Massive black hole binary mergers within sub-pc scale gas discs

J. Cuadra1, P. J. Armitage1,2, R. D. Alexander3,1, M. C. Begelman1,2
1JILA, University of Colorado and National Institute of Standards and Technology, Boulder, CO 80309-0440, USA
2Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder CO 80309, USA
3Leiden Observatory, Universiteit Leiden, Niels Bohrweg 2, 2300 RA, Leiden, the Netherlands

ABSTRACT
We study the efficiency and dynamics of supermassive black hole binary mergers driven by angular momentum loss to small-scale gas discs. Such binaries form after major galaxy mergers, but their fate is unclear since hardening through stellar scattering becomes very inefficient at sub-parsec distances. Gas discs may dominate binary dynamics on these scales, and promote mergers. Using numerical simulations, we investigate the evolution of the semi-major axis and eccentricity of binaries embedded within geometrically thin gas discs. Our simulations directly resolve angular momentum transport within the disc, which at the radii of interest is likely dominated by disc self-gravity. We show that the binary decays at a rate which is in good agreement with analytical estimates, while the eccentricity grows. Saturation of eccentricity growth is not observed up to values $e > 0.35$. Accretion onto the black holes is variable, and is roughly modulated by the binary orbital frequency. Scaling our results, we analytically estimate the maximum rate of binary decay that is possible without fragmentation occurring within the surrounding gas disc, and compare that rate to an estimate of the stellar dynamical hardening rate. For binary masses in the range $10^5 M_\odot < M < 10^8 M_\odot$ we find that decay due to gas discs may dominate for separations below $a \sim 10^{-3}$ pc, in the regime where the disc is optically thick. The minimum merger time scale is shorter than the Hubble time for $M < 10^7 M_\odot$. This implies that gas discs could commonly attend relatively low mass black hole mergers, and that a significant population of binaries might exist at separations of a few hundredths of a pc, where the combined decay rate is slowest. For more massive binaries, where this mechanism fails to act quickly enough, we suggest that scattering of stars formed within a fragmenting gas disc could act as a significant additional sink of binary angular momentum.

Key words: black hole physics – accretion: accretion discs – galaxies: nuclei – galaxies: active

1 INTRODUCTION
The assembly of present-day galaxies occurs via the hierarchical merger of lower mass progenitors, many of which seem likely to have supermassive black holes (SMBHs) in their nuclei. Following a merger of two galaxies, each harbouring a black hole, we expect the SMBHs to sink toward the centre of the merger product and eventually form a black hole binary. What happens subsequently is less clear. There is no doubt that at sufficiently small separations ($a \sim 10^{-3}$ pc) loss of angular momentum to gravitational radiation occurs rapidly enough to effect coalescence. However, at larger separations an as-yet undetermined mix of angular momentum loss to stars and to gas is required to bring the black holes into the gravitational radiation inspiral regime. Observations provide scant information. Detecting black hole binaries on pc or sub-pc scales (where the black holes’ gravity dominates the galactic potential) is extremely difficult, with the closest confirmed binary in the radio galaxy 0402+379 (Rodriguez et al., 2006) having a projected separation of 7.3 pc. A variety of circumstantial evidence – of which the most compelling is probably the small observed scatter in the relationships between black hole mass and galaxy properties (Ferrarese et al., 2006; Gebhardt et al., 2000) – suggests that binaries do merge, though these arguments do

* Present address: Shanghai Astronomical Observatory, Shanghai 200030, China; jcuadra@shao.ac.cn

© 2008 RAS
not constrain the time scale or the mechanism that facilitates mergers.

One longstanding motivation for considering the role of gas in hastening binary mergers has come from the fact that stellar dynamical processes may simply fail. Although dynamical friction against the stellar background is rapid for large separations (Milosavljević & Merritt, 2003), for pc-scale binaries the rate slows dramatically as the black holes eject the finite number of stars that have orbits that come close enough to the centre to interact with the binary. Replenishment of stars on these loss cone orbits occurs on the relaxation time, which typically exceeds the Hubble time (Begelman et al., 1980; Yu, 2002; Merritt, 2006). This 'last parsec' problem occurs for spherical or axisymmetric nuclei, and is particularly severe for massive galaxies which are observed to possess rather low density stellar cores (Faber et al., 1997). More recently, interest in this long standing problem has been further piqued by the realization that if gas is responsible for driving mergers, then it is likely that some of that material will survive in the intermediate vicinity of the holes at the moment of coalescence. Plausibly, some small fraction of the energy released during the merger may go into heating the gas, producing an electromagnetic precursor (Armitage & Natarajan, 2002) or afterglow (Milosavljević & Phinney, 2005; Lippai et al., 2008; Schnittman & Krolik, 2008; Shields & Bonning, 2008) to the gravitational wave event. Detection of an electromagnetic counterpart would greatly improve the accuracy with which space-based gravitational wave observatories such as LISA can localize mergers, and allow much more information to be gathered about the properties of the host galaxy (Kocsis et al., 2007). It may also be possible to determine whether gas discs typically attend mergers by searching for periodic electromagnetic signals produced by binaries at larger separations, independent of any gravitational wave information Haiman et al. (2008).

Theoretically, it is well established that the dissipative nature of gas allows inflow down to pc scales in galactic nuclei following galactic mergers (Mihos & Hernquist, 1996; Springel et al., 2005; Mayer et al., 2007; Levine et al., 2008). Perhaps more surprisingly, smaller amounts of gas appear to be able to penetrate deep within the black hole's sphere of influence even in galaxies such as the Milky Way that are far removed from any significant merger activity. Evidence for such inflows comes from observations of young stars with disc-like kinematics close to Sgr A* (Paumard et al., 2006; Lu et al., 2008), which may well have formed in situ from gas at sub-pc scales (Levin & Beloborodov, 2003; Nayakshin & Cuadra, 2005). Taken together, these lines of argument suggest that it is plausible to invoke the existence of gas discs at small radii as a common feature of galactic nuclei, though the typical masses and structure of such discs are subject to considerable uncertainty. Prior theoretical work has established that gas discs are likely to be important for SMBH binary mergers, provided that the mass of gas in the disc is roughly of the same order of magnitude as the mass in black holes (Gould & Rix, 2000; Armitage & Natarajan, 2002; Escala et al., 2005; Dotti et al., 2007; MacFadyen & Milosavljević, 2008; Hayasuki, 2008).

In this paper we revisit the interaction between a SMBH binary and a relatively low mass circumbinary gas disc. Our main innovation is to study this interaction using numerical simulations that directly resolve the physical mechanism of angular momentum transport within the gas. At the radii of greatest interest for the SMBH binary merger problem (typically tens of a pc, where the stellar dynamical time scales are longest for many galaxies) it is expected that self-gravity will dominate the dynamics of surrounding gas discs (Shlosman et al., 1990; Goodman, 2003). The onset of self-gravity in an accretion disc can result either in a quasi-stable disc with outward angular momentum transport, or in fragmentation of the disc into stars. Either possibility is of great interest for its effect of SMBH binary mergers. Here, we study directly the first regime in which the disc remains stable and angular momentum is transported by the action of self-gravitating spiral arms. Physically this occurs if the cooling time in the disc exceeds the local dynamical time (Gammie, 2001). We use the simulations to study the rate at which the gas drives the binary toward merger, and the effect that the gas disc has on the eccentricity of the binary. Previous simulations that have used either artificial viscosity or a Navier-Stokes formulation to model angular momentum transport have found that eccentricity growth (both of the binary, and within the gas disc) attends the decay of the binary semi-major axis (Artymowicz et al., 1991; Armitage & Natarajan, 2005; MacFadyen & Milosavljević, 2008). The time dependent accretion signatures that result may permit the identification of black hole binaries within gas rich systems prior to the final coalescence phase.

2 INITIAL CONDITIONS AND NUMERICAL METHOD

In this paper we present simulations that follow the evolution of a binary that initially has a total mass $M$, separation $a_0$, and orbital frequency $\Omega_0 = \sqrt{GM/a_0^3}$. The mass ratio between the two components is set to $q = M_1/M_2 = 3$, as typically expected for a major galaxy merger (Volonteri et al., 2003). We try two different initial orbital configurations for the binary: a circular orbit of radius $a_0$ and an eccentric orbit with semi-major axis size $a_0$ and eccentricity $e = 0.1$.

Around the binary we set a co-rotating circular disc of gas. The disc is aligned with the orbital angular momentum of the binary, has a total mass of $M_\text{d} = 0.2M$ and extends initially from $2a_0$ to 5$a_0$. The surface density of the disc decreases with radius $R$ (measured from the binary centre of mass) as $\Sigma(R) \propto R^{-1}$. The disc dimensionless thickness is initially constant $H/R = 2M_\text{d}/M$. We set the initial internal energy of the gas accordingly as $u = (H/R)^2 v_K^2/(\gamma(\gamma-1))$, where $v_K = \sqrt{GM/R}$ is the Keplerian velocity around the binary and $\gamma = 5/3$ is the adiabatic index.

Although the use of a Navier-Stokes viscosity is an improvement over a reliance on purely numerical effects to transport angular momentum, it too may lead to unphysical behaviour. In particular, it is known that a disc subject to only a Navier-Stokes shear viscosity is unstable to the growth of eccentricity even in the absence of perturbations (Ogilvie, 2001). In general, there is no reason to assume that a ‘turbulent viscosity’ resulting from a physical process such as self-gravity or the magnetorotational instability (Balbus & Hawley, 1998; Balbus & Papaloizou, 1999) will behave in the same way as a microscopic fluid viscosity.
Figure 1. Logarithmic maps of the disc column density (in units of $M_{\odot}^{-2}$) at different times during the simulation. The panel at $t = 0$ shows the smooth initial conditions. Until $t \approx 350\Omega_{-10}$, material piles up at $R \approx 3a_0$ as a result of the torques shown in Fig. 3, forming a dense ring. The ring breaks due to its self-gravity (see Fig. 4), spreading the gas approximately over the original radial range. A spiral pattern develops, and the disc stays in that state until at least $t \approx 1200\Omega_{-10}$, when the simulation ends.

abatic index of the gas. The disc stability to self-gravity is measured by the Toomre parameter $Q = \Omega c_s/ (\pi G \Sigma)$, which translates to $Q \approx (H/R)(M/M_d)$ for a disc in hydrostatic equilibrium. In our case, the thickness of the disc ensures that the disc is not initially unstable to self-gravity, as the Toomre parameter is $\approx 2$. Only after cooling has affected the gas, will the disc get thinner and become unstable.

To calculate the evolution of the system we use the smoothed particle hydrodynamics (SPH; e.g., Monaghan, 1992) code GADGET-2 (Springel, 2005). The code solves for the adiabatic hydrodynamics and gravitational forces of the system, and includes a term for artificial viscosity, necessary to treat shocks.

On top of the basic hydrodynamics set-up we add cooling. This is implemented by defining a cooling time and setting the radiative cooling term for each gas particle to be $(du/dt)_{\text{cooling}} = -u/t_{\text{cool}}$. The cooling time is set to be proportional to the orbital time of the gas around the central binary, $t_{\text{cool}} = \beta/\Omega$, with $\beta$ a constant and $\Omega = \sqrt{GM/F^3}$. This prescription is commonly used in the literature (e.g., Gammie, 2001; Rice et al., 2005; Nayakshin et al., 2007; Alexander et al., 2008), since the condition for disc fragmentation corresponds to a constant $\beta$. As we will discuss
later, in a real disc changes to the surface density and opacity with radius result in a value of $\beta$ that varies with radius – we include this physics in our analytic estimates but not in the numerical work. In this paper we concentrate on the case where the cooling is not fast enough to make the disc fragment, therefore we set the value of $\beta = 10$. We defer the regime of faster cooling and subsequent star formation, $\beta \lesssim 5$, to future study.

The gravity among gas particles is calculated using a Barnes–Hut tree, as implemented in GADGET-2. However, since higher accuracy is required when calculating the evolution of the binary orbit we compute the gravitational forces on the SMBHs directly, by summing exactly the forces from all gas particles instead of using a tree approximation. Maintaining symmetry, the gravitational attraction from the SMBHs is added directly to each particle.

While most of the gas is expected to remain in the disc, accretion streams allow matter to flow from the inner edge of the disc onto the black holes, where it becomes tightly bound. To prevent the computation from halting due to the accretion of particles that pass within the sink radius of $0.1 a_0$ of either black hole (Cuadra et al., 2006). The mass and momentum of the accreted gas are added to the respective black hole, thereby conserving the linear and angular momentum of the system $^2$. When we discuss the ‘accretion rate’ onto the black holes, the reader should be aware that it is this accretion of particles that pass within the sink radius that is the actual quantity being computed. The physical accretion rate would track the numerical one provided that the residence time of gas through the (unmodeled) disc at smaller radii is shorter than the characteristic time scale of accretion variations.

We run each calculation for more than a thousand binary dynamical times using 2 million SPH particles to model the gas. In addition, to test the numerical convergence of our results, we run shorter simulations with 8 million particles.

3 EVOLUTION OF THE DISC

All simulations display very similar disc evolution, so in this section we concentrate on the run with 2 million gas particles and the binary on an initially circular orbit. Figures 1 and 2 show the evolution of the surface density $\Sigma$ of the disc. Figure 1 shows maps of $\Sigma$ at 6 selected times, while Fig. 2 shows the azimuthally-averaged $\Sigma(R)$ profiles at the initial conditions and every $100 \Omega_0^{-1}$ in the interval $50 \leq t \Omega_0 \leq 550$.

The disc has initially a density profile $\Sigma \propto R^{-1}$, as seen in the solid line of Fig. 2. The sharp boundaries are quickly relaxed and become smoother. Figure 3 shows the different contributions to the torque per unit radius acting at this early stage in the simulation. The gravitational potential of the binary produces a torque that oscillates as a function of radius, as described in detail by MacFadyen & Milosavljević (2008). On top of that there is a large component produced by the disc self-gravity, and smaller contributions from gas pressure and artificial viscosity. The total torque has a minimum at $R/a_0 \approx 3$, where material piles up and eventually a dense ring forms. There are other minima in $dG/dR(R)$ – notably one at $R/a_0 \approx 2$ – but they occur at locations where not much gas is present, preventing the formation of other density enhancements. From the figure it is clear that the torques produced by hydrodynamical forces (both pressure forces, which could be physically important, and artificial viscosity which is a numerical concern) are negligible compared with the gravitational torques. The artificial viscosity torque becomes somewhat important only at the edges of the disc, due to the fact that shocks form at these locations.

As seen in Fig. 2, the gas piles up at $R \approx 3a_0$, until the

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]

\[ \frac{dG}{dR} \]
column density becomes so high that the Toomre parameter $Q = \Omega_c/(\pi G \Sigma)$, plotted azimuthally averaged in Fig. 4, becomes smaller than unity and the ring collapses due to its self-gravity. That process can be seen in the panels 2–4 of Fig. 1, where the ring forms, develops asymmetric structure, and finally collapses after reaching values $Q \approx 0.5$. The gas that was in the ring then disperses in radius and reaches a quasi-steady state characterised by a complicated spiral pattern and a value of $Q \approx 1.5$. It is interesting to note that the critical value of $Q$ is different when the gas accumulates in a ring, at the early stage of the simulations, compared to the case where the gas is spread over a larger radial range, later in the evolution of the disc. This shows that the disc behaviour is not completely determined by its (azimuthally averaged) local properties. Instead, its stability seems to depend on the properties of the gas at neighbouring annuli and on its azimuthal distribution (see, e.g., Lodato, 2007).

To characterise the spiral structure that forms in the disc, we measure its Fourier modes, defined as

$$C_m = \frac{1}{M_d} \int_{R_m}^{R_{out}} R dR \int_0^{2\pi} d\phi \Sigma(R, \phi) \exp(-im\phi).$$

Figure 5 shows the evolution of the $m = 1, 2, 3, 4$ modes as a function of time. No mode dominates the structure of the disc after the initial transient period.

We also measure the eccentricity of the disc, defined as

$$e(R) = \frac{\int d\phi \Sigma(R, \phi) v_R \exp(i\phi)}{\int \Sigma(R, \phi) v_\phi}.$$  

Figure 6 shows the eccentricity profile both at the beginning of the simulation, and when we stopped it, 1200 time units later. Both profiles are built averaging over 20 snapshots, to minimise the random fluctuations. The initial conditions of the simulation had the gas rotating in circular orbits around the centre of mass of the system. However, the non-Keplerianity of the potential makes the gas rapidly acquire a small non-zero eccentricity, an effect that is obviously larger closer to the binary. After the collapse of the ring, the eccentricity of the disc reaches values $e \approx 0.1$, but then it decays again to smaller values. As shown in the figure, the eccentricity profile remains approximately the same as initially until the end of the run, $t \approx 1200\Omega_0^{-1}$, even though the binary reaches $e \approx 0.1$ (Sect. 4). In a different run where the binary reached $e \approx 0.3$ the disc did not acquire a significant eccentricity either. These results are consistent with those of MacFadyen & Milosavljević (2008), whose disc only became eccentric after a few thousand binary orbital times.

To illustrate the long-term evolution of the disc shape, Fig. 7 shows a column density map averaging 100 snapshots at $t \approx 700\Omega_0^{-1}$, in a frame corotating with the binary. The
due to self-gravity as a viscosity which is a power-law on surface density and radius,

$$\nu \propto \Sigma^A R^B.$$  \hspace{1cm} (3)

The parameters $A; B$ are determined by the cooling mechanism in the disc. In our case the cooling time is proportional to the dynamical time-scale of the disc, $t_{cool} = \beta/\Omega$ (Section 2), which sets the viscosity parameter to $\alpha = 0.4/\beta$, (e.g., Rice et al., 2005). That, together with the condition $Q \approx Q_0 \approx 1.5$ (Fig. 4), gives $A = 2, B = 9/2$, independent of the values of $\beta$ and $Q_0$.

The other important parameter is

$$S = \frac{2\pi a^2 \Sigma_0(a)}{M_2},$$  \hspace{1cm} (4)

which is roughly the mass ratio between the disc and the secondary black hole. $\Sigma_0$ is defined as the surface density the disc would have in the absence of the secondary. Over the relatively short time scales of the simulations self-gravity results in relatively modest (less than an order of magnitude) changes to the disc surface density. To a reasonable approximation we can then extrapolate the $\Sigma(R)$ profile we use as initial conditions to $R = a$ and get $\Sigma_0(a) \approx 0.01 M a^{-2}$. That, together with the mass of the secondary, $M_2 = 0.25 M$, gives $S \approx 1/4$.

With the parameters obtained above, we follow Ivanov et al. (1999) calculation for the evolution of the surface density of a circumbinary disc, under the assumption that the torque of the binary only affects a narrow ring in the inner part of the disc. We obtain that the time-scale for the black holes to merge is, for the case of a self-gravitating disc,

$$t_{disc} = \frac{1}{\sqrt{0.933} \ 0.81\alpha Q_0^{1/2} G^{1/2} \Sigma_0^2(a_0 a_0^3)^{5/2}} \approx \frac{1}{\sqrt{M^2} 19.3\alpha Q_0^{-2} G^{1/2} \Sigma_0^2(a_0 a_0^3)^{7/2}}.$$  \hspace{1cm} (5)

Using the parameters of our model, we estimate that $t_{disc} \approx 3 \times 10^4 \Omega_0^{-1}$.

### 4.2 Numerical results

Figure 8 shows the evolution of the binary orbital elements from our simulations. The orbit of the binary shrinks initially at a rate $da/dt \sim -10^{-4} a_0 \Omega_0$. The orbital decay gets somewhat stalled during the accumulation of material in a ring ($t \approx 350 \Omega_0^{-1}$, see Fig. 1), but it recovers after the ring breaks. Practically the same rate of change is found for a binary that is initially in an orbit with $e = 0.1$. That particular simulation was continued until $t \approx 3300 \Omega_0^{-1}$; from that run it is clear that in the long term the rate of decay slows down, reaching $da/dt \sim -2 \times 10^{-5} a_0 \Omega_0$. We attribute this to the fact that the gas disc continuously absorbs angular momentum from the binary and in average moves farther away from it – as this process goes on, it is increasingly hard for the gas to influence the binary. The order of magnitude of the decay rate is in good agreement with the theoretical expectations from the previous section. \footnote{Due to the long computing times required to complete these simulations, we have not yet explored the parameter space of different}

---

**Figure 7.** Density map averaged over $t\Omega_0 = 650–750$, well after the initial transient evolution, plotted in a frame co-rotating with the binary. The disc shows no persistent asymmetry.

Image shows that the disc develops no persistent asymmetry, except for the streams taking material from the inner part of the disc to the black holes.

---

### 4 EVOLUTION OF THE ORBIT

The binary starts in a circular orbit of radius $a_0$. However, as the binary exerts a positive total torque on the gaseous disc, it transfers its angular momentum to it, making the binary shrink. Less obviously, the coupling between the disc and the binary results in the excitation of binary eccentricity, even (as in our runs) in the case where the gas disc itself remains approximately circular.

#### 4.1 Analytic expectations

The rate of angular momentum transfer between a binary and a surrounding accretion disc can be calculated analytically (Pringle, 1991). Most relevant for our purposes is the work of Syer & Clarke (1995) and Ivanov et al. (1999), who calculated the evolution of a binary on a circular orbit, under the effect of an circum-binary $\alpha$ disc. They neglected the self-gravity of the disc, and assumed that the mass of the secondary black hole to be much smaller than the primary. Although our conditions violate these assumptions, their calculations can provide us with an order-of-magnitude estimate of the effect we expect to obtain.

In this section we follow Ivanov et al. (1999, their section 4.1) to estimate the time-scale for binary shrinking. We first need to express the angular momentum transport in the disc.
 Massive black hole binary mergers within sub-pc scale gas discs

Figure 8. Evolution of the binary orbit elements $a$ (top panel) and $e$ (bottom) for the different simulations. Solid lines show simulations with the binary in an initially circular orbit, while dashed lines show binaries with $e = 0.1$ initially. In both cases the eccentricity grows, and the orbital decay proceeds at the same rate. Thicker lines show shorter simulations with higher resolution that agree remarkably with the standard runs.

We also study the evolution of the binary eccentricity. When the orbit of the binary is initially circular, it remains so during the first part of the calculation, during which the disc is still mostly symmetric (Fig. 5). However, after the disc becomes more asymmetric, the eccentricity starts growing at a rate $de/dt \sim 1.5 \times 10^{-4} \Omega_0$. In a different simulation, where the binary orbit is initially eccentric, the eccentricity growth of the disc starts earlier, at $t \approx 150\Omega_0^{-1}$, but proceeds at approximately the same rate. As in the case of the binary size, the eccentricity growth slows down with time. It is interesting however to study the relation between eccentricity and binary separation. Figure 9 shows that $de/da$ remains approximately constant in the long simulation. We then conclude that the slower eccentricity growth at later times is mainly produced by the dearth of material to absorb the binary angular momentum. It is not due to the onset of any intrinsic damping mechanism, such as might occur if resonant damping grows to balance resonant excitation of eccentricity (Moorhead & Adams, 2008). If eccentricity growth saturates in our case, it does only for $e \gtrsim 0.35$.

The standard resolution of 2 million particles was adopted based on a comparison between the physical torques due to self-gravity and numerical torques due to code artificial viscosity. It is not obvious that a resolution that minimizes this numerical effect is also adequate to reproduce eccentricity growth accurately. Accordingly, we tested the numerical convergence of these results by running shorter simulations with 8 million particles, shown with thick lines in Figs. 8 and 9. These runs give the same results as our 2 million particle runs. We are then confident that our simulations capture properly the interaction between the disc and the binary.

4.3 Maximum rate of orbital decay

Ultimately, the main question that we wish to address is whether gas discs absorb binary angular momentum efficiently enough to drive the black holes toward coalescence on a short time scale. Evidently this will depend on the amount of gas that is typically present on very small scales within galactic nuclei. This quantity is not well known, either theoretically or observationally. We can, however, estimate the maximum rate of orbital decay that can be driven by a surrounding thin accretion disc. Such a limit can be derived because, although more gas results in more rapid orbital decay, the amount of gas in the disc cannot be increased arbitrarily. Eventually, a very massive disc will fragment into stars rather than remaining as a fluid structure, destroying the hydrodynamic interaction that leads to shrinkage\textsuperscript{4}.

To estimate the maximum rate of decay we relax the constant $\beta$ cooling time assumption that we introduced for

\textsuperscript{4} The newly formed stars, of course, could refill the loss cone and thereby help to shrink the binary via stellar dynamical processes. We defer a calculation of that process to future work.
numerical convenience. In a real disc the cooling rate will be set by the physical conditions of the disc. Some discs will cool faster and fragment into stars, failing to transport angular momentum as a disc. We can then compute the maximum decay rate by finding the disc solution that maximizes $|\text{d}T/\text{d}t|$, subject to the constraint that the disc is everywhere stable against fragmentation. We emphasize that the resulting disc model – which is marginally stable against fragmentation at every radius – is deliberately fine tuned and is not what one would expect as the outcome of gas inflow toward the nuclei. It suffices to yield a limit, but that limit is unlikely to be exactly realized in nature.

From the analyses of Syer & Clarke (1995) and Ivanov et al. (1999), it is clear that the orbital decay will be faster for discs that are more massive ($\dot{\Sigma}_{\text{disc}} \propto \Sigma^{-5/2}$, see eq. 5). The first step is therefore to calculate the maximum surface density, as a function of radius, that a disc can have without the first step is therefore to calculate the maximum surface density, as a function of radius, that a disc can have without fragmenting. We follow Levin (2007), and note that the fragmentation boundary in effect represents the maximum stress that can be sustained in a quasi-equilibrium self-gravitating disc, and can be expressed in terms a maximum allowed value of $\alpha$ (Rice et al., 2005). If in addition the disc opacity is a function of temperature only, then there is an analytic solution for a critically self-gravitating disc with $Q = 1$. Each value of the temperature corresponds to unique values of $\Omega$ and $\Sigma$, and together these specify the critical solution. We obtain

$$ T = \frac{2m_p}{k_b} \left( \frac{\pi G \Sigma}{\Omega} \right)^2, \quad (6) $$

$$ \tau = \frac{\kappa(T) \Sigma}{2}, \quad (7) $$

$$ \alpha = \frac{8\sigma_{\text{eq}}}{9} \left( \frac{2m_p}{k_b} \right)^4 \left( \frac{\pi G}{\Omega} \right)^6 \frac{\tau}{1 + \tau^2} \Sigma^5 \Omega^{-7}, \quad (8) $$

where the opacity $\kappa(T)$ is taken from the Semenov et al. (2003) opacity tables. The calculation is stopped when the temperature reaches $\sim 1300$ K, as the opacity drops sharply there, allowing in principle unreasonably large disc masses. Moreover, the critical value of $\alpha$ can be much larger in this regime (e.g., Johnson & Gammie, 2003), rendering this analysis invalid.

Once a maximum $\Sigma(R)$ is established, we can calculate the maximum rate of binary shrinking using eq. 5. Figures 10 and 11 show in solid lines this time-scale for systems with binary masses $M = 3 \times 10^5, 3 \times 10^6, 3 \times 10^7 M_\odot$, and mass ratios 1, 3, 9, as a function of binary separation. At large separations, the mass needed to make the disc fragment is small, and therefore the removal of angular momentum from the binary is not efficient. At short separations, much more mass is needed to make the disc fragment, so the process can be much faster.

The sharp break that appears in every curve is due to the transition from the optically thick regime at large separations to the optically thick one. At large distances and low temperatures, the opacity is dominated by ice grains and goes roughly as $\kappa = \kappa_{\text{ice}} T^2$, with $\kappa_{\text{ice}} \approx 2 \times 10^{-4} \text{cm}^2 \text{g}^{-1} \text{K}^{-2}$ (e.g., Bell & Lin, 1994). In addition, for a self-gravitating disc the disc properties are linked by $\tau \propto (\Sigma/\Omega)^2$, which gives an optical depth $\tau \propto \Sigma^5/\Omega^7$. Looking at eq. 8, it is clear that the relation between surface density and orbital frequency for a fixed $\alpha$ goes as $\Sigma \propto \Omega^{11/10}$ for optically thin discs, while there is no constraint for optically thick discs, as the $\Sigma$ dependence cancels out. This explains the sharp decline of the time-scale at the point where the disc becomes optically thick – arbitrarily large masses are in principle allowed, making the angular momentum transfer through the disc very efficient. This behaviour, however, breaks down once the gas reaches $T \approx 166$ K and the opacity law changes to a shallower power-law on temperature. Further changes of slope are seen in the curves at even shorter separations, which are due to substructure in $\alpha(T)$.

Overall, the time-scale for binary merger becomes shorter as the distance between the black holes is smaller, reaching a value of only $\sim 3 \times 10^6$ yr for a $3 \times 10^6 M_\odot$ binary at a separation of $\sim 0.02$ pc. The calculation is stopped at this point, as explained above, because the unperturbed disc would reach $T \approx 1300$ K.

For comparison we also show in Figs. 10 and 11 the time-scale for orbital decay produced by scattering stars, $t_{\text{stellar}}$, as estimated by Milosavljević & Merritt (2003), who...
of the binary orbital frequency,
\[ \Omega_{\text{hub}} = \left( \frac{4\pi^2 m_p \pi G}{k} \right)^{2/3} \frac{1}{Q_0^{2/3}} \left( \frac{4\sigma_{SB}}{9 \Omega_{\text{Edd}}^2} \right)^{1/3} \]
\[ = 7.89 \times 10^{-10} s^{-1} \left( \frac{\tau}{10} \right)^{2/3} \left( \frac{Q_0}{1.5} \right)^{-8/3} \frac{1}{0.1} \]
which corresponds to a period of 250 yr. The most common type of binary would then exhibit phase-dependent variations of the accretion rate on a time scale that is too long to be readily observable. Rarer binaries closer to merger would make better candidates for detection via accretion variations. A binary of total mass M will reach the hang-up point at a separation \( a_{\text{hub}} = (GM/\Omega_{\text{hub}}^2)^{1/3} \).

\[ a_{\text{hub}}(M) = \frac{M^{1/3}}{G^{1/9}} \left( \frac{10}{4\pi^2 m_p \pi G} \right)^{4/9} \frac{Q_0^{16/9}}{9 \Omega_{\text{Edd}}^2} \left( \frac{16/9}{0.1} \right)^{2/9} \]

We finally evaluate the stellar scattering time-scale at that position to obtain the hang-up time-scale,
\[ t_{\text{hub}}(M) = 1.74 \times 10^{10} \text{yr} \times \frac{M}{(1 + q^2)^{(3 \times 10^6 M_\odot)^{1.21}}} \frac{10}{Q_0^{16/9}} \frac{1.5}{\alpha_{\text{ice}}^{16/9}} \frac{1}{0.1} \frac{1}{2/9} \]

These analytical approximations are plotted as dashed lines in Fig. 12, and provide a good order-of-magnitude estimate.

5 ACCRETION ONTO THE BLACK HOLES

Gas that gets within 0.1\( a_0 \) of either black hole is taken away from the simulation to prevent the very short time-steps that would be necessary to follow it. This numerical trick is also useful to study the accretion onto each black hole, in particular its variability and behaviour as a function of the orbital phase. Figure 13 shows the accretion rate onto each black hole as a function of time. Overall, more gas is accreted by the secondary black hole, which is closer to the gaseous disc and has a greater specific angular momentum. Both curves show a large peak at the beginning, which is a result of the initial conditions – gas that was situated inside the tidal truncation radius. The accretion rates increase again at \( t \approx 400\Omega_{\text{Edd}}^{-1} \), after the disc loses its ring-like shape and a lot of radial mixing occurs.

The accretion rates are strongly variable, especially on the time-scale of the orbital frequency. Once a \( M \sim 3 \times 10^6 M_\odot \) binary reaches a separation of 0.01 pc, the orbital frequency gets to an interesting observable range \( \sim 0.1 \text{ yr}^{-1} \). Observation of variability on these frequencies will thus help identify the presence of binaries in such close orbits, which would otherwise be extremely difficult to detect. One should note, however, that the maximal accretion rate that a self-gravitating disk can support is typically a small fraction of the Eddington limit\(^5\). Additional gas infall to small radii

\(^5\) The mass accretion rate for a critically fragmenting disc is \( \dot{M} \sim 3\pi^3aQ_0 G^2(\Sigma(\Omega)/\Omega)^3 \). In the optically thin regime this is only \( \dot{M} \sim 10^{-6} M_\odot \text{ yr}^{-1} \), roughly independent of the binary mass and separation. For smaller separations the accretion rate can grow larger as the disc becomes optically thick, but remains sub-Eddington for the whole range of our study.
(King & Pringle, 2007) – bypassing the region most vulnerable to star formation – would be needed to produce very luminous sources.

It should be noticed that the accretion rates are somewhat influenced by numerical resolution. With 2 million SPH particles, an accretion rate of $10^{-5}M\Omega_0$ corresponds to only 100 accreted particles per orbital dynamical time. A shorter simulation with 8 million particles indeed shows up to factor 2 discrepancies, however, the long-term trends and short-term variability are mostly the same.

6 DISCUSSION

In this paper we studied the interaction of a supermassive binary black hole with a self-gravitating circumbinary gaseous disc. Our focus has been on sub-parsec scales, where binary hardening due to stellar dynamical processes is thought to be slow, and where efficient radiative cooling makes it likely that the gas will form a geometrically thin accretion structure. Studies of galaxy mergers, and indirect evidence from our own Galactic Centre, suggest that gas flows to such small scales may be a common feature of galactic nuclei, particularly following mergers. If present, the existence of a gas disc around the binary can be a dominant factor in determining whether the binary merges and the time-scale for this process. The presence of a disc will influence the binary orbit, as it will transport the angular momentum of the binary outward. At the radii of greatest interest the evolution of such gas discs will be dominated by their self-gravity, which will alter the structure of the disc. Our simulations resolve angular momentum transport due to self-gravity physically, without requiring any ad hoc prescription to mimic viscosity.

Our simulations show that the disc, after a transient phase due to the initial conditions, develops a quasi-steady spiral pattern that lasts for at least hundreds of dynamical times. The binary couples gravitationally to the disc, and loses angular momentum. This process makes the orbit of the binary decay at a rate $da/dt \sim -\text{few} \times 10^{-5}a_0\Omega_0$ (for our simulation parameters), in good agreement with theoretical calculations for non-self-gravitating discs. The eccentricity of the binary also increases, at a rate $de/dt \sim 10^{-4}\Omega_0$.

Due to the computational expense of the simulations, we have not followed the system for long enough to see a large change in the orbital parameters. Our direct numerical experiments show only a modest (20%) decrease in the binary separation, accompanied by a growth of eccentricity to values of around $e \approx 0.3$. The evolution of both $a$, $e$ appears to slow down with time, approaching asymptotic values $a \approx 0.8a_0$; $e \approx 0.4$. However, we attribute this trend to the fact that the gaseous disc absorbs angular momentum and moves away from the binary, making the interaction less efficient. This is consistent with a viscous time-scale $t_{\text{dec}} \sim (R/H)^2\eta_{\text{visc}}/\alpha \sim 3000\Omega_0^{-1}$ at $R \sim 3a_0$, where most of the disc mass is located early in the simulation. On the other hand, the ratio $de/da$ remains roughly constant throughout our calculations, suggesting that the mechanism for eccentricity growth has not saturated. We speculate then that, given enough influx of external material (which could easily occur on the viscous time-scale described above), the orbit will continue its decay and the eccentricity will keep growing, perhaps reaching the very large values needed to influence the decay in the relativistic regime.

We also studied the accretion onto the black holes during the orbital decay, and found that it is highly variable, especially on the time-scale of the orbital frequency. This variability pattern, if observed in AGN, could help us to identify binaries at sub-parsec separations.

Motivated by the good agreement between the decay rate obtained from our simulations, and that predicted analytically by Ivanov et al. (1999), we combine their formalism with that of Levin (2007) to calculate the maximum decay rate that can be obtained due to a self-gravitating disc. We find that the decay produced by the disc can dominate over stellar scattering once the binary separation is 0.01–0.1 pc. The time-scale for decay at the critical separation where gas disc and stellar processes are equally important is shorter than the Hubble time for binary masses $M \lesssim 10^7 M_\odot$. This implies that geometrically thin gas discs could provide a solution to the ‘last-parsec’ problem for binaries in this range, whereas gas discs are unable to drive mergers within galaxies hosting the most massive black holes with masses of $10^9 M_\odot$ and above. However, even for masses where gas discs can in principle result in mergers, the time scales are typically rather long. At low redshift, where the typical time scale between major mergers for galaxies hosting $10^7 M_\odot$ black holes is several Gyr (Volonteri et al., 2003), we would then expect that many galaxies would host binaries with separations close to the hang-up radius of a few hundredths of a parsec. At higher redshift the interval between mergers is much shorter – less than a Gyr – and interactions between three or more black holes in a single system could not be neglected.

If we accept that most galactic mergers result in black hole mergers, the results presented here suggest that geometrically thin gas discs could in principle drive mergers on a short enough time scale for relatively low mass black holes at low redshift. Additional processes, either stellar dynamical or hydrodynamic, are needed if very massive binaries are to merge, and to avert ubiquitous 3-body interactions (that might result in ejections) at high redshift. One logical alternative that we did not explore here is that provided...
by even more massive discs. Such discs will fragment and form stars very close to the binary, and if those stars can be scattered by the binary prior to exploding as supernovae they will absorb the binary angular momentum almost as efficiently as if the disc had remained fluid. Given slower scattering, only low mass stars and stellar remnants will be able to contribute to binary decay. The stellar mass function is evidently critical in determining whether such a scenario is viable, as is the stellar dynamics of stars formed in a disc close to a (probably eccentric) binary black hole.

ACKNOWLEDGEMENTS

We thank Volker Springel for his advice on modifying GADGET-2, Patricia Arévalo for her help analysing the accretion rate variability, and Giuseppe Lodato for pointing us to the work of Ivanov et al. (1999). This research was supported by NASA under grants NNG04GL01G, NNX070A08G and NNG05GI92G, and by the NSF under grant AST 0407040. RDA acknowledges support from the Netherlands Organisation for Scientific Research (NWO) through VIDI grants 639.042.404 and 639.042.607. This research was supported in part by the National Science Foundation through TeraGrid resources provided by the National Center for Supercomputing Applications under grants AST070025N and MCA08X017, and utilised the Xeon Linux Cluster (tungsten).

REFERENCES

Alexander R. D., Armitage P. J., Cuadra J., Begelman M. C., 2008, ApJ, 674, 927
Armitage P. J., Natarajan P., 2002, ApJL, 567, L9
Armitage P. J., Natarajan P., 2005, ApJ, 634, 921
Artymowicz P., Clarke C. J., Lubow S. H., Pringle J. E., 1991, ApJL, 370, L35
Balbus S. A., Hawley J. F., 1998, Reviews of Modern Physics, 70, 1
Balbus S. A., Papaloizou J. C. B., 1999, ApJ, 521, 650
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Bell K. R., Lin D. N. C., 1994, ApJ, 427, 987
Cuadra J., Nayakshin S., Springel V., Di Matteo T., 2006, MNRAS, 366, 358
Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956
Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152
Faber S. M., Tremaine S., Ajhar E. A., et al., 1997, AJ, 114, 1771
Ferrarese L., Côté P., Dalla Bontà E., et al., 2006, ApJL, 644, L21
Gammie C. F., 2001, ApJ, 553, 174
Gebhardt K., Bender R., Bower G., et al., 2000, ApJL, 539, L13
Goodman J., 2003, MNRAS, 339, 937
Gosold A., Rix H.-W., 2000, ApJL, 532, L29
Haiman Z., Kocsis B., Menou K., 2008, ArXiv e-prints 0807.4697
Hayasaki K., 2008, ArXiv e-prints 0805.3408
Ivanov P. B., Papaloizou J. C. B., Polnarev A. G., 1999, MNRAS, 307, 79
Johnson B. M., Gammie C. F., 2003, ApJ, 597, 131
King A. R., Pringle J. E., 2007, MNRAS, 377, L25
Kocsis B., Haiman Z., Menou K., 2007, ArXiv e-prints, 712
Levin Y., 2007, MNRAS, 374, 515
Levin Y., Beloborodov A. M., 2003, ApJ, 590, L33
Levine R., Gnedin N. Y., Hamilton A. J. S., Kravtsov A. V., 2008, ApJ, 678, 154
Lippai Z., Frei Z., Haiman Z., 2008, ApJL, 676, L5
Lodato G., 2007, Nuovo Cimento Rivista Serie, 30, 293
Lu J. R., Ghez A. M., Hornstein S. D., Morris M. R., Becklin E. E., Matthews K., 2008, ArXiv e-prints, 808
MacFadyen A. I., Milosavljević M., 2008, ApJ, 672, 83
Mayer L., Kazantzidis S., Madau P., Colpi M., Quinn T., Wadsley J., 2007, Science, 316, 1874
Merritt D., 2006, Reports of Progress in Physics, 69, 2513
Mihos J. C., Hernquist L., 1996, ApJ, 464, 641
Milosavljević M., Merritt D., 2003, ApJ, 596, 860
Milosavljević M., Phinney E. S., 2005, ApJL, 622, L93
Monaghan J. J., 1992, ARA&A, 30, 543
Moorhead A. V., Adams F. C., 2008, Icarus, 193, 475
Nayakshin S., Cuadra J., 2005, A&A, 437, 437
Nayakshin S., Cuadra J., Springel V., 2007, MNRAS, 379, 21
Ogilvie G. I., 2001, MNRAS, 325, 231
Paumard T., Genzel R., Martins F., et al., 2006, ApJ, 643, 1011
Pringle J. E., 1991, MNRAS, 248, 754
Rice W. K. M., Lodato G., Armitage P. J., 2005, MNRAS, 364, L56
Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, ApJ, 664, 49
Schnittman J. D., Krolik J. H., 2008, ArXiv e-prints, 802
Semenov D., Henning T., Helling C., Illner M., Sedlmayr E., 2003, A&A, 410, 611
Shields G. A., Bonning E. W., 2008, ArXiv e-prints, 802
Shlosman I., Begelman M. C., Frank J., 1990, Nature, 345, 679
Springel V., 2005, MNRAS, 364, 1105
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Syer D., Clarke C. J., 1995, MNRAS, 277, 758
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
Yu Q., 2002, MNRAS, 331, 935

© 2008 RAS, MNRAS 000, 1–11