Massive Black Hole Binary Evolution

David Merritt
Rochester Institute of Technology

Milos Milosavljevic
California Institute of Technology

Follow this and additional works at: http://scholarworks.rit.edu/article

Recommended Citation

This Article is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Articles by an authorized administrator of RIT Scholar Works. For more information, please contact ritsscholarworks@rit.edu.
Massive Black Hole Binary Evolution

David Merritt
Rochester Institute of Technology
Rochester, NY USA
e-mail:David.Merritt@rit.edu
http://www.rit.edu/~drmsps/

Miloš Milosavljević
Theoretical Astrophysics
California Institute of Technology
Pasadena, CA USA
e-mail:milos@tapir.caltech.edu
http://www.tapir.caltech.edu/~milos/

Abstract

Coalescence of binary supermassive black holes (SBHs) would constitute the strongest sources of gravitational waves to be observed by LISA. While the formation of binary SBHs during galaxy mergers is almost inevitable, coalescence requires that the separation between binary components first drop by a few orders of magnitude, due presumably to interaction of the binary with stars and gas in a galactic nucleus. This article reviews the observational evidence for binary SBHs and discusses how they would evolve. No completely convincing case of a bound, binary SBH has yet been found, although a handful of systems (e.g. interacting galaxies; remnants of galaxy mergers) are now believed to contain two SBHs at projected separations of \( \lesssim 1 \) kpc. \( N \)-body studies of binary evolution in gas-free galaxies have reached large enough particle numbers to reproduce the slow, “diffusive” refilling of the binary’s loss cone that is believed to characterize binary evolution in real galactic nuclei. While some of the results of these simulations – e.g. the binary hardening rate and eccentricity evolution – are strongly \( N \)-dependent, others – e.g. the “damage” inflicted by the binary on the nucleus – are not. Luminous early-type galaxies often exhibit depleted cores with masses of \( \sim 1 - 2 \) times the mass of their nuclear SBHs, consistent with the predictions of the binary model. Studies of the interaction of massive binaries with gas are still in their infancy, although much progress is expected in the near future. Binary coalescence has a large influence on the spins of SBHs,
even for mass ratios as extreme as 10 : 1, and evidence of spin-flips may have been observed.

1 Introduction

With an ever-increasing number of secure detections, supermassive black holes (SBHs) have evolved, in the span of a few years, from exotic possibilities to well-established components of galaxies. While it was understood since the 1960's that the energy sources of quasars must be gravitational [185], it was thirty years before the existence of SBHs was firmly established, through measurements of the Keplerian rise in the rotation velocity of stars or gas at the very centers of galactic nuclei [106]. It is now generally accepted that the formation and evolution of galaxies and SBHs are tightly intertwined, from the early phases of proto-galactic formation [200], through hierarchical build-up in CDM-like cosmogonies [83], to recent galaxy mergers [150].

SBHs appear to be linked in fundamental ways to the dynamics of the stellar component in galaxies, both on large and small scales. An astonishingly tight correlation exists between SBH mass and the central velocity dispersion of the stellar component, $M_\bullet \sim \sigma^\alpha$, $\alpha \approx 4.5$ [51]; the correlation with the velocity dispersion averaged over kiloparsec scales is weaker but still impressive [65, 49]. Similar correlations exist between SBH mass and bulge luminosity [131, 129] and central concentration of the light [78, 38], indicating that SBHs “know” about the depth of the gravitational potential well in which they live. These tight correlations probably reflect a degree of feedback in the growth of SBHs [200].

On small scales, SBHs are embedded in stellar cusps, parsec-scale regions where the stellar density increases approximately as a power law with distance from the SBH into the smallest resolvable radii [29, 52, 138, 66]. Faint galaxies have steep nuclear density profiles, $\rho \sim r^{-\gamma}$, $1.5 \lesssim \gamma \lesssim 2.5$, while bright galaxies typically have weaker cusps, $\gamma \lesssim 1$. Steep cusps form naturally as the growth of the SBH pulls in stars [163]. In small dense galaxies where the star-star relaxation time is shorter than $10^{10}$ yr, steep cusps may also form via collisional relaxation [8, 171]. Weak cusps may be remnants of strong cusps that were destroyed by binary SBHs during galaxy mergers; in fact the structure and kinematics of galactic nuclei are now believed to be fossil relics of the merger process [135].

Larger galaxies grow through the agglomeration of smaller galaxies and protogalactic fragments. If more than one of the fragments contained a SBH, the SBHs will form a bound system in the merger product [18, 187]. This scenario has received considerable attention because the ultimate coalescence of such a pair would generate an observable outburst of gravitational waves [209]. The evolution of a binary SBH can be divided into three phases [18]: 1. As the galaxies merge, the SBHs sink toward the center of the new galaxy via dynam-
ical friction where they form a binary. 2. The binary continues to decay via gravitational slingshot interactions [191] in which stars on orbits intersecting the binary are ejected at velocities comparable to the binary’s orbital velocity, while the binary’s binding energy increases. 3. If the binary’s separation decreases to the point where the emission of gravitational waves becomes efficient at carrying away the last remaining angular momentum, the SBHs coalesce rapidly.

The transition from (2) to (3) is understood to be the bottleneck of a SBH binary’s path to coalescence, since the binary will quickly eject all stars on intersecting orbits, thus cutting off the supply of stars. This is called the “final parsec problem” [152]. But there are other possible ways of continuing to extract energy and angular momentum from a binary SBH, including accretion of gas onto the binary system [4] or refilling of the loss cone via star-star encounters [228, 151] or triaxial distortions [143]. Furthermore there is circumstantial evidence that efficient coalescence is the norm. The X-shaped radio sources [32] are probably galaxies in which SBHs have recently coalesced, causing jet directions to flip. The inferred production rate of the X-sources is comparable to the expected merger rate of bright ellipticals, suggesting that coalescence occurs relatively quickly following mergers [137]. If binary SBHs failed to merge efficiently, uncoalesced binaries would be present in many bright ellipticals, resulting in 3- or 4-body slingshot ejections when subsequent mergers brought in additional SBHs. This would produce off-center SBHs, which seem to be rare or non-existent, as well as (perhaps) too much scatter in the $M - \sigma$ and $M - L_{\text{bulge}}$ relations [83].

While the final approach to coalescence of binary SBHs is not well understood, much of their dynamical effect on the surrounding nucleus takes place very soon after the binary forms. The binary quickly (less than a galactic crossing time) ejects from the nucleus a mass in stars of order its own mass [177, 150] significantly lowering the central density on parsec scales. There is reasonably quantitative agreement between this model and the observed structure of nuclei: the “mass deficit” – the stellar mass that is “missing” from the centers of galaxies, assuming that they once had steep cusps like those observed at the centers of faint ellipticals – is of order the black hole mass [153, 181, 77].

While the binary SBH model is compelling, there is still not much hard evidence in its support. Observationally, no bona fide binary SBH (i.e. gravitationally bound pair of SBHs) has definitely been detected, although there is circumstantial evidence (precessing radio jets; periodic outburst activity) for SBH binaries in a number of active galaxies, as reviewed briefly below (see [104] for a more complete review of this topic). But the binary SBH model has one great advantage: the postulated effects are accessible to observation, since they extend to scales of $1 - 100$ pc, the distance out to which a binary SBH can significantly influence stellar motions. Much of the recent theoretical work in this field has been directed toward understanding the influence of a binary SBH on its stellar surroundings and looking for evidence of that influence in the distribution of light at the centers of galaxies.
Following the definition of terms and time scales in §2, we present a brief overview of the observational evidence for binary SBHs in §3. Interaction of a binary SBH with stars is discussed in §4. The possibility of multiple SBHs in galactic nuclei, and the implications for coalescence, are discussed in §5. §6 summarizes N-body work on the evolution of binary SBHs, with an emphasis on the question of binary wandering. Observational evidence for the destruction of nuclear density cusps is reviewed in §7. In some galaxies, the predominant source of torques leading to decay of the binary may be gas; this topic is reviewed in §8. Finally, the influence of binary coalescence on SBH spins is summarized in §9.

2 Preliminaries

We write \( m_1 \) and \( m_2 \) for the masses of the two components of a binary SBH, with \( m_2 \leq m_1 \), \( q \equiv m_2/m_1 \), and \( m_{12} \equiv m_1 + m_2 \). (We also sometimes write \( M \) for the mass of the single SBH that forms via coalescence of two SBHs of combined mass \( m_{12} \).) The semi-major axis of the binary’s Keplerian orbit is \( a \) and \( e \) is the orbital eccentricity. The binary’s binding energy is

\[
|E| = \frac{G m_1 m_2}{2a} = \frac{G \mu m_{12}}{2a} \quad (1)
\]

with \( \mu = m_1 m_2/m_{12} \) the reduced mass. The orbital period is

\[
P = 2\pi \left( \frac{a^3}{G m_{12}} \right)^{1/2} = 9.36 \times 10^3 \text{ yr} \left( \frac{m_{12}}{10^8 M_\odot} \right)^{-1/2} \left( \frac{a}{1 \text{ pc}} \right)^{3/2} \quad (2)
\]

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

\[
V_{\text{bin}} = \sqrt{\frac{G m_{12}}{a}} = 658 \text{ km s}^{-1} \left( \frac{m_{12}}{10^8 M_\odot} \right)^{1/2} \left( \frac{a}{1 \text{ pc}} \right)^{-1/2} \quad (3)
\]

A binary is “hard” when its binding energy per unit mass, \( |E|/m_{12} = G\mu/2a \), exceeds \( \sim \sigma^2 \), where \( \sigma \) is the 1D velocity dispersion of the stars in the nucleus. The precise meaning of “hard” is debatable when talking about a binary whose components are much more massive than the surrounding stars [87, 177]. For concreteness, we adopt the following definition for the semi-major axis of a hard binary:

\[
a \leq a_h \equiv \frac{G\mu}{4\sigma^2} \approx 2.7 \text{ pc} \left( 1 + q \right)^{-1} \left( \frac{m_2}{10^8 M_\odot} \right) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-2} \quad (4)
\]

At distances \( r \gg a \), stars respond to the binary as if it were a single SBH of mass \( M \). The gravitational influence radius of a single SBH is defined as the
distance within which the force on a test mass is dominated by the SBH, rather than by the stars. A standard definition for $r_{\text{infl}}$ is

$$r_{\text{infl}} = \frac{GM}{\sigma^2} \approx 10.8 \text{ pc} \left(\frac{M}{10^8 M_\odot}\right) \left(\frac{\sigma}{200 \text{ km s}^{-1}}\right)^{-2}. \quad (5)$$

Thus $r_{\text{infl}} = 4(M/\mu)a_h$. For an equal-mass binary, $r_{\text{infl}} \approx 16a_h$, and for a more typical mass ratio of $q = 0.1$, $r_{\text{infl}} \approx 50a_h$. An alternative, and often more useful, definition for $r_{\text{infl}}$ is the radius at which the enclosed mass in stars is twice the black hole mass:

$$M_*(r < r_{\text{infl}}) = 2M. \quad (6)$$

This definition is appropriate in nuclei where $\sigma$ is a strong function of radius; it is equivalent to equation (5) when the density of stars satisfies $\rho(r) = \sigma^2/2\pi Gr^2$, the “singular isothermal sphere,” and when $\sigma$ is measured well outside of $r_{\text{infl}}$.

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 [165]

$$t_{gr} = \frac{5}{256F(e) G \mu m_{12}^{2/3}} c^5 a^4,$$

$$F(e) = \left(1 - e^2\right)^{7/2} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4\right). \quad (7)$$

This can be written

$$t_{gr} = \frac{5}{164F(e) \sigma^8 m_{12}^{2/3}} \left(\frac{a}{a_h}\right)^4$$

$$\approx 3.07 \times 10^8 \text{yr} \left(\frac{q^3}{(1+q)^6}\right) \left(\frac{m_{12}}{10^8 M_\odot}\right) \left(\frac{\sigma}{200 \text{ km s}^{-1}}\right)^{-8} \left(\frac{a}{10^{-2}a_h}\right)^4. \quad (8)$$

This relation can be simplified by making use of the tight empirical correlation between SBH mass and $\sigma$, the “$M - \sigma$ relation.” Of the two forms of the $M - \sigma$ relation in the literature [51, 65], the more relevant one [51] is based on the velocity dispersion measured in an aperture centered on the SBH, which is approximately the same quantity $\sigma$ defined above; the alternative form [65] defines $\sigma$ as a mean value along a slit that extends over the entire half-light radius of the galaxy. In terms of the central $\sigma$, the best current estimate of the $M - \sigma$ relation is [50]

$$\left(\frac{M}{10^8 M_\odot}\right) = (1.66 \pm 0.24) \left(\frac{\sigma}{200 \text{ km s}^{-1}}\right)^{\alpha}. \quad (9)$$
with \( \alpha = 4.86 \pm 0.43 \). Combining equations (8) and (9) and setting \( M = m_{12}, F = 1 \) gives

\[
\begin{align*}
\tau \approx & \quad 5.0 \times 10^8 \text{yr} \frac{q^3}{(1+q)^6} \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^{-3.14} \left( \frac{a}{10^{-2} a_h} \right)^4 \\
\approx & \quad 7.1 \times 10^8 \text{yr} \frac{q^3}{(1+q)^6} \left( \frac{m_{12}}{10^8 M_\odot} \right)^{-0.65} \left( \frac{a}{10^{-2} a_h} \right)^4.
\end{align*}
\]

(10)

Coalescence in a Hubble time (\( \sim 10^{10} \text{ yr} \)) requires \( a < \sim 0.05 a_h \) for an equal-mass binary and \( a < 0.15 a_h \) for a binary with \( q = 0.1 \). Inducing a SBH to decay from a separation \( a \approx a_h \approx 10^9 \text{ pc} \) to a separation such that \( \tau \lesssim 10^{10} \text{ yr} \) is called the “final parsec problem” [152]. Much of the theoretical work on massive black hole binary evolution has focused on this problem.

3 Observations of Binary Supermassive Black Holes

3.1 External Galaxies

If a binary SBH is defined as two SBHs separated by a distance \( a \lesssim a_h \), then no completely convincing example of such a binary has yet been found. Here we briefly review the small set of cases in which clear evidence is seen for two, widely separated SBHs in a single system (“dual SBHs”), as well as the still-circumstantial evidence for true binary SBHs. For a more complete review of this topic, see [104].

3.1.1 Dual SBHs

Figure 1 shows what was probably the first clear example of two SBHs in one “system,” in this case a pair of interacting galaxies near the center of the galaxy cluster Abell 400. The associated radio source 3C75 consists of a pair of twin radio lobes originating from the radio cores of the two galaxies; the projected separation of the cores is \( \sim 7 \text{ kpc} \) [161]. Such double-jet systems are expected to be rare given the small fraction of giant elliptical galaxies that are associated with luminous radio sources.

“Binary” quasars are common but most are believed to be chance projections or lensed images [155, 102]. Among the binary quasars for which lensing can be ruled out, the smallest projected separation belongs to LBQS 0103-2753 at \( z = 0.85 \), with an apparent spacing between centers of 2.3 kpc [95]. However the two quasar spectra show a \( \Delta z \) of 0.024 suggesting a chance projection.

Galaxies in the late stages of a merger are the most plausible sites for dual SBHs and many of these exhibit double nuclei in the optical or infrared [80, 23]. However few show unambiguous evidence of AGN activity in both nuclei,
indicative of SBHs. One clear exception is NGC 6240 (Fig. 2), for which both nuclei exhibit the flat X-ray spectra characteristic of AGNs [105]. The projected separation is 1.4 kpc. Another likely case is Arp 299 [12].

Interestingly, there are no known dual SBHs with separations below $\sim 1$ kpc, even though a 1 kpc separation would be resolvable to distances of several hundred Mpc.

### 3.1.2 Evidence for Binary SBHs

Many active galaxies exhibit periodic variability with periods of days or years, consistent with the orbital periods of true binary SBHs having $a \lesssim a_h$. Undoubtedly the clearest example is OJ 287, a “blazar,” i.e. an active galaxy in which the jet is believed to be orientated nearly parallel to the line of sight, at $z = 0.306$. Optical variability of OJ 287 has been recorded since 1890 [175, 206] and has a strict period of 11.86 yr ($\sim 9$ yr in the galaxy’s rest frame); the last major outburst was observed (on schedule) in 1994. The outbursts are gener-
Figure 2: Chandra X-ray image of the starburst galaxy NGC 6240, showing the two nuclear sources. Projected separation of the nuclei is about 1.4 kpc. Image courtesy of NASA/CXC/MPE/S.Komossa et al.

ally double-peaked with the peaks separated by about a year; the second peak is accompanied by enhanced radio emission. Models to explain the periodicity usually invoke a second SBH with \( q \lesssim 0.1 \). In one class of model, the variability reflects true changes in the source luminosity due to variations in the accretion rate as the smaller SBH passes through the accretion disk surrounding the larger SBH [201, 117, 213]. In these models, the observed variability period is equal to the binary orbital period, and the binary orbit is highly eccentric \( (e \approx 0.7) \), implying a relatively short \( (\lesssim 10^5 \text{ yr}) \) time scale for orbital decay via gravitational radiation. The lag between primary and secondary peaks may be due to the time required for the disturbance induced by the passage through the accretion disk to propagate down the jet [213]. Alternatively, the luminosity
variations may reflect changes in the jet direction resulting from precession of the accretion disk, the latter induced by torques from the second SBH [96]. In this model, the binary orbital period is much less than 9 yr, and the secondary maxima could be due to a “nodding” motion of the accretion disk [96].

Many other examples of variability in AGN at optical, radio and even TeV energies are documented [223], with periods as short as \( \sim 25 \) days [85]. Indeed evidence for variability has even been claimed for the Milky Way SBH, at radio wavelengths; the ostensible period is 106 days [231]. However none of these examples exhibits as clear a periodicity as OJ287.

Table 3.1.2 gives a list of active galaxies for which periodic variability has been claimed.

<table>
<thead>
<tr>
<th>Source</th>
<th>Period (yr)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mkn 421</td>
<td>23.1</td>
<td>[121]</td>
</tr>
<tr>
<td>PKS 0735+178</td>
<td>14.2</td>
<td>[46]</td>
</tr>
<tr>
<td>BL Lac</td>
<td>14</td>
<td>[47]</td>
</tr>
<tr>
<td>ON 231</td>
<td>13.6</td>
<td>[122]</td>
</tr>
<tr>
<td>OJ 287</td>
<td>11.9</td>
<td>[175]</td>
</tr>
<tr>
<td>PKS 1510-089</td>
<td>0.92</td>
<td>[225]</td>
</tr>
<tr>
<td>Sgr A*</td>
<td>0.290</td>
<td>[231]</td>
</tr>
<tr>
<td>3C 345</td>
<td>10.1</td>
<td>[230]</td>
</tr>
<tr>
<td>AO 0235+16</td>
<td>5.7</td>
<td>[180]</td>
</tr>
<tr>
<td>3C 66A</td>
<td>0.175</td>
<td>[110]</td>
</tr>
<tr>
<td>Mkn 501</td>
<td>0.065</td>
<td>[85]</td>
</tr>
<tr>
<td>3C 273</td>
<td>0.0026</td>
<td>[224]</td>
</tr>
</tbody>
</table>

Table 1: Sources with periodic variation in the nuclear emission

Radio lobes in active galaxies provide a fossil record of the orientation history of the jets powering the lobes. Many examples of sinusoidally or helically distorted jets are known, and these observations are often interpreted via a binary SBH model. The wiggles may be due to physical displacements of the SBH emitting the jet (e.g., [188]) or to precession of the larger SBH induced by orbital motion of the smaller SBH (e.g., [186]). In the radio galaxy 3C 66B, the position of the radio core shows well-defined elliptical motions with a period of just 1.05 yr [204], implying \( t_{\text{pr}} \lesssim 10^{3} \) yr.

About a dozen radio galaxies exhibit abrupt changes in the orientation of their radio lobes, producing a “winged” or X-shaped morphology [115]. While originally interpreted via a precession model [36], a more likely explanation is that the SBH producing the jet has undergone a spin flip, due perhaps to capture of a second SBH [137, 233].

A number of quasars show the peaks of their broad emission lines at very
different redshifts from their narrow emission lines, or two displaced emission line peaks, which might be attributed to orbital motion of the SBHs associated with the line emitting regions [63, 64, 203, 232]. This interpretation has fallen out of favor however since the candidate systems do not show the predicted radial velocity variations [37].

A number of other possibilities exist for detecting binary SBHs, including the use of space interferometers to measure the astrometric reflex motion of AGN photocenters due to orbital motion of the jet-producing SBHs [219]; measurement of periodic shifts in pulsar arrival times due to passage of gravitational waves from binary SBHs [123]; and, of course, direct detection of gravitational waves by space-based interferometers.

3.2 Limits on the Binarity of The Milky Way Black Hole

The likely longevity of binary SBHs motivates the question whether the closest and best-studied SBH at the center of the Milky Way galaxy is a binary. Monitoring of the proper motion of stars orbiting the SBH has led to a precise measurement of its mass, \( M \sim 4 \times 10^6 M_\odot \) [69, 193]. The Milky Way SBH is coincident with the compact (< 1 AU) radio source Sagittarius (Sgr) A* [4]. If the Milky Way SBH were a binary, the radio source would probably be associated with the more massive of the two binary components. Limits on the masses of the components could be placed by measuring astrometric reflex motion of the radio source relative to distant quasars [6, 184]. Such measurements have recovered the magnitude of the solar reflex motion in the galaxy but have so far yielded no evidence for a binary SBH. The most recent upper limits on the mass of a binary companion of Sgr A* are \( M_2 \lesssim 10^4 M_\odot \) for binaries with semimajor axes \( 10^3 \text{ AU} < a < 10^5 \text{ AU} \) [183]. This places any companion that may exist in the class of “intermediate-mass” black holes (IBHs). The parameter space of SBH-IBH binaries at the Galactic center is illustrated in Figure 3.

IBHs have been suggested as a possible explanation for ultraluminous X-ray sources; however their existence is not widely accepted. It has been suggested that the center of the Milky Way is a place where IBHs might naturally form via the runaway merging of massive stars in the young, dense star clusters ([59] and references therein). Two such clusters, the Arches and the Quintuplet, are presently located in the Galactic center region. The segregation of massive stars to the cluster center accelerates the “core collapse” in which the stellar density at the center of the cluster increases drastically. Collapse time can be shorter than the life time of the most massive stars; in this case runaway stellar coalescence ensues resulting in the formation of a supermassive star at the cluster center. If the star survives mass loss through winds and avoids exploding as a pair-instability supernova, it collapses to form an IBH [220]. Dynamical friction in the background stellar cusp of the Galactic bulge subsequently drags the IBH toward the SBH until two black holes form a hard binary. This process might explain the puzzling presence of early-type stars [67, 69] deep inside the sphere.
Figure 3: A crude illustration of the parameter space for a SBH-IBH binary at the Galactic center. Assuming a circular orbit around a SBH of $3 \times 10^6 M_\odot$, a IBH with mass $M_{\text{IBH}}$ and semi-major axis $a$ can be ruled out by measurement of an astrometric wobble of the radio image of Sgr A$^*$. The shaded regions show the detection thresholds for astrometric resolutions of 0.01, 0.1 and 1 milliarc-seconds, respectively, assuming a monitoring period of 10 years. The dashed lines indicate coalescence due to gravitational radiation in $10^6$ and $10^7$ years, respectively (From [84], see also [229]).

of influence of the SBH at the Galactic center ([84], but see [99]).
4 Interaction of Binary Black Holes with Stars

4.1 Dynamics of a Massive Binary in a Fixed Stellar Background

Stars passing within a distance $\sim 3a$ of the center of mass of a hard binary undergo a complex interaction with the two black holes, followed almost always by ejection at velocity $\sim \sqrt{\mu/m_{12}V_{\text{bin}}}$, the “gravitational slingshot” [191]. Each ejected star carries away energy and angular momentum, causing the semimajor axis, eccentricity, orientation, and center-of-mass velocity of the binary to change and the local density of stars to drop. If the stellar distribution is assumed fixed far from the binary and if the contribution to the potential from the stars is ignored, the rate at which these changes occur can be computed by carrying out scattering experiments of massless stars against a binary whose orbital elements remain fixed during each interaction [89, 187, 87, 88, 13, 148, 177, 133, 134]. Figure 4 shows an example of field star velocity changes in a set of scattering experiments.

Consider an encounter of a single field star of mass $m_\star$ with the binary. Long before the encounter, the field star has velocity $v_0$ with respect to the center of mass of the field star-binary system and its impact parameter is $p$. A long time after the encounter, the velocity $v$ of the field star attains a constant value. Conservation of linear momentum implies that the change $\delta V$ in the velocity of the binary’s center of mass is given by

$$\delta V = -\frac{m_\star}{m_{12}}\delta v.$$  \hspace{1cm} (11)

The velocity change results in a random walk of the binary’s center-of-mass momentum, as discussed in more detail below. The energy of the field star-binary system, expressed in terms of pre-encounter quantities, is

$$E_0 = \frac{1}{2}m_\star v_0^2 + \frac{1}{2}m_{12}V_0^2 - \frac{Gm_1m_2}{2a_0}$$

$$= \frac{1}{2}m_\star \left(1 + \frac{m_\star}{m_{12}}\right)v_0^2 - \frac{Gm_1m_2}{2a_0}$$  \hspace{1cm} (12)

with $V_0 = -(m_\star/m_{12})v_0$ the initial velocity of the binary’s center of mass and $a_0$ the binary’s initial semi-major axis. After the encounter,

$$E = \frac{1}{2}m_\star \left(1 + m_\star/m_{12}\right)v^2 - \frac{Gm_1m_2}{2a}$$  \hspace{1cm} (13)

and $E = E_0$, so that

$$\delta \left(\frac{1}{a}\right) = \frac{m_\star(v^2 - v_0^2)}{Gm_1m_2} \left(1 + \frac{m_\star}{m_{12}}\right)$$

$$\approx \frac{m_\star(v^2 - v_0^2)}{Gm_1m_2}.$$  \hspace{1cm} (14)
Figure 4: Distribution of field star velocity changes for a set of scattering experiments in which the field star’s velocity at infinity relative to the binary was $v_{f0} = 0.5V_{bin} \hat{e}_x$. The binary’s mass ratio was 1 : 1, and the orientation of the binary’s orbital plane with respect to the $x$-axis was varied randomly between the scattering experiments. Each plot represents $5 \times 10^4$ scattering experiments within some range of impact parameters $[p_1, p_2]$ in units of $a$. (a) $[6, 10]$ (b) $[2, 4]$ (c) $[0.6, 1]$ (d) $[0.4, 0.6]$. Solid lines in (a) and (b) are the distributions corresponding to scattering off a point-mass perturber. In (c) and (d), the mean of this distribution (which is very narrow) is indicated by the arrows. The gravitational slingshot is apparent in the rightward shift of the $\delta v$ values when $p$ is small, due to the randomization of ejection angles (from [133]).

Averaged over a distribution of field-star velocities and directions, equation (4.1) gives the binary hardening rate $(d/dt)(1/a)$ [87, 88, 148, 177].

The angular momentum of the field star-binary system about its center of mass, expressed in terms of pre-encounter quantities, is

$$L_0 = m_* \left(1 + \frac{m_*}{m_{12}} \right) \ell_0 + \mu \ell_{60}$$

where $\ell_0 \equiv p v_0$ and $\ell_{60} \equiv L_{60}/\mu$ with $L_b$ the binary’s orbital angular momentum. Conservation of angular momentum during the encounter gives

$$\delta \ell_b = -\frac{m_*}{\mu} \left(1 + \frac{m_*}{m_{12}} \right) \delta \ell$$

13
\[ \approx -\frac{m_\ast}{\mu} \delta \ell. \] (16)

Changes in \(|\ell_b|\) correspond to changes in the binary’s orbital eccentricity \(e\) via the relation \(e^2 = 1 - \ell_b^2/Gm_{12}a^2\) [148, 177]. Changes in the direction of \(\ell_b\) correspond to changes in the orientation of the binary [134].

The results of the scattering experiments can be summarized via a set of dimensionless coefficients \(H, J, K, L, \ldots\) which define the mean rates of change of the parameters characterizing the binary and the stellar background. These coefficients are functions of the binary mass ratio, eccentricity and hardness but are typically independent of \(a\) in the limit that the binary is very hard. The hardening rate of the binary is given by

\[ \frac{d}{dt} \left( \frac{1}{a} \right) = H \frac{G\rho}{\sigma} \] (17)

with \(\rho\) and \(\sigma\) the density and 1D velocity dispersion of stars at infinity. The mass ejection rate is

\[ \frac{dM_{ej}}{d\ln(1/a)} = J m_{12} \] (18)

with \(M_{ej}\) the mass in stars that escape the binary. The rate of change of the binary’s orbital eccentricity is

\[ \frac{de}{d\ln(1/a)} = K. \] (19)

The diffusion coefficient describing changes in the binary’s orientation is

\[ \langle \Delta \vartheta^2 \rangle = L \frac{m_\ast}{m_{12}} \frac{G\rho a}{\sigma} \] (20)

with \(m_\ast\) the stellar mass. Brownian motion of the binary’s center of mass is determined by the coefficients \(A\) and \(C\) which characterize the Chandrasekhar diffusion coefficients at low \(V\):

\[ \langle \Delta v_\parallel \rangle = -AV; \]
\[ \langle \Delta v_\parallel^2 \rangle = 1/2 \langle \Delta v_\perp^2 \rangle = C. \] (21)

The mean square velocity of the binary’s center of mass is \(\langle V^2 \rangle = C/2A\).

The binary hardening coefficient \(H\) reaches a constant value of \(\sim 16\) in the limit \(a \ll a_h\), with a weak dependence on \(q\) [87, 148, 177]. In a fixed background, equation (17) therefore implies that a hard binary hardens at a constant rate:

\[ \frac{1}{a(t)} - \frac{1}{a_h} \approx H \frac{G\rho}{\sigma} (t - t_h), \quad t \geq t_h, \quad a(t_h) = a_h. \] (22)
This is sometimes taken as the definition of a “hard binary.” The time to reach zero separation is

\[
\begin{align*}
t - t_h &= \frac{1}{HG^2 \rho a_h} \sigma = \frac{2q(1 + q)2}{H G \rho m_{12}} \\
&\approx 5.2 \times 10^5 \text{ yr} q (1 + q) \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^3 \left( \frac{\rho}{10^3 M_\odot \text{ pc}^{-3}} \right)^{-1} \\
&\times \left( \frac{m_{12}}{10^8 M_\odot} \right)^{-1}.
\end{align*}
\]

Orbital shrinkage would occur quite rapidly in the environment of a galactic nucleus if the properties of the stellar background remained fixed.

However if the binary manages to shrink to a separation at which \( t_{gr} \lesssim 10^{10} \) yr, the changes it induces in its stellar surroundings will be considerable. The mass ejected by the binary in decaying from \( a_h \) to \( a_{gr} \) is given by the integral of equation (18):

\[
M_{ej} = \int_{a_{gr}}^{a_h} \frac{J(a)}{a} da.
\]

Figure 5 shows \( M_{ej} \) as a function of the mass ratio \( q \) for \( \sigma = 200 \) km s\(^{-1} \) and various values of \( t_{gr} \). The mass ejected in reaching coalescence is of order \( m_{12} \) for equal-mass binaries, and several times \( m_2 \) when \( m_2 \ll m_1 \). A SBH that grew to its current size through a succession of mergers should therefore have displaced a few times its own mass in stars. If this mass came mostly from stars that were originally in the nucleus, the density within \( r_{\text{infl}} \) would drop drastically and the hardening would stop. Without some way of replenishing the supply of stars (and in the absence of other mechanisms for extracting angular momentum from the binary, e.g., torques from gas clouds; cf. §8), decay would stall at a separation much greater than \( a_{gr} \).

In the case of extreme binary mass ratios, \( m_2 \ll m_1 \), the assumption that all stars passing a distance \( \sim a \) from the binary will be ejected is likely to be incorrect; many such stars will pass through the binary system without being appreciably perturbed by the smaller black hole. The concept of “ejection” in extreme binary mass ratios may be misleading; since the typical amount of energy transferred in stellar pericenter passages is small, most of the slingshot stars are not ejected from the nucleus and remain bound to the larger black hole. Figure 5 should thus be interpreted with caution in the regime \( m_2 \ll m_1 \).

The extreme-mass-ratio regime is poorly understood but deserves more study in view of the possibility that intermediate-mass black holes may exist having masses of \( \sim 10^3 M_\odot \) [43].

Changes in the binary’s orbital eccentricity (eq. 19) are potentially important because the gravity wave coalescence time drops rapidly as \( e \to 1 \) (eq. 8). For a hard binary, scattering experiments give \( K(e) \approx K_0 e (1 - e^2) \), with \( K_0 \approx 0.5 \) for an equal-mass binary [148, 177]. The dependence of \( K \) on \( m_2/m_1 \)
is not well understood and is an important topic for further study. The implied changes in $e$ as a binary decays from $a = a_h$ to zero are modest, $\Delta e \lesssim 0.2$, for all initial eccentricities.

### 4.2 Evolution in an Evolving Background

The scattering experiments summarized above treat the binary’s environment as fixed and homogeneous. In reality, the binary is embedded at the center of an inhomogeneous and evolving galaxy, and the supply of stars that can interact with it is limited.

In a fixed spherical galaxy, stars can interact with the binary only if their pericenters lie within $\sim R \times a$, where $R$ is of order unity. Let $L_{lc} = Ra \sqrt{2[E - \Phi(Ra)]} \approx \sqrt{2Gm_{12}Ra}$, the angular momentum of a star with pericenter $Ra$. The “loss cone” is the region in phase space defined by $L \leq L_{lc}$. The mass of stars in the loss cone is

$$M_{lc}(a) = m * \int dE \int_0^{L_{lc}} dL N(E, L^2)$$
\[ m_\star \int dE \int_0^{L_2} dL^2 \frac{4\pi^2}{12} f(E, L^2) P(E, L^2) \]
\[ \approx 8\pi^2 Gm_{12} m_\star R a \int dE f(E) P_{\text{rad}}(E). \]  

(25)

Here \( P \) is the orbital period, \( f \) is the number density of stars in phase space, and \( N(E, L^2) dEdL \) 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 \( M_{lc}(a_h) \), the mass within the loss cone when the binary first becomes hard; this is an upper limit since some stars 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 \( \Phi(a_h) \), equation (25) implies

\[ M_{lc}(a_h) \approx 3R\mu. \]

(26)

We can compute the change in \( a \) that would result if the binary interacted with this entire mass, by using the fact the mean energy change of a star interacting with a hard binary is \( \sim 3G\mu/2a \) [177]. Equating the energy carried away by stars with the change in the binary’s binding energy gives

\[ \frac{3}{2} G\mu \frac{dM}{a} \approx \frac{Gm_1 m_2}{2} \left( \frac{1}{a} \right) \]  

or

\[ \ln \left( \frac{a_h}{a} \right) \approx 3 \frac{\Delta M}{m_{12}} \approx \frac{9R\mu}{m_{12}} \approx \frac{9R}{1 + q^2} \]

(28)

if \( \Delta M \) is equated with \( M_{lc} \). Only for very low mass ratios \( (q \lesssim 10^{-3}) \) is this decay factor large enough to give \( t_{gr} < 10^{10} \text{ yr} \) (eq. 8), but the time required for such a small black hole to reach the nucleus is likely to exceed a Hubble time [132]. Hence even under the most favorable assumptions, the binary would not be able to interact with enough mass to reach gravity-wave coalescence.

But the situation is even worse than this, since not all of the mass in the loss cone will find its way into the binary. The time scale for the binary to shrink is comparable with stellar orbital periods, and some of the stars with \( r_{peri} \approx a_h \) will only reach the binary after \( a \) has fallen below \( \sim a_h \). We can account for the changing size of the loss cone by writing

\[ \frac{dM}{dt} = \int_{E_0(t)}^{\infty} \frac{1}{P(E)} \frac{dM_{lc}}{dE} \]
\[ = 8\pi^2 Gm_{12} m_\star R a(t) \int_{E_0(t)}^{\infty} f_i(E) dE, \]

(29)

where \( M(t) \) is the mass in stars interacting with the binary and \( f_i(E) \) is the initial distribution function; setting \( P(E_0) = t \) reflects the fact that stars on
orbits with periods less than $t$ have already interacted with the binary and been ejected. Combining equations (27) and (29),

$$\frac{d}{dt} \left( \frac{1}{a} \right) \approx 24\pi^2 R G m_* \int_{E_0(t)}^{\infty} f_i(E) dE. \quad (30)$$

Solutions to equation (30) show that a binary in a singular isothermal sphere galaxy stalls at $a_h/a \approx 2.5$ for $m_2 = m_1$, compared with $a_h/a \approx 10$ if the full loss cone were depleted (equation 28). In galaxies with shallower central cusps, decay of the binary would stall at even greater separations.

### 4.3 Collisional Loss-Cone Replenishment

A binary black hole depletes its loss cone very quickly since stars within the loss cone need only a few close encounters with the binary to be ejected. Whether the binary can continue to exchange energy with stars depends on the efficiency with which stars are re-supplied to the loss cone.

The most commonly invoked mechanism for loss cone re-filling is two-body scattering of stars. A small angular momentum perturbation, for instance from a passing star, can deflect a star with $L > L_{\text{lc}}$ into the loss cone. This process has been studied in detail in the context of scattering of stars into the tidal disruption sphere of a single black hole [56, 119, 27]. The basic equations are similar in the case of scattering into a binary SBH, except that the critical angular momentum increases by a factor $\sim \sqrt{a/r}$, where $r_t$ is the tidal disruption radius. (Other differences are discussed below.) If the binary parameters are assumed fixed, a steady-state flow of stars into the loss cone will be achieved on roughly a two-body relaxation time scale $T_R$, and the distribution function near $L_{\text{lc}}$ will have the form

$$f(E, L) \approx \frac{1}{\ln(1/R_{\text{lc}})} \overline{f}(E) \ln \left( \frac{R}{R_{\text{lc}}} \right), \quad (31)$$

where $R$ is a scaled angular momentum variable, $R \equiv L^2/L_c^2(E)$, $L_c(E)$ is the angular momentum of a circular orbit of energy $E$, and $\overline{f}$ is the distribution function far from the loss cone, assumed to be isotropic. The mass flow into the central object is $m_* \int \mathcal{F}(E) dE$, where

$$\mathcal{F}(E) dE = 4\pi^2 L_c^2(E) \left\{ \int \frac{d\nu}{v_r} \lim_{R \to 0} \frac{(\Delta R)^2}{2R} \right\} \overline{f} \ln(1/R_{\text{lc}}) dE. \quad (32)$$

The quantity in brackets is the orbit-averaged diffusion coefficient in $R$.

A crude estimate of the collisional re-supply rate is given by

$$\dot{M}_* \approx \frac{M_*(r < r_{\text{crit}})}{T_R(r_{\text{crit}})} \quad (33)$$
where $M_*(r)$ is the mass in stars within radius $r$ and $r_{\text{crit}}$ is the critical radius at which stars scatter into the loss cone in a single orbital period; beyond $r_{\text{crit}}$, the diffusion rate drops rapidly with radius. Estimates based on simple galaxy models give $r_{\text{crit}} \approx 10 - 100a$. To get an idea of the scattering rate, we consider the nucleus of the Milky Way. Equation (4) gives $a_h \approx 0.32(1 + q)^{-1}$ pc. Assuming $a = a_h$, $q = 0.1$ and $r_{\text{crit}} \approx 30a_h$ gives $r_{\text{crit}} \approx 1.1$ pc. The mass within this radius is $\sim 3 \times 10^6 M_\odot [67]$ and the relaxation time at this radius, assuming stars of a solar mass, is $\sim 2 \times 10^9$ yr. The scattered mass over $10^{10}$ yr is then $\sim 10^7 M_\odot$. This is comparable to the mass of the Milky Way SMBH, $M \approx 3 \times 10^6 M_\odot [193]$, but the scattering rate would drop as the binary shrinks, suggesting that scattered stars would contribute only modestly to refilling of the loss cone. In more massive galaxies, the nuclear density is lower, relaxation times are longer, and collisional refilling would be even less important. A more detailed calculation of the collisional refilling rates in real galaxies [228] concludes that few if any binary SBHs could reach coalescence via this mechanism. The author took the presently observed luminosity profiles of galaxies as initial conditions. The stellar density profiles must have been steeper before the binary SBH formed [150], leading to substantially more rapid decay early in the life of the binary.

Another criticism of standard loss cone theory is its assumption of a quasi-steady-state distribution of stars in phase space near $L_{\text{lc}}$ [151]. This assumption is appropriate at the center of a globular cluster, where relaxation times are much shorter than the age of the universe, but is less appropriate for a galactic nucleus, where relaxation times almost always greatly exceed a Hubble time [44]. (The exceptions are the nuclei of small dense systems like the bulge of the Milky Way.) The distribution function $f(E, L)$ immediately following the formation of a hard binary is approximately a step function,

$$f(E, L) \approx \begin{cases} f(E), & L > L_{\text{lc}} \\ 0, & L < L_{\text{lc}}, \end{cases}$$  \hspace{1cm} (34)$$
much steeper than the $\sim \ln L$ dependence in a collisionally relaxed nucleus (equation 31). Since the transport rate in phase space is proportional to the gradient of $f$ with respect to $L$, steep gradients imply an enhanced flux into the loss cone. Figure 6 shows the evolution of $N(E, R)$ at a single $E$ assuming that the loss cone is empty initially within some $R_{\text{lc}} \equiv L_{\text{lc}}/L_{\text{c}}(E)$ and that $N(E, R, t = 0)$ is a constant function of $R$ outside of $R_{\text{lc}}$. (The loss cone boundary is assumed static; in reality it would shrink with the binary.) Also shown is the collisionally-relaxed solution of equation (31). The phase-space gradients decay rapidly at first and then more gradually as they approach the steady-state solution. The total mass consumed by the binary, shown in the lower panel of Fig. 6, is substantially greater than would be computed from the steady-state theory, implying greater cusp destruction and more rapid decay of the binary. This time-dependent loss cone refilling might be particularly effective in a nucleus that continues to experience mergers or accretion events, in such a way
that the loss cone repeatedly returns to an unrelaxed state with its associated steep gradients.

Figure 6: (a) Slices of the density \( N(E,R,t) \) at one, arbitrary \( E \), recorded, from left to right, at \( 10^0, 10^1, 10^2, 10^3 \), and \( 10^4 \) Myr (solid curve). Initially, \( N(E,R,t) = 0 \) for \( R \leq R_{lc} \) and \( N(E,R,t) = \text{const} \) for \( R > R_{lc} \). We also show the equilibrium solution of equation (31) (dot-dashed curve). (b) The total number of stars consumed by the loss cone as a function of time (solid curve). The scale has been set to galaxy M32 with initial separation between the MBHs of 0.1 pc. (From [151])

There are other differences between loss cones around single and binary SBHs. A star that interacts with a massive binary generally remains inside the galaxy and is available for further interactions. In principle, a single star can interact many times with the binary before being ejected from the galaxy or falling outside the loss cone; each interaction takes additional energy from the binary and hastens its decay. Consider a simple model in which a group of \( N \) stars in a spherical galaxy interact with the binary and receive a mean energy increment of \( \langle \Delta E \rangle \). Let the original energy of the stars be \( E_0 \). Averaged over a single orbital period \( P(E) \), the binary hardens at a rate

\[
\frac{d}{dt} \left( \frac{Gm_1m_2}{2a} \right) = m_* \frac{N(\Delta E)}{P(E)}. \tag{35}
\]
In subsequent interactions, the number of stars that remain inside the loss cone scales as $L_{\Delta}^2 \propto a$ while the ejection energy scales as $\sim a^{-1}$. Hence $N(\Delta E) \propto a^4a^{-1} \propto a^0$. Assuming the singular isothermal sphere potential for the galaxy, one finds

$$\frac{a_{h}}{a(t)} \approx 1 + \frac{\mu}{m_{12}} \ln \left[ 1 + \frac{m_{12}N(\Delta E)}{2\mu\sigma^2} \frac{t - t_{h}}{P(E_0)} \right]$$

(36)

[151]. Hence the binary’s binding energy increases as the logarithm of the time, even after all the stars in the loss cone have interacted at least once with the binary. Re-ejection would occur differently in nonspherical galaxies where angular momentum is not conserved and ejected stars could miss the binary on their second passage. However there will generally exist a subset of orbits defined by a maximum pericenter distance $\lesssim a$ and stars scattered onto such orbits can continue to interact with the binary.

As these arguments suggest, the long-term evolution of a binary SBH due to interactions with stars may be very different in different environments. (We stress that the presence of gas may substantially alter this picture; cf. §8.) There are three characteristic regimes (we stress that the presence of gas may substantially alter this picture; cf. section 8) [151].

1. **Collisional.** The relaxation time $T_R$ is shorter than the lifetime of the system and the phase-space gradients at the edge of the loss cone are given by steady-state solutions to the Fokker-Planck equation. The densest galactic nuclei may be in this regime. Resupply of the loss cone takes place on the time scale associated with scattering of stars onto eccentric orbits. The decay time of a binary SBH scales as $|a/\dot{a}| \sim m_{*}^{-1} \sim N$ with $N$ the number of stars. In the densest galactic nuclei, collisional loss cone refilling may just be able to drive a binary SBH to coalescence in a Hubble time. For sufficiently small $T_R$, scattering refills the loss cone in less than an orbital period (“full loss cone”) and the decay follow $a^{-1} \sim t$. $N$-body studies are typically in this regime, as discussed below.

2. **Collisionless.** The relaxation time is longer than the system lifetime and gravitational encounters between stars can be ignored. The low-density nuclei of bright elliptical galaxies are in this regime. The binary SBH quickly interacts with stars whose pericenters lie within its sphere of influence; in a low-density (spherical or axisymmetric) nucleus, the associated mass is less than that of the binary and the decay tends to stall at a separation too large for gravitational wave emission to be effective. However evolution can continue due to re-ejection of stars that lie within the binary’s loss cone but have not yet escaped from the system. In the spherical geometry, re-ejection implies $|a/\dot{a}| \sim (1 + t/t_0)/a$, leading to a logarithmic dependence of binary hardness on time. Re-ejection in galactic nuclei may contribute a factor of $\sim$ a few to the change in $a$ over a Hubble time.

3. **Intermediate.** The relaxation time is of order the age of the system or somewhat longer. While gravitational encounters contribute to the re-population
Table 2: Physical Regimes for Long-Term Decay of Massive Black Hole Binaries

<table>
<thead>
<tr>
<th>Form of Decay</th>
<th>Regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^{-1} \propto \text{const}$</td>
<td>Collisionless</td>
</tr>
<tr>
<td>$\propto t/N$</td>
<td>Collisional (diffusion)</td>
</tr>
<tr>
<td>$\propto t + \text{const}$</td>
<td>Collisional (full loss cone)</td>
</tr>
<tr>
<td>$\propto \ln(1 + t/t_0) + \text{const}$</td>
<td>Re-ejection</td>
</tr>
</tbody>
</table>

of the loss cone, not enough time has elapsed for the phase space distribution to have reached a collisional steady state. Most galactic nuclei are probably in this regime. The flux of stars into the loss cone can be substantially higher than predicted by the steady-state theory, due to strong gradients in the phase space density near the loss cone boundary produced when the binary SBH initially formed. This transitory enhancement would be most important in a nucleus that continues to experience mergers or infall, in such a way that the loss cone repeatedly returns to an unrelaxed state with its associated steep gradients.

Table 1 summarizes the different regimes. The evolution of a real binary SBH may reflect a combination of these and other mechanisms, such as interaction with gas. There is a close parallel between the final parsec problem and the problem of quasar fueling: 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.

4.4 Non-axisymmetric nuclei

The estimates made above were based on spherical models of nuclei. The total number of stars in a full loss cone can be much larger if the nucleus is flattened and axisymmetric [124], when only one component of the angular momentum is conserved. In very flattened nuclei (with ellipticities $\epsilon \sim 0.5$), single emptying of an initially full loss cone can in some cases be sufficient to drive the binary to coalescence [228]. However loss cone dynamics can be qualitatively different in non-axisymmetric (triaxial or bar-like) potentials, since a much greater number of stars may be on "centrophilic" – box or chaotic – orbits which take them arbitrarily near to the SBH(s) [159, 68, 196, 212, 228]. Stars on centrophilic orbits of energy $E$ experience pericenter passages with $r_{\text{peri}} < d$ at a rate $\sim A(E)d$ [143]. If the fraction of stars on such orbits is appreciable, the supply of stars into the binary’s loss cone will remain essentially constant, even in
the absence of collisional loss-cone refilling. Such models need to be taken seriously given recent demonstrations [169, 90, 170] that galaxies can remain stably triaxial even when composed largely of centrophilic orbits. Furthermore imaging of galaxy centers on parsec scales reveals a wealth of features in the stellar distribution that are not consistent with axisymmetry, including bars, nuclear spirals, and other misaligned features [221, 164, 40].

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

$$\frac{dM_\star}{dt} \approx R_a \int A(E) M_c(E) dE$$

(37)

where $M_c(E)dE$ is the mass on centrophilic orbits in the energy range $E$ to $E + dE$. In a nucleus with $\rho \sim r^{-2}$, the implied feeding rate into a radius $r_{\text{inf}}$ is roughly

$$\dot{M}_\star \approx f_c \sigma^3 \frac{\sigma}{G}$$

(38)

$$\approx 2500 M_\odot \text{yr}^{-1} f_c \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^3,$$

(39)

where $f_c$ is the fraction of stars on centrophilic orbits. If this rate were maintained, the binary would interact with its own mass in stars in a time of only $\sim 10^5$ yr, similar to the decay time estimated above (eq. (23) for a binary in a fixed background. In fact, the feeding rate would decline with time as the centrophilic orbits were depleted. Solving the coupled set of equations for $a(t)$ and $M_c(t)$, one finds that at late times, the binary separation in a $\rho \propto r^{-2}$ nucleus varies as [170]

$$\frac{a_h}{a} \approx 3 \times 10^{w} f_c^2 \left( \frac{\sigma}{200 \text{ km s}^{-1}} \right)^3 \left( \frac{m_{12}}{10^8 M_\odot} \right)^{-1} \left( \frac{t}{10^{10} \text{yr}} \right).$$

(40)

Comparison with Table 1 shows that this is the same time dependence as for the “full loss cone” regime of spherical nuclei. Placing just a few percent of a galaxy’s mass on centrophilic orbits is sufficient to overcome the final parsec problem and induce coalescence, if the stellar density profile is steep and if the chaotic orbits are present at all energies. This example is highly idealized, but shows that departures from axial symmetry in galactic nuclei can greatly affect the rate of decay of a binary SBH.

5 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 center. The lighter SBH is ejected with a velocity roughly \( 1/3 \) the relative orbital velocity of the binary \([191, 92]\), and the binary recoils with a speed that is lower by \( m_3/(m_1 + m_2) \). Each close interaction of the smaller SBH with the binary increases the latter’s binding energy by \( \langle E/E \rangle \approx 0.4 m_3/(m_1 + m_2) \) \([89]\). 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 gravity 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 \([166]\) and assuming a fixed potential for the galaxy \([214, 147, 215]\). 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), J. Makino and collaborators mention two \( N \)-body simulations of triple SBH systems at the centers 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 \( e \) compared with a circular-orbit binary with the same semi-major axis.

In a wide, hierarchical triple, \( m_3 \ll (m_1, m_2) \), the eccentricity of the dominant binary oscillates through a maximum value of \( \sim \sqrt{1 - 5 \cos^2 i/3}, |\cos i| < \sqrt{3/5} \), with \( i \) the mutual inclination angle \([107]\). One study \([21]\) estimates that the coalescence time of the dominant binary in hierarchical triples can be reduced by factors of \( \sim 10 \) via the Kozai mechanism.

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 \([217]\) 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 a significant fraction of nuclei could be left with no SBH, with an offset SBH, or with a SBH whose mass is lower than expected based on the $M - \sigma$ or $M - L_{\text{bulge}}$ relations.

There is a need for simulations of multiple-SBH systems that include both gravitational loss terms, accurate (regularized) interactions between the SBH particles, and the interactions of SBH particles with stars.

6 N-Body Studies of Binary Black Hole Evolution

The interaction of a massive binary with point perturbers at the center of a galaxy is a straightforward problem for $N$-body simulation. In principle, $N$-body studies can reveal both the long-term evolution of the binary, as well as the effect of the binary on its stellar surroundings. The latter can be compared with observed nuclear density profiles as a test of the theory (§7).

Unfortunately, unless great care is taken, $N$-body studies are likely to give misleading results. This follows from the result (Section 4.3) that time scales for two-body scattering of stars into the binary’s loss cone are of order $10^{10}$ yr or somewhat longer in real galaxies. In $N$-body simulations, relaxation times are shorter by factors of $\sim N/10^{11}$ than in real galaxies, hence the long-term evolution of the binary is likely to be dominated by spurious loss cone refilling, wandering of the binary, and other noise-driven effects.

$N$-body studies are most useful at characterizing the early stages of binary formation and decay, or simulating the disruptive tidal effects of a SBH on the nucleus of an infalling galaxy. Due to algorithmic limitations – primarily the difficulty of integrating galaxy models with high central concentrations – most such studies [35, 76, 126, 125, 158, 157, 86, 25, 128] have been based on galaxy models with unrealistically large cores.

Figure 7 is from the first [136] $N$-body simulation of galaxy mergers in which the pre-merger galaxies contained power-law nuclear cusps as well as massive particles representing the SBHs. These simulations were run using GADGET [202], a tree code with inter-particle softening, and were not able to accurately follow the formation and decay of the massive binary. The SBH in the larger galaxy was found to tidally disrupt the steep cusp in the infalling galaxy, producing a remnant with only slightly higher central density than that of the giant galaxy initially. This result helps to explain the absence of dense cusps in bright galaxies [55], and suggests that the central structure of galaxies can only be understood by taking into account the destructive influence of SBHs on the stellar distribution during mergers. Additional results, using a similar $N$-body code and a variety of mass ratios for the merging galaxies, were reported in [141].

Had these simulations been extended to longer times using a more accurate
Figure 7: Final density profiles from a set of 10:1 merger simulations in which each galaxy contained a black hole (a-d) and in which neither galaxy contained a black hole (e-g) [136]. The four thin curves in each frame correspond to four different pre-merger orbits. (a), (e) Space density of stars initially associated with the secondary galaxy; thick curves are the initial density profile. (b), (f) Space density of stars initially associated with the primary galaxy; thick curves are the initial density profile. (c), (g) Space density of all stars. Lower thick curves are the initial density profile of the primary galaxy, and upper thick curves are the superposition of the initial density profiles of the primary and secondary galaxies. Lines of logarithmic slope $-1$ and $-2$ are also shown. (d), (h) Logarithmic slope of the surface density profiles of the merger remnants. Thick curves correspond to the initial primary galaxy.

N-body code, the massive binary would have ejected stars via the gravitational slingshot and lowered the central density still more. This was first demonstrated [150] in an N-body study that used a tree code for the early stages of the merger, and NBODY6, a high-precision, direct-summation code [1], for the later stages, when the binary separation fell below the tree code's softening length. The pre-merger galaxies had steep, $\rho \sim r^{-2}$ density cusps and the mass ratio
was 1:1. These simulations were continued until the binary separation had decayed by a factor of $\sim 10$ below $a_h$. The initially steep nuclear cusps were converted to shallower, $\rho \sim r^{-1}$ profiles shortly after the SBH particles had formed a hard binary; thereafter the nuclear profile evolved slowly toward even shallower slopes as the massive binary ejected stars. As Fig. 5 illustrates, the stellar density around the binary drops very quickly after the binary reaches a separation $a \approx a_h$.

Figure 8: Lagrangian radii around each of the two SBH particles in an equal-mass merger simulation [150]. From bottom to top, the radii enclose $10^{-4}$, $10^{-3.5}$, $10^{-3}$, $10^{-2.5}$, $10^{-2}$, $10^{-1.5}$ and $10^{-1}$ in units of the mass of one galaxy before the merger. The binary becomes “hard” at $t \approx 11$, and very rapidly heats the surrounding stellar fluid, lowering the local density.

The hardening rate of the binary in these simulations was found not to be strongly dependent on the number of particles. This result was subsequently shown [151] to be due to the small $N$: stars were resupplied to the loss cone via collisions at a higher rate than they were being kicked out by the binary, ensuring a continuous supply of stars and allowing the binary to continue to shrink. While a qualitatively similar evolution may take place in some galaxies – for instance, loss cones in non-axisymmetric potentials can be continuously repopulated by stars on centrophilic orbits (§4.4) – collisional loss cone refilling is very unlikely to achieve anything like a full loss cone except in very small, dense galaxies. The long-term evolution of the binary in almost all published $N$-body simulations are therefore not representative of what one would expect in real galaxies.
Figure 9: 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 ([20]). 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.

The relevant dimensionless parameter is

$$q(E) \equiv \frac{(\delta J)^2}{J_{lc}^2},$$

where $\delta J$ is the change over one radial period in the angular momentum of a star on a low-$J$ orbit, and $J_{lc}$ is the angular momentum of an orbit at the edge of the binary’s loss cone. A value $q(E) \gg 1$ implies that the loss cone orbits at energy $E$ are re-populated at a much higher rate than they are depopulated by the binary, and the loss cone remains nearly full. A value $q \ll 1$ implies that the loss cone is essentially empty, and repopulation must take place diffusively, as stars scatter in from $J \gtrsim J_{lc}$. In real galaxies, $N$ is large and $\delta J$ is small, implying $q \ll 1$. Achieving $q \ll 1$ in N-body simulations requires large particle numbers, and/or a model for the galaxy that has an unrealistically low central density, so that the star-star relaxation time is long. Figure 9 shows a recent set of N-body simulations that does both [20]. The
massive binary was embedded in a Plummer [168] galaxy model, which has a core radius comparable to its half-mass radius; this model is very different from real galaxies but its very low degree of central concentration implies a long relaxation time and low rate of collisional loss-cone refilling. Large particle numbers were achieved, without sacrificing accuracy, by running the simulation on a parallel GRAPE cluster. The $N$-dependence of the binary’s hardening rate is clear; in the simulations with binary mass $M_1 = M_2 = 0.02 M_{gal}$, the $N$-dependence of the hardening rate is $s \equiv (d/dt)(1/a) \approx N^{-0.8}$, almost as steep as the $N^{-1}$ dependence predicted for an “empty” loss cone [151]. A similar study [128], based on King-model galaxies, found a similar result.

The Plummer-model initial conditions used in the simulations of Figure 9 were identical to those adopted in two other $N$-body studies based on a more approximate $N$-body code [178, 25]. Contrary to Figure 9, Chatterjee, Hernquist & Loeb (2003) found that the binary hardening rate “saturated” at values of $N \gtrsim 2 \times 10^5$, remaining constant up to $N \approx 4 \times 10^5$. They speculated that this was due to a kind of Brownian-motion-mediated feedback, in which the binary maintains a constant supply rate by modulating the local density of stars. However no supporting evidence for this model was presented; for instance, it was not demonstrated that the central density was actually regulated by the binary, or that the amplitude of the Brownian wandering increased with $N$ in the manner postulated. Furthermore these authors provided no plots showing the claimed $N$-dependence of the hardening rate. Chatterjee et al.’s conclusion, that “a substantial fraction of all massive binaries in galaxies can coalesce within a Hubble time,” is not substantiated by the more accurate $N$-body simulations shown in Figure 9.

While Brownian motion probably does affect the decay rate of binaries in $N$-body simulations [151], it is doubtful that the effect is significant in real galaxies. The Brownian velocity of single black holes is found in $N$-body integrations to be [114]

$$1/2 M \langle V^2 \rangle \approx 3/2 m_* \dot{\sigma}^2 \quad (42)$$

where $\dot{\sigma}^2$ is the 1D, mean square stellar velocity within a region $r \lesssim 0.5 r_h$ around the black hole (and includes the influence of the black hole on the stellar motions), and $m_*$ is the stellar mass. In the case of the Milky Way black hole, equation (42) implies $V_{\text{rms}} \approx 0.2$ km s$^{-1}$ (assuming $m_* = M_\odot$) and an rms displacement of $\lesssim 0.1$ pc. Brownian motion of a massive binary is larger than that of a single black hole, but only by a modest factor [133, 150, 128]. The rms displacement of a binary from its otherwise central location would therefore be very small in a real galaxy, probably even less than the separation between the two components of the binary.

The goal of $N$-body studies is to simulate binary evolution in galaxies with realistic density profiles, and with large enough particle numbers that repopulation of the binary’s loss cone takes place diffusively, as in real galaxies. Two avenues are open for making further progress in this area. 1. One could
Figure 10: Results from a set of $N$-body integrations of a massive binary in a
galaxy with a $\rho \sim r^{-0.5}$ density cusp [205]. Each curve is the average of a set of
integrations starting from different random realizations of the same initial conditions.
(a) Evolution of the “mass deficit” (Eq. 43), i.e. the mass in stars ejected
by the binary. For a given value of binary separation $a$, the mass deficit is nearly
independent of particle number $N$, implying that one can draw conclusions from
observed mass deficits about the binary that produced them. (b) Evolution of bi-
nary eccentricity. The eccentricity evolution is strongly $N$-dependent and tends
to decrease with increasing $N$, suggesting that the eccentricity evolution in real
binaries would be modest.

combine a Monte-Carlo treatment of stellar encounters [57, 58] with a lookup
table, derived from scattering experiments, of energy and angular momentum
changes experienced during close passages of stars to the binary [178, 133]. Such
a hybrid algorithm would allow one to adjust the degree of collisionality at will
and record the effects on both the binary’s evolution, and the influence of the
binary on the stellar distribution. This approach would be difficult to generalize
to non-spherical geometries however. 2. A straightforward $N$-body approach is
also feasible, but particle numbers in excess of $\sim 10^7$ are required [151]. Such
large particle numbers are just now becoming feasible for direct-summation $N$-
body codes, by combining the GRAPE accelerator boards [127] with a parallel
architecture [33]. Indeed just such an approach was used for the integrations of
Figure 9, although the galaxy models in that study were rather unphysical.

Figure 10 shows a promising early step in this direction. The initial galaxy
models had $\rho \sim r^{-0.5}$ density cusps; integrations were carried out on a GRAPE-
6 computer, limiting the total particle number to $\sim 256K$, but the motion of the
black hole binary (of mass $M_1 = M_2 = 0.005M_{gal}$) and nearby stars was
carried out using the Mikkola-Aarseth chain regularization algorithm [145, 146, 3]. Because of the models’ higher central density and limited particle numbers, the binary’s loss cone was only partially empty, $q \gtrsim 1$, and the $N$-dependence of the hardening rate was shallower than expected in real galaxies, $(d/dt)(1/a) \sim N^{-0.4}$. Figure 10 (a) shows that the “damage” inflicted by the binary on the nucleus is not strongly dependent on $N$, as expected (cf. Eq. 18). This is an encouraging result since it implies that one can hope to learn something definite about pre-existing binaries by comparing $N$-body simulations with observations of the centers of current-day galaxies (§7). On the other hand, Figure 10 (b) suggests that the evolution of the binary’s eccentricity is strongly $N$-dependent. This may explain the rather disparate results on eccentricity evolution in past $N$-body studies [150, 86, 2].

Much progress on this problem is expected in the next few years.

7 Evidence for Cusp Destruction

A potentially powerful constraint on models of binary SBH evolution is the observed central density structure of galaxies. Figure 5 shows that a massive binary must eject order its own mass in reaching a separation at which $t_{gr} \lesssim 10^{10}$ yr if $m_2 \approx m_1$, or several times $m_2$ if $m_2 \ll m_1$. These numbers should be interpreted with caution since: (1) binaries might not decay this far – they may stall – or the final stages of decay might be driven by gas dynamics rather than energy exchange with stars; (2) the definition of “ejection” used in Figure 5 is escape of a star from an isolated binary, and does not take into account the confining effect of the nuclear potential; (3) the effect of repeated mergers on nuclear density profiles, particularly mergers involving very unequal-mass binaries, is poorly understood. Nevertheless, even the initial formation of a hard binary displaces a mass of order $m_2$ (Fig. 8).

The luminosity profile data can probably be used to rule out one model of binary evolution. In a “collisionless” galaxy (Table 1), the binary’s loss cone never refills, and decay of the binary would stall. The binary carves out a “hole” in both phase space and configuration space; the radius of the latter would be $\sim 3a_h$ [233]. While central minima may have been seen in the luminosity profiles of a few galaxies [113], these are likely due to dust obscuration, and the great majority of galaxies show a clearly rising stellar density into radii $\lesssim r_{infl}$. The non-existence of true “cores” suggests either that some degree of loss-cone refilling occurs, or that the final decay of the binary takes place via a more efficient process than ejection of stars.

Nevertheless there is a well-defined trend for the central densities of bright galaxies to decrease with increasing luminosity [52, 138, 44, 66, 79]. Nuclear densities in elliptical galaxies and spiral bulges with $M_V \lesssim -20$ follow $\rho \sim r^{-\gamma}$, $\gamma \lesssim 1$, while in fainter spheroids, $1 \lesssim \gamma \lesssim 2.5$. A natural interpretation is that the brightest galaxies – which presumably formed via one or more mergers –
have experienced more cusp destruction than fainter galaxies. (An alternative possibility, discussed below, is that the nuclei in faint galaxies re-formed after being destroyed.)

In practice, this hypothesis is difficult to test since it requires knowledge of the pre-merger density profiles. A reasonable guess is that all galaxies originally had steep power-law density cusps, since these are generic in the faintest galaxies known to harbor SBHs. For instance, both M32 and the bulge of the Milky Way have $\rho \sim r^{-1.5}$ at $r \lesssim r_{\text{infl}}$ and $\rho \sim r^{-2}$ just outside [112, 67].

Figure 11: Observed surface brightness profile of NGC 3348. The dashed line is the best-fitting Sersic model to the large-radius data. Solid line is the fit of an alternative model, the “core-Sersic” model, which fits both the inner and outer data well. The mass deficit is illustrated by the area designated “depleted zone” and the corresponding mass is roughly $3 \times 10^8 M_\odot$ [77].

The “mass deficit” [153] is defined as the difference in integrated mass between the observed density profile and the primordial (pre-merger) profile. For instance, if the primordial profile is a power law of index $\gamma_0$ inward of some radius $r_b$, then

$$M_{\text{def}} = 4\pi \int_0^{r_b} \left[ \rho(r_b) \left( \frac{r}{r_b} \right)^{-\gamma_0} - \rho(r) \right] r^2 dr.$$  \hspace{1cm} (43)
Mass deficits in samples of bright elliptical galaxies were computed in three recent studies [153, 181, 77]. In the first two studies, the authors assumed power-laws of various slopes for the pre-merger profiles, and found $\langle M_{\text{def}}/M_\bullet \rangle \approx 1$ for $\gamma_0 = 1.5$ with $M_\bullet$ the current mass of the SBH. The latter study made use of the fact that the light profiles of bright galaxies show an abrupt downward deviation relative to a Sersic [197] profile fit to the outer regions (Fig. 11). Mass deficits inferred in this study were slightly larger, $M_{\text{def}}/M_\bullet \approx 2$.1 These numbers are within the range predicted by the binary SBH model, particularly given the uncertainties associated with the effects of multiple mergers.

In small dense galaxies, a destroyed cusp would be expected to re-form via the Bahcall-Wolf [8, 171] process, on a timescale of order the star-star relaxation time measured at $r_{\text{infl}}$. This time is of order $10^9$ yr in the Milky Way bulge and the nucleus of M32. This may be the explanation for the steep power-law profiles observed at the centers of these galaxies. Alternatively, the steep cusps may be due to star formation that occurred after the most recent merger [97].

More rigorous tests of the binary SBH model will require a better understanding of the expected effect of massive binaries on stellar density profiles. As discussed above, while the best current $N$-body simulations suggest $\rho \sim r^{-1}$ following binary formation [150], the simulations are dominated by noise over the long term.

A number of other processes could compete with binary SBHs in the destruction of nuclear density cusps. A population of three or more SBHs in a galactic nucleus would undergo a complicated set of close encounters resulting ultimately in coalescence and/or ejection of some or all of the SBHs (§5). In the process, the stellar background would be heated and a mass of order five times the combined mass in SBHs removed [139]. This model reproduces the observed time dependence of core radii in globular clusters very well [142] but its relevance to galactic nuclei is less clear; the model requires binary coalescence times long enough that an uncoalesced binary is present when a third SBH falls in [217]. If a binary SBH does eventually coalesce, the gravitational radiation carries a linear momentum leading to a recoil of the coalesced hole [19, 53]. Recoil velocities are estimated to be as large as $\sim 400$ km s$^{-1}$ [48, 140], although with considerable uncertainty. A SBH ejected from a galactic nucleus with a velocity of $\sim 10^2$ km s$^{-1}$ would quickly fall back to the center, but its displacement and infall would heat the stellar fluid and lower its density. Figure 12 shows the effects of ejection on nuclear density profiles. Mass deficits produced by this mechanism can be comparable in amplitude to those predicted by the binary SBH model.

A major focus of future work should be to calculate the evolution of $\rho(r)$ as predicted by the various scenarios for binary decay discussed in this article.

1The author of this study presents his mass deficits as significantly smaller than those found in the earlier studies. However he bases his comparison on values of $M_{\text{def}}$ computed exclusively using $\gamma_0 = 2$. \[1\]
Figure 12: Effect on the nuclear density profile of SBH ejection. The initial galaxy model (black line) has a $\rho \sim r^{-1}$ density cusp. (a) Impulsive removal of the SBH. Tick marks show the radius of the black hole’s sphere of influence $r_{\text{infl}}$ before ejection. A core forms with radius $\sim 2r_{\text{infl}}$. (b) Ejection at velocities less than escape velocity. The black hole has mass 0.3% that of the galaxy; the galaxy is initially spherical and the black hole’s orbit remains nearly radial as it decays via dynamical friction. The arrow in this panel marks $r_{\text{infl}}$ in the initial galaxy. [140].

8 Interaction of Binary Black Holes with Gas

Interstellar gas might play an important role in the dynamical evolution of a binary SBH. Interactions with gas complement interactions with the stellar environment (§4) and with other SBHs (§5). Any gas situated close to a binary is disturbed by the SBHs and exerts gravitational force on them, thereby affecting their orbit. Furthermore, if SBH coalescence is accompanied by the presence of gas, an observable electromagnetic afterglow might follow coalescence.

The collisional, dissipative nature of interstellar gas gives rise to a behavior fundamentally different from that of the point-mass dynamics of stellar systems. It is natural to distinguish between two classes of flows in dynamical systems containing gas. In hot flows the gas temperature is comparable to the virial temperature of the system, while in cold flows the gas temperature is significantly below the virial temperature. The virial temperature can be defined as $T_{\text{vir}} = GM_1(\mu m_p/2ak)$, where $\mu$ is the mean particle mass in units of the proton mass $m_p$, and $k$ is the Boltzmann constant. The prototype of a hot flow is the
spherical, “Bondi” accretion onto a single black hole, in which the accreting gas is supported by pressure against free infall toward the accretor. The prototype of a cold flow is a thin disk, in which the gas is rotationally supported against infall. Even in hot flows rotational support is realized close to the accretor when the gas has nonzero net angular momentum (e.g. [109]).

The angular momentum barrier is central to SBH formation theories. Any model for how material is channeled into an accreting black hole must describe the mechanism by which angular momentum is removed from the material. Whatever this mechanism may be, it is expected that it operates universally during the epoch in which SBHs grew to their present masses by rapidly accreting material onto pre-existing black hole “seeds.” This is also the period when galaxy merging peaks [97, 83, 222, 216]. While still elusive to astronomical probes due to severe obscuration [190], the nuclei of merging galaxies, which are also the sites for the formation of binary SBHs [18], are expected to contain the largest concentration of dense gas anywhere in the universe. The inevitable abundance of gas motivates an inquiry into the role of gas dynamics as an alternative to stellar dynamics in the process of SBH coalescence. Some of the mechanisms that remove angular momentum of interstellar gas and thus channel it into the neighborhood of SBHs include the torquing of gas flow by the rapidly-fluctuating potential of merging galaxies [144] and by nested stellar bars [199], angular momentum transport by hydrodynamical turbulence that might be driven by the onset of self-gravity [198, 61, 74] or by supernovae embedded within a large-scale toroidal circumnuclear flow [218], angular momentum extraction by magnetohydrodynamical turbulence [11] or by magnetic braking [22], and more speculatively, by Rossby vortex instabilities [118].

Astronomical observations offer abundant evidence for both hot and cold gas flows in the immediate vicinity of SBH candidates. The origin and the dynamical impact of the two classes of gas flow are distinct and are discussed here separately.

8.1 Interaction with Hot Gas

Hot gas permeates the interstellar space in galaxies and the intergalactic space in groups of galaxies and galaxy clusters. Virial temperatures range between $10^6$ K – $10^8$ K and the hot gas is almost completely ionized. Primordial and secondary sources contribute to the pool of hot gas. During the early stages of galaxy formation, intergalactic space contains partially ionized gas inherited from the pregalactic, early universe. Hydrogen recombines at redshifts $z \sim 1000$ and is reionized at redshifts $z \sim 10$ by the radiation emitted by the earliest structures. The partially ionized gas cools within the confining gravitational potential of dark matter halos and filaments. Cold gas accelerates toward the halos’ centers of gravity and is shock-heated to about the virial temperature. Some of the coldest inflowing gas escapes heating by accreting along narrow channels that reach deep inside the primary halo. Cooling times in the halo
centers where the gas is the densest are short compared to the dynamical time and thus most of the primordial gas is consumed in starbursts on a dynamical time scale.

Tenuous gas that remains after the cooling time has exceeded the dynamical time in the nascent galaxy might still be plentiful enough to feed a massive black hole growing at an Eddington-limited rate. The residual number density at the radius of influence of the SBH is

\[
n \approx \frac{\sigma^3 kT}{GMA} \approx 20 \text{ cm}^{-3} \mu \left( \frac{M}{10^8 M_\odot} \right)^{0.11} \left( \frac{\Lambda}{2 \times 10^{-23} \text{ erg cm}^3 \text{s}^{-1}} \right)^{-1},
\]

(44)

where \( T \) is the virial temperature of the galaxy, \( \mu \) is the average atomic mass in units of the proton mass, \( \Lambda \) is the cooling function \([30]\), and we have employed the \( M - \sigma \) relation (Eq. 9) to relate the virial temperature to the black hole mass. The thermal stability limit could in principle be exceeded if the gas kept at the Compton temperature by a continuum flux from an unobscured AGN \([45]\).

This so-called “cooling flow model of quasar fueling” \([26, 160]\) is however plagued by many problems \([108]\ and references therein). Most of the gas left over from star formation might be blown out by the mechanical feedback associated with the radiative and mechanical output of the accreting massive black hole \([200, 100, 156]\). A small amount of angular momentum in the gas results in circularization and settling into an accretion disk. This disk may be susceptible to fragmentation, thereby converting most of the gas mass into stars and effectively cutting off the supply of gas to the SBH \([208]\).

The geometry of the flow of a hot, magnetized gas near a binary black hole is unknown. Assuming spherical, non-rotating accretion, the time scale on which the hot gas is captured by the SBH is

\[
t_{\text{capt}} \equiv \frac{M}{\dot{M}} \approx f_b G^2 M \frac{\sigma^3}{\mu m_p n} \approx 10^8 \text{ yr} f_b \mu^{-2} \left( \frac{M}{10^8 M_\odot} \right)^{-0.44} \left( \frac{\Lambda}{2 \times 10^{-23} \text{ erg cm}^3 \text{s}^{-1}} \right),
\]

(45)

where \( f_b \sim 1 - 10 \) is a numerical factor that depends on the equation of state of the gas.

If a binary black hole is present, gravitational torques from the gas induce decay of the binary’s semi-major axies on approximately the same time scale. This crude estimate is based on an analogy with binary-star interactions: the binary must eject of order its own mass in stars to decay an \( e \)-folding in separation. Hot gas torquing the binary might be ejected in an outflow and thus the
actual rate at which gas is accreting onto individual binary components might be severely suppressed compared to the accretion expected in an isolated black hole.

Galactic nuclei also contain hot gas produced by secondary sources. For example, observations with the Chandra X-ray Observatory have revealed tenuous ($n \approx 10 - 100$ cm$^{-3}$), hot ($T \approx 1$ keV) plasma within a parsec of the $\sim 4 \times 10^6 M_\odot$ Milky Way SBH [7]. This plasma is being generated by the numerous massive, evolved stars in the galactic region [67] through stellar wind and supernova activity. Since its temperature is higher than the virial, most ($> 99\%$) of the plasma escapes the neighborhood of the SBH [176]. While the hot gas densities in active galaxies might be transiently larger than that at the Galactic center, the tendency of the hot plasma to escape the neighborhood of the SBH reduces the likelihood that large quantities of virialized gas would remain enmeshed with the binary’s orbit long enough to affect its dynamical evolution.

Recently, Escala et al. [41, 42] carried out smoothed particle hydrodynamical (SPH) simulations of binary point masses interacting with a massive, spherical cloud of hot gas initially centered on the binary. Gravitational drag from the gas induces decay in the binary’s orbit. The relevance of spherical, hot initial conditions is contingent on the astrophysical plausibility that a compressed accumulation of hot gas comparable in mass to the SBH can be sustained.

8.2 Interaction with Cold Gas

The specific angular momentum of a cold flow might easily exceed that of the binary. The gas then tends to settle into rotationally supported, geometrically thin rings and disks (recall that “cold” gas is colder than the virial temperature but can be hot enough to be ionized).

Observations offer abundant evidence for the presence of dense gas in galactic nuclei. Thin, Keplerian molecular disks on scales $0.1 \text{ pc} - 0.5 \text{ pc}$ have been seen in the water maser emission in the nuclei of Seyfert galaxies [154, 60, 82]. The Galactic nucleus contains a $4 \times 10^6 M_\odot$ black hole surrounded by a $\sim (10^4 - 10^5) M_\odot$ molecular gas torus at distances $> 1 \text{ pc}$ from the SBH [94]. Compact stellar disks on scales $\geq 20 \text{ pc}$, which are fossil evidence of past gas circularization, are evident in the nuclei are of many galaxies [167]. Massive accretion disks must be present in quasars and the Narrow-Line Seyfert I nuclei to account for what appears to be rapid accretion onto the central SBHs in these systems. However, the structure of these disks at radii comparable to the size of a hard SBH binary is unknown. The principal challenges to constructing extended disk models are the instabilities related to incomplete ionization and the susceptibility to gravitational fragmentation [103, 198, 149, 75].

If a disk surrounding a binary SBH is initially inclined with respect to the binary’s orbital plane, the quadrupole component of the binary’s gravitational
potential causes differential precession in the disk at the rate [111]

\[ \Omega_{\text{prec}}(r) = \frac{3}{4} \frac{q}{(1 + q)^2} \frac{(GM_{12})^{1/2}a^2}{r^{7/2}}, \]  

which results in a warping of the disk. As in the Bardeen-Petterson mechanism [15, 173, 194], the warp either dissipates, or smears around the binary, resulting ultimately in a nearly axisymmetric disk in the binary’s orbital plane.

Interest in co-planar, circumbinary disks stems from their ability to extract a binary’s angular momentum via a form of tidal coupling. Two interrelated questions might be posed: What is the response of a circumbinary disk to the binary’s tidal forcing? and: How does such a disk affect the evolution of the binary’s orbit?

Existing attempts to answer these questions have employed ad hoc models for the form of the binary-disk torque coupling [172, 93], or have been restricted to binaries with components of very unequal mass [4], where an array of neighboring Lindblad resonances facilitate binary-disk coupling [72], much like the coupling between a massive planet and its natal gas disk [73]. Early numerical simulations of circumbinary disks with nearly equal masses [5], however, suggested that the disks are truncated exterior to the resonances, which was interpreted as a consequence of a collisionless nonlinear parametric instability [189, 39]. Fluid dynamical theory of circumbinary disk truncation is still lacking.

In a circular binary the outer Lindblad resonances (OLR) are located at radii \( r_m = (1 + 1/m)^{2/3} \), where \( m = 1, 2, \ldots \) is the order in the decomposition of the binary’s gravitational potential into multipoles:

\[ \varphi(r, \theta) = \sum_{m=0}^{\infty} \varphi_m(r) \cos[m(\theta - \Omega_{\text{bin}}t)]. \]  

The outermost OLR is located at \( r \approx 1.6a \). The resonances are radii in the disk where the natural, epicyclic frequency of radial oscillations in the disk is an integer multiple of the rate at which a packet of disk gas receives tidal “kicks” by the binary. The forcing near a resonance, as well as at a radius where surface density in the disk exhibits a large gradient, excites nonaxisymmetric propagating disturbances, or “density waves,” in the disk.

The gravitational potential of eccentric binaries contains low-frequency components that are absent in circular binaries. These low-frequency components activate resonances located at larger radii than in the circular case, and might lead to mutual excitation and reinforcement of the binary and the disk eccentricities [162, 71]. Many extrasolar planets, which are thought to form in circumstellar disks, are notably eccentric\(^2\), suggesting that dynamical coupling between a binary point mass (a star and a planet, or a pair of black holes) and a gas disk is conducive to eccentricity growth. The observed circumbinary

\(^2\)http://exoplanets.org
disks in young stellar binaries such as GG Tau, which are typically eccentric, are truncated at radii a few times the semimajor axis [130], which lends support to this hypothesis. Eccentricity in SBH binaries accelerates coalescence due to gravitational wave emission (Eq. 7) and might be detectable in gravitational wave trains.

Density waves transport angular momentum outward through the circumbinary disk. Angular momentum flux carried by the waves is extracted from the binary’s angular momentum. The binary experiences a negative torque equal and opposite to the total angular momentum flux transferred to the disk. The location of the inner edge of the disk reflects a balance between the angular momentum flux deposited into the disk, and the angular momentum flux transported through the disk by another, possibly viscous mechanism. Wave momentum is deposited into the disk material via a form of dissipative damping. The location in the disk where the waves are damped can be separated by many wavelengths from the location where they are excited. The damping could take place in the nonlinear steepening and the breaking of wave crests [192, 179]. In marginally optically thick disks, radiation damping might also play a role [24]. Yet another form of damping could be due to the dissipation of wave shear if the disk is strongly viscous [207]. The amplitude of the density waves is a steeply decreasing function of the radius of excitation. The amplitude is diminished if the waves are nonlinear at excitation and damp in situ, but then one expects the inner edge to recede where in situ damping shuts off.

The intricate and insufficiently understood nature of binary-disk interactions calls for grid-based hydrodynamical simulations with a shock-capturing capability. The necessity that the radial wavelength, which is smaller than the vertical scale height of the disk, be resolved by multiple cells, places severe demands on the computational resources, especially if a three dimensional representation of the disk is required. It should also be noted that the radiative and thermal structure of accretion disks around single SBHs are not adequately understood on any radial scale.

As a binary’s semimajor axis decreases due to stellar, gas dynamical, or gravitational radiation processes, a circumbinary disk’s inner edge spreads inward viscously while maintaining constant edge-to-semimajor axis ratio, e.g., \( r_{\text{edge}}/a \sim 2 \). In the final stages of the gravitational radiation-driven inspiral, however, the time scale on which the semimajor axis decays becomes shorter than the viscous time scale, and the disk can no longer keep up with the binary, resulting in binary-disk detachment. On the relevant length scales the disk might be dominated by radiation pressure and the electron scattering opacity; the structure and the stability of such disks is an active research area [210].
9 Spin Evolution during Mergers

Coalescence of a binary black hole results in a spinning remnant.\footnote{In this section, we set $G = c = 1$.} Angular momentum conservation implies

$$S_1 + S_2 + L_{\text{orb}} = S + J_{\text{rad}}$$

(48)

where $S_1$ and $S_2$ are the spin angular momenta of the two SBHs just before the final plunge, $L_{\text{orb}}$ is the orbital angular momentum of the binary before the plunge, $S$ is the spin of the resulting black hole, and $J_{\text{rad}}$ is the angular momentum carried away by gravitational waves during and after the coalescence [54]. The simplest case to treat is extreme mass ratio mergers, $q \equiv m_2/m_1 \ll 1$, for which the binary can be described as a test particle of mass $m_2 \equiv m$ orbiting a black hole of mass $m_1 \equiv M \gg m$, and both $S_2$ and $J_{\text{rad}}$ can be ignored. The change in the larger hole’s spin is computed by adding the smaller hole’s energy and orbital angular momentum at the last stable orbit (LSO). The latter varies from $L_{\text{LSO}}/m = \sqrt{12}M$ for circular equatorial orbits around a non-spinning hole to $L_{\text{LSO}} = M(9M)$ for prograde (retrograde) orbits around a maximally-spinning hole, $S_1 = M^2$. The much larger value of $L_{\text{LSO}}$ in the case of retrograde capture implies that a rapidly-rotating hole will typically spin down if capture occurs from random directions [34, 70, 226, 227, 91]. The change in spin assuming $q \ll 1$ is

$$\delta \hat{a} = q \left( -2\hat{a} + \hat{L}_{\text{LSO},z} \right) + O(q^2)$$

(49)

where $\hat{a} \equiv |S_1|/M^2$, $L_z$ is the orbital angular momentum parallel to $S$, and $\hat{L} \equiv L/mM$. The first term in equation (49) describes conservation of spin angular momentum of the larger hole as its mass grows, $\hat{a} \propto M^{-2}$, while the second term describes the increase in spin due to torquing by the smaller body.

The change in spin after a single coalescence is illustrated in Fig. 13 as a function of $q$ and initial spin; the upper (lower) curves represent prograde (retrograde) captures from equatorial orbits, and the dashed lines are for capture over the pole. The bias toward spin-down is evident; retrograde capture from the equatorial plane produces a nearly (but never completely) non-spinning remnant when $q \approx 2.5\hat{a}$, $q \leq 0.23$, and rapid final rotation ($\hat{a} \gtrsim 0.9$) requires both a large initial spin and a favorable inclination. On the other hand, if the larger hole is slowly rotating initially, $\hat{a} \lesssim 0.5$, mass ratios $q \gtrsim 0.3$ always result in spin-up. The oft-repeated statement that “mergers spin down black holes” reflects a preconception that SBHs are likely to be formed in a state of near-maximal rotation (e.g., [14, 62]).\footnote{Note the error in Figure 1 of Hughes & Blandford (2003), which shows the change in spin for mergers with mass ratio $q = 0.5$: the darkest contour on that plot should be labelled $\hat{a} = 0.5$, not $\hat{a} = 0$.}

Successive mergers from random directions with fixed $q$ (i.e. secondary mass grows proportionately to primary mass) lead to a steady-state spin distribution
Figure 13: Final spin $\hat{a}$ of a remnant black hole in terms of its original spin, for mass ratios $q = 0.1$ (red), 0.3 (green) and 0.5 (blue). The change in spin was computed using the test-particle approximation for $L_{\text{LSO}}$ [16, 91, 227]. Upper (lower) curves correspond to prograde(retrograde) capture from the equatorial plane; dashed curves are for capture over the pole. Capture of a low-mass secondary is likely to spin down the larger hole unless the latter is slowly rotating initially. Capture of a massive secondary results in spinup unless infall is nearly retrograde or the original spin is large.

$N(\hat{a})$ that is uniquely determined by $q$. For small $q$, this distribution can be derived from the Fokker-Planck equation [91]:

$$N(\hat{a})d\hat{a} \approx N_0\hat{a}^2 e^{-3\hat{a}^2/2\hat{a}_{\text{rms}}^2} d\hat{a}, \quad \hat{a}_{\text{rms}} \approx 1.58\sqrt{q}.$$  (50)

Figure 14 shows $N(\hat{a})$ for various values of $q$, computed via Monte-Carlo experiments (not from the Fokker-Planck equation) using the test-mass approximation for $L_{\text{LSO}}$. The Gaussian form of equation (50) is seen to be accurate only for $q \lesssim 0.1$. For $q \gtrsim 1/8$, the distribution is skewed toward large spins.

Accurate calculation of spin-up during a merger of comparably massive black holes requires a fully general-relativistic numerical treatment. Adopting various approximations for the radius of the innermost stable circular orbit (ISCO) for comparably massive binaries [28, 17, 81, 31], and assuming that mass and
Figure 14: Steady-state spin distributions produced by successive capture from random directions at fixed mass ratio $q$, for $q = (1/32, 1/16, 1/8, 1/4, 1/2)$. Curves were generated using Monte-Carlo experiments based on the test-particle approximation, $q \ll 1$; hence the curve for $q = 1/2$ should be viewed as illustrative only.

Angular momenta are conserved during coalescence, gives a remnant spin in equal-mass mergers of $\dot{a} \approx 0.8 - 0.9$. Baker et al. [9, 10] present full numerical calculations of equal-mass mergers with and without initial spins. In the absence of initial spins, 3% of the system’s mass-energy and 12% of its angular momentum are lost to gravitational radiation, and the final spin is $\dot{a} \approx 0.72$. Coalescence of initially spinning holes from circular orbits in the equatorial plane yields $\dot{a} \approx 0.72 + 0.32\dot{s}$ with $\dot{s}$ the initial spin parameter of the two holes (assumed equal); Baker et al. considered initial spins in the range $-0.3 \leq \dot{s} \leq 0.2$, where negative/positive values indicate spins aligned/couteraligned with the orbital angular momentum. Extrapolating this result toward $\dot{s} = 1$ suggests that prograde mergers of black holes with initial spins $\dot{s} \geq 0.85$ will result in a maximally-spinning remnant.

Confronting these predictions with observation is problematic for a number of reasons: merger histories of observed SBHs are not known, SBH spins are difficult to determine observationally, and other mechanisms, such as gas accretion, can act efficiently to spin up SBHs [14]. However there is circumstantial evidence that mergers played a dominant role in determining the spins of at least some SBH. If SBH spins were the product of gas accretion, the jets in active
galaxies should point nearly perpendicularly to the disks of their host galaxies [182]. In fact, there is almost no correlation between jet direction and galaxy major axis in Seyfert galaxies [211, 101, 195]. Among the possible explanations [101] for the misalignment, perhaps the most natural is that SBH spins were determined by the same merger events that formed the bulge, long before the formation of the gaseous disk, and that subsequent spin-up by gas accretion from the disk plane has been minimal [134].

Coalescence of two black holes during a merger should result in a “spin-flip,” a reorientation of the spin axis of the more massive black hole. In the test-particle limit, the reorientation angle is

$$\delta \theta = \frac{q}{a} \hat{L}_{LSO} \sqrt{1 - \mu^2} + O(q^2)$$ (51)

with $\mu$ the cosine of the angle between the orbital angular momentum vector and the spin axis of the larger hole. When $q \geq 0.2$, the spin orientation is overwhelmed by the plunging body in a retrograde merger, even if the initial spin of the larger hole was close to maximal. Hence, even “minor mergers” (defined, following galactic dynamicists, as mergers with $q \leq 0.3$) are able to produce a substantial reorientation. In fact there is a class of active galaxies which exhibit radio lobes at two, nearly-orthogonal orientations, and in which the production of plasma along the fainter lobes appears to have ceased [115, 32]. These “X-shaped” or “winged” radio galaxies, of which about a dozen are known, are plausible sites of recent (within the last $\sim 10^8$ yr) black-hole coalescence [137, 233]. Furthermore the implied coalescence rate is roughly consistent with the expected merger rate for the host galaxies of luminous radio sources [137]. Alternative models have been proposed for the X-shaped sources, including a warping instability of accretion disks [174], backflow of gas along the active lobes [116], and binary-disk interactions before coalescence [120].

It is likely that all of these mechanisms are active at some level and that the time scale for realignment influences the radio source morphology, with the most rapid realignments producing the classical X-shaped sources, while slower realignment would cause the jet to deposit its energy into a large volume, leading to an S-shaped FRI radio source [137].

If the black holes are spinning prior to coalescence, they will experience spin-orbit precession, on a time scale that is intermediate between $t_{gr}$ and the orbital

---

5Liu (2004) criticized the black hole coalescence model on the grounds that “calculations based on general relativity show that the change in inclination of a rotating central SMBH is negligible in a minor merger and a significant reorientation of the active SMBH requires a comparatively rare major merger (Hughes & Blandford 2003).” This erroneous statement probably had its origin in the final sentence of the Hughes & Blandford paper, which states that “An abrupt change in inclination...requires a comparatively rare major merger.” Hughes & Blandford defined a “major merger” as having a mass ratio $q \geq 0.1$, in conflict with the standard definition among galactic dynamicists, $q \geq 0.3$. In fact Hughes & Blandford conclude, in agreement with Merritt & Ekers (2002), that mass ratios exceeding $\sim 0.2$ can result in spin-flips.
period. To PN2.5 order, the spin angular momentum of either hole evolves as

$$\frac{dS_1}{dt} = \frac{1}{a^3} \left[ \left( 2 + \frac{3m_2}{2m_1} \right) L_{\text{orb}} - S_2 + 3 (\hat{n} \cdot S_2) \hat{n} \right] \times S_1$$  \hfill (52)

where $\hat{n}$ is a unit vector in the direction of the displacement vector between the two black holes. The evolution equation for $S_2$ is given by interchanging the indices. The magnitude of each spin vector remains fixed (to this order), and each spin precesses around the total angular momentum vector $J = L_{\text{orb}} + S_1 + S_2$. When the two black holes are comparably massive, the orbital angular momentum greatly exceeds the spin angular momentum of either hole until just prior to coalescence. As the binary shrinks, the spins have a tendency to unalign with $L_{\text{orb}}$. Ignoring spin-spin effects, the precession rate for equal-mass holes in a circular orbit is

$$\Omega_1 = \frac{7}{2} \frac{m_1^{1/2}}{a^{5/2}} \mu.$$ \hfill (53)

The precession rate is lower than the orbital frequency:

$$\frac{\Omega_{\text{orb}}}{\Omega_1} = 2 \frac{a}{7 \mu} = \frac{16}{7} \frac{a}{R},$$ \hfill (54)

where $R = 2m_{12}$, but higher than the radiation reaction time scale:

$$\frac{1}{2\pi} \Omega_1 t_{\text{gr}} = \frac{35}{1024\pi} \frac{a^{3/2}}{m_{12}^{3/2}} = \frac{35\sqrt{2}}{512\pi} \left( \frac{a}{R} \right)^{3/2},$$ \hfill (55)

unless the binary is close to coalescence. If the dominant source of energy loss during the late stages of infall is gravitational radiation, the spin direction will undergo many cycles of precession before the black holes coalesce. This does not seem to happen in the X-shaped radio sources, based on the apparently sudden change in jet direction; if the coalescence model for the X-sources is correct, the final stages of infall must occur on a shorter time scale than $t_{\text{gr}}$. This might be seen as evidence that shrinkage of the binary is usually driven by gas dynamics, not gravitational radiation losses, prior to the final coalescence.

When the black hole masses are very different, $q \ll 1$, the ratio of spin of the larger hole to $L_{\text{orb}}$ is

$$\frac{S_1}{L_{\text{orb}}} = \frac{f_1}{q} \left( \frac{m_1}{a} \right)^{1/2} = \frac{f_1}{q} \left( \frac{R}{a} \right)^{1/2}$$ \hfill (56)

where $S_1 = f_1m_1^2$. The two quantities are approximately equal when the separation measured in units of the larger hole’s Schwarzschild radius is equal to $q^{-2}$. When this separation is reached, the binary orbit rapidly changes its plane, and a new regime is reached where the spin of the smaller black hole precesses
about the spin of the larger hole. The precession rate of the larger hole is given by
\[ \Omega_1 = \frac{2m_1^{1/2}m_2}{a^{3/2}} \] (57)
in both regimes, and
\[ \frac{\Omega_{\text{orb}}}{\Omega_1} = \frac{2a}{R}, \quad \frac{1}{2\pi} \Omega_1 t_{\text{gr}} = \frac{5}{128\pi} \left( \frac{a}{R} \right)^{3/2}. \] (58)

The spin direction of a black hole formed via binary coalescence is also affected by torques that reorient the binary prior to coalescence. The role of torques from gaseous accretion disks was discussed above; another source of torques is perturbations from passing stars or gas clouds [134]. A single star that passes within a distance \( \sim 3a \) of the binary will exchange orbital angular momentum with it, leading both to a change in the binary’s orbital eccentricity as well as a change in the orientation of the binary’s spin axis, as discussed above. Referring to equations (17) and (20), the reorientation rate is related to the hardening rate via
\[ \langle \Delta \theta^2 \rangle = \frac{L}{H m_{12}} t^{-1}_{\text{harden}} \] (59)
where \( t^{-1}_{\text{harden}} = a(d/dt)(1/a) \). Scattering experiments [134] give \( L/H \approx 4 \) for a hard, equal-mass binary. The implied change in the binary’s orientation after shrinking from \( a \approx a_h \) to \( a \approx 10^{-3}a_h \) is
\[ \delta \theta \approx \sqrt{\frac{30m_*}{m_{12}}}. \] (60)

The reorientation begins to be significant if \( m_*/m_{12} \gtrsim 10^{-3} \), which may be the case for intermediate-mass black holes.

10 Summary

1. No completely convincing case of a bound, binary SBH has yet been discovered. “Dual” SBHs, i.e. two, widely-separated SBHs in a single system, have been seen in some interacting galaxies and binary quasars. However none has a projected separation less than \( \sim 1 \) kpc. Strong limits can be placed on the binarity of the Milky Way SBH.

2. The evolution of binary SBHs in gas-poor galaxies is dominated by the gravitational slingshot ejection of stars that pass near the binary, carrying away energy and angular momentum. Once the binary has ejected all stars on intersecting orbits, continued hardening depends on a refilling of the binary’s “loss cone.” Possible mechanisms for loss-cone refilling include star-star gravitational scattering, chaotic orbits in non-axisymmetric potentials, and perturbations from additional massive objects.
3. Most $N$-body simulations of binary evolution have been based on such small particle numbers that the binary’s loss cone was refilled at a spuriously high rate by gravitational encounters, Brownian motion of the binary and other finite-$N$ effects. As a result, many results from these studies cannot usefully be extrapolated to the large-$N$ regime; in particular, they do not make useful predictions about the binary hardening rates expected in real galaxies.

4. $N$-body studies do appear to make robust predictions about the “mass deficit,” the mass in stars ejected from a galactic nucleus by an evolving binary. Observed mass deficits are of order $1 - 2$ times the SBH mass, consistent with $N$-body predictions.

5. Studies of the interaction of binary SBHs with gas are in their infancy. Major uncertainties are the amount, distribution and thermodynamic state of gas very near the centers of galaxies containing massive binaries.

6. Binary coalescence can have a large influence on SBH spins; even mass ratios as extreme as $10:1$ can substantially spin up or reorient a SBH. Evidence for spin-flips may have been observed in the so-called $X$-shaped radio sources.

11 Acknowledgements

This work was supported by grants AST-0071099, AST-0206031, AST-0420920 and AST-0437519 from the NSF, grant NNG04GJ48G from NASA, and grant HST-AR-09519.01-A from STScI. M. M. was supported at Caltech by a post-doctoral fellowship from the Sherman Fairchild Foundation.
References


58


