

Modelling globular clusters in the TNG50 simulation: predictions from dwarfs to giant galaxies

Jessica E. Doppel¹   Laura V. Sales¹  Dylan Nelson²  Annalisa Pillepich³  Mario G. Abadi^{4,5}  Eric W. Peng^{6,7} Federico Marinacci⁸  Jill Naiman⁹  Paul Torrey¹⁰  Mark Vogelsberger¹¹  Rainer Weinberger¹²  and Lars Hernquist¹³

¹ University of California, Riverside, 900 University Ave, Riverside, CA 92521, USA

² Institut für theoretische Astrophysik, Zentrum für Astronomie, Universität Heidelberg, Baden-Württemberg, D-69120 Heidelberg, Germany

³ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

⁴ Instituto de Astronomía Teórica y Experimental, CONICET-UNC, Laprida 854, X5000BGR, Córdoba, Argentina

⁵ Observatorio Astronómico, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina

⁶ Department of Astronomy, Peking University, Beijing 100871, China

⁷ Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China

⁸ Department of Physics and Astronomy 'Augusto Righi', University of Bologna, I-40129 Bologna, Italy

⁹ School of Information Sciences, University of Illinois, 501 E. Daniel St, Champaign, IL 61820, USA

¹⁰ Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611, USA

¹¹ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

¹² Canadian Institute for Theoretical Astrophysics, 60 St George Street, Toronto, ON M5S 3H8, Canada

¹³ Institute for Theory and Computation, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

Accepted 2022 September 27. Received 2022 September 26; in original form 2022 July 14

ABSTRACT

We present a post-processing catalogue of globular clusters (GCs) for the 39 most massive groups and clusters in the TNG50 simulation of the IllustrisTNG project (virial masses $M_{200} = [5 \times 10^{12} - 2 \times 10^{14}] M_{\odot}$). We tag GC particles to all galaxies with stellar mass $M_* \geq 5 \times 10^6 M_{\odot}$, and we calibrate their masses to reproduce the observed power-law relation between GC mass and halo mass for galaxies with $M_{200} \geq 10^{11} M_{\odot}$ (corresponding to $M_* \sim 10^9 M_{\odot}$). Here, we explore whether an extrapolation of this $M_{\text{GC}}-M_{200}$ relation to lower mass dwarfs is consistent with current observations. We find a good agreement between our predicted number and specific frequency of GCs in dwarfs with $M_* = [5 \times 10^6 - 10^9] M_{\odot}$ and observations. Moreover, we predict a steep decline in the GC occupation fraction for dwarfs with $M_* < 10^9 M_{\odot}$ that agrees well with current observational constraints. This declining occupation fraction is due to a combination of tidal stripping in all dwarfs plus a stochastic sampling of the GC mass function for dwarfs with $M_* < 10^{7.5} M_{\odot}$. Our simulations also reproduce available constraints on the abundance of intracluster GCs in Virgo and Centaurus A. These successes provide support to the hypothesis that the $M_{\text{GC}}-M_{200}$ relation holds, albeit with more scatter, all the way down to the regime of classical dwarf spheroidals in these environments. Our GC catalogues are publicly available as part of the IllustrisTNG data release.

Key words: galaxies: clusters: intracluster medium – galaxies: dwarf – galaxies: general – galaxies: star clusters.

1 INTRODUCTION

The formation of globular clusters (GCs) in connection to galaxies and their dark matter haloes is still unclear. Currently, the most successful models link the formation of GCs (or their early progenitors) to baryonic processes in the interstellar medium (ISM) of galaxies. These processes are connected to star formation in high density/high pressure environments (Kruijssen & Cooper 2012; Kruijssen 2015; Elmegreen 2017) and best sampled in mergers and early stages of galaxy formation (Kravtsov & Gnedin 2005; Prieto & Gnedin 2008; Li & Gnedin 2014; Renaud, Bournaud & Duc 2015). However, GCs have also been hypothesized to form at the centers of their

own low-mass dark matter haloes before reionization (Peebles 1984; Boylan-Kolchin 2017), later infalling on to larger galaxies and groups and cluster haloes to form the clustered GC distributions typically found in these systems (Diemand, Madau & Moore 2005; Creasey et al. 2019). Although this scenario predicts older ages and lower metallicities for GCs than current measurements (Lotz, Miller & Ferguson 2004; Bastian et al. 2020), the discovery of a few very metal poor GCs in M31, the Milky Way, and JWST observations may provide some support to such pristine formation scenarios playing at least some role in building the population of GCs observed in galaxies today (Larsen et al. 2020; Martin et al. 2022; Errani et al. 2022; Mowla et al. 2022).

Observationally, the mass in GCs is found to be a power-law function of inferred halo mass for galaxies with stellar mass $M_* \geq 10^{10} M_{\odot}$ (Blakeslee, Tonry & Metzger 1997; Peng et al. 2008; Spitler

* E-mail: jdoppel01@ucr.edu

& Forbes 2009; Georgiev et al. 2010; Harris, Harris & Alessi 2013; Hudson, Harris & Harris 2014; Harris, Harris & Hudson 2015). At face value, this relation may encode important information on the formation scenario of GCs. Theoretical models suggest that a quasi-linear power-law relation between GC mass and halo mass may arise naturally in hierarchical formation scenarios as the result of consecutive mergers, serving more as a confirmation of the hierarchical assembly of galaxies rather than shedding light on the formation mechanism of GCs themselves (El-Badry et al. 2019). However, in the regime of dwarf galaxies ($M_* \leq 10^9 M_\odot$), there are fewer mergers with GC-bearing companions, offering a clearer window for the study of GC formation mechanisms than in more massive galaxies. It is therefore important to extend the study of the GC mass–halo mass relation to lower mass galaxies.

Theoretical models linking the formation of GCs to the ISM of galaxies seem to suggest a downturn in the efficiency of GC formation in dwarfs, departing from the extrapolation of the GC mass–halo mass relation measured on more massive galaxies (El-Badry et al. 2019; Choksi & Gnedin 2019; Bastian et al. 2020). The lower efficiency of GC formation per halo mass in dwarf galaxies is naturally expected due to the lower baryonic content in low-mass haloes, which limits the available gas to form stellar clusters in merger and accretion events. On the other hand, a scenario where GCs are linked to dark matter mini-haloes would imply a single power-law relation between GC mass and halo mass in the regime of dwarfs, due to the self-similarity of subhalo mass in Lambda cold dark matter (Λ CDM; e.g. Creasey et al. 2019). Although current observational constraints on the radial distribution combined with the abundance of GCs in MW-mass galaxies limits the fraction of GCs formed in mini-haloes to ~ 30 per cent (Creasey et al. 2019) for such hosts, the importance of the mini-halo formation scenario for GCs in the regime of dwarfs remains largely unconstrained.

Measuring the relation between GC mass and halo mass on the scale of dwarfs is, however, very challenging. First, while there are several methods to estimate halo mass from observables in more massive galaxies (lensing, rotation curves, abundance matching), halo mass estimates in the scale of dwarf galaxies are more scarce and uncertain. Secondly, GC numbers are lower in low-mass galaxies, meaning that completeness and contamination in GC surveys impact more heavily low-mass dwarfs than estimates for high-mass galaxies. There are, however, several observational efforts to constrain the GC content in dwarf galaxies. Most notably, Forbes et al. (2018) finds that dwarfs in the Local Volume are consistent with an extrapolation of the power-law relation between GC mass and halo mass observed in more massive galaxies, where halo masses for the dwarf galaxies are estimated using gas kinematics (see Burkert & Forbes 2020 for a similar discussion using GC numbers rather than mass). However, other work cautions that this might be biased to include only dwarfs that have at least one GC, while including all dwarfs of a given mass in the average could lead to a departure downwards from the power-law extrapolation (Bastian et al. 2020).

In light of this discussion, another important diagnostic emerges as a potential constraint: the ability of galaxies of a given mass to host at least one GC, or the GC occupation fraction. Observations in the Virgo cluster suggest that all dwarfs with $M_* > 10^9 M_\odot$ have GCs, but that fraction declines quite steeply for lower mass objects, finding 50 per cent occupation in dwarfs with $M_* \sim 10^{7.5} M_\odot$ (Sánchez-Janssen et al. 2019), which is similar to the conclusion presented in Eadie, Harris & Springford (2022) using a compilation of available data for dwarfs. Recently, a comparable occupation fraction was reported for dwarf galaxy satellites of MW-like primaries in the local volume (Carlsten et al. 2022). However, the available constraints

involve mostly satellite dwarfs, or dwarf galaxies embedded in the gravitational potential of larger hosts, meaning that tidal stripping and other environmental effects might have influenced their original GC content, preventing a simple interpretation. Unfortunately, surveys of GCs in field dwarfs are scarce and still insufficient to constrain GC occupation fractions (e.g. Georgiev et al. 2010).

An interesting path forward is to use cosmological simulations of dwarf galaxies in high-density environments to understand the connection between GCs, dwarf galaxies, and their dark matter haloes. This is particularly appealing since hydrodynamical cosmological simulations of representative volumes of the Universe have been powerful tools to understand and model the evolution of satellite dwarfs and their properties – such as colour, mass content, morphology – in the environments of groups and clusters (Sales et al. 2015; Yun et al. 2019; Joshi et al. 2020; Vogelsberger et al. 2020; Donnari et al. 2021a; Engler et al. 2021a; Joshi et al. 2021) creating a realistic population of satellite dwarfs in good agreement with observations (Donnari et al. 2021b; Engler et al. 2021b; Riggs et al. 2022). However, the spatial and mass resolution of such simulations is too coarse to directly resolve the process of GC formation.

While employing idealized galaxy and galaxy merger set-ups (Bekki & Chiba 2002; Kruijssen et al. 2012; Renaud et al. 2015; Lahén et al. 2019, 2020, 2021) or cosmological zoom-in of galaxies at high redshifts (Kim et al. 2017; Ma et al. 2020; Sameie et al. 2022) have shown important successes on simulating the formation and evolution of GCs and their connection to the ISM of the host galaxy, these techniques are currently unable to sample the evolutionary history of galaxies until the present day and within high-density environments, where most of the GC observational data are available today.

To circumvent this limitation, in this paper we develop a GC catalogue added in post-processing via a particle tagging technique to make predictions on the abundance, distribution and kinematics of surviving GCs in the environments of groups and clusters at $z = 0$. This technique is inspired by the successes of particle-tagging for studying stellar halo science (Bullock & Johnston 2005; Pefiarrubia, Navarro & McConnachie 2008; Cooper et al. 2010; Laporte et al. 2013) and it has been shown to have success in modeling the *surviving* population of GCs in cosmological simulations of galaxy clusters (Ramos et al. 2015; Mistani et al. 2016; Ramos-Almendares et al. 2018, 2020; Doppel et al. 2021).

Tagging techniques of this kind mentioned above are complementary to more detailed methods where GCs formation sites are identified in hydrodynamical simulations and followed in time by a set of subgrid prescriptions to model their evolution until the present day (e.g. Kruijssen et al. 2011; Mistani et al. 2016; Li et al. 2017; Pfeffer et al. 2018; Keller et al. 2020; Trujillo-Gomez et al. 2021; Chen & Gnedin 2022; Reina-Campos et al. 2022). Note that most of these works also require fairly high-resolution simulations and have been mostly focused on the scale of MW-like galaxies so far. Instead, the less computationally intensive modeling associated with particle tagging methods offer the opportunity to compile theoretical predictions for the GC content, their positions and velocities for a large number of galaxies and dwarfs with realistic properties within high-density environments such as simulated groups and galaxy clusters.

Here, we extend the particle-tagging method applied in Ramos-Almendares et al. (2020) for Fornax and Virgo mass galaxy clusters ($M_{200} \geq 8 \times 10^{13} M_\odot$) that was implemented in the Illustris simulations (Vogelsberger et al. 2013, 2014a,b; Genel et al. 2014) to lower mass dwarf galaxies using the highest resolution hydrodynamical run of the TNG50 simulation (Pillepich et al. 2019; Nelson et al.

2019b). Thus, the tagged GCs allow us to study the GC content of a variety of galaxy groups and clusters consistent with mass estimates of Centaurus A, Fornax, Hydra, and Virgo, where observations of GCs are abundant. This work presents one of the largest studies of its kind, containing 39 groups and clusters including their associated 5000+ galaxies with $M_* \geq 5 \times 10^6 M_\odot$, and 196 611 GCs. The GC catalogues generated for this work are made publicly available¹ (see Data Availability section for accessibility information).

The paper is organized as follows. In Section 2, we present the simulation and GC tagging technique. In Section 3, we show our results on the intracluster GC component and benchmark our catalogue using current observations. Our main results on the content of GCs in dwarf galaxies are shown in Sections 4 and 5. We summarize our main findings in Section 6.

2 METHODS

2.1 The TNG50 simulation

We use the highest resolution hydrodynamical run of the TNG50 simulation (Pillepich et al. 2019; Nelson et al. 2019b), which allows us to relate the properties of the tagged GCs directly to the properties of galaxies, galaxy groups or galaxy clusters that they belong to. TNG50 is an unprecedentedly high-resolution cosmological hydrodynamical simulation for its volume, with a box size of 51.7 Mpc per side with 2160^3 gas and dark matter particles, allowing for a mass resolution of, on average, $8.4 \times 10^4 M_\odot$ for baryons and a fixed mass resolution of $4.5 \times 10^5 M_\odot$ for dark matter. The simulation has a gravitational softening length of 288 pc for stars and dark matter at $z = 0$. TNG50 assumes a flat, Λ CDM cosmology and uses cosmological parameters from Planck Collaboration XIII (2016). Its galaxy formation model follows star formation in moderately dense ISM conditions, stellar evolution, and chemical enrichment via supernovae, primordial, and metal line cooling of gas, as well as heating from the background radiation field, the seeding and subsequent growth of supermassive black holes as well as AGN feedback at both low and high accretion rates, and galactic winds (Weinberger et al. 2017; Pillepich et al. 2018a). The TNG50 simulation is part of the larger IllustrisTNG project (Naiman et al. 2018; Pillepich et al. 2018b; Nelson et al. 2018; Springel et al. 2018; Marinacci et al. 2018; Nelson et al. 2019a).

2.2 Galaxy selection

We tag GCs in all TNG50 host haloes with a virial mass $M_{200} \geq 5 \times 10^{12} M_\odot$ (where M_{200} refers to the mass within the virial radius r_{200} defined as the radius enclosing an average density equal to 200 times the critical density of the Universe). This selection results in 39 groups and clusters with a virial mass distribution shown in Fig. 1. The high-mass end is roughly on par with lower estimated virial masses of the Virgo cluster $M_{200} \sim 10^{14} M_\odot$ (lime green circle and errorbar Karachentsev & Nasonova 2010; Weinmann et al. 2011) and Hydra 1 (brown circle and errorbar Tamura et al. 2000), Fornax cluster $M_{200} \sim 10^{14} M_\odot$ (cyan circle and errorbar Drinkwater, Gregg & Colless 2001), down to Centaurus A with estimated $M_{200} \leq 10^{13} M_\odot$ (dark purple circle and errorbar van den Bergh 2000; Karachentsev et al. 2007) and less massive elliptical systems closer to the lower mass cut.

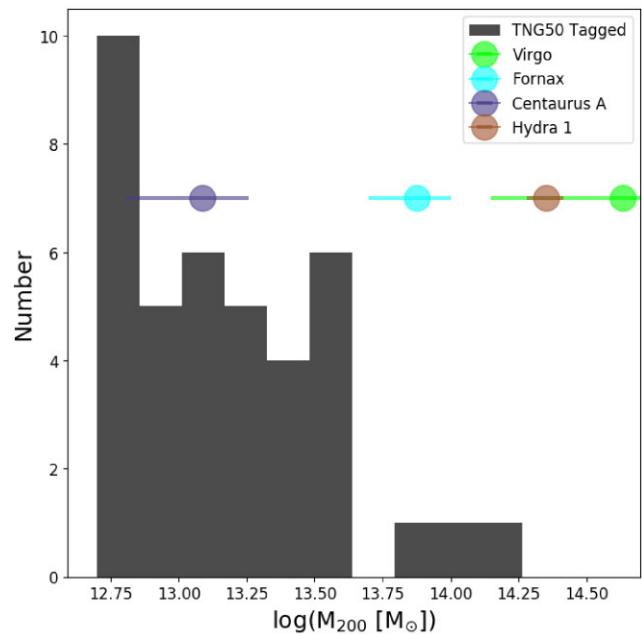


Figure 1. The distribution of TNG50 $z = 0$ virial masses (M_{200}) of the 39 most massive galaxy groups and clusters within the simulation to which we tag GCs (black histogram). We cover a wide range of masses, from Centaurus A on the low-mass end (dark purple circle and errorbar van den Bergh 2000; Karachentsev et al. 2007), to Fornax (cyan circle and errorbar Drinkwater et al. 2001), and to Hydra 1 (brown circle and errorbar Tamura et al. 2000) and low end mass estimates of Virgo (lime green circle and errorbar Karachentsev & Nasonova 2010; Weinmann et al. 2011) on the high-mass end. The ~ 1.5 dex range of virial masses allows us to study potential effects environment might play in their $z = 0$ GCs.

We identify all galaxies that interacted with each of these groups (defined here as being part of their merger tree) and achieved a maximum stellar mass $M_{*,\max} \geq 5 \times 10^6 M_\odot$ during their lifetime as candidates to host the tagged GCs. For each of our selected galaxies, we calculate their infall time t_{inf} , defined here by following their main branch progenitors in the Sublink merger tree (Rodriguez-Gomez et al. 2015) to the last time that the progenitor was its own central. This corresponds to the snapshot before they begin interacting with their current host galaxy group or cluster, or any lower mass halo that eventually merges with the group or cluster (i.e. in pre-processing Benavides, Sales & Abadi 2020; Joshi et al. 2021). Here, we also impose a minimum of 100 dark matter particles to remove spurious objects in the subhalo catalogue. In the case of the central galaxy in each of our 39 groups, following Ramos-Almendares et al. (2020), we define the infall time as the snapshot when the main progenitor branch reaches 5 per cent its $z = 0$ value.

The target selection process gives us 8746 progenitor galaxies to be tagged with GCs at their infall time of which 6415 survive to $z = 0$. For our study of GCs associated with satellite galaxies, our final sample includes 5453 satellite galaxies in groups and clusters with $M_* \geq 5 \times 10^6 M_\odot$ at $z = 0$, which guarantee well resolved galaxies with at least ~ 60 stellar particles at $z = 0$.

2.3 GC tagging

The method to tag GCs in our cosmological simulation follows mostly from the one already introduced in the Illustris simulations by Ramos-Almendares et al. (2020), with some modifications and improvements to extend the model to lower mass galaxies. The

¹www.tng-project.org/doppel22

method ‘tags’ GCs to a set of dark matter particles, selected to have a given energy distribution (enforced through a specific distribution function) that matches observational properties of GCs systems at $z = 0$. In principle, one could choose to tag on any collisionless-type particle, for example stars. We instead favour the tagging of dark matter particles to ensure that all galaxies have enough available particles with the desired distribution function to select from when assigning GCs. In particular, GCs systems observed in galaxies are dispersion dominated and typically more extended than the stellar component. Using the dark matter component to search for suitable tracers in energy-space ensures that we maximize the number of candidate particles to host GCs since dark matter is always dispersion dominated (unlike stars in disks) and more extended than the stars. This is particularly important in the regime of dwarfs, where the stellar content is low resulting on a low number of stellar particles overall and even lower beyond the inner central regions (see for instance declining stellar halo fractions predicted in dwarf galaxies, fig. 5 in Elias et al. 2018). Following Ramos-Almendares et al. (2020), the tagging is done only once (at infall time) for each galaxy, after which the particle ID is used to identify those tagged GCs in the $z = 0$ snapshot.

The first step is to identify, for each object, the maximum subset of dark matter particles that are candidates to be GCs, defined as those that are consistent with a specified distribution in energy adopted for the GCs. We assume that the dark matter follows a NFW profile (Navarro, Frenk & White 1996):

$$\rho_{\text{NFW}}(r) = \frac{\rho_{\text{NFW}}^0}{(r/r_{\text{NFW}})(1+r/r_{\text{NFW}})^2} \quad (1)$$

which we find by best fit to the density distribution of dark matter particles following Lokas & Mamon (2001) at infall time. We assume $r_{\text{NFW}} = r_{\text{max}}/\alpha$, where r_{max} is the radius of maximum circular velocity and $\alpha = 2.1623$ (Bullock et al. 2001). We calculate r_{max} for each galaxy at their time of infall.

GCs are assumed to follow a Hernquist profile (Hernquist 1990):

$$\rho_{\text{HQ}}(r) = \frac{\rho_{\text{HQ}}^0}{(r/r_{\text{HQ}})(1+r/r_{\text{HQ}})^3}. \quad (2)$$

Two sets of GCs are tagged, one corresponding to a more extended metal poor or ‘blue’ component, and one more concentrated and metal rich, or ‘red’ component, with relative fraction of red to blue component following observations in Harris et al. (2015). We assume that $r_{\text{HQ}} = \beta r_{\text{NFW}}$, where $\beta_{\text{blue GCs}} = 3.0$ and $\beta_{\text{red GCs}} = 0.5$. The remaining parameter ρ_{HQ}^0 is fit such that the number of resultant candidate particles is maximized. The procedure as well as the assumed parameters is the same as introduced in Ramos-Almendares et al. (2020) using the Illustris simulations.

For reference, the resulting radial distributions of red and blue GCs are typically within the tidal radius of surviving satellites, which are estimated to be $r_{\text{tidal}} \sim 5\text{--}100$ kpc in our sample using analytical calculations for our highest and lowest host and satellite masses (Binney & Tremaine 2008; Springel et al. 2008). This is confirmed by a very high fraction of tagged GCs remaining bound at $z = 0$, which show medians 96 per cent and 85 per cent for red and blue GCs, respectively. Note that although the sample as a whole shows large bound fractions, a minority of individual objects may lose most or in some cases *all* of their GCs for specific orbits or accretion histories, introducing scatter in some of the relations explored in Section 4.

We numerically compute the distribution function of each of these three components (dark matter NFW, blue GCs and red GCs) as

Binney & Tremaine (2008):

$$f_i(\epsilon) = \frac{1}{8\pi} \left[\int_0^\epsilon \frac{d^2\rho_i}{d\psi^2} \frac{d\psi}{\sqrt{\epsilon - \psi}} + \frac{1}{\sqrt{\epsilon}} \left(\frac{d\rho_i}{d\psi} \right) \Big|_{\psi=0} \right], \quad (3)$$

where ρ_i is the density profile of $i = \text{DM, GCs, blue GCs}$, Ψ is the relative gravitational potential, and ϵ is the relative energy. Since the potential is not recorded in every snapshot, it should be noted that the potential of the dark matter particles is calculated for each progenitor subhalo via a tree gravity for computational efficiency. Then, in equally spaced bins of relative energy, we select a fraction of the particles $f_{\text{HQ},i}/f_{\text{NFW}}$ for $i = \text{red and blue GCs}$ to be the GC candidate particles. We impose an additional radius cut of $r_h/3$, as suggested by Yahagi & Bekki (2005) and implemented in Ramos-Almendares et al. (2020), where r_h is the half-mass radius of the entire halo at its infall time. This is the final set of GC *candidate* particles.

The next step is to populate galaxies with a total mass in GCs, or M_{GC} . This is the key assumption of the method: galaxies follow a power-law relation between the mass of their total GC systems and their virial mass M_{200} at infall. We thus calibrate the model such that after evolving in the cluster of host potential (tidal stripping, stellar evolution), they reproduce the observed power-law $M_{\text{GC}} - M_{200}$ relation at $z = 0$. More specifically, from Harris et al. (2015),

$$M_{\text{GC},z=0} = a M_{\text{halo},z=0}^b, \quad (4)$$

where $a = 2.6 \times 10^{-8}$ and 4.9×10^{-5} for red and blue GCs, respectively, and the slopes $b = 1.2$ and 0.96 for red and blue GCs. As done in Harris et al. (2015), $M_{\text{halo},z=0}$ is calculated using abundance matching parameters from Hudson et al. (2015) to assign halo masses to satellites. To calibrate this relation, we select from our satellite sample described in Section 2.2, only those that survived to $z = 0$, and calculate the fraction of the candidates GC particles that are still bound to the galaxy at $z = 0$: $f_{\text{bound}} = N_{\text{candidates}(z=0)}/N_{\text{candidates}(z_{\text{inf}})}$. We consider a GC candidate still bound to a subhalo at present day if its corresponding dark matter particle is considered bound to the subhalo via Subfind. We then make the assumption that the relationship between $M_{\text{GC}} - M_{\text{halo}}$ also followed a power law at infall such that

$$M_{\text{GC},\text{inf}} = \frac{1}{f_{\text{bound}}} M_{\text{GC},z=0} = a_{\text{inf}} M_{\text{halo},\text{inf}}^{b_{\text{inf}}}. \quad (5)$$

We find the best-fitting $a_{\text{inf}} = 2.6 \times 10^{-7}$ and 7.3×10^{-5} and $b_{\text{inf}} = 1.14$ and 0.98 for red and blue GCs respectively. The infall GC mass of each galaxy is then calculated using their virial mass from this best-fitting infall relation at t_{inf} . The result of this calibration is shown in Fig. 2. Blue and red points represent the resulting present-day blue and red GC mass, respectively, for each galaxy with a given M_{200} . For reference, the magenta and cyan lines show the results from Harris et al. (2015) for red and blue GCs, respectively. Note that, despite all galaxies starting from a scatter-free infall $M_{\text{GC}} - M_{200}$ relation, the variations in infall time, tidal stripping and stellar evolution of the galaxies (which might influence the calculation of M_{200} from abundance matching) results in a present-day $M_{\text{GC}} - M_{200}$ relation with scatter, in agreement with observations (see Ramos-Almendares et al. 2018, 2020, for more detailed discussions).

Most importantly, the calibration to determine M_{GC} is done using only more massive galaxies, where observational constraints on the GC-halo mass relation are available. In particular, only satellites with estimated $M_{200} \geq 10^{11} \text{ M}_\odot$, which roughly corresponds to $M_* \geq 10^9 \text{ M}_\odot$ using Hudson et al. (2015) are used to calibrate the model. For dwarf galaxies with $M_{200} < 10^{11} \text{ M}_\odot$, the calculated M_{GC} is a

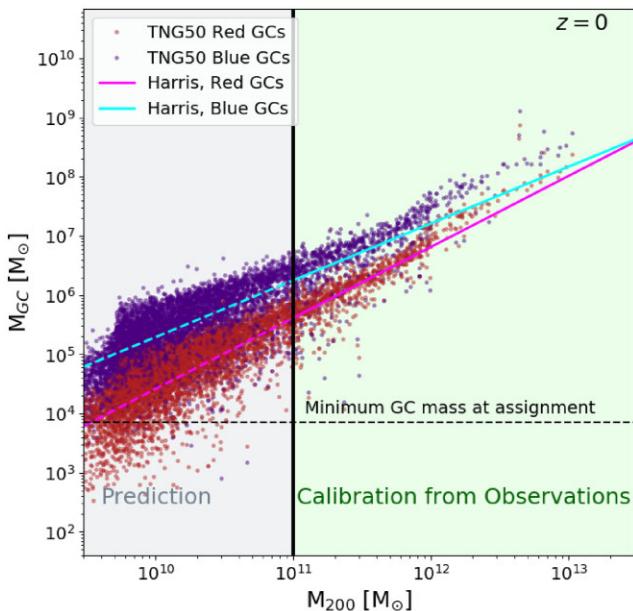


Figure 2. M_{GC} as a function of halo mass M_{200} at present-day from the TNG50 simulation + GC tagging model that shows the result of the mass calibration process. Individual dots show simulated galaxies for the blue (indigo) and red (red) components. The observed GC mass–halo mass relation from Harris et al. (2015) is shown in cyan and magenta solid lines for blue and red GCs, respectively. The extrapolation of those results to dwarf galaxies is indicated with the same colours but using short dashed lines. For galaxies with $M_{200} > 10^{12} \text{ M}_\odot$, we plot the virial mass corresponding to the simulation value rather than calculated from Hudson et al. (2015) due to large discrepancies between simulations and the abundance matching model in that regime. The calibrated red and blue GCs follow a power law with a slope in good agreement with observations and predict a variable scatter that increases towards the low-mass end. Most importantly, results for $M_{200} < 10^{11} \text{ M}_\odot$ are a prediction of the model since the calibration is done only using systems more massive than this cut-off. The horizontal dashed line shows $M_{\text{GC}} = 7 \times 10^3 \text{ M}_\odot$, our minimum individual GC mass considered to assign mass to the tagged GC particles. Galaxies below this mass are not populated with GCs in our model.

prediction of the model, assuming they follow the same relation as their more massive counterparts.

2.4 Assigning individual GC masses

As explained above, the tagging method first selects as many GC candidate particles as possible, by identifying all dark matter particles with matching energies to the intended GC distribution (see Section 2.3). After the mass calibration is carried out and M_{GC} is defined at infall (see equations 4 and 5), the mass weight of each candidate GC particle is simply calculated by dividing M_{GC} into the identified number of candidate GC particles. This means that the weight of a given tagged GC particle could be smaller than the mass of a full GC. While working with the full set of candidate GC particles provides the most complete representation of the possible phase space for GC systems, it is convenient to define a ‘realistic GC catalogue’, where a subset of the tagged GC candidate particles are selected to match the *number* of GCs expected. We take this approach in what follows, as it allows a consistent comparison to observational data.

In previous iterations of this GC tagging model, we have taken the approach of assigning all realistic GCs the same, average mass ($m_{\text{GC}} = 1 \times 10^5 \text{ M}_\odot$; see Ramos-Almendares et al. 2020; Doppel

et al. 2021). While this approach was correct for the more massive galaxy sample presented in these previous works, the GC luminosity function in observed early-type galaxies changes with the stellar mass of the host (Jordán et al. 2007), an effect that becomes particularly important when evaluating the GC content of lower mass dwarf galaxies (Forbes et al. 2018). Because of our increased resolution and the selection of dwarf galaxies below $M_* \sim 10^8 \text{ M}_\odot$, we enter the regime in which a more detailed mass modelling for the GCs is required.

We model the GC population of each galaxy at infall assuming a Gaussian distribution in luminosity [we assume a mass to light ratio (M/L) ($\text{M}_\odot / \text{L}_\odot$) = 1 in the z band], with a dispersion of z -band GC luminosities (σ_z) that reproduces the relationship with the M_B of their host galaxies at $z = 0$ as measured in Jordán et al. (2007; see Appendix A for a more detailed discussion of this calibration). We note that we still assume a constant mean luminosity $\sim 2 \times 10^{5} \text{ L}_\odot$ for all GC luminosity functions, independent of the mass of the host galaxy, but we limit the maximum mass that a GC can sample to 1/100 the stellar mass of the host following observations of the most massive GCs in dwarfs (Kruijssen & Cooper 2012). We also employ a uniform low-mass (or luminosity) cut-off for individual GCs $= 7 \times 10^3 \text{ M}_\odot$ and an upper mass/luminosity cut-off equal to $5 \times 10^6 \text{ M}_\odot$ to ensure that we are excluding massive objects that could be nuclear GCs (Kruijssen & Cooper 2012).

For each galaxy, we proceed to randomly draw individual GC masses from the resulting Gaussian distribution until the sum of all realistic GC candidates adds up to the estimated mass in GCs at infall. This steps concludes with a corresponding number of realistic GCs for each galaxy, $N_{\text{GC, inf}}$. This number is always smaller than the number of particles identified as candidate GC particles in the step described in Section 2.3. We then subsample $N_{\text{GC, inf}}$ from the list of all the GC candidate particles identified for each galaxy (enforcing that they follow the same relative energy distribution function) and we assign them one of the drawn GC masses, building one possible realization of the realistic GC catalogue for each given galaxy. Note that further versions of the realistic catalogue might be constructed by repeating the sampling of the GC luminosity function and the selection of the GC candidate particles, if so desired. In this work, we employ only one realization per galaxy, but see Doppel et al. (2021) for an example of how multiple realizations per object might be used to assess the impact of low number statistics in the determination of galaxy velocity dispersion from GC tracers.

As highlighted before, the individual mass assignment to GCs is performed at infall, and particle IDs are tracked onwards to $z = 0$. Since the tagging technique is meant to model the *surviving* GCs at $z = 0$, we do not make assumptions about the shape of the *initial mass function* of GCs, nor do we take into account mass-loss for individual GCs or the total destruction of GCs (see also Ramos-Almendares et al. 2020 for a detailed discussion). We instead use observational results on the evolved GC luminosity function presented in Jordán et al. (2007) to assign final masses to the tagged surviving GCs. We present in Appendix B estimates of the dynamical friction expected for the tagged GCs and demonstrate that the results presented here are not strongly affected by dynamical friction.

2.5 The GC population of TNG50 group and cluster members

The top panel of Fig. 3 shows the average GC mass functions at $z = 0$ binned in ranges of stellar mass of the host galaxy in TNG50. Notice that while the GC mass sampling and assignment is performed at infall for all galaxies, this plot shows present-day results for surviving satellite galaxies, which means individual distributions of GCs could have been affected by tidal stripping. The top panel shows

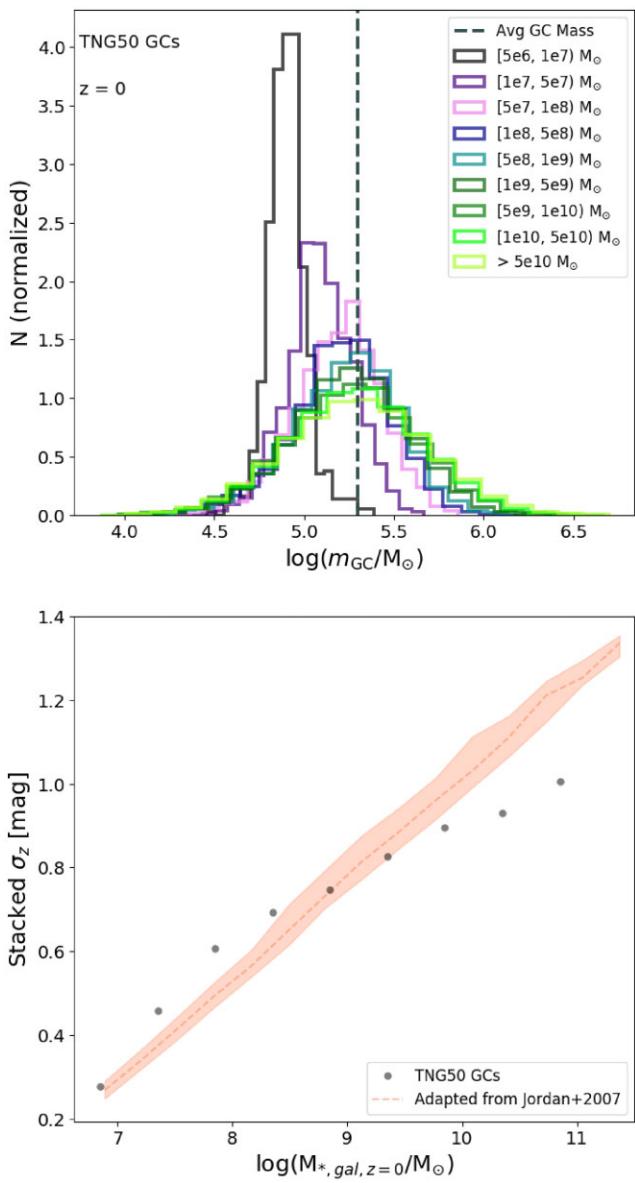


Figure 3. *Top:* Stacked mass functions of individual GCs in narrow bins of host galaxy stellar mass at $z = 0$, as labelled. We can quantitatively see the expected decrease in dispersion for decreasing stellar mass. We note that the downward shift of the median GC mass is due to the upper limit of $\min(5 \times 10^6 M_\odot, M_{*,\text{infall}}/100)$, which plays a large role for lower mass galaxies. *Bottom:* Relation between the dispersion in the z -band GC luminosity function, σ_z , and host galaxy stellar mass. Gray points show the measured dispersion of the stacked luminosity functions of our model in TNG50 shown in the top panel, and the orange shaded region shows observational expectations based on results reported in Jordán et al. (2007) adapted using the simulations to convert their B -magnitudes to M_* (median shown in dashed line, and shading corresponds to 25–75 per cent scatter in each stellar mass bin). Additionally, due to the high end mass cut of $5 \times 10^6 M_\odot$ for a single GC mass, the best-fitting luminosity function dispersions of high-mass systems are somewhat underestimated, but we do find reasonable agreement for low-mass systems.

a significant drop in the average GC mass for dwarf galaxies with $M_* < 10^8 M_\odot$, which in our model is attributed to the upper mass cut-off to sample GC mass (set to $1/100 M_{*,\text{infall}}$ for each galaxy), and confirms the importance of taking this into account when dealing with GC content in low-mass dwarf galaxies (Forbes et al. 2018).

The bottom panel of Fig. 3 shows in cyan symbols the dispersion in the z -band magnitudes of simulated GCs associated to each galaxy, σ_z , and how it compares to the one measured in observations (orange shaded region Jordán et al. 2007). Note that this relation is an extrapolation below $M_* < 10^8 M_\odot$. While the overall agreement is good, there is a flattening in σ_z for our tagged GCs in high-mass galaxies, which we attribute to our absolute upper limit in the z -band luminosity/mass of individual GCs corresponding to $5 \times 10^6 M_\odot$.

We showcase some examples of our GC catalogue with the final tagging results in Fig. 4. Pink and light blue dots indicate our tagged realistic GCs overplotted on to the stellar density predicted by TNG50, shown in the background greyscale. To create some intuition on the range of simulated objects included in our sample, we show several systems on different mass scales, from a Virgo-like galaxy cluster in the top left of the figure, a Fornax mass galaxy cluster in the top right, and a Centaurus A mass group in the bottom left. Interestingly, it is not uncommon to find substructures of GCs in our catalogue: the bottom right-hand panel shows a set of simulated GCs that appear to be following a tidal stream in the stellar component of a disrupting host galaxy. We also see correct behaviour of the GCs as a whole – the red GCs are more spatially concentrated around their host galaxies than the blue GCs. While this is partially imposed by design in the model, the more centrally concentrated tagging for the red component is done at infall, while Fig. 4 shows that it is mostly preserved until $z = 0$ despite tidal stripping events and interactions with the host environment. In agreement with previous version of this tagging technique (Ramos-Almendares et al. 2018, 2020), the model predicts the formation of an intracluster GC component, or GCs that exist in the space between the galaxies, which is built mostly from the disruption and merging of early accreted satellite galaxies, a topic that we return to in Section 3.

3 BUILD-UP OF THE INTRACLUSTER GC COMPONENT

Observationally, the presence of GCs in the intracluster regions (or ICGCs) has been detected and surveyed in several nearby galaxy groups and clusters such as Fornax (Bassino et al. 2003; Schuberth et al. 2008), Coma (Peng et al. 2011; Madrid et al. 2018), Abell 1689 (Alamo-Martínez & Blakeslee 2017), Virgo (Lee et al. 2010; Durrell et al. 2014; Ko et al. 2017; Longobardi et al. 2018), and Centaurus A (Taylor et al. 2017). Similarly to GCs in the halo of the MW (see e.g. Keller et al. 2020), ICGC studies hold the promise to help unravel the accretion history of their host haloes and important properties of the progenitor galaxies building the intracluster light of the groups and clusters (e.g. Villaume et al. 2020; Ko et al. 2022).

One of the predictions of our GC model is the formation of such an accreted ICGC component, built from a combination of GCs previously associated with galaxies that have merged to the group or cluster host and also from the stripping of surviving satellite galaxies. Such a component is not directly ‘tagged’ or calibrated for in our simulations, but instead is the result of the hierarchical assembly of structures in Λ CDM. More specifically, while some GCs are tagged to the central galaxies in each group, this occurs when they reach a very small fraction of their final virial mass (5 per cent, see Section 2 for details), resulting in those GCs assigned to the central galaxy being largely subdominant (about a \sim dex less

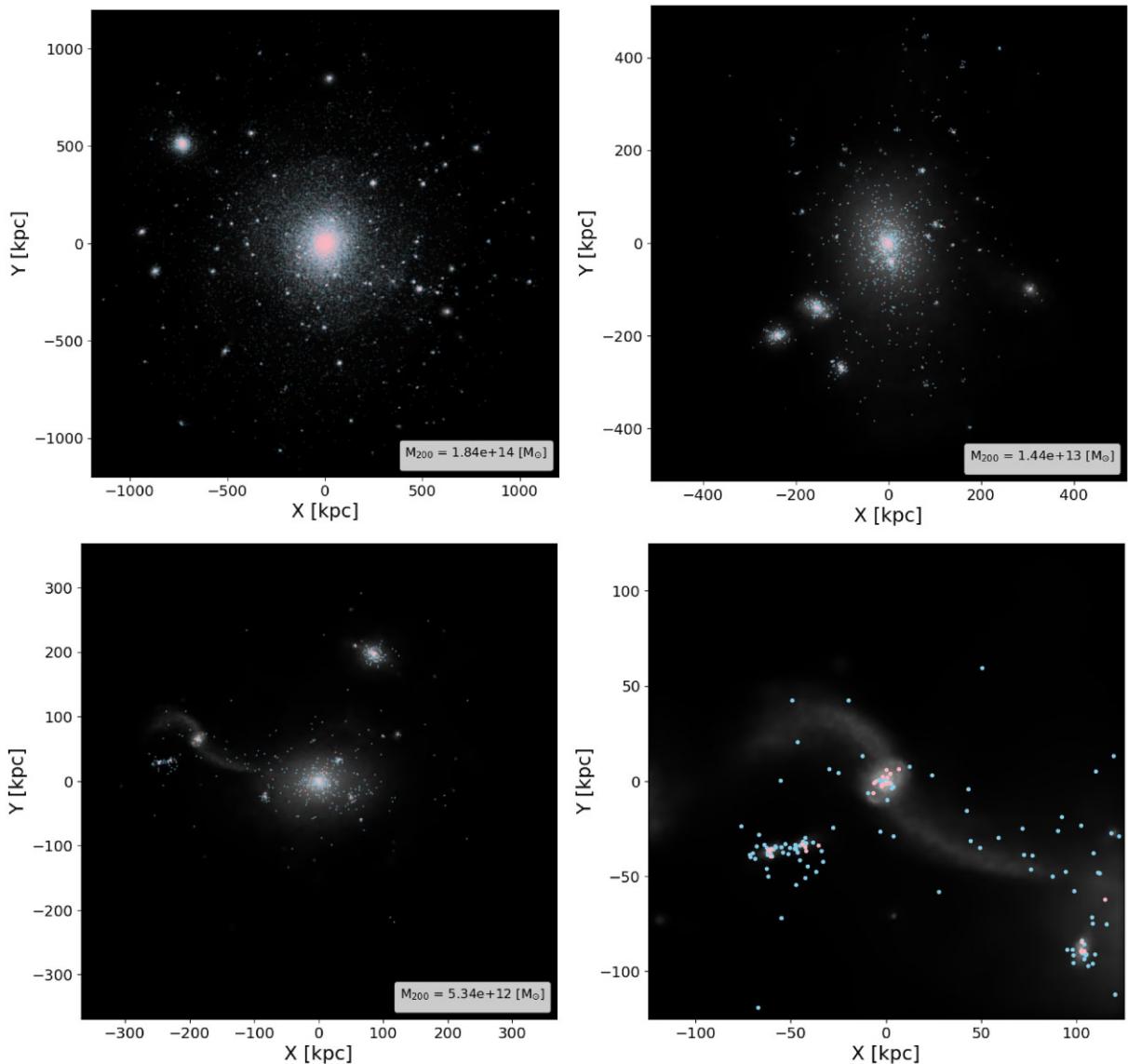


Figure 4. Spatial maps of one realization of our GC catalogue (pink and light blue points) overplotted on a visualization of the stellar density (background greyscale) for the most massive galaxy group (a Virgo or Hydra 1 analogue) (top left), a galaxy group with a viral mass around $1.5 \times 10^{13} M_{\odot}$ (top right), and one low-mass galaxy group with a virial mass $\sim 5 \times 10^{12} M_{\odot}$ (bottom left). The bottom right shows a zoom-in of the GC particles associated with the stellar stream in the bottom left image. We find that the GCs distribute as expected, with the red population more spatially concentrated about their hosts and the blue component more spatially extended. We also find the presence of intracluster GCs, see Section 3 for a more detailed discussion.

in GC numbers) compared to the accreted ICGCs acquired from tidal stripping and merging of the satellite galaxies. The study of the ICGC component is therefore an important benchmark of our GC model.

In this work, we define ICGCs to be GCs within the virial radius of a group or cluster host that are not currently gravitationally associated with any satellite as measured using Subfind. We note that while this differs from observational methods of determining GC membership to the ICL, which includes fitting profiles to distinguish the ICGCs from the GCs associated with the BCG (e.g. Taylor et al. 2017), employing radial cuts to remove the contribution of GCs of satellite galaxies (e.g. Lee et al. 2010), or using kinematic data of GCs when available (Longobardi et al. 2018), this is a definition that is best physically motivated for our purposes. We have explicitly checked that using different radius cuts for satellite galaxies to distinguish

between the ICGCs from the GCs of satellites, as done in some observational studies, does not substantially change the properties of the ICGCs reported here.

The top row of Fig. 5 shows projections of GCs (associated with galaxies and part of the ICGCs) tagged in the second most massive group in our sample (FoF group 1), with $M_{200} \sim 9 \times 10^{13} M_{\odot}$, comparable to the Virgo or Fornax clusters. As before, pink and light blue dots correspond to tagged red or blue GCs and the greyscale indicates the stellar component. Because GCs are now assigned individual masses (see Section 2.4), we can create different maps mimicking different luminosity (or mass) cuts: the left-hand panel shows all tagged GCs in FoF 1 (or equivalent all GCs above a mass cut $7 \times 10^3 M_{\odot}$), while the right-hand panel illustrates what would be observed in a shallower survey only able to map GCs more massive than $\geq 10^6 M_{\odot}$.

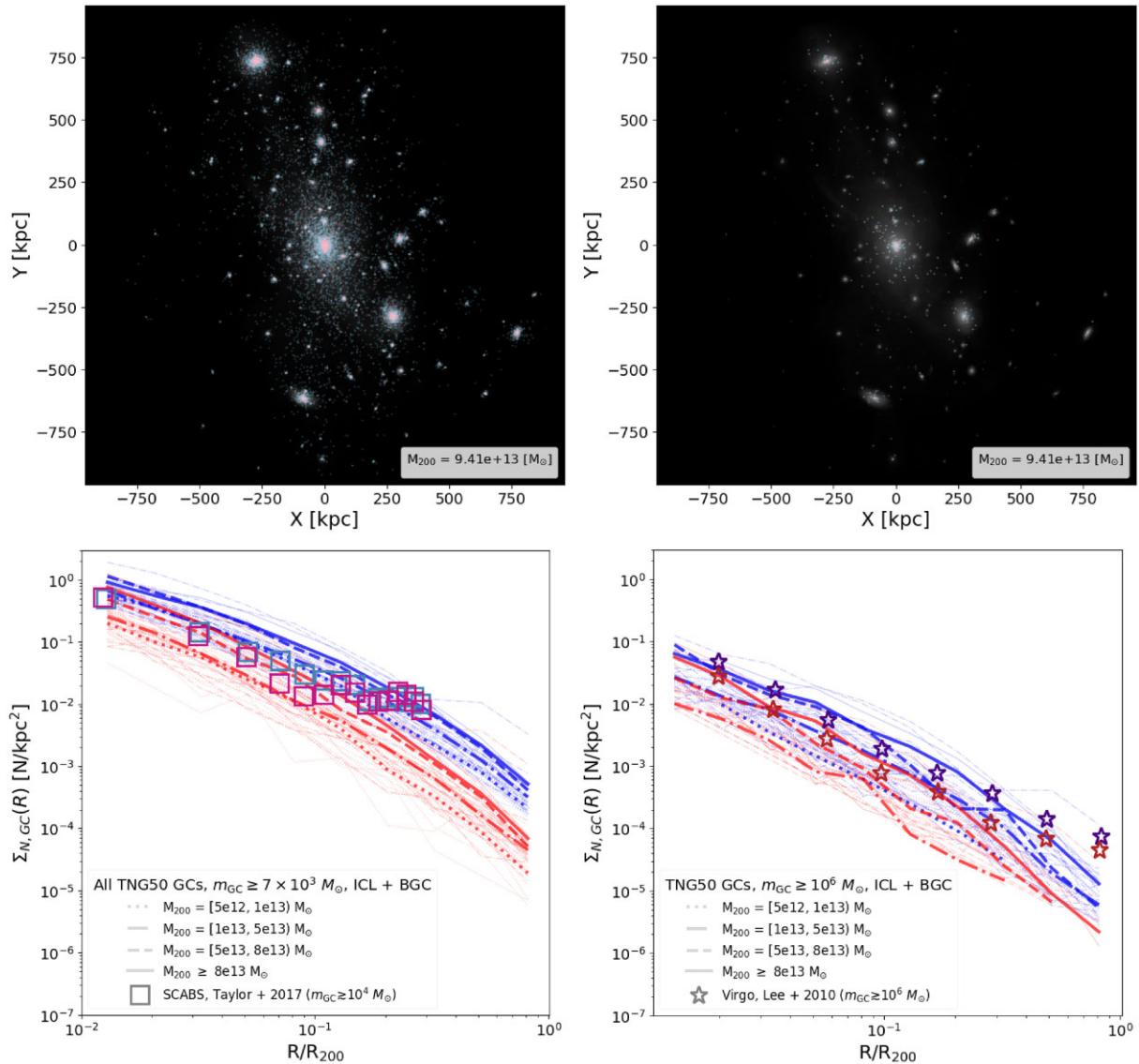


Figure 5. Top: Projections of the stars (background greyscale) and GCs (pink and skyblue points) for FoF group 1 all the realistic GC particles associated to the group (left) and the most massive (and thus the brightest) GC particles, defined to be those with individual GC mass $m_{\text{GC}} \geq 1 \times 10^6 M_{\odot}$ (right-hand panel). Bottom: Radial surface number density profiles for the GCs for all groups (low transparency curves) and medians for various mass bins (high alpha curves) compared to observations for the Virgo ICGCs from Lee, Park & Hwang (2010, red and blue stars,) and SCABS (pink and cobalt squares, Taylor et al. 2017). The bottom left shows the profiles using all realistic GC particles and the bottom right shows the profiles using only massive GC particles, as defined for the top row. We see that this mass cut puts the predictions of the model much more in-line with what is shown in the observations from Lee et al. (2010). This visually illustrates the effects of brightness cuts in observations of the ICGCs.

As expected, the number of GCs decreases in the right-hand panel due to the lower availability of more rare massive GCs. Interestingly, the substructure mapping should be different between these two images, as more massive GCs are preferentially formed in more massive galaxies (see top panel of Fig. 3), leaving dwarf galaxies underrepresented in the right-hand panel compared to the left. The extension of our model to include the masses of individual GCs makes the current GC catalogue especially useful for exploring how completeness and magnitude limits might impact observational results.

We quantify the ICGCs via their projected number density profile as a function of projected radius (normalized to the virial radius of the host) in the bottom panels of Fig. 5. The left-hand and right-hand panels correspond again to all GCs and GCs more massive than 10^6

M_{\odot} , respectively. Individual thin lines (red or blue to refer to the red or blue GCs) indicate the projected radial profiles in each of our 39 groups, while thick curves show the resulting medians when splitting our sample in four virial mass bins roughly consistent with: Virgo mass objects ($M_{\text{vir}} \geq 8 \times 10^{13} M_{\odot}$), Fornax mass objects ($5 \times 10^{13} M_{\odot} \leq M_{200} < 8 \times 10^{13} M_{\odot}$), higher-end mass estimates of Centaurus A ($1 \times 10^{13} M_{\odot} \leq M_{200} < 5 \times 10^{13} M_{\odot}$), and lower-end halo mass estimates of Centaurus A as well as massive elliptical systems ($5 \times 10^{12} \leq M_{\text{vir}} < 1 \times 10^{13} M_{\odot}$).

There is a weak dependence of the ICGC number density on host mass, with smaller mass systems having lower number densities, but the object to object scatter is large. The red ICGCs have a slightly steeper radial distribution than the blue one, as expected from the differential stripping due to their initially more biased distribution

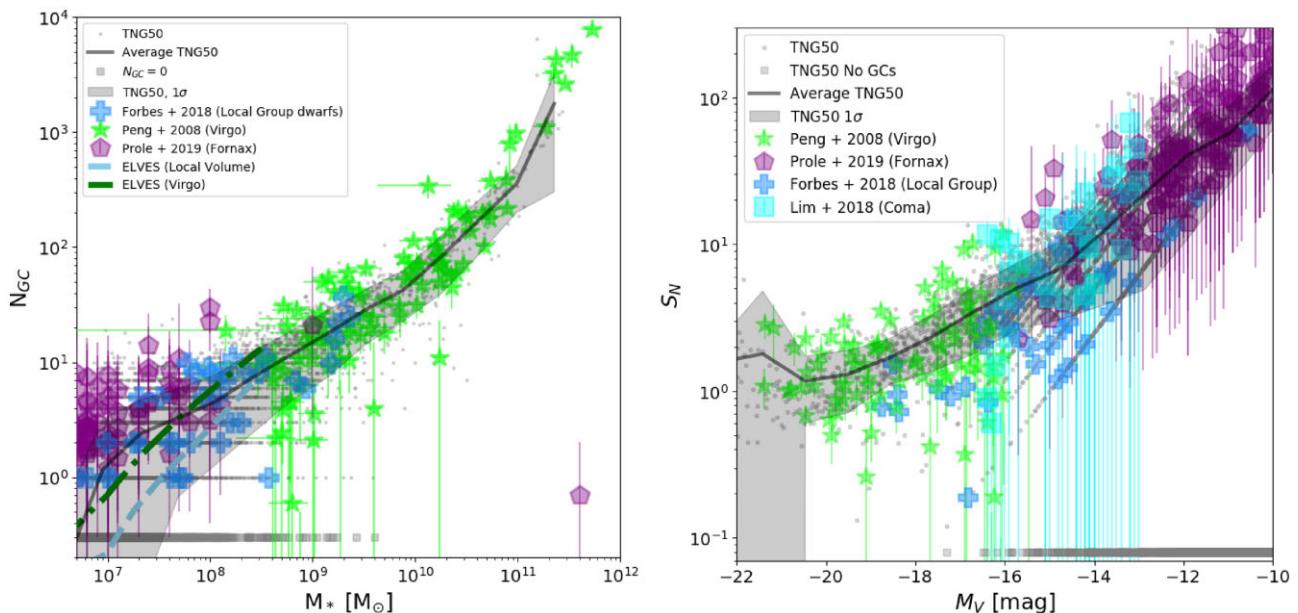


Figure 6. *Left:* Number of GCs, N_{GC} , as a function of host galaxy stellar mass. Simulation points are plotted as grey dots, with the average shown as the solid black line and 1σ variation shown as the grey shaded region. Simulated galaxies with no GCs after the mass function selection are shown as grey squares at $N_{\text{GC}} = 0.2$. Observational data for observed galaxies are plotted as solid lime green stars (Virgo, Peng et al. 2008), purple pentagons (Fornax, Prole et al. 2019), and blue crosses (Local Group, Forbes et al. 2018). Average results from ELVES-II in Virgo- and Local Volume-like environments are shown in dark green dot-dashed and light blue dashed lines, respectively. The simulation points tend to follow the trend and scatter of the observational data. *Right:* The specific frequency, S_N , as a function of host galaxy V -band magnitude, M_V . M_V and S_N have been corrected to correspond to the mass-to-light ratio observed for Virgo (see Appendix A). Coloured shapes correspond to the same observations as before, with the addition of cyan squares (Coma Lim et al. 2018). Galaxies with $S_N = 0$ are shown as grey squares at $S_N = 0.07$. The agreement of both measures of GC abundance with observations in terms of shape and scatter suggests that the assumption that GC mass scales with halo mass holds to a reasonable extent, even into the dwarf regime.

towards the centers of their host galaxies at infall, but the effect is rather small.

Global GC surveys are very challenging observationally for external and distant systems. As a result the available data is scarce. We compare our predictions with two available constraints: GCs in Cen A from the SCABS survey that correspond to a minimum GC mass of $\sim 10^4 M_{\odot}$ (Taylor et al. 2017, red and blue squares) and GCs in the Virgo cluster from Lee et al. (2010). In the bottom left-hand panel, we see that our model shows an overall good agreement with measurements in Cen A, although we predict a steeper red ICGC component than the SCABS results. Here we are assuming a virial mass $M_{200} = 10^{13} M_{\odot}$, which corresponds to $R_{200} \sim 450$ kpc. The flattening observed in Cen A beyond $R \sim 0.2R_{200}$ might be associated with the ring-like structure detected in this system (Taylor et al. 2017) and might not necessarily be present in our sample, although we do find some interesting cases where simulations also predict a flattening. We defer this study to future work.

The observational data in Virgo corresponds to a brightness cut-off 21.3 mag in the i band, which means that only the brightest ~ 13 per cent of the GCs in the Virgo cluster are detected (Lee et al. 2010). We therefore show in the bottom right-hand panel of Fig. 5 the number density profiles of GCs more massive than $10^6 M_{\odot}$, which is a better match to the shallower GC survey in Virgo using SDSS data. Here, we assume a 1700 kpc virial radius, which corresponds to a virial mass $\sim 5 \times 10^{14} M_{\odot}$ following (Kashibadze, Karachentsev & Karachentseva 2020). We find a good agreement in normalization and slope of our simulated GC catalogue and these observations in Virgo, with the differentiation between blue and red GCs improved with respect to Ramos-Almendares et al. (2020), mostly driven by the improved numerical resolution in our simulations.

While the overall objective is not to reproduce in detail the observations of individual systems, it is reassuring to see that the predictions of our GC tagging method for ICGCs number densities are well in the ballpark of the observations available to date. This is particularly important given that this component is not directly tagged in the simulations, but instead is naturally built by the assembly process of groups and clusters. A more detailed study of the ICGC component and its relation with the build-up of the intracluster light will be presented in future work (Ahvazi et al. in preparation).

4 GC CONTENT IN DWARFS TO GIANT GALAXIES

The GC tagging model calibrates the total mass of GC systems in galaxies at $z = 0$, using the $M_{\text{GC}} - M_{200}$ power-law relation from Harris et al. (2015). As explained in detail in Section 2.3, only simulated haloes with calculated $M_{200} > 10^{11} M_{\odot}$ participate in the calibration, while lower mass objects are assumed to follow an extrapolation of that power law. The GC content of dwarf galaxies in haloes less massive than $M_{200} = 10^{11} M_{\odot}$, corresponding to $M_* \sim 10^9 M_{\odot}$ in the stellar-halo mass relation of TNG50, is therefore a prediction of the model under this assumption. We explore in this section how the results obtained in the regime of dwarf galaxies compare to current observational constraints.

4.1 Number and specific frequency of GCs

Fig. 6 shows in the left-hand panel the relation between the number of GCs, N_{GC} , and the host galaxy stellar mass, M_* , in our simulated systems (grey symbols). Individual galaxies are shown in grey points,

with the median relation (including galaxies with $N_{\text{GC}} = 0$) shown as the solid black line, with the 25–75 per cent dispersion shown as the grey-shaded region. Our results agree well in both overall shape and dispersion with available constraints from observations shown here in coloured symbols: green stars from galaxies in the Virgo cluster (Peng et al. 2008), magenta pentagons for dwarfs in the Fornax cluster (Prole et al. 2019), and additional low-mass galaxies from the Local Volume in sky-blue crosses (Forbes et al. 2018). We also indicate the average results from ELVES-II reported in Carlsten et al. (2022), showing a low-mass selection of Virgo cluster dwarfs (green dot-dashed curve) and dwarfs in the Local Volume (dashed light blue). While our sample does not include low density environments such as the Local Volume, our average values for the lowest mass objects resolved in our sample track well the slope of the average number of GCs per system observed in ELVES-II.

For completeness, we also show the related quantity, specific frequency of GCs or S_N , as a function of V -band magnitude in the right-hand panel of Fig. 6. We calculate the specific frequency following Harris & van den Bergh (1981):

$$S_N = N_{\text{GC}} 10^{0.4(M_V + 15)}, \quad (6)$$

where N_{GC} is number of GCs and M_V is the V -band absolute magnitude of the host galaxy. For most galaxies, we take M_V directly from the simulation, except for the high-mass galaxies ($M_* > 10^9 M_\odot$), where we adopt a fixed mass-to-light ratio equal to 3.6 to convert from mass to luminosity following observations in the Virgo cluster (Peng et al. 2008).

The colour coding on the right-hand panel of Fig. 6 is the same as introduced for the left-hand panel, with our simulated galaxies shown in grey and a set of available observational constraints using colour symbols with error bars. We show galaxies with $S_N = 0$ as grey squares with $S_N = 0.07$ so that they are visible on the log scale. The median and 25–75 percentiles are calculating not including galaxies with $S_N = 0$.

Simulated S_N values overlap well with observational constraints, in particular in the regime of dwarf galaxies, where typical S_N values of several dozens to a few hundreds become common for dwarfs fainter than $M_V \sim -13$. The inclined lines seen for simulated galaxies with $M_V > -16$ correspond to discrete numbers of GCs (galaxies with 1, 2, 3 GCs) and seem to represent well several of the dwarf galaxies in the Forbes et al. (2018) sample.

While, to a certain degree, the agreement in the high-mass end of Fig. 6 might be expected because of the calibration of our model to follow the $M_{\text{GC}}-M_{200}$ relation, it is not fully guaranteed due to the following factors: (i) our method tags the satellite population at infall and not at $z = 0$, (ii) we tag based on halo mass and not M_* as shown here where galaxies continue to evolve their M_* and M_V after infall, and (iii) we tag on total GC mass, M_{GC} , not specifically in GC number. Most importantly, our simulations compare well with measurement of GC numbers in dwarf systems below those used to calibrate the $M_{\text{GC}}-M_{200}$ relation, offering support to the hypothesis that this power-law relation between GC mass and halo mass extends at least to objects with $M_* \sim 5 \times 10^6 M_\odot$.

Interestingly, the left-hand panel of Fig. 6 shows that the average number of GCs continues to decrease with smaller M_* in the full range explored here (when including zeros). This is relevant because it helps rule out more extreme, ‘purely stochastic’ models where the number of GCs is simply a random number in the low-mass end (e.g. El-Badry et al. 2019). We note that this purely stochastic model is not proposed as physically motivated, but instead used in El-Badry et al. (2019) as an interesting extreme behaviour to explore the slope of the relation between halo and GC mass. Such purely stochastic models,

while being able to reproduce the high-mass end of the power-law relation $M_{\text{GC}}-M_{200}$ due to mergers and hierarchical assembly, would provide a much shallower or constant average number of GCs with M_* in the low-mass regime where stochasticity starts to dominate. Instead, our results agree well with the conclusions presented in Forbes et al. (2018), where the slope and scatter of the GC content is consistent with a model where dwarf haloes lay on an extrapolation of the GC mass–halo mass relation measured for more massive systems.

4.2 Radial extent of GCs

We show in Fig. 7 our predictions for the (3D) radial extent of the tagged GC systems as a function of stellar mass. We use the half-number radius $r_{h, \text{GC}}$ to characterize the radial extent of the GC systems, which we calculate by rank-ordering the GCs associated with each galaxy in increasing distance to their host and finding the radius of the GC that divides the sample in two. It is expected that the accuracy of this estimate scales with the number of GCs, with dwarf galaxies having the largest uncertainties given their low number of GCs. In this figure, we include only simulated galaxies with $N_{\text{GC}} \geq 3$ (grey circles), which allows for the determination of $r_{h, \text{GC}}$ (this cut in N_{GC} might not necessarily apply in observations (purple pentagons and green triangles), where the half number radius is determined via profile-fitting, see below). Projected sizes in observations have been converted to 3D by multiplying the reported values by a (4/3) factor, which assumes a spherical distribution (Somerville et al. 2018).

Given the relatively high spatial resolution of TNG50 (~ 290 pc at $z = 0$), the radial extents of the GC systems considered here are numerically well resolved. Their typical sizes increase from a few kpc for dwarfs with $M_* \sim 10^7 M_\odot$ to $r_{h, \text{GC}} \sim 40$ kpc for our largest satellite galaxy with $M_* \sim 10^{11} M_\odot$, with a significant object-to-object scatter, in particular at the low-mass end. The median trend is highlighted by the black solid line, with shaded regions indicating the 25–75 percentiles in our sample. In agreement with observations, simulated GCs are typically more extended than the stellar component in galaxies, which is indicated by the grey-green shaded curve and shaded area showing the median and 25–75 percentiles of the half-mass radius of the stars in the same galaxies. On average, GCs are a factor of ~ 2 –3 times more extended than the stars in galaxies, with a hint at a smaller ratio for low-mass galaxies.

We reproduce well typical sizes for GC systems in MW-mass galaxies, predicting $r_{h, \text{GC}} \sim 10$ kpc for galaxies with $M_* = 5 \times 10^{10} M_\odot$, and an increasing size with mass, in good agreement with data from Hudson & Robison (2018). This is not completely surprising since the scale parameters in the Hernquist profiles used to tag the red and blue GC components at infall were partially chosen in the original model (see Ramos-Almendares et al. 2020) to reproduce typical GC distributions in these scales.

It is interesting, however, to explore what predictions arise from extrapolating the same scaling towards low-mass galaxies. We show with purple pentagons data from dwarfs in the Fornax cluster, taken from Prole et al. (2019). While our systems overlap with the observed dwarfs, simulated galaxies seem to have systematically larger half-number radii than observations. However, we caution that the exact size measured is very sensitive to the definition chosen in systems dominated by low-number statistics, like GCs in dwarfs.

The bottom panel of Fig. 7 shows a different approach, often used in observations of dwarf galaxies: determining the size of GCs based on profile-fitting of the resulting stacked GC profile (instead of individual GC counting in each galaxy as in the upper panel). We show the stacked projected number density profile of GCs for dwarf galaxies with $5 \times 10^6 < M/M_\odot < 10^{8.5}$ as a function of (projected)

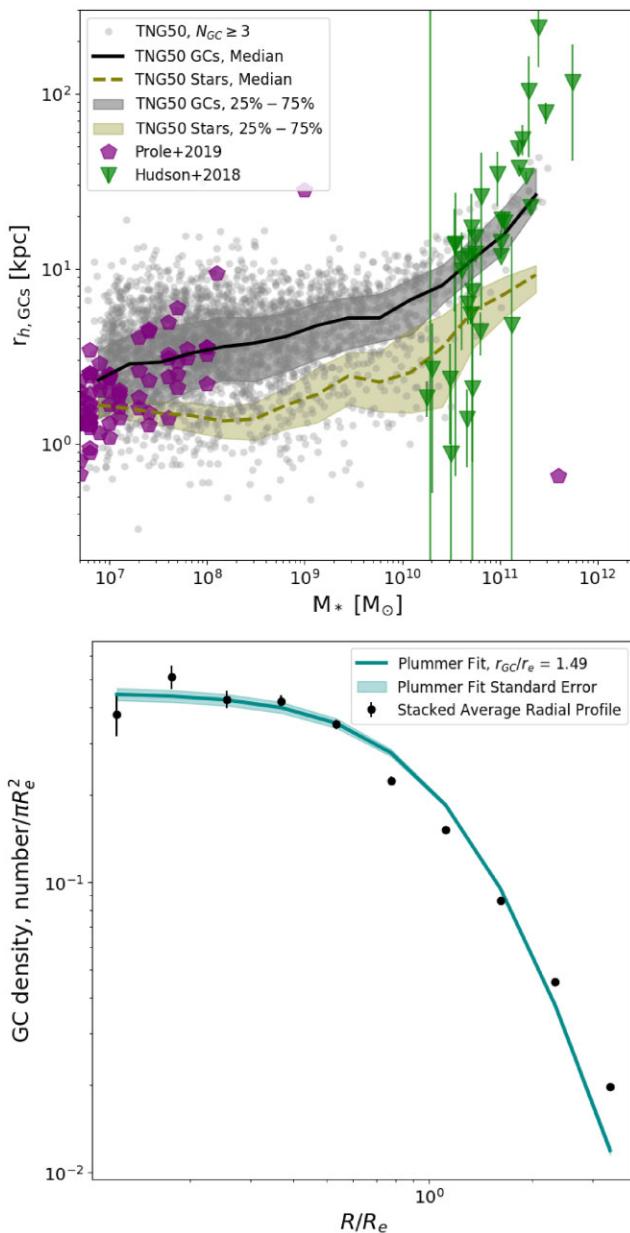


Figure 7. *Top:* The half number radius of GCs around TNG50 galaxies (grey points), low surface brightness galaxies from Prole et al. (2019) (purple pentagons), and higher-mass galaxies from Hudson & Robison (2018) (green triangles). We show simulated galaxies with $N_{GC} \geq 3$. We find good agreement with observations for higher mass galaxies, but find a flatter slope than the observations for dwarf mass galaxies. We do note however that the scatter in the simulated points covers the range seen in observations. *Bottom:* Low number of GCs in dwarfs may favour using alternative methods to measure a half-number radii than individual counts. Inspired by observational methods of calculating R_h, GCs in dwarfs, we show the average *stacked* radial profiles for the GCs of dwarf galaxies with stellar masses between $5 \times 10^6 M_\odot$ and $10^{8.5} M_\odot$. The solid dark cyan line shows the best fitting Plummer profile for the stacked GCs. The error bars and the shaded region are obtained via bootstrapping. Our best-fitting profile recovers the expected factor of ~ 1.5 that relates the stellar half light radius, R_e , with the GC half number radius, R_{GC} , which is consistent with observational estimates.

radius normalized to the effective radius for each dwarf (see for instance Carlsten et al. 2022). Errorbars and the shaded region are calculated via bootstrapping and correspond to the rms values from those realizations. The dark cyan line shows the best-fitting Plummer profile for our simulated dwarfs, which suggests that the half number radius of GCs in these systems is ~ 1.5 times the half-mass radius,² which is in good agreement with observational estimates (Georgiev et al. 2010; Carlsten et al. 2022).

Since in our model, the scaling of the GC tagging depends only on the dark matter half-mass radius (through the calculation of the best-fitting NFW profile at infall), the good agreement with the scaling of GCs and the *stellar* component of galaxies is, again, not guaranteed and an interesting feature of our catalogue. It also points to another puzzling link between GCs and dark matter haloes, in this case through radial extent instead of total mass that may shed light on the origin and formation of GCs.

4.3 Dependence on environment

Recently, Carlsten et al. (2022) reported a higher GC content for dwarf galaxies in the environment of Virgo compared to dwarf satellites of the same mass in lower density environments of the Local Volume. This finding has been interpreted as an extension of a radial trend in the Virgo and Coma cluster where dwarf galaxies near the centre (and therefore on higher density regions) have on average a larger specific frequency compared to those located further out (Peng et al. 2008; Lim et al. 2018). Such a trend has been explained as a natural consequence of dwarfs with inner orbits and higher environmental densities having formed their stars earlier on, with more intense star formation histories leading to the formation of GCs with a higher specific frequency (Peng et al. 2008; Mistani et al. 2016) than objects in the field.

Additionally, since dwarfs stopped forming stars in high density environments earlier than those in the field, comparing them at fixed M_* today means that the dark matter haloes of those in high-density environments are biased high. This follows since quiescent dwarfs today should have continued forming stars reaching higher luminosities at the present day had they stayed in the field (Mistani et al. 2016). Such an effect would also lead to a higher GC content for early type dwarfs in groups and clusters.

Our GC catalogues sample a relatively narrow range of environments, including groups and low-mass clusters with $M_{200} = [5 \times 10^{12} - 2 \times 10^{14}] M_\odot$ and no dwarfs around MW-type galaxies (such as those in the low density regions of ELVES) or directly in the field. However, we have explicitly checked that, within the range of environments of our sample, we find no significant difference in the predicted GC number or specific frequency for simulated dwarfs in low-mass versus high-mass host haloes, nor do we find a trend with cluster-centric radii.

Our method is unable to link the GCs to the star formation histories (only infall virial mass is used to tag the GCs on to our galaxies). However, the second effect (related to the higher halo mass for dwarf galaxies in higher density environments) is naturally taken into account in our catalogue. We find no significant difference in

²We have explicitly checked that on the high-mass end, provided that the chosen profile provides a good fit to the individual GC distributions, the half mass radius computed via GC counting or via profile fitting are within statistical uncertainty of each other, being therefore less of an issue for massive galaxies with a numerous GC population than in low mass dwarfs with only a few GCs.

the infall mass or infall times of the surviving dwarf population between our simulated groups, which partially explains the lack of correlation between N_{GC} or S_N with environment seen in our sample (see Appendix C).

Noteworthy, in agreement with our predictions, dwarfs in the Fornax cluster also show no enhancement in GC number of specific frequency (Prole et al. 2019) when compared to dwarfs in the Georgiev et al. (2010) field sample. This might suggest that while the mode of star formation and differences in halo masses may imprint an excess of GCs for dwarfs in higher density environments, those effects set in at higher density environments (closer to those of Virgo and Coma clusters, $M_{200} > 5 \times 10^{14} M_{\odot}$) than those simulated here. Surveys of dwarfs in intermediate-mass groups and low-mass clusters are needed to confirm this hypothesis and determine whether or not our GC tagging model might benefit in the future from including additional GC formation channels. For instance, an increased number of GCs forming in starburst events associated with pericenter passages have been shown successful at explaining a cluster-centric radial gradient in GC content for cluster dwarf galaxies (e.g. Mistani et al. 2016) and the higher GC content in ultra-diffuse galaxies (e.g. Carleton et al. 2021).

5 GC OCCUPATION FRACTION

While all massive galaxies appear to have associated GCs, the same is not true for low-mass galaxies, some of which are observed to host no GCs. The GC occupation fraction, defined here as the fraction of galaxies at fixed stellar mass that host at least 1 GC, is an important constraint on GC formation scenarios and is fundamental to determine the minimum galaxy mass able to form GCs that survive until the present day.

As discussed in Section 1, observationally, the GC occupation fraction is found to be close to one for galaxies with stellar masses $M_* \geq 10^9 M_{\odot}$, and to sharply decrease for lower mass galaxies (Sánchez-Janssen et al. 2019; Carlsten et al. 2022; Eadie et al. 2022). An important caveat of these studies is that the low-mass galaxies included are mostly satellite objects, although the host mass varies from the Virgo cluster to satellites of $\sim L^*$ hosts in the Local Volume. We can use our GC catalogue to compare with these observations and to determine the role of tidal stripping in satellite galaxies in establishing such a trend.

Black starred symbols in Fig. 8 show the median $z = 0$ GC occupation fraction in our simulated galaxies as a function of host galaxy stellar mass for GCs with individual masses $m_{\text{GC}} > 10^5 M_{\odot}$ to mimic the brightness cut-off from Sánchez-Janssen et al. (2019) in the Virgo cluster. In agreement with observations, our catalogue predicts a decreasing occupation fraction for dwarfs with $M_* \leq 10^9 M_{\odot}$, while all galaxies more massive than that are expected to host GCs. We find a weak dependence of the occupation fraction with cluster-centric radius, with occupation fraction being only slightly lower when considering satellites in the inner regions of simulated groups and clusters ($r < r_{200}/2$, solid black stars, short dashed curve) compared to including all satellites within the virial radius (open black stars, dotted curve). Shaded regions indicate 25–75 percentiles of our sample.

Encouragingly, our present-day occupation fraction agrees well with available measurements. For instance, green-shaded area corresponds to galaxies within $\sim R_{200}/3 - R_{200}/2$ of the Virgo cluster (Sánchez-Janssen et al. 2019), while the occupation fraction in satellite dwarfs within the Local Volume is shown in magenta (Carlsten et al. 2022). We find little variation in occupation fraction with environment across the 39 simulated groups in TNG50, which

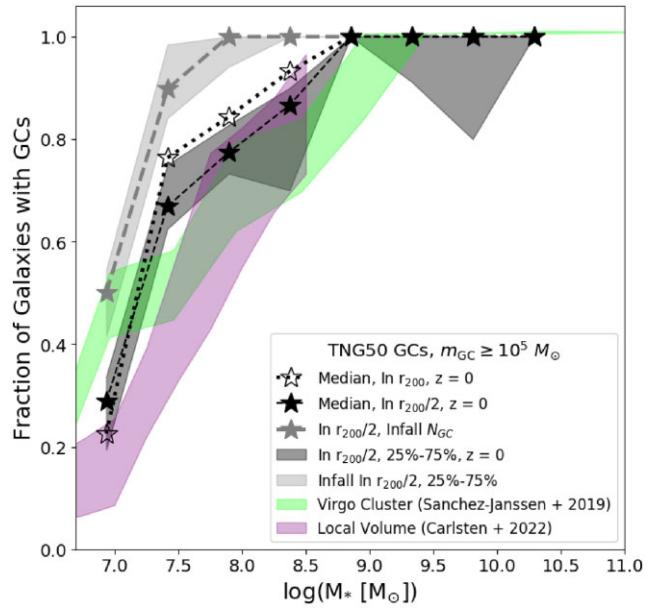


Figure 8. A look at GC occupation fraction (defined as the fraction of galaxies that have at least one GC associated to them) as a function of stellar mass for galaxies within $r_{200}/2$ of our selected groups. The figure shows the occupation fraction from TNG50 by infall number of GCs, as the grey dotted line with stars (with the grey shaded region showing the 25–75 per cent spread between environments); $z = 0$ values are shown within r_{200} by the unfilled black stars and dotted line and within $r_{200}/2$ as the filled black stars and dashed line (with the 25–75 per cent spread between environments shown as the black shaded region). Observed occupation fractions from Virgo (Sánchez-Janssen et al. 2019) and the Local Volume (Carlsten et al. 2022) are shown as lime green and magenta shaded regions, respectively. The difference between the dim grey and the black filled stars shows that tidal stripping has a sizable effect on setting the occupation fraction in dwarfs with $M_* < 10^9 M_{\odot}$. Lower-mass dwarfs with $M_* < 10^{7.5} M_{\odot}$ have additionally a 50 per cent occupation fraction already at infall, which we explain through their low total GC mass together with the stochastic sample of the GC mass function.

agrees well with findings reported in the two environments explored by the ELVES survey (Carlsten et al. 2022).

For comparison, we show in grey the ‘initial’ occupation fraction, e.g. the occupation fraction measured at the infall time for our simulated dwarfs within the $r_{200}/2$ sample (grey starred symbols and dashed curve for the median, shading indicating 25–75 percentiles). Differences between the grey curve at infall and the black curve today is a direct measure of the impact of tidal stripping of GCs by the host groups and clusters, which seems to be substantial for dwarfs with $10^{7.5} < M_*/M_{\odot} < 10^9$.

In particular, our model predicts that all dwarfs with $M_* \sim 10^8 M_{\odot}$ should host at least one GC with mass $\sim 10^5 M_{\odot}$ in the field, while such dwarfs have only 75 per cent occupation fraction on average when observed in groups and clusters. This is a testable prediction that might be confirmed or refuted when large observational samples of field dwarfs with their GCs become available.

On the other hand, for stellar masses lower than $M_* \sim 10^8 M_{\odot}$, the prediction for the infall GC occupation fraction is already lower than 1. For instance, our model predicts that only half of the dwarfs with $M_* \sim 10^7 M_{\odot}$ hosted at least one GC with $M_* \geq 10^5 M_{\odot}$ at infall.

Dwarfs with $M_* \sim 10^7 M_{\odot}$ have a ‘halo mass’ $M_{200} \sim 7 \times 10^9 M_{\odot}$ [calculated following abundance matching from Hudson et al. (2015) as described in Section 2.3] and Fig. 2 shows that for such objects the median mass in GCs is $M_{\text{GC}} \sim 2.6 \times 10^5 M_{\odot}$. This seems

above our GC mass $= 10^5 M_\odot$ considered for this occupation fraction calculation, raising the question of why the occupation fraction is lower than 1 at infall.

We find that the scatter around the $M_{\text{GC}} - M_{200}$ relation coupled to the stochastic sampling of the GC mass function (see Section 2.4) makes the chances for dwarfs of this mass to host a GC with $M = 10^5 M_\odot$ about half. Indeed, the maximum GC mass in our model is set to be one hundredth of the mass of the dwarf (limited inspired by observations of dwarfs in the Local Group), placing a GC with mass $10^5 M_\odot$ close to the upper limit of the mass distribution and therefore relatively unlikely from a random normal draw (see purple histogram on the upper panel of Fig. 3 for the typical mass function of GCs in this mass range).

We have explicitly checked that removing the $0.01M_*$ cut for the sampling of GCs increases the occupation fraction slightly on the low-mass end while leaving it unchanged for $M_* > 10^8 M_\odot$. For instance, in our lowest mass bin the occupation fraction increases by a factor of ~ 2 as a result of a less restrictive mass distribution from which to draw the individual cluster masses. Given the observational uncertainties and variations between the Virgo and Local Volume measurements, our predictions for $M_* \sim 10^7 M_\odot$ remain consistent with observations.

While stochasticity explains the initial low occupation fraction, we note that at present day, the occupation fraction has additionally dropped to 25 per cent for dwarfs with $M_* \sim 10^7 M_\odot$, which is due, similarly to more massive satellites, to tidal stripping from the host. This value is in good agreement with results from the Local Volume, but it is slightly lower than that measured for the Virgo cluster. Observations of dwarfs in the field for this mass range might also help constrain if our model is initially underpredicting the occupation fraction.

Another possibility is that projection effects in high-density environments such as the Virgo cluster could artificially be increasing the occupation fraction of low-mass dwarfs by assigning GCs from the intracluster component or from neighboring galaxies to these dwarfs. Occupation fraction being a requirement of only 1 GC is certainly subject to significant Poisson noise, which worsens in environments with a high background component such as clusters. We will use our catalogue to explore projection effects in future work.

We highlight that the numbers presented in this section should be taken as upper limits assuming no additional GC destruction mechanism is at play after the tagging time at infall. This might not always apply, in particular in cases where dynamical friction timescales might be short, for instance, for low-mass galaxies. We show in Appendix B that considering the effects of dynamical friction does not significantly change our results. We conclude that the occupation fraction predicted by our model is in reasonable agreement with current observational constraints and that additional data from other environments, and more specifically, from the field, would help verify (or reject) the predictions of our model.

6 SUMMARY

In this work, we present a catalogue of GCs tagged to the 39 most massive groups in the TNG50 simulation. Our systems have virial masses in the range $M_{200} = [5 \times 10^{12} - 2 \times 10^{14}] M_\odot$ providing simulated analogues of massive ellipticals in the field to low-mass galaxy clusters. Known systems in this range may include Cen A, Fornax, Hydra-I, or the Virgo cluster, where GC data are abundant. Our GC tagging technique follows from the one already applied to galaxy clusters with $M_{200} \geq 10^{14} M_\odot$ in the Illustris simulation (Ramos-Almendares et al. 2020), with improvements to take full

advantage of the increase in resolution and the inclusion of lower mass dwarfs in our sample.

Briefly, our GCs are tagged to any satellite galaxy identified in the merger tree to have a maximum stellar mass $M_{*,\text{max}} \geq 5 \times 10^6 M_\odot$ and that has ever interacted with our host groups. For each satellite, we identify dark matter particles in its subhalo at infall with a given energy distribution that is consistent with the phase-space that we choose for the GC systems. All galaxies are tagged at infall, after which the dynamics of their assigned GCs is followed by the simulation until the present day. This enables the prediction of GC content in galaxies from dwarfs to giant ellipticals with stellar masses in the range: $M_* = [5 \times 10^6 - 6 \times 10^{11}] M_\odot$. GCs are tagged to more than 8000 simulated galaxies across time, of which more than 5000 survive in our sample at $z = 0$.

We include a new modeling of the GC mass function that allows us to assign individual GC mass to each tagged particle. This is a necessary improvement over the previous model in Ramos-Almendares et al. (2020) that assigns all tagged particles equal GC mass. As discussed in Sections 2.4 and 4.1, this addition is fundamental to reproducing the GC content in dwarf galaxies.

The GC tagging method relies on only one strong assumption: galaxies at infall follow a power-law relation between mass in GCs and halo mass, with a normalization and slope that is calibrated to reproduce the present-day $M_{\text{GC}} - M_{200}$ relation from Harris et al. (2015). Most importantly, this relation is known to hold only for galaxies with stellar mass $M_* \sim 10^9 M_\odot$ and above. We therefore consider only galaxies with halo mass $M_{200} > 10^{11} M_\odot$ (or equivalently, $M_* \sim 10^9 M_\odot$) to participate in the calibration, while applying the calibrated relation to lower mass galaxies as well. This approach allows us to make predictions on the GC systems of dwarfs with $M_* < 10^9 M_\odot$ under the assumption that they follow an extrapolation of the same power law of more massive systems. In this paper, we compare these predictions with available observational data on GCs of dwarf galaxies. Our main results can be summarized as follows:

(i) The GC tagging method naturally gives rise to the formation of an intracluster GC (ICGC) component which is in good agreement with the currently available data. Our individual GC-mass modelling allows the construction of mock observations of GCs at different brightness/mass cutoffs, which might prove a very useful tool for theory/observation comparison once more ICGC systems are mapped in groups and clusters.

(ii) The predicted number (N_{GC}) and specific frequency (S_N) of GCs in dwarf galaxies with $M_* = [5 \times 10^7 - 10^9] M_\odot$ are consistent with observations of dwarfs in the Local Volume as well as in clusters such as Virgo and Fornax. This provides support to the idea that low-mass dwarfs lay in an extrapolation of the GC mass–halo mass relation of more massive counterparts, in agreement with conclusions from Forbes et al. (2018). In particular, the average number of GCs as a function of galaxy mass seems in agreement with that reported for the ELVES survey in low-mass objects (Carlsten et al. 2022) and it is different from one where the number of GCs is simply a random draw in the low-mass end.

(iii) The radial distribution of GCs around satellites in a wide range of masses is also well reproduced in our catalogue, with median values ranging from $r_{h,\text{GC}} \sim 2$ kpc for low-mass dwarfs with $M_* \sim 10^7 M_\odot$ to ~ 25 kpc for $M_* = 2 \times 10^{11} M_\odot$. A closer inspection to the GCs in dwarf galaxies indicates that the low number of GCs expected might bias high the estimates of the half-number radius obtained by simply rank-ordering the identified GCs in distance. When stacking GCs of similar-mass dwarfs and finding a best-fitting profile, as often performed in observations, we find that the half

number radius of GCs in dwarfs is closely related to that of the stars, $r_{h,GC} \sim 1.5r_{h,*}$, which is a common assumption in the literature. This is substantially smaller than the factor of $\sim 3\text{--}5$ between GCs and the size of the stellar component in more massive galaxies like the MW and giant ellipticals.

(iv) We predict a steeply declining GC occupation fraction for dwarfs with $M_* < 10^9 M_\odot$, which is in reasonable agreement with current constraints from Virgo (Sánchez-Janssen et al. 2019) and the Local Volume (Carlsten et al. 2022). In our model, tidal stripping plays a significant role at lowering the occupation fraction for all dwarf galaxies, and this effect cannot be neglected when interpreting occupation fraction data in observations. For instance, we predict almost 100 per cent occupation for dwarfs in the field with $M_* = 10^8 M_\odot$, e.g. hosting at least 1 GC with stellar mass $10^5 M_\odot$, while in group and cluster environments the fraction is ~ 75 per cent, in agreement with observations. For lower mass dwarfs, stochasticity in the sampling of the GC mass function coupled to their low-GC mass content (set by their low halo mass) results in the expectation of only one in two dwarfs with $M_* \sim 10^7 M_\odot$ hosting a $\sim 10^5 M_\odot$ GC at infall. For comparison, tidal stripping effects lower this to one in four for the group and cluster environments.

Our GC tagging method is linked to an empirical calibration of the GC mass–halo mass relation and does not specifically model the formation of GCs. However, some of the results might be used to shed light on GC formation mechanisms. For example, our model naturally predicts the scaling of the size of GC systems across all masses to the dark matter halo distribution (through the half-mass radius in dark matter). The fact that we find a good agreement with observations on the typical GC system sizes from dwarfs to large galaxies suggests another puzzling link between dark matter haloes and GCs, besides the scaling on mass. An interesting link between the GC sizes and the estimated virial radius has been observationally found in galaxies with mass comparable to the MW and above (Hudson & Robison 2018). Our results suggest that a tight link between these two radii extends all the way into the dwarfs regime.

In particular, the GC radial extent in the regime of dwarf galaxies seems in agreement with the predictions from the model where GCs form at the centers of their own dark matter haloes, or mini haloes, as first suggested by Peebles (1984). While this is not true for more massive galaxies, where such a ‘cosmological’ origin of GCs would predict radial distributions that are too extended compared to MW-like galaxies (Creasey et al. 2019), in the regime of dwarfs, the hierarchical clustering of these primordial mini-haloes is of order few kpc, which is in good agreement with observations and predictions of our model (see fig. 4 in Creasey et al. 2019).

This suggests that, if GCs can form in their own mini-haloes and hierarchically assemble in the haloes of galaxies today, the best sites to look for such objects might be dwarf galaxies, where a larger fraction of GCs would be consistent with a cosmological origin. Ultimately, only measurements of individual GC ages and metallicities would be able to fully differentiate between a primordial GC formed in its own dark matter halo, from a GC formed via baryonic processes in the ISM of galaxies (Bastian et al. 2020). Targetting GCs around dwarf galaxies with $M_* \sim 10^7\text{--}10^8 M_\odot$ might give us the best opportunity to narrow down GC origins.

More broadly, the GC catalogue presented in this work is a useful resource to study the 6D properties of GCs in groups and clusters, environments where the ab initio formation of GCs in cosmological simulations is not yet feasible. Targeting ~ 40 systems allows the study of halo-to-halo variations and the understanding of the link between GC properties and particular assembly history of each group; a goal that we will pursue in future work. The GC catalogue

created herein is made publicly available as part of the IllustrisTNG data release.

ACKNOWLEDGEMENTS

We would like to thank the referee, Oleg Gnedin, for a constructive and useful report that helped improved an earlier version of this manuscript. The authors would like to thank Chervin Laporte and Felipe Ramos-Almendares for useful discussions in the early phases of this project and Volker Springel for guidance and early access to TNG50 results. JED and LVS are grateful for financial support from the NSF-CAREER-1945310 and NASA ATP-80NSSC20K0566 grants. JED also acknowledges support to the DAAD through their short term research grant and hospitality from research and administrative staff members at the Max Planck for Astrophysics during a 3-month exchange visit. DN acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1). PT acknowledges support from NSF AST-1909933, NSF AST-2008490, and NASA ATP grant 80NSSC20K0502. MV acknowledges support through NASA ATP 19-ATP19-0019, 19-ATP19-0020, 19-ATP19-0167, and NSF grants AST-1814053, AST-1814259, AST-1909831, AST-2007355, and AST-2107724. RW is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference CITA 490888-16. This research was supported in part by the National Science Foundation under grant no. NSF PHY-1748958.

DATA AVAILABILITY

The realistic GC catalogues are available to the public. They can be downloaded here (www.tng-project.org/doppel22) or in the TNG50 public data release (www.tng-project.org/data) upon acceptance of this paper. The catalogue of all GC candidates or additional realizations of the realistic GC catalogues will also be made available upon request.

REFERENCES

Alamo-Martínez K. A., Blakeslee J. P., 2017, *ApJ*, 849, 6
 Bassino L. P., Cellone S. A., Forte J. C., Dirsch B., 2003, *A&A*, 399, 489
 Bastian N., Pfeffer J., Kruijssen J. M. D., Crain R. A., Trujillo-Gomez S., Reina-Campos M., 2020, *MNRAS*, 498, 1050
 Bekki K., Chiba M., 2002, *ApJ*, 566, 245
 Benavides J. A., Sales L. V., Abadi M. G., 2020, *MNRAS*, 498, 3852
 Binney J., Tremaine S., 2008, *Galactic Dynamics: Second Edition*. Princeton University Press, Princeton, New Jersey
 Blakeslee J. P., Tonry J. L., Metzger M. R., 1997, *AJ*, 114, 482
 Boylan-Kolchin M., 2017, *MNRAS*, 472, 3120
 Bullock J. S., Johnston K. V., 2005, *ApJ*, 635, 931
 Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, *MNRAS*, 321, 559
 Burkert A., Forbes D. A., 2020, *ApJ*, 159, 56
 Carleton T., Guo Y., Munshi F., Tremmel M., Wright A., 2021, *MNRAS*, 502, 398
 Carlsten S. G., Greene J. E., Beaton R. L., Greco J. P., 2022, *ApJ*, 927, 44
 Chen Y., Gnedin O. Y., 2022, *MNRAS*, 514, 4736
 Choksi N., Gnedin O. Y., 2019, *MNRAS*, 488, 5409
 Cooper A. P. et al., 2010, *MNRAS*, 406, 744
 Creasey P., Sales L. V., Peng E. W., Sameie O., 2019, *MNRAS*, 482, 219
 Diemand J., Madau P., Moore B., 2005, *MNRAS*, 364, 367
 Donnari M. et al., 2021a, *MNRAS*, 500, 4004
 Donnari M., Pillepich A., Nelson D., Marinacci F., Vogelsberger M., Hernquist L., 2021b, *MNRAS*, 506, 4760

Doppel J. E., Sales L. V., Navarro J. F., Abadi M. G., Peng E. W., Toloba E., Ramos-Almendares F., 2021, *MNRAS*, 502, 1661

Drinkwater M. J., Gregg M. D., Colless M., 2001, *ApJ*, 548, L139

Durrell P. R. et al., 2014, *ApJ*, 794, 103

Eadie G. M., Harris W. E., Springfield A., 2022, *ApJ*, 926, 162

El-Badry K., Quataert E., Weisz D. R., Choksi N., Boylan-Kolchin M., 2019, *MNRAS*, 482, 4528

Elias L. M., Sales L. V., Creasey P., Cooper M. C., Bullock J. S., Rich R. M., Hernquist L., 2018, *MNRAS*, 479, 4004

Elmegreen B. G., 2017, *ApJ*, 836, 80

Engler C. et al., 2021a, *MNRAS*, 500, 3957

Engler C. et al., 2021b, *MNRAS*, 507, 4211

Errani R. et al., 2022, *MNRAS*, 514, 3532

Forbes D. A., Read J. I., Gieles M., Collins M. L. M., 2018, *MNRAS*, 481, 5592

Genel S. et al., 2014, *MNRAS*, 445, 175

Georgiev I. Y., Puzia T. H., Goudfrooij P., Hilker M., 2010, *MNRAS*, 406, 1967

Harris W. E., Harris G. L. H., Alessi M., 2013, *ApJ*, 772, 82

Harris W. E., Harris G. L., Hudson M. J., 2015, *ApJ*, 806, 36

Harris W. E., van den Bergh S., 1981, *AJ*, 86, 1627

Hernquist L., 1990, *ApJ*, 356, 359

Hudson M. J. et al., 2015, *MNRAS*, 447, 298

Hudson M. J., Harris G. L., Harris W. E., 2014, *ApJ*, 787, L5

Hudson M. J., Robison B., 2018, *MNRAS*, 477, 3869

Jordán A. et al., 2007, *ApJS*, 171, 101

Joshi G. D., Pillepich A., Nelson D., Marinacci F., Springel V., Rodriguez-Gomez V., Vogelsberger M., Hernquist L., 2020, *MNRAS*, 496, 2673

Joshi G. D., Pillepich A., Nelson D., Zinger E., Marinacci F., Springel V., Vogelsberger M., Hernquist L., 2021, *MNRAS*, 508, 1652

Karachentsev I. D. et al., 2007, *AJ*, 133, 504

Karachentsev I. D., Nasonova O. G., 2010, *MNRAS*, 405, 1075

Kashabidze O. G., Karachentsev I. D., Karachentseva V. E., 2020, *A&A*, 635, A135

Keller B. W., Kruijssen J. M. D., Pfeffer J., Reina-Campos M., Bastian N., Trujillo-Gomez S., Hughes M. E., Crain R. A., 2020, *MNRAS*, 495, 4248

Kim J.-h. et al., 2017, *MNRAS*, 474, 4232

Ko Y. et al., 2017, *ApJ*, 835, 212

Ko Y. et al., 2022, *ApJ*, 931, 120

Kravtsov A. V., Gnedin O. Y., 2005, *ApJ*, 623, 650

Kruijssen J. M. D., 2015, *MNRAS*, 454, 1658

Kruijssen J. M. D., Cooper A. P., 2012, *MNRAS*, 420, 340

Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Bastian N., Icke V., 2012, *MNRAS*, 421, 1927

Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Icke V., 2011, *MNRAS*, 414, 1339

Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2019, *ApJ*, 879, L18

Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., Steinwandel U. P., Moster B. P., 2020, *ApJ*, 891, 2

Laporte C. F. P., White S. D. M., Naab T., Gao L., 2013, *MNRAS*, 435, 901

Larsen S. S., Romanowsky A. J., Brodie J. P., Wasserman A., 2020, *Science*, 370, 970

Lee M. G., Park H. S., Hwang H. S., 2010, *Science*, 328, 334

Li H., Gnedin O. Y., 2014, *ApJ*, 796, 10

Li H., Gnedin O. Y., Gnedin N. Y., Meng X., Semenov V. A., Kravtsov A. V., 2017, *ApJ*, 834, 69

Li H., Vogelsberger M., Bryan G. L., Marinacci F., Sales L. V., Torrey P., 2021, *MNRAS*, 514, 265

Lim S., Peng E. W., Côté P., Sales L. V., den Brok M., Blakeslee J. P., Guhathakurta P., 2018, *ApJ*, 862, 82

Lokas E. L., Mamon G. A., 2001, *MNRAS*, 321, 155

Longobardi A. et al., 2018, *ApJ*, 864, 36

Lotz J. M., Miller B. W., Ferguson H. C., 2004, *ApJ*, 613, 262

Ma X. et al., 2020, *MNRAS*, 493, 4315

Madrid J. P., O'Neill C. R., Gagliano A. T., Marvil J. R., 2018, *ApJ*, 867, 144

Marinacci F. et al., 2018, *MNRAS*, 480, 5113

Martin N. F. et al., 2022, *Nature*, 601, 45

Mistani P. A. et al., 2016, *MNRAS*, 455, 2323

Mowla L. A. et al., 2022, *ApJ*, 937, L35

Naiman J. P. et al., 2018, *MNRAS*, 477, 1206

Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563

Nelson D. et al., 2018, *MNRAS*, 475, 624

Nelson D. et al., 2019a, *Comput. Astrophys. Cosmol.*, 6, 2

Nelson D. et al., 2019b, *MNRAS*, 490, 3234

Peebles P. J. E., 1984, *ApJ*, 277, 470

Peñarrubia J., Navarro J. F., McConnachie A. W., 2008, *ApJ*, 673, 226

Peng E. W. et al., 2008, *ApJ*, 681, 197

Peng E. W. et al., 2011, *ApJ*, 730, 23

Pfeffer J., Kruijssen J. M. D., Crain R. A., Bastian N., 2018, *MNRAS*, 475, 4309

Pillepich A. et al., 2018a, *MNRAS*, 475, 648

Pillepich A. et al., 2018b, *MNRAS*, 475, 648

Pillepich A. et al., 2019, *MNRAS*, 490, 3196

Planck Collaboration XIII, 2016, *A&A*, 594, A13

Prieto J. L., Gnedin O. Y., 2008, *ApJ*, 689, 919

Prole D. J. et al., 2019, *MNRAS*, 484, 4865

Ramos F., Coenda V., Muriel H., Abadi M., 2015, *ApJ*, 806, 242

Ramos-Almendares F., Abadi M., Muriel H., Coenda V., 2018, *ApJ*, 853, 91

Ramos-Almendares F., Sales L. V., Abadi M. G., Doppel J. E., Muriel H., Peng E. W., 2020, *MNRAS*, 493, 5357

Reina-Campos M., Trujillo-Gomez S., Deason A. J., Kruijssen J. M. D., Pfeffer J. L., Crain R. A., Bastian N., Hughes M. E., 2022, *MNRAS*, 513, 3925

Renaud F., Bournaud F., Duc P.-A., 2015, *MNRAS*, 446, 2038

Riggs S. D., Loveday J., Thomas P. A., Pillepich A., Nelson D., Holwerda B. W., 2022, *MNRAS*, 514, 4676

Rodriguez-Gomez V. et al., 2015, *MNRAS*, 449, 49

Sales L. V. et al., 2015, *MNRAS*, 447, L6

Sameie O. et al., 2022, preprint ([arXiv:2204.00638](https://arxiv.org/abs/2204.00638))

Sánchez-Janssen R. et al., 2019, *ApJ*, 878, 18

Schuberth Y., Richtler T., Bassino L., Hilker M., 2008, *A&A*, 477, L9

Somerville R. S. et al., 2018, *MNRAS*, 473, 2714

Spitler L. R., Forbes D. A., 2009, *MNRAS*, 392, L1

Springel V. et al., 2008, *MNRAS*, 391, 1685

Springel V. et al., 2018, *MNRAS*, 475, 676

Tamura T., Makishima K., Fukazawa Y., Ikebe Y., Xu H., 2000, *ApJ*, 535, 602

Taylor M. A., Puzia T. H., Muñoz R. P., Mieske S., Lançon A., Zhang H., Eigenthaler P., Bovill M. S., 2017, *MNRAS*, 469, 3444

Trujillo-Gomez S., Kruijssen J. M. D., Reina-Campos M., Pfeffer J. L., Keller B. W., Crain R. A., Bastian N., Hughes M. E., 2021, *MNRAS*, 503, 31

van den Bergh S., 2000, *AJ*, 119, 609

Villaume A., Foreman-Mackey D., Romanowsky A. J., Brodie J., Strader J., 2020, *ApJ*, 900, 95

Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518

Vogelsberger M. et al., 2014b, *Nature*, 509, 177

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, *Nature Rev. Phys.*, 2, 42

Weinberger R. et al., 2017, *MNRAS*, 465, 3291

Weinmann S. M., Lisker T., Guo Q., Meyer H. T., Janz J., 2011, *MNRAS*, 416, 1197

Yahagi H., Bekki K., 2005, *MNRAS*, 364, L86

Yun K. et al., 2019, *MNRAS*, 483, 1042

APPENDIX A: MASS-TO-LIGHT CALIBRATIONS

Simulated galaxies in TNG50 have stellar masses and corresponding luminosities calculated in several bands, including information on the V -band magnitudes necessary, for example, for computing the specific frequency S_N in Fig. 6. However, the simulated luminosities include only evolution due to stellar population models and might

neglect important effects, such as dust attenuation. We therefore compute the V -band luminosities of our sample by using a mass-to-light ratio calibration fit to the Virgo cluster data (using stellar masses and V -band absolute magnitudes from Peng et al. 2008).

This is shown in Fig. A1, where grey symbols indicate the results directly from the simulations and green stars are data from Virgo. Thin coloured lines indicate constant mass-to-light ratios, as labelled, while the thick black solid line highlights the conversion used in this paper. As expected, the calibrated relation differs from the simulated values mostly at the high-mass end, where dust effects might be playing a more important role. While this correction does not significantly impact any of the results in this paper, considering a mass-to-light ratio equal to 3.6 for more massive galaxies (instead of ~ 2 as suggested by the simulation) improves the agreement with S_N data reported in Section 4.1.

The evolution in mass-to-light ratio and changes in star formation rate once a galaxy becomes a satellite make necessary an additional calibration in our model. This calibration is related to the dispersion in the luminosity (or mass) function of individual GC masses, σ_z , described in Section 2.4 in our main article and shown in Fig. 1 as a function of stellar mass M_* . This z -band luminosity dispersion σ_z is observationally constrained at $z = 0$ as a function of the *present-day* B -band magnitude (Jordán et al. 2007). However, the GC tagging and mass assignment in our model is done at infall (and not present day), requiring of an adjustment at the moment to perform the GC tagging to reproduce the desired results at $z = 0$.

For illustration, we show in Fig. A2 the B -band luminosity evolution in all our galaxies from infall to $z = 0$. To compensate for this evolution, we first calculate the ‘target’ relation between σ_z and stellar mass M_* (shown in orange in the bottom panel of Fig. 3), where M_* is calculated as the median M_* in our simulated galaxies with a

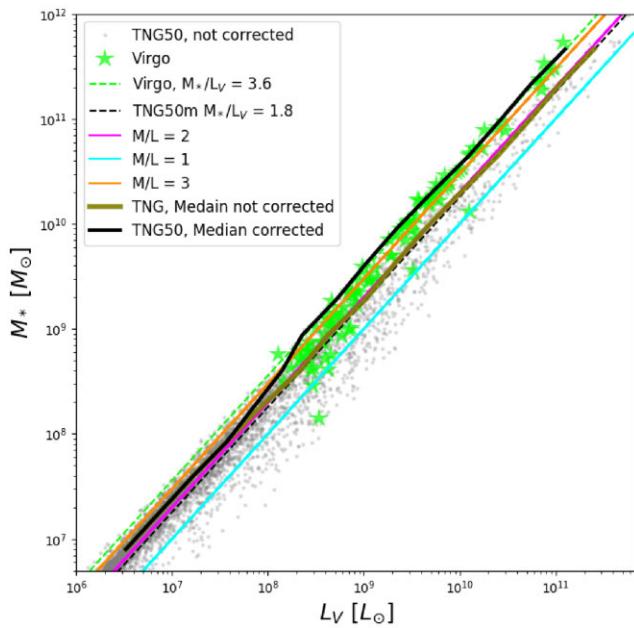


Figure A1. Mass-to-light ratio for TNG50 (grey points that show individual galaxies, with the grey-green line showing the median bins of stellar mass and the black dotted line showing the best fit) compared to that of Virgo galaxies overplotted as lime green stars (with the best fit shown as the dotted lime green line). At fixed stellar mass, Virgo galaxies tend to be less luminous at higher masses than simulated objects. This discrepancy at face value in mass-to-light ratio between TNG50 and Virgo for high-mass galaxies causes a discrepancy in both M_V and S_N for those masses. We therefore adopt a ‘corrected’ mass-to-light ratio (shown in black line) to compute S_N in our results.

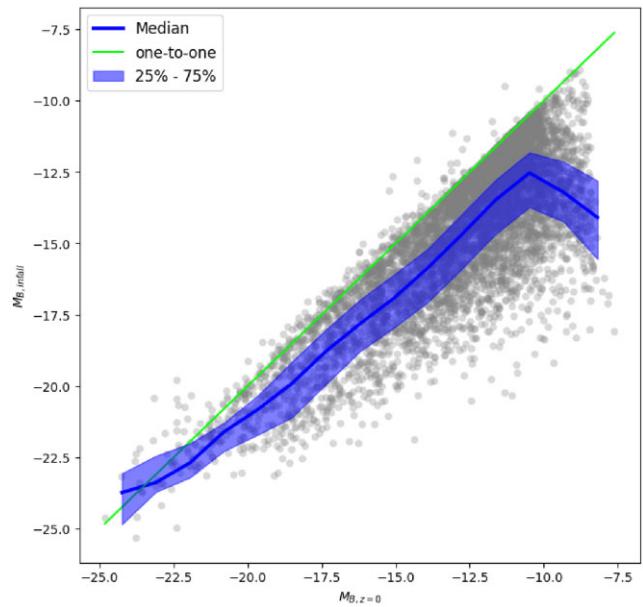


Figure A2. Comparison between galaxy M_B at infall and at $z = 0$. Since observational results are available at $z = 0$ but our tagging occurs at infall, we require a calibration that seeks to take the brighter B -luminosities at infall into account when calculating GC luminosity function.

given B -band luminosity, all at $z = 0$. Next, we correct the initial σ_z (e.g. at infall time) by calculating the σ_z that would correspond to each galaxy assuming their infall stellar mass and then multiply that by a constant factor: $\sigma_{z,\text{inf}} = \alpha * \sigma_z(M_*)$, where $\sigma_z(M_*)$ is our target relation at $z = 0$ as described before. After experimenting with different values, we find $\alpha = 0.75$ a reasonable choice, in particular to reproduce the median σ_z at $z = 0$ observed in low-mass galaxies, which is the main focus of this work.

APPENDIX B: POTENTIAL EFFECTS OF DYNAMICAL FRICTION

Massive objects such as GCs can experience dynamical friction, as they move within the gravitational potential of the smoothly distributed mass in the host galaxy. Our tagging method does not self-consistently follow this effect since we tag them on to the dark matter particles and all components (dark matter, baryons and GCs) have similar particle mass in our simulations. By default, our GC catalogue ignores dynamical friction effects since the method is tailored to tag only the ‘surviving’ population of GCs and not the initial one. However, it is important to double-check that after tagging our GCs they would not be substantially affected by dynamical friction and expected to coalesce to the centre of the galaxies and be dissolved.

To gain some intuition, we estimate analytically the typical timescales for dynamical friction in our systems following equation 8.17 in Binney & Tremaine (2008):

$$t_{\text{fric}} = \frac{2.7 \text{ Gyr}}{\ln \Lambda} \frac{r_i}{30 \text{ kpc}} \left(\frac{\sigma_H}{200 \text{ km/s}} \right)^2 \left(\frac{100 \text{ km/s}}{\sigma_{\text{GC}}} \right)^3,$$

where σ_H is the typical velocity dispersion in the host, σ_{GC} is the velocity dispersion of the GC, both as proxies for mass, r_i is the initial radius of the GC orbit, and $\ln \Lambda = 5.8$ is assumed as a typical Coulomb logarithm. We vary the velocity dispersion of the host assuming $\sigma_H = 800, 200, 50, 20$, and 10 km s^{-1} corresponding roughly to the scales of a cluster, an MW-like galaxy and dwarfs with $M_* \sim 10^9, 10^8$, and $10^{6.5} \text{ M}_\odot$, respectively. For the GCs, we

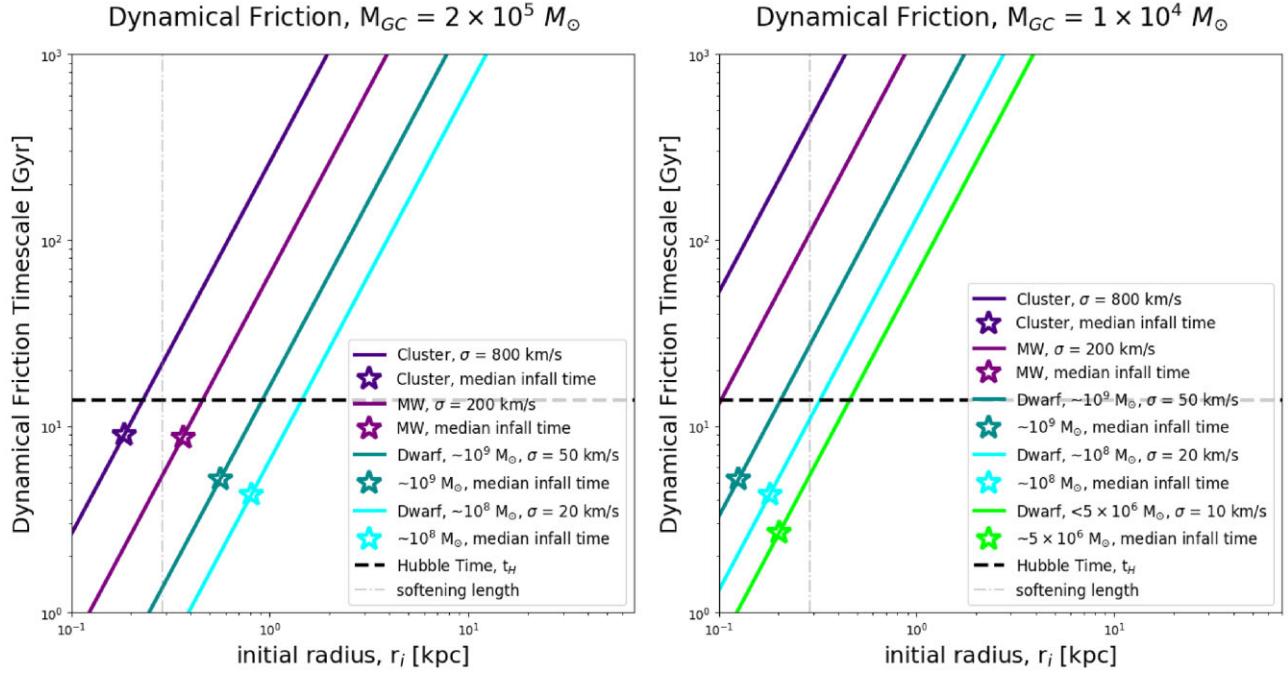


Figure B1. Dynamical friction timescales for different types of GC host systems, assuming the median GC mass of $2 \times 10^5 M_\odot$ (left) and $10^4 M_\odot$ (right). The typical time-scales associated with dynamical friction are longer than the age of the Universe for most initial radii and in particular for high-mass hosts. We also indicate the median infall time for galaxies in each mass range with a starry symbol. For dwarf galaxies, dynamical friction timescales might be lower than a Hubble time only for GCs at very small radii $r < 0.5\text{--}1.0$ kpc, depending on GC mass, but comparable to the time since their infall time, when GCs are tagged. We therefore expect not a significant change in any of the results when including dynamical friction. Notice that our least massive dwarfs do not have GCs as massive as $2 \times 10^5 M_\odot$ and therefore are not included on the left-hand panel.

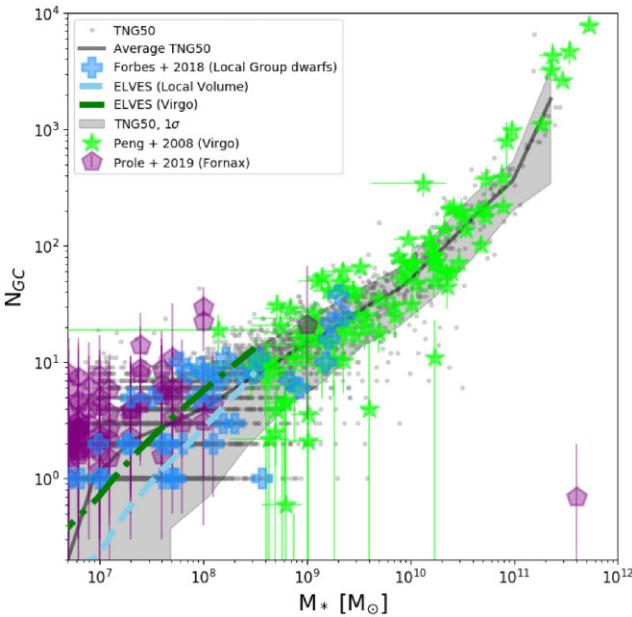


Figure B2. The same as the left-hand panel of Fig. 6, but including an estimation of the removal of GCs by dynamical friction. We see very little change in the overall behaviour of the GC abundances with stellar mass when including dynamical friction.

compare the effects on two scales: a $2 \times 10^5 M_\odot$ (typical GC mass) and $1 \times 10^4 M_\odot$ (our lower limit and common value in low-mass galaxies). We assume a half-mass radius $r_h = 3$ pc to translate GC mass into velocity dispersion σ_{GC} . Finally, we consider the radius

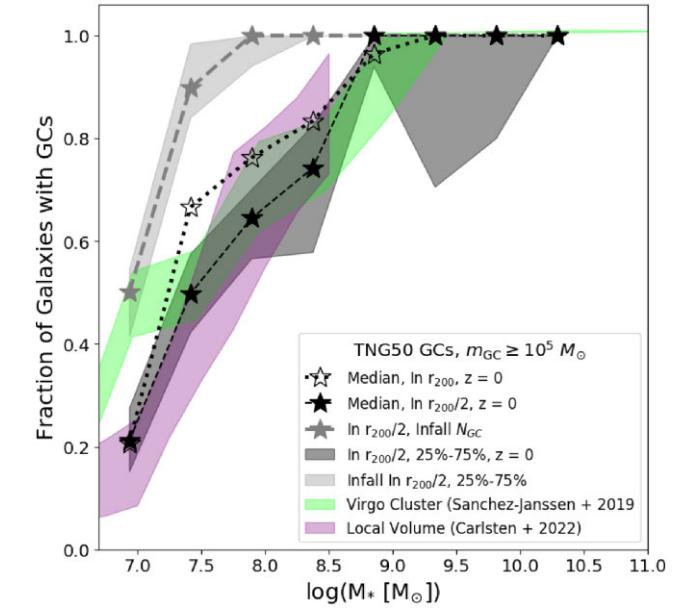


Figure B3. The same as Fig. 8, but including an estimation for removal of GCs by dynamical friction. We see that the dwarf galaxy stellar mass bins that previously sat above observed values fall nicely within the range of observations when including this effect.

r_i as the distance of the GC to the centre of the host at infall (the moment of the tagging).

We show the results in Fig. B1, where the dynamical friction timescales are shown as a function of the distance of the GC. For reference, we indicate the age of the Universe with a thick dashed

horizontal line, areas where t_{fric} is above the Hubble time t_H indicates that dynamical friction effects are unimportant. As expected, the dynamical friction timescales increase with radius, meaning that only GCs in the very inner regions are potentially affected. Fig. B1 also shows that t_{fric} is shorter for more massive GCs, as expected, but even in this case only GCs within ~ 1 kpc have the potential to decay and coalesce due to dynamical friction forces. In the case of a lighter GC, as the one shown on the right-hand panel, the relevant distance where dynamical friction effects might be important shrinks to ~ 0.5 kpc.

Reassuringly, the distances where dynamical friction might be a factor of concern are quite small compared to the typical GCs radial extension (see Fig. 7) and suggest that dynamical friction effects are not important in our sample. Moreover, the time of relevance is not the age of the Universe but the time since infall, when the GC is tagged. Those are highlighted with a starry symbol in Fig. B1 and correspond to the median infall times of galaxies of a given stellar mass in our sample. On average, dynamical friction effects are negligible and if present, may impact only the lowest mass galaxies in the sample.

Next, using the same equation above, we compute a dynamical friction time individually for each tagged GC and comparing t_{fric} to the particular infall time of that host galaxy we can individually assess whether GCs are expected to decay or not. We flag all GCs where $t_{\text{fric}} < (t_H - t_{\text{inf}})$ as ‘merged’, and remove them from our sample at $z = 0$. Figs B2 and B3 show that this would have no significant consequences for our main results, including the number of GCs per galaxy or the occupation fraction, respectively. We therefore conclude that while dynamical friction might impact a few of our GC on an individual basis none of the statistical results presented here changes appreciably. In our released catalogue, we provide a dynamical friction flag to allow the user to decide whether to include these objects or not in their calculations.

APPENDIX C: ENVIRONMENTAL EFFECTS

Inspired by observations of galaxies in higher density environments showing a higher GC abundance (e.g. Peng et al. 2008; Carlsten et al. 2022), we have checked if this phenomenon was present in our tagged GCs catalogue in TNG50. We split the environments in bins of virial mass, and within those bins, we computed the median and 25–75 per cent range of GC abundance in bins of host galaxy stellar mass. Fig. C1 shows the result of this test. There is little if any variation in both the median and the scatter between the different virial mass bins. Running the same check on S_N shows the same lack of dependence with the host.

The GC tagging model employed in this work relies on the infall virial mass of a galaxy; thus we checked to see how infall virial mass varies across the tagged environments in Fig. C2. Binning again in host environment virial mass, we calculated the median infall virial mass in bins of *present-day* host galaxy stellar mass, $M_{*,z=0}$. We find a weak environmental dependence on the infall halo masses at fixed $z = 0$ stellar mass that goes in the direction expected: galaxies of a given stellar mass today had a larger infall virial mass for high-density environments (see e.g. Mistani et al. 2016). However, we had had to include low-mass host haloes with present-day $M_{200} \geq 10^{11} M_\odot$ (which are not included in our catalogue) in order to observe the effect. Limiting the host halo mass to the ones included in this study ($M_{200} > 5 \times 10^{12} M_\odot$) shrinks the effect appreciably, explaining why our GC catalogue shows no significant dependence with environment. Thus, it may be necessary to study a much wider range of host halo masses in order to see the observed environmental dependence on GC abundance.

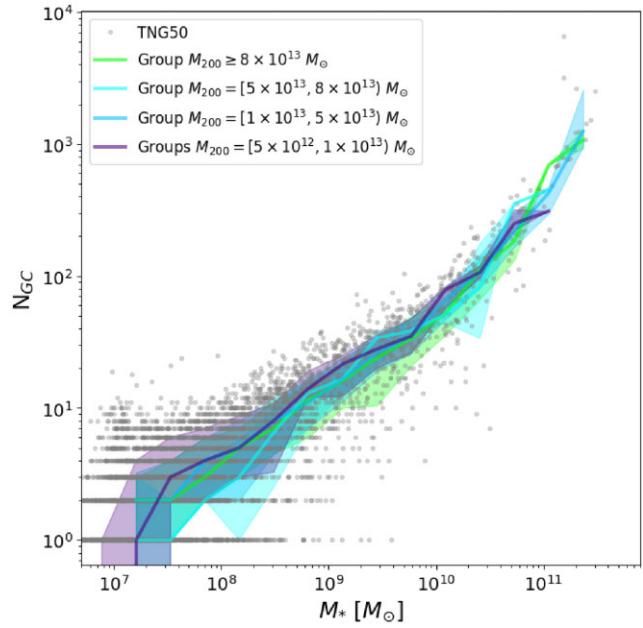


Figure C1. Number of GCs, N_{GC} as a function of host galaxy stellar mass M_* binned in host cluster virial mass. The solid lines show the median in bins of host galaxy stellar mass with the shaded region showing the 25–75 per cent scatter in each bin. We find no pronounced dependence on GC abundance with host group or cluster environment.

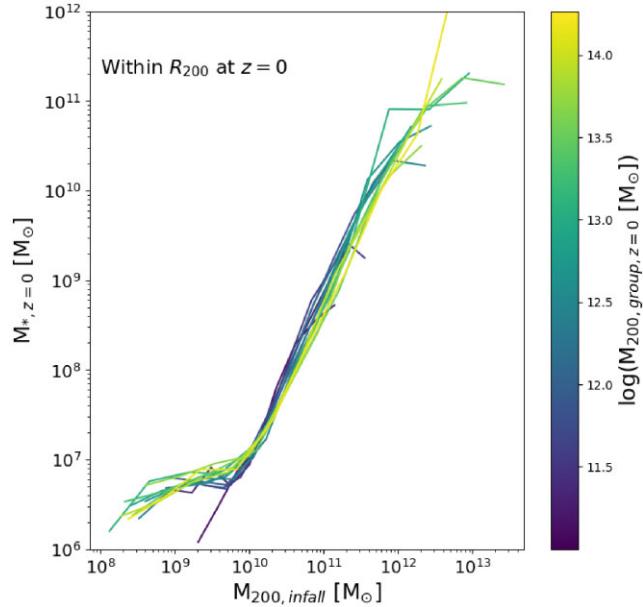


Figure C2. M_* at $z = 0$ as a function of infall virial mass $M_{200,\text{infall}}$ for galaxies within R_{200} of their $z = 0$ host environment in TNG50. Medians are colored by $z = 0$ host virial mass (colour bar on the right). There is a weak but systematic trend for satellites with fixed stellar mass today to have a larger infall virial mass in more massive hosts, in particular for $M_* > 10^7 M_\odot$. Note that we extend the calculation to host virial masses $M_{200} = 10^{11} M_\odot$, which is well below our minimum host halo mass tagged, in order to clearly see the effect. For hosts with $M_{200} > 5 \times 10^{12} M_\odot$, as studied here, there is not enough difference in satellite infall masses to lead to any environmental trend on GC content.