A New Fokker–Planck Approach for the Relaxation-driven Evolution of Galactic Nuclei

Eugene Vasiliev^{1,2}

¹ Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford, OX1 3NP, UK; eugvas@lpi.ru

² Lebedev Physical Institute, Leninsky prospekt 53, Moscow, 119991, Russia

Received 2017 May 2; revised 2017 September 11; accepted 2017 September 12; published 2017 October 5

Abstract

We present an approach for simulating the collisional evolution of spherical isotropic stellar systems based on the one-dimensional Fokker–Planck equation. A novel aspect is that we use the phase volume as the argument of the distribution function instead of the traditionally used energy, which facilitates the solution. The publicly available code PHASEFLOW implements a high-accuracy finite-element method for the Fokker–Planck equation, and can handle multiple-component systems, optionally with the central black hole and taking into account loss-cone effects and star formation. We discuss the energy balance in the general setting, and in application to the Bahcall–Wolf cusp around a central black hole, for which we derive a perturbative solution. We stress that the cusp is not a steady-state structure, but rather evolves in amplitude while retaining an approximately $\rho \propto r^{-7/4}$ density profile. Finally, we apply the method to the nuclear star cluster of the Milky Way, and illustrate a possible evolutionary scenario in which a two-component system of lighter main-sequence stars and stellar-mass black holes develops a Bahcall–Wolf cusp in the heavier component and a weaker $\rho \propto r^{-3/2}$ cusp in the lighter, visible component, over the period of several Gyr. The present-day density profile is consistent with the recently detected mild cusp inside the central parsec, and is weakly sensitive to initial conditions.

Key words: galaxies: nuclei – galaxies: star clusters: general – Galaxy: center – methods: numerical

1. Introduction

The dynamical evolution of many classes of stellar systems is driven by two-body (collisional) relaxation. While a rigorous theoretical description of this phenomenon needs to take into account spatial inhomogeneities and gravitational polarization effects (e.g., Heyvaerts 2010; Chavanis 2012), this leads to extremely complicated equations, and in practice, simpler approaches are usually taken. One of the commonly used approximations is that the overall effect of two-body relaxation may be described as a sequence of uncorrelated pairwise weak encounters of a test star moving through a uniform infinite medium of field stars. The evolution of the distribution function (DF) of these test stars is described in terms of a Fokker–Planck equation, with advection (drift) and diffusion coefficients for a velocity computed from the DF of the field stars (Rosenbluth et al. 1957). Furthermore, recognizing that the cumulative effect of perturbations is small over the dynamical timescale, the Fokker–Planck equation is written in the orbit-averaged way for the DF expressed in terms of integrals of motion. Finally, identifying the DF of test stars with that of the field stars and mandating that the gravitational potential is determined by the density computed from the DF itself, one arrives at the coupled system of Fokker–Planck and Poisson equations describing the evolution of a stellar system under two-body relaxation. This was first accomplished by Hénon (1961) for a homologous spherical isotropic model, and later by Cohn (1979) for a spherical system with a DF f expressed in terms of energy E and angular momentum L on a suitable grid in this two-dimensional space. In his approach, the DF of field stars, entering the expressions for diffusion coefficients, was taken to be the isotropized version of $f(E, L)$, i.e., a function of E only. This approximation greatly simplifies the calculations and is justified by the fact that the diffusion coefficients are integrals over the DF, hence should be relatively insensitive to its moderate deviation from isotropy. Furthermore, the

solution itself was found to depend on L rather weakly. Recognizing this fact, Cohn (1980) simplified the equations even further by considering the isotropic one-dimensional Fokker–Planck equation for $f(E)$, which substantially increased numerical accuracy.

The next decade has seen great progress in applying this approach to the dynamics of star clusters and galactic nuclei. Several groups have developed independent implementations of the one-dimensional isotropic Fokker–Planck equation, adding various layers of complexity: two or more mass components (Inagaki & Wiyanto 1984), strong scattering (Goodman 1983), stellar mergers and three-body heating (Lee 1987; Quinlan & Shapiro 1989), stellar mass loss and tidal escape (Chernoff & Weinberg 1990), the loss of stars into the central black hole (Murphy et al. 1991, following the earlier work of Cohn & Kulsrud 1978). Later, new methods for solving the two-dimensional equation were presented in Takahashi (1995), Drukier et al. (1999), and Einsel & Spurzem (1999). One-dimensional approximation is sufficient for many problems of interest, and agrees reasonably well with the more general two-dimensional treatment (Cohn 1985). At the same time, alternative approaches were developed: spherical Monte Carlo (e.g., Shapiro 1985) and gaseous (e.g., Louis & Spurzem 1991) models, and a large industry of N -body simulations. Fokker–Planck models still enjoy some popularity, and owing to their low computational demand, allow quickly exploring a large parameter space. So far, none of the existing codes were made public, which certainly limits their usage; we intend to fill this gap with a new implementation that is described in this paper.

The novel aspect of our method is the use of phase volume as the argument of DF, which is a one-dimensional projection of the three-dimensional action space. The use of actions in Fokker–Planck calculations is conceptually cleaner (e.g., Binney & Lacey 1988) and simplifies the recomputation of gravitational potential (the DF evolves adiabatically and hence

is unchanged when expressed in terms of actions). Nevertheless, for unclear reasons, this has never been attempted before, except for the unpublished thesis by Girash (2009). Another novel feature of our implementation is the use of the finite-element method for the Fokker–Planck equation; with higher-order polynomial basis elements, this substantially increases the accuracy of spatial discretization compared to the traditionally employed Chang & Cooper (1970) finite-difference scheme. A similar approach was used in Takahashi (1993), although it was not labeled as such.

We present the basic formalism for dealing with phase volume and the moments of DF in Section 2, introduce the Fokker–Planck equation with all auxiliary ingredients in Section 3, and discuss its conservation properties in Section 4. Then we revisit the classical problem of the cusp formation around a central massive black hole in Section 5, and derive the first-order correction to the scale-free Bahcall–Wolf solution, demonstrating that it does not remain stationary, but rather changes its amplitude while approximately retaining the functional form. In Section 6 we validate our Fokker–Planck code against more sophisticated modeling methods in a test case of a regrowing cusp. Finally, we apply our code to the nuclear star cluster (NSC) of our Galaxy in Section 7 and demonstrate that the present-day structure agrees with the observations under quite general assumptions. Section 8 summarizes our results. Technical details are deferred to the Appendix.

2. Definitions

We consider a spherically symmetric stellar system, consisting of one or several species of stars (components), each one described by a DF $f_c(\mathbf{x}, \mathbf{v})$, $c = 1 \dots N_{\text{comp}}$. We use the convention that the integral $\iiint d^3x \iiint d^3v f_c(\mathbf{x}, \mathbf{v}) = M_c$, the total mass of stars in this component. The mass of an individual star of each species is denoted as m_c . The total gravitational potential is denoted by $\Phi(r)$, and its inverse function is $r_{\text{max}}(E)$, the maximum radius accessible to a star with energy E (so that $\Phi(r_{\text{max}}(E)) = E$).

According to the Jeans theorem, a steady-state DF $f(\mathbf{x}, \mathbf{v})$ must be a function of integrals of motion—in a spherical system, these are the energy $E \equiv \Phi(|\mathbf{x}|) + \frac{1}{2}|\mathbf{v}|^2$ and the angular momentum $L \equiv |\mathbf{x} \times \mathbf{v}|$; in the isotropic case, f may only depend on E .

Instead of E , we use the phase volume h as the argument of the DF. It is defined as the volume of phase space enclosed by the energy hypersurface:

$$h(E) \equiv \iiint d^3x \iiint d^3v \begin{cases} 1 & \text{if } \Phi(|\mathbf{x}|) + |\mathbf{v}|^2/2 < E, \\ 0 & \text{otherwise} \end{cases} \\ = \int_0^{r_{\text{max}}(E)} 4\pi r^2 dr \int_0^{\sqrt{2[E-\Phi(r)]}} 4\pi v^2 dv \quad (1a)$$

$$= \frac{16\pi^2}{3} \int_0^{r_{\text{max}}(E)} r^2 \{2[E - \Phi(r)]\}^{3/2} dr \quad (1b)$$

$$= 16\pi^2 \int_{\Phi(0)}^E dE' \int_0^{r_{\text{max}}(E')} r^2 \sqrt{2[E' - \Phi(r)]} dr \quad (1c)$$

$$= 4\pi^2 \int_0^{L_{\text{circ}}^2(E)} J_r(E, L) dL^2. \quad (1d)$$

Its derivative by energy is called the density of states³:

$$g(E) \equiv \frac{dh(E)}{dE} \\ = 16\pi^2 \int_0^{r_{\text{max}}(E)} r^2 \sqrt{2[E - \Phi(r)]} dr \quad (2a)$$

$$= 4\pi^2 \int_0^{L_{\text{circ}}^2(E)} T_{\text{rad}}(E, L) dL^2. \quad (2b)$$

Here $J_r(E, L) \equiv \int_{r_-}^{r_+} v_r dr$ is the radial action, $T_{\text{rad}}(E, L) \equiv 2 \int_{r_-}^{r_+} dr/v_r = \partial J_r / \partial E$ is the radial period (its dependence on L at a fixed E is usually weak), and $L_{\text{circ}}(E)$ is the angular momentum of a circular orbit with energy E . We note that in the case of a two-dimensional anisotropic Fokker–Planck equation, one could use two action variables J_r and L as arguments of the DF; h is the counterpart 1(d) of these action variables in the one-dimensional case. Cohn (1979) and later studies did express the DF in terms of J_r and L during the recomputation of potential, but still used $f(E, L)$ in the Fokker–Planck equation itself.

The correspondence between energy E and phase volume h is determined by the potential $\Phi(r)$; since both h and g are monotonically increasing functions of E , this is an invertible transformation, so one may equivalently express $E(h)$ and $g(h)$. Conversely, given $E(h)$, one may find the potential $\Phi(r)$ (or, rather, the inverse function $r_{\text{max}}(\Phi)$) using the Abel transform:

$$r_{\text{max}}^3(\Phi) = \frac{3}{8\pi^3} \int_{\Phi(0)}^{\Phi} \frac{dE}{\sqrt{2(\Phi - E)}} \frac{dg(E)}{dE} \\ = \frac{3}{8\pi^3} \int_0^{h(\Phi)} \frac{dh}{\sqrt{2[\Phi - E(h)]}} \frac{dg(h)}{dh}. \quad (3)$$

In a power-law potential $\Phi(r) = \Phi_0 + Cr^{2-\gamma}$ (which corresponds to a density profile $\rho \propto r^{-\gamma}$), the functions g and h also have a power-law behavior: $g(E) \propto \mathcal{E}^{(8-\gamma)/(4-2\gamma)}$, $h(E) \propto \mathcal{E}^{(12-3\gamma)/(4-2\gamma)}$, where $\mathcal{E} \equiv E - \Phi_0$ if the potential is finite at the origin ($\gamma < 2$), or $\mathcal{E} \equiv -E$ otherwise. In both cases, $g(h) \propto h^{(8-\gamma)/(12-3\gamma)}$. In particular, for a Keplerian potential $\Phi = -GM/r$ (either in the vicinity of the central black hole, or at large radii where the density is negligible),

$$h = \frac{2\sqrt{2} \pi^3 (GM)^3}{3(-E)^{3/2}}, \quad g = \frac{(3h)^{5/3}}{4\pi^2 (GM)^2} = -\frac{3}{2} \frac{h}{E}. \quad (4)$$

For a given DF $f(h)$, one may introduce several derived functions of phase volume:

$$I_0(h) \equiv \int_{E(h)}^0 f(E') dE' = \int_h^\infty \frac{f(h')}{g(h')} dh', \quad (5)$$

$$K_g(h) \equiv \int_{E(0)}^{E(h)} f(E') g(E') dE' = \int_0^h f(h') dh', \quad (6)$$

$$K_h(h) \equiv \int_{E(0)}^{E(h)} f(E') h(E') dE' = \int_0^h \frac{f(h') h'}{g(h')} dh', \quad (7)$$

³ Cohn (1980) and other studies use two similarly related quantities: $p \equiv g/(4\pi^2)$, $q \equiv h/(4\pi^2)$.

$$K_E(h) \equiv \int_{E(0)}^{E(h)} f(E') g(E') E' dE' = \int_0^h f(h') E(h') dh'.$$

It is easy to demonstrate that $K_g(h)$ is the mass of stars with energies lower than E (or enclosed by phase volume h), $K_h(h)$ equals to $2/3$ times the kinetic energy of stars enclosed by phase volume h , and $K_E(h)$ measures the total energy (sum of kinetic and potential energies) of stars inside this volume. For instance, writing the kinetic energy as $\iiint d^3x \iiint d^3v f(x, v) |v|^2/2$ and transforming the integration volume to $\int_{\Phi(0)}^E dE' \int_0^{r_{\max}(E')} dr$, as in 1(c), one obtains Equation (7).

A physically valid system must have finite mass, thus $f(h)$ must drop faster than h^{-1} as $h \rightarrow \infty$, and rise slower than h^{-1} as $h \rightarrow 0$. Additionally, the requirement for the energy to be finite imposes a stricter constraint for the inner DF slope in the case of a singular potential ($\gamma \geq 2$): $f(h)$ should grow no faster than $h^{(5\gamma-16)/(12-3\gamma)}$; in the *Keplerian* case ($\gamma = 3$), this reads $f \lesssim Ch^{-1/3} \propto \sqrt{-E}$, which implies that the density profile must be shallower than $\rho \propto r^{-2}$.

The advantage of using h as the argument of the DF is that it is conserved under adiabatic changes of the potential (e.g., Young 1980), which will be important for the Fokker–Planck equation. The density is related to the DF via

$$\begin{aligned} \rho(r) &= 4\pi \int_{\Phi(r)}^0 dE f(E) \sqrt{2[E - \Phi(r)]} \\ &= 4\pi \int_{h[\Phi(r)]}^{\infty} dh' \frac{f(h')}{g(h')} \sqrt{2[E(h') - \Phi(r)]}, \end{aligned} \quad (8)$$

and determines the potential through the Poisson equation, which in the spherically symmetric case yields

$$\Phi(r) = -4\pi G \left[\frac{1}{r} \int_0^r dr' r'^2 \rho(r') + \int_r^{\infty} dr' r' \rho(r') \right]. \quad (9)$$

3. Fokker–Planck Equation

The one-dimensional orbit-averaged Fokker–Planck equation describing the diffusion in energy space can be written in the following flux-conservative form (e.g., Cohn 1980):

$$\frac{\partial [f(E, t) g(E)]}{\partial t} = - \frac{\partial \mathcal{F}(E, t)}{\partial E}, \quad (10a)$$

$$- \mathcal{F}(E, t) \equiv D_{EE}(E) \frac{\partial f(E, t)}{\partial E} + D_E(E) f(E, t). \quad (10b)$$

We take h as the independent variable instead of E , transforming the derivatives as $\frac{\partial}{\partial E} = g(h) \frac{\partial}{\partial h}$. We also add source s and sink $-v f$ terms to this equation, and write it separately for each species c :

$$\frac{\partial f_c(h, t)}{\partial t} = - \frac{\partial \mathcal{F}_c(h, t)}{\partial h} + s_c(h, t) - \nu_c(h, t) f_c(h, t), \quad (11a)$$

$$- \mathcal{F}_c \equiv A_c f_c + D \frac{\partial f_c}{\partial h}. \quad (11b)$$

Here $\mathcal{F}$ is the flux through the phase volume, and the advection and diffusion coefficients are given by

$$A_c(h) = \Gamma m_c \sum_i K_{g,i}(h), \quad (12a)$$

$$D(h) = \Gamma g(h) \sum_i m_i [h I_{0,i}(h) + K_{h,i}(h)], \quad (12b)$$

$$\Gamma \equiv 16\pi^2 G^2 \ln \Lambda, \quad (12c)$$

where $\ln \Lambda$ is the Coulomb logarithm, and the functions $I_{0,i}$, $K_{g,i}$ and $K_{h,i}$ are given by (5)–(7) for each species. Note that the diffusion coefficient D is the same for all species, while the advection coefficient A_c is proportional to the mass of a single star for each species.

The source term $s_c(h)$ may represent the star formation rate per unit phase volume; it is not localized in the real space, but rather spread according to the density generated by a function of h (8)—in this case, the rate of increase of $f(h)$ with time. A δ -function source at h_0 corresponds to the star formation rate $\dot{\rho}(r) \propto \sqrt{E(h_0) - \Phi(r)}$, i.e., somewhat concentrated toward the origin.

The sink term may describe the loss of stars captured by the central black hole. Of course, the capture or tidal disruption occurs when the star passes the pericenter of its orbit at a distance smaller than r_{LC} from the black hole, so any description in terms of orbit-averaged Fokker–Planck equation is necessarily approximate. By returning to the local (non-orbit-averaged) equation, Cohn & Kulsrud (1978) derived a suitable boundary condition for a one-dimensional orbit-averaged diffusion equation in angular momentum, neglecting the diffusion in energy. It is expressed in terms of the loss-cone filling factor q , the ratio between the mean-square change of angular momentum per one orbital period to the width of the loss-cone boundary:

$$q(E) \equiv \frac{\mu T_{\text{rad}}(E, L=0)}{\mathcal{R}_{\text{LC}}}, \quad (13a)$$

$$\mathcal{R}_{\text{LC}}(E) \equiv \frac{L_{\text{LC}}^2(E)}{L_{\text{circ}}^2(E)} = \frac{2 G M_* r_{\text{LC}}}{L_{\text{circ}}^2(E)}, \quad (13b)$$

$$\mu(E) \equiv \frac{8\pi^2}{g(E)} \int_0^{r_{\max}(E)} \frac{\langle \Delta v_{\perp}^2 \rangle r^2 dr}{\sqrt{2[E - \Phi(r)]}}, \quad (13c)$$

$$\langle \Delta v_{\perp}^2 \rangle \equiv \sum_i \Gamma m_i \left[\frac{4}{3} I_{0,i}(E) + 2 J_{1/2,i} - \frac{2}{3} J_{3/2,i} \right], \quad (13d)$$

$$J_n(E, r) \equiv \int_{\Phi(r)}^E dE' f_i(E') \left(\frac{E' - \Phi(r)}{E - \Phi(r)} \right)^n. \quad (13e)$$

μ is the orbit-averaged diffusion coefficient in angular momentum, computed from the local diffusion coefficient in velocity 13(d) via double integration; unlike for the similar coefficients A and D , here the averaging cannot be reduced to a single integral (except for the first term I_0).⁴ These expressions rely on the standard relaxation theory, neglecting the effect of resonant relaxation in a nearly Keplerian potential (Rauch & Tremaine 1996). However, as shown by recent detailed calculations, the overall impact of resonant relaxation on the loss rates is surprisingly moderate, because the enhancement of relaxation rate at intermediate angular momenta is compensated by a suppression of relaxation at very low angular momenta due to rapid relativistic precession (Merriitt 2015; Alexander 2017).

⁴ Most early studies neglected the second term in this integral, corresponding to scattering by stars with higher binding energies; this underestimates the diffusion coefficient at low $|E|$, but the error is rather minor in the region of peak flux ($\lesssim 20\%$), and moreover, low $|E|$ corresponds to the full loss-cone regime where the value of μ does not matter anyway.

Therefore, we retain the classical expressions for the diffusion coefficient in angular momentum.

The steady-state solution to the diffusion equation in angular momentum has a nearly logarithmic profile, determined by the boundary condition (Lightman & Shapiro 1977). If $q \ll 1$ (the empty loss-cone regime), $f \approx 0$ at the loss-cone boundary $\mathcal{R}_{\text{LC}}$, but since $\mathcal{R}_{\text{LC}} \ll 1$, $f(E, L)$ does not vary substantially over most part of the angular-momentum range, and in the opposite, full loss-cone regime, f is even closer to a constant (L -averaged) value. The timescale for establishing the steady-state profile is also much shorter than the timescale for the diffusion in energy, unless the initial distribution was strongly non-uniform over the entire range of L . Therefore, we may approximate the effect of a loss cone in the two-dimensional Fokker–Planck equation for $f(E, L)$ by an energy-dependent loss term $-\nu(E)f(E)$ in the one-dimensional Equation 11(a) for $f(E)$, responsible for the steady-state flux in the angular-momentum direction (e.g., Merritt 2013).⁵ The loss rate is given by

$$\nu = \frac{\mu}{\alpha + \ln(1/\mathcal{R}_{\text{LC}})}, \quad \alpha \approx (q^2 + q^4)^{1/4}. \quad (14)$$

In the absence of the loss-cone effects, the evolution of the system is invariant w.r.t. simultaneous rescaling of time and stellar mass; however, the difference in boundary conditions between empty and full loss-cone regimes breaks this invariance.

The evolution of the entire stellar system is described by a coupled set of Fokker–Planck⁶ and Poisson equations, with the diffusion coefficients computed self-consistently from the DF itself. As in previous studies (e.g., Cohn 1980), we solve them in turn, first advancing the evolution of the DF f_c of all components (the Fokker–Planck step), and then computing the overall density profile and the potential (the Poisson step). The advantage of using h as the independent variable for the DF is important in the Poisson step, where the small adjustment of the potential preserves the DF (adiabatic invariance). Previous studies also used this scheme, but the DF was first expressed in terms of the action variables and then converted back into energy, which is completely unnecessary. Of course, we still need to construct the mapping between Φ and h in the updated potential, since it indirectly enters the expressions for the diffusion coefficients (5)–(7) and (12) through $g(h)$. In some contexts, it may be useful to evolve the system in a fixed external potential (e.g., the Keplerian potential of the central black hole); in this case, the recomputation of $\rho(r)$, $\Phi(r)$ and $h(\Phi)$ is omitted, but the diffusion coefficients still need to be updated in the course of evolution, which is mandatory for the energy conservation. More details about the numerical implementation are given in the Appendix.

4. Conservation Laws

The flux-conservative formulation of the Fokker–Planck equation and its discretized version preserve the mass exactly

⁵ The quantity denoted by α here corresponds to q/ξ in that paper, but the approximation for $\xi(q)$ quoted there actually refers to α , not ξ (clearly $\xi < 1$ from its definition).

⁶ As discussed by Chavanis (2013), the name “Fokker–Planck equation,” strictly speaking, refers to a linear parabolic PDE with diffusion coefficients arising from an external thermal bath, whereas our Equation 11(a) is a nonlinear equation with the evolving DF itself entering the expressions for the diffusion coefficients. However, it is traditionally known by this name in the stellar-dynamical context.

(up to roundoff errors). Since we express f as a function of h and keep it fixed when solving the Poisson equation, it also conserves the mass exactly.

The energy conservation is a more subtle property. In a fixed potential and without source or sink terms, the Fokker–Planck equation alone implies the following evolution equation for the energy density $E(h)f(h, t)$, where in the multicomponent case $f \equiv \sum_c f_c$:

$$\frac{\partial[E(h)f(h, t)]}{\partial t} = -E(h) \sum_c \frac{\partial \mathcal{F}_c(h, t)}{\partial h} = -\frac{\partial \mathcal{F}_E(h, t)}{\partial h}, \quad (15a)$$

$$-\mathcal{F}_E \equiv \sum_c \left[-E \mathcal{F}_c - \frac{D}{g} f_c + A_c I_{0,c} \right], \quad (15b)$$

and the mass flux $\mathcal{F}_c$ of each component is defined by Equation 11(b). This rather weird-looking conservation law results from the fact that the diffusion coefficients contain integrals over the DF itself. Taking the derivative of $\mathcal{F}_E$ by h , the first term yields $E(h) \partial \mathcal{F}(h)/\partial h + \mathcal{F}(h)/g(h)$, and the remaining terms neutralize the second half of this expression, leaving only the product of E and df/dt . The first term in the energy flux corresponds to advective transport (energy carried by particles moving through phase space), and the remainder is the conductive flux representing non-local energy exchange through collisional relaxation.

In the case of evolving potential, it seems impossible to derive a local conservation law, but one may demonstrate the conservation of the total energy, generalizing the derivation presented in the Appendix of Cohn (1979) to the multicomponent case with an external potential. Define $\rho(r)$ to be the density generated by the combined DF of all stellar components f via (8), and let $\Phi_*(r)$ be the potential corresponding to this density via the Poisson Equation (9). We will specialize to the case when an external potential represents the central black hole of mass M_* , but the derivation is similar for an arbitrary external distributed mass profile. Hence, the total potential is given by $\Phi(r) = \Phi_*(r) - G M_*/r$.

Define the total kinetic T and potential energy W as

$$T \equiv \frac{3}{2} \int_{\Phi(0)}^0 dE f(E) h(E) = \frac{3}{2} \sum_c K_{h,c}(h = \infty), \quad (16)$$

$$W \equiv \frac{1}{2} \int_0^\infty dr 4\pi r^2 \rho(r) \Phi_*(r) + M_* \Phi_*(0). \quad (17)$$

For the potential energy to be finite, the density must be shallower (steeper) than $r^{-5/2}$ at small (large) radii, although in fact it must drop faster than r^{-3} at large radii for the total mass to be finite. In the presence of the central black hole, this condition is stricter—the stellar potential must be finite at origin (i.e., the density must be shallower than r^{-2} —same requirement as for K_E to be finite).

When the stellar density $\rho(r, t)$ evolves with time, so does its associated gravitational potential $\Phi_*(r, t)$; hence the rate of

change of the potential energy can be written as

$$\begin{aligned} \frac{dW}{dt} &= \frac{d}{dt} \left[\frac{1}{2} \int_0^\infty dr \, 4\pi r^2 \, \rho(r, t) \, \Phi_*(r, t) + M_* \, \Phi_*(0, t) \right] \\ &= \frac{1}{2} \int_0^\infty dr \, 4\pi r^2 \left\{ \rho(r, t) \frac{\partial \Phi_*(r, t)}{\partial t} + \frac{\partial \rho(r, t)}{\partial t} \Phi_*(r, t) \right\} + \frac{d [M_* \, \Phi_*(0, t)]}{dt} \\ &= \int_0^\infty dr \, 4\pi r^2 \, \rho(r, t) \frac{\partial \Phi_*(r, t)}{\partial t} + \frac{d [M_* \, \Phi_*(0, t)]}{dt}, \end{aligned} \quad (18)$$

where we have integrated the second term in curly braces by parts twice, using the Poisson equation, demonstrating that the two terms have equal contribution to dW/dt .

The sum of energies of all stars is given by

$$\begin{aligned} \mathcal{E} &\equiv \sum_c K_{E,c}(h = \infty) = \int_{\Phi(0)}^0 dE \, f(E) \, g(E) \, E \\ &= \int_{\Phi(0)}^0 dE \, f(E) \, E \, 4\pi \int_0^{r_{\max}(E)} dr \, 4\pi r^2 \sqrt{2[E - \Phi(r)]} \\ &= \int_0^\infty dr \, 4\pi r^2 \, 4\pi \int_{\Phi(r)}^0 dE \, f(E) \sqrt{2[E - \Phi(r)]} \, E. \end{aligned} \quad (19a)$$

Now if we substitute $E = \Phi_*(r) - \frac{GM_*}{r} + \frac{1}{2}v^2$ in the last line, and recall the definition of ρ in Equation (8), then the integral splits into three parts. The first part is twice the potential energy of self-interaction between stars, i.e., the first term in Equation (17) that involves only the product of $\rho(r)\Phi_*(r)$. The last part is the kinetic energy of all stars T , Equation (16). The middle term can be integrated by parts twice to yield $M_* \, \Phi_*(0)$. Hence

$$\mathcal{E} = T + 2W - M_* \, \Phi_*(0). \quad (19b)$$

On the other hand, the total energy of the system is given by

$$U \equiv T + W = \mathcal{E} - W + M_* \, \Phi_*(0). \quad (20)$$

In a virial equilibrium, $T = -W/2 = -U$, and hence $\mathcal{E} = 3U - M_* \, \Phi_*(0)$. We now consider the case when the stellar potential evolves together with the DF, so that the correspondence between h and E depends on time, and there are possibly source or sink terms in the right hand side (rhs) of Equation 11(a). First,

$$\begin{aligned} \frac{d\mathcal{E}}{dt} &= \frac{d}{dt} \int_0^\infty dh \, f(h, t) \, E(h, t) \\ &= \int_0^\infty dh \, \frac{\partial f(h, t)}{\partial t} \, E(h, t) + \int_0^\infty dh \, f(h, t) \, \frac{\partial E(h, t)}{\partial t}. \end{aligned}$$

The first term is converted into the energy flux through the boundary, using Equation 15(a), plus the integral term describing the energy change associated with source or sink terms:

$$\mathcal{S}_E \equiv \int_0^\infty dh \, E(h, t) \sum_c [s_c(h, t) - \nu_c(h, t)f_c(h, t)]. \quad (21)$$

In the second term we expand $\rho(r)$ according to Equation (8), exchange the order of integration in r and h , and replace $\partial E/\partial t$

with the time derivative of the total potential;

$$\begin{aligned} \frac{d\mathcal{E}}{dt} &= -\mathcal{F}_E(h, t) \Big|_{h=0}^\infty + \mathcal{S}_E \\ &\quad + \int_0^\infty dr \, 4\pi r^2 \, \rho(r, t) \left[\frac{\partial \Phi_*(r, t)}{\partial t} - \frac{G}{r} \frac{dM_*}{dt} \right]. \end{aligned}$$

Noting that $\Phi_*(0) = -\int_0^\infty dr \, 4\pi G r \rho(r)$ and invoking Equation (18), we obtain

$$\frac{d\mathcal{E}}{dt} = -\mathcal{F}_E(h, t) \Big|_{h=0}^\infty + \mathcal{S}_E + \frac{dW}{dt} - M_* \frac{d\Phi_*(0, t)}{dt}. \quad (22)$$

Finally, taking the time derivative of U (20), we obtain

$$\begin{aligned} \frac{dU}{dt} &= \frac{d\mathcal{E}}{dt} - \frac{dW}{dt} + \frac{d [M_* \, \Phi_*(0, t)]}{dt} \\ &= -\mathcal{F}_E(h, t) \Big|_{h=0}^\infty + \mathcal{S}_E + \frac{dM_*}{dt} \, \Phi_*(0). \end{aligned} \quad (23)$$

In most cases, the energy flux $\mathcal{F}_E$ tends to zero as $h \rightarrow 0$ or $h \rightarrow \infty$, hence the total energy of the system is conserved in the absence of source/sink terms or changes in M_* ; however, we will later see that the situation is different in the case of a Bahcall–Wolf cusp.

5. The Bahcall–Wolf Cusp Revisited

We now reconsider the classical problem of a steady-state stellar distribution around a central black hole of mass M_* , focusing on the region inside the black hole sphere of influence (defined as the radius r_{infl} enclosing the mass of stars equal to $2M_*$). Hence the potential is determined only by the black hole ($\Phi(r) = -GM_*/r$), and the correspondence between energy and phase volume is given by Equation (4).

Let us first neglect the loss of stars into the black hole, and consider a one-component system described by $f(h)$. In this case, as demonstrated by Bahcall & Wolf (1976), the physically relevant solution has zero flux of mass $\mathcal{F}$, Equation 11(b), which corresponds to $f(E) \propto (-E)^{1/4}$ or $f(h) \propto h^{-1/6}$. However, this solution cannot extend all the way to large radii because the total mass is infinite; hence, we need to consider the effect of small deviations from a pure power law. Let

$$f(h) = C_0 h^{-1/6} + C_1 h^\mu \quad (24)$$

be the perturbed Bahcall–Wolf solution at small h ; we mandate that $\mu > -1/6$ since the first term should dominate as $h \rightarrow 0$. Then, working out the functions I_0 , K_g , K_h and eventually the flux $\mathcal{F}$, we obtain that the dominant term is $\mathcal{F}(h) \propto C_1 h^{\mu+1} f(h)$, hence from Equation 11(a) it follows that the time derivative of f is $\propto C_1 h^\mu f(h)$. If the solution needs to stay self-similar, then $\partial f/\partial t \propto f$, and hence $\mu = 0$. In this case, the mass and energy fluxes are, to the leading order,

$$\mathcal{F}(h) \approx -\frac{29}{20} \Gamma m C_1 h f(h) \propto -C_1 h^{5/6}, \quad (25a)$$

$$\begin{aligned} \mathcal{F}_E(h) &\approx \frac{64}{25} 3^{1/3} \pi^2 G^2 M_*^2 \Gamma m C_0^2 + 5 E(h) \mathcal{F}(h) \\ &\propto \text{const} + C_1 h^{1/6}, \end{aligned} \quad (25b)$$

and the corresponding density profile is

$$\rho(r) = \frac{2^{15/4} 3^{1/6} \pi^2 (G M_\bullet)^{5/4}}{21 \Gamma(\frac{3}{4})^2} \frac{C_0}{r^{7/4}} + \frac{2^{7/2} \pi (G M_\bullet)^{3/2}}{3} \frac{C_1}{r^{3/2}}. \quad (25c)$$

As expected, the mass flux is proportional to C_1 , i.e., it vanishes in the case of a pure Bahcall–Wolf cusp. However, in the perturbed DF the mass flux is small but finite, and is directed inward (if $C_1 > 0$) or outward (if $C_1 < 0$); correspondingly, the DF increases or decreases with time while maintaining approximately the same functional form $\propto h^{-1/6}$. The energy flux, on the other hand, has a constant term that depends on the amplitude of the unperturbed DF (C_0) and corresponds to the energy being pumped into the system through the boundary at $h = 0$. In other words, the black hole acts as a heat source, but the rate of energy production is determined by the stellar distribution itself. This was noted already in the early papers (e.g., Bahcall & Wolf 1976, Section IIIe; Lightman & Shapiro 1977, Section IIa), but attributed to the (negative) energy being carried into the black hole by captured stars (and therefore extracted from the system). As we see from Equation 25(b), the energy flux remains finite even without any absorption of stars, i.e., even if $C_1 = 0$. It may be also understood as the heat conduction flux: the existence of the black hole mandates that the velocity dispersion (temperature) rises toward small radii as $r^{-1/2}$, and hence the energy is transported outward. In this context, the black hole itself acts as a heat bath because its own binding energy is formally infinite. The advective component of the energy flux ($E\mathcal{F} \propto h^{1/6}$) is subdominant to the conductive flux at high $|E|$ —this is an essential feature of the Bahcall–Wolf solution (the dominant term in the mass flux that would be proportional to $h^{2/3}$ vanishes identically, leaving only the next-order correction $\propto h^{5/6}$).

Of course, in reality the black hole imposes an absorbing boundary at a small but finite $h_{\min}$, so that $f(h_{\min}) = 0$, and the flux of mass through the boundary is nonzero. Figure 1 demonstrates that the DF is close to the solution without an absorbing boundary over a large range of h , but has a distinct hump at $h_{\min} < h \lesssim 10^3 h_{\min}$ (top panel, green curve). The mass flux $\mathcal{F}$ is negative in this range, corresponding to the flow toward the black hole; however, outside this region, the flux is still directed outward, as in the case without absorption (middle panel). Moreover, the value of the energy flux $\mathcal{F}_E$ is virtually identical in these two situations (bottom panel), although the importance of different terms in Equation 15(b) varies between these cases. In the case without an absorbing boundary, $E(h)\mathcal{F}(h) \propto h^{1/6}$, whereas the other two terms tend to a constant limit as $h \rightarrow 0$: the energy is transported by conduction, instead of being carried by the mass flow (advection). In the case of an absorbing boundary, the last two terms vanish at $h_{\min}$, since $f(h_{\min}) = 0$ and $A \propto M(h < h_{\min}) = 0$, and the remaining term $\mathcal{F}_E = E(h_{\min})\mathcal{F}(h_{\min})$ corresponds to the energy of captured stars removed from the system. However, in both cases $\mathcal{F}_E$ is ultimately determined by the maximum rate at which the energy can be transported into the outer parts of the system. The asymptotic expressions for the DF, fluxes and density in Equations (24) and (25), match the numerical solution very well in the range of $h \lesssim 0.1 h_{\text{infl}}$, where h_{infl} corresponds to the potential at the radius of influence.

Figure 1. Bahcall–Wolf cusp in three variants: without captures (red solid curves), with an absorbing boundary at $h_{\min} \approx 10^{-11}$ (green dashed curves), and with loss-cone captures at all energies (blue dotted curves). $h = 1$ corresponds to the energy $E = \Phi(r_{\text{infl}})$, where r_{infl} is the influence radius that encloses the mass of stars equal to twice the black hole mass $M_\bullet$. Top panel: normalized DF $f(h)h^{1/6}$; middle panel: mass flux $\mathcal{F}(h)$, Equation 11(b); bottom panel: energy flux $\mathcal{F}_E(h)$, Equation 15(b), and separately the advective part of the energy flux $E\mathcal{F}$ in thinner lines. The initial model had a Plummer DF and a black hole with mass $M_\bullet = 0.1$ times the stellar mass.

Finally, if we account for the loss-cone effects by adding a loss term, Equation (14), the DF becomes somewhat suppressed inside $\sim 0.1 h_{\text{infl}}$, and the mass flux is directed inward in this range (however, it is still two orders of

magnitude lower than the maximum outward flux at roughly the half-mass radius). The rate of energy extraction from the system by captured stars is spread across the entire cusp range, hence the outward energy flux gradually decreases toward $h_{\min}$, but the total energy change rate dU/dt is roughly the same as without the loss-cone effects. In all cases, the advective component of the energy flux is subdominant, as illustrated in the bottom panel of Figure 1.

It is important to note that the relative change of the total energy of the stellar system U is far larger than the change of the total mass (the latter is zero without an absorbing boundary, practically negligible in the case of absorption at $h_{\min}$, and still fairly small if loss-cone effects are included). Hence, the system gradually expands, responding to the energy source at its center. Following Hénon (1975), the long-term asymptotic behavior of an isolated stellar system without mass loss may be described by a self-similar (homologous) model. If we denote the characteristic radius by $\tilde{r}$, the total energy $U \propto GM/\tilde{r}$, and the relaxation time $T_{\text{rel}} \propto \tilde{r}^{3/2} \propto |U|^{-3/2}$. Since the heat production rate in the cusp is determined by the maximum energy flux that can be transported through the system, $\mathcal{F}_E \propto U/T_{\text{rel}}$, the total energy decreases with time as $dU/dt = \mathcal{F}_E \propto |U|^{5/2}$, and hence $U(t) \propto t^{-2/3}$.

Interestingly, the amplitude of the cusp may either increase or decrease with time, depending on the sign of C_1 in (24). Figure 2, top panel, shows the evolution of density profiles of a model with a Plummer DF and a central black hole of mass $M_* = 10^{-3} M$. At first, the Bahcall–Wolf cusp develops at small radii and gradually fades into the cored density profile. As the core radius shrinks and its density increases, so does the amplitude of the cusp, roughly maintaining the $r^{-7/4}$ profile inside a fraction of influence radius. At the time when an equivalent isolated system would reach a core collapse, the density in the cusp attains a maximum, but its profile is much less steep than in a core-collapsed model (e.g., Baumgardt et al. 2005). The subsequent evolution follows a self-similar profile with a gradually decreasing density and proportionally increasing radius. The ratio between the cusp radius and the core radius is $\propto M_*^{1/4}$ for sufficiently small M_* , as shown by Hoggie et al. (2007) from simple dimensional arguments (Figure 2, bottom panel); however, this only holds in the post-collapse phase (i.e., this ratio is clearly not constant in the top panel). The fact that the Bahcall–Wolf cusp evolves in amplitude has not been observed in the early papers, which produced a steady-state solution by fixing the value of the DF at the outer cusp boundary. In reality, this external heat bath must be replaced by a real physical system that responds to the heat source at the center.

6. Validation of the Fokker–Planck Approach

To demonstrate how well the Fokker–Planck description matches the actual evolution of a stellar system, we compare it to two other methods: the stellar-dynamical Monte Carlo code RAGA (Vasiliev 2015), and the N -body code ϕ GRAPEch (Harfst et al. 2008). Both methods represent the system as a collection of discrete particles, as opposed to the description in terms of smooth functions $f(h)$ and $\Phi(r)$ in the Fokker–Planck approach. However, the evolution is treated quite differently in these codes. ϕ GRAPEch is a conventional direct-summation code with GPU acceleration provided by the SAPPORO library (Gaburov et al. 2009) and chain regularization for an accurate treatment of particle encounters with the central black hole. The

Figure 2. Top panel: evolution of density profile in a Bahcall–Wolf cusp. Shown is the density multiplied by r^2 as a function of radius. The gray dotted-dashed line is the original profile—a Plummer sphere in virial units, with a central black hole of mass $M_* = 10^{-3}$ of the total mass; its presence adiabatically modifies the density profile at $r \lesssim 10^{-3}$ to form a weaker cusp $\rho \propto r^{-3/2}$. The blue dashed line is the profile just after the Bahcall–Wolf cusp has formed, at $t \approx 0.25 T_{r,h}$, where $T_{r,h}$ is the half-mass relaxation time. The purple dot-dot-dash line corresponds to $t = 7.5 T_{r,h}$ and the red solid line to $t = 15 T_{r,h}$, at which point the cusp amplitude reaches its maximum. For comparison, a model without a black hole reaches a core collapse at this time, and its density profile is shown as the dotted green line. The orange dashed line is the profile at $t = 30 T_{r,h}$, where the cusp has decreased in amplitude as the energy is being pumped into the system, pushing the mass outward. Bullets mark the influence radius (containing the mass of stars equal to $2M_*$), and crosses denote the cusp radius (defined as $r_{\text{cusp}} \equiv GM_*/\sigma^2$, where σ is the velocity dispersion at this radius) at corresponding times. Bottom panel: density profiles of models with different black hole masses in the self-similar (post-core-collapse) regime. From left to right, models have $M_* = 10^{-5}$ – 10^{-1} ; for comparison, the core-collapsed model without a black hole (same as in the top panel) is shown by a dotted curve. Radii are normalized to the virial radius, i.e., all models have the same total energy. Bullets and crosses mark the influence and cusp radii.

potential computed from particles represents both the smooth global profile and the fluctuations driving the collisional relaxation. By contrast, in the Monte Carlo method the global potential is represented as a smooth function of radius, which is computed from particle positions, but has a much lower noise thanks to several spatial and temporal smoothing techniques. The effect of two-body relaxation is simulated by adding perturbations to particle velocities as they move in the smooth potential; the amplitude of these perturbations follows the same

prescription as in the Fokker–Planck approach, but without orbit averaging. The actual DF represented by particles does not need to be isotropic in the Monte Carlo method; however, in computing the diffusion coefficients, an isotropic approximation is employed.

Vasiliev (2015) demonstrated that the growth of the Bahcall–Wolf cusp is well described by the Monte Carlo approach, in comparison to the direct N -body simulation. We now augment this comparison to include the loss-cone effects. Particles approaching to within a given distance r_{LC} from the black hole are captured, and a certain fraction of their mass is added to the black hole mass; in this test we adopt the accretion fraction of 100%, even though in reality it is likely much smaller than unity (Metzger & Stone 2016). Thus the capture boundary is given in physical space in both the Monte Carlo and the N -body approaches, whereas the Fokker–Planck formulation adopts a more approximate prescription in terms of angular-momentum boundary $L_{\text{LC}} \equiv \sqrt{2GM_*} r_{\text{LC}}$ and a steady-state expression for the loss-cone flux.

We take the initial density to be described by a general double power-law profile (Zhao 1996):

$$\rho(r) \equiv \rho_0 \left(\frac{r}{r_0} \right)^{-\gamma} \left[1 + \left(\frac{r}{r_0} \right)^\alpha \right]^{(\gamma-\beta)/\alpha}, \quad (26a)$$

$$\rho_0 \equiv \frac{M}{4\pi r_0^3} \frac{\alpha \Gamma\left(\frac{\beta-\gamma}{\alpha}\right)}{\Gamma\left(\frac{3-\gamma}{\alpha}\right) \Gamma\left(\frac{\beta-3}{\alpha}\right)}. \quad (26b)$$

In the *Keplerian* potential, an isotropic DF cannot have a density profile shallower than $r^{-1/2}$. For this test, we adopt $\gamma = 0.6$, $\beta = 5$, $\alpha = 2$, the black hole mass $M_* = 0.1 M$, and the capture radius $r_{\text{LC}} = 10^{-5} r_0$. The N -body simulation has $N = 65,535$ equal-mass particles plus the black hole, and to compare the evolution rate with the other methods, we set the Coulomb logarithm to $\ln \Lambda \simeq \ln(M_*/m_*) \approx 9$. The values of N and r_{LC} , of course, are far from realistic for galactic nuclei. If both the relaxation time and the total simulation time are multiplied by K and the capture radius by K^{-1} , this preserves the loss-cone filling factor q , Equation 13(a), hence the captured mass per relaxation time remains almost the same, up to a logarithmic correction in Equation (14). In fact, the adopted values roughly correspond to values of the Milky Way nucleus models from the next section, after rescaling by $K = 350$ (i.e., taking $N = 4 \times 10^7$, $\ln \Lambda = 15$, $r_{\text{LC}} = 3 \times 10^{-8} r_0$ and setting $r_0 = 5 \text{ pc}$, which makes one time unit equivalent to 0.9×10^7 year).

Figure 3 compares the density profiles at different times and the capture rates between three methods. It is clear that the Fokker–Planck and Monte Carlo approaches result in a very similar evolution, which is not surprising because both are based on the same prescription for relaxation. The approximate treatment of the loss cone in the Fokker–Planck method appears to be sufficiently accurate. By contrast, in the N -body system the cusp does not extend all the way to the center, although the density profiles match those of the other methods at radii $\gtrsim 0.03 r_{\text{infl}}$. At smaller radii, the enhancement of angular-momentum diffusion due to resonant relaxation leads to a more rapid loss of stars, preventing the growth of the cusp. However, the total capture rate increases only moderately ($\lesssim 50\%$), in line with the earlier studies (Rauch & Tremaine 1996; Hopman & Alexander 2006). Moreover, in our N -body simulations we neglected relativistic effects, which quench the

Figure 3. Comparison of three methods for studying the evolution of a stellar cusp around a massive black hole: direct N -body simulation (solid blue curves), the Monte Carlo code RAGA (dashed green), and the Fokker–Planck approach from this paper (dotted red). The initial system had a density profile, Equation (26), with the inner slope $\gamma = 0.6$, total mass $M = 1$, and a black hole of mass $M_* = 0.1$, with a capture radius $r_{\text{LC}} = 10^{-5}$. The top panel shows the density at three different moments of time: initial (gray), $t = 1000$ (thinner and darker curves), and $t = 2500$, when the Bahcall–Wolf cusp is fully in place (thicker and lighter curves). The agreement between the Monte Carlo and Fokker–Planck models is very good, while in the N -body model the steep cusp only extends down to $r \simeq 0.02 \simeq 0.03 r_{\text{infl}}$, and becomes shallower farther inward. This could be attributed to the resonant relaxation, which is not accounted for in the other two methods. The bottom panel shows the evolution of the capture rate as a function of time: it increases roughly twofold when the cusp is in place, and then slowly drops as the density continues to decrease in amplitude. Again, the N -body model has a somewhat higher rate owing to the resonant relaxation.

resonant relaxation at high eccentricities and counteract its impact on the density profiles (Merritt 2015).

Overall, the agreement between the approaches is satisfactory, taking into account various approximations made in the Monte Carlo and Fokker–Planck methods. It should be noted that the N -body simulation took a few days, the Monte Carlo simulation a few CPU hours, and the Fokker–Planck run only a few minutes.

7. The Milky Way NSC

We now apply the Fokker–Planck method to construct evolutionary models of the NSC of our Galaxy. At present, it hosts a central black hole of mass $M_* \approx 4 \times 10^6 M_\odot$ (Boehle et al. 2016), and the surface brightness profile of the old stellar

population moderately rises toward the center (Schödel et al. 2017). The slope of the density profile is somewhat lower than expected for a steady-state Bahcall–Wolf profile, but on the other hand, it does not seem to have a central depression (core), as inferred in earlier studies (e.g., Do et al. 2009).

We consider one- and two-component models with an initial density profile described by Equation (26). The total mass of the NSC is taken to be $2.5 \times 10^7 M_\odot$ (Schödel et al. 2014), and we keep it fixed throughout the evolution, postponing the role of star formation for a later study (A. Genozov et al. 2017, in preparation). For the one-component models, we assume equal-mass stars with $m_* = 1 M_\odot$, and for the two-component models, we take the stellar-mass black holes of $m_h = 10 M_\odot$ to contribute 1% to the total mass of the NSC (Alexander 2005), distributed initially with the same profile. The capture radius is set to $r_{LC} = (M_*/m_*)^{1/3} r_* \approx 3.6 \times 10^{-6}$ pc for solar-type stars, and $r_{LC} = 8GM_*/c^2 \approx 1.5 \times 10^{-6}$ pc for black holes; to facilitate the comparison between models, we neglect the growth of M_* due to accreted mass.

We explore two one-parameter families of models, with the inner cusp slope $\gamma = 1/2$ or $3/2$, varied scale radius r_0 , and other parameters fixed to $\alpha = 2$, $\beta = 5$. The former choice is the shallowest possible cusp slope for an isotropic model, and the initial DF is very strongly suppressed at all energies inside the sphere of influence; this type of initial conditions mimics the depletion of the cusp resulting from a previously existing binary SBH (e.g., Merritt & Szell 2006; Merritt 2010). The other choice corresponds to an adiabatically grown cusp around an SBH (it is very similar to the density profile obtained by embedding an SBH into a preexisting Plummer model and adiabatically readjusting the potential while keeping $f(h)$ fixed, e.g., Young 1980; Quinlan et al. 1995). The two sets of models are intended to bracket more realistic cases: in the former case, we expect that the regrowth of the cusp takes a significant time, while in the latter case the relaxation time is shortest and roughly constant inside r_{infl} .

The evolution of all models follows a similar route: first a Bahcall–Wolf cusp grows from the outside in, and subsequently, the system reaches a self-similar expansion regime powered by energy transfer from the SBH. The density profiles at the latter stage have a nearly universal shape: as its amplitude gradually decreases, the characteristic scale proportionally increases. In single-component models with an initially shallow density profile ($\gamma = 1/2$), the formation of the cusp takes longer than the Hubble time, unless the initial scale radius was unrealistically small. By contrast, in two-component models the evolution proceeds much faster as a result of rapid mass segregation, and they reach a self-similar regime much earlier, regardless of the initial density slope. The density of the heavy species (stellar-mass black holes) roughly follows a $\rho \propto r^{-7/4}$ profile at smallest radii, becoming steeper farther out when they cease to dominate in the energy exchange rate (the so-called strong mass segregation regime, Alexander & Hopman 2009; Preto & Amaro-Seoane 2010), while the lighter stars follow a somewhat shallower $\rho \propto r^{-3/2}$ profile inside r_{infl} .

Figure 4 shows the density profiles of several models at the time $t = 10$ Gyr compared to the observations from Schödel et al. (2017). The classical Bahcall–Wolf cusp is clearly excluded by the data, as it produces much too steep projected density profiles at all radii inside r_{infl} . Taken face value, a single-component model with an initially shallow profile and a suitably chosen initial scale radius matches the data best: the

Figure 4. Density profiles of the Milky Way nuclear star cluster. Top panel: surface density profile as a function of projected radius. The gray curve shows the observational data from Schödel et al. (2017), normalized to have the three-dimensional density $\rho = 1.1 \times 10^5 M_\odot \text{ pc}^{-3}$ at $r = 1$ pc; other lines show models that have the same density at 1 pc at time $t = 10$ Gyr. The red solid curve is the one-component model with a steep initial density profile, which has formed a Bahcall–Wolf cusp in less than 1 Gyr. The green dashed line is a one-component model with an initially depleted profile that did not have enough time to regrow the cusp. The blue dotted-dashed line is a similar two-component model, in which the heavy stellar-mass black holes accelerate the formation of the quasi-stationary profile. Middle panel: three-dimensional density profile of the same models, multiplied by r^2 to compress the dynamic range. Additionally, the density profile of the heavy species in the two-component model is shown by the black dotted curve. Bottom panel: evolution of the stellar density at $r = 1$ pc for several two-component models with different initial cusp slopes and scale radii. Dots mark the approximate formation time of the cusp, after which the evolution proceeds self-similarly (i.e., different models have the almost the same profiles, but they were attained at different times).

radius of this shallow “core” shrinks with time and reaches a value $\sim 0.1\text{--}0.2$ pc, just enough to produce a good fit to the surface density in the entire range $10^{-2}\text{--}10$ pc. This was the conclusion reached by Merritt (2010), who used a somewhat different initial profile, also with a shallow core. However, single-component models are not particularly realistic. When we consider two-component models with only 1% contribution of stellar black holes, the situation is very different: the cusp regrows and a self-similar stage is achieved in a much shorter time, less than 10 Gyr, unless the initial scale radius was larger than ~ 5 pc (but in that case the present-day density is too low anyway). There is much less difference between models with initially shallow ($\gamma = 1/2$) or steep ($\gamma = 3/2$) cusps: the bottom panel of Figure 4 demonstrates that they all reach the same self-similar asymptotic regime, although at different moments of time (e.g., the model with initial $r_0 = 4$ pc and $\gamma = 1/2$ has the same density profile at $t = 8$ Gyr as the model with $r_0 = 3$ pc, $\gamma = 1/2$ at $t = 10$ Gyr, or the model with $r_0 = 3$ pc, $\gamma = 3/2$ at $t = 11$ Gyr, and their evolution is almost identical at all later times). As the top panel shows, their projected density profile is still too steep at $r \lesssim 0.1$ pc compared to the observations. Our preliminary tests indicate that adding a moderate amount of continuous star formation at $r \gtrsim 1$ pc drives the resulting density profile very close to the observed profile; we do not report these models here, but defer them to a separate paper.

Overall, the models with the present-day density at $r = 1$ pc in the range $(0.8\text{--}1.5) \times 10^5 M_\odot \text{ pc}^{-3}$ agree rather well with the observed density profile (suitably scaled in amplitude), especially in the middle of this range. The enclosed mass within 1 pc is $(0.6\text{--}1) \times 10^6 M_\odot$, and the half-mass radius is 4–4.5 pc, again in good agreement with observations (e.g., Schödel et al. 2014). As the middle panel of Figure 4 shows, the density of stellar-mass black holes is higher than that of stars inside $r \lesssim 10^{-2}$ pc, which has important implications for the rate of extreme mass-ratio inspirals in Milky Way-sized galactic nuclei (Amaro-Seoane & Preto 2011). The present-day rates of black hole captures and stellar tidal disruptions are close to 5×10^{-6} and 6×10^{-5} events per year, respectively; these values are probably more sensitive to the model assumptions (spherical symmetry, neglect of resonant relaxation) than the inference about the global structure and density profile.

8. Discussion and Conclusions

We have reviewed the Fokker–Planck approach for studying the evolution of spherical isotropic stellar systems driven by collisional relaxation, and presented the publicly available⁷ code PHASEFLOW, which can handle multicomponent systems with star formation and loss-cone effects. A novel aspect in this work is the use of phase volume instead of energy as the argument of the DF, which facilitates the solution. We discussed the energy conservation and transport properties of the system and constructed a perturbative solution for the distribution of stars around a central massive black hole (the Bahcall–Wolf cusp). Despite being commonly labeled as a “steady-state solution,” it actually evolves with time, following the energy transfer from the black hole to the stellar system. At late times, the system approaches a self-similar expansion

regime powered by the central heat source (e.g., Marchant & Shapiro 1980; Merritt 2009).

We applied the method to the NSC of the Milky Way and demonstrated that in the presence of two mass components (lighter stars and heavier stellar-mass black holes), mass segregation leads to accelerated formation of the cusp within a few gigayears, and the subsequent evolution occurs in a self-similar regime with little dependence on initial conditions. The present-day density in the model matches the observed surface brightness profile in the range 0.1–10 pc quite well, although is somewhat steeper farther in. We conjecture that the inclusion of star formation predominantly concentrated at radii $\gtrsim 1$ pc (e.g., Aharon & Perets 2015) would bring the present-day profile into better agreement with observations. The recent N -body model of Baumgardt et al. (2017), which included star formation and realistic mass spectrum, looks quite similar to our simplified two-component Fokker–Planck models, which take only a few seconds to a few minutes to run, allowing a comprehensive exploration of parameter space.

Of course, the one-dimensional Fokker–Planck description is only valid for spherically symmetric systems with isotropic velocity distribution. How serious this limitation is depends on the problem. For instance, angular-momentum relaxation in axisymmetric systems occurs faster, leading to a few-fold increase in the loss-cone flux, as demonstrated by Vasiliev & Merritt (2013) using a two-dimensional Fokker–Planck equation, neglecting the diffusion in energy and assuming a fixed potential of the black hole. Resonant relaxation also leads to a faster diffusion in angular momentum, but taking into account its suppression through relativistic precession of high-eccentricity orbits, the overall impact on the evolution and the loss-cone capture rate is rather moderate (Merritt 2015). In computing the loss-cone flux, we assumed a steady-state profile of the DF in angular momentum, which is established after a small fraction of the relaxation time; at earlier times, the flux may be higher or lower, depending on the initial anisotropy profile (Wang & Merritt 2004; Lezhnin & Vasiliev 2015; Stone et al. 2017). The loss of stars into the black hole has very little impact on the global evolution of the system, which is driven by energy relaxation, therefore these details are immaterial unless we are interested specifically in the rate of tidal disruption events. The DF anisotropy may also be important in the context of tidal mass loss, necessitating a full two-dimensional Fokker–Planck treatment (e.g., Takahashi & Baumgardt 2012).

Fokker–Planck codes for axisymmetric systems are limited to the two-integral case (Goodman 1983; Einsel & Spurzem 1999), and none exist for triaxial systems. On the other hand, collisional relaxation in these systems can be studied using the more general Monte Carlo approach (Vasiliev 2015), which, however, is much more computationally demanding, or still more expensive direct N -body simulations. We believe that the Fokker–Planck method still remains valuable and could be used as a quick tool to explore a large variety of models and determine general trends, complementing the more elaborate approaches. We hope that the software described in this paper and provided to the community will facilitate the applications of this method in various contexts.

This work was supported by the European Research council under the 7th Framework programme (grant No. 321067) and by NASA (grant No. NNX13AG92G). I thank A. Generozov for valuable comments on the early draft and continuous feedback on the code.

⁷ The Fokker–Planck solver is provided as part of the AGAMA library for galaxy modeling (E. Vasiliev 2018, in preparation), available at <https://github.com/GalacticDynamics-Oxford/Agama>.

Appendix Numerical Method

Here we present our formulation of finite-element method for the Fokker–Planck equation and show its relation to the classical Chang & Cooper scheme (which is also included as a special case).

We start by defining a scaled spatial coordinate x instead of h , such that the integrals involving the DF are written as $\int f(x) \mu(x) dx$, where $\mu \equiv dh(x)/dx$. More generally, we define the inner product in the space of all functions of x in the domain $x_- \leq x \leq x_+$ as

$$\langle f(x), g(x) \rangle \equiv \int_{x_-}^{x_+} f(x) g(x) \mu(x) dx. \quad (27)$$

We work with a finite-dimensional subspace of functions $\tilde{f}(x)$ that can be represented in a discretized form:

$$\tilde{f}(x) \equiv \sum_{j=1}^B f_j \hat{e}_j(x), \quad (28)$$

where $\hat{e}_j(x)$ are fixed basis function and f_j are expansion coefficients. To find the coefficients f_j that best describe the discretized counterpart of an arbitrary continuous function $f(x)$, we demand that

$$\begin{aligned} \mathcal{P}_i\{f\} &\equiv \langle f(x), \hat{e}_i(x) \rangle = \langle \tilde{f}(x), \hat{e}_i(x) \rangle \\ &= \sum_{j=1}^B M_{ij} f_j \text{ for all } i = 1 \dots B, \text{ where } M_{ij} \equiv \langle \hat{e}_i, \hat{e}_j \rangle; \end{aligned} \quad (29)$$

in other words, the projection of function $f(x)$ onto each basis vector is the same as the projection of its discrete counterpart. This linear system may be written more compactly as $\mathbf{M}\mathbf{f} = \mathcal{P}\{f\}$ (denoting vectors with boldface and matrices with sans-serif font).

The Fokker–Planck Equation 11(a) for a given species reads

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} &= -\frac{1}{\mu(x)} \frac{\partial \mathcal{F}(x, t)}{\partial x} + s(x, t) - \nu(x, t) f(x, t), \\ -\mathcal{F}(x, t) &\equiv A(x) f(x, t) + \frac{D(x)}{\mu(x)} \frac{\partial f(x, t)}{\partial x}, \end{aligned} \quad (30)$$

where A and D are the advection and diffusion coefficients, Equation (12), s is the source term (star formation rate), ν is the loss-cone draining rate, Equation (14). We now apply the Galerkin projection operator $\mathcal{P}_i\{\circ\} \equiv \langle \circ, \hat{e}_i \rangle$ to both sides of this equation and replace $f(x, t)$ with its discretized representation $\tilde{f}(x, t) \equiv \sum_j f_j(t) \hat{e}_j(x)$:

$$\begin{aligned} \sum_{j=1}^B M_{ij} \frac{df_j(t)}{dt} &= -\int \hat{e}_i(x) \frac{\partial \mathcal{F}(x, t)}{\partial x} dx + s_i(t) - \sum_{j=1}^B V_{ij}(t) f_j(t), \\ s_i &\equiv \mathcal{P}_i\{s\}, \quad V_{ij} \equiv \mathcal{P}_i\{\nu \hat{e}_j\}. \end{aligned} \quad (31)$$

The first term may be integrated by parts to yield

$$\begin{aligned} &[-\mathcal{F}(x, t) \hat{e}_i(x)]|_{x_-}^{x_+} + \int_{x_-}^{x_+} \mathcal{F}(x, t) \frac{d\hat{e}_i(x)}{dx} dx \\ &= [-\mathcal{F}(x, t) \hat{e}_i(x)]|_{x_-}^{x_+} + \sum_{j=1}^B R_{ij} f_j, \\ R_{ij} &\equiv -\int_{x_-}^{x_+} \left(A(x) \frac{d\hat{e}_i(x)}{dx} \hat{e}_j(x) + \frac{D(x)}{\mu(x)} \frac{d\hat{e}_i(x)}{dx} \frac{d\hat{e}_j(x)}{dx} \right) dx. \end{aligned} \quad (32)$$

The expression in square brackets contains the flux through the boundaries of the integration region. For simplicity, we consider only two cases: (a) Neumann boundary condition with $\mathcal{F} = 0$, in which case this term vanishes, or (b) Dirichlet boundary condition with $f(x_-, t) = 0$. For our choice of basis functions (see below), $\hat{e}_i(x_-) = 1$ if $i = 1$ and 0 otherwise; hence the coefficient f_1 is identically zero and may be excluded from the linear system, whereas for $i > 1$ the boundary term vanishes again.

In the matrix form, the Fokker–Planck equation is a first-order differential equation for the vector of coefficients $\mathbf{f}$:

$$\mathbf{M} \frac{d\mathbf{f}(t)}{dt} = (\mathbf{R} - \mathbf{V})\mathbf{f}(t) + \mathbf{s}(t). \quad (33)$$

The key point of the finite-element method is that the basis functions $\hat{e}_j(x)$ are nonzero only in a narrow range of x each, so that their products entering the matrices M , R , V are nontrivial only if $|i - j| \leq N$, where N is on the order of a few. Then the matrices can be efficiently inverted using LU -decomposition with $\mathcal{O}(N^2 B)$ operations (a familiar special case is the tridiagonal matrix algorithm for the case $N = 1$). A suitable choice for the basis are B-spline functions—piecewise polynomials of degree N , defined by a set of grid knots x_k , $k = 1 \dots K$; the total number of basis elements is $B = K + N - 1$, each function is nonzero on at most $N + 1$ consecutive intervals between knots, and has $N - 1$ continuous derivatives at each knot. In the case $N = 1$, the basis element $\hat{e}_k$ is a $\wedge$ -shaped function spanning two adjacent grid cells (from x_{k-1} to x_{k+1}), and the expansion coefficients coincide with the values of the function at each knot ($\tilde{f}(x_k) = f_k$), but in general, this does not hold. To compute the integrals entering the projection operator $\mathcal{P}$ and matrix elements, it is convenient to use the Gauss–Legendre quadrature with $N + 1$ points per each segment $x_k \dots x_{k+1}$, which gives an exact result if the integrand is a polynomial of degree $\leq 2N + 1$ (in particular, a product of two basis functions).

The conventional finite-difference scheme may also be reformulated in the form of the matrix Equation (33) as follows. The basis functions are non-overlapping $\square$ -shaped blocks spanning intervals $x_{k-1/2} \dots x_{k+1/2}$ around each grid node x_k , where the half-indexed points are centers of grid cells $x_{k+1/2} \equiv (x_k + x_{k+1})/2$. The coefficients f_k correspond to the nodal function values $\tilde{f}(x_k)$, the projection operator $\mathcal{P}_k\{f\} = (x_{k+1/2} - x_{k-1/2})f(x_k)$, and the matrices M , V are diagonal with elements $M_{ii} = (x_{i+1/2} - x_{i-1/2})\mu(x_i)$ and $V_{ii} = M_{ii} \nu(x_i)$. The combined advection/diffusion matrix R is tridiagonal, and its elements are computed using the Chang & Cooper (1970) prescription:

$$\begin{aligned} R_{i-1} &= W_{i-1/2}^- C_{i-1/2}, \\ R_{ii} &= -W_{i+1/2}^- C_{i+1/2} - W_{i-1/2}^+ C_{i-1/2}, \\ R_{i+1} &= W_{i+1/2}^+ C_{i+1/2}, \\ C_{i+1/2} &\equiv \frac{1}{(x_{i+1} - x_i)} \frac{D(x_{i+1/2})}{\mu(x_{i+1/2})}, \\ w &\equiv \frac{A(x_{i+1/2})}{C_{i+1/2}}, \quad W_{i+1/2}^- \equiv \frac{w}{\exp w - 1}, \\ W_{i+1/2}^+ &\equiv W_{i+1/2}^- + w. \end{aligned}$$

The idea behind these expressions is that in a near-equilibrium system, the advection and diffusion terms in the flux $\mathcal{F}(x_{i+1/2})$ nearly cancel each other; to make a finite-difference estimate more accurate, we take a suitably weighted combination of f_i and f_{i+1} instead of a simple-minded equal-weight average (see Park & Petrosian 1996 for mathematical details and a comparison of methods). In the finite-element method with stencil width $N > 1$, the flux is estimated more accurately, and these intricacies are unnecessary.

It remains to devise a time integration strategy for Equation (33). The time derivative in the left-hand side is replaced with the finite-difference approximation $df/dt = (f^{\text{new}} - f^{\text{old}})/\Delta t$, but in the right-hand side we may use any combination of the old and the new function values. However, even though the discretized Fokker–Planck equation conserves the mass exactly (in the absence of fluxes through boundaries), it does not automatically conserve energy. Consider the change in total energy $\mathcal{E}$ in one timestep:

$$\begin{aligned} \Delta \mathcal{E} &\equiv \int_0^\infty (f^{\text{new}}(h) - f^{\text{old}}(h)) E(h) dh \\ &= -\Delta t \int_0^\infty \frac{\partial \mathcal{F}(h)}{\partial h} E(h) dh \\ &= -\Delta t [\mathcal{F}(h) E(h)]_0^\infty \\ &\quad + \Delta t \int_0^\infty \frac{\mathcal{F}(h)}{g(h)} dh. \end{aligned}$$

For simplicity, we consider a single-component system with zero-flux boundary conditions. In the numerical approximation of the flux $\mathcal{F}$ we may use a certain linear combination f^E of the old and new DF values for the evolving function, and another combination f^R to compute the relaxation matrix R (note that since the advection/diffusion coefficients linearly depend on the DF, so does the relaxation matrix). Using Equations (5)–(7) and Equation (12), we obtain

$$\begin{aligned} \frac{\Delta \mathcal{E}}{\Delta t} &= -\int_0^\infty dh \left[\frac{f^E(h)}{g(h)} \int_0^h dh' f^R(h') \right] \\ &\quad - \int_0^\infty dh \left[\frac{\partial f^E(h)}{\partial h} \int_0^\infty dh' \frac{f^R(h') \min(h, h')}{g(h')} \right]. \end{aligned}$$

We integrate the second term by parts and again note that the boundary term is zero.

$$\begin{aligned} \frac{\Delta \mathcal{E}}{\Delta t} = & - \int_0^\infty dh \int_0^h dh' \frac{f^E(h) f^R(h')}{g(h)} \\ & + \int_0^\infty dh \int_h^\infty dh' \frac{f^E(h) f^R(h')}{g(h')}. \end{aligned}$$

Finally, we exchange the order of integration in the second term and then switch h and h' :

$$\frac{\Delta \mathcal{E}}{\Delta t} = \int_0^\infty dh \int_0^h dh' \frac{f^E(h') f^R(h) - f^E(h) f^R(h')}{g(h)}. \quad (34)$$

To cancel the energy error, we may take $f^R = f^E$. If both are equal to f^{old} , this corresponds to the explicit Euler method, which is only conditionally stable, however, and in practice would require very short timesteps. Epperlein (1994) suggested an elegant linearization for an implicit Euler method: replace the rhs of Equation (33) with $R^{\text{old}} f^{\text{new}} + R^{\text{new}} f^{\text{old}} - R^{\text{old}} f^{\text{old}}$. The energy error resulting from the first and the second term thus cancels because f^E and f^R are exchanged in Equation (34), and the third term does not introduce any error. Since R^{new} depends linearly on f^{new} , this results in an ordinary linear equation system to be solved at each step; however, its matrix is dense, not band-diagonal. We instead opt to retain the band-diagonal structure of the system by replacing R^{new} with a linear extrapolation constructed from the previous timestep, and solve for the vector f only. We also experimented with a Crank–Nicolson scheme having an equal-weight symmetric combination of old and new R and f in the rhs but found it to be only marginally better in terms of energy conservation, and prone to instability.

The joint evolution of the DF and the potential is followed using the operator-splitting approach: first we advance the DF for a timestep Δt using the Fokker–Planck equation in a fixed potential Φ , and then recompute the stellar density, Equation (8), and potential, Equation (9), in the Poisson step, while keeping $f(h)$ fixed. Since the potential Φ and the mapping between Φ and h also enter the integral for the density, most previous studies updated it in several iterations. Instead we predict the potential $\tilde{\Phi}$ at the end of the timestep by linearly extrapolating its evolution from the previous timestep, recompute the density using this predicted $\tilde{\Phi}$ and its associated mapping between E and h , and then use this density to update the potential. We have checked that this procedure is sufficiently accurate so that further iterations do not significantly improve it, and we perform it after each Fokker–Planck step. As shown in the Appendix of Cohn (1979), the combination of Fokker–Planck and Poisson steps conserves the energy to within $\mathcal{O}(\Delta t^2)$ per timestep, and the error arising from performing only one iteration in our predictor/corrector scheme is also $\mathcal{O}(\Delta t^2)$.

We typically use a uniform grid in the scaled variable $x \equiv \ln h$ covering a sufficiently large range ($\gtrsim 20$ orders of magnitude) with a few hundred points (higher-order finite-element methods need fewer points), and extrapolate the DF outside the grid as a power law in h . Other studies that used $f(E)$ typically employed some scaling transformations to increase the dynamical range and resolution, tailored to the specific problem, whereas in the case of $f(h)$, a uniform grid in $\ln h$ is always a reasonable choice: a significant change in the properties of the system (e.g., the slope of the density profile) is always accompanied by a significant change in h , even if it occurs in a relatively narrow range of E . Density and potential are computed on a logarithmically spaced grid in radius with ~ 100 nodes, which covers the extent of the grid in h , but needs not coincide with it. Various quantities such as $\Phi(r)$, $h(\Phi)$, and $I_0(h)$ are represented by quintic splines, constructed from independently computed values and derivatives of the relevant function at grid nodes, which provide substantially higher interpolation accuracy than cubic splines (well below 10^{-8} with $\lesssim 10$ nodes per decade).

References

- Aharon, D., & Perets, H. 2015, *ApJ*, **799**, 185
 Alexander, T. 2005, *PhR*, **419**, 65
 Alexander, T. 2017, *J. Phys. Conf. Series*, **840**, 012019
 Alexander, T., & Hopman, C. 2009, *ApJ*, **697**, 1861
 Amaro-Seoane, P., & Preto, M. 2011, *CQGra*, **28**, 094017
 Bahcall, J., & Wolf, R. 1976, *ApJ*, **209**, 214
 Baumgardt, H., Amaro-Seoane, P., & Schödel, R. 2017, *A&A*, submitted (arXiv:1701.03818)
 Baumgardt, H., Makino, J., & Hut, P. 2005, *ApJ*, **620**, 238
 Binney, J., & Lacey, C. 1988, *MNRAS*, **230**, 597
 Boehle, A., Ghez, A., Schödel, R., et al. 2016, *ApJ*, **830**, 17
 Chang, J., & Cooper, G. 1970, *JCoPh*, **6**, 1
 Chavanis, P.-H. 2012, *PhyA*, **391**, 3680
 Chavanis, P.-H. 2013, *A&A*, **556**, 93
 Chernoff, D., & Weinberg, M. 1990, *ApJ*, **351**, 121
 Cohn, H. 1979, *ApJ*, **234**, 1036
 Cohn, H. 1980, *ApJ*, **242**, 765
 Cohn, H. 1985, in IAU Symp. 113, *Dynamics of Star Clusters*, ed. J. Goodman & P. Hut (Cambridge: Cambridge Univ. Press), 161
 Cohn, H., & Kulsrud, R. 1978, *ApJ*, **226**, 1087
 Do, T., Ghez, A., Morris, M., et al. 2009, *ApJ*, **703**, 1323
 Drukier, G., Cohn, H., Lugger, P., & Yong, H. 1999, *ApJ*, **518**, 233
 Einsel, C., & Spurzem, R. 1999, *MNRAS*, **302**, 81
 Epperlein, E. 1994, *JCoPh*, **112**, 291
 Gaburov, E., Harfst, S., & Portegies Zwart, S. 2009, *NewA*, **14**, 630
 Girash, J. 2009, PhD thesis, Harvard Univ.
 Goodman, J. 1983, PhD thesis, Princeton Univ.
 Harfst, S., Gualandris, A., Merritt, D., & Mikkola, S. 2008, *MNRAS*, **389**, 2
 Heggie, D., Hut, P., Mineshige, S., Makino, J., & Baumgardt, H. 2007, *PASJ*, **59**, L11
 Hénon, M. 1961, *AnAp*, **24**, 369
 Hénon, M. 1975, in IAU Symp. 69, *Dynamics of Stellar Systems*, ed. A. Hayli (Cambridge: Cambridge Univ. Press), 133
 Heyvaerts, J. 2010, *MNRAS*, **407**, 355
 Hopman, C., & Alexander, T. 2006, *ApJ*, **645**, 1152
 Inagaki, S., & Wiyanto, P. 1984, *PASJ*, **36**, 391
 Lee, H. M. 1987, *ApJ*, **319**, 801
 Lezhnin, K., & Vasiliev, E. 2015, *ApJL*, **808**, L5
 Lightman, A., & Shapiro, S. 1977, *ApJ*, **211**, 244
 Louis, P., & Spurzem, R. 1991, *MNRAS*, **251**, 408
 Marchant, A., & Shapiro, S. 1980, *ApJ*, **239**, 685
 Merritt, D. 2009, *ApJ*, **694**, 959
 Merritt, D. 2010, *ApJ*, **718**, 739
 Merritt, D. 2013, *CQGra*, **30**, 244005
 Merritt, D. 2015, *ApJ*, **810**, 2

- Merritt, D., & Szell, A. 2006, [ApJ](#), **648**, 890
- Metzger, B., & Stone, N. 2016, [MNRAS](#), **461**, 948
- Murphy, B., Cohn, H., & Durisen, R. 1991, [ApJ](#), **370**, 60
- Park, B., & Petrosian, V. 1996, [ApJS](#), **103**, 255
- Preto, M., & Amaro-Seoane, P. 2010, [ApJL](#), **708**, L42
- Quinlan, G., Hernquist, L., & Sigurdsson, S. 1995, [ApJ](#), **440**, 554
- Quinlan, G., & Shapiro, S. 1989, [ApJ](#), **343**, 725
- Rauch, K., & Tremaine, S. 1996, [NewA](#), **1**, 149
- Rosenbluth, M., MacDonald, W., & Judd, D. 1957, [PhRv](#), **107**, 1
- Schödel, R., Feldmeier, A., Kunneriath, D., et al. 2014, [A&A](#), **566**, 47
- Schödel, R., Gallego-Cano, E., Dong, H., et al. 2017, [A&A](#), in press (arXiv:1701.03817)
- Shapiro, S. 1977, [ApJ](#), **217**, 281
- Shapiro, S. 1985, in IAU Symp. 113, Dynamics of Star Clusters, ed. J. Goodman & P. Hut (Cambridge: Cambridge Univ. Press), 373
- Stone, N., Genozov, A., Vasiliev, E., & Metzger, B. 2017, arXiv:1709.00423
- Takahashi, K. 1993, [PASJ](#), **45**, 233
- Takahashi, K. 1995, [PASJ](#), **47**, 561
- Takahashi, K., & Baumgardt, H. 2012, [MNRAS](#), **420**, 1799
- Vasiliev, E. 2015, [MNRAS](#), **446**, 3150
- Vasiliev, E., & Merritt, D. 2013, [ApJ](#), **774**, 87
- Wang, J., & Merritt, D. 2004, [ApJ](#), **600**, 149
- Young, P. 1980, [ApJ](#), **242**, 1232
- Zhao, H.-S. 1996, [MNRAS](#), **278**, 488