Nonequilibrium behaviour of finite gravitating systems

Douglas C. Heggie
School of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, UK
E-mail: d.c.heggie@ed.ac.uk

Abstract. The behaviour of \( N \) equal point masses with an inverse square law of attraction is one of the fundamental problems of statistical physics, because of its numerous applications in astrophysics, and its simplicity. But the simplicity is deceptive. From a theoretical point of view this problem is one of the hardest because it is scale-free, the interaction is long-range, and the interaction exhibits a short-range divergence. Therefore theoretical information is best developed for systems with artificial cutoffs at large and small distances. From the point of view of simulations, the problem is hard because the computational effort grows roughly as \( N^3 \), and because of fundamental problems in simulating a chaotic system.

This talk reviews the relationship between these two approaches, with particular emphasis on simulations of isolated systems (i.e. with no boundary). We emphasise the range of time scales on which different non-equilibrium phenomena operate, and focus on those which are driven by relaxation. We discuss the characteristics of core collapse and gravothermal oscillations, where both basic results of statistical mechanics and phenomenological toy models are particularly instructive. We also review the long-term fate of finite isolated systems.

1. Introduction
The classical gravitational \( N \)-body problem may be defined as the task of understanding the solutions of the system

\[
\ddot{\mathbf{r}}_i = - \sum_{j=1,j \neq i}^{N} G m_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3},
\]

where \( \mathbf{r}_i \) is the position of the \( i \)th body with respect to an inertial frame, \( m_i \) is its mass, and \( G \) is the universal constant of gravitation. As L. Spitzer, Jr, once said, “These equations have an appealing – if deceptive – simplicity”. In fact they present us with a never-ending, self-renewing series of challenges, on all fronts: astrophysical, computational, mathematical and theoretical.

In this review we follow two complementary approaches to this task: that of statistical mechanics, and that of simulation. Then we consider the non-equilibrium behaviour of these systems, with particular regard to the time-scales of non-equilibrium behaviour, the phenomena of core collapse and gravothermal oscillations, and finally the long-term evolution of \( N \)-body systems. We shall consider only the most idealised problems of this kind, neglecting the fact that, in applications, the particles may be stars, with their own internal evolution, non-gravitational effects, and external perturbations (Fig.1).
There are ten integrals of the motion, of which the most important is the energy

$$E = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{r}_i^2 - \sum_{1 \leq i < j \leq N} \frac{G m_i m_j}{|r_i - r_j|},$$

and of course the total mass $M = \sum_{i=1}^{N} m_i$ is also conserved. Statistical mechanics and thermodynamics thrive on studying how the behaviour of a system is changed when $E$ changes, but the scale-free nature of gravitation removes this parameter: by scaling, we can use units such that $G = M = 1$ and $E = -1, 0$ or $+1$, and so only the sign of $E$ matters. Indeed there are great qualitative differences between bound ($E < 0$) and unbound ($E > 0$) systems, but we concentrate in this review on the behaviour of bound systems, and so we may take $E = -1$. The only remaining parameter is $N$.

2. Classical Statistical Mechanics
Gravity is a long-range force, which immediately presents a difficulty for statistical mechanics, as energy is non-extensive. (Padmanabhan (1990) gives an excellent introduction to these issues.) Two neighbouring self-gravitating systems of energy $E_{1,2}$ make up a composite system whose
total energy is $E_{1+2} = E_1 + E_2 + W_{12}$, where the last term is the mutual gravitational potential energy of the two systems. Unfortunately there is no useful limit in which $W_{12}$ is negligible, which is a consequence of the long-range nature of gravity.

At short range there is a new obstacle: the singularity in the interaction potential. The normal approach to studying the microcanonical ensemble (i.e. systems of a given energy) is to compute the entropy as

$$S(E) = \ln g(E),$$

where $g(E)$ is the integral representing the volume of the energy hypersurface (in phase space) of all states of a system with energy $E$. Unfortunately this integral diverges at small pair separations, if $N \geq 2$.

Because of these difficulties at short and long range, progress in understanding the problem is made, paradoxically enough, by destroying the scale-invariance, and imposing barriers at short and long range. Then it is found that the microcanonical ensemble has three regimes: (i) a hot regime, in which the particles are kept in by the long-range barrier; (ii) a cold, condensed regime, in which they are held against gravity by the short-range barrier; and (iii) an intermediate regime in which the specific heat is negative, unlike the other two. The canonical ensemble replaces (iii) with a phase transition. Other toy models exhibit the same phenomena (e.g. Lynden-Bell & Lynden-Bell 1977).

The negative specific heat of bound self-gravitating systems in dynamical equilibrium is familiar to astronomers. For a bound system the total energy $E < 0$, while the virial theorem can be used to show that the kinetic energy is $K = -E$ (apart from fluctuations). Therefore as temperature (and $K$) increases, $E$ decreases, i.e. the system has negative specific heat. By a similar argument, the size of the system decreases.

3. Mean-Field Treatment of Self-Gravitating Systems

Within the astrophysics community, much work on self-gravitating systems starts by assuming that a system can be described by the single-particle distribution function, which can be thought of as the number-density of particles in the 6-dimensional phase space whose coordinates are the position and velocity of a particle. The mean field limit of statistical mechanics uses the same description. One can then prove, for example, that equilibrium is isothermal. Then the distribution of velocities is Maxwellian, and in spherical symmetry the structure is well known. This isothermal sphere has infinite radius and mass, and so we often consider (again) a system enclosed in a reflecting wall. In this setting Antonov (1962) made the influential discovery that the isothermal equilibrium is not a local entropy maximum if the density contrast between the centre of the system and the wall is sufficiently large.

Lynden-Bell & Wood (1968) gave a physical explanation of this result. Let the temperature at the centre and at the edge be denoted by $T_c, T_e$, respectively. Imagine a situation of high density contrast, in which there is a tiny central core surrounded by a massive but tenuous halo. If $T_c > T_e$ then heat flows outwards. $T_e$ increases, but only slightly: the halo is held in by the surrounding wall, and so has positive specific heat, but its mass gives it a large heat capacity. $T_c$ also increases, because the core is self-gravitating and has negative specific heat; it also has small mass, and so $T_c$ rises more than $T_e$. This accentuates the flow of heat, which develops a “gravothermal” runaway. Similarly, if we start with a situation in which $T_c < T_e$, these processes run in reverse, a fact whose significance was first appreciated by Sugimoto & Bettwieser (1983). The flow of heat, incidentally, is mediated by small-angle gravitational scatterings between individual pairs of particles.

4. Simulation

Beyond insights such as these, computer simulation provides much of our information on the behaviour of self-gravitating $N$-body systems. The problem is a demanding one, however. The
effort taken by computation of all forces scales as $N^2$, while the evolution of a system in near-dynamic equilibrium, which takes place on the time scale of the Chandrasekhar relaxation time (Chandrasekhar 1942), varies approximately as $N$ (in appropriate units; see Sec.5). Hence the cost of each simulation varies roughly as $N^3$. Various means of accelerating this task have been devised. In software, there are techniques for computing forces (e.g. hierarchical [tree] codes) which can reduce the complexity by about one power of $N$ (e.g. Dehnen 2000). Acceleration can also be done in hardware, and the GRAPE project at the University of Tokyo has succeeded spectacularly in providing researchers with very fast, very cost-effective hardware accelerators for inverse-square forces (Makino et al 2003).

More fundamental difficulties for simulation are caused by the chaotic nature of the solutions of eq.(1), and the resulting exponential divergence of neighbouring orbits (Miller 1964, Goodman et al. 1993). For example, in a typical simulation with particles of equal mass, an initial rounding error of order $10^{-16}$ grows to $O(1)$ after about 5 typical time units, where simulations may easily last $10^3$ units or more. In fact it was realised early in the development of $N$-body simulations in astronomy that results of the same calculation performed with different hardware and/or software differ grossly in detail (Lecar 1968). Those working in this field believe that the statistical results of careful calculations are, nevertheless, reliable, though little has been done to justify this assertion. An example of what is meant by a statistical result is the mass-weighted median radius of the particles in a nearly spherical system, which astronomers refer to as the “half-mass radius”.

5. Non-Equilibrium Behaviour of the Gravitational $N$-Body Problem

5.1. Initial conditions

For the remainder of the paper we shall summarise the different phases in the evolution of a particular kind of $N$-body system. The initial conditions are specified by some realisation of the single-particle distribution function, $f$, which is taken to be an $n = 5$ polytrope\(^2\) Referred to in the astronomical literature as “Plummer’s model” (Fig.1), it leads to a relation $\rho \propto |\phi|^5$ between the density and potential, in the mean field approximation.

This model is in approximate dynamic equilibrium, in a sense which we now discuss. Let $R$ be a measure of the length scale (such as the initial half-mass radius). Then the time scale of orbital motions is the crossing time, which is of order

$$t_{cr} \sim \sqrt{\frac{R^3}{GM}}.$$ 

On this time scale, the assertion is that the statistical properties of the model are constant, within fluctuations.

5.2. Core collapse

Though the model may be in dynamical equilibrium, it is not in thermal equilibrium, and we expect the gravothermal runaway to occur. The time scale on which this occurs is the relaxation time scale, $t_{rel}$, and, as already noted, we have

$$t_{rel} \sim N t_{cr}$$

(see, e.g., Spitzer 1987). The runaway manifests itself as a rise in the central density (Fig.2), $\rho_c$. Simplified models (in which the evolution of $f$ is modelled by a Fokker-Planck equation, for

1 Actually the exponential growth of errors saturates a little below this value (Hut & Heggie 2002)
2 Strictly, this refers to a self-gravitating gas (Chandrasekhar 1939), and we should refer to the appropriate analogue (see, e.g., Binney & Tremaine 1987).
example) in fact imply that $\rho_c$ tends to infinity as time $t$ approaches a finite value, $t_{\text{coll}}$. Indeed this is an example of finite-time blow-up, which is common in nonlinear evolution equations. As often happens in such situations the late stages of this evolution are nearly self-similar, i.e. the spatial structure remains the same except for time-dependent scalings. Here, for example, the radius of the core (defined to be that part of the system where the space density exceeds half its central value) is related to $\rho_c$ by $\rho_c \propto r_c^{-2.2}$ approximately (Lynden-Bell & Eggleton 1980). Thus $r_c \rightarrow 0$ as $t \rightarrow t_{\text{coll}}$, and it also follows that the number of particles in the core, $N_c$, has the same behaviour. For these reasons the process is referred to as core collapse. At any instant in the approach to core collapse, the density profile also reflects these scalings.

In an $N$-body model the density does not become infinite, and the collapse ceases when three-body processes become competitive with two-body scattering. The relevant three-body processes are exothermic on average. For example, three single particles may come together to form a binary (whose internal degrees of freedom have a negative energy) and a single particle which acts as a catalyst. The kinetic energy associated with translational degrees of freedom therefore increases. This energy is “created” inside the core, whose collapse therefore stops, eventually.

The time scale of binary formation is of order $N^2 t_{\text{cr}}$, but for processes in the core we replace $N$ by $N_c$. Then a comparison with eq.(2) shows that this time scale becomes comparable with the relaxation time scale when $N_c$ is of order unity. Since $N_c \propto \rho_c r_c^3$, we see that $\rho_c \propto N_c^{-2.8}$ approximately. The initial central density of a Plummer model is independent of $N$, and so the central density at the end of core collapse should scale as $\rho_c \propto N^{2.8}$ approximately, as is borne out by simulations.

5.3. Post-collapse evolution

At the end of core collapse there is, in physical terms, a balance between heating by binaries and collapse by two-body encounters. We may suppose that thermodynamic equilibrium is
approached locally, first in the core and subsequently in a surrounding zone of increasing radius. (In practice this has not been observed, as the temperature gradients are masked by fluctuations.) When the radius of the isothermal zone becomes large enough, this zone becomes unstable to the gravothermal instability. If fluctuations give rise to a slight temperature inversion, in which the core is cooler than its surroundings, then the core will expand. This continues until the zone in which there is a temperature inversion comes into contact with the cooler outlying envelope of the system. Then the temperature inversion disappears, and the system is free to undergo core collapse once again, much as it did when starting from the initial conditions.

The resulting cycle of collapse, binary formation, temperature inversion, expansion and recollapse is referred to as a “gravothermal oscillation”. This behaviour was discovered by Sugimoto & Bettwieser (1983) using a simplified model in which the system is treated as a self-gravitating gas, with a thermal conductivity designed to transport thermal energy on the time scale of the relaxation time. They were subsequently confirmed in $N$-body simulations by Makino (1996), and have also been investigated using a Fokker-Planck equation (e.g. Cohn, Hut & Wise 1989) and even a toy model described by a low-order dynamical system (Allen & Heggie 1992).

One reason for the long gap between the discovery of gravothermal oscillations and their confirmation in an actual $N$-body simulation is their dependence on $N$ (the one scale-free parameter in our problem). For sufficiently small $N$ (a few thousand or less) they do not occur, and their confirmation had to await the development of sufficiently fast hardware to simulate a large enough system. For the same reason, the way in which the nature of the oscillations changes in $N$-body simulations as $N$ becomes even larger is only now beginning to be revealed.

Though what has been found remains conjectural until it is established by $N$-body simulations, the simplified models have been very useful for investigating the way in which gravothermal oscillations vary with $N$. Small-amplitude oscillations have been studied with particular care (Goodman 1987). If the overall post-collapse expansion of the system (see below) is scaled out, this phase of evolution becomes an equilibrium, and it has been shown that it becomes unstable through a period-doubling bifurcation as $N$ increases through a critical value. As $N$ increases further, the oscillations evolve along the usual period-doubling route to chaos. A property of well-developed oscillations that has not yet been studied adequately is their apparent self-similarity (Fig.4).

5.4. Late stages of evolution; escape

With or without gravothermal oscillations (depending on the size of $N$), the release of energy by binaries causes a steady exansion of the system (see end of Sec.2). The time scale is not determined by conditions in the core, however, but by the flow of energy in the bulk of the system (Hénon 1975). Conventionally, researchers measure such quantities at the median radius of the particles (the “half-mass” radius). The time scale of the expansion is of order 10 times the relaxation time at that radius.

On this time scale the distribution of velocities becomes increasingly anisotropic, with a preference for radial motions towards or away from the densest part of the system. The profiles of both anisotropy and density (i.e. $\rho(r)$) approach what appears to be almost universal shapes, independent of $N$ (Baumgardt et al 2002), after time-dependent scaling. The reason for this universal structure is unknown.

On an even longer time scale, particles escape from the bound system. For example, for a system with $N = 8192$ bound members initially, at a time when half of the particles are escaping, it has already expanded by a factor of over $10^3$. There appear to be three mechanisms for escape. Most escapers are created by the classical mechanism of two-body encounters, in which one particle gains enough energy to escape. Another class of escapers (including escaping binary pairs) arises from three-body interactions mainly taking place in the core. Finally, as
Gravothermal oscillations in a “gas” model, in which the thermal conductivity is chosen to transport energy on the relaxation time scale. The behaviour of binaries is mimicked with a heating term proportion to the theoretical rate of formation of binaries in three-body encounters if the number of particles in the system were of order $4 \times 10^5$. The central density is plotted against time. The right-hand diagram shows an expanded version of one of the maxima in the figure on the left. This plot resembles a scaled version of the section of the left-hand diagram from about $t = 2$ to $t = 2.8$.

these two kinds of escapers cause a decrease in the mass of the system, outlying loosely bound single particles become unbound without any few-body interaction.

It is usually assumed that this process of escape continues until there is one binary left, surrounded by an expanding cloud of single particles and binaries, and perhaps some higher-order stable groups such as hierarchical triple particles (in which a distant companion is in a bound orbit about a close binary). In fact this picture, though plausible, is not established, as useful rigorous criteria for escape do not exist. What is usually done is to assert that a particle will escape if (a) it lies at a distance exceeding some multiple of the half-mass radius, and (b) its energy is positive, and (c) its radial velocity is positive. According to one estimate (Heggie & Hut 2003), the lifetime of an isolated $N$-body system varies as $N^k t_{cr}$, where $k \approx 13$.

Acknowledgments
I am indebted to my colleagues S.J. Aarseth, H. Baumgardt, P. Hut and J. Makino for many discussions and collaborations on the topics of this review. I warmly thank the symposium organisers for making it possible to attend.

References
Allen FS, Heggie DC. 1992 MNRAS 257 245-56
Antonov VA. 1962 Vest. Leningrad Univ. 7, 135; english transl. in Goodman J, Hut P. 1985 Dynamics of Star Clusters IAU Symp.113, Reidel, Dordrecht, 525-40
Baumgardt H., Hut P., Heggie D. C., 2002, MNRAS, 336, 10
Binney J, Tremaine S. 1987 Galactic Dynamics Princeton University Press
Chandrasekhar S. 1939 *An Introduction to the Study of Stellar Structure* University of Chicago Press

Chandrasekhar S. 1942 *Principles of Stellar Dynamics* University of Chicago Press

Cohn H, Hut P, Wise M. 1989 *ApJ* 342 814-22

Dehnen W. 2000 *ApJ* 536 L39-L42

Goodman J. 1987 *ApJ* 313 576-95

Goodman J, Heggie D, Hut P. 1993 *ApJ* 415 715-33

Heggie DC, Hut P. 2003 *The Gravitational Million Body Problem* Cambridge University Press

Hénon M. 1975 *IAUS* 69 133-49

Hut P, Heggie DC. 2002 *JSP* 109 1017-1025

Lecar M. 1968 *Bull Astron* 3 91-104

Lynden-Bell D, Eggleton PP. 1980, *MNRAS* 191 483-98

Lynden-Bell D, Lynden-Bell RM. 1977 *MNRAS* 181 405-19

Lynden-Bell D, Wood R. 1968 *MNRAS* 138 495-525

Makino J. 1996 *ApJ* 471 796-803

Makino J, Fukushima T, Koga M, Namura K. 2003 *PASJ* 55 1163-87

Miller RH. 1964 *ApJ* 140 250-6

Padmanabhan T. 1990 *Phys. Rept.* 188 285–

Spitzer L., Jr. 1987 *Dynamical Evolution of Globular Clusters* Princeton University Press

Sugimoto D, Bettwieser E. 1983 *MNRAS* 204P 19-22