

# MOCCA: dynamics and evolution of single and binary stars of multiple stellar populations in tidally filling and underfilling globular star clusters

Arkadiusz Hypki <sup>1</sup>★, Mirek Giersz <sup>1</sup>, Jongsuk Hong,<sup>2</sup> Agostino Leveque <sup>1</sup>, Abbas Askar <sup>3</sup>,  
Diogo Belloni <sup>4</sup> and Magdalena Otulakowska-Hypka <sup>5</sup>

<sup>1</sup>*Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, PL-00-716 Warsaw, Poland*

<sup>2</sup>*Korea Astronomy and Space Science Institute, Daejeon 34055, Republic of Korea*

<sup>3</sup>*Lund Observatory, Department of Astronomy, and Theoretical Physics, Lund University, Box 43, SE-221 00 Lund, Sweden*

<sup>4</sup>*National Institute for Space Research, Av. dos Astronautas, 1758, 12227-010 São José dos Campos, SP, Brazil*

<sup>5</sup>*Astronomical Observatory Institute, Faculty of Physics, A. Mickiewicz University, Słoneczna 36, PL-60-286 Poznań, Poland*

Accepted 2022 September 27. Received 2022 September 23; in original form 2022 May 4

## ABSTRACT

We present an upgraded version of the MOCCA code for the study of dynamical evolution of globular clusters (GCs) and its first application to the study of evolution of multiple stellar populations. We explore initial conditions spanning different structural parameters for the first (FG) and second generation of stars (SG) and we analyse their effect on the binary dynamics and survival. Here, we focus on the number ratio of FG and SG binaries, their spatial variation, and the way their abundances are affected by various cluster initial properties. We find that present-day SG stars are more abundant in clusters that were initially tidally filling. Conversely, FG stars stay more abundant in clusters that were initially tidally underfilling. We find that the ratio between binary fractions is not affected by the way we calculate these fractions [e.g. only main-sequence binaries (MS) or observational binaries, i.e. MS stars  $>0.4 M_{\odot}$  mass ratios  $>0.5$ ]. This implies that the MS stars themselves are a very good proxy for probing entire populations of FG and SG. We also discuss how it relates to the observations of Milky Way GCs. We show that MOCCA models are able to reproduce the observed range of SG fractions for Milky Way GCs for which we know these fractions. We show how the SG fractions depend on the initial conditions and provide some constraints for the initial conditions to have more numerous FG or SG stars at the Hubble time.

**Key words:** methods: numerical – software: data analysis – stars: chemically peculiar – galaxies: star clusters: general.

## 1 INTRODUCTION

Globular clusters (GCs) are old and dense stellar systems that play an important role in the study of stellar evolution, stellar dynamics, and stellar populations resulting from complex dynamical interactions. They are very efficient laboratories to study all types of stellar astronomical objects, even the most exotic ones, like cataclysmic variables, X-ray binaries, blue straggler stars (BSSs), black holes (BHs), or intermediate-mass black holes (IMBHs).

Initially, GCs were considered to host only a single stellar population that had to be formed from a single cloud upon the cluster formation. This implied that they should have the same chemical composition and age for all the stars in the cluster. It was indeed observed in the first colour–magnitude diagrams (CMDs) of GCs that they showed a common picture: the main sequence (MS), sub-giant, red giant, and horizontal branches appeared as single lines. The first direct observational proofs of multiple stellar populations (MSPs) were given by Lee et al. (1999) for  $\omega$  Cen. Extensive observational study of *HST* results on GCs abundances variations and MSP was presented by Piotto et al. (2015) for many GCs (see

also Milone et al. 2017a). In turn, Renzini et al. (2015) compared the possible formation scenarios of MSPs with regards to the new observational facts. Now, it appears that nearly all Galactic GCs show evidence of MSPs, e.g. by multiple MS populations in their CMDs, chemical variations, anticorrelations of elements, or extreme helium abundance (see Bastian et al. 2013; Bastian & Lardo 2018; Gratton et al. 2019, and references therein). Sometimes, there are more than two populations present, e.g. NGC 2808 with at least five populations (Milone et al. 2015c; Simioni et al. 2016) or M2 with seven (Milone et al. 2015a). MSPs are also found in GCs outside of our Galaxy, e.g. in NGC 1846, 1806, and 1783 in Large Magellanic Cloud (LMC; Mackey et al. 2008), in LMC intermediate age clusters (Milone et al. 2009, 2015b), or in M31 (Cezario et al. 2013).

There are a number of observational techniques that allow to identify the multiple populations. Thanks to *HST* photometry of 59 GCs it was revealed that first (FG), and second generation of stars (SG) appear as distinct sequences in many regions of the CMDs (Milone et al. 2017a). Together with photometric observations also the spectroscopic methods were used to study MSP revealing differences in the abundances of p-capture elements (Gratton 2004; Carretta et al. 2009; Gratton, Carretta & Bragaglia 2012; Gratton et al. 2019). All of this demonstrates that MSPs are rather common in Galactic GCs. The number of multiple populations for GCs varies quite substantially – from two populations for low-

\* E-mail: [ahypki@camk.edu.pl](mailto:ahypki@camk.edu.pl)

mass clusters up to several populations in massive clusters (e.g.  $\omega$  Cen). However, there are studies showing that there might be some GCs with no apparent signs of MSPs, which adds some complexity to this picture, e.g. Ruprecht 106 (Villanova et al. 2013).

Spatial and kinematic differences have been observed in several GCs (e.g. Richer et al. 2013; Bellini et al. 2015; Cordero et al. 2017; Milone et al. 2018; Dalessandro et al. 2021) while in some no differences were found (e.g. Cordonì et al. 2020). The spatial differences in velocities are consistent with formation models proposed by (e.g. Calura et al. 2019, and references therein). A number of studies have investigated the formation (Decressin et al. 2007; D’Ercole et al. 2008; Bekki 2010, 2011; Gieles et al. 2018; Calura et al. 2019; Wang et al. 2020; McKenzie & Bekki 2021) and different aspects of the dynamical evolution of MSPs (Vesperini et al. 2013, 2018, 2021; Hénault-Brunet et al. 2015; Miholics, Webb & Sills 2015; Tiongco, Vesperini & Varri 2019; Sollima 2021). The effects of the tidal field on the formation of MSP were investigated e.g. by Vesperini et al. (2021) and Sollima (2021).

Some of the properties of MSPs among GCs have a wide spread and can differ strongly between clusters. For example, the fraction of SG stars have values as low as 35 per cent for some GCs (e.g. M 71) and as high as 90 per cent for others ( $\omega$  Cen) that poses a great challenge for interpretation. Also the radial distributions of FG and SG presents some interesting properties. There are GCs that have SG more radially concentrated than the FG, e.g.  $\omega$  Cen (Sollima et al. 2007), 47 Tuc (Cordero et al. 2014), and there are GCs for which radial profiles of the two populations suggest no differences, e.g. NGC 6752 (Nardiello et al. 2015), M 5 (Lee 2017), or NGC 6362 (Dalessandro et al. 2014).

There is no significant correlation between GCs masses and the galaxy orbital parameters (Milone et al. 2017b). However, there is some sign that larger fraction of FG stars is characteristic for the GCs with larger perigalactic radii (Zennaro et al. 2019; Milone et al. 2020a). The possible role of mass-loss due to stellar escape from two-body relaxation in producing this trend is discussed also in Vesperini et al. (2021). Strong correlation between fraction of SG with the cluster total mass was reported by Carretta et al. (2010) and Milone et al. (2017a). This suggests that the host galaxies have some kind of influence on the GCs evolution after all. Baumgardt & Hilker (2018) report also that simple-population GCs seem to have initial masses smaller than  $\sim 1.5 \cdot 10^5 M_{\odot}$ , whereas MSPs are present in more massive ones. However, this conclusion is also challenged by e.g. NGC 419 (Li et al. 2020) which is a massive cluster but with no apparent signs of MSPs.

We can trace formation and dynamical history of stars with the use of numerical simulations, and their physical nature with various astronomical observations. Star clusters are often used to test stellar evolution and population synthesis models as well as MSPs within the GCs. The aspect of MSPs and their formation is especially important in the context of our understanding of the formation and evolution of GCs and their host galaxies.

Binaries play an important role in the dynamical evolution of GCs, and they are also one of the important sources of the production of exotic astronomical objects. Since it is very difficult to determine their populations and binarity together in the photometric observations, the nature and dynamics of binary stars in the MSPs have been relying on the theoretical approaches (Vesperini et al. 2011; Hong et al. 2015, 2016, 2019; Khalaj & Baumgardt 2015). Especially, Sollima et al. (2022) investigated with Monte Carlo codes the fractions of SG binaries surviving to the Hubble time. They claim

that the fraction of SG binaries depends only on the ratio between the total cluster mass and the initial size of the SG. A remarkable attempt to distinguish the multiple population binaries has been made by Milone et al. (2020b). They obtained the distribution of populations among the binaries using *chromosome maps* and proved the existence of mixed binaries composed of one FG and one SG stars. Another interesting observational study for the binaries in the multiple populations was done by Dalessandro et al. (2018). They found an inflation of FG/SG ratio of velocity dispersion at the large radii in the fully mixed cluster NGC 6362 whose velocity dispersions are supposed to be identical for the different population. It turned out, that the inflation is due to the internal motion of binaries with the fact that the FG binaries are more abundant at the large radii compared to the SG binaries which are preferentially disrupted at the central regions (Hong et al. 2019; Mastrobuono-Battisti & Perets 2021).

There are proposed multiple explanations for how the SG could be formed from the leftover gas from the FG formation and enriched: by asymptotic giant branch (AGB) stars (Ventura et al. 2001), fast rotating massive stars and massive binaries (Decressin et al. 2007; Krause et al. 2013), supermassive stars (Denissenkov & Hartwick 2014; Gieles et al. 2018), accretion discs (Bastian et al. 2013), or non-conservative mass-transfer in massive binaries (de Mink et al. 2009). There is also a scenario which assumes that star clusters could form as a merger of protoclusters (Elmegreen 2017; Howard et al. 2019). However, there is no conclusive evidence for which of those scenarios are responsible for SG formation. In this paper, we are working within the scenario described in D’Ercole et al. (2008) and Calura et al. (2019). In those models SG forms from AGB ejecta and pristine gas reaccreted from the external medium. SG form in the inner regions as a result of a cooling flow and gas flowing to the inner regions where SG stars are formed. Then, with time the evolution of a cluster erases the spatial differences (e.g. Tiongco et al. 2019, and references within).

This paper is organized as follows. In Section 2, there is an introduction to new features which were implemented into MOCCA code in order to prepare new simulations of GCs with MSPs (Section 2.1). Section 2.2 provides summary of the MOCCA models which were computed so far and which are planned to be finished in a near future. In this section, we present also new, public version of the BEANS software, which is a general, web-based solution for distributed data analysis of huge data sets (Section 2.3). In Section 3, there is a description of the results obtained from the simulations, together with some general validation tests showing that the current version of MOCCA code is consistent with the previous version. In Section 4, we discuss how the results from the MOCCA simulations explain (or are consistent) with the already available observations of Milky Way GCs. In particular, we analyse how the findings of this work could help to constrain the initial conditions of the star clusters. In Section 5, we summarize our findings about MSPs, and we present our future plans and projects. In Appendix A, we describe how we treat in MOCCA code the mixed populations and we present briefly the output files. And finally, Appendix B contains an example script which shows how one can analyse in parallel all MOCCA simulations at once.

## 2 NUMERICAL SIMULATIONS

This section presents the new features implemented into MOCCA code, the initial conditions for the simulations and tools used to perform distributed data analysis.

## 2.1 MOCCA – new features

The numerical simulations were performed with the MOCCA<sup>1</sup> code (Giersz 1998; Hypki & Giersz 2013; Giersz et al. 2014). MOCCA is currently one of the most advanced codes to perform full stellar and dynamical evolution of real size star clusters. MOCCA can be used to study various aspects of star cluster evolution, like formation and evolution of compact binaries, e.g. BHs binaries (Hong et al. 2020), exotic objects like IMBHs (Giersz et al. 2015), BSSs (Hypki & Giersz 2017), or can be used to model even the most massive star clusters, e.g. 47 Tuc, (Giersz & Heggie 2011). MOCCA follows the stellar and dynamical evolution of all stars in the system closely to  $N$ -body codes and provides almost as much information about single and binary stars (Wang et al. 2016). In MOCCA stellar evolution for single and binary stars was performed with the SSE/BSE code (Hurley, Pols & Tout 2000; Hurley, Tout & Pols 2002) which was strongly updated by Belloni et al. (2017a, b). Among the most important upgrades are the ones which concern common-envelope phase. Summary of the stellar evolution features one can find in Kamlah et al. (2022). Strong dynamical interactions in MOCCA are performed with FEWBODY code (Fregeau et al. 2004; Fregeau & Rasio 2007).

The most noticeable new feature which was added to MOCCA code, since the first MOCCA-SURVEY-1 (Askar et al. 2017), is the support for MSPs. MOCCA is now able to follow the full dynamical evolution of MSPs, and to some extent, also the stellar evolution for different populations. MOCCA can support up to 10 different stellar populations in one simulation.

A few technical aspects of MSPs implementation require additional description. Every star in MOCCA simulation has its own population id (assigned as an initial parameter at time  $T = 0$ ). After that, the star can have various interactions with other stars and binaries which might result in a change of its mass (e.g. mass transfers, dynamical mergers). If the star happens to have such an event with a star coming from the same population, the result of the interaction is performed as usual. However, if the interacting star comes from a different population, the star is marked as so-called *mixed population* star. In MOCCA, we currently use stellar evolution codes which do not support mixing between stars with different chemical abundances, thus we are only marking these stars. For events which concern stars from two different populations we take as a current metallicity the one which comes from the more massive star. This simplification allows us to evolve mixed population stars as normal stars, but one has to keep in mind that the masses, luminosities, colours for mixed populations can be slightly inaccurate. More details on how the mixed population id is build for stars with complex histories is described in Appendix A.

Initial conditions for all MOCCA-SURVEY-2 simulations were generated using the MCLUSTER code<sup>2</sup> (Leveque, Giersz & Paolillo 2021) which is an upgraded version of the original one<sup>3</sup> (Küpper et al. 2011). Among the most important additions to the original MCLUSTER code is the capability to generate initial conditions for star clusters for up to 10 different stellar populations. Additionally, the initial conditions for the multiple populations are set in such a way that the whole star cluster is in virial equilibrium by solving Jeans equations (Kamlah et al. 2022). Every population can have its own set of initial values. One can choose initial number of single and binary stars for every population, initial model (homogeneous sphere, Plummer, or King 1966). Models can be initially segregated or not

**Table 1.** The table shows selected initial conditions of MOCCA-SURVEY-2 – a set of MOCCA simulations performed with the upgraded version of the code. All of the initial conditions were chosen to test the mixing between different stellar populations for star clusters of a real size. FG means the first generation, and SG the second one. Some of the parameters are different for these two stellar populations (e.g.  $N$ ) and some of them are the same in both cases (e.g.  $R_h$ ). The meaning of the initial parameters is following:  $N$  – initial number of objects;  $W_0$  – King model parameters;  $M_{\max}$  – upper mass limit for a single star;  $m_{\text{func}}$  – stellar mass function;  $fb$  – binary fraction;  $R_t$  – tidal radius [pc];  $R_h$  – half-mass radius for the whole star cluster;  $\text{conc}_{\text{pop}}$  – concentration parameter between two populations (e.g. value 0.1 means that SG is 10 times more concentrated than FG);  $\text{evol}$  – stellar evolution on/off or point mass (1, 0, -1, respectively). For details, see description in Section 2.2.

Parameter	FG	SG
$N$	400 k	200 k, 400 k, 600 k
$W_0$	6	6, 8
$M_{\max} [M_{\odot}]$	150	50, 150
$m_{\text{func}}$		0, 1
$fb$		0.0, 0.1, 0.95
$R_t$ [pc]		60, 120
$R_h$ [pc]		0.6, 1.2, 2.0, 4.0, 6.0
$\text{conc}_{\text{pop}}$		0.1, 0.5, 1.0, 1.5
$\text{evol}$		-1, 0, 1

(Baumgardt, De Marchi & Kroupa 2008), the stars can be initially fractaled (Goodwin & Whitworth 2004), virial ratio can be set to every population separately too. Stellar mass functions can be set to equal masses, Kroupa (2001) mass function, multi power law (based on MUFU by L. Subr), optimal sampling following the prescriptions of either Kroupa (2011), or Maschberger & Clarke (2012). Pairing for binary components and semimajor axis distributions can be also set to a few different procedures [e.g. random pairing; ordered pairing; uniform mass ratio ( $0.1 < q < 1.0$ ) distribution for binaries with component masses larger than  $5 M_{\odot}$  according to Kiminki & Kobulnicky 2012; Sana et al. 2012; Kobulnicky et al. 2014 – selected option for all MOCCA simulations presented in Section 2.2]. Eigenevolution can be switched off, can follow Kroupa (1995), or can follow a new eigenevolution and mass feeding algorithm (Kroupa et al. 2013; Belloni et al. 2017b). Concentration parameter can be set up for  $n$ -th population with respect to the first population. At last, the metallicity can be also set up separately for every stellar population too. Only the tidal radius can be set to the whole star cluster.

Some details about the output files which are produced by MOCCA simulations, and which data are being stored, are presented in Appendix A.

## 2.2 Initial conditions

We present here MOCCA simulations performed with the newest version of the MOCCA code (see Section 2.1). The new simulations form MOCCA-SURVEY-2 – a new major upgrade to the existing MOCCA-SURVEY-1, which has been successfully used in a number of projects already. MOCCA-SURVEY-2 consists currently of 257 models.

The initial conditions for the MOCCA-SURVEY-2 simulations are shown in the Table 1. Parameter  $N$  denotes initial number of bound

<sup>1</sup><https://moccacode.net>

<sup>2</sup><https://github.com/agostinolev/mcluster>

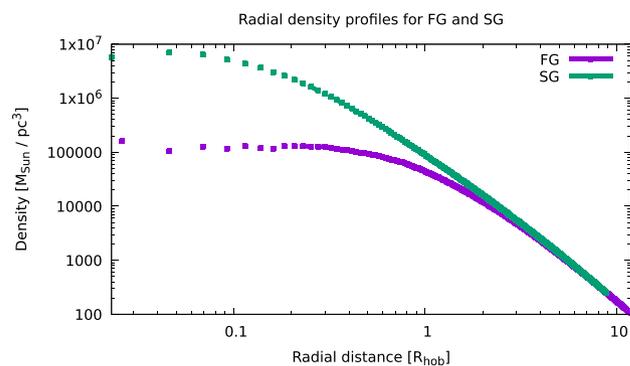
<sup>3</sup><https://github.com/ahwkuepper/mcluster>

objects for the first population (FG) and the second population (SG), respectively. A bound object is either a single star or a binary. The parameter  $W_0$  is the King model parameter, and can be specified for both populations separately. The parameter  $M_{\max}$  is the upper mass limit for single stars for both populations separately. For the first population the upper limit is the same –  $150 M_{\odot}$ . For the second population, it is  $150 M_{\odot}$  or  $50 M_{\odot}$  (the latter limit mimics a scenario for which the second population contains less massive stars). The next parameter  $m_{\text{func}}$  is stellar mass functions. For MOCCA-SURVEY-2, we use two values: the value 0 is for equal mass simulations, and value 1 for Kroupa (2001) mass function. The next parameter  $fb$  is binary fraction for both populations. The number of binaries is defined as  $fb_1 * N_1$  for FG and similarly for SG. The parameter  $R_t$  is the tidal radius for the entire star cluster in parsecs, and  $R_h$  is half-mass radius also for the whole star cluster. The parameter  $\text{conc}_{\text{pop}}$  is the concentration parameter but set to all MSPs with respect to the FG. It is defined as  $R_{h1}/R_{h1}$  – the ratio between the half-mass radius of the  $i$ -th population to the first population. The parameter  $\text{evol}$  defines whether the stellar evolution is on, off, or the stars are point masses. For stellar evolution switched off we assign every star its ZAMS radius to keep stellar collisions and mergers possible. For the case of point-mass stars the radii are not assigned. This parameter is used only for some simulations to compare them with simulations already presented in the literature and for testing.

The rest of the initial parameters are common for all simulations. The semimajor axes distributions follow Kroupa (1995), eigenevolution (Kroupa 1995), reviewed by Belloni et al. (2017b), is switched on for MOCCA-SURVEY-2 by default (with only a few exceptions for some point mass models). The minimum semimajor axis is set up automatically in such a way that it will cause no immediate merger. The maximum semimajor axis is set to 100 au. Tidal field is due to a point-mass galaxy with mass inside circular cluster orbit and constant rotation velocity equal to  $220 \text{ km s}^{-1}$ . Metallicity for both populations is set to 0.001 (solar metallicity is  $Z = 0.02$ ). Mass of neutron stars (NSs) and BHs is determined according to rapid supernova model which includes mass fallback Fryer et al. (2012). Pair-instability supernovae (PSNs) to entirely disrupt massive stars, and pair-instability pulsation supernovae (PPSNs) are set according to Belczynski et al. (2016). Electron capture supernovae are also switched on by default for all models. Kick mechanism for NS/BH formation is set to standard momentum conservation (NS and BH have the same kick distributions) and the velocities are set to Maxwellian distribution ( $\sigma = 265.0 \text{ km s}^{-1}$  – some BH kick velocities are reduced because of mass fallback Fryer et al. 2012). White dwarfs are set not to get any kick velocities upon formation.

Fig. 1 presents an example spatial distribution of two populations for one of the MOCCA simulations from MOCCA-SURVEY-2 as a function of observational half-mass radius ( $R_{\text{hob}}$ ). This MOCCA simulation consists of  $N_1 = 400 \text{ k}$ ,  $N_2 = 400 \text{ k}$ ,  $fb = 0.95$ ,  $R_t = 60 \text{ pc}$ ,  $R_h = 1.2 \text{ pc}$ , and  $\text{conc}_{\text{pop}} = 0.5$  (green SG is two times more densely concentrated than violet FG). We include this plot to emphasize what are the scales of the distributions of the two populations. It means that most of the total mass is very much in the centre, and the cluster has a lot of space to expand. In many aspects, it behaves like an isolated cluster. The plot shows also that SG dominates in the centre, whereas FG in the outskirts of the cluster.

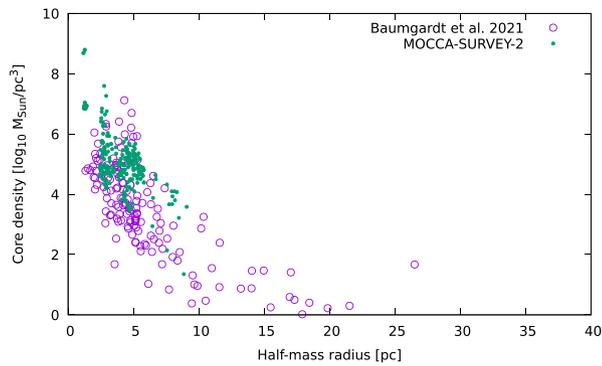
It is important to stress out that the initial conditions were chosen to test the possible mixing between different stellar populations for GCs of realistic sizes. Thus, we have chosen rather dense models in order to make the mixing process more apparent. Moreover, the



**Figure 1.** An example radial density profile of FG and SG for one of the MOCCA simulation ( $N_1 = 400 \text{ k}$ ,  $N_2 = 400 \text{ k}$ ,  $fb = 0.95$  for both populations,  $R_t = 60 \text{ pc}$ ,  $R_h = 1.2 \text{ pc}$ ) as a function of observational half-mass radius ( $R_{\text{hob}}$ ). The concentration parameter ( $\text{conc}_{\text{pop}}$ ) of the SG (green) with respect to the FG (violet) is equal to 0.5, which means that the second population is two times more densely concentrated than the FG.

initial conditions were chosen in such a way that for some of the models FG is denser, for other SG, and on top of that there are some simulations where the distributions of FG and SG are the same. Additionally, we have included also the models in which the SG is more numerous (the sole reason for that was testing, because otherwise it would be rather hard to justify why SG could have initially more stars than FG). Moreover, it is important to point out that the paper is mainly focused on the models based on the scenario described in Calura et al. (2019). In this scenario, SG forms in the very cluster centre, after some time delay (usually a few dozens of Myr), from gas lost due to stellar winds of AGB stars and pristine gas reaccrued by GC during its movement through gas cloud left after GC formation. There are, however, some deviations in our models to this scenario. We wanted to simplify the physical picture of SG formation. We assumed that there is no time delay for SG formation. Thus, the cluster potential was deeper than for only FG stars. To account for the fast FG cluster expansion (before SG formation) due to strong mass-loss of massive stars we assumed larger maximum mass for SG stars. This, a bit artificial assumption, increases overall cluster expansion, keeping it possibly close to expansion of only FG star cluster. Also, because of stronger mass segregation of SG stars, it keeps the SG overall more concentrated with respect to the FG as it is suggested by Calura et al. (2019). The hydrodynamic simulations tell us that the SG is relatively dense and concentrated in the very cluster centre, so newly formed massive SG stars are prone to collisions and mergers which will produce stars even more massive (in runaway collisions it is probable to form stars with masses larger than even  $100 M_{\odot}$ ). Our assumption of large maximum IMF mass is a very simple proxy for such massive stars. Additionally, in dense environment, a significant fraction of massive stars can be kicked out of the system due to dynamical interactions before SNe explosions. The large SG fraction, of the order of 1/3, was assumed (together with large maximum IMF mass) to mimic FG (without SG) expansion. The stronger mass-loss from SG and stronger energy generation in dense and massive SG help in stronger expanding the cluster.

Although the purpose of choosing such initial conditions was to test the mixing between populations and we did not plan to simulate real populations of Milky Way GCs it turned out that MOCCA-SURVEY-



**Figure 2.** Central density as a function of half-mass radius from MOCCA-SURVEY-2 simulations for 12 Gyr (green dots) together with the Milky Way GCs parameters taken from Baumgardt & Vasiliev (2021, violet circles). A plot with cluster mass as a function of half-mass radius one can find in Maliszewski et al. (2022).

2 covers the latter surprisingly well. In Fig. 2, we present the global parameters (central density as a function of time, total cluster mass as a function of time one can find in Maliszewski et al. 2022) on top panel, and cluster total mass on bottom panel) for MOCCA-SURVEY-2 simulations at the time 12 Gyr together with the Milky Way GCs parameters determined by Baumgardt & Vasiliev (2021) and Vasiliev & Baumgardt (2021). It is interesting to see that even with our initially dense models we cover the real GCs parameters very well. The other parts of the plots, especially those with the higher half-mass radii ( $R_h$ ), are not covered simply because we did not have wide enough initial values (e.g. MOCCA-SURVEY-2 does not have models with low masses). The initial conditions degenerate over time and the models which initially had different degree of tidal filling can end up at 12 Gyr having similar masses and  $R_h$ . It is unexpected to see that even starting with quite high densities (of the order of  $10^7 M_\odot \text{pc}^{-3}$ , up to  $\sim 10^8 M_\odot \text{pc}^{-3}$ ) we eventually get the clusters which cover MW GCs properties quite well (in the given mass regime of our chosen initial conditions). Some of our models have very high initial central densities which may present a challenge to justify them as physically reasonable. However, there are cases where such initial extreme densities are considered as the one which were present at the formation of some GCs (e.g.  $\approx 10^8 M_\odot \text{pc}^{-3}$  for  $\omega$  Cen; Marks, Kroupa & Dabringhausen 2021). The central densities (top panel in Fig. 2) fall especially well within the real GCs values – it is very unexpected result but gives additional confidence that the findings presented in this work can be applied and discussed with respect to the MW GCs. It has additionally some interesting implications for observations discussed later in Section 4.

The MOCCA-SURVEY-2 was designed to test a new major version of MOCCA code. The plan is to extend this survey further with new models. We especially plan to include models which will cover better and more comprehensively the global parameter space of real MW GCs. We also plan to include simulations with gravitational kicks between dynamically merging stars, simulations with STARTRACK evolutionary code (Belczynski, Kalogera & Bulik 2002; Belczynski et al. 2008), simulations with the use of TSUNAMI code (Trani, Fujii & Spera 2019), as a replacement of the internal integration code in FEWBODY (Fregeau et al. 2004), and simulations with hierarchical systems too. TSUNAMI code is especially important because it includes post-Newtonian

terms up to order 2.5 and tidal forces (equilibrium and dynamical tides).

### 2.3 Data analysis

Data analysis for the purpose of this paper is done with BEANS software<sup>4</sup> (Hypki 2018). BEANS is open-source, web-based software for interactive and distributed data analysis of huge data sets.

The MOCCA simulations take a lot of disc space. There are many MOCCA simulations already done and every one of them contains millions of rows with stellar evolution and dynamical events. Thus, there is a strong need to have a tool for distributed data analysis which would simplify data examination as much as possible. BEANS helps to solve all these problems. It allows to create self-explanatory, shareable notebooks which provide the full data analysis for the need of this project.

Access to all MOCCA simulations is available for all interested parties upon requests online through the web page <http://beans.camk.edu.pl>. It is our working BEANS instance which has access to all MOCCA simulations computed for (older) MOCCA-SURVEY-1 and the current MOCCA-SURVEY-2. The web page provides necessary computing power and storage to analyse the MOCCA simulations in any possible way.

BEANS is ready for use in production. It already has ability to write scripts in Apache Pig<sup>5</sup> (high level language for Apache Hadoop platform<sup>6</sup>), AWK, and Python. In order to make it useful for as many users as possible there is a plan to write a number of additional plugins. One of them will be a plugin to work with Virtual Observatory.<sup>7</sup> In this way BEANS will gain access to a huge amount of observational data, and to other simulations.

One of the example scripts written in BEANS is described in details in the Appendix B. It shows how one can analyse all of the MOCCA-SURVEY-2 simulations in parallel with an easy to understand script written in Apache Pig.

## 3 RESULTS

The new version of the MOCCA code was profoundly tested. We have checked a huge number of global parameters of star clusters, snapshots of the system, we compared the results with previous version of the code. We have also performed full technical tests to make sure that there are no problems with the MOCCA simulations.

### 3.1 Comparison with the $N$ -body models

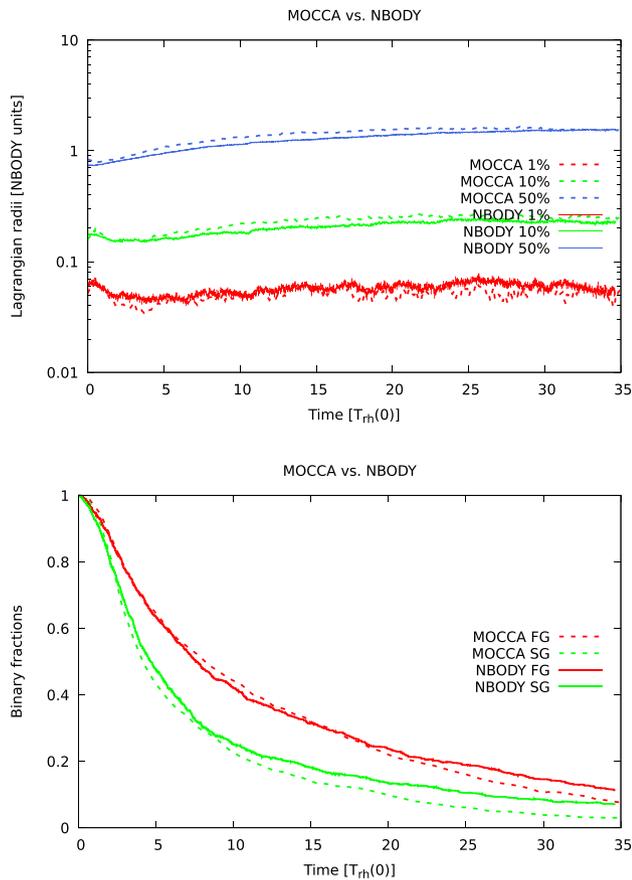
In order to test and to validate the new version of the MOCCA code, it was tested and compared with existing NBODY models. We have decided to use a few models presented in Hong et al. (2015, 2016), and we are presenting the results for one of them: initial King model  $W_0 = 7$ ,  $N = 20$  k, binary fraction = 0.1,  $R_{h,FG}/R_{h,SG} = 5$ , and the initial binary hardness parameter  $\chi_{g,0} = 20$  (for definitions see section 2 in Hong et al. 2015). All the particles in the simulation have equal-mass, all the binaries from FG and SG have the same binding energy, and the stellar evolution is switched off. The MOCCA simulation was started with exactly the same initial conditions (e.g.

<sup>4</sup><https://beanscode.net/>

<sup>5</sup><https://pig.apache.org>

<sup>6</sup><https://hadoop.apache.org>

<sup>7</sup><https://ivoa.net>



**Figure 3.** Comparison between one selected simulation from Hong et al. (2015, 2016) (King model  $W_0 = 7$ ,  $N = 20$  k, binary fraction = 0.1,  $R_{h,FG}/R_{h,SG} = 5$ , and the initial hardness parameter  $\chi_{g,0} = 20$ , see definitions in Section 2 therein) and MOCCA for exactly the same initial conditions. Top panel presents Lagrangian radii 1 per cent, 10 per cent, and 50 per cent for both codes for all objects, whereas the bottom panel presents binary fractions for FG and SG for the two codes. Both panels present the results as a function of time scaled by the initial half-mass relaxation time ( $T_{rh}(0)$ ). For details, see Section 3.1.

positions, velocities, initial mass) and with the most up-to-date version to make the comparison most accurate.

The comparison between results of the one selected NBODY simulation from Hong et al. (2015) and MOCCA run is presented in Fig. 3. The top panel presents Lagrangian radii 1 per cent, 10 per cent, and 50 per cent in NBODY units as a function of time scaled by the initial half-mass relaxation times ( $T_{rh}(0)$ ). The comparison is truncated to the maximum time the NBODY simulation was computed. The bottom panel presents binary fractions for FG and SG for the two codes, also as a function of time scaled by  $T_{rh}(0)$ . The binary fractions were computed taking into account all bound objects (binaries and single stars) and were scaled by the initial number of objects. The comparison demonstrates that MOCCA reproduces the global evolution of low- $N$  clusters ( $N \sim 20$  000) for the different Lagrangian radii (top panel). MOCCA follows also the evolution of the binary fractions for both populations very accurately (bottom panel). The only difference seems to be present for 50 per cent Lagrangian radius. This is due to the fact that MOCCA cannot follow the stage with initial violent relaxation phase as accurately as NBODY. Moreover, the initial model was not in virial equilibrium (see Hong et al. 2016). NBODY simulation regained the equilibrium quickly ( $\sim$ dynamical

scale), which MOCCA could not follow. The detailed comparison of the older versions of MOCCA (the Monte Carlo approach) with NBODY (direct  $N$ -body code) one can find in Giersz, Heggie & Hurley (2008) which was done for M67 star cluster, in Giersz et al. (2013) which presents comparison with  $N$ -body systems up to  $N = 200$  k particles, or in Wang et al. (2016) where MOCCA simulations were compared with  $N$ -body simulations of four massive GCs with  $10^6$  stars and 5 per cent primordial binaries.

### 3.2 Multiple stellar populations in tidally filling and underfilling clusters

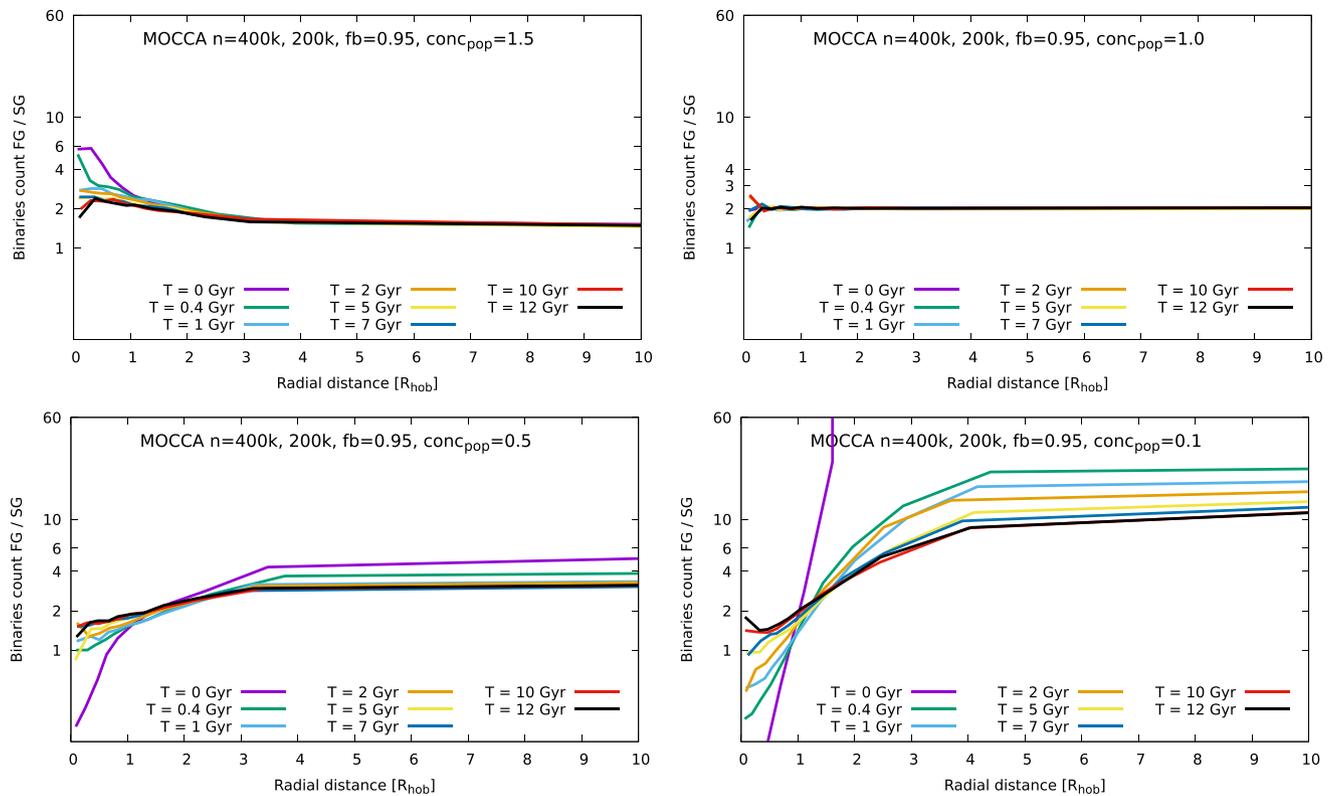
Initial conditions for MSPs are expected to have an influence into the dynamical evolution, in particular mixing, between different populations. Thus, this was taken as the first task to consider while evaluating the results of the MOCCA-SURVEY-2 simulations.

We start the description of the results using number of binaries from both populations. These are pure values taken from the MOCCA simulations, not changed in any way, and degraded by any procedure mimicking the real observations. Later, we will discuss how the found results are influenced by the procedure to selected binaries. We will show whether and if the findings are changed by selected e.g. only main-sequence binaries, observational binaries, or binaries together with single stars. Moreover, the number of binaries, or binary ratios of the two populations are result of complex interplay between the dynamics, binary dissolutions and ejections, which we discuss later.

Fig. 4 shows the time evolution of the radial profile of the ratio of the number of FG to SG binaries for a few representative simulations with different initial concentration parameters ( $\text{conc}_{\text{pop}}$ ). All MOCCA simulations presented in this figure start with the same initial conditions and differ only in the  $\text{conc}_{\text{pop}}$  parameter. In our case, having only two stellar populations,  $\text{conc}_{\text{pop}}$  defines the ratio between the half-mass radius of the SG to the half-mass radius of the FG. The highest  $\text{conc}_{\text{pop}} = 1.5$  is for the MOCCA simulation presented on the top left plot, the lowest  $\text{conc}_{\text{pop}} = 0.1$  is on the bottom right plot. The plots are ordered by decreasing  $\text{conc}_{\text{pop}}$  values. It is important to stress out that the only parameter different for these four simulations is just the concentration parameter between two populations. The main purpose of this figure is to show, in general, how the ratio between the number of binaries from FG and SG evolves with time for different concentration parameters. Unless other specified, every MOCCA simulation presented since this place, have the initial conditions:  $N = 400$  k,  $200$  k,  $W_0 = 6, 6$ ,  $M_{\text{max}} = 150, 150 M_{\odot}$ ,  $m_{\text{func}} = 1, 1$ ,  $fb = 0.95, 0.95$ ,  $R_t = 60$ , and  $\text{evol} = 1$ .

The two top panel in Fig. 4 are included here only for testing because  $\text{conc}_{\text{pop}} = 1.5$  (spatial distribution of SG is more extended than FG), and  $\text{conc}_{\text{pop}} = 1.0$  (FG and SG have the same initial spatial distribution) are hard to justify as a real scenario for the SG formation. However, the top left panel shows that if the FG is more concentrated then it naturally loses more binaries as a result of dynamical interactions, thus the ratio drops with time and eventually at Hubble time drops closer to 1.0. Also, for  $\text{conc}_{\text{pop}} = 1.5$  the ratio between binaries count for the distances larger than  $R_{\text{hob}} = 1$  does not change significantly over time (there is no apparent mixing happening in the outskirts of the cluster). The top right panel shows that if both populations start with the same spatial distribution there is actually no change over the Hubble time for the FG/SG ratio. It is expected behaviour which shows that MOCCA code is working correctly. The slightly scattered values in the centre are simply the fluctuations.

For the MOCCA simulations with deeper concentrated second populations (two bottom plots) one can see that there appears a



**Figure 4.** Ratios between all binary stars count from FG to SG for four selected MOCCA simulations, with initial conditions summarized in the titles, as a function of radial distance scaled by  $R_{\text{hob}}$ . Each plot consists of eight lines (for different times), and the colours are consistent between panels. The only difference between MOCCA models is the  $\text{conc}_{\text{pop}}$  with values from 1.5 (top left), up to 0.1 (bottom right). For details, see Section 3.2.

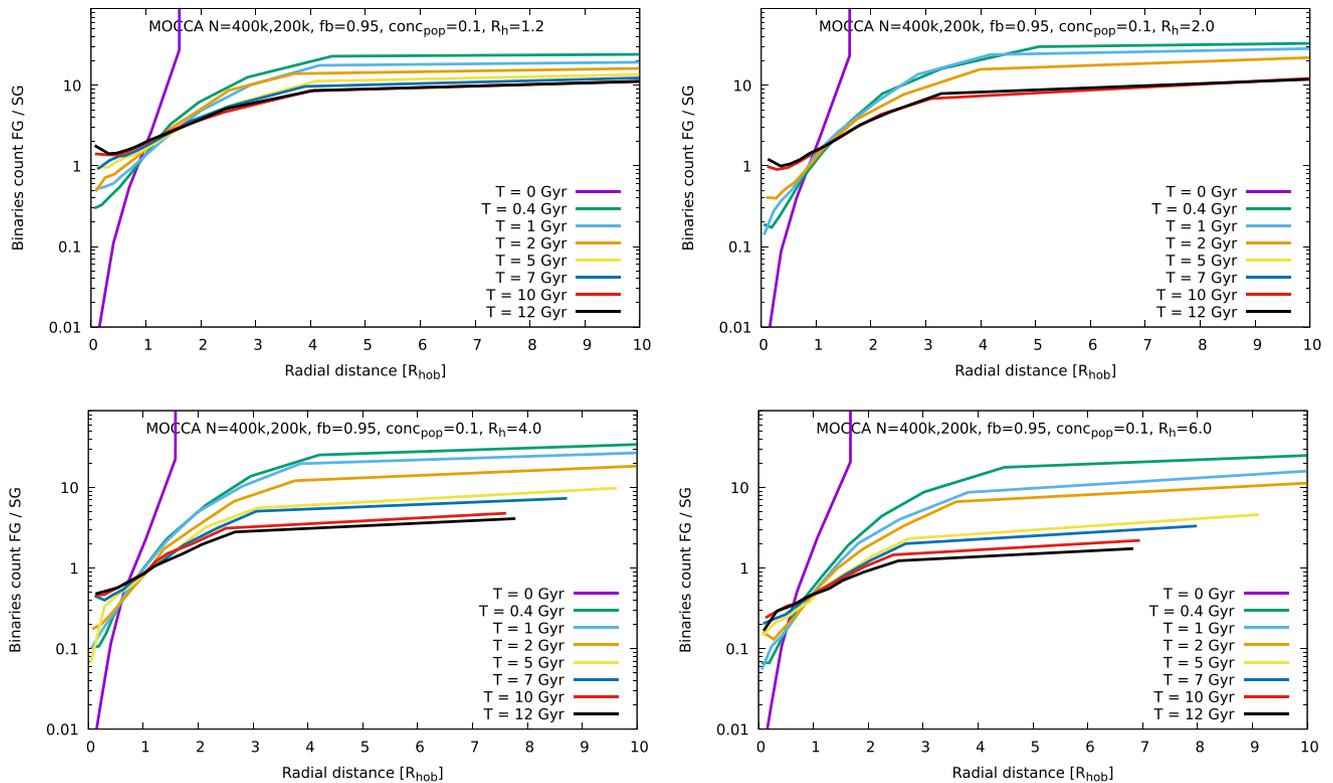
very distinct feature. The ratio between the number of binaries from FG to SG for the central regions below  $R_{\text{hob}}$  (i.e. values on the X-axis smaller than 1.0) for the beginning of the simulations (e.g. 400 Myr, 1 Gyr) have small values. This is because initially the second population is more deeply concentrated. Then, with time this ratio is getting larger, and eventually it becomes larger than one. This means that in the central Lagrangian radii there are now less binaries from SG. These binaries, initially more concentrated, are being destroyed or ejected. Eventually, the ratio throughout all Lagrangian radii is larger than 1.0 – FG gets more numerous than SG for the entire cluster in general agreement with the findings of Hong et al. (2015), Hong et al. (2016), Vesperini et al. (2011), and Sollima et al. (2022).

Another distinct feature observed in Fig. 4, for  $\text{conc}_{\text{pop}} < 1.0$  models, is that the ratio is getting larger for the central regions (essentially below the half-mass radius) and is getting lower for the outskirts of star clusters (further than  $R_{\text{hob}}$ ). Initial shape of the ratio between FG and SG flattens for the entire cluster which means that there is mixing in all regions of the star cluster. For  $\text{conc}_{\text{pop}} = 0.1$  the initial ratio  $\sim 0.3$  (blue colour) at 400 Myr increased to  $\sim 2.0$  at 12 Gyr for the central regions and dropped from  $\sim 15.0$  to  $\sim 8.0$  for the outskirts of the star cluster ( $> 4 \times R_{\text{hob}}$ ).

The two bottom plots in Fig. 4 reveal interesting time evolution of the binary ratio for the two populations for two models which differ in by how much the SG is more deeply concentrated with respect to FG. We have decided to investigate it more deeply and in Fig. 5 we present also the binary ratios between FG and SG but for four models which have the same initial conditions, and only different  $R_{\text{h}}$  (from 1.2 pc which is tidally underfilling model, up

to 6.0 pc which is model close to be tidally filling). The overall evolution of the ratios for times from  $T = 0$ , up to the Hubble time present similar evolution. At the time,  $T = 0$  SG is 10 times more deeply concentrated than FG ( $\text{conc}_{\text{pop}} = 0.1$ ). With time, this ratio increases for all models but in principle only in the central regions of the cluster (where dynamical interactions play much more significant role). In the outer regions of the cluster this ratio drops for all models. However, there is a number of interesting differences for tidally filling and underfilling clusters. For tidally underfilling models (e.g.  $R_{\text{h}} = 1.2$  pc), the binary ratio FG/SG gets larger than 1.0 after a few Gyr. This means that for such dense models, where SG was deeply concentrated and more numerous in the centre, there is a large number of SG binaries which were destroyed or ejected from the centre. In such models, SG population quickly burns out and as a result at the Hubble time the ratio is larger than 1.0 for the entire star cluster. In turn, for the models which are tidally filling (e.g.  $R_{\text{h}} = 6.0$  pc), the initial ratio in the centre is also much smaller than 1.0 (SG dominates in the centre). In such models, SG is also being burned by dynamical interactions, however, on much less scale. Thus, with time the ratio increases (there is more and more SG being destroyed), but the ratio does not go above the value 1.0. For tidally filling clusters, the SG still dominates in the centre of the GC, whereas in the outskirts of the GC, the FG dominates. Additionally, for tidally filling clusters the binary ratios FG/SG decreases more significantly in the outskirts of the cluster. This is a consequence of the fact that tidally filling clusters have more FG binaries closer to the tidal radius and lose a larger fraction of FG single and binary stars.

The ratio between FG to SG binaries during 12 Gyr of evolution changes much more profoundly for tidally filling clusters (bottom



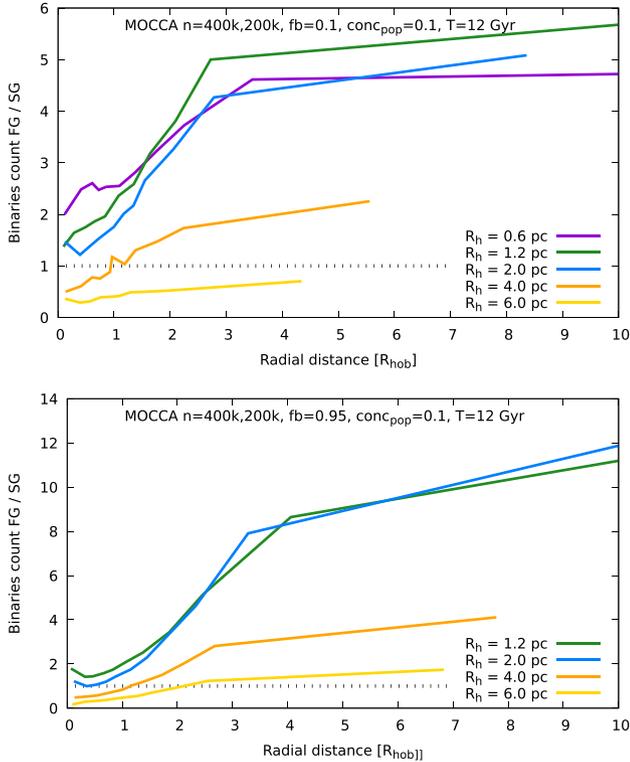
**Figure 5.** Description as in Fig. 4. The MOCCA models differ only with  $R_h$ . For details, see text.

right panel in Fig. 5). The ratio at 0.4 Gyr are of the order of 20.0 at a distance of several  $R_{\text{hob}}$ , and after Hubble time the ratio drops to a value just slightly larger than 1.0 for our model. For tidally filling models SG is burning in the centre, whereas FG are preferentially removed or destroyed in the cluster outskirts – this changes the mass function (Vesperini et al. 2021). For tidally underfilling clusters (top left panel in Fig. 5), the ratio changes less significantly. Such relation could be a valuable tool for observations to try to infer some boundaries for the initial conditions of the GCs.

One of the most prominent features of the evolution of ratios between binaries from FG to SG are for times  $T = 12$  Gyr (black curves in Figs 4 and 5). They reveal how the ratio behaves for tidally filling and underfilling clusters. In order to investigate this more closely, Fig. 6 shows the ratios between binaries FG/SG as a function of radial distance (scaled by  $R_{\text{hob}}$ ). All the ratios come from  $T = 12$  Gyr and were collected from a set of MOCCA simulations which differ only in the initial half-mass radii (the smaller it is the more tidally underfilling and more dense the cluster is initially). On the top panel in Fig. 6, there are MOCCA simulations for 10 per cent of initial binary fractions, and the bottom panel for 95 per cent. The rest of MOCCA initial conditions are summarized in the captions of the plots (notice that all of the simulations have  $\text{conc}_{\text{pop}} = 0.1$  which means that SG is 10 times more concentrated than FG). On both panels, a dashed black line at value 1.0 is plotted for reference. It represents radial distances at which the number of binaries from FG equals SG. All points above this line denote that there are more FG binaries, and all points below the line represent radial distances at which the SG is more numerous. All of the MOCCA simulations in Fig. 6 have initially more numerous FG than SG by factor of 2. The figure shows that the simulations for which FG, at the time  $T = 12$  Gyr, is still more numerous is only for the clusters

which initially were tidally underfilling (smaller  $R_h$  values). Deeper investigation of the internal structure of these models revealed that the clusters, because of their high initial densities, evolve very quickly towards core collapse, and after that the cluster just expands nearly homogeneously and continuously with time (all Lagrangian radii increase in the same way with time, even 1 per cent Lagrangian radius). For these clusters there was plenty of space to expand and preserve the more numerous FG. Only in the central regions, the ratio FG to SG is significantly lower, but yet still larger than 1.0. For tidally underfilling models the initial ratio is well below 1.0 (see violet curve on top left panel in Fig. 5). SG is significantly more numerous at the time  $T = 0$ , but because of the high density SG is quickly burned out and eventually for tidally underfilling clusters FG gets more numerous even in the central regions of the cluster. With larger  $R_h$  values (tidally filling models), the ratio between FG and SG binaries looks different. At some point (in our models  $R_h > 2.0$  pc) SG gets more numerous at the time 12 Gyr at least for the central regions of the cluster. Initially, the number of SG binaries is dominant only in the central regions (for  $R_h = 4.0$  pc for radial distances less than one  $R_h$ ), whereas in the outskirts of the star cluster FG is still more numerous. But at some value (in our case  $R_h = 6.0$  pc), the cluster is initially tidally filling and that changes the resulting number of FG binaries in comparison to SG. For such a model  $R_h$  increases because of the initial mass-loss of the massive stars, and then because of the energy which is being generated in the core. In turn,  $R_t$  decreases due to mass-loss and ejections of stars. The cluster loses more and more FG binaries, because they are closer to the tidal field, and eventually the number of SG binaries begins to dominate on all radial distances.

Fig. 6 shows that the ratio between the number of FG and SG binaries presents the same features for 10 per cent and 95 per cent



**Figure 6.** Ratios between all binary stars count from FG to SG for a few selected MOCCA simulations for the time 12 Gyr for different initial half-mass radii ( $R_h$ ) as a function of  $R_{\text{hob}}$ . Top panel contains MOCCA simulation with initial 10 per cent of binaries, the bottom panel 95 per cent. Other initial parameters are summarized in the captions. The  $\text{conc}_{\text{pop}}$  between both populations is the same for all models and equals 0.1 (SG is initially 10 times more concentrated than FG). Additionally, artificial line at value 1.0 is plotted as a reference, dividing a plot into two regions: above the line the number of FG binaries dominates, below FG binaries. See Section 3 for details.

initial binary fractions. In the centre of clusters, the ratio is smaller than 1 ( $R_h = 4.0, 6.0$ ), and in the outskirts the ratio is large for both of the binary fractions. It seems that the results presented in the figure are not influenced significantly by the initial binary fraction and the ratios between FG and SG behave similarly.

Fig. 7 shows, as an example, the number of escaped (left-hand panel) and destroyed binaries (right-hand panel) for the two populations in two MOCCA simulations (one underfilling –  $R_h = 1.2$  pc, and one filling –  $R_h = 6.0$  pc) as a fraction of the initial binaries number for any given population. The figure explains why the number of binaries changes like e.g. in Fig. 6 differently for filling and underfilling clusters. For the model tidally underfilling the fraction of destroyed binaries is several times larger than for the model tidally filling (SG is being destroyed the most, and at the time 12 Gyr 80 per cent of all SG binaries are eventually dissolved that way). In turn, for the model tidally filling, the more important mechanism for removal of binaries is due to escapers (around 0.5 of FG binaries escaped after 12 Gyr, and only 0.1 of the SG – this is because more FG binaries are less concentrated than the SG). There are two additional features visible in Fig. 7. The ratio between destroyed FG and SG binaries is noticeably larger for the tidally filling model than for the underfilling one. Also, what is interesting and unexpected, the number of escaped binaries is larger for SG than FG for the tidally underfilling model. This suggests that escapes connected with strong

dynamical binary interactions play an important role in the binary removal process.

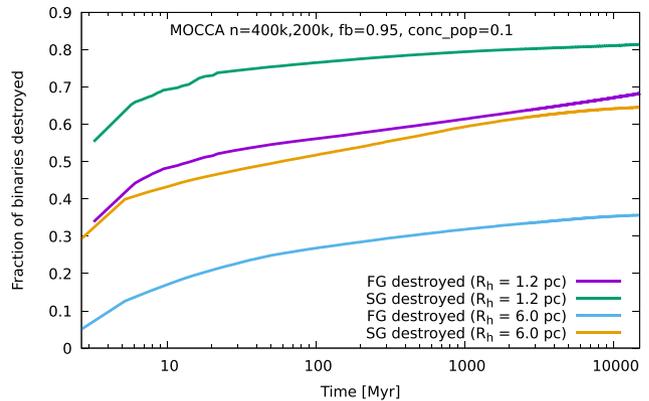
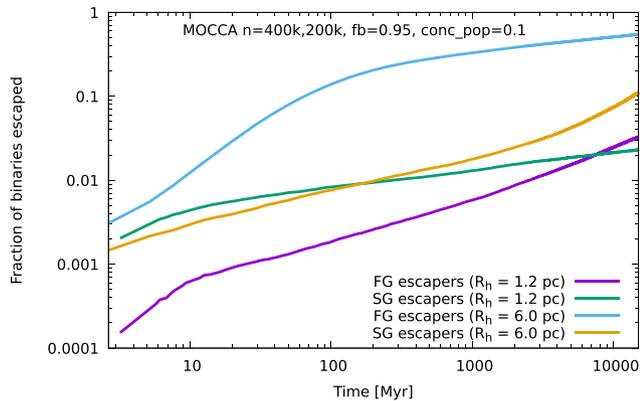
For completeness let’s examine the behaviour for the ratio between binaries from FG to SG also from the point of view of the  $\text{conc}_{\text{pop}}$  parameter. Fig. 8 shows the ratio between binary stars FG to SG at the time 12 Gyr, but the simulations in this case differ only by the  $\text{conc}_{\text{pop}}$  parameter. All other initial parameters are the same and summarized in the caption. All of the MOCCA models in this figure are underfilling ( $R_h = 1.2$  pc). The figure clearly shows that for the smaller  $\text{conc}_{\text{pop}}$  values we observe the same result for the ratio between binaries count FG to SG like in the previous Fig. 6. Initially, the overall number of FG is larger than SG for the whole cluster, but not in the central regions. Because SG was initially more concentrated the ratio was below value 1.0 in the centre. With time, and because SG binaries were quickly dissolved or ejected, this ratio reverts and as a result the FG/SG ratio gets larger than 1.0 even for the central regions of the cluster. Fig. 8 shows that the more the SG is concentrated, the larger the ratio between FG to SG is. This can have an important implications for observations. For  $\text{conc}_{\text{pop}}$  equal to 1.0 (FG is initially distributed statistically in the same way as SG), the ratio is rather flat around value 2.0, so in principle the ratio between initially more numerous FG populations is preserved. There is only a slight bump in this ratio for the central regions (below  $R_{\text{hob}} < 0.5$  pc) which is connected with fluctuations.

The model, which initially consists of less concentrated SG stars ( $R_{\text{hSG}}/R_{\text{hFG}} = 1.5$ ) is shown here for completeness despite the fact it is rather hard to justify initial concentration (there is no reason why SG could be less concentrated than FG). However, it is interesting to see that for such a model the initial radial distribution of FG and SG are basically saved. FG is more numerous in the central regions, whereas SG is more numerous in the outskirts of the cluster. After 12 Gyr, this distribution is still preserved and binaries count reverts from a value  $>2.0$  to a value  $<2.0$  indeed around  $R_{\text{hob}}$ . Here again, there is just small decrease in the number of FG in the very centre ( $R_{\text{hob}} < 0.2$ ) which is connected with FG being preferentially more often dissolved or ejected due to dynamical interactions.

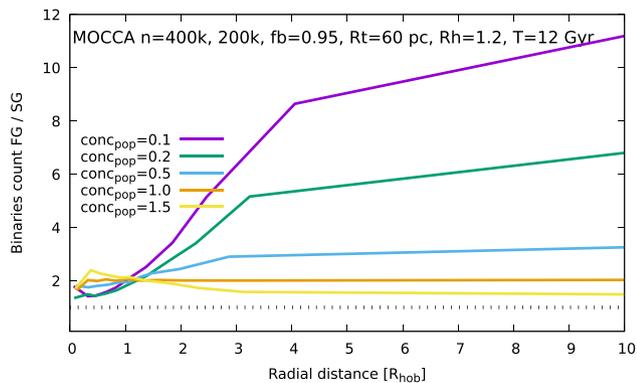
### 3.3 Different ways of computing binaries count for FG and SG

We have checked whether the changes of the ratio FG/SG (e.g. as in Fig. 4) are also visible for the binaries if we take into account binaries computed in a different way e.g. only MS binaries.

Fig. 9 shows the ratios between FG and SG binaries as a function of the radial distance scaled by  $R_{\text{hob}}$ , but computed in a few different ways than in Fig. 4 for  $T = 12$  Gyr. There are two MOCCA simulations presented on the plot. Their common initial properties are summarized in the caption. They differ only on the  $R_h$ . The five first dashed curves on the plot concern MOCCA simulation with initial  $R_h = 1.2$  pc (small  $R_h$ ), which represents a class of underfilling models. In turn, the five bottom solid lines represent MOCCA model with initial  $R_h = 6.0$  pc, which represents filling model (large  $R_h$ ). Each five lines represents (in that order) all binaries (of any masses), observational binaries, only MS binaries, binaries selected based on Lucatello et al. (2015, L15), and binaries with at least one red giant (RG). The observational binary is a binary if it fulfills the following conditions: (1) consists of two MS stars, (2) more massive star is  $>0.4 M_{\odot}$ , (3) mass ratio  $q > 0.5$ , 4. more massive star is less than  $M_{\text{turnoff}}$ . The conditions roughly follow the prescriptions provided by Milone et al. (2012). The MS binary is simply a binary which consists of two MS stars of any mass. And finally, the criteria for L15 are following:  $m_1 > 0.7 M_{\odot}$ ,  $\log P/d > 2$ , and  $q > 0.3$ . These criteria have been chosen from the reported detection efficiency as a function



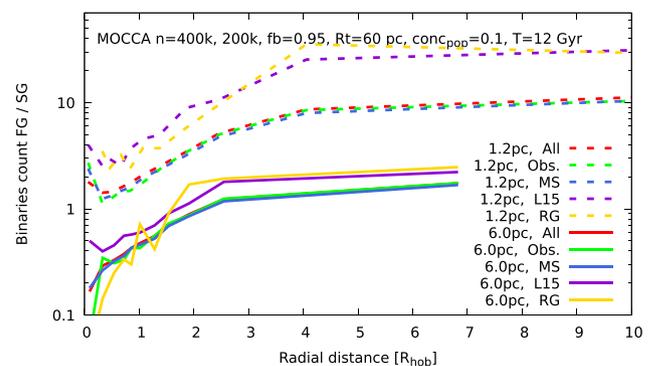
**Figure 7.** Number of binaries escaped from a star cluster (left-hand panel) and number of binaries dissolved due to dynamical interactions and as a result of stellar evolution (right-hand panel) for FG and SG binaries for two selected MOCCA simulations as a fraction of the initial number of binaries. MOCCA simulations have parameters summarized in the titles, but the only difference between them is  $R_h$  with values 1.2 – tidally underfilling cluster, and 6.0 – tidally filling cluster. For every MOCCA simulation, there are two lines on each panel for FG and SG binaries. For details, see Section 3.2.



**Figure 8.** The ratio between binaries count of FG and SG for a set of MOCCA simulations which differ in  $\text{conc}_{\text{pop}}$  values only (e.g.  $\text{conc}_{\text{pop}} = 0.1$  means that SG is initially 10 times more concentrated than FG) as a function of radial distance scaled by  $R_{\text{hob}}$ . The other initial parameters are summarized in the caption. Additionally, artificial line at value 1.0 is plotted as a reference (above the line the number of FG binaries dominates). The model with  $\text{conc}_{\text{pop}} = 1.5$ , in which SG is initially less concentrated than FG, is rather hard to explain physically. However, it is presented here for completeness (MOCCA-SURVEY-2 is a test survey and some of the models had such concentrations for SG, see Section 2 for details).

of the orbital period reported by these authors (Fig. 4; Lucatello et al. 2015) and to impose the presence of a massive primary component.

Fig. 9 shows that the ratio of binaries from different populations looks very similar for any method of computing the number of binaries if we take some subset of MS binaries. The shapes of the lines are the same as in Fig. 6. For the models which are tidally underfilling ( $R_h = 1.2$  pc) the ratio has the clear low values in the central regions with the ratios above 1.0 (FG dominates). The higher values are in the outskirts of star clusters. In turn, for the tidally filling clusters ( $R_h = 6.0$  pc) the lowest values are also in the centre but lower than 1.0, which means that SG is more numerous. At some distance, this ratio gets larger than 1.0 and FG starts to dominate. The same features and the same physical processes are responsible for the shape of these curves as it was described in Section 3.2 (e.g. Fig. 6). Some differences appear for the cases of L15 and for RG binaries. For them the shape is the same, but the ratios are slightly larger, especially for underfilling model (yellow, and



**Figure 9.** Binaries number computed in different ways for a set of MOCCA simulations for two different  $R_h$  and for time 12 Gyr as a function of the radial distance scaled by  $R_{\text{hob}}$ . Each line corresponds to one MOCCA simulation whose basic parameters are summarized in the caption. The simulations belong to two categories: underfilling cluster ( $R_h = 1.2$  pc, dashed lines), and filling cluster ( $R_h = 6.0$  pc, solid lines). The binaries are computed for the two groups in five different ways: all binaries, observational binaries, MS binaries, based on Lucatello et al. (2015, L15), and for red giants. For detailed definitions and a discussion, see Section 3.3.

violet lines on Fig. 9). Lucatello et al. (2015) report binary fraction of 4.9 per cent  $\pm$  1.3 per cent for FG and 1.2 per cent  $\pm$  0.4 per cent for SG (with remark that the spectroscopic targets are located in a non-uniform way between  $0.5 - 3 \times R_{\text{hob}}$ ). We have computed them for models MOCCA  $n = 400k, 200k, f_b = 0.95, R_t = 60$  pc,  $\text{conc}_{\text{pop}} = 0.1$  for  $R_h = 1.2, 2.0, 4.0,$  and  $6.0$  pc and the results, taken cautiously, are consistent within  $2\sigma$  with all the above models, and within  $1\sigma$  with the model tidally underfilling.

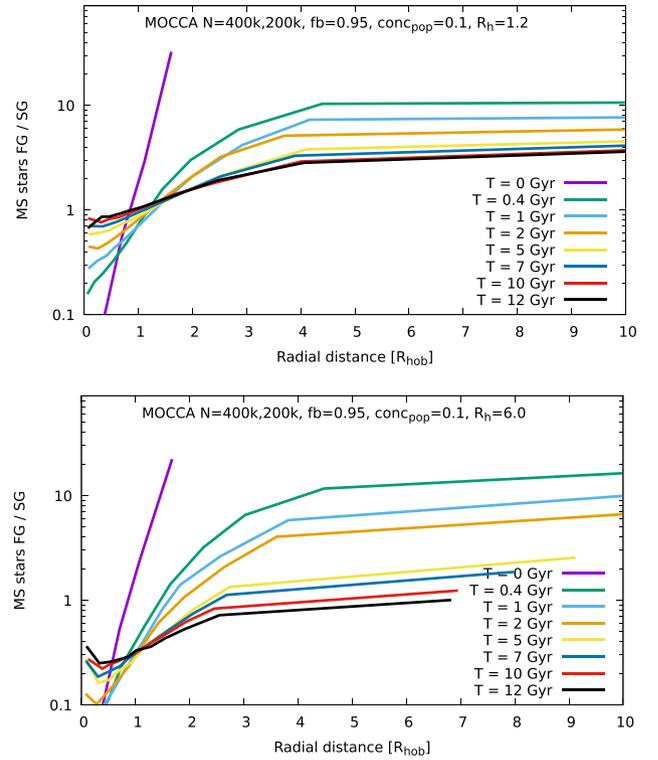
The tidally underfilling models give a better agreement with Lucatello et al. (2015) for the SG binary fractions (see Fig. 9). On the other hand, the tidally filling models suggest that they are better at obtaining the fraction of FG to SG stars from Milone et al. (2017a) (see Section 4 for details). It mildly suggests that this discrepancy in binary fraction and overall fraction of SG stars may indicate that the SG stars are indeed born with a very high concentration, as suggested by Gratton et al. (2019) and Sollima et al. (2022). This high concentration of SG stars (within an overall tidally filling cluster) would result in increased destruction of SG binaries, while at the same time the overall fraction of SG stars would also increase

**Table 2.** Number of FG binaries computed in a different way for one selected MOCCA simulation (underfilling model with  $N_1 = 400\text{ k}$ ,  $N_2 = 200\text{ k}$ ,  $\text{fb} = 0.95$ ,  $R_h = 1.2\text{ pc}$ ,  $R_1 = 60\text{ pc}$ ,  $R_{h\text{SG}}/R_{h\text{FG}} = 0.1$ ,  $T = 12\text{ Gyr}$  for FG – the same as one of the models from Fig. 9). The first column is Lagrangian radius, the second column contains the number of all binaries (with any mass and types, in brackets is given the number single stars), and the next columns the number of observational binaries, MS binaries only (and number of MS single stars in brackets), and the number of binaries based on Lucatello et al. (2015, L15). The numbers of binaries are given for every Lagrangian radius separately, these values are not cumulative.

Lagr. rad.	All bin.	Obs. bin.	MS bin.	L15
1 per cent	52 (181)	19	43 (122)	8
10 per cent	1695 (8303)	391	1230 (6226)	186
20 per cent	2135 (11 405)	496	1605 (8673)	229
30 per cent	4174 (23 125)	962	3289 (18 371)	350
40 per cent	4074 (23 317)	813	3261 (19 056)	321
50 per cent	6499 (37 835)	1210	5368 (31 682)	419
60 per cent	8181 (47 927)	1411	6942 (41 199)	492
70 per cent	12 055 (69 087)	1929	10 295 (60 691)	685
80 per cent	13 206 (70 470)	1953	11 393 (62 606)	684
90 per cent	23 352 (11 0171)	3231	20 264 (99 045)	1221
100 per cent	23 100 (93 287)	3082	20 015 (84 252)	1208

as we will preferentially lose FG stars. We point out, however, that Lucatello et al. (2015) results’ are averaged over 6 GCs, and the number of examined stars is relatively small. Moreover, the average was done for the selected GCs for which Milone et al. (2017a) values of  $N_{\text{FG}}/N_{\text{Tot}}$  range from 0.175 (for NGC 104) up to 0.542 (for NGC 288). Thus, this conclusion is based on limited information and has to be treated cautiously.

It is interesting to see that these different ways of computing binaries follow all of the binaries so well. From simulations, one can have full information about binaries, their properties, number, and spatial positions. However, for observations this is not possible. It is even more surprising if one takes into account how many of the binaries are in these groups. Table 2 shows the number of binaries for one selected MOCCA simulation (tidally underfilling model with  $R_h = 1.2\text{ pc}$  from Fig. 9) for a number of Lagrangian radii. Additionally, for two columns the table contains the number of single stars in equivalent categories for comparison. The table shows that the number of observational binaries are significantly less numerous than all binaries in the system. But yet, they follow the internal structure of the cluster in a very similar way, and thus they also follow the mixing between populations very well too. It is surprising to see this because the number of all binaries (all stellar types, all masses) for the model is still around 100 k, but the number of e.g. observational binaries is around 14 k only. Thus, the observational binaries should mimic the photometric observations and one can see that even such low number of observational objects can follow the global properties of the cluster. Moreover, the visible fraction of binaries varies between different regions of the cluster which additionally should disturb the picture: for Lagrangian radius 1 per cent there are 19 observational binaries from 52 all binaries which give 36 per cent, whereas for the last Lagrangian radius only 3082 binaries are visible from all 23 100 binaries, which give only 13 per cent. However, even with such low numbers of binaries, it is still possible to reflect the global properties of the whole cluster. It seems that observational binaries are sufficiently good proxy to probe the entire cluster. This result is probably connected with the fact that at late evolution times there is only a small difference between



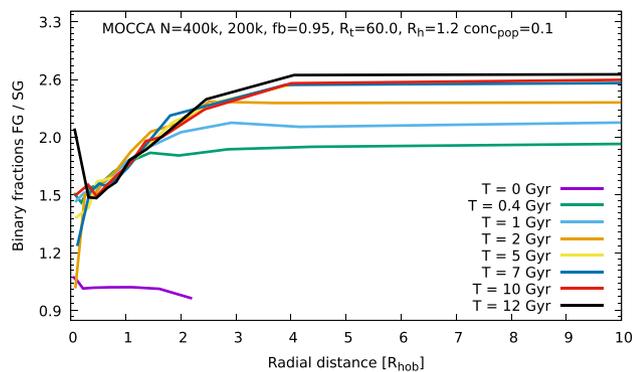
**Figure 10.** Description as in Fig. 5 and also with the same MOCCA models. The only difference is that only two MOCCA models are shown here (for smallest and largest  $R_h$ ) and plots are prepared based on MS single stars only (not binaries). For details, see Section 3.3.

masses of all stars/binaries and also the number of non-MS binaries is relatively small.

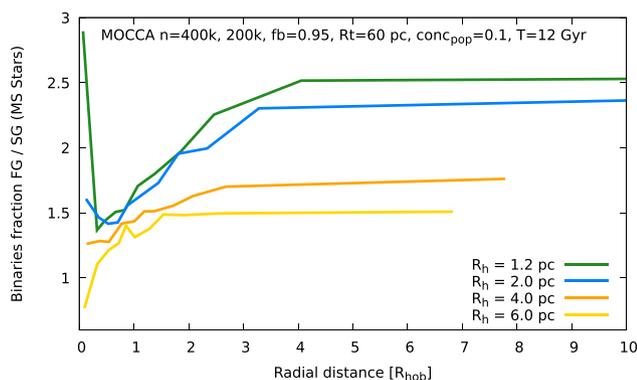
And finally, Fig. 10, presents two MOCCA simulations – the same as from Fig. 5 but only with two initial  $R_h = 1.2, 6.0$  and prepared with MS single stars (not binaries). The plot shows that even taking only MS single stars into account (not binaries), to track the mixing between the two populations, seems to be enough. MS single stars present the same features as the binaries described in details in Section 3.2. The only noticeable difference is that for MS single stars the ratios between FG and SG seem to have values consequently closer to 1.0. In Fig. 5 for MOCCA simulation with  $R_h = 1.2$  the ratio between FG and SG binaries below  $R_{\text{hob}}$  has the value about 2.0, whereas for MS single stars in Fig. 10 this ratio is about 1.0. In turn, for MOCCA model with  $R_h = 6.0$  it is the ratio with value about 0.2 for binaries, and for MS stars about the value 0.3. We did not investigate these differences closely. From now on in this paper FG and SG mean both FG, SG binaries, or FG, SG stars (unless specified otherwise).

### 3.4 Binary fractions

The same trend as in Fig. 4 for binaries count one can be obtained also with the binary fractions. A binary fraction is defined here simply as a ratio between the number of binaries to the number of binaries and single stars (for any given population). The Fig. 11 shows radial profile of ratios between binary fractions from FG and SG for a few selected time-steps for one MOCCA simulation. The concentration parameter  $\text{conc}_{\text{pop}}$  between populations equals to 0.1. As expected, the initial binary ratio is equal to 1.0 for the entire star cluster (violet line in Fig. 11). Later with time the binary ratio between FG and



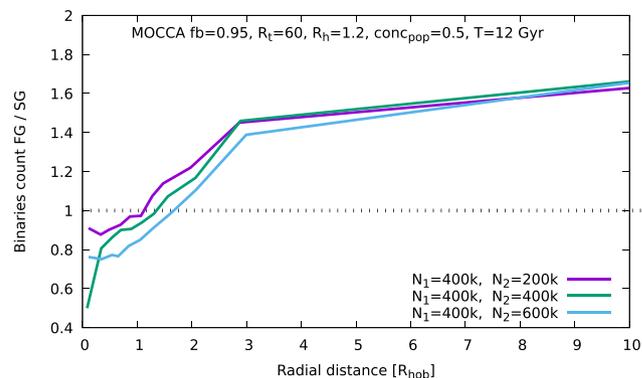
**Figure 11.** Radial profile of ratios between binary fractions from FG and SG for a few selected times for one MOCCA simulation with concentration parameter between populations equals to 0.1 ( $\text{conc}_{\text{pop}} = 0.1$ ) as a function of radial distance scaled by  $R_{\text{hob}}$ . For details, see Section 3.3.



**Figure 12.** Binary fractions between FG and SG (only MS stars) for a set of MOCCA simulations which differ in half-mass radii (from 1.2 pc – tidally filling, up to 6.0 pc – filling models) as a function of radial distance scaled by  $R_{\text{hob}}$ . The rest of the initial parameters are summarized in the caption. The binary fractions between FG and SG show the same evolution of the mixing between two populations as computed from the number of all binaries (see Fig. 11, or Fig. 9).

SG increases, and eventually after Hubble time there is significantly more binaries from FG than the SG in the central regions of the cluster – the same evolution of populations is observed with the larger binary fraction from FG and SG visible on the Fig. 4, or Fig. 5. It is also interesting to see that the ratio increases quickly over about first 2 Gyr, and after that the ratio increases much more slowly. This is because for underfilling models for substantial part of their evolution they behave as isolated clusters – they nearly freely expand. FG was initially more extended than SG, so it expands a bit faster. Additionally, because of increase of  $R_{\text{hob}}$  the same value on X-axis means actually larger physical distance. SG binaries need more time to access the same physical distance as FG binaries. This leads to increase of the ratio between FG and SG binaries. Later, when cluster becomes tidally filling the described process slows down and SG binaries easier can catch FG binaries.

Similar evolution of the binary fraction ratios stands valid also for e.g. MS binaries will be taken into account only. Fig. 12 shows the ratio between binary fractions FG and SG for a few selected MOCCA simulations which have the same initial conditions (given in the caption), but have different  $R_{\text{h}}$ . The ratios between the binary fractions of FG and SG are computed here using only MS binaries



**Figure 13.** Ratio between binaries number of FG to SG for three MOCCA simulations for 12 Gyr as a function of radial distance scaled by  $R_{\text{hob}}$ . All MOCCA simulations have  $\text{fb} = 0.95$ ,  $R_t = 60$ ,  $R_h = 1.2$ ,  $\text{conc}_{\text{pop}} = 0.5$  which means that SG is two times more concentrated than FG. They differ in the initial number of second population ( $N_2 = 200\text{k}, 400\text{k}, 600\text{k}$ ), thus their initial mass differs too. The ratios are scaled by the initial number of binaries for easier comparison. For details, see Section 3.3.

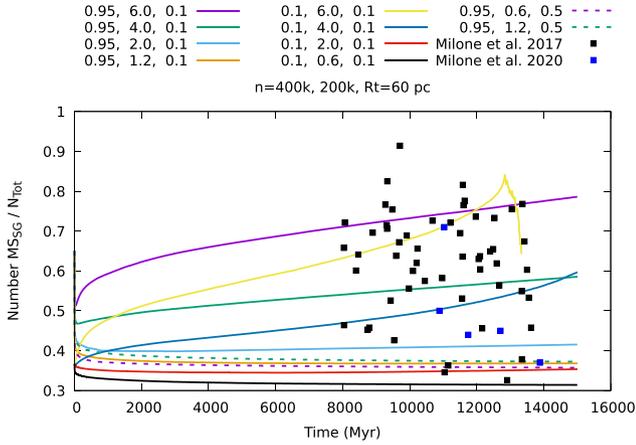
and stars. The models with small  $R_{\text{h}}$  are tidally underfilling, whereas the one with large  $R_{\text{h}}$  values are filling. One can see that the evolution of the ratio between two populations is basically the same as it was already discussed in Section 3.2. The ratio between the binaries fractions have a peak in the centre of the cluster for the tidally filling models (small  $R_{\text{h}} = 1.2$  pc) which is caused by quickly burned out SG and as a result FG starts to dominate (the same evolution like black curve in Fig. 11). In turn, for tidally filling models, there the ratio between the binary fractions is smaller than 1.0 for central regions (SG dominates), and at some distance (around  $R_{\text{hob}}$ ) the ratio gets larger than 1.0. It means that FG dominates, and the ratio raises to the value around 1.5 which is very close to the initial one.

Fig. 12 shows also that the ratio between binary fractions FG to SG is smaller than 1.0 only in the very centre of a cluster (much below  $R_{\text{hob}}$ ) for tidally filling clusters (large initial  $R_{\text{h}}$  values). Only there SG binary fractions are larger than for FG. It seems that SG for tidally filling clusters is not dissolved so efficiently like for much more dense tidally underfilling clusters. This could be potentially important imprint in observational data which could help to narrow down the initial conditions for star clusters.

The initial mass of the model does not change the ratios of binaries number of FG to SG too. Fig. 13 presents three MOCCA simulations ( $\text{fb} = 0.95$ ,  $R_t = 60$ ,  $R_h = 1.2$ ,  $\text{conc}_{\text{pop}} = 0.5$ ) at 12 Gyr as a function of radial distance scaled by  $R_{\text{hob}}$ . They have different only the initial number of SG stars ( $N_2 = 200\text{k}, 400\text{k}, 600\text{k}$ ). The ratios are scaled by the initial number of binaries in order to compare them on one plot. As a result the models have different initial masses. The plot shows that the evolution of the ratio between FG and SG binaries is not affected much by the initial mass. Only the very centre of the cluster shows larger differences between different models, but this is most likely due to fluctuations of small number of binaries. This can have rather important implications for the observations because it shows that the mixing does not depend much on the initial mass of the star cluster. It seems that the initial degree of tidal filling or underfilling matters more.

## 4 DISCUSSION

The analysis of the MOCCA-SURVEY-2 simulations revealed a number of interesting features of MSPs which could be helpful to understand



**Figure 14.** Number of MS SG as a fraction of all objects for a number of selected MOCCA simulations presented with lines as a function of time. The black and blue dots represent values for real Milky Way GCs (Milone et al. 2017b, 2020b) and are spread in time randomly between 8 and 14 Gyr for clarity. The MOCCA simulations are characterized with three values: binary fraction, half-mass radius, and  $\text{conc}_{\text{pop}}$ , respectively. They represent models which are tidally underfilling and strongly tidally filling.

some of the observational properties of the real GCs. In this section, we will discuss the results of our analysis especially with respect to the boundaries of the initial conditions of GCs. We would like to remind that we are working in a scenario in which SG was formed from an FG ejecta and pristine gas in the central regions of the FG cluster. We also focus in this work mainly on the MOCCA simulations with  $N_{\text{FG}} = 400$  k, and  $N_{\text{SG}} = 200$  k (see Table 1) because SG as it is expected from theory and observations is less numerous than FG and this allows to compare to some extent MOCCA simulations with observations.

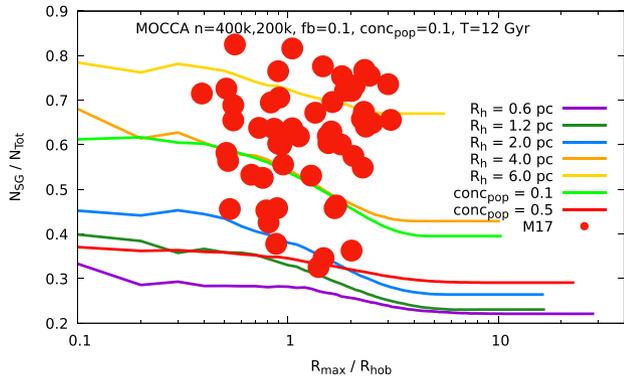
We emphasize that the primary goal of this study was to explore the mixing process of FG and SG binary stars and the range of models explored was not meant to provide a comprehensive survey of initial conditions mimicking the Milky Way GCs. However, as it turned out, the models cover the properties of the Galaxy GCs surprisingly well (see Fig. 2). Fig. 14 shows MS SG as a fraction of the total number of objects for a number of MOCCA simulations (as lines) and for real Milky Way GCs (Milone et al. 2017b) as black dots. Each MOCCA simulation (400 k FG, 200 k SG,  $R_t = 60$  pc) is summarized with caption consisting of three values: binary fraction, half-mass radius and  $\text{conc}_{\text{pop}}$ . One can see that MOCCA-SURVEY-2 models cover surprisingly well the real GCs population in terms of the fraction of SG stars of total number of objects. The models of MOCCA-SURVEY-2 analysed in this paper can reproduce the whole spectrum of MS SG/N<sub>Tot</sub> fractions – the ones for which SG dominates and also the ones with FG more numerous. The models can reproduce basically any value from the observations (from  $\sim 0.3$  up to  $\sim 0.8$ , the point  $\sim 0.9$  concerns  $\omega$ Cen which might not be a good representative of GCs) despite the fact that they differ in  $R_h$  and  $\text{conc}_{\text{pop}}$  initial values only (we already described in Section 3.2 that initial binary fraction does not influence much the results). It is quite unexpected finding, but gives high confidence that the results of MOCCA-SURVEY-2 can be compared with the real observations. Even with our limited variety of initial conditions (e.g. limited initial mass range for models) we cover MW GCs well, and thus we can discuss all of the observational features presented in the literature.

MOCCA-SURVEY-2 models can well reproduce fractions of SG stars for the known range of observed values in the real GCs. The main

thing which seems to be varying the models from MOCCA-SURVEY-2 from the point of view of these fractions is the tidal filling of the cluster. Fig. 14 shows a clear division of models into two groups: the ones for which number of SG increases (tidally filling models), and the ones for which it decreases (tidally underfilling). In the case of MOCCA-SURVEY-2 models, this division is around  $R_h = 2.0$  pc. Among GCs which have the lowest SG stars there is e.g. M 4 (NGC 6838) with  $\sim 35$  per cent of SG with respect to all MS stars (Table 1; Milone et al. 2020b). The MOCCA models which are consistent with low SG fractions are the models which are tidally underfilling (low  $R_h$  values in Fig. 14). One can see that the initial SG fraction drops for them significantly in the beginning of cluster evolution due to fast mass-loss, which expands the cluster (due to high densities many dynamical interactions eject a lot of objects). Then the ratio drops further and settle around the values 0.3–0.4. These are the models which are initially very dense, go to a core collapse very quickly and after the collapse they expand roughly homogeneously. Moreover, SG binaries burn out very quickly because it is initially very dense, the binaries are being destroyed or ejected more than for FG. In turn, the GCs which have the highest values of SG are e.g. 47 Tuc (NGC 104) or M13 (NGC 6205) which have  $\sim 80$  per cent of SG stars (Table 1; Milone et al. 2020b). These are old, still massive and large GCs. Their SG fraction is consistent with MOCCA models which were initially tidally filling, and thus also not very dense clusters (high  $R_h$  values in Fig. 14). These are the models in which  $R_{h\text{FG}}$  is large and thus many of FG stars are close to the tidal radius. This in turn causes that the models lose preferentially FG stars, and thus leaving the cluster eventually with overpopulated SG stars. This is only a hint about the initial GC conditions. To get better constraints and comparison with MW GCs, we will need to check also other cluster properties which can be different than the ones following the present MOCCA-SURVEY-2 models.

The results of MOCCA-SURVEY-2 simulations work within the scenario in which it is assumed that SG was formed from an FG ejecta and pristine gas in the central regions of the FG cluster. In that scenario, it is expected that SG should be concentrated in the centre of the cluster and should have higher densities than FG stars. These are the models which initially have SG less numerous than FG and the  $\text{conc}_{\text{pop}}$  values are below 1.0 (SG is more concentrated than FG). A number of such models are shown as lines in Fig. 14. It is important to point out that starting with such models we are able to reproduce all cases when it comes to the fraction of SG with respect to total number of stars.

MOCCA simulations provide some interesting implications for the so-called mass-budget problem. In the literature, there are known GCs for which SG is more numerous (see Section 1). This creates a challenge for interpretation because it is hard to expect that the formation of multiple population would create more SG than FG stars. Up to now one of the most controversial implications is that GCs should be much more massive during formation (Ventura et al. 2014). Then, GC has to lose a lot of FG stars, which would contribute to host galaxy formation, and also they would provide some contribution to the reionization of the Universe (e.g. Renzini et al. 2015). We find from MOCCA-SURVEY-2 models that the main parameter which matters in that respect is how the initial model is tidally filling its potential with respect to the host galaxy and also the concentration parameter. For the models which are initially tidally filling there is basically no problem to have at 12 Gyr for SG to dominate the cluster in terms of the number of stars. It is in agreement with the results of previous study from Vesperini et al. (2021) and Sollima et al. (2022). Fig. 14 shows how numerous are MW GCs which have SG larger than FG. It seems that these are the clusters which were



**Figure 15.** Fractions of SG of the total number of stars ( $N_{SG}/N_{Tot}$ ) for Milky Way GCs (Table 2; Milone et al. 2017a, abbreviation M17) as a function of the maximum radius of the observations normalized to the half-light radius ( $R_{max}/R_{hob}$ ). Additionally, on the plot, there are a few MOCCA simulations with the fractions of SG stars as a function of the distance from the cluster centre in the units of half-light radius. Five first MOCCA simulations have common initial properties, summarized in the caption, and different only initial half-mass radii. There are two additional MOCCA simulations:  $conc_{pop} = 0.1$  with  $R_h = 4.0$  pc,  $fb = 0.95$ , and  $conc_{pop} = 0.5$  with  $R_h = 1.2$  pc,  $fb = 0.95$  given for comparison. Different  $R_h$  represent star clusters initially tidally underfilling (e.g.  $R_h = 0.6$  pc), and tidally filling (e.g.  $R_h = 6.0$  pc). The data present FG fractions only for 12 Gyr. For details, see text.

formed close to be tidally filling. In order to have stronger claims about the mass-budget problem, we need more models and we need to check if other GCs observational parameters are also recovered. However, our models give some support for other authors claiming that the mass budget can be overcome.

The potential presence of an IMBH (Baumgardt, Makino & Ebisuzaki 2004; Trenti et al. 2007) does not seem to influence the findings of this paper (for details of the IMBH formation in MOCCA simulations in MOCCA-SURVEY-1; see Giersz et al. 2015). The IMBH influences the dynamical interactions, and the core properties of star clusters, but yet it seems not to have any significant influence on the fractions of FG or SG populations. There are models in MOCCA-SURVEY-2 which initially are tidally filling, or tidally underfilling (see e.g. Fig. 8) and which produced (or not produce) a massive IMBH, but the evolution of the fraction of the population seems to be unaffected by an IMBH. Moreover, we observe in some of the models with an IMBH, the homogeneous expansion for the most dense clusters (Lagrangian radii increase at a steady pace with time). In such models, the IMBH works like an energy source. After core collapse it behaves like a set of binaries in the centre of a cluster and prevents further collapse of the core. After that, we also observe the homogeneous expansion of the cluster described in the Section 3.2. We plan to investigate IMBH influence on the multiple population more deeply in the next papers.

Fractions of SG with respect to the overall number of stars can vary significantly in different radial parts of a star cluster (e.g. Fig. 12). We have checked how the fraction of SG changes from the central parts of a cluster to the outskirts and compared a few MOCCA simulations with available observations. Fig. 15 shows the fractions of SG for Milky Way GCs (Table 2; Milone et al. 2017a). For comparison, the figure shows a set of MOCCA simulations which have the initial parameters summarized in the caption and only different initial half-mass radii. The MOCCA simulations show SG fractions only for 12 Gyr for a number of Lagrangian radii scaled by

$R_{hob}$ . Fig. 15 shows that the values from MOCCA-SURVEY-2 cover the real observations very well. MOCCA-SURVEY-2 is able in principle to reproduce the whole observed range from observations. The fraction of SG depending on the maximum distance of observations from the centre of the cluster ( $R_{max}$ ) show a very distinct evolution for different initial concentrations with respect to the tidal radius. The first group has lower changes of the fraction of SG with respect to  $R_{max}$ . Violet curve for tidally underfilling cluster ( $R_h = 0.6$  pc) have in the centre SG fraction  $\sim 0.3$ , and in the outskirts of the cluster value around  $\sim 0.25$ . This is a difference of the order of 0.05. In turn, for the clusters which started initially tidally filling the SG fraction for low  $R_{max}$  is high (SG dominates strongly), then with larger  $R_{max}$  the fraction drops more significantly. Eventually for the model with  $R_h = 6.0$  pc it drops below 0.7 ( $>0.1$  difference), and for the model with  $R_h = 4.0$  pc the difference is even larger (from 0.7 to 0.45). For tidally filling cluster the ratio drops more significantly with  $R_{max}$ . It would be very informative to compute  $N_{SG}/N_{Tot}$  values for MW GCs for a number of radii up to  $R_{max}$  and then compare them with the simulations. The fraction of SG stars, if computed for different values of  $R_{max}$ , could be an interesting tool for probing the initial conditions of the clusters.

Interestingly, MSP distribution may also help to narrow down the initial conditions for GCs. Fig. 15 shows that the models from MOCCA-SURVEY-2 with initial  $R_h = 0.6$  pc seem to have initial conditions too dense. The models with such densities do not cover any of the known MW GCs. This is very helpful e.g. in determining the initial conditions for the next set of MOCCA simulations.

The work in this paper is a natural extension of the work done in Vesperini et al. (2021) – but e.g. with the most up-to-date version of the MOCCA code and for many more models. In Vesperini et al. (2021), it was found that in order to have  $M_{SG}/M_{tot}$  values observed  $\sim 0.35$ – $0.9$  one has to lose preferentially FG stars. This occurs during the first Gyr when system expands and responds to the mass-loss of FG which is less concentrated. In our study, we connect this by how much the initial model is settled with respect to the tidal radius. Moreover, in Vesperini et al. (2021), the final values for  $M_{SG}/M_{tot}$  are between 0.53 and 0.8. In this work, we are also able to reproduce lower values – characteristic for underfilling clusters.

The results are also consistent with Sollima (2021), who find that SG observed today depends uniquely on the ratio between the initial cluster mass and the half-mass radius of SG. The efficiency of SG being destroyed decreases with time after the cluster initial mass-loss. Thus the SG fraction observed today is strongly sensitive to the initial size of the SG. They use this relation to constrain the initial conditions of GCs and concluded (with a few assumptions) that SG must have formed with a typical half-mass radius smaller than 0.5–1 pc. In turn, FG is more extended and more exposed to the tidal field during the cluster evolution. They find that the final FG binary fraction depends on the strength of the tidal field which determines the variation of cluster mass and size. We find the same conclusions but in our models we see it through how much the models are filling or underfilling – this determines the final fractions of SG as we already described in previous paragraphs.

## 5 CONCLUSIONS

The main purpose of the paper was to test mixing in dense GCs for various initial conditions for two populations of stars (FG and SG) for real size GCs – the goal difficult to obtain currently with direct  $N$ -body codes. We work within scenario by Calura et al. (2019), in which SG forms in the very cluster centre, after some time delay (usually a few dozens of Myr), from gas lost due to stellar winds of AGB stars

and pristine gas reaccrated by GC during its movement through gas cloud left after formation (some small deviations are described in Section 2.2). The main conclusion of the paper concerns how initially tidally filling or underfilling clusters influence the relative number of FG to SG stars. The results of this study are the following:

(i) We find that the most important factor affecting the relative fraction of FG or SG is the initial tidal filling of a cluster. In order to have FG more numerous than SG, a star cluster has to be tidally underfilling since the cluster can expand without losing FG stars during the initial expansion phase. On the other hand, tidally filling clusters tend to have more SG stars than FG due to the preferential loss of FG stars through the tidal radius in accordance with their initial position. Binaries also have the same trend, tidally filling clusters have more SG binaries and underfilling clusters have more FG binaries (Fig. 6). But the differential evolution of binaries between FG and SG is mostly driven by the dynamical interactions at the central regions. The binaries from SG are much more efficiently destroyed or ejected due to the dynamical interactions at the dense central regions, especially in an underfilling cluster (see Section 3.2). In turn, to have a more numerous SG binaries, a star cluster has to be initially tidally filling system whose central density is not very high to destroy or kick SG binaries efficiently but FG binaries can escape from the cluster easily.

(ii) Even though the MOCCA-SURVEY-2 models presented in this paper were not designed to mimic Milky Way GCs, their cover observational parameters very well (see Section 2.2). Our models can reproduce the observed range of fractions of SG stars with respect to the total number of stars for known values of MW GCs (see Fig. 14). We find that the MW GCs which have the lowest values of SG to total number of stars (e.g. NGC 6496 with 33 per cent of SG; see Table 2; Milone et al. 2017b) are consistent with our MOCCA models of tidally underfilling cases ( $R_h = 0.6$  pc in Fig. 14). In turn, MW GCs with the highest SG fractions (e.g. 47 Tuc, M13 which has 80 per cent of SG stars) are consistent with MOCCA models initially tidally filling. They are old, still massive and large GCs. For these models,  $R_{hFG}$  are large and thus many of FG stars are closer to the tidal radius. Thus, these models lose preferentially FG stars, and as a result, SG dominates at the present day, which is consistent with the conclusions of Vesperini et al. (2021). We note also, that MOCCA-SURVEY-2 have a minimum initial number of SG/FG of 0.5, and it is not clear whether very large final fractions of SG/FG could be reproduced assuming smaller initial fractions (Bastian & Lardo 2015). This has also an implication to so-called mass-budget problem. MW GCs with more numerous SG create a challenge for the interpretation because it is hard to explain why SG can be more abundant. Some scenarios (Ventura et al. 2014; Renzini et al. 2015) tried to interpret the mass-budget problem, but there has been no definitive proof yet. Our MOCCA simulation results can shed a light on solving the mass-budget problem by reproducing any fraction of SG value of the real MW GCs without any unrealistic assumptions or postulates. However, we need to run more models, and check if other parameters of our models are in good agreement with MW GCs properties. We are aware that we get good agreement with respect to one parameter but we will have to check also other ones in order to make more conclusive statements.

(iii) we examined the fractions of SG with respect to the overall number of stars (Fig. 15) as a function of maximum radial distance ( $R_{max}$ ) for which they were measured (Table 2; Milone et al. 2017a). MOCCA models can reproduce the whole range of values from observations, but we find that  $R_{max}$  can vary significantly within one GC if we take  $R_{max}$  to only a fraction of  $R_h$  or a few half-mass radii. For example, tidally underfilling models the fraction  $N_{SG}/N_{Tot}$

varies from  $\sim 0.7$  in the very centre to about  $\sim 0.45$  in the outskirts (see details in Section 4).

(iv) We find also that the ratio between binary count from FG to SG which takes into account only MS stars is a very good proxy for the entire cluster. It provides the same conclusions for mixing as those coming from all binaries, or only observational ones (see Section 3.2). The same behaviour of the ratio between two populations let us believe that even by observing a small fraction of visible MS binaries one can actually probe the whole star cluster (i.e. all binaries) at least from the point of view of mixing populations.

(v) The conclusions of this paper are consistent between the models with the initial binary fraction 10 per cent and 95 per cent. It is known, that for the latter case many binaries are weak from the point of view of the mean binding energy of binaries and many of them are quickly destroyed. But still, the number of primordial binaries has a profound effect on the star cluster evolution. It is interesting to see that the initial binary fraction does not affect much the ratio between number of binaries from two populations.

(vi) Preliminary examination indicates that a presence of IMBH in some models does not affect the conclusions of this work. The IMBHs, even with masses  $> 1000.0 M_\odot$ , do not affect significantly the ratio between binary fractions of two populations. From this point of view, any try to infer constrains on the initial star clusters concentrations should be safe too.

In this paper, we have concentrated on the number of FG and SG stars through the prism of tidally filling or underfilling clusters. There are more parameters which differ simulations within MOCCA-SURVEY-2 simulations. We plan to explore their influence on the FG and SG populations in follow-up papers (e.g. IMBH and BH subsystem influence, time delay between FG and SG formation). We did not check e.g. whether different  $M_{max}$  values (Table 1) have any influence on the FG, SG populations, or their relative number.

## SOFTWARE

MOCCA code is open source<sup>8</sup> for our collaborators. We are open to start new projects, in which one could use already existing MOCCA simulations, or start new numerical simulations.

BEANS<sup>9</sup> software is open source and it is freely available for anyone.

## ACKNOWLEDGEMENTS

We would like to thank Enrico Vesperini for very insightful comments and suggestions that helped to significantly improve this paper. We thank also the referee for all the comments and suggestions, which improved the paper even more.

This research has been partly financed by the Polish National Science Centre grants: 2016/23/B/ST9/02732, 2018/30/A/ST9/00050, 2019/32/C/ST9/00577. AA acknowledges support from the Swedish Research Council through the grant 2017-04217.

## DATA AVAILABILITY

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

<sup>8</sup><https://moccacode.net/license/>

<sup>9</sup><https://beanscode.net/download/>

## REFERENCES

- Askar A., Szkudlarek M., Gondek-Rosińska D., Giersz M., Bulik T., 2017, *MNRAS*, 464, L36
- Bastian N., Lardo C., 2015, *MNRAS*, 453, 357
- Bastian N., Lardo C., 2018, *ARA&A*, 56, 83
- Bastian N., Lamers H. J. G. L. M., de Mink S. E., Longmore S. N., Goodwin S. P., Gieles M., 2013, *MNRAS*, 436, 2398
- Baumgardt H., Hilker M., 2018, *MNRAS*, 478, 1520
- Baumgardt H., Vasiliev E., 2021, *MNRAS*, 505, 5957
- Baumgardt H., Makino J., Ebisuzaki T., 2004, *ApJ*, 613, 1143
- Baumgardt H., De Marchi G., Kroupa P., 2008, *ApJ*, 685, 247
- Bekki K., 2010, *ApJ*, 724, L99
- Bekki K., 2011, *MNRAS*, 412, 2241
- Belczynski K., Kalogera V., Bulik T., 2002, *ApJ*, 572, 407
- Belczynski K., Kalogera V., Rasio F. A., Taam R. E., Zezas A., Bulik T., Maccarone T. J., Ivanova N., 2008, *ApJS*, 174, 223
- Belczynski K. et al., 2016, *A&A*, 594, A97
- Bellini A. et al., 2015, *ApJ*, 810, L13
- Belloni D., Giersz M., Rocha-Pinto H. J., Leigh N. W. C., Askar A., 2017a, *MNRAS*, 464, 4077
- Belloni D., Askar A., Giersz M., Kroupa P., Rocha-Pinto H. J., 2017b, *MNRAS*, 471, 2812
- Calura F., D’Ercole A., Vesperini E., Vanzella E., Sollima A., 2019, *MNRAS*, 489, 3269
- Carretta E. et al., 2009, *A&A*, 505, 117
- Carretta E., Bragaglia A., Gratton R. G., Recio-Blanco A., Lucatello S., D’Orazi V., Cassisi S., 2010, *A&A*, 516, A55
- Cezario E., Coelho P. R. T., Alves-Brito A., Forbes D. A., Brodie J. P., 2013, *A&A*, 549, A60
- Cordero M. J., Pilachowski C. A., Johnson C. I., McDonald I., Zijlstra A. A., Simmerer J., 2014, *ApJ*, 780, 94
- Cordero M. J., Hénault-Brunet V., Pilachowski C. A., Balbinot E., Johnson C. I., Varri A. L., 2017, *MNRAS*, 465, 3515
- Cordoni G., Milone A. P., Mastrobuono-Battisti A., Marino A. F., Lagioia E. P., Tailo M., Baumgardt H., Hilker M., 2020, *ApJ*, 889, 18
- D’Ercole A., Vesperini E., D’Antona F., McMillan S. L. W., Recchi S., 2008, *MNRAS*, 391, 825
- Dalessandro E. et al., 2014, *ApJ*, 791, L4
- Dalessandro E. et al., 2018, *ApJ*, 864, 33
- Dalessandro E., Raso S., Kamann S., Bellazzini M., Vesperini E., Bellini A., Beccari G., 2021, *MNRAS*, 506, 813
- de Mink S. E., Pols O. R., Langer N., Izzard R. G., 2009, *A&A*, 507, L1
- Decressin T., Meynet G., Charbonnel C., Prantzos N., Ekström S., 2007, *A&A*, 464, 1029
- Denissenkov P. A., Hartwick F. D. A., 2014, *MNRAS*, 437, L21
- Elmegreen B. G., 2017, *ApJ*, 836, 80
- Fregeau J. M., Rasio F. A., 2007, *ApJ*, 658, 1047
- Fregeau J. M., Cheung P., Portegies Zwart S. F., Rasio F. A., 2004, *MNRAS*, 352, 1
- Fryer C. L., Belczynski K., Wiktorowicz G., Dominik M., Kalogera V., Holz D. E., 2012, *ApJ*, 749, 91
- Gieles M. et al., 2018, *MNRAS*, 478, 2461
- Giersz M., 1998, *MNRAS*, 298, 1239
- Giersz M., Heggie D. C., 2011, *MNRAS*, 410, 2698
- Giersz M., Heggie D. C., Hurley J. R., 2008, *MNRAS*, 388, 429
- Giersz M., Heggie D. C., Hurley J. R., Hypki A., 2013, *MNRAS*, 431, 2184
- Giersz M., Leigh N., Marks M., Hypki A., Askar A., 2014, preprint ([arXiv:1411.7603](https://arxiv.org/abs/1411.7603))
- Giersz M., Leigh N., Hypki A., Lützgendorf N., Askar A., 2015, *MNRAS*, 454, 3150
- Goodwin S. P., Whitworth A. P., 2004, *A&A*, 413, 929
- Gratton R. G., 2004, *Mem. Soc. Astron. Ital.*, 75, 274
- Gratton R. G., Carretta E., Bragaglia A., 2012, *A&AR*, 20, 50
- Gratton R., Bragaglia A., Carretta E., D’Orazi V., Lucatello S., Sollima A., 2019, *A&AR*, 27, 8
- Hénault-Brunet V., Gieles M., Agertz O., Read J. I., 2015, *MNRAS*, 450, 1164
- Hong J., Vesperini E., Sollima A., McMillan S. L. W., D’Antona F., D’Ercole A., 2015, *MNRAS*, 449, 629
- Hong J., Vesperini E., Sollima A., McMillan S. L. W., D’Antona F., D’Ercole A., 2016, *MNRAS*, 457, 4507
- Hong J., Patel S., Vesperini E., Webb J. J., Dalessandro E., 2019, *MNRAS*, 483, 2592
- Hong J., Askar A., Giersz M., Hypki A., Yoon S.-J., 2020, *MNRAS*, 498, 4287
- Howard C. S., Pudritz R. E., Sills A., Harris W. E., 2019, *MNRAS*, 486, 1146
- 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
- Hypki A., 2018, *MNRAS*, 477, 3076
- Hypki A., Giersz M., 2013, *MNRAS*, 429, 1221
- Hypki A., Giersz M., 2017, *MNRAS*, 466, 320
- Kamlah A. W. H. et al., 2022, *MNRAS*, 511, 4060
- Khalaj P., Baumgardt H., 2015, *MNRAS*, 452, 924
- Kiminki D. C., Kobulnicky H. A., 2012, *ApJ*, 751, 4
- King I. R., 1966, *AJ*, 71, 64
- Kobulnicky H. A. et al., 2014, *ApJS*, 213, 34
- Krause M., Charbonnel C., Decressin T., Meynet G., Prantzos N., 2013, *A&A*, 552, A121
- Kroupa P., 1995, *MNRAS*, 277, 1491
- Kroupa P., 2001, *MNRAS*, 322, 231
- Kroupa P., 2011, in Alves J., Elmegreen B. G., Girart J. M., Trimble V., eds, *Computational Star Formation, Proceedings of the International Astronomical Union, IAU Symposium, Volume 270*. p. 141
- Kroupa P., Weidner C., Pflamm-Altenburg J., Thies I., Dabringhausen J., Marks M., Maschberger T., 2013, in Oswalt T. D., Gilmore G., eds, *Stellar Systems and Galactic Structure, Vol.5*, Springer, The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations. Springer Science+Business Media, Dordrecht, p. 115
- Küpper A. H. W., Maschberger T., Kroupa P., Baumgardt H., 2011, *MNRAS*, 417, 2300
- Lee Y. W., Joo J. M., Sohn Y. J., Rey S. C., Lee H. C., Walker A. R., 1999, *Nature*, 402, 55
- Lee J.-W., 2017, *ApJ*, 844, 77
- Leveque A., Giersz M., Paolillo M., 2021, *MNRAS*, 501, 5212
- Li C., Wang Y., Tang B., Milone A. P., Yang Y., Ji X., 2020, *ApJ*, 893, 17
- Lucatello S., Sollima A., Gratton R., Vesperini E., D’Orazi V., Carretta E., Bragaglia A., 2015, *A&A*, 584, A52 (L15)
- Mackey A. D., Broby Nielsen P., Ferguson A. M. N., Richardson J. C., 2008, *ApJ*, 681, L17
- McKenzie M., Bekki K., 2021, *MNRAS*, 507, 834
- Maliszewski K., Giersz M., Gondek-Rosińska D., Askar A., Hypki A., 2022, *MNRAS*, 514, 5879
- Marks M., Kroupa P., Dabringhausen J., 2022, *Astronomy & Astrophysics*, 659, A96
- Maschberger T., Clarke C., 2012, in Capuzzo-Dolcetta R., Limongi M., Tornambè A., eds, *ASP Conf. Ser. Vol. 453, Advances in Computational Astrophysics: Methods, Tools, and Outcome*. Astron. Soc. Pac., San Francisco, p. 367
- Mastrobuono-Battisti A., Perets H. B., 2021, *MNRAS*, 505, 2548
- Miholics M., Webb J. J., Sills A., 2015, *MNRAS*, 454, 2166
- Milone A. P., Bedin L. R., Piotto G., Anderson J., 2009, *A&A*, 497, 755
- Milone A. P. et al., 2012, *A&A*, 540, A16
- Milone A. P. et al., 2015a, *MNRAS*, 447, 927
- Milone A. P. et al., 2015b, *MNRAS*, 450, 3750
- Milone A. P. et al., 2015c, *ApJ*, 808, 51
- Milone A. P. et al., 2017a, *MNRAS*, 464, 3636
- Milone A. P. et al., 2017b, *MNRAS*, 469, 800
- Milone A. P., Marino A. F., Mastrobuono-Battisti A., Lagioia E. P., 2018, *MNRAS*, 479, 5005
- Milone A. P. et al., 2020a, *MNRAS*, 491, 515
- Milone A. P. et al., 2020b, *MNRAS*, 492, 5457
- Nardiello D., Milone A. P., Piotto G., Marino A. F., Bellini A., Cassisi S., 2015, *A&A*, 573, A70
- Piotto G. et al., 2015, *AJ*, 149, 91

- Renzini A. et al., 2015, *MNRAS*, 454, 4197
- Richer H. B., Heyl J., Anderson J., Kalirai J. S., Shara M. M., Dotter A., Fahlman G. G., Rich R. M., 2013, *ApJ*, 771, L15
- Sana H. et al., 2012, *Science*, 337, 444
- Simioni M., Milone A. P., Bedin L. R., Aparicio A., Piotto G., Vesperini E., Hong J., 2016, *MNRAS*, 463, 449
- Sollima A., 2021, *MNRAS*, 502, 1974
- Sollima A., Ferraro F. R., Bellazzini M., Origlia L., Straniero O., Pancino E., 2007, *ApJ*, 654, 915
- Sollima A., Gratton R., Lucatello S., Carretta E., 2022, *MNRAS*, 512, 776
- Tiongco M. A., Vesperini E., Varri A. L., 2019, *MNRAS*, 487, 5535
- Trani A. A., Fujii M. S., Spera M., 2019, *ApJ*, 875, 42
- Trenti M., Ardi E., Mineshige S., Hut P., 2007, *MNRAS*, 374, 857
- Vasiliev E., Baumgardt H., 2021, *MNRAS*, 505, 5978
- Ventura P., D'Antona F., Mazzitelli I., Gratton R., 2001, *ApJ*, 550, L65
- Ventura P., di Criscienzo M., D'Antona F., Vesperini E., Tailo M., Dell'Agli F., D'Ercole A., 2014, *MNRAS*, 437, 3274
- Vesperini E., McMillan S. L. W., D'Antona F., D'Ercole A., 2011, *MNRAS*, 416, 355
- Vesperini E., McMillan S. L. W., D'Antona F., D'Ercole A., 2013, *MNRAS*, 429, 1913
- Vesperini E., Hong J., Webb J. J., D'Antona F., D'Ercole A., 2018, *MNRAS*, 476, 2731
- Vesperini E., Hong J., Giersz M., Hypki A., 2021, *MNRAS*, 502, 4290
- Villanova S., Geisler D., Carraro G., Moni Bidin C., Muñoz C., 2013, *ApJ*, 778, 186
- Wang L. et al., 2016, *MNRAS*, 458, 1450
- Wang L., Kroupa P., Takahashi K., Jerabkova T., 2020, *MNRAS*, 491, 440
- Zennaro M., Milone A. P., Marino A. F., Cordoni G., Lagioia E. P., Tailo M., 2019, *MNRAS*, 487, 3239

## APPENDIX A: OUTPUT FILES

The output files produced for every MOCCA simulation consist of around 20 files. They were designed to store global properties of star clusters, profiles, snapshots, and various events which occur for and between stars (e.g. stellar evolution, dynamical interactions, merger). The density of the stored information depends on the type of the information. Global star cluster parameters are computed for every time-step, snapshots are written by default every 50 Myr and stellar events (e.g. mass-loss, dynamical interactions, natal kicks properties) are all written to the output files. The overall density for the output was designed to be as compact as possible but providing as much information as possible for the versatile needs of any user.

An additional explanation is needed on how different interactions are treated for some multiple populations. For mass transfers, we consider marking a star as a mixed population if the star gained some mass larger than  $1.0 e^{-7} M_{\odot}$  (an arbitrary chosen value). The star which lost the mass keeps its population id. When it comes to dynamical collisions, if two stars are coming from two different populations, we always mark the star as a mixed population.

A mixed population id for a star is designed in such a way that it holds compact and partial information about the history of mixing between different populations. In other words, the mixed population id is larger for a star for which the mixing (mass transfers or dynamical collisions) happened more times. The mixed population id is a concatenation of two populations ids of interacting stars and it is done according to the following rules:

(i)  $\text{pop1}+\text{pop2} \rightarrow (01)(01) \rightarrow$  gives the id 101 (first two digits in the right brackets tell us how many merger stars come from population 1 and two next digits from the next brackets – how many

stars come from the population 2). In this case, there is one star from population 1 and one from population 2.

(ii)  $\text{pop2}+101 \rightarrow (01+01)(01) \rightarrow$  number 201 – in this case there are in total two stars (or mass transfers) from population 2 and one star from population 1.

(iii)  $\text{pop1}+201 \rightarrow (02)(01+01) \rightarrow$  number 202 – in this case there are in total two stars (or mass transfers) from population 2 and two stars from population 1.

(iv) further population ids are created accordingly.

The following scheme is shown for simulation consisting of two populations. For the simulations started with more simulations there is only a larger number of significant digits [e.g.  $(02)(01)(03) \rightarrow$  two stars or mass transfers from population 3, one from population 2, and three from population 1]. This mixed population id is designed only to help to see how complex the mixing for a particular star was.

It is worth to mention that this mixed population id is only a compact information which provides a glimpse into a history of a given star – the higher the number, the more events the star had with different populations. However, if needed, with MOCCA output files one can recover the full history for any given star in the system (all stellar evolutionary events, all dynamical interactions, exchanges, disruptions, collisions, etc.).

The main output file which gives information about global properties of a star cluster is called *system*. The file contains over 500 columns with global parameters of the star cluster computed on every time-step (e.g. total mass, total potential, and kinetic energies, number of MSs, WDs, BHs, and many more). The next important files concern events which can occur for any given binary or a single star (i.e. stellar evolution step, binary-single, binary-binary interactions, binary formation, physical collision between two single stars). All such events are carefully saved with all possible properties (stars' parameters before and after the interactions) which might be useful later: masses, radii, luminosities, semimajor axes, eccentricities, impact parameters, velocities, positions in the cluster and more. The next MOCCA files contain information about binary and single stars which escaped from the cluster, parameters during any kind of kick which a star could get (e.g. supernova kick). There are Lagrangian radii computed for all populations for every time-steps, a number of different spatial profiles (e.g. surface brightness profile), and finally snapshot files which contain a number of parameters for every star and binary in a system.

The list of MOCCA output files changes between versions. The complete and always up-to-date list of files, columns and their descriptions one can find on the MOCCA web page.<sup>10</sup>

## APPENDIX B: BEANS SCRIPT

In Appendix B, we present one of the scripts which was used to analyse MOCCA simulations within BEANS software for the purpose of this paper. The goal of describing this script here is to show how one can perform a complex data analysis for large MOCCA data sets with relatively easy to understand Apache Pig scripts. The following script produces data with number of stars or binaries from different populations measured in different ways (e.g. all binaries from the populations, only observational binaries).

<sup>10</sup><https://moccacode.net/output/>

The script is written in a language Apache Pig,<sup>11</sup> which is high level language for Apache Hadoop<sup>12</sup> platform used for distributed data analysis of Big Data. The script is described here line by line. Every line of interest is denoted with a comment line and a number (lines starting with characters - -). In the script there are places with characters [...] which only purpose is to shorten the script and to make it easier to read. These places consist of lines which are similar to the lines which are around characters [...] (e.g. list of columns are shorten with [...]). The keywords of Apache Pig like *load*, *using*, are case insensitive.

```
-- <1>
snap = LOAD ' DATASETS=' 'MOCCA' '
TABLES=' 'snapshot' '
    FILTER=' '(timenr == 0)
    OR (tphys>400.0 tphys<430.0)
    [...]
    OR (tphys>12000.0 tphys<12030.0) ' ' '
    USING BeansTable ();

-- <2>
sys = LOAD ' DATASETS=' 'MOCCA' '
TABLES=' 'system' ' '
    USING BeansTable ();

-- <3>
aux = LOAD ' DATASETS=' 'Auxiliary MOCCA tables' '
TABLES=' 'Additional lagrangian radii
(survey2, projection,
selected snapshots)' ' '
    USING BeansTable ();

-- <4>
aux = FOREACH aux GENERATE *,
    DSID(tbid1) as dsid;

-- <5>
snap = FOREACH snap GENERATE tbid as tbidSnap,
    DSID(tbid) as dsidSnap, timenr,
    tphys, r, ik1, ik2, sm1, sm2,
    popId1, popId2, idd1;

-- <6>
sys = FOREACH sys GENERATE tbid as tbidSys,
    DSID(tbid) as dsidSys,
    timenr, r.h, rhob, rtid, sturn, sturnm,
    pop1oc, pop2oc, pop1b, pop2b;

-- <7>
sys0 = FILTER sys BY timenr == 0;
-- <8>
sys = JOIN sys BY dsidSys, sys0 BY dsidSys;
-- <9>
sys = FOREACH sys GENERATE
    sys::tbidSys as tbidSys,
    sys::dsidSys as dsidSys,
    sys::timenr as timenr,
    sys::r.h as r.h,
```

```
[...]
sys0::pop1b as pop1b,
sys0::pop2b as pop2b;

-- <10>
snapSys = JOIN snap BY (dsidSnap, timenr),
    sys BY (dsidSys, timenr);

-- <11>
snap = FOREACH snapSys GENERATE
    snap::tbidSnap as tbidSnap,
    snap::dsidSnap as dsid,
    snap::timenr as timenr,
    snap::tphys as tphys,
    snap::r as r,
    snap::idd1 as idd1,
    (snap::popId1 == 1 ? 1 : 0) AS pop11,
    (snap::popId1 == 1
    AND snap::ik1 <= 1 ? 1 : 0) AS pop11ms,
    [...]
    sys::pop1b as pop1b0,
    sys::pop2b as pop2b0;

-- <12>
snapJoined = JOIN snap BY (dsid, timenr),
    aux BY (dsid, timenr);

-- <13>
snap = FOREACH snapJoined GENERATE
    snap::tbidSnap AS tbidSnap,
    snap::timenr AS timenr,
    snap::tphys AS tphys,
    snap::idd1 AS idd1,
    [...]
    snap::pop1b0 as pop1b0,
    snap::pop2b0 as pop2b0,
    (CASE
    WHEN snap::r < aux::lagrLumR1 THEN 1
    WHEN snap::r < aux::lagrLumR10 THEN 10
    [...]
    WHEN snap::r < aux::lagrLumR90 THEN 90
    ELSE 100 END) AS lumLagr,
    (CASE
    WHEN snap::r < aux::lagrLumR1 THEN
aux::lagrLumR1
    WHEN snap::r < aux::lagrLumR10 THEN
aux::lagrLumR10
    [...]
    WHEN snap::r < aux::lagrLumR90 THEN
aux::lagrLumR90
    ELSE 10000.0 END) AS lumLagrR;

-- <14>
snapGr = GROUP snap BY (tbidSnap, timenr, lum-
Lagr);

-- <15>
snapFl = FOREACH snapGr GENERATE
    SUM(snap.pop11) + SUM(snap.pop21) as pop1c,
    SUM(snap.pop11ms) + SUM(snap.pop21ms) as
pop1cms,
    SUM(snap.pop12) + SUM(snap.pop22) as pop2c,
```

<sup>11</sup><https://pig.apache.org>

<sup>12</sup><https://hadoop.apache.org>

```

SUM(snap.pop12ms) + SUM(snap.pop22ms) as
pop2cms,
[...]
SUM(snap.sinpop2ms) as sinpop2cms,
MIN(snap.idd1) as minId,
FLATTEN(snap);

-- <16>
snapF1 = FILTER snapF1 BY snap::idd1 == minId;

-- <17>
snapF1 = FOREACH snapF1 GENERATE pop1c, pop1cms,
pop2c,
[...]
snap::tbidSnap as tbidSnap,
DSID(snap::tbidSnap) as dsid,
snap::timenr as timenr,
FLOOR(snap::tphys, 100.0) as tphys,
snap::idd1 as idd1,
(snap::lumLagrR < 9999.0 ?
snap::lumLagrR : snap::rtid) as r,
[...]
snap::pop1b0 as pop1b0,
snap::pop2b0 as pop2b0;

-- <18>
STORE snapF1 INTO 'NAME=''set01 obs projec-
tion''
USING BeansTable();

```

In principle, the commands <1> - <6> read data from different tables and store them in so-called *relations* for later use.

The first command (<1>) reads data from BEANS software from all data sets with the name *mocca*, from all tables with the name *snapshot*. The snapshot tables in MOCCA simulations contain the list of all stars and binaries in the system for a number of time-steps (every 50 Myr). This instruction uses search engine and looks for every table which fulfils these query parameters. Because MOCCA-SURVEY-2 contains around 250 MOCCA simulations, this command will read this number of snapshot files. The term *USING BeansTable()* means that the script should use custom reader called *BeansTable*. This is a custom reader, part of BEANS software, which can read multiple tables in parallel and speed up the entire script significantly. The data read from all MOCCA snapshot files will be ‘stored’ under the relation *snap* (equivalent to a variable in languages like C). Later, this relation will be used to perform further operations like filtering or grouping. Apache Pig internally uses a series of low-level MapReduce jobs, which means that the data will be processed as a stream of data, in parallel. There is no need for Apache Pig to read entire tables to memory. In the command <1> there is also a *FILTER* subcommand. This is used internally by MOCCA-BEANS plugin and allows to speed reading data by retrieving from MOCCA output only the rows which fulfil this *FILTER* query.

The second command (<2>) uses also the *BeansTable* reader, but in this case the data are read from *system* tables from MOCCA simulations. The *system* tables in MOCCA simulations contain several hundreds of global parameters of star clusters like total mass, core radius, half-mass radius, total number of binary-single, or binary-binary interactions, total mass of WDs and many more. A few columns (e.g. turn-off mass) from these tables will be needed later to compute e.g. the observational number of binaries. Under the relation name *sys* there are stored all rows from all *system* tables.

The third command (<3>) reads one table (identified by a query *Additional lagrangian radii...*) from a different notebook identified by a query *Auxiliary MOCCA tables*. This is the way how one can merge within one BEANS script tables coming from different notebook and compute some new values. The notebook *Auxiliary MOCCA tables* contain some auxiliary tables computed based on MOCCA simulations which might be useful for any projects. One of the auxiliary tables are Lagrangian radii computed based on projected positions and luminosities of stars – this is a table which is being read in the command <3>. This auxiliary table is stored in the relation *aux*.

The next command (<4>) goes over every line from relation *aux*, rewrites all columns (statement \*), and adds additional column (*dsid*). The column is added using a function *DSID*. Every table within BEANS is described by a unique ID, called hereafter *TableId*. The subcommand *DSID(tbid1)* takes a column *tbid1* from relation *aux*, returns what is the ID of a data set for this table and saves this value under a column name *dsid*. This *dsid* column will be needed later to join different tables coming from different MOCCA simulations together.

The next command (<5>) goes over every row from relation *snap* and extracts only a few columns from it (*TableId* and saved it as *tbidSnap*, time number *timenr*, physical time *tphys*, position in the cluster *r*, stellar types of stars *ik1*, *ik2*, masses *sm1*, *sm2*, population ids *pop1d1*, *pop1d2* and id of the first star *idd1*). There is also new column added here, just like in the previous command, *dsidSnap*. This is ID of a data set for every snapshot line (one MOCCA simulation is stored in one data set, so their IDs differ). This column will be used later to join the rows in relation *snap* with the rows of other relations.

The next command (<6>) is very similar to the row <5> but in this command some number of columns are extracted from *system* tables (e.g. half-mass radius *r\_h*, observational half-mass radius *rhob*, tidal radius *rtid*, turn-off mass *sturn*, total number of objects from FG *pop1oc*, SG *pop2oc* and total number of binaries from FG *pop1b*, and SG *pop2b*).

The commands <7> - <8> append to every line from MOCCA *system* tables (relation *sys*) a few columns from the same tables but from physical time 0 Myr.

The next command (<7>) read all rows from relation *sys* (table *system* from MOCCA simulations) and leaves only these lines which come from the physical time 0 Myr (*timenr* == 0, *timenr* is an integer number which enumerates time-steps in MOCCA simulations). The rows, coming from time 0 Myr, are stored separately in relation *sys0*.

The next command (<8>) uses *JOIN* statement to join rows from two relations *sys* and *sys0*. Join operator works in such a way that for all rows from the first relation (*sys*) it looks for a corresponding row from the second relation (*sys0*). In our case, the rows are matched using the column *dsidSys* for both relations. This is the column which stored IDs of data sets in which particular tables belong to. In other words, with the command <8> we add to all rows from *system* tables (*sys*) a few values from the same table but from time  $T = 0$  Myr (relation *sys0*).

The command <9> only rewrites the columns names to a simpler names – Apache Pig during the last join command added prefixed like *sys::* and *sys0::* because there were the same columns in both relations.

In the command <10> there is also a join operator applied, but this time for *snapshot* rows (relations *snap*) and rows with global parameters of star clusters (relation *sys*). The purpose of this join is also to attach to snapshot rows some values of star clusters. The output of this join is stored in the relation *snapSys*. The command <11> takes as an input relation from the previous command (*FOREACH*

*snapSys*) and for every row from that relation it computes some additional values. The purpose of the entire script is to compute the number of stars, binaries for Lagrangian radii taking into account e.g. different definitions of binaries. Thus, in the command `<11>` one can see different statements computing some values. For example, the expression `(snap::pop1d1 == 1 AND snap::ik1 <= 1? 1 : 0) AS pop11ms` checks if a first star in the *snapshot* row belongs to FG and is MS star (ik1 is stellar type in SSE/BSE code). If a star is indeed MS star from FG the number 1 is returned. There is a number of similar, quite self-explanatory expressions, like this in the command `<11>`. All of these values will be needed later to compute occurrences of such stars or binaries at different Lagrangian radii. Please, notice that at the end of command `<11>` there are rewritten some values from relation *sys*, like half-mass radius ( $R_h$ ), observational half-mass radius (*rhob*). These values were attached to *snapshot* rows in the previous command using join operator.

The result of the long command `<11>` is stored in the relation *snapJoined* and it is used in command `<12>` for another join operator. This time we join the *snapshot* rows with Auxiliary MOCCA table which holds the Lagrangian radii for the entire system (see command `<3>`). Here, the interesting part is that the join is performed for two relations *snap* and *aux* but with two columns: first by *dsid* and then by *timenr*. This, join has to be done over two columns because first one has to divide all rows by the data set ID, and then all rows from one data set have to be divided into separate time-steps.

The next command `13` only rewrites some columns by simplifying the names from e.g. `snap::pop11` to `pop11`. There are only two additional statements in this command at the end which checks for every object from *snapshot* rows to which Lagrangian radii the star or binary belongs (statement `CASE...WHEN...ELSE`). The statement saves the Lagrangian radii as an integer number (1, 10, 20...) and as a real position in star clusters.

The commands `<14>` - `18` in principle perform the last task of the script – they calculate how many objects from different populations are in different Lagrangian radii.

The command `<14>` performs a group operation. This time the grouping is done over three columns: *snapshot* table ID (*tbidSnap*), time-step (*timenr*), and at last over the Lagrangian radius computed

by the luminosity (*lumLagr*). In this way, all rows from *snapshot* files coming from all MOCCA simulations are divided into smaller groups. In command `<15>` every group is aggregated. For every Lagrangian radius, it is computed e.g. how many individual stars belong to FG ( $SUM(snap.pop11) + SUM(snap.pop21)$ ), how many MS individual stars belong to FG ( $SUM(snap.pop11ms) + SUM(snap.pop21ms)$  as *pop1cms*), the same for SG, then how many binaries, observational binaries, observational MS binaries belong to FG and SG, etc.

The command `<16>` is used to filter many *snapshot* lines produced by the previous rows and leave only one row per one data set, per one time-step, and per one Lagrangian radii. The command `<17>` only simplifies the columns names, computes a few new columns (data set ID as *dsid*, simplifies physical time to multiplications of the value 100.0) and replaces the last Lagrangian radius by the tidal radius (`snap::lumLagrR < 9999.0? snap::lumLagrR : snap::rtid`).

This example BEANS script might be a little overwhelming at first, especially for the readers not familiar with such way of data analysis, but Apache Pig is considered as a scripting language with a steep learning curve. After spending some time on writing scripts in Apache Pig, everything is much easier to understand, and the possibilities of a smooth and fast data analysis of huge data sets reward every time spend on learning this tool. The description was kept short and if a reader wishes to find more details about Apache Pig language, one is encouraged to read official documentations<sup>13</sup> or tutorials on the BEANS web page.<sup>14</sup> BEANS is the software written in a general form, and can be used to analyse any tabular data sets. For flat, plan text files it works out-of-the box, for more complex data sets, it might be easier to write a plugin for BEANS. The most important note is that BEANS can work as a central repository for a huge number of data sets coming from any sources with the ability to analyse them together in one script.

This paper has been typeset from a  $\text{\TeX/L\AA\TeX}$  file prepared by the author.

<sup>13</sup><http://pig.apache.org/docs/r0.17.0/>

<sup>14</sup><https://beanscode.net/tutorials/>