Dynamics of galaxy cores and supermassive black holes

David Merritt

Department of Physics, Rochester Institute of Technology, Rochester, NY 14623, USA

E-mail: merritt@astro.rit.edu

Received 25 May 2006
Published 7 August 2006
Online at stacks.iop.org/RoPP/69/2513

Abstract

Recent work on the dynamical evolution of galactic nuclei containing supermassive black holes is reviewed. Topics include galaxy structural properties, collisionless and collisional equilibria, loss-cone dynamics and dynamics of binary and multiple supermassive black holes. (Some figures in this article are in colour only in the electronic version)
Contents

1. Introduction 2515
2. Characteristic length and time scales 2518
3. Structural properties of galaxies and nuclei 2522
4. Collisionless equilibria 2527
   4.1. Spherical nuclei 2528
   4.2. Axisymmetric nuclei 2530
   4.3. Non-axisymmetric nuclei 2532
   4.4. The 'adiabatic growth' model 2537
5. Collisional equilibria 2538
   5.1. The Bahcall–Wolf solution 2539
   5.2. Multi-mass equilibria 2542
   5.3. Cusp regeneration 2544
6. Loss-cone dynamics 2545
   6.1. Classical loss-cone theory 2546
   6.2. Application to galactic nuclei 2550
   6.3. Time-dependent loss cone dynamics 2553
   6.4. Non-axisymmetric nuclei 2555
   6.5. Black-hole-driven expansion 2556
   6.6. Constraining the consumption rate 2558
7. Dynamics of binary black holes 2559
   7.1. Early evolution of binary SBHs and the generation of mass deficits 2560
   7.2. Late evolution of binary SBHs and the ‘final-parsec problem’ 2563
   7.3. Multiple black hole systems 2570
Acknowledgments 2571
References 2571
1. Introduction

Galaxy cores are the hosts of supermassive black holes (SBHs), the engines of quasars and of active galactic nuclei. There is increasing evidence that SBHs play an important role in the formation and global evolution of galaxies and of the intergalactic medium (IGM) (Merloni et al. 2005). The energy released by the formation and growth of SBHs must have had a major impact on how gas cooled to form galaxies and galaxy clusters (Silk and Rees 1998). While the detailed history of SBH growth is still being debated, much work has focused on the possibility that the seeds of SBHs were black holes of much smaller mass—either remnants of the first generation of stars, so-called ‘Population III black holes’ (Madau and Rees 2001), or the (still speculative) ‘intermediate-mass black holes’ (IMBHs), remnants of massive stars that form in dense star cluster via physical collisions between stars (Portegies Zwart et al. 2004).

These exciting developments have led to a renewed interest in the dynamics and evolution of galactic nuclei. Early theoretical studies (Spitzer and Saslaw 1966, Spitzer and Stone 1967, Colgate 1967, Sanders 1970, Shields and Wheeler 1978) emphasized stellar encounters and collisions as the dominant physical processes. In these models, the density of a compact ($\rho \gtrsim 10^6 M_\odot pc^{-3}$) stellar system gradually increases as energetic stars are scattered into elongated orbits via two-body (star-star) gravitational encounters. The increase in density leads to a higher rate of physical collisions between stars; when collision velocities exceed $\sim 10^3$ km s$^{-1}$, roughly the escape velocity from a star, collisions liberate gas that falls to the centre of the system and condenses into new stars which undergo further collisions. Begelman and Rees (1978) argued that the evolution of a dense nucleus would lead inevitably to the formation of a massive black hole at the centre, either by runaway stellar mergers or by the creation of a massive gas cloud which collapses. Subsequent studies (Duncan and Shapiro 1983, David et al. 1987a, 1987b, Quinlan and Shapiro 1987, Quinlan and Shapiro 1989) included ‘seed’ BHs which grow via accretion of stars or gas liberated by stellar collisions or tidal disruptions.

A fundamental time scale in these models is the relaxation time determined by the stars, or

$$T_r \approx \frac{v_{\text{rms}}^3}{G^2 m_* \rho \ln \Lambda}$$

(Spitzer 1987), where $v_{\text{rms}}$ is the stellar velocity dispersion, $m_*$ and $\rho$ are the stellar mass and mass density and $\ln \Lambda \approx 10^1$ is the Coulomb logarithm. In a time $T_r$, gravitational (not physical) encounters between stars can exchange orbital energy and angular momentum, and stars in the high-velocity tail of the distribution will be ejected. The result, after a time $T_c \approx 10^2 T_r$, is ‘core collapse’: the core shrinks to zero size and infinite density. In order for many of the evolutionary models cited above to be viable, $T_c$ must be shorter than the age of the universe, i.e. $T_c$ must be less than $\sim 10^8$ yr, implying very high stellar densities. Physical collisions between stars also feature prominently in many of the early models. Collisions occur on a time scale that is longer than $T_c$ by a factor $(\ln \Lambda)\Theta^2/(1 + \Theta) \approx 10^1$ where $\Theta$, the ‘Safronov number’ (Safronov 1960), is of order unity for stars in a galactic nucleus.

The necessity of attaining high densities for these evolutionary models to work—much higher densities than could be confirmed via direct observation at the time (or indeed now)—was clearly recognized. For instance, Saslaw (1973) noted that

It is an extrapolation from the observations of galaxies we have discussed to the idea that even more dense stellar systems exist...Yet this follows naturally enough from the observations of quasars and the realization that the central density of massive compact stellar systems increases with age.
Spitzer (1971) remarked

...the rate of dynamical evolution will depend on how compact is the stellar system resulting from initial gas inflow. If this rate of evolution is slow, activity will not begin for a long time. In fact, in some systems there might be a wait of $10^{12}$ years before the fireworks begin.

A number of developments since the 1970s have led to a qualitatively different picture of the dynamics and evolution of galactic nuclei.

- SBHs are now known to be ubiquitous components of galactic nuclei and, at least in the most massive galaxies, to have been present with roughly their current masses ($\sim 10^9 M_\odot$) since very early times, as soon as $\sim 10^9$ yr after the Big Bang (Fan et al. 2001, Fan et al. 2003). Evolution of galactic nuclei during and after the era of peak quasar activity therefore took place with the SBHs already in place and processes such as core collapse and the buildup of massive stars via collisions could not have occurred after this time due to the inhibiting effect of the SBH’s gravitational field.

- Observations with the Hubble Space Telescope have elucidated the run of stellar density and velocity dispersion near the centres of nearby galaxies (Crane et al. 1993, Ferrarese et al. 1994, Lauer et al. 1995). In the majority of galaxies massive enough to contain SBHs, the implied central relaxation time is much greater than the age of the universe, due both to the (relatively) low stellar densities and also to the presence of a SBH, which increases $v_{\text{rms}}$ (Faber et al. 1997, Ferrarese et al. 2006a). Only the smallest galaxies known to harbour SBHs have nuclear relaxation times shorter than $10^{10}$ yr. The bulge of the Milky Way is in this category, as is the nearby dwarf elliptical galaxy M32; both have nuclear relaxation times $T_r \approx 10^9.5$ yr short enough that gravitational encounters between stars could have influenced the stellar distribution near the SBH. But in the majority of galaxies, the long relaxation times imply that nuclear structure will still reflect to a large extent the details of the nuclear formation process.

- Since the 1980s, the dominant model for the formation of elliptical galaxies and bulges—the stellar systems that contain SBHs—has been the merger model (Toomre 1977). Galaxy mergers are complex phenomena, but an almost certain consequence of a merger is the infall of the progenitor galaxies’ SBHs into the nucleus of the merged system, resulting in the formation of a binary SBH (Begelman et al. 1980). Such a massive binary would inject a substantial amount of energy into the stellar motions, enough to determine the structure of the remnant core. Indeed this process is probably responsible for the low densities, and long relaxation times, at the centres of the brighter elliptical galaxies.

- Quasars and active nuclei are now believed to be powered by the infall of gas onto a pre-existing SBH. The bulk of the gas is believed to originate far from the nucleus and to be driven inwards by gravitational torques (e.g. Shlosman et al. 1990). The fact that quasar activity peaks at approximately the same epoch as galaxy mergers is seen as strong support for this picture, since mergers are efficient at driving gas into a nucleus (e.g. Kauffmann and Haehnelt 2000). Capture or disruption of stars by the SBH are now generally believed to be energetically insignificant, at least in a time-averaged sense.

While the circumstantial evidence for massive central singularities in galaxies was considered by many to be already compelling in the 1970s, it took much longer for SBHs to be acknowledged as fundamental components of nuclei. A review paper on nuclear dynamics written as recently as twelve years ago (Gerhard 1994) hardly mentions SBHs, even though its author had done pioneering work a decade earlier on the orbital dynamics of galaxies containing massive central objects (Gerhard and Binney 1985). Such conservatism was justifiable given the poor quality of the dynamical evidence for SBHs prior to the late 1990s.
A recurring theme of this paper will be the distinction between collisionless nuclei—which have central relaxation times longer than $10^{10}$ yr—and collisional nuclei in which $T_r \lesssim 10^{10}$ yr (section 3). As noted above, the majority of observed nuclei are ‘collisionless’ in this sense. The morphology of a collisionless nucleus is constrained only by the requirement that the stellar phase–space density $f$ satisfies Jeans’s theorem, i.e. that $f$ be constant along orbits, as they are defined in the combined gravitational potential of the SBH and the stars. This weak condition is consistent with a wide variety of possible equilibrium configurations (section 4), including non-axisymmetric nuclei and nuclei in which the majority of orbits are chaotic. In a nucleus with $T_r \lesssim 10^{10}$ yr, on the other hand, the stellar distribution will have had time to evolve to a more strongly constrained, collisionally relaxed steady state (section 5). For a single-mass population of stars moving in the Keplerian potential of a black hole, the steady-state density profile is $\rho(r) \approx \rho_0 r^{-7/4}$, the so-called Bahcall–Wolf (1976) solution. The distribution of stars near the Milky Way SBH appears to be consistent with this collisionally-relaxed form (Schödel et al. 2006), but there is no evidence for Bahcall–Wolf cusps in any other galaxies.

The recent explosion of data from the Galactic centre has motivated a large and ever-expanding number of theoretical studies of the central star cluster and its evolution due to collisional processes (as reviewed by Alexander 2005). It is important to keep in mind that the Galactic centre is extreme, not only in the sense of harbouring the nearest SBH, but also in the sense of having the smallest SBH with a well-determined mass (Ferrarese and Ford 2005). The bulge of the Milky Way is also one of the faintest systems known to contain a SBH, and its atypically high central density and short relaxation time are consistent with its low luminosity given the parameter correlations defined by more luminous systems (section 3). As important as the Galactic centre is, it is not typical of the majority of nuclei known to contain SBHs.

The distinction between collisionless and collisional nuclei is important in the context of star–SBH interactions. The ‘loss cone’ of a SBH, i.e. the set of orbits that intersect its event horizon or tidal disruption sphere, is rapidly depopulated in a spherical or axisymmetric nucleus, and continued feeding of stars to the SBH requires some mechanism for loss-cone repopulation; typically, this is assumed to be gravitational encounters between stars. Classical loss cone theory (Frank and Rees 1976, Lightman and Shapiro 1977, Cohn and Kulsrud 1978) was worked out in the late 1970s in the context of massive black holes in globular clusters. Globular clusters are many relaxation times old, and loss cone dynamics was approached from the point of view that the distribution of stars would reach a steady state under the competing effects of gravitational scattering and capture. The same theory was later applied, with only minor modifications, to galactic nuclei, most of which however are far too young for this steady state to have been achieved. In a collisionless nucleus, SBH feeding rates can be much higher than in a relaxed nucleus, e.g. if the nucleus is triaxial and many of the orbits are ‘centrophilic;’ or much lower, e.g. if orbits around the loss cone were substantially depeleted by a binary SBH before its coalescence. Considerations like these have prompted a reinvestigation of the nuclear loss cone problem, as reviewed in section 6.

In a collisionless nucleus, the distribution of stars around the SBH should still reflect the energy input from the binary SBH that preceded the current, single hole. (Some nuclei might contain uncoalesced binaries, particularly if they experienced a recent merger, although the observational evidence for binary SBHs is still largely circumstantial; see Komossa 2003.) The effect of binary SBHs on the morphology and kinematics of galactic nuclei is discussed in section 7. An important success of the binary SBH model is its ability to quantitatively explain the low central densities of bright elliptical galaxies, i.e. the ‘cusp-core dichotomy.’ Previously, this dichotomy had been widely interpreted to mean that dwarf and giant elliptical galaxies had experienced very different formation histories (e.g. Kormendy 1985).
The literature on galactic nuclei and SBHs is enormous and exponentially expanding, and the writing of this review paper was greatly facilitated by the recent appearance of a number of other papers that emphasize different aspects of this broad topic. A comprehensive review of observational evidence for SBHs is given by Ferrarese and Ford (2005). Alexander 2003, 2005 reviews the physics of star–star and star–SBH interactions, with an emphasis on dissipative processes and on the nucleus of the Milky Way; as noted above, such processes are of less importance in the nuclei of most galaxies known to contain SBHs. Komossa 2002, 2003 reviews the observational evidence for the interaction of single and binary SBHs with stars and gas in galactic nuclei. (For a more theoretically oriented review of binary SBHs, see Merritt and Milosavljević (2005).) While intermediate mass black holes (IMBHs) are still seen as speculative by many astrophysicists, many of the models proposed for their formation are closely similar to early models for the dissipative evolution of galactic nuclei, as discussed by Miller and Colbert (2004) and van der Marel (2004). The prospect of detecting gravitational waves from coalescing black holes in galactic nuclei is a prime motivation behind much recent work on nuclear dynamics. While no single paper can do justice to this exciting topic, the reviews by Schutz (2002) and Hughes (2003) are excellent places to start. Dark matter is almost certainly not a dominant component of galactic nuclei, but its density there is likely to be higher than elsewhere in galaxies, making galactic nuclei ideal targets for so-called ‘indirect-detection’ studies, as reviewed by Bertone and Merritt (2005). Finally, some early review papers on the evolution of galactic nuclei, while outdated in some respects, are still very much worth reading today, including those by Spitzer (1971), Saslaw (1973), Rees (1984) and Gerhard (1992,1994).

2. Characteristic length and time scales

The radius of the event horizon of a non-rotating hole of mass $M_\bullet$ is

$$r_S = \frac{2GM_\bullet}{c^2} \approx 9.6 \times 10^{-6} \text{ pc} \left( \frac{M_\bullet}{10^8 M_\odot} \right).$$

However, the dynamical influence of a SBH extends far beyond $r_S$. In a stellar nucleus with 1D velocity dispersion $\sigma$, the SBH’s ‘influence radius’ is customarily defined as

$$r_h = \frac{GM_\bullet}{\sigma^2} \approx 11 \text{ pc} \left( \frac{M_\bullet}{10^8 M_\odot} \right) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-2}.$$

In a singular isothermal sphere (SIS) nucleus, $\rho = \sigma^2/2\pi Gr^2$, the velocity dispersion is constant with radius in the absence of the SBH, and the stellar mass within $r_h$ is $2M_\bullet$. This is often taken as an alternate definition of $r_h$: the influence radius is the radius containing a stellar mass equal to $2M_\bullet$ or

$$M_\star (r \leq r_h) = 2M_\bullet.$$

The second definition of $r_h$ is generally to be preferred since $\sigma$ is a function of radius near the SBH, and unless otherwise stated, this is the definition that will be adopted below. In the Milky Way, $r_h \approx 3$ pc according to either definition.

The tidal disruption radius $r_t$ is the distance from the SBH where tidal forces can pull a star apart. Strictly, $r_h$ depends on the structure of the star and on the shape of its orbit, but it is of order

$$r_t \approx r_\star \left( \frac{M_\bullet}{m_\star} \right)^{1/3} \approx 1.0 \times 10^{-5} \text{ pc} \left( \frac{M_\bullet}{10^8 M_\odot} \right)^{1/3},$$

where $m_\star$ is the mass of the star.
where the latter expression assumes stars of Solar mass and radius. In terms of the radius of the event horizon,
\[
r_{\text{e}} \approx \left( \frac{M_*}{10^8 M_\odot} \right)^{-2/3} r_s \tag{6}
\]
so that SBHs more massive than \( \sim 10^8 M_\odot \) ‘swallow stars whole.’

The relaxation time is defined as the time for (mostly distant) gravitational encounters between stars to establish a locally Maxwellian velocity distribution. Assuming a homogenous isotropic distribution of equal-mass stars, the relaxation time is approximately
\[
T_r \approx \frac{0.34 \sigma^3}{G^2 \rho m_* \ln \Lambda} \tag{7a}
\approx 0.95 \times 10^{10} \text{ yr} \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^3 \left( \frac{\rho}{10^6 M_\odot \text{pc}^{-3}} \right)^{-1} \left( \frac{m_*}{M_\odot} \right)^{-1} \left( \frac{\ln \Lambda}{15} \right)^{-1} \tag{7b}
\]
(Spitzer 1987). The Coulomb logarithm, \( \ln \Lambda \), is a ‘fudge factor’ that accounts approximately for the divergent total perturbing force in an infinite homogeneous medium. Within the SBH’s sphere of influence, \( N \)-body experiments (Preto et al. 2004) suggest that
\[
\ln \Lambda \approx \ln \left( r_{\text{h}} \sigma^2 / 2 G m_* \right) \approx \ln \left( M_* / 2 m_* \right) \approx \ln (N_*/2) \tag{8}
\]
with \( N_* \equiv M_* / m_* \), the number of stars whose mass makes up \( M_* \). For \( m_* = M_\odot \) and \( M_* = (0.1, 1, 10) \times 10^8 M_\odot \), \( \ln \Lambda \approx (15, 18, 20) \). In the gravitational field of a SBH, i.e. at \( r \lesssim r_{\text{h}} \), encounters lead to a steady-state distribution of orbital energies in a time \( \sim T_r \), as discussed in section 5.

Figure 1 shows estimates of \( T_r \), measured at \( r = r_{\text{h}} \), in a sample of early-type galaxies and bulges. The relaxation time at \( r_{\text{h}} \) is almost always in excess of \( 10^{10} \) yr, although there is a clear trend with luminosity, suggesting that \( T_r (r_{\text{h}}) \) drops below \( 10^{10} \) yr for spheroids fainter than absolute magnitude \( M_V \approx -18 \), roughly the luminosity of the Milky Way bulge. Furthermore, in the handful of Local Group galaxies for which the SBH influence radius is well resolved, \( T_r \) continues to decrease inside of \( r_{\text{h}} \): in the case of the Milky Way, to \( \sim 6 \times 10^9 \) yr at 0.2\( r_{\text{h}} \) and \( \sim 3.5 \times 10^9 \) yr at \( \sim 0.1 r_{\text{h}} \). At radii \( r \ll r_{\text{h}} \) where \( \sigma^2 \propto r^{-1} \), the relaxation time varies as \( T_r \sim r^{\gamma-3/2} \) if \( \rho \propto r^{-\gamma} \). Several Local Group galaxies are known to have \( r_{\text{h}} \sim 1.5 \) at \( r \lesssim r_{\text{h}} \) (Lauer et al. 1998) implying that \( T_r \) is approximately constant in the SBH in these galaxies. Relaxation times in the nuclei of brighter galaxies are probably always longer than a Hubble time, and the stellar distribution at \( r \lesssim r_{\text{h}} \) in these galaxies should still reflect the details of their formation.

In what follows, collisional nuclei are defined as those which have \( T_r \lesssim 10^{10} \) yr at \( r < r_{\text{h}} \), while collisionless nuclei have \( T_r \gtrsim 10^{10} \) yr.

A number of other physical processes have time scales that are related to \( T_r \). The core-collapse time \( T_{cc} \) is \( \sim 10^5 T_r \); in a time of order \( T_{cc} \), a nucleus lacking a SBH develops a high-density core and a power-law envelope, \( \rho \sim r^{-2.2} \). The core collapse time is interestingly short at the centres of M33 and NGC205: both have \( T_c \lesssim 10^8 \) yr (Hernquist et al. 1991, Valluri et al. 2005) and both lack dynamical signatures of a SBH. However, core collapse is probably unimportant in galaxies with SBHs, both because \( T_r \) is \( \gtrsim 10^9 \) yr and because the collapse would be inhibited by the presence of the hole. Another relaxation-driven mechanism is diffusion of stars into the tidal disruption sphere of the SBH. The rate is approximately \( \dot{N} \approx N / (\ln(2/\theta_{lc}^2)) T_r \), where \( \theta_{lc} \approx \sqrt{r_{\text{c}}/r_{\text{h}}} \) is the angular size subtended by the sphere \( r = r_{\text{c}} \) as seen from \( r = r_{\text{h}} \) and \( N \) is the number of stars within \( r_{\text{h}} \) (Frank and Rees 1976, Lightman and Shapiro 1977).
Physical collisions between stars take place in a time $T_{\text{coll}}$ where

$$T_{\text{coll}} = \left[ 16 \sqrt{\frac{\pi}{3}} n \sigma^2 r^2 (1 + \Theta) \right]^{-1}$$

or

$$\approx 1.1 \times 10^{11} \text{yr} \left( \frac{n}{10^6 \text{pc}^{-3}} \right)^{-1} \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-1} \left( \frac{r}{R_\odot} \right)^{-2} \left( \frac{1 + \Theta}{3} \right)^{-1}$$

and $\Theta$ is the Safronov number:

$$\Theta = \frac{G m_\star}{2 \sigma^2 r_\star}$$

or

$$\approx 2.38 \left( \frac{m_\star}{M_\odot} \right) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{2} \left( \frac{r_\star}{R_\odot} \right)^{-1}$$

The ratio of the collision time to the relaxation time is

$$\frac{T_{\text{coll}}}{T_r} \approx 0.8 \ln \Lambda \frac{\Theta^2}{1 + \Theta} \approx 0.3 \left( \frac{r}{0.01 \text{ pc}} \right)^4$$

the right-hand expression assumes Solar-type stars around the Milky Way SBH and shows that even in the very dense environment of the Galactic centre, physical collisions are significant only at very small radii, $r \lesssim 0.02 \text{ pc} \approx 10^{-2} r_\odot$ for Solar-type stars. Stellar collisions may be responsible for the depletion of luminous, late-type giant stars at the Galactic centre (Phinney 1989, Alexander 1999, Alexander 2003) or for the formation of the so-called S-stars.
However, there is little evidence that stellar collisions have affected the stellar populations at the centres of any of the other Local Group galaxies (e.g. Lauer et al. 1998). From the point of view of nuclear dynamics, stellar collisions are probably always of minor importance (e.g. Freitag and Benz 2002) and they will not be discussed further in this paper.

The spheroids that host SBHs are believed to have formed via mergers, and in many cases, the merging galaxies would have contained pre-existing SBHs. The result is a binary SBH (Begelman et al. 1980). A number of additional length and time scales are associated with the binaries. Let $M_1$ and $M_2$ be the masses, respectively, of the larger and smaller of the two SBHs, with $q = M_2/M_1 \leq 1$ the mass ratio, $M_{12} = M_1 + M_2$ the total mass and $\mu = M_1 M_2/M_{12}$ the reduced mass. Two SBHs form a gravitationally bound pair when their separation falls below the influence radius defined by the larger hole. Approximating the relative orbit as Keplerian, i.e. ignoring the force perturbations from stars, the binary’s binding energy is

$$|E| = \frac{G M_1 M_2}{2a} = \frac{G \mu M_{12}}{2a}$$

with $a$ the semi-major axis and the orbital period is

$$P = 2\pi \left( \frac{a^3}{GM_{12}} \right)^{1/2} = 9.36 \times 10^3 \text{ yr} \left( \frac{M_{12}}{10^8 M_\odot} \right)^{-1/2} \left( \frac{a}{1 \text{ pc}} \right)^{3/2}.$$  

The relative velocity of the two SBHs, assuming a circular orbit, is

$$V_{\text{bin}} = \sqrt{\frac{GM_{12}}{a}} = 658 \text{ km s}^{-1} \left( \frac{M_{12}}{10^8 M_\odot} \right)^{1/2} \left( \frac{a}{1 \text{ pc}} \right)^{-1/2}.$$  

A massive binary is called ‘hard’ when its binding energy per unit mass, $|E|/M_{12} = G \mu/2a$, exceeds $\sim \sigma^2$. (The motivation for this definition is given in section 7.) A standard definition for the semi-major axis of a hard binary is

$$a_h = \frac{G \mu}{4 \sigma^2} \approx 2.7 \text{ pc} (1 + q)^{-1} \left( \frac{M_2}{10^8 M_\odot} \right) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-2}.$$  

Defining $r_h$ as $GM_{12}/\sigma^2$ then implies

$$a_h = \frac{\mu}{4M_{12}} r_h = \frac{1}{4} q (1+q)^2 r_h.$$  

For an equal-mass binary, $a_h \approx 0.06 r_h$, and for a more typical mass ratio of $q = 0.1$, $a_h \approx 0.02 r_h$.

If the binary’s semi-major axis is small enough that its subsequent evolution is dominated by emission of gravitational radiation then $\dot{a} \propto -a^{-3}$ and coalescence takes place in a time $t_{gr}$, where (Peters 1964)

$$t_{gr} = \frac{5}{256 F(e) G^3 \mu M_{12}^3} a^4,$$

$$F(e) = (1 - e^2)^{3/2} \left( 1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right).$$

This can be written as

$$t_{gr} \approx \frac{5}{16^4 F(e) \sigma^8 M_{12}^3} \left( \frac{a}{a_h} \right)^4 \approx 3.07 \times 10^8 \text{ yr} \frac{q^3}{(1+q)^6} \left( \frac{M_{12}}{10^8 M_\odot} \right) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-8} \left( \frac{a}{10^{-2} a_h} \right)^4.$$
with $e$ the orbital eccentricity. Separations much less than $a_h$ are required in order for emission of gravitational waves to induce coalescence in $10^{10}$ yr.

Many of these expressions can be simplified by making use of the empirical correlations between SBH mass and galaxy properties. The tightest of these is the $M - \sigma$ relation, which has two extant forms, based either on the velocity dispersion $\sigma_c$ measured in an aperture centred on the nucleus (but large compared with $r_h$) (Ferrarese and Merritt 2000) or on the global rms stellar velocity (Gebhardt et al. 2000). The first form is relevant here since $\sigma$ as used above is defined near the centre of a galaxy; the alternative form (Gebhardt et al. 2000) defines $\sigma$ as a mean value along a slit that extends over the entire half-light radius of the galaxy and can be strongly influenced by rotation, inclination and other factors extraneous to nuclear dynamics. (The Gebhardt et al. form of the relation also exhibits substantially more scatter.) In terms of $\sigma_c$, the best current determination of the $M - \sigma$ relation is

$$
\left( \frac{M_*}{10^8 M_\odot} \right) = (1.66 \pm 0.24) \left( \frac{\sigma_c}{200 \text{ km s}^{-1}} \right)^{4.86},
$$

with $\alpha = 4.86 \pm 0.43$ (Ferrarese and Ford 2005). The intrinsic scatter in this relation is consistent with zero (Merritt and Ferrarese 2001c, Ferrarese and Ford 2005), although the number of galaxies with well-determined SBH masses is still very small, of order a dozen or less (Valluri et al. 2004). The relations between $M_*$ and the mass or luminosity of the host spheroid (elliptical galaxy or spiral galaxy bulge) appear to be less tight; the mean ratio of $M_*$ to $M_{gal}$ is $\sim 1.3 \times 10^{-3}$ (Merritt and Ferrarese 2001a, Marconi and Hunt 2003).

Using equation (19), one can write

$$
r_h \approx 18 \text{ pc} \left( \frac{\sigma_c}{200 \text{ km s}^{-1}} \right)^{2.86}
$$

or

$$
\approx 13 \text{ pc} \left( \frac{M_*}{10^8 M_\odot} \right)^{0.59},
$$

and

$$
a_h \approx 4.5 \text{ pc} \left( \frac{q}{1 + q} \right)^2 \left( \frac{\sigma_c}{200 \text{ km s}^{-1}} \right)^{2.86},
$$

or

$$
\approx 3.2 \text{ pc} \left( \frac{q}{1 + q} \right)^2 \left( \frac{M_*}{10^8 M_\odot} \right)^{0.59}.
$$

Setting $M = M_{12}$, $F(e) = 1$ in equation (18) gives

$$
t_{12} \approx 5.0 \times 10^8 \text{ yr} \left( \frac{q^3}{1 + q} \right)^{3.14} \left( \frac{\sigma_c}{200 \text{ km s}^{-1}} \right)^{-3.14} \left( \frac{a}{10^{-2} a_h} \right)^4,
$$

or

$$
\approx 7.1 \times 10^8 \text{ yr} \left( \frac{M_{12}}{10^8 M_\odot} \right)^{-0.65} \left( \frac{a}{10^{-2} a_h} \right)^4.
$$

3. Structural properties of galaxies and nuclei

Knowledge of the distribution of mass and light near the centres of galaxies is limited by the $\sim 0.1''$ angular resolution of the Hubble Space Telescope (although ground-based adaptive optics are starting to improve on this, e.g. Melbourne et al. 2005, Davidge et al. 2005). The corresponding linear scale is $\sim 5 \text{ pc} (D/10 \text{ Mpc})$; at the distance of the Virgo cluster ($D \approx 16 \text{ Mpc}$), resolving $r_h$ is only possible in galaxies with $M_* \gtrsim 10^8 M_\odot$ (equation (20b)). While a great deal is known about the morphology of bright ($M_B \lesssim -20$) galaxies on scales
$r < r_h$, the central structure of fainter spheroids is poorly understood, and what little we know comes almost entirely from Local Group galaxies (including the bulge of the Milky Way). Not surprisingly, parametric functions that were developed to fit the luminosity profiles of distant galaxies often fail to describe the well-resolved centres of Local Group galaxies. In this section, an attempt is made to synthesize recent work on the light distributions in both bright (distant) and fainter (mostly nearby) galaxies.

Among the various functional forms that have been proposed to fit galaxy luminosity profiles, the most generally successful have been those based on S´ersic (1968) law,

\[ \ln I(R) = \ln I_e - b(n_S) \left( \frac{R}{R_e} \right)^{1/n_S} - 1 \]  

(23)

The constant $b$ is normally chosen such that $R_e$ is the projected radius containing one-half of the total light. The shape of the profile is then determined by $n_S$; $n_S = 4$ is the de Vaucouleurs (1948) law, which is a good representation of bright elliptical galaxies, while $n_S = 1$ is the exponential law, which approximates the luminosity profiles of dwarf galaxies (Binggeli et al. 1984). Deviations from the best-fitting S´ersic law are typically 0.05 mag rms and, at least in the case of galaxies outside the Local Group, the fits are often good over the full observed range, typically two to three decades in radius (Graham and Guzmán 2003, Graham et al. 2003, Trujillo et al. 2004). An alternative way to write equation (23) is

\[ \frac{d \ln I}{d \ln R} = - \frac{b}{n_S} \left( \frac{R}{R_e} \right)^{1/n_S}, \]  

(24)

i.e. the logarithmic slope varies as a power of the projected radius, falling to zero at the centre. While there is no consensus on why the S´ersic law is such a good representation of galactic spheroids, a possible hint comes from the dark-matter halos produced in $N$-body simulations of hierarchical structure formation, which have density profiles that are also well described by equation (24) (Navarro et al. 2004, Graham et al. 2005). This functional form may be characteristic of systems that form via chaotic, collisionless relaxation.

The S´ersic index $n_S$ correlates reasonably well with galaxy absolute magnitude (Graham and Guzmán 2003) and less well with other structural parameters such as $R_e$ (Caon et al. 1993, Graham and Guzmán 2003). In a recent comprehensive study (Ferrarese et al. 2006a), the Advanced Camera for Surveys on the Hubble Space Telescope was used to measure luminosity profiles of 100 early-type galaxies in the Virgo cluster. The best-fitting relations $n_S(M_B)$ and $R_e(M_B)$ were found to be

\[ \log n_S \approx \log(3.89) - 0.10(M_B + 20), \]  

(25)

\[ \log R_e \approx \log(17.8) - 0.055(M_B + 20), \]  

(26)

where $R_e$ is in units of kpc.

In galaxies resolved on scales of order a few $r_h$ or better, systematic deviations from the S´ersic law often appear near the centre. These deviations are of two kinds. Galaxies fainter than $M_B \approx -20$ are generally observed to have higher surface brightnesses at small radii than predicted by S´ersic’s law. These galaxies are sometimes called ‘power-law’ galaxies since $I(R)$ can be reasonably well approximated as a power law at $R < R_e$. Galaxies brighter than $M_B \approx -20$ generally exhibit central deficits in the intensity, inside of a ‘break’ radius $R_b$ that is of order a few times $r_h$. These are sometimes called ‘core’ galaxies.

The surface brightness surpluses at small radii are most clearly defined in Local Group galaxies such as the Milky Way, M31 and M32. Each of these galaxies exhibits an approximately power-law dependence of $I$ on $R$ into the innermost resolved radius, $I \sim R^{-\gamma}$, implying a spatial density $\rho \sim r^{-\gamma}$. $\gamma \approx 16$ (figure 2). In the bulge of the Milky Way,
stellar density has been derived from number counts that extend down to \( \sim 0.005 \text{ pc} \approx 10^{-3} r_h \); the result is

\[
\rho(r) \approx 1.8 \times 10^5 M_\odot \text{ pc}^{-3} \left( \frac{r}{0.38 \text{ pc}} \right)^{-\alpha}.
\]

with \( \alpha \approx 2.0 \) at \( r \geq 0.38 \text{ pc} \) and \( \alpha \approx 1.4 \) at \( r < 0.38 \text{ pc} \) (Genzel et al. 2003, Schödel et al. 2006). M31 and M32 have central density profiles very similar to that of the Milky Way, both in slope and normalization (Lauer et al. 1998), although \( I(R) \) in both galaxies cannot be measured on scales much smaller than \( r_h \). Deviations from a Sérsic law in M32 appear inside of \( R \approx 10'' \approx 40 \text{ pc} \approx 20r_h \) (Tonry 1984, Graham 2002), which suggests that the central surplus cannot be ascribed solely to the dynamical influence of the SBH.

Figure 2. Mass density profiles near the centres of the Milky Way and M32 (Schödel et al. 2006, Lauer et al. 1998). Dashed line is \( \rho \propto r^{-1.5} \). Both galaxies contain SBHs with masses \( \sim 3 \times 10^6 M_\odot \) and with influence radii \( r_h \approx 3 \text{ pc} \).
Dynamics of galaxy cores

Figure 3. (Left Panel) Mass of the ‘central massive object’ (CMO) plotted against absolute blue magnitude of the host galaxy (or bulge for spiral galaxies). Stellar nuclei (defined as an excess with respect to the best-fitting Sérsic profile) are shown as red squares. SBHs in elliptical and spiral galaxies are shown as filled and open circles respectively. Upper limits on the SBH mass are shown for NGC205 and M33. (Middle Panel) CMO mass as a function of velocity dispersion of the host galaxy. (Right Panel) CMO mass plotted against galaxy mass. The solid red and black lines show the best-fit relations fit to the nuclei and to the SBH samples, respectively, with 1σ confidence levels on the slope shown by the dotted lines. In the middle panel, the dashed line is the $M_r-\sigma$ relation. In the right panel, the dashed line is the fit obtained for the combined nuclei+SBH sample. (Adapted from Ferrarese et al. 2006b.)

Very recently, a possible connection has been found between the central luminosity excesses in low-luminosity galaxies and the SBHs in brighter galaxies (Ferrarese et al. 2006b, Wehner and Harris 2006). If the ‘nucleus’ is defined as the interior light or mass in excess of the best-fitting Sérsic profile then the ratio of ‘nuclear’ mass to total galaxy mass is found to be roughly equal to the ratio $M_*/M_{gal}$ in galaxies with SBHs or $\sim 10^{-3}$ (figure 3). This has led to the suggestion that galaxies always form a ‘central massive object’ (CMO): either a SBH or a compact stellar nucleus. A possible objection to this appealing picture is that it fails to describe some of the best-resolved galaxies, e.g. M31, M32 and the Milky Way, all of which contain SBHs as well as ‘nuclei,’ i.e. central excesses with respect to the best-fitting Sérsic law (e.g. Graham 2002). It is possible that the ‘nuclei’ in some of the unresolved galaxies in figure 3 are similar to the power-law cusps observed in the well-resolved Local Group galaxies, rather than distinct, compact cores such as in NGC 205.

At the other extreme of luminosity, bright elliptical galaxies generally show central deficits in $I(R)$ or ‘cores.’ The cores extend outwards to a break radius $R_b$ of order a few times $r_h$. The profiles at $R \lesssim R_b$ are well described as power laws but with small slopes, $\Gamma \lesssim 0.3$, and furthermore the profiles often show a distinct inflection or change in slope, at $R_b$ (Lauer et al. 1995, Trujillo et al. 2004, Ferrarese et al. 2006a). Such cores are apparent in elliptical galaxies

region dominated by the gravitational force from the SBH, say $r \lesssim 0.2r_h$, processes such as adiabatic contraction (Young 1980) and collisional relaxation (Bahcall and Wolf 1976) can produce power-law profiles, but these mechanisms are ineffective at $r \gtrsim r_h$. Very recently, a possible connection has been found between the central luminosity excesses in low-luminosity galaxies and the SBHs in brighter galaxies (Ferrarese et al. 2006b, Wehner and Harris 2006). If the ‘nucleus’ is defined as the interior light or mass in excess of the best-fitting Sérsic profile then the ratio of ‘nuclear’ mass to total galaxy mass is found to be roughly equal to the ratio $M_*/M_{gal}$ in galaxies with SBHs or $\sim 10^{-3}$ (figure 3). This has led to the suggestion that galaxies always form a ‘central massive object’ (CMO): either a SBH or a compact stellar nucleus. A possible objection to this appealing picture is that it fails to describe some of the best-resolved galaxies, e.g. M31, M32 and the Milky Way, all of which contain SBHs as well as ‘nuclei,’ i.e. central excesses with respect to the best-fitting Sérsic law (e.g. Graham 2002). It is possible that the ‘nuclei’ in some of the unresolved galaxies in figure 3 are similar to the power-law cusps observed in the well-resolved Local Group galaxies, rather than distinct, compact cores such as in NGC 205.

At the other extreme of luminosity, bright elliptical galaxies generally show central deficits in $I(R)$ or ‘cores.’ The cores extend outwards to a break radius $R_b$ of order a few times $r_h$. The profiles at $R \lesssim R_b$ are well described as power laws but with small slopes, $\Gamma \lesssim 0.3$, and furthermore the profiles often show a distinct inflection or change in slope, at $R_b$ (Lauer et al. 1995, Trujillo et al. 2004, Ferrarese et al. 2006a). Such cores are apparent in elliptical galaxies
brighter than $M_V \approx -19.5$, but this number may be resolution-dependent since fainter galaxies are mostly unresolved on scales of $r_h$.

One model, discussed in detail below, attributes the cores to the dynamical influence of binary SBHs. If this model is correct, the ‘mass deficit’—the amount of mass that was removed in creating the core—is a measure of the time-integrated effect of the binary on the nucleus. The mass deficit is defined (Milosavljević et al. 2002) as the difference in integrated mass between the deprojected density profile $\rho(r)$ and an inwards extrapolation of the outer, deprojected profile $\rho_{\text{fit}}(r)$:

$$M_{\text{def}} \equiv 4\pi \int_0^{r_b} [\rho_{\text{fit}}(r) - \rho(r)] r^2 dr.$$ (28)

Here, $r_b$ is the (spatial) radius where the luminosity profile departs from the fitted profile. Figure 4 illustrates the computation of $M_{\text{def}}$ in NGC 3348. Note that the pre-existing profile was assumed to be a Sérsic law; under this assumption, the mass deficit is found to be roughly equal to $M_\star$, the mass currently in the SBH (Graham 2004). This is consistent with the mass displaced by an infalling SBH (section 7).

The idea that the cores of bright elliptical galaxies are created during mergers provides a nice solution to a long-standing puzzle: why dwarf and giant elliptical galaxies appear to occupy two distinct families in terms of their central properties (Jerjen and Binggeli 1997, Graham and Guzmán 2003). Dwarf ellipticals define a continuous sequence spanning ten magnitudes, such that the central surface brightness increases linearly with absolute magnitude. In galaxies brighter than $M_B \approx -20.5$, the central surface brightness declines with increasing luminosity, leading Kormendy (1985) to suggest that ‘dwarf elliptical galaxies are very different from the sequence of giant ellipticals.’ But $M_B \approx -20.5$ is also roughly where cores appear, and if the cores are ‘removed’ as in figure 4 by fitting a smooth profile to the surface brightness data at $R > R_b$, one finds that bright ellipticals smoothly continue the sequence defined by dwarf ellipticals of central density increasing with luminosity. Figure 5 illustrates this for a sample of early-type galaxies in Virgo. Even after this adjustment, the central densities of

**Figure 4.** Surface brightness profile in the R band of NGC 3348, a ‘core’ galaxy. The dashed line is the best-fitting Sérsic model; the observed profile (points, and solid line) falls below this inside of a break radius $r_b \approx 0.35''$. (From Graham 2004.)
the ‘core’ galaxies in figure 5 should probably be interpreted as lower limits, since the pre-binary-SBH nuclei might have had central excesses such as those observed in many fainter galaxies.

Our current understanding of nuclear cusps and cores presents an interesting contrast to the earlier view, in which cores were seen as generic (e.g. Tremaine 1997), and the presence of power-law inner profiles was attributed to some secondary process such as adiabatic contraction driven by growth of the SBH (e.g. van der Marel 1999a). Nowadays, power-law profiles are deemed ‘natural’ and special explanations are sought for cores. Preconceptions aside, the existence of a core begs the question: what determines the core radius? A reasonable answer in the case of galactic nuclei, as discussed in section 7, is that the size of the core is determined by the radius at formation of the binary SBH that preceded the current, single SBH. (A similar explanation appears to work for cores of globular clusters; see Merritt et al 2004.) No comparably compelling explanation currently exists either for the inner power-law profiles or the compact nuclei observed in fainter elliptical galaxies, although it is intriguing that the latter appear to be present only in galaxies that lack a SBH.

4. Collisionless equilibria

Collisionless nuclei have relaxation times greater than the age of the universe (section 2), so the distribution of stars around the SBH still reflects to some extent the details of the nuclear formation process, including possibly the effects of binary SBHs (section 7). While little can be said \textit{ab initio} about the expected form of $\rho(r)$, a collisionless steady state must satisfy the
coupled equations
\[ \rho(r) = \int \int \int f(E, I_2, I_3) d^3 v, \]  
\[ \nabla^2 \Phi = 4\pi G \rho(r). \]

Here \( f \) is the number density of stars in phase space; \( \Phi(r) \) is the gravitational potential, which includes contributions both from the SBH and from the stars; \( E = v^2/2 + \Phi(r) \) is the orbital energy per unit mass; and \( I_2 \) and \( I_3 \) are additional isolating integrals of the motion in \( \Phi \), if they exist. For instance, in an axisymmetric nucleus, \( I_2 = J_z \), the angular momentum about the symmetry axis. Jeans’s theorem states that a steady-state \( f \) must be expressible in terms of the isolating integrals of the motion in the potential \( \Phi \), or equivalently that \( f \) must be independent of phase on every invariant torus.

A useful way to think about the self-consistency problem is to view each set of integral values \((E, I_2, I_3)\) as defining a single, time-averaged orbital density; the total density \( \rho(r) \) must then be representable as a superposition of orbits with non-negative weights (Schwarzschild 1979, Vandervoort 1984). Roughly speaking, self-consistency requires that there exist at least as many distinct orbit families as there are dimensions in \( \rho \); for instance, a triaxial mass distribution requires three isolating integrals (Schwarzschild 1981). This condition places almost no restrictions on the form of \( \rho \) in spherical or axisymmetric geometries but can be an important constraint in non-axisymmetric nuclei (see Merritt 1999 for a review).

4.1. Spherical nuclei

The most general form of \( f \) that preserves spherical symmetry is \( f = f(E, L^2) \) with \( L = r \times v \) the angular momentum per unit mass. If the velocity distribution is assumed to be isotropic, \( f = f(E) \), \( f \) is determined uniquely by \( \rho \) and \( \Phi \) via equation (29a). For example, if the stellar density follows a power law near the SBH, \( \rho(r) = \rho_0 r^{-\gamma} \), and assuming that \( r \ll r_h \) so that the contribution to \( \Phi \) from the stars can be ignored, then

\[ f(E) = \frac{3 - \gamma}{8} \sqrt{\frac{2}{\pi^3}} \frac{\Gamma(\gamma + 1)}{\Gamma(\gamma - \frac{3}{2})} \frac{M_* \phi_0^{3/2}}{m_* (GM_*)^3} \left( \frac{|E|}{\phi_0} \right)^{\gamma - 3/2}, \quad \gamma > 1/2. \]

Here \( m_* \) is the stellar mass, \( \phi_0 = GM_* r_h \), and \( r_h \) is the radius containing a mass in stars equal to twice \( M_* \) (section 2).

For \( \gamma \leq 1/2 \), equation (30) states that \( f(E) \) is undefined; the reason is that the low-\( L \) orbits at each \( E \) force the density to increase faster than \( v^{-1/2} \) toward the centre. Achieving a steady state in this case requires a depopulation of the eccentric orbits, i.e. a velocity ellipsoid that is biased toward circular orbits. The amplitude of the anisotropy required can be estimated by repeating the derivation above after setting \( f = f(E, L^2) = KL^{-2\beta} |E|^\beta \), which imposes a constant degree of anisotropy, \( \sigma_\parallel^2/\sigma_\perp^2 = 1 - \beta \); here \( \sigma_\parallel \) and \( \sigma_\perp \) are the velocity dispersions perpendicular and parallel to the radius vector. The result is

\[ f(E, L^2) = \frac{3 - \gamma}{2^{3-\beta}} \sqrt{\frac{2}{\pi^3}} \frac{\Gamma(\gamma + 1 - 2\beta)}{\Gamma(1 - \beta) \Gamma(\gamma - \frac{1}{2} - \beta)} \frac{M_* \phi_0^{3/2}}{m_* (GM_*)^3} \left( \frac{L^2}{L_0^2} \right)^{-\beta} \left( \frac{|E|}{\phi_0} \right)^{\gamma - 3/2 - \beta}, \]

where \( L_0^2 = GM_* r_h \). In this more general case, a non-negative \( f \) implies \( \beta < \gamma - 1/2 \), i.e. \( \sigma_\parallel^2/\sigma_\perp^2 > 3/2 - \gamma \).

This result is relevant to the ‘core’ galaxies (section 3), some of which have essentially flat or even centrally-decreasing densities within \( r_h \) (e.g. Lauer et al. 2002). The velocity distribution of stars near the SBHs in these galaxies must be biased toward circular orbits. The effect has probably been seen in M87 (figure 6; Merritt and Oh 1997,
Figure 6. Velocity dispersions and velocity anisotropy in a spherical model of the centre of M87, a ‘core’ galaxy, under three different assumptions about the mass of the SBH: (a) $M_\bullet = 1.0 \times 10^9 \, M_\odot$, $r_h \approx 0.3''$; (b) $M_\bullet = 2.4 \times 10^9 \, M_\odot$, $r_h \approx 0.8''$; (c) $M_\bullet = 3.8 \times 10^9 \, M_\odot$, $r_h \approx 1.3''$. The individual components $\sigma_r, \sigma_t$ of the velocity dispersion tensor were derived from the observed run of line-of-sight velocity dispersion with radius, under the assumption that the mass-to-light ratio of the stars is constant. Models with $M_\bullet \gtrsim 1 \times 10^9 \, M_\odot$ are characterized by tangential anisotropy, $\sigma_t > \sigma_r$ at $r \lesssim r_h$, a consequence of the very flat stellar density profile and the presence of a central mass. The best current estimate of $M_\bullet$ is $3.6 \pm 1 \times 10^9 \, M_\odot$ (Macchetto et al 1997). (From Merritt and Oh 1997.)

An example of a formation mechanism that produces an anisotropic core is ejection of stars on radial orbits by a binary SBH (section 7).

Even in spherical nuclei without cores, Jeans’s theorem permits large anisotropies, both toward circular orbits ($\sigma_\theta > \sigma_r$) and radial orbits ($\sigma_r > \sigma_t$) (Merritt 1985, Dejonghe 1989). Circularly-biased velocity distributions are generically stable on dynamical time...
scales (Barnes et al. 1986, Dejonghe and Merritt 1988); very radial velocity distributions, like that of a proposed, SBH-free model for the nucleus of M87 (Binney and Mamon 1982, Newton and Binney 1984), are unstable to non-spherical modes that convert the nucleus into a triaxial spheroid (Merritt 1987). However, it is hard to see how strong radial anisotropies would develop in the first place. A moderate radial anisotropy has been claimed for stars within \( \sim 0.1 \) pc of the Milky Way SBH based on proper motion measurements (Schödel et al. 2003).

Figure 6 illustrates an important point about anisotropic models for galactic nuclei. In general, one expects a degeneracy in the inferred gravitational potential as a result of the non-uniqueness of \( f \). For any choice of \( \Phi(r) \), there are many 2D functions \( f(E, L_z^2) \) that can precisely reproduce the 1D function \( \rho(r) \). The same turns out to be true if additional moments of the velocity distribution function are measured, e.g. the line-of-sight velocity dispersion: there will generally exist a range of functions \( \Phi(r) \) such that a non-negative \( f(E, L_z^2) \) can be found which reproduces a finite set of observed moments exactly. In practice, the range of allowed \( \Phi \)'s is often extremely wide (e.g. Dejonghe and Merritt 1992, Merritt 1993). This means that one cannot expect to infer a unique value for \( M_\bullet \), the mass of the SBH, in spherical nuclei unless substantially more data are available than the low-order velocity moments; for instance, each of the three models illustrated in figure 6 makes identical predictions about the observed velocity dispersion profile. A similar indeterminacy afflicts axisymmetric models, as discussed below.

Anisotropy in the stellar velocity distribution at \( r < r_h \) has consequences for the rate of interaction of stars with the SBH (section 7). It also complicates inferences, based on the observed kinematics, about the mass of the SBH, as discussed in more detail below.

### 4.2. Axisymmetric nuclei

Axisymmetric nuclei have a density \( \rho = \rho(\sigma, z) \) where \( z \) is parallel to the symmetry axis and \( \sigma^2 = x^2 + y^2 \). The axisymmetric analogue of an isotropic spherical model is a so-called ‘two-integral’ model, \( f = f(E, L_z) \) with \( L_z \) the component of the angular momentum about the symmetry axis. In two-integral models, the functional form of \( f \) is determined uniquely once \( \rho(\sigma, z) \) and \( \Phi(\sigma, z) \) are specified; the only freedom that remains is the choice of which sign to attach to \( L_z \) for each orbit, i.e. the degree of streaming about the symmetry axis (Lynden-Bell 1962, Dejonghe 1986, Hunter and Qian 1993). Since \( f \) depends on \( \sigma_\phi \) and \( u_z \) through the symmetric combination \( v^2_\sigma + v^2_z \), the velocity distribution at every position is forced to be the same with respect to \( \sigma \) and \( z \), e.g. \( \sigma_\phi = \sigma_z \) (sometimes called ‘isotropy’). As the flattening of the mass distribution increases, the tensor virial theorem demands that \( \sigma \) decrease, and so \( \sigma_\phi \) falls as well, i.e. the velocity distribution becomes more biased toward circular motions; in the disc limit, two-integral models contain only circular orbits.

Two-integral models are consistent with virtually every oblate-spheroidal mass distribution, i.e. the \( f \) inferred from \( \rho \) and \( \Phi \) is non-negative at every \( (E, L_z) \). However prolate mass models tend to require negative \( f \)'s unless they are nearly spherical (Batsleer and Dejonghe 1993, Dejonghe 1986) or have unrealistic isodensity contours (Jiang and Moss 2002), implying that prolate or barlike nuclei are dependent on a third integral.

The ease with which two-integral models can be constructed has made them popular for modelling the central parts of galaxies (e.g. Binney et al. 1990, Kuijken 1995, Verolme et al. 2002), and the same approach is sometimes used when estimating the mass of the SBH from kinematical data (e.g. van der Marel et al. 1994, Dehnen 1995, Magorrian et al. 1998). However, it is dangerous to base estimates of \( M_\bullet \) on two-integral modelling, since one has no freedom to adjust \( f \) once the galaxy mass model has been specified; the kinematical data are not used at all aside from determining the fraction of orbits that rotate in the two
directions about the symmetry axis. The best-fit $M_\bullet$ is determined by the choice mandated for $f$ (van der Marel 1999b). Indeed most of the putative SBH detections based on two-integral models (e.g. Magorrian et al. 1998) are now believed to be spurious since they were based on data that failed to resolve the SBH’s sphere of influence, sometimes by as much as two orders of magnitude (Merritt and Ferrarese 2001b).

Numerical integrations reveal that the majority of orbits in realistic axisymmetric potentials are regular, i.e. they respect a third isolating integral $I_3$ (e.g. Richstone 1982, Evans 1994). The main family of regular orbits are the tubes, which fill a torus-shaped region having the same symmetries as the potential. Varying $I_3$ at fixed $E$ and $L_z$ is sufficiently well constrained to varying the extent of the orbit in the $z$ direction. Since the self-consistency problem ($29a$, $29b$) generally has a solution for the restricted form $f = f(E, L_z)$, allowing the additional freedom of a third integral typically results in a large degeneracy of solutions, just as in the spherical geometry with $f = f(E, L)$. This freedom is exploited by modellers to construct self-consistent models which reproduce not only $\rho(\sigma, z)$, but also whatever additional kinematical information is available, e.g. the rotation curve near SBH (van der Marel et al., 1998, Bower et al. 2001, Gebhardt et al. 2003, Verolme et al. 2002, Valluri et al. 2005).

Originally it was hoped that kinematical information could eliminate the degeneracy inherent in three-integral modelling and permit well-constrained estimates of $M_\bullet$ in nearby galaxies. Instead, it was found that in many galaxies, variations in the assumed value of $M_\bullet$ could be compensated for by variations in $f(E, L_z, I_3)$ without changing the predicted line-of-sight kinematics at all; indeed models with $M_\bullet = 0$ often provide fits to the nuclear kinematics that are of precisely the same quality as models with a putative ‘best-fit’ value of $M_\bullet$ (e.g. in NGC 4342, Cretton and van den Bosch 1999; in NGC 3379, Gebhardt et al. 2000). This degeneracy is similar to that which characterizes the spherical geometry (figure 6): the extra dimensionality in the data (1D $\to$ 2D) is matched (crudely speaking) by the extra freedom in $f(E, L_z^2) \to f(E, L_z, I_3)$. The degeneracy is illustrated in figure 7 for the galaxy M32, one of the best observed and best resolved of the SBH candidate galaxies; each of the models whose properties are illustrated there is an equally good fit to the kinematical data even though the assumed value of $M_\bullet$ varies from 1.4 to $4.8 \times 10^6 M_\odot$ (Valluri et al. 2004).

While M32 is currently the only galaxy for which such comprehensive modelling has been carried out, it is likely that $M_\bullet$, as derived from stellar kinematics in other galaxies, is comparably degenerate, if not more so, since many of these galaxies were observed at lower effective resolutions than M32. Indeed a conservative, but justifiable, view is that no best-fit value of $M_\bullet$ has been derived from stellar kinematics in any galaxy aside from the Milky Way. Even in the case of the Milky Way, it is instructive to recall that until about 2004, stellar-kinematical estimates of $M_\bullet$ varied by a factor of two, from $\sim 1.8 \times 10^6 M_\odot$ (Chakrabarty and Saha 2001) to $\sim 3.3 \times 10^6 M_\odot$ (Genzel et al. 2000), in spite of the availability of velocity data that were resolved into a distance of $\sim 10^{-3} r_h$ from the SBH. It was only after the orbits of individual stars, some with pericentre distances smaller than $10^{-4} r_h$, had been traced that the degeneracy was removed; the best current determination is $M_\bullet = 3.7 \pm 0.2 \times 10^6 M_\odot$ (Ghez et al. 2005). No external galaxy has stellar kinematical data of quality remotely comparable with that of the Galactic centre, even ca 2000, and so a factor of four degeneracy in $M_\bullet$ in a galaxy like M32 (the nearest external galaxy, which is resolved on a scale of $\sim 0.1 r_h$) is not surprising.

It is common practice to ‘remove’ the degeneracy in spherical or axisymmetric modelling by imposing ad hoc regularization on the solutions, e.g. ‘maximum entropy’

\footnote{In fact the degeneracy in $M_\bullet$ was even greater, since these masses were all derived from models that imposed isotropy on the stellar motions.}
Figure 7. Internal kinematics of six, three-integral axisymmetric models of the galaxy M32. Each panel shows the three velocity dispersions $\sigma_r, \sigma_\phi, \sigma_\theta$ versus distance along the major axis in seconds of arc ($1'' \approx 3$ pc). The assumed value of $M_\bullet$ increases from $1.4 \times 10^6 M_\odot$ (Model A; $r_h \approx 0.4''$) to $4.8 \times 10^6 M_\odot$ (Model F; $r_h \approx 1''$). Each of these models provides an equally good fit to the HST/FOS and ground-based kinematical data for M32. Models with small $M_\bullet$ have radially-biased velocity ellipsoids ($\sigma_r > \sigma_\phi, \sigma_\theta$) at $r \lesssim r_h$, while models with large $M_\bullet$ have tangentially-biased velocities; in other words, the indeterminacy in $M_\bullet$ implies an indeterminacy also in the nuclear kinematics (Valluri et al 2004).

The indeterminacy in stellar-dynamical estimates of $M_\bullet$ is currently the biggest impediment to refining our understanding of SBH demographics, and it also severely limits the inferences that can be drawn about the stellar kinematics of nuclei (e.g. figures 6 and 7). In collisional nuclei like that of M32, the indeterminacy might be reduced by requiring the stellar distribution function to be collisionally relaxed in the sense defined below (section 5). However the majority of galaxies with claimed SBH detections have collisionless nuclei, and the indeterminacy in stellar-dynamical estimates of $M_\bullet$ in these galaxies will probably persist for the foreseeable future.

4.3. Non-axisymmetric nuclei

The degeneracy inherent in axisymmetric models becomes even more marked in triaxial models, and this fact, plus the complexity of dealing with an additional degree of freedom, has kept most galaxy modellers from venturing beyond the axisymmetric paradigm. But...
there are compelling reasons for considering non-axisymmetric models of galactic nuclei. Imaging of the centres of galaxies reveals a wealth of features in the stellar distribution on scales $r \approx r_h$ that are not consistent with axisymmetry, including bars, bars-within-bars and nuclear spirals (e.g. Shaw et al. 1995, Peng et al. 2002, Erwin and Sparke 2002). Even if these features are transient, they may be present for a significant part of a galaxy’s lifetime. Since orbits in non-axisymmetric potentials do not conserve any component of the angular momentum, ‘centrophilic’ orbits like the boxes (Schwarzschild 1979) are allowed, which can pass arbitrarily close to the centre after a finite time. This can imply chaos in the motion (Gerhard and Binney 1985), as well as much greater rates of interaction of stars with the SBH (Norman and Silk 1983).

Triaxiality is usually defined via the index $T$ where

$$T \equiv \frac{a^2 - b^2}{a^2 - c^2} \quad (32)$$

and $(a, b, c)$ are the scale lengths of the long $(x)$, intermediate $(y)$, and short $(z)$ axes respectively. Oblate spheroids have $T = 0$, prolate spheroids have $T = 1$, and $T = 1/2$ is the ‘maximally triaxial’ case.

In a triaxial nucleus containing a SBH, the character of the orbits depends on the distance from the centre. (1) In the region near the SBH, $r \lesssim r_h$, the motion is nearly Keplerian and the forces from the stars constitute a perturbation, causing orbits to precess typically without destroying their integrability (Sridhar and Touma 1999, Sambhus and Sridhar 2000). Nearly-circular orbits are converted into tubes similar to the tube orbits in axisymmetric potentials. Nearly-radial orbits are converted into ‘pyramids,’ Keplerian ellipses with one focus lying near the SBH and which precess in $x$ and $y$ (figure 8; Poon and Merritt 2001). A symmetric pair of pyramid orbits oriented above and below the $x − y$ plane looks similar to a classical box orbit (Schwarzschild 1979), although with the opposite orientation, i.e. along the short axis of the figure, making it less useful for reconstructing the stellar density. A number of other, minor orbit families associated with resonances can be identified (figure 8). (2) At intermediate radii, the SBH acts as a scattering centre, rendering almost all the boxlike orbits chaotic (Gerhard and Binney 1985). This ‘zone of chaos’ extends outward from a few times $r_h$ to a radius where the enclosed stellar mass is roughly $10^2$ times the mass of the SBH (Valluri and Merritt 1998, Papaphilippou and Laskar 1998, Poon and Merritt 2001, Kalapotharakos et al. 2004, Kalapotharakos and Voglis 2005). Integrable tube orbits continue to exist at these energies, as do resonant families like the 2 : 1 ‘banana’ orbits that avoid the centre (Miralda-Escude and Schwarzschild 1989). (3) At still larger energies, the phase space is a complex mixture of chaotic and regular orbits, including resonant boxlike orbits that remain stable by avoiding the centre (Merritt and Valluri 1999).

Little is apparently known about the influence of figure rotation on the structure of orbits in triaxial black-hole nuclei. One study (Valluri 1999) found that figure rotation tends to increase the degree of orbital chaos, apparently because the Coriolis forces broaden orbits that would otherwise be thin, driving them into the destabilizing centre.

The importance of these various orbit families is a function of how useful they are for solving the triaxial self-consistency problem. Following Schwarzschild’s 1979, 1993 pioneering work, it was generally assumed that triaxial nuclei would need to be supported by the regular (non-chaotic) orbit families like the pyramids and bananas (Kuijken 1993, Syer and Zhao 1998, Zhao et al. 1999, Jalali and de Zeeuw 2002). The first full self-consistency studies of triaxial nuclei (Poon and Merritt 2002, Poon and Merritt 2004) revealed that this was only partly correct. When only regular orbits were included in the orbital libraries, solutions were found in the nearly oblate and ‘maximally’ triaxial geometries; the dominant
orbits were the tubes circulating around the short axis, and the pyramid orbits. These models had power-law ($\rho \propto r^{-1}, r^{-2}$) radial density profiles and extended outwards to $\sim 5r_h$. But when chaotic orbits were also included, at least 40%, and as much as 75%, of the mass was found to be assigned to these orbits. $N$-body integrations of these self-consistent models confirmed their stability, at least for several crossing times. Highly prolate models could not be constructed, with or without chaotic orbits.

This work showed for the first time that chaotic orbits could be major components of galactic nuclei. In retrospect, this need not have been surprising. Chaotic orbits fill a volume defined by an equipotential surface. Far from the centre of a galaxy, the equipotential surfaces are nearly spherical, and so chaotic orbits at large energies are not very useful for reconstructing
an elongated triaxial figure. Near the centre of a galaxy with a power-law density profile, however, equipotentials are only slightly rounder than equidensities (figure 9). Add to this the fact that many of the regular orbit families in triaxial nuclei tend to be oriented counter to the figure (Poon and Merritt 2001), and it follows that self-consistent solutions (if they exist) will draw heavily from the chaotic orbits.

Jean's theorem in its usual form (e.g. Binney and Tremaine 1987) equation (29a); seems to exclude chaotic orbits. However, a little thought confirms that chaotic orbits are perfectly acceptable components of steady-state galaxies, as long as they are populated with a uniform phase–space density throughout the accessible part of phase space (Kandrup 1998); indeed chaos is considered almost a requirement for defining a steady state in statistical mechanics (e.g. Sinai 1963). In triaxial black-hole nuclei, the time scale for achieving a uniform population of chaotic phase space—the ‘chaotic mixing’ time—would be very short, of the order of a few crossing times (Merritt and Valluri 1996, Valluri and Merritt 2000, Kandrup and Sideris 2002, Kandrup and Siopis 2003).

Chaotic orbits have well-defined statistical properties. Let $N_E(r_p < d)$ be the number of pericentre passages per unit time for a chaotic orbit of energy $E$, such that the distance from the SBH at pericentre $r_p$ is less than $d$. Numerical integrations show (e.g. figure 10) that

$$N_E(r_p < d) \approx A(E) \times d,$$

(33)

i.e. the number of close encounters with the SBH is an approximately linear function of the encounter distance. The function $A(E)$, in a singular isothermal sphere ($\rho \propto r^{-2}$) nucleus, is roughly

$$A(E) \approx \frac{\sigma^5}{G^2 M_*} e^{-(E - E_h)/\sigma^2}$$

(34)

and in a triaxial nucleus with $\rho \propto r^{-1}$,

$$A(E) \approx \sqrt{\frac{G M_*}{r_h^3}} \left(\frac{E}{E_h}\right)^{-1.4}$$

(35)

with $E_h \equiv \Phi(r_h)$ (Merritt and Poon 2004). These relations, together with estimates of the fraction of chaotic orbits at each energy (e.g. from the self-consistent solutions), can be used to estimate the total rate at which stars are ‘scattered’ by the triaxial potential into the SBH, or into its tidal disruption sphere (section 6). Here we emphasize that this mechanism of feeding stars to the SBH is collisionless, i.e. independent of gravitational encounters.

**Figure 9.** Principal-plane cuts through the equipotential (black/thick) and equidensity (red/thin) surfaces of a triaxial nucleus containing a SBH. The density falls off as $r^{-3/2}$ and the density axis ratios are $c/a = 0.5$, $b/a = 0.7$. The plots extend to $\sim 5$ times the SBH’s influence radius $r_h$ in each direction. The equipotential surfaces are only moderately rounder than the equidensity surfaces. Chaotic orbits, which fill equipotential surfaces, can therefore be useful building blocks in the triaxial self-consistency problem.
Figure 10. Left panel: distribution of close approaches of a chaotic orbit to the central SBH in two triaxial nucleus models. Vertical axis shows the cumulative number of approaches having pericentre distances less than $d$; the SBH’s tidal disruption radius is of order $10^{-7}$ in these units. The radial density profile is $\rho \propto r^{-\gamma}$ with $\gamma = 2$ (top panel) and $\gamma = 1$ (bottom panel). Both orbits have energies $\sim E_h \equiv \Phi(r_h)$. The solid lines have unit slope. Right panel: the function $A(E)$ (equation (33)) that describes the cumulative rate of pericentre passages in four triaxial nucleus models. The triaxiality index is $T \equiv (a^2 - b^2)/(a^2 - c^2)$; $T = 0.25$ is a nearly-oblate model and $T = 0.5$ is a ‘maximally triaxial’ model. Points are from integrations of chaotic orbits; lines show fits; arrows indicate $E_h$. (From Merritt and Poon 2004.)

Poon and Merritt’s (2004) study is still the only attempt to solve the self-consistency problem for triaxial, black-hole nuclei on scales $r \lesssim r_h$. There is however a large body of work addressing the large-scale effects of central mass concentrations on triaxial spheroids or bars (e.g. Norman et al. 1985, Hasan et al. 1993, Dubinski 1994, Merritt and Quinlan 1998, Holley-Bockelmann et al. 2002, Kalapotharakos et al. 2004, Kalapotharakos and Voglis 2005, Athanassoula et al. 2005). These studies generally proceed by first constructing an $N$-body model for the bar or triaxial spheroid, then a compact mass is inserted or grown at the centre and the model is integrated forward. Typically the model isophotes evolve toward rounder and/or more axisymmetric shapes on scales $\gtrsim r_h$; the evolution can be very striking when the mass of the central object exceeds $\sim 1\%$ of the total galaxy mass (compared with $\sim 0.1\%$ for real SBHs) and when the figure is elongated. When the mass of the SBH is smaller, $M_\bullet \lesssim 10^{-3}M_{\text{gal}}$, evolution is often still observed but the final shape can still be non-axisymmetric. The sudden evolution for large $M_\bullet$ is likely due to conversion of the box orbits to chaotic orbits and the ensuing chaotic mixing. It is currently unclear whether Nature would select stable triaxial configurations for galactic nuclei like those constructed by Poon and Merritt (2004), or whether the presence of a SBH would mitigate against such equilibria, as it seems to do on larger scales.

Statler et al. (2004) used the measured streaming velocities of stars in NGC 4365 to model its intrinsic shape; they concluded that the galaxy was strongly triaxial, $T \approx 0.45$, and fairly elongated, $c/a \approx 0.6$. This result implies that SBH’s of mass $M_\bullet \approx 10^{-3}M_{\text{gal}}$ do not impose large-scale axisymmetry in galaxies. The velocity data on which this conclusion was based were only resolved on scales $r \gtrsim 10r_h$ however and so do not constrain the triaxiality of the nucleus.

Due to the near-Keplerian nature of the potential at $r < r_h$, orbits like the pyramids are not centred on the SBH. Sridhar and Touma (1999) noted that off-centre orbits can persist even
in nuclei where the SBH itself is offset from the centre of the stellar spheroid; furthermore they identified one family of loop orbits for which the offset was in the same direction as that of the spheroid. Salow and Statler (2001) and Sambhus and Sridhar (2002) used this result to construct self-consistent, planar, lopsided models for the nucleus of M31, and Jacobs and Sellwood (2001) showed via \textit{N}-body simulations that the lopsided models could be relatively long-lived.

4.4. The ‘adiabatic growth’ model

It is tempting to try to derive the distribution of stars around a SBH from first principles (Peebles 1972a). For instance, if the stellar velocity distribution is assumed to be Maxwellian, \( f(v) \propto e^{-v^2/2\sigma^2} \), with constant \( \sigma \), then Jeans’s theorem implies \( f(E) \propto e^{-E} \propto e^{-(v^2/2\Phi(r))/\sigma^2} \) and equation (29a) gives for the stellar density near the SBH

\[
\rho(r) \propto \int_0^{\sqrt{GM_*/r}} e^{-[v^2/2\Phi(r)]/\sigma^2 v^2} dv \propto e^{GM_*/\sigma^2 r}, \; r \ll GM_*/\sigma^2. \tag{36}
\]

This expression implies an exponentially divergent stellar mass within \( r \). It has other unphysical features as well: for instance, the fact that the velocity dispersion \( \sigma \) is constant implies that typical kinetic energies near the SBH are much smaller than binding energies, hence most stars must be near their apocentres.

A slightly more sophisticated approach is to start with an ‘isothermal’ nucleus, \( f \propto e^{-E} \), \textit{without} a central mass, then increase the value of \( M_\bullet \) from zero and ask what happens to \( f \) and \( \rho \). The initial model has a constant-density core; as the black hole grows, stars are pulled in and the density increases. In the limit that the rate of change of \( M_\bullet \) is slow compared with orbital periods, this ‘adiabatic growth’ model yields for the final stellar distribution at \( r \ll r_h \)

\[
f(E) = \text{const.}, \rho(r) \propto \left( \frac{L}{r_h} \right)^{-3/2}. \tag{37}
\]

(Peebles 1972b, Young 1980). This model and variations (non-spherical or rotating nuclei, etc) has been very widely investigated (see Merritt 2004 for a comprehensive review).

At first sight, equation (37) is very promising: as figure 2 shows, the stellar densities near the centres of the Local Group galaxies M32 and the Milky Way increase as \( \rho \sim r^{-3/2} \) at \( r \lesssim r_h \). But there is a consensus that the adiabatic growth model is probably not relevant to the structure of galaxy cores, for the following reasons.

- Both galaxies in figure 2 have collisional nuclei, i.e. the relaxation time at \( r < r_h \) is shorter than the age of the galaxy. The nuclear density profiles in these galaxies must have been strongly modified by energy exchange between stars (section 5).
- The postulates of the adiabatic growth model—that the SBH grew in mass at a fixed location via spherically-symmetric accretion—are extremely unlikely; almost all models for growth of SBHs invoke strong departures from spherical symmetry in order to remove angular momentum from the infalling gas (e.g. Shlosman et al. 1990).
- Density profiles in the Milky Way, M32 and many other galaxies are steep power laws, \( \rho \sim r^{-2} \), at \( r \gtrsim r_h \). If a SBH grows at the centre of a galaxy with a pre-existing power-law cusp, the profile at \( r \lesssim r_h \) is an even steeper power law: an initial profile \( \rho \propto r^{-\gamma_0} \) becomes \( \rho \propto r^{-\gamma}, \; \gamma = 2 + (4 - r_h^{-1})^{-1} > 2 \) (figure 11). Such steep profiles are not observed (although they might be present, but unresolved, in some galaxies).
- The adiabatic growth model cannot explain the flat inner density profiles of the ‘core’ galaxies (section 3) without ad hoc assumptions (e.g. van der Marel 1999b). The existence of the cores is now generally attributed to ejection of stars by a binary SBH following a
Figure 11. Influence of the adiabatic growth of a black hole on its nuclear environment in a spherical, isotropic galaxy. (a) Density profiles after growth of the black hole. Initial profiles were power laws, $\rho_i \propto r^{-\gamma_0}$, with $\gamma_0$ increasing upwards in steps of 0.25. The radial scale is normalized to $r_h$ as defined in the initial galaxy (equation (4)). The slope of the final profile at $r < r_h$ is almost independent of the initial slope. (b) Velocity anisotropies after growth of the black hole. A slight bias toward circular motions appears at $r < r_h$.

Recently there has been a revival of interest in the adiabatic growth model in the context of the dark matter distribution at the centre of the Milky Way and other galaxies (Gondolo and Silk 1999). Some of the objections just raised to the adiabatic growth model for stellar nuclei do not apply to dark matter (Bertone and Merritt 2005).

5. Collisional equilibria

Most galaxies with well-determined SBH masses have central relaxation times much longer than $10^{10}$ yr (figure 1): their nuclei are ‘collisionless.’ In a collisionless nucleus, the distribution of stars near the SBH will reflect the details of the nuclear formation process, and many steady-state configurations are possible, as discussed in section 4. But figure 1 reveals a clear trend of $T_r(r_h)$ with luminosity, such that galactic spheroids fainter than $M_V = -18$ have relaxation times at $r_h$ shorter than $10^{10}$ yr. The Milky Way nucleus has $T_r(r_h) \approx 5 \times 10^{10}$ yr but the steep density profile implies shorter relaxation times at smaller radii: $\approx 6 \times 10^9$ yr at $0.2 r_h$ (0.6 pc) and $\approx 3.5 \times 10^9$ yr at $0.1 r_h$ (0.3 pc) (assuming Solar-mass stars). The nucleus of the Milky Way is therefore ‘collisional’ in the sense defined in section 2. Three other Local Group galaxies, M32, M33 and NGC 205, also have collisional nuclei (Lauer et al. 1998, Hernquist et al. 1991, Valluri et al. 2005) although M32 is the only one of these to exhibit dynamical evidence for a SBH (Merritt et al. 2001, Valluri et al. 2005.)

Beyond the Local Group, essentially all of the galaxies for which the SBH’s influence radius is spatially resolved are ‘core’ galaxies (section 3) with low nuclear densities and long relaxation times. Still, it is reasonable to suppose that collisional nuclei are present in at
least some galaxies with spheroid luminosities below the value at which the cores appear, $M_V = -20$. Furthermore, the nuclei of the ‘core’ galaxies may have been much denser before the cores were created by binary SBHs (section 7).

In a collisional nucleus, an approximately steady-state distribution of stars is set up around the SBH in a time $\sim T_s(r_h)$. In the case of a single stellar mass, the steady-state density is $\rho \propto r^{-7/4}$; if there is a mass spectrum, the heavier stars will concentrate to the centre (‘mass segregation’). ‘Core collapse,’ the runaway increase in density that occurs in isolated stellar systems after $\sim 10^2 T_s$, does not occur in nuclei containing SBHs, because the time required is too long and because the presence of the SBH inhibits the runaway (Marchant and Shapiro 1980).

Continued loss of stars to the SBH implies that no precisely steady-state equilibrium can exist; for instance, the nucleus will slowly expand due to the effective heat input as stars are destroyed (Shapiro 1977). This and related effects are discussed in more detail in section 6.

### 5.1. The Bahcall–Wolf solution

Gravitational encounters drive the local velocity distribution toward a Maxwellian, but a Maxwellian velocity distribution implies an exponentially divergent mass near the SBH (equation 36). The existence of a region close to the hole where stars are captured or destroyed prevents the nucleus from reaching thermal equilibrium. The density must drop to zero on orbits that intersect the SBH’s event horizon at $r = r_S$, or that pass within the tidal disruption sphere at $r_t$; the latter radius is most relevant since galaxies with collisional nuclei probably always have $M_* < 10^8 M_\odot$ (section 2). Ignoring for the moment a possible dependence of the stellar evolution of $f$ can be approximated via the isotropic orbit-averaged Fokker–Planck equation,  

$$4\pi^2 p(E) \frac{\partial f}{\partial t} = -\frac{\partial F_E}{\partial E} \tag{38a}$$  

$$F_E(E, t) = -D_{EE} \frac{\partial f}{\partial E} - D_E f \tag{38b}$$

(Spitzer 1987). Here $p(E)$ is a phase–space volume element, $p(E) = 2^{-3/2} \pi G^3 M_*^2 |E|^{-5/2}$ near the SBH, $F_E$ is the flux of stars in energy space, and $D_E$ and $D_{EE}$ are diffusion coefficients that describe the effects of small-angle scattering:

$$D_{EE}(E) = 64\pi^4 G^2 m^2 \ln \Lambda \left[ q(E) \int_{-\infty}^E \mathrm{d}E' f(E') + \int_{E}^0 \mathrm{d}E' q(E') f(E') \right],$$  

$$D_E(E) = -64\pi^4 G^2 m^2 \ln \Lambda \int_{E}^0 \mathrm{d}E' p(E') f(E') \tag{39}$$

with $\ln \Lambda \approx \ln(M_\odot/m)$ the Coulomb logarithm and $q(E) = (2^{1/2} \pi/6) G^2 M_*^2 |E|^{-3/2}$. The boundary conditions are $f(E_v) = 0$ and $f(0) = f_0$; $E_v = GM_*/r_t$ is the energy at which stars are lost to the SBH and $f_0$ is the phase space density at $E = 0$. (The inner boundary condition is only approximate since the condition for a star to pass inside $r_t$ is angular-momentum dependent, as discussed in more detail below. The outer boundary condition is also approximate since a non-zero $f$ at large radii implies a population of stars whose contribution to the potential has been ignored.)

Bahcall and Wolf (1976) first presented numerical solutions to equations (38a) and (38b). They found that a steady state is reached after roughly one relaxation time at $r_h$. If $|E_i| \gg GM_*/r_h$, i.e. if the disruption radius $r_t$ is much smaller than $r_h$ (which is the case in real nuclei), the steady-state solution is close to a power law,  

$$f(E) = f_0 |E|^{1/4}, \quad \rho(r) = \rho_0 r^{-7/4}, \quad |E| \ll |E_i|, r_t \ll r. \tag{40}$$
Figure 12. Evolution of the stellar distribution around a SBH due to energy exchange between stars. These curves were computed from the isotropic, orbit-averaged Fokker–Planck equation (equations (38a), (38b)) with boundary condition $f = 0$ at $|E| = 6$. Left panel: phase-space density $f$; right panel: configuration-space density $\rho$. The initial distribution (shown in bold) had $\rho \propto r^{-7/4}$ near the SBH; thin curves show $f$ and $\rho$ at times of $(0.2, 0.4, 0.6, 1.0)$ in units of the relaxation time at the SBH's initial influence radius $r_h$. Dashed lines show the ‘zero-flux’ solution $f \propto |E|^{1/4}, \rho \propto r^{-7/4}$. The steady-state density is well approximated by the zero-flux solution at $r \lesssim 0.2r_h$.

Equation (40) is a ‘zero-flux’ solution, i.e. it implies $F_E = 0$. (An ‘isothermal’ distribution, $f \sim e^{E/\sigma_0^2}$, also implies zero flux but is unphysical for the reasons discussed above.) Figure 12 illustrates the evolution of $f(E,t)$ in the case of a cluster that extends beyond the SBH’s influence radius. The Bahcall–Wolf cusp rises above the pre-existing density inside a radius $\sim 0.2r_h$.

In the numerical solutions, the steady-state flux is found to be small but non-zero, of order

$$F(E) \approx \frac{n(r_t)r_t^3}{T_r(r_t)} \propto r_t.$$  

In other words, the flux is determined by the rate at which stars can diffuse into the disruption sphere at $r_t$. This flux is ‘small’ in the sense that the one-way flux of stars in or out through a surface at $r_t \ll r \lesssim r_h$ is much greater; except near $r_t$, the inward and outward fluxes almost cancel. Thus, the steady-state flux is limited by the ‘bottleneck’ at $r = r_t$. As $r_t$ is reduced, the flux approaches zero and the numerical solution approaches the power-law form of equation (40).

At the centre of the Milky Way, the flux implied by the Bahcall–Wolf solution would only be of order $\sim 10^{-12}$ stars yr$^{-1}$. Frank and Rees (1976) pointed out that the actual loss rate to a black hole would be dominated by changes in angular momentum, not energy, implying a much higher flux. (This is discussed in more detail in section 6.) Bahcall and Wolf (1977) included these loss-cone effects heuristically, by finding steady-state solutions of the modified equation 4

$$4\pi^2 p(E) \frac{\partial f}{\partial t} = -\frac{\partial F_E}{\partial E} - \rho_h(E, t)$$  

where the term $\rho_h$ is an approximate representation of the true loss rate into the SBH (Lightman and Shapiro 1977). They found that the addition of the loss term had only a small effect on the steady-state form of $f(E)$ and $\rho(r)$ even though it substantially increased the implied loss rate.

Solutions of the full, anisotropic Fokker–Planck equation, including a careful treatment of loss-cone dynamics as well as physical collisions between stars, were first presented by
Dynamics of galaxy cores 2541

Figure 13. Evolution of the mass in stars within a distance $0.1 r_h$ from a massive body (‘black hole’) in a set of $N$-body simulations with different initial conditions and different particle numbers. A chain-regularization algorithm was used to handle close encounters between stars and the black hole. Smooth curves are solutions to the Fokker–Planck equation (38a), (38b). (From Preto et al. 2004.)

Cohn and Kulsrud (1978). (The parameters in this study were chosen to mimic the distribution of stars around a black hole in a globular cluster.) The logarithmic derivative of the steady-state density was found to be $d \ln \rho / d \ln r \approx -1.65$ for $r$ in the range $10^{-3} < r / r_h < 10^{-1}$, compared with the Bahcall–Wolf value of $-1.75$. The velocity anisotropy was found to be close to zero for $r \gtrsim 10^{-3} r_h$. The Bahcall–Wolf solution has been verified in a number of other studies based on fluid (Amaro-Seoane et al. 2004) or Monte-Carlo (Marchant and Shapiro 1980, Duncan and Shapiro 1983, Freitag and Benz 2002) approximations to the Fokker–Planck equation.

Most recently, advances in computer hardware (e.g. Makino et al. 2003) and software (e.g. Mikkola and Aarseth 1990, Mikkola and Aarseth 1993) have made it possible to test the Bahcall–Wolf solution via direct $N$-body integrations, avoiding the approximations of the Fokker–Planck formalism (Preto et al. 2004, Baumgardt et al. 2004a, Merritt and Szell 2005). Figure 13 shows a set of $N$-body simulations of collisional cusp growth compared with the predictions of the isotropic Fokker–Planck equation. In these simulations, there was no loss of stars and so a true ‘zero-flux’ equilibrium was established.

The Milky Way is probably the only galaxy in which a Bahcall–Wolf cusp could currently be detected: its central relaxation time is shorter than $10^{10}$ yr and its nuclear density profile
is resolved on scales $\ll r_h$. Figure 2 suggests that the bright stars have $\rho \sim r^{-1.5}$ at $10^{-3} \lesssim r/r_h \lesssim 10^{-1}$. This is slightly shallower than the Bahcall–Wolf ‘zero-flux’ prediction but probably consistent given the uncertainties in the observed profile (Schödel et al 2006). The presence of a mass spectrum also implies a smaller slope ($\S$ 5.2). In addition, the time required to reach a steady state at the Galactic centre may be $\gtrsim 10^{10}$ yr (Merritt and Szell 2005).

5.2. Multi-mass equilibria

Galaxies contain stars with a range of masses. The initial mass function, i.e. the distribution of masses at the time of formation, is believed to be roughly a power-law, $n(m) \propto m^{-\alpha}$, $\alpha \approx 2$ (Salpeter 1955, Miller and Scalo 1979), but the mass function changes with time as stars lose mass and as new stars are formed. Mass functions near the centres of galaxies are difficult to constrain observationally; this is true even at the Galactic centre due to crowding and obscuration. Evolutionary models assuming a constant rate of star formation (e.g. Alexander 2005) suggest that $\sim 75\%$ of the mass after 10 Gyr would be in the form of ‘live’ stars with $(m) \approx 0.5 M_\odot$, $\sim 20\%$ in white dwarves with $0.6 M_\odot \lesssim m \lesssim 1.1 M_\odot$ and a few per cent in neutron stars ($m = 1.4 M_\odot$) and black holes ($m \approx 10 M_\odot$). At the centre of the Milky Way there is also known to be a population of more massive stars, $3 M_\odot \lesssim m \lesssim 15 M_\odot$, inside $\sim 10^{-2} r_h$, which probably could not have formed in situ due to tidal stresses from the SBH (Ghez et al 2003, Eisenhauer et al 2005).

Exchange of energy between stars with different masses tends to establish local equipartition of kinetic energy, leading to spatial segregation (Spitzer and Shull 1975): the more massive stars congregate closer to the centre.

The mass-segregation time scale can be estimated by considering two mass groups $m_1$ and $m_2$. Assuming Maxwellian velocity distributions, the mean rate of change of kinetic energy for stars of mass $m_1$ due to encounters with stars of mass $m_2$ is

$$\frac{de_i}{dt} = 8 (6\pi)^{1/2} G^2 m_1 \rho_1 \Lambda (v_{1,\text{rms}}^2 + v_{2,\text{rms}}^2)^{3/2} (\epsilon_2 - \epsilon_1),$$

(43)

where $\epsilon_i = (1/2)(m_i v_{i,\text{rms}}^2)$; the same expression with indices interchanged gives $\frac{de_2}{dt}$ (Spitzer 1987). Suppose that $m_2 \ll m_1$ and assume $v_{1,\text{rms}} \approx v_{2,\text{rms}}$, appropriate shortly after the nucleus forms. Then

$$T_1 \equiv \frac{1}{\epsilon_1} \frac{de_1}{dt} = \frac{0.0814 v_{1,\text{rms}}^4}{G^2 m_1 \rho_1 \Lambda},$$

$$T_2 \equiv \frac{1}{\epsilon_2} \frac{de_2}{dt} = \frac{\rho_2}{\rho_1} T_1.$$  

(44)

Consider first the case that the heavier stars dominate the total density and have a relaxation time $T_{r,1}$. Then $T_2 \approx T_{r,1}$ and the light stars reach equipartition on the same time scale that the heavy stars establish a collisional steady-state. On the other hand, if the light stars dominate, with relaxation time $T_{r,2}$, then $T_1 \approx (m_2/m_1) T_{r,2}$, and the heavy stars lose energy to the light stars very rapidly compared with $T_{r,2}$. The first case describes a nucleus containing stars and particle dark matter, while the latter describes a nucleus containing stars and a population of massive remnants. In either case, the time for the sub-dominant component to reach equipartition with the dominant component is of order $T_r$ for the dominant component or less.

The single-mass Fokker–Planck equation can be generalized to the multi-mass case by defining $f(E, m, t) dm$ as the number density of stars in phase space with masses in the range $m$ to $m + dm$. Then (e.g. Merritt 1983)

$$4\pi^2 p(E) \frac{\partial f}{\partial t} = - \frac{\partial F_E}{\partial E},$$

$$F_E(E, m, t) = - D_{EE} \frac{\partial f}{\partial E} - m D_E f,$$

(45)
Figure 14. Evolution of a population of low-mass objects due to heating from a dominant, high-mass population and to scattering into the SBH. $f(E, t)$ and $\rho(r, t)$ are the phase–space and configuration-space densities, respectively, of the low-mass population. Times shown are 2, 4, ... 20 in units of time $T_2$ defined in equation (44). The space density evolves from $\rho \sim r^{-2.3}$ to $\sim r^{-1.5}$. (Adapted from Merritt 2004.)

with diffusion coefficients

$$D_{EE}(E) = 64\pi^4 G^2 \ln \Lambda \left[ q(E) \int_{-\infty}^{E} dE'h(E') + \int_{E}^{0} dE'q(E')h(E') \right],$$

$$D_E(E) = -64\pi^4 G^2 \ln \Lambda \int_{E}^{0} dE'p(E')g(E'),$$

and $g$ and $h$ are moments over mass of $f$:

$$g(E, t) = \int_{0}^{\infty} f(E, m, t)m \, dm,$$

$$h(E, t) = \int_{0}^{\infty} f(E, m, t)m^2 \, dm.$$ (47)

Considering again the case of a nucleus containing stars with just two masses, $m_1 \gg m_2$, the evolution equation for the lighter component may be found by taking the first moment over mass of equation (45), restricting the integral to the mass range which characterizes the lighter stars. The result is

$$4\pi^2 \rho(E) \frac{\partial g_2}{\partial t} = \frac{\partial}{\partial E} \left( D_{EE} \frac{\partial g_2}{\partial E} \right),$$ (48a)

$$D_{EE}(E) = 64\pi^4 G^2 \ln \Lambda \left[ q(E) \int_{-\infty}^{E} dE'h_1(E') + \int_{E}^{0} dE'q(E')h_1(E') \right]$$ (48b)

with $g_2$ the phase–space mass density of the lighter stars. (These equations apply also to the more general case of a distribution of light-star masses; $g_2$ is then the total mass density of the lighter population.) The steady-state solution is obtained by setting $\partial g_2/\partial E = 0$, yielding a density profile for the light component $g_2 \propto r^{-3/2}$ (cf equation (30)), independent of the $f$ that describes the heavier stars. Figure 14 illustrates the evolution towards this state. If the heavier objects dominate the density, then their steady-state density is $\rho \propto r^{-7/4}$; in other words, the lighter component is less centrally concentrated than the heavier component. Note however the difference in density slopes is fairly small even in this extreme case.

Bahcall and Wolf (1977) solved the coupled Fokker–Planck equations for systems containing stars of two masses around a SBH. They assumed that the two populations had
similar densities at $r \gtrsim r_h$. The heavier component was found to always attain a steady-state density with the slope close to the single-mass value $-7/4$, while the lighter component was less centrally condensed, with index

$$-\frac{d \ln \rho_2}{d \ln r} \approx \left( \frac{m_2}{4m_1} + \frac{3}{2} \right).$$

(49)

This expression gives the expected results $\rho_2 \propto r^{-7/4}$ when $m_2 = m_1$ and $\rho_2 \propto r^{-3/2}$ when $m_2 \ll m_1$.

Murphy et al. (1991) incorporated realistic mass spectra into their time-dependent solutions of the isotropic Fokker–Planck equation; they also included physical collisions as well as mass loss from stars. They found that the more massive component always attained a power-law slope close to $-7/4$, the least massive component had a slope near $-3/2$ and intermediate mass groups scaled approximately as equation (49). Baumgardt et al. 2004b, 2005 carried out $N$-body simulations of star clusters containing a massive black hole and a range of stellar masses; they observed mass segregation but did not compare their results in detail with the Bahcall–Wolf formula. Baumgardt et al. found that the density profile of the ‘bright stars’ in their simulations (i.e. the particles representing giants) was very flat after a relaxation time, with no hint of a power-law cusp. This puzzling result appears not to have been satisfactorily explained. Most recently, Freitag et al. (2006) presented Monte-Carlo models of the evolution of the Galactic centre star cluster including a Kroupa et al. (1993) mass function. The heaviest mass group in their simulations consisted of the $\sim 10 M_\odot$ stellar black holes; in models designed to mimic the Galactic centre cluster, they found that the black holes segregated to the centre in $\sim 5$ Gyr, after which loss of stars into the SBH drove a general expansion (section 6.5). The stellar black holes were found to dominate the mass density within $\sim 0.2$ pc $\approx 0.1 r_h$ of the centre. The density profile of the black holes was found to be ‘compatible’ (modulo noise) with the Bahcall–Wolf single mass form, $d \log \rho/d \log r = -7/4$, while the main-sequence stars near the centre had $d \log \rho/d \log r \approx 1.3 - 1.4$. Freitag et al.’s simulations were targeted towards nuclei with $SBHs in the mass range $10^3 M_\odot \leq M_s \leq 10^7 M_\odot$, even though there is no firm evidence for SBHs with masses below $\sim 10^6 M_\odot$, since these would be the SBHs of most interest to gravitational-wave physicists. Their results suggest that mass segregation is likely to be of marginal importance in the majority of nuclei with confirmed SBHs due to the long relaxation times (section 2).

5.3. Cusp regeneration

As discussed below (section 7), binary SBHs formed during galaxy mergers are efficient at destroying dense nuclei. Essentially all stellar spheroids are believed to have experienced such events in the past, and evidence for the ‘scouring’ effect of binary SBHs is seen in the flat, central density profiles or ‘mass deficits’ of bright elliptical galaxies (figure 4). An important question is whether the existence of dense cusps at the centres of galaxies such as the Milky Way and M32 implies that no binary SBH was ever present, or whether a collisional cusp could have spontaneously regenerated after being destroyed. Merritt and Szell (2005) used $N$-body integrations to simulate cusp destruction by a binary SBH, then combined the two ‘black holes’ into one and continued the integrations until a Bahcall–Wolf cusp had formed around the single massive particle. Figure 15 shows the results of one such integration for mass ratio $q = 0.5$ and initial density profile $\rho \sim r^{-1.5}$. Cusp regeneration was found to require roughly one relaxation time as measured at the SBH’s influence radius $r_h$; for reasonable mass ratios ($M_2/M_1 \lesssim 0.5$), the binary hardly affects the density at this radius (figure 15). Growth of the cusp is preceded by a stage in which the stellar velocity dispersion evolves towards isotropy.
Figure 15. Regeneration of a Bahcall–Wolf cusp around a black hole after destruction by a massive binary. The dotted line is the initial density profile of an N-body model containing a massive central particle (the 'black hole') of mass 0.01, in units where the total galaxy mass is one. The lower solid line shows the density after infall of a second 'black hole' of mass 0.005 has destroyed the cusp. Upper solid lines show the evolving density after the two black hole particles were combined into one, simulating coalescence. The final time would be roughly $10^{10}$ yr if scaled to the Galactic centre. The dashed line has logarithmic slope of $-7/4$. (Adapted from Merritt and Szell 2005.)

and away from the tangentially-anisotropic state induced by the binary. When scaled to the Galactic centre, these experiments suggest that a dense cusp could have been regenerated in $10^{10}$ yr, although evolution towards the steady-state profile might still be occurring.

As discussed above (section 3), ‘mass deficits,’ or cores, are observed to disappear in galaxies fainter than $M_V \approx -19.5$. This might be due in part to cusp regeneration in these galaxies, although central relaxation times are almost always too long for this explanation to be convincing (figure 1). An alternative explanation is suggested by figure 1: galaxies fainter than $M_V \approx -19.5$ are mostly unresolved on scales of $r_h$, which is also approximately the size of a core created by a binary SBH. The lack of mass deficits in galaxies with $M_V \gtrsim -19.5$ probably just reflects a failure to resolve the cores in these galaxies.

6. Loss-cone dynamics

As discussed in section 5.1, the existence of a region $r \leq r_c$ close to the SBH where stars are captured or destroyed can have a significant influence on the steady-state distribution of stars even at radii $r \gg r_c$, since it precludes the formation of an ‘isothermal’ distribution of velocites. Loss of stars is also important because of its observational consequences: tidally disrupted stars are expected to produce X- and UV radiation with luminosities of $\sim 10^{44}$ erg s$^{-1}$, potentially outshining their host galaxies for a period of days or weeks (Rees 1990, Kobayashi et al 2004, Khokhlov and Melia 1996, Kochanek 1994). A handful of x-ray flaring events have been observed that have the expected signature (Komossa 2002, Komossa et al 2004, Halpern et al...
2004), and the number of detections is crudely consistent with theoretical estimates of the event rate (Donley et al. 2002, Wang and Merritt 2004). Tidal flaring events may dominate the x-ray luminosity function of AGN at $L_X \lesssim 10^{44}$ erg s$^{-1}$ (Milosavljević et al. 2006).

Compact objects (neutron stars or stellar-mass black holes) can remain intact at much smaller distances from the SBH; these objects would emit gravitational waves at potentially observable amplitudes before spiralling in and may dominate the event rate for low-frequency gravitational wave interferometers like LISA (Sigurdsson and Rees 1997, Gair et al. 2004, Hopman and Alexander 2006). (The other major categories of low-frequency gravitational wave events are compact, stellar binaries and binary SBHs; the latter are discussed in section 7.)

A central SBH acts like a sink, removing stars on ‘loss-cone’ orbits, i.e. orbits that intersect the capture or disruption sphere $r = r_t$. In a spherical galaxy, this removal is complete after just one galaxy crossing time, and continued supply of stars to the SBH requires some mechanism for loss-cone repopulation. The most widely discussed mechanism is gravitational encounters, which drive a diffusion in energy ($E$) and angular momentum ($L$). The latter dominates the loss rate (Frank and Rees 1976, Lightman and Shapiro 1977). Roughly speaking, many of the stars within the SBH’s influence radius will be deflected into $r_t$ in one relaxation time, i.e. the loss rate is roughly $(M_*/m_*)T_e(r_t)^{-1}$. In a collisional nucleus (section 2) with $M_* \approx 10^6 M_\odot$, this is $\sim 10^6/(10^{10}$ yr $) \approx 10^{-4}$ yr$^{-1}$.

Classical loss cone theory (Bahcall and Wolf 1976, Lightman and Shapiro 1977, Ipser 1978, Cohn and Kulsrud 1978) was directed towards understanding the observable consequences of massive black holes at the centres of globular clusters. Globular clusters are many relaxation times old, and this assumption was built into the theory, by requiring the stellar phase space density $f$ near the black hole to have reached an approximately steady state under the influence of gravitational encounters. One consequence of the much longer ($\gtrsim 10^{10}$ yr) relaxation times in galactic nuclei is that the stellar density profile near the SBH need not have the collisionally-relaxed, Bahcall–Wolf form (section 5.1). Another is that nuclei can be strongly non-spherical or non-axisymmetric, which permits the existence of centrophilic (box or chaotic) orbits (section 4.3); stars on centrophilic orbits can easily dominate the loss rate. In addition, depending on the details of the nuclear formation process, the stellar phase–space density in a galactic nucleus can depart strongly from its steady-state form near the loss cone, even in a precisely spherical galaxy. For instance, if the current SBH was preceded by a binary SBH, stars on orbits with pericentre distances $\sim a_h \gg r_t$ will have been ejected by the binary, and until these orbits are repopulated (on a time scale $\sim T_e$), the loss rate to the SBH can be much smaller than in the steady state. Other initial conditions, e.g. an ‘adiabatic’ cusp around the SBH (section 4.4), could result in higher loss rates than predicted by the steady-state theory.

6.1. Classical loss-cone theory

Stars are swallowed or tidally disrupted after coming within a certain distance of the hole. The tidal disruption radius is

$$r_t = \left( \frac{\eta^2 M_*}{m_*} \right)^{1/3} r_\ast. \tag{50}$$

Here $m_*$ and $r_\ast$ are the stellar mass and radius, and $\eta$ is a form factor of order unity; $\eta = 0.844$ for an $n = 3$ polytrope. Tidal disruption occurs outside of the hole’s event horizon for a solar-type star when $M_* \gtrsim 10^8 M_\odot$. In what follows, $r_t$ will be used indiscriminately to denote the radius of capture or disruption, and stars will be assumed to vanish instantaneously when $r < r_t$. Galaxies with the highest feeding rates are expected to be those with the smallest SBHs, $M_* \lesssim 10^7 M_\odot$, and in these galaxies $r_t > r_\ast$. 

In a spherical galaxy, an orbit that just grazes the tidal disruption sphere has angular momentum

$$L_{lc}^2 = 2r_c^2 [E - \Phi(r_c)] \approx 2GM_\bullet r_c; \quad (51)$$

the latter expression assumes $|E| \ll GM_\bullet /r_c$, i.e. that stars are on nearly-radial orbits with apocentres much larger than $r_c$.

An upper limit to the consumption rate in a spherical galaxy comes from assuming that stars are instantaneously replaced on their original orbits after being consumed by the hole. In this ‘full loss cone’ model, stars are consumed at a constant rate

$$\dot{N}_{FLC} \approx \int \frac{N_{lc}(E)}{P(E)} dE = \int F(E)^{FLC} dE, \quad (52)$$

where $P(E)$ is the period of a nearly-radial orbit of energy $E$ and $N_{lc}(E)dE$ is the number of stars at energies $E$ to $E + dE$ on orbits with $L \leq L_{lc}$. In a spherical galaxy,

$$N(E, L) dE = \frac{8\pi}{2L_f(E, L)P(E, L)} dE dL \quad (53)$$

(\text{Spitzer 1987}) and

$$\dot{N}_{FLC} \approx 4\pi^2 \int f(E) L_{lc}^2(E) dE \approx 8\pi^2 r_c^2 \int f(E)(E - \Phi_1) dE, \quad (54)$$

where $\Phi_1 \equiv \Phi(r_c)$ and the stellar velocity distribution has been assumed to be isotropic. Using equation (30) for $f(E)$ near the SBH in a power-law ($\rho \propto r^{-\gamma}$) nucleus, this expression can be evaluated:

$$\dot{N}_{FLC} \approx (3 - \gamma) \sqrt{\frac{2}{\pi}} \Gamma(\gamma + 1) \left( \frac{M_\bullet}{m_*} \right) \left( \frac{r_c}{r_h} \right)^{3/2 - \gamma} \left( \frac{GM_\bullet}{r_h^2} \right)^{1/2}. \quad (55)$$

For $\gamma = 3/2$, which is close to the value observed near the Milky Way SBH (figure 2), the time $T_\bullet$ required to consume $N_\bullet = M_\bullet /m_*$ stars is independent of $r_c$:

$$T_\bullet \approx \frac{2\sqrt{2}}{9} \left( \frac{GM_\bullet}{r_h} \right)^{-1/2} r_h \approx 0.3 \frac{r_h}{\sigma}, \quad (56)$$

i.e. the hole consumes its mass in stars in roughly one crossing time at $r_h$.

Early discussions of SBH feeding (e.g. Hills 1975, Oneremy 1976, Shields and Wheeler 1978) sometimes assumed a full loss cone; for instance, in the ‘black tide’ model for quasar fuelling (Young et al. 1977, Young 1977b), gas from tidally disrupted stars radiates as it spirals into the SBH, and the hole shines as a quasar until its mass reaches $\sim 10^8 M_\odot$ at which point stars are swallowed whole and the quasar fades. A number of recent studies (Zhao et al. 2002, Miralda-Escudé and Kollmeier 2005, Holley-Bockelmann et al. 2005, Holley-Bockelmann and Sigurdsson 2006) have revived the full loss cone idea in the context of single or binary SBHs. At least in spherical or axisymmetric galaxies, feeding rates as large as (55) would only persist for single orbital periods, long enough for stars on loss-cone orbits to be consumed; presumably these orbits would have been depleted already during the chaotic events associated with SBH formation. Various justifications have been invoked for assuming that $\dot{N}$ remains of order $N_{FLC}^{FLC}$ for much longer times; perhaps the most plausible is that the nucleus is strongly non-axisymmetric and contains centrophilic orbits (section 6.4).

Under the full-loss-cone assumption, feeding rates are high enough that the mass accumulated in 10 Gyr can be sufficient to reproduce observed SBH masses (Zhao et al. 2002), although presumably only a fraction of the mass liberated when $r_c > r_S$ would find its way into the hole.

A more common assumption is that loss cone orbits were depleted at some early time, and that continued supply of stars to the SBH is limited by the rate at which these orbits can...
be re-populated. The most widely discussed mechanism for orbital repopulation is two-body relaxation. Consider again a spherical galaxy. Star–star encounters induce changes in orbital energy and angular momentum. The former were considered above (section 5) and shown to imply very low consumption rates. Much larger rates are implied by changes in orbital angular momentum (Frank and Rees 1976, Lightman and Shapiro 1977, Young 1977b). In a time \( T_r(E) \), the two-body relaxation time for orbits of energy \( E \), the typical change in \( L^2 \) is of order \( \sim L^2_2(E) \approx G^2 M_2^2/(2E) \), the squared angular momentum of a circular orbit. In a single orbital period \( P \ll T_r \), the rms angular momentum change is roughly

\[
(\delta L)^2 \approx (P/T_r) L^2_2.
\]  

Near the hole, orbital periods are short, and a star will complete many orbits before being deflected in. Far from the hole, on the other hand, orbital periods are long enough that a star can be deflected in and out of the loss cone in a single orbital period. For stars in this outer (‘pinhole’) region, consumption by the hole has almost no effect on the orbital population, and the feeding rate is equal to the full loss cone rate defined above. The energy separating the inner, diffusive region from the outer, full-loss-cone region is \( E_{crit} \), defined as energy such that

\[
\delta L \approx L_{lc}.
\]

A more quantitative definition of \( E_{crit} \) is in terms of the orbit-averaged quantity \( q(E) \):

\[
q(E_{crit}) = 1,
\]

where

\[
q(E) \equiv \frac{1}{R_{lc}(E)} \int \frac{dR}{v_R} \lim_{R \rightarrow 0} \frac{\langle (\Delta R)^2 \rangle}{2R} = \frac{P(E)\mu(E)}{R_{lc}(E)},
\]

here \( R \equiv L^2/L_{lc}(E)^2 \), a dimensionless angular momentum variable, and \( \langle (\Delta R)^2 \rangle \) is defined in the usual way as the sum, over a unit interval of time, of \( (\Delta R)^2 \) due to encounters. As the second expression in equation (59) shows, \( q(E) \) is the ratio of the orbital period to the orbit-averaged time \( (\bar{\mu}/R_{lc})^{-1} \) for angular momentum to change by \( L_{lc} \), with \( \bar{\mu}(E) \) the orbit-averaged diffusion coefficient. In terms of \( q \), the ‘diffusive’ loss cone regime has \( q(E) < 1 \) and the ‘pinhole’ regime has \( q(E) > 1 \).

Typically, the total consumption rate is dominated by stars with \( q(E) \lesssim 1 \), i.e. by stars that diffuse gradually into the loss cone. For \( q \ll 1 \), there is almost no dependence of density on orbital phase since changes in \( L \) take much longer than an orbital period. In this regime, it is appropriate to describe the diffusion in terms of the Fokker–Planck equation (Lightman and Shapiro 1977). Ignoring changes in \( E \), the Fokker–Planck equation describing diffusion in angular momentum is

\[
\frac{\partial N}{\partial t} = \frac{1}{2} \frac{\partial}{\partial R} \left[ \langle (\Delta R)^2 \rangle \frac{\partial N}{\partial R} \right],
\]

where the dependence of \( N \) on \( E \) is understood. Taking the limit \( R \rightarrow 0 \) and averaging over one orbital period, this becomes

\[
\frac{\partial N}{\partial t} = \bar{\mu} \frac{\partial}{\partial R} \left( R \frac{\partial N}{\partial R} \right),
\]

where \( \bar{\mu}(E) \) is the orbit-averaged diffusion coefficient defined above. Equation (61) has the same form as the heat conduction equation in cylindrical coordinates, as can be seen by changing variables to \( j = R^{1/2} \):

\[
\frac{\partial N}{\partial t} = \frac{\bar{\mu}}{4j} \frac{\partial}{\partial j} \left( j \frac{\partial N}{\partial j} \right);
\]

\( j \) plays the role of a radial variable and \( \bar{\mu}/4 \) is the diffusivity. Figure 16 illustrates the geometry. The loss cone is sometimes called a ‘loss cylinder’ by virtue of this analogy with the heat conduction problem.
The steady-state solution to equation (61) is

\[ N(R; E) = \frac{\ln(R/R_{lc})}{\ln(1/R_{lc}) - 1} \tilde{N}(E), \]  

where

\[ \tilde{N}(E) = \int_{R_{lc}}^{1} N(E, R) dR \]  

or, in terms of the phase space density \( f \),

\[ f(R; E) = \frac{\ln(R/R_{lc})}{\ln(1/R_{lc}) - 1} \tilde{f}(E). \]

The phase space density tends logarithmically to zero at the loss cone boundary \( R = R_{lc} \). The implied (number) flux of stars into the loss cone is

\[ F(E) dE = \frac{4\pi^2 P(E) L^2(E) \tilde{\mu}(E) \tilde{f}(E)}{\ln(1/R_{lc}) - 1} dE \approx \frac{\tilde{N}(E) \tilde{\mu}(E)}{\ln(1/R_{lc})} dE \]

where the latter expression assumes \( L_{lc} \ll L \) and \( P(E, R) \approx P(E) \). This is lower than the full-loss-cone feeding rate by a factor \( \sim q / \ln(1/R_{lc}) \approx q / \ln(GM_*/4|E|r_t) \). Equation (66b) implies that a large fraction of the stars in the diffusive regime will migrate into the hole in one relaxation time; the consumption rate depends only logarithmically on the ‘size’ \( r_t \) of the tidal disruption sphere (Frank and Rees 1976, Lightman and Shapiro 1977). In the 1D (energy-dependent) problem, the consumption rate is much smaller, scaling as \( r_t^{-1} \) (section 5).

This analysis breaks down where \( q \gtrsim 1 \), since changes in \( L \) over one orbital period are of order \( L \), and the separation of time scales that is implicit in the orbit-averaged treatment does not apply. Cohn and Kulsrud (1978) developed an approximate scheme to deal with this situation. In the presence of relaxation, some stars will enter and exit the loss cone \( R \ll R_{lc} \) in one orbital period, thus evading disruption. Cohn and Kulsrud (1978) wrote the \( r \)-dependent Fokker–Planck equation for stars near the loss cone and computed how the density changes.
along orbits due to the competing effects of capture at $r < r_i$ and relaxation-driven repopulation at $r > r_i$. They found that the $L$-dependence of the steady-state phase–space density near the loss could be be approximated by replacing $R_{lc}$ in equation (65) by $R_0$, where

$$R_0(E) = R_{lc}(E) \times \begin{cases} \exp(-q), & q(E) > 1 \\ \exp(-0.186q - 0.824\sqrt{q}), & q(E) < 1 \end{cases}$$

(67)

Near the hole, $q \ll 1$ and $R_0 \approx R_{lc}$: relaxation effects are small and the phase–space density falls to zero just at the loss-cone boundary. Far from the hole, relaxation dominates, and $f$ only falls to zero for orbits with $L < L_{lc}$. Cohn and Kulsrud (1978) then computed the flux by assuming that equations (65) and (66b) still applied, even in the ‘pinhole’ regime where the diffusion approximation breaks down; this assumption gives the reasonable result that the flux approaches the full loss cone value far from the hole.

Cohn and Kulsrud (1978) included this ‘boundary layer’ prescription into a fully time-dependent calculation of the evolution of $f(E, R, t)$ for stars around a black hole. They were primarily interested in the case of a $\sim 10^3 M_\odot$ black hole in a globular cluster, and so their solution was required to match on to a constant-density (isothermal) core at large radii. The contribution of the stars to the gravitational potential was ignored. Cohn and Kulsrud (1978) also assumed that the star cluster was many relaxation times old, and so they evolved their time-dependent equations until a steady state was reached. The stellar density profile evolved approximately to the Bahcall–Wolf form, $\rho \sim r^{-7/4}$, near the hole and the feeding rate attained a constant value. This value was substantially larger than had been found by earlier authors who adopted more approximate descriptions of the loss-cone boundary, by factors ranging from $\sim 2$ (Shapiro and Marchant 1978) to $\sim 15$ (Lightman and Shapiro 1977), showing the sensitive dependence of the feeding rate on the assumed form of $f$ near the loss cone.

6.2. Application to galactic nuclei

Loss cone theory as developed in the classic papers of Frank and Rees (1976), Lightman and Shapiro (1977) and Cohn and Kulsrud (1978) can be used to estimate feeding rates for SBHs in galactic nuclei. A number of possible complications arise however:

- Much of the total consumption occurs from orbits that extend beyond $\sim r_h$, hence the contribution of the stars to the gravitational potential cannot be ignored.
- Galactic nuclei are often much less than one relaxation time old (section 2). This means that the stellar density near the SBH need not have the Bahcall–Wolf (1976) steady-state form. In the most luminous galaxies, the nuclear density profile is in fact known to be much flatter than $\rho \sim r^{-7/4}$ at $r \lesssim r_h$ (section 3).
- The dependence of $f$ on $L$ near the loss cone boundary can also be very different from its steady-state form, regardless of the form of $\rho(r)$. For instance, if the current SBH was preceded by a massive binary, almost all low-angular-momentum stars will have been ejected by the binary.
- Galactic nuclei need not be spherical or even axisymmetric. In a triaxial nucleus containing centripelic orbits, the mass in stars on orbits that intersect the SBH’s capture sphere can be enormous, much greater than $M_\bullet$, so that the loss cone is never fully depleted. Even in a spherical nucleus, the velocity distribution can be anisotropic (section 4).
- Galactic nuclei sometimes undergo catastrophic changes, due to galaxy mergers, infall of star clusters or black holes, star formation, etc all of which can substantially affect the feeding rate on both the short and long terms.
Starting with Murphy et al. (1991), a standard approach has been to consider only the first two of these complications when computing feeding rates of SBHs in galactic nuclei, i.e. to allow the density profile around the SBH to have whatever form is implied by the observations (which however often do not resolve $r_h$) and to include the contribution of the stars to the gravitational potential when computing the angular-momentum diffusion coefficient (e.g. Murphy et al. 1991, Syer and Ulmer 1999, Magorrian and Tremaine 1999, Wang and Merritt 2004). The energy dependence of the stellar distribution function is then fixed by $\rho(r)$ and by the assumed (or measured) value of $M_*$; the $L$-dependence of $f$ near the loss cone is taken from the steady-state theory.

Figure 17 shows tidal disruption rates computed in this way for a sample of early-type galaxies (Wang and Merritt 2004)\(^3\). There is a weak net dependence of $\dot{N}$ on $M_*$, in the sense that nuclei with smaller SBHs tend to have higher feeding rates. In the smallest galaxies with well-determined SBH masses, e.g. M32, $\dot{N}$ is predicted to exceed $10^{-4}$ yr\(^{-1}\). A similar rate would be predicted for the Milky Way, which is almost an exact copy of M32 in terms of its

\(^3\) These estimates supersede those in Magorrian and Tremaine (1999) who took most of their SBH masses from Magorrian et al. (1998).
central density profile and SBH mass. It follows that, at the Galactic centre, the most recent tidal disruption event should have occurred just a few thousand years ago.

The galaxies in figure 17 with the highest feeding rates have density profiles $\rho \sim r^{-2}$ at $r \gtrsim r_h$. In the case of the Milky Way bulge, which is the only spheroid in this category near enough to be well resolved on scales $\ll r_h$, we know that $\rho(r)$ flattens from $\sim r^{-2}$ to $\sim r^{-1.5}$ at $r < r_h$ (Genzel et al. 2003). However much of the loss cone flux comes from stars with apocentres $\gtrsim r_h$, and so it is reasonable to approximate $\rho(r)$ as a singular isothermal sphere (SIS), $\rho \propto r^{-2}$, in these low-luminosity galaxies. Figures 18 and 19 show the predicted loss rates in SIS nuclei. The total $\dot{N}$ depends on the two parameters $(M_*/m_*, r_h/r_t)$; adopting (50) for $r_t$, the second one of these parameters can be written

$$\frac{r_h}{r_t} = \frac{2\Theta}{\eta^{2/3}} \left( \frac{M_*/m_*}{} \right)^{2/3} \left( \frac{\sigma}{100 \text{ km s}^{-1}} \right)^{-2} \left( \frac{m_*}{M_\odot} \right) \left( \frac{r_*}{R_\odot} \right)^{-1}$$

with $\Theta = Gm_*/2\sigma^2 r_*$ the Safronov number; $\eta$ in (50) has been set to 0.844. In figure 19, two different assumptions have been made about the relation between $M_*$ and $\sigma$: constant $\sigma$, and the $M_* - \sigma$ relation. The stellar mass and radius were set equal to the Solar values. In both cases, the results are well fit by the simple expression

$$\dot{N} \approx 7.1 \times 10^{-4} \text{ yr}^{-1} \left( \frac{\sigma}{70 \text{ km s}^{-1}} \right)^{7/2} \left( \frac{M_*}{10^6 M_\odot} \right)^{-1}.$$
The normalization constant in equation (69) was chosen to reproduce $\dot{N}$ exactly for $\sigma = 70 \text{ km s}^{-1}$, $M_\bullet = 10^6 M_\odot$; the consumption rate scales as $m^{-1/3} \Phi_{1/4}$ for non-Solar stars. Equation (69), combined with the $M_\bullet - \sigma$ relation, implies $\dot{N} \sim M_\bullet^{0.25}$. Equation (69) is consistent with the rates computed directly from the observed luminosity profiles in the fainter galaxies (figure 17).

Tidal disruption rates as high as $\sim 10^{-4} \text{ yr}^{-1}$ in nuclei with $M_\bullet \approx 10^6 M_\odot$ imply a liberated mass of $\sim M_\bullet$ after 10 Gyr. This is not necessarily a problem since only a fraction of the gas removed from stars is expected to find its way into the hole (Rees 1990). Nevertheless, the high values of $\dot{N}$ in low-luminosity galaxies suggest that matter tidally liberated from stars might contribute substantially to SBH masses in these galaxies.

The faintest systems in which there is solid kinematical evidence for SBHs are M32 and the bulge of the Milky Way ($L \approx 10^9 L_\odot$, $M_\bullet \approx 10^6.5 M_\odot$). However there is strong circumstantial evidence for massive black holes in fainter systems (e.g. Filippenko and Ho 2003) and for intermediate mass black holes (IMBHs) in starburst galaxies and star clusters (e.g. van der Marel 2004). If nuclear black holes are common in the so-called dE (dwarf elliptical) galaxies and in the bulges of late-type spiral galaxies, these systems would dominate the total tidal flaring rate, due both to their large numbers and to their high individual event rates. A simple calculation (Wang and Merritt 2004) suggests that the tidal flaring rate due to dwarf galaxies in the Virgo cluster alone would be of the order of $10^{-1} \text{ yr}^{-1}$. Non-detection of flares after a few years of monitoring would argue against the existence of IMBHs in dwarf galaxies.

6.3. Time-dependent loss cone dynamics

The majority of galaxies with detected SBHs have collisionless nuclei (section 2). In these galaxies, the assumption that the stellar phase space density has reached an approximate steady state under the influence of gravitational encounters breaks down. One consequence is that the stellar density near the SBH need not have the Bahcall–Wolf $r^{-7/4}$ form. As discussed above, it is straightforward to modify the classical loss-cone treatment for arbitrary $\rho(r)$. But the fact that galactic nuclei are not collisionally relaxed also has implications for the more detailed...
Figure 20. Dependence of the loss-cone flux on energy and time in the galaxy NGC 4168 assuming a binary mass ratio of $q = 0.1$. Curves are labelled by their binding energy; the value of the gravitational potential at $r = r_h$ in this galaxy is 1.76. The thick black curve is the total flux, in units of stars per year. (From Merritt and Wang 2005.)

form of the phase space density near the loss cone boundary. For instance, in a nucleus that once contained a binary SBH, stars on orbits such that $L \lesssim L_{\text{bin}} = (2GM_1a_h)^{1/2}$ will have been ejected, where $M_1$ is the binary mass and $a_h \approx G\mu/4\sigma^2$ is the ‘hard’ binary separation (equation (15)). Since $a_h \gg r_h$, there will be a gap in angular momentum space around the single SBH that subsequently forms, corresponding to stars with $L \lesssim L_{\text{bin}}$ that were ejected by the binary. Before the single SBH can begin to consume stars at the steady-state rate computed above, this gap needs to be refilled. At the other extreme, one can imagine a nucleus that formed in such a way that the gradients in $f$ near the loss cone boundary are much greater than their steady state values, implying larger feeding rates.

Equation (61) implies a characteristic time to set up a steady-state distribution in angular momentum near the loss cone:

$$t_L \approx \frac{R}{\mu} \approx \frac{L_c^2}{L_{\text{lc}}} T_r.$$  (70)

Setting $L^2 \approx 2GM_1a_h$—the appropriate value for a loss cone that was emptied by a binary SBH—and $L_c^2 \approx GM_2r_h$, appropriate for stars at a distance $\sim r_h$ from the hole,

$$\frac{t_L}{T_r(r_h)} \approx \frac{a_h}{r_h} \approx \frac{M_2}{M_1}$$  (71)

with $M_2/M_1 \leq 1$ the mass ratio of the binary that created the gap. Since $T_r(r_h)$ can be much greater than $10^{10}$ yr in the bright elliptical galaxies that show evidence of cusp destruction (figure 1), even large mass ratio binaries can open up phase–space gaps that would not be refilled in a galaxy’s lifetime, implying much lower rates of SBH feeding than in the steady-state theory.

Figure 20 illustrates this for a particular giant elliptical galaxy, NGC 4168, assuming a binary mass ratio $M_2/M_1 = 0.1$ (Merritt and Wang 2005). Equation (61) was integrated forward assuming an initially sharp phase–space cutoff at $L = L_{\text{lc}}$. Stars with energies near $\Phi(r_h)$ are the first to be scattered into the hole; the total flux reaches 1%, 10%, 50% and 90% of its steady state value in a time of 4.5, 9.8, 17 and 97 Gyr. Figure 21 shows the results of a
Figure 21. Two characteristic times associated with loss-cone refilling in a sample of elliptical galaxies, assuming that a phase–space gap was created by a binary SBH with mass ratio \( q \). \( t_0 \) is the elapsed time before the first star is scattered into the single, coalesced hole and \( t_{1/2} \) is the time for the loss-cone flux to reach \( 1/2 \) of its steady-state value. The solid lines are the approximate fitting function, equation (72). (From Merritt and Wang 2005.)

A similar calculation for each of the ‘core’ galaxies from figure 18; the time for the total flux to reach \( 1/2 \) of its steady-state value is roughly

\[
\frac{t_{1/2}}{10^{11} \text{ yr}} \approx \frac{q}{(1 + q)^2} \frac{M_\star}{10^8 M_\odot}
\]

with \( q \leq 1 \) the mass ratio of the pre-existing binary. While highly idealized, calculations like these demonstrate how different the feeding rates in collisionless nuclei can be from the predictions of steady-state theory. Hopefully, more progress on this important problem can be expected in the near future.

6.4. Non-axisymmetric nuclei

A qualitatively different kind of SBH feeding can occur in non-axisymmetric (triaxial or barlike) nuclei. Orbits in non-axisymmetric potentials do not conserve any component of the angular momentum and certain orbits, the so-called centrophilic orbits, have ‘filled centres’: they pass arbitrarily close to the potential centre after a sufficiently long time (Norman and Silk 1983, Gerhard and Binney 1985). In the presence of a massive central object like a SBH, centrophilic orbits tend to be unstable (chaotic) (e.g. Merritt and Valluri 1999) and this fact was long taken to imply that triaxiality could not be maintained in galaxies containing SBHs. However as discussed in section 4, recent work suggests that nuclei can remain stably triaxial even when most of their stars are on chaotic orbits. Even if long-lived triaxial equilibria are not possible, the high frequency of barlike distortions observed at the centres of galaxies suggests that transient departures from axisymmetry are common.
The mass associated with stars on centrophilic orbits in triaxial galaxies can easily be \( \gg M_\bullet \), greatly exceeding the mass on loss-cone orbits in the spherical or axisymmetric geometries. A lower limit on the feeding rate in the triaxial geometry comes from ignoring collisional loss cone refilling and simply counting the rate at which stars on centrophilic orbits pass within a distance \( r_t \) from the centre as they move along their orbits. For a single orbit, this rate is \( A(E)r_t \) (section 4.3) where the function \( A(E) \) can be determined by numerical integrations (Merritt and Poon 2004, Holley-Bockelmann and Sigurdsson 2006). If \( N_c(E)dE \) is the number of stars on centrophilic orbits in the energy range \( E \) to \( E + dE \), then the loss rate to the SBH (ignoring orbital depletion) is

\[
\dot{N} \approx r_t \int A(E)N_c(E)dE.
\]

(73)

In a SIS (\( \rho \sim r^{-2} \)) nucleus, if the fraction \( f_c(E) \) of chaotic orbits at energy \( E \) is initially independent of \( E \), the accumulated mass after time \( t \) is

\[
\Delta M \approx 1 \times 10^8 M_\odot f_c \left( \frac{r_t}{r_s} \right)^{1/2} \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{5/2} \left( \frac{M_\bullet}{10^8 M_\odot} \right)^{1/2} \left( \frac{t}{10^{10} \text{ yr}} \right)^{1/2}
\]

(Merritt and Poon 2004); the \( t^{1/2} \) dependence reflects the reduction in the feeding rate as centrophilic orbits are depleted. Even for modest values of \( f_c \) (\( \sim 0.1 \)), this collisionless mechanism can supply stars to the SBH at higher rates than collisional loss-cone repopulation, particularly in galaxies with \( M_\bullet \gtrsim 10^7 M_\odot \) in which relaxation times are very long. Indeed after \( 10^{10} \) yr, equation (74) predicts an accumulated mass

\[
\frac{M_\bullet}{10^8 M_\odot} \approx f_c \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^5.
\]

(75)

This is remarkably similar in form to the \( M_\bullet - \sigma \) relation, and even the normalization is of the right order if \( f_c \approx 1 \).

Hills pointed out already in 1975 that full-loss-cone feeding rates in spherical galaxies could grow \( \sim 10^8 M_\odot \) holes in \( 10^{10} \) yr, and Zhao et al (2002) noted that the accumulated mass would satisfy a relation like (75) between \( M_\bullet \) and \( \sigma \). Although there are many problems with this simple idea—for instance, when \( M_\bullet \lesssim 10^8 M_\odot \), most of the mass liberated from tidally disrupted stars would be lost from the nucleus—the idea that supply rates can plausibly approach the (spherical) full-loss-cone rate in a triaxial nucleus appears to be quite solid. An important next step will be to extend the Poon and Merritt (2004) self-consistency studies of triaxial black-hole nuclei to full galaxy models, using both orbital-superposition and \( N \)-body techniques to ensure that the models are long-lived.

6.5. Black-hole-driven expansion

Steady-state solutions like the ones described above can only be approximate descriptions of nuclei, since the supply of stars in a galaxy is finite, and destruction of stars by the SBH will eventually cause the stellar density to drop. In the equilibrium models of Bahcall and Wolf (1976), Cohn and Kulsrud (1978) and others, this effect is absent since the stellar distribution function is fixed far from the hole, enforcing an inward flux of stars precisely large enough to replace the stars being destroyed or consumed by the hole. In reality, the relaxation time beyond a certain radius would be so long that the encounter-driven flux of stars could not compensate for losses near the hole, forcing the density to drop. Expansion occurs for two reasons. (1) Stars are physically destroyed, reducing their numbers. (2) Disrupted stars are those most tightly bound to the hole, on loss-cone orbits, and to achieve such an orbit a star must have given up energy to other stars. In effect, the black hole acts as a
heat source (Shapiro 1977, Dokuchaev 1989), in much the same way that hard binary stars inject energy into a post-core-collapse globular cluster and cause it to re-expand (Hénon 1961, Hénon 1965).

A very simple model that produces self-similar expansion of a black-hole nucleus can be constructed by simply changing the outer boundary condition in the Bahcall–Wolf (1976) problem (section 5.1) from \( f(0) = f_0 \) to \( f(0) = 0 \). Equations (38a) and (38b) then describe the evolution of a finite cluster of stars as they diffuse to lower energies and are consumed by the hole. One finds that the evolution after \( \sim \)one relaxation time can be described as \( \rho(r,t) = \rho_*(r) \rho^*(r) \), with \( \rho^*(r) \) slightly steeper than the \( \rho \sim r^{-7/4} \) Bahcall–Wolf form; the normalization drops off as \( \rho_* \propto t^{-1} \) at late times (Merritt, unpublished). Figure 22 shows the results of a slightly more realistic calculation in a model designed to mimic M32. After reaching approximately the Bahcall–Wolf steady-state form, the density drops in amplitude with roughly fixed slope for \( r \lesssim r_h \). This example suggests that the nuclei of galaxies like M32 or the Milky Way might have been \( \sim \)a few times denser in the past than they are now, with correspondingly higher rates of stellar tidal disruption and stellar collisions.

Expansion due to a central black hole has been observed in a handful of studies based on fluid (Amaro-Seoane et al. 2004), Monte-Carlo (Shapiro and Marchant 1978, Marchant and Shapiro 1980, Freitag et al. 2006), Fokker–Planck (Murphy et al. 1991) and \( N \)-body (Baumgardt et al. 2004a, 2004b) algorithms. All these studies allowed stars to be lost into or destroyed by the black hole; however most adopted parameters are more suited to
globular clusters than to nuclei, e.g. a constant-density core. Murphy et al. (1991) applied the isotropic, multi-mass Fokker–Planck equation to the evolution of nuclei containing SBHs, including an approximate loss term in the form of equation (42) to model the scattering of low-angular-momentum stars into the hole. Most of their models had what would now be considered unphysically high densities and the evolution was dominated by physical collisions between stars. However in two models with lower densities, they reported observing significant expansion over $10^{10}$ yr; these models had initial central relaxation times of $T_r \lesssim 10^9$ yr when scaled to real galaxies, similar to the relaxation times near the centres of M32 and the Milky Way. The $\rho \sim r^{-7/4}$ form of the density profile near the SBH was observed to be approximately conserved during the expansion. Freitag et al. (2006) carried out Monte-Carlo evolutionary calculations of a suite of models containing a mass spectrum, some of which were designed to mimic the Galactic centre star cluster. After the stellar-mass black holes in their models had segregated to the centre, they observed a strong, roughly self-similar expansion. Baumgardt et al. (2004a) followed core collapse in $N$-body models with and without a massive central particle; ‘tidal destruction’ was modelled by simply removing stars that came within a certain distance of the massive particle. When the ‘black hole’ was present, the cluster expanded almost from the start and in an approximately self-similar way.

These important studies notwithstanding, there is a crucial need for more work on this problem. Establishing the self-similarity of the expansion in the Fokker–Planck or fluid descriptions would be a good start; such studies (e.g. Lynden-Bell and Inagaki 1983, Heggie 1985, Heggie and Stevenson 1988) were an important complement to numerical simulations in understanding the post-core-collapse evolution of star clusters.

6.6. Constraining the consumption rate

Calculations of SBH feeding rates in real galaxies are subject to many uncertainties, particularly in collisionless nuclei, due to the wide range of possible geometries (section 6.4) and initial conditions (section 6.3). Additional uncertainties include the form of the nuclear density profile at $r \ll r_h$ and, of course, the mass of the SBH. Plots like figure 18, which was based on spherical, steady-state loss cone theory, should probably be seen as little more than order-of-magnitude estimates of the true tidal flaring rate. In such an uncertain situation, it makes sense to look for observational evidence of tidal disruptions as a constraint on the theory. The ROSAT all-sky survey detected soft x-ray outbursts from a number of galaxies with no previous history of nuclear activity. Roughly half a dozen of these events had the properties of a tidal disruption flare (Komossa 2002 and references therein), and follow-up optical spectroscopy of the candidate galaxies confirmed that at least two were subsequently inactive (Gezari et al. 2003). The mean event rate inferred from these outbursts is roughly consistent with theoretical predictions (Donley et al. 2002). Some fraction of the x-ray luminosity function of active galaxies (Hasinger et al. 2005) must also be due to stellar tidal disruptions. Convolving equation (69) for the disruption rate with the SBH mass function, and assuming that individual tidal disruption events have an $L_X \propto t^{-5/3}$ time dependence as predicted in ‘fallback’ models (Li et al. 2002), one concludes (Milosavljević et al. 2006) that tidal disruptions can account for the majority of x-ray selected AGN with soft x-ray luminosities below $\sim 10^{43}$–$10^{44}$ erg s$^{-1}$. Nearer home, it might be possible to search for ‘afterglows’ of the most recent tidal disruption event at the Galactic centre, which could plausibly have occurred

---

4 Trenti et al. (2005) observed expansion in $N$-body models of a star cluster containing a massive black hole. However since the number of stars was conserved in their simulations, the expansion must have been due to binary heating, either star–star or star–BH binaries. This example illustrates the subtleties involved in extrapolating from small-$N$ systems to the much larger $N$ of galactic nuclei; stellar binaries are energetically insignificant in nuclei.
as little as \( \sim 10^3 \) yr ago. Possible examples of such signatures include x-ray fluorescence of giant molecular clouds (Sunyaev and Churazov 1998) and changes in the surface properties of irradiated stars (Jimenez et al. 2006).

7. Dynamics of binary black holes

Galaxies are believed to grow through the agglomeration of smaller galaxies and protogalactic fragments. If more than one of the fragments contained an SBH, the two SBHs will form a bound system in the merger product (Begelman et al. 1980). This scenario has received considerable attention since the ultimate coalescence of a binary SBH would generate an observable outburst of gravitational waves (Thorne and Braginskii 1976) and possibly electromagnetic radiation as well (Milosavljević and Phinney 2005). Binary SBHs are also increasingly invoked to explain the properties of active galaxies, including AGN variability (Valt aoja et al. 2000, Xie 2003), the bending and precession of radio jets (Roos et al. 1993, Romero et al. 2000), X- and Z-shaped radio lobes (Merritt and Ekers 2002, Gopal-Krishna et al. 2003) and the correlation of radio loudness with galaxy morphology (Wilson and Colbert 1995, Balmaverde and Capetti 2006).

Komossa et al. (2003) reviews the observational evidence for binary SBHs. A handful of galaxies exhibit two active nuclei with separations as small as \( \sim 1 \) kpc (Komossa 2003, Ballo et al. 2004, Hudson et al. 2006), much greater however than the parsec-scale separations that characterize true binaries. Very recently (Rodriguez et al. 2006), VLBA observations of an elliptical galaxy at \( z = 0.055 \) with two compact central radio sources were used to infer the presence of the first, true binary SBH; the projected separation is only \( \sim 7 \) pc and the inferred total mass is \( \sim 1.5 \times 10^8 M_\odot \).

A binary SBH with \( a \approx a_h \) (equation (15)) contains a fraction \( \sim (M_1 + M_2)/M_{gal} \approx 10^{-3} \) of the total gravitational energy of its host galaxy, and such a large binding energy implies a significant change in the distribution of stars, gas or dark matter at the centre of the galaxy when the binary forms. If stars are the dominant component, formation of the binary results in a low-density core, with a displaced mass of order the mass of the binary. As discussed in sections 3 and 7.3, luminous elliptical galaxies always contain such cores, with masses and sizes that are roughly consistent with predictions of the binary SBH model (Merritt 2006). Other dynamical effects associated with binary SBHs include high-velocity ejection of stars via the gravitational slingshot (Hills 1988), chaos induced in stellar orbits by the time-dependent potential of the binary (Kandrup et al. 2003) and the ‘gravitational rocket’ effect, the kick imparted to a coalescing binary due to anisotropic emission of gravitational waves (Favata et al. 2004, Baker et al. 2006), which can displace a coalesced SBH from its central location and potentially eject it into intergalactic space. These mechanisms all go in the direction of ‘heating’ the nucleus, and collectively, they set the initial conditions for evolution of the nucleus after the two SBHs coalesce.

Binary SBHs may fail to coalesce in some galaxies because their evolution stalls at a separation much greater than required for the efficient emission of gravitational waves (Valtonen 1996). In a galaxy where this occurs, the binary may be present when a third SBH (or a second binary) falls in, resulting in a complicated interaction between the multiple SBHs (Mikkola and Valtonen 1990) and possibly ejection of one or more from the nucleus or even from the galaxy. Such events cannot be too frequent or the tight correlations observed between SBH mass and galaxy properties would be violated. Furthermore the total mass density in SBHs in the local universe is consistent with that inferred from high-redshift AGN (Merritt and Ferrarese 2001a, Yu and Tremaine 2002), implying that only a small fraction of SBHs could have been ejected from galaxies in the intervening period.
In what follows, the focus will be on the interaction of binary and multiple SBHs with (point-mass) stars in galactic nuclei and on the implications for nuclear dynamics. Merritt and Milosavljević (2005) present a more general review of the astrophysics of massive black hole binaries.

7.1. Early evolution of binary SBHs and the generation of mass deficits

In galaxy mergers in the local universe, typical mass ratios are believed to be large, of order 10 : 1 (e.g. Sesana et al. 2004). To a good approximation, the initial approach of the two SBHs can therefore be modelled by assuming that the galaxy hosting the smaller SBH spirals inward under the influence of dynamical friction from the fixed distribution of stars in the larger galaxy. Modelling both galaxies as singular isothermal spheres ($\rho \sim r^{-2}$) and assuming that the smaller galaxy spirals in on a circular orbit, its tidally-truncated mass is $\sim \sigma_s^3 r/2G\sigma_g$, with $\sigma$ and $\sigma_g$ the velocity dispersions of large and small galaxies, respectively (Merritt 1984).

Chandrasekhar's (1943) formula then gives for the orbital decay rate and infall time

$$\frac{dr}{dt} = -0.30 \frac{Gm_2}{\sigma_1^2} \ln \Lambda, \quad t_{\text{infall}} \approx 3.3 \frac{r_0 \sigma}{\sigma_g^3},$$

where $\ln \Lambda$ has been set to 2. Using equation (19) to relate $\sigma$ and $\sigma_g$ to the respective SBH masses $M_1$ and $M_2$, this becomes

$$t_{\text{infall}} \approx 3.3 \frac{r_0}{\sigma} \left( \frac{M_1}{M_2} \right)^{0.62},$$

i.e. $t_{\text{infall}}$ exceeds the crossing time of the larger galaxy by a factor $\sim q^{-0.62}$. Thus for mass ratios $q \gtrsim 10^{-3}$, infall requires less than $\sim 10^9 T_{\text{cr}} \approx 10^{10}$ yr (Merritt 2000). This mass ratio is roughly the ratio between the masses of the largest ($\sim 10^{9.5} M_\odot$) and smallest ($\sim 10^6 M_\odot$) known SBBs and so it is reasonable to assume that galaxy mergers will almost always lead to formation of a binary SBH, i.e. a bound pair, in a time less than 10 Gyr.

Equation (76) begins to break down when the two SBHs approach more closely than $R_{12} \approx r_h$, the influence radius of the larger hole, since the orbital energy of $M_2$ is absorbed by the stars, lowering their density and reducing the frictional force. Figure 23 illustrates this via $N$-body integrations of a $q = 0.1$ binary. In spite of this slowdown, the separation between the two SBHs continues to quickly drop until $a \approx a_h \approx (q/4)r_h$ (equation (15)) at which separation the binary’s binding energy per unit mass is $\sigma^2$ and the binary is ‘hard’—it ejects stars that pass within a distance $\sim a$ with velocities large enough to remove them from the nucleus (Mikkola and Valtonen 1992, Quinlan 1996).

What happens next depends on the density and geometry of the nucleus. In a spherical or axisymmetric galaxy, the mass in stars on orbits that intersect the binary is small, $\lesssim M_{12}$, and the binary rapidly interacts with and ejects these stars. Once this has occurred, no stars remain to interact with the binary and its evolution stalls. In non-axisymmetric (e.g. triaxial) nuclei, on the other hand, the mass in stars on centrophilic orbits can be much larger, allowing the binary to continue shrinking past $a_h$. And in collisional nuclei of any geometry, gravitational scattering of stars can repopulate depleted orbits. These different cases are discussed independently below.

The stalling that is predicted to occur in spherical galaxies can be reproduced via $N$-body simulations if $N$ is large enough to suppress two-body relaxation. One finds

$$\frac{a_{\text{stall}}}{r_h} \approx 0.2 \frac{q}{(1 + q)^2}.$$  (Merritt 2006) with $r_h$ the influence radius of the larger SBH, defined as the radius containing a stellar mass equal to twice $M_{12}$ after infall of the smaller SBH has lowered the nuclear
Dynamics of galaxy cores

Figure 23. Upper panel: thick (black) line shows the $N$-body evolution of a massive binary with $M_2/M_1 = 0.1$: the initial galaxy model had a $\rho \sim r^{-1}$ density cusp. Thin (red) line is the evolution predicted by the Chandrasekhar’s dynamical friction formula assuming a fixed galaxy. Horizontal lines indicate $r_h$ and $a_h$. The inset shows the evolution of the inverse semi-major axis of the binary in this integration and in a second integration with roughly one-half the number of particles; the latter curve lies above the former, i.e. the decay occurs more rapidly for smaller $N$ when $a \lesssim a_h$, due to the higher rate of star–star encounters. In the large-$N$ limit of real galaxies, the binary hardening rate would drop to zero at $a \approx a_h$. Lower panel: evolution of the mass deficit in the same two $N$-body integrations. Lines show least-squares fits to $t \geq 120$. (Adapted from Merritt 2006.)

This relation was established using galaxy models with initial, power-law density profiles, $\rho \sim r^{-\gamma}$, $0.5 \leq \gamma \lesssim 1.5$; the coefficient in equation (78) depends only weakly on $\gamma$. Table 1 gives predicted stalling radii for binary SBHs at the centres of the brightest Virgo cluster galaxies, based on the galaxy structural data of Ferrarese et al. (2006a); all of these galaxies exhibit large, low-density cores which might have been formed by binary SBHs.

If the binary does stall at $a \approx a_{\text{stall}}$, it will have given up an energy

$$\Delta E \approx - \frac{G M_1 M_2}{2r_h} + \frac{G M_1 M_2}{2a_{\text{stall}}}$$  \hspace{1cm} (79)

$$\approx - \frac{1}{2} M_2 \sigma^2 + 2 M_1 \sigma^2$$  \hspace{1cm} (80)

$$\approx 2 M_1 \sigma^2$$  \hspace{1cm} (81)
to the stars in the nucleus, i.e. the energy transferred from the binary to the stars is roughly proportional to the combined mass of the two SBHs. The reason for this counter-intuitive result is the $a_{\text{stall}} \sim M_2$ dependence of the stalling radius (equation (78)): smaller infalling holes form tighter binaries. Detailed $N$-body simulations (Merritt 2006) verify this prediction: the mass deficit (equation (28)) is found to be

$$\frac{M_{\text{def}}}{M_2} \approx 0.70q^{0.2}$$

for nuclei with initial density slopes $1 \lesssim \gamma \lesssim 1.5$. Thus, in galaxies with pre-existing density profiles similar to those currently observed at the centres of the Milky Way and M32, equation (82) implies that mass deficits generated by stalled binaries should lie in the relatively narrow range

$$0.4 \lesssim \frac{M_{\text{def}}}{M_2} \lesssim 0.6, 0.05 \lesssim q \lesssim 0.5.$$  

Observed mass deficits are somewhat larger than this (Merritt and Milosavljević 2002, Ravindranath et al. 2002, Graham 2004, Ferrarese et al 2006a). Typical values are $M_{\text{def}} \approx M_\bullet$ and some galaxies have $M_{\text{def}} / M_\bullet$ as large as $\sim 4$ (table 1, figure 24). These numbers should be interpreted with caution since both $M_{\text{def}}$ and $M_\bullet$ are subject to systematic errors, the former from uncertain $M/L$ corrections, the latter from various difficulties associated with SBH mass estimation (section 4). On the other hand, bright elliptical galaxies like M49 and M87 (the first two galaxies in table 1) have probably undergone numerous mergers and the mass deficit should increase after each merger, even if expressed as a multiple of the final (accumulated) SBH mass. If the stellar mass displaced in a single merger is $\sim 0.5M_2$, then—assuming that the two SBHs always coalesce before the next SBH falls in—the mass deficit following $N$ mergers with $M_2 \ll M_1$ is $\sim 0.5N M_\bullet$. $N$-body simulations verify this prediction (Merritt 2006, figure 25).

---

5 The first $N$-body simulation to follow the destruction of a power-law density cusp by a massive binary was that of Milosavljević and Merritt (2001), however the number of particles used was so small that the evolution was dominated by spurious loss-cone refilling and the estimates of $M_{\text{def}}$ accordingly uncertain. Subsequent studies (Henssengrund et al. 2002, Chatterjee et al. 2003, Makino and Funato 2004) achieved nearly 'empty loss cones' around the binary by using mean-field algorithms and/or larger $N$, but these investigations were all based on galaxy models with pre-existing cores, again making it difficult to draw useful conclusions about the values of $M_{\text{def}}$ to be expected in real galaxies.
Mass deficits in the range $0.5 \lesssim M_{\text{def}}/M_* \lesssim 1.5$ therefore imply $1 \lesssim N \lesssim 3$, consistent with the number of gas-free mergers expected for bright galaxies (Haehnelt and Kauffmann 2002). This correspondence constitutes strong evidence that the cores of bright elliptical galaxies are due to heating by binary SBHs.

Mass deficits as large as those in M49 and M87 ($3 \lesssim M_{\text{def}}/M_* \lesssim 4$; table 1) are harder to explain in this way, although a number of additional mechanisms associated with binary SBHs can ‘heat’ a nucleus and increase $M_{\text{def}}$. (a) If two SBHs fail to efficiently coalesce a binary will be present when a third SBH falls in. This scenario is conducive to smaller values of $M_*$, since one or more of the SBHs could eventually be ejected by the gravitational slingshot (Mikkola and Valtonen 1990); and to larger values of $M_{\text{def}}$, since multiple SBHs are more efficient than a binary at displacing stars (Merritt et al 2004). Multiple-SBH interactions are discussed in more detail in section 7.3. (b) The gravitational-wave rocket effect is believed capable of delivering kicks to a coalescing binary as large as $\sim 150 \text{ km s}^{-1}$ (Favata et al 2004, Blanchet et al 2005, Herrmann et al 2006, Baker et al 2006) and possibly much higher (Redmount and Rees 1989). The stellar density drops impulsively when the SBH is kicked out, and again when its orbit decays via dynamical friction. Mass deficits produced in this way can be as large as $\sim M_*$ (Merritt et al 2004, Boylan-Kolchin et al 2004). (c) Binaries might continue to harden beyond $R_{12} \approx a_h$, as discussed below, although it is not completely clear what the net effect on $M_{\text{def}}$ would be.

7.2. Late evolution of binary SBHs and the ‘final-parsec problem’

The rapid phase of binary evolution discussed above is tractable using $N$-body codes since the mechanisms that extract angular momentum from the binary (dynamical friction from the stars, gravitational-slingshot ejection of stars by the binary) depend essentially on the mass density of the objects (stars) interacting with the binary and not on their individual masses (at least for $m_* \ll M_*$. Evolution of the binary beyond $a \approx a_h$ can be qualitatively different, and in fact $N$-body simulations often find that the hardening rate $(d/a)/(1/a)$ is a decreasing function of $N$, the number of ‘star’ particles in the simulation, once $a$ drops below $\sim a_h$.
Figure 25. Density profiles (upper) and mass deficits (lower) in $N$-body simulations of multi-stage mergers. The initial galaxy (dotted lines, top panels) contained a central point of mass $0.01 M_{\text{gal}}$; subsequent curves show density profiles after repeated infall (followed by coalescence) of a second ‘black hole’ of mass $0.005$ (a), $0.0025$ (b) and $0.001$ (c). In the lower panels, points show $M_{\text{def}}/M_*$ at $t_{\text{stop}}$ and dotted lines show $M_{\text{def}}/M_*=0.5N$, where $M_*$ is the accumulated central mass. (From Merritt 2006.)

(Makino and Funato 2004, Berczik et al 2005, Merritt et al 2005). Figure 26 illustrates the $N$-dependence for equal-mass binaries in spherical galaxy models. A straightforward extrapolation of results like these to real galaxies implies that binary SBHs would stall at separations of order $10^9$ pc (table 1)—the “final-parsec problem”.

Stalling is expected in the large-$N$ limit since, in a fixed, smooth potential, the number of stars on orbits intersecting the binary is limited, and in fact decreases with time as the binary shrinks. Stars can interact with the binary only if their pericenters lie within $\sim R \times a$, where $R$ is of order unity. Let $L_{\text{lc}} = R a \sqrt{2[E - \Phi(R a)]} \approx \sqrt{2GM_{12}Ra}$, the angular momentum of a star with pericentre $Ra$. The binary’s ‘loss cone’ is the region in phase space defined by $L \leq L_{\text{lc}}$. In a spherical galaxy, the mass of stars in the loss cone is

$$M_{\text{lc}}(a) = m_\ast \int dE \int_0^{L_{\text{lc}}} dL N(E, L^2)$$
$$= m_\ast \int dE \int_0^{L_{\text{lc}}^2} dL^2 4\pi^2 f(E, L^2) P(E, L^2)$$
$$\approx 8\pi^2 GM_{12} m_\ast R a \int dE f(E) P(E). \quad (84)$$

Here $P$ is the orbital period, $f$ is the number density of stars in phase space and $N(E, L^2) dE dL$ is the number of stars in the integral-space volume defined by $dE$ and $dL$. In the final line, $f$ is assumed isotropic and $P$ has been approximated by the period of a radial orbit of energy $E$. An upper limit to the mass that is available to interact with the binary is $\sim M_{\text{lc}}(a_0)$, the mass within the loss cone when the binary first becomes hard; this is an upper limit since some stars
Figure 26. Long-term evolution of the binary semi-major axis (a) and hardening rate (b) in a set of high accuracy N-body simulations; the initial galaxy model was a low-central-density Plummer sphere. Units are $G = M_{\text{gal}} = 1$, $E = -1/4$, with $E$ the total energy. (a) Dashed lines are simulations with binary mass $M_1 = M_2 = 0.005$ and solid lines are for $M_1 = M_2 = 0.02$, in units where the total galaxy mass is one. (b) Filled (open) circles are for $M_1 = M_2 = 0.005$ ($0.02$). Crosses indicate the hardening rate predicted by a simple model in which the supply of stars to the binary is limited by the rate at which they can be scattered into the binary’s influence sphere by gravitational encounters. The simulations with largest $(M_1, M_2)$ exhibit the nearly $N^{-1}$ dependence expected in the ‘empty loss cone’ regime that is characteristic of real galaxies. (Adapted from Berczik et al. 2005.)

That are initially within the loss cone will ‘fall out’ as the binary shrinks. Assuming a singular isothermal sphere for the stellar distribution, $\rho \propto r^{-2}$, and taking the lower limit of the energy integral to be $\Phi_1(a_h)$, equation (84) implies

$$M_{\text{lc}}(a_h) \approx 3R\mu.$$  \hfill (85)

The change in $a$ that would result if the binary interacted with this entire mass can be estimated using the mean energy change of a star interacting with a hard binary, $\sim 3G\mu/2a$ (Quinlan 1996). Equating the energy carried away by stars with the change in the binary’s binding energy gives

$$\frac{3}{2} \frac{G\mu}{a} dM \approx \frac{G M_1 M_2}{2} d\left(\frac{1}{a}\right)$$  \hfill (86)

or

$$\ln \left(\frac{a_n}{a}\right) \approx 3 \frac{\Delta M}{M_{12}} \approx \frac{9R\mu}{M_{12}} \approx 9R \frac{q}{(1+q)^2}.$$  \hfill (87)

if $\Delta M$ is equated with $M_{\text{lc}}$. Equation (87) suggests that the mass available to the binary would allow it to shrink by only a modest factor below $a_n$. In reality, the time scale for the binary to shrink is comparable with stellar orbital periods, and some of the stars with pericentre distances of order $a_0$ will only reach the binary after it has shrunk to smaller separations. And in most galaxies the density profile is shallower than $\rho \sim r^{-2}$ implying smaller $M_{\text{lc}}$ and less of a change in $a$.

The final parsec problem is a ‘problem’ because it implies a low rate of SBH coalescence, which is disappointing to physicists hoping to detect the gravitational radiation emitted during the final plunge (Folkner 1998). It is also a ‘problem’ in the sense that many circumstantial lines of evidence suggest that SBH binaries do efficiently coalesce. (a) Only one, reasonably compelling case for a true binary SBH exists (Rodriguez et al. 2006), even though binaries
with the same projected separation (∼7 pc, which is the expected stalling radius for an $M_{12} \approx 10^9 M_\odot$ binary; see table 1) could be easily resolved in many other galaxies by radio interferometry. (b) Jets in the great majority of radio galaxies do not show the wiggles expected if the SBH hosting the accretion disc were orbiting or precessing. (c) Jets from Seyfert galaxies are randomly oriented with respect to the discs in their host galaxies (Ulvestad and Wilson 1984). This is naturally understood if SBH spins were randomized at an earlier epoch by binary coalescences during the merger events that formed the bulges (Merritt 2002, Kendall et al 2003, Saitoh and Wada 2004). (d) If binary SBHs are common, mergers will sometimes bring a third SBH into a nucleus containing an uncoalesced binary, resulting in three-body ejection of one or more holes. But the total mass density of SBHs in the local universe is consistent with that inferred from high-redshift AGN (Merritt and Ferrarese 2001a, Yu and Tremaine 2002) implying that ejections are rare. (e) Tight, empirical correlations between SBH mass and galaxy properties (Ferrarese and Merritt 2000, Graham et al 2001, Marconi and Hunt 2003) would also be weakened if SBH ejections were common. (f) The frequency of SBH binary coalescences estimated from X-shaped radio source statistics (Merritt and Ekers 2002) is roughly consistent with the galaxy merger rate, implying that the time for coalescence is short compared with the time between galaxy mergers.

A number of mechanisms have been proposed for efficiently extracting angular momentum from binary SBHS and avoiding the ‘final parsec problem.’

7.2.1. Non-axisymmetric geometries  As discussed above (section 4), steady-state, non-axisymmetric (e.g. triaxial) configurations are possible for nuclei, even in the presence of chaos induced by the SBH, and in fact departures from axisymmetry are often invoked to enhance fueling of AGN by gas (Shlosman et al 1990). Many orbits in a triaxial nucleus are centrophilic, passing arbitrarily close to the centre after a sufficiently long time (Norman and Silk 1983, Gerhard and Binney 1985). This implies feeding rates for a central binary that can be higher than in the spherical geometry and, more importantly, independent of $N$ in the large-$N$ limit (Merritt and Poon 2004).

The total rate at which stars pass within a distance $R_a$ of a central binary is

$$\dot{M} \approx a \int A(E)M_c(E)dE,$$

where $M_c(E)dE$ is the mass on centrophilic orbits in the energy range $E$ to $E + dE$, and $A(E)$ was defined above (equation (33)): $A(E) \times d$ is the rate at which a single star on a centrophilic orbit of energy $E$ experiences near-centre passages with pericentre distances $\leq d$. The implied hardening rate is

$$s \equiv \frac{d}{dt} \left( \frac{1}{a} \right) \approx \frac{2\langle C \rangle}{aM_c} \int \dot{M}(E)dE.$$  

Here, $\langle C \rangle \approx 1.25$ is the average value of the dimensionless energy change during a single star–binary encounter, $C \equiv [M_e/2m_\star](\Delta E/E)$. This expression can be evaluated for a $\rho \sim r^{-2}$ galaxy using equation (34) for $A(E)$. The result is

$$s \approx \frac{4\sqrt{6}}{9} \frac{\langle C \rangle K_{T_c}}{\sigma r_h^2} \int e^{-(E-E_h)/2\sigma^2} dE \approx 2.5\overline{f}_c\sigma r_h.$$  

Here $\overline{f}_c$ is an energy-weighted, mean fraction of centrophilic orbits, and the lower integration limit was set to $E_h \equiv \Phi(r_h)$. For the triaxial galaxy modelled in figure 27, the implied value of $s$ is $\sim 40\overline{f}_c$, consistent with the measured peak value of $s \approx 20$.

This simple calculation, like the $N$-body models on which figure 27 was based, is highly idealized, and it would be premature to draw strong conclusions about the behaviour of binary
SBHs in more realistic models of merging galaxies. However these results do convincingly demonstrate that binary hardening can be much more efficient in triaxial geometries than in spherical geometries due to the qualitatively different character of the stellar orbits.

7.2.2. Collisional loss cone repopulation. The theory of loss cones around single black holes (section 6) can be applied, with only minor changes, to binary SBHs. The orbital separation $a$ of the SBHs in a binary is much larger than the tidal disruption radius around a single SBH:

$$a \approx 10^5 \times \eta^{-2/3} \left( \frac{M_*}{10^6 M_\odot} \right)^{-1/3} \left( \frac{m_*}{M_\odot} \right)^{1/3} \left( \frac{r_\ast}{R_\odot} \right)^{-1} \left( \frac{a}{1\text{pc}} \right).$$

(91)

Since the physical scale of the loss cone is so large for a binary, the angular deflection of a star over one orbital period will almost always be small compared with the angular size subtended by the loss cone, implying that most stars will wander ‘diffusively’ into the binary (Milosavljević and Merritt 2003a). In the diffusive regime, the flux of stars into a sphere of radius $r_t$ scales only logarithmically with $r_t$ (equation (66b)), and so the rate of supply of stars to a massive binary will be of the same order as the loss rate into a single SBH of the same mass, or (very roughly) $\sim M_{12}/T(r_h)$. In the bright elliptical galaxies that exhibit clear evidence for the ‘scouring’ effect of binary SBHs, nuclear relaxation times are always extremely long, $\gtrsim 10^{14}$ yr (figure 1). In these galaxies, the mass in stars scattered into a central binary in $10^{10}$ yr would be completely negligible compared with $M_{12}$. Encounter-driven loss cone repopulation is only likely to be significant in galaxies with central relaxation times shorter than $\sim 10^{10}$ yr, since in these galaxies the mass scattered into the binary over the lifetime of the universe can be comparable with the binary’s mass. Figures 1 and 17 suggest that the only galaxies in this regime are those—such as the Milky Way and M32—which exhibit steep, power-law density profiles at $r \lesssim r_h$. If binary SBHs were ever present in these galaxies, the low-density cores which they produced have since presumably ‘filled in’ via encounters (section 5.3) or via new star formation; hence the rate of scattering of stars into the binary’s sphere of
influence would have been much lower in the past than inferred from the current density profiles.

These arguments were largely confirmed by Yu (2002) in her study of the influence of collisional loss-cone refilling on binary SBH evolution in a sample of nearby galaxies. Yu (2002) made a number of simplifying assumptions. (1) The galaxy’s density profile was assumed to be fixed in time; the ‘scouring’ effect of the binary was ignored. (2) The distribution of stars around the binary’s loss cone was assumed to be in a collisional steady state, even in galaxies with nuclear relaxation times \( \gg 10^{10} \) yr. (3) The influence of stars on the binary was computed using the results of scattering experiments on isolated binaries. Thus the hardening rate of the binary was computed from

\[ \frac{d}{dt} \left( \frac{1}{a} \right) = H \frac{G \rho}{\sigma} \]  

(92)

where \( \rho \) and \( \sigma \) are the stellar density and velocity dispersion and \( H = H(a, e, q) \) a dimensionless rate coefficient (Hills 1983, 1992, Mikkola and Valtonen 1992, Quinlan 1996, Merritt 2001)\(^6\). (4) Only changes in orbital angular momenta of the stars were considered, even though the stellar energy distribution would also change significantly in the time \( \sim T_r \) required for significant feeding of the binary. Under these assumptions, Yu (2002) found that binary hardening time scales were longer than a Hubble time in almost all galaxies, with little dependence on binary mass ratio. Yu’s more detailed conclusions about the properties of ‘stalled’ binaries are probably too rough to be very useful because of her neglect of the influence of the binary on the density profile.

\( N \)-body techniques would seem to be better suited to the collisional-loss-cone problem since they can easily deal with the strong, early effects of the binary on the stellar distribution, and, when coupled with regularization schemes, can accurately treat binary–star interactions without the need for rate coefficients derived from scattering experiments. An example of a mechanism that is not reproduced in scattering experiments is the ‘secondary slingshot,’ the repeated interaction of a star with a binary (Milosavljević and Merritt 2003a). But unless \( N \) is very large, relaxation in \( N \)-body simulations is so rapid that the binary’s loss cone remains essentially full, and the diffusive loss cone repopulation expected in real galaxies will not be reproduced (Milosavljević and Merritt 2003a).

Figure 28 shows a promising early step in this direction, using \( N \)-body models with \( N \) up to \( 0.25 \times 10^6 \); the Mikkola–Aarseth chain regularization algorithm was used for close SBH–SBH and SBH–star interactions (Mikkola and Aarseth 1990, 1993, Aarseth 2003b). The initial galaxy models had \( \rho \sim r^{-0.5} \) near the centre, somewhat shallower than expected in real galaxies, although steeper than the constant-density cores assumed in most modelling studies (e.g. Henssendorf et al. 2002, Chatterjee et al. 2003, Makino and Funato 2004). Adopting a steeper initial profile would have produced models with essentially full loss cones; in the models of figure 28, one can show that the loss cone defined by the binary was only partially filled for the largest \( N \) considered (Merritt et al. 2005). The \( N \)-dependence of the hardening rate was found to be \( (d/dt)(1/a) \sim N^{-0.4} \), shallower than the \( \sim N^{-1} \) dependence expected in real galaxies. Simulations like these can therefore still not be scaled to real galaxies, but they are useful in exploring the \( N \)-dependence of the changes induced by the binary on the galaxy. For instance, figure 28 (a) shows that mass deficits are not strongly \( N \)-dependent when expressed as a function of the binary separation \( a \). On the other hand, figure 28 (b) suggests that

\( N \)

Rate coefficients derived from scattering experiments can only be legitimately applied to binary SBHs in circumstances where nature would contrive to continue supplying stars to the binary; otherwise the binary will rapidly eject all stars on intersecting orbits and decay will halt, as discussed above. Unfortunately it is common practice to apply equations like (92) without providing justification of this sort.
the evolution of the binary’s eccentricity is strongly $N$-dependent. This may explain the rather disparate results on eccentricity evolution in various $N$-body studies (Milosavljević and Merritt 2001, Hemsendorf et al. 2002, Aarseth 2003a).

7.2.3. Summary of late evolution regimes. Table 2 summarizes the different regimes of binary evolution in stellar nuclei. ‘Collisionless’ refers here (as elsewhere in this article) to the large-$N$ limit in which star–star gravitational encounters are inactive and stars move along fixed orbits, until interacting with the binary. Almost all galaxies are expected to be in this regime. The evolution of a real binary SBH may reflect a combination of these and other mechanisms, such as interaction with gas (Merritt and Milosavljević 2005). There is a close parallel between the final parsec problem and the problem of quasar fuelling: both require that of order $10^8 M_\odot$ be supplied to the inner parsec of a galaxy in a time shorter than the age of the universe. Nature clearly accomplishes this in the case of quasars, probably through gas flows driven by torques from stellar bars. The same inflow of gas could contribute to the decay of a binary SBH in a number of ways: by leading to the renewed formation of stars which subsequently interact with the binary, by inducing torques which extract angular momentum from the binary, through accretion, increasing the masses of one or both of the SBHs and reducing their separation, etc.

Table 2. Regimes for long-term evolution of binary SBHs.

| Geometry        | Loss-Cone Regime       | Decay                        |
|-----------------|------------------------|------------------------------|
| Spherical/      | Collisionless          | $a^{-1} \propto \text{const (stalls)}$ |
| Axisymmetric    |                        |                              |
| Spherical/      | Collisional (diffusive)| $a^{-1} \propto t/N$          |
| Axisymmetric    |                        |                              |
| Spherical/      | Collisional (full loss cone) | $a^{-1} \propto t + \text{const}$ |
| Axisymmetric    |                        |                              |
| Triaxial        | Collisionless          | $a^{-1} \propto t + \text{const}$ |
7.3. Multiple black hole systems

If binary decay stalls, an uncoalesced binary may be present in a nucleus when a third SBH, or a second binary, is deposited there following a subsequent merger. The multiple SBH system that forms will engage in its own gravitational slingshot interactions, eventually ejecting one or more of the SBHs from the nucleus and possibly from the galaxy and transferring energy to the stellar fluid.

If the infalling SBH is less massive than either of the components of the pre-existing binary, \( M_3 < (M_1, M_2) \), the ultimate outcome is likely to be ejection of the smaller SBH and recoil of the binary, with the binary eventually returning to the galaxy centre. The lighter SBH is ejected with a velocity roughly \( 1/3 \) the relative orbital velocity of the binary (Saslaw et al. 1974, Hut and Rees 1992), and the binary recoils with a speed that is lower by \( \sim M_3/(M_1+M_2) \). Each close interaction of the smaller SBH with the binary increases the latter's binding energy by \( \langle \Delta E/E \rangle \approx 0.4 M_3/(M_1+M_2) \) (Hills and Fullerton 1980). If \( M_3 > M_1 \) or \( M_3 > M_2 \), there will most often be an exchange interaction, with the lightest SBH ejected and the two most massive SBHs forming a binary; further interactions then proceed as in the case \( M_3 < (M_1, M_2) \).

During the three-body interactions, both the semi-major axis and eccentricity of the dominant binary change stochastically. Since the rate of gravitational wave emission is a strong function of both parameters (\( \dot{E} \propto a^{-4}(1-e^2)^{-7/2} \)), the timescale for coalescence can be enormously shortened. This may be the most promising way to coalesce SBH binaries in the low-density nuclei of massive galaxies, where stalling of the dominant binary is likely.

This process has been extensively modelled using the PN2.5 approximation to represent gravitational wave losses (Peters and Mathews 1963) and assuming a fixed potential for the galaxy (Valtonen et al. 1989, Mikkola and Valtonen 1990, Valtonen et al. 1994). In these studies, there was no attempt to follow the pre-merger evolution of the galaxies or the interaction of the binary SBHs with stars. In two short non-technical contributions (submissions for the IEEE Gordon Bell prizes in 2001 and 2002), Makino and collaborators mention two N-body simulations of triple SBH systems at the centres of galaxies using the GRAPE-6 and (apparently) a modified version of NBODY1. Relativistic energy losses were neglected and the SBH particles all had the same mass. Plots of the time evolution of the orbital parameters of the dominant binary show strong and chaotic eccentricity evolution, with values as high as 0.997 reached for short periods. Such a binary would lose energy by gravity wave emission very rapidly, by a factor \( \sim 10^8 \) at the time of peak eccentricity.

In a wide, hierarchical triple, \( M_3 \ll (M_1, M_2) \), the eccentricity of the dominant binary oscillates through a maximum value of \( \sim 1 - 5 \cos^2 i/3 \), \( \cos i \ll \sqrt{3}/5 \), with \( i \) the mutual inclination angle (Kozai 1962). Blaes et al. (2002) estimated that the coalescence time of the dominant binary in hierarchical triples can be reduced by factors of \( \sim 10 \) via the Kozai mechanism; Iwasawa et al. (2005) recently observed Kozai oscillations in N-body simulations of galaxy models containing equal-mass triples.

If the binary SBH is hard when the third SBH falls in, the ejected SBH can gain enough velocity to escape the galaxy. If the three masses are comparable, even the binary can be kicked up to escape velocity. One study (Volonteri et al. 2003) estimates (based on a very simplified model of the interactions) that the recoil velocity of the smallest SBH is larger than galactic escape velocities in 99% of encounters and that the binary escapes in 8% of encounters. Thus

---

7 This idea was developed very thoroughly as a model for two-sided radio galaxies; the radio lobes were associated with matter entrained with the ejected SBHs (Valtonen and Heinämäki 2000 and references therein). While the ‘slingshot’ model for radio galaxies has not gained wide acceptance, these papers are still an extremely valuable resource for information about the statistics of multiple-SBH interactions.
a significant fraction of nuclei could be left with no SBH, with an offset SBH or with an SBH whose mass is lower than expected based on the $M_\bullet-\sigma$ or $M_\bullet-L_{\text{bulge}}$ relations.

There is a need for comprehensive $N$-body simulations of multiple-SBH systems that include gravitational loss terms, accurate (regularized) interactions between the SBH particles and self-consistent treatment of the stars.

Acknowledgments

My understanding of the topics discussed here has benefited enormously over the years from discussions with many of my colleagues, including (but not limited to) D Axon, H Baumgardt, D Heggie, P Hut, S Komossa, S McMillan, M Milosavljevic, S Portegies Zwart, G Quinlan, F Rasio, A Robinson, and R Spurzem. This work was supported by grants AST-0206031, AST-0420920 and AST-0437519 from the NSF, grant NNG04GJ48G from NASA, and grant HST-AR-09519.01-A from STScI.

References
Peters P C and Mathews J 1963 Phys. Rev. 131 435–40
Phinney E S 1989 IAU Symp. 136: The Center of the Galaxy ed M Morris pp 543–54
Pizzella A, Corsini E M, Morelli L, Sarzi M, Scarlata C, Stiavelli M and Bertola F 2002 Astrophys. J. 573 131–7
Plummer H C 1911 Mon. Not. R. Astron. Soc. 71 460–70
Poon M Y and Merritt D 2001 Astrophys. J. 549 192–204
Poon M Y and Merritt D 2002 Astrophys. J. (Lett.) 568 L89–92
Poon M Y and Merritt D 2004 Astrophys. J. 606 774–87
Portegies Zwart S F, Baumgardt H, Hut P, Makino J and McMillan S L W 2004 Nature 428 724–6
Preto M, Merritt D and Spurzem R 2004 Astrophys. J. 613 L109
Pringle J E 1991 Mon. Not. R. Astron. Soc. 248 754–9
Pringle J E 1992 Mon. Not. R. Astron. Soc. 258 811–8
Pringle J E 1996 Mon. Not. R. Astron. Soc. 281 357–61
Pullin J 1998 Gravitation and relativity: At the turn of the millennium Proc. GR-15 Conf. IUCCA (Pune) ed N Dadhich and J Narlikar pp 87–106
Quataert E, Narayan R and Reid M J 1999 Astrophys. J. Lett. 517 L101–4
Quinlan G D 1996 New Astron. 1 35–56
Quinlan G D and Hernquist L 1997 New Astron. 3 533–54
Quinlan G D and Shapiro S L 1987 Astrophys. J. 321 199–210
Quinlan G D and Shapiro S L 1989 Astrophys. J. 343 725–49
Rafikov R R 2002 Astrophys. J. 569 997–1008
Raiteri C M et al 2001 Astron. Astrophys. 377 396–412
Ravindranath S, Ho L C and Filippenko A V 2002 Astrophys. J. 566 801–8
Read J and Gilmore G 2003 Mon. Not. R. Astron. Soc. 339 949–56
Redmount I H and Rees M J 1989 Comm. Astron. 14 165–73
Rees M J 1984 Annu. Rev. Astron. Astrophys. 22 471–506
Rees M J 1990 Science 247 817–23
Reid M J, Readhead ACS, Vermeulen R C and Treuhaft R N 1999 Astrophys. J. 524 816–23
Richstone D O 1982 Astrophys. J. 252 496–507
Richstone D O and Tremaine S 1988 Astrophys. J. 327 82–8
Rodriguez C, Taylor G B, Zavala R T, Peck A B, Pollack L K and Romani R W 2006 Astrophys. J. 646 49–60
Romero G D, Chajet L, Abraham Z and Fan J H 2000 Astron. J. 119 1617–25
Roos N 1981 Astron. Astrophys. 104 218–28
Roos N, Kaasstra J S and Hummel C A 1993 Astrophys. J. 409 130–3
Rudak B and Paczynski B 1983 Acta Astron. 31 13–24
Safronov V S 1960 Ann. Astrophysique 23 979–82
Saitoh T R and Wada K 2004 Astrophys. J. 615 L93–6
Salo R M and Stutler T S 2001 Astrophys. J. (Lett.) 551 L49–52
Salpeter E E 1955 Astrophys. J. 121 161–7
Sambhus N and Sridhar S 2000 Astrophys. J. 542 143–60
Sambhus N and Sridhar S 2002 Astron. Astrophys. 388 766–70
Sanders D B, Soifer B T, Elias J H, Madore B F, Matthews K, Neugebauer G and Scoville N Z 1988 Astrophys. J. 325 74–91
Sanders R H 1970 Astrophys. J. 162 791–810
Saslaw W C 1973 Pub. Astron. Soc. Pacific 85 5–23
Saslaw W C, Valtonen M J and Aarseth S J 1974 Astrophys. J. 190 253–70
Schödel R, Eckart A, Genzel R, Merritt D, Alexander T, Ott T and Straubmeier C 2006 in preparation
Schödel R, Ott T, Genzel R, Eckart A, Mouawad N and Alexander T 2003 Astrophys. J. 596 1015–34
Schmitt H R, Pringle J E, Clarke C J and Kinney A L 2002 Astrophys. J. 575 150–5
Schutz B F 2002 Light-houses of the universe: The most luminous celestial objects and their use for cosmology Proceedings of the MPA/ESO ed M Gilfanov, R Sunyaev and E Churazov p 207 and 207–224
Schwarzchild M 1979 Astrophys. J. 232 236–47
Schwarzchild M 1981 Structure and Evolution of Normal Galaxies pp 43–53
Schwarzchild M 1993 Astrophys. J. 409 563–77
Sérscic J L 1968 Atlas de Galaxias Austrasle Cordoba, (Argentina: Observatorio Astronomico)
Sesana A, Haardt F, Madau P and Volonteri M 2004 Astrophys. J. 611 623–32
Shapiro S L 1977 Astrophys. J. 217 281–6
Shapiro S L and Marchant A B 1978 Astrophys. J. 225 603–24
Shaw M, Axon D, Probst R and Gatley I 1995 Mon. Not. R. Astron. Soc. 274 369–87
Dynamics of galaxy cores

Volonteri M, Haardt F and Madau P 2003 Astrophys. J. 582 559–73
Volonteri M, Madau P and Haardt F 2003 Astrophys. J. 593 661–6
Wang J and Merritt D 2004 Astrophys. J. 600 149–61
Wehner E and Harris W 2006 Astrophys. J. 644 L17–20
Wehrle A E, Unwin S C, Jones D L, Meier D L and Piner B G 2003 Interferometry in Space S Michael Proc. SPIE 4852 pp 152–60
Weidinger M, Moller P and Fynbo J P U 2004 Nature 430 999–1001
Wilson A S and Colbert E J M 1995 Astrophys. J. 438 62–71
Woosley S E, Heger A and Weaver T A 2002 Rev. Mod. Phys. 74 1015–1071
Wyithe J S B and Loeb A 2002 Astrophys. J. 581 886–94
Xie G 2003 Publ. Yunnan Observ. 95 107–116
Xie G Z, Li K H, Zhang X, Bai J M and Liu W W 1999 Astrophys. J. 522 846–62
Xie G Z, Liang E W, Zhou S B, Li K H, Dai B Z and Ma L 2002 Mon. Not. R. Astron. Soc. 334 459–70
Young P 1980 Astrophys. J. 242 1232–7
Young P J 1976 Phys. Rev. D 14 3281–9
Young P J 1977a Astrophys. J. 212 227–33
Young P J 1977b Astrophys. J. 215 36–52
Young P J, Shields G A and Wheeler J C 1977 Astrophys. J. 212 367–82
Yu Q 2002 Mon. Not. R. Astron. Soc. 331 935–58
Yu Q and Tremaine S 2002 Mon. Not. R. Astron. Soc. 335 965–76
Yu Q and Tremaine S 2003 Astrophys. J. 599 1129–38
Zhang X, Xie G Z and Bai J M 1998 Astron. Astrophys. 330 469–73
Zhao H, Carollo C M and de Zeeuw P T 1999 Mon. Not. R. Astron. Soc. 304 457–64
Zhao H, Hachnelt M G and Rees M J 2002 New Astron. 7 385–94
Zier C and Biermann P L 2002 Astron. Astrophys. 396 91–108