Diffusion and Mixing in Globular Clusters

Yohai Meiron and Bence Kocsis

Institute of Physics, Eötvös University, Pázmány P. s. 1/A, Budapest, 1117, Hungary

Abstract

Collisional relaxation describes the stochastic process with which a self-gravitating system near equilibrium evolves in phase-space due to the fluctuating gravitational field of the system. The characteristic timescale of this process is called the relaxation time. In this paper, we highlight the difference between two measures of the relaxation time in globular clusters: (1) the diffusion time with which the isolating integrals of motion (i.e., energy E and angular momentum magnitude L) of individual stars change stochastically and (2) the asymptotic timescale required for a family of orbits to mix in the cluster. More specifically, the former corresponds to the instantaneous rate of change of a star's E or L, while the latter corresponds to the timescale for the stars to statistically forget their initial conditions. We show that the diffusion timescales of E and L vary systematically around the commonly used half-mass relaxation time in different regions of the cluster by a factor of $\sim$10 and $\sim$100, respectively, for more than 20% of the stars. We define the mixedness of an orbital family at any given time as the correlation coefficient between its E or L probability distribution functions and those of the whole cluster. Using Monte Carlo simulations, we find that mixedness converges asymptotically exponentially with a decay timescale that is $\sim$10 times the half-mass relaxation time.

Key words: diffusion – globular clusters: general – stars: kinematics and dynamics

1. Introduction

Star clusters evolve under many external and internal factors. Externally, depending on the star cluster's location in the host galaxy, gravitational perturbations (in the form of a tidal field), an occasional collision with a giant molecular cloud or another star cluster, or a supermassive black hole all affect the structure of clusters on timescales that are fairly short compared to the age of the universe. Internally (after a short phase in which gas dynamics and violent relaxation dominate, ending in a cluster that is nearly gas free and spherical), stellar evolution causes individual stellar masses to generally decrease and may provide ambient gas for the formation of a second generation of stars. Energy and angular momentum exchange due to gravitational interactions between stars (single, binary, or multiple) leads to a plethora of phenomena such as mass segregation, core collapse, and cluster evaporation (Spitzer 1987; Binney & Tremaine 2008; Merritt 2013; and references therein). This so-called collisional evolution contrasts with the evolution of systems on galactic or cosmological scales due to the large-scale gravitational field. Each of these factors has been studied independently, and all of them in tandem in more modern star cluster modeling, which includes realistic prescriptions for much of the physics involved (Sippel & Hurley 2013; Heggie 2014; Wang et al. 2016).

Relaxation is the idea at the heart of collisional evolution. It is conceptually useful to split the gravitational potential of the star cluster into an approximately time-independent smooth potential and a fluctuating component that accounts for time-dependent finite number effects (i.e., dynamical two-body encounters and resonant[1] multibody interactions; see, e.g., Fouvry & Bar-Or 2017). In spherical star clusters, the smooth component allows four independent isolating integrals (energy and three angular momentum vector components) to exist.

The fluctuating component drives slow variations in their values, a process in which the six-dimensional phase-space distribution function evolves toward the maximum entropy configuration. This effect is described approximately by diffusion (Chandrasekhar 1942 and many references thereafter; cf. Bar-Or et al. 2013, who discuss anomalous diffusion in galactic centers).

The modern idea of relaxation was first introduced in thermodynamics by Maxwell (1866). Jeans (1913) applied this to stellar dynamics by assuming that stars, like molecules in a gas, are subjected to thermal agitation. He estimated the timescale associated with relaxation based on deflection angles and the idea of mean free paths. Around the same time, Karl Schwarzschild was working on the velocity distribution in the Galaxy. He investigated how a Maxwellian velocity distribution may be produced in stellar systems. He also derived the timescale for this to occur based on perturbations to the orbital energy due to successive stellar encounters (Schwarzschild 1924; published posthumously[2]). Chandrasekhar (1942) extended those ideas and provided a rigorous evaluation of those timescales (based on both deflection angles and energy); his scattering theory is discussed in more detail in Section 3. These early authors considered a very simplified model for a star cluster: infinite and homogeneous. In this case, the rate of diffusion is of course the same everywhere. More modern kinetic approaches have been developed that account for spatial inhomogeneity and collective effects (Heyvaerts 2010; Chavanis 2012, 2013; Sridhar & Touma 2016).

Fundamental differences between the intermolecular forces in gas and the gravitational force (such as its long range and always-attractive nature) lend star clusters very different thermal properties. Nevertheless, models of a star cluster as a gravitating gaseous sphere (analogous to a star; sometimes called fluid-dynamical models) were developed in the late

---

[1] Resonant in the sense of the commensurability condition in the mean field potential (e.g., Merritt 2015, and reference therein).

[2] The paper fragment (in German) is also found in Voigt (1992) with a foreword by R. Wielen.

1970s and were quite successful in investigating core collapse (Hachisu et al. 1978; Lynden-Bell & Eggleton 1980; for the original connection see Lynden-Bell & Wood 1968). More elaborate Fokker–Planck models followed (e.g., Cohn 1979, following from earlier works such as Hénon 1961 and Kuzmin 1957), based on orbit-averaged diffusion coefficients, which required fewer assumptions than the gaseous models. These more accurate models made it possible to study the time evolution of star clusters, in the continuum (large N) limit, with a relatively modest computational effort. They considered a cluster's inhomogeneous density profile, where now the relaxation time may vary considerably between the inner and outer parts of the cluster.

These previous works have devoted much attention to the collisional evolution of the star cluster as a whole, but not much focus has been given to the evolution of particular orbital families, i.e., stars with similar (initial) values of total energy E and angular momentum magnitude L with respect to the cluster's center. Furthermore, multiple interpretations have been offered for the term *relaxation time*—namely, that it is the (mean) time for a quantity to change by an order of itself, or that it is the timescale for a star to statistically "forget its initial conditions." Under the circumstances relevant for a star cluster, however, those definitions are not the same as the diffusion timescale. While the diffusion time is an instantaneous timescale, those definitions describe a long-term behavior we call *mixing* and discuss further and quantify in Sections 2 and 4.

Since nonresonant two-body relaxation is the main process responsible for both E and L exchange in star clusters, in this paper we use the term *relaxation time* for the timescales associated with the diffusion of either E or L. The direction of the angular momentum vector is also a constant of motion that is affected by two-body relaxation; in this case, however, vector resonant relaxation (Rauch & Tremaine 1996) may play a role as well. The relative importance of vector resonant versus two-body relaxation in star clusters (specifically, a globular cluster lacking a central singularity) will be investigated in a future paper (Y. Meiron & B. Kocsis 2018, in preparation).

In this paper, rather than studying the collisional evolution of the star cluster as a whole, we focus on the evolution (in the statistical sense) of particular orbital families, i.e., stars that are initially in a small neighborhood of a point in (E, L)-space. We use an idealized isolated Plummer model (that is isotropic) as an example, where all stars have the same mass, their masses are constant in time, and there are no binaries. The two aspects of this study are to find the E- and L-relaxation times as a function of E and L and to statistically follow representative orbital families in time and quantify their degree of mixing using a quantity we call mixedness.

In Section 2, we discuss diffusion in general terms; in Section 3, we quantify the relaxation time and calculate it for different orbital families in a Plummer model; in Section 4, we discuss the concept of mixing and how it is quantified by mixedness, which we measure for representative orbital families in a Plummer model; finally, we discuss general and astrophysical significance in Section 5.

2. Short- and Long-term Behavior

Diffusion due to two-body encounters is the dominant cause of change of the energy and angular momentum of individual particles in a system in equilibrium.[3] A secondary cause is the gradual change of the global potential (which in our case is indirectly due to two-body encounters but in the general case could be due to other reasons, e.g., a change of the tidal field due to the cluster's motion through a galaxy). A timescale associated with the diffusion time of any constant of motion is commonly referred to as the relaxation time.

The rate of diffusion, expressed by the diffusion coefficients, is not uniform and depends on phase-space coordinates. Since individual particles are generally not stationary and move in phase-space even without diffusion, the concept of a relaxation time is meaningful if it is associated with some kind of averaged diffusion coefficient. Even so, the relaxation time is only an instantaneous timescale (Chandrasekhar 1942) akin to the local slope of a curve. In this paper, we identify the relaxation time with the instantaneous ensemble-average of the diffusion time of the energy or angular momentum magnitude of individual stars. We calculate the diffusion time as the average initial rate of the square change of energy and angular momentum due to two-body encounters with other stars in the cluster.

The mixing of a property x is the tendency of a distribution of x of any subpopulation in the cluster to evolve toward the distribution in the whole cluster (which itself may be changing in time due to collisional evolution); it occurs due to diffusion and therefore is not a separate physical process. This then describes the long-term behavior of a system as opposed to the instantaneous diffusion time. The subpopulation is a set of particles with a very narrow (initial) distribution of x, and x in our case is a constant of motion (i.e., a quantity that would not change in the absence of diffusion). In particular, for an approximately spherically symmetric cluster, it is a set of orbits with nearly the same (initial) semimajor axis and eccentricity[4] (but arbitrary orientations of the orbital plane, orbital phase, etc.) Consider the energy (for example) distribution of an orbital family: it is initially very narrow but widens with time due to diffusion. We expect that it will approach asymptotically to the energy distribution of all particles in the system. We quantify the amount of mixing using a quantity called mixedness, defined in Section 4.

In the following sections we explore the concepts of relaxation and mixing in a more detailed way, using the Plummer model to illustrate each one.

3. Relaxation Time

3.1. Basic Concepts

The following formula is often used to estimate the relaxation time in a stellar system:

$$t_{\text{relax}} = \frac{\alpha \sigma^3}{G^2 m \rho \ln \Lambda}, \tag{1}$$

---

[3] Collisionless equilibrium; see chapter 4 of Binney & Tremaine (2008).

[4] In non-Keplerian spherically symmetric potentials, orbits are planar "rosettes" rather than ellipses, but one can still define orbital elements geometrically. The pericenter r_p and apocenter r_a are the radial turning points of motion that satisfy $E = \Phi_{\text{eff}}(r, L)$, where $\Phi_{\text{eff}}(r, L) = \Phi(r) + \frac{1}{2}(L/r)^2$ is the effective radial potential, $\Phi(r) = \sum_{i=1}^{N} Gm/|\mathbf{r}_i - \mathbf{r}|$ is the potential, and E and L are respectively the energy and angular momentum per unit mass that are approximately conserved in an approximately spherical cluster. The semimajor axis and eccentricity are defined as $a = \frac{1}{2}(r_p + r_a)$ and $e = (r_a - r_p)/(2a)$.

where $\sigma \equiv \sqrt{\langle v^2 \rangle / 3}$ is the one-dimensional velocity dispersion, m is a particle's mass, ρ is the particle mass density, and G is the gravitational constant. The dimensionless factors α and $\ln\Lambda$ hide much of the complexity of the problem; they vary depending on the exact definition and will be discussed in the following paragraph. This formula has proved quite useful, but it is important to understand its caveats. The derivation of this type of formula (Chandrasekhar 1942; Spitzer 1987; Binney & Tremaine 2008) requires making several assumptions. The first and perhaps most critical is that diffusion can be adequately described by a superposition of independent two-body interactions. Additional assumptions include the uniform spatial density (which implies the neglect of self-gravity), the isotropy of the velocity field, and the fact that the velocities follow the Maxwell–Boltzmann distribution characterized by σ. Those assumptions are really valid only in a hypothetical infinite and homogeneous medium, where the mean-field gravity is neglected. To apply this to the case of star clusters as an average, global quantity, additional approximations are made: the density is equal to the average, density within the half-mass–radius $r_{\rm h}$, and $\sigma = \beta\sqrt{GmN/r_{\rm h}}$ (based on the virial theorem, with β an order of unity constant). The result is what is commonly called the *half-mass relaxation time*:

$$t_{\rm rh} = \frac{\gamma}{\ln\Lambda}\sqrt{\frac{Nr_{\rm h}^3}{Gm}} = \frac{\gamma N t_{\rm dyn}}{\ln\Lambda}, \qquad (2)$$

where $\gamma = 8\pi\alpha\beta^3/3$, N is the total number of stars in the cluster, and $t_{\rm dyn}^2 \equiv r_{\rm h}^3/GM$ is the dynamical time ($M = mN$ is the total cluster mass). Spitzer (1987) got a value of $\gamma = 0.138$ from simple considerations. In the following sections we use $t_{\rm rh}$ as a reference time since it is a very simple estimate to make, and despite the many approximations it retains the correct scaling with the number of particles. Thus, the results discussed throughout this paper are independent of N.

Chandrasekhar (1942, and references therein) proposed a definition for relaxation time based on energy, leading to an expression of the form of Equation (1), where the relaxation time is said to have been reached when the cumulative square change of energy $(\Delta E)^2$ becomes of the same order as the square of the initial kinetic energy. The average $\langle (\Delta E)^2 \rangle$ per unit time (the diffusion coefficient) is meticulously calculated by considering the root mean square (rms) energy change due to two-body encounters sampled independently from a homogeneous uniform medium within a given minimum and maximum impact parameter and a Maxwell–Boltzmann velocity distribution. This gives $\alpha = 9/(16\sqrt{\pi}) \approx 0.317$.

The approach taken by Spitzer (1987)[5] and Binney & Tremaine (2008, and references therein) is essentially identical with respect to summing up individual encounters, but they compute the change in velocity components during the encounters rather than energy. Since as noted above the

medium is approximated as infinite and homogeneous, the energy and the velocity magnitude are interchangeable.[6] The mean change in the parallel component of the velocity $\Delta v_\parallel$, its square $(\Delta v_\parallel)^2$, and the square change of the perpendicular component $(\Delta v_\perp)^2$ are calculated per unit time to yield diffusion coefficients. One advantage in this method is that it gives the expression for dynamical friction for free through the so-called drift term $\langle \Delta v_\parallel \rangle$. Another advantage is that up to the second order, the diffusion of coefficients of any quantity (such as energy) can be written as a linear combination of the velocities. We make use of this property to derive the local angular momentum diffusion coefficients in Appendix A. An additional difference to Chandrasekhar (1942)'s approach is that instead of assuming a foreground velocity distribution equal to the background velocity distribution and averaging over both, these authors more simply substitute a typical value for the velocity for the test star, which gives a value of $\alpha \approx 0.340$ (the exact value can be written as a complicated expression involving the error function).

The factor $\ln\Lambda \approx \ln(b_{\max}/b_{\min})$ is the Coulomb logarithm, which crops up in the derivation due to the divergence of the integral over the impact parameters. This divergence occurs on small scales due to the fact that the small-angle deflection approximation mishandles strong collisions, and it also occur on a large scale due to the local approximation (i.e., the neglect of inhomogeneity). Chandrasekhar (1942) interpreted $b_{\max}$ as the interparticle distance, but Cohen et al. (1950) point out that it should be the order of the size of the system (or the region that contains most of the particles). Due to the uncertainty, in star clusters investigations $\ln\Lambda$ is generally set to $\ln(\lambda N)$, where the value of λ could be empirically determined from N-body simulations (Giersz & Heggie 1994 got $\lambda \approx 0.11$).

Larson (1970) derives a timescale for the collisional evolution in a substantially different way. His work is based on reorganizing the Fokker–Planck equation as a set of moment equations. The velocity moments, representing different kinds of deviations from a Maxwellian velocity distribution, are shown to decay exponentially with timescales similar (i.e., up to an order of unity factor) to the relaxation time defined by Chandrasekhar (1942; Equation (1)), with an apparent tendency for the higher moments to relax more slowly than the lower ones.

While Equation (2) is a very useful timescale parameter for a star cluster, it hides the very important information on how the diffusion timescale depends on the location within the cluster (e.g., the central regions versus the outskirts). Equation (1) is a bit more general and one could in principle substitute for a spherically symmetric star cluster $\sigma(r)$ and $\rho(r)$ corresponding to a particular model and obtain an expression for the relaxation time that is a function of the radius. This is somewhat an abuse of Equation (1) as it is derived under the assumption that σ and ρ are constants and the medium is infinite. In addition, this kind of calculation will not yield the dependence on the test star's eccentricity. In the two following subsections, we calculate the relaxation time (for both energy and angular momentum) for each orbital family in a Plummer model by using Chandrasekhar (1942)'s basic scattering theory but not making further assumptions apart from the isotropy of the model (which is justified for a Plummer model).

---

[5] Both Chandrasekhar (1942) and Spitzer (1987) give one additional definition each for the relaxation time. The former's is based on the deflection angle rather than the energy change, and the latter's is the time-dimensioned constant of the encounter term in the Fokker–Planck equation.

[6] If the mean field is spatially homogeneous, then the three velocity components are three integrals of motion that change due to the stochastic fluctuating component of the potential.

3.2. Methods

We calculate the E- and L-diffusion coefficients and the corresponding relaxation timescales for a specific (E, L) orbital family due to stellar scattering using the local diffusion approximation. This amounts to adding up the contributions of incoherent local two-body flyby encounters, assuming that (1) the flyby events have a short duration relative to the orbital timescale and (2) that the encounters are predominantly local, where the density of scatterers is approximately homogeneous. The changes in E and L are accumulated incoherently over the unperturbed orbit in the cluster with a given E and L, and the relaxation times follow from there.

Rosenbluth et al. ([1957](#)) gave implicit expressions for the average change per unit time of a test particle's velocity components $\langle \Delta v_i \rangle$, as well as $\langle \Delta v_i \Delta v_j \rangle$, due to two-body encounters, which depend on the target particle's velocity and the background particles' velocity distribution. Under the assumption of an isotropic velocity field, these expressions could be simplified to yield three useful functions expressing the average change parallel to the original direction of velocity $\langle \Delta v_\parallel \rangle$ as well as the square change in the parallel direction $\langle (\Delta v_\parallel)^2 \rangle$ and perpendicular $\langle (\Delta v_\perp)^2 \rangle$ to the original velocity vector (note that isotropy of the velocity field necessitates $\langle \Delta v_\perp \rangle = 0$). Remember that these are changes per unit time, despite the notation.

Given these functions, we can express (up to the second order) the average square change per unit time of the energy and angular momentum:

$$\langle (\Delta E)^2 \rangle = v^2 \langle (\Delta v_\parallel)^2 \rangle, \tag{3}$$

$$\langle (\Delta L)^2 \rangle = \frac{r^2}{v^2} \Big[v_t^2 \langle (\Delta v_\parallel)^2 \rangle + \tfrac{1}{2} v_r^2 \langle (\Delta v_\perp)^2 \rangle \Big], \tag{4}$$

where r is the radial coordinate (the cluster center is at the origin of the coordinate system), v_r is the velocity component in the radial direction, v_t is the tangential velocity component, and $v^2 = v_r^2 + v_t^2$. Equation ([3](#)) is very simple to derive: one just needs to remember that due to assumption (1) above (short-duration encounters), the potential energy does not change during an encounter and velocity terms with powers higher than two are neglected (Spitzer [1987](#), Equations (2)–(51) therein). Equation ([4](#)), on the other hand, is more difficult to derive. The full derivation is given in Appendix [A](#).

In Appendix [B](#), we substitute Rosenbluth's expressions for $\langle (\Delta v_\parallel)^2 \rangle$ and $\langle (\Delta v_\perp)^2 \rangle$ (both scale linearly with density and are otherwise functions of v only) and perform orbital averaging. The result is two functions denoted by $D_{E^2}(E, L)$ and $D_{L^2}(E, L)$ called the orbitally averaged diffusion coefficients. Cohn & Kulsrud ([1978](#)) have similarly derived the diffusion coefficients, but instead of L they considered the square relative angular momentum and performed orbital averaging under the assumption of Keplerian orbits (see also Merritt [2015](#)).

It is natural to define the relaxation time for any property x simply as x^2 / D_{x^2}, but this may lead to strange results. For example, if we defined the E-relaxation time as E^2 / D_{E^2}, then it would *decrease* from the center of the cluster outward, which goes against the intuition that diffusion is more important in the innermost regions. Further, on physical grounds we may note that the gravitational scattering process, assuming local short-duration encounters, changes the instantaneous velocity directly, hence the instantaneous kinetic energy, while the

potential energy with respect to the cluster is fixed. Therefore we define the energy relaxation time specifically with respect to the average *kinetic* energy for a star on an (E, L) orbit $\overline{E_k}$. While the kinetic energy is not a constant of motion, its orbital average (by definition) is. Unlike in the Keplerian case where $\overline{E_k} = |E|$, in star clusters $\overline{E_k}$ strongly differs from this result especially in the inner regions of the cluster where it is much smaller than $|E|$. It generally depends on both E and L through the potential.

A related issue for L-relaxation is that if it were defined as L^2 / D_{L^2}, then any nearly radial orbit would have an asymptotically zero relaxation time. Indeed, if the direction of angular momentum vector is nearly a null-vector, then its direction physically changes by an arbitrarily high rate for nearly radial orbits due to any finite torque. However, since we are mainly interested in the long-term evolution of the magnitude of the angular momentum vector, we choose the reference angular momentum to be the maximal (circular) angular momentum L_c corresponding to the given energy of the orbit. Using this definition, the L-relaxation time is simply proportional to the inverse net torque exerted on the orbit due to two-body encounters, which is finite even for $L = 0$. This definition represents an upper limit of the actual angular momentum diffusion time.

Thus, the relaxation times are defined as follows:

$$t_{\mathrm{rx},E} \equiv \overline{E_k}^2 / D_{E^2} \tag{5}$$

and

$$t_{\mathrm{rx},L} \equiv L_c^2 / D_{L^2}, \tag{6}$$

where except for L_c, which is only a function of E, all other variables are functions of both E and L.

### 3.3. Results

In Figures [1](#) and [2](#), we present the variation of the diffusion coefficients in (E, L)-space for a Plummer model (see Appendix [B](#)). The units of E and L in the figures are given in Hénon units[7] for a Plummer model that is normalized such that its virial radius is unity (giving a Plummer radius of $3\pi/16$). The half-mass–radius in this case is $r_h \approx 0.7686$ Hénon length units. The diffusion coefficients are presented on a logarithmic scale as dimensionless quantities that are independent of the number of particles or the choice of Coulomb logarithm. In order to achieve that, we normalize the diffusion coefficients by multiplying them by

$$\frac{N}{\langle x^2 \rangle \ln \Lambda}, \tag{7}$$

where x is either E or L and $\langle x^2 \rangle$ indicates its mean square for the entire cluster. We can obtain those cluster averages by performing the appropriate integrals on the distribution function (Equation ([27](#)) in the Appendix). In Hénon units for our given model, $\langle E^2 \rangle = 704/(105\pi^2) \approx 0.68$ and $\langle L^2 \rangle = 9\pi^2/256 \approx 0.35$.

---

[7] Also known as N-body units, where the model's total mass and the gravitational constant G are set to unity, and the total energy of the model is set to $-1/4$. For a cluster with a total mass of M and a Plummer radius of r_0 (specified in whatever physical units), the Hénon energy and angular momentum (per unit mass) units are $[E] = \frac{3\pi}{16}\frac{GM}{r_0}$ and $[L] = \sqrt{\frac{16}{3\pi}GMr_0}$, respectively.
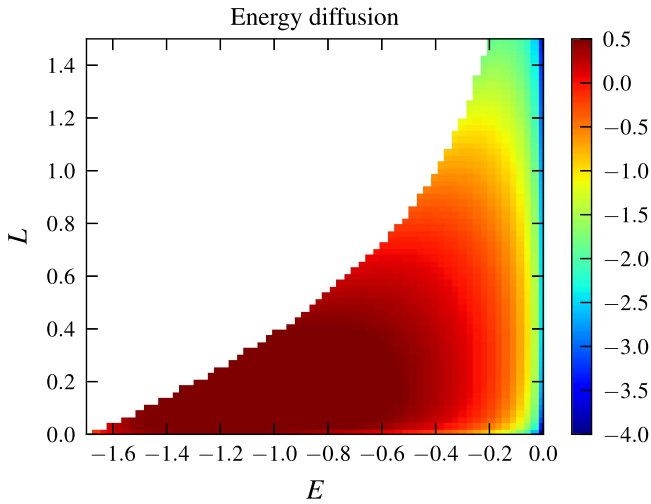
Figure 1. Energy diffusion coefficient for a Plummer model, calculated by orbital averaging for each point on a grid in (E, L)-space. The color scale is $\log_{10}[ND_{E^2}/(\ln\Lambda\langle E^2\rangle)]$ where N is the number of particles, $\ln\Lambda$ is the Coulomb logarithm, and $\langle E^2\rangle$ is the mean square energy of particles in a Plummer model. This normalization guarantees dimensionlessness and independence of the number of particles. The axes are in Hénon units for a Plummer model with a virial radius of one Hénon length unit (see the text for details).
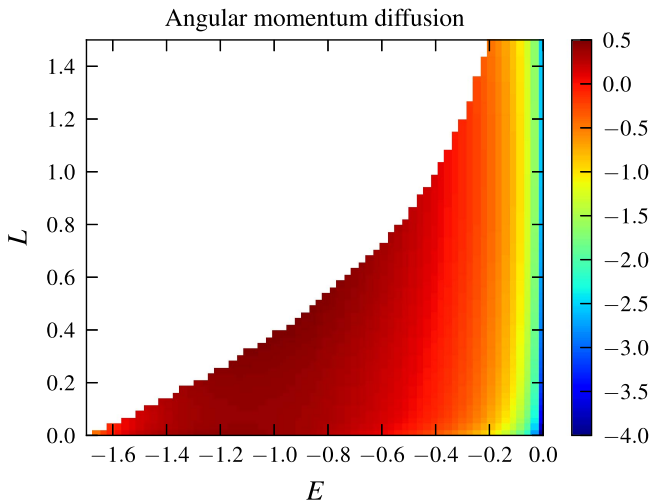
**Figure 2.** Same as Figure 1, but for the angular momentum diffusion coefficient.

We can get an analytical approximation of the relaxation time as a function of radius by substituting $\sigma(r)$ and $\rho(r)$ for a Plummer model in Equation (1). These are given by (e.g., Binney & Tremaine 2008)

$$\sigma^2(r) = \frac{GM}{6r_0}\left[1 + \left(\frac{r}{r_0}\right)^2\right]^{-1/2} \tag{8}$$

and

$$\rho(r) = \frac{3M}{4\pi r_0^3}\left[1 + \left(\frac{r}{r_0}\right)^2\right]^{-5/2}. \tag{9}$$

The analytical approximation appears as the dotted black line in Figure 3. This figure also shows the relaxation times calculated numerically according to Section 3.2 as a function of the semimajor axis for various eccentricity values. The energy and angular momentum relaxation times are represented by the
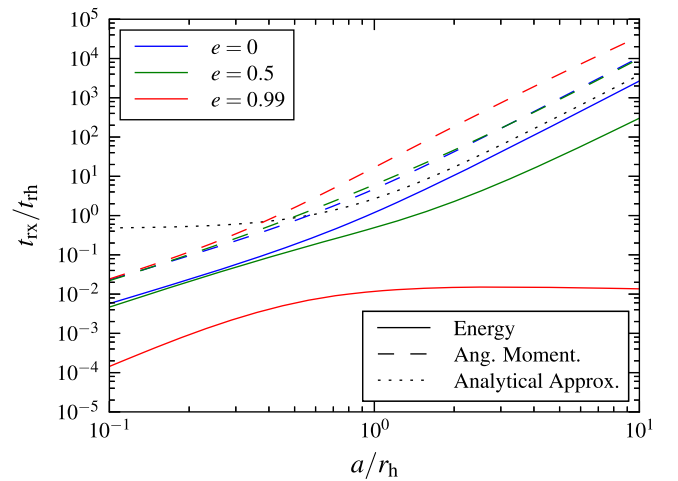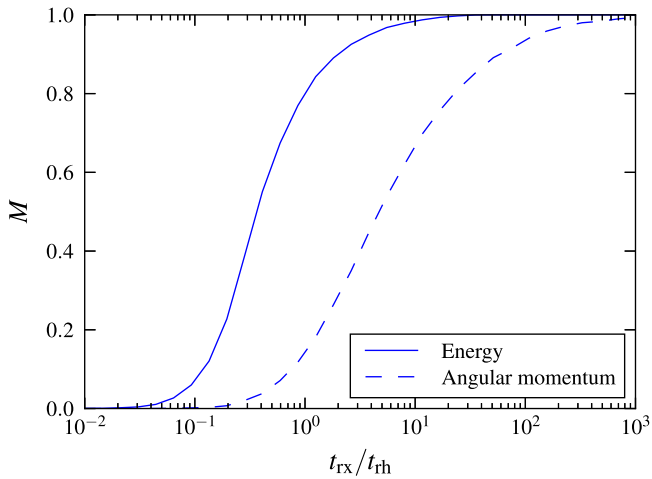

Figure 3. The energy (solid lines) and angular momentum (dashed lines) relaxation times as functions of the semimajor axis for different eccentricity cases in a Plummer model. The blue, green, and red lines represent zero, moderate ($e = 0.5$), and high ($e = 0.99$) eccentricities, respectively. The dotted black line represents the analytical approximation in Equation (1). The times are normalized by t_{rh} (given by Equation (2)) while the semimajor axis is normalized by the half-mass–radius.

solid and dashed lines, respectively, while the colors represent the eccentricity: circular orbits (blue), an intermediate eccentricity of $e = 0.5$ (green), and a high eccentricity of 0.99 (red). The relaxation time is normalized by t_{rh} given by Equation (2) so it is independent of N and $\ln\Lambda$ while the semimajor axis is normalized by the half-mass–radius r_h.

In particular, for a circular orbit with $a = r_h$ the energy relaxation time is $t_{rx,E} = 1.2t_{rh}$ while the angular momentum relaxation time is $t_{rx,L} = 4.8t_{rh}$ and the analytical approximation gives $2.7t_{rh}$. As noted above, the dependence on N and $\ln\Lambda$ is normalized out of these results by presenting them in units of t_{rh}, but the exact numbers do depend on the particular choice of γ and for the analytical approximation they depend on α as well. One may in fact go backward from those results and tune the dimensionless parameters of t_{rh} to get a better analytical estimate for a Plummer sphere. Generally, Figure 3 shows that the approximation using Equations (8)–(9) overestimates the relaxation time for a small a. For a large a, the dependence of both $t_{rx,L}$ and $t_{rx,E}$ on a is the same as the analytical approximation's dependence on the radius (with the exception of $t_{rx,E}$ in the $e = 0.99$ case)—namely $\sim r^{7/2}$. For the highest eccentricity case shown, the E-relaxation time seem to be independent of a for $a \gtrsim r_h$. For large values of a, the analytical approximation underestimates $t_{rx,L}$ but overestimates $t_{rx,E}$. This discrepancy is only $\sim30\%$ for circular orbits but is more than a factor of 10 in the moderately eccentric case.

Figure 4 shows the cumulative mass with energy (solid line) and angular momentum (dashed line) relaxation times that are shorter than t_{rx}. Half of the stars in the cluster have energy relaxation times shorter than $0.37t_{rh}$, while 90% have energy relaxation times shorter than $2.0t_{rh}$. The corresponding numbers for the angular momentum relaxation time are $4.7t_{rh}$ and $58t_{rh}$, respectively. This shows that the relaxation times are broadly distributed. The angular momentum magnitude diffusion is systematically slower by an order of magnitude than energy diffusion, which is mostly due to our particular definition of diffusion times, e.g., normalized to the average kinetic energy and circular angular momentum, respectively.

Figure 4. The cumulative mass normalized to the total cluster mass (or the cumulative number of stars normalized to the total number of stars) with energy (solid line) and angular momentum (dashed line) relaxation times shorter than t_{rx} in units of t_{rh}.

Table 1
Four Representative Initial Regions of Orbits in (E, L) Space Examined for Mixing in a Plummer Model in Section 4

Region	E	L	ΔE	ΔL	a	e	$t_{rx,E}$	$t_{rx,L}$
(I)	−1.21	0.16	0.01	0.01	0.2	0.5	0.1	0.8
(II)	−0.78	0.41	0.01	0.01	0.4	0.5	0.4	4.0
(III)	−0.30	0.30	0.02	0.02	1.0	0.9	0.3	83
(IV)	−0.30	1.20	0.02	0.02	1.0	$\lesssim 0.3$	14	64

Note. The corresponding semimajor axis a and eccentricity e for the (E, L) values are given to a one-digit accuracy, while in the the case of region (IV) the spread of eccentricities is larger than in the other regions. All parameters are in Hénon units for a Plummer model with a virial radius of one (see text for details). The relaxation times for the energy and angular momentum are calculated according to the procedure described in Section 3.2 and are given in units of t_{rh} (Equation (2)).

4. Mixing

4.1. Representative Regions

To investigate mixing in a star cluster, we choose four representative initial orbital families and follow their collisional evolution. These four families specified in Table 1 correspond to very small regions in (E, L)-space. In each region, the energy and angular momenta fall between $E \pm \Delta E$ and $L \pm \Delta L$, respectively, and they are centered at the semimajor axis and eccentricity given in Table 1:

(I) represents the inner region, low energy, and intermediate (relative) angular momentum particles. The energy range is selected so that approximately 90% of the particles have a higher energy than the middle of the range; the angular momentum range is selected so that the corresponding eccentricity is around 0.5.

(II) represents the intermediate region with the most typical particles in the system, in the sense that the middle of the range is selected close to the geometric median of all (E, L) values.

(III) represents traversing orbits between the outer and inner regions with high energy and low relative angular

momentum. The energy range is selected so that approximately 90% of the particles have a lower energy than the middle of the range; the angular momentum range is selected so that the corresponding eccentricity is around 0.9.

(IV) represents the outer region, high energy, and high relative angular momentum. It has the same energy range as region (III) but the angular momentum range is selected so that the corresponding eccentricity is lower than 0.3 (in all other regions the semimajor axis as well as the eccentricity have narrow distributions).

In the two following subsections we describe the Monte Carlo simulations we performed and how we used them to follow the widening of the E and L probability distributions of particles in these regions in time to investigate the long-term collisional behavior.

4.2. Methods

In order to follow statistically the collisional evolution of the four selected orbital families, we performed a series of simulations using the MOCCA code (MOnte Carlo Cluster simulAtor; Giersz et al. 2013). This code is based on the orbit-averaged Monte Carlo method of Hénon (1971) that was later substantially improved by Stodółkiewicz (1986). The basic idea is that changes in each star's energy and angular momentum from one state to the next (successive states of the system are separated by a time step, which is a fraction of t_{rh}) are computed by randomly selecting the position of the star on its orbit, randomly choosing another star, letting the two interact, and multiplying the effect by an appropriate factor. While MOCCA is a very sophisticated code, capable of realistically simulating globular clusters including physical effects such as stellar evolution and the accurate integration of few-body subsystems, we turned most of these features off and integrated very basic models using the code's dynamics capabilities only. Our models were 1024 separate particle realizations (with different random seeds) of a Plummer model with 16k particles each (k = 1024). The models were evolved for about $10t_{rh}$.

The relatively large number of models is required since as noted in the previous subsection we look at very small regions of (E, L)-space. The fraction of particles in these regions is as low as 5×10^{-4} (for region (III)). Thus for an $N = 16\,k$ model, only a handful of particles per model have the desired initial E and L values. We therefore superimpose the particle population at each region from 1024 such simulations. In addition, to increase the statistics of the background, we superimpose 64 of these simulations for a total of 1M particles (M = 2^{20}), at each snapshot. Despite the large number of models, this type of simulation is computationally inexpensive by modern standards, and the whole model set can be run on a desktop computer within less than a day.

4.3. Results

The top panel of Figure 5 shows the evolution of the energy probability distribution function[8] (PDF; left) and the angular momentum PDF (right) of the whole system due to collisional

---

[8] Note that the probability distribution function is different from the distribution function (DF). The latter is defined in such a way that $f(\boldsymbol{u})d^3\boldsymbol{x}d^3\boldsymbol{v}$, where $\boldsymbol{u}$ is a combination (or several combinations) of the phase-space coordinates, is the number of particles (or the mass) inside the six-dimensional cube of volume $d^3\boldsymbol{x}d^3\boldsymbol{v}$ around any point $(\boldsymbol{x}, \boldsymbol{v})$ in phase-space corresponding to $\boldsymbol{u}$. In contrast, $p(\boldsymbol{u})d\boldsymbol{u}$ (where $p(\boldsymbol{u})$ is the PDF of some of the quantity $\boldsymbol{u}$) is the fraction of particles in the range $d\boldsymbol{u}$ around $\boldsymbol{u}$ for any $\boldsymbol{x}$ and $\boldsymbol{v}$.
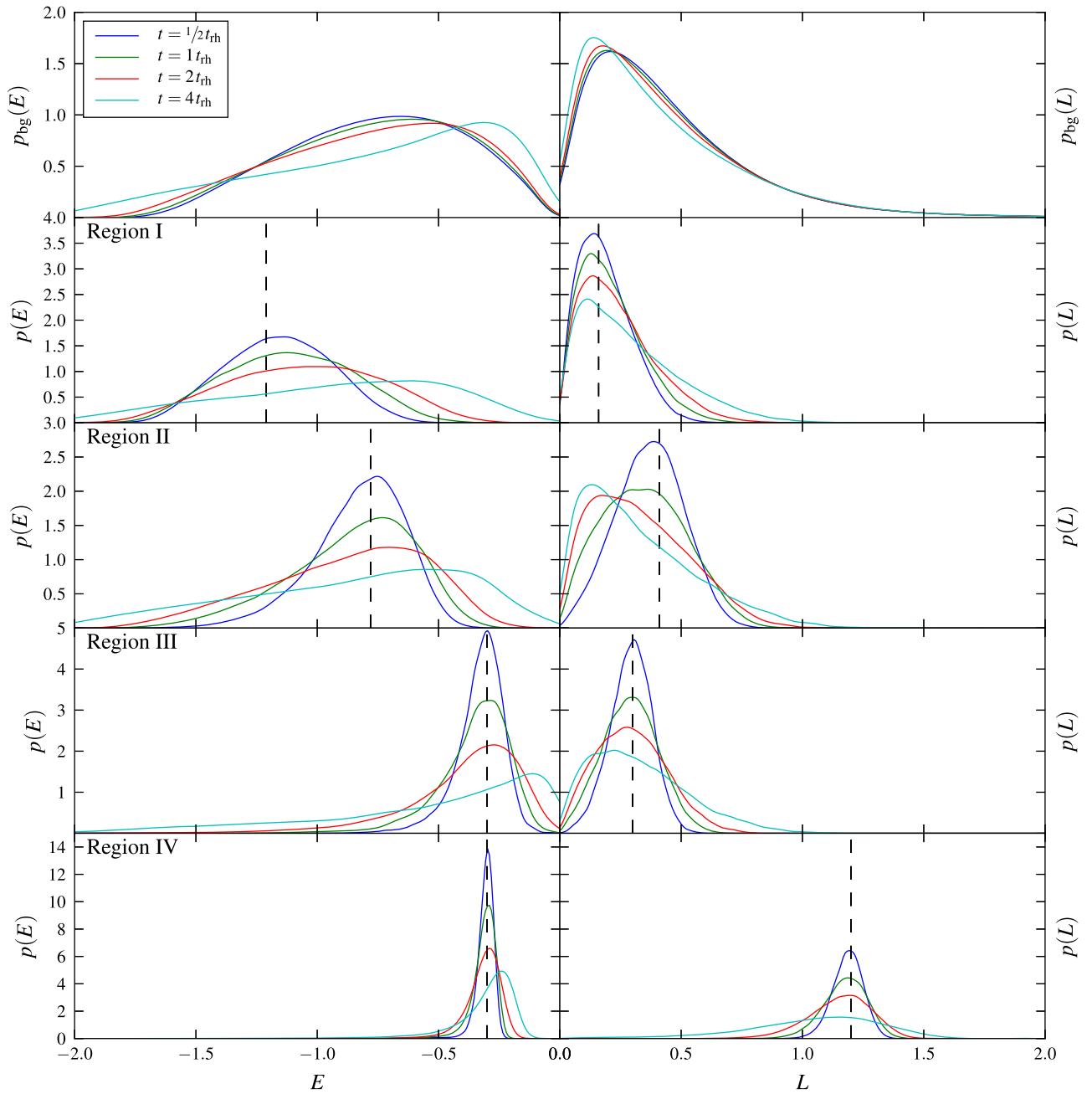
Figure 5. Top panel shows the evolution of the energy PDF (left) and the angular momentum PDF (right) of the whole system due to collisional relaxation and the gravitational response. The four rows of panels below show the evolution of the four representative regions (see Table 1). For clarity, only four time epochs are shown: the dashed black line represents the initial energy or angular momentum value (the distributions at $t = 0$ resemble Dirac delta-functions), while the blue, green, red, and cyan are respectively the distributions at 0.5, 1, 2, and 4 times the half-mass relaxation time.

relaxation and the gravitational response. The four rows of panels below show the evolution of the four representative regions. In the language of quantum mechanics, those graphs show different projections of the system's propagator, which is the probability amplitude for a particle to transition from one state to another in a given time. For clarity, only four time epochs are shown: the dashed black line represents the initial energy or angular momentum value (the distributions at $t = 0$ resemble Dirac delta functions), while the blue, green, red, and cyan are respectively the distributions at 0.5, 1, 2, and 4 times the half-mass relaxation time defined in Equation (2).

In all cases we find that the E- and the L-distribution of each of the four regions asymptotically approach the system's

distribution, which is itself slowly changing in time, and thus they mix toward a fully mixed state. To quantify the degree of mixing, we define the mixedness, denoted as c, of a subpopulation with respect to $\boldsymbol{u}$ through the correlation coefficient of its PDF $p(\boldsymbol{u})$ with that of the fully mixed configuration $p_{\text{bg}}(\boldsymbol{u})$ as

$$c = \frac{\langle p, p_{\text{bg}} \rangle}{\|p\| \|p_{\text{bg}}\|}. \tag{10}$$

Here $\langle A, B \rangle \equiv \int A(\boldsymbol{u}) B(\boldsymbol{u}) d\boldsymbol{u}$ is the scalar product on the space of PDFs and $\|A\| \equiv \sqrt{\langle A, A \rangle}$. In this paper we restrict our attention to mixing in one dimension only. The PDFs are one

dimensional and are denoted by $p(E)$ and $p(L)$, for energy and angular momentum, respectively. The correlation coefficient between distributions varies between zero and unity. $c = 0$ represents the completely uncorrelated case, where $p(\boldsymbol{u})$ and $p_{\rm bg}(\boldsymbol{u})$ have disjoint support sets (in practice this is the case only when $p(\boldsymbol{u})$ is a Dirac delta function). $c = 1$ represents the fully correlated case where $p(\boldsymbol{u})$ and $p_{\rm bg}(\boldsymbol{u})$ are proportional.[9]

The motivation for this definition is based on a stochastic random-walk model of relaxation introduced by Kocsis & Tremaine ([2015]). In that model, each star's actions change randomly in each time step according to a given transition probability function. In that case, it can be shown that the evolution is governed by a linear operator. Decomposing the PDFs of the actions in the orthonormal eigenfunctions of this linear operator shows that each such mode decays independently exponentially in time with distinct decay constants. There is one mode whose decay constant is zero, which represents the fully mixed steady-state distribution. The projection of the PDF on the steady-state distribution given by Equation ([10]) is the natural way to define mixedness in such models.

Estimating c from discrete data can be difficult. One has to first estimate the continuous functions p and $p_{\rm bg}$ from two discrete sets of values. This could be done by a variety of methods such as kernel density estimation and clustering analysis; most of these have one free parameter or more. Here again we turn to the simplest approach, in this case data binning (histogram). The freedom in this method is to choose the size and position of all the bins. When the number of data points in both sets is extremely large, it is expected that one can produce smooth and fiducial PDFs (with any reasonable density estimation method), but from numerical experiments we found that the relevant data set sizes are not large enough. A second problem is that even a small bias in the estimation of c may lead to a big systematic error in the derivation of a timescale when analyzing the dependence of c on time due to the asymptotic approach to unity as the stellar system is evolving toward a fully mixed state. In Appendix C, we describe the numerical procedure to estimate c.

Figure [6] shows $1-c(E)$ (left panel) and $1-c(L)$ (right panel) for the four regions. It is evident that the level of mixedness approaches unity asymptotically exponentially in time. Because in any real star cluster the number of particles is finite and the number of particles in any small region of (E, L)-space is likewise small, it is expected that the target distribution becomes statistically indistinguishable from the background in a finite amount of time. While the relaxation time is shorter for region (I), it is evident that region (II) mixes earlier than the others both in energy and in angular momentum. Another conclusion from this figure is that the value of $c(L)$ initially approaches unity faster than $c(E)$ for regions (I), (II), and (III). The opposite happens for region (IV). This may also be in part related to the proximity of the initial L values to the system's median. Note that the angular momentum mixedness curve for region (II) may be saturating due to the numerical problem with the estimator described above; the relative error in $1-c$ may be very large when c approaches unity.
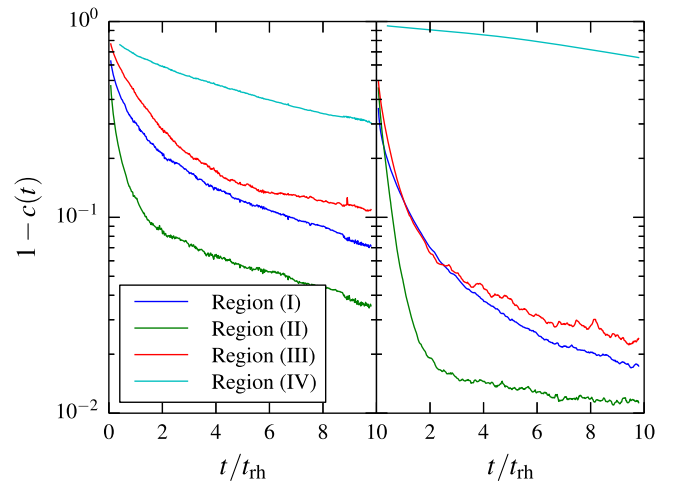
Figure 6. One minus the energy (left panel) and angular momentum mixedness (right panel) as a function of time of the four representative orbital families in (E, L)-space described in Section [4.1].

It is not easy to measure a timescale from the noisy mixedness curves of Figure [6]. One possibility is to choose a threshold (e.g., 90% or 99%) and define the mixing time as the time at which the curve (or an extrapolation thereof) crosses that threshold. Another way is to assume that the mixedness curves approach unity exponentially and to define the mixing time as the decay time of this exponential function. Indeed, from the left panel we see that for all curves, $\log_{10}[1 - c(t)]$ is roughly linear after about $t = 4t_{\rm rh}$; moreover, all four curves are roughly parallel, implying a shared underlying mixing timescale. By fitting C and $t_{\rm mix}$ assuming $c = 1 - C\exp(-t/t_{\rm mix})$ we find that the energy mixing time is between $t_{\rm mix,E} \sim 9t_{\rm rh}$ (for region (I)) and $15t_{\rm rh}$ (for region (III)). The exact values depend on the time interval where the fit is made, and they are constrained to within $\sim 20\%$. A similar picture is seen in the right panel regarding the angular momentum. In this case the green curve seems to almost stall (implying a very long e-folding time). However, this is possibly attributed to the numerical problem mentioned above. The timescale derived for regions (I) and (III) is $t_{\rm mix,L} \sim 9t_{\rm rh}$ and $\sim 20t_{\rm rh}$ for region (IV).

5. Discussion

5.1. Relaxation

In Section [3.3] we used scattering theory to calculate the energy E and angular momentum L relaxation times for different orbital families in a Plummer sphere. We showed that the half-mass relaxation time $t_{\rm rh}$ (Equation ([2])) provides a decent estimate for the order of magnitude of the relaxation times for orbits with a semimajor axis that equals the half mass–radius and is consistent with the more rigorous calculations (for both E and L) to within a factor of ~ 5 for mildly eccentric orbits. This number is similarly consistent with the E- and L-relaxation times of half of the cluster mass within the same factor, as shown in Figure [4]. Our more rigorous calculations are also based on some assumptions (e.g., an isotropic velocity distribution) but importantly attempt taking into account the non-Maxwellian nature of the velocity distribution and the nonuniform spatial density. The distribution function of a Plummer sphere is proportional to $(-E)^{7/2}$ (the full expression is given in Equation ([27]) in the Appendix). By writing the energy at a fixed radius r_0, we can see that the

distribution of velocity magnitudes at any given position is proportional to $v^2[-v^2-2\Phi(r_0)]^{7/2}$, where $\Phi(r_0)$ is a negative constant (the potential). This distribution drops to zero at the escape velocity, while the Maxwell–Boltzmann distribution has an infinite tail of high velocities. For circular orbits with $a > r_h$, most of the conditions for the standard approximation are met and Equation (1) gives a very good approximation for the energy relaxation despite the somewhat different functional form of the velocity distribution. The discrepancy is most evident for eccentric orbits and for circular orbits as well when $a < r_h$.

Measuring the diffusion coefficients directly from N-body simulations is a better way to find the relaxation time as a function of the orbital elements that does not depend on any assumptions. This could in principle be done by measuring the rms change of E (or L) denoted $\langle(\Delta E)^2\rangle$ of particles in a small bin in (E, L)-space over a short period of time Δt; however, the problem is determining this Δt. The forces acting on a particle are correlated on very short timescales, and it is only on longer timescales that the random walk-like behavior is revealed. However, on yet longer timescales, particles starting from a small bin in (E, L)-space may be scattered throughout this space, and the instantaneous rate of their energy diffusion would be affected by their new (E, L) values rather than the initial ones, where we are interested in measuring the diffusion rate. This means that $\langle(\Delta E)^2\rangle$ as a function of Δt is expected to be quadratic at short time intervals, transition to linear at longer intervals, and saturate to a constant value when the distribution becomes fully mixed. The problem measuring the diffusion coefficients on large values of Δt becomes less severe for a large N because the longer local relaxation time everywhere means that particles deviate more slowly from their initial position in (E, L)-space. Furthermore, it is not guaranteed in general that there exists such an intermediate timescale that is long enough for the correlated behavior to disappear but not too long that the energies scatter too far from the original value. This could be the case for low-N systems such as an open cluster, where the gravitational field is dominated by finite-number effects and the mixing time is shorter than the orbital period. In globular clusters, however, this is unlikely to be the case as we have shown.

Theuns (1996) directly measured the energy diffusion coefficients, as a function of energy only, in King models from direct-summation N-body simulations of up to 32k particles. He defined individual particles as being in different "states" between two local maxima of the $E(t)$ curve of each particle, where ΔE is defined by the difference between adjacent maxima and Δt is the time interval between them. Doing so for both angular momentum and energy in tandem requires a larger number of particles but is easily achievable with modern computers. Diemand et al. (2004) similarly measured the mean energy relaxation times in the context of cosmological simulations, but instead of as a function of E they considered different radial bins of a Hernquist model. They chose Δt from different considerations, requiring that most particles spend most of the time interval in the same radial bin. Measuring the diffusion time from astronomical observation is a much bigger challenge. By measuring the positions of young (bright) white dwarfs in the globular cluster 47 Tucanae, Heyl et al. (2015) were able to calculate a diffusion rate consistent with a core relaxation time of about ∼70 Myr.

The results presented in Section 3.3 can also be used to roughly estimate the diffusion coefficients (or relaxation times) for the selected orbital families, but this is only meaningful as a sanity check or as a validation of the MOCCA code. This is because unlike an N-body code that needs only assume Newtonian physics and gravity, the MOCCA code is essentially already programmed with scattering theory. More specifically, it is a statistical way of solving the Fokker–Planck equation, under the additional assumption of spherical spatial symmetry (velocity anisotropy, however, can be accommodated). Therefore, measuring the diffusion from these results would be circular. However, measuring the rate of mixing from these results is meaningful in the sense that it is a result of the long-term stochastic behavior that is reasonably well described by the Fokker–Planck equation. Measuring mixing from an N-body simulation may show additional effects not described by the Fokker–Planck equation or the approximate solution provided by MOCCA. For example, scalar resonant relaxation, if present, could be captured in the mixedness curve of L, but this could not be revealed in the present study, which assumes spherical symmetry.

5.2. Energy versus Angular Momentum

For processes such as loss cone refilling it is more appropriate to consider $t_{rx,L}$, which can differ from $t_{rx,E}$ and t_{rh} considerably under most circumstances. We found[10] that the difference at the half-mass–radius is only a factor of a few for circular orbits but is more than an order of magnitude for mildly eccentric orbits. However, comparing the diffusion rates of E and L is more difficult than comparing $t_{rx,E}$ and $t_{rx,L}$. The diffusion coefficients cannot be compared directly simply because they have different dimensionality. The diffusion coefficients are converted to relaxation times by choosing a reference E and L, but as already discussed in Section 3.2, there is some unavoidable arbitrariness to that choice. The statement we can make from the figures in Section 3.3 is that the kinetic energy diffusion rate is faster than the angular momentum diffusion rate relative to the circular angular momentum. The underlying reason can be easily understood by considering an orbit with high eccentricity (e.g., the red curves in Figure 3) and a large semimajor axis. This orbit, being only weakly bound, has a small kinetic and total energy (with respect to the central potential); being almost radial, it also has a small angular momentum (with respect to the circular value for that energy). Changing this orbit's average kinetic energy by its own amount (which should take approximately $t_{rx,E}$) would transform it generally to another weakly bound orbit, with a slightly different semimajor axis. Changing this orbit's angular momentum by the reference angular momentum, the circular angular momentum (which should take approximately $t_{rx,L}$) would transform it from a radial to a circular orbit. This would require many more scatterings, and thus $t_{rx,L} > t_{rx,E}$ for this kind of orbit and in general.

5.3. Mixing

Phase mixing occurs on the dynamical timescale for the angles, but mixing of the integrals of motion takes much longer. It has been shown (Goodman et al. 1993; Hemsendorf & Merritt 2002) that divergence in this space is exponential on a

---

[10] For single-mass clusters without a central black hole.

timescale (i.e., the inverse of the Liapunov exponent) proportional to the crossing time, with a factor of $(\ln \ln N)^{-1}$ leading to an extremely weak dependence on N. This derivation makes assumptions similar to that of scattering theory: the interactions between stars are relatively discrete, separate encounters—in other words, incoherent and uncorrelated—and the spatial distribution is uniform. Mixing in (E, L)-space is a different aspect of the chaotic nature of the N-body problem that describes the statistical spreading of constants of motion due to collisions. We quantified this process by defining mixedness, which is a relative measure of the width of the E or L distribution of a subpopulation with respect to the global population. Quantitatively, it is a measure of the mean (averaging is assumed over different realizations of the initial conditions) correlation coefficient between the PDF starting from a small confined region in (E, L)-space and the PDF of the whole cluster. This is defined on all timescales, and its rate varies in time from very fast initially to a constant. Asymptotically at later times, mixing converges exponentially in time with a characteristic decay timescale. The mixing time scales like the relaxation time(s) with N because it is driven by the same physical process of collisional diffusion.

We measured mixedness by conducting Monte Carlo simulations, necessitating the use of a statistical estimator on a discrete data set, which has a bias that we attempted to correct and that worked to a certain degree (i.e., on long timescales c saturates at a value close to unity). Solving the Fokker–Planck equation directly, in both E and L with two interacting components (background and subpopulation in a certain region), may have possibly resulted in more accurate estimates of the mixedness in some aspects.

Curiously, the determination of the mixing timescale as $\sim 10 t_{\rm rh}$ is reminiscent of a result obtained by Bar-Or et al. (2013). They derived a timescale for a small initial perturbation superimposed at a specific energy on a system at a steady state, to reach that steady state (mixing, in our terminology, although they referred to this as relaxation). They showed that this timescale was equal to roughly 10 times the energy diffusion time. Despite the fact that they considered a power-law cusp in the galactic center context rather than a Plummer model (which has a flat center) and used very simple Fokker–Planck analysis to derive this, it appears to be consistent with our result (cf. Appendix B of Madigan et al. 2011).

We demonstrated the most basic manifestation of mixing using a single-mass population in a self-gravitating Plummer sphere. The adopted definition is applicable to more general systems. In the context of spherical (i.e., globular) star clusters, multiple stellar populations are often observed. This is revealed in both spectroscopic studies, which show stellar populations characterized by different chemical abundances (Gratton et al. 2001; Marino et al. 2008; Carretta et al. 2009), and photometric studies in which different populations form distinguishable sequences in color–magnitude diagrams (Lee et al. 1999; Bedin et al. 2004; Piotto et al. 2007). There is no consensus regarding the formation of such a secondary population. The two leading models are ad hoc formation of the second-generation stars from the gas accumulated from the external interstellar medium and a minor merger of clusters with an age difference of a few hundred million years (which could be quite rare; see Lee 2015). Hong et al. (2017) carried out numerical simulations based on these two formation scenarios and found that both of them reproduce the observed

radial trend of the ratio between the stellar populations. While the spatial mixing of different populations has been studied (Decressin et al. 2008; Vesperini et al. 2013; Hénault-Brunet et al. 2015; Miholics et al. 2015), the details of the isolating integrals mixing process may help distinguish among formation models.

Finally, an important utility for the mixing time is within hybrid collisional-collisionless N-body codes. In this kind of scheme, the evolution of a stellar system is computed in such a way that only a fraction of the stars experiences two-body encounters. Fiducially simulating such a system requires the advance knowledge of how to divide (e.g., in (E, L)-space) the system into collisional and collisionless components and of how long it is possible to simulate before reassigning particles into the two groups (cf. Y. Meiron et al. 2018, in preparation).

Appendix A
Local Diffusion of Angular Momentum Magnitude

In this section we write the mean square change in the angular momentum vector's magnitude during a short encounter as a function of the mean square velocity changes parallel and perpendicular to the original velocity direction. In other words, we express $\langle (\Delta L)^2 \rangle$ as a function of $\langle (\Delta v_\parallel)^2 \rangle$ and $\langle (\Delta v_\perp)^2 \rangle$. Since we are only computing the average change during a single short encounter ("local diffusion") our expressions will depend on phase-space coordinates (r, v, and v_r). In the next step we integrate over them to get the orbital averaged coefficients. Also note that we are interested in the square change in the vector's magnitude, *not* the square magnitude of the difference vector, thus $(\Delta L)^2 \equiv (|L_2| - |L_1|)^2$.

We start by writing the angular momentum vector before the encounter $L_1 = r \times v_1$. Since the encounter occurs over a very short period, r does not change and therefore does not need to be subscripted. After the encounter, the angular momentum vector is $L_2 = r \times v_2$ with

$$v_2 = v_1 + (\Delta v_\parallel)\hat{v}_1 + (\Delta v_\perp)\hat{u}_1, \tag{11}$$

where $\hat{v}_1 = v_1/v_1$ a unit vector in the direction of $\hat{v}_1$ and $\hat{u}_1$ is an unknown unit vector perpendicular to it. The new angular momentum following some simple algebra is

$$L_2 = \left[1 + \frac{(\Delta v_\parallel)}{v_1}\right] L_1 + (\Delta v_\perp)(r \times \hat{u}_1) \tag{12}$$

and its square magnitude is

$$L_2^2 = \left[1 + \frac{(\Delta v_\parallel)}{v_1}\right]^2 L_1^2 + (\Delta v_\perp)^2 |r \times \hat{u}_1|^2$$
$$+ 2\left[1 + \frac{(\Delta v_\parallel)}{v_1}\right](\Delta v_\perp) L_1 \cdot (r \times \hat{u}_1). \tag{13}$$

The last term could be simplified as follows:

$$L_1 \cdot (r \times \hat{u}_1) = (r \times v_1) \cdot (r \times \hat{u}_1)$$
$$= (r \cdot r)(v_1 \cdot \hat{u}_1) - (r \cdot \hat{u}_1)(v_1 \cdot r) = -r^2 u_{1r} v_{1r}, \quad (14)$$

where we have used that $v_1 \cdot \hat{u}_1 = 0$ by definition and defined v_{1r} and u_{1r} as the radial components of v_1 and $\hat{u}_1$, respectively.

We are interested in the quantity

$$(\Delta L)^2 = (L_2 - L_1)^2 = L_2^2 - 2L_2 L_1 + L_1^2, \quad (15)$$

which contains the parallel and perpendicular velocity changes under a square root in the middle term. Since we only consider those changes up to the second order, we can write the L_2 as a Taylor series. The result is

$$L_2 = \sqrt{\cdots} = L_1 + \frac{L_1}{v_1}(\Delta v_\parallel) - \frac{r^2 v_{1r} u_{1r}}{L_1}(\Delta v_\perp)$$
$$- \frac{1}{2}\left[\frac{r^4 v_{1r}^2 u_{1r}^2}{L_1^3} - \frac{|r \times \hat{u}_1|^2}{L_1}\right](\Delta v_\perp)^2$$
$$+ \text{higer order terms} \quad (16)$$

and therefore

$$(\Delta L)^2 = \frac{L_1^2}{v_1^2}(\Delta v_\parallel)^2 + \frac{r^4 v_{1r}^2 u_{1r}^2}{L_1^2}(\Delta v_\perp)^2$$
$$- \frac{2r^2 u_{1r} v_{1r}}{v_1}(\Delta v_\parallel)(\Delta v_\perp). \quad (17)$$

We immediately see that the last term does not contribute to the average because $\langle (\Delta v_\perp) \rangle = 0$. In addition, since the vector $\hat{u}_1$ is independent of the change in velocity, the average is (now dropping the subscript 1)

$$\langle (\Delta L)^2 \rangle = \frac{L^2}{v^2}\langle (\Delta v_\parallel)^2 \rangle + \frac{r^4 v_r^2}{L^2}\langle u_r^2 \rangle \langle (\Delta v_\perp)^2 \rangle. \quad (18)$$

It is relatively easy to geometrically show that

$$u_r = \frac{v_t}{v}\cos\beta, \quad (19)$$

where $v_t^2 = v^2 - v_r^2$ and β is a random angle. Since $\langle \cos^2 \beta \rangle = 1/2$ and $L = rv_t$, we finally get

$$\langle (\Delta L)^2 \rangle = \frac{r^2}{v^2}\left[v_t^2 \langle (\Delta v_\parallel)^2 \rangle + \frac{1}{2}v_r^2 \langle (\Delta v_\perp)^2 \rangle\right]. \quad (20)$$

This result is in agreement with Equation (88) of Bar-Or & Alexander (2016) but not in agreement with Equation (21) of Spitzer & Shapiro (1972). The $1/2$ factor in the right term, which comes from the square cosine averaging, is not present there.

Appendix B
Orbital Averaging

We finalize the calculation of the diffusion coefficients writing $\langle (\Delta v_\parallel)^2 \rangle$ and $\langle (\Delta v_\perp)^2 \rangle$ as functions of velocity with the Rosenbluth potentials as substituting into Equations (3) and (4):

$$\frac{\langle (\Delta E)^2 \rangle}{\Delta t} = \frac{8\pi\Gamma v^3}{3}[F_4(v) + E_1(v)] \quad (21)$$

and

$$\frac{\langle (\Delta L)^2 \rangle}{\Delta t} = \frac{8\pi\Gamma r^2}{3v}$$
$$\times \left[\left(v^2 - \frac{3}{2}v_r^2\right)F_4(v) + \frac{3}{2}v_r^2 F_2(v) + v^2 E_1(v)\right], \quad (22)$$

where $\Gamma = 4\pi G^2 m^2 \ln\Lambda$ and

$$F_n(v) \equiv \int_0^v \left(\frac{v'}{v}\right)^n f(v')dv' \quad (23)$$

and

$$E_n(v) \equiv \int_v^\infty \left(\frac{v'}{v}\right)^n f(v')dv' \quad (24)$$

are the Rosenbluth potentials. We use the fact that $v'd\,v' = dE'$ and change the variable of integration to the energy, since f for the Plummer model is given as a function of only energy:

$$F_n(E, r) \equiv v^{-n}(E) \int_{\Phi(r)}^E v'^{n-1} f(E')dE' \quad (25)$$

and

$$E_n(E, r) \equiv v^{-n} \int_E^\infty v'^{n-1} f(E')dE' \quad (26)$$

given a spherically symmetric potential $\Phi(r)$ derived from $f(E)$. For a Plummer model (Aarseth et al. 1974),

$$f(E) = \frac{24\sqrt{2}}{7\pi^3}\frac{r_0^2 N}{G^5 M^5}(-E)^{7/2} \quad (27)$$

and

$$\Phi(r) = -\frac{GM}{\sqrt{r^2 + r_0^2}}, \quad (28)$$

where r_0 is the Plummer radius, M is the total mass, and G is the gravitational constant. $f(E) = 0$ outside the range $\Phi(0) < E < 0$. Note that we write the distribution function such that $\int f d^3x\,d^3v = N$; often it is normalized to the total mass rather than the number of particles.

We can forget about the velocity dependence because both the total velocity and its radial component can be written as functions of E, L, and r. The relations are

$$v = \sqrt{2[E - \Phi(r)]} \quad (29)$$

and

$$v_r = \sqrt{2[E - \Phi(r)] - L^2/r^2}, \quad (30)$$

and similarly v' relates to E' in Equations (25) and (26).

Now that the local diffusion coefficients are in a form where the only phase-space coordinate they depend on is r, we can proceed to the orbital averaging for a fixed E and L. The Rosenbluth potentials (only E_1, F_2, and F_4 are needed) are tabulated for the given E and radii between the pericenter r_p and the apocenter r_a (which are also functions of E and L). The orbit-averaging integrals are (Spitzer 1987, chapter 2b)

$$D_{E^2}(E, L) = \frac{2}{P_r}\int_{r_p}^{r_a} \frac{\langle (\Delta E)^2 \rangle}{\Delta t}\frac{dr}{v_r} \quad (31)$$

and

$$D_{L^2}(E, L) = \frac{2}{P_r} \int_{r_\mathrm{p}}^{r_\mathrm{a}} \frac{\langle (\Delta L)^2 \rangle}{\Delta t} \frac{dr}{v_r}, \tag{32}$$

where P_r is the radial orbital period (in rosette-type orbits, this is not the same as the angular period). To numerically perform the integrals, v_r from Equation (30) is substituted, and we solve the potential to find r_p, r_a, and P_r for these values of E and L.

For circular orbits, $v_r = 0$ and there is no need to perform orbital averaging as r and v are constant along the orbit. The result is

$$D_{E^2} = \frac{8\pi \Gamma v^3}{3} [F_4(E, r) + E_1(E, r)] \tag{33}$$

and

$$D_{L^2} = \frac{r^2}{v^2} D_{E^2}. \tag{34}$$

The energy and circular velocity are related as follows to the radius

$$E = \tfrac{1}{2} r \Phi'(r) + \Phi(r) \tag{35}$$

and

$$v = \sqrt{r \Phi'(r)}, \tag{36}$$

where $\Phi'(r)$ is the gradient of the potential in the radial direction. The functions E_1 and F_4 still need to be evaluated as before (albeit at a single point) because they represent scattering contribution from the field particles, not orbital averaging.

B.1. Numerical Integration

The radial velocity v_r in the denominator in the orbital average integrals approaches zero at the apsides, causing the integrands to diverge at the integration limits. Despite the finiteness of the integrals, this poses a numerical problem that is mitigated as follows. A reasonable approximation could be made by writing the reciprocal of the problematic term as a Taylor series around each apsis and analytically calculating its integral up to a small distance ϵ, where numerical integration is easier.

$$\int_{r_\mathrm{p}}^{r_\mathrm{p}+\epsilon} v_r^{-1} dr = \left[\frac{2\epsilon}{-\Phi'(r_\mathrm{p}) + L^2/r_\mathrm{p}^3} \right]^{1/2} \tag{37}$$

and

$$\int_{r_\mathrm{a}-\epsilon}^{r_\mathrm{a}} v_r^{-1} dr = \left[\frac{2\epsilon}{\Phi'(r_\mathrm{a}) - L^2/r_\mathrm{a}^3} \right]^{1/2}. \tag{38}$$

We choose $\epsilon = 10^{-2}(r_\mathrm{a} - r_\mathrm{p})$.

Appendix C
Estimation of the Mixedness

Here we describe the numerical procedure to estimate

$$c = \frac{\int p(x) p_\mathrm{bg}(x) dx}{\sqrt{\int p^2(x) dx} \sqrt{\int p_\mathrm{bg}^2(x) dx}} \tag{39}$$

from discrete data. We assume that $p_\mathrm{bg}(x)$ is a smooth function[11] and that the data $G = \{x_1, \ldots, x_N\}$ are a random realization of the PDF $p(x)$. The estimator $\hat{c}$ is calculated by the following numerical procedure. The data set G is divided into n bins by first throwing out the innermost and outermost ϵN data points (this gets rid of outliers). The smallest and largest x-values of the remaining data points are the limits of the histogram. Then the integral in the numerator is calculated through the trapezoidal rule where instead of $p(x)$ we use the number of data points in each bin, $p_\mathrm{bg}(x)$, which is evaluated at the center of the bin. The left integral in the denominator is evaluated in the same way, and the right integral is calculated analytically or integrated numerically in some other way. This gives us a *biased* estimator $\hat{c}$. Below we describe the numerical experiments we performed to attempt to correct the bias and evaluate the statistical error.

The correction factor $\zeta = c/\hat{c}$ is a function of the number of particles and the uncorrected estimator $\hat{c}$ of the projection. The number of bins used in the procedure may also play a role, but it is marginalized by choosing an optimal number of bins n as a function of the number of data points N. The correction factor (as well as the optimal bin number) may very well depend on the exact functional form of both p and p_bg, but we assume for the sake of the numerical experiments that both are normal distributions centered at the origin, with widths of σ and 1, respectively. The true value of the projection is given analytically in this case by

$$c = \sqrt{\frac{2\sigma}{1 + \sigma^2}}, \tag{40}$$

which can be inverted to find $\sigma(c)$. The idea behind the numerical experiments is that for different values of N (which determines the number of bins n) and c (which determines σ), we make many (2^{18}) realizations and calculate the biased estimator $\hat{c}$ for each one. This gives us a distribution of values; the correction factor ζ is the ratio $c/\langle \hat{c} \rangle$ between the real projection and the average of the biased estimators. The width of the distribution helps determine the error.

The optimization of the number of bins to be used in the procedure is done first. It seems that the best values of $\hat{c}$ (i.e., the closest to c) are obtained roughly when

$$n = \mathrm{round}(1.584 \, N^{0.38}). \tag{41}$$

The best number of bins depends on c somewhat as well, but Equation (41) marginalizes over that with some bias toward lower values of c. The lowest number of data points we consider is $N = 32$, which gives $n = 6$ bins. With $n(N)$ fixed for the procedure and the choice $\epsilon = 1/64$, we proceed by generating realizations G for (c, N) pairs and calculating the $\hat{c}$ distribution. For each value of N we choose 39 values of c equally distributed between 0.05 and unity. We find that the width (representing the error) strongly depends on N but weakly on c. At a fixed N, the correction factor itself as a function of $\hat{c}$ can be somewhat approximated by a 3-parameter function with the same functional form as a Sérsic profile. We only use this fitting to evaluate the sensitivity of ζ to N by

---

[11] This assumption means that the total number of particles in the system is large enough for the background distribution to be considered smooth.

looking at how the fit parameters change with N. We find that they vary strongly with N only when the threshold to throw away another data point at the tails of the distribution is passed (e.g., between $N = 191$ and 192); therefore we make sure that those transitions are included in the grid.

The correction is generally small. The deviation of $\hat{c}$ from c is the largest for the smallest number of particles ($N = 32$ in our experiments) and the smallest projection value ($c = 0.05$ in our experiments). In this case it is still less than 15%. It is, however, critical to take this projection into account for the purpose of this study, because a small bias in the mixedness value may cause a very large deviation in the mixing timescale. This ad hoc correction is computed for the simple case where both the background distribution p_{bg} and the target distribution p are normal, and moreover their centroids are the same. The results may very well depend on the functional form, but a more general or elegant derivation of the bias correction is outside the scope of this work.

ORCID iDs

Yohai Meiron ⓘ https://orcid.org/0000-0003-3518-5183
Bence Kocsis ⓘ https://orcid.org/0000-0002-4865-7517

References

Aarseth, S. J., Henon, M., & Wielen, R. 1974, A&A, 37, 183
Bar-Or, B., & Alexander, T. 2016, ApJ, 820, 129
Bar-Or, B., Kupi, G., & Alexander, T. 2013, ApJ, 764, 52
Bedin, L. R., Piotto, G., Anderson, J., et al. 2004, ApJL, 605, L125
Binney, J., & Tremaine, S. 2008, Galactic Dynamics (2nd ed.; Princeton, NJ: Princeton Univ. Press)
Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2009, A&A, 505, 117
Chandrasekhar, S. 1942, Principles of Stellar Dynamics (Chicago, IL: Univ. Chicago Press)
Chavanis, P.-H. 2012, PhyA, 391, 3680
Chavanis, P.-H. 2013, A&A, 556, A93
Cohen, R. S., Spitzer, L., & Routly, P. M. 1950, PhRv, 80, 230
Cohn, H. 1979, ApJ, 234, 1036
Cohn, H., & Kulsrud, R. M. 1978, ApJ, 226, 1087
Decressin, T., Baumgardt, H., & Kroupa, P. 2008, A&A, 492, 101
Diemand, J., Moore, B., Stadel, J., & Kazantzidis, S. 2004, MNRAS, 348, 977
Fouvry, J.-B., & Bar-Or, B. 2017, MNRAS, submitted (arXiv:1712.05862)
Giersz, M., & Heggie, D. C. 1994, MNRAS, 268, 257
Giersz, M., Heggie, D. C., Hurley, J. R., & Hypki, A. 2013, MNRAS, 431, 2184
Goodman, J., Heggie, D. C., & Hut, P. 1993, ApJ, 415, 715
Gratton, R. G., Bonifacio, P., Bragaglia, A., et al. 2001, A&A, 369, 87
Hachisu, I., Nakada, Y., Nomoto, K., & Sugimoto, D. 1978, PThPh, 60, 393
Heggie, D. C. 2014, MNRAS, 445, 3435
Hemsendorf, M., & Merritt, D. 2002, ApJ, 580, 606
Hénault-Brunet, V., Gieles, M., Agertz, O., & Read, J. I. 2015, MNRAS, 450, 1164
Hénon, M. 1961, AnAp, 24, 369
Hénon, M. H. 1971, Ap&SS, 14, 151
Heyl, J., Richer, H. B., Antolini, E., et al. 2015, ApJ, 804, 53
Heyvaerts, J. 2010, MNRAS, 407, 355
Hong, J., de Grijs, R., Askar, A., et al. 2017, MNRAS, 472, 67
Jeans, J. H. 1913, MNRAS, 74, 109
Kocsis, B., & Tremaine, S. 2015, MNRAS, 448, 3265
Kuzmin, G. G. 1957, PTarO, 33, 75
Larson, R. B. 1970, MNRAS, 147, 323
Lee, J.-W. 2015, ApJS, 219, 7
Lee, Y.-W., Joo, J.-M., Sohn, Y.-J., et al. 1999, Natur, 402, 55
Lynden-Bell, D., & Eggleton, P. P. 1980, MNRAS, 191, 483
Lynden-Bell, D., & Wood, R. 1968, MNRAS, 138, 495
Madigan, A.-M., Hopman, C., & Levin, Y. 2011, ApJ, 738, 99
Marino, A. F., Villanova, S., Piotto, G., et al. 2008, A&A, 490, 625
Maxwell, J. C. 1866, RSPS, 15, 167
Merritt, D. 2013, Dynamics and Evolution of Galactic Nuclei (Princeton, NJ: Princeton Univ. Press)
Merritt, D. 2015, ApJ, 804, 52
Miholics, M., Webb, J. J., & Sills, A. 2015, MNRAS, 454, 2166
Piotto, G., Bedin, L. R., Anderson, J., et al. 2007, ApJL, 661, L53
Rauch, K. P., & Tremaine, S. 1996, NewA, 1, 149
Rosenbluth, M. N., MacDonald, W. M., & Judd, D. L. 1957, PhRv, 107, 1
Schwarzschild, K. 1924, in Festschrift für Hugo von Seeliger (zum 75. Geburtstag), ed. H. Kienle (Berlin: Springer)
Sippel, A. C., & Hurley, J. R. 2013, MNRAS, 430, L30
Spitzer, L. 1987, Dynamical Evolution of Globular Clusters (Princeton, NJ: Princeton Univ. Press)
Spitzer, L., Jr., & Shapiro, S. L. 1972, ApJ, 173, 529
Sridhar, S., & Touma, J. R. 2016, MNRAS, 458, 4143
Stodółkiewicz, J. S. 1986, AcA, 36, 19
Theuns, T. 1996, MNRAS, 279, 827
Vesperini, E., McMillan, S. L. W., D'Antona, F., & D'Ercole, A. 2013, MNRAS, 429, 1913
Voigt, H.-H. 1992, Karl Schwarzschild. Collected Works, Vol. 1 (Berlin: Springer)
Wang, L., Spurzem, R., Aarseth, S., et al. 2016, MNRAS, 458, 1450