



CHAOTIC ACCRETION AND MERGING SUPERMASSIVE BLACK HOLES

Christopher Nixon

Submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy

May 2012

Theoretical Astrophysics Group
Department of Physics and Astronomy
University of Leicester

Chaotic accretion and merging
supermassive black holes

by

Christopher Nixon

The main driver of the work in this thesis is the idea of chaotic accretion in galaxy centres. Most research in this area focuses on orderly or coherent accretion where supermassive black holes or supermassive black hole binaries are fed with gas always possessing the same sense of angular momentum. If instead gas flows in galaxies are chaotic, feeding occurs through randomly oriented depositions of gas. Previous works show that this chaotic mode of feeding can explain some astrophysical phenomena, such as the lack of correlation between host galaxy structure and the direction of jets. It has also been shown that by keeping the black hole spin low this feeding mechanism can grow supermassive black holes from stellar mass seeds. In this thesis I show that it also alleviates the ‘final parsec problem’ by facilitating the merger of two supermassive black holes, and the growth of supermassive black holes through rapid accretion. I also develop the intriguing possibility of breaking a warped disc into two or more distinct planes.

The work in this thesis has been published in the Monthly Notices of the Royal Astronomy Society:

- 1) “Retrograde accretion and merging supermassive black holes”
Nixon et al. (2011a) MNRAS Volume 412, Issue 3, pp. 1591-1598
- 2) “The final parsec problem: aligning a binary with an external accretion disc”
Nixon et al. (2011b) MNRAS Volume 417, Issue 1, pp. L66-L69
- 3) “Broken discs: warp propagation in accretion discs”
Nixon and King (2012) MNRAS Volume 421, Issue 2, pp. 1201-1208
- 4) “Rapid AGN accretion from counterrotating discs”
Nixon et al. (2012) MNRAS Volume 422, Issue 3, pp. 2547-2552
- 5) “Stable counteralignment of a circumbinary disc”
Nixon (2012) MNRAS in press

Acknowledgments

I thank my supervisor Prof. Andrew King for providing everything required and more during my PhD. I thank Prof. Jim Pringle for his continued guidance and for initiating my interest in astrophysics. I thank the rest of my collaborators, particularly Daniel Price for meeting all of my SPH needs. I thank the members of the Theory Group (past and present) for all the help, advice and discussions I have had over the years, in particular Richard Alexander. I also thank Lisa Brant for all the help with admin and travel and Gary Gilchrist and the rest of RCS for computing support.

I thank my family and friends for support and encouragement, and I thank all the people who have helped me to get to where I am.

I also thank the University of Leicester, STFC and BIS for providing more computer power than I could consume, and the STFC for the studentship that funded this work.

I also acknowledge the use of SPLASH (Price, 2007) which I used to generate all of the figures in this thesis.

Contents

Abstract	i
Acknowledgments	iii
1 Introduction	1
1.1 Prologue	2
1.2 Supermassive black holes	3
1.2.1 Observations	4
1.2.2 Growing supermassive black holes	5
1.3 The final parsec problem	9
1.3.1 Dynamical friction	10
1.3.2 Gas discs	11
1.3.3 Gravitational waves	12
1.3.4 Summary	12
1.4 Accretion discs	13
1.4.1 Thin disc equations	14
1.4.2 Warped discs	17

1.5	Numerical methods	19
1.5.1	Accretion disc ring-code	20
1.5.2	Smoothed particle hydrodynamics	22
1.5.2.1	Calculating densities	23
1.5.2.2	Smoothing kernels	23
1.5.2.3	Consistent h - ρ calculation	25
1.5.2.4	Equations of motion	26
1.5.2.5	Dissipative terms	27
1.5.2.6	Time-stepping	29
2	Binary-Disc Alignment	30
2.1	Abstract	31
2.2	Introduction	31
2.3	The binary-disc torque	34
2.4	Co- or counter-alignment?	36
2.5	Discussion	38
3	Stable counteralignment	44
3.1	Abstract	45
3.2	Introduction	45
3.3	Simultaneous co- <i>and</i> counter-alignment?	47
3.4	Simulation	49
3.4.1	Setup	50

3.4.2	Geometry	51
3.4.3	Stable counteralignment	51
3.5	Conclusions	58
4	Retrograde Accretion	59
4.1	Abstract	60
4.2	Introduction	60
4.3	Prograde vs. Retrograde	62
4.4	Where does the mass go?	64
4.5	Eccentricity growth	66
4.6	Orbital evolution of the binary	69
4.7	Simulations	71
4.7.1	Code setup	72
4.7.2	Where does the mass go?	73
4.7.2.1	Secondary capture	74
4.7.2.2	Primary capture	76
4.7.2.3	Dual Capture	78
4.7.3	Eccentricity growth	81
4.8	Discussion	84
5	Broken Discs	87
5.1	Abstract	88
5.2	Introduction	88

5.3	Numerical method	92
5.3.1	Evolution equation	92
5.3.2	Generating the constrained effective viscosities	95
5.4	Simulations	101
5.4.1	Does the disc break?	101
5.4.1.1	$\theta = 10^\circ$	102
5.4.1.2	$\theta = 30^\circ$	103
5.4.1.3	$\theta = 45^\circ$	103
5.4.1.4	$\theta = 60^\circ$	103
5.5	Discussion	109
6	Rapid AGN accretion	113
6.1	Abstract	114
6.2	Introduction	114
6.3	Counterrotating discs	116
6.4	Simulations	118
6.4.1	Code Setup	118
6.4.2	Tilted disc evolution	120
6.4.2.1	$\theta = 180^\circ$	129
6.4.2.2	Effect of the gas equation of state	132
6.5	Discussion	135
7	Conclusions	138

7.1	Misaligned circumbinary discs	139
7.2	Stable counteralignment	139
7.3	Retrograde circumbinary discs	140
7.4	Broken discs	141
7.5	Rapid accretion	142
7.6	Future work	142
	References	150

List of Figures

2.1	Binary–disc alignment: warped disc shape	40
3.1	Stable counteralignment: tilt angle	53
3.2	Stable counteralignment: twist angle	54
3.3	Stable counteralignment: simulation face–on	56
3.4	Stable counteralignment: simulation edge–on	57
4.1	Retrograde accretion: simulation of secondary accretion	75
4.2	Retrograde accretion: simulation of primary accretion	77
4.3	Retrograde accretion: simulation of dual accretion face on	79
4.4	Retrograde accretion: simulation of dual accretion tilted	80
4.5	Retrograde accretion: eccentricity growth	83
5.1	Broken discs: effective viscosity coefficient Q1	98
5.2	Broken discs: effective viscosity coefficient Q2	99
5.3	Broken discs: effective viscosity coefficient Q3	100
5.4	Broken discs: $\theta = 10^\circ$	105
5.5	Broken discs: $\theta = 30^\circ$	106

5.6	Broken discs: $\theta = 45^\circ$	107
5.7	Broken discs: $\theta = 60^\circ$	108
6.1	Rapid accretion: accretion rates	121
6.2	Rapid accretion: simulation with $\theta = 30^\circ$	123
6.3	Rapid accretion: simulation with $\theta = 60^\circ$	124
6.4	Rapid accretion: simulation with $\theta = 120^\circ$	127
6.5	Rapid accretion: simulation with $\theta = 150^\circ$	128
6.6	Rapid accretion: simulation with $\theta = 180^\circ$	131
6.7	Rapid accretion: simulation with an isothermal equation of state . . .	133
6.8	Rapid accretion: simulation with an alternative thermal treatment . .	134

Dedicated to my parents

1

Introduction

1.1 Prologue

In this thesis I explore the evolution of gas discs (e.g. Pringle, 1981) around spinning supermassive black holes (SMBH), or SMBH binaries. This is an important ingredient in understanding the evolution of SMBH, governing both how they grow and how they merge. SMBH and SMBH binaries possess significant amounts of angular momentum, as well as mass. This complicates the disc physics somewhat, but in this thesis I show that SMBH and SMBH binaries encourage their own growth and mergers through their effects on surrounding gas.

In this Chapter I provide the required background to my work. In Chapter 2 I examine the effect of a binary on a misaligned external gas disc, finding that both alignment and counteralignment of the disc are possible. In Chapter 3 I show that counteralignment of a circumbinary disc is stable and thus prolonged periods of accretion from retrograde circumbinary discs are to be expected. In Chapter 4 I show that such accretion can lead to significant evolution of the binary, specifically increasing the binary eccentricity to unity once the smaller mass SMBH has interacted with approximately its own mass. Once the binary eccentricity is large enough gravitational waves will complete the coalescence. This provides a mechanism for merging SMBH within a Hubble time.

In Chapter 5 I explore the evolution of a misaligned disc around a single SMBH. In this case the black hole spin plays an important role. In previous work the assumed form of the disc viscosity has been fairly simple. In this work we use the full effective viscosities expected in a warped disc (Ogilvie, 1999), finding as a result that the disc may break into two distinct planes. This suggests a mode of feeding the SMBH where there is a large internal disc tilt, allowing cancellation of angular momentum. In Chapter 6 I explore this possibility and quantify the increase in

accretion rate associated with the enhanced mixing of the gas orbits.

Finally in Chapter 7 I provide the conclusions from my work and suggest possible future research.

1.2 Supermassive black holes

Generally, black holes are objects which possess an ‘event horizon’ inside which their gravitational field is so strong that even light cannot escape it. For the purposes of this thesis we can describe a black hole by two parameters: mass and angular momentum. Black holes themselves do not emit any observable electromagnetic radiation and thus they are potentially difficult to detect. However with so much mass in a small volume, their compactness generates a gravitational field with large amounts of potential energy available to infalling material. As this energy is liberated the black holes become indirectly observable (see Section 1.2.1).

Black holes are thought to provide other forms of radiation. The most relevant here, and the focus of much current research, is gravitational wave radiation. This is released between orbiting massive objects, however its efficiency depends very strongly on the mass and separation of the objects. Therefore it is only critical when objects of extreme mass are orbiting very close to each other. This is an important aspect in coalescing SMBH (Section 1.3.3)

The physics of black holes can become very complex. However, if we are only interested in the dynamics occurring far from the innermost stable circular orbit (isco) we can neglect much of the complexity associated with General Relativity. We can parametrise the main effects as perturbations to a Newtonian potential, making tractable problems that would otherwise inhibit progress. For a non-spinning black

hole we treat it as a point mass with a Newtonian gravitational potential,

$$\phi(r) = -\frac{GM}{r}, \quad (1.1)$$

where G is the gravitational constant, M is the mass of the black hole and r is the radial distance from the black hole. This is a spherically symmetric potential which exerts a force per unit mass of

$$\mathbf{f}(r) = -\nabla\phi = -\frac{GM}{r^3}\mathbf{r} \quad (1.2)$$

where \mathbf{r} is the radial vector from the point mass.

The angular momentum or spin of the black hole can play a significant role in the dynamics – even far from the isco. The effect of the spin can be approximated as a perturbation to the orbits of particles in a Newtonian potential (Wilkins, 1972). This perturbation causes the orbits to precess around the black hole spin vector. This effect vanishes for circular orbits in the plane of the black hole spin, either coaligned or counteraligned. This precession is known as the Lense–Thirring effect (Lense and Thirring, 1918; Mashhoon et al., 1984) and plays an important role in the evolution of a misaligned disc around a black hole (see Chapter 5).

1.2.1 Observations

Astronomers now generally agree that the centre of every reasonably large galaxy contains an SMBH. The strongest evidence is for the SMBH at the centre of our own galaxy, with the most clear evidence being the observation of stellar orbits in the central parsec (e.g. Schödel et al., 2002; Ghez et al., 2005). These high-resolution astrometric imaging observations follow full orbits of stars around a central mass

concentration which is only plausibly explained by the presence of an SMBH.

Evidence also exists for the presence of SMBH in other galaxies such as stellar and gas kinematics (e.g. Kormendy and Richstone, 1995). For example Keplerian rotation profiles in observed in maser discs. Also the detection of a gravitationally redshifted Iron $K\alpha$ emission line suggests the presence of SMBH in galaxy centres (Tanaka et al., 1995).

It is also generally accepted that galaxies follow a hierarchical structure formation where small galaxies merge to produce larger ones (e.g. Springel et al., 2005). Thus galaxy mergers play a significant role in galaxy evolution. In this picture it is important to understand how the two SMBH evolve. It is expected that they will sink rapidly to the centre of the merged galaxy and form a binary, however there is currently no strong evidence for the existence of binary SMBH on \lesssim parsec scales. Therefore we expect the binary does merge, however it is not clear how this is achieved (Begelman et al., 1980). I shall return to this point in more detail in Section 1.3.

1.2.2 Growing supermassive black holes

We still do not fully understand how SMBH can grow to be so large so quickly. Near-infrared observations of quasars provide evidence for SMBHs of mass $\gtrsim 10^9 M_\odot$ at a redshift of $z \simeq 6$ (e.g. Willott et al., 2003). This provides a significant challenge for current models. In any physically realistic scenario, gas in a galaxy possesses angular momentum which is a barrier to accretion. The gas must spiral slowly inwards through a disc, transporting angular momentum outwards by the action of a viscosity (see Section 1.4). This process is slow. The timescale for a disc of gas at a radius of 1 pc to accrete on to an SMBH of mass $10^8 M_\odot$ approaches a Hubble

time. We therefore need an alternative mechanism for driving the accretion of gas on to SMBH. One natural idea is that accretion occurs through chaotic deposition of gas into the galaxy centre. This is reasonable, as the nuclei of galaxies are very chaotic places, with random injections of energy and momentum from star formation, supernovae and stellar winds. Moreover, there is no reason to assume that gas falling into the galactic centre has any prior knowledge of the SMBH angular momentum. This gas must settle into a disc, generally with a random angle of misalignment between the disc and the SMBH angular momentum. The disc is probably strongly warped because of the chaotic nature of the infalling gas, and warped further still by interaction with the spinning SMBH via the Lense–Thirring effect (e.g. Bardeen and Petterson, 1975). In Chapters 5 & 6 we shall explore the evolution of such discs. A chaotic accretion scenario was originally suggested by King and Pringle (2006) as a mechanism for keeping the spin of SMBH low, because accretion occurs from both retrograde and prograde discs at various epochs. This allows the growth of SMBH from stellar mass seeds (King and Pringle, 2006, 2007; King et al., 2008).

Soltan (1982) suggests that there is on average a $10^8 M_\odot$ SMBH per reasonable size galaxy. This argument equates the power in the observed background radiation field with the power emitted by quasars¹. King and Pringle (2006) showed that this can place limits on the growth of SMBH. If we assume SMBH undergo luminous accretion then the accretion luminosity cannot exceed the Eddington luminosity

$$L_{\text{Edd}} = \frac{4\pi GMc}{\kappa} \simeq 10^{47} M_9 \text{ erg s}^{-1} \quad (1.3)$$

where $M = 10^9 M_9 M_\odot$ and κ is the electron scattering opacity. This limit on the

¹Quasars are assumed to be powered by the emission from accretion discs around SMBH.

luminosity sets a limit on the rate at which the black hole can accrete as

$$L_{\text{Edd}} = \epsilon \dot{M}_{\text{acc}} c^2 \quad (1.4)$$

where ϵ is the accretion efficiency, determined by the fractional binding energy of the isco of the black hole (see King and Pringle 2006 for more detail). Therefore the growth rate of the black hole accreting at the Eddington limit is given by

$$\dot{M} = (1 - \epsilon) \dot{M}_{\text{acc}} \quad (1.5)$$

and therefore

$$\dot{M} = \frac{1 - \epsilon}{\epsilon} \frac{M}{t_{\text{Edd}}} \quad (1.6)$$

with

$$t_{\text{Edd}} = \frac{\kappa c}{4\pi G} = 4.5 \times 10^8 \text{yr}. \quad (1.7)$$

King and Pringle (2006) further argue that although ϵ varies as accretion proceeds, its minimum value ϵ_{min} sets a limit on the mass M to which the black hole can grow from an initial seed mass M_0 . By integrating (1.6) they show that

$$\frac{M}{M_0} < \exp \left[\left(\frac{t}{t_{\text{Edd}}} \right) \left(\frac{1}{\epsilon_{\text{min}}} - 1 \right) \right] \quad (1.8)$$

Now let us remember that there is observational evidence for an SMBH of mass $\gtrsim 10^9 M_{\odot}$ at a redshift of $z \simeq 6$ (e.g. Willott et al., 2003). At a redshift of $z \simeq 6$, $t/t_{\text{Edd}} \simeq 2$. Therefore if we assume that the black hole is maximally spinning (dimensionless Kerr parameter $a \simeq 1$), and thus $\epsilon_{\text{min}} = 0.43$ we have $M/M_0 \lesssim 20$. Clearly there is no prospect of growing black holes in this manner. However, if we have $\epsilon_{\text{min}} \simeq 0.06$, as appropriate for $a \simeq 0.3 - 0.4$, then $M/M_0 \sim 4 \times 10^{13}$. Thus we

see that it is clearly possible to grow the black holes rapidly if we can keep the spin low². Note that as the spin can vary and accretion is not continuous one does not expect SMBH of mass $\sim 10^{13}M_{\odot}$.

King and Pringle (2006) go on to discuss the spin evolution of SMBH. Using the results of King et al. (2005) we expect that the fraction of randomly oriented accretion events that counteralign with respect to the hole's spin (the rest coalign) to be

$$f = \frac{1}{2} \left[1 - \frac{J_d}{2J_h} \right], \quad (1.9)$$

where J_d and J_h are the magnitudes of the disc and hole angular momenta respectively. Therefore there are two modes of accretion relevant here. If $J_d \lesssim J_h$ then accretion episodes will alternate between coaligned and counteraligned, however if $J_d \gtrsim J_h$ then accretion proceeds in a prograde manner, consistently spinning up the hole. There is an additional complicating effect that the specific angular momentum of prograde and retrograde accretion are distinct. The radius of the isco is largest for $a = -1$ and smallest for $a = 1$, suggests that spindown is considerably more effective than spinup (King and Pringle, 2006).

King et al. (2008) explored this effect in detail by calculating the effect of randomly oriented accretion events on to an SMBH. They argue that the self-gravity of the disc limits the mass in each event to be a small fraction of that of the hole $M_{\text{sg}} \lesssim (H/R) M$ and thus both coaligned and counteraligned discs occur. The spin parameter performs a random walk and King et al. (2008) suggest that the final spins from this mode of accretion are $a \simeq 0.3$ (see Eq. 19 of that paper). An important factor discussed in their paper is the effect of black hole coalescences which may significantly increase or decrease the spin. Both accretion (King et al., 2008) and

²The hole's spin will vary as it accretes, however it must spend some time with low spin to grow efficiently.

coalescences (Hughes and Blandford, 2003) statistically lead to spindown, however any individual episode is capable of significantly increasing the spin. Whilst the hole spin is low it may accrete efficiently and hence grow its mass to the observed values.

1.3 The final parsec problem

The ‘final parsec problem’ (Begelman et al., 1980) is born from a dichotomy between observations and theory. Begelman et al. (1980) showed that it is theoretically difficult to drive the merging of an SMBH binary. However, observationally we know that galaxy mergers are common in the universe and that galaxies host one (and apparently only one³) SMBH, suggesting that the process which merges SMBH binaries is rather more efficient than the theory suggests.

When two galaxies merge it is expected that the two SMBH sink towards the centre of the merged galaxy through dynamical friction with background gas and stars and begin to orbit each other at some radius. Now one of two things must happen: either the galaxy can remove angular momentum and energy from the binary orbit efficiently enough to facilitate a merger, or the binary will ‘stall’ and therefore not coalesce. There are several mechanisms that have been discussed in the literature to extract angular momentum from the binary. I briefly summarise the main ones here.

³It is not yet possible to rule out binary SMBH completely, and there are claimed observations of dual SMBH galaxies, however these are usually at much larger separations than the sub–parsec scale required here (Merritt and Milosavljević, 2005), however some close binaries may exist on \sim pc scales (e.g. ~ 7 pc separation, Rodriguez et al., 2006).

1.3.1 Dynamical friction

Dynamical friction with background stars is a very efficient mechanism at driving the SMBH binary from \sim kpc scales to \sim 1 pc. On the larger scales the motion of the SMBH through a background of stars leads to an over-density of stars behind each SMBH through gravitational focusing. This over-density causes a drag force on the binary which removes angular momentum and energy from its orbit (see e.g. Binney and Tremaine, 2008).

At smaller separations (but still \gtrsim 1 pc) there are many stars on orbits that produce an interaction with the binary. During this three-body encounter angular momentum and energy are exchanged between the components resulting in the ejection of one of the objects. It is highly favourable for the lowest mass object to be ejected, as it is moving fastest (requiring the exchange of the least angular momentum) and thus the star is ejected carrying off some energy and angular momentum. This process is efficient at shrinking the binary until separations of \sim 1 pc. At this point the binary has scattered away most of the stars it can interact with and refilling of these orbits is apparently too slow (Begelman et al., 1980).

Refilling these orbits may be possible if the galaxy potential is triaxial (Merritt and Poon, 2004). Triaxiality in the potential induces box-orbits, passing arbitrarily close to the centre, which can interact with the binary. Berczik et al. (2006) show that triaxial potentials can induce the merger of SMBH binaries on timescales of \lesssim 10 Gyr. It is not clear whether this works in practise: the degree of triaxiality in dark matter haloes is not known, and the destruction of these box-orbits by the SMBH binary should cause a back-reaction on the halo which reduces its triaxiality. For the purposes of this thesis I will assume that dynamical friction with background stars can only shrink the binary to \sim 1 pc (Begelman et al., 1980).

1.3.2 Gas discs

Gas is expected to play a role in facilitating an SMBH binary merger. We know that galaxies supply their central SMBH with large amounts of gas during the relatively quiet phases between mergers. Therefore during a merger, when the entire galaxy is stirred up, we expect large amounts of gas to fall to the galaxy centre and interact with the binary. Note that the total mass in the binary is a small fraction of the mass in gas in the galaxy and therefore only a small fraction of the total gas is required.

Previous work in this area has focused on the evolution of prograde gas discs around an SMBH binary. A prograde circumbinary disc does not evolve as a usual accretion disc. Instead the rotation of the binary acts as a source of angular momentum generating a decretion disc⁴. The angular momentum of the binary is communicated to the disc through resonances, these occur at radii where the orbits in the disc line up with the orbit of the binary, i.e. where

$$\Omega^2 = m^2 (\Omega - \omega)^2 \quad (1.10)$$

or equivalently

$$\Omega(R) = \frac{m\omega}{m \pm 1} \quad (1.11)$$

where Ω is the disc angular velocity and ω is the binary orbital frequency and $m = 1, 2, \dots$ is the wave mode number (Papaloizou and Pringle, 1977).

The resonances close to the binary have the strongest effect. To a good approximation the disc will move inwards viscously seemingly unperturbed until it reaches the 1:2 resonance. At this point the disc is strongly affected by the binary and is

⁴An accretion disc moves mass inwards, however a decretion disc moves mass outwards.

held at this radius. In holding the disc out the binary slowly shrinks. The timescale for this exchange of angular momentum is long and indeed is found to be too long to provide a reasonable solution to the final parsec problem (Lodato et al., 2009; Cuadra et al., 2009). The main reason for this lack of efficiency is the gas has to carry away the angular momentum and thus moves outwards, reducing the exchange of angular momentum.

However gas can still play a major role in the evolution of an SMBH binary. So far we have only considered prograde discs. In Chapter 4 we shall see that retrograde discs can be more efficient at merging the binary.

1.3.3 Gravitational waves

Gravitational wave emissions play a crucial role in merging an SMBH binary. At all separations the binary is emitting gravitational waves, however the efficiency of this process is dependent on the binary separation to the fourth power (e.g. Eq. 7 in Lodato et al., 2009). Therefore when the binary is at separations $\gtrsim 0.1$ pc this mechanism will take too long to coalesce the binary ($\sim 10^{11}$ yrs), however if the binary can reach separations $\lesssim 0.01$ pc then gravitational waves will complete the coalescence in $\sim 10^7$ yrs (e.g. Lodato et al., 2009).

This is the only mechanism which gets stronger at smaller separations, and is therefore a crucial element in the final stages of an SMBH merger.

1.3.4 Summary

Dynamical friction is capable of shrinking the binary to separations ~ 1 pc. If we can drive the binary to separations ~ 0.01 pc, then gravitational waves will complete the coalescence. Prograde gas discs are not very efficient at driving this evolution, in

Chapter 4 we propose a possible solution to this problem – retrograde circumbinary discs.

1.4 Accretion discs

In general gas falling towards a black hole (or any point mass) possesses angular momentum, i.e. gas is not perfectly aimed to collide directly with the object. Instead it orbits the black hole. In reality the orbit is not a closed ellipse but forms a rosette as the potential is not perfectly Keplerian due to nearby stars or gas self-gravity etc. On this rosette gas orbits intersect and shock, leading to dissipation of energy through heating or radiation. However angular momentum must be conserved during this process. Therefore the gas settles into the orbit of lowest energy for the given angular momentum – a circle. In general gas will not all possess the same specific angular momentum and so settles into a disc, rather than a ring. More generally gas will have angular momentum with a spread of directions, and so settles into a warped disc (see Section 1.4.2).

The disc then evolves under the action of a viscosity. For the purposes of this thesis ‘viscosity’ is defined as whatever process drives the Shakura & Sunyaev viscosity $\nu = \alpha c_s H$ (Shakura and Sunyaev, 1973). In the literature there is much debate on its origin. The most accepted is the magneto-rotational instability (MRI) suggested by Balbus and Hawley (1991). This instability requires a weakly magnetised disc and the angular velocity to decrease outwards as in Keplerian and near-Keplerian discs. The cartoon picture of the instability is two parcels of gas, at different radii, joined by a magnetic field loop. The inner parcel rotates faster and thus the loop becomes stretched. The magnetic field acts to slow the inner parcel and speed up the outer parcel, thus transferring angular momentum from the inner parcel to the

outer parcel. The net effect is the inner parcel moves inwards through the disc, and the outer parcel moves outwards. For the purposes of this thesis we assume that the disc viscosity can be parametrised by the Shakura & Sunyaev α .

The parametrisation is a simple, but powerful, way of characterising the angular momentum transport in a disc. For a viscosity we need a length and a velocity. In a disc the typical length scale is the disc thickness H , for example the largest length scale for any turbulence in the disc $\sim H$. Any supersonic motions in the disc drive shocks and thus rapidly dissipate, therefore the maximal signal velocity in the disc is approximately the disc sound speed c_s . Therefore the maximal viscosity expected in a disc is $\nu = c_s H$ and so by adding the efficiency parameter we have the Shakura & Sunyaev viscosity $\nu = \alpha c_s H$.

1.4.1 Thin disc equations

Now that we have a prescription for the viscosity in a disc we can derive its equations of motion. To make things simple, we assume (for the moment) that the disc can be modelled in 2D polars (R, ϕ) with surface density $\Sigma(R, t)$, radial velocity $v_R(R, t)$ and angular velocity $\Omega(R, t)$. Now we consider the annulus of gas between R and $R + \Delta R$. The mass in this annulus is $2\pi R \Delta R \Sigma$. Conservation of mass then allows us to equate the rate of change of mass in the annulus to the net flow of mass into or out of the annulus

$$\frac{\partial}{\partial t} (2\pi R \Delta R \Sigma) = (2\pi R v_R \Sigma)_R - (2\pi R v_R \Sigma)_{R+\Delta R} \quad (1.12)$$

If we rearrange and take the limit as $\Delta R \rightarrow 0$ then we arrive at the continuity equation (mass conservation) for a disc

$$R \frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial R} (R \Sigma v_R) = 0. \quad (1.13)$$

Similarly we can derive the equation expressing angular momentum conservation. The angular momentum of an annulus is $2\pi R \Delta R \Sigma R^2 \Omega$ and the rate of change of angular momentum is the net flux of angular momentum plus the net torques

$$\begin{aligned} \frac{\partial}{\partial t} (2\pi R \Delta R \Sigma R^2 \Omega) &= (2\pi R \Sigma R^2 \Omega v_R)_R - (2\pi R \Sigma R^2 \Omega v_R)_{R+\Delta R} \\ &+ G(R + \Delta R) - G(R) \end{aligned} \quad (1.14)$$

and again by rearranging and taking the limit $\Delta R \rightarrow 0$ we have the angular momentum equation

$$R \frac{\partial}{\partial t} (\Sigma R^2 \Omega) + \frac{\partial}{\partial R} (R \Sigma v_R R^2 \Omega) = \frac{1}{2\pi} \frac{\partial G}{\partial R} \quad (1.15)$$

where $G(R, t)$ is the internal disc torque.

$G(R, t)$ is the result of the disc viscosity. The viscous force is proportional to the rate of shearing

$$F = 2\pi R H \mu R \frac{d\Omega}{dR} \quad (1.16)$$

where $R d\Omega/dR$ is the rate of shear and μ is the dynamic viscosity – related to the kinematic viscosity by $\mu = \rho \nu$. Therefore, using $\rho = \Sigma/H$ the force is

$$F = 2\pi R \nu \Sigma R \frac{d\Omega}{dR}. \quad (1.17)$$

Note that this force acts in the azimuthal direction, and thus when the cross product

with the radial vector is taken to generate the torque, we can see that the torque acts in the correct direction (i.e. on the z-component of angular momentum - note we are only considering a flat disc in the R - ϕ plane so far). So we can write the internal torque due to viscosity as

$$G = 2\pi R\nu\Sigma R\Omega'R \quad (1.18)$$

and so the equation expressing angular momentum conservation is

$$R\frac{\partial}{\partial t}(\Sigma R^2\Omega) + \frac{\partial}{\partial R}(R\Sigma v_R R^2\Omega) = \frac{\partial}{\partial R}(\nu\Sigma R^3\Omega'). \quad (1.19)$$

By multiplying Eq. 1.13 by $R^2\Omega$ and then subtracting from Eq. 1.19 we can rearrange for v_R to get

$$v_R = \frac{\frac{\partial}{\partial R}(\nu\Sigma R^3\Omega')}{R\Sigma\frac{\partial}{\partial R}(R^2\Omega)}. \quad (1.20)$$

Substituting this into (1.13) gives

$$\frac{\partial\Sigma}{\partial t} = \frac{1}{R}\frac{\partial}{\partial R}\left(\frac{\frac{\partial}{\partial R}[\nu\Sigma R^3(-\Omega')]}{\frac{\partial}{\partial R}(R^2\Omega)}\right) \quad (1.21)$$

which for the Keplerian potential, $\Omega = \sqrt{GM/R^3}$, is

$$\frac{\partial\Sigma}{\partial t} = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}(\nu\Sigma R^{1/2})\right]. \quad (1.22)$$

This is essentially a diffusion equation, if we know the viscosity ν then we know how a distribution $\Sigma(R, t)$ will evolve.

If we assume the disc is in hydrostatic balance we can also derive the vertical struc-

ture by equating the vertical pressure force with the relevant component of gravity

$$\frac{1}{\rho} \frac{\partial P}{\partial z} = - \frac{GM}{R^2 + z^2} \frac{z}{\sqrt{R^2 + z^2}} \quad (1.23)$$

where the last factor is the geometrical sine term. If the disc is thin, then we can use $|z| \ll R$, so

$$\frac{1}{\rho} \frac{\partial P}{\partial z} = - \frac{GMz}{R^3}. \quad (1.24)$$

Now if we assume an equation of state, e.g. isothermal $P = c_s^2 \rho$ then

$$c_s^2 \frac{\partial \ln \rho}{\partial z} = - \frac{GMz}{R^3}. \quad (1.25)$$

We can then integrate to get

$$\rho = \rho_0 \exp\left(\frac{-GMz^2}{2c_s^2 R^3}\right) = \rho_0 \exp\left(-\frac{z^2}{2H^2}\right) \quad (1.26)$$

where we have introduced the disc scale–height $H = c_s/\Omega$.

From this we can see that a disc in vertical hydrostatic equilibrium has $H/R = c_s/v_\phi$, so for thin discs the flow is supersonic.

1.4.2 Warped discs

In the previous section we derived the equations relevant to a flat disc, however in general the disc may also be warped. Warping disturbances can propagate in two distinct modes: wave–like or diffusive. If $\alpha < H/R$ then waves propagate efficiently, however if $\alpha > H/R$ then diffusion dominates (Papaloizou and Pringle, 1983). In this thesis we focus on the latter case relevant for discs around black holes. Here I detail a simple derivation of the equation of motion for a diffusive warped disc.

The local angular momentum density is $\mathbf{L}(R, t) = \Sigma R^2 \Omega \boldsymbol{\ell}(R, t)$, where $\boldsymbol{\ell}(R, t)$ is a unit vector determining the direction of the angular momentum of the ring. The mass conservation equation is the same as for the flat disc (1.13).

The angular momentum conservation is expressed as

$$\begin{aligned} \frac{\partial}{\partial t} (2\pi R \Sigma R^2 \Omega \Delta R \boldsymbol{\ell}) &= (2\pi R \Sigma R^2 \Omega v_R \boldsymbol{\ell})_R \\ &\quad - (2\pi R \Sigma R^2 \Omega v_R \boldsymbol{\ell})_{R+\Delta R} \\ &\quad + \mathbf{G}(R + \Delta R) - \mathbf{G}(R) \\ &\quad + \Delta \mathbf{T} \end{aligned} \quad (1.27)$$

where \mathbf{G} is the viscous torque and $\Delta \mathbf{T}$ is any other (external) torque acting on ΔR .

The viscous torque has two obvious components, the (R, ϕ) stress contributes a torque acting in the direction of $\boldsymbol{\ell}$ (1.18)

$$\mathbf{G}_1 = 2\pi R \nu_1 \Sigma R \Omega' R \boldsymbol{\ell}, \quad (1.28)$$

where ν_1 is the azimuthal shear viscosity. If we consider two neighbouring rings with $\boldsymbol{\ell}$ and $\boldsymbol{\ell} + \Delta \boldsymbol{\ell}$ then the (R, z) stress acts to communicate $\Delta \boldsymbol{\ell}$ between the rings and thus the torque acts in the direction $\partial \boldsymbol{\ell} / \partial R$. The (R, z) torque is then

$$\mathbf{G}_2 = 2\pi R \frac{1}{2} \nu_2 \Sigma R^2 \Omega \frac{\partial \boldsymbol{\ell}}{\partial R}, \quad (1.29)$$

where ν_2 is the vertical shear viscosity and the factor of a half comes from integrating $\cos^2 \phi$ across the ring (Papaloizou and Pringle, 1983).

Therefore, once again rearranging and taking the limit $\Delta R \rightarrow 0$ we have (Papaloizou

and Pringle, 1983)

$$\begin{aligned} \frac{\partial}{\partial t} (\Sigma R^2 \Omega \ell) + \frac{1}{R} \frac{\partial}{\partial R} (\Sigma v_R R^3 \Omega \ell) &= \frac{1}{R} \frac{\partial}{\partial R} (\nu_1 \Sigma R^3 \Omega' \ell) \\ &+ \frac{1}{R} \frac{\partial}{\partial R} \left(\frac{1}{2} \nu_2 \Sigma R^3 \Omega \frac{\partial \ell}{\partial R} \right). \end{aligned} \quad (1.30)$$

Pringle (1992) showed that this equation can be combined with (1.13) to eliminate v_R and give an equation governing the evolution of the total angular momentum vector

$$\begin{aligned} \frac{\partial \mathbf{L}}{\partial t} &= \frac{1}{R} \frac{\partial}{\partial R} \left\{ \frac{(\partial/\partial R) [\nu_1 \Sigma R^3 (-\Omega')]}{\Sigma (\partial/\partial R) (R^2 \Omega)} \mathbf{L} \right\} \\ &+ \frac{1}{R} \frac{\partial}{\partial R} \left[\frac{1}{2} \nu_2 R |\mathbf{L}| \frac{\partial \ell}{\partial R} \right] \\ &+ \frac{1}{R} \frac{\partial}{\partial R} \left\{ \left[\frac{\frac{1}{2} \nu_2 R^3 \Omega |\partial \ell / \partial R|^2}{(\partial/\partial R) (R^2 \Omega)} + \nu_1 \left(\frac{R \Omega'}{\Omega} \right) \right] \mathbf{L} \right\}. \end{aligned} \quad (1.31)$$

The derivation of this equation is somewhat simplistic, so to check its validity Ogilvie (1999) derived the equation from the full hydrodynamical equations of motion. Ogilvie (1999) found that the equation is correct up to two important points. First the equation is missing a term which makes neighbouring annuli precess if they are tilted with respect to each other (cf. Eq 5.3), and second the viscosities in the disc are related and are uniquely determined once the α parameter and warp amplitude are known. The consequences of these relations are studied in Chapter 5.

1.5 Numerical methods

In this section I provide a brief description of the numerical methods employed in this thesis. The two methods used are an accretion disc ring-code (Pringle, 1992)

and Smoothed Particle Hydrodynamics.

1.5.1 Accretion disc ring-code

In this section I give a description of the code used in Chapter 5. The method is described in full in Pringle (1992) and here I repeat the salient points. It is a 1D code⁵ which integrates the evolution of rings of gas due to various internal and external torques assuming diffusive (viscous) communication of angular momentum. Each ring of gas is assumed to be in centrifugal balance, i.e. the rings are rigid and concentric, thus obviating the need to integrate over the dynamical timescale. The gas is implicitly assumed to be in equilibrium on the dynamical timescale, with any changes occurring on the much longer viscous timescale. This will clearly miss any dynamical behaviour, for example excluding any dynamical infall of gas. However this method is very powerful as it allows us to integrate for a viscous time or longer to determine the secular evolution of the disc, and we can also perform many simulations exploring a wide parameter space.

The full evolution equation for the accretion disc is Eq. 1.31 (Pringle, 1992) where the first term on the right hand side describes the usual viscous diffusion of mass. This is governed by the azimuthal shear viscosity ν_1 . The second term is also diffusive, governed by the vertical viscosity ν_2 , and is responsible for diffusing the disc tilt. The third term is an advective torque. Depending on its sign this advects angular momentum (and hence mass) inwards or outwards through the disc.

The main variable is the local angular momentum density of the ring, $\mathbf{L}(R, t)$. From this we can work out the surface density of the ring as

$$\Sigma = \frac{|\mathbf{L}|}{h} \quad (1.32)$$

⁵1D implies that all quantities only depend on radius

where $h(R, t) = R^2\Omega$ is the specific angular momentum of the ring of gas. We can also find the unit tilt vector for the ring as

$$\boldsymbol{\ell}(R, t) = \frac{1}{\Sigma h} \mathbf{L}(R, t). \quad (1.33)$$

From the tilt vector we can derive the angles for the tilt and twist of each ring. If we define a frame of reference (usually the angular momentum vector for the central object or the total angular momentum vector) then the tilt angle is the inclination of the ring relative to the frame of reference. The twist angle defines the azimuth of the tilt. The tilt and twist are the Euler angles in spherical polars and can be seen in Fig. 1 of Ogilvie (1999). Thus for an inclination angle $\beta(R, t)$ and a twist angle of $\gamma(R, t)$ we have the unit tilt vector as (Pringle, 1996, 1997)

$$\boldsymbol{\ell}(R, t) = (\ell_x, \ell_y, \ell_z) = (\cos \gamma \sin \beta, \sin \gamma \sin \beta, \cos \beta). \quad (1.34)$$

We can calculate the quantities β and γ from this by

$$\beta = \cos^{-1}(\ell_z) \quad (1.35)$$

and

$$\gamma = \tan^{-1}\left(\frac{\ell_y}{\ell_x}\right). \quad (1.36)$$

This code will be used in Chapter 5 to determine the evolution of a tilted accretion disc around a spinning black hole. This scenario has been extensively studied (Bardeen and Petterson, 1975; Pringle, 1992; Scheuer and Feiler, 1996; Natarajan and Pringle, 1998; Armitage and Natarajan, 1999; Natarajan and Armitage, 1999; Lodato and Pringle, 2006, etc.). However these investigations assume simple forms for the viscosities ν_1 and ν_2 . Ogilvie (1999) determines the relation between the

coefficients of the two viscosities for a particular value of the dimensionless disc viscosity parameter α (Shakura and Sunyaev, 1973) and a general warped disc. A relation between the viscosities was derived in the linear regime by Papaloizou and Pringle (1983)

$$\alpha_2 = \frac{1}{2\alpha_1} \quad (1.37)$$

where $\nu_1 = \alpha_1 c_s H$, $\nu_2 = \alpha_2 c_s H$ and $\alpha = \alpha_1$ is the Shakura & Sunyaev viscosity parameter. However Ogilvie (1999) showed that it was possible to derive the full nonlinear viscosity coefficients for a warped disc analytically (although the final solutions must be found numerically). Ogilvie (1999) showed that in the linear regime the analysis of Papaloizou and Pringle (1983) produces the correct relation (Eq. 1.37), but for strong warp amplitudes the viscosities can be markedly different. In Chapter 5 we determine the evolution of a tilted disc around a spinning black hole using the full effective viscosities derived in Ogilvie (1999).

1.5.2 Smoothed particle hydrodynamics

Smoothed Particle Hydrodynamics (SPH) is a Lagrangian numerical method for integrating the equations of motion for a fluid. It was first introduced by Lucy (1977) and Gingold and Monaghan (1977). SPH is a simple, intuitive and fast way of solving the equations of motion for fluid dynamics. This has led to SPH being the focus of much attention in the literature, its methodology being developed for over 30 years. There are now many good review papers (e.g. Monaghan 1992; Rosswog 2009; Springel 2010; Price 2012). In particular Price (2012) provides an excellent introduction to the basics and subtleties of the method. Rather than repeat much of the good work that is printed in these reviews I instead provide a brief description of the basic principles which underpin the method, focusing on the parts of greatest

relevance to this thesis. There are many variations of SPH, but the one used in this thesis follows Price (2012).

1.5.2.1 Calculating densities

The most fundamental part of SPH is the density calculation. For a set of particles with positions \mathbf{r}_j and masses m_j the density field can be approximated as

$$\rho(\mathbf{r}_i) = \sum_{j=1}^N m_j W(\mathbf{r}_i - \mathbf{r}_j, h_i) \quad (1.38)$$

where W is a smoothing kernel (see Section 1.5.2.2), whose size depends on the smoothing length⁶ h_i , and the sum is taken over all the N particles found within the range of the smoothing kernel. It is clear from the form of (1.38) that ρ is independent of time and thus depends only on the instantaneous particle positions and masses, and not on their histories. Also we can see that the properties of the density estimate will be inherited from the properties of the smoothing kernel W .

1.5.2.2 Smoothing kernels

For a good density estimate the kernel must satisfy at least the following properties:

- 1) $W(\mathbf{r} - \mathbf{r}_i, h) > 0$ for all \mathbf{r}_i , decreases monotonically from \mathbf{r} and has a smooth first derivative (continuous second derivative).
- 2) Isotropy. The kernel should be dependent only on the magnitude of the separation of the particles and the smoothing length, i.e. $W(\mathbf{r}_i - \mathbf{r}_j, h) = W(\mathbf{r}_j - \mathbf{r}_i, h) = W(|\mathbf{r}_i - \mathbf{r}_j|, h)$. Thus the estimate is independent of rotations and translations.

⁶Note: this is one of two existing formalisms. Here we define h_i as the smoothing length and the kernel size as ζh_i where ζ is dependent on the kernel function. Another formalism defines h_i as the kernel size.

3) Flat inner regions to prevent large fluctuations in the density estimate by a small movement of a close neighbour.

We now denote the kernel as $W(q_{ij})$ where $q_{ij} = |\mathbf{r}_i - \mathbf{r}_j|/h_i$.

The obvious choice for a kernel is a Gaussian:

$$W(q_{ij}) = \frac{\sigma}{h^d} \exp[-q_{ij}^2], \quad (1.39)$$

where d is the number of spatial dimensions and $\sigma = [1/\sqrt{\pi}, 1/\pi, 1/(\pi\sqrt{\pi})]$ is a normalisation factor in $[1, 2, 3]$ dimensions. However, it is clear that this kernel is non-zero across the whole domain. This requires that all particles contribute to the density estimate (and all other calculated quantities) leading to a numerical method that is $\mathcal{O}(N^2)$. This is restrictive computationally and so a kernel which goes to zero at a finite distance is usually employed.

The most commonly used class of kernels are the B-spline kernels (Monaghan and Lattanzio, 1985), of which the most used is the M_4 (cubic) spline truncated at $2h$ (i.e. the kernel size in units of the smoothing length $\zeta = 2$). This is given by $W(q) \equiv \frac{1}{h^d} w(q)$, where

$$w(q) = \sigma \begin{cases} \frac{1}{4}(2-q)^3 - (1-q)^3, & 0 \leq q < 1; \\ \frac{1}{4}(2-q)^3, & 1 \leq q < 2; \\ 0. & q \geq 2, \end{cases} \quad (1.40)$$

and σ is a normalisation constant given by $\sigma = [2/3, 10/(7\pi), 1/\pi]$ in $[1, 2, 3]$ dimensions.

If greater accuracy in the estimated quantities is required then higher order kernels can be used, such as the M_5 (quartic) or the M_6 (quintic). As the order of the spline kernels increases they approach the Gaussian.

1.5.2.3 Consistent h - ρ calculation

After choosing a smoothing kernel, Eq. 1.38 gives the density ρ if the smoothing length h is known. To take advantage of the Lagrangian nature of the method we should allow for variable h so that regions with a high number density of particles are not excessively smoothed. Thus we require a constraint on h :

$$h(\mathbf{r}_i) = \eta \left(\frac{m_i}{\rho_i} \right)^{1/d} \quad (1.41)$$

where η is a parameter (to be chosen by the user) which specifies the smoothing length in units of the mean particle spacing. Typically η is between 1.2 – 1.5 (Rosswog, 2009) and in this thesis we choose $\eta = 1.2$ – this value is stable to the pairing instability for the B-spline kernels (Price, 2012). It is clear that we should choose $\eta \gtrsim 1$ so that the kernel function has not significantly decreased before reaching the nearest neighbours, however we do not want to choose $\eta \gg 1$ as we will smooth out regions that could otherwise be resolved.

We can convert η into an average number of neighbours by

$$h \approx \eta \left(\frac{m}{\rho} \right)^{1/d} \approx \eta \left(\frac{m}{m N_{\text{neigh}}/V} \right)^{1/d} \quad (1.42)$$

where V is the volume of the kernel. This can then be rearranged to get

$$N_{\text{neigh}} \approx \eta^d \frac{V}{h^d} \quad (1.43)$$

which in three dimensions with $V = \frac{4}{3}\pi(\zeta h)^3$ is

$$N_{\text{neigh}} \approx \frac{4}{3}\pi(\zeta * \eta)^3. \quad (1.44)$$

For the M_4 cubic spline ($\zeta = 2$) with $\eta = 1.2$ this corresponds to ≈ 58 neighbours.

We now have two equations constraining both the density (1.38) and the smoothing length (1.41). To calculate these quantities self-consistently we must iterate between the two equations. This is done via a Newton–Raphson technique.

1.5.2.4 Equations of motion

We now have a method for self-consistently calculating the density and smoothing length of the particles. To evolve them in time we require an equation of motion. Following Price (2012) we can derive an equation of motion simply from the density (1.38) and the Lagrangian. The Lagrangian for a set of discrete particles is

$$L = \sum_j m_j \left[\frac{1}{2} v_j^2 - u_j(\rho_j, s_j) \right] \quad (1.45)$$

where m_j , v_j , u_j , ρ_j and s_j are the mass, velocity, internal energy, density and entropy respectively of particle j and the sum is over all particles.

To find the equations of motion we use the principle of least action to derive the Euler–Lagrange equations:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \mathbf{v}_i} \right) - \frac{\partial L}{\partial \mathbf{r}_i} = 0. \quad (1.46)$$

This shows explicitly that the method requires the density, velocity and thermal energy to be differentiable⁷, and thus we shall see below that extra terms are required to deal with discontinuities. Also we must specify the thermodynamics of the gas, i.e. how the internal energy is related to the density and pressure. As entropy is conserved, the first law of thermodynamics gives $du = PdV$, where the change in

⁷The density is differentiable as long as the kernel is differentiable.

particle volume is given by $dV = -m/\rho^2 d\rho$. We now also assume $P = K\rho^\gamma$, or more simply $P = c_s^2\rho$ for the pressure.

The equations of motion are then (Price, 2012)

$$\rho_i = \sum_j m_j W(q_{ij}), \quad (1.47)$$

$$\frac{d\mathbf{v}_i}{dt} = - \sum_j m_j \left[\frac{P_i}{\Omega_i \rho_i^2} \frac{\partial W_{ij}(q_{ij})}{\partial \mathbf{r}_i} + \frac{P_j}{\Omega_j \rho_j^2} \frac{\partial W_{ji}(q_{ji})}{\partial \mathbf{r}_i} \right], \quad (1.48)$$

$$\frac{du_i}{dt} = \frac{P_i}{\Omega_i \rho_i^2} \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \frac{\partial W_{ij}(q_{ij})}{\partial \mathbf{r}_i}. \quad (1.49)$$

Note that in the case of an isothermal equation of state the temperature of the gas remains constant and hence $du_i/dt = 0$.

1.5.2.5 Dissipative terms

We have discussed above the need for the Lagrangian to be differentiable. Therefore the current set of equations will break down in the presence of discontinuities in the flow, such as shocks. Therefore we include dissipation terms to smooth the velocities ('viscosity') and the internal energies ('conductivity').

The viscosity takes the form

$$\left(\frac{d\mathbf{v}_i}{dt} \right)_{\text{diss}} = \sum_j m_j \frac{\alpha_{\text{SPH}} v_{\text{sig}} (\mathbf{v}_i - \mathbf{v}_j) \cdot \hat{\mathbf{r}}_{ij}}{\bar{\rho}_{ij}} \hat{\mathbf{r}}_{ij} \bar{F}_{ij}. \quad (1.50)$$

where $\bar{F}_{ij} = [F_{ij}(h_i) + F_{ij}(h_j)]/2$ with $\nabla_i W_{ij}(q_{ij}) = \hat{\mathbf{r}}_{ij} F_{ij}(h_i)$ and the signal velocity is

$$v_{\text{sig}} = \begin{cases} \frac{1}{2}[c_{s,i} + c_{s,j} - \beta \mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij}]; & \mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij} \leq 0; \\ 0; & \mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij} > 0. \end{cases} \quad (1.51)$$

The thermal conductivity (combined with the term due to the viscosity) gives

$$\left(\frac{du_i}{dt}\right)_{\text{diss}} = - \sum_j \frac{m_j}{\bar{\rho}_{ij}} \left[\frac{1}{2} \alpha_{\text{SPH}} v_{\text{sig}} (\mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij})^2 + \alpha_u v_{\text{sig}}^u (u_i - u_j) \right] \bar{F}_{ij} \quad (1.52)$$

where v_{sig}^u need not be the same signal velocity as 1.51 (Price, 2008). In this thesis we do not use the conductivity term and thus set the coefficient $\alpha_u = 0$ throughout.

In principle these dissipative terms can be switched off when the flow is continuous, and so there have been many attempts to reduce the unwanted dissipation away from shocks and discontinuities (e.g. Morris and Monaghan, 1997; Price and Monaghan, 2005; Cullen and Dehnen, 2010). However in this thesis we take a different approach, where we use the viscosity term to represent a Shakura & Sunyaev disc viscosity.

Lodato and Price (2010) found that with suitable adjustments, the viscosity term (1.50) can provide a viscosity which “shows remarkable agreement with the analytical theory both in the linear and nonlinear regime”. To generate the disc viscosity from (1.50) the viscosity should be applied to both approaching and receding particles, switches on the viscosity should not be used, i.e. α_{SPH} should be constant in time and space, and finally the term should be multiplied by a factor $h/|r_{ij}|$ similar to the Monaghan (1992) viscosity scheme.

It can then be shown that this form of SPH viscosity leads to a Shakura & Sunyaev viscosity of the form (Murray, 1996; Lodato and Price, 2010)

$$\alpha = \frac{1}{10} \alpha_{\text{SPH}} \frac{\langle h \rangle}{H} \quad (1.53)$$

where $\langle h \rangle$ is the azimuthally averaged (strictly shell averaged for a warped disc) smoothing length.

1.5.2.6 Time-stepping

Finally, to integrate the equations of motion, we need to determine a suitable timestep for the integration. The two conditions imposed on the timestep are the Courant condition and a force condition. The Courant condition is

$$dt_C = \frac{h_i}{v_{\text{sig},i}}, \quad (1.54)$$

and the force condition is

$$dt_f = \left(\frac{h_i}{|a_i|} \right)^{1/2}, \quad (1.55)$$

where $|a_i|$ is the total acceleration of the particle. The final timestep is then chosen as the minimum of the two with a stability parameter $C \sim 0.25$. Thus

$$dt = C \min \{dt_C, dt_f\}. \quad (1.56)$$

For individual particle timesteps this is set for each particle. However if a global timestep is used then the minimum of (1.56) is taken over all particles.

2

Binary–Disc Alignment

2.1 Abstract

We consider the interaction between a binary system (e.g. two supermassive black holes or two stars) and an external accretion disc with misaligned angular momentum. This situation occurs in galaxy merger events involving supermassive black holes, and in the formation of stellar–mass binaries in star clusters. We work out the gravitational torque between the binary and disc, and show that their angular momenta $\mathbf{J}_b, \mathbf{J}_d$ stably counteralign if their initial orientation is sufficiently retrograde, specifically if the angle θ between them obeys $\cos\theta < -J_d/2J_b$, on a time short compared with the mass gain time of the central accretor(s). The magnitude J_b remains unchanged in this process. Counteralignment can promote the rapid merger of supermassive black hole binaries (see Chapter 4), and possibly the formation of coplanar but retrograde planets around stars in binary systems.

2.2 Introduction

Galaxy mergers are commonly thought to be the main mechanism driving the co-evolution of galaxies and their central supermassive black holes (SMBH). In such a merger we expect the formation of a SMBH binary in the centre of the merged galaxy. Gravitational waves quickly drive the binary to coalesce if the orbital separation can be shortened to $\lesssim 10^{-2}\text{pc}$. The binary may stall at a separation greater than this if the interaction with the merged galaxy is not efficient enough in extracting orbital angular momentum and energy. For the stellar component of the galaxies this occurs at approximately a parsec, creating “the final parsec problem” (Begelman et al., 1980; Milosavljević and Merritt, 2001). There have been many papers exploring potential solutions to this problem, for example a sling-shot mechanism involving a triple SMBH system (Iwasawa et al., 2006), efficient refilling of

the binary loss cone by angular momentum exchange between stellar orbits and a triaxial dark matter halo (Berczik et al., 2006) and also the evolution of the binary with a prograde accretion disc (circumbinary discs: Armitage and Natarajan 2005; MacFadyen and Milosavljević 2008; Lodato et al. 2009; Cuadra et al. 2009 and embedded discs: Escala et al. 2005; Dotti et al. 2007, 2009). In a recent paper (Nixon et al., 2011a, Chapter 4) we explored the evolution of a binary interacting with a *retrograde* circumbinary accretion disc. We showed that this is more efficient than a prograde disc in removing binary orbital angular momentum and energy. This is simply because there are no orbital resonances between the binary and the disc and thus there is direct accretion of retrograde gas on to the binary.

Here we consider the alignment process between a binary system and an external misaligned accretion disc. This situation can arise in at least two astronomical contexts. A merger event between galaxies can produce an SMBH binary in the centre of the merged galaxy, and this or a later accretion event may surround the hole with a disc of accreting gas. A similar situation arises during the formation of stars in a cluster. A binary system may form, but also capture gas into an external disc.

In both of these cases, there is no compelling reason to assume that the binary and disc rotation are initially parallel or even roughly coaligned (cf. King and Pringle 2006). As we shall see, the gravitational interaction between the binary and the disc generates differential precession in the disc gas, and thus viscous dissipation. This gives a dissipative torque which vanishes only when the binary and disc angular momenta $\mathbf{J}_b, \mathbf{J}_d$ are either parallel or antiparallel. In all such cases, the torque diffuses the tilt or warp through the disc (cf. Pringle 1992, 1999; Wijers and Pringle 1999) driving the system to one of these equilibria. The existence of a warp makes the precise definition of disc angular momentum \mathbf{J}_d quite subtle and we return to

this point in Section 2.5.

The binary–external disc interaction is very similar to the effect of the Lense–Thirring (LT) precession on an accretion disc around a spinning black hole (Bardeen and Petterson, 1975; Pringle, 1992; Scheuer and Feiler, 1996; Natarajan and Pringle, 1998; Armitage and Natarajan, 1999; Natarajan and Armitage, 1999; Nelson and Papaloizou, 2000; Lodato and Pringle, 2006, etc.) if we replace \mathbf{J}_b by the hole spin angular momentum \mathbf{J}_h . For some years it was thought that the LT interaction always led to co–alignment (i.e. \mathbf{J}_b and \mathbf{J}_d parallel). However King et al. (2005) (hereafter KLOP) showed on general grounds that counteralignment does occur, if (and only if) the initial angle θ between \mathbf{J}_d and \mathbf{J}_h satisfies $\cos \theta < -J_d/2J_h$, where $J_d = |\mathbf{J}_d|$ and $J_h = |\mathbf{J}_h|$. Scheuer and Feiler (1996) had implicitly assumed $J_d \gg J_h$ and so enforced co–alignment. With this restriction lifted, King and Pringle (2006, 2007) and King et al. (2008) showed that accretion from a succession of randomly–oriented discs leads to spindown of the supermassive black hole, allowing rapid mass growth.

Here we examine the alignment process for a binary and an external disc. We show that the argument of KLOP is generic, and that the disc and binary counteralign if and only if $\cos \theta < -J_d/2J_b$. As a result it is quite possible for SMBH binaries to be surrounded by a completely retrograde disc which strongly promotes coalescence (cf. Nixon et al. 2011a, Chapter 4). In the case of a newly–formed stellar binary, the presence of a counteraligned disc can lead to the formation of planets with retrograde orbits.

2.3 The binary–disc torque

We consider a binary system with masses M_1, M_2 and a circular orbit, with the binary angular momentum vector pointing along the z -axis of cylindrical polar coordinates (R, ϕ, z) . For simplicity we assume $M_2 \ll M_1$ and place M_1 at the origin, with M_2 orbiting at radius a in the (R, ϕ) plane (our conclusions are not affected by this assumption). The orbit has angular velocity

$$\Omega_b = \left[\frac{G(M_1 + M_2)}{a^3} \right]^{1/2}. \quad (2.1)$$

Now we consider a disc particle in an orbit about the binary at radius $R \gg a$. If both the small quantities M_2/M_1 and a/R actually vanished, the particle’s orbit would be a circle, with angular velocity $(GM_1/R^3)^{1/2}$. When these quantities are small but finite they induce various perturbations in the orbit. Some of these perturbations have (inertial–frame) frequency $2\Omega_b$ and higher multiples. These are oscillatory, and have no long–term secular effect. Long–term effects on the orbit, and hence eventually on the disc, come from the zero–frequency (azimuthally symmetric $m = 0$) term in the binary potential. This point is considered in more detail in Bate et al. (2000), who considered the related problem of a disc around the primary mass M_1 (i.e. $M_2 \ll M_1$, but $R \ll a$).

Physically this $m = 0$ term is given by replacing the orbiting mass M_2 with the same mass spread uniformly over its orbit, i.e. a ring of mass M_2 and radius a in the (R, ϕ) plane. Adding in the potential from the fixed point mass M_1 at the origin we find the effective gravitational potential experienced by a disc particle as

$$\Phi(R, z) = -\frac{GM_1}{(R^2 + z^2)^{1/2}} - \frac{GM_2}{2\pi} \int_0^{2\pi} \frac{d\phi}{r} \quad (2.2)$$

where r is the distance between the particle position and a point on the ring at (a, ϕ, z) , i.e.

$$r^2 = R^2 + a^2 + z^2 - 2Ra \cos \phi. \quad (2.3)$$

We now expand Eq. 2.2 in powers of a/R and z/R , keeping terms only up to second order. This gives

$$\Phi(R, z) = -\frac{G(M_1 + M_2)}{R} + \frac{GM_2 a^2}{4R^3} + \frac{G(M_1 + M_2)z^2}{2R^3} - \frac{9GM_2 a^2 z^2}{8R^5} + \dots \quad (2.4)$$

The orbital frequency Ω of the particle subject to this potential is given by

$$\Omega^2 = \frac{1}{R} \frac{\partial \Phi}{\partial R} \quad (2.5)$$

and its vertical oscillation frequency ν by

$$\nu^2 = \frac{\partial^2 \Phi}{\partial z^2}, \quad (2.6)$$

both evaluated at $z = 0$. The nodal precession frequency is $\Omega_p = \Omega - \nu$ and we find

$$\Omega_p(R) = \frac{3}{4} \left[\frac{G(M_1 + M_2)}{R^3} \right]^{1/2} \frac{M_2}{M_1 + M_2} \frac{a^2}{R^2}. \quad (2.7)$$

This frequency is very similar to that for LT precession around a spinning black hole (e.g. Scheuer and Feiler 1996), which goes as R^{-3} rather than the $R^{-7/2}$ here. Equation 2.7 is formally almost identical to the precession frequency found by Bate et al. (2000) for a disc around the primary, although derived for $R \gg a$ rather than $R \ll a$. The same argument as in that paper shows that if the disc and binary axis are misaligned by an angle θ (called δ in Bate et al. 2000) with $0 \leq \theta \leq \pi/2$, the precession frequency is just multiplied by $\cos \theta$. The opposite case with the

disc somewhat counteraligned (i.e. $\theta > \pi/2$) is equivalent to the $\theta < \pi/2$ case with the binary angular momentum reversed. But this reversal leaves the precession frequency unchanged, since we are dealing only with the $m = 0$ part of the potential. So for all θ with $0 < \theta < \pi$ the precession frequency is

$$\Omega_p(\theta) = \Omega_p |\cos \theta|. \quad (2.8)$$

2.4 Co– or counter–alignment?

We have shown above that the effect of the binary potential on the disc is to induce precession of the disc orbits. This precession is strongly dependent on radius: rings of gas closer to the binary precess faster. The differential precession creates a dissipative torque between adjacent rings of gas tending to make $\theta \rightarrow 0, \pi$ so that the precession ultimately vanishes.

The precession timescale in the disc increases with radius (cf. Eq. 2.7). The torque therefore acts faster at smaller radii to co– or counter–align disc orbits with the binary plane. This leads to the creation of a warp in the disc, where the inner parts are co– or counter–aligned and the outer parts are still misaligned (cf. Fig. 2.1). This warp propagates outwards, eventually co– or counter–aligning the entire disc with the binary plane. This effect was solved numerically for discs warped under the LT effect by Lodato and Pringle (2006).

Now we argue as in KLOP that since each ring feels a precession, the resultant back–reaction on the binary is a sum of precessions, which is just a precession. This argument is equivalent to that presented in Bate et al. (2000) who argue that because the binary potential is symmetric about the plane of the binary, the disc–binary torque cannot have a component in the direction of \mathbf{J}_b . Accordingly \mathbf{J}_b can

only precess. These arguments show that we can write the torque on the binary in the same form as the LT–induced torque on a spinning black hole considered in KLOP, i.e.

$$\frac{d\mathbf{J}_b}{dt} = -K_1[\mathbf{J}_b \wedge \mathbf{J}_d] - K_2[\mathbf{J}_b \wedge (\mathbf{J}_b \wedge \mathbf{J}_d)]. \quad (2.9)$$

Here K_1, K_2 are coefficients depending on disc properties. The first term gives the magnitude and sign of the torque inducing the precession. It does not change the alignment angle θ . The second term describes the torque which changes θ . The same arguments as in KLOP for the LT case, and Bate et al. (2000) for a disc around the primary, show that dissipation in the disc requires K_2 to be a positive quantity. Its magnitude depends on the properties of the disc and the binary. The one difference from the LT case is that the $|\cos \theta|$ dependence means that the sign of the coefficient K_1 can be either positive or negative. But this difference has no effect on the conditions under which the disc and binary co– or counter–align. These are formally identical with the ones for the LT case derived by KLOP, with the binary angular momentum \mathbf{J}_b replacing the hole spin angular momentum \mathbf{J}_h . The process has a different timescale specified by the different magnitude of K_2 .

The same arguments as in KLOP now show that the magnitude J_b of the binary angular momentum remains constant, while the direction of \mathbf{J}_b aligns with the total angular momentum $\mathbf{J}_t = \mathbf{J}_b + \mathbf{J}_d$, which is of course a constant vector. During this process the magnitude of J_d^2 decreases because of dissipation (KLOP). Counteralignment ($\theta \rightarrow \pi$) occurs if and only if $J_b^2 > J_t^2$. By the cosine theorem

$$J_t^2 = J_b^2 + J_d^2 - 2J_b J_d \cos(\pi - \theta), \quad (2.10)$$

so this is equivalent to

$$\cos \theta < -\frac{J_d}{2J_b}. \quad (2.11)$$

Thus counteralignment of a binary and an external disc is possible, and requires

$$\theta > \pi/2, \quad J_d < 2J_b. \quad (2.12)$$

2.5 Discussion

So far we have avoided fully spelling out the meaning of the disc angular momentum J_d . This is complicated because the binary torque falls off very strongly with radius, and so a large contribution to the angular momentum in a distant part of the disc may be irrelevant to the alignment process, or affect this process in a time–dependent way (cf. Lodato and Pringle 2006). Sections 3 and 4 of KLOP discuss these questions in more detail. Effectively J_d can be thought of as the disc angular momentum inside the warp radius, and therefore a time–dependent quantity.

At early times J_d is small, as only a fraction of the total gas interacts with the binary. Counter–alignment may occur if $\theta > \pi/2$, but at later times, as J_d grows and more gas is able to interact with the binary, alignment eventually happens (when $J_d > 2J_b$). So if $\theta > \pi/2$, even for $J_d > 2J_b$ we expect $\sim 2J_b |\cos \theta|$ of disc angular momentum to counteralign with the binary before the outer disc comes to dominate and enforce coalignment (cf. Lodato and Pringle 2006).

The typical timescale for co– or counter–alignment for a SMBH binary is

$$t_{\text{binary}} \simeq \frac{J_b}{J_d(R_w)} \frac{R_w^2}{\nu_2} \quad (2.13)$$

where R_w is the warp radius, $J_d(R_w)$ is the disc angular momentum within R_w , and ν_2 is the vertical disc viscosity. This equation describes the transfer of angular momentum in the disc to the hole, this is mediated by the ν_2 viscosity flattening

tilted rings of gas into the plane of the hole, which occurs at the warp radius R_w . This is identical to the formal expression for LT alignment of a spinning black hole if we replace the spin angular momentum J_h with J_b (cf. Scheuer and Feiler 1996). The warp radius is given by equating the precession time $1/\Omega_p(R)$ to the vertical viscous time R^2/ν_2 . Inside this radius the precession timescale is short and the disc dissipates and co- or counter-aligns with the binary plane. Outside this radius the disc is not dominated by the precession and so maintains its misaligned plane. The connecting region therefore takes on a warped shape shown in Fig. 2.1. As time passes the warp propagates outwards and co- or counter-aligns the entire disc with the binary plane.

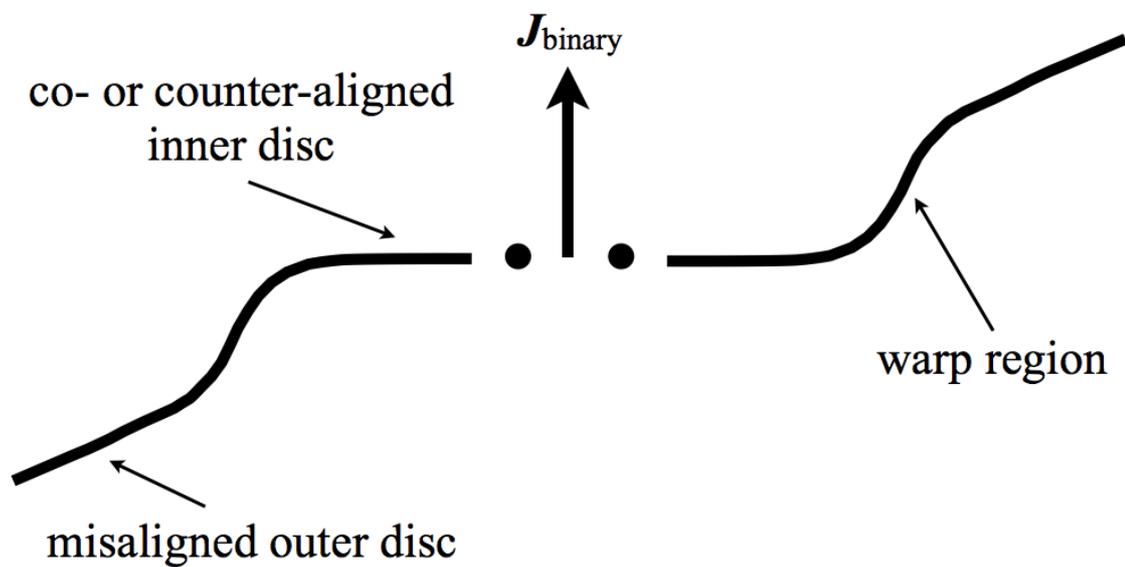


Figure 2.1: The warped disc shape expected after the inner disc co- or counter-aligns with the binary plane but the outer disc stays misaligned. Eventually the entire disc will co- or counter-align with the binary plane, depending on the global criterion (Eq. 2.12). Note that in practice precession makes the warp non-axisymmetric.

Approximating the disc angular momentum as

$$J_d(R_w) \sim \pi R_w^2 \Sigma (GM R_w)^{1/2} \quad (2.14)$$

with Σ the disc surface density and $M = M_1 + M_2$, and using the steady–state disc relation $\dot{M} = 3\pi\nu\Sigma$ we find

$$t_{\text{binary}} \sim 3 \frac{M_2}{M_1} \left(\frac{a}{R_w} \right)^{1/2} \frac{\nu_1}{\nu_2} \frac{M}{\dot{M}}, \quad (2.15)$$

where we have also used (e.g. Frank et al., 2002)

$$J_b = M_1 M_2 \left(\frac{Ga}{M} \right)^{1/2}. \quad (2.16)$$

Since $\nu_1 < \nu_2$ (Papaloizou and Pringle, 1983), $a \ll R_w$ and $M_2 < M_1$, we see that alignment takes place on a timescale shorter than the mass growth of the central accretor(s).

The timescale (2.15) is directly analogous to the expression

$$t_{\text{LT}} \sim 3a_* \left(\frac{R_s}{R_w} \right)^{1/2} \frac{\nu_1}{\nu_2} \frac{M}{\dot{M}} \quad (2.17)$$

for alignment under the LT precession, where $a_* < 1$ is the Kerr spin parameter and R_s the Schwarzschild radius of the spinning hole. Evaluating R_w in the two cases we find

$$\frac{t_{\text{LT}}}{t_{\text{binary}}} \sim \frac{3^{1/2}}{2} \left(\frac{a_*}{M_2/M_1} \right)^{1/2} \left(\frac{a}{R_s} \right)^{1/4}. \quad (2.18)$$

Thus in general, provided we assume that the ratio ν_1/ν_2 is similar in the two cases and that the hole spin is not rather small ($a_* < (R_s/a)^{1/2}(M_2/M_1)$), then the binary–disc alignment is rather faster than the corresponding process for spinning

black holes.

Our result has significant consequences for SMBH binaries. For random orientations, Eq. 2.12 shows that initial disc angles leading to alignment occur significantly more frequently than those giving counteralignment only if $J_d > 2J_b$. (In the LT case this fact leads to a slow spindown of the hole, because retrograde accretion has a larger effect on the spin, King et al. 2008.) A number of studies (Armitage and Natarajan 2005; MacFadyen and Milosavljević 2008; Cuadra et al. 2009; Lodato et al. 2009) have shown that prograde external discs are rather inefficient in shrinking SMBH binaries and solving the final parsec problem. This is essentially because of resonances within the disc. In contrast, the slightly rarer retrograde events have a much stronger effect on the binary. These rapidly produce a counterrotating but coplanar accretion disc external to the binary, which has no resonances. We note that Nixon et al. (2011a, Chapter 4) show that the binary gradually increases its eccentricity as it captures negative angular momentum from the disc, ultimately coalescing once this cancels its own. A non-zero binary eccentricity changes the detailed form of the perturbing potential from that in Eq. 2.4, but cannot change the precessional character leading to the torque equation (2.9). Our results remain unchanged, particularly the counteralignment condition (2.12), apart from minor modifications of the timescale (2.15).

Thus in a random sequence of accretion events producing external discs, the prograde events have little effect, and the retrograde ones shrink the binary. In particular, a sequence of minor retrograde events with $J_d < J_b$ has a cumulative effect and must ultimately cause the binary to coalesce once the total retrograde $\sum J_d = J_b$. This is important, since the disc mass is limited by the onset of self-gravity to $M_d \lesssim (H/R) M_1$ (cf. King, Pringle and Hofmann 2008). Coalescence will then occur once the retrograde discs have brought in a total mass M_2 , i.e. once a sequence

of $\gtrsim (M_2/M_1)(R/H)$ retrograde discs have accreted (see Chapter 4). For minor mergers this requires at most a few randomly oriented accretion disc events, rising to a few hundred for major mergers ($q > 0.1$).

We note finally that similar considerations apply in planet–forming discs around stellar binary systems, which can also be initially misaligned (Bate et al., 2010). This may offer a way of making retrograde planets in binaries, as recently suggested for ν Octantis (Eberle and Cuntz, 2010).

3

Stable counteralignment

3.1 Abstract

In general, when gas accretes on to a supermassive black hole binary it is likely to have no prior knowledge of the binary angular momentum. Therefore a circumbinary disc forms with a random inclination angle θ to the binary. It is known that for $\theta < 90^\circ$ the disc will coalign with respect to the binary. If $\theta > 90^\circ$ the disc wholly counteraligns if it satisfies $\cos\theta < -J_d/2J_b$, where J_d and J_b are the magnitudes of the disc and binary angular momentum vectors respectively. If however $\theta > 90^\circ$ and this criterion is not satisfied the same disc may counteralign its inner regions and, on longer timescales, coalign its outer regions. I show that for typical disc parameters, describing an accretion event on to a supermassive black hole binary, a misaligned circumbinary disc is likely to wholly co- or counter-align with the binary plane. This is because the binary angular momentum dominates the disc angular momentum. However with extreme parameters (binary mass ratio $M_2/M_1 \ll 1$ or binary eccentricity $e \sim 1$) the same disc may simultaneously co- *and* counter-align. It is known that coplanar prograde circumbinary discs are stable. I show that coplanar retrograde circumbinary discs are also stable. A chaotic accretion event on to an SMBH binary will therefore result in a coplanar circumbinary disc that is either prograde or retrograde with respect to the binary plane.

3.2 Introduction

Galaxy mergers are commonly thought to be the main mechanism driving the co-evolution of galaxies and their central supermassive black holes (SMBH). An SMBH binary is likely to form in the centre of the merged galaxy and subsequently accrete from a circumbinary disc. It is reasonable to expect that the angular momentum of the binary and that of the accreting gas are uncorrelated, and that each has no

prior knowledge of the other. We can therefore expect a random distribution of orientations for such circumbinary discs.

Understanding the evolution of a misaligned circumbinary disc is needed if we are to understand the evolution of SMBH binaries. In a recent paper (Nixon et al., 2011a, Chapter 4) we showed that a retrograde circumbinary disc can be very efficient in extracting angular momentum from the binary orbit. This may offer a solution to the final parsec problem (Begelman et al. 1980; Milosavljević and Merritt 2001).

In another recent paper (Nixon et al., 2011b, hereafter NKP, Chapter 2) we showed that the dominant effect of the binary potential, on a misaligned circumbinary disc, is to induce radially-dependent precessions of the misaligned disc particle orbits. This differential precession is known to induce warping of the disc as rings of gas dissipate energy through viscosity. The precession vanishes only when the orbit is in the plane of the binary (either prograde or retrograde) and thus we expect that both prograde and retrograde orbits are possible equilibria for the gas. NKP showed that the evolution of a misaligned circumbinary disc is formally similar to the evolution of a misaligned disc around a spinning compact object. In this case the evolution is driven by the Lense–Thirring effect (e.g. Bardeen and Petterson 1975; Pringle 1992; Scheuer and Feiler 1996; Lodato and Pringle 2006; Nixon and King 2012; Chapter 5). This implies that the analysis of King et al. (2005), which calculates the conditions for co- or counter-alignment, holds for circumbinary discs as well. The disc co- or counter-aligns depending on the magnitudes and directions of \mathbf{J}_b and \mathbf{J}_d , the angular momentum of the binary and disc respectively. The whole disc counteraligns with the binary if the initial inclination angle of the disc, θ , and the magnitudes of the disc and binary angular momentum, J_d and J_b respectively, satisfy

$$\cos \theta < -\frac{J_d}{2J_b}. \quad (3.1)$$

NKP only considered the zero-frequency (azimuthally symmetric $m = 0$) term in the binary potential. This is a reasonable approach as all other terms induce oscillatory effects which cancel out on long timescales.

It is natural to assume that coaligned discs are stable, however in the past counteraligned discs have been incorrectly found to be unstable (Scheuer and Feiler, 1996). I therefore use a full three dimensional hydrodynamic approach to check the assumptions in NKP are valid and to confirm that co- and counter-alignment are stable. In Section 3.3 I discuss the possibility of both co- and counter-alignment of the same disc (cf. Lodato and Pringle 2006). In Section 3.4 I report a simulation of a counteraligning disc and give my conclusions in Section 3.5.

3.3 Simultaneous co- and counter-alignment?

Lodato and Pringle (2006) considered the alignment of a disc and a spinning black hole. They showed that for a disc where $\theta > \pi/2$ and $J_d > 2J_h$, initial counteralignment of the inner disc occurs, followed by subsequent coalignment of the outer disc. During the alignment of the outer disc the inner disc retains its retrograde nature and so a warp of significant amplitude is achieved ($\Delta\theta \sim \pi$). This scenario produces a disc which is simultaneously counteraligned (in the inner parts) and coaligned (in the outer parts). For the subsequent evolution of such a disc see Nixon et al. (2012, Chapter 6).

If however the angular momenta are such that $J_d \lesssim J_h$ then the disc can only wholly co- or counter-align and the above process is impossible. In this section we discuss whether it is feasible to have circumbinary discs such that both co- and counter-alignment are simultaneously possible. This possibility is constrained as the mass of the disc is limited by self-gravitational collapse and the radius of the disc is limited

by the gravitational sphere of influence of the binary. Therefore there must be a maximum feasible disc angular momentum.

The disc simultaneously co- and counter-aligns if (and only if) $\theta > \pi/2$ and the condition (3.1) does not hold. So for randomly aligned discs, simultaneous co- and counter-alignment is only possible if $J_d \gtrsim 2J_b$ (assuming $\cos \theta \sim -1$). Thus I derive the condition for $J_d \gtrsim 2J_b$. The disc angular momentum is

$$J_d \sim M_d \sqrt{GM R_d} \quad (3.2)$$

where M_d is the mass in the disc, M is the total binary mass, R_d is a characteristic radius for the disc and G is the gravitational constant. We note that the definition of J_d in this case is simply the total angular momentum in the disc. The exact definition of J_d is usually more subtle because the disc takes time to communicate its angular momentum.

The angular momentum of a binary with eccentricity e is

$$J_b = \mu \sqrt{GM a (1 - e^2)} \quad (3.3)$$

where M is the total mass of the binary and μ is the reduced mass ($\mu \sim M_2$ for $M_2 \ll M_1$).

So for simultaneous co- and counter-alignment we require

$$M_d \sqrt{GM R_d} \gtrsim 2\mu \sqrt{GM a (1 - e^2)}. \quad (3.4)$$

Self-gravity limits $M_d \lesssim (H/R)M$, so we get

$$(H/R)M \sqrt{GM R_d} \gtrsim 2\mu \sqrt{GM a (1 - e^2)}. \quad (3.5)$$

After a bit of algebra this tells us that we can only get simultaneous co- and counter-alignment if

$$\frac{R_d}{a} \gtrsim 4 \left(\frac{R}{H}\right)^2 \left(\frac{M_2}{M}\right)^2 (1 - e^2), \quad (3.6)$$

where we have used the approximation $\mu \sim M_2$.

For the most optimistic parameters we have $a \sim 0.1$ pc and the sphere of influence of the binary ~ 10 pc. This gives the left hand side as $\lesssim 100$. For typical disc thickness ($H/R \sim 10^{-3}$) it is clear that binaries with a low eccentricity ($e \approx 0$) we cannot get simultaneous co- and counter-alignment unless the mass ratio is extreme ($M_2/M_1 \lesssim 10^{-2}$). However if the disc is very thick then this is possible. For small binary mass-ratios the disc mass may be greater than the mass of the secondary, therefore the hydrodynamic drag on the binary may become significant (Ivanov et al., 1999). If the binary is significantly eccentric, as predicted for binaries that have accreted through a retrograde disc (Nixon et al., 2011a, Chapter 4) then simultaneous co- and counter-alignment may be possible. I shall return to these possibilities in future work. However these arguments suggest that it is reasonable to assume an SMBH binary must wholly co- or counter-align the disc with the binary plane.

3.4 Simulation

To confirm the stable counteralignment of a circumbinary disc I perform one simple simulation. I use the SPH code PHANTOM, a low-memory, highly efficient SPH code optimised for the study of non-self-gravitating problems. This code has performed well in related simulations. For example Lodato and Price (2010) simulated warped accretion discs and found excellent agreement with the analytical work of Ogilvie (1999) on the nature of the internal accretion disc torques.

The implementation of accretion disc α -viscosity (Shakura and Sunyaev, 1973) in PHANTOM is described in Lodato and Price (2010). Specifically, we use the ‘artificial viscosity for a disc’ described in Sec. 3.2.3 of Lodato and Price (2010), similar to earlier SPH accretion disc calculations (e.g. Murray, 1996). The main differences compared to standard SPH artificial viscosity are that the disc viscosity is applied to both approaching and receding particles and that no switches are used. The implementation used here differs slightly from Lodato and Price (2010) in that the β^{AV} term in the signal velocity is retained in order to prevent particle interpenetration. The disc viscosity in PHANTOM was extensively calibrated against the 1D thin α -disc evolution in Lodato and Price (2010) (c.f. Fig. 4 in that paper) and the disc scale heights employed here are similar. We use $\alpha^{\text{AV}} = 1$ and $\beta^{\text{AV}} = 2$ which at the employed resolution (see below) corresponds to a physical viscosity $\alpha_{\text{SS}} \approx 0.05$. Note that the initial viscosity is slightly smaller than this, however during the simulation the disc spreads and particles are accreted and thus the spatial resolution decreases which increases the viscosity (see Eq. 38 of Lodato and Price 2010).

3.4.1 Setup

The simulation has two equal mass sink particles representing the binary (of total mass unity in code units), on a circular orbit with separation 0.5 (in code units). The sink particles effect the gas only by Newtonian gravity. Initially the circumbinary gas disc is flat and composed of 1 million SPH particles, in hydrostatic equilibrium from 1.0 to 2.0 in radius, and surface density distribution $\Sigma \propto R^{-1}$, set up by means of the usual Monte-Carlo technique. This places particles in random positions following the required surface density profile and with a Gaussian vertical density profile. The vertical hydrostatic equilibrium corresponds to $H/R = 0.05$ at $R = 1$. The

equation of state for the gas is isothermal. The disc is initially tilted at 170° to the binary plane. Any gas which falls within a radius of 0.5 from the binary centre of mass is removed as it no longer has any affect on the alignment of the circumbinary disc. The disc mass is negligible in comparison to the binary mass and thus the gravitational back-reaction of the gas on the binary is not included.

3.4.2 Geometry

The binary is taken to orbit in the x - y plane with binary angular momentum vector in the z -direction. We define the ‘tilt’ and the ‘twist’ of the disc with respect to the binary using Euler angles (e.g. Bardeen and Petterson 1975; Pringle 1996), where the unit angular momentum vector in the disc is described at any radius by

$$\boldsymbol{\ell} = (\cos \gamma \sin \beta, \sin \gamma \sin \beta, \cos \beta) \quad (3.7)$$

with $\beta(R, t)$ the local angle of disc tilt with respect to the z -axis and $\gamma(R, t)$ the local angle of disc twist measured from the x -axis.

3.4.3 Stable counteralignment

As predicted by NKP the dominant effect of the binary on the disc is to induce precessions in the gas orbits. As the radial range of the disc in this simulation is only a factor of two, the precession rate changes little between the inner and outer parts of the disc. However there is still a differential precession across the disc, which leads to a twist. This causes dissipation between rings of gas and so a small amplitude warp in the disc (cf. Fig. 3.1). The disc angular momentum vector precesses around the binary angular momentum vector for the duration of

the simulation. In Fig. 3.2 the twist angle in the disc (at unit radius) is plotted against time. As the twist angle is defined between $\pm 180^\circ$ the plot has a ‘saw-tooth’ shape, as the disc precesses until the twist is -180° which is then equivalent to 180° .

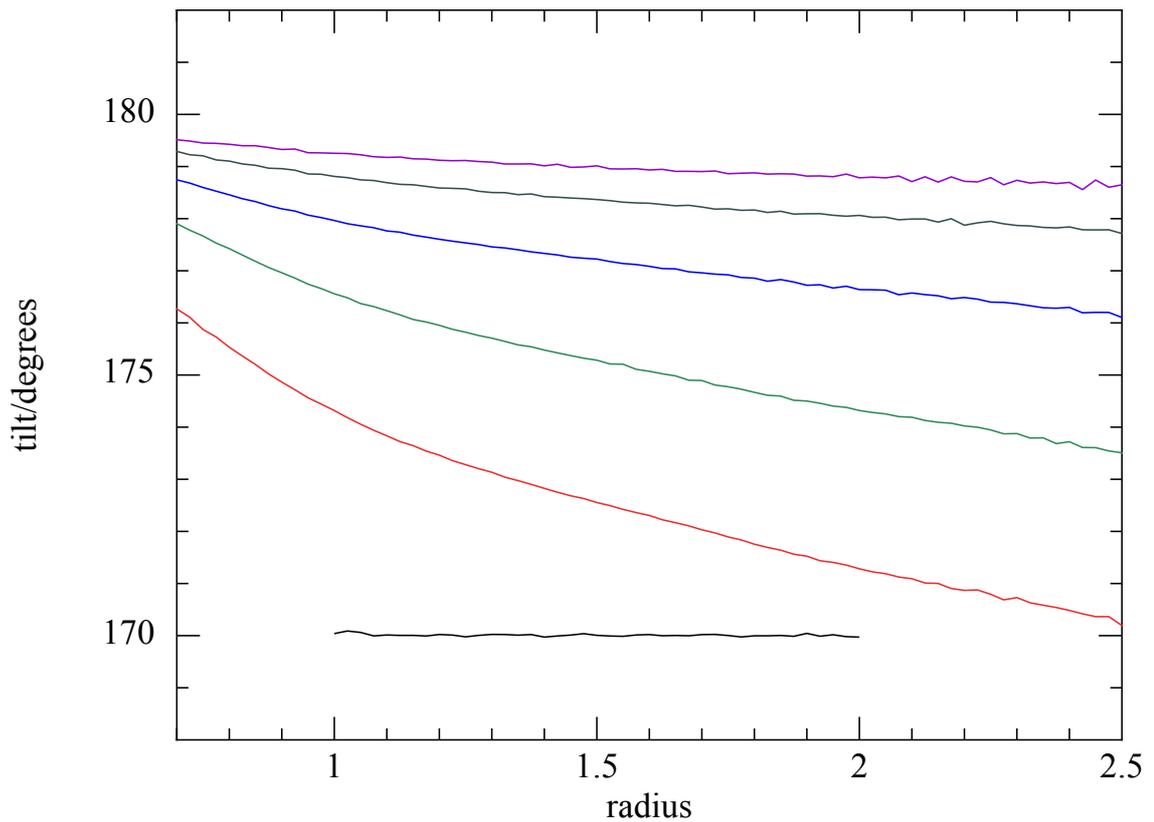


Figure 3.1: The tilt angle in the disc as a function of radius. This is plotted at six different times corresponding to (in units of the dynamical time at $R = 1$) $t = 0$ (black), $t = 500$ (red), $t = 1000$ (green), $t = 1500$ (blue), $t = 2000$ (grey), $t = 2500$ (purple). Initially the disc is flat with a global tilt of 170° . In time the disc spreads and counteraligns with respect to the binary plane.

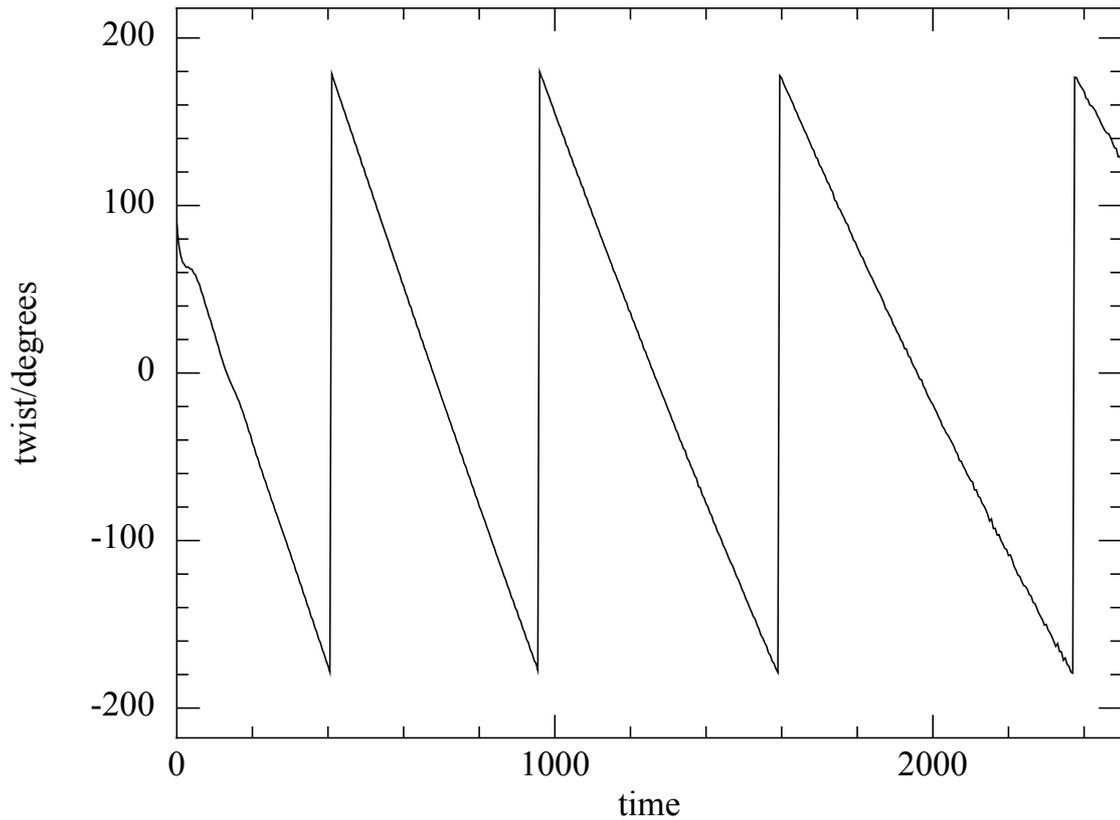


Figure 3.2: The twist angle (angle from x -axis to the line of nodes) at $R = 1$ in the disc as a function of time. Initially the line of nodes is in the y -direction and hence the angle is 90° . The binary potential induces precession of the gas orbits. The twist is calculated between $\pm 180^\circ$ which generates the ‘saw-tooth’ structure in the plot.

Figs. 3.3 & 3.4 show the disc structures at various times. Fig. 3.3 shows the disc column density viewed face on, and Fig. 3.4 shows it edge on. In Fig. 3.3 we see the usual viscous spreading of the disc, and also the lack of any resonances in the disc. This is expected in a counterrotating circumbinary accretion disc (Papaloizou and Pringle 1977; Nixon et al. 2011a; Chapter 4). In Fig. 3.4 we can see the tilt of the disc. Initially the disc is tilted by 170° , but then precesses and counteraligns with the binary plane.

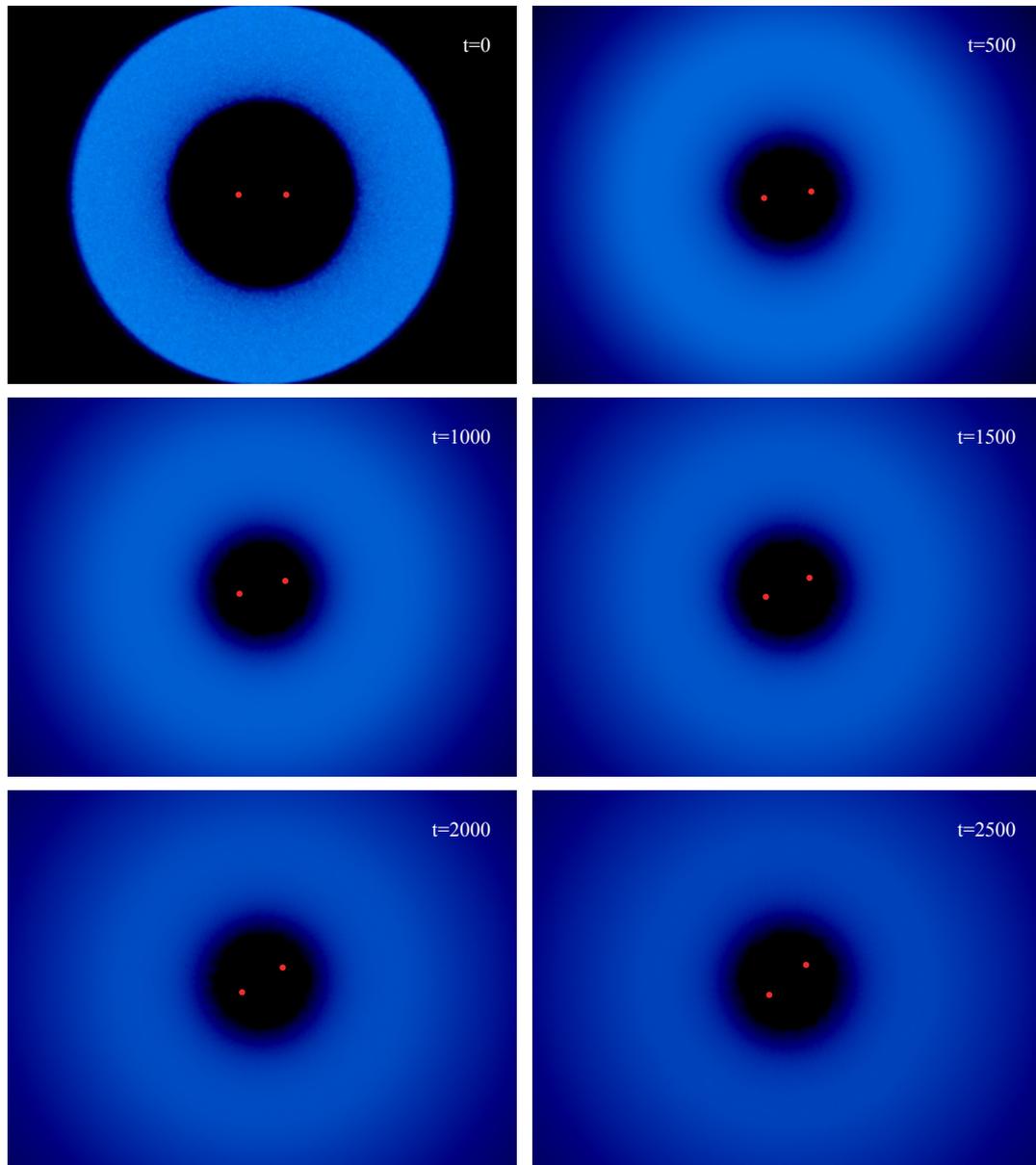


Figure 3.3: Face-on column density rendering of the disc at various times in the simulation. Each component of the binary is represented by a red filled circle. Over time the disc spreads and precesses around the binary. The precession induces dissipation in the disc which aligns the disc with the binary plane.

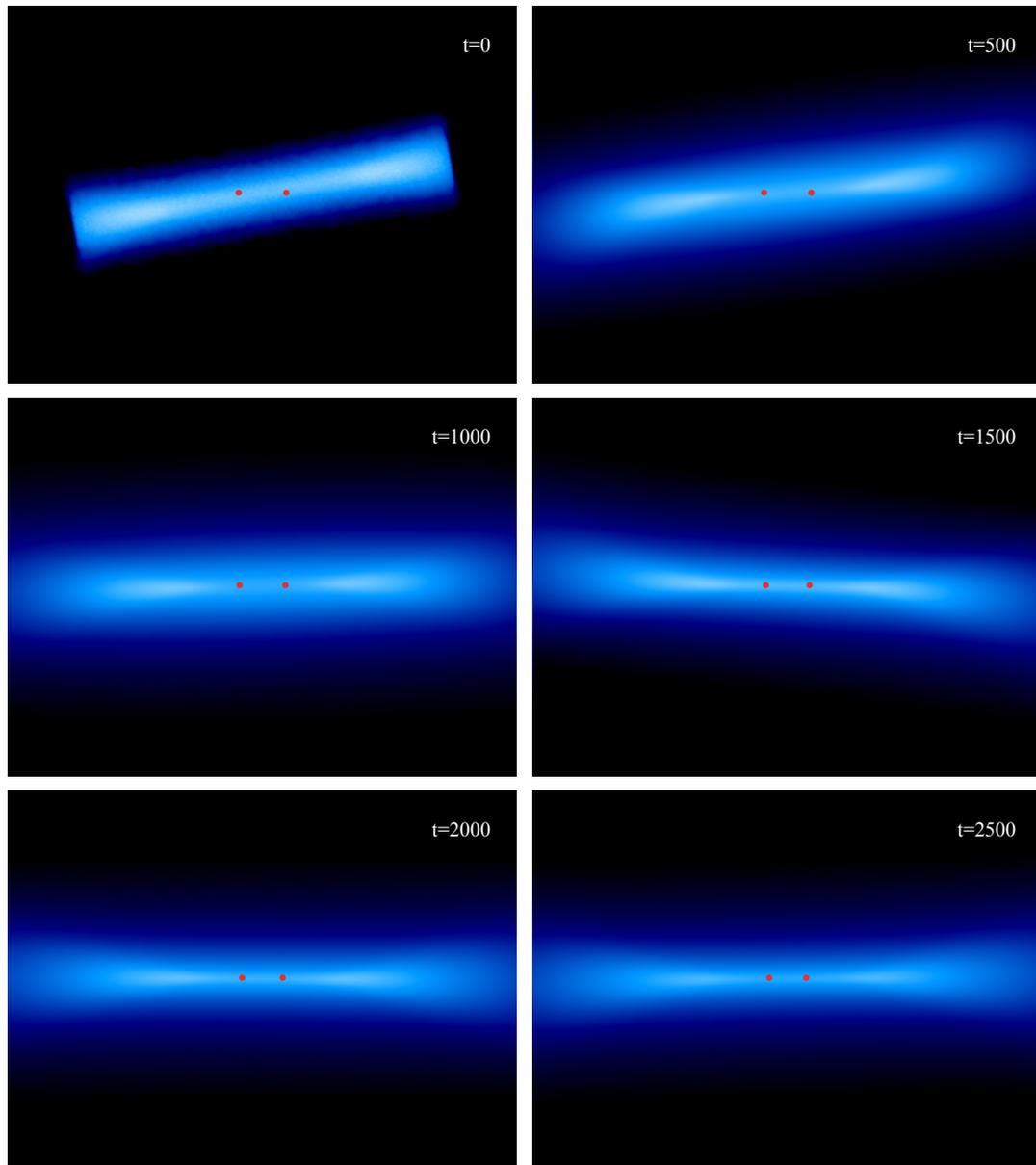


Figure 3.4: Edge-on column density rendering of the disc at various times in the simulation. Each component of the binary is represented by a red filled circle. Over time the disc spreads and precesses around the binary. The precession induces dissipation in the disc which aligns the disc with the binary plane.

I have also performed simulations (not illustrated) of a moderately eccentric binary ($e = 0.3$) and an unequal mass ratio binary ($M_2/M_1 = 0.5$) which both display the same behaviour.

3.5 Conclusions

I have shown that for realistic parameters a circumbinary disc must usually wholly co- or counter-align with a binary. However for extreme mass ratios or high eccentricities the binary may be dominated by the disc angular momentum (cf. Eq. 3.6). In this case evolution similar to Fig. 9 of Lodato and Pringle (2006) is expected with simultaneous co- and counter-alignment of the disc.

I have shown that a disc with an initial inclination of 170° to the binary plane stably counteraligns. Circumbinary discs, with $J_d \ll J_b$ and an initial misalignment angle of $> 90^\circ$ counteralign with respect to the binary. If the disc angular momentum is not negligible, King et al. (2005) showed that the condition for counteralignment is

$$\cos \theta < -\frac{J_d}{2J_b}. \quad (3.8)$$

A counteraligned circumbinary disc is efficient at shrinking the binary as it directly absorbs negative angular momentum when capturing gas into circumprimary or circumsecondary discs (Nixon et al., 2011a, Chapter 4). This interaction increases the binary eccentricity. Nixon et al. (2011a, Chapter 4) show that the timescale to increase the eccentricity from zero to unity is $\sim M_2/\dot{M}$ where \dot{M} is the mass inflow rate through the retrograde circumbinary disc. Once the eccentricity is high enough gravitational wave losses will drive the binary to coalescence.

4

Retrograde Accretion

4.1 Abstract

We investigate whether a circumbinary gas disc can coalesce a supermassive black hole binary system in the centre of a galaxy. This is known to be problematic for a prograde disc. We show that in contrast, interaction with a *retrograde* circumbinary disc is considerably more effective in shrinking the binary because there are no orbital resonances. The binary directly absorbs negative angular momentum from the circumbinary disc by capturing gas into a disc around the secondary black hole, or discs around both holes if the binary mass ratio is close to unity. In many cases the binary orbit becomes eccentric, shortening the pericentre distance as the eccentricity grows. In all cases the binary coalesces once it has absorbed the angular momentum of a gas mass comparable to that of the secondary black hole. Importantly, this conclusion is unaffected even if the gas inflow rate through the disc is formally super-Eddington for either hole. The coalescence timescale is therefore always $\sim M_2/\dot{M}$, where M_2 is the secondary black hole mass and \dot{M} the inflow rate through the circumbinary disc.

4.2 Introduction

Astronomers now generally agree that the centre of every reasonably large galaxy contains a supermassive black hole (SMBH). Moreover the mass of this hole correlates (at least at low redshift) with properties of the host galaxy (Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Häring and Rix, 2004). In the hierarchical picture of structure growth, small galaxies merge to produce large ones, promoting accretion on to their central SMBHs, and apparently causing these holes to coalesce. The favoured mechanism for driving the holes closer is dynamical friction. However

it is unclear that this can bring them close enough for gravitational wave losses to complete the coalescence, since the frictional process itself scatters away the stars causing it, and refilling of the loss cone is apparently too slow. This is often called the ‘final parsec problem’, as dynamical friction typically stalls at such separations between the holes (Begelman et al., 1980).

A possible way of overcoming this problem is interaction with gas orbiting in a disc just outside the SMBH binary (Armitage and Natarajan, 2005; MacFadyen and Milosavljević, 2008; Lodato et al., 2009; Cuadra et al., 2009). There has also been much discussion of cases where an SMBH binary is embedded in a disc (Escala et al., 2005; Dotti et al., 2007, 2009). It is implicitly assumed that dissipative torques make the disc coplanar with the binary. Studies of the circumbinary disc problem have so far considered prograde discs, i.e. those rotating in the same sense as the binary. Then tidal interaction with the binary turns the disc into a decretion disc, which transports angular momentum outward, but with little inward mass transport (Lodato et al., 2009). If the disc mass is large enough to carry away the binary angular momentum, a decretion disc is vulnerable to the self-gravitational instability (Lodato et al., 2009). This can rob the disc of the gas it needs to drive further angular momentum loss, halting the binary shrinkage.

This makes it doubtful that a prograde disc can ever in practice shrink the binary separation from $a \sim 1$ pc to the point ($a \sim 10^{-2}$ pc) where gravitational wave losses can drive it to coalescence. However the separation of the SMBH binary is much smaller than the interacting galaxies themselves so it is highly unlikely that the central gas flows are always prograde. These flows also receive randomly-directed injections of energy and momentum from star formation and supernovae, suggesting that retrograde flows are as likely as prograde. This kind of chaotic accretion gives a plausible picture of the mass and spin evolution of a central accreting black hole

(King and Pringle, 2006; King et al., 2008). It also implies that if nothing else manages to drive an SMBH binary to coalescence, it is highly likely that at some point there will be a retrograde coplanar disc surrounding the binary. This situation would arise if for example an earlier episode of prograde gas accretion failed to coalesce the binary, and was followed by a retrograde accretion event. As we will see, retrograde discs behave quite differently from prograde ones, and may offer a solution to the final parsec problem.

4.3 Prograde vs. Retrograde

We start by contrasting the main qualitative features of the prograde and retrograde cases. These stem from the physics of the interaction between the binary and the disc, where dissipative torques try to share the angular momenta. If the disc is prograde, this interaction shrinks the binary but moves the inner edge of the disc outwards, reducing the torque shrinking the binary. If instead the disc is retrograde, the effect is to shrink the binary, but also to move the inner edge of the disc *inwards*.

A prograde disc becomes a decretion disc, transporting angular momentum outwards with little net mass transport (Lodato et al., 2009). A retrograde disc instead remains an accretion disc, transporting angular momentum outwards and mass inwards. As we shall see, the long-term evolution of the disc–binary system is radically different in the two cases.

In particular, the disc–binary torque is quite different. For a prograde disc the tidal interaction occurs mainly through resonances. These occur when

$$\Omega^2 = m^2(\Omega - \omega)^2, \quad (4.1)$$

where ω is the binary orbital frequency, $\Omega(R)$ is the disc angular velocity and $m = 1, 2, \dots$ is the wave mode number (for example see the analysis in Papaloizou and Pringle 1977). So there are resonances at radii where

$$\Omega(R) = \frac{m\omega}{m \pm 1}, \quad (4.2)$$

where Ω and ω have the same sign. Resonances outside the binary orbit (i.e. with $\Omega < \omega$) correspond to the positive sign in the denominator of Eq. 4.2 and so appear at radii where

$$\frac{\Omega(R)}{\omega} = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots \quad (4.3)$$

The dominant interaction then involves the 2:1 (more strictly 1:2) resonance.

By contrast, in a retrograde disc Ω and ω have opposite signs, and Eq. 4.1 requires $|\Omega| > |\omega|$, so there are no resonances in a circumbinary retrograde disc. The disc–binary interaction is direct, as the inner edge of the disc starts to impinge on the secondary black hole. A retrograde circumbinary disc remains an accretion disc whose material is gravitationally captured by the binary, directly reducing its angular momentum. This is inherently more promising for shrinking the binary towards coalescence than the prograde case, where the binary dams up the disc.

It is important to understand that ‘capture’ simply means that the gas orbits a particular hole, and so has added its (negative) angular momentum to the binary orbit. It does *not* imply that the relevant hole must actually accrete this gas (although it may). Once captured the gas is bound to the hole and thus may be treated as a single body. Some or all of this captured gas can be expelled, for example by radiation pressure. Provided that this process is isotropic in the frame of the hole this does not change its orbital angular momentum, and so has very little effect on the orbital dynamics and eventual coalescence (cf. Eq. 4.22 below).

4.4 Where does the mass go?

The effect of a retrograde circumbinary disc differs in detail depending on how the captured mass is distributed between the two black holes. Accordingly we look at the reaction of test particles to the binary. To make things simple we first consider a circular binary with a low mass ratio, i.e. $M_2/M_1 \lesssim 0.1$. Then to first order we can treat the primary as fixed and the secondary as following a circular orbit of radius a around it with velocity $V = (GM_1/a)^{1/2}$. The circumbinary disc gas is on circular orbits with velocity $\simeq (GM_1/R)^{1/2}$ at each radius R . We assume the disc gas spirals slowly inwards from a large radius by the usual viscous evolution. As it is counter-rotating we can ignore all resonant effects and assume everything is ballistic until orbits begin to cross and fluid effects appear. We define the effective radii R_1, R_2 of the two holes as the radii within which gas particles are captured by each hole, e.g. by forming a disc around one or other of them, so that this captured gas has the same net specific angular momentum as the relevant hole. Evidently R_1, R_2 cannot in practice be larger than the individual Roche lobes for each hole.

For the disc particles orbiting closest to M_2 the interaction with M_2 is initially hyperbolic and so we can use the impulse approximation. Here the relative velocity is approximately $2V$. If the disc edge is at a radius $R = a + b$, where $b \ll a$, the impulse approximation shows that the disc particle acquires an inward radial velocity

$$U_R = \left(\frac{GM_2}{b^2} \right) \times \frac{2b}{2V} = \frac{GM_2}{bV}. \quad (4.4)$$

The inner edge of the circumbinary disc is not significantly perturbed if its distance b from the secondary's orbit is large enough that $U_R \lesssim V$, i.e. $V^2 \gtrsim GM_2/b$, or equivalently $b \gtrsim (M_2/M_1)a$. From this we conclude that the secondary cleanly pulls

the gas from the unperturbed inner edge of the disc at b provided that $R_2 \gtrsim b$, i.e.

$$\frac{R_2}{a} \gtrsim \frac{M_2}{M_1}. \quad (4.5)$$

Comparing this with the Roche lobe constraint

$$\frac{R_2}{a} \lesssim 0.4f \left(\frac{M_2}{M_1} \right)^{1/3}, \quad (4.6)$$

where $f \lesssim 1$ is a dimensionless factor, we see that the secondary captures almost all the gas for mass ratios

$$q = \frac{M_2}{M_1} \lesssim q_{\text{crit}} = 0.25f^{3/2} \simeq 0.25. \quad (4.7)$$

Equivalently we can define a Safronov number (Safronov, 1972) as used in discussing accretion of planetesimals from a protoplanetary disc:

$$\Theta = \frac{v_{\text{esc}}^2}{2v_{\text{orb}}^2} \simeq \frac{GM_2}{R_2} \frac{a}{GM_1} = \frac{M_2}{M_1} \frac{a}{R_2} \quad (4.8)$$

where $v_{\text{esc}}, v_{\text{orb}}$ are the escape velocity from the secondary's effective radius, and its orbital velocity respectively. This measures how much the gas is gravitationally perturbed before being captured. Small Θ implies little perturbation. We see that the criterion for secondary capture is just $\Theta \lesssim 1$, that is, the gas is cleanly captured without significant perturbation.

For larger mass ratios the flow becomes more complex and it is likely that some gas falls towards the binary centre of mass, producing some form of primary capture. In numerical simulations (Section 4.7) we will find that even in this case most of the gas is captured by the secondary. So in general the secondary captures most of the mass, and only in a major merger with $q > q_{\text{crit}}$ does the primary capture a

non-negligible amount from a retrograde circumbinary disc.

4.5 Eccentricity growth

A retrograde circumbinary disc can decrease both the energy and angular momentum of the SMBH binary, and so change its eccentricity. A simple argument shows how this happens. We consider a slightly eccentric binary orbit (the orbit is never exactly circular if it is shrinking) and again for simplicity assume that the mass ratio is sufficiently extreme that we can regard the primary black hole as effectively fixed, and only the secondary as interacting with the disc.

Momentum conservation shows that capture of disc gas always reduces the secondary's orbital velocity (see Eq. 4.18). This holds both for a direct collision, or (more commonly) if the secondary captures gas into a bound disc around itself. The secondary's mass cannot decrease in these interactions. Its specific orbital energy therefore always drops, so that the binary semimajor axis a decreases. But for mass capture into a symmetrical disc around the secondary near apocentre, the new orbit must retain the same apocentre as the old one (it must pass through this point, and the radial velocity remains zero there). Given the decrease of the semimajor axis, this means that capture near apocentre tends to increase the orbital eccentricity e (since $a[1 + e]$ remains constant). The eccentricity is evidently

$$e \sim \frac{\Delta M}{M_2} \quad (4.9)$$

where ΔM is the amount of mass the secondary has captured. Exactly the same reasoning shows that for mass capture near pericentre, the quantity $a(1 - e)$ has to stay constant despite a further decrease in a – in other words, the eccentricity must

decrease here by about the same amount ($\Delta M/M_2$) it increased at apocentre.

The secondary obviously captures at all points in between apo- and pericentre, but the effects are opposed for significant times. If the eccentricity is initially small these times are nearly equal, and e stays small as the orbit shrinks, provided that $\Delta M/M_2$ is below a certain value (4.12).

If on the other hand the orbit is initially quite eccentric, or the mass grows significantly in one orbit, the pericentre may be too small to allow disc interaction. Then e grows with $a(1+e) = a_0 \simeq \text{constant}$, and the pericentre distance goes as

$$p = a(1 - e) = 2a - a_0. \quad (4.10)$$

The binary therefore coalesces once the original semimajor axis has halved. We show below (Eq. 4.22) that this occurs once the secondary has absorbed the (negative) angular momentum of disc gas with mass comparable to its own. If gas flows inwards through the circumbinary disc at the rate \dot{M} the timescale for coalescence is

$$t_{\text{co}} \simeq \frac{M_2}{\dot{M}} \quad (4.11)$$

We stress again that this does *not* require either hole to accrete this gas but only to capture the gas into a bound orbit around the hole. In particular, \dot{M} can formally exceed the Eddington accretion rate for either hole without affecting the orbital shrinkage. If the accretion rates on to either hole were super-Eddington then the captured gas would be blown away by radiation pressure from the disc(s) around the secondary (and possibly primary) black hole, without significantly changing the binary orbital evolution (cf. Eq. 4.22).

The critical eccentricity separating cases where the binary remains almost circular

from those of growing eccentricity depends on the surface density distribution of the circumbinary disc. In a real three-dimensional disc with scale height H and aspect ratio H/R the surface density tails off over a length-scale $H(a) \sim (H/R)a$. This suggests that the critical eccentricity dividing these two cases is just

$$e_{\text{crit}} \sim \frac{H}{R} \quad (4.12)$$

Thus any binary starting with $e > e_{\text{crit}}$, or achieving it by capturing a large mass (comparable to the secondary's) in one orbit, must become very eccentric. We note that even a preceding episode of prograde accretion can leave the binary with an eccentricity exceeding e_{crit} (e.g. Cuadra et al. 2009), so that growth to high eccentricity is very likely if accretion is chaotic. Our conclusions about eccentricity growth agree with those of Dotti et al. (2009), who considered a related but different problem. A secondary black hole was injected with significant eccentricity into a pre-existing dense circumnuclear disc surrounding a primary black hole. The secondary's orbit was initially retrograde with respect to this interior disc. However the secondary was able to interact with enough gas in less than one orbit that it cancelled all of its angular momentum. The secondary then briefly had approximately zero angular momentum before capturing more gas and so changing its angular momentum to prograde. Here we restrict ourselves to accretion events on much smaller scales than in Dotti et al. (2009). This difference in lengthscales is important. At the smaller scales we consider, a disc with mass $M_d \gg M_2$ would probably be self-gravitating, and this change of angular momentum sign is unlikely to occur. Our simulations agree with this conclusion: once interaction with the disc has cancelled the orbital angular momentum (and thus much of the orbital energy) of the secondary, the binary coalesces.

4.6 Orbital evolution of the binary

We can now make analytic estimates of the orbital evolution as the binary interacts with an exterior retrograde disc. For simplicity we again assume that $q = M_2/M_1 \ll 1$, so that the secondary has specific orbital energy and angular momentum E, J^2 where

$$E = -\frac{GM_1}{2a} = \frac{1}{2}v(r)^2 - \frac{GM_1}{r} \quad (4.13)$$

and

$$J^2 = GM_1 a(1 - e^2) \quad (4.14)$$

We have seen above that in many cases the binary eccentricity grows quite strongly. We take the limit where the secondary interacts with the disc only very near apocentre $r = a(1 + e)$. Here its velocity v_{ap} is purely azimuthal, with

$$[a(1 + e)v_{\text{ap}}]^2 = J^2 \quad (4.15)$$

which by Eq. 4.14 gives

$$v_{\text{ap}}^2 = \frac{GM_1}{a} \frac{1 - e}{1 + e} \quad (4.16)$$

Near apocentre the secondary interacts with disc material moving with azimuthal velocity $v_{\text{disc}} < 0$, with

$$v_{\text{disc}}^2 = \frac{GM_1}{r} = \frac{GM_1}{a(1 + e)} \quad (4.17)$$

We assume that a mass ΔM of disc matter is captured into orbit about the secondary near apocentre, as we discussed above, so that all of its orbital angular momentum is transferred to the secondary. To allow for mass loss from the subsequent accretion process on to this hole (if for example this is super-Eddington, or mass interacts gravitationally with the secondary but is all accreted by the primary) we assume

that the effective mass of the hole plus disc becomes $M_2 + \alpha\Delta M$, with $0 \leq \alpha \leq 1$.

Then conservation of linear momentum gives

$$M_2 v_{\text{ap}} - \Delta M v_{\text{disc}} = (M_2 + \alpha\Delta M)u \quad (4.18)$$

where u is the new apocentre velocity of the secondary plus its captured gas disc.

The changes $\Delta E, \Delta a$ in orbital specific energy and semi-major axis are given by

$$\frac{GM_1}{2a^2} \Delta a = \Delta E = \frac{1}{2}u^2 - \frac{1}{2}v_{\text{ap}}^2 \quad (4.19)$$

Combining Eqs. 4.16,4.17,4.18 and 4.19 gives

$$\frac{\Delta a}{a^2} = \frac{-2}{a(1+e)} \frac{\Delta M}{M_2} [(1-e)^{1/2} + \alpha(1-e)] \quad (4.20)$$

to lowest order in ΔM . We noted above that if the interaction is confined to the immediate vicinity of apocentre then the apocentre distance $a_0 = a(1+e)$ stays constant in the subsequent evolution. Then $1+e = a_0/a$ and $1-e = 2 - a_0/a$, so that the left hand side of Eq. 4.20 is simply $\Delta(1-e)/a_0$ and therefore Eq. 4.20 becomes

$$\Delta(1-e) = -\frac{2\Delta M}{M_2} [(1-e)^{1/2} + \alpha(1-e)]. \quad (4.21)$$

Using $M_2 = M_{20} + \alpha M$, where M_{20} was the mass of the secondary hole when e was zero, and M is the total mass since transferred from the disc, this integrates to give

$$1-e = \left(\frac{M_{20} - M}{M_{20} + \alpha M} \right)^2, \quad (4.22)$$

Hence in this approximation the binary coalesces (i.e. $1-e=0$) once the disc

has transferred a mass equal to the secondary black hole, i.e. after a time t_{co} (cf. Eq. 4.11), *independently of the fraction α* . This means that mass loss has no effect in slowing the inspiral. We note that Eq. 4.20 implies that the energy dissipated in the shrinkage of an eccentric binary is

$$-M_2 \Delta E \simeq \frac{GM_1 \Delta M}{2a} \quad (4.23)$$

per binary orbit, which is less than that produced by viscous dissipation in the disc (this has to pull in a mass larger than ΔM on each orbit).

All other cases give similar timescales $\sim t_{\text{co}}$ for coalescence. For example if the orbit stays circular we can set $e = 0$ in Eq. 4.20 and find

$$\frac{\Delta a}{a} = -2(1 + \alpha) \frac{\Delta M}{M_2}, \quad (4.24)$$

so that $a \propto M_2^{-2(1+\alpha)}$. If the secondary gains the transferred mass we have $\alpha = 1, a \propto M_2^{-4}$, while if the transferred mass (but not the angular momentum) ends up on the primary we have $\alpha = 0, a \propto M_2^{-2}$. Shrinking the binary from $a = 1$ pc to $a = 10^{-2}$ pc, where gravitational radiation rapidly coalesces it, requires the transfer of between 2 and 9 times the original mass of the secondary in the two cases. This is larger than in the eccentric case because the torque on the binary now decreases with a (indeed the required mass would be formally infinite without gravitational radiation).

4.7 Simulations

We now use Smoothed Particle Hydrodynamics (SPH) simulations to compare with the analytical arguments above.

4.7.1 Code setup

We use a fully 3D conservative Lagrangian implementation of the SPH algorithms, for example Springel and Hernquist (2002); Rosswog (2009); Price (2012). We neglect gas self-gravity and so are able to use linked-lists of particles rather than the usual tree for neighbour finding (Deegan, 2009). We assume an isothermal disc, i.e. $P = c_s^2 \rho$, where the temperature and hence sound speed c_s is constant for all particles at all times. We make this choice for simplicity, and anticipate that the interaction of gas with the binary is not greatly affected by it. In particular we noted in Section 4.6 above (cf. Eq. 4.23) that binary shrinkage does not greatly increase the heating of the disc. We integrate a ‘live’ binary which feels the back reaction of the gas and generates self-consistent orbits for the binary and the gas.

Our aim here is to understand the dynamical interaction of the disc with the binary, rather than long-term behaviour such as the timescales for coalescence, which we know depends on the total mass transferred through the disc (see Section 4.6 above). Accordingly we use only the standard SPH artificial viscosity (Rosswog, 2009) and do not include a physical viscosity. Our simulations run only for a few tens of binary orbits, so that the viscosity scheme employed should not affect our results.

Our code units are dimensionless and take the initial separation and mass of the binary as unity. The period of the binary in its initial configuration is thus 2π . The general disc setup is gas, of mass M_d , spread from R_{in} to R_{out} with a surface density following a power law in radius, i.e. $\Sigma = \Sigma_0 (R/R_0)^{-p}$ with p typically $= 1$. The initial vertical structure in the disc is a Gaussian. We set the sound speed in the gas to a value ensuring that our neglect of self-gravity is justified. The relevant quantity here is the Toomre parameter $Q = c_s \Omega / \pi G \Sigma$ (Toomre, 1964), which describes the stability of a disc to gravitational collapse. For $Q < 1$ the disc

self-gravity overwhelms the disc pressure resulting in collapse. Thus we arrange that Q exceeds a minimum value (> 1) throughout the disc. Typically we arrange $Q > 5$ for a disc of mass $M_d = 10^{-2}$. The particles are all on initially circular orbits. The simulations detailed below were repeated with half a million, one million and two million particles and were deemed to have converged. All of the simulations detailed below initially contain one million particles.

4.7.2 Where does the mass go?

Two main parameters govern the gas flow for a retrograde disc–binary system: the binary mass ratio and the ‘capture’ radius of each binary component. As we discussed in Section 4.4, with mass ratios $q \simeq 1$, both the primary and secondary can interact with the inner edge of the disc so we should expect gas capture by both objects. However if $q \ll 1$, an inward–moving disc reaches the secondary before the primary. Then the secondary’s effective radius may be large enough for it to capture all of the gas. If instead this radius is small the gas may be perturbed towards the primary rather than captured. In Section 4.4 we estimated the critical value dividing these cases (Eq. 4.5). We study the effects of changing parameters here.

It is of course currently impossible to follow gas accretion down to the innermost stable circular orbits of the two black holes while simultaneously following the gas flow from the circumbinary disc. The analytic work above suggests that the secondary hole is always surrounded by a gas disc denser than the circumbinary disc and any gas flow from it, which therefore captures any gas impacting it. If initially there had been no gas around the secondary, gas entering $R_{\text{capture}} = aM_2/2M_1$ would be captured (but not necessarily accreted). This gas would spread to form a disc around the secondary, on a timescale short with respect to the mass transfer

timescale in the circumbinary disc. The secondary's disc then forms an obstacle for further inflow from the circumbinary disc. As further mass is captured in this way, this disc can become at most as large as some fraction of the Roche lobe (cf. Eq. 4.6). In the language of planetary dynamics this condition is formally equivalent to taking the secondary's Safronov number, defined as $(\text{planetary escape velocity}/\text{orbital velocity})^2/2$, (e.g. Safronov 1972; Hansen and Barman 2007) to be of order $2.5q^{2/3}$. From this analysis we expect the capture radius of the holes to be of order or smaller than the Roche lobe. For completeness we also consider the possibility of still smaller capture radii. Once a particle moves within the capture radius it is assumed to have impacted upon the disc assumed to be present inside the capture radius. This gas could be accreted by the black hole or expelled by radiation pressure from the innermost parts of the disc. In either case it is clear that once it has been 'captured' it plays no significant further role in the binary dynamics. Accordingly we add its mass and momentum to the relevant hole and remove it from the simulation.

4.7.2.1 Secondary capture

The first simulation has a mass ratio of $q = 0.1$ and effective radii 0.15, 0.2 for the secondary and primary. Eq. 4.5 implies that the secondary should capture the inner edge of the circumbinary disc without significantly perturbing the remaining disc particles. From Eq. 4.8 the secondary's Safronov number is $2/3$.

We start with the retrograde disc extending from 1 to 2 in radius, and a circular binary.

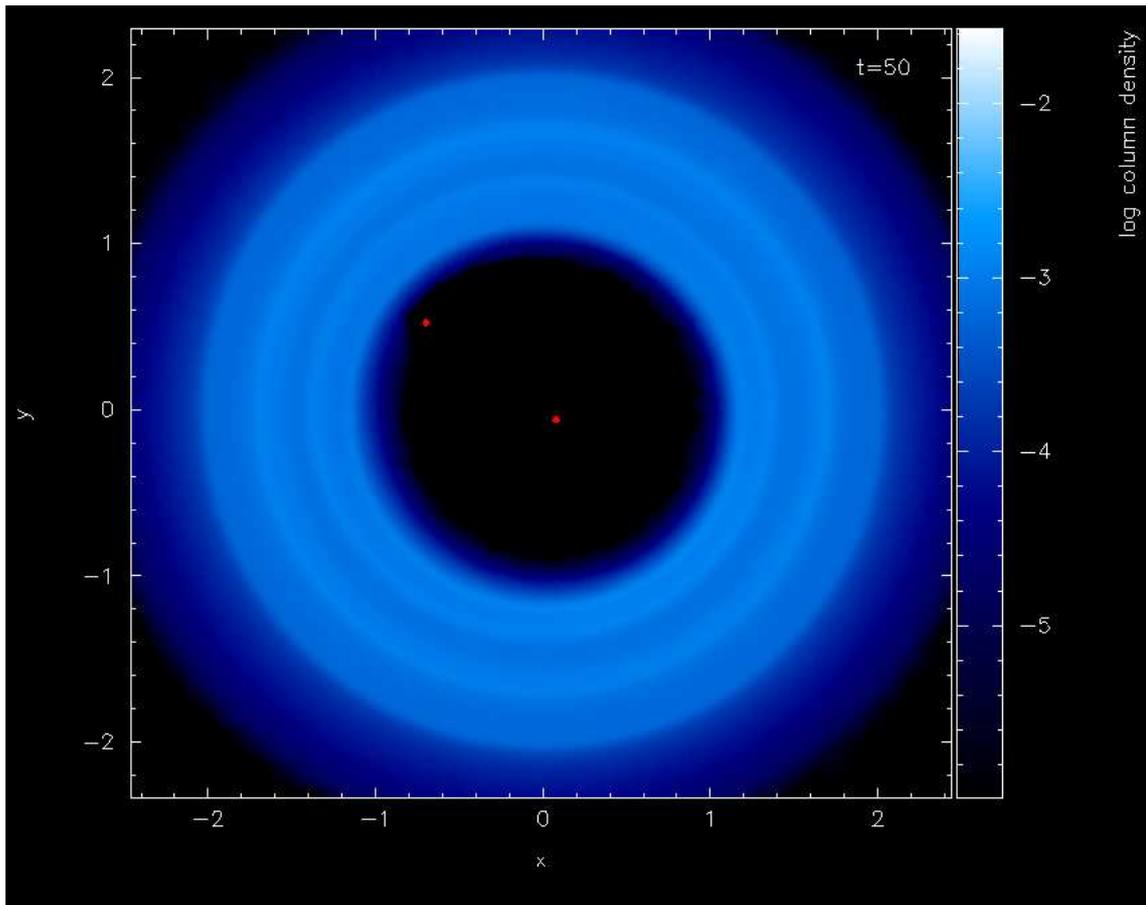


Figure 4.1: Image of the simulation from Section 4.7.2.1 at time $t = 50$. The binary is represented by the two dots. The axes are in code units with the log of the column density given by the bar.

At first the disc inner edge is slightly perturbed, but not enough to provide any noticeable accretion onto the primary. After one binary orbit the system settles and the secondary smoothly captures the inner edge of the disc. Fig. 4.1 shows the state of the simulation at $t = 50$. At this time, of the particles accreted, the secondary has captured $\sim 100\%$ (149759 particles) and the primary has captured $\sim 0\%$ (3 particles).

4.7.2.2 Primary capture

The second simulation also has a mass ratio of $q = 0.1$, but this time we use an effective radius of 0.05 for the secondary with again 0.2 for the primary capture radius, so that its Safronov number is $\Theta = 2$. Eq. (4.5) implies that the secondary should perturb the disc particles significantly without directly accreting all of them.

The initial disc is exactly the same as for the simulation in Section 4.7.2.1. The binary is again initially circular.

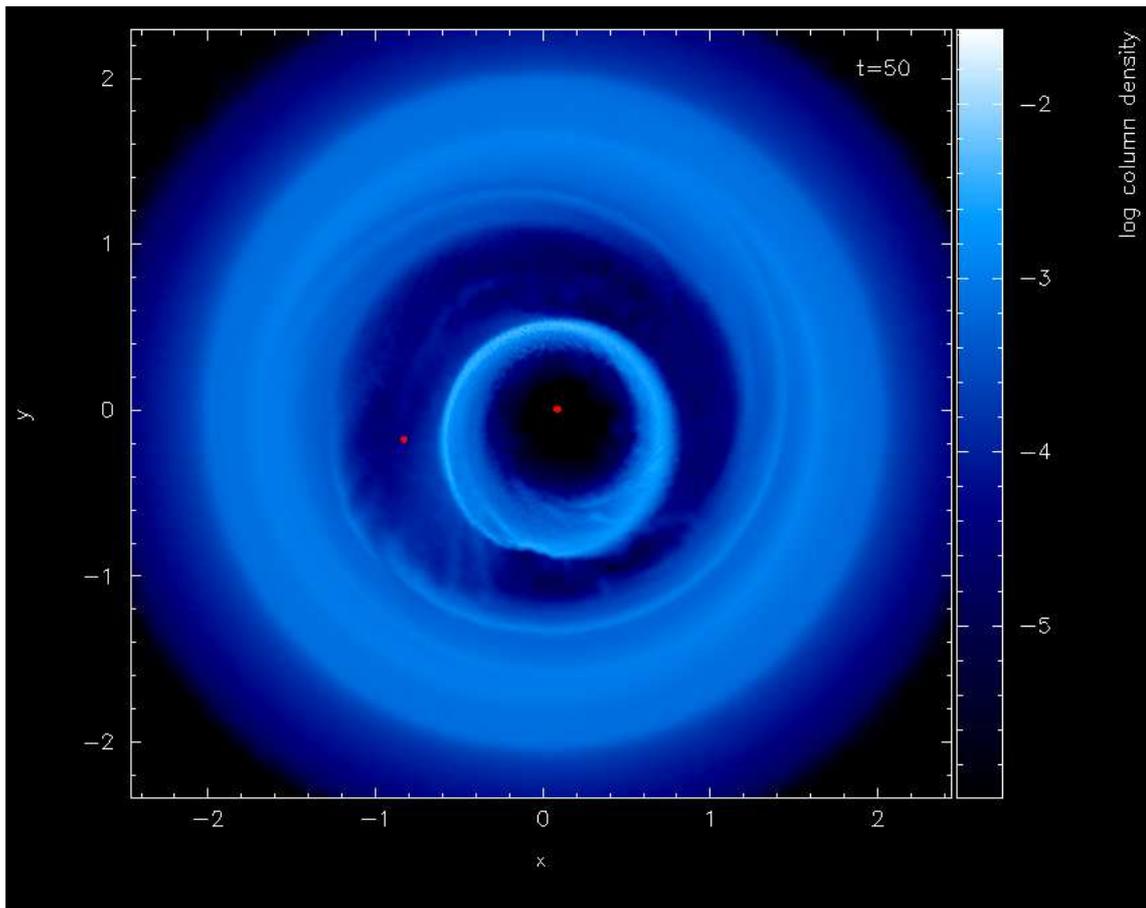


Figure 4.2: Image of the simulation from Section 4.7.2.2 at time $t = 50$.

At first, the disc inner edge is significantly perturbed, with some particles on orbits passing close to the primary. In the first 3 – 4 orbits the gas flow is very chaotic. After this the system settles into a quasi-steady state. Both holes capture from the retrograde circumprimary disc (as shown in Fig. 4.2). The secondary both disturbs and captures particles from the outer edge of the circumprimary disc, and perturbs more particles into it. The circumprimary disc is warped, eccentric and precessing. This is probably a result of particle noise destroying the symmetry about the orbital plane, as might indeed happen in a realistic situation.

At time $t = 50$ the secondary has captured $\sim 95\%$ (147642) particles and the primary has captured $\sim 5\%$ (8193 particles). So even in this case the secondary still takes most of gas captured from the circumbinary disc. The binary probably coalesces once the disc has transferred a mass $\sim M_2$, so if the disc is more massive than this, the coalesced hole eventually accretes the remainder. Clearly if the secondary’s mass or its capture radius is made arbitrarily small, it would accrete very little, and most of the mass would be captured by the primary. However it is clear that this requires extreme choices of these parameters.

4.7.2.3 Dual Capture

Here we look at the interaction of an near equal-mass binary with the disc. We use $q = 0.5$ and the same disc as in Sections 4.7.2.1 and 4.7.2.2. Again the binary is initially circular. Here we use capture radii of 0.01 for both holes to allow formation of circumprimary and circumsecondary discs.

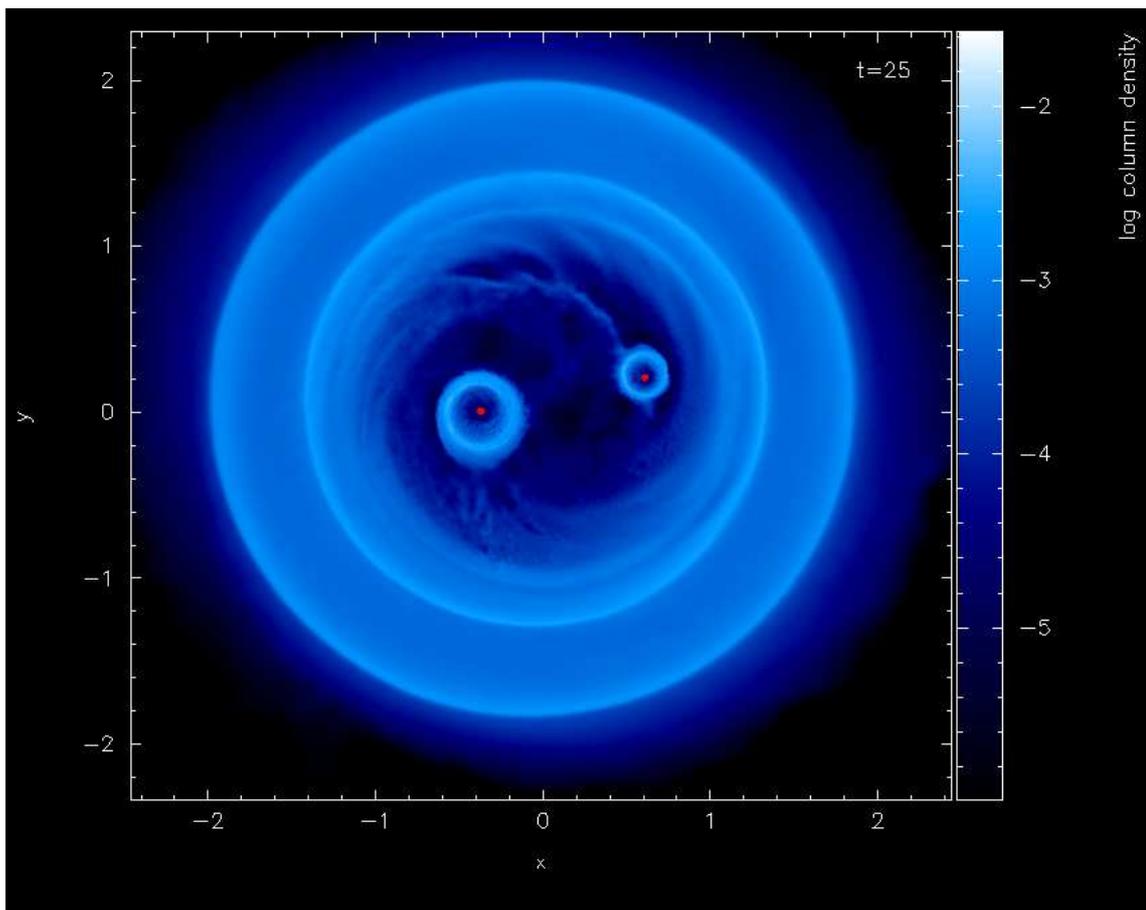


Figure 4.3: Face on image of the simulation from Section 4.7.2.3 at time $t = 25$.

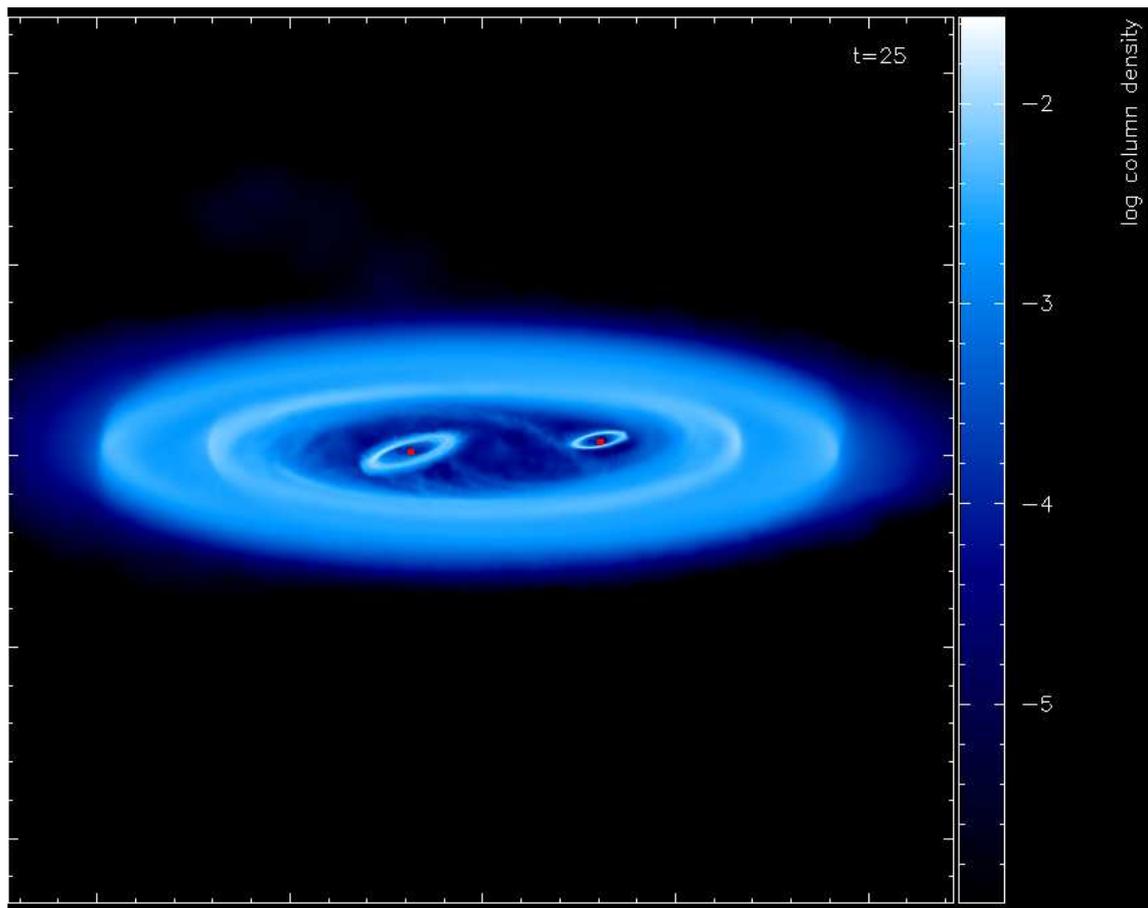


Figure 4.4: Image of the simulation from Section 4.7.2.3 at time $t = 25$ viewed at an angle of 15° to the plane of the disc.

During the initial 3 – 4 binary orbits the flow is very chaotic, with mass captured by both holes. As more gas is captured (but this time not ‘accreted’) by both holes, circumprimary and circumsecondary discs are formed which persist throughout the simulation. The discs are supplied by streams from the circumbinary disc. We show the simulation at time $t = 25$ in Figs. 4.3 & 4.4. At this time the primary has captured 48% (57969) and the secondary 52% (62189) of the accreted particles. Together with the previous simulation, this shows that the primary can gain significant mass only if the mass ratio is close to unity. We note again that the discs are not planar, and show a significant tilt with respect to the binary plane. Again this is probably a result of particle noise removing the symmetry about the orbital axis. Globally angular momentum is conserved but locally the streams that supply these discs need not be planar.

4.7.3 Eccentricity growth

The analytic arguments of Section 4.5 suggest that capture from a retrograde circumbinary disc at apocentre and pericentre of the binary orbit increase and decrease its eccentricity respectively. Here we show three SPH experiments exploring this.

The first simulation has an initially circular binary, with $q = 10^{-3}$, where the secondary is embedded just inside the inner edge of a retrograde disc. In code units the disc is spread radially from 0.8 to 1.5. The disc has an initial mass $M_d = 10^{-2}$.

The second simulation also starts with the secondary embedded in the disc. However this time the inner edge of the disc extends much further in, to 0.1.

The third simulation starts with a retrograde disc interior to the binary, i.e. a circumprimary disc extending from 0.1 to 1.0. In this case we start with initial eccentricity 0.5 and the binary at apocentre – so the secondary begins outside the

circumprimary disc but plunges into it before it reaches pericentre.

All of these simulations have capture radii 0.1, 0.01 for the primary and secondary respectively.

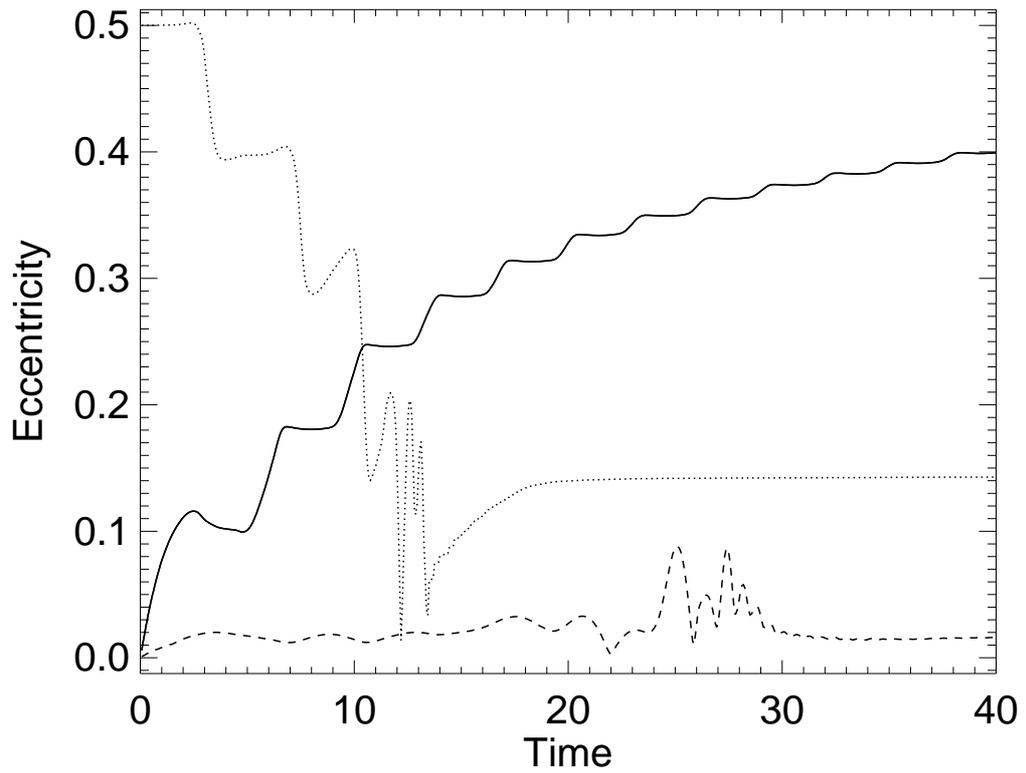


Figure 4.5: Eccentricity growth and decay of the simulations in Section 4.7.3. Time is in code units. The solid line is the first simulation (capture at apocentre but not at pericentre), the dashed line is the second simulation (capture at both pericentre and apocentre) and the dotted line is the third simulation (capture at pericentre and not apocentre)

We show in Fig. 4.5 how the eccentricity of the binary evolves in the three cases. In the first simulation the binary captures enough gas at apocentre on the first orbit to plunge inside the inner edge of the disc. This means that the eccentricity growth at apocentre cannot be moderated by decay at pericentre. The eccentricity therefore grows as shown in Fig. 4.5. In the second simulation the secondary is still capturing when it reaches pericentre and so the eccentricity decays. It is notable that the eccentricity never returns precisely back to zero in this case. This happens because the secondary captures unequal amounts of gas at apo- and pericentre. This is reasonable, as its velocity is higher at pericentre and so it has less time to interact with the gas. In addition the mass and angular momentum at each radius is not constant.

In the third simulation the secondary can initially only capture from the circumprimary disc at pericentre and so the eccentricity decays. After approximately 3 – 4 orbits the secondary has damped out most of its initial eccentricity to an almost circular orbit. After a short interaction with the inner parts of the disc it reaches its inner edge. Here there is a brief interval of eccentricity growth as it captures from this inner edge of a (now) circumbinary disc.

All of the results of these simulations agree with the analytic arguments of Section 4.5.

4.8 Discussion

We have seen that accretion from a retrograde circumbinary disc can be considerably more effective in shrinking a supermassive black hole binary system than accretion from a corresponding prograde disc. Coalescence occurs if the secondary black hole captures angular momentum from a gas mass comparable to its own.

The reason for this effectiveness is the absence of orbital resonances in a retrograde disc. This cannot react to tidal torques from the binary and become a decretion disc, which is what tends to slow the evolution in the prograde case (Lodato et al., 2009; Cuadra et al., 2009; Dotti et al., 2007; MacFadyen and Milosavljević, 2008). Instead, the disc directly feeds negative orbital angular momentum to the secondary black hole. An important aspect here is that there is no restriction on the rate at which this can occur. In particular gas can flow inwards at rates higher than Eddington for even the primary black hole without any significant effect on the amount of gas required to coalesce the binary.

Capture to the primary hole is negligible unless the mass ratio is very close to unity, and is never dominant (even if ‘capture’ actually leads to ‘accretion’, which is not required). This may be important in interpreting attempts to observe merger events. In many cases the binary may develop very high eccentricity before gravitational wave emission coalesces it. The latter rapidly damps the eccentricity during the final inspiral, but leaves a residual value which is likely to be significant in any merger event detectable by gravitational wave observatories such as LISA. If the secondary hole actually accretes the captured mass it will acquire a large spin (Kerr parameter $a \simeq 1$) antiparallel to its binary orbit, which would be potentially detectable in the LISA waveform. Moreover if the mass ratio is very close to unity it is conceivable that the primary hole might do the same. This could in principle favour high black hole spin as a result of major mergers (e.g. in giant ellipticals), as sometimes proposed. However we should recall that the holes may well *not* gain much mass, particularly if this is captured at super-Eddington rates, and the primary definitely does not gain mass (and thus spin up) unless the mass ratio is very close to unity. For these reasons it seems unlikely that coalescences induced by retrograde gas flows produce rapidly spinning merged holes except in rare cases.

We caution finally self-gravity can deplete the circumbinary disc of gas and reduce its ability to shrink the binary (cf. Lodato et al. 2009 in the prograde case). Since coalescence by a retrograde disc requires $M_d \gtrsim M_2$, and self-gravity effects appear unless $M_d \lesssim (H/R)(M_1+M_2)$, where H/R is the disc aspect ratio, it appears that the final parsec problem is so far alleviated only for mergers with $M_2/M_1 \lesssim H/R$. This work does suggest how coalescence might work for larger mass ratios. The secondary hole has to absorb negative angular momentum from a gas mass comparable to its own, preferably in an eccentric orbit. Although gas self-gravity is inevitably important in such an event, a retrograde flow has the advantage that there is no limit on the mass inflow rate. A sequence of randomly oriented accretion events, aligning both prograde and retrograde, may provide a solution to the final parsec problem.

5

Broken Discs

5.1 Abstract

We simulate the viscous evolution of an accretion disc around a spinning black hole. In general any such disc is misaligned, and warped by the Lense–Thirring effect. Unlike previous studies we use effective viscosities constrained to be consistent with the internal fluid dynamics of the disc. We find that nonlinear fluid effects, which reduce the effective viscosities in warped regions, can promote the breaking of the disc into two distinct planes. This occurs when the Shakura & Sunyaev dimensionless viscosity parameter α is $\lesssim 0.3$ and the initial angle of misalignment between the disc and hole is $\gtrsim 45^\circ$. The break can be a long-lived feature, propagating outwards in the disc on the usual alignment timescale, after which the disc is fully co- or counter-aligned with the hole. Such a break in the disc may be significant in systems where we know the inclination of the outer accretion disc to the line of sight, such as some X-ray binaries: the inner disc, and so any jets, may be noticeably misaligned with respect to the orbital plane.

5.2 Introduction

In realistic astrophysical situations the angular momentum of a spinning black hole may be significantly misaligned with respect to a surrounding accretion disc. This is expected in black hole X-ray binaries where the hole may have received a significant kick during formation (Shklovskii 1970; Sutantyo 1978; Arzoumanian et al. 2002; Hobbs et al. 2005). It is also likely for accretion on to a supermassive black hole (SMBH) in the centre of a galaxy. The scale of the SMBH is so small compared with the galaxy that any process which drives gas down to the galactic centre is unlikely to have any special orientation with respect to the SMBH spin. The spin direction is set by the SMBH's accretion history and not by whatever causes the

next accretion event. Processes such as star formation, merging of satellite galaxies, gas cloud collisions and other galactic phenomena likely to drive gas inwards are chaotic, so it is reasonable to expect a random spread of angles between the angular momentum of infalling gas and that of the hole (cf. King and Pringle 2006, 2007; King et al. 2008). This gas is likely to form a ring around the SMBH which then viscously spreads into a misaligned accretion disc.

The Lense–Thirring (hereafter LT) effect of a spinning black hole causes non–equatorial orbits to precess at rates dependent on the distance from the hole (Lense and Thirring, 1918). In an accretion disc, this differential precession causes a warp which propagates through the disc. This is the Bardeen–Petterson effect (Bardeen and Petterson, 1975). Here we restrict our attention to viscous Keplerian accretion discs with negligible self–gravity. Papaloizou and Pringle (1983) linearised the equations of fluid dynamics to derive the first consistent equation governing the evolution of a warped disc of this type, assuming a tilt angle smaller than the disc angular semi–thickness. They introduced two effective kinematic viscosities controlling the angular momentum transport in the disc. In principle these two quantities depend on the nature of the warp: ν_1 governs the usual radial communication of the component of angular momentum perpendicular to the plane of the disc due to differential rotation (for a flat disc this is the Shakura & Sunyaev alpha viscosity), while ν_2 governs the radial communication of the component of angular momentum parallel to the local orbital plane; this acts to flatten tilted rings of gas. Papaloizou and Pringle (1983) also showed that conservation laws require a relation between ν_1 and ν_2 . For small warps this is

$$\nu_2 = \frac{\nu_1}{2\alpha^2}, \quad (5.1)$$

where α is the usual dimensionless viscosity parameter for discs (Shakura and Sun-

yaev, 1973) assumed $\ll 1$ by Papaloizou and Pringle (1983). This gives

$$\alpha_2 = \frac{1}{2\alpha}, \quad (5.2)$$

(again for small α) where α_2 is an equivalent dimensionless viscosity parameter for ν_2 .

Pringle (1992) derived an evolution equation valid for larger warps expressing global angular momentum conservation for a disc composed of rings of gas with no internal degrees of freedom. Pringle (1992) went on to develop a numerical technique for integrating this equation over rings of gas to study the viscous evolution of a time-dependent warped disc. Lodato and Pringle (2006) used this to study the Bardeen–Petterson effect. They assumed constant effective viscosities, related as expected for small amplitude warps (Eq. 5.1). This treatment suggests that a disc warp must always straighten itself out on a timescale shorter than the accretion timescale (assuming only the LT torque and the internal disc torques, i.e. no other external torques), because (5.1) implies $\nu_2 \gg \nu_1$ for typical values of α . The realignment process enhances the accretion rate by inducing extra dissipation (Lodato and Pringle, 2006).

The analysis of Pringle (1992) did not allow for any internal degrees of freedom within the gas rings, and was described as the ‘naive approach’ in Papaloizou and Pringle (1983). However Ogilvie (1999) derived equations of motion directly from the full three-dimensional fluid dynamical equations with an assumed isotropic viscosity. This confirmed that the equations of Pringle (1992) are formally valid only when two differences are taken into account (or can be safely neglected). The first is that the Pringle (1992) equations omit a torque causing rings to precess if they are tilted with respect to their neighbours. In a viscous Keplerian disc where $\alpha \ll 1$

but not close to zero, this torque is much smaller than the usual viscous torques included by Pringle (1992) and so can be neglected in time-dependent calculations. (however for completeness we include the torque in this work). Our main focus, however, is on the second extension. The quantities ν_1 and ν_2 are functions of the disc structure and so depend on the warp amplitude $|\psi| = R |\partial\ell/\partial R|$, where ℓ is the unit angular momentum vector and R is the spherical radius coordinate. Ogilvie (1999) used the equations of fluid dynamics to determine the relation between the effective viscosities for a general warp amplitude. In contrast to an assumption of constant effective viscosities, this analysis suggests that these quantities *drop* in a warped region, making it much harder for the disc to straighten any twisted rings of gas. Indeed, if the warp becomes large enough the disc may break into distinct planes with only a tenuous connection. We expect this to occur in large warps, as nonlinear corrections to the fluid flow are significant for warp amplitudes $|\psi| \gtrsim \alpha$ (Ogilvie, 1999).

There have been several theoretical investigations which show evidence of breaking discs. Larwood and Papaloizou (1997) report an example in Smoothed Particle Hydrodynamics (SPH) simulations of a circumbinary disc with a strong radial density decrease. This is suggestive, since the evolution of a misaligned circumbinary disc is qualitatively similar to that of a misaligned disc around a spinning black hole (Nixon et al., 2011b, Chapter 2). SPH simulations by Lodato and Price (2010) also found disc breaking. They assumed the presence of a large-amplitude warp and followed its viscous development. Their results agreed with those expected for the constrained viscosities derived in Ogilvie (1999), and showed that the disc could break if the warp amplitude was large enough (note that this paper did not attempt to show that the LT precession could *produce* such a warp).

Here we explore disc evolution under the LT effect with general warp amplitudes, and

the effective viscosities constrained as in Ogilvie (1999). We explore the evolution for a range of α values and different degrees of misalignment between the disc and hole. We aim to discover if the internal torques in an accretion disc can sustain a smooth warped configuration under the LT torque, or whether instead the disc breaks into distinct planes.

5.3 Numerical method

5.3.1 Evolution equation

The problem of disc breaking is complex, and to make progress we adopt the simplest approach. We use a 1D Eulerian ring code of the type described in Pringle (1992) and used in Lodato and Pringle (2006). Our code has been adapted to include the third effective viscosity and the nonlinear fluid dynamical constraints found by Ogilvie (1999). Here ‘1D’ means that the properties assigned to each ring depend only on its radius. Section 2.1 of Lodato and Pringle (2006) gives a detailed discussion of the validity of this approach. The main restriction is that the evolution of a disc warp must be diffusive rather than wave-like. Papaloizou and Pringle (1983) show that this holds if $\alpha > H/R$, which is usually satisfied for accretion discs around compact objects.

Our 1D ring code describes the evolution of warped accretion discs by evolving the

angular momentum density vector of each ring $\mathbf{L}(R, t)$ through the equation

$$\begin{aligned}
\frac{\partial \mathbf{L}}{\partial t} = & \frac{1}{R} \frac{\partial}{\partial R} \left\{ \frac{(\partial/\partial R) [\nu_1 \Sigma R^3 (-\Omega')]}{\Sigma (\partial/\partial R) (R^2 \Omega)} \mathbf{L} \right\} \\
& + \frac{1}{R} \frac{\partial}{\partial R} \left[\frac{1}{2} \nu_2 R |\mathbf{L}| \frac{\partial \ell}{\partial R} \right] \\
& + \frac{1}{R} \frac{\partial}{\partial R} \left\{ \left[\frac{\frac{1}{2} \nu_2 R^3 \Omega |\partial \ell / \partial R|^2}{(\partial/\partial R) (R^2 \Omega)} + \nu_1 \left(\frac{R \Omega'}{\Omega} \right) \right] \mathbf{L} \right\} \\
& + \frac{1}{R} \frac{\partial}{\partial R} \left(\nu_3 R |\mathbf{L}| \ell \times \frac{\partial \ell}{\partial R} \right) \\
& + \boldsymbol{\Omega}_p \times \mathbf{L}
\end{aligned} \tag{5.3}$$

where ν_1 , ν_2 and ν_3 are the effective viscosities, $\Omega(R)$ is the local azimuthal angular velocity, $\Sigma(R)$ is the disc surface density, $\ell(R)$ is the unit angular momentum vector, and $\boldsymbol{\Omega}_p(R)$ is the precession frequency induced in the disc by the LT effect (defined below). We note that $\mathbf{L} = \Sigma R^2 \Omega \ell$.

Equation (5.3) shows that there are five independent torques acting on the rings of gas. The first four terms on the right hand side represent the internal disc torques responsible for communicating angular momentum. The last term represents the LT torque. Here we briefly discuss these terms.

The first term on the right hand side of (5.3) describes the usual viscous diffusion of mass. This is governed by the azimuthal shear viscosity ν_1 . The second term is also diffusive. This term, governed by the vertical viscosity ν_2 , is responsible for diffusing the disc tilt. The third term is an advective torque. Depending on its sign this advects angular momentum (and hence mass) inwards or outwards through the disc.

The three terms discussed above are exactly those derived straightforwardly from the conservation of mass and angular momentum (Pringle, 1992). The fourth term

is a precessional torque found by the analysis in Ogilvie (1999). This torque causes rings to precess when they are tilted with respect to their neighbours. The direction of the precession is dependent on α and $|\psi|$ (cf. Fig. 5 in Ogilvie, 1999). This term has exactly the same form as the second term in (5.3) with the substitutions $0.5\nu_2 \rightarrow \nu_3$ and $\partial\ell/\partial R \rightarrow \boldsymbol{\ell} \times \partial\ell/\partial R$. This transformation shows that we can integrate this term using the technique described in Pringle (1992) for integrating the second term.

The fifth term is the LT torque arising from the gravitomagnetic interaction of the disc with the spinning hole. This induces precessions in the disc orbits, and hence causes the angular momentum vector for each ring to precess around the spin vector for the black hole with a frequency

$$\boldsymbol{\Omega}_p = \frac{2G\mathbf{J}_h}{c^2 R^3} \quad (5.4)$$

with c the speed of light and G the gravitational constant. The magnitude of \mathbf{J}_h is given by $J_h = acM(GM/c^2)$ where a is the dimensionless spin parameter and M is the mass of the hole (Kumar and Pringle, 1985). We include the back-reaction on the hole angular momentum vector \mathbf{J}_h by summing the effect of the precessions over all radii, giving the torque on the hole as:

$$\frac{d\mathbf{J}_h}{dt} = -2\pi \int \boldsymbol{\Omega}_p \times \mathbf{L} R dR. \quad (5.5)$$

Since $\boldsymbol{\Omega}_p \propto \mathbf{J}_h$ this equation makes it clear that the magnitude of \mathbf{J}_h is conserved and hence \mathbf{J}_h can only precess on a sphere, simply because the LT torque has no component in the direction of \mathbf{J}_h for any disc structure. Equivalently the LT torque induces only precessions in the disc, so that the reaction back on the hole is simply a sum of precessions, which is itself a precession.

We integrate both (5.3) and (5.5) using a simple forward Euler method (Pringle, 1992) on a logarithmically spaced grid. However we adopt a different implementation of the boundary conditions. Rather than implementing a sink over a few grid cells we simply remove all of the angular momentum that reaches the inner- and outer-most cells at the end of each integration step. This allows angular momentum (and hence mass) to flow freely across the boundary, giving the usual $\Sigma = 0$, torque-free, accreting boundary condition. This is the form adopted in Lodato and Price (2010). We note that this boundary condition does not conserve angular momentum exactly. However this small effect does not affect the inclination of any of the rings and hence does not alter our results. Once the warp reaches the outer boundary it flows freely off the computational domain and the disc is flattened (i.e. we do not impose a boundary condition at the outer edge of the disc to maintain the disc warp).

5.3.2 Generating the constrained effective viscosities

We use the effective viscosities derived by Ogilvie (1999) for a locally isothermal disc. These are constrained to be consistent with the internal fluid dynamics of the disc, assuming an isotropic physical viscosity. The physics of these constraints is simply local conservation of mass and angular momentum in the disc.

The effective viscosities which describe the transport of angular momentum in a warped disc take the usual form of a Shakura & Sunyaev disc viscosity. The constraints determine the nonlinear coefficients (Q_1, Q_2, Q_3) of the effective viscosities (cf. equations 5.10, 5.11 and 5.12). These coefficients all decrease in magnitude in the presence of a strong warp, where the disc rings are highly inclined to each other. Communication of angular momentum is reduced because the rings of gas are no longer in perfect contact. We expect angular momentum to be communicated through perturbations of the particles from circular orbits. The rate of communi-

cation, and hence the magnitude of the viscosities, therefore depend on how easily these particles can interact. This is harder for particles on inclined orbits, and once these reach a critical inclination this can only happen where the rings cross. We note that for inviscid discs, which we do not consider here, this interpretation is no longer valid, see Ogilvie (1999) for further details.

We now detail the calculation of the effective viscosities from the disc properties. The notation in Ogilvie (1999) differs from that in Pringle (1992). Our Eq. 5.3, and equation 122 from Ogilvie (1999) relate the effective viscosities (ν_1 , ν_2 and ν_3) to the nonlinear viscosity coefficients (Q_1 , Q_2 and Q_3) by

$$\nu_1 = \frac{Q_1 \mathcal{I} \Omega^2}{\Sigma R \frac{d\Omega}{dR}}, \quad (5.6)$$

$$\nu_2 = \frac{2Q_2 \mathcal{I} \Omega}{\Sigma} \quad (5.7)$$

and

$$\nu_3 = \frac{Q_3 \mathcal{I} \Omega}{\Sigma}, \quad (5.8)$$

where \mathcal{I} is given by

$$\mathcal{I} = \frac{1}{2\pi} \int_0^{2\pi} \tilde{\mathcal{I}} d\phi = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} \rho z^2 dz d\phi. \quad (5.9)$$

Assuming Σ constant in azimuth, and a locally isothermal disc with sound speed c_s , we can write this as $\mathcal{I} \propto \Sigma c_s^2 / \Omega^2$. For a point-mass potential with $\Omega = (GM/R^3)^{1/2}$ we get after some algebra

$$\nu_1 = -\frac{2}{3} Q_1 [(H/R)^2 R^2 \Omega], \quad (5.10)$$

$$\nu_2 = 2Q_2 [(H/R)^2 R^2 \Omega] \quad (5.11)$$

and

$$\nu_3 = Q_3 [(H/R)^2 R^2 \Omega], \quad (5.12)$$

where $H = c_s/\Omega$ is the disc thickness.

For simplicity we assume that

$$(H/R)^2 R^2 \Omega = \text{const.} \quad (5.13)$$

This removes the radial dependence of the viscosities, making them functions only of $|\psi|$ and α and so allowing a direct comparison with the constant viscosity simulations of Lodato and Pringle (2006). The assumption (5.13) requires that the sound speed $c_s \propto R^{-3/4}$. This compares with the usual steady-state disc, where $c_s \propto R^{-3/8}$ (e.g. Frank et al. 2002).

We note that for a disc without a warp, ν_1 reduces to exactly the usual α disc viscosity (Shakura and Sunyaev, 1973), since

$$\nu_1 = \alpha c_s H = \alpha (H/R)^2 R^2 \Omega \quad (5.14)$$

and

$$Q_1(|\psi| = 0) = -3\alpha/2. \quad (5.15)$$

The effective viscosities are now fixed by the nonlinear coefficients $Q_1(\alpha, |\psi|)$, $Q_2(\alpha, |\psi|)$ and $Q_3(\alpha, |\psi|)$. We calculate these coefficients in the same way as described in Ogilvie (1999) with $\Gamma = 1$ and $\alpha_b = 0$ using the code kindly provided by Gordon Ogilvie. Figs. 5.1, 5.2 & 5.3 show the coefficients plotted against the warp amplitude for $\alpha = 0.2$.

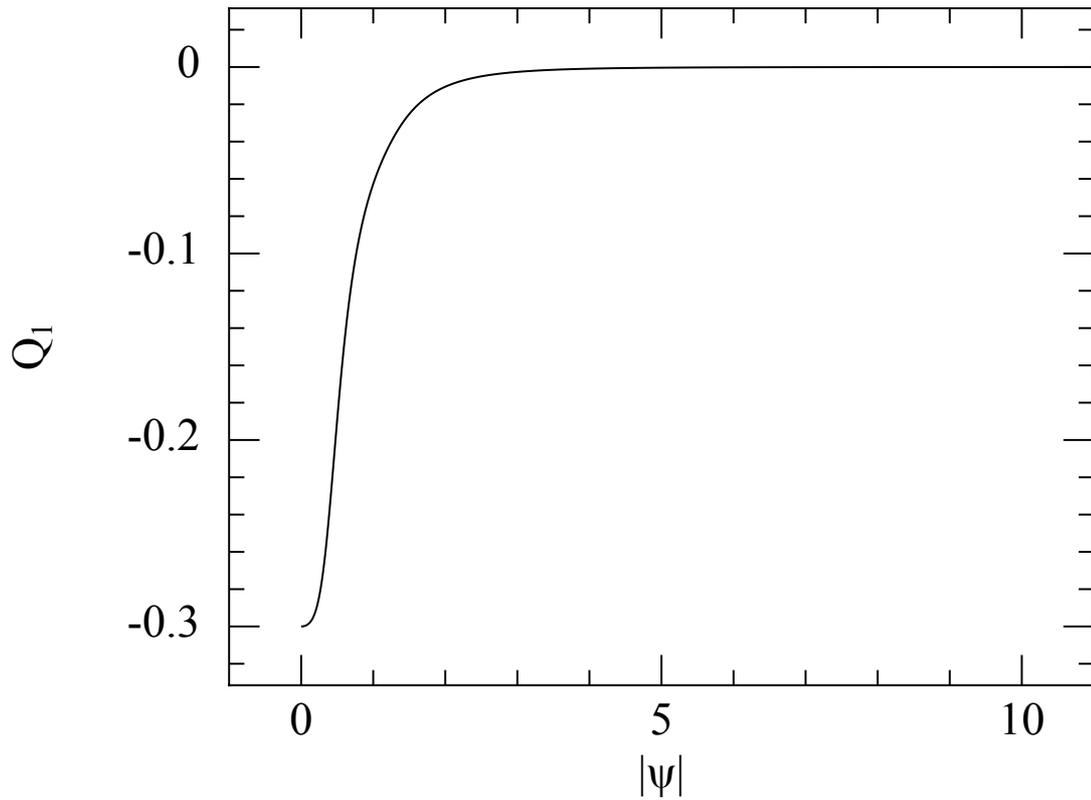


Figure 5.1: The effective viscosity coefficient Q_1 plotted against warp amplitude $|\psi|$, assuming that $\alpha = 0.2$, $\alpha_b = 0$ and $\Gamma = 1$. The coefficients change significantly for $0 < |\psi| \lesssim 5$, after which they are approximately constant at a small but non-zero value

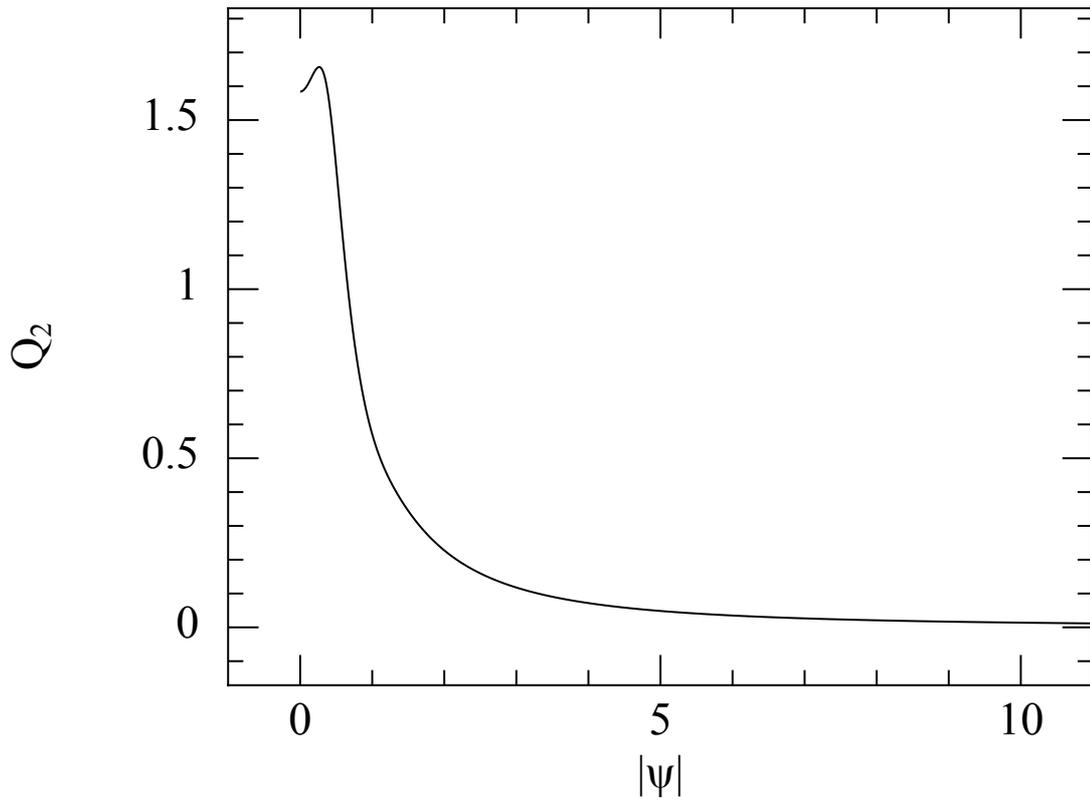


Figure 5.2: The effective viscosity coefficient Q_2 plotted against warp amplitude $|\psi|$, assuming that $\alpha = 0.2$, $\alpha_b = 0$ and $\Gamma = 1$. The coefficients change significantly for $0 < |\psi| \lesssim 5$, after which they are approximately constant at a small but non-zero value

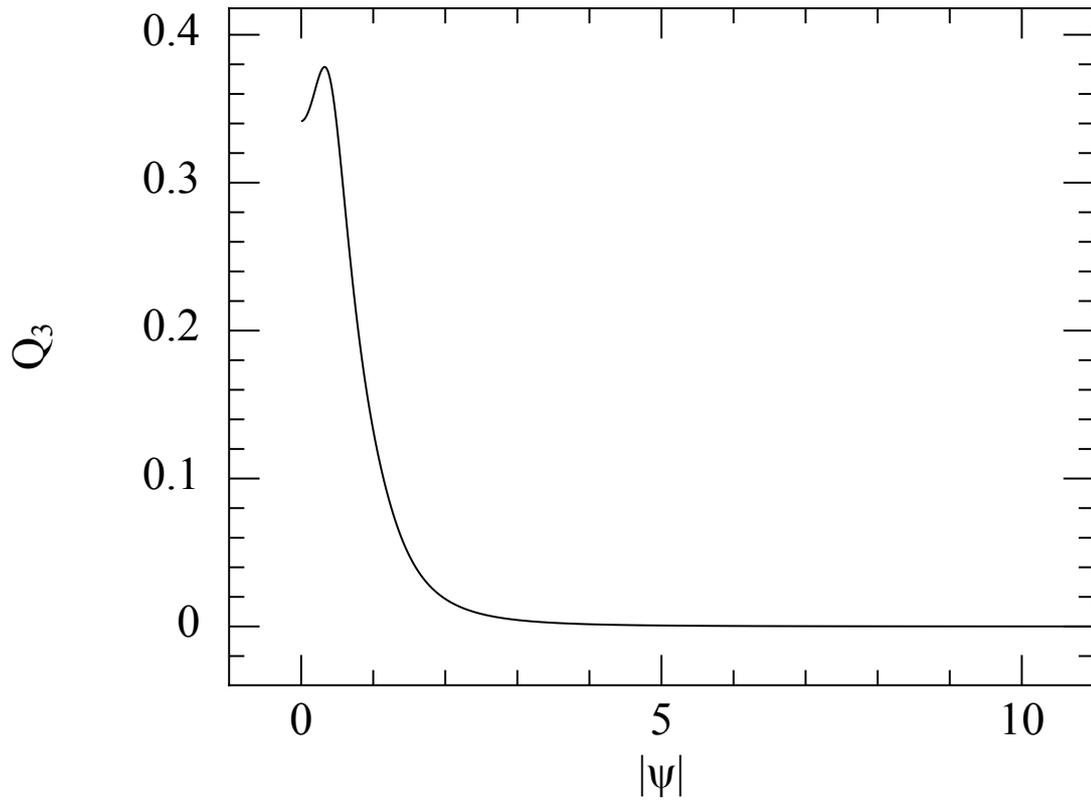


Figure 5.3: The effective viscosity coefficient Q_3 plotted against warp amplitude $|\psi|$, assuming that $\alpha = 0.2$, $\alpha_b = 0$ and $\Gamma = 1$. The coefficients change significantly for $0 < |\psi| \lesssim 5$, after which they are approximately constant at a small but non-zero value

5.4 Simulations

We use the method described above to simulate the warping of a misaligned accretion disc under the LT effect. We set up our initially planar but misaligned disc on a logarithmically spaced grid from $R_{\text{in}} = 10R_{\text{isco}}$ to $R_{\text{out}} = 1000R_{\text{isco}}$ using 100 grid cells (where $R_{\text{isco}} = 6GM/c^2$ is the radius of the innermost stable circular orbit around a Schwarzschild black hole). The disc surface density is initially a Gaussian ring peaked at $R_0 = 500R_{\text{isco}}$ with width $50R_{\text{isco}}$. We initialise each ring with a constant misalignment angle, θ_0 , such that the initial disc is flat but tilted to the plane of the black hole.

Lodato and Pringle (2006) suggest that the only governing parameters for simulations of this type are the ratio of disc to hole angular momenta ($J_{\text{d}}/J_{\text{h}}$), the ratio of the warp radius $R_{\text{w}} = \Omega_{\text{p}}R^3/\nu_2$ to R_0 , and the ratio of the effective viscosities ν_2/ν_1 (note that they neglect ν_3). However, as we use the constrained effective viscosities, ν_2/ν_1 is now determined once we have chosen α .

The parameters that we are free to choose in these simulations are α , H/R , R_0 , θ_0 and $J_{\text{d}}/J_{\text{h}}$. For black hole accretion discs $H/R \sim 10^{-3}$ and so we assume this throughout. Lodato and Pringle (2006) explored the effect of changing R_{w}/R_0 , so we fix R_0 as above (note that this does not fix the ratio R_{w}/R_0 . However this ratio varies little throughout our simulations.) To examine disc breaking we choose $J_{\text{d}}/J_{\text{h}} \ll 1$ as the simplest case and then vary α and θ_0 .

5.4.1 Does the disc break?

We perform several simulations for $0.2 \leq \alpha \leq 0.5$ to determine the various nonlinear responses of the disc to warping. We do not simulate $\alpha < 0.2$ as this is where the viscosity ν_1 is predicted to become negative in strong warps. We simulate a

range of θ_0 from 10° to 60° . This allows us to explore the two regimes: $|\psi| \lesssim \alpha$, where nonlinear effects should be negligible, and $|\psi| \gtrsim \alpha$, where nonlinear effects should be important. Our setup corresponds to $J_d/J_h \approx 0.02$, with R_w/R_0 varying between ~ 0.95 and ~ 1.20 as α changes. Note that we evaluate (5.13) at R_0 with $H/R = 10^{-3}$.

We perform these simulations for two cases. In the first we use constant effective viscosities. This acts as a control so we can isolate the effect of introducing nonlinear dynamics in the second set of simulations, where we use the full effective viscosities. The constant effective viscosities are simply calculated using the method described in Section 5.3.2 with $|\psi| = 0$.

For the simulations involving constant effective viscosities we expect to recover the Bardeen–Petterson effect as seen by several other authors (e.g. Bardeen and Petterson 1975; Pringle 1992; Scheuer and Feiler 1996; Lodato and Pringle 2006 etc). However the simulations involving the full nonlinear effective viscosities should modify this somewhat.

5.4.1.1 $\theta = 10^\circ$

In Fig. 5.4 (left column) we show the disc structures for simulations with $\theta_0 = 10^\circ$ and the constant effective viscosities at 0.01, 0.1 and 1 t_{visc} . Here we define the viscous timescale as $t_{\text{visc}} = R_0^2/\nu_1$ with ν_1 evaluated for $|\psi| = 0$ and α as defined in the particular case. Fig. 5.4 shows the behaviour for different values of the disc viscosity parameter α . In this case the behaviour is similar for all simulations, with lower α aligning in a shorter fraction of the viscous time (although in reality more slowly because the viscous timescale is longer).

In the right hand column of Fig. 5.4 we show the same simulation as in the left

hand column, however this time we use the full effective viscosities. We expect the evolution here to be similar, as the warp is never able to achieve a large amplitude, making the viscosities similar in both cases. This is indeed the case, as there is no visible difference between left and right hand columns of Fig. 5.4.

5.4.1.2 $\theta = 30^\circ$

In Fig. 5.5 we show the disc structures for simulations with $\theta_0 = 30^\circ$. Again there is very little difference between the disc structures obtained using the constant effective viscosities and those with the full effective viscosities. However the disc profile steepens in the early stages of the $\alpha = 0.2$ simulation. This suggests that for $\theta = 30^\circ$ we would have to go to $\alpha \ll 0.2$ to get disc breaking.

5.4.1.3 $\theta = 45^\circ$

In Fig. 5.6 we show the disc structures for simulations with $\theta_0 = 45^\circ$. There is now a noticeable difference between the constant effective viscosity and full effective viscosity simulations. Even for large α the disc profile is steepened and for $\alpha \sim 0.2$ the disc breaks into two distinct planes. This break persists for more than a tenth of the viscous timescale for the disc. The disc aligns when the break propagates out to the edge of our grid which occurs at $t = 0.13t_{\text{visc}}$.

5.4.1.4 $\theta = 60^\circ$

In Fig. 5.7 we show the disc structures for simulations with $\theta_0 = 60^\circ$. There is again a noticeable difference between the constant effective viscosity and full effective viscosity simulations. At $t = 0.01t_{\text{visc}}$ the disc profile is steepened noticeably more than in the $\theta = 45^\circ$ case and by $t = 0.1t_{\text{visc}}$ all of the simulations have maintained

a break. We can also see that at $t = 0.01t_{\text{visc}}$ in the constant effective viscosity simulation with $\alpha = 0.2$ it is possible for the disc to break without the nonlinear effects.

These simulations suggests two main results of using the full effective viscosities. First, discs can break for higher values of α , and second, discs can break for lower inclination angles.

There are two features of the simulations which we have not discussed so far. The first is the wiggles in the disc at the top of the warp in the full effective viscosity simulations with $\theta = 45, 60$ (Figs. 5.6 & 5.7). In these simulations the disc appears to try to break into more than just two distinct planes. It may be that for smaller α or larger $|\psi|$ the disc can break into more than two planes. However we deem this beyond the scope of the present investigation.

Second, the torques on the disc in the strong warp regions generate rapid mass flow rates into the ring inside the warp. This results in prominent spikes in $\Sigma(R)$ inside regions with large $|\psi|$. Increasing resolution only resolves these spikes better. We therefore leave the investigation of this effect to future work, which will require a different numerical method.

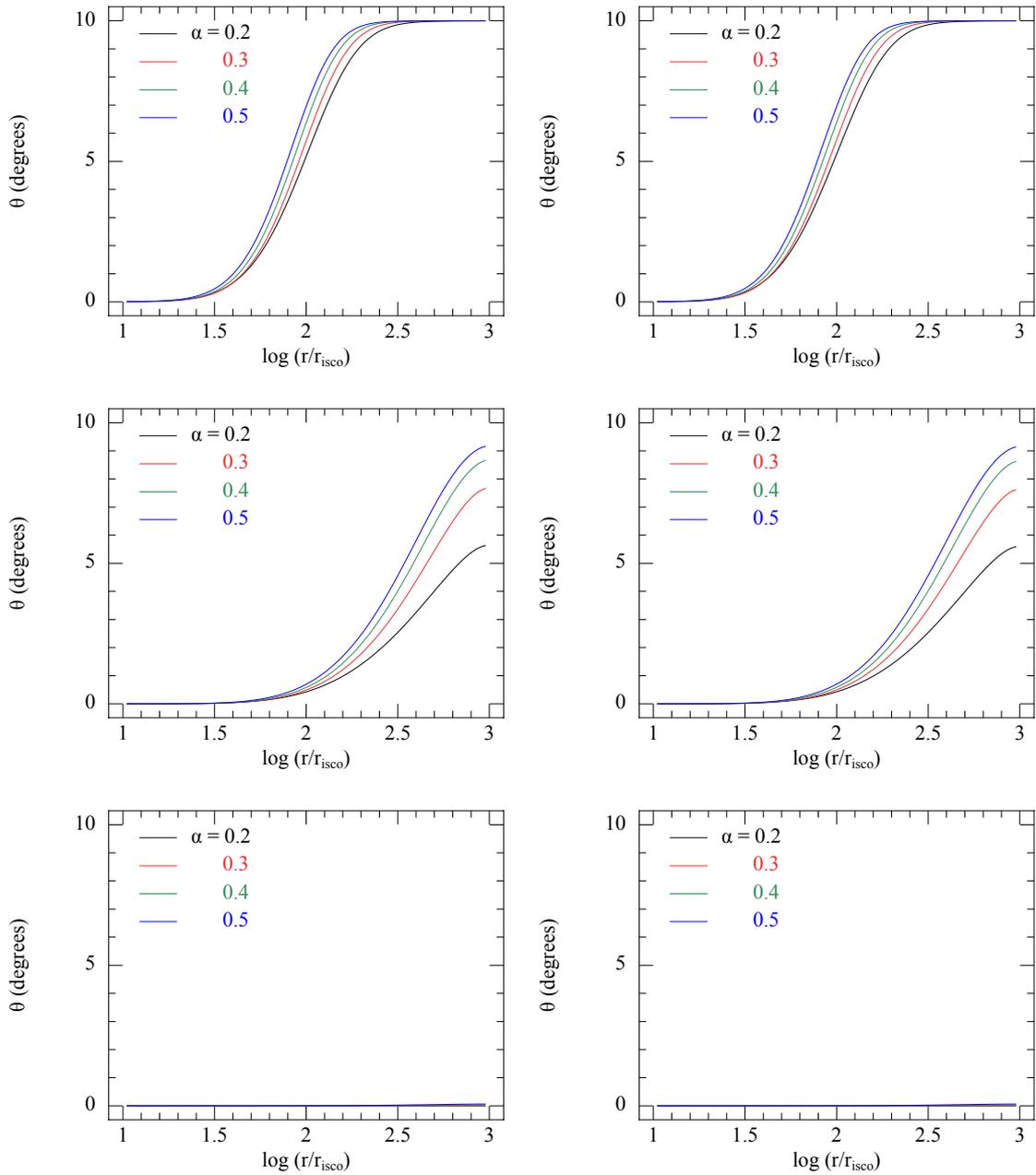


Figure 5.4: Disc structures given by plotting the tilt angle (θ) between the disc and the hole against the log of the radius. The initial misalignment of the disc is $\theta_0 = 10^\circ$. From top to bottom the panels correspond to $t = 0.01, 0.1$ and 1 viscous times after the start of the calculation. The left hand column shows the constant effective viscosity simulations, whereas the right hand column shows the full effective viscosity simulations. On each plot the legend gives the value of α used in the simulation. By $t = 1$ the warp has moved to the edge of our grid and the disc aligns. This shows that for small inclinations the nonlinear effects are negligible.

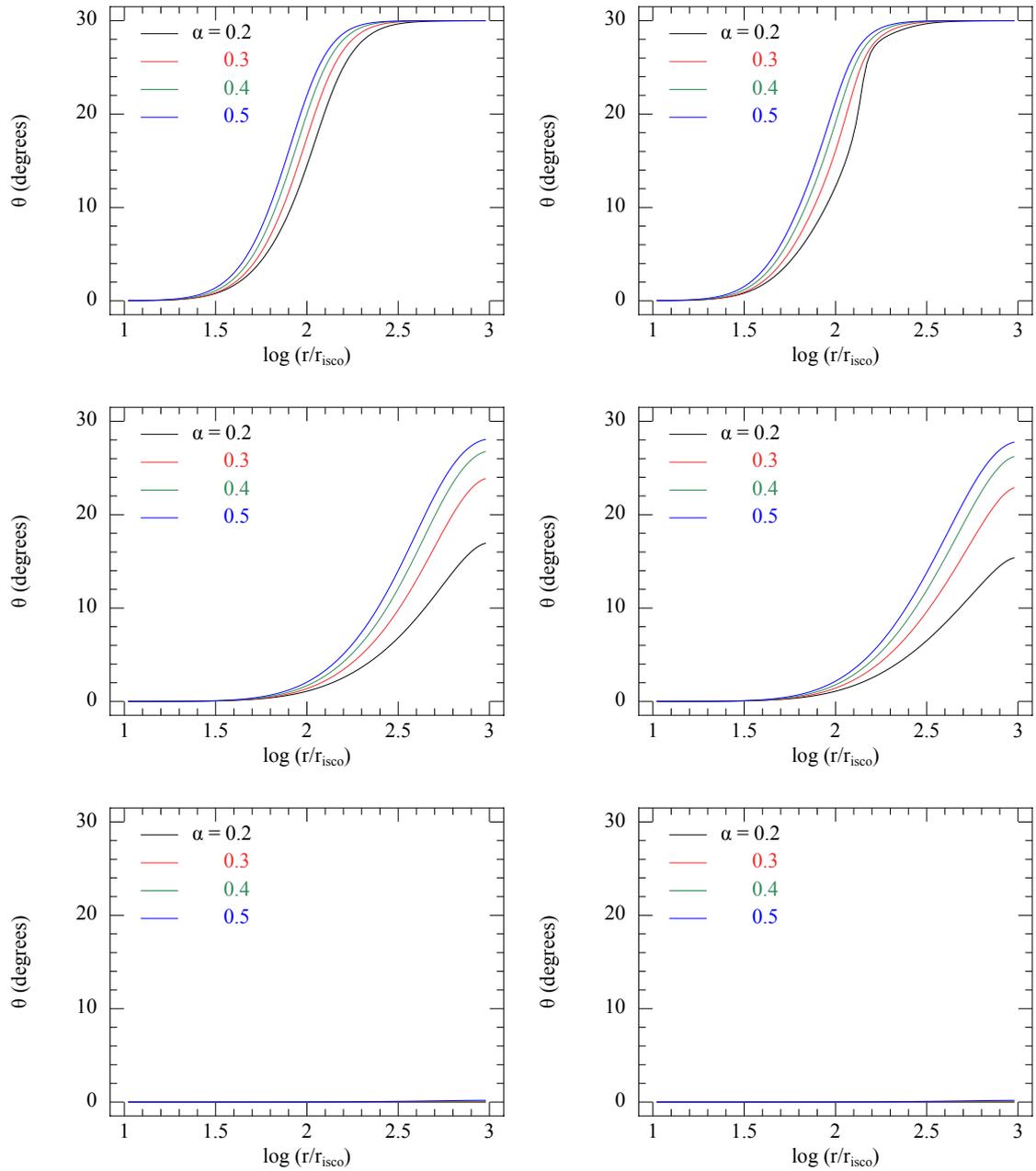


Figure 5.5: Disc structures as in Fig. 5.4, but this time for initial misalignment $\theta = 30^\circ$. From top to bottom the panels correspond to $t = 0.01$, 0.1 and 1 viscous times after the start of the calculation. The left hand column shows the constant effective viscosity simulations, whereas the right hand column shows the full effective viscosity simulations. This figure shows that even for a 30° misalignment angle the effect of the full effective viscosities can be small, however there is a steepening of the disc profile at early times for the $\alpha = 0.2$ simulations. This suggests that for the nonlinear effects to be important at $\theta_0 = 30^\circ$ we need $\alpha \ll 0.2$.

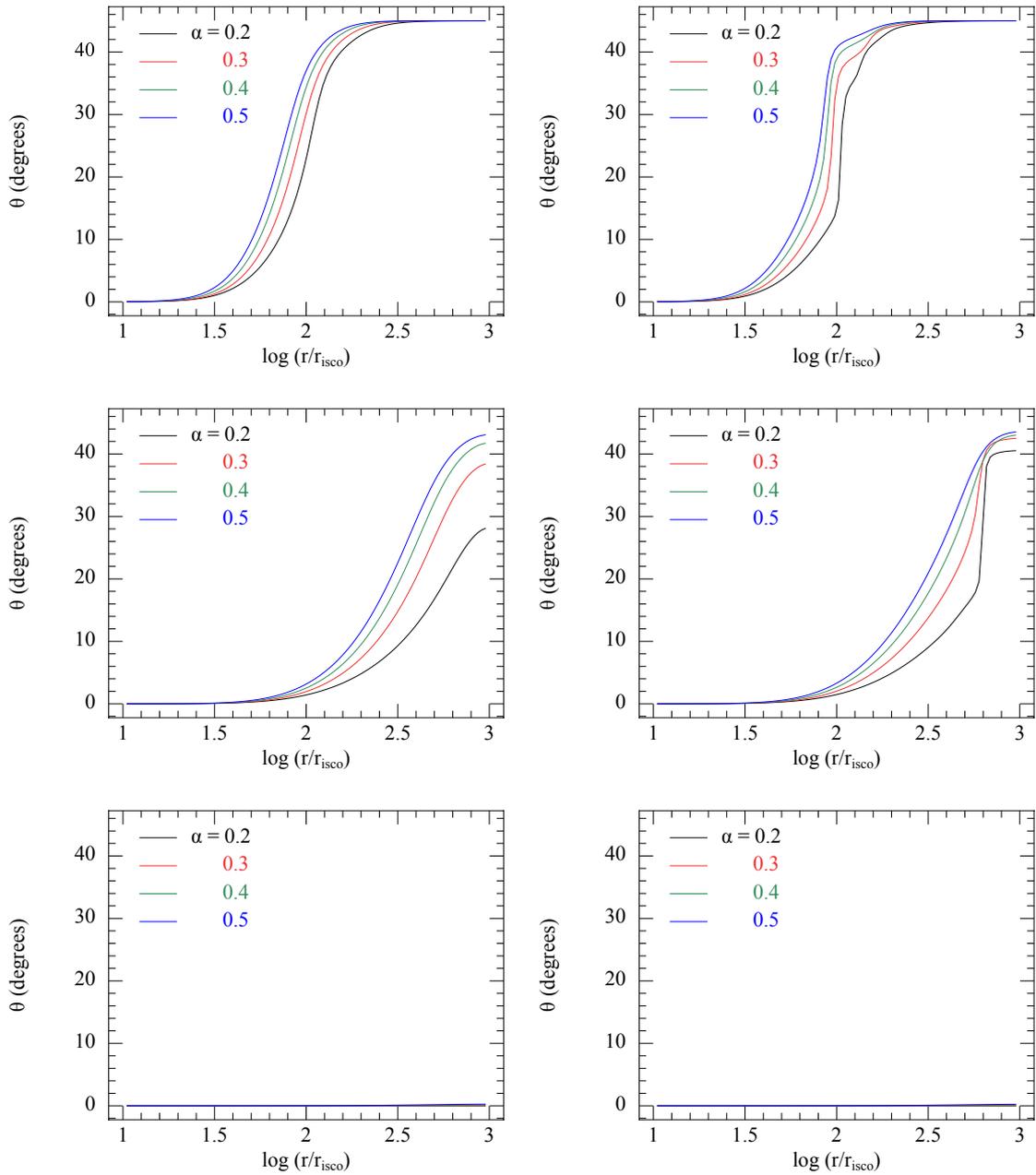


Figure 5.6: Disc structures as in Fig. 5.4, but this time for initial misalignment $\theta = 45^\circ$. From top to bottom the panels correspond to $t = 0.01$, 0.1 and 1 viscous times after the start of the calculation. The left hand column shows the constant effective viscosity simulations, whereas the right hand column shows the full effective viscosity simulations. This figure shows that the nonlinear effects can significantly influence the evolution of the disc. Here the profile is noticeably steepened at $t = 0.01t_{\text{visc}}$ for all α . For $\alpha = 0.2$ the break persists until reaching the outer edge of the grid. For larger α the disc is smoothed before it aligns.

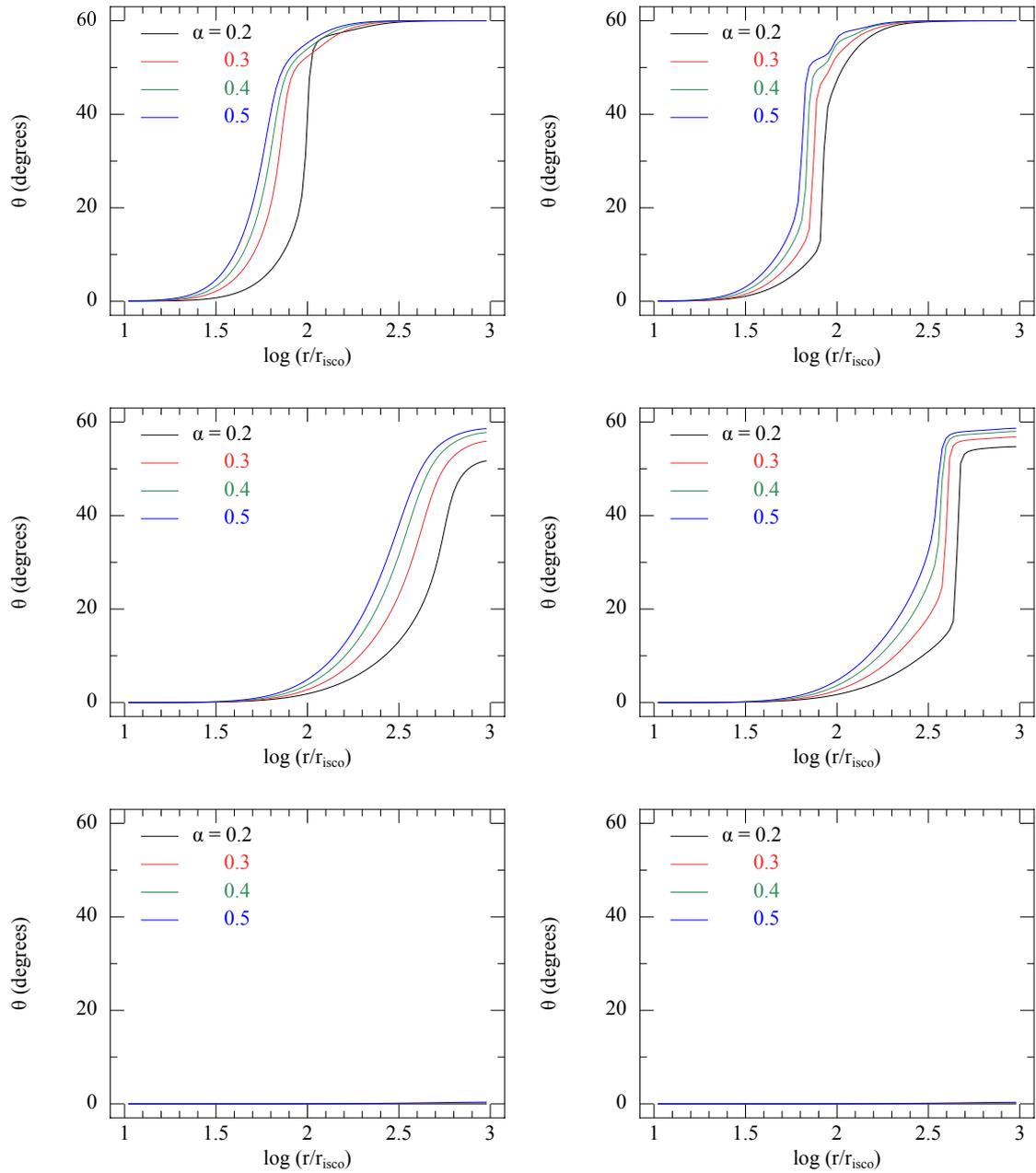


Figure 5.7: Disc structures as in Fig. 5.4, but this time for initial misalignment $\theta = 60^\circ$. From top to bottom the panels correspond to $t = 0.01$, 0.1 and 1 viscous times after the start of the calculation. The left hand column shows the constant effective viscosity simulations, whereas the right hand column shows the full effective viscosity simulations. This figure shows that the nonlinear effects can significantly influence the evolution of the disc. Initially the disc profile is significantly steepened for all α and the disc breaks for all α . This break propagates outwards in the disc until reaching the outer edge of the grid at which point the disc aligns.

5.5 Discussion

We have simulated the evolution of a misaligned accretion event on to a spinning black hole. Many previous simulations of this type assumed that the effective viscosities governing the angular momentum transport were constant. Here we have used effective viscosities constrained to be consistent with the internal fluid dynamics of the disc. We compare these results with those of equivalent simulations which assume constant effective viscosities. In the latter case we find that for small enough α the precession induced by the LT effect can cause the disc to break into two distinct planes if the misalignment is large enough. We also find that if we allow nonlinear effects (Ogilvie, 1999) the disc can break for higher values of α and smaller inclination angles. This suggests that nonlinear effects significantly change the evolution of strongly warped discs. The main effect is in reducing the magnitude of the effective viscosities in the warped region, noticeably reducing ν_1 more than ν_2 . This reduces the disc's ability to smooth out the warp induced by the LT torque and so prevent a break. This is essentially a modified form of the Bardeen–Petterson effect, where the warp is steepened by the weakened disc response.

For small values of α , $\nu_2 \gg \nu_1$. Then the effective viscosity that tries to flatten tilted rings of gas is much stronger than the effective viscosity responsible for transporting angular momentum radially. However the ν_1 and ν_2 torques responsible for communicating angular momentum are proportional to ℓ and $\partial\ell/\partial R$ respectively. So to break the disc we also need a large warp. As $|\psi|$ increases the ν_2 torque becomes comparatively stronger than the ν_1 torque. When the discrepancy between the magnitudes of the two torques is large enough the disc is effectively unable to communicate radially. As the differential precession induced by the LT effect acts faster at smaller radii, this allows the vertical effective viscosity (ν_2) to flatten each ring in turn into the plane of the spinning black hole. To maintain a smooth (small

amplitude) warp the disc must be able to communicate radially on a timescale comparable to the flattening of rings into the plane of the black hole. When this does not hold the disc breaks.

For $\alpha \sim 1$ the disc can efficiently maintain a smooth warp. However for $\alpha \sim 0.2-0.4$ and a large misalignment we show there is a strong tendency for the disc to break under an external torque such as the LT torque. We see no reason for this behaviour to change when $\alpha \leq 0.1$. However there may be other nonlinear effects which become important here. The next step is a full 3D hydrodynamical numerical investigation to determine the disc response in this regime (e.g. Nelson and Papaloizou 2000; Lodato and Price 2010).

When the disc breaks it splits into a co- or counter-aligned inner disc and a misaligned outer disc. This break can propagate outwards and (dependent on α and θ_0) may persist until it reaches the outer regions of the disc. At this point the whole disc is co- or counter-aligned with the hole spin. This suggests that a break in the disc can be a long-lived feature, lasting for roughly the alignment timescale. We note that although we have not included any simulations of counter-aligning discs, it is already known that for $J_d/J_h \ll 1$ there is a symmetry in the behaviour of discs of this type around $\theta = 90^\circ$ (King et al. 2005 and Lodato and Pringle 2006). Therefore the same behaviour would be seen in discs that are counter-aligning with the black hole (i.e. a misalignment of 120° would produce the same dynamics with a counter-aligning disc as the 60° case produces here).

For observational purposes it is important to compare the alignment timescale t_{al} with the lifetime t_{life} of the system. If $t_{\text{al}} > t_{\text{life}}$, the centre of the disc is never aligned with its warped outer plane. In an accreting stellar-mass black-hole binary this would mean that jets (which propagate along the axis of the central parts of the disc, and so the spin of the black hole) are in general misaligned with respect to the

orbital plane. This is seen in the black–hole binary GRO J1655-40 (Martin et al. 2008 and references therein). Similarly the continuum–fitting method for estimating black–hole spins (Kulkarni et al. 2011 and references therein) uses the assumption that the black hole spin is aligned with the orbital axis in estimating the radiating disc area and hence the ISCO radius. It therefore implicitly requires $t_{\text{al}} < t_{\text{life}}$. In our simulations it is clear that discs are able to remain warped or broken for significant fractions of their viscous timescale. However we note that for strong warps there may be other nonlinear effects not included in the analysis of Ogilvie (1999) or additional dissipation caused by fluid instabilities (cf. Gammie et al. 2000). These may impact upon the alignment timescale.

It is also possible that the disc is never able to align fully; for example an accretion disc in a misaligned black hole binary system may be twisted one way by the LT effect and a different way by the binary torque (e.g. Martin et al., 2009). We therefore suggest careful consideration before assuming that black hole accretion discs are aligned – particularly as Lodato and Pringle (2006) suggest that the accretion rate through warped discs is significantly enhanced, which suggests the possibility that strongly accreting discs are also strongly warped.

Disc breaking can have some important observational consequences. For example accretion discs in binary systems often have an outer disc aligned to the binary plane. However the inner disc (and so the direction of any jets) must be aligned with the black hole spin, which may be significantly misaligned with respect to the binary axis. Similarly breaks in protoplanetary discs may occur through the internal disc response to warping during formation of the disc, rather than being driven by the presence of a planet.

We have used a 1D Eulerian ring code to calculate the evolution of misaligned accretion events on to a spinning black hole. We caution that there is still much

work to do before we can fully understand the evolution of such a disc. For example our treatment requires gas to orbit on circular rings, inherently excluding any effects which might break the cylindrical symmetry in the local disc plane. We also assume that the disc responds viscously, with no wave-like behaviour. The method also prevents us from examining exactly what happens in the disc break: by its nature, it occurs over only over a small number of rings and is therefore poorly resolved – however this does not affect our conclusion that the disc can break. The most frustrating restriction is the difficulty in simulating small enough α : the simulations we present here suggest that for small α nonlinear fluid effects can be particularly important. We will address these restrictions in future investigations.

We remark finally that although we have considered alignment effects induced by the LT effect on an accretion disc around a spinning black hole, the torque between a misaligned binary and an external accretion disc has a very similar form (Nixon et al., 2011b, Chapter 2). Thus we expect similar phenomena to appear in that case. The enhanced accretion rates through the disc due to the dissipation in the warp (Lodato and Pringle, 2006) would increase the mass flow rate on to the binary. Therefore the coalescence timescale for the binary would be reduced further (Nixon et al., 2011a, Chapter 4).

6

Rapid AGN accretion

6.1 Abstract

Accretion in the nuclei of active galaxies may occur chaotically. This can produce accretion discs which are counterrotating or strongly misaligned with respect to the spin of the central supermassive black hole (SMBH), or the axis of a close SMBH binary. Accordingly we consider the cancellation of angular momentum in accretion discs with a significant change of plane (tilt) between inner and outer parts. We estimate analytically the maximum accretion rate through such discs and compare this with the results of Smoothed Particle Hydrodynamics (SPH) simulations. These suggest that accretion rates on to supermassive black holes may be larger by factors $\gtrsim 100$ if the disc is internally tilted in this way rather than planar. This offers a natural way of driving the rapid growth of supermassive black holes, and the coalescence of SMBH binaries.

6.2 Introduction

Astronomers still do not know how the supermassive black holes (SMBH) in galaxy centres accrete gas and grow. The problem lies in the very small angular momentum this gas must have in order to accrete on a reasonable timescale. For example, the viscous time of a standard accretion disc of ~ 1 pc radius accreting on to a SMBH of $10^8 M_\odot$ approaches a Hubble time. A natural remedy for this is to assume that on small scales near the hole, accretion occurs in episodes whose angular momenta are randomly oriented with respect to each other. This chaotic picture predicts that the spin of the hole remains low, allowing rapid mass growth provided that there is an adequate mass supply (King and Pringle 2006; King and Pringle 2007; King et al. 2008; Fanidakis et al. 2011). A characteristic feature of this type of accretion is that the disc flow on to the SMBH is often retrograde (King et al., 2005) with respect to

the hole spin, or to the rotation of an SMBH binary resulting from a galaxy merger. In Chapter 4 (Nixon et al., 2011a) we showed that a retrograde coplanar external gas disc can be markedly more efficient in shrinking an SMBH binary than a prograde one. A retrograde disc can cancel the binary orbital angular momentum directly, rather than being slowed by the resonances which always arise in a prograde circumbinary disc. It is natural to ask if this angular momentum cancellation can work if we replace the SMBH binary with a pre-existing gas disc surrounding a single SMBH. If so, this may be a mechanism promoting much faster delivery of gas on to the black hole. This situation may arise naturally from misaligned accretion events, for example where an accretion event forms a disc around the black hole, and a second event forms an outer disc which is misaligned with respect to the first one. Similar cases occur as a misaligned accretion disc attempts to reach an axisymmetric (co- or counter-aligned) configuration around a spinning black hole (King et al. 2005; Lodato and Pringle 2006), and also during the closely related co- or counter-alignment of an accretion disc around a central binary (Nixon et al., 2011b, Chapter 2).

In all these cases integrating the full dynamics of the accretion disc is complex and time-consuming because the component discs can warp before interacting. To study the mixing of distinct or opposing disc angular momenta we investigate a simple form of this kind of accretion. We assume an accretion disc with initially distinct planes in its inner and outer parts. We let these two parts interact viscously and determine how much the accretion is enhanced compared with a disc lying in a single plane. In Section 6.3 we give simple analytical estimates of this. In Section 6.4 we perform global 3D hydrodynamical simulations to examine the disc's dynamical behaviour in the light of these estimates. We interpret our results in Section 6.5.

6.3 Counterrotating discs

We consider an inner and outer disc counterrotating with respect to each other but not coplanar, that is, the angle θ between the disc angular momentum vectors obeys $\pi/2 < \theta < \pi$.

If the discs were in perfect contact the gas would share opposed angular momenta from almost the same radii. This may be unphysical (or imply a very rapid evolution away from the initial state) since the discs may attempt to open a gap of some size, as we discuss later. For generality we choose an arbitrary gap between the discs of size a .

Now we assume that the gas efficiently shares angular momentum across this gap and that an equal mass of gas from each disc takes part in the interaction. The resultant angular momentum of the interacting gas is

$$L_{\text{new}} = \frac{1}{2} [L(R_{\text{gap}} + a) + L(R_{\text{gap}})] \quad (6.1)$$

$$L_{\text{new}} \approx \frac{1}{2} \sqrt{GM R_{\text{gap}}} \left(1 + \frac{1}{2} \frac{a}{R_{\text{gap}}} + \cos \theta \right) \quad (6.2)$$

to first order in a/R_{gap} , where R_{gap} is the outer edge of the inner disc (so that $R_{\text{gap}} + a$ is the inner edge of the outer disc), G is the gravitational constant, M is the mass of the central object, and we have assumed a Keplerian potential for the gas.

The circularisation radius for this interacting gas is

$$R_{\text{circ}} = \frac{1}{4} R_{\text{gap}} \left(1 + \cos \theta + \frac{1}{2} \frac{a}{R_{\text{gap}}} \right)^2 \quad (6.3)$$

We expect the size of the gap between the two discs to be roughly the scale of the

epicyclic motions in the gas disc as this is where fluid orbits cross and interact. This implies a gap size $\sim H(R_{\text{gap}})$, where H is the semi-thickness of the disc, so that

$$R_{\text{circ}} = \frac{1}{4} \left(1 + \cos \theta + \frac{1}{2} \frac{H}{R} \right)^2 R_{\text{gap}}. \quad (6.4)$$

For typical AGN disc parameters, $H/R \sim 10^{-3}$, and there are two distinct cases. For θ close to π , we can have $1 + \cos \theta \ll H/R$, so that

$$R_{\text{circ}} \approx \frac{1}{16} (H/R)^2 R_{\text{gap}}. \quad (6.5)$$

This suggests that the gas orbits would decrease by ~ 7 orders of magnitude in radius which implies direct accretion of the gas as long as its path to the hole is clear.

If however θ is not close to π we have $1 + \cos \theta \gg H/R$, and

$$R_{\text{circ}} \approx \frac{1}{4} (1 + \cos \theta)^2 R_{\text{gap}}. \quad (6.6)$$

Then the gas orbit decreases least when $\theta \sim \pi/2$, where $R_{\text{circ}} \sim \frac{1}{4} R_{\text{gap}}$. The orbit scale (6.4) varies smoothly between these two extreme values.

The accretion rate through a disc is approximately

$$\dot{M} \sim \frac{M_{\text{disc}}}{\tau_{\text{visc}}} \quad (6.7)$$

where M_{disc} is the disc mass and τ_{visc} is the viscous timescale

$$\tau_{\text{visc}} \sim \frac{R^2}{\nu} \quad (6.8)$$

where R is a characteristic radius for the disc and ν is a measure of the disc viscosity.

Thus for constant viscosity¹ the accretion rate is $\propto R^{-2}$. This suggests that for $\theta > \pi/2$, where the gas orbits decrease by a factor > 4 , there is potential for the accretion rate to increase by a factor > 16 . For large $\theta \approx \pi$ the cancellation can lead to direct accretion of gas on a dynamical timescale. This allows accretion on a local disc filling timescale at the radius of the gap. This timescale is much shorter than the viscous timescale as we only need move gas to R_{gap} and not right on to the hole. Note that the rate at which mass can be supplied to R_{gap} is clearly an upper bound on the possible accretion rate from this process.

If the gas does not accrete directly then the viscous timescale is shortened to

$$\tau_{\text{visc, circ}} \sim (R_{\text{circ}}/R_{\text{gap}})^2 \tau_{\text{visc, gap}}. \quad (6.9)$$

Although these estimates are suggestive, we caution that we have assumed that the gas shares angular momentum efficiently, and that an equal mass of gas interacts from each disc. In reality neither of these simplifications may be valid. Accordingly we use Smoothed Particle Hydrodynamics (SPH) to simulate the interactions between the discs.

6.4 Simulations

6.4.1 Code Setup

We use PHANTOM, a low-memory, highly efficient SPH code optimised for the study of non-self-gravitating problems. This code has performed well in related simulations. For example Lodato and Price (2010) performed simulations of warped

¹This is only approximate as one expects ν to increase with radius.

accretion discs and found excellent agreement with the analytical work of Ogilvie (1999) on the nature of the internal accretion disc torques in response to warping.

The implementation of accretion disc α -viscosity (Shakura and Sunyaev, 1973) in PHANTOM is described in Lodato and Price (2010). Specifically, we use the ‘artificial viscosity for a disc’ described in Sec. 3.2.3 of Lodato and Price (2010), similar to earlier SPH accretion disc calculations (e.g. Murray, 1996). The main differences compared to standard SPH artificial viscosity are that the disc viscosity is applied to both approaching and receding particles and that no switches are used. Our implementation also differs slightly from Lodato and Price (2010) in that we retain the β^{AV} term in the signal velocity in order to prevent particle interpenetration from occurring in high-velocity stream-disc collisions. The disc viscosity in PHANTOM was extensively calibrated against a 1D thin α -disc evolution in Lodato and Price (2010) (c.f. Fig. 4 in that paper) and the disc scale heights employed here are similar. As the exact value of α is unimportant given the dynamical nature of the simulations, we simply use $\alpha^{\text{AV}} = 1$ and $\beta^{\text{AV}} = 2$ which at the employed resolution (see below) corresponds to a physical viscosity $\alpha_{\text{SS}} \approx 0.02\text{--}0.06$.

For these simulations we model the gravity of the accretor with a Newtonian point mass potential, adopting an accretion radius of 0.1 in code units. We start with a flat disc of gas, composed of 10 million SPH particles, in hydrostatic equilibrium between 1.0 to 2.0 in radius, and surface density distribution $\Sigma \propto R^{-1}$, setup using the usual Monte-Carlo technique. We also choose a vertical density profile corresponding to $H/R = 0.02$ at $R = 1$ and employ an isothermal equation of state. To produce the two distinct planes for the disc we rotate the outer half (in radius) of the disc by an angle θ . We then perform simulations with different values of θ to find how the accretion rate changes with inclination angle.

6.4.2 Tilted disc evolution

We perform seven simulations corresponding to inclination angles $\theta = 30n^\circ$ where $n = 0, 1, \dots, 6$. The $\theta = 0^\circ$ simulation gives a flat-disc accretion rate to which we compare the other simulations. We note that the 180° case is a rather unrealistic setup as it would require perfectly anti-parallel accretion events. Even in the co- and counter-aligned discs predicted by the numerical simulations of Lodato and Pringle (2006) the discs never achieve a configuration where θ is precisely 180° .

The simulations were run with an isothermal equation of state, i.e. the temperature of each particle was held constant for each simulation (see Section 6.4.2.2 below for a discussion of the effects of different equations of state). This implies that the pressure, determined by $P = c_s^2 \rho$, is simply proportional to ρ , and the sound speed c_s is constant in time and the same for all particles.

In Fig. 6.1 we report the accretion rates achieved for the different values of θ . This shows that for discs inclined by less than 90° the accretion rates are similar to that of a flat disc for the duration of our simulations. In this case the interaction between inclined gas orbits is relatively weak with the gas staying on near circular orbits, with planes slowly varying from the inner to outer disc. In contrast the simulations with $\theta \geq 90^\circ$ show a dramatic increase in the accretion rate. This is caused by the direct cancellation of orbital angular momentum and energy where the orbits of the disc gas cross.

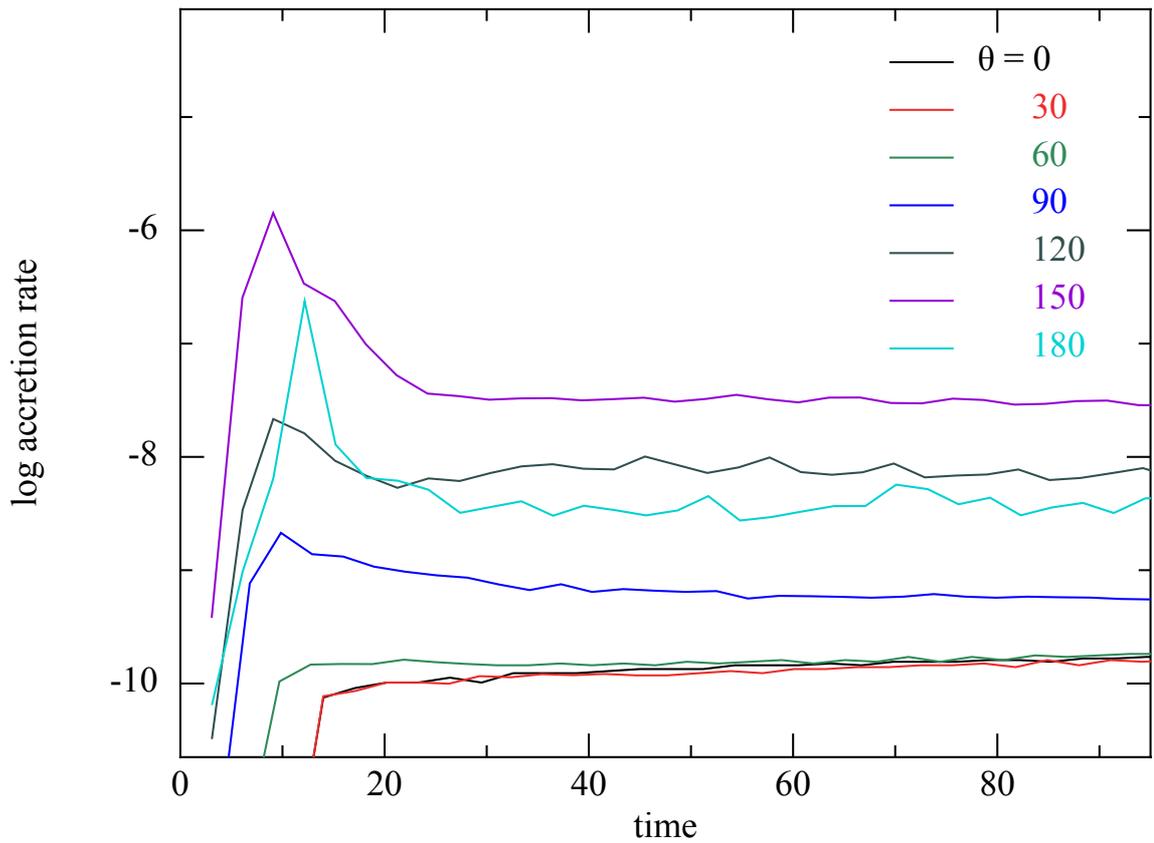


Figure 6.1: The accretion rates for the isothermal simulations with different inclination angles θ . Time is in units of the dynamical time at $R = 1$. The discs with a tilt $\geq 90^\circ$ are much more efficient at driving accretion, as these involve a degree of counterrotation and so direct cancellation of angular momentum.

In Fig. 6.2 we show the disc structure for the simulation where $\theta = 30^\circ$. Here the two discs initially join to form a coherent disc with a warped region joining the still misaligned inner and outer discs. Gas is quickly depleted from the warp region, leaving a clear drop (of 2-3 orders of magnitude) in projected density in the warp region. This break between the discs persists throughout the simulations as the two discs slowly align with each other, reaching an angular separation of $\sim 10^\circ$ by the time the simulation ends. The large gradient in specific angular momentum in the warp region causes gas to quickly move on past the warp. This causes the low density in the warp region which is the disc break (Nixon and King, 2012, Chapter 5). The disc therefore maintains two distinct planes. We still resolve the rings of gas between the two discs but this is clearly of much lower density than the inner and outer discs (cf. Lodato and Price 2010). In Fig. 6.3 we show the similar disc structure for the simulation where $\theta = 60^\circ$.

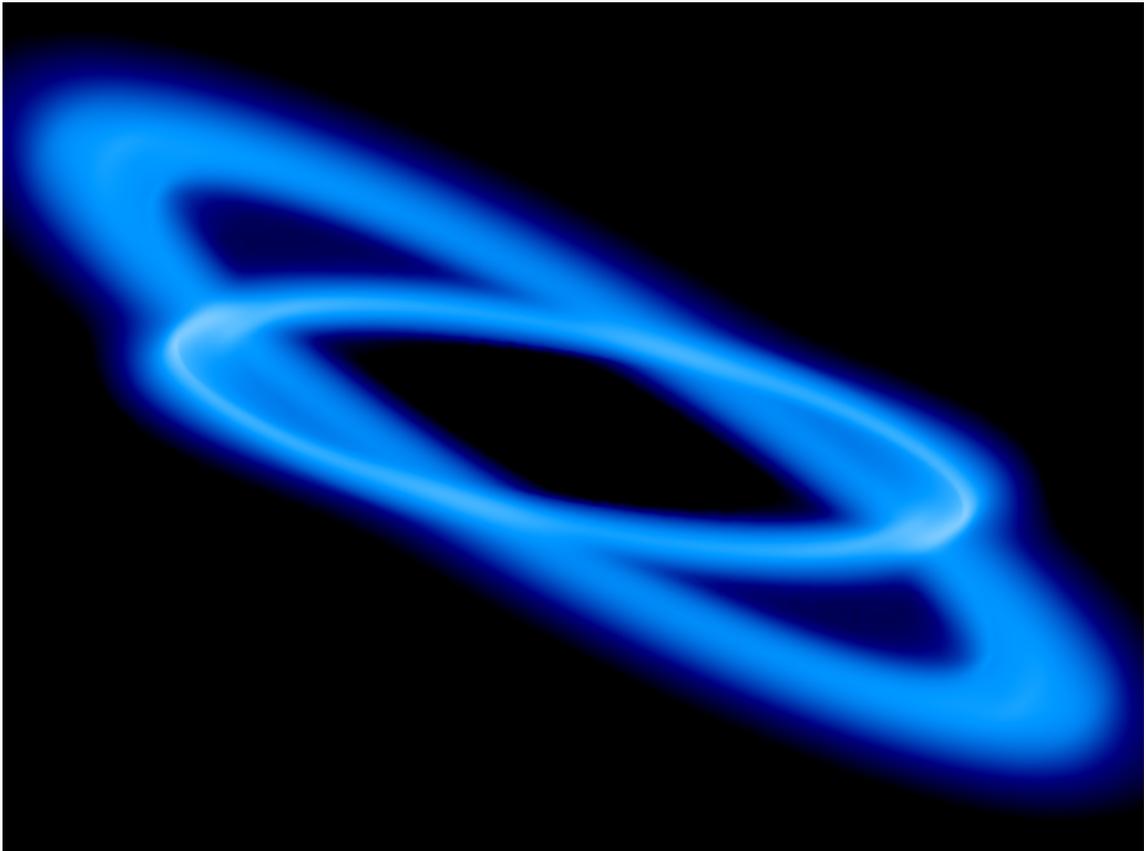


Figure 6.2: Column density projection showing the disc structure for the $\theta = 30^\circ$ simulation. The disc has evolved for ~ 50 dynamical times (~ 10 orbits).

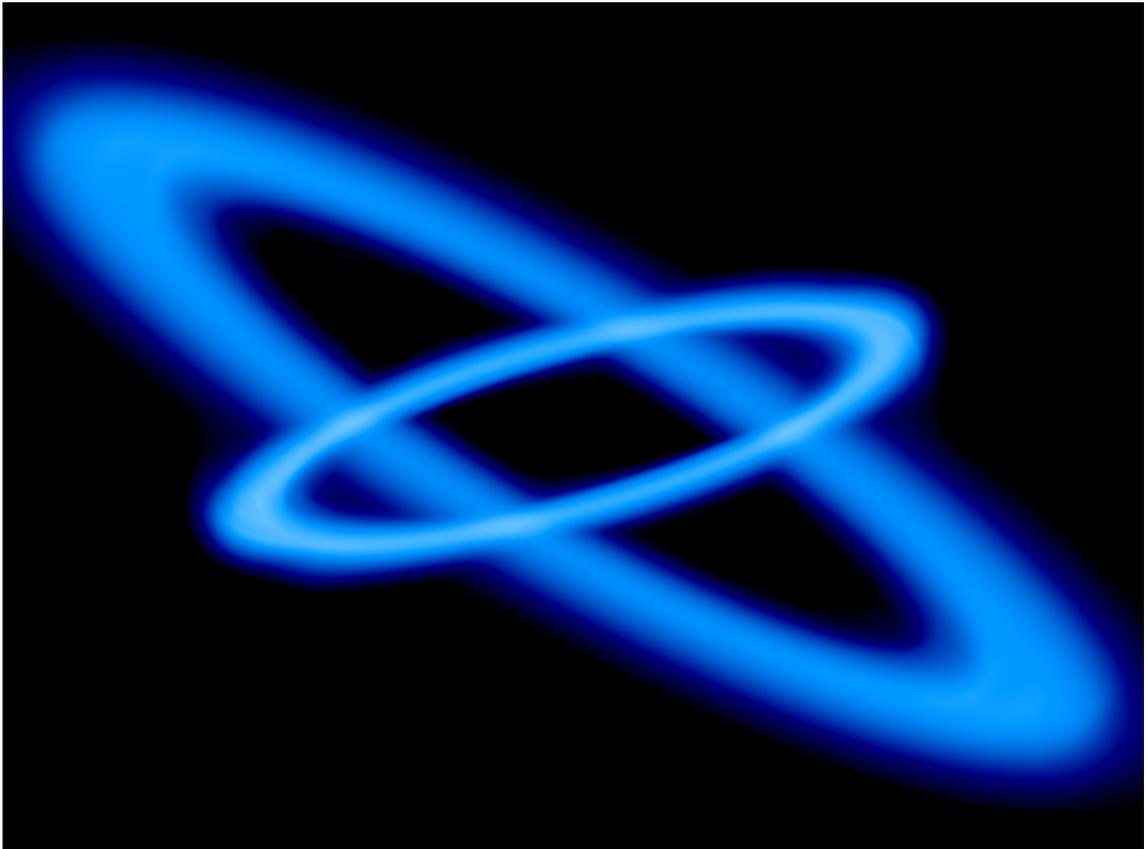


Figure 6.3: Column density projection showing the disc structure for the $\theta = 60^\circ$ simulation. The disc has evolved for ~ 50 dynamical times (~ 10 orbits).

The $\theta = 90^\circ$ simulation (not illustrated) has an initial period of chaotic flow between the two discs, until the inner disc moves towards an inclination of $\sim 85^\circ$ after mixing with some of the gas from the outer disc. At this point a strong warp is set up and the flow continues in a similar way to $\theta = 60^\circ$.

The evolution of the $\theta = 120^\circ$ and $\theta = 150^\circ$ simulations proceed in a dramatically distinct way to those with smaller θ . This is simply because the sharing of angular momentum now causes significant changes to the gas orbits. In the $\theta = 120^\circ$ simulation gas falls from the region between the two discs and circularises at a radius inside the inner accretion disc. In this process some of gas is also accreted, as it carries a spread of angular momentum. Clearly this gas does not have precisely the angular momentum of the inner disc; it is a mix of gas from both discs. Therefore the quasi-steady state for this simulation is an outer disc causing cancellation with the outer edge of the original inner disc. This gas falls and feeds both direct accretion and the new innermost disc (white region of high density in Fig. 6.4). This innermost disc is strongly warped with respect to the original inner disc and so there is again extra dissipation between these two discs as they try to straighten. The disc structure can be seen in Fig. 6.4.

In the $\theta = 150^\circ$ simulation the gas falls to a much smaller radius from the region between the two discs. Initially a large amount of gas is directly accreted on to the sink particle representing the central accretor. The strong dissipation occurring early in the simulation also causes the inner disc to spread all the way in to the accretion radius. Some of the falling gas tries to circularise at a radius where it is forced to impact upon the inner disc. As it still has the sense of angular momentum of the outer disc this causes more cancellation of angular momentum, driving more accretion. For these reasons this simulation generates the highest accretion rates – this is simply because it causes the greatest mixing of angular momentum in the

gas. The disc structure for this simulation can be seen in Fig. 6.5.

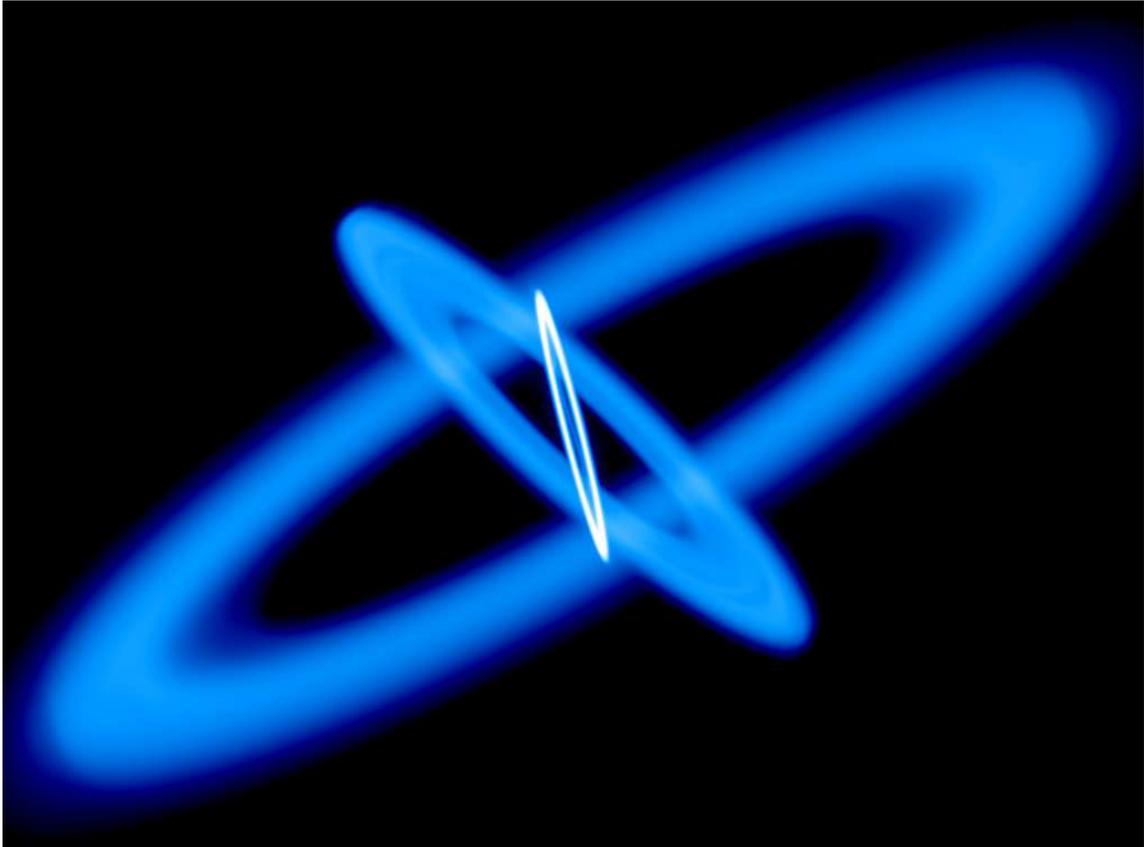


Figure 6.4: Column density projection showing the disc structure for the $\theta = 120^\circ$ simulation. The disc has evolved for ~ 50 dynamical times (~ 10 orbits). Note the new innermost disc (of high density and therefore white) which has formed from the dynamical infall of gas from the warp region.

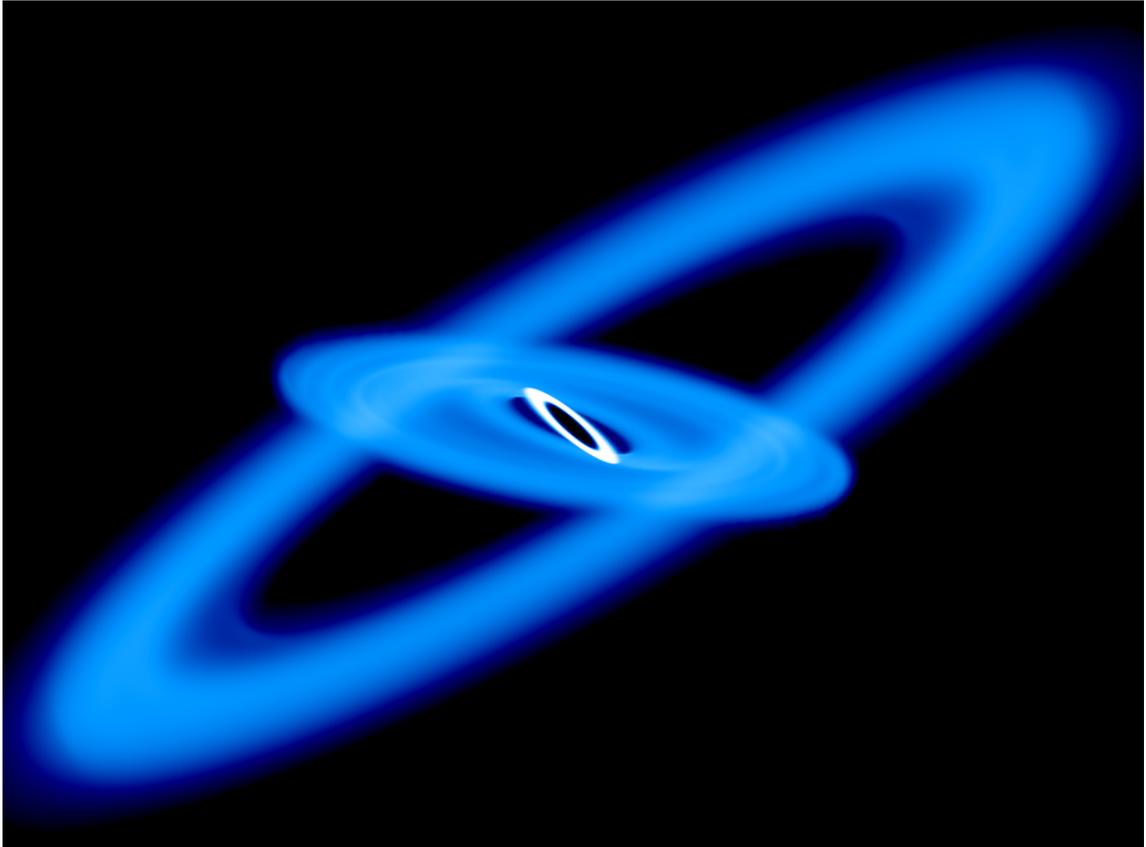


Figure 6.5: Column density projection showing the disc structure for the $\theta = 150^\circ$ simulation. The disc has evolved for ~ 50 dynamical times (~ 10 orbits). Note the new innermost disc (of high density and therefore white) which has formed from the dynamical infall of gas from the warp region.

6.4.2.1 $\theta = 180^\circ$

The $\theta = 180^\circ$ case does not initially appear a very realistic initial setup. However something close to this configuration may appear in the evolution of accretion discs around spinning black holes (cf. King et al. 2005; Lodato and Pringle 2006) and also during the evolution of an accretion disc around a binary system (cf. Nixon et al. 2011b; Chapter 2).

Initially while the two discs (inner and outer) are in contact there is great cancellation of angular momentum for the gas. This gas then attempts to fall towards the central sink particle. It therefore impacts upon the inner disc. The pressure of the inner disc gas is enough to force the falling gas out of the plane of the disc where it can continue on a low angular momentum orbit towards the centre. It then circularises inside the inner disc (with some accretion) after colliding with gas on similar orbits and losing orbital energy in shocks. This new innermost disc clearly has the same sense of angular momentum as the outer disc since this has the larger specific angular momentum. Once the original inner disc has spread inwards (and this new innermost disc has spread outwards), exactly the same process starts again between these two discs. See Fig. 6.6 for the disc structure.

Angular momentum cancellation thus proceeds as follows. Imagine the two discs are not in contact. They spread viscously until gas from the discs can interact, sharing and so destroying angular momentum. Now consider two test particles (one from the inner disc and one from the outer disc) which interact. The particle from the inner disc adopts a slightly eccentric orbit, which passes through its parent disc. It rejoins its own disc, which shrinks slightly to accommodate the loss of angular momentum. The particle from the outer disc also adopts a slightly eccentric orbit (inside its original orbit). But this forces it to interact more with the ‘hostile’ inner

disc. This causes runaway angular momentum loss for the outer disc particle, almost to the point of free-fall to the centre. Now the particle's free-fall energy is enough to move it out of the disc plane (helped by the pressure force from the inner disc). It thus moves over and then inside the inner disc. The overall effect is to shrink the inner disc and move the outer disc particles on to orbits that pass inside the inner disc.

At this point one of three things can happen to the outer disc particle:

- 1) it may have cancelled enough angular momentum to accrete directly.
- 2) it may have cancelled enough to fall inside the inner edge of the discs where it crosses the plane of the inner disc and meets other particles on similar orbits: the particles shock and dissipate energy, forming a new disc at much smaller radii. This disc should be planar, but any small perturbations in the initial discs will be amplified here at smaller radii and could produce a disc with a substantial warp.
- 3) The particle's new orbit may take it through the inner (initial) disc. If this is the case it again interacts with the (hostile) disc and falls further. There are again two possibilities here. Either the particle bounces all the way in past the hostile disc, or it has so many interactions that it adopts the sense of rotation of the hostile disc. So the particle either accretes or drives the inner disc to shrink, and hence drives accretion through the inner disc.

All sequences here greatly enhance the central accretion rate while the discs are trying to mix their angular momenta.

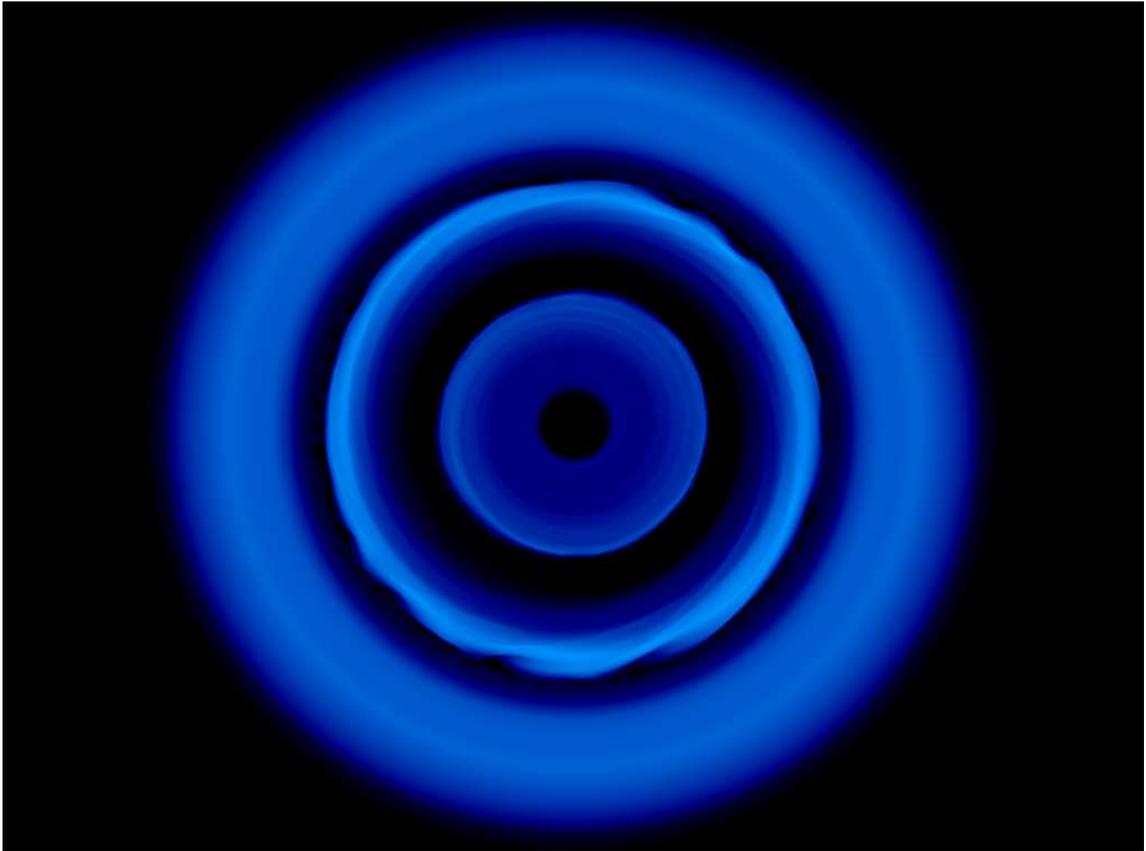


Figure 6.6: Column density projection showing the disc structure for the $\theta = 180^\circ$ simulation. The disc has evolved for ~ 50 dynamical times (~ 10 orbits). The innermost disc is formed of gas falling from the gap between the original inner disc (now the middle disc) and the outer disc. In this picture the innermost and outermost discs are rotating clockwise and the middle disc rotates counter-clockwise.

6.4.2.2 Effect of the gas equation of state

We have also investigated the effect of changing the thermal treatment of the gas. In the simulations detailed above we used an isothermal equation of state. The gas was assumed to radiate away any dissipative heating instantly. The other extreme is to assume that any dissipation feeds the internal energy of the particles, which are not allowed to cool by radiative losses. Therefore we also simulate the $\theta = 150^\circ$ case using an adiabatic equation of state (with $\gamma = 5/3$), transferring the kinetic energy dissipated by the viscous terms into internal energy. In all other ways this simulation is precisely the same as the corresponding isothermal simulation. In Figs. 6.7 & 6.8 we show the disc evolution after ~ 20 dynamical times. In the isothermal case the gas cancels angular momentum and falls to circularise at a new radius. In comparison Fig. 6.8 shows that the equation of state can play a significant role in the hydrodynamics. In this case the gas is significantly heated when it shocks and thus initially expands violently. This process drives more mixing of the gas and thus the accretion is now almost a factor of two higher here than in the isothermal case. The gas is still bound to the black hole and so returns on chaotic orbits to eventually accrete.

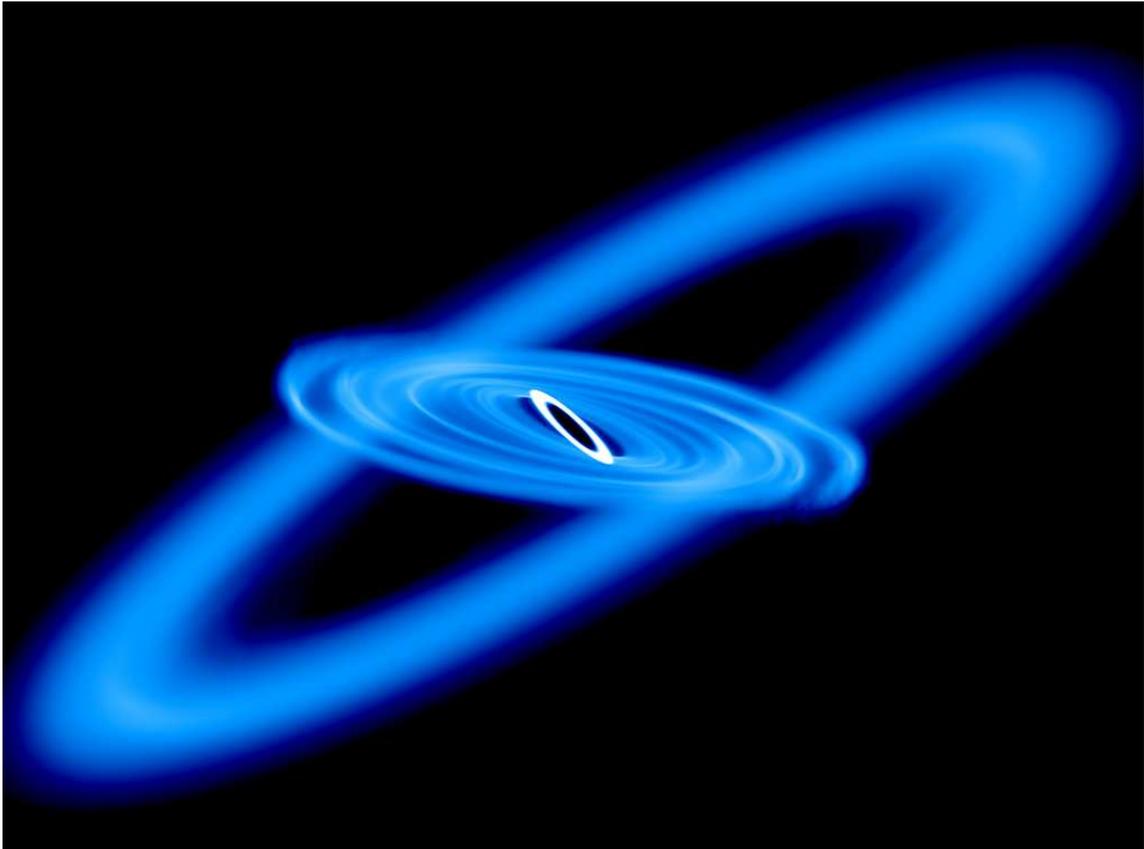


Figure 6.7: Column density projection showing the disc structure for the $\theta = 150^\circ$ simulation with an isothermal equation of state. The disc has evolved for ~ 20 dynamical times. Note that the gas is not heated by shocks, and thus can fall to form a smaller disc. This figure should be compared to Fig. 6.8 where the thermal treatment allows shock heating.

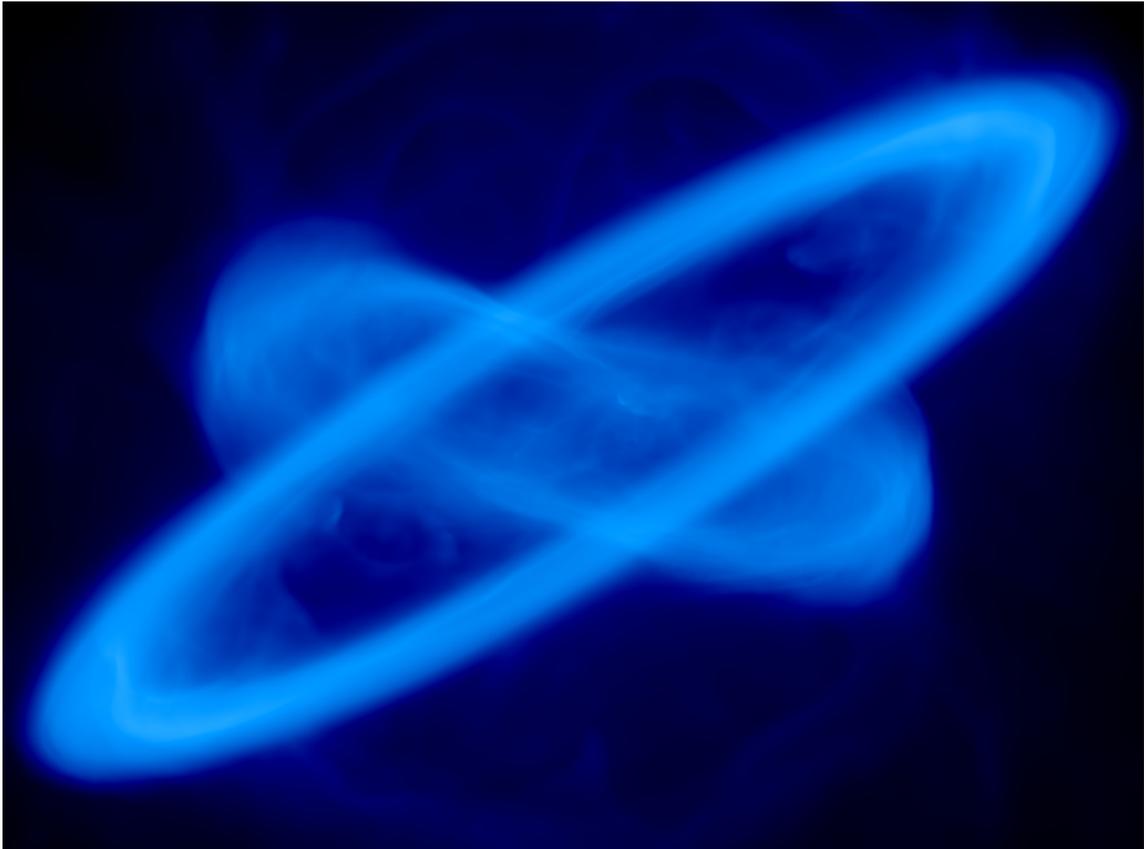


Figure 6.8: Column density projection showing the disc structure for the $\theta = 150^\circ$ simulation with an “adiabatic” equation of state (see the text for a full definition). The disc has evolved for ~ 20 dynamical times. Note that the gas no longer can cool to form a smaller disc (cf. Fig. 6.7), but instead forms outflows which return on chaotic orbits generating a lot of cancellation of angular momentum.

6.5 Discussion

Retrograde circumbinary accretion discs are natural in the chaotic accretion picture (King and Pringle 2006; King and Pringle 2007), and offer the possibility of cancelling the binary orbital angular momentum (Nixon et al., 2011a, Chapter 4). We have performed numerical simulations of accretion discs with two initially distinct planes. In the region between the two discs angular momentum is shared, driving accretion, particularly if $\theta > \pi/2$.

The highest accretion rates come from discs with inclination $\theta \approx 150^\circ$. Here the gas cancels angular momentum when the discs meet, allowing this gas to fall directly to the centre without the need to interact with the inner disc. Clearly this gas would in reality have some residual angular momentum and so would circularise at a new smaller radius. In the simulations however the gas falls inside our accretion radius for the central sink particle and so is directly accreted. Gas falling from between the discs always has angular momentum with the same sense of direction as the outer disc. Therefore if the gas tries to circularise outside the accretion radius it impacts the inner disc, cancelling more angular momentum and falling further until it is accreted.

The accretion rates found here are > 100 times those of a flat planar disc. This increase comes from the mixing of angular momenta and consequent dynamical infall of gas to smaller radii. As gas is depleted from the gap the disc tries to spread and fill the gap, promoting more cancellation of angular momentum and so more accretion. The timescale for gas accretion is shortened (cf. Eq. 6.9). If enough angular momentum is cancelled then the gas accretes directly on a dynamical timescale.

Our simulations show that discs with an initial inclination less than 90° evolve into

coherent discs joined by a warp. Over time, dissipation from this warp brings the two discs into a common plane. This extra dissipation enhances accretion, the effect increasing with inclination angle (although negligible in comparison to the enhanced accretion when $\theta > \pi/2$). The discs appear to break, in the sense that the surface density of the disc is greatly reduced in the warp. However in these high-resolution simulations we still resolve the orbits of the disc particles in the warp. This disc breaking is predicted by the numerical simulations of warp propagation in accretion discs by Nixon and King (2012) (Chapter 5) using the constrained viscosities of Ogilvie (1999), and has been also been seen in the SPH simulations of Larwood and Papaloizou (1997) and Lodato and Price (2010).

Inclinations greater than 90° cancel rather than sharing angular momentum between particles. This leads to particles falling on to smaller orbits, and so the accretion rates in these simulations are significantly higher. The accretion rate for $\theta = 180^\circ$ is lower than that of $\theta = 150^\circ$. This is probably because gas must be pushed out of the plane of the discs before it can leap over the inner disc.

The numerical simulations do not predict accretion rates as high as the analytical approach (6.4). This was based on two assumptions: that the angular momentum cancellation is perfect, and an equal mass from each disc interacts. The simulations suggest instead that the gas falling from the gap has a spread of angular momentum, where some accretes and some falls to a smaller radius.

The main result is that internally tilted accretion discs can generate enhanced accretion rates of up to $\sim 10^4$ times that of a plane disc (cf. Fig. 6.1). In a quasi-steady state (such discs never reach a full steady state) we see accretion rates that are $\gtrsim 100$ times those of a planar disc. As remarked above, discs tilted in this way arise naturally during the alignment of a disc around a spinning black hole or a central binary (cf. Lodato and Pringle 2006 & Nixon et al. 2011b; Chapter 2). We also

expect such structures to occur during chaotic gas infall into galactic centres. This is a very efficient way of driving gas right down to the vicinity of an SMBH or an SMBH binary and thus causing accretion. Nested discs suggest a way of transferring gas from large radii (approaching galaxy scales) to small radii.

We note finally that our result, that tilted and nested discs imply enhanced accretion, has implications for simulations where subgrid models are used to mimic feedback from the SMBH. The dynamics of misaligned accretion events may imply much larger accretion rates than predicted by the simple assumption of a planar disc with a fixed direction of angular momentum.

7

Conclusions

In this thesis I have presented mechanisms which may be important ingredients to understanding two astrophysical phenomena. The first is the final parsec problem, which requires angular momentum to be extracted from an SMBH binary orbit to allow the SMBH to coalesce. The second is the growth of supermassive black holes, which requires rapid accretion of gas. Here I summarise the main conclusions from each chapter.

7.1 Misaligned circumbinary discs

In Chapter 2 we explored the effect of a binary on an external gas disc. We used the dominant zero-frequency (azimuthally symmetric $m = 0$) term of the binary potential to derive the main secular effect which is simply that misaligned gas orbits in the disc will precess about the binary. This is similar to the Lense–Thirring effect from a spinning black hole on a misaligned accretion disc. Using results that were originally derived with the latter case in mind, we determined the resulting evolution. The disc must counteralign with the binary if the inclination angle between the two is such that $\cos \theta < -J_d/2J_b$. Otherwise the disc coaligns with the binary. This suggests that for randomly oriented accretion events on to an SMBH binary we can expect to form both prograde and retrograde circumbinary discs.

7.2 Stable counteralignment

In Chapter 3 I tested whether our assumptions in Chapter 2 were correct by simulating a misaligned disc around a binary. This simulation showed the precession induced in the disc by the binary and the subsequent counteralignment of the disc that was initially tilted by 170° to the binary plane. This shows that counteralign-

ment is stable and can occur for up to half of all randomly oriented accretion events.

7.3 Retrograde circumbinary discs

In Chapter 4 we explored the subsequent evolution of a retrograde circumbinary gas disc. We show that it can be considerably more effective in shrinking an SMBH binary than accretion from a corresponding prograde disc. Coalescence occurs once the binary has lost all of its angular momentum, i.e. once the secondary black hole captures a gas mass comparable to its own.

The reason for this effectiveness, in comparison to the prograde case, is twofold. First the disc has negative angular momentum, and second there are no orbital resonances in a retrograde disc. The disc cannot react to tidal torques from the binary and become a decretion disc, which is what slows the evolution in the prograde case (Lodato et al., 2009; Cuadra et al., 2009). Instead, the disc directly feeds negative orbital angular momentum to the secondary black hole. An important aspect here is that there is no restriction on the rate at which this can occur. In particular gas can flow inwards at rates higher even than Eddington for the primary black hole, without any significant effect on the amount of gas required to coalesce the binary. This happens because once the gas is captured, it has given up its (negative) angular momentum to the binary – a subsequent Eddington wind would most likely be isotropic and thus not affect the binary orbit.

We caution finally that self-gravity can deplete the circumbinary disc of gas and reduce its ability to shrink the binary (cf. Lodato et al. 2009 in the prograde case). Since coalescence by a retrograde disc requires $M_d \gtrsim M_2$, and self-gravity effects appear unless $M_d \lesssim (H/R)(M_1 + M_2)$, where H/R is the disc aspect ratio, it appears that the final parsec problem is so far alleviated only for mergers with

$M_2/M_1 \lesssim H/R$. This work does suggest how coalescence might work for larger mass ratios. The secondary hole has to absorb negative angular momentum from a gas mass comparable to its own, preferably in an eccentric orbit. Although gas self-gravity is inevitably important in such an event, a retrograde flow has the advantage that there is no limit on the mass inflow rate. One can also envisage a scenario where multiple randomly oriented accretion events interact with the binary, in this case the binary will merge when the total mass of retrograde accretion is comparable to the secondary black hole mass.

7.4 Broken discs

In Chapter 5 we looked at the effect of the full effective viscosities (Ogilvie, 1999) on the evolution of a misaligned accretion disc around a spinning black hole. With constant effective viscosities the disc is able to maintain a coherent smooth warp between the aligned inner part and the misaligned outer part. However, if we include the full effective viscosities the evolution is changed. As the full effective viscosities decrease sharply in magnitude in a strong warp, it becomes possible to break the disc into two (possibly more) distinct planes. This hints at a possible mode of accretion where a sharp tilt in the disc is able to induce counterrotation in the gas and hence subsequent cancellation of angular momentum. It also suggests that observations of accretion discs around black holes may be problematic as it is unclear how the disc surface is oriented, even if the outer disc plane is known.

7.5 Rapid accretion

In Chapter 6 we explored the evolution of a disc with a significant internal tilt. We found that discs with a tilt less than 90° have similar accretion rates. This is probably because the expected increase in the disc torques (cf. Eq. 5.3) is offset by the decrease in the effective viscosities (cf. Figs. 5.1, 5.2 and 5.3). However, if the tilt in the disc is greater than 90° the gas is able to cancel angular momentum directly, and so fall to a much smaller radius. The mass flow rates through such discs, can exceed those in the corresponding flat disc by factors $\sim 100 - 10000$. We also show that this result does not depend on the thermal treatment for the gas. If we assume the gas is isothermal it can readily fall to smaller radii as it is not heated in shocks. However, if the gas retains the shock heat, the flow becomes more chaotic, driving more mixing of the gas angular momentum and hence more accretion. The difference in accretion rate between these two extremes is only a factor of ~ 2 .

7.6 Future work

I would like to simulate a full accretion episode on to an SMBH binary, from the capture of the initial gas into a circumbinary disc, the alignment or counteralignment of the disc and then the subsequent evolution of the binary. This could be extended to include multiple accretion events with random angular momentum (both magnitude and orientation). Another intriguing possibility is the prospect of simultaneous co- and counter-alignment of the same disc. This is almost certainly likely to drive enhanced accretion and it would be useful to quantify how high the accretion rates can be, both for the merging of an SMBH binary and also for the growth of individual SMBH.

I would also like to perform simulations of tilted accretion discs around spinning black holes. Previous investigations have been limited by resolution and also focused on small warps. From the investigations in this thesis it appears obvious that in the nonlinear, large warp, regime the disc can evolve in an extraordinary way.

References

- Armitage, P. J. and Natarajan, P.: 1999, *ApJ* **525**, 909
- Armitage, P. J. and Natarajan, P.: 2005, *ApJ* **634**, 921
- Arzoumanian, Z., Chernoff, D. F., and Cordes, J. M.: 2002, *ApJ* **568**, 289
- Balbus, S. A. and Hawley, J. F.: 1991, *ApJ* **376**, 214
- Bardeen, J. M. and Petterson, J. A.: 1975, *ApJ* **195**, L65+
- Bate, M. R., Bonnell, I. A., Clarke, C. J., Lubow, S. H., Ogilvie, G. I., Pringle, J. E., and Tout, C. A.: 2000, *MNRAS* **317**, 773
- Bate, M. R., Lodato, G., and Pringle, J. E.: 2010, *MNRAS* **401**, 1505
- Begelman, M. C., Blandford, R. D., and Rees, M. J.: 1980, *Nature* **287**, 307
- Berczik, P., Merritt, D., Spurzem, R., and Bischof, H.-P.: 2006, *ApJ* **642**, L21
- Binney, J. and Tremaine, S.: 2008, *Galactic Dynamics: Second Edition*, Princeton University Press
- Cuadra, J., Armitage, P. J., Alexander, R. D., and Begelman, M. C.: 2009, *MNRAS* **393**, 1423
- Cullen, L. and Dehnen, W.: 2010, *MNRAS* **408**, 669

-
- Deegan, P.: 2009, *Ph.D. thesis*, University of Leicester, March 2009. Awarded 18 December 2009.
- Dotti, M., Colpi, M., Haardt, F., and Mayer, L.: 2007, *MNRAS* **379**, 956
- Dotti, M., Ruszkowski, M., Paredi, L., Colpi, M., Volonteri, M., and Haardt, F.: 2009, *MNRAS* **396**, 1640
- Eberle, J. and Cuntz, M.: 2010, *ApJ* **721**, L168
- Escala, A., Larson, R. B., Coppi, P. S., and Mardones, D.: 2005, *ApJ* **630**, 152
- Fanidakis, N., Baugh, C. M., Benson, A. J., Bower, R. G., Cole, S., Done, C., and Frenk, C. S.: 2011, *MNRAS* **410**, 53
- Ferrarese, L. and Merritt, D.: 2000, *ApJ* **539**, L9
- Frank, J., King, A., and Raine, D. J.: 2002, *Accretion Power in Astrophysics: Third Edition*
- Gammie, C. F., Goodman, J., and Ogilvie, G. I.: 2000, *MNRAS* **318**, 1005
- Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., Richstone, D., and Tremaine, S.: 2000, *ApJ* **539**, L13
- Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., and Duchêne, G.: 2005, *ApJ* **620**, 744
- Gingold, R. A. and Monaghan, J. J.: 1977, *MNRAS* **181**, 375
- Hansen, B. M. S. and Barman, T.: 2007, *ApJ* **671**, 861
- Häring, N. and Rix, H.-W.: 2004, *ApJ* **604**, L89

-
- Hobbs, G., Lorimer, D. R., Lyne, A. G., and Kramer, M.: 2005, *MNRAS* **360**, 974
- Hughes, S. A. and Blandford, R. D.: 2003, *ApJ* **585**, L101
- Ivanov, P. B., Papaloizou, J. C. B., and Polnarev, A. G.: 1999, *MNRAS* **307**, 79
- Iwasawa, M., Funato, Y., and Makino, J.: 2006, *ApJ* **651**, 1059
- King, A. R., Lubow, S. H., Ogilvie, G. I., and Pringle, J. E.: 2005, *MNRAS* **363**, 49
- King, A. R. and Pringle, J. E.: 2006, *MNRAS* **373**, L90
- King, A. R. and Pringle, J. E.: 2007, *MNRAS* **377**, L25
- King, A. R., Pringle, J. E., and Hofmann, J. A.: 2008, *MNRAS* **385**, 1621
- Kormendy, J. and Richstone, D.: 1995, *ARA&A* **33**, 581
- Kulkarni, A. K., Penna, R. F., Shcherbakov, R. V., Steiner, J. F., Narayan, R., Sä
Dowski, A., Zhu, Y., McClintock, J. E., Davis, S. W., and McKinney, J. C.: 2011,
MNRAS **414**, 1183
- Kumar, S. and Pringle, J. E.: 1985, *MNRAS* **213**, 435
- Larwood, J. D. and Papaloizou, J. C. B.: 1997, *MNRAS* **285**, 288
- Lense, J. and Thirring, H.: 1918, *Phys. Z.* **19**, 156
- Lodato, G., Nayakshin, S., King, A. R., and Pringle, J. E.: 2009, *MNRAS* **398**,
1392
- Lodato, G. and Price, D. J.: 2010, *MNRAS* **405**, 1212
- Lodato, G. and Pringle, J. E.: 2006, *MNRAS* **368**, 1196

-
- Lucy, L. B.: 1977, *AJ* **82**, 1013
- MacFadyen, A. I. and Milosavljević, M.: 2008, *ApJ* **672**, 83
- Martin, R. G., Pringle, J. E., and Tout, C. A.: 2009, *MNRAS* **400**, 383
- Martin, R. G., Tout, C. A., and Pringle, J. E.: 2008, *MNRAS* **387**, 188
- Mashhoon, B., Hehl, F. W., and Theiss, D. S.: 1984, *General Relativity and Gravitation* **16**, 711
- Merritt, D. and Milosavljević, M.: 2005, *Living Reviews in Relativity* **8**, 8
- Merritt, D. and Poon, M. Y.: 2004, *ApJ* **606**, 788
- Milosavljević, M. and Merritt, D.: 2001, *ApJ* **563**, 34
- Monaghan, J. J.: 1992, *ARA&A* **30**, 543
- Monaghan, J. J. and Lattanzio, J. C.: 1985, *A&A* **149**, 135
- Morris, J. P. and Monaghan, J. J.: 1997, *Journal of Computational Physics* **136**, 41
- Murray, J. R.: 1996, *MNRAS* **279**, 402
- Natarajan, P. and Armitage, P. J.: 1999, *MNRAS* **309**, 961
- Natarajan, P. and Pringle, J. E.: 1998, *ApJ* **506**, L97
- Nelson, R. P. and Papaloizou, J. C. B.: 2000, *MNRAS* **315**, 570
- Nixon, C. J.: 2012, *MNRAS* p. 3054
- Nixon, C. J., Cossins, P. J., King, A. R., and Pringle, J. E.: 2011a, *MNRAS* **412**, 1591
- Nixon, C. J. and King, A. R.: 2012, *MNRAS* **421**, 1201

-
- Nixon, C. J., King, A. R., and Price, D. J.: 2012, *MNRAS* p. 2780
- Nixon, C. J., King, A. R., and Pringle, J. E.: 2011b, *MNRAS* **417**, L66
- Ogilvie, G. I.: 1999, *MNRAS* **304**, 557
- Papaloizou, J. and Pringle, J. E.: 1977, *MNRAS* **181**, 441
- Papaloizou, J. C. B. and Pringle, J. E.: 1983, *MNRAS* **202**, 1181
- Price, D. J.: 2007, *Publ. Astron. Soc. Aust.* **24**, 159
- Price, D. J.: 2008, *Journal of Computational Physics* **227**, 10040
- Price, D. J.: 2012, *Journal of Computational Physics* **231**, 759
- Price, D. J. and Monaghan, J. J.: 2005, *MNRAS* **364**, 384
- Pringle, J. E.: 1981, *ARA&A* **19**, 137
- Pringle, J. E.: 1992, *MNRAS* **258**, 811
- Pringle, J. E.: 1996, *MNRAS* **281**, 357
- Pringle, J. E.: 1997, *MNRAS* **292**, 136
- Pringle, J. E.: 1999, in J. A. Sellwood & J. Goodman (ed.), *Astrophysical Discs - an EC Summer School*, Vol. 160 of *Astronomical Society of the Pacific Conference Series*, pp 53–+
- Rodriguez, C., Taylor, G. B., Zavala, R. T., Peck, A. B., Pollack, L. K., and Romani, R. W.: 2006, *ApJ* **646**, 49
- Rosswog, S.: 2009, *New Astron. Rev.* **53**, 78
- Safronov, V. S.: 1972, *Evolution of the protoplanetary cloud and formation of the earth and planets.*

-
- Scheuer, P. A. G. and Feiler, R.: 1996, *MNRAS* **282**, 291
- Schödel, R., Ott, T., Genzel, R., Hofmann, R., Lehnert, M., Eckart, A., Mouawad, N., Alexander, T., Reid, M. J., Lenzen, R., Hartung, M., Lacombe, F., Rouan, D., Gendron, E., Rousset, G., Lagrange, A.-M., Brandner, W., Ageorges, N., Lidman, C., Moorwood, A. F. M., Spyromilio, J., Hubin, N., and Menten, K. M.: 2002, *Nature* **419**, 694
- Shakura, N. I. and Sunyaev, R. A.: 1973, *A&A* **24**, 337
- Shklovskii, I. S.: 1970, *Soviet Ast.* **13**, 562
- Soltan, A.: 1982, *MNRAS* **200**, 115
- Springel, V.: 2010, *ARA&A* **48**, 391
- Springel, V. and Hernquist, L.: 2002, *MNRAS* **333**, 649
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F.: 2005, *Nature* **435**, 629
- Sutantyo, W.: 1978, *Ap&SS* **54**, 479
- Tanaka, Y., Nandra, K., Fabian, A. C., Inoue, H., Otani, C., Dotani, T., Hayashida, K., Iwasawa, K., Kii, T., Kunieda, H., Makino, F., and Matsuoka, M.: 1995, *Nature* **375**, 659
- Toomre, A.: 1964, *ApJ* **139**, 1217
- Wijers, R. A. M. J. and Pringle, J. E.: 1999, *MNRAS* **308**, 207
- Wilkins, D. C.: 1972, *Phys. Rev. D* **5**, 814

Willott, C. J., McLure, R. J., and Jarvis, M. J.: 2003, *ApJ* **587**, L15