around \( q \approx 0.4 \) for high-\( Z \) systems, the latter reflecting the peak in the initial BH binary \( q \)-distribution. Also note that \( dF_{\text{merger}}/dq \) can be quite different for different \( m_{12} \) ranges. For example, merging BH binaries with \( m_{12} > 42 \, M_\odot \) are only produced in low-\( Z \) systems, and \( dF_{\text{merger}}/dq \) peaks around \( q \approx 0.3 \) for \( m_{12} \in [42, 67] \, M_\odot \), and is roughly uniform between \( q \approx 0.2 \) and 1 for \( m_{12} \approx 67 \, M_\odot \).

We emphasize that these results for \( dF_{\text{merger}}/dq \) refer to the representative triple systems studied in Sections 2–4, and thus should be considered for illustrative purposes only. As noted above, the merger fraction \( \eta_{\text{merger}} \) depends on various parameters of the triple systems. While we have not attempted to quantify \( \eta_{\text{merger}} \) for all possible triple system parameters, it is clear that the principal finding of Section 4 (i.e. \( \eta_{\text{merger}} \) increases with decreasing \( q \)) applies only for systems with sufficiently strong octupole effects. In fact, from Figs 9 and 10, we can estimate that the octupole-induced feature in \( \eta_{\text{merger}} \) becomes prominent only when \( \epsilon_{\text{out}} \gtrsim 0.005 \), or equivalently

\[
\frac{a}{a_{\text{out}, \text{eff}}} \gtrsim 0.005 \left( \frac{1 + q}{1 - q} \right)^{1/2} \frac{e_{\text{out}}^2}{e_{\text{out}}} \approx 0.01 \frac{e_{\text{out}}}{\epsilon_{\text{out}}},
\]  

where in the second step, we have used \( q \sim 0.5 \) and \( e_{\text{out}} \sim 0.6 \). When this condition is satisfied, the inner binary can usually also undergo a one-shot merger (see equation 26), leading to strong dependence of the merger fraction on \( q \). For triple systems with \( a_{\text{out}, \text{eff}} < 0.1 \) (such as the case when the tertiary is a supermassive BH with \( m_s \gtrsim 10^6 \, m_{12} \)), the octupole effect is unimportant (see the discussion following equation 2), and we expect the merger fraction to be almost independent of \( q \). Indeed, an analytical fitting formula for BH mergers induced by pure quadrupole-ZLK effect shows \( \eta_{\text{merger}} \propto (1 - q)^{1/3} \) (see equation 53 of Liu & Lai 2018, or equation 26 of Liu & Lai 2021). For such systems, we expect \( dF_{\text{merger}}/dq \) to be mainly determined by the initial \( q \)-distribution of BH binaries at their formation.

### 6 SUMMARY AND DISCUSSION

We have studied the dynamical formation of merging BH binaries induced by a tertiary companion via the ZLK effect, focusing on the expected mass-ratio distribution of merging binaries. The octupole potential of the tertiary, when sufficiently strong, can increase the inclination window and probability of extreme eccentricity excitation, and thus enhance the rate of successful binary mergers. Since the octupole strength \( \epsilon_{\text{out}}(1 - q)/(1 + q) \) (see equation 2) increases with decreasing binary mass ratio \( q \), it is expected that ZLK-induced BH mergers favour binaries with smaller mass ratios. We quantify the dependence of the merger fraction/probability on \( q \) using a combination of numerical integrations and analytical calculations, based on the secular evolution equations for hierarchical triples. We develop new analytical criteria (Section 3.2) that allow us to determine, without full numerical integrations, whether an initial BH binary can undergo a ‘one-shot merger’ or a more gradual merger under the influence of a tertiary companion. These allow us to compute the merger probability semi-analytically by only studying non-dissipative (i.e. no GWs) triple systems (see equation 24). We show that for hierarchical triples with semimajor axial ratio \( a_{\text{out}, \text{eff}} \gtrsim 0.01–0.02 \) (see equation 38), the BH binary merger fraction \( \eta_{\text{merger}} \) can increase by a larger factor (up to \( \sim 20 \)) as \( q \) decreases from unity to 0.2 (see Figs 9–13). When combined with a reasonable estimate of the mass-ratio distribution of the initial BH binaries (Section 5.1), our results for the merger fraction suggest that the final merging BH binaries have an overall mass-ratio distribution that peaks around \( q \sim 0.3 \) or 0.4, although very different distributions can be produced when restricting to specific ranges of total binary masses (see Figs 15 and 16).

Taking our final results (Figs 15 and 16) at face value, we tentatively conclude that the mass-ratio distribution \( dF_{\text{merger}}/dq \) of BH binary mergers induced by a comparable-mass companion is inconsistent with the current LIGO/VIRGO result (see Fig. 1), suggesting that such tertiary-induced mergers may not be the dominant formation channel for the majority of the detected LIGO/VIRGO events. However, there are at least two important issues/caveats to keep in mind:

(i) \( dF_{\text{merger}}/dq \) depends strongly on the initial mass-ratio distribution of BH binaries at their formation (\( dF/dq \)), which is uncertain and depends sensitively on the metallicity of the binary formation environment (see Section 5.1). It is also possible that the initial BH binary mass-ratio distribution is much more skewed towards equal masses than what we found in Section 5.1 (e.g. if stellar binaries with significantly asymmetric masses become unbound due to mass-loss and supernova kicks as their components become BHs). Such a distribution was found by population synthesis studies that include octupole-order ZLK effects and models of stellar evolution (e.g. Hamers et al. 2013; Toonen, Perets & Hamers 2018). These studies find that ZLK oscillations in stellar binaries with small \( q \) can experience mass transfer and merge without forming a compact object binary; as a result, most compact object binaries form with large mass ratios. The prevalence of this phenomenon likely depends on the initial semimajor axes of the inner binaries. Further study would be required to understand the competition between this primordial large-\( q \) enhancement and the elevated merger fractions for small \( q \) found in this study in an astrophysically realistic population.

Figure 14. Completeness of the semi-analytical merger fraction, defined as \( f_{\text{merger}}/f_{\text{merger}} \), as a function of the integration time used for the non-dissipative simulations, in the fiducial parameter regime while \( e_{\text{out}} \) is fixed at a few values. The thin grey lines indicate the completeness for particular combinations of \( (q, e_{\text{out}}) \), and the thick black line denotes their average. We see that completeness is still increasing as the integration time approaches 2000\( ZLK \approx 3 \) Gyr.
Mass-ratio distribution

Figure 15. Mass-ratio distributions of the initial BH binaries (solid lines) and merging BH binaries (dotted lines) when using $\alpha = 2.35$ for the MS stellar initial mass function (see Sections 5.1 and 5.2). Top panel: distribution of binary mass ratio at formation and merger for all possible total binary BH masses. Each BH mass is obtained from the MS mass using the fitting formula of Spera & Mapelli (2017) for metallicities of $2 \times 10^{-4}$ (low $Z$) and 0.02 (high $Z$), while the merger fraction of BH binaries is given by equation (37). To produce these distributions, $10^5$ initial MS binaries are used for each metallicity, and the number of merging BH binaries has been scaled up by a factor of 10 for visibility. The counts refer to the number per $\Delta q = 0.05$ bin. Bottom four panels: same as the top panel but with specific ranges of $m_{12}$, the total BH mass of the binary (as labeled). Note that low-$m_{12}$ systems are mainly produced from high-$Z$ MS binaries, while high-$m_{12}$ systems are mainly produced in low-$Z$ MS binaries.

(ii) When the tertiary mass $m_3$ is much larger than the BH binary mass $m_{12}$, as in the case of a supermassive BH tertiary, dynamical stability of the triple requires $a_{out} \gg a$, which implies that the octupole effect is negligible ($\epsilon_{oct} \ll 1$). For such triple systems, we expect the merger fraction to depend very weakly on the mass ratio, and the final $dF_{\text{merger}}/dq$ to depend entirely on the initial $dF/dq$. Although the merger fraction of such ‘pure quadrupole’ triples is small ($\lesssim 6$ per cent; see equation 53 of Liu & Lai 2018), additional ‘external’ effects can enhance the merger efficiency significantly [e.g. when the outer orbit experiences quasi-periodic torques from the galactic potential (Petrovich & Antonini 2017; see also Hamers & Lai 2017), or from the spin of a supermassive BH (Liu et al. 2019b)].

Near the completion of this paper, we became aware of the simultaneous work by Martinez, Rodriguez & Fragione (2021), who study a similar topic using a population synthesis approach.

ACKNOWLEDGEMENTS

We thank the anonymous referee whose detailed review and comments greatly improved this paper. YS thanks Jiseon Min for useful discussions. This work has been supported in part by NSF grant AST1715246. YS is supported by the NASA FINESST grant 19-ASTRO19-0041. BL gratefully acknowledges support from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 847523 ‘INTERACTIONS’.

DATA AVAILABILITY

The data referenced in this paper will be shared upon reasonable request to the corresponding author.
REFERENCES

Naoz S., 2016, ARA&A, 54, 441
Sana H. et al., 2012, Science, 337, 444

APPENDIX A: ORIGIN OF OCTUPOLE-INACTIVE GAP

We investigate the origin of the ‘octupole-inactive gap’, an inclination range near \( I_0 \approx 90^\circ \) for which \( \epsilon_{\text{max}} \) does not attain \( \epsilon_{\text{lim}} \) despite being in between two octupole-active windows. This gap was first identified in Section 2.3, and is seen in both the non-dissipative and full simulations with GW dissipation (see Figs 4–8).

To better understand this gap, we first review the mechanism by which extreme eccentricity excitation occurs. In the test-particle limit, Katz et al. (2011) showed that \( K_2 \) oscillates over long time-scales when \( \omega \), the argument of pericentre of the inner orbit, is circulating. This then leads to orbit flips (and extreme eccentricity excitation) between prograde and retrograde inclinations when \( K_2 \) changes signs: since \( \epsilon \) is non-negative, the sign of \( K_2 \) determines the sign of \( \cos I \). Katz et al. (2011) obtained coupled oscillation equations in \( K_2 \) and \( \Omega_2 \), the azimuthal angle of the inner eccentricity vector in the inertial reference frame. The amplitude of oscillation of \( K_2 \) can then be analytically computed, and the octupole-active window (the range of \( I_0 \) over which orbit flips occur) is the region for which the range of these oscillations encompasses \( K_2 = 0 \) (Katz et al. 2011).

\(< K_2 = 0 \) (Katz et al. 2011).

When \( \omega \) is librating instead, \( \Omega_2 \) jumps by \( \approx 180^\circ \) every ZLK cycle, and the oscillations in \( K_2 \) are suppressed.

In the finite-\( \eta \) case, we commented in Section 2.3 that the relation between \( K_2 \) oscillations and extreme eccentricity excitation (and orbit flipping) can be generalized even when \( \eta \) is non-zero. \( K_2 \) still oscillates over time-scales \( \gg t_{\text{ZLK}} \) when \( \omega \) is circulating, and if its range of oscillation contains \( K_2 \equiv -\eta/2 \), then the inner orbit flips, in the process attaining extreme eccentricities. To be precise, orbit flips are...
defined to be when the range of inclination oscillations changes from $(\cos I_0)_- < \cos I < \cos I_{0, \lim}$ to $\cos I_{0, \lim} < \cos I < (\cos I_0)_+$ or vice versa, where $(\cos I_0)_{\pm}$ are given by equation (10) and $I_{0, \lim}$ satisfies equation (8).

However, the range of oscillation of $K$ is more complex than it is in the test-particle limit. Fig. A1 compares the behaviour of $e_{\max}$ in the non-dissipative simulations (top panel; reproduced from the top panel of Fig. 6) to the range of oscillations in $K$ (bottom panel). Denote the centre of the gap $I_{0, \text{gap}}$ (shown as the vertical black line in both panels of Fig. A1). Near $I_{0, \text{gap}}$, $K$ oscillates about $K(I_{0, \text{gap}})$, which is positive, and the oscillation amplitude goes to zero at $I_{0, \text{gap}}$. On the other hand, orbit flips (and extreme eccentricity excitation) are possible when the range of oscillation of $K$ encloses $K_c$ (i.e. $K_{\text{min}} < K_c < K_{\text{max}}$).

The purple shaded regions in both panels of Fig. A1 illustrate this equivalence, as they show both the $e_{\text{max}}$-attaining inclinations in the top panel and the inclinations where $K_{\text{min}} < K_c < K_{\text{max}}$ in the bottom panel. But since $K(I_{0, \text{gap}}) > 0$ while $K_c < 0$, there will always be a range of $I_0$ about $I_{0, \text{gap}}$ for which the oscillation amplitude is smaller than $K(I_{0, \text{gap}}) - K_c$, and orbit flips are impossible in this range. This range then corresponds to the octupole-inactive gap.

This analysis has simply pushed our lack of understanding on to a new quantity: Why are $K$ oscillations suppressed in the neighborhood of $I_{0, \text{gap}}$? A quantitative answer to this question is beyond the scope of this paper, but for a qualitative understanding, we can examine the evolution of a system in the octupole-inactive gap. The left-hand panel of Fig. A2 shows the same simulation as Fig. 3 but with an additional panel showing $\Omega_c$, while the right-hand panel shows a simulation with the same parameters except $I_0 = 88^\circ$, which is near $I_{0, \text{gap}}$ (see Fig. A1). The oscillations in $K$ (third panels) are much smaller for $I_0 = 88^\circ$ than for $I_0 = 93.5^\circ$, and no orbit flips occur. Most interestingly, the fourth panel shows that the evolution of $\Omega_c$ is much less smooth than in Fig. 3, jumping at almost every other eccentricity maximum. Katz et al. (2011) have already pointed out that jumps in $\Omega_c$ occur when $\omega$ is librating, rather than circulating.

When the octupole-order terms are neglected, the circulation-libration boundary is a boundary in the $e$–$\omega$ space: As long as the ZLK separatrix exists in the $e$–$\omega$ plane and $e_0 > 0$, then an initial $\omega_0 = 0$ causes $\omega$ to circulate, while an initial $\omega_0 = \pi/2$ causes $\omega$ to librate (e.g. Kinoshita 1993; Shevchenko 2016). However, when including octupole-order terms, this picture breaks down. To illustrate this, for a range of $I_0$ and both $\omega_0 = 0$ and $\pi$, we evolve the fiducial system parameters for a single ZLK cycle, using $g = 0.2$ as is used for Figs A1 and A2, and consider both the dynamics with and without the octupole-order terms. Fig. A3 gives the resulting changes in $\Omega_c$ over a single ZLK period when the octupole-order effects are neglected (top panels) and when they are not (bottom panels). Two observations can be made: (i) $I_{0, \text{gap}}$ is approximately where $\Delta \Omega_c = 0$ for circulating initial conditions when neglecting octupole-order terms, and (ii) the inclusion of the octupole-order terms seem to cause $\Omega_c$ to exclusively vary slowly ($|\Delta \Omega_c| \ll 180^\circ$) except for $I_{0, \text{gap}} < I_0 < I_{0, \lim}$. The former is plausible: If $K(I_{0, \text{gap}})$ is the location of an equilibrium in the $K$–$\Omega_c$ space, then it must satisfy $\Delta \Omega_c = 0$. The latter suggests that the assumption of circulation of $\omega$ in Katz et al. (2011) may be satisfied for many more initial conditions than...
the quadrupole-level analysis suggests, as long as they are not in octupoleinactive gap.

Finally, examination of the bottom panel of Fig. A1 suggests that the oscillation amplitude in $K$ grows roughly linearly with $|I_0 - I_{0, \text{gap}}|$ in the vicinity of $I_{0, \text{gap}}$ (this may be because, in the test-particle limit, librating $\omega$ gives oscillation amplitudes in $K$ that are higher order in $K$ and $1/\Omega_1 e$, as pointed out by Katz et al. 2011). Assuming this, the gap width can then be given by

$$\text{Gap Width} = 2 \left( I_{0, \lim} - I_{0, \text{gap}} \right).$$

This explains why the gap does not exist in the test-particle regime, as $I_{0, \lim} = I_{0, \text{gap}} = 90^\circ$ by symmetry of the equations of motion.

It is clear from the preceding discussion and Fig. A3 that the octupole-order, finite-$\eta$ dynamics are complex, and our discussion can only be considered heuristic. Nevertheless, in the absence of a closed form solution to the octupole-order ZLK equations of motion or a full generalization of the work of Katz et al. (2011), they provide a preliminary understanding of the octupole-inactive gap.
Figure A3. Plot of $\Delta \Omega_e$, the change in $\Omega_e$ over a single ZLK cycle, for $q = 0.2$ and the fiducial parameters using different initial conditions. In the top panel, octupole-order terms are neglected, while in the lower panel, they are not. The solid and dashed vertical black lines denote $I_{0, \text{gap}}$ and $I_{0, \text{lim}}$, respectively.

This paper has been typeset from a TEX/LATEX file prepared by the author.