Mon. Not. R. Astron. Soc. **000**, 0–0 (2000) Printed 4 November 2018 (MN plain TEX macros v1.6)

Overmerging in N-body simulations

Eelco van Kampen

Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, evk@roe.ac.uk

Accepted ... Received ...; in original form ...

ABSTRACT

The aim of this paper is to clarify the notion and cause of overmerging in N-body simulations, and to present analytical estimates for its timescale. Overmerging is the disruption of subhaloes within embedding haloes due to *numerical* problems connected with the discreteness of N-body dynamics. It is shown that the process responsible for overmerging is particle-subhalo two-body heating. Various solutions to the overmerging problem are discussed.

Key words: galaxies: evolution - dark matter - large-scale structure of Universe

1 INTRODUCTION

In the study of the formation and evolution of large-scale structure in the universe, clusters of galaxies, galaxies, and many other gravitational systems, haloes play an important rôle. A *halo* is defined as a collapsed and virialized density maximum. A halo can contain several smaller haloes, denoted as *subhaloes*. We denote a halo that contains subhaloes as an *embedding halo*. The existence of subhaloes is an important issue in cosmology, especially for galaxies, and groups and clusters of galaxies. For example, hierarchical structure formation scenarios for CDM-like spectra predict many more dwarf galaxies (or satellites) than observed (Klypin et al. 1999b, Moore et al. 1999). Also, a spiral galaxy cannot retain its disk if there is an high abundance of subhaloes within its halo (Moore et al. 1999). Thus, the question is whether the initial density fluctuation spetrum is such that not many dwarf galaxies form in the first place, or that they are easily destroyed within our Galaxy, and hard to find outside it. On larger scale, clusters of galaxies do contain an abundance of subhaloes, i.e. its member galaxies, which are often distorted and stripped, but probably not destroyed.

N-body simulations are routinely used to model the formation, evolution, and clustering of galaxy and galaxy cluster haloes. However, the N-body simulation method does have its limitiations, and care should be taken with the interpretation of the simulation results. One such limitation arises from the use of particles to represent the mass distribution whose evolution one tries to simulate. If the physical mass distribution is effectively collisionless, as is often the case, the use of particles gives rise to artificial collisional effects within the numerical mass distribution, especially though two-body interaction. Two-body encounters between simulation particles, either close or distant, deflect their orbits significantly, while they should behave like test particles, and respond only to the mean potential. This is especially a problem for subhaloes, which are easily destroyed in an N-body simulation with insufficient resolution (White et al. 1987; Carlberg 1994; van Kampen 1995). We use the term *overmerging* for the numerical processes that artificially merge haloes and subhaloes in an N-body model, usually by disrupting the subhalo. Thus, the term *merging* only denotes merging due to physical processes.

However, there seems to be some confusion in the literature over the nature, cause, and importance of the overmerging problem. This paper attempts to clarify the difference between the three most important two-body processes operating on subhaloes, and provide estimates for their associated timescales. The first process is *two-body evaporation*, which is an *internal* process operation within any halo or subhalo. It is due to two-body interactions between the particles within the halo or subhalo. The second process is *particle-subhalo two-body heating*, which is the heating of 'cold' subhaloes by particles from the 'hot' embedding halo through two-body interactions. The third process is *particle-subhalo tidal heating*, where the subhalo is considered collisionless, and increases its kinetic energy through tidal interactions with particles from the embedding halo. All three process are illustrated in Fig. 1.

Besides the two-body processes, the use of *softened* particles, in order to minimize two-body effects, can cause overmerging as well by artificially enhancing physical processes like merging and disruption by tidal forces (van Kampen 1995; Moore et al. 1996). Groups of softened particles are not as compact as real haloes, and their artificially larger sizes make N-body groups more prone to tidal disruption.

Carlberg (1994) proposed particle-subhalo two-body heating as the main cause for overmerging. He gave a timescale for this process, but no derivation. This was provided by van Kampen (1995). It has been shown before (Carlberg 1994; van Kampen 1995) that the two-body heating time-scale for small subhaloes orbiting an embedding halo is short enough to result in their complete destruction and dispersion. Subsequent authors, including Moore et al. (1996) and Klypin et al. (1999a), referenced Carlberg (1994)

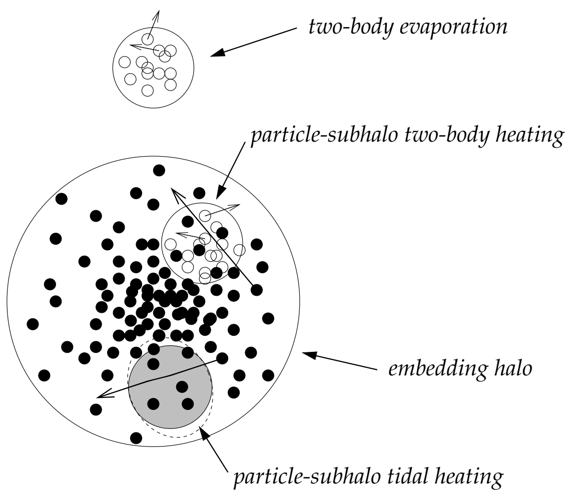

Figure 1. Graphical illustration of the main numerical disruption processes that cause overmerging. Open circles represent the 'cold' particles of an isolated halo or a subhalo within an embedded halo, whose 'hot' particles are indicated by filled circles.

as saying that 'particle-halo heating' is at the root of the problem. Moore et al. (1996) then claim that the process is not important for the resolution of the simulation performed by Carlberg (1994) because the timescale is too long. However, the process Moore et al. (1996) actually describes and derives a timescale for is a different one, driven by tidal encounters between simulation particles and perfectly collisionless subhaloes, while Carlberg (1994) and van Kampen (1995) clearly had a collisonal process in mind, driven by two-body encounters between individual subhalo particles and particles from the embedding halo. This paper shows that the latter process has a much shorter timescale, and therefore, along with excessive softening, is the main cause for overmerging.

2 TWO-BODY EFFECTS IN N-BODY SIMULATIONS

Two-body effects become dominant for systems modelled by small numbers of particles. This is usually quantified by the two-body relaxation timescale, which is defined as the time it takes, on average, for a particle to change its velocity by of order itself. After this time a system is denoted as *relaxed*. The relaxation timescale is defined in terms of the *half-mass radius* r, the *typical velocity* v (usually taken to be equal to the velocity dispersion), and the average change Δv per *crossing time* r/v:

$$t_{\rm relax} \equiv \frac{v^2}{(\Delta v)^2}\frac{r}{v} \ . \qquad (1)$$

For an *isolated* virialized system of N point particles, it is easy to show that this is of order $0.1N/\ln N$ crossing times (eg. Binney & Tremaine 1987). If the time interval one tries to cover for a particular problem is larger than the two-body relaxtion timescale, the problem becomes artificially collisional.

Although many systems are likely to endure physical mechanisms like violent relaxation and phase mixing during some stage of their evolution, an obvious worry is that such physical mechanisms might not actually be important in a given situation, so that two-body interactions are very much unwanted as they might mimic the effects of physical mechanisms. A different problem is that a system might completely evaporate through two-body interactions.

The problem can be alleviated somewhat by softening the particles, which reduces the two-body relaxation timescale to $0.1N/\ln\Lambda$ crossing times (see van Kampen (1995) and references theirein), where $\Lambda = {\rm Min}(R/4\epsilon, N)$, with R the effective size of the system, which we take to be twice the half-mass radius r, and ϵ the softening length of the N-body particles. However, ϵ is necessarily a function of both N and r, as softened particles should not overlap too much. Too large a choice for ϵ will prevent particles from clustering properly and produce haloes which are too extended and too cold (that is, the velocity dispersion is too small). Given that $N/2$ particles reside within r by definition, the mean particle number density within r is $3N/(8\pi r^3)$. The maximum mean particle density desirable is set by the minimum mean nearest neighbour distance for the particles: $n_{\rm max} = 3/(4\pi r_{\rm nn}^3)$. Most often used is Plummer softening, which just means that particles have a Plummer density profile, $\rho(r) \sim (r^2+\epsilon^2)^{5/2}$. The effective force resolution, defined as the separation between two particles for which the radial component of the softened force between them is half its Newtonian value, is $\approx 2.6\epsilon$ for Plummer softening (Gelb & Bertschinger 1994), so we want $r_{\rm nn} \gtrsim 2.6\epsilon$, which gives $n_{\rm max} \approx 0.014/\epsilon^3$. Thus, we find a maximum realistic softening length

$$\epsilon \approx \frac{r}{2N^{\frac{1}{3}}} \ . \qquad (2)$$

For this ϵ the relaxation time becomes $0.3N/\ln N$ crossing times, i.e. three times larger than for the point particles case.

Even though softening alleviates the problem of two-body effects somewhat, softened particle groups are more extended and less strongly bound (van Kampen 1995). This makes them more vulnerable to two-body disruption processes, which are more efficient for larger subhaloes, as shown below. Furthermore, the timescales for *physical* disruption processes are effected. Subhalo-subhalo tidal heating has a timescale inversely proportional to the subhalo size (van Kampen 2000), and is therefore slower, although the lower binding energy might compensate for this. Tidal stripping and disruption will be artificially enhanced, however, because of the larger subhalo size and the weaker binding of the particles inside the group (van Kampen 1995). Because the enhanced tidal disruption due to softening has the same net effect as two-body disruption, which is also enhanced due to the larger subhalo size, the two disruption processes accelerate each other. In the next section we derive two-body disruption timescales without taking into account tidal disruption, and then treat these timescales as upper limits.

3 DISRUPTION TIMESCALES

3.1 Two-body evaporation

This process is internal to haloes and subhaloes in other words, it is a self-disruption process. Two-body interactions between particles within the same (sub)halo change their orbits and velocities, thus every once in a while the velocity will be larger than the escape velocity, and a particle will 'leak' out of the (sub)halo. The timescale for this process is about a hundred times the relaxation timescale (e.g. Binney & Tremaine 1987),

$$t_{\rm dis} \equiv 30 \frac{N}{\ln N} \frac{r}{v} \ , \qquad (3)$$

so for small N this becomes important. As an example, for galaxy haloes, evaporation becomes an issue for $N < 10$, as the crossing time for most galaxy haloes is larger than 0.2 Gyr, independent of their mass.

Moore et al. (1996) tested whether this process gets enhanced for subhaloes within embedding haloes due to the influence of the mean tidal field of the embedding halo. They simulated a collisional group of particles within a *smooth*, and therefore collisionless, isothermal system. They found that the evaporation rate was similar to that for an isolated group, and concluded that *"relaxation effects are not important at driving mass loss from haloes within current simulations"*. Their conclusion is incorrect, however, as they did not consider the particle-subhalo two-body heating process, which we discuss next.

3.2 Particle-subhalo two-body heating

Particles within a subhalo do not just interact amongst themselves (driving the evaporation process describe above), but also with the particles of the embedding halo. As the latter are usually hotter than those of the subhalo, velocity changes to the subhalo particles will always be positive. The process very much resembles the kinetic heating of a cold system that is introduced into a hot bath: an embedding halo 'boils' the subhalo into dissolution. A derivation for the disruption timescale of this process is given by van Kampen (1995, his eq. 15, which is erroneous by a factor of two):

$$t_{\rm dis} \approx \frac{v_{\rm s}^2}{v_{\rm h}^2} \frac{N_{\rm h}}{12 \ln(r_{\rm h}/2\epsilon)} \frac{r_{\rm h}}{v_{\rm h}} \ . \qquad (4)$$

Here N denotes the number of particles, and the subscripts h and s denote embedding halo and subhalo respectively. A similar expression was given earlier by Carlberg (1994, his eq. 13 with his indices c and g swapped, no derivation given):

$$t_{\rm dis} \approx \frac{v_{\rm s}^2}{v_{\rm h}^2} \frac{N_{\rm h}}{8 \ln(r_{\rm h}/\epsilon)} \frac{r_{\rm h}}{v_{\rm h}} \ . \qquad (5)$$

We can rewrite eq. (4), using eq. (2) and the virial theorem for both halo and subhalo, as

$$t_{\rm dis} \approx \frac{N_{\rm s}}{4 \ln N_{\rm h}} \frac{r_{\rm h}}{r_{\rm s}} \frac{r_{\rm h}}{v_{\rm h}} \approx \frac{N_{\rm s}^{\frac{2}{3}}}{8 \ln N_{\rm h}} \frac{r_{\rm h}^2}{v_{\rm h}} \epsilon^{-1} \ . \qquad (6)$$

This timescale is shorter than that for two-body evaporation, by a factor of (van Kampen 1995)

$$100 \frac{v_{\rm h}}{v_{\rm s}} \frac{r_{\rm s}^2}{r_{\rm h}^2} \frac{\ln N_{\rm h}}{\ln N_{\rm s}} \ , \qquad (7)$$

where the virial theorem, $v^2 \sim N/r$, is used for both systems.

Because $r_{\rm h}$ is at least several times $r_{\rm s}$, the disruption time (6) is at least $\approx N_{\rm s}/\ln(N_{\rm h})$ embedding halo crossing times, which covers the range $0.05 - 0.15 N_{\rm s}$ crossing times for $N_{\rm h} \approx 10^3 - 10^9$. Thus, it is a much faster process than two-body evaporation.

3.3 Particle-subhalo tidal heating

A different cause for overmerging was proposed by Moore et al. (1996): the tidal heating of subhaloes by particles of their embedding haloes. Subhaloes are taken to be *collisionless*, and get disrupted through an increase of their internal kinetic energy by tidal distortion from passing N-body particles, which are artificially large as compared to the true dark matter halo particles.

The time-scale for this process as given by Moore et al. (1996; their eq. (3), which is eq. (7-67) of Binney & Tremaine 1987 with the assumption that the *r.m.s.* radius is equal to the half-mass radius) reads

$$t_{\rm dis} \approx 0.03 \frac{v_{\rm h}}{G n_{\rm p}} \frac{m_{\rm s}}{m_{\rm p}^2} \frac{r_{\rm p}^2}{r_{\rm s}^3} \ , \qquad (8)$$

where the subscript p stands for *perturber*. The perturber is an N-body particle of the embedding halo with mass $m_{\rm p}$ and size $r_{\rm p}$, at a distance q from the centre of the embedding halo. Note that the impulse approximation implies $v^2/\Delta v^2 = E/\Delta E$. Moore et al. (1996) then assume the embedding halo to be isothermal, so that $n_{\rm p} \approx v_{\rm h}^2/2\pi G m_{\rm p} q^2$, set the half-mass radius of the subhalo equal to the tidal radius, $q v_{\rm s}/(3 v_{\rm h})$, and assume the subhalo to be virialized. This gives

$$t_{\rm dis} \approx 94 \Big(\frac{v_{\rm h}}{1000 \ {\rm km \ s^{-1}}} \Big) \Big(\frac{r_{\rm p}}{10 \ {\rm kpc}} \Big)^2 \Big(\frac{10^9 {\rm M_\odot}}{m_{\rm p}} \Big) {\rm Gyr} \ . \qquad (9)$$

Relation (8) was originally derived by Spitzer (1958) for the disruption by giant molecular clouds of open star clusters. An important assumption in its derivation is the tidal approximation, which is only valid for impact parameters $b > b_{\rm min}$. Aguilar & White (1985) found that $b_{\rm min}$ should be at least five times the size of *both* the perturber and the perturbed system. Binney & Tremaine (1987) use the tidal approximation down to $b_{\rm min} = r_{\rm cluster} < r_{\rm cloud}$, and introduce a correction factor $g = 3$ to take into account the encounters for which the tidal approximation fails.

Moore et al. (1996) take this result and apply it to tidal interactions between the N-body particles of an embedding halo and its subhaloes theirin. Thus, they set $r_{\rm p}$ to the gravitational softening length ϵ. However, as the size of the perturbers is now *smaller* than the size of the perturbed subhaloes, the tidal approximation, even with the correction factor g included, is only valid for $b > r_{\rm s}$. Therefore, setting $r_{\rm p} = \epsilon$ in eq. (9) is incorrect; instead, one should set $r_{\rm p} = r_{\rm s}$. This means that the timescale becomes $(r_{\rm s}/\epsilon)^2$ times longer. Using eq. (2), the time-scale becomes $4 N_{\rm s}^{2/3}$ times larger than proposed by Moore et al. (1996).

But there is another change to be made, as it is in fact the close encounters of halo particles that do the most damage to the subhaloes. According to Binney & Tremaine (1987), a good estimate for the disruption timescale can be had from an interpolation between the approximations for tidal encounters and for penetrating ($b = 0$) encounters. For each tidal encounter (Binney & Tremaine 1987, their eq. 7-55),

$$(\Delta E)_{\rm tid} = \frac{4 G^2 m_{\rm p}^2 m_{\rm s} r_{\rm s}^2}{3 v_{\rm h}^2} \frac{1}{b^4} \ , \qquad (10)$$

while for each penetrating encounter (Binney & Tremaine

1987, their eq. 7-57)

$$(\Delta E)_{\rm pen} = \frac{4\pi G^2 m_{\rm p}^2}{v_{\rm h}^2} \int_0^\infty \frac{R^3}{(R^2+\epsilon^2)^2} \Sigma_{\rm s}(R) dR\ . \qquad (11)$$

If the perturbed subhalo is an isothermal sphere, i.e. $\Sigma_{\rm s}(R) = v_{\rm s}^2/(6GR) \approx 0.2 m_{\rm h}/(r_{\rm h} R)$, we find

$$(\Delta E)_{\rm pen} \approx \frac{2G^2 m_{\rm p}^2 m_{\rm s} r_{\rm s}^2}{v_{\rm h}^2} \frac{1}{\epsilon r_{\rm s}^3}. \qquad (12)$$

Interpolating contributions from the tidal and penetrating encounters, i.e.

$$\Delta E = \frac{4G^2 m_{\rm p}^2 m_{\rm s} r_{\rm s}^2}{3 v_{\rm h}^2} \frac{1}{b^4 + \frac{2}{3}\epsilon r_{\rm s}^3}\ , \qquad (13)$$

finally allows us to integrate over *all* encounters. Following the procedure of Binney & Tremaine (1987), we simply find eq. (8) with $r_{\rm p}$ $(= \epsilon)$ replaced by $0.52(\epsilon r_{\rm s}^3)^{1/2}$. Following Moore et al. (1996) again we get the same functional form as eq. (9), but the timescale is approximately $2(r_{\rm s}/\epsilon)^{3/2} \approx 5N_{\rm s}^{1/2}$ times *longer* than estimated by Moore et al. (1996).

By definition, $m_{\rm s}/m_{\rm p} = N_{\rm s}$, and we use eq. (2) to get

$$t_{\rm dis} \approx 4.5 N_{\rm s}^{\frac{5}{6}} \frac{r_{\rm h}}{r_{\rm s}} \frac{r_{\rm h}}{v_{\rm h}} \approx 2.2 N_{\rm s}^{\frac{1}{2}} \frac{r_{\rm h}^2}{v_{\rm h}} \epsilon^{-1}\ . \qquad (14)$$

As $r_{\rm h}$ is at least a few times $r_{\rm s}$, the disruption time is at least $20 N_{\rm s}^{5/6}$ embedding halo crossing times. It is also a factor of $18 N_{\rm s}^{-1/6} \ln N_{\rm h} \approx 100$ times longer than the particle-subhalo two-body disruption timescale.

4 CONCLUSIONS AND DISCUSSION

Overmerging is the numerical disruption of subhaloes within embedding haloes. Of the three main two-body disruption processes, particle-subhalo two-body heating is clearly identified as the cause for overmerging. Its timescale is shown to be much shorter than that for the two other processes, two-body evaporation and particle-subhalo tidal heating. Note that softened particles form into more extended subhaloes than is realistic, so they are more vulnerable to these disruption processes (van Kampen 1995), and to possible physical disruption processes as well.

Recently several research groups used simulations with a very high resolution in order to resolve the overmerging problem (Klypin et al. 1999a; Ghinga et al. 1998, 1999; Moore et al. 1999). Unfortunately, different group finders and different definitions for disruption times were used, so a direct comparison of the results is not straightforward. Still, the consensus is that increasing the number of particles overcomes, at least partially, the overmerging problem. However, the resolution needs to be rather high: for N-body simulations on a cosmological scale, this requires the use of at least 10^9 particles, which is not very practical. Furthermore, for the smallest groups the overmerging problem simply remains.

Another option is to include a baryonic component. With the addition of dissipative particles, haloes should be more compact and have a higher central density for the same numbre of particles. However, as Klypin et al. (1999a) remark, there is a limit to this as some fraction of the baryons tend to end up in rotationally supported disks. A more practical problem with dissipative particles is the actual simulation techniques needed, which usually is some form of smoothed particle hydrodynamics (SPH). The resolution of SPH codes is typically not as high as that of N-body codes, so for the purpose of resolving the overmerging problem it is not a useful alternative at present.

A third option is to use halo particles, which prevents overmerging by construction (van Kampen 1995). The idea is that a group of particles that has collapsed into a virialised system is replaced by a single halo particle. Local density percolation, also called adaptive friends-of-friends, is adopted for finding the groups. This is designed to identify the embedded haloes that the traditional percolation group finder links up with their parent halo. By applying the algorithm several times during the evolution, merging of already-formed galaxy haloes is taken into account as well. Once a halo particle is formed, more N-body particles will group around it at later times. If such a group can virialize, it is replaced by a more massive halo particle. This will usually happen in the field. However, for halo particles that end up in overdense regions, the particles that swarm around a halo particles will be stripped quite rapidly.

Once the overmerging problem is resolved down to the subhalo mass-scale one is interested in, the physical processes can be studied. This is becoming feasible for current simulations. However, whether the physical processes themselves are properly modelled using N-body simulations has yet to be proven. The problem of artificially large subhaloes due to softening needs to be solved, for example. Another problem might be the modelling of dynamical friction, which requires a very smooth distribution of particles in the embedding halo in order to produce the wake that generates the drag force.

ACKNOWLEDGEMENTS

I am grateful to John Peacock and Ben Moore for useful discussion, suggestions and comments.

REFERENCES

Aquilar L.A., White S.D.M., 1985, ApJ, 295, 374
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton
Carlberg R.G., 1994, ApJ, 433, 468
Gelb J., Bertschinger E., 1994, ApJ, 436, 467
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1999, astro-ph/9910166
Klypin A.A., Gottlöber S., Kravtsov A.V., Khokhlov A.M., 1999a, ApJ, 516, 530
Klypin A.A., Kravtsov A.V., Valenzuela O., Prada F., 1999b, astro-ph/9901240
Moore B., Katz N., Lake G., 1996, ApJ, 457, 455
Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Spitzer L., 1958, ApJ, 127, 17
van Kampen E., 1995, MNRAS, 273, 295
White S.D.M., Davis M., Efstathiou G., Frenk C.S., 1987, Nat, 330, 451