

Statistical study for binary star evolution in dense embedded clusters

Wenjie Wu^{1,*}, Pavel Kroupa^{1,2}, and Vikrant V. Jadhav²

¹ Helmholtz-Institut für Strahlen- und Kernphysik, Universität Bonn, Nussallee 14-16, 53115 Bonn, Germany

² Astronomical Institute, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Praha 8, Czech Republic

Received 4 September 2025 / Accepted 23 January 2026

ABSTRACT

Context. The dynamical evolution of binary populations in embedded star clusters shapes the statistical properties of binaries observed in the Galactic field. Accurately modelling this process requires resolving both the early cluster dynamics and binary interactions.

Aims. We aim to characterize the early dynamical evolution of primordial binaries in embedded clusters and identify the key parameters that govern binary survival and disruption.

Methods. We performed a set of direct N -body simulations starting from 100% primordial binaries in a time-varying gas potential of a gas-embedded cluster. To describe the evolution of binary orbital properties, we defined the empirical dynamical operators for the period, binding energy, and mass ratio. We then calibrated them across the simulated ensemble.

Results. The binding energy and orbital period evolve in a consistent, sigmoidal fashion. Their dynamical operators reveal that hard binaries heat the cluster and suppress wide binary formation, while a small residual population of soft binaries survives. The evolution of the mass-ratio distribution is less directly linked to dynamical processing and more shaped by internal processes such as stellar physics process in the pre-main sequence phase. High- q systems tend to be enhanced, while low- q systems are prone to disruption.

Conclusions. Binary evolution in clusters is primarily governed by binding energy and orbital period. Our model displays an improvement over previous parameterizations of the dynamical operator by allowing for the existence of wide binaries and incorporating the embedded cluster phase. For individual clusters, direct N -body modelling remains the only reliable approach. On Galactic scales, population synthesis methods based on the stellar dynamical operator approach developed in this work remain essential.

Key words. binaries: general – stars: kinematics and dynamics – open clusters and associations: general – galaxies: star clusters: general

1. Introduction

A key factor in the dynamical modelling of stellar population is the characterization of the binary population (BP), which comprises the distributions of binary parameters such as primary mass, orbital period, eccentricity, and mass ratio. Numerous studies have shown that the dynamical evolution of star clusters is significantly influenced by the presence and properties of binary systems (e.g. Heggie 1975; Kroupa 1995a,b; Parker et al. 2009; Kroupa & Petr-Gotzens 2011). In the context of stellar population synthesis, particularly when accounting for unresolved binaries, access to a reliable model for the BP is crucial for accurately reconstructing the stellar mass function (e.g. Kroupa et al. 1991; Marks & Kroupa 2011; Pang et al. 2024; Kroupa 2025).

It is widely accepted that the Galactic field stellar population originates primarily from the dissolution of embedded star clusters that form in molecular cloud clumps (e.g. Kroupa 1995a,b; Lada & Lada 2003; Bressert et al. 2010; Lada 2010; Kroupa 2011; Marks & Kroupa 2011; Kroupa 2025). Following the expulsion of residual gas, an embedded cluster rapidly loses its dynamical equilibrium, leading to the loss of a significant fraction of single and multiple stellar systems into the Galactic field. As shown in studies such as Kroupa et al. (2001) and Boily & Kroupa (2002), the cluster undergoes a phase of rapid expansion followed by re-virialization, ultimately reaching a gas-free state. These

post-gas-expulsion phases are less dynamically active than the deeply embedded phase, suggesting that the BP's dynamical evolution prior to gas expulsion warrants detailed investigation.

Direct observations of the BP remain challenging. Prior to gas expulsion, the residual gas and dust obscures stellar light, making direct observation nearly impossible. During the rapid expansion phase, stars become observable, and several techniques may be employed. For close binaries with semi-major axes of $a < 10$ AU, spectroscopic monitoring and eclipsing binary analysis can be used to infer binary properties (e.g. Sana et al. 2012; Kounkel et al. 2019); however, these methods require long observational times and are less sensitive to low mass ratios. Interferometry, sparse aperture masking, high-resolution imaging, and astrometry all have the capacity to detect binaries at intermediate separations ($a \approx 1$ –300 AU), depending on the distance to the system (e.g. Sana et al. 2014; De Furio et al. 2019). However, these techniques are often infeasible for distant star-forming regions. For systems with high mass ratios or wide separations, analyses using colour–magnitude diagrams (CMDs) and proper motion data are effective (e.g. Tokovinin & Briceño 2020; Jadhav et al. 2021), although such methods also have clear limitations. As a result, performing detailed modelling procedures of the BP using N -body simulations becomes essential.

Kroupa (1995a,b) introduced the concept of inverse dynamical population synthesis by performing a series of direct N -body

* Corresponding author: wwu@astro.uni-bonn.de

simulations with 200 primordial binaries using the `nbody5` code (Aarseth 1999a). In these calculations, the binaries were initially embedded in a bound stellar cluster, and their dynamical evolution was followed self-consistently. This work led to the formulation of the initial binary star distribution functions. The resulting changes in the BP invoked by the stellar-dynamical processing of the BP were subsequently encapsulated through operators acting on the corresponding distribution functions, as derived from the simulation outcomes (see also Kroupa 2002). Building upon this approach, Marks et al. (2011) extended the range of initial cluster parameters, such as the initial half-mass radius and cluster mass (up to $10^{3.5} M_{\odot}$, corresponding to approximately 6000 stars or 3000 binaries), and investigated how these parameters influence the dynamical operators. These studies laid the groundwork for the development of the binary population synthesis code BiPoS1 (Marks et al. 2012; Dabringhausen et al. 2022), which enables binary populations to be inferred on galactic scales.

Modelling the dynamical evolution of binary-rich, high-mass star-forming regions such as the Orion Nebula Cluster (ONC), which contains on the order of 10^4 stars and brown dwarfs, remains a major challenge for direct N -body simulations. Moreover, the deeply embedded phase of open clusters likely had significantly higher masses than observed today. For instance, Kroupa et al. (2001, see also Safaei et al. 2025) suggest that the initial mass in stars and brown dwarfs of the Pleiades was between 3000 and 4000 M_{\odot} , compared to its present-day mass of roughly 800 M_{\odot} . Since the dynamical evolution of binary systems is primarily driven by close encounters rather than by the overall cluster potential, a direct N -body method needs to be employed to quantify the dynamical processing of the BP (see e.g. Spurzem & Kamlah 2023).

As discussed in Wang et al. (2020a), direct N -body codes such as `nbody6` and its parallelized extension `nbody6++GPU` (Wang et al. 2015) face significant performance bottlenecks when simulating binary-rich dense clusters, due to inefficient parallelization in handling close encounters. Consequently, previous applications of inverse dynamical population synthesis have been limited to clusters with masses below $10^{3.5} M_{\odot}$. Wang et al. (2020a) introduced a new direct N -body code, PeTar, which incorporates highly efficient parallel algorithms for integrating close systems. As a result, PeTar enables a realistic and scalable modelling of binary-rich clusters in regimes previously inaccessible to other direct N -body frameworks.

The primary objective of this paper is to perform direct N -body simulations using PeTar, incorporating the most up-to-date findings from studies of the initial binary population (IBP). We considered an analytical model that isolates the dominant dynamical effects of residual gas during the deeply embedded phase of massive open clusters. The analysis was conducted within the framework of inverse dynamical population synthesis. We began by examining the distributions of commonly studied binary parameters (i.e. orbital period, semi-major axis, eccentricity, and mass ratio). In addition, we extended the analysis to investigate the dependence of binary properties on primary mass and the binary fraction across different mass regimes.

The paper is organized as follows. Section 2 outlines the initial conditions and N -body set-up and Sect. 3 introduces the binary distribution functions and dynamical operators. The main results are presented in Sect. 4, followed by our conclusions in Sect. 5. The appendices provide supporting analyses of the period distribution, parameter fits and correlations, and a comparison with the analytic framework of Marks et al. (2011).

2. Method

In this study, we employed the state-of-the-art N -body code PeTar (Wang et al. 2020a) to conduct N -body simulations. In this section, we describe the details of the set-up of the initial state and the N -body integration.

2.1. Setup of the initial configuration

The initial configurations for our model were generated using the `McLuster` code (Küpper et al. 2011), which supports a wide range of initial conditions suitable for generating stellar distributions for N -body simulations. Below, we describe the specific configuration used in our study.

The cluster mass is set to $M_{\text{ecl}} = 5000 M_{\odot}$ (the actual value varies slightly; see details below) and the half-mass radius is set to $R_{\text{h,m}} = 0.3$ pc. This value is motivated by the empirical $M_{\text{ecl}}-R_{\text{h}}$ relation from Marks et al. (2012), which yields $R_{\text{h}} \approx 0.303$ pc. In practice, this 1% deviation cannot be implemented exactly due to the finite number of stars. To obtain statistically reliable results, we generated 100 realizations of the cluster using different random seeds, treating them as samples from the same model. We note that the set-up procedure introduces deviations between the input and actual values of cluster parameters (e.g. $R_{\text{h,m}}$, M_{ecl}). We consistently used the actual ones calculated directly from the generated snapshot parameters.

2.1.1. Initial mass function

We adopted the canonical initial mass function (IMF) $\xi(m) = dN/dm$ from Kroupa (2001), where dN is the infinitesimal number of stars in the stellar mass range m to $m + dm$. It follows a two-part power-law form (up to a normalization constant), expressed as

$$\xi(m) \propto \begin{cases} m^{-\alpha_1} & m_{\text{min}} \leq m < m_1, \\ k_{\xi} m^{-\alpha_2} & m_1 \leq m \leq m_{\text{max}}, \end{cases} \quad (1)$$

where $\alpha_1 = 1.3$, $\alpha_2 = 2.3$, $m_{\text{min}} = 0.08 M_{\odot}$, and $m_1 = 0.5 M_{\odot}$. The constant $k_{\xi} = m_1^{\alpha_2 - \alpha_1}$ ensures continuity at the break mass, m_1 .

For the upper mass limit, m_{max} , we adopted the empirical $m_{\text{max}}-M_{\text{ecl}}$ relation from Weidner & Kroupa (2006) and Pflamm-Altenburg & Kroupa (2007, see also Yan et al. 2023) giving $m_{\text{max}} = 89.73 M_{\odot}$. To suppress any stochastic variations in the number of massive stars, which can significantly affect the dynamical evolution (see e.g. Wang & Jerabkova 2021), we used the optimal sampling method introduced by Kroupa et al. (2013). In conventional random sampling, a uniform random variable $X \in (0, 1]$ can be drawn and the corresponding stellar mass, m_{\star} , is then found by solving

$$\int_{m_{\text{min}}}^{m_{\star}} \xi(m) dm = X.$$

Here, X represents the quantile corresponding to m_{\star} . In contrast, optimal sampling uses equal spaced values of $X \in (0, 1]$, with the number of grid points determined by $N = M_{\text{cl}}/\langle m \rangle$, where $\langle m \rangle$ is the average stellar mass computed from Eq. (1). Thus, an optimal sampling ensures that the resulting mass array is deterministic and identical for a fixed cluster mass. A complete physical motivation and theoretical discussion can be found in Kroupa et al. (2013, 2026); Gjergo et al. (2026); here, we focus on its practical application, rather than its underlying theory.

2.1.2. Initial binary population

We adopted a primordial binary fraction of 100%. As discussed in Kroupa (1995a) and emphasized by Thies et al. (2015) and Kroupa (2025), this assumption provides a consistent basis for exploring whether a single birth configuration can evolve into a binary population similar to that observed in the Galactic field and in star-forming regions. We followed the same hypothesis to ensure consistency across our model. The IBP was assigned separately for low-mass binaries (primary mass of $<5 M_{\odot}$) and high-mass binaries (primary mass of $>5 M_{\odot}$). Prior to pairing stars, a single mass-array was generated from the IMF (see Sect. 2.1.1). It is of crucial importance to draw only one primary and secondary component mass from the mass array to avoid affecting the IMF.

For the low-mass binaries, we adopted the prescription of Belloni et al. (2017). The primary mass, m_p , and secondary mass, m_s , were randomly and independently selected from the stellar mass-array, considering only stars with $m < 5 M_{\odot}$ for pairing. For each cluster model, the pairs differ, despite the models having the same mass array. The orbital period, P , can be drawn from the distribution in Kroupa (1995c, i.e. Eq. (8) in that work), while the eccentricity, e , is then drawn from a thermal distribution ($\propto 2e$), while the orbital phase and inclination are sampled from uniform distributions on $[0, 2\pi)$ and the unit sphere, respectively.

For high-mass binaries, motivated by the observed dynamical ejections of OB stars (Beccari et al. 2017; Kroupa et al. 2018; Jerabkova et al. 2019), a separate pairing procedure was adopted following Kroupa et al. (2018), as implemented by Long Wang in the updated version of McLUster (see Wang et al. 2019). The mass ratios are generated by choosing the secondary from the mass-array, so that the mass-ratio distribution is consistent with the observational data (see Sana et al. 2012), based on O-star observations in six nearby Galactic open clusters. The period and eccentricity are drawn from the empirical distributions provided in Oh et al. (2015) and Belloni et al. (2017).

We also adopted the pre-main sequence eigenevolution model introduced by Kroupa (1995b) and updated in Belloni et al. (2017). This process modifies the mass ratio and eccentricity (and, accordingly, the period), accounting for mass transfer and circularization during the pre-main sequence phase (spanning about first 10^5 yr). The pre-main sequence eigenevolution modifies the binary population on a system-by-system basis, transforming the birth binary population into the IBP. However, this procedure breaks the analytical properties of the original mass distribution and introduces stochasticity. As a result, even under optimal sampling, the final mass array differs slightly between realizations and the actual cluster mass, M_{ecl} , might end up exceeding the nominal input value.

2.1.3. Initial phase-space distribution

We initialized the positions and velocities of all binary centre-of-mass (c.m.) systems, using a Plummer model (Plummer 1911) and treating each binary as a single particle. The positions and velocities of individual stars were then computed from their binary orbital parameters and the c.m. positions and velocities. In this study, we did not assume any primordial mass segregation. After generating each cluster, we rescaled the velocities (only the c.m. velocities for binaries) to ensure virial equilibrium (following Aarseth 1999b) with the gas potential (see Sect. 2.2), such that the virial ratio is $Q = 0.5$.

We note that our set-up assumes the embedded cluster to be initially in a state of virial equilibrium. Recent studies on BPs

have explored non-equilibrium initial conditions, such as fractal structures (e.g. Dorval et al. 2017), showing that the environment can significantly influence the dynamical evolution of BPs. However, as pointed out by Kroupa et al. (2001), the initial conditions of embedded clusters are subject to physical constraints. In reality, star formation proceeds over about a Myr until gas expulsion. Once a star forms, it either falls towards the cluster's potential minimum or it is already engaged in a 'virial-equilibrium' orbit (see also Moeckel & Clarke 2011). In either case, the accumulating stellar system is continuously virializing as new stars are added, since the crossing timescale of an embedded cluster (see Sect. 4.1) is much shorter than its formation timescale. Therefore, a young embedded cluster is most likely to be close to virial equilibrium.

2.2. The N -body integration

PeTar combines several advanced computational techniques to handle interactions across different distance scales: the Barnes-Hut tree algorithm (Barnes & Hut 1986) for long-range interactions, a fourth-order Hermite integrator with block time-step (e.g. Aarseth 2003) for intermediate-range interactions, and the slow-down algorithmic regularization method (Wang et al. 2020c) for short-range interactions.

PeTar integrates the SSE/BSE code (Hurley et al. 2000, 2002; Banerjee et al. 2020) to model both single and binary stellar evolution. The remnant mass is determined based on the rapid supernova scenario from Fryer et al. (2012). The calculations also include contributions from pair-instability (Belczynski et al. 2016) and electron-capture supernovae (Belczynski et al. 2008). For all models, we assumed a solar metallicity of $Z = 0.02$.

The galpy library (Bovy 2015) has been integrated into PeTar as an extension to provide various types of external fields. To simplify the model, the model clusters can be set to orbit on a round orbit around a point-mass ($M_{\text{gal}} = 5 \times 10^{10} M_{\odot}$) at a distance $D = 8.5$ kpc.

The gas was modelled as a static Plummer potential in the cluster-centric frame (see e.g. Kroupa et al. 2001; Banerjee & Kroupa 2013). The gas mass was set to the mass of the embedded cluster divided by the star formation efficiency (SFE). Following the previous N -body modelling process and to reduce the number of parameters, we assumed $\text{SFE} = 1/3$. Both the stellar and gas potentials share the same Plummer radius. As discussed in Kroupa et al. (2001), the gas expulsion starts with a delay after the star formation phase, which is typically 0.6 Myr. Therefore, we chose the end time of all simulations to be 0.6 Myr, because the cluster expansion emerging thereafter largely freezes the binary population. Gas expulsion drives substantial cluster expansion, while the dynamical environment becomes less favourable for further binary processing beyond this point. Thus, limiting the integration time to 0.6 Myr allows us to isolate the effects of early cluster dynamics prior to the onset of gas-driven expansion.

We acknowledge that this model omits detailed star-gas interactions, such as dynamical friction with the gas and feedback-driven gas removal. In particular, gas dynamical friction may influence the evolution of high-mass binaries, potentially leading to tighter configurations or enhanced disruption rates. Recent studies have shown that gas interactions can harden binaries and extend the traditional soft-hard boundary (Rozner & Perets 2024), while gas-assisted capture provides an efficient channel for the formation of new binaries, especially among massive stars in gas-rich environments (Rozner et al. 2023). Such effects tend to be especially pronounced for massive or compact

binaries. However, high-mass binaries constitute a minor fraction of the population in this study. As shown in Sect. 4.4, most dynamical modifications occur among low-mass binaries. Therefore, we conclude that our gas model is sufficient to capture the dominant dynamical effects relevant to the scope of this work.

All computations were performed using 8 CPU cores with SIMD acceleration on the MARVIN cluster, hosted by the University of Bonn. By leveraging the HPC system's capability to run multiple samples in parallel, we completed the entire set of simulations within approximately one week.

3. Binary distribution functions and the dynamical operator

We adopted the framework established by Kroupa (1995a) and Marks et al. (2011) to describe the dynamical evolution of binary star populations. In this section, we summarize the mathematical definitions and notations.

At any given time, t , the distribution of binary parameters (i.e. the absolute value of the binding energy, E_b , orbital period, P , semi-major axis, a , eccentricity, e , primary mass, m_p , secondary mass, m_s , and mass ratio, $q = m_s/m_p$) is described by the individual binary distribution function (BDF), denoted as $\Phi_{\pi_i}^t$. Here, π_i represents one of the binary parameters. The BDF characterizes the distribution of binary parameters within a population, but is not equivalent to a probability density function (PDF); specifically, it is not normalized to unity. Instead, the integral of the BDF over the entire domain of π_i yields the total binary fraction, f_{bin}^t , of the cluster at time, t :

$$\int_{\pi_{i,\text{min}}}^{\pi_{i,\text{max}}} \Phi_{\pi_i}^t(\pi_i') d\pi_i' = f_{\text{bin}}^t. \quad (2)$$

The time evolution of the BDF from an initial time t_0 to a later time, t_{end} , is modelled using a dynamical operator $\Omega_{\pi_i}(\pi_i; \omega_j)$, where ω_j is a set of fitting parameters that characterize the dynamical processes, such as \mathcal{E}_{cut} and \mathcal{A} , as defined in Eq. (20) of Marks et al. (2011). The evolved BDF at time, t_{end} , is given by

$$\Phi_{\pi_i}^{t_{\text{end}}}(\pi_i; \omega_j) = \Omega_{\pi_i}(\pi_i; \omega_j) \cdot \Phi_{\pi_i}^{t_0}(\pi_i). \quad (3)$$

While a BDF shares some formal similarities with a PDF, the key distinction lies in normalization: a PDF integrates to unity, whereas a BDF integrates to the binary fraction. To empirically determine the BDF, we first estimated the empirical PDF and then scaled it by the observed binary fraction. Consequently, the empirical dynamical operator can be directly computed as the ratio of the empirical BDFs at t_{end} and t_0 .

The functional form of the dynamical operator, Ω_{π_i} is chosen empirically, typically guided by the shape of the measured empirical operator. Physically, $\Omega_{\pi_i}(\pi_i)$ can be interpreted as an importance function, representing the survival probability of binaries with parameter, π_i , after dynamical processing.

A closed-form expression for the initial BDF is generally not available, primarily due to the inclusion of pre-main sequence eigenevolution. This process alters parameters such as eccentricity based on the birth pericentre, which itself depends on the birth period and eccentricity. Consequently, the final BDF, $\Phi_{\pi_i}^{t_{\text{end}}}$, is difficult to derive analytically, even when the form of Ω_{π_i} is known. This complexity renders standard maximum likelihood estimation impractical for fitting ω_j . To overcome this challenge, we employed the Kullback–Leibler importance estimation procedure to determine the parameters, ω_j . The theoretical basis for

this method is detailed in, for example, Sugiyama et al. (2007). In brief, we maximized the following objective function,

$$\ell(\omega_i) = \frac{1}{N_{t_{\text{end}}}} \sum_{k=1}^{N_{t_{\text{end}}}} \log \left(\Omega_{\pi_i}(\pi_{i,t_{\text{end}}}^{(k)}; \omega_j) \right), \quad (4)$$

subject to the normalization constraint:

$$\frac{1}{N_{t_0}} \sum_{k=1}^{N_{t_0}} \Omega_{\pi_i}(\pi_{i,t_0}^{(k)}; \omega_j) = f_{\text{bin}}^{t_{\text{end}}}. \quad (5)$$

Here, $\pi_{i,t_0}^{(k)}$ and $\pi_{i,t_{\text{end}}}^{(k)}$ denote individual binary systems sampled at times, t_0 and t_{end} , respectively, with N_{t_0} and $N_{t_{\text{end}}}$ representing the number of binaries in the corresponding snapshots. This approach avoids the need for explicit analytical forms of the BDFs, thereby reducing the complexity of dynamical effect modelling.

4. Results and discussion

To analyse the simulation results, we first defined the cluster boundary. In the presence of a point-mass galactic tidal field, the tidal radius serves as a natural choice; it marks the region where the galactic potential overtakes the cluster's own potential. However, the inclusion of gas as part of the cluster's gravitational potential significantly increases the tidal radius compared to the gas-free case, covering a large volume where the stellar density is very low (less than $1 M_{\odot} \text{pc}^{-3}$). A system in this space has a very small probability of interacting with other systems and is therefore very dynamically inactive. To define a more meaningful boundary, we used local stellar density as the criterion instead.

Following Casertano & Hut (1985), we define the local density for each star as the mass density within a sphere that includes its six nearest neighbours. Stars with local densities below $1 M_{\odot} \text{pc}^{-3}$ were excluded in the calculation of the cluster's centre and the boundary of the cluster. The cluster centre was then computed as the local-mass-density-weighted average position of the remaining stars, while the cut-off radius, R_{cut} , is defined as the distance from this centre to the furthest remaining star. This approach assumes that, under isotropy, the local density reflects the global density profile. In our simulations, R_{cut} typically lies between 2 and 3 pc. It can be determined from the final snapshot (at 0.6 Myr), which is the same value used for earlier snapshots, with the cluster centre recalculated accordingly. During the evolution, stars are ejected from the cluster due to close encounters and the cluster expands slightly (see Sect. 4.1). Thus, the R_{cut} value from the final snapshot is large enough to cover all dynamical active systems during the calculation. In the analysis of binary populations, we only considered stars within R_{cut} . At 0 Myr (i.e. the initial time), there are 30–50 stars out of R_{cut} for each sample. However, this will not affect the results, because we introduced no correlation between the position and binary parameters in the initial set-up. The half-mass radius and related quantities are determined using all stars, including those located beyond R_{cut} , with respect to the updated cluster centre.

Binary systems were identified following the method in Kroupa (1995a,b), where we computed the binding energy between each star and its nearest neighbour. If the binding energy is negative, the pair is considered a binary. To detect higher order multiples, we treated the identified binaries as single particles (using their c.m. properties) and repeated the procedure to find

triple and quadruple systems. In this study, we included up to quadruples, but such systems are rare (no more than five per snapshot) and, thus, they were excluded from further analysis. In the following analysis, a ‘system’ denotes either a single star or a binary.

In this work, we did not aggregate the simulation outputs into a single combined dataset. Our focus is on the evolution of binary populations within individual star clusters, where local conditions and individual cluster evolution can lead to significant variation across realizations. Combining all models into one sample would smooth out these differences, making it difficult to isolate the physical mechanisms responsible for binary evolution on a cluster-by-cluster basis. Such an aggregation procedure poses the risk of introducing biases or masking dynamical signatures that are only apparent when each cluster is treated as an independent system. To preserve the individuality of each model and accurately capture robust dynamical trends, we analysed all realizations separately, while identifying parameters that exhibited consistent behaviour across the ensemble. We began by visually inspecting the distribution of various parameters across individual realizations, selecting those that displayed consistent behaviour. We then applied different binning methods to test whether the trends persist. For each candidate parameter, we defined a dynamical operator and apply importance sampling. Only parameters with consistently behaving operators across all samples were retained. Those not shown were either strongly correlated with others or too noisy to be modelled reliably.

4.1. Timescale relevant to binary evolution

Our model evolves the cluster only up to 0.6 Myr. A natural question arises regarding whether this short timespan is sufficient for significant changes in the BP. As demonstrated by Marks et al. (2011), binary populations can evolve substantially on timescales comparable to the crossing time t_{cr} , defined as the ratio of cluster diameter to the typical stellar velocity.

We adopt the half-mass radius, $R_{\text{h,m}}$, as the characteristic size and the root mean square of a single velocity component (i.e. 1D velocity dispersion, σ_{ecl} , of systems within that radius as the velocity scale. The crossing time is thus:

$$t_{\text{cr}} = \frac{2R_{\text{h,m}}}{\sigma_{\text{ecl}}}. \quad (6)$$

In our analysis, t_{cr} is calculated using the initial values of $R_{\text{h,m}}$ and σ_{ecl} . In our model, the gas potential increases the velocity dispersion, so we compute t_{cr} directly from Eq. (6) with the velocity dispersion being calculated using the c.m. velocities of systems within the half-mass radius. We also show the half-mass two-body relaxation time, t_{rh} . The definition follows from e.g. Meylan & Heggie (1997),

$$\frac{t_{\text{rh}}}{t_{\text{cr}}} \approx 0.138 \left(\frac{R_{\text{h,m}}}{2R_{\text{vir}}} \right)^{3/2} \frac{N}{\ln \Lambda}, \quad (7)$$

where R_{vir} is the virial radius of the cluster, in our case (Plummer sphere) $R_{\text{h,m}} \approx 0.77R_{\text{vir}}$, N is the number of systems within $R_{\text{h,m}}$, and $\ln \Lambda$ is the Coulomb logarithm. We used $\Lambda = 0.02N$ based on the measurement of Giersz & Heggie (1996).

Figure 1a shows the distribution of t_{cr} across all samples. Most clusters have $t_{\text{cr}} \approx 0.1$ Myr. Over 0.6 Myr, the system undergoes approximately five to six crossing times. As discussed

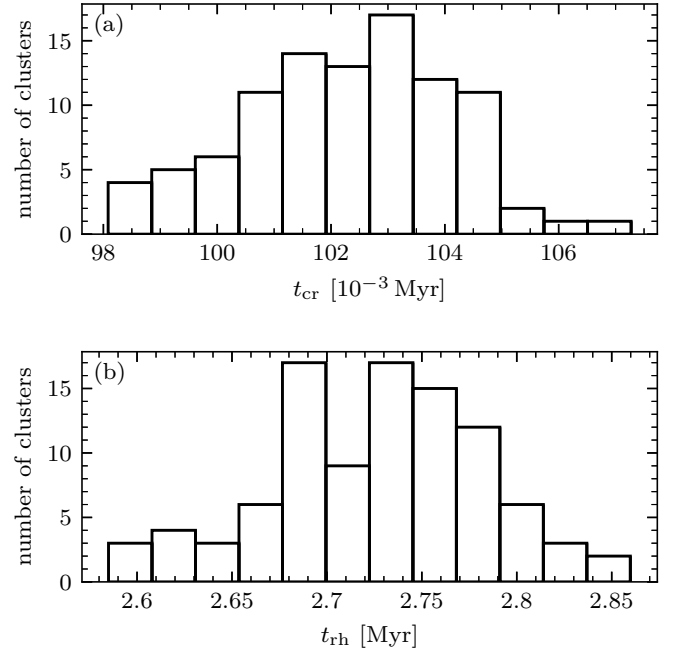


Fig. 1. Histogram of the crossing time, t_{cr} , (panel a) and the half-mass relaxation time, t_{rh} , (panel b) for all model clusters in the sample.

in Sect. 2.1, our set-up introduces minor sample-to-sample variations, which are reflected in the distribution of t_{cr} . The histogram exhibits a mean and median of 0.102 Myr. The skewness is -0.07 , with a Fisher kurtosis of -0.33 . Both values indicate that the distribution of t_{cr} is approximately normal, exhibiting a very slight left skew and a mildly flatter shape than the standard normal curve. In *McLuster*, the half-mass radius is constrained to match the specified value; that is, $R_{\text{h,m}} \equiv 0.3$ pc at the initial time. However, due to the random nature of the velocity distribution, each realization exhibits fluctuations in the velocity dispersion. Figure 1b also shows the histogram for t_{rh} . For t_{rh} , the number of systems within the half-mass radius (i.e. the mean system mass) introduces an additional source of dispersion. These types of dispersions arising from our initial set-up were not considered in the present study, but their potential influence on the dynamical evolution of the cluster and the binary populations should be explored in future works.

Figure 2 shows the time evolution of the half-mass radius, half-number radius, (1D) velocity dispersion, and binary fraction. The half-mass radius remains nearly constant, while the half-number radius increases by about 10%, indicating an evolution towards mass segregation. The velocity dispersion increases by 5%, further suggesting that the cluster structure remains largely unchanged over this timespan. However, the binary fraction decreases significantly (dropping to around 50%) indicating that the binary population evolves much faster than the global cluster structure because the relaxation time is ≈ 2.7 Myr.

4.2. Dynamical evolution of binding energy and orbital period

The binding energy is widely used to characterize dynamical processing in star clusters. For a binary, the ratio between its binding energy and the mean kinetic energy of stars in the environment provides a natural measure of its hardness. In this section we model the dynamical operators for both binding energy and the orbital period.

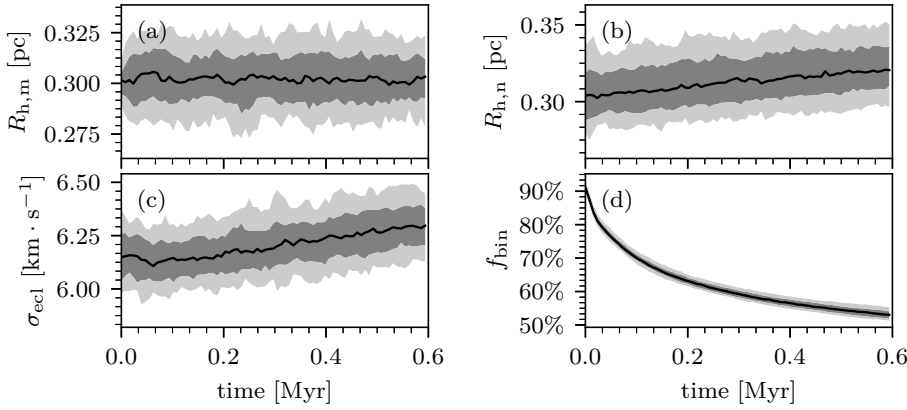


Fig. 2. Time evolution of global cluster properties: half-mass radius, $R_{h,m}$ (a); half-number radius, $R_{h,n}$ (b); 1D velocity dispersion within $R_{h,m}$, σ_{ecl} (c); and binary fraction (d). Black lines show the median. Dark and light grey areas mark the 16th–84th (68%) and 2.5th–97.5th (95%) percentiles, respectively.

The absolute value of binding energy of a binary is defined as

$$E_b = G \frac{m_p m_s}{2a}. \quad (8)$$

Marks et al. (2011) introduced a dynamical operator for E_b denoted Ω_{E_b} , parameterized by a sigmoid-like function in log-space,

$$\Omega_{E_b}(E_b) = \frac{\mathcal{A}'_E}{1 + \exp[\mathcal{S}'_E(\log_{10} E_b - \mathcal{E}'_{cut})]} - \frac{\mathcal{A}'_E}{2}, \quad (9)$$

for $\log_{10} E_b > \mathcal{E}'_{cut}$

with $\Omega_{E_b}(E_b) \equiv 0$ for $\log_{10} E_b \leq \mathcal{E}'_{cut}$. Here, \mathcal{A}'_E quantifies the total variation of the operator, \mathcal{S}'_E sets the steepness of the transition, and \mathcal{E}'_{cut} defines a sharp cutoff below which the operator vanishes. In this work, we adopted a similar functional form, but generalized it to allow smoother transitions and a finite asymptote at low binding energies, expressed as

$$\Omega_{E_b}(E_b) = \frac{\mathcal{A}_E}{1 + \exp[\mathcal{S}_E(\log_{10} E_b - \mathcal{E}_{cut})]} + \mathcal{B}_E. \quad (10)$$

Here, \mathcal{A}_E sets the overall amplitude, while \mathcal{S}_E and \mathcal{E}_{cut} control the slope and location of the transition. Unlike the formulation of Marks et al. (2011), we introduce an additional parameter, \mathcal{B}_E , which allows the operator to asymptotically approach a non-zero value at low E_b . This modification removes the sharp cutoff and accounts for the continuous presence of wide (soft) binaries, which are short-lived yet persist in quasi-equilibrium due to their continual replenishment through dynamical interactions.

Figure 3b shows the empirical (i.e. as measured from the N -body output data) dynamical operator Ω_{E_b} . At $E_b \gtrsim 10^3 \text{ M}_\odot \text{ pc}^2 \text{ Myr}^{-2}$, the number of binaries becomes very small (see Fig. 3a), and the inferred trends in this high-energy regime are unreliable. Such strongly bound (i.e. hard, systems) are either extremely compact (short-period) or contain very massive components; in both cases, stellar evolution dominates over dynamical interactions. We therefore restricted the fitting of Eq. (10) to the range $E_b \in [0, 10^3] \text{ M}_\odot \text{ pc}^2 \text{ Myr}^{-2}$. For each of the 100 realizations, parameter estimation was carried out independently. To illustrate the results clearly, we show only the fit obtained from the median parameter values as a solid grey line in Fig. 3b. The full set of fitting results is provided in Appendix B. For comparison with the formulation of Marks et al. (2011), we also performed a regression using Eq. (9), with the results and discussion presented in Appendix C.

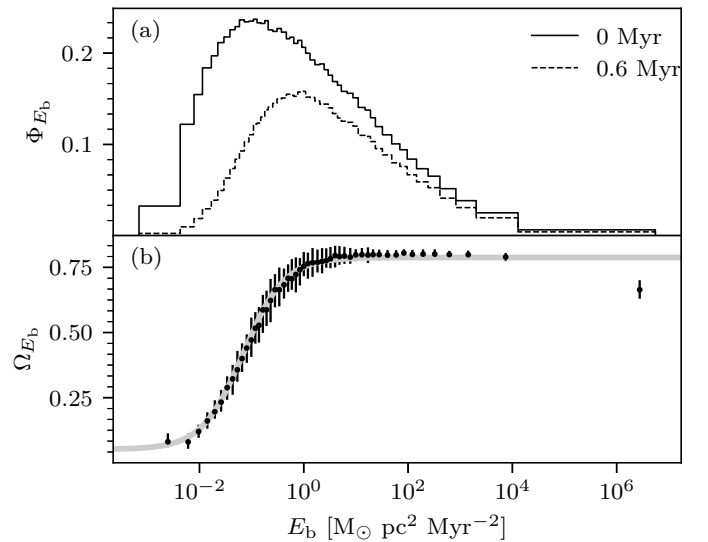


Fig. 3. (a) BDF of the absolute value of binding energy E_b at 0 Myr (solid) and 0.6 Myr (dashed). For all samples the same binning is used; curves show the median across all realizations. (b) Dynamical operator, Ω_{E_b} , as defined from the N -body data. Black dots indicate the median values, with error bars marking the 16th–84th percentiles. The solid grey line shows the median fit of Eq. (10).

We note that different combinations of \mathcal{E}_{cut} and \mathcal{S}_E could yield similar objective function values. This degeneracy arises because a larger $|\mathcal{S}_E|$ can compensate for a smaller \mathcal{E}_{cut} . In practice, however, the new model (Eq. (10)) exhibits greater robustness, since both the minimum and maximum values of Ω_{E_b} are explicitly constrained. We further find that performing an initial χ^2 fit to the histogram data (and using its outcome as a starting point for direct importance estimation) leads to more stable parameter recovery and reduces the risk of convergence to local minima.

As shown by Heggie (1975) and Hills (1975), three-body encounters drive a systematic evolution of binary hardness, encapsulated in the Heggie–Hills law: hard binaries become harder, soft binaries become softer. The classification into hard or soft is inherently relative. In a single encounter, a binary is considered soft if its binding energy is smaller than the kinetic energy of the incoming single star; in this case, the single star is slowed down and the binary’s binding energy decreases in magnitude. Conversely, a binary is considered to be hard if its binding energy exceeds that of the kinetic energy; after the interaction, the single star is accelerated and the binary becomes more tightly bound. In stellar clusters, this criterion is thus usually applied

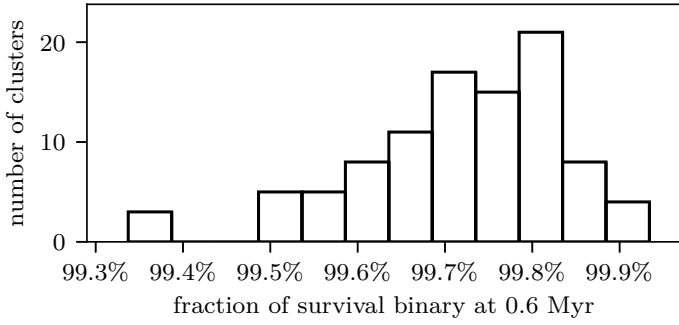


Fig. 4. Histogram of the survival fraction of binaries at 0.6 Myr for all model clusters. The survival fraction is defined as the number of binaries present at both 0 and 0.6 Myr, divided by the total number of binaries.

with respect to the average kinetic energy of single stars, such that the hard–soft boundary depends on the cluster velocity dispersion and evolves over time. Hard binaries act as an energy source, heating the cluster through single–binary interactions that accelerate stars. In contrast, soft binaries do not provide a sustained cooling effect (see e.g. Kroupa et al. 1999): after only a few encounters (typically within a crossing time), they end up disrupted. Taking the Heggie–Hills law into account, the dynamical operator is shown to effectively capture the net outcome of hardening, softening, mergers, disruptions, binary formation, and escape. To qualitatively assess the relative importance of these processes, we go on to examine each process separately below.

Formation. Each binary is assigned a unique identifier (as implemented in the PeTar analysis tool). A binary present at both 0 and 0.6 Myr is classified as a surviving binary. Figure 4 shows the survival fraction at 0.6 Myr, defined as the number of surviving systems divided by the total number of binaries. The result indicates that more than 99.6% of binaries are survivors, implying that binary formation during this period is negligible, as also shown by Kroupa & Burkert (2001). More complex processes, such as component exchange (where one member of a binary is replaced during an encounter) or disruption followed by subsequent re-formation, are subsumed under the formation channel in this work. These events are difficult to track in the simulations because resolving them would require output time intervals much shorter than the typical orbital periods of binaries. However, since fewer than 0.7% of binaries at 0.6 Myr are non-surviving systems, their impact on the overall results is negligible.

Merging. Mergers are treated by the stellar evolution routines in PeTar (SSE/BSE). From the log files we find, across all model clusters, a maximum of 25 merger events and an average of about 15 per simulation. Mergers occur only in very hard binaries with extremely large binding energies. These events account for the decline of Ω_{E_b} in the range 10^4 – 10^6 $M_\odot \text{pc}^2 \text{Myr}^{-2}$. Given that the maximum number of mergers (25) corresponds to only $\approx 0.6\%$ of the initial binary population (over 4000 systems), the overall impact of mergers is negligible. The merges are, however, of astrophysical interest as they lead to post-merger stars that do not fit for single star evolution synthesis and are likely to also show abnormal circumstellar material and magnetic activity (e.g. Korčáková et al. 2025; Dvořáková et al. 2024).

Hardening and softening. Since more than 99.3% of binaries at 0.6 Myr are survivors, the final BDF is shaped almost entirely

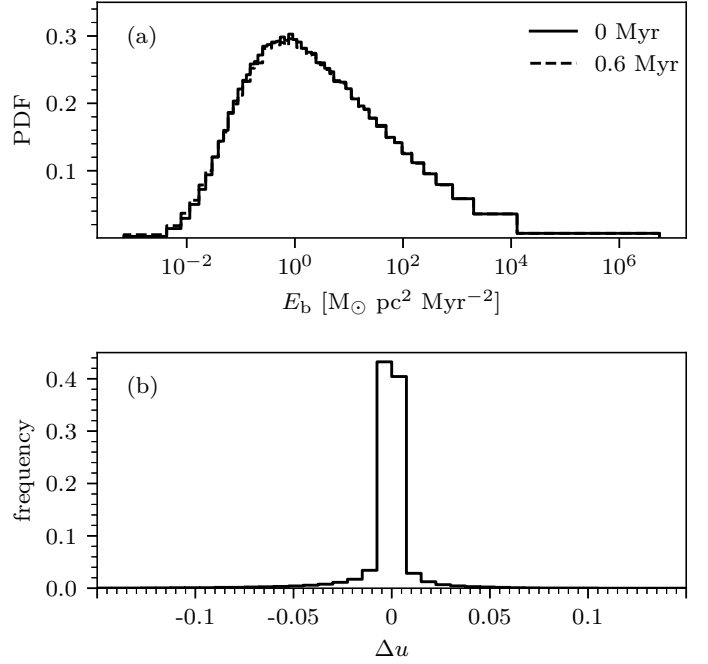


Fig. 5. (a) Empirical PDFs of the absolute value of the binding energy E_b for surviving binaries at 0 Myr (solid) and 0.6 Myr (dashed). The same binning is applied to all models, and the plotted histograms show, for each bin, the median value across all clusters. The two distributions are nearly indistinguishable. (b) Histogram of quantile shifts (Eq. (11)); see text for definition), where the bar heights represent the median values across all clusters.

by the survivor population at 0.6 Myr. In Fig. 5a, we compare the initial and evolved PDFs of E_b for the same set of surviving binaries; the two distributions are indistinguishable. A two-sample Kolmogorov–Smirnov test performed for each model cluster yields a minimum p -value of 0.96, indicating no statistically significant difference. To further examine whether substantial but mutually cancelling hardening and softening could nevertheless occur at the ‘individual’ level, we introduced a per-binary ‘quantile shift’. For each cluster, we first construct a single empirical cumulative distribution function (ECDF), $F(E_b)$, from the survivor sample. Given a binary with binding energy E'_b , its ECDF value $F(E'_b)$ corresponds to the fraction of systems with binding energy less than or equal to E'_b . By construction, $F(E_b)$ lies in $[0, 1]$ and represents the percentile rank of that binary within the overall distribution. We can then define the quantile shift between 0 and 0.6 Myr as

$$\Delta u \equiv F(E_b^{0.6 \text{ Myr}}) - F(E_b^{0 \text{ Myr}}), \quad (11)$$

so that $\Delta u > 0$ denotes hardening (a binary moves upward in the distribution) and $\Delta u < 0$ denotes softening. This quantile-based measure is insensitive to the absolute scale of E_b and remains robust across its wide dynamic range. As shown in Fig. 5b, the distribution of Δu is narrowly peaked around zero, with the overwhelming majority of binaries showing $|\Delta u| < 0.05$. If strong but balanced hardening and softening were present, we would expect a broad, symmetric spread of Δu values (many binaries shifting significantly upward and downward, even if the net PDF appears unchanged). The absence of such a spread rules out the possibility that substantial but compensating changes are hidden in the stable global PDF. We therefore conclude that, over 0–0.6 Myr, hardening and softening contribute negligibly to the evolution of the BDF.

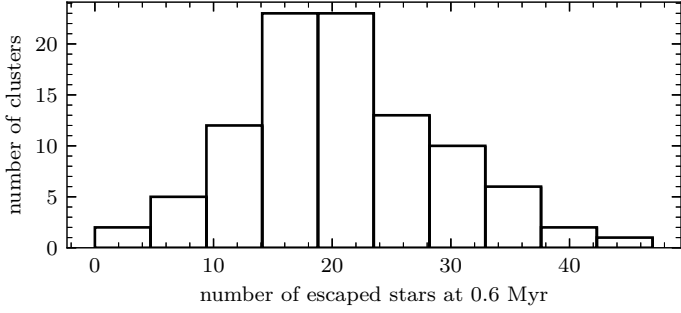


Fig. 6. Histogram for number of escaped stars at 0.6 Myr for all model clusters. The number is calculated as the difference between the number of stars within R_{cut} at $t = 0$ Myr and $t = 0.6$ Myr.

To directly and clearly visualize the hardening-softening evolution, Fig. 7 shows a two-dimensional (2D) density map of the absolute value of the binding energy for surviving binaries at 0 Myr and 0.6 Myr. Most surviving binaries maintain nearly unchanged binding energies, suggesting that dynamical evolution has only a minor impact on the overall orbital energy distribution. In contrast, binaries with relatively low binding energies exhibit more pronounced dynamical evolution, with both hardening and softening occurring more frequently than in tightly bound systems. The diagram further indicates that softening events are more common than hardening in our simulations. We expect that decreasing M_{ecl} would favour hardening over softening, since the lower velocity dispersion in such environments would enable more binaries to become hard systems.

Escape. Because the gas potential is included as part of the cluster, the overall gravitational potential is deeper than in the gas-free case modelled by Marks et al. (2011), and only a small number of escapers are expected. We define escapers as stars located beyond R_{cut} at 0.6 Myr. Across all model clusters, the number of escapers never exceeds 47 (Fig. 6). Compared to the total binary population of several thousand, this number is negligible. We therefore conclude that stellar escape does not play a significant role in shaping the evolution of the binary distribution at this stage.

Disruption. From the preceding discussion, we find that formation, merging, escape, and hardening-softening all have negligible impacts. Consequently, binary disruption emerges as the dominant dynamical effect within the first 0.6 Myr. Disruption efficiently removes soft binaries from the population and therefore shapes the low-binding-energy tail of the distribution. It is important to note, however, that our survival-probability-based approach imposes intrinsic limitations. By construction, it can only track binaries that remain bound and thus cannot account for the appearance of new wide systems. In other words, this method constrains the outcome within a restricted binding-energy range, but cannot reproduce the continuous replenishment of wide binaries through dynamical encounters. As a result, the present ansatz is best interpreted as valid only between lower and upper limits in binding energy, outside of which additional physics (particularly the formation and re-formation of wide binaries) must be invoked.

In summary, among the various dynamical processes considered (i.e. formation, merging, escape, and hardening-softening) only disruption has a significant impact. Accordingly, $\Omega_{E_b}(E_b)$ can be interpreted as the survival probability of binaries at a given binding energy, consistent with the formulation of Marks et al. (2011).

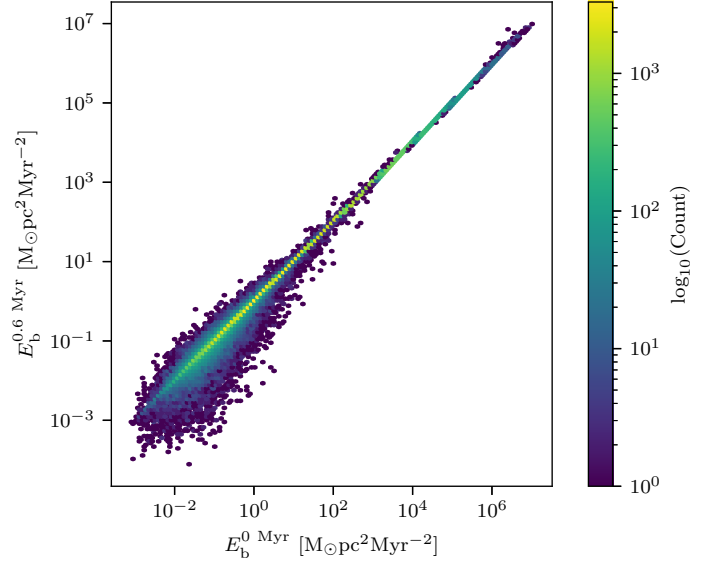


Fig. 7. 2D density map of the absolute value of the binding energy for surviving binaries at 0 Myr and 0.6 Myr. Colours indicate the number of binaries on a logarithmic scale. The main axes use an equal aspect ratio (identical scaling on both axes).

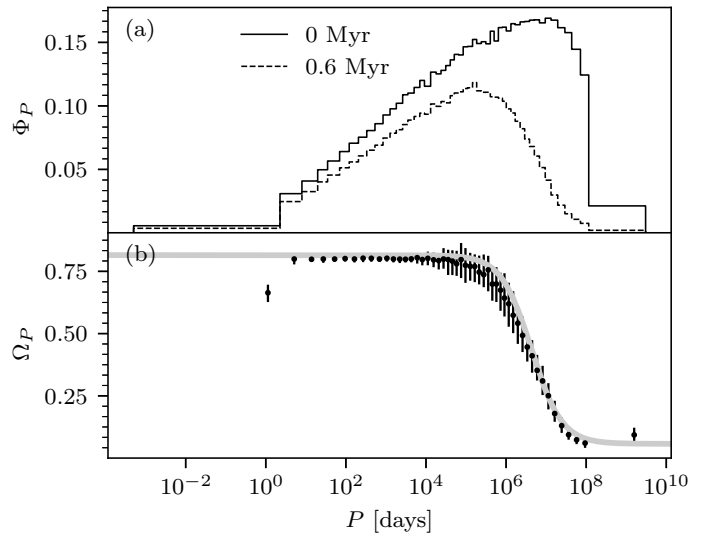


Fig. 8. Empirical BDF and dynamical operator for the orbital period, P . The panel layout is analogous to Fig. 3. A clear cut-off is seen around 10^7 days; model fitting is restricted to $P > 2$ days, where dynamical effects dominate over stellar evolution. A slight deviation is visible around $E_b \approx 10^5$ – 10^6 , reflecting a limitation of our model in accurately capturing the evolution of binaries in this period range.

Another key observable is the orbital period, P , which can be directly constrained from optical observations. Figure 8 shows the empirical BDF and dynamical operator for P , where a cut-off is evident around 10^7 days. We adopted the same functional form as in Eq. (10) for the dynamical operator, but with distinct parameters: \mathcal{A}_P , \mathcal{S}_P , \mathcal{P}_{cut} , and \mathcal{B}_P . For the fitting, we restricted the sample to $P > 2$ days, since systems with shorter periods are extremely tight and their evolution is dominated by stellar evolution rather than dynamics. In addition, we compare the initial and evolved PDFs of P for surviving binaries, as well as their quantile shift (see Appendix A). The results mirror those obtained for the binding energy: the distributions remain essentially unchanged, and disruption is the only significant dynamical

effect on the orbital-period distribution. Thus the period operator essentially encodes the survival probability in P , analogous to the binding-energy case.

Both the positive offsets \mathcal{B}_E and \mathcal{B}_P in the fitted dynamical operators imply that soft binaries are not disrupted instantaneously. Instead, they can persist transiently and, in some cases, they can escape intact if they form near the cluster periphery. Such escaping wide binaries might naturally contribute to the very wide binary population observed in the Galactic field.

4.3. Mass-ratio evolution

Unlike binding energy and orbital period, where the survival probability offers a relatively direct diagnostic of dynamical processing, the evolution of the mass-ratio distribution reflects a more complex interplay between internal stellar evolution, dynamical encounters, and pre-main sequence eigenevolution.

Figure 9 shows the BDF and empirical dynamical operator for q . Owing to the substantial scatter among different realizations, we do not attempt a parametric fit to the operator. Nevertheless, a clear trend emerges: binaries with low mass ratios ($q \lesssim 0.2$) are strongly suppressed, whereas those with nearly equal masses ($q \approx 1$) are preferentially retained, with the operator reaching values of ≈ 0.6 . In the intermediate regime, the operator remains approximately constant at ≈ 0.55 . This behaviour can be understood in the framework of pre-main sequence eigenevolution, in which the birth mass ratio, q_{bir} , of low-mass binaries is modified to the initial ratio q_{ini} according to (as described by Kroupa 1995a,b)

$$q_{\text{ini}} = \begin{cases} q_{\text{bir}} + \rho(1 - q_{\text{bir}}) & \rho \leq 1, \\ 1 & \rho > 1, \end{cases} \quad (12)$$

where

$$\rho = \left(\frac{\lambda R_{\odot}}{r_{\text{peri}}} \right)^{\chi}, \quad (13)$$

with $\lambda = 28$, $\chi = 3/4$, and r_{peri} the pericentre distance in units of solar radii. The corresponding modification of eccentricity is given by

$$\ln e_{\text{ini}} = -\rho + \ln e_{\text{bir}}. \quad (14)$$

These expressions indicate that long-period low-mass binaries experience minimal change in eccentricity due to pre-main sequence eigenevolution, while the mass ratio is strongly altered for close systems. We generate a large number of such systems with different primary masses ($m_p = 4.3, 2.0, 1.0, 0.21 M_{\odot}$) to illustrate how the initial period distribution depends on q . As shown in Fig. 10, binaries with $q \approx 1$ concentrate strongly at very short periods, where they may rapidly merge. Conversely, systems with very low q are limited to long periods (from thousands of years to about 1 Myr), which fall within the range of strong dynamical processing (cf. Fig. 8). Hence, we expect dynamical effects to preferentially modify low- q binaries. Furthermore, we find that as the primary mass decreases, the separation between period distributions corresponding to intermediate q values narrows, while the difference at very low q persists. This can be attributed to the limited range of physically possible q values for low-mass primaries due to the lower mass cut-off in the IMF. Overall, these results suggest that the apparent role of the mass ratio arises mainly from its correlation with binding energy established during pre-main sequence eigenevolution, rather than from any intrinsic dynamical effect of q itself.

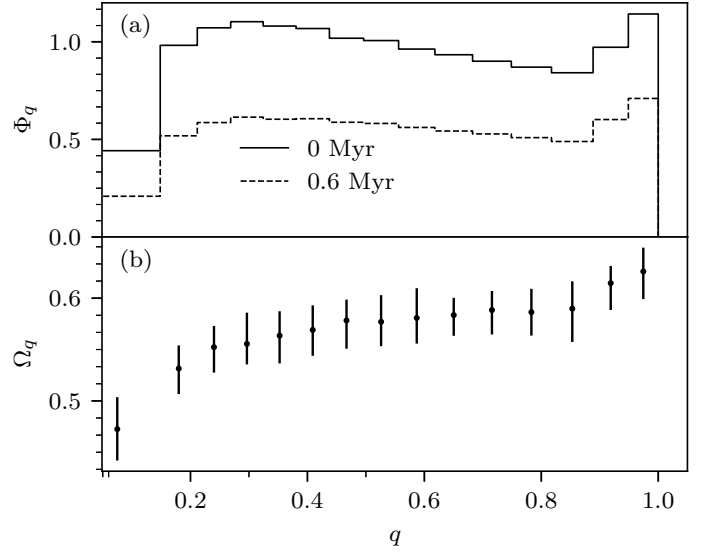


Fig. 9. (a) BDF of the mass ratio, q , at $t = 0$ Myr (solid) and $t = 0.6$ Myr (dashed). For each sample the same binning is used; curves show the median across all clusters. (b) Empirical dynamical operator, Ω_q . Black dots indicate the median values, with error bars marking the 16th–84th percentile range.

4.4. Mass dependence of the binary fraction

Observational studies (e.g. Kroupa et al. 2003; De Rosa et al. 2014; Offner et al. 2023) have consistently shown that the observed binary fraction of Galactic field stars depends strongly on the primary stellar mass. To quantify this relation, we define

$$f_{\text{bin}}(m_p) = \frac{N_{\text{bin}}(m_p)}{N_{\text{bin}}(m_p) + N_{\text{single}}(m_p)}, \quad (15)$$

where $N_{\text{bin}}(m_p)$ and $N_{\text{single}}(m_p)$ denote the number of binaries and single stars with primary mass m_p , respectively. Here we consider only binary systems, and neglect higher-order multiples (triples, quadruples, etc.). This definition is directly comparable to observations: when a sample is constructed by selecting stars within a given mass range and binaries are classified according to their primary mass, the measured binary fraction corresponds to $f_{\text{bin}}(m_p)$ as defined above. In this sense, $f_{\text{bin}}(m_p)$ represents the probability that a star of mass m_p has a companion of lower mass ($q \leq 1$). By construction of our initial conditions, all stars are assumed to reside in binaries, such that $f_{\text{bin}}(m_p) \equiv 1$ at $t = 0$ Myr. The subsequent time evolution of f_{bin} therefore directly reflects the combined influence of dynamical processes, and provides a natural basis for comparison with observational constraints.

Figure 11 shows the measurement of $f_{\text{bin}}(m_p)$ from the model clusters at 0.6 Myr. A clear dependence on primary mass is evident. For $m_p > 5 M_{\odot}$, binaries are dynamically stable and rarely disrupted, since their large absolute value of binding energies place them firmly in the hard-binary regime. As discussed in Sect. 4.2, such systems are only affected by mergers. This result is consistent with observations indicating that nearly all massive stars reside in binaries (see e.g. Offner et al. 2023). For $m_p < 5 M_{\odot}$, the binary fraction rises steeply with increasing mass, saturating near 1 M_{\odot} and approaching an asymptotic value of $\approx 65\%$. Above $\approx 0.3 M_{\odot}$, however, in our model results, the growth of f_{bin} slows considerably, indicating a turnover in the low-mass regime where the binary fraction decreases to very low values.

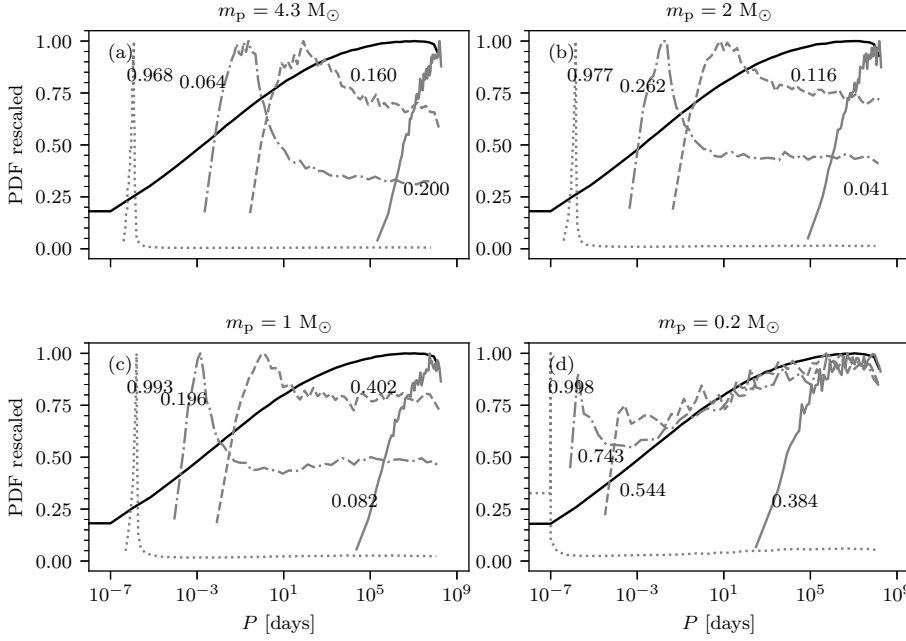


Fig. 10. Initial (i.e. after pre-main sequence eigenevolution) period distributions for binaries with different mass ratios and primary masses. Each panel corresponds to a different primary mass (as indicated in the title). The black solid line shows the overall period distribution for that primary mass, while grey lines (with different styles) indicate distributions in individual q bins. The label next to each gray line denotes the left edge of the corresponding q bin; all bins have a fixed width of 0.001. All distributions are normalized to their peak values to facilitate visual comparison across different scales. We note that the extremely short orbital period appearing here ($\approx 10^{-7}$ days, i.e. on the order of milliseconds) is unphysical. Such systems are rarely generated and would merge almost immediately after the start of the simulation.

To characterize the trend for $m_p < 5 M_\odot$, we adopt the same sigmoid-like functional form as in Eq. (9), but without a hard cut-off,

$$f_{\text{bin}}(m_p) = \frac{\mathcal{A}_b}{1 + \exp[\mathcal{S}_b(\log_{10} m_p - \mathcal{M}_{\text{cut}})]} - \frac{\mathcal{A}_b}{2}. \quad (16)$$

Here, $\mathcal{A}_b/2$ sets the saturation level of f_{bin} for intermediate-mass primaries, while $10^{\mathcal{M}_{\text{cut}}}$ defines the characteristic mass scale below which the binary fraction declines towards zero.

We interpret $f_{\text{bin}}(m_p)$ probabilistically, and model the binary indicator variable with a Bernoulli distribution. The corresponding log-likelihood function for the parameters \mathcal{A}_b , \mathcal{S}_b , and \mathcal{M}_{cut} is

$$\ln \mathcal{L}(\mathcal{A}_b, \mathcal{S}_b, \mathcal{M}_{\text{cut}}) = \sum_i \left[y_i \cdot \ln f_{\text{bin}}(m_p^{(i)}) + (1 - y_i) \cdot \ln (1 - f_{\text{bin}}(m_p^{(i)})) \right], \quad (17)$$

where $y_i = 1$ if the i -th system is a binary with $q \leq 1$, and $y_i = 0$ otherwise. For each sample we perform maximum-likelihood estimation restricted to $m_p < 5 M_\odot$. The solid curve in Fig. 11 shows the median fit across all samples.

Comparison with the observational results for G-, K-, M-, and A-type stars in the Galactic field (Kroupa et al. 2003; De Rosa et al. 2014), and shown in Fig. 11, reveals that our models predict a higher fraction of low-mass binaries. For A-type stars, in contrast, the model reproduces the observations well. The binary fraction at a given mass can be reduced in two ways: (i) by the disruption of binaries whose primaries fall in that mass range, and (ii) indirectly, through the disruption of higher-mass binaries. $f_{\text{bin}}(m_p)$ is defined with respect to the primary mass; it is calculated by counting all single stars and binary systems with primary mass m_p within a given mass range and identifying the fraction of binary systems. When a binary system is disrupted, its contribution to the binary population at the primary mass m_p is removed, while the former secondary becomes a single star at its own mass bin. As a result, the number of binary systems

at the secondary mass m_s is not reduced by this disruption however the number of single stars at m_s is increased by 1 and thus the binary fraction also reduced¹. If, at later evolutionary stages, disruption becomes inefficient for binaries with $m_p \geq 1 M_\odot$, then the further decline of the binary fraction at lower masses must be driven primarily by the preferential dynamical disruption of low-mass systems; namely, channel (i). It is possible that post-gas-expulsion evolution continues to play an important role for binaries with $m_p \lesssim 1 M_\odot$, which are characterized by low binding energies and are particularly vulnerable to perturbations. This effect is expected to differ between the Galactic field and cluster environments, potentially leading to systematic differences in the binary population properties of the two. A more complete treatment will require the modelling of the subsequent dynamical evolution of low-mass binaries beyond the embedded phase.

4.5. Future works

Future works will investigate several key factors that may influence the dynamical evolution of binary populations in embedded clusters. First, variations in the total cluster mass will be examined in combination with different star formation efficiencies (SFE). These two parameters jointly determine the velocity dispersion of the system, which, in turn, sets the critical binding energy threshold distinguishing hard and soft binaries. Marks et al. (2011) have demonstrated that the dynamical evolution of binary populations depends primarily on the initial density of the cluster. By varying the cluster's initial mass and half-mass radius, they showed that the fitting parameters describing binary processing correlate strongly with cluster density. It will be explored whether similar trends emerge in the present models.

Secondly, hardening and softening effects are expected to be more prominent in low-mass clusters, where the velocity dispersion is relatively low. In such environments, binary encounters

¹ For example, we can consider a binary system composed of a G-type primary and a M-type secondary. If such a system is disrupted, the binary fraction of G-type stars is reduced and simultaneously the binary fraction of M-type stars is also decreased because one more single stars is added into the M-type stars.

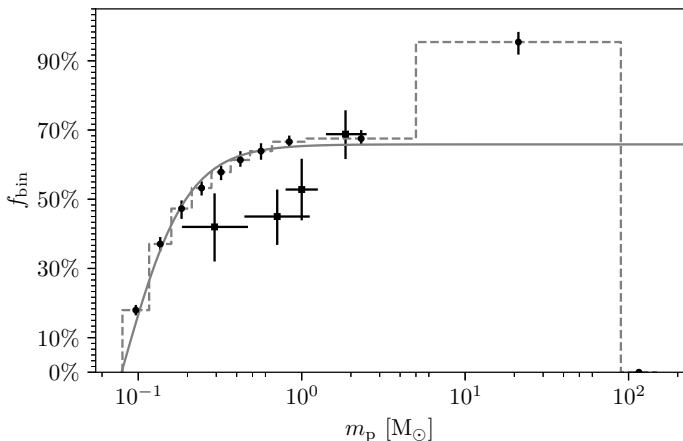


Fig. 11. Binary fraction as a function of primary mass, m_p . Points show the median across all samples, with error bars marking the 16th–84th percentile range. The dashed gray line indicates the binning scheme and median values. The last bin spans 89.73–149.17 M_\odot ; its upper edge corresponds to the most massive stars found in our models, which form through mergers and are therefore single. The solid gray curve shows the best-fit model obtained via maximum-likelihood estimation. Observational results are over-plotted as open squares: Kroupa et al. (2003) for M-, K- and G-type stars (left three points) and De Rosa et al. (2014) for A-type stars (right-most point). We note that all observational results come from the Galactic field.

are less energetic and thus more likely to result in gradual energy exchange rather than immediate disruption. Consequently, binaries tend to evolve through cumulative interactions (either hardening or softening) rather than being promptly destroyed, as is more common in high-velocity environments where encounters are more violent.

In addition to the above factors, a third aspect worth attention concerns the role of high-mass binaries. The present work does not explore in detail the dynamical evolution of high-mass binary systems. Due to the shape of the IMF, such systems are intrinsically rare in the models. However, in more massive environments such as young globular clusters, the number of high-mass binaries can be significantly larger, making their dynamical processing and resulting populations more relevant, particularly for the emergence of multiple population (see e.g. Wang et al. 2020b). Recent observational studies have highlighted the importance of high-mass binaries; for instance, a correlation between their properties and metallicity has been reported (Sana et al. 2025), and pulsar observations have revealed unusually wide binary systems possibly linked to the evolution of massive binaries (Lian et al. 2025). These findings suggest that the formation and survival of high-mass binaries remain an open question and warrant further investigation.

5. Summary and conclusions

In this study, we investigated the early dynamical evolution of binary populations in embedded massive star clusters using direct N -body simulations with PeTar. By systematically comparing binary distribution functions (BDFs) at $t = 0$ and $t = 0.6$ Myr, we quantified the impact of stellar dynamics through empirical dynamical operators. Constructing such operators has proven to be highly essential, as our results show that only the binding energy and orbital period provide robust dynamical diagnostics, both of which can be described by sigmoid-like operators with interpretable parameters. This highlights the

necessity of direct N -body simulations for modelling embedded clusters, as no simple analytic substitute exists.

Compared to the model of Marks et al. (2011), our formulation introduces an additional offset term, enabling the operator to account for a residual population of wide binaries. This refinement is crucial to capture the persistence of wide systems, which may escape the cluster before disruption and plausibly contribute to the very wide binaries observed in the Galactic field. At the same time, in line with Marks et al. (2011), it has not been explicitly tested whether the binding energy alone already provides a complete description of binary evolution.

Other binary properties exhibit weaker dynamical signatures. The shape of the mass-ratio distribution shows no strong evolutionary trend, being shaped primarily by the initial conditions and pre-main sequence eigenevolution rather than cluster dynamics. At 0.6 Myr, the binary fraction as a function of primary mass, $f_{\text{bin}}(m_p)$, shows a strong mass dependence that can be modeled by a sigmoid-like function: high-mass stars ($m_p \gtrsim 5, M_\odot$) almost universally retain companions, while the fraction declines towards lower masses. This trend is consistent with observations of A-type stars, although our models predict a slightly larger fraction of M-, L-, and G-type star-binaries than in the Galactic field.

Taken together, these results suggest a coherent picture in which disruption dominates the early dynamical evolution, primarily affecting soft, low-binding-energy binaries, while the shape of the mass ratio distribution is largely set by pre-main sequence eigenevolution prior to dynamical interactions. Thus, the key dynamical information is effectively encoded in binding energy and orbital period, with other parameters playing a secondary role.

For modelling the binary population of an individual cluster, direct N -body simulations remain indispensable. On galactic scales, however, where the field population originates from the dissolution of many embedded clusters with diverse properties, population synthesis approaches in the spirit of Marks et al. (2011) remain valuable; that is, provided realistic cluster-scale dynamics, such as those quantified here, are incorporated. Future extensions of this model will further clarify the connection between early cluster dynamics and the present-day binary population.

Acknowledgements. P.K. acknowledges support through the DAAD Eastern European Exchange Programme between the University of Bonn and Charles University in Prague as well as through grant 26-21774S from the Czech Grant Agency.

References

- Aarseth, S. J. 1999a, *PASP*, 111, 1333
- Aarseth, S. J. 1999b, *Celest. Mech. Dyn. Astron.*, 73, 127
- Aarseth, S. J. 2003, *Gravitational N-body Simulations* (Cambridge: Cambridge University Press)
- Banerjee, S., & Kroupa, P. 2013, *ApJ*, 764, 29
- Banerjee, S., Belczynski, K., Fryer, C. L., et al. 2020, *A&A*, 639, A41
- Barnes, J., & Hut, P. 1986, *Nature*, 324, 446
- Beccari, G., Petr-Gotzens, M. G., Boffin, H. M. J., et al. 2017, *A&A*, 604, A22
- Belczynski, K., Kalogera, V., Rasio, F. A., et al. 2008, *ApJS*, 174, 223
- Belczynski, K., Heger, A., Gladysz, W., et al. 2016, *A&A*, 594, A97
- Belloni, D., Askar, A., Giersz, M., Kroupa, P., & Rocha-Pinto, H. J. 2017, *MNRAS*, 471, 2812
- Boily, C., & Kroupa, P. 2002, *ASP Conf. Ser.*, 285, 141
- Bovy, J. 2015, *ApJS*, 216, 29
- Bressert, E., Bastian, N., Gutermuth, R., et al. 2010, *MNRAS*, 409, L54
- Casertano, S., & Hut, P. 1985, *ApJ*, 298, 80
- Dabringhausen, J., Marks, M., & Kroupa, P. 2022, *MNRAS*, 510, 413

- De Furio, M., Reiter, M., Meyer, M. R., et al. 2019, *ApJ*, **886**, 95
- De Rosa, R. J., Patience, J., Wilson, P. A., et al. 2014, *MNRAS*, **437**, 1216
- Dorval, J., Boily, C. M., Moraux, E., & Roos, O. 2017, *MNRAS*, **465**, 2198
- Dvořáková, N., Korčáková, D., Dinnbier, F., & Kroupa, P. 2024, *A&A*, **689**, A234
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, **749**, 91
- Giersz, M., & Heggie, D. C. 1996, *MNRAS*, **279**, 1037
- Gjergo, E., Zhang, Z.-Y., Kroupa, P., et al. 2026, *Res. Astron. Astrophys.*, **26**, 025003
- Heggie, D. C. 1975, *MNRAS*, **173**, 729
- Hills, J. G. 1975, *AJ*, **80**, 809
- 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
- Jadhav, V. V., Roy, K., Joshi, N., & Subramaniam, A. 2021, *AJ*, **162**, 264
- Jerabkova, T., Beccari, G., Boffin, H. M. J., et al. 2019, *A&A*, **627**, A57
- Korčáková, D., Dvořáková, N., Bermejo Lozano, I., et al. 2025, *Galaxies*, **13**, 46
- Kounkel, M., Covey, K., Moe, M., et al. 2019, *AJ*, **157**, 196
- Kroupa, P. 1995a, *MNRAS*, **277**, 1491
- Kroupa, P. 1995b, *MNRAS*, **277**, 1522
- Kroupa, P. 1995c, *MNRAS*, **277**, 1507
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Kroupa, P. 2002, Habilitation Thesis, University of Kiel
- Kroupa, P. 2011, *IAU Symp.*, **270**, 141
- Kroupa, P. 2025, *Contrib. Astron. Observ. Skal. Pleso*, **55**, 217
- Kroupa, P., & Burkert, A. 2001, *ApJ*, **555**, 945
- Kroupa, P., & Petr-Gotzens, M. G. 2011, *A&A*, **529**, A92
- Kroupa, P., Tout, C. A., & Gilmore, G. 1991, *MNRAS*, **251**, 293
- Kroupa, P., Petr, M. G., & McCaughrean, M. J. 1999, *New A*, **4**, 495
- Kroupa, P., Aarseth, S., & Hurley, J. 2001, *MNRAS*, **321**, 699
- Kroupa, P., Bouvier, J., Duchêne, G., & Moraux, E. 2003, *MNRAS*, **346**, 354
- Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2013, in *Planets, Stars and Stellar Systems, Galactic Structure and Stellar Populations*, eds. T. D. Oswalt, & G. Gilmore (Berlin: Springer), 5, 115
- Kroupa, P., Jeřábková, T., Dinnbier, F., Beccari, G., & Yan, Z. 2018, *A&A*, **612**, A74
- Kroupa, P., Gjergo, E., Jerabkova, T., & Yan, Z. 2026, in *Encyclopedia of Astrophysics* (Berlin: Springer), 2, 173
- Küpper, A. H. W., Maschberger, T., Kroupa, P., & Baumgardt, H. 2011, *MNRAS*, **417**, 2300
- Lada, C. J. 2010, *Phil. Trans. R. Soc. London Ser. A*, **368**, 713
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
- Lian, Y., Pan, Z., Zhang, H., et al. 2025, *ApJS*, **279**, 51
- Marks, M., & Kroupa, P. 2011, *MNRAS*, **417**, 1702
- Marks, M., Kroupa, P., & Oh, S. 2011, *MNRAS*, **417**, 1684
- Marks, M., Kroupa, P., Dabringhausen, J., & Pawlowski, M. S. 2012, *MNRAS*, **422**, 2246
- Meylan, G., & Heggie, D. C. 1997, *A&A Rev.*, **8**, 1
- Moeckel, N., & Clarke, C. J. 2011, *MNRAS*, **410**, 2799
- Offner, S. S. R., Moe, M., Kratter, K. M., et al. 2023, *ASP Conf. Ser.*, **534**, 275
- Oh, S., Kroupa, P., & Pflamm-Altenburg, J. 2015, *ApJ*, **805**, 92
- Pang, X., Liao, S., Li, J., et al. 2024, *ApJ*, **966**, 169
- Parker, R. J., Goodwin, S. P., Kroupa, P., & Kouwenhoven, M. B. N. 2009, *MNRAS*, **397**, 1577
- Pflamm-Altenburg, J., & Kroupa, P. 2007, *MNRAS*, **375**, 855
- Plummer, H. C. 1911, *MNRAS*, **71**, 460
- Rozner, M., & Perets, H. B. 2024, *ApJ*, **968**, 80
- Rozner, M., Generozov, A., & Perets, H. B. 2023, *MNRAS*, **521**, 866
- Safaei, G., Haghi, H., Zonoozi, A. H., & Kroupa, P. 2025, *MNRAS*, **541**, 1753
- Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Science*, **337**, 444
- Sana, H., Le Bouquin, J. B., Lacour, S., et al. 2014, *ApJS*, **215**, 15
- Sana, H., Shenar, T., Bodensteiner, J., et al. 2025, *Nat. Astron.*, **9**, 1337
- Spurzem, R., & Kamlah, A. 2023, *Liv. Rev. Comput. Astrophys.*, **9**, 3
- Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P., & Kawanabe, M. 2007, in *Advances in Neural Information Processing Systems*, eds. J. Platt, D. Koller, Y. Singer, & S. Roweis, (New York: Curran Associates, Inc.)
- Thies, I., Pflamm-Altenburg, J., Kroupa, P., & Marks, M. 2015, *ApJ*, **800**, 72
- Tokovinin, A., & Briceño, C. 2020, *AJ*, **159**, 15
- Wang, L., & Jerabkova, T. 2021, *A&A*, **655**, A71
- Wang, L., Spurzem, R., Aarseth, S., et al. 2015, *MNRAS*, **450**, 4070
- Wang, L., Kroupa, P., & Jerabkova, T. 2019, *MNRAS*, **484**, 1843
- Wang, L., Iwasawa, M., Nitadori, K., & Makino, J. 2020a, *MNRAS*, **497**, 536
- Wang, L., Kroupa, P., Takahashi, K., & Jerabkova, T. 2020b, *MNRAS*, **491**, 440
- Wang, L., Nitadori, K., & Makino, J. 2020c, *MNRAS*, **493**, 3398
- Weidner, C., & Kroupa, P. 2006, *MNRAS*, **365**, 1333
- Yan, Z., Jerabkova, T., & Kroupa, P. 2023, *A&A*, **670**, A151

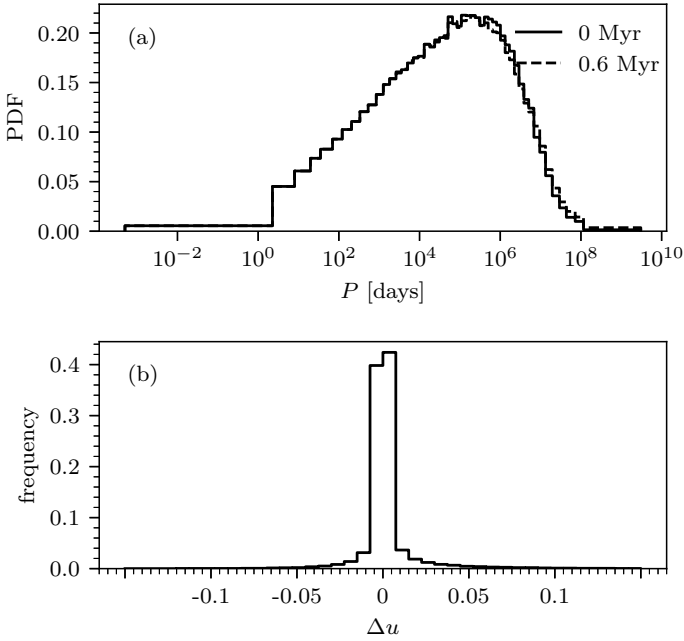


Fig. A.1. (a) Empirical PDFs of the period P for surviving binaries at 0 Myr (solid) and 0.6 Myr (dashed). The same binning is applied to all models, and the plotted histograms show, for each bin, the median value across all clusters. The two distributions are nearly indistinguishable. (b) Histogram of quantile shifts, where the bar heights represent the median values across all clusters.

Appendix A: PDF and quantile shift distribution of Period

Figure A.1 shows the initial and evolved PDFs, together with the quantile shifts for the periods of surviving binaries. The result is consistent with that obtained for the absolute value of binding energy: the dynamical operator can be interpreted as a disruption probability. Since orbital period and binding energy are strongly correlated, binding energy alone provides a sufficient and physically well-motivated basis for population synthesis models. The analysis of the period distribution therefore mainly serves as a cross-check of the binding-energy results.

Appendix B: Results on fitting parameters

Tables B.1, B.2 and B.3 list several percentiles of the fitted parameters across all models, highlighting their statistical properties. To further illustrate both the scatter among different realizations and the correlations between parameters, we visualize their distributions as corner plots in Figs. B.1, B.2, and B.3.

We emphasize that the corner plots shown here should not be interpreted as posterior distributions. Instead, each point represents the best-fit parameter set from a single simulation. All simulations share the same physical set-up, but differ in their random initial conditions. The use of corner plots therefore serves two purposes: (i) to demonstrate how randomness in the initial cluster configuration leads to variation in the fitted parameters, and (ii) to provide an intuitive visualization of possible correlations among them.

The corner plots reveal a strong anti-correlation between $\mathcal{A}_{E/P}$ and $\mathcal{B}_{E/P}$ (Pearson correlation coefficient < -0.9), suggesting that an abundance of hard binaries suppresses the survival or appearance of wide systems. A plausible explanation is that

Table B.1. Fitting results for Ω_{E_b} .

Parameter	2.5%	16%	50%	84%	97.5%
\mathcal{A}_E	0.686	0.714	0.736	0.753	0.780
\mathcal{S}_E	-3.016	-2.756	-2.535	-2.360	-2.172
\mathcal{E}_{cut}	-1.235	-1.204	-1.163	-1.119	-1.072
\mathcal{B}_E	0.015	0.037	0.052	0.074	0.102

Notes. Parameters are defined in Eq. 10. For all models combined, we report the 2.5% (-2σ), 16% (-1σ), 50% (median), 84% ($+1\sigma$), and 97.5% ($+2\sigma$) percentiles of each parameter, representing their statistical distributions. These values correspond to the points shown in the corner plots (Fig. B.1).

Table B.2. Fitting results for Ω_p .

Parameter	2.5%	16%	50%	84%	97.5%
\mathcal{A}_p	0.734	0.741	0.758	0.775	0.792
\mathcal{S}_p	2.213	2.456	2.642	2.862	3.059
\mathcal{P}_{cut}	6.545	6.585	6.635	6.697	6.742
\mathcal{B}_p	0.030	0.043	0.057	0.075	0.088

Notes. Same as Table B.1, but for the fitting parameters of Ω_p . The corresponding corner plots are shown in Fig. B.2.

Table B.3. Fitting results for $f_{\text{bin}}(m_p)$.

Parameter	2.5%	16%	50%	84%	97.5%
\mathcal{A}_b	1.262	1.295	1.317	1.342	1.367
\mathcal{S}_b	-5.706	-5.411	-5.059	-4.787	-4.590
\mathcal{M}_{cut}	-1.103	-1.103	-1.102	-1.101	-1.100

Notes. Parameters are defined in Eq. 16. Same as Table B.1, but for the fitting parameters of $f_{\text{bin}}(m_p)$. The corresponding corner plots are shown in Fig. B.3.

frequent encounters with hard binaries heat the cluster and disrupt wide pairs. Binary–binary encounters can directly destroy wide systems, while binary–single interactions harden the hard binaries and inject energy into the cluster. The resulting heating increases the velocity dispersion and reduces the chance for wide binaries to survive or form. In contrast, we find no significant correlations between other parameter pairs.

Appendix C: Comparison to Marks et al. (2011)

To directly compare with previous work, we apply the dynamical operator given by Eq. (9) to fit the evolution of the absolute value of binding-energy distribution under our updated IBP prescription and including the gas potential. For this comparison we adopt a simple χ^2 fit to the histograms, which avoids the additional complexity that a hard cut-off would introduce in a direct importance-estimation framework.

To enable a more intuitive comparison, we map our fitted model parameters back to the framework of Marks et al. (2011). In their formulation, the dynamical operator is explicitly parameterized as a function of the initial cluster density ρ_{ecl} . By applying their Eqs. (21)–(23), we transform each of our fitted parameters into an equivalent cluster density, which we denote as $\rho_{\text{ecl},M}$. This mapping provides a common basis for comparison: instead of contrasting operator parameters directly (which depend on our modified functional form), we can identify

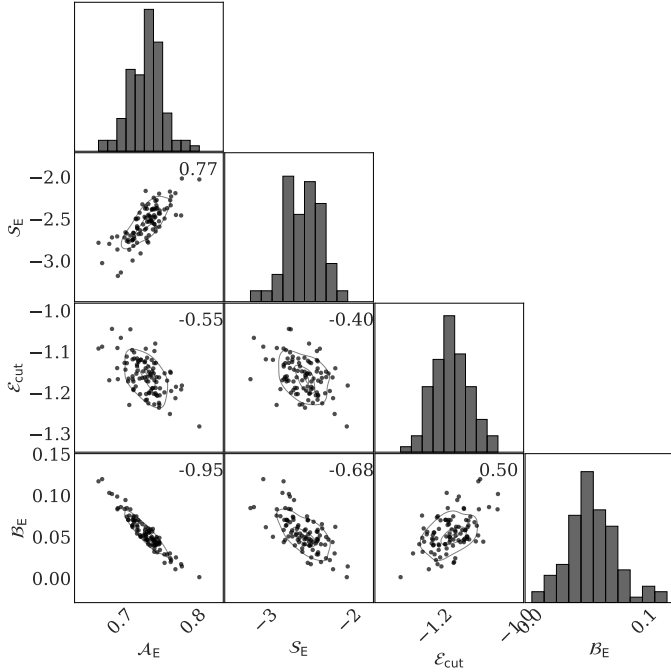


Fig. B.1. Corner plots of the fitted parameters \mathcal{A}_E , \mathcal{S}_E , \mathcal{E}_{cut} , and \mathcal{B}_E . Contours in the scatter plots indicate the 39.3% (1σ), 86.5% (2σ), and 98.9% (3σ) confidence levels. Numbers in the upper right of each panel give the Pearson correlation coefficient between the corresponding parameters. (a) Parameters \mathcal{A}_E , \mathcal{S}_E , \mathcal{E}_{cut} , and \mathcal{B}_E . Each point corresponds to one simulation, not posterior samples.

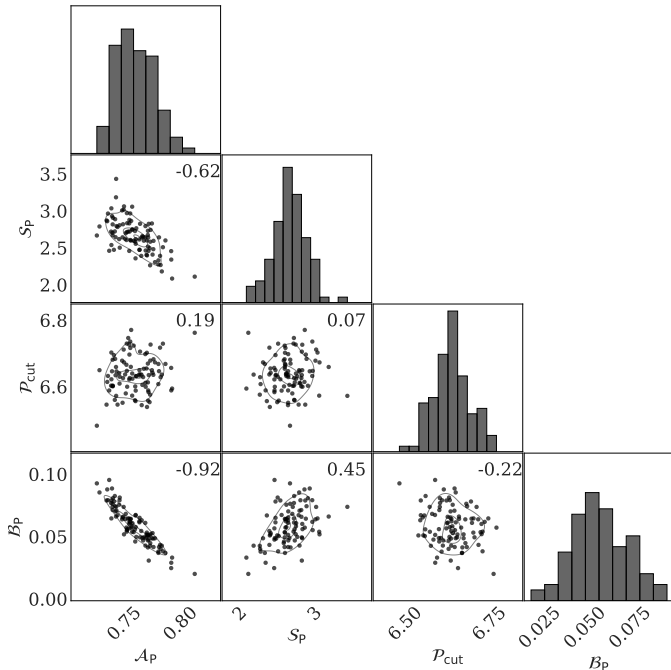


Fig. B.2. Corner plots of the fitted parameters \mathcal{A}_p , \mathcal{S}_p , \mathcal{P}_{cut} , and \mathcal{B}_p . Plot formatting is consistent with Fig. B.1.

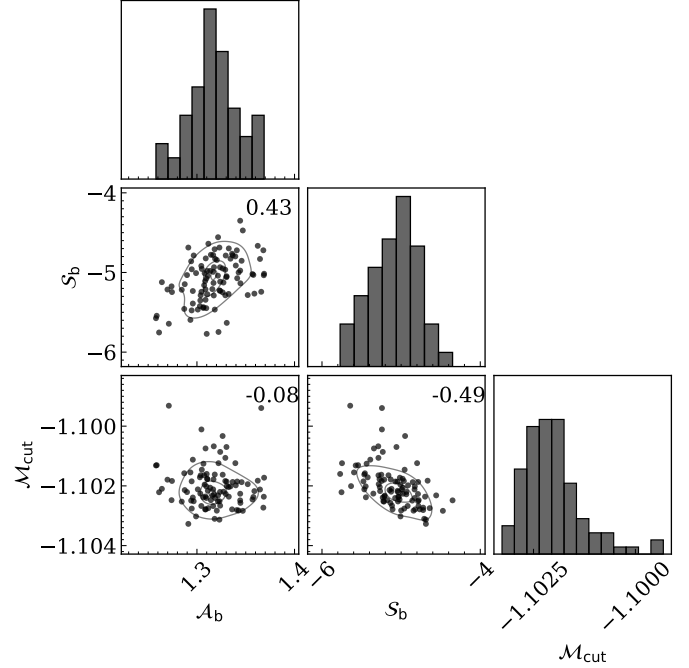


Fig. B.3. Corner plot of the fitting parameters \mathcal{A}_b , \mathcal{S}_b , and \mathcal{M}_{cut} . Plot formatting is consistent with Fig. B.1.

which effective $\rho_{\text{ecl},M}$ values in the Marks et al. (2011) framework are dynamically equivalent to our embedded cluster models, i.e. reproduce the behaviour of our simulations. In other words, this approach allows us to directly place our results onto the ρ_{ecl} scale used by Marks et al. (2011), thereby highlighting both agreements and systematic differences.

Figure C.1 presents the outcome of this comparison. Our simulations have a physical cluster density in stars of $\rho_{\text{ecl}} \approx 10^{4.5} \text{ M}_{\odot} \text{ pc}^{-3}$. For \mathcal{A}'_E , the transformed value corresponds to $\rho_{\text{ecl},M} \approx 10^5 \text{ M}_{\odot} \text{ pc}^{-3}$, i.e. stronger binary disruption than expected at the actual density. This is naturally explained by the inclusion of the gas potential in our modelling, which raises the velocity dispersion and hence the encounter rate. Similarly, for $\mathcal{E}'_{\text{cut}}$, the inferred $\rho_{\text{ecl},M}$ is also higher than the physical ρ_{ecl} , again reflecting the higher velocity dispersion. In contrast, \mathcal{S}'_E maps to a lower $\rho_{\text{ecl},M}$, corresponding to a steeper transition; this indicates that the peak position of the operator does not shift strongly with $\mathcal{E}'_{\text{cut}}$.

We further note that adopting the 1, 3, and 5 Myr operators from Marks et al. (2011) yields different values of $\rho_{\text{ecl},M}$. In our simulations, owing to the presence of the gas potential and the relatively short timescale (≤ 0.6 Myr), the clusters do not undergo significant structural evolution. In contrast, the high-density clusters modelled by Marks et al. (2011) expand substantially due to energy release from binary hardening (i.e. so-called binary burning; see their Fig. 4). As a result, their binary populations follow evolutionary trajectories that diverge from our result.

Overall, the $\rho_{\text{ecl},M}$ mapping provides a direct, parameter-independent way to compare our results with previous models. The systematic offsets we find underscore the importance of including the gas potential in embedded-phase simulations, and highlight that the link between operator parameters and cluster density is both time-dependent and sensitive to structural evolution.

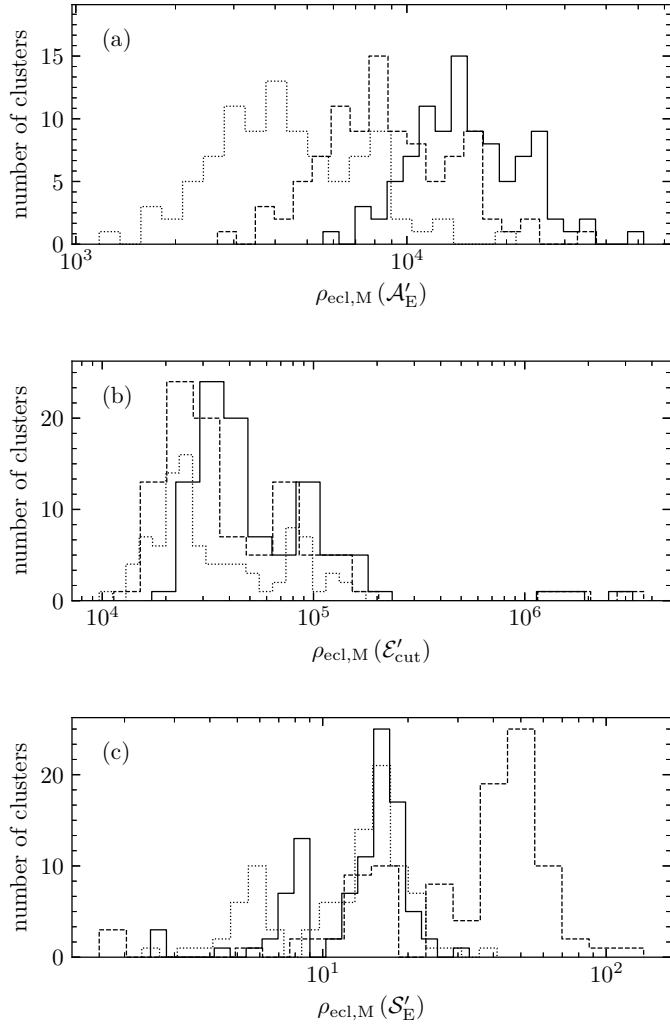


Fig. C.1. Equivalent cluster densities $\rho_{\text{ecl},M}$, shown as histograms. Solid, dashed, and dotted lines correspond to values obtained using the 1, 3, and 5 Myr models of Marks et al. (2011), respectively. Each panel shows the distribution of $\rho_{\text{ecl},M}$ derived from one fitted parameter.