ARE GLOBULAR CLUSTERS THE REMNANT NUCLEI OF PROGENITOR DISK GALAXIES?

TORSTEN BÖKER
European Space Agency, Department RSSD, Keplerlaan 1, 2200 AG Noordwijk, Netherlands; tboeker@rssd.esa.int

Received 2007 August 15; accepted 2007 November 21; published 2007 December 18

ABSTRACT

The globular cluster system of a typical spheroidal galaxy makes up about 0.25% of the total galaxy mass. This is roughly the same mass fraction as contained in the nuclear star cluster (or stellar nucleus) present in most nearby low-mass galaxies. Motivated by this “coincidence,” this Letter discusses a scenario in which globular clusters of present-day galaxies are the surviving nuclei of the dwarf galaxies that—according to the hierarchical merging paradigm of galaxy formation—constitute the “building blocks” of present-day massive galaxies. This scenario, which was first suggested by Freeman, has become more attractive recently in light of studies that demonstrate a complex star formation history in a number of massive globular clusters.

Subject heading: globular clusters: general

1. INTRODUCTION

The origin of globular cluster (GC) systems of present-day galaxies continues to provide a rich topic for investigation, speculation, and debate. Enabled mostly by the spatial resolution and precision photometry of the Hubble Space Telescope (HST), recent studies have established that GCs are not uniform in their properties, but display a rather wide range in color and metallicity. The color distribution of GCs in early-type galaxies nearly always shows a double-peaked structure (Peng et al. 2006), motivating a commonly accepted view of GC systems as “bimodal.” The situation in disk-dominated galaxies is less clear, partly because the analysis is limited by the smaller number of GCs in less massive galaxies. Whether or not bimodality is present in all galaxies, the broad range in GC color makes it clear that GCs within a given galaxy have formed neither coevally nor under the same physical conditions.

Moreover, the ongoing formation of young massive clusters (YMCs) observed in nearby starbursts such as M82 (McCrady & Graham 2007) or NGC 4038/9 (Mengel et al. 2002), as well as the presence of some intermediate-age GCs in, e.g., the Magellanic Clouds (Goudfrooij et al. 2006) or M33 (Chandar et al. 2006), suggest that GC formation was not confined to the early universe, and possibly continues until the present day. Although it seems plausible that individual YMCs can evolve into GC-like structures, it is less clear that entire GC systems can be created in this way. It may well be that nature has more than one method to make GCs.

The recent, comprehensive review of Brodie & Strader (2006) summarizes the proposed scenarios for GC formation. The most widely used are (1) in situ formation through multiphase dissipational collapse (Forbes et al. 1997), (2) massive star formation triggered by major disk-disk mergers (Ashman & Zepf 1992), and (3) dissipationless accretion of GC systems during the assembly of massive galaxies through mergers at high redshift (Côté et al. 1998).

These models all have considerable problems in fully explaining the observed properties of GCs. For example, the multiphase dissipational collapse model invokes two distinct phases of GC formation to explain color bimodality, but cannot readily provide a physical mechanism that would explain the dormancy between them. The major-merger model, on the other hand, fails to account for the lack of intermediate-age GC systems and the systematically higher specific frequency in gE galaxies which are presumed to be the end products of major disk-disk mergers. Finally, the dissipationless accretion scenario (Côté et al. 1998), as well as more recent studies employing semianalytic models of galaxy formation (e.g., Beasley et al. 2002; Rhode et al. 2005), appear able to explain the observed color distributions of GC systems. However, these models do not explain how the GCs of the first galaxies formed, and thus merely shift the problem to earlier times.

Most importantly, none of these scenarios can explain the observed complexity in the star formation history of some GCs. While most “typical” GCs studied so far show no obvious signs of multiple stellar populations (e.g., Sarajedini et al. 2007), the situation is different for the most massive GCs. For example, the high mass and multiple stellar populations of ω Cen have long triggered speculation that it is the remnant nucleus of a satellite galaxy which was destroyed when merging with the Milky Way (Norris et al. 1997; Lee et al. 1999). Moreover, recent observations have established that ω Cen is not a unique case, and that the high-mass end of the Milky Way GC system commonly shows evidence for multiple stellar populations (Moehler & Sweigart 2006; Piotto et al. 2007; Caloi & D’Antona 2007). These results have challenged the traditional view that all GCs are simple stellar populations born in a single, short-lived (“instantaneous”) burst of star formation.

This Letter revisits and expands on an alternative formation scenario for GCs which was first suggested by Freeman (1993). The underlying idea is that GCs form as the nuclei of dwarf galaxies in the early universe, and are accreted over time as their hosts merge into larger structures. The Letter is structured as follows: § 2 briefly describes the Freeman (1993) scenario, and mentions some recent observational developments that are relevant in this context. Section 3 reviews what is known about the formation of NCs in late-type disk galaxies. Section 4 then discusses some quantitative aspects of the proposed scenario and presents a first-order plausibility check. Finally, § 5 concludes with some open issues and predictions.

2. GLOBULAR CLUSTERS AS DWARF NUCLEI

Based on the observed lack of an abundance gradient in the GC system of the Milky Way outer halo, Searle & Zinn (1978) suggested that GCs originate in small “fragments” (i.e., protogalaxies) which are accreted over an extended period of time into a larger host. However, they did not discuss the specifics of where and how the GC precursors form inside the fragments.

Building on this scenario, Freeman (1993) suggested that nucleated dwarf galaxies and their compact stellar cores might
be the present-day analogs of those Searle-Zinn fragments that escaped disruption in a merger event. He also pointed out that chemical inhomogeneities in present-day GCs might be explained by self-enrichment due to the strong stellar winds observed in starbursting galaxy nuclei. The general idea behind this scenario has recently received increased attention (see, e.g., Lee et al. 2007) because it might provide a natural explanation for the observed multiple stellar populations in some GCs, especially when taking into account two other recent observational results:

1. Stellar nucleation is not limited to dE galaxies, but is also common in low-mass disk galaxies, as shown by recent surveys with HST (Carollo et al. 1997; Bóker et al. 2002). These nuclear star clusters (NCs) are structurally very similar to GCs, and nearly always contain multiple stellar populations. They are therefore attractive candidates for the precursors of at least those GCs that have complex star formation histories. In fact, it is difficult to explain how a stellar cluster can accrete (or retain) enough gas for multiple star formation episodes if it is not located at the bottom of a potential well such as the nucleus of a galaxy.

2. The mass of a GC system appears to be proportional to that of its host galaxy (McLaughlin 1999). As discussed further in § 4, this is expected if GCs indeed form as the stellar nuclei of “building blocks” which merge into massive present-day galaxies.

The Freeman (1993) scenario can thus be modified as follows: present-day GCs have formed as NCs in the centers of low-mass disk galaxies in the early universe where they grew in mass through a series of discrete star formation events as long as the disk structure of their host remained undisturbed by any merger activity. This growth process ends when the host galaxy disk is disrupted in a merger or tidal interaction. Due to its large stellar density, the (former) NC survives the merger, but because it is no longer located in a potential well, it experiences no further gas inflow. No further star formation takes place in the NC, and it will simply age passively as a GC in the merger product. It should be noted explicitly that the presence of multiple stellar populations is not a necessary consequence of this process: if a NC was accreted early enough, it might have experienced only a single star formation episode and thus contain only a single generation of stars.

3. NUCLEAR STAR CLUSTERS

Over the last decade, a number of studies—both via imaging and spectroscopic observations—have contributed to the following picture of present-day NCs:

1. NCs are common: the fraction of galaxies with an unambiguous NC detection is 75% in late-type (Scd-Sm) spirals (Bóker et al. 2002), 58% in earlier type (S0-Sbc) spirals (Balcells et al. 2007), and 70% in spheroidal (E and S0) galaxies (Graham & Guzmán 2003; Côté et al. 2006).\(^1\)

2. The structural properties of NCs in late-type disks are indistinguishable from those of GCs, and thus certainly do not rule out an evolutionary connection: typical half-light radii are 2–5 pc (Bóker et al. 2004). The size distribution of NCs in spheroidal galaxies also shows a large overlap with that of Milky Way GCs, although some present-day (d)E nuclei are clearly larger than the “typical” GC (Geha et al. 2002; Côté et al. 2006).

3. Despite their compactness, present-day NCs are rather massive: their typical dynamical mass is \(10^6–10^7 M_\odot\) (Walcher et al. 2005), which is at the high end of the GC mass function. Because of their very high stellar densities, there is little doubt that NCs are difficult to disrupt, and that they will remain structurally intact even after the disruption of their host galaxy.

4. The star formation history of NCs in disk galaxies is complex, as evidenced by the fact that they have stellar populations composed of multiple generations of stars (Walcher et al. 2006; Rossa et al. 2006). In the NCs of late-type disks, the youngest stellar generation is nearly always younger than 100 Myr, which explains why present-day NCs are much more luminous than GCs (Bóker et al. 2002; Côté et al. 2006). Nevertheless, NCs always contain an old (~Gyr scale) stellar population, suggesting that they have been in place for a long time.

Taken together, these facts suggest that NCs have built up the bulk of their present-day mass through a series of star formation episodes (Walcher et al. 2006). This repetitive “rejuvenation” can only occur as long as gas is funneled toward the nucleus, and thus no longer takes place after the NC has been removed from the galaxy nucleus or after the gas supply has otherwise been disrupted, e.g., through transformation of a late-type disk into a dE galaxy.

Three recent and independent studies of NCs in different galaxy types (Rossa et al. 2006; Wehner & Harris 2006; Côté et al. 2006) have established that NCs obey similar scaling relationships with host galaxy properties as do supermassive black holes. This has led Ferrarese et al. (2006) to suggest that NCs and black holes might constitute two different incarnations of a “compact massive object,” the formation of which is inherent to galaxy assembly, and which contains ~0.2% of the host galaxy (bulge) mass. Although Balcells et al. (2007) have questioned the simple linearity, they confirm that any NC typically contains 0.1%–0.3% of the bulge mass.

A possible theoretical explanation for a similar proportionality also in truly bulgeless disk galaxies has recently been offered by Emsellem & van de Ven (2008) who point out the compressive nature of tidal forces in a shallow gravitational potential, e.g., in an exponential disk. Based on their simulations, a NC formed through this mechanism should contain ~0.1%–0.5% of the host galaxy mass. In this model, matter infall onto the nucleus occurs naturally in bulgeless disks, possibly explaining why NCs with recent star formation have only been found in late-type spirals.

The following section investigates the possibility that many, if not all, GCs are the surviving nuclei of dwarf galaxies that were “swallowed” during the assembly of present-day galaxies.

4. A SIMPLISTIC SCENARIO

A prerequisite for the proposed scenario is that NCs are present already in the first galaxies, i.e., the “building blocks” for galaxy assembly. This is not an implausible proposition. For example, Cen (2001) has suggested that the reionization of the universe produces inward, convergent shocks in dark matter “minihalos.” These shocks compress the baryonic gas which in turn becomes self-gravitating and undergoes star formation, forming a dense stellar cluster in the process. These “seed clusters” naturally define a center of gravity for the assembly of a gas disk. As Cen (2001) points out, the NCs of high-redshift minihalos formed through this process are plausible progenitors of present-day GCs (see also Moore et al. 2006).

The subsequent evolution of these “primordial” NCs in their
parent halos and the expected tidal fields has been studied in detail by Mashchenko & Sills (2005a, 2005b). These authors made the a priori assumption that 0.88% of the total halo mass is concentrated in a NC, a somewhat higher fraction than what has been found for present-day NCs. These authors conclude, in agreement with Cen (2001), that the predicted properties of such NCs match those of present-day GCs, independent of the exact value of the initial NC mass fraction.

They also point out that current observations are insufficient to rule out significant amounts of dark matter in GCs which might be expected if GCs indeed form at the centers of minihalos. So far, however, dark matter has not been convincingly detected in GCs, although the flat velocity dispersion profile of some GCs found by Scarpa et al. (2007) appears consistent with the presence of dark matter (or, as those authors prefer, with modified Newtonian dynamics). In any case, even the apparent absence of dark matter in GCs does not fully invalidate the idea, because Mashchenko & Sills (2005b) show that, depending on the density profile of the halo, dark matter can be lost almost completely by tidal stripping.

In this Letter, the Cen (2001) scenario is modified and extended by allowing for the expected secular evolution of NCs that occurs as long as they remain in the center of their parent halo. This secular evolution offers a natural explanation for the diversity in stellar populations within and across GCs. For a first-order assessment, the following simplifying assumptions are made:

1. All building block galaxies are identical in that their barionic matter is organized in a gas-rich disk of \( \sim 10^8 \text{--} 10^9 M_\odot \). This mass range is plausible because it makes dynamical friction an efficient mechanism to merge small galaxies into a larger halo (see eq. [7-27] in Binney & Tremaine 1987). It is also consistent with models of structure formation (e.g., Moore et al. 2006), as well as with observed masses of dwarf galaxies in the local universe (e.g., Geha et al. 2006).

2. The center of each progenitor disk is occupied by a NC with \( \sim 0.2\% \) of the host mass. As long as the disk escapes merger activity, its NC grows in mass via a series of nuclear starbursts fueled by gas flow from the disk onto the nucleus, similar to those observed in nearby galaxies (e.g., Schinnerer et al. 2003, 2006, 2007).

3. Once the disk experiences a destructive merger, the NC is captured in the halo of the merger product. From then on, it ages passively as a GC. Therefore, the later an galaxy is accreted, the more massive the newly added GC, and the more complex its star formation history.

4. Stars and gas of the merging disk will be incorporated into the merger products’ halo and disk, respectively, so that the total mass of the merger product is simply the sum of all the constituents.

5. The potential for cluster formation induced by the merging process is ignored for the moment, i.e., GCs are only added through accretion of NCs.

In this simplistic scenario, the total number of GCs is equal to the number of accreted building blocks. Because any intermediate merger products are created from identical building blocks as described in assumption 1, the number of GCs simply scales linearly with assembled galaxy mass, regardless of the details of the merger tree. This offers a natural explanation for the empirical result that all but the latest type spirals (see § 5) show a constant specific frequency of GCs: the observed T-value (Zepf & Ashman 1993) is \( \sim 2 \) GCs per \( 10^9 M_\odot \) (Goudfrooij et al. 2003; Rhode et al. 2007). This value is likely a lower limit because (i) not all GCs will survive and (ii) the faintest GCs might be missed in surveys. Taken at face value, however, it follows that the building blocks should have masses of \( \leq 5 \times 10^8 M_\odot \), entirely consistent with assumption 1 above.

The primordial NC of such a galaxy should then contain 0.002 \( \times 5 \times 10^9 M_\odot \), i.e., \( \sim 10^6 M_\odot \), in good agreement with the smallest observed NC mass of \( 8 \times 10^3 M_\odot \) (Walcher et al. 2005). This apparent minimum NC mass is unlikely to be a selection effect, because Böker et al. (2002) found a faint-end cutoff in the NC luminosity function in a much larger, unbiased sample of NCs. Note that NC masses in this range lie well above the critical mass \( m_c \sim 2 \times 10^4 M_\odot \) at which the GC mass function changes slope because less massive clusters are prone to disruption (Fall & Zhang 2001). It is therefore reasonable to expect that any such clusters will have survived until the present day.

Nevertheless, depending on the details of their orbit and initial density profile, the captured clusters will experience two-body relaxation and/or tidal stripping. These effects can easily reduce the cluster mass by a factor of 2--5 over \( \sim 10 \) Gyr (e.g., Vesperini & Heggie 1997). An even higher mass loss has been reported in some Milky Way GCs (de Marchi et al. 2006; de Marchi & Pulone 2007), making dynamical evolution a likely explanation for the fact that many GCs today have masses below \( 10^5 M_\odot \) (McLaughlin & van der Marel 2005).

If present-day GC systems are indeed made from NCs accumulated throughout the merging history of a galaxy, then the mass of any GC system should—to first order—contain the same mass fraction as each individual NC (because mass is conserved according to assumption 3 above), i.e., \( \sim 0.2\% \) of the total host galaxy mass. Intriguingly, this appears entirely consistent with observations: McLaughlin (1999) concludes that the formation efficiency of GCs is universal, at least for ellipticals and the spheroidal component of disk galaxies (including the halo of the Milky Way), with a value of \( \epsilon_{\text{clus}} \equiv M_{\text{clus}}/(M_{\text{gal}} + M_{\text{halo}}) \sim 0.0026 \). It should be noted, however, that the situation is less clear when entire disk galaxies (i.e., including the gas-rich thin disk) are considered (Brodie & Strader 2006).

Assuming for the moment that \( \epsilon_{\text{clus}} \) is indeed universal and had the same value at all redshifts, the argument can be turned around: the putative GC system of a \( 10^9 M_\odot \) building block galaxy should then contain a mass of only \( 5 \times 10^3 M_\odot \). Unless this mass is concentrated in very few (one or two) clusters, none of the GCs in that system will have survived until the present day. Note that this also represents a challenge for the dissipationless accretion scenario, because in order to produce significant numbers of stellar clusters in the halos of early-universe galaxies with enough mass to survive until \( z = 0 \), the value of \( \epsilon_{\text{clus}} \) must have been significantly higher in the past.

5. OPEN ISSUES AND PREDICTIONS

The toy model described in the last section is clearly simplistic, and there are a few points that require further investigation. For example, it has been assumed throughout the Letter that the mass fraction of NCs remains constant over time at 0.2% of the host galaxy, as observed in present-day galaxies. For this assumption to hold, the gas accretion rate onto the host galaxy must match the mass growth of the NC. To first order, such a link does not appear implausible because there are a number of mechanisms that transport gas from the outskirts of a galaxy toward the center. However, it is presently unclear whether galaxy evolution models can indeed explain such “fine tuning.”

Moreover, there is no doubt that the stellar populations of GCs are generally old. For example, the Milky Way GCs have ages of \( \geq 8 \) Gyr with a relatively small spread of \( <3 \) Gyr (De Angeli et al. 2005). Because in the proposed scenario, the age of the youngest stellar population in a GC denotes the time when its former host was disrupted, this implies that the merging of disk
galaxies into the Milky Way was mostly complete 8 Gyr ago. Any NCs accreted after that time must have been residing in dE-type galaxies, where they had already been aging passively for some time because there is no gas supplied to the nucleus in these galaxies. While this conclusion does not seem to be contradicted by any current observations, more detailed simulations of the Milky Way formation are required to test its plausibility.

A third issue that is not explained by the proposed scenario is the double-peaked color distribution of many GC systems. In massive galaxies, where this effect is most pronounced, it may be possible to invoke additional cluster formation during the merger of two massive spirals (each of which has formed its own GC system according to the proposed scenario), but not in any minor mergers that led to the buildup of the disk galaxies themselves. That star formation is indeed enhanced in major mergers is undisputed and has been observed in a number of nearby objects (e.g., Whitmore et al. 1993; Whitmore & Schweizer 1995). However, the fact that there are hardly any GCs known with intermediate ages (a few Gyr) requires that the majority of these mergers be completed by $z \sim 2$, possibly a challenge to models of galaxy formation. On the other hand, it may not even be necessary to invoke the presence of two GC subpopulations with intrinsically different metallicities to explain color bimodality (Yoon et al. 2006).

One prediction that the proposed scenario can be tested against is that “pure” disk galaxies should have a systematically lower $T$-value. This is because they apparently had an uneventful merger history which enabled them to avoid the formation of a bulge. Because these galaxies likely assembled a larger fraction of their history which enabled them to avoid the formation of a bulge, they should have gathered proportionally fewer NCs. Generally speaking, little is known about the GC systems of the latest type spirals, although a few studies have been undertaken which seem to confirm this prediction. For example, van den Bergh & Harris (1982) report the absence of GCs in the edge-on spiral NGC 891 which, however, might not be a good example of a bulgeless system since it is classified as Sb. Moreover, Olsen et al. (2004) find only a handful of confirmed GCs in most late-type spirals of the Sculptor group, and report that the specific frequency of GCs is lower in late-type spirals than in bulge-dominated galaxies. Unfortunately, these authors do not present $T$-values, which is a more robust diagnostic because of potential color variations between galaxy types.

It is still uncertain when exactly NCs formed in the universe and what the duty cycle of rejuvenation in those early nuclei was. In principle, one would expect that a large fraction of GCs show some level of complexity in their stellar populations. However, the earlier a NC was acquired, the fewer stellar generations it should contain, and the smaller the expected age differences between those populations are. As mentioned in § 2, single-population GCs are certainly expected, especially if there was a considerable delay between the original collapse of the Cen (2001) model and the onset of “rejuvenation” of the NC.

On average, however, more massive GCs should have been accreted more recently than low-mass ones. Because they spent a longer time in a galaxy nucleus, they experienced a higher number of star formation events and contain both younger and more complex stellar populations. The fact that multiple stellar populations so far have mostly been detected in rather massive GCs such as NGC 2808 and ω Cen thus is not surprising.

In summary, the speculative scenario for GC formation laid out in this letter does not seem implausible, and offers a natural explanation for the observed mass of GC systems ($M_{\text{GCs}} \approx 0.0025M_{\odot}$) and the presence of multiple stellar populations in massive GCs.

I am grateful to C. J. Walcher, R. van der Marel, G. de Marchi, and M. Perryman for helpful comments. The anonymous referee is acknowledged for a thorough and critical review, as well as a number of detailed suggestions which greatly improved this Letter.

REFERENCES

Ashman, K. M., & Zepf, S. E. 1992, ApJ, 384, 50
Balcas, M., Graham, A. W., & Peletier, R. 2007, ApJ, 665, 1084
Beasley, M. A., et al. 2002, MNRAS, 333, 383
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Böker, T., Laine, S., van der Marel, R. P., Sarzi, M., Rix, H.-W., Ho, L., & Shields, J. C. 2002, AJ, 123, 1389
Böker, T., et al. 2004, AJ, 127, 105
Brodie, J. P., & Strader, J. 2006, ARA&A, 44, 193
Caloi, V., & D’Antona, F. 2007, A&A, 463, 949
Carollo, C. M., Stiavelli, M., de Zeeuw, P. T., & Mack, J. 1997, AJ, 114, 2366
Cen, R. 2001, ApJ, 560, 592
Chandar, R., et al. 2006, ApJ, 646, L107
Côté, P., Marzke, R. O., & West, M. J. 1998, ApJ, 501, 554
Côté, P., et al. 2006, ApJS, 165, 57
De Angeli, F., et al. 2005, AJ, 130, 116
de Marchi, G., & Pulone, L. 2007, A&A, 467, 107
de Marchi, G., Pulone, L., & Paresce, F. 2006, A&A, 449, 161
Emsellem, E., & van de Ven, G. 2008, ApJ, in press
Fall, S. M., & Zhang, Q. 2001, ApJ, 561, 751
Ferrarese, L., et al. 2006, ApJ, 644, L21
Forbes, D. A., Brodie, J. P., & Grillmair, C. J. 1997, AJ, 113, 1652
Fremantle, K. 1993, in ASP Conf. Ser. 48, The Globular Cluster–Galaxy Connection, ed. G. H. Smith & J. P. Brodie (San Francisco: ASP), 608
Geha, M., Blanton, M. R., Masjedi, M., & West, A. A. 2006, ApJ, 653, 240
Geha, M., Guzmán, R., & Guhathakurta, P., & van der Marel, R. 2002, AJ, 124, 3073
Goudfrooij, P., et al. 2003, MNRAS, 343, 665
———. 2006, MNRAS, 372, 1888
Graham, A. W., & Guzmán, R. 2003, AJ, 125, 2936
Lee, Y.-W., et al. 1999, Nature, 402, 55
———. 2007, preprint (arXiv:0707.1705)