Compact groups from semi-analytical models of galaxy formation – II: Different assembly channels

E. Díaz-Giménez1,2★, A. Zandivarez1,2, G. A. Mamon3

1 Universidad Nacional de Córdoba (UNC). Observatorio Astronómico de Córdoba (OAC). Córdoba, Argentina
2 CONICET. Instituto de Astronomía Teórica y Experimental (IATE). Córdoba, Argentina
3 Institut d’Astrophysique de Paris (UMR 7095: CNRS & Sorbonne Université), Paris, France

ABSTRACT
We study the formation of over 6000 compact groups (CGs) of galaxies identified in mock redshift-space galaxy catalogues built from semi-analytical models of galaxy formation (SAMs) run on the Millennium Simulations. We select CGs of 4 members in our mock SDSS galaxy catalogues and, for each CG, we trace back in time the real-space positions of the most massive progenitors of their 4 galaxies. By analysing the evolution of the distance of the galaxy members to the centre of mass of the group, we identify 4 channels of CG formation. The classification of these assembly channels is performed with an automatic recipe inferred from a preliminary visual inspection and based on the orbit of the galaxy with the fewest number of orbits. Most CGs show late assembly, with the last galaxy arriving on its first or second passage, while only 10-20 per cent form by the gradual contraction of their orbits by dynamical friction, and only a few per cent forming early with little subsequent contraction. However, a SAM from a higher resolution simulation leads to earlier assembly. Assembly histories of CGs also depend on cosmological parameters. At similar resolution, CGs assemble later in SAMs built on parent cosmological simulations of high density parameter. Several observed properties of mock CGs correlate with their assembly history: early-assembling CGs are smaller, with shorter crossing times, and greater magnitude gaps between their brightest two members, and their brightest galaxies have smaller spatial offsets and are more passive.

Key words: galaxies: groups: general – galaxies: statistics – methods: numerical

1 INTRODUCTION
Isolated Compact Groups (CGs) of galaxies constitute an extreme environment, and a unique test-bed for galaxy interactions. Indeed, these high apparent density and low velocity dispersions systems (Hickson et al. 1992) make them ideal laboratories for galaxy interactions and in particular mergers (Mamon 1992). In turn, these interactions induce important transformations that have a strong impact on the history of their galaxy members.

For years, astronomers have been puzzled on how these small systems of galaxies can survive long periods of time and not ending up merging into a single massive galaxy. Numerous N-body simulations have been carried out to answer this question (e.g. Carnevali, Cavaliere & Santangelo 1981; Barnes 1985; Mamon 1987; Navarro, Mosconi & García-Lambas 1987; Bode, Cohn & Lugger 1993; Athanassoula, Makino & Bosma 1997), and generally agree that the galaxies rapidly merge over a timescale of 0.5 to 2 Gyr (see Mamon 1990 for a quantitative comparison of the results of the early simulations). If we observe CGs that do not survive long, a possible explanation for their existence today could be that the group membership is replenished by newly arriving galaxies (Diaferio, Geller & Ramella 1994), as expected from Press-Schechter (Press & Schechter 1974) theory (Mamon 2000). Nevertheless, under proper conditions, the merging process can be slowed down and therefore the longevity of these systems is not necessarily an issue.

Still, it is not clear how such dense isolated galaxy systems form in the first place. If CGs seen in the local Universe were as dense when they formed as they appear now, they should have formed at early times (z > 7 according to Appendix A or z > 3 if the groups are double the size in real space as they appear in projection). Alternatively, CGs may be galaxy systems that only very recently have acquired their fourth luminous member (CGs are defined to have at least 4
More accurate values for the simulations can be found at http://gavo.mpa-garching.mpg.de/Millennium/Help/simulation of compact groups identified in the mock lightcones (\(r<17.77\)). The amplitude of the power spectrum of primordial density simulation on which the SAM had been run. The space density and nature of CGs depend strongly on the cosmology of parent simulation constructed from five different SAMs (Diaz-Gimenez et al. 2010; Zandivarez et al. 2014b; Diaz-Gimenez & Mamon 2010; Zandivarez et al. 2014b; Diaz-Gimenez & Taverna 2018), as well as their mass assembly and their frequency versus time (Wiens et al. 2019). The nature of CGs has also been analysed from mock CGs extracted from hydrodynamical simulations (Hartssuiker & Ploeckinger 2020).

The assembly history of CGs is best explored with a very comprehensive numerical study. A suitable tool for such a study are the mock galaxy catalogues constructed using a semi-analytical model of galaxy formation (SAM) on top of a N-body cosmological simulation. During the last decade, such mock catalogues built from SAMs have been widely used to understand the frequency, nature and identification of CGs (McConnachie, Ellison & Patton 2008; Diaz-Gimenez & Mamon 2010; Zandivarez et al. 2014b; Diaz-Gimenez & Zandivarez 2015; Taverna et al. 2016; Diaz-Gimenez, Zandivarez & Taverna 2018), as well as their mass assembly (Farhang et al. 2017) and their frequency versus time (Wiens et al. 2019). The nature of CGs has also been analysed from mock CGs extracted from hydrodynamical simulations (Hartssuiker & Ploeckinger 2020).

Tzanavaris et al. (2019) analysed the evolutionary history of a handful of groups identified in 3D as small, dense and isolated systems inspired in the observational criteria of CGs. Despite their small number of groups (one quintet, one quartet and 8 triplets), they visualised a possible variety of evolutionary paths suggesting that CGs form via different assembly channels, where all galaxies end up lying within 0.5 Mpc from the most massive galaxy within the last 9 to 1 Gyr.

We have recently performed a comprehensive study of the frequency and nature of CGs identified in nine mock lightcones constructed from five different SAMs (Diaz-Gimenez et al. 2020, hereafter Paper I). We found that the frequency and nature of CGs depend strongly on the cosmological parameters of the parent dissipationless cosmological simulation on which the SAM had been run. The space density of physically dense CGs increases as a strong power of the amplitude of the power spectrum of primordial density fluctuations on the comoving scale of 8 h Mpc, \(\sigma_8\). Furthermore, the fraction of CGs that are not physically dense systems of at least 4 galaxies of comparable luminosity, but instead caused by chance alignments of galaxies along the line of sight, increases with the cosmological density parameter, \(\Omega_m\). In currently favoured cosmologies (e.g. WMAP9, Hinshaw et al. 2013 and Planck, Planck Collaboration et al. 2020), only roughly 35 per cent of CGs are physically dense (in rough agreement with early estimates by Mamon 1986, 1987). But higher-resolution N-body cosmological simulations capture better the formation of physically dense CGs, which now represent roughly half of all CGs. In Paper I, we also found that intrinsic differences in the SAM recipes also lead to differences in the frequency and nature of CGs.

In the present article, we address the following questions. Which is the principal formation channel of CGs? did they assemble early and keep their density, did they gradually increase in density, or did the last galaxy arrive just now on its first or 2nd passage? Is there a mix of CG assembly scenarios? How do cosmological parameters affect CG formation? If CG formation channels are mixed, are there links between specific channels and observational properties of CGs?

To answer these questions, we explore the formation channels of CGs by analysing the mock CG samples we built from the Millennium Simulations (Springel et al. 2005; Boylan-Kolchin et al. 2009) in Paper I. Using the galaxy merger trees from the SAMs, we then follow back in time the most massive progenitors of each of the member galaxies to study the spatial evolution of CGs.

The layout of this work is as follows. In Sect. 2, we present the SAMs and the lightcone construction from these SAMs. We explain in Sect. 3 how we built the mock CG sample. Evolutionary tracks are illustrated and classified in Sect. 4. The statistics of the formation scenarios are presented in Sect. 5, while their links to observational properties are explored in Sect. 6. We conclude and discuss our results in Sect. 7.

### 2 Mock Galaxy Lightcones

We used four SAMs run on the Millennium simulation (MS I, Springel et al. 2005): those of De Lucia & Blaizot (2007) and

---

**Table 1. Semi-analytical models analysed in this work**

| Authors | acronym | Cosmology | Simulation | Lightcone |
|---------|---------|-----------|------------|-----------|
|         | (1)     | (2)       | (3)    | (4)    | (5) | (6) | (7)    | (8)    | (9)    | (10) | (11) |
| De Lucia & Blaizot (2007) | DLB | WMAP1 | 0.25 | 0.73 | 0.90 | MS I | 500 | 3757143 | 3387 | 1908 |
| Guo et al. (2011) | G11 | WMAP1 | 0.25 | 0.73 | 0.90 | MS I | 500 | 3149024 | 3175 | 1570 |
| Guo et al. (2011) | G11 | WMAP1 | 0.25 | 0.73 | 0.90 | MS II | 100 | 3214602 | 2558 | 1195 |
| Guo et al. (2013) | G13 | WMAP7 | 0.27 | 0.70 | 0.81 | MS I | 500 | 2982462 | 1682 | 1010 |
| Henriques et al. (2015) | H15 | Planck | 0.31 | 0.67 | 0.83 | MS I | 480 | 3087401 | 1291 | 764 |

Notes: The columns are: (1): authors; (2): acronym of SAM; (3): cosmology of parent simulation; (4): density parameter of parent simulation; (5): dimensionless z=0 Hubble constant of parent simulation; (6): standard deviation of the (linearly extrapolated to z=0) power spectrum on the scale of \(8 h^{-1}\) Mpc; (7): name of parent simulation (MS I = Millennium Simulation I, MS II = Millennium Simulation II); (8): periodic box size of parent simulation \([h^{-1}\text{Mpc}]\); (9): number of galaxies in the mock lightcone; (10): total number of compact groups identified in the mock lightcones \((r<17.77)\); (11): the number of compact groups with only four galaxy members. More accurate values for the simulations can be found at http://gavo.mpa-garching.mpg.de/Millennium/Help/simulation.

* H15 is run on a re-scaled version of the original Millennium simulation (hence, the different box size).
We, hereafter, limit our analyses to CGs of 4 galaxies (hereafter, CG4s), because assembly histories are simpler when the final number of galaxies is the same. Table 1 indicates that CGs of four galaxies constitute 47% to 60% of the total sample of CGs (decreasing with improved resolution). The relative differences in the number of CG4s in each SAM resembles those of the CGs (including higher richness groups) previously analysed in Paper I.

The Hickson-like automatic search identifies systems that can hardly be considered as dense structures in 3D space (McConnachie et al. 2008; Díaz-Giménez & Mamon 2010). These "Fake" groups have a large maximum 3D separation between their galaxies, \( z = 0 > 1 \, h^{-1} \text{Mpc} \) or they have no close pairs in 3D space (within 200 \( h^{-1} \text{kpc} \)), which makes them unphysical systems. Therefore, Fake CG4s will not be followed back in time to trace their history, and we will consider them as a separate class when needed.

4 CG ASSEMBLY: A TEST SAMPLE

Before analysing all 6447 CG4s, we restricted our analysis to a test sample of 342 CG4s, built by selecting \( \sim 5\% \) of the CG4s of each SAM. In the test sample, we excluded the Fake CGs.

In our analyses, we tracked the evolutionary path of each \( z=0 \) galaxy member of the test sample. From the MS database\(^2\), we retrieved the merger history of each of these galaxies using the query example G2 provided in the database. We then selected only the main progenitor of each \( z=0 \) galaxy (main branch of the merger tree) up to the redshift where the galaxy is formed.

At each snapshot, CG4s are characterised by:

- the centre of stellar mass \( r_{\text{cm}} = \frac{\sum_{i=1}^{4} M_i r_i}{\sum_{i=1}^{4} M_i} \), where \( r_i \) is the Cartesian vector from the origin of the box to each galaxy member;
- the module of the comoving 3D distance of each member to the centre of mass. We name this parameter \( d_i \); and
- the module of the physical 3D distance to the centre of mass computed as \( p_i = \frac{d_i}{1 + z_{\text{snapshot}}} \).

4.1 Visual analysis of CG assembly histories

We observed four different types of CG assembly channels:

Late Assembly: the fourth galaxy has just arrived in the CG for the first time (Late).

Late Second Pericentre: the fourth galaxy has arrived in the CG on its second passage (Second).

---

\(^1\) In our samples of CGs, we deal with two different numbers of member galaxies: the observed number and the real number. The observed number of members takes into account the blending of galaxies that occurs when two particles in the simulation are too close in projection. For example, a CG with five members may appear as a CG of four observed members if one galaxy is blended with another. In this work, we use the real number of members when defining CG4s.

\(^2\) http://gavo.mpa-garching.mpg.de/Millennium/
Gradual Contraction: the four galaxies might or might not be together since an early epoch, but they have already completed two or more orbits together, becoming gradually closer with each orbit (Gradual).

Early Assembly: the four galaxies are together since an early epoch (hereafter Early).

Figure 1 displays examples of these four CG assembly classes extracted from the test sample. The first example, **Late Assembly**, shows that the last galaxy (in black) arrived from a physical distance of roughly 1 $h^{-1}$ Mpc. The second example, **Late Second Pericentre**, shows a dense triplet (all galaxies but the black one) forming around 8 Gyr ago ($z = 1.2$) and gradually becoming denser with time, while the last arrival entered the group 3.5 Gyr ago (from a physical distance of 500 $h^{-1}$ kpc), bounced out to nearly 200 $h^{-1}$ kpc and fell back into the group at $z = 0$, lying 40 $h^{-1}$ kpc from the centre. The third example, **Gradual Contraction**, shows a CG that forms between 6 and 5 Gyr ago, and the orbits gradually decay. The fourth example, **Early Assembly**, shows that the CG is already in place 9 Gyr ago, keeping its size ($r_{\text{max}} \approx 100 h^{-1} $ kpc) since that time.

From the visual classification of the test sample CG4s, we found that 32% are Late, 33% are Second, 24% Gradual, and 11% Early.

Figure 2 shows the frequency of the four CG assembly classes among the test sample of CG4s (blue solid lines), separately for our five considered SAMs. This first attempt of classification shows a tendency to a predominant percentage of recently formed CG4s in those SAMs run on the MS I. In the GII SAM run on the MS II, the fractions of CG4s classified as Early and Gradual are enhanced by a factor two, and the Gradual class is dominant. However, the fractions of assembly classes found with the visually-classified test sample should be taken with caution, because of its small number of CGs.

### 4.2 Automatic classification of CG assembly histories

Our sample of 6000 CG4s was too large to fully classify by visual means. Instead, we performed an automatic classification of CG assembly histories. We first measured the pericentres and apocentres of each galaxy’s orbit relative to the centre of mass of the group. We defined pericentres as those points in the profiles where the steep changes from negative to positive, but also we only considered “deep” pericentres, i.e., there is a pericentre in the snapshot $t$ if

\[
\frac{p_j(t_{j+2}) - p_j(t_{j})}{p_j(t_j)} > 0.2 \quad \text{AND} \quad \frac{p_j(t_{j+2}) - p_j(t_{j})}{p_j(t_j)} > 0.2 \ .
\]

The apocentres are the highest values of $p_j(t)$ in the profiles between two consecutive pericentres.

The profiles of each galaxy member are characterized by the following ten parameters:

(i) $t_{1p}$: time of the first (earliest) pericentre;
(ii) $h_{1p}$: $p_1$ at the time of the first pericentre;
(iii) $t_{lp}$: time of the last pericentre;
(iv) $h_{lp}$: $p_1$ at the time of the last pericentre;
(v) $t_{1a}$: time of the first apocentre;
(vi) $h_{1a}$: $p_1$ at the time of the first apocentre;
(vii) $t_{la}$: time of the last apocentre;
(viii) $h_{la}$: $p_1$ at the time of the last apocentre;
(ix) $n_p$: total number of deep pericentres (defined in eq. (1));
(x) $p(t_0)/p(t_1)$: ratio between the $p_i$ in the two latest snapshots.

Following the insights from the observations of the individual galaxy profiles, we characterize the assembly history...
Formation and evolution of CGs

5 RESULTS ON COMPACT GROUP FORMATION CHANNELS

Given the high success rate achieved by the automatic classification, we applied it to the entire sample of CG4s. Figure 4 shows the percentage of CG4s within each formation channel. Roughly 20% of the CG4s are classified as fake. This figure shows some differences with those from the smaller test sample (dashed lines in Fig. 2), which suffers from worse statistics. However, regardless of the SAMs, the tendency of the predominant recent formation of CG4s is confirmed and the percentage of CGs diminishes towards the earlier assembly channels. Between one third and one half of the CG4s are Early systems. We, therefore, conclude that the predictions obtained from the automatic recipe reproduce the visual classifications of test CG4s with excellent fidelity.

5.1 Physically dense and chance alignments CGs

We split the samples of CG4s into physically dense (Reals) and chance alignments (CAs) following the classification of Díaz-Giménez & Mamon (2010). Real CGs are those whose maximum comoving 3D separation between the four closest galaxies is less than \(100\, h^{-1}\) kpc, or less than \(200\, h^{-1}\) kpc if the the ratio of line-of-sight to transverse sizes in real space.
Figure 4. Percentages of $z=0$ 4-galaxy CG assembly channels in the full sample, for the five different semi-analytical models, using the automatic classification. The percentages of CGs that are Fake are also shown. Error bars are the 95% binomial confidence intervals.

Table 2. Number of $z=0$ CGs with only four members for each class

| SAM  | CG4 | Reals | CAs |
|------|-----|-------|-----|
|      | Total | %     | Total Fake |
| All  | 6447 | 2024  | 31  | 4423 | 1191 |
| DLB  | 1908 | 588   | 31  | 1320 | 376  |
| G11  | 1570 | 530   | 34  | 1040 | 250  |
| G13  | 1010 | 300   | 30  | 710  | 202  |
| H15  | 764  | 185   | 24  | 579  | 171  |
| GII  | 1195 | 421   | 35  | 774  | 192  |

is less than 2. The remaining CGs are classified as CAs. In this scenario, the Fake CGs defined in Sect. 3 belong to the CA subsample.

The numbers of different classes of CG4s identified in each lightcone are listed in Table 2. One first notices that the fraction of Reals among CG4s is typically 30 per cent lower than the corresponding fraction among all CGs given in the top panel of fig. 8 of Paper I. In particular, among CG4s, the fraction of Reals never exceeds 35 per cent (reached for GII).

In addition, in Fig. 5, we show the percentage of Reals and CAs within each formation channel. Regardless of the SAM, over three-quarters of the Late forming CGs are CAs; CAs are also the dominant population within the Second channel, while the Reals dominate the Early assembly CGs. In Appendix B1, we focused on the CA CG4s by analysing their internal substructures. Among the few CAs that formed earlier, the CAs formed by a triplet plus a some-

Figure 5. Percentage of Reals and CAs (including Fakes) within each formation channel.

what distant galaxy are the main contributors to the long lived classes (Fig. B1).

Figure 6 displays the percentages of different CG assembly channels among Reals and CAs. The assembly histories of Real CGs depends somewhat on the SAM. Real CGs in the DLB and GII SAMs are mainly assembled gradually, while in G11 and G13, Real CGs present a homogeneous mixture of Late, Second and Gradual classes. Finally, Real CGs in the H15 SAM are dominated by recently assembled CGs.

Figure 6 indicates that, regardless of the SAM, CAs assemble significantly more recently than Real CG4s. Indeed, the Late class of assembly history accounts for a significantly higher fraction of CAs (typically over 40 per cent) than of Reals (typically under 30 per cent).

5.2 Dependence on cosmology and SAM recipes

Following Paper I, we explore the dependence of our results with the underlying cosmological parameters of the SAMs and with their physical recipes. Figure 7 shows the fraction of a composite assembly class named Earlys, which combines Gradual and Early classes, as a function of cosmological parameters and the fraction of orphan galaxies in each SAM. We estimate the power-law relations for the fraction of Earlys: $\Omega_{m}^\alpha$, $\sigma_8^\beta$, and $f_\text{orph}^\gamma$. The fits were obtained using all the SAMs with the same resolution (MS I). The best-fitted values of the exponents are quoted in the figure.

We first noted some trends of CG assembly classes with cosmological parameters. SAMs built in denser universes (G13 and especially H15) have lower fractions of Earlys. On the other hand, there is only a weak positive trend of the assembly history as a function of the amplitude of mass fluctuations ($\sigma_8$). We discuss these trends in Sect. 7 below.
Contreras et al. (2013) and Pujol et al. (2017) observed that differences in the treatment of orphan galaxies in the SAMs impact on the small scale clustering of galaxies, which are the relevant scales for CGs. Orphan galaxies are those whose subhaloes became no longer sufficiently massive to be realistic or no longer visible at all. The SAMs can no longer use the subhaloes to follow their positions and rely instead on crude recipes for their orbital evolution. Hence, the treatment of orphans implemented in each SAM affects the abundance of this particular population of satellite galaxies.

In Paper I, we noticed that orphan galaxies are frequent in CGs, especially Real ones. We also found that the fraction of CGs that are physically dense (i.e. Reals) decreases with the fraction of orphans in the SAMs for given cosmological resolution. We therefore analysed the fraction of Earlys as a function of the fraction of orphan galaxies in the lightcones built from each SAM (right panels of Fig. 7).

Considering SAMs obtained from the MS I simulation (i.e., all with the similar mass resolution), Fig. 7 shows that the fraction of Earlys increases with the fraction of orphan galaxies. This trend appears to be statisticallly significant in Reals and even more pronounced in CAs. As noted in Paper I, lower fractions of orphan galaxies are found in mocks from SAMs run on the higher resolution MS II simulation. Indeed, such simulations allow following the orbits of low-mass galaxies, which tend to live in lower subhalo masses, instead of having to predict them with uncertain dynamical recipes. From this figure, it can be seen that despite the lower fraction of orphans, a higher fraction of Earlys is found in the MS II, similar to the fraction of Earlys in the comparable G11 SAM. This suggests that orphans play a minor role in comparison with $\Omega_m$ in setting the fraction of CGs assembling early.

Our classification is based on the orbits of key galaxies, some of which require analytical recipes when the subhalo has too few particles or disappears entirely, i.e. for orphan galaxies. We therefore examined the impact of orphan key galaxies within each formation channel, and we show the main results in Fig. 8.

First, we found that, summing over all assembly channels, roughly one-third (or less depending on the SAM) of the key galaxies become orphans at some time in their evolutionary history (top labels). Furthermore, the occurrence of orphan key galaxies increases from the Late to the Early channel. For the MS I based SAMs, the percentage of key galaxies that are orphans rises from $\sim 13\%$ in the Late CGs to $\sim 85\%$ in the Early CGs. For the better resolved GII SAM, this increase is considerably lower.

From the merger trees of the orphan key galaxies, it is possible to determine the look-back-time and the physical distance to the mass centre when they became orphans. The median values of these two features are shown in Fig. 8. Typically, the key galaxies in the Second and Gradual channels become orphans during the last 5 Gyr, when they are less than $100 \, h^{-1}$ kpc from the CG centre (typically less than $50 \, h^{-1}$ kpc for the better resolved GII SAM). The key galaxies in the Early channel have also become orphans when they are close to the centre, but this occurs earlier. In these three assembly classes, galaxies become orphans at distances to the CG centre-of-mass of order the half-maximum size of the system ($200 \, h^{-1}$ kpc) or less (in particular for the better resolved GII SAM, but also for some other cases). Therefore, the assembly into the small CG size typically occurs when the galaxies are not yet orphans, i.e. before the time when the SAMs use analytical recipes to predict galaxy orbits.

On the other hand, the key galaxies in the Late channel became orphans only recently, but at larger distances to the centre. However, the percentage of Late CGs that are classified using an orphan galaxy is very low in all the SAMs. We conclude that the presence of orphan galaxies as key galaxies does not modify the results found in this work. We also checked whether the visual classification was influenced with the key galaxy being an orphan, but found no clear such influence.

6 ARE OBSERVED CG PROPERTIES LINKED TO THEIR ASSEMBLY HISTORY?

One may wonder whether observational properties of CGs may retain signatures of their particular assembly history.

Figure 9 shows the distribution of different properties of CGs in the GII SAM in form of boxplot diagrams (similar plots for CGs in the other four SAMs are shown in Figure B2). The dimensionless group crossing time and group virial theorem mass-to-light ratio are measured as Diaz-Giménez et al. (2012): $H_0 \tau_{cr} = 50 (\pi/\sqrt{3}) (h d_{ij}/\sigma_p)$ and $M_{vir}/L = 3\pi \left(d_{ij}^2\right)^{-1} \sigma^2/G(L)$, where $d_{ij}$ is the median inter-galaxy separation projected on the sky.

Analysing the median values obtained from this figure\(^3\),
some dependence of these properties on the assembly class is observed. CG4s in the Early class show the largest magnitude gaps, as well as the lowest projected sizes, crossing times and brightest galaxy relative projected offsets. Those attributes are actually a good reflection of their early formation times. On the other hand, CG4s recently formed show the opposite behaviour, with the Late class showing the smallest magnitude gaps, as well as the largest projected sizes, crossing times, and brightest galaxy relative projected offsets.

In general, there seems to be a progression in the median properties between the four group assembly channels, following the trend: Early Assembly, Gradual Contraction, Late Second Pericentre, Late Assembly.

We have also included in this figure the properties of those CG4s classified as Fake to have the complete picture of all CG4s in the main sample. In most of the properties, with 95% confidence that the corresponding medians are different (McGill et al. 1978; Krzywinski & Altman 2014).

Fake CG4s behave similarly to the recently formed CG4s, except for the crossing times where Fake CG4s have lower values.

While Fig. 9 shows this analysis only for the GII SAM, roughly the same trends are observed for all the remaining SAMs (see Figure B2). Moreover, Figs. 9 and B2 show that all SAMs display weaker but apparently significant trends for Early assembly histories for CGs of higher group velocity dispersion and higher brightest galaxy stellar mass-to-luminosity ratio. In contrast, contradictory trends among the CG4s built from the different SAMs are found for group luminosity and virial-theorem-mass-to-luminosity ratio.

7 SUMMARY AND DISCUSSION

In this work, we studied the different assembly channels that give rise to the compact groups of galaxies (CGs) of galaxies identified at present using an automatic algorithm that follows Hickson’s well-known recipes (Hickson 1982; Hickson et al. 1992). Following our previous work (Diaz-Gimenez...
et al. 2020), we identified CGs in several mock lightcones constructed from the Millennium Simulations using different SAMs. Since our study involves following each member of a CG back in time, we simplified our analysis by only selecting CGs with four members (CG4s), and by only following their most massive progenitors. We thus analysed the assembly histories of over 6000 CG4s in the outputs of five SAMs.

To start with, we performed a preliminary analysis, where we visually inspected a small subset (~350 CG4s) of CG assembly histories, using the evolution of the 3D physical distances of galaxy members to the group centre of mass. This allowed us to differentiate four different basic assembly channels that we named Early Assembly, Gradual Contraction, Late Assembly, and Late Second Pericentre (Fig. 1). The first two channels encompass relatively old systems, while the last two are characterised by rather late assembly. We notice that there is a wide spectrum of possible variations for the evolution of CGs, and some of them evolve in ways that are at the boundaries between channels. Therefore, our classification is most probably a simplification. From the analysis of the test sample, we observed a tendency of CG4s to be predominantly recently formed (Fig. 2).

We then extended this analysis to the full sample of CG4s, by defining an automatic classification using some of the features measured on the evolution of the physical 3D distance to the group centre of mass of the galaxy with the fewest significant pericentres (see Fig. 3). This automatic

![Figure 8. Median (over CGs) positions and times when key galaxies become orphan (they lose their subhalos and the SAMs treat their orbits within analytical recipes). Top labels quote the percentages of CG4s (excluding Fake CGs) in each SAM whose key galaxies became orphans. Each panel corresponds to the different assembly channels. Different symbols display the different SAMs, as in Fig. 7.Inset legends quote the percentage of CG4s within each assembly channel whose key galaxy became an orphan at some time (from top to bottom: DLB, G11, G13, H15 and GII).](image1)

| Channel       | DLB(35%) | G11(34%) | G13(36%) | H15(35%) | GII(25%) |
|---------------|----------|----------|----------|----------|----------|
| Early         | 11%      | 12%      | 16%      | 11%      | 13%      |
| Gradual       | 73%      | 76%      | 82%      | 65%      | 39%      |
| Late          | 13%      | 28%      | 40%      | 32%      | 18%      |
| Second        | 33%      | 28%      | 40%      | 32%      | 18%      |
| t_orph [Gyr]  | 1000     | 1000     | 1000     | 1000     | 1000     |

![Figure 9. Distribution of the observable properties of z=0 mock 4-galaxy CGs (with no splitting between Real and CAs) as a function of the assembly class for the GII SAM (second to last columns). We also include the properties for the sample of CGs previously classified as Fake CAs (first column). The properties, from top to bottom, are distance (median line-of-sight velocity), group luminosity, projected radius of the minimum enclosing circle, group line-of-sight velocity dispersion, dimensionless group crossing time, group virial theorem mass-to-light ratio, dimensionless brightest galaxy projected group-centric distance, dimensionless brightest galaxy line-of-sight velocity offset, r-band magnitude gap between the two most luminous galaxies, and brightest galaxy stellar mass-to-light ratio. The waists, tops and bottoms of the coloured boxes indicate the median values, and the 75th and 25th percentiles, respectively, the widths of the boxes are proportional to the numbers of CGs, while notches around the waists show the 95% confidence interval on the medians.](image2)
classification achieved a 96 per cent average success rate in recovering the visual classification.

Applying our automatic classification on the full sample of 6000 CGs, we found that recently formed CGs are the dominant assembly class, regardless of the SAMs. When analysing CGs classified as Reals according to the classification scheme of Díaz-Giménez & Mamon (2010), we observed that they have, on average, an important component of older systems, whereas CGs that are Chance Alignments according to the Díaz-Giménez & Mamon classification are preferentially dominated by recently assembled systems (Fig. 6).

One may wonder what is the meaning of assembly classes of CGs that are chance alignments. There is no bi-modality in the 3D properties of CGs: as seen in Fig. 6 of Díaz-Giménez & Mamon (2010), our distinction between Reals and CAs is based on a somewhat arbitrary cut in the smooth distribution of mock CGs in the alignment versus size plane. Many CAs barely missed being classified as Reals. Therefore, there is no strong reason to exclude CAs from the analysis.

Among CGs, Reals have greater number density and presumably greater mass density than the CAs. The shorter assembly time for the Reals relative to the CAs is thus presumably greater mass density than the CAs. The shorter crossing times), with the largest magnitude gap between the two brightest members, while more recently formed CGs show the opposite behaviour (Figs. 9 and B2).

There are similarities between the correlation of assembly times and observational properties for CGs and for groups in general. Since the pioneering work of Jones et al. (2003), old galaxy groups have been directly associated with systems showing a large magnitude gap between the two brightest galaxies. This association has been confirmed using galaxies in mock catalogues built from SAMs, not only using the gap between the two brightest galaxies (e.g., Dariush et al. 2007; Díaz-Giménez, Muriel & Mendes de Oliveira 2008), but also using the gap between the first and fourth brightest galaxies (e.g., Dariush et al. 2010; Kangasukka, Díaz-Giménez & Zandivarez 2016). Other physical quantities have been found to correlate with the epoch of group assembly. Khosroshahi, Ponman & Jones (2007) found that older galaxy systems (with large magnitude gap) have a more concentrated halo compared with those recently formed. In the same vein, Ragone-Figueroa et al. (2010) found that more concentrated groups form earlier in cosmological hydrodynamical simulations. In our work, we found that the four galaxy members in early formed CGs are clearly confined to a smaller region of space than those in recently formed CGs.

Raouf et al. (2014) argued that group assembly history is best measured with a combination of observational parameters: they considered the group and brightest group galaxy (BGG) luminosities and the physical offset of the BGG from the luminosity centroid. Farhang et al. (2017) found that CGs acquire half their mass earlier than normal groups, but later than Fossil groups. Our results have shown that the findings valid for early-forming galaxy groups also hold for early-assembly CGs (at least for median values), which are denser (shorter crossing times) and whose BGGs have greater magnitude gaps, and lie closer to the luminosity centroid, than later-assembly CGs.
However, our analysis is different from all these studies, because we do not measure the epoch when half of the CG mass is assembled. Instead, we measure the spatial assembly history, which we split into four formation channels. We found that these channels represent themselves an evolutionary sequence that are correlated with observational properties in accordance to what was found for the mass assembly history by the studies mentioned above.

Unfortunately, the median observational properties of CGs split into each of the four formation channels are not sufficiently different to reliably infer the formation channel of individual CGs. Some issues remain to be clarified. What prevents Early Assembly CGs to gradually shrink in size by dynamical friction and inelastic galaxy-galaxy encounters? Will our results differ when we analyze mock catalogues built from cosmological hydrodynamical simulations (as rapidly done by Hartsuiker & Ploeckinger 2020), which now have sufficient spatial resolution to achieve better small-scale physics compared to SAMS? Do the formation channels of 4-member CGs differ from those of other (less dense) groups of 4 members?

ACKNOWLEDGEMENTS

We thank the anonymous referee for useful comments, as well as Matthieu Tricottet. We also thank the authors of the SAMS for making their models publicly available. The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory (GAVO). This work has been partially supported by Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET) and the Secretaría de Ciencia y Tecnología de la Universidad de Córdoba (SeCyT).

DATA AVAILABILITY

The data underlying this article were accessed from http://gavo.mpa-garching.mpg.de/Millennium/. Galaxy light-cones built in Paper I and used in this work are accessed from https://doi.org/10.7910/DVN/WGOPCO (Diaz-Gimenez et al. 2021). The derived data generated in this research will be shared on reasonable request to the corresponding authors.

REFERENCES

Athanasoula E., Makino J., Bosma A., 1997, MNRAS, 286, 825
Barnes J., 1985, MNRAS, 215, 517
Bode P. W., Cohn H. N., Lugger P. M., 1993, ApJ, 416, 17
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Carnevali P., Cavaliere A., Santangelo P., 1981, ApJ, 249, 449
Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Dariush A., Khosroshahi H. G., Ponman T. J., Pearce F., Raychaudhury S., Hartley W., 2007, MNRAS, 382, 433
Dariush A. A., Raychaudhury S., Ponman T. J., Khosroshahi H. G., Benson A. J., Bower R. G., Pearce F., 2010, MNRAS, 405, 1873
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Diaferio A., Geller M. J., Ramella M., 1994, AJ, 107, 888
Díaz-Giménez E., Mamon G. A., 2010, MNRAS, 409, 1227
Díaz-Giménez E., Zandivareh A., 2015, A&A, 578, A61
Díaz-Giménez E., Muriel H., Mendes de Oliveira C., 2008, A&A, 490, 965
Díaz-Giménez E., Mamon G. A., Pacheco M., Mendes de Oliveira C., Alonso M. V., 2012, MNRAS, 426, 296
Díaz-Giménez E., Zandivareh A., Taverna A., 2018, A&A, 618, A157
Díaz-Giménez E., Taverna A., Zandivareh A., Mamon G. A., 2020, MNRAS, 492, 2588
Díaz-Giménez E., Taverna A., Zandivareh A., Mamon G., 2021, Galaxy Lightcones, doi:10.7910/DVN/WGOPCO, https://doi.org/10.7910/DVN/WGOPCO
Farhang A., Khosroshahi H. G., Mamon G. A., Dariush A. A., Raouf M., 2017, ApJ, 840, 58
Guo Q., et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Hartsuiker L., Ploeckinger S., 2020, MNRAS, 491, L66
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hickson P., 1982, ApJ, 255, 382
Hickson P., Mendes de Oliveira C., Huchra J. P., Palumbo G. G., 1992, ApJ, 399, 353
Hinshaw G., et al., 2013, ApJS, 208, 19
Iodotou D., Thomas P. A., Henriques B. M., Sargent M. T., Higlop J. M., 2019, MNRAS, 489, 3609
Jones L. R., Ponman T. J., Horton A., Babul A., Ebeling H., Burke D. J., 2003, MNRAS, 343, 627
Kanagasaki M. J., Díaz-Giménez E., Zandivareh A., 2016, A&A, 586, A40
Khosroshahi H. G., Ponman T. J., Jones L. R., 2007, MNRAS, 377, 595
Knebe A., et al., 2015, MNRAS, 451, 4029
Komatsu E., et al., 2011, ApJS, 192, L8
Krzywinski M., Altman N., 2014, Nature Methods, 11, 119
Mamon G. A., 1986, ApJ, 307, 426
Mamon G. A., 1987, ApJ, 321, 622
Mamon G. A., 1990, in Sulpentic J. W., Keel W. C., Telesco C. M., eds, NASA Conference Publication Vol. 3098, NASA Conference Publication. p. 609
Mamon G. A., 1992, ApJ, 401, L3
Mamon G. A., 2000, in Valtonen M. J., Flynn C., eds, Astronomical Society of the Pacific Conference Series Vol. 209, IAU Colloq. 174: Small Galaxy Groups. pp 217–225
McConnachie A. W., Ellison S. L., Patton D. R., 2008, MNRAS, 387, 1281
McGill R., Tukey J. W., Larsen W. A., 1978, The American Statistician, 32, 12
Navarro J. F., Molaeinezhad A., Tavasoli S., 2014, MNRAS, 442, 1578
Planck Collaboration et al., 2016, A&A, 594, A13
Planck Collaboration et al., 2020, A&A, 641, A6
Ponman T. J., Bourner P. D. J., Ebeling H., Bohringer H., 1996, MNRAS, 283, 690
Press W. H., Schechter P., 1974, ApJ, 187, 425
Raouf M., Khosroshahi H. G., Ponman T. J., Dariush A. A., Molaeinezhad A., Tavassoli S., 2014, MNRAS, 442, 1578
Schneider D. P., Gunn J. E., 1982, ApJ, 263, 14
Spergel D. N., et al., 2003, ApJS, 148, 175
Springel V., et al., 2005, Nature, 435, 629

Formation and evolution of CGs 11
APPENDIX A: COMPACT GROUP VIRIALIZATION REDSHIFT IF THEY REMAIN AS DENSE UNTIL PRESENT

We estimate here the redshift \( z_{\text{vir}} \) when CGs virialize if they remain at the same physical density until present. The mean density of the CG is

\[
\bar{\rho} = \rho_{\text{CG, crit}}(0) = \rho_{\text{crit}}(z_{\text{vir}}) \rho_{\text{crit}}(z_{\text{vir}}),
\]

where \( \rho_{\text{crit}}(z) = 3H^2(z)/(8\pi G) \) is the critical density. The CG virialization redshift is therefore the solution of

\[
\Delta(z) E^2(z) \rho(z) = \Delta_{\text{CG}}.
\]

where \( E(z) = H(z)/H_0 = \sqrt{\Omega_m(1+z)^3 + 1 - \Omega_m} \) for a flat Universe with current density parameter \( \Omega_m \). According to the values listed in table 5 of Díaz-Giménez et al. (2012), the overdensity of CGs is \( \Delta_{\text{CG}} = 3(M_{VT}/L_R)(4\pi r_p^3) = 24000 \) and 28000, for the median CGs in the 2MASS catalogue (Díaz-Giménez et al. 2012) and mock CGs (from the G11 SAM) respectively (all considered after velocity filtering). Here, we measured the CG density within the sphere of radius equal to the smallest circumscribed radius (this assumes that the typical CG is spherical). Solving equation (A2), we deduce \( z_{\text{vir}} = 7.1 \) (2MCG) and 7.6 (mock CGs). However, CGs may be extended along the line-of-sight than viewed in projection. Adopting a CG radius of 2\( \rho_p \) instead of \( r_p \), the overdensities are 8 times lower, and the virialization redshifts are now \( z_{\text{vir}} = 3.0 \) and 3.2 respectively.

APPENDIX B: ANALYSIS OF CGS IN ALL SEMI-ANALYTICAL MODELS

B1 Assembly histories of CA CGs

For a more detailed analysis, we defined CA sub-classes as follows:

- **3+1**: \( r_{\text{max}} < D_s \) and \( r_{\text{min}} < d_s \) and \( r_{\text{max}} \leq d_s \),
- **2+2**: \( r_{\text{max}} < D_s \) and \( r_{\text{min}} < d_s \) and \( r_{\text{max}} > d_s \) and \( r_{ij} \leq d_s \),
- **2+1+1**: \( r_{\text{max}} < D_s \) and \( r_{\text{min}} < d_s \) and \( r_{\text{max}} > d_s \) and \( r_{ij} > d_s \),
- **Fakes**: \( r_{\text{max}} \geq d_s \) or \( r_{\text{min}} \geq d_s \),

where \( r_{\text{max}} \) and \( r_{\text{min}} \) are the maximum and minimum inter-particle separation using the 4 members of the group, respectively, \( r_{\text{max}} \) is the maximum inter-particle separation between the 3 closest members of the group, \( r_{ij} \) is the inter-particle separation between the 2 members that do not define the pair with \( r_{\text{min}} \), and the thresholds are \( D_s = 1 \ h^{-1} \) Mpc and \( d_s = 200 \ h^{-1} \) kpc. Table B1 displays the percentages of the different CA sub-classes for each SAM.

We checked how the assembly histories differ among the different sub-classes of CAs (excluding Fakes). Figure B1 shows the percentages of CGs for the CA sub-classes as a function of assembly class. All the CA sub-classes roughly follow the general trend previously observed for the full sample of CAs, with the main contribution to **Gradual** and **Early** classes of the CAs coming from the dominant “3+1” sub-class. The sub-classes with one or two galaxy pairs are mainly recently formed in all SAMs. In summary, the richer the subsystem, the earlier it tends to assemble, from pairs (2+2 or 2+1+1 CAs) to triplets (3+1 CAs) to quartets (Reals).

Interestingly, the key galaxy of a CA, which by definition is the galaxy with the fewest orbits, is not necessarily an outlier at \( z = 0 \) (for 3+1 CAs) or one of them (for 2+2 CAs).

| Table B1. Number of \( z=0 \) CAs and percentages within each CA subclass |
|---------------------------------------------------------------|
| **SAM** | **#** | **% Fake** | **% 3+1** | **% 2+2** | **% 2+1+1** |
| All | 4423 | 27 ± 1 | 52 ± 3 | 7 ± 2 | 13 ± 1 |
| DLB | 1320 | 28 ± 2 | 52 ± 3 | 7 ± 2 | 13 ± 2 |
| G11 | 1040 | 24 ± 4 | 54 ± 3 | 7 ± 2 | 15 ± 2 |
| G13 | 710 | 28 ± 4 | 48 ± 4 | 9 ± 2 | 15 ± 3 |
| H15 | 579 | 30 ± 4 | 47 ± 4 | 8 ± 2 | 15 ± 3 |
| GII | 774 | 25 ± 3 | 56 ± 3 | 9 ± 2 | 10 ± 2 |

Note: Each percentage \( p \) is quoted as \( p \pm ci \) where \( ci \) is the 95% binomial confidence interval computed as \( \pm (1.96 \times \sqrt{f(1-f)/N_{\text{CG}}}) \times 100 \) where \( f = p/100 \).

Figure B1. Same as Fig. 6, but only for the 4-galaxy CGs that are chance alignments, split by sub-class.

van den Bosch F. C., 2002, MNRAS, 331, 98
or 2+1+1 CAs). So the classification of CA assembly histories, which is a function of the orbit of the key galaxy, is not necessarily tied to an outlier but to one of the galaxies of the physical subsystem (i.e. triplet in a 3+1 CA). However, the key galaxy of a CA is more likely to be an outlier galaxy, since to first order, galaxies further out live in lower mean densities, hence have the longest orbital times and the fewest number of orbits. Nevertheless, the fraction of CAs classified as Late or Second is less than two-thirds. On the other hand, finding that the 3+1 CAs are the main contributor to Earlies is expected, because 3+1 CAs have only one outlier and are thus less likely to have their key galaxy be an outlier than 2+2 or 2+1+1 CAs.

B2 Observational properties versus SAM

Figure B2 shows the properties of CGs in each assembly class for the SAMs of DLB, G11, G13 and H15.
Figure B2. Same as Figure 9, but for the CGs built from the four other semi-analytical models.