ADVENTURES WITH PLANETS
& BINARIES IN ACCRETION DISCS

Alex Dunhill

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

October 1, 2013

Theoretical Astrophysics Group
Department of Physics and Astronomy
University of Leicester

Abstract

Adventures with planets
& binaries in accretion discs

by

Alex Dunhill

The primary idea behind the work in this thesis is that accretion discs interacting with astrophysical bodies, from planets to supermassive black holes (SMBHs), can strongly affect the dynamical behaviour of those bodies. While this idea is by no means new, observational and theoretical developments in recent years provide fresh motivation to consider this effect across a number of astrophysical contexts.

Of particular relevance to the work here are three observational measurements which I attempt to reconcile with theory by invoking interactions with accretion discs. Firstly, observations of giant exoplanets show that they prefer to inhabit eccentric orbits, which is unexpected given the predictions of planet formation theory. Conversely, *Kepler*'s discovery of planets with low eccentricities around moderately eccentric binaries goes against theoretical expectation that their orbits should be eccentric. In galactic centres, binary supermassive black holes are not observed despite theoretical expectations that their evolution should drive them to ~ parsec separations and leave them there.

In this thesis I use high-resolution smoothed particle hydrodynamics simulations to investigate each of these problems involving planet- and binary-disc interactions. I show that these interactions are unable to solve the problem of eccentric giant exoplanets, but that they can cause damping of circumbinary planetary eccentricity and so are able to explain *Kepler*'s circumbinary planets. I use this latter to place a limit on the surface density in which Kepler-16b in particular can have formed. I also show that a disc formed from a gas cloud moving prograde with respect to a SMBH binary will fragment to form stars sooner than a similar retrograde disc. Consequently, a retrograde disc is able to drive stronger binary evolution than is the prograde disc. Allowing that a large number of such encounters would be expected in the aftermath of the galaxy merger that formed the binary, this process may be able to solve the 'last parsec problem.'

Two of the chapters in this thesis have been published or accepted for publication in the journal *Monthly Notices of the Royal Astronomical Society*:

Chapter 2:
"A limit on eccentricity growth from global 3D simulations of disc-planet interactions"
A.C. Dunhill, R.D. Alexander & P.J. Armitage, 2013, *MNRAS*, **428**, 3072.

Chapter 3:
"The curiously circular orbit of Kepler-16b"
A.C. Dunhill & R.D. Alexander, 2013, *MNRAS*, **435**, 2328.

Acknowledgements

Contents

List of Figures

List of Tables

Dedicated to my parents Enid & Tony Dunhill,
and also to my grandparents Joy & Eddie Dunhill.

1

Introduction

Of the many diverse concepts in astrophysics, one of the most widely applicable is that of the accretion disc. These discs occur in situations where gaseous material is gravitationally bound to a central body, but has significant angular momentum and cannot simply fall on to it. If the gas is dissipative or is otherwise able to radiate its energy away angular momentum conservation results in a circular disc orbiting the central body. As the disc is able to slowly transport its own angular momentum outward it accretes on to the object at the centre – this is an accretion disc.

Astrophysical accretion discs are observed or inferred in a number of contexts. Ranging from the formation of stars and planets to the aftermath of huge galaxy mergers or Roche-lobe overflow in binary star systems, accretion discs are ubiquitous. Much of the detailed structure of an individual accretion disc is set by the environment it finds itself in, but can also be significantly affected by the presence of perturbing bodies at the centre or embedded within the disc itself. While classical accretion disc theory considers only a central body and an exterior disc, it has now been long recognised that many astrophysical phenomena can be explained by discs perturbed in this manner. In general, these problems are not analytically solvable and so numerical simulations are used to investigate how these processes evolve in a given system and under a given set of physical laws and other assumptions.

In this thesis I consider problems involving accretion discs across different astrophysical contexts, both as a vehicle for shaping planetary orbits and as a driver of supermassive black hole binary evolution. Therefore I summarise the key observational constraints upon them and their place in the wider astrophysical picture in Section 1.1. A comprehensive theoretical understanding of the physics of accretion discs is not necessary to understand much of the work described in later chapters, but a familiarity with the key ideas will be beneficial so I cover the salient derivations in Section 1.2. Similarly, in Section 1.3 I briefly discuss disc-satellite interactions, including planet gap-opening criteria. The bulk of this thesis describes numerical simulations using smoothed-particle hydrodynamics (SPH), so in Section 1.4 I derive the basic algorithms behind the technique and describe the non-standard modifications implemented in the code used for simulations here.

The later chapters are arranged as follows. In Chapter 2 I examine the effect on a giant planet's eccentricity of torques from its parent protoplanetary disc. Chapter 3 considers the situation of a young planet migrating through a circumbinary disc, motivated by recent *Kepler* observations of circumbinary planets with low eccentricities, and applied specifically to the case of Kepler-16b. In Chapter 4 I turn to supermassive black hole accretion and consider the effect of a gas cloud falling on to a binary system, and

how the initial configuration of the cloud changes the subsequent evolution. In Chapter 5 I provide conclusions from my work and consider possible future avenues for research. Finally, Appendix A contains derivations of the main fluid equations for SPH.

1.1 Contexts for accretion discs

1.1.1 Protoplanetary discs

The bulk of the scientific work presented in this thesis (Chapters 2 and 3) considers an accretion disc around a young star (or stars), composed of material left over from the initial collapse of the molecular cloud from which the star(s) formed. These discs are variously termed 'protoplanetary' and 'circumstellar', and although these terms carry different connotations when discussing the evolution of young stellar objects (YSOs), it will become clear that they are interchangeable for the stage in this evolution of interest for this thesis. Therefore unless otherwise noted, these terms are to be understood to be synonymous[1].

To see why an accretion disc might be expected to form around a young star, we can consider the smallest scales on which a giant molecular cloud (GMC) can collapse to form a star. At this scale λ, the pressure P of the cloud is just able to balance with gravity to support itself – or equivalently, at this size the sound crossing time ($t_{\rm sc} \sim \lambda/c_{\rm s}$ for sound speed $c_{\rm s}$) is comparable to the free-fall time of the cloud, $t_{\rm ff} \sim (G\rho)^{-1/2}$ (where G is the gravitational constant and ρ is the cloud's mean density). This typical length-scale for collapse, the Jeans length $\lambda_{\rm J}$, is therefore

$$\lambda_{\rm J} \simeq \frac{c_{\rm s}}{\sqrt{G\rho}}. \tag{1.1}$$

The mass associated with this length-scale can be approximated as the Jeans mass $M_{\rm J} \sim \rho\lambda_{\rm J}^3$:

$$M_{\rm J} \simeq \frac{c_{\rm s}^3}{\sqrt{G^3\rho}}. \tag{1.2}$$

If we now make some approximations about the thermal state of the gas we can estimate $\lambda_{\rm J}$ for a given $M_{\rm J}$. If the gas is isothermal, then the sound speed at a given temperature T is given by $c_{\rm s}^2 = k_{\rm B}T/\mu m_{\rm P}$, where $k_{\rm B}$ is the Boltzmann Constant, $\mu \simeq 2.3$ is the mean molecular weight of the gas (assumed to be mostly Hydrogen) and $m_{\rm p}$ is the mass

---

[1] Other important distinctions are between 'circumstellar' and 'protostellar' discs, and between 'circumstellar' and 'circumbinary' discs. These distinctions are noted later in the text.

of a proton. Taking a typical temperature for the cloud of $T \sim 10$ K, a collapsing cloud of one solar mass ($M_J = M_\odot$) has a Jeans length $\lambda_J \sim 0.1$ pc.

Observations of collapsing molecular cores using ammonia line emission as a tracer show that they typically have sizes 0.05 pc $\lesssim \lambda_{core} \lesssim 0.5$ pc (Goodman et al., 1993), which fits nicely with this back-of-the-envelope estimate. These same observations indicate that the specific angular momentum j (angular momentum per unit mass) of the clouds is in the region $10^{21} \lesssim j_{core} \lesssim 10^{22}$ g^2 cm^{-1} (Goodman et al., 1993).

As the cloud collapses, its potential will be dominated by the the mass of the densest region of its core, $M_\bullet$. In this case the outer regions will begin to take approximately Keplerian orbits at radii R. These orbits have specific angular momenta $j_K = \sqrt{GM_\bullet R}$. Equating this with the observed angular momenta of GMC cores gives $10^{16} \lesssim R \lesssim 10^{18}$ cm for $M_\bullet \sim 1 \, M_\odot$.

We can therefore see that if angular momentum is conserved through the entire collapse then the gas will not condense into an object smaller than approximately 10^3 au across. As stars are much smaller than this ($R_\odot \sim 5 \times 10^{-3}$ au) there must be some mechanism to remove the angular momentum of the gas as it collapses. Therefore we would expect an accretion disc to form around a young star as it forms.

1.1.1.1 Protostellar evolution: the formation of protoplanetary discs

It is helpful at this point to paint a broad picture of how protoplanetary discs form. This process is inescapably tied to the formation and subsequent evolution of protostellar bodies and I here give a simplified overview of the entire process. This overview is based on the simplified picture of, e.g. Shu et al. (1987), known as the Singular Isothermal Sphere. I shall gloss over many gaps in the theory such as, for example, the inability of this simple picture to account for binary stars, the divergent density profile required and the neglect of angular momentum (e.g. Pringle, 1989; Whitworth et al., 1996).

We start then with a typical GMC in the galaxy. As these structures are known to be heavily turbulent (e.g. Larson, 1981) it is no surprise that over-densities will occur. As outlined in Section 1.1.1, if these over-densities have a certain size and mass then they will begin to contract under their own self-gravity, forming first extended filamentary structures (e.g. Schneider et al., 2013; Hacar et al., 2013) and then bound clumps. These objects continue to collapse from the inside-out – that is, the central region forms a collapsed core which the next outer-most region collapses on to, followed by the next region, and so on (e.g. Tsamis et al., 2008; Klaassen et al., 2012). As the collapse con-

tinues, conservation of angular momentum spins up the central region and a protostellar[2] disc forms perpendicular to the net angular momentum vector of the collapsing material, with torques from the disc acting to prevent the protostar from spinning up so much that it breaks up (Lin et al., 2011). At this stage, it is common for a jet or outflow to form close to or along this axis (e.g. Klaassen et al., 2012). We now have a central protostar with a disc and possibly a jet, surrounded by the remainder of the original bound clump in the form of a gaseous envelope.

Over time the envelope will either be accreted on to the disc or else dispersed, leaving the 'naked' star and disc directly observable (e.g. Shu et al., 1987; Hartmann, 2009; Williams and Cieza, 2011). Although this thesis is only concerned with this stage of protostellar evolution, to which we shall return, the system itself will of course continue to evolve.

The next stage in the life of the system occurs when the gaseous circumstellar[3] disc becomes dispersed. The processes typically credited with causing this is a combination of accretion through the disc and photoevaporation by the star (e.g. Alexander et al., 2006a; Owen et al., 2010), but other processes such as magnetically-launched winds can play an important role (e.g. Königl and Salmeron, 2011). Although the bulk of the gaseous material is now removed from the system, any rocky bodies that were embedded therein will remain. This forms a planetary debris disc, where the detritus from the planet formation process is laid bare. Eventually most of this rocky material will agglomerate into planets, form analogues to the asteroid belt or else be removed from the system, leaving a young star and associated planetary system fully formed.

This simplified view of the process of star and planet formation is depicted in Figure 1.1. The time and length-scales shown in each panel are approximate and there is significant overlap between the ages objects observed at different stages.

Traditionally, YSOs such as those in panels (c) to (e) are classified according to the slope of their infrared (IR) spectral energy distribution (SED), examples of which are shown in Figure 1.2. This slope is defined as:

$$\alpha_{\mathrm{IR}} = \frac{d \log{(\lambda F_\lambda)}}{d \log{(\lambda)}} \tag{1.3}$$

where λ is the wavelength (typically measured between $2 - 25\mu$m) and F_λ is the measured

---

[2] At this stage the stellar mass is typically a fraction of its final mass, with the remainder of the mass still in the disc. The disc is therefore termed 'protostellar' as it will mostly go on to form part of the star.

[3] The star has now accreted most of its final mass and so the disc is no longer protostellar but now merely circumstellar.
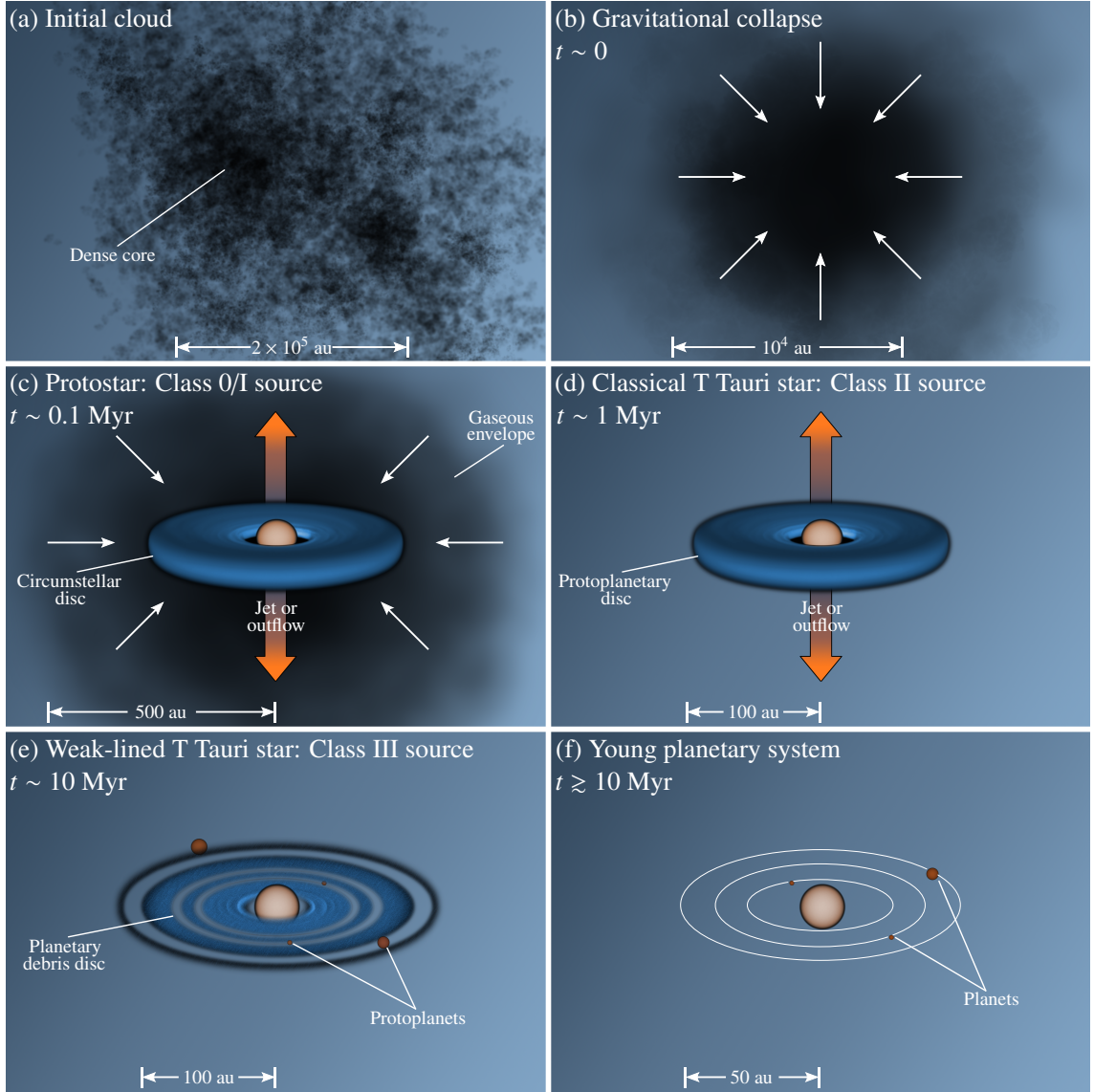
Figure 1.1: Cartoon showing the evolutionary phases of a star forming from a Giant Molecular Cloud. Panel (a) shows the formation of an initial core from a transient overdensity in a turbulent Giant Molecular Cloud. Panel (b) shows the collapse of the core. Panel (c) shows the young stellar object and its attendant disc embedded in a thick gaseous envelope. Panel (d) shows the young protostar and protoplanetary disc after the envelope has dispersed. Panel (e) shows the post-gas disc phase where protoplanets orbit in a debris disc, and panel (f) shows the final stellar system. While the work in Chapters 2 & 3 of this thesis only considers the stage depicted in panel (d), the broader context shown in this figure is useful to bear in mind. This sequence is heavily simplified and ignores many of the details of the evolution shown, and both time- and length-scales are approximate only. Figure based on that of Greene (2001), after Shu et al. (1987).

---

flux at a given wavelength λ (Williams and Cieza, 2011). According to the scheme proposed by Lada (1987), objects with $\alpha_{IR} > 0$ are termed Class I. Class II objects are those
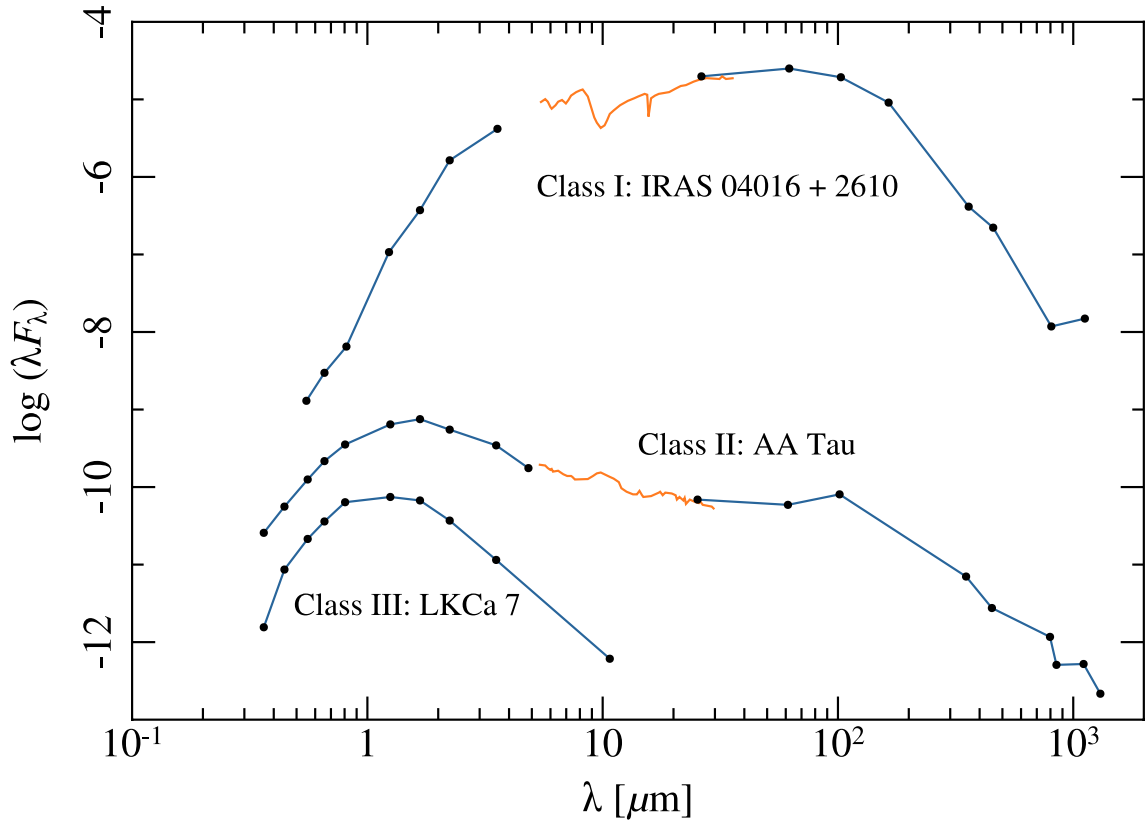
Figure 1.2: Example SEDs for Class I – III objects. The *y*-axis plots the flux at the Earth in arbitrary units multiplied by wavelength, and each object is systematically offset in the *y* direction. The *x*-axis plots wavelength in μm. It is clear that as the object class progresses from I to III the extent of the IR excess beyond ∼ 10 μm becomes less extreme, with Class III well approximated by a black-body spectrum. Points and blue lines are data taken from Kenyon and Hartmann (1995), while orange lines are *Spitzer* spectra from Furlan et al. (2006). Figure adapted from Hartmann (2009).

_____

with $\alpha_{\mathrm{IR}} \lesssim 0$, and Class III objects are well approximated by a blackbody with typical stellar effective temperatures. A further category, Class 0, was added to this scheme by André et al. (1993). This comprises those pre-stellar objects invisible at IR wavelengths longer than 10 μm but which are visible further into the mm/sub-mm.

An alternative classification for certain protostellar bodies are the two catagories of T Tauri stars (TTs), named after the prototypical T Tau, which rely on the optical and X-ray properties of the objects. Classical T Tauri stars (CTTs) are young, solar-type stars with strong emission lines such as Hα (e.g. Bertout, 1989; Barrado y Navascués and Martín, 2003) and are also strong X-ray sources, with an ultraviolet (UV) excess not seen in typical stellar SEDs. A second catagory of TTs, known as Weak-lined T Tauris (WTTs) have similar stellar properties and strong X-ray emission (e.g. Feigelson and Montmerle,

1999; Glassgold et al., 2000), but lack much of the strength in emission lines and the UV excess characteristic of the former class (e.g. Walter et al., 1988; Williams and Cieza, 2011).

A physical interpretation of TTs is that CTTs are protostars still undergoing accretion from a circumstellar disc, which is the source of the emission lines and the UV excess – allowing direct measurements of the accretion rates onto the star (and by extension, through the disc; e.g. Calvet and Gullbring, 1998; Gullbring et al., 1998, 2000). WTTs lack these spectral characteristics as they have lost their discs and therefore see do not see significant accretion (Walter et al., 1988). The strong X-ray signature common to both types of TTs is a signature of their young ages, as they have very strong magnetic fields.

We can therefore see that CTTs and WTTs are approximately equated with Class II and Class III protostars respectively, as indicated in panels (d) and (e) of Figure 1.1, although there is some overlap in ages and the correspondence is not always exact. Class 0/I objects do not have corresponding T Tauri types as they are not usually visible in the optical.

1.1.1.2   Observational properties of protoplanetary discs

As this thesis deals primarily with theoretical discs corresponding to Class II/CTT objects, I shall summarise the main observational constraints on these discs here. Figure 1.3 shows a schematic of a protoplanetary disc along with the main emission mechanisms, and the region of the disc they originate from.

As noted in Section 1.1.1.1, protostellar objects bearing protoplanetary discs have three distinct observational properties, with well-established physical origins:

- There is a strong excess above a stellar black-body in the IR and into the sub-mm. Stellar light is absorbed by dust grains in the disc and re-emitted at longer wavelengths.

- They have a stellar optical spectrum with strong emission lines (e.g. Hα) and a large UV excess above the stellar blackbody. These lines are directly from the accretion columns of disc material moving along magnetic field lines (see Figure 1.3), which shock as they hits the stellar surface producing the UV excess.

- They are strong X-ray sources, due to the strong magnetic field young stars possess. This results in very active coronae and chromospheres, which emit strongly in the X-ray.
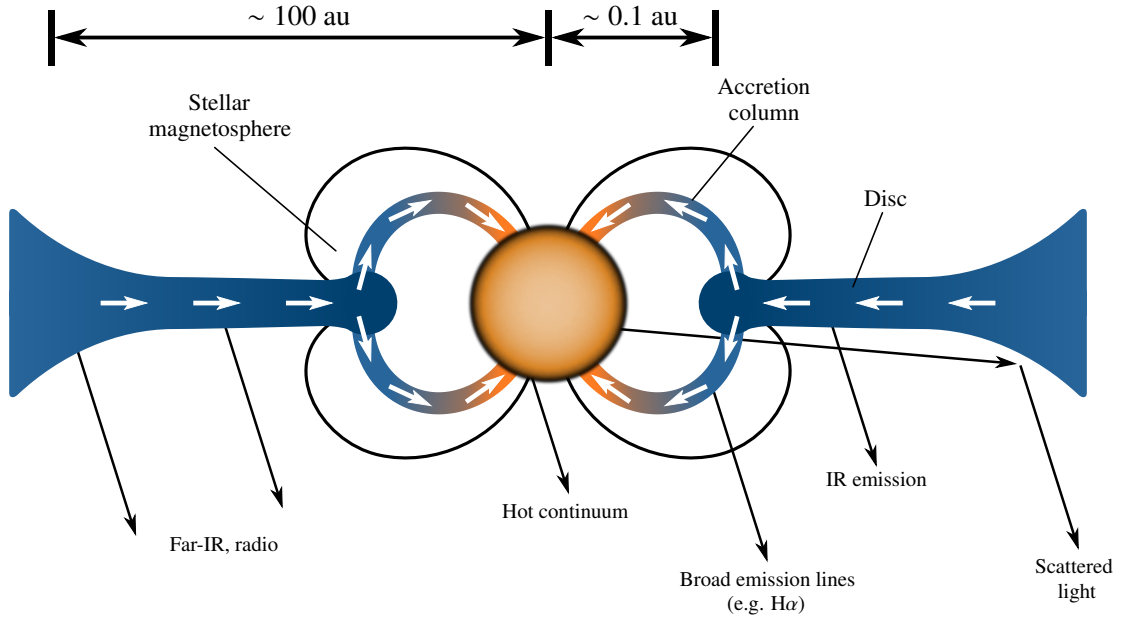
Figure 1.3: Schematic diagram showing a CTT and its disc (not to scale). Labelled are the major physical emission mechanisms and the approximate region of the disc they originate from. Disc material is funnelled out of the plane of the disc along magnetic field lines as it approaches a few stellar radii, effectively truncating the disc at this radius. The funnelled material forms accretion columns which emit heavily as they hit the stellar surface in an accretion shock. Figure adapted from Hartmann (2009).

Although individual properties can vary strongly between CTTs due to variations in intrinsic mass, age, and environment among other factors, statistical samples of observations combined with detailed models are a useful tool for divining the evolution of protoplanetary discs.

A simple deduction about the lifetimes of protoplanetary discs can be made from observations of nearby star forming regions such as Taurus. By assuming that stellar ages are approximately uniform in any given region (i.e. that the duration over which the stars formed is much shorter than the time since the last star formed) and simply counting those objects with an IR excess indicative of an accretion disc, a rough estimate of disc lifetime can be made from the decrease of disc fraction with age. This method yields a typical disc lifetime of $t \sim 5 - 10$ Myr (e.g. Haisch et al., 2001), although given that cluster ages are not certain to within a factor of three or so (Hillenbrand, 2009) these are approximate at best.

There is some evidence that disc lifetimes are systematically different around binary stars than around their solitary counterparts. Kraus et al. (2012) found that although

wide binaries (with projected separations greater than ~ 40 au) follow the same age-disc frequency trend as single stars, closer binaries seem to show a different trend. They found that these close binaries are far more likely to lose their disc early in their lifetimes (disc frequencies are as low as a third at $1 - 2$ Myr; a similar statistic was found by Harris et al., 2012), but those which do survive are likely to live for as long as $5 - 10$ Myr. Alexander (2012) used 1D models of disc evolution and photoevaporation (similar to that described in Section 1.4.2) around binary systems and indeed found that they live longer than discs around a single star. The low survival fraction is therefore likely to be due to tidal effects of the binary on the disc as it forms.

Although disc lifetimes are long, there is strong evidence that they end abruptly. Models suggest that the transition between CTT and WTT (or equivalently between Class II and Class III) occurs when the accretion rate in the disc drops below some critical level compared to the rate at which the star is able to photoevaporate the surface layers by high-energy photons (either X-rays or EUV; Alexander et al., 2006a,b; Owen et al., 2010). Observations of the small number of discs observed to be undergoing this inside-out clearing suggest that the process takes $\lesssim 0.5$ Myr (e.g. Williams and Cieza, 2011), and this is well matched by the models.

By the time they become visible in the optical, typical T Tauri stars have masses between 0.1 and 2 $M_\odot$. Typical disc masses are of the order $M_{\rm d} \sim 0.01 M_\star$, from sub-mm fluxes (e.g. Andrews and Williams, 2005; Andrews et al., 2010, 2013). However, the disc masses evolve significantly over the lifetime of the disc as it accretes on to the star. Typical accretion rates are of the order $\dot{M} \sim 10^{-8} M_\odot {\rm yr}^{-1}$ (e.g. Hartmann et al., 1998), although the scatter covers several orders of magnitude.

Again, there is tentative evidence that circumbinary discs differ from others, in this case being less massive than their circumstellar counterpars at wide binary seperations. According to a sample of visual binaries identified in Taurus-Auriga by Kraus et al. (2011, 2012) typical discs around wide (semimajor axis $a_{\rm b} \gtrsim 10$ au) binaries have $M_{\rm d} \leqslant 10^{-3} M_\odot$. For close binaries ($a_{\rm b} \ll 1$ au) the masses seem more typical of a standard protoplanetary disc (e.g. Rosenfeld et al., 2012). In contrast, Harris et al. (2012) found that circumbinary discs around close binaries are systematically brighter than other discs at mm wavelengths, and this may be due to the discs being more massive. These discrepancies may be related to the clearing of an inner disc cavity by the binary and how this relates to the orbital properties of the binary, but this is unclear. It is also expected that a circumbinary accretion disc will have a steeper radial surface density profile than a standard disc in the inner regions (e.g. Pringle, 1991), but observations are again not sensitive to this region.

Unfortunately, when dealing with accretion disc theory we are often interested in the disc surface density as a function of radius rather than simply the total disc mass (see Section 1.2). The same mm/sub-mm observations can give some information about the surface density in the outer regions (beyond a few 10s of au; e.g. Andrews et al., 2010, 2011), and while the resolution here is currently of the order of 10 au, *ALMA* and other future telescopes may be able to increase this to ~ au scales (e.g. Wolf and D'Angelo, 2005). At smaller radii, the surface density profiles of protoplanetary discs are essentially unconstrained. While some theoretical limits can be made these are only loosely based in observation and I postpone discussing them until Section 1.2.3.

A key insight into the form of these discs was found by both Kenyon and Hartmann (1987) & Adams et al. (1987) who modelled how much reprocessing of stellar light can occur for a given disc geometry and fitted the results to observed SEDs. This is an important parameter in disc models as the disc geometry is set by the temperature of the gas, which in turn controls the disc viscosity in standard model accretion discs (see Section 1.2). Kenyon and Hartmann (1987) found that most T Tauri discs in their sample can be well fitted by a flaring disc model at large radii – that is the disc thickness is an increasing function of radius. Models such and these and later models by Chiang and Goldreich (1997) provide excellent fits to the observations in most cases.

There is also some evidence that the dust in protoplanetary discs is strongly vertically stratified – larger grains sink towards the disc midplane, depleting the surface layers of dust that would otherwise emit in the mid-IR (D'Alessio et al., 2006; Furlan et al., 2006). Further details of the vertical structure of discs (e.g. the presence of 'dead zones' close to the disc midplane) are essentially unconstrained by observations and I will not discuss them in any great detail. I instead refer the reader to the recent comprehensive review by Armitage (2011).

1.1.2   Supermassive black hole accretion

While the evidence for circumstellar accretion discs is abundant and direct, the opposite is the case for accretion discs around Supermassive Black Holes (SMBHs). These are objects of many millions of solar masses which exist at the centre of most large galaxies. The most direct evidence for their existence comes from observations following stellar orbits around the central object of our own galaxy, Sgr A* (e.g. Ghez et al., 2008). These point to a central mass of approximately $M_{BH} \sim 4 \times 10^6 \, M_\odot$, and the only explanation for a mass so large and yet unseen lying in that small a volume is a SMBH. There is now wide agreement among astronomers that these objects are common and are found in most

galactic centres, with evidence coming from sources such as stellar kinematics (e.g. van den Bosch et al., 2012).

The existence of SMBH accretion discs is inferred from the unified model of Active Galactic Nuclei (AGN). In this paradigm, the great variety of powerful astronomical activity seen originating from the centres of galaxies are all different manifestations of an accretion event on to the galaxy's SMBH seen from different angles and obscured by different media (e.g. Antonucci, 1993; Urry and Padovani, 1995). In some rare cases the accretion discs themselves are observable using maser emission to trace keplerian orbits on sub-parsec scales (e.g. Greene et al., 2013). The presence of AGN radio jets is another indication of the existence of an accretion disc, as one possibility is that they are launched by the disc (e.g. Blandford and Payne, 1982; Livio et al., 1999).

While black holes are famously unobservable directly, they are actually among the brightest objects in the universe. The removal of gravitational energy from a parcel of gas as it accretes inwards from the innermost stable orbit (ISCO) is very efficient as it will fall almost directly on to the black hole, and this energy is radiated away as X-rays. This process is very well understood from decades of observation and theory of accretion on to stellar-mass black holes (e.g. Frank et al., 2002), and in the context of SMBHs the amount of gravitational potential energy available to infalling gas is orders of magnitude greater.

Therefore the observation of powerful luminous activity observed in the central regions of galaxies, independently known to harbour SMBHs, is easily attributed to an accretion event. Estimates of the lifetimes of these outbursts are of the order $\sim 10 - 100$ Myrs (e.g. Bird et al., 2008), and an accretion disc is the only viable way to continue feeding the accretion over this period (see Section 1.2.2 for approximations of the viscous time-scales in these discs).

1.1.2.1 The last parsec problem

The question of how such incredibly massive bodies can grow has been mostly answered. Large galaxies are known to form and grow via hierarchical mergers (as first suggested by Searle and Zinn, 1978), and these mergers provide the means for growing 'seed' black holes to such extreme masses (e.g. Mihos and Hernquist, 1994; King and Pringle, 2007), as they allow both the feeding of SMBHs by new gas from the merger in addition to the direct merging of SMBHs from each of the merging galaxies.

However, there are a few kinks in this simple picture. Firstly, such a process should leave a population of binary SMBHs in the process of coalescing, but these are not observed. It is well established that dynamical friction and stellar encounters are capable of

driving a SMBH from a merging galaxy to ~ pc distances from the centre of the galaxy it is merging with, but at this distance the process stalls (Begelman et al., 1980). Once the binary reaches a separation of ~ 10^{-2} pc then gravitational wave radiation is extremely efficient at removing energy from the system and allows it to coalesce within ~ 10 Myr (e.g. Lodato et al., 2009), but beyond these separations the process is very inefficient. This is therefore known as the 'last parsec problem' – how do binary SMBHs cross this final parsec after stellar interactions cease to have a strong effect before coalescence can be driven by gravitational waves?

The problem is only exacerbated by evidence from observations of high-redshift ($z = 6 - 7$) SMBHs with masses of the order 10^{8-9} $M_\odot$ (e.g. Mortlock et al., 2011; Willott et al., 2013), indicating that nature's solution to the final parsec problem is capable of driving SMBH growth to huge masses on very short time-scales – at these redshifts, the universe is approximately a Gyr old.

A potential solution to the problem is that an accretion disc may be able to influence the dynamics of the system sufficiently to allow coalescence. However, it has been shown that a standard gas disc requires a mass on the order of that of the SMBHs themselves in order to achieve this on the required time-scales (e.g. Lodato et al., 2009). Despite this it is still possible that accretion discs play a vital role in the process. Recent progress in this area concerning mis- or counter-aligned accretion discs show that they are a promising and capable mechanism for affecting the orbital parameters of the binary, as they are are able to exchange angular momentum with the SMBHs with greater ease than can a standard pro-grade accretion disc (e.g. Nixon et al., 2011a,b, 2012). This is because retrograde accretion discs are not affected by the resonant interactions with the binary described in Section 1.3. Accretion discs are also able to drive eccentricity growth in the binary (Cuadra et al., 2009), and this can hasten the point at which gravitational radiation begins to extract significant energy from the system – the key parameter of the orbit for this is the periastron distance rather than mean separation or semimajor axis, so an eccentric orbit will lose more energy in this way than a circular one.

Under the chaotic accretion paradigm of SMBH evolution, SMBHs grow primarily from a large number of small, randomly orientated accretion events (King and Pringle, 2006). This is required because consideration of black hole spin is important for the maximum accretion rate, and a low spin is needed to explain high-redshift SMBHs. Low-spin black holes accrete more efficiently than high-spin holes, and the short time-scales allowed by observations for early SMBHs to have formed in indicate that the accretion must indeed have been very efficient. An approximately isotropic distribution of a large

number of small accretion events is able to keep the spin relatively low compared to a small number of large events (King and Pringle, 2006).

If an infalling gas cloud is closer to a retrograde orbit than a prograde orbit, then the resulting disc will counter-align with to the black hole spin on some time-scale[4] (e.g. King et al., 2005). This means that the gas accreted by the hole will have the opposite sense of angular momentum than the black hole, allowing its spin to decrease[5].

Chaotic accretion is therefore an attractive potential solution to the last parsec problem, as it allows us to invoke misaligned accretion discs that may be able to help coalesce binary SMBHs from ~ pc separations down to scales where gravitation waves take over the process and allow merger. I shall investigate such an accretion event on to an SMBH binary in Chapter 4.

1.2   Accretion disc theory

Much of the behaviour and evolution of accretion discs is described primarily by one equation, known as the 1D viscous diffusion equation. It is a simple but powerful description of a disc that is assumed to be thin (that is, the disc thickness H is much less than the radius R) and axisymmetric.

Consider a vertically infinitesimal fluid element of mass dm orbiting a a central mass M at radius R with orbital frequency Ω in 2D polar coordinates R and ϕ, as shown in Figure 1.4. The element has specific angular momentum $j = dL/dm = R^2\Omega$. Allowing that the element has radial extent ΔR and subtends an angle $d\phi$ then it has surface density $\Sigma = dm/(R\Delta R d\phi)$. The mass m and angular momentum L of a full annulus of such elements are given by integrating

$$m = \int_0^{2\pi} dm = \int_0^{2\pi} R\Delta R\Sigma \, d\phi$$
$$= 2\pi R\Delta R\Sigma \tag{1.4}$$

and

$$L = \int_0^{2\pi} dL = \int_0^{2\pi} R^2\Omega \, dm = \int_0^{2\pi} R^2\Omega R\Delta R\Sigma \, d\phi$$
$$= 2\pi R\Delta R\Sigma R^2\Omega \tag{1.5}$$

---

[4] More correctly, which component aligns with which depends on the ratio of angular momenta – but the important point is that the components end in an aligned configuration.

[5] Note that this refers to an accretion event on to a *single* SMBH – the connection to binaries is simply that chaotic accretion is independently expected from theory.
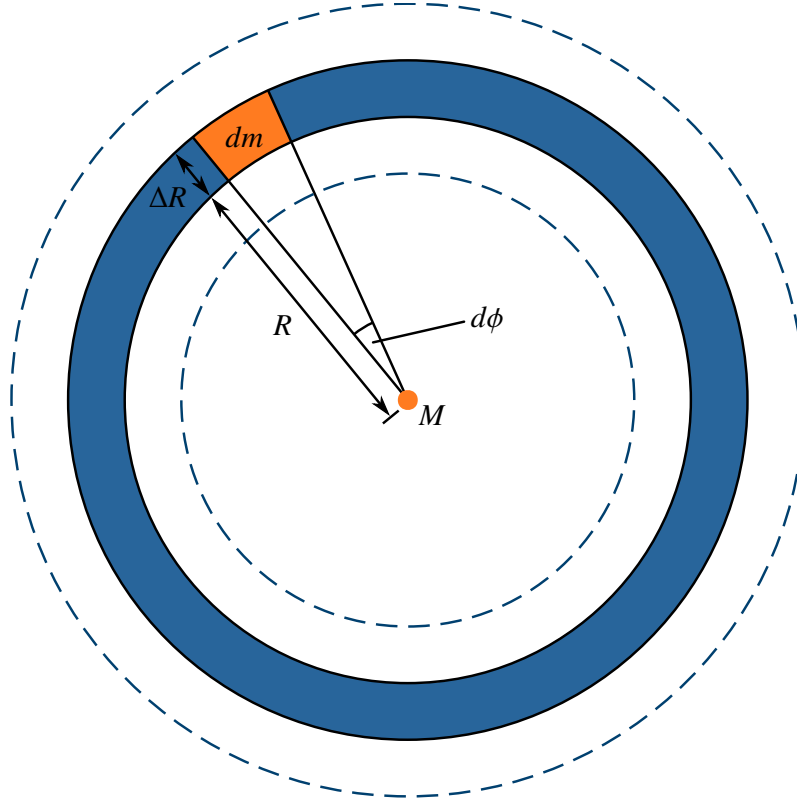
Figure 1.4: Schematic diagram showing the situation from which the 1D viscous diffusion equation for an accretion disc (Equation 1.14) is derived. A fluid element of mass dm (orange element) orbits a central mass M (solid orange circle) at radius R. The element has radial extent ΔR and subtends angle $d\phi$. Integrating with respect to ϕ yields an annulus of such elements (blue annulus). Dashed lines show the adjacent exterior and interior annuli. Viscous torques between annuli cause them to repel each other and allow them to exchange angular momentum. This is the essence of the 1D viscous accretion disc.

_____

Constructing a disc of such annuli with the possibility of mass flow between them allows us to define a radial drift velocity between annuli, u_R. Across the boundary between an annulus at R and its neighbour at $R + \Delta R$, the rate of mass flow across annuli is given by

$$\frac{\partial}{\partial t}(2\pi R \Delta R \Sigma) = (u_R 2\pi R \Sigma)_{(R)} - (u_R 2\pi R \Sigma)_{(R+\Delta R)}$$

$$\simeq -2\pi \Delta R \frac{\partial}{\partial R}(R\Sigma u_R). \tag{1.6}$$

Therefore in the limit of $\Delta R \rightarrow 0$, we obtain the equation of mass conservation

$$R\frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial R}(R\Sigma u_R) = 0. \tag{1.7}$$

When considering the equivalent flow of angular momentum, the effect of viscous torques Γ_ν between annuli must be accounted for. Again taking annuli at R and $R + \Delta R$,

$$\frac{\partial}{\partial t}(2\pi R\Delta R\Sigma R^2\Omega) = (u_R 2\pi R\Sigma R^2\Omega)_{(R)} - (u_R 2\pi R\Sigma R^2\Omega)_{(R+\Delta R)} + \Gamma_{\nu(R+\Delta R)} - \Gamma_{\nu(R)}$$

$$\simeq -2\pi\Delta R\frac{\partial}{\partial R}(R\Sigma u_R R^2\Omega) + \Delta R\frac{\partial\Gamma_\nu}{\partial R} \tag{1.8}$$

and again in the limit of $\Delta R \to 0$ this becomes

$$R\frac{\partial}{\partial t}(\Sigma R^2\Omega) + \frac{\partial}{\partial R}(R\Sigma u_R R^2\Omega) = \frac{1}{2\pi}\frac{\partial\Gamma_\nu}{\partial R}. \tag{1.9}$$

The internal disc torque Γ_ν is the result of viscous stresses between annuli. The viscous force F_ν generated per unit length around the circumference of an annulus is related to the rate of shear, $R(\partial\Omega/\partial R)$ by

$$F_\nu = \nu\Sigma R\frac{\partial\Omega}{\partial R} \tag{1.10}$$

where ν is the kinematic viscosity, which is related to the shear viscosity η via the (volume) density ρ by $\eta = \rho\nu$. The total viscous torque between two annuli is then

$$\Gamma_\nu = 2\pi R\nu\Sigma R^2\frac{\partial\Omega}{\partial R}. \tag{1.11}$$

Multiplying Equation 1.7 by $R^2\Omega$ and expanding Equation 1.9 with the product rule gives

$$R\Sigma u_R\frac{\partial}{\partial R}(R^2\Omega) = \frac{1}{2\pi}\frac{\partial\Gamma_\nu}{\partial R} \tag{1.12}$$

if we assume that $\partial\Omega/\partial t = 0$ (i.e. the central mass M is fixed in space and time). Combining Equations 1.7, 1.11 and 1.12 yields

$$R\frac{\partial\Sigma}{\partial t} = -\frac{\partial}{\partial R}\left[\frac{1}{2\pi\frac{\partial}{\partial R}(R^2\Omega)}\frac{\partial}{\partial R}\left(2\pi R\nu\Sigma R^2\frac{\partial\Omega}{\partial R}\right)\right]. \tag{1.13}$$

For a Keplerian orbit, $\Omega = \sqrt{GM/R^3}$ so $\partial\Omega/\partial R = -3/2\sqrt{GM/R^5}$ and $\partial(R^2\Omega)/\partial R = 1/2\sqrt{GM/R}$, so Equation 1.13 reduces to

$$\frac{\partial\Sigma}{\partial t} = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}(\nu\Sigma R^{1/2})\right]. \tag{1.14}$$

This is the 1D viscous diffusion equation for accretion discs. Practical use of this equation requires a prescription for the viscosity ν in the disc, and this is discussed in

Section 1.2.2.

Although we have hereto assumed that the disc is vertically infinitesimal, this treatment can be expanded to the vertical structure of the disc if we assume that the disc is thin, with $R \gg z$. Assuming hydrostatic equilibrium there is no flow in the vertical equilibrium, so vertical gravity must be balanced by the pressure P in the disc. We therefore have

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -\frac{GM}{R^2 + z^2}\frac{z}{\sqrt{R^2 + z^2}}. \tag{1.15}$$

As $R \gg z$, $\sqrt{R^2 + z^2} = R$ and $R^2 + z^2 = R^2$ to first order in z/R, so this becomes

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -\frac{GMz}{R^3}, \qquad R \gg z. \tag{1.16}$$

Taking an isothermal equation of state for the gas, the sound speed $c_{\mathrm{s}} = \sqrt{P/\rho}$. Equation 1.16 then becomes

$$\frac{c_{\mathrm{s}}^2}{\rho}\frac{\partial \rho}{\partial z} = c_{\mathrm{s}}^2\frac{\partial \ln \rho}{\partial z} = -\frac{GMz}{R^3}. \tag{1.17}$$

Rearranging and integrating this gives

$$\rho(z) = \rho_0 \exp\left(-\frac{GMz^2}{2c_{\mathrm{s}}^2 R^3}\right) = \rho_0 \exp\left(-\frac{z^2\Omega^2}{2c_{\mathrm{s}}^2}\right) \tag{1.18}$$

for some midplane density ρ_0 at $z = 0$. Defining the disc scale height $H = c_{\mathrm{s}}/\Omega$ this reduces to

$$\rho(z) = \rho_0 \exp\left(-\frac{z^2}{2H^2}\right). \tag{1.19}$$

An isothermal disc therefore has a Gaussian vertical density structure. The introduction of H allows us to see from the disc aspect ratio $H/R = c_{\mathrm{s}}/u_{\mathrm{K}}$ (where $u_{\mathrm{K}} = \Omega R$ is the Keplerian orbital velocity) that requiring the disc to be thin ($H/R \ll 1$) is equivalently a statement that the flow must be very supersonic (i.e. $c_{\mathrm{s}} \ll u_{\mathrm{K}}$).

1.2.1   Azimuthal velocity profile

Although the disc has so far been assumed to follow Keplerian orbits, this is not necessarily going to be true for a given set of disc parameters. A more accurate expression for the orbital velocity u_ϕ can be found by considering that the disc will reach hydrostatic

equilibrium radially as well as vertically:

$$\frac{u_\phi^2}{R} = \frac{1}{\rho}\frac{\partial P}{\partial R} + \frac{GM}{R^2}.$$ (1.20)

For a power-law radial pressure profile of the form $P = P_0(R/R_0)^{-\beta}$ with a normalisation factor $P_0 = \rho_0 c_s$ at $R = R_0$, this becomes

$$u_\phi = u_K\left[1 - \beta\frac{c_s^2}{u_K^2}\right]$$ (1.21)

where u_K is the Keplerian orbital velocity. $\beta \sim 3$ is typical for circumstellar accretion discs (e.g. Armitage, 2010), and in that case the disc orbits with only slightly sub-Keplerian velocities, $u_\phi \simeq 0.996\,u_K$. The more pressure-dominated a disc is, the slower it rotates as the pressure is able to provide the support lost by lowering the rotation rate.

1.2.2 Disc viscosity: sources of angular momentum transport

From Equation 1.11, we can see that the viscous torque between two annuli is proportional to the radial gradient in Ω. For a Keplerian disc (or any shearing disc where Ω is an decreasing function of R) the direction of angular momentum transport due to viscosity will be outward.

However, the exact form of the viscous torque is an important parameter in modelling an accretion disc, and without it we are limited to basic statements about the direction of the flow (e.g. mass will flow inwards, while angular momentum will flow outwards). The simplest treatment is to assume a constant ν across all R, t (see Section 1.4.2 for an example of this treatment), but this is self-evidently unphysical.

A more physically-motivated model would be to postulate that the disc viscosity is simply a molecular viscosity. This can be disproved by simple order-of-magnitude estimates. The typical time-scale for accretion is given by the viscous time-scale

$$t_\nu \sim \frac{R^2}{\nu}$$ (1.22)

For a typical viscous fluid, the dynamical viscosity is of the order $\eta \sim 10^{-2}$ kg m^{-1} s^{-1}. Taking a typical disc density for a protoplanetary disc to be $\rho \sim M_d/R^2H$ with $M_d \sim 0.1M_\odot$, $R \sim 100$ au and $H \sim 0.1\,R$, this gives a viscous time-scale of $t_\nu \sim 10^9$ yr at 1 au. This is orders of magnitude longer than even the longest estimated disc lifetimes

(see Section 1.1.1.2), so modelling angular momentum transport as a molecular viscosity does not work.

It is thought that instead the angular momentum in a disc can be transported via turbulence of some form. Although this form is extremely uncertain, in practice it is possible to paramaterize ν in such a way that it is not necessary to know the exact nature of its source, as shown by Shakura and Sunyaev (1973). They argued that the strength of the turbulence is limited by two properties: the sound speed c_s and the scale height H. If we accept that the turbulent eddies are somewhat subsonic and are limited to sizes smaller than H, then

$$\nu = \alpha \, c_s H = \alpha \, H^2 \Omega. \tag{1.23}$$

This equation usefully sidesteps our ignorance of the turbulence itself. Further, it allows the viscosity to depend only on local rather than global disc conditions, as c_s and H are both functions of temperature.

The exact source of the viscosity in an accretion disc is not fully understood, but is most often attributed to the Magneto-rotational instability (MRI; Balbus and Hawley, 1991, 1998). In this magnetohydrodynamic instability, elements of ionized disc material orbiting a central mass M in a weak magnetic field behave as if attached by springs (see Figure 1.5; Balbus, 2011). The innermost element m_i orbits faster than the outermost m_o, stretching the spring. This pulls the innermost mass back, and the outermost forwards. In a rotating frame, this removal of angular momentum from m_i and addition of the same to m_o causes the former to sink closer to the centre and the latter to move outwards. This is an unstable situation as the problem has only been made worse, setting up a runaway instability. In the framework of the MRI, the 'springs' are magnetic field lines which become stretched before breaking into turbulent regions. The net angular momentum is conserved in the linear phase before the fluid becomes turbulent, but the onset of turbulence and the mixing it causes allows angular momentum to be transported non-conservatively. This mixing of fluid elements in the turbulence then allows the process to start again.

Magneto-hydrodynamic (MHD) simulations show that MRI turbulence is capable of driving turbulence with an effective Shakura-Sunyaev αs in the range $10^{-3} \lesssim \alpha \lesssim 10^{-1}$ (e.g. Hawley et al., 1995; Stone et al., 1996; Hirose et al., 2009). Observational evidence, while weak, indicates that the true values of α are of the order 10^{-1} for black hole accretion discs (King et al., 2007), and 10^{-2} for protoplanetary discs (Hartmann et al., 1998). Using these typical values for the Shakura-Sunyaev parameterisation of the disc viscosity, viscous time-scales for circumstellar accretion discs and SMBH accretion discs can be estimated from Equation 1.22. For a circumstellar disc with a radius of 100 au and
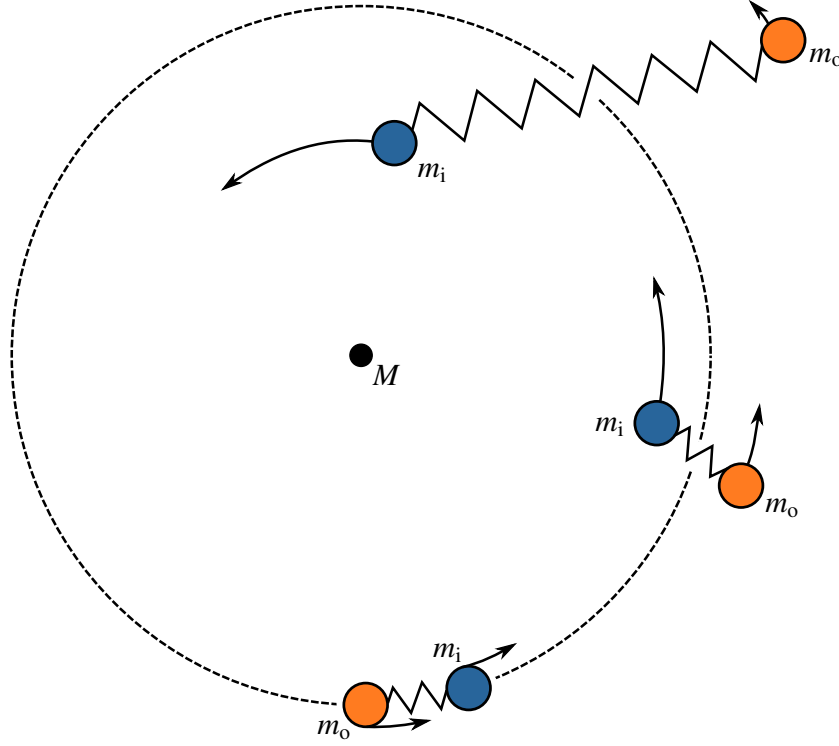
Figure 1.5: Cartoon diagram of the driving of the magneto-rotational instability (MRI). A magnetic field in the disc causes ionized fluid elements to become bound as if by a spring. Here two fluid elements orbit a central mass M. The innermost element m_i orbits faster than the outermost m_o, and is pulled backward by the 'spring', while m_o is pulled forward. This causes m_i to lose angular momentum, which is gained by m_o. Although the force would seem to be causing the elements to move closer, this change in angular momentum causes m_i to orbit closer to M while m_o moves outwards. This exacerbates the problem, driving a runaway instability. Figure adapted from Balbus (2011).

$H/R = 0.05$ around a solar mass star, $t_\nu \sim 6$ Myr. For a SMBH accretion disc with a 1 pc radius and $H/R = 0.01$ around a $10^7 \, M_\odot$ black hole $t_\nu \sim 500$ Myr. These values are approximately consistent with the expected lifetimes of the discs.

In the circumstellar case, the use of MRI to drive turbulence may be problematic. The invocation of an ionized disc is only likely to be a justified assumption in the surface layers of the disc, which are heated mainly by X-rays from the central star (and also possibly from nearby young stars; Glassgold et al., 2000) but also by cosmic rays. Deeper towards the midplane, Gammie (1996) suggested that there may exist so-called MRI 'dead-zones' which is heavily shielded enough that the ionization fraction is too low for the MRI to act effectively. This is an area of much active research (e.g. Salmeron and Wardle, 2008; Gressel et al., 2012; Martin et al., 2013) but is beyond the scope of this thesis: suffice it to bear in mind that values of α are extremely uncertain!

1.2.3   Disc surface density

A popular model used to describe the likely surface density profile for the inner regions of a protoplanetary disc is the Minimum-Mass Solar Nebula (MMSN; Weidenschilling, 1977; Hayashi, 1981). It is constructed simply by assuming that the mass which forms the planets in our solar system was originally smeared out into rings around their current orbits, and assuming that the gas disc originally had solar composition. Although there are a myriad assumptions convolved into this simple model and the precise values vary among the many different formulations, a commonly adopted form is that of Weidenschilling (1977), given by

$$\Sigma(R) \simeq 5 \times 10^3 \, \text{g cm}^{-2} \left(\frac{R}{1 \, \text{au}} \right)^{-3/2} . \tag{1.24}$$

Assuming a disc that extends from 0.1 au to 100 au, this gives a disc mass $M_\text{d} \sim 0.02 M_\star$. Although this gives a disc mass consistent with observational measurements by Andrews et al. (2010) of $M_\text{d} \sim 0.01 M_\star$, the radial power-law slope is somewhat steeper than measured by observations. However, it should be noted that these observations probe different regions of the disc ($\gtrsim 50$ au in most cases) than the radii from which the MMSN is constructed from (i.e. the scale of the Solar System, ~ 40 au).

Another theoretical constraint on the surface density of a protoplanetary disc can be made from measured accretion rates, and an assumption that accretion through the disc has reached a steady-state. Taking the conservation Equations 1.7 and 1.9, the steady state assumption can be applied by taking $\partial/\partial t = 0$. Thus Equation 1.7 gives

$$R\Sigma u_R = \text{constant}. \tag{1.25}$$

As the disc is accreting inwards, $u_R < 0$ must be true, and so the flow of mass through the annulus at R is

$$\dot{M} = -2\pi R \Sigma u_R. \tag{1.26}$$

By the same process, Equation 1.9 yields

$$R\Sigma u_R R^2 \Omega = \frac{\Gamma_\nu}{2\pi} + \frac{C}{2\pi} \tag{1.27}$$

where C is some constant, relating to the rate at which angular momentum is exchanged

between the star and the inner edge of the disc. Using Equation 1.11[6], this gives

$$C = 2\pi R^3 \left[\Sigma u_R \Omega - \nu\Sigma \frac{\partial\Omega}{\partial R} \right].$$

(1.28)

A simple requirement for the star to remain intact is that its rotation rate must be less than Keplerian, or else its outer layer would become unbound. Therefore we can say that at the surface of the star $R_\star$, $\Omega < \Omega_K(R_\star)$. There must therefore be some radial point at which Ω decouples from the Keplerian profile and begins to decrease with increasing R. Denoting the thickness of this region as b, it can be stated that at $R = R_\star + b$, $\partial\Omega(R_\star + b)/\partial R = 0$. Assuming that the $b \ll R_\star$, we have

$$\Omega(R_\star + b) = \sqrt{\frac{GM_\star}{R_\star^3}} \left[1 + O\left(\frac{b}{R_\star}\right) \right]$$

(1.29)

where $O(b/R_\star)$ simply denotes terms of order $b/R_\star$. Evaluating Equation 1.28 at $R = R_\star + b$ gives

$$C = 2\pi(R_\star + b)^3 \Sigma u_R \Omega(R_\star + b)$$

(1.30)

Under the assumption that $b \ll R_\star$, $(R_\star + b)^3 \simeq R_\star$ and $\Omega(R_\star + b) \simeq \sqrt{GM_\star/R_\star^3}$ are valid approximations, so this becomes

$$C = -\dot{M}\sqrt{GM_\star R_\star}$$

(1.31)

by recalling Equation 1.26. This can now be substituted back into Equation 1.28 to give an expression for Σ in terms of $\dot{M}$ and ν (Assuming once again that the disc is Keplerian outside of the boundary region):

$$\nu\Sigma = \frac{\dot{M}}{3\pi} \left[1 - \sqrt{\frac{R_\star}{R}} \right]$$

(1.32)

For typical disc parameters and an α viscosity law this expression gives a surface density Σ at a radius of 1 au of $10 - 1000$ g cm^{-2} for $10^{-9} \lesssim \dot{M} \lesssim 10^{-7}$ $M_\odot$ yr^{-1}.

Far from the edge of the disc (where $R \gg R_\star$), $\Sigma \propto \nu^{-1}$. For a disc with a constant α that has a temperature profile of the form $T \propto R^{-1/2}$ (consistent with observations of

---

[6] Note that this analysis may not be strictly applicable to a T Tauri star, where the inner edge of the disc is not in fact at the stellar surface but is held out by the strong magnetosphere (see Figure 1.3). In practice this is likely to apply an additional torque at the disc edge and so using Equation 1.11 to calculate Γ_ν will not give the correct answer. However, to order of magnitude the result should be similar, and so the approximation is still useful.

flaring discs; Kenyon and Hartmann, 1987) and a constant $\dot{M}$, this gives $\Sigma \propto R^{-1}$. Typical parameters give a 'canonical' protoplanetary disc surface density profile

$$\Sigma(R) \simeq 100 \, \text{g cm}^{-2} \left(\frac{R}{1 \, \text{au}} \right)^{-1}, \tag{1.33}$$

which is a 'flatter' (Σ scales with a lower power of R) and lower surface density profile than given by the MMSN model (compare this with Equation 1.24).

1.3 Disc-satellite interactions

While disc models in themselves are an active area of research in astrophysics, of equal interest is the effect that embedded bodies may have on the disc in which they live, and vice versa. Although this is also relevant in the context of other types of discs than gas accretion discs (e.g. planetary rings and galactic discs), the treatment here will concentrate on a satellite to the central body about which the disc is orbiting (e.g. a young protoplanet in a protoplanetary disc, or a stellar or black hole binary with unequal masses)

The theoretical framework for understanding the effect that the satellite has on the disc goes back to pioneering work by Goldreich and Tremaine (1979, 1980), and although it has been expanded and updated a number of times since (e.g. by Tanaka et al., 2002, to consider a full 3D disc), the original work has stood the test of time exceptionally well. In this section I will outline the main concepts of their work.

First, consider a satellite embedded in a Keplerian disc orbiting a central mass M_c moving on a circular orbit with orbital frequency Ω_s. The satellite will excite a resonance at radii r in the disc where the ratio of orbital periods between satellite and disc is $m : m \pm 1$ for positive integer values of m. In terms of orbital frequencies, this becomes

$$m \left(\Omega_d(r) - \Omega_s \right) = \pm \Omega_d(r). \tag{1.34}$$

where $\Omega_d(r)$ is the orbital frequency of disc material at radius r. Resonances that fulfil this first condition are called Lindblad resonances (LRs).

A further resonance is excited where the condition

$$\Omega_d(r) = \Omega_s, \tag{1.35}$$

is fulfilled, called the corotation resonance (CR). Resonances that take a positive sign on the right hand side of Equation 1.34 are termed Outer Lindblad resonances (OLRs), while
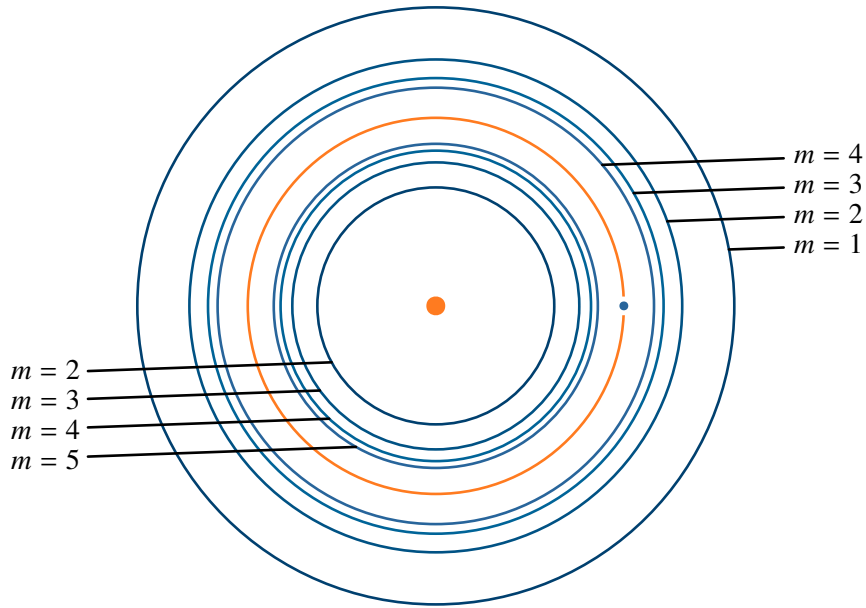
Figure 1.6: Schematic showing Lindblad and Corotation resonances (blue and orange lines respectively) excited in a Keplerian disc by a satellite on a circular orbit (solid blue circle) around a central mass (solid orange circle). Only the lowest 4 m resonances on each side of the satellite's orbit are shown here[7] – higher m resonances become more closely spaced as they approach the satellite's orbital radius. Figure adapted from Armitage (2007).

those taking a negative sign are Inner Lindblad resonances (ILRs). Using the Keplerian approximation we have made (i.e. that $\Omega = \sqrt{GM_c/r^3}$), we can see that LRs occur at radii r_L:

$$r_L = \left(1 \pm \frac{1}{m}\right)^{2/3} a_s \tag{1.36}$$

and the CR at r_C:

$$r_C = a_s \tag{1.37}$$

where a_s is the semimajor axis of the satellite's orbit. Figure 1.6 shows a schematic of the orbits at which these resonances occur.

Thus far I have made the picture far simpler by assuming a circular orbit for the satellite and Keplerian orbits for the disc material. However, Goldreich and Tremaine (1979) showed that if these assumptions are not made, the picture becomes instantly orders of magnitude more complex. Firstly, orbits are no longer solely defined by their orbital

---

[7] The reason why ILRs start at $m = 2$ can be seen from Equation 1.36: $m = 1$ gives $r_L = 0$.

frequency Ω, but also by their epicyclic frequency

$$\kappa^2 = \frac{2\Omega}{r} \frac{d}{dr} \left(r^2 \Omega \right) \tag{1.38}$$

for both the satellite and the disc material.

Goldreich and Tremaine (1979) also found that instead of resonances being excited by the single frequency Ω_s, resonances are also excited by harmonic components of the satellite's orbit. It is necessary to expand the potential of the planet ϕ^s into a fourier series, separating out the components that vary in time t and azimuthal angle θ from those that are functions of r:

$$\phi^s(r, \theta, t) = \sum_{l=-\infty}^{\infty} \sum_{m=0}^{\infty} \phi_{l,m}^s(r) \cos \left(m\theta - \left[m\Omega_s + \frac{(l+m)}{m}\kappa_s \right] t \right). \tag{1.39}$$

Goldreich and Tremaine (1980) found that for the satellite's eccentricity $e_s \ll 1$, $\phi_{l,m}^s$ is proportional to $e_s^{|l-m|}$. Therefore if we consider only up to first order in e_s we can discard all components except for $l = m - 1, m, m + 1$.

Equation 1.39 can be understood as each $\phi_{l,m}^s$ representing a harmonic of the planet's potential, which moves with pattern speed $\Omega_{l,m}$, given by

$$\Omega_{l,m} = \Omega_s + \frac{(l-m)}{m}\kappa_s \tag{1.40}$$

and this allows us to simplify Equation 1.39 into

$$\phi^s(r, \theta, t) = \sum_{l=-m}^{m} \sum_{m=0}^{\infty} \phi_{l,m}^s(r) \cos \left[m(\theta - \Omega_{l,m}t) \right]. \tag{1.41}$$

Using linear perturbation theory, Goldreich and Tremaine (1979) found that the response of the disc to excitation by ϕ^s now becomes discontinuous when either of the following conditions are satisfied:

$$m\left(\Omega_d(r) - \Omega_{l,m} \right) = \pm\kappa_d \tag{1.42}$$

or

$$\Omega_d(r) = \Omega_{l,m}. \tag{1.43}$$

For the circular satellite orbit and Keplerian disc case considered previously, $\kappa = \Omega$ and so these become identical to Equations 1.34 and 1.35. Therefore, these represent the general

conditions for Lindblad and Corotation resonances for any satellite and any disc.

This more general treatment shows us that instead of the simple case shown in Figure 1.6, the true picture for a non-circular satellite orbit is far more complicated. In fact, we can see that for each value of $m > 0$ [8], there are 9 separate resonances excited: each value of $l = m - 1, m, m + 1$ in the satellite's potential has an ILR, a CR and an OLR (Goldreich and Sari, 2003). Resonances excited by the $l = m$ component are called the principal resonances, while those with $l = m \pm 1$ are first-order resonances. $l = m + 1$ resonances are often termed 'fast' first-order resonances, as their pattern speed $\Omega_d = \Omega_{m+1,m} + \kappa_d/m$ (that is, the pattern speed of the resonance is faster than that of the potential component that excites it), and by the same token $l = m - 1$ resonances are called 'slow' first-order resonances (Masset and Ogilvie, 2004).

1.3.1   Torque strengths

The effect of these resonances is the excitation of spiral density waves in the disc, which generate torques between the components of the system – the exchange of angular momentum between the perturbing satellite and the disc material. The effect this has on the orbital elements on the satellite depends strongly upon the sign of the torques Γ generated by each component.

In the case of LRs, angular momentum can be thought of as being transported down the angular velocity gradient (Goldreich and Sari, 2003). As Ω is generally a decreasing function of radius, angular momentum will be removed from an ILR and deposited with the perturbing potential, and OLRs will take angular momentum from the perturber. Defining positive torques as those which decrease the angular momentum of the planet, $\Gamma_{\text{ILR}} < 0$ and $\Gamma_{\text{OLR}} > 0$. This does not help us for the case of a CR however, as it occurs at the same angular velocity as the perturbing component. In this case the important consideration is whether the interaction is dominated by material outside or inside of corotation – disc material on either side of co-rotation will move towards it. The effect is that Γ_{CR} takes the opposite sign to $d(\Sigma/B)/dr$, where B is Oort's constant

$$B = \frac{1}{2r}\frac{d}{dr}(r^2\Omega) = \frac{\kappa^2}{4\Omega}. \tag{1.44}$$

---

[8] Note that the case of $m = 0$ is special, as $\phi^s_{l,m=0}$ has no non-axisymmetric component (i.e. in Equation 1.41, $\cos[m(\theta - \Omega_{l,m}t)] = 1$). Instead we can think of $\phi^s_{l,m=0}$ as representing the axisymmetric 'smearing' of the planet's mass along the length of its orbit (Nixon et al., 2011b).

The net effect of these torques is a simple summation over all of the resonances:

$$\Gamma_{\text{net}} = \sum_{m=1}^{\infty} \Gamma_{\text{OLR}} + \sum_{m=2}^{\infty} \Gamma_{\text{ILR}} + \sum_{m=1}^{\infty} \Gamma_{\text{CR}}. \tag{1.45}$$

In practice, it is found that all of the terms are of similar magnitude, and finding the small net residual requires a number of assumptions to be made regarding the thermodynamics and hydrodynamics of the disc (e.g. Tanaka et al., 2002). Given the dependence of Γ_{CR} on surface density gradients, this generally becomes a highly nonlinear problem, and so numerical simulations are often required (e.g. de Val-Borro et al., 2006; Armitage, 2010).

1.3.2   Gap opening and migration types

It is a well-known result that a satellite embedded in an accretion disc will open a gap at the radius of its orbit if the mass ratio q between it and the central mass is high enough. This is most often seen in the context of embedded protoplanets. This is simple to see from applying the criteria given above for the signs of Lindblad torques to a circularly orbiting planet: Γ_{ILR} is negative and so gives angular momentum to the planet. Disc material at the resonance therefore has less angular momentum and its orbit shrinks, moving it away from the planet. Conversely, Γ_{OLR} is positive and receives angular momentum, so disc material moves outwards, again away from the planet. As this process repeats, the result is the formation of a gap in the disc.

However, this process is not always effective for a given planet in a given disc – instead, it is determined both by q and by the disc viscosity ν. To see how this is the case, it is convenient to have an approximation for the order of magnitude of Γ. Rather than follow the lengthy process of Goldreich and Tremaine (1979, 1980) to derive this, I instead use the impulse approximation approach adopted by Lin and Papaloizou (1979a,b)[9].

In a frame co-rotating with the planet, consider a fluid element with mass m passing close to a planet of mass M_{p} with some impact parameter b at relative velocity u, on a parallel orbit to the planet. The force perpendicular to the motion of the fluid, $F_{\perp}$ is given by

$$F_{\perp} = \frac{GmM_{\text{p}}}{b^2} \left[1 + \left(\frac{ut}{b} \right)^2 \right]^{-3/2} \tag{1.46}$$

where t is time, defined so that $t = 0$ corresponds to the moment the element passes the

---

[9] The derivation more closely follows the form of Armitage (2010), but the assumptions and results are identical.

planet. The impulse $\delta u_\perp$ felt by the planet due to the element is found by integrating

$$|\delta u_\perp| = \int_{-\infty}^{\infty} \frac{F_\perp}{m} dt = \frac{2GM_p}{bu}. \tag{1.47}$$

Assuming conservation of kinetic energy across the interaction, we can equate the initial and final velocity squares:

$$u^2 = |\delta u_\perp|^2 + (u - \delta u_\parallel)^2. \tag{1.48}$$

Assuming that the deflection is small, we can say that $u^2 + u_\parallel^2 \simeq u^2$, so

$$\delta u_\parallel \simeq \frac{|\delta u_\perp|^2}{2u} = \frac{1}{2u}\left(\frac{2GM_p}{bu}\right)^2. \tag{1.49}$$

For a planet semimajor axis a, we can approximate the order of magnitude of the specific angular momentum exchange Δj as

$$\Delta j \simeq a\,\delta u_\parallel = \frac{2G^2M_p^2 a}{b^2 u^3}. \tag{1.50}$$

The total torque on the planet in this approximation can be found by integrating over the whole disc. The mass dm of an annulus with surface density Σ close to the planet between b and $b + db$ (so that $a \simeq b$) is given by

$$dm \simeq 2\pi a \Sigma db. \tag{1.51}$$

The annulus and planet have angular frequencies Ω and Ω_p respectively, and the time Δt for the full annulus to interact with the planet is simply

$$\Delta t = \frac{2\pi}{|\Omega - \Omega_p|} \simeq \frac{4\pi}{3\Omega_p}\frac{a}{b} \tag{1.52}$$

using a first order binomial expansion to approximate $|\Omega - \Omega_p| \simeq 3\Omega_p b/2a$. This can also be used to approximate the relative velocity as $u = |u_{ann} - u_p| \simeq a|\Omega - \Omega_p| \simeq 3\Omega_p b/2$. Using this, the total torque from the annulus is $dL_{ann}/dt \simeq \Delta j\, dm/\Delta t$:

$$\frac{dL_{ann}}{dt} = -\frac{3G^2 M_p^2 a \Sigma \Omega_p}{bu^3}\,db = -\frac{8}{9}\frac{G^2 M_p^2 a \Sigma}{\Omega_p^2}\frac{db}{b^4}, \tag{1.53}$$

and the total torque $\Gamma = dL/dt$ from gas exterior to the planet is finally given by integrating

outwards to infinity from some minimum impact parameter b_{min}:

$$\frac{dL}{dt} = \int_{b_{\mathrm{min}}}^{\infty} \frac{dL_{\mathrm{ann}}}{dt} = -\frac{8}{27} \frac{G^2 M_{\mathrm{p}}^2 a \Sigma}{\Omega_{\mathrm{p}}^2 b_{\mathrm{min}}^3}. \tag{1.54}$$

We are now in a position to evaluate the gap-opening criterion for an embedded planet. The characteristic length-scale in a disc is its scale height H, so the gap must be approximately this wide to be maintained for a significant time. The angular momentum ΔL required to clear gas out of an annulus between a and $a + H$ is

$$\Delta L = 2\pi a H \Sigma \cdot \left. \frac{dj}{dr} \right|_a \cdot H \tag{1.55}$$

where j is the specific angular momentum of the gas, given by $j = \sqrt{GM_\star r}$ in the case of Keplerian orbits. The time-scale for the process, t_{open} can be estimated from

$$t_{\mathrm{open}} = \frac{\Delta L}{dL/dt}. \tag{1.56}$$

if we take $b_{\mathrm{min}} = H$ in Equation 1.54. As it is the disc viscosity which opposes the opening of the gap, we can equate the gap closing time-scale with the viscous time-scale (Equation 1.22):

$$t_{\mathrm{close}} = \frac{H^2}{\nu} \tag{1.57}$$

where ν can be taken to be the Shakura and Sunyaev prescription: $\nu = \alpha c_{\mathrm{s}} H = \Omega H^2$. By equating these time-scales and after some rearranging, we arrive at the critical mass ratio $q_{\mathrm{crit}} = M_{\mathrm{p}}/M_\star$ required to open and maintain a gap in a disc:

$$q_{\mathrm{crit}} = \left(\frac{27\pi}{8} \right)^{1/2} \left(\frac{H}{R} \right)^{5/2} \alpha^{1/2}. \tag{1.58}$$

An additional criterion for gap-opening can be found by considering the disc scale height and the planet's Hill radius – the radius at which gravitational interaction is dominated by the planet rather than the central star. This is given by

$$R_{\mathrm{Hill}} = a \left(\frac{q}{3} \right)^{1/3} \tag{1.59}$$

A planet can only open a gap when $R_{\text{Hill}} \gtrsim H$, so a new limit on q_{crit} is that

$$q_{\text{crit}} \gtrsim 3 \left(\frac{H}{R} \right)^3 . \tag{1.60}$$

For typical protoplanetary disc parameters ($H/R = 0.05$, $\alpha = 0.01$, $M_\star = M_\odot$) a gap can be opened by a planet of approximately Saturn mass. This neatly splits the population into two bins: those which are able to open a gap, and those which are not. The latter are refered to as being in the 'Type I migration regime'. As they are unable to maintain a gap in the disc, they migrate under the influence of their disc torques: In the isothermal case traditionally considered, the outer Lindblad torques are generally stronger, resulting in rapid inward migration on short time-scales. This is Type I migration.

Gap-opening planets instead migrate at much slower rates. As they are able to hold back the viscously accreting gas, the gap moves radially inwards at the same rate as that at which the gas accretes. They therefore migrate on viscous time-scales, and this 'Type II' migration is orders of magnitude slower than Type I. In Chapter 2 I address how planets in the Type II regime interact with the disc and how this affects the planet's orbital eccentricity, and discuss how the problem differs from that where planets are in the Type I regime. In Chapter 3 I test how the problem changes when we consider a planet and disc around a binary star, with specific focus on the circumbinary planet Kepler-16b.

1.4 Numerical techniques

The scientific results presented in Chapters 2 to 4 of this thesis are based on the results of numerical simulations using smoothed particle hydrodynamics (SPH). As the name suggests, this is a particle-based method for computing fluid dynamics which is widely used in astrophysics. Rather than give a detailed review of SPH and the many flavours and variants in use throughout the literature, I will provide a basic description of the principles and concepts involved and derive the basic SPH equations (Sections 1.4.1, 1.4.1.1 and Appendix A), before describing the non-standard modifications made to the code used to perform the simulations presented in later chapters (Section 1.4.1.2). The more interested reader is referred to works by Cossins (2010) and especially Price (2012) for thorough introduction and reviews of the method.

1.4.1 Smoothed particle hydrodynamics

SPH is a particle-based Lagrangian technique for calculating fluid dynamics. It has been widely used in astrophysics since its inception nearly 40 years ago (Gingold and Monaghan, 1977; Lucy, 1977), as it has a number of attractive properties. Firstly, resolution follows the flow automatically, so there is no need to pre-define a simulation box or region – and no time is wasted calculating fluid quantities in regions with no fluid. Another major point in favour of SPH is that it is naturally conservative. Therefore it is particularly well suited to problems where angular momentum conservation is of key importance. While not without its problems (e.g. sometimes a low density region needs to be well resolved, and SPH in practice tends towards being more viscid than other numerical methods), the shortcomings of SPH are generally well understood and characterized (e.g. Price, 2012).

The beating heart of any SPH code is the density estimator – the very basis of the method lies in being able to estimate the fluid density based on the positions of a distribution of particles, which are used as interpolation points of the underlying fluid. For an individual particle a with position $\boldsymbol{r}_a$ and mass m_a in a field of similar particles with positions $\boldsymbol{r}_b$ and masses m_b one estimates the density as

$$\rho(\boldsymbol{r}_a) = \sum_b^{N_{\text{neigh}}} m_b W(|\boldsymbol{r}_a - \boldsymbol{r}_b|, h_a) \tag{1.61}$$

where W is a weighting function used to smooth the distribution, called the smoothing kernel, and N_{neigh} is some number of neighbouring particles over which we sum, which is set by the smoothing length associated with the kernel, h_a.

A good choice of smoothing kernel will satisfy three essential criteria:

- It must be non-negative for all r, decrease monotonically with r and have a smooth derivative.

- It must be symmetric in r, so that $W(|\boldsymbol{r}_a - \boldsymbol{r}_b|, h_a) \equiv W(|\boldsymbol{r}_b - \boldsymbol{r}_a|, h_a)$.

- It must be flat for small r, so that small-scale changes in position do not strongly affect the density estimate.

A familiar function that fulfils these is the Gaussian, but this has the drawback that it never formally goes to zero – therefore the sum in Equation 1.61 would be over all particles, which is computationally very expensive. Instead, a family of functions called the B-spline kernels are most often used, primarily the M_4 cubic spline, and it is this kernel that

is used in the simulations presented in later chapters. The form of this kernel is [10]

$$W(r, h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3 & \text{for } 0 \leqslant \frac{r}{h} \leqslant \frac{1}{2}, \\ 2\left(1 - \frac{r}{h}\right)^3 & \text{for } \frac{1}{2} < \frac{r}{h} \leqslant 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.62}$$

where $r = |\boldsymbol{r}_a - \boldsymbol{r}_b|$. Here we can see how the smoothing length h sets N_{neigh} in Equation 1.61 – it is simply the number of particles within radius h of particle a. In practice, N_{neigh} is a constant of the simulation used to set the smoothing length h for each particle, and in the simulations described in this thesis we set $N_{\text{neigh}} = 50$. The relationship between the density and smoothing length is set by enforcing

$$\frac{4\pi}{3} h_a^3 \rho_a = m_a N_{\text{neigh}}. \tag{1.63}$$

From Equation 1.61, it is possible to estimate the the value of any arbitrary field just from the positions and smoothed densities. Consider a field F with N particles as before with positions $\boldsymbol{r}_a$ and densities ρ_a where the field $F(\boldsymbol{r} = \boldsymbol{r}_a) = F_a(\boldsymbol{r}_a)$ is known. The field at positions $\boldsymbol{r}' \neq \boldsymbol{r}_a$ can be estimated as the smoothed field $F(\boldsymbol{r}' \neq \boldsymbol{r}_a) = F_{\text{s}}(\boldsymbol{r}')$ by noting that

$$F_{\text{s}}(\boldsymbol{r}') = \int_0^\infty F(\boldsymbol{r}') W(|\boldsymbol{r}_a - \boldsymbol{r}'|, h_a) \, d\boldsymbol{r}'$$

$$\simeq \sum_a^{N_{\text{neigh}}} \frac{m_a}{\rho_a} F_a(\boldsymbol{r}_a) W(|\boldsymbol{r}_a - \boldsymbol{r}'|, h_a). \tag{1.64}$$

One further useful feature of SPH is that the spatial derivative for any field estimated in this manner can be found simply by taking the derivative of the kernel itself. To see that this is true we can consider the same field F and take the derivative $\nabla_a F = \partial F / \partial \boldsymbol{r}_a$:

$$\nabla_a F_{\text{s}}(\boldsymbol{r}') = \frac{\partial}{\partial \boldsymbol{r}_a} \int_0^\infty F(\boldsymbol{r}') W(|\boldsymbol{r}_a - \boldsymbol{r}'|, h_a) \, d\boldsymbol{r}'. \tag{1.65}$$

As the only part of the right hand side of this to depend on $\boldsymbol{r}_a$ is the kernel, we can follow

---

[10] Note that the standard in SPH literature is for W to vanish at $r = 2h$ rather than at $r = h$. This is a peculiarity of GADGET-2 (Springel, 2005), the code used here, and where I make comparison to the literature I will use this definition of the standard SPH smoothing length rather than that used by GADGET-2.

the same step taken in Equation 1.64 to estimate the integral as a sum, giving

$$\nabla_a F_s(\boldsymbol{r}') \simeq \sum_a^{N_{\text{neigh}}} \frac{m_a}{\rho_a} F_a(\boldsymbol{r}_a) \nabla_a W(|\boldsymbol{r}_a - \boldsymbol{r}'|, h_a). \tag{1.66}$$

This makes SPH calculations very efficient because there is no need to take derivatives during the simulation – ∇W is known simply from the choice of W. The same is true for the divergence and curl of vector fields following the same arguments.

In order to implement these methods into a simulation we need a set of equations of motion that describe the hydrodynamics: the continuity, momentum and energy equations. The first of these is given simply by the SPH density estimator (Equation 1.61), which enforces conservation of mass. The momentum equation for SPH is given by

$$\frac{d\boldsymbol{u}_a}{dt} = -\sum_b^{N_{\text{neigh}}} m_b \left[\frac{P_a}{\Lambda_a \rho_a^2} \nabla_a W_{ab}(h_a) + \frac{P_b}{\Lambda_b \rho_b^2} \nabla_a W_{ab}(h_b) \right], \tag{1.67}$$

where

$$\Lambda_a = 1 + \frac{h_a}{3\rho_a} \sum_a^N m_a \frac{\partial W_{ab}}{\partial h_a} \tag{1.68}$$

is a term accounting for smoothing-length gradients.

In the formulation of SPH used in this thesis a particle's internal energy ϵ_a is not directly evolved. Instead an entropic function $A_a(\epsilon_a)$ is used, from which the internal energy is computed. In this way we can strictly ensure that entropy is never inadvertently decreased through errors caused by SPH estimation methods. In the simulations described in Chapters 2 and 3, a locally isothermal equation of state is enforced, so that A_a (and therefore ϵ_a) depend only on a particle's radius R from the barycentre of the simulation, allowing the disc scale height H to be explicitly set as an input parameter. In Chapter 4, I use

$$P_a = A_a \rho_a^\gamma, \qquad A_a = \frac{(\gamma - 1)\epsilon_a}{\rho_a^{(\gamma-1)}} \tag{1.69}$$

where γ is the polytropic index of the gas. A_a is then evolved according to

$$\frac{dA_a}{dt} = \frac{\gamma - 1}{\rho^{\gamma-1}} \left(\frac{d\epsilon_a}{dt} \right), \tag{1.70}$$

with contributions to the term $d\epsilon_a/dt$ described explicitly in Section 1.4.1.1 and Chapter 4.

1.4.1.1 Artificial viscosity

These equations describe a dissipation-less flow where entropy and momentum are naturally conserved, thanks to the Lagrangian nature of SPH. However, in a real astrophysical fluid this is not necessarily the case. Converging flows will shock against one another, dissipating energy by heating. Dissipation-less SPH also experiences problems of particle-penetration, where particles become disordered and pass through one another.

In these cases, 'artificial' terms must be added to the SPH equations to mimic the real behaviour of an astrophysical fluid, known as artificial viscosity. There are a number of different forms of this, and the one implemented in the code used here is that of Morris and Monaghan (1997).

In this scheme, the additional term

$$\frac{d\boldsymbol{u}_a}{dt}\bigg|_{\text{AV}} = -\sum_b^{N_{\text{neigh}}} m_b \, \Pi_{ab} \, \nabla_a \overline{W}_{ab} \tag{1.71}$$

is added to Equation 1.67, where $\overline{W}_{ab}$ is the arithmetic mean of the kernels $W_{ab}(h_a)$ and $W_{ba}(h_b)$. Similarly the dissipation term in Equation 1.70 is given by

$$\frac{d\epsilon_a}{dt}\bigg|_{\text{AV}} = -\sum_b^{N_{\text{neigh}}} m_b \, \Pi_{ab} \, \boldsymbol{u}_a \cdot \nabla_a \overline{W}_{ab}. \tag{1.72}$$

Π_{ab} is the actual artificial viscosity term, and is given by

$$\Pi_{ab} = \begin{cases} (-\overline{\alpha}_{ab} \, \overline{c}_{s_{ab}} \mu_{ab} + \overline{\beta}_{ab} \mu_{ab}^2)/\overline{\rho}_{ab} & \text{for } \boldsymbol{u}_{ab} \cdot \boldsymbol{r}_{ab} < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{1.73}$$

with

$$\mu_{ab} = \frac{\overline{h}_{ab} \, \boldsymbol{u}_{ab} \cdot \boldsymbol{r}_{ab}}{|\boldsymbol{r}_{ab}|^2 + \xi \overline{h}_{ab}^2}. \tag{1.74}$$

In this notation, barred terms $\overline{Q}_{ab}$ refer to the arithmetic mean of quantity Q between particles a and b, $\overline{Q}_{ab} = (Q_a + Q_b)/2$. c_{s_a} is the sound speed of particle a given by $c_{s_a} = dP_a/d\rho_a$. Un-barred vector terms $\boldsymbol{Q}_{ab} = \boldsymbol{Q}_a - \boldsymbol{Q}_b$, and the term $\xi = 10^{-4}$ prevents the solution from becoming divergent for particles with very small $\boldsymbol{r}_{ab}$.

α_a and β_a are terms giving the level of artificial viscosity felt for each particle a. In the 'standard' formulation of SPH artificial viscosity, α is a constant of the simulation and is therefore identical for each particle at all times, and typically $\beta = 2\alpha$. The Morris

and Monaghan (1997) formulation introduces individual α_a for each particle and allows them to evolve in time, but $\bar{\beta}_{ab} = 2\bar{\alpha}_{ab}$ is still typically kept, and this is the form used throughout the simulations presented in this thesis.

Following the method advocated by Price (2004), α_a is evolved according to

$$\frac{d\alpha_a}{dt} = \frac{\alpha_{\min} - \alpha_a}{\tau_a} + (\alpha_{\max} - \alpha_a) S_a \tag{1.75}$$

where $\alpha_{\min}$ and $\alpha_{\max}$ are parameters of the simulation, τ_a is a decay time-scale given by

$$\tau_a = \frac{h_a}{2l c_{s_a}} \tag{1.76}$$

where l is a term typically $0.1 - 0.2$ (taken to be 0.1 throughout this thesis) and

$$S_a = \max\{-\nabla \cdot \boldsymbol{u}_a, 0\} \tag{1.77}$$

is the source term, which acts to increase the viscosity in regions of convergent flow. In the SPH simulations presented in Chapters 2 and 3, $\alpha_{\min} = 0.01$, but is increased to $\alpha_{\min} = 0.1$ in Chapter 4. This is due to the absence of an explicit Navier-Stokes viscosity in the latter simulations (see Section 1.4.1.2), requiring a higher level of artificial viscosity to prevent particle penetration. $\alpha_{\max} = 2$ is kept throughout.

It is useful to consider the form of the artificial viscosity 'switch' employed here (i.e. the conditions under which the terms in Equations 1.73 and 1.77 become non-zero). Ideally, the aim is to detect the presence of a shock, which the algorithms here interpret as any convergent flow. In Equation 1.73 this is true for the condition $\boldsymbol{u}_{ab} \cdot \boldsymbol{r}_{ab} < 0$ and similarly for $\nabla \cdot \boldsymbol{u}_a < 0$ in Equation 1.77.

However, both of these will also be true in the case of a shear flow (e.g. in an accretion disc), where the divergence is negative but the flows are not in actual fact converging. This problem can be reduced, but not eliminated, by employing a method devised by Balsara (1995). In this case we multiply Π_{ab} in Equation 1.73 by a factor $\overline{f_{ab}^{\text{Bal}}} = (f_a^{\text{Bal}} + f_b^{\text{Bal}})/2$ where

$$f_a^{\text{Bal}} = \frac{|\nabla \cdot \boldsymbol{u}_a|}{|\nabla \cdot \boldsymbol{u}_a| + |\nabla \times \boldsymbol{u}_a| + \xi(c_{s_a}/h_a)}. \tag{1.78}$$

This serves to suppress the artificial viscosity in the case where the vorticity (curl of the velocity field, $\nabla \times \boldsymbol{u}_a$) is dominant over the convergence. As in equation 1.74, the factor ξ avoids numerical problems with very small numbers. This method is employed throughout the simulations in this thesis.

1.4.1.2   Code modifications

In addition to the artificial viscosity described above, a number of non-standard modifications to the code have been made which I will now describe.

Firstly, a Navier-Stokes viscosity has been applied on top of the artificial viscosity described above. This was done to more accurately describe a Shakura and Sunyaev α-disc by allowing α to be an explicit input parameter of the simulation. This was done following the method of Lodato and Price (2010), using the 'two first derivatives' parameterisation described therein. To do this, the SPH momentum equation (Equation 1.67) is altered to include the shear stress tensor $\mathcal{S}$. Introducing indices i, j, k as vectors in Einstein summation notation, the momentum equation becomes

$$\frac{du_a^i}{dt} = -\sum_b^{N_{\text{neigh}}} m_b \left[\frac{P_a + \mathcal{S}_a^{ij}}{\Lambda_a \rho_a^2} \nabla_a^j W_{ab}(h_a) + \frac{P_b + \mathcal{S}_b^{ij}}{\Lambda_b \rho_b^2} \nabla_a^j W_{ab}(h_b) \right] \tag{1.79}$$

where the stress tensor $\mathcal{S}_a^{ij}$ is given by

$$\mathcal{S}_a^{ij} = \eta_a \left[\frac{\partial u_a^i}{\partial r_a^j} + \frac{\partial u_a^j}{\partial r_a^i} \right] - \frac{2}{3} \delta^{ij} \eta_a \left(\frac{\partial u_a^k}{\partial r_a^k} \right). \tag{1.80}$$

η_a is again the shear viscosity at particle a, related to the kinematic shear viscosity ν by $\nu = \eta/\rho$. Note that the additional artificial viscosity term given by Equation 1.71 is still included on top of this, as the lack of a bulk viscosity component here does not prevent particle penetration. A derivation and fuller description of the implementation is described in Appendix A. Calibration of the parameterisation is described in Chapter 2.

The version of the GADGET-2 code used includes so-called 'sink'-particles, following the method of Cuadra et al. (2006). These particles remove gas particles which lie within a given sink radius and absorb their mass and momentum. These are used in this thesis to describe solid or compact objects which only interact gravitationally and not hydrodynamically (specifically stars and planets in Chapters 1 and 2 and super-massive black holes in Chapter 4).

As standard, GADGET-2 uses a Barnes-Hut tree-code to estimate gravitational forces for all particles in the simulation. In this method, distant particles are grouped together into cells and their contribution to the potential is approximated at the cell centre of mass plus a number of multipole expansions. For gas particles this is sufficient, as they are well distributed at high N and can be accurately approximated with this method.

However, for the sink particles, this is not good enough. The effect of the gas on the

dynamics of the sink particles is of primary interest in these simulations, so potentially large errors introduced by poorly approximating their gravitational contribution is highly undesirable. To this end, I have modified the code to remove the sink particles from the gravity tree, and their forces are instead summed directly, following the method of Cuadra et al. (2009). As this is only done for 2 or 3 particles in each simulation, the additional computational expense involved is not great.

As mentioned a number of times, in Chapters 2 and 3 I impose a locally isothermal equation of state. This is done by calculating the radius R (in the $x - y$ plane) from the centre of mass of all the sink particles (the total mass of which I denote as $M_\star$), and setting the sound speed c_s at particle a such that

$$c_s = K_0 \left(\frac{R_a}{R_0} \right)^{-1/4} \tag{1.81}$$

where K_0 is set so that the scale height H/R is normalized to some value $(H/R)_0$ at the radius R_0,

$$K_0 = \left(\frac{H}{R} \right)_0 R_0 \sqrt{\frac{GM_\star}{R_0^3}}, \tag{1.82}$$

where the exact normalisation factors are given in the relevant chapters.

As the sound speed and Temperature T are related by $c_s^2 \propto T$, this $c_s \propto R^{-1/4}$ profile gives a temperature profile that goes as $T \propto R^{1/2}$. Remembering from Section 1.2 that the disc scale height H is defined as

$$H = c_s \Omega \tag{1.83}$$

this gives a disc that flares slightly, as $H/R \propto R^{1/4}$. These temperature and scale height proportionalities are consistent with a linear viscosity law (such as the α-disc model; Hartmann et al., 1998) and with observations of flared protoplanetary discs (Kenyon and Hartmann, 1987).

1.4.2   1D spreading ring

In Chapter 2, the SPH implementation of a Navier-Stokes viscosity described in Section 1.4.1.2 is calibrated using a spreading ring test. A 3D annulus of SPH particles is allowed to orbit a point mass and spread viscously, and the level of spreading is then compared to that of a similar disc modelled in 1D with a constant viscosity.

A numerical code is used to evolve the 1D viscous diffusion equation (Equation 1.14). While this equation does have an analytic solution for a constant viscosity ν, it involves
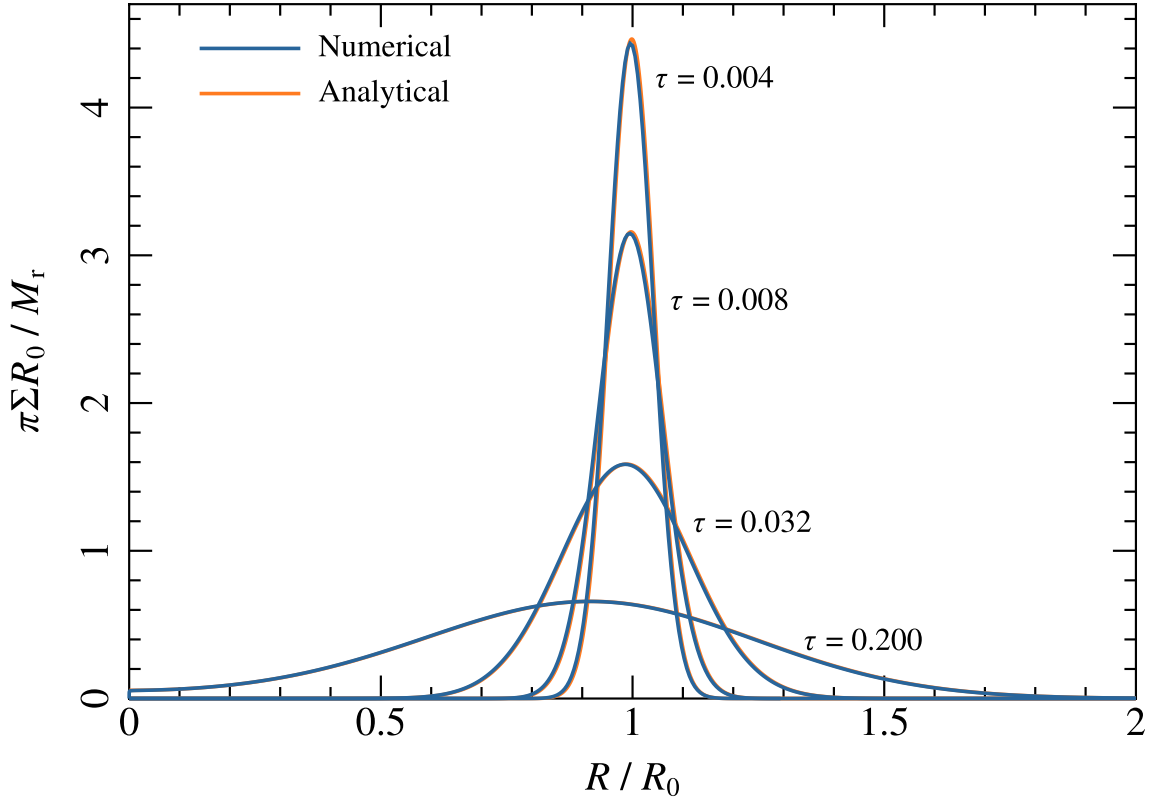
Figure 1.7: Analytical (orange) and numerical (blue) solutions to the spreading ring problem with constant kinematic viscosity ν. Dimensionless surface density $\pi\Sigma R_0/M_r$ (where M_r is the ring mass at R) is plotted against dimensionless radius R/R_0 at various dimensionless times $\tau = 12\nu t R_0^{-2}$. The slight disparity between the solutions at early times ($\tau = 0.004$) is due to the different initial conditions: the analytic solution is initially a delta-function at $R = R_0$ while the numerical solution requires a finite initial width. The exact solution was taken from the source code to SPLASH (Price, 2007).

———————————————————————————————————————

a modified Bessel function of the first kind, and calculating it is a complicated process (I refer the more interested reader to the original derivation by Lynden-Bell and Pringle, 1974, or the somewhat less opaque review by Pringle 1981). Instead, using the substitutions $X = 2R^{1/2}$, $Y = 3\nu\Sigma R^{1/2}$ and $S = \Sigma R^{3/2}$, Equation 1.14 can be rearranged into

$$\frac{\partial S}{\partial t} = \frac{\partial^2 Y}{\partial X^2}. \tag{1.84}$$

This differential equation can then be easily integrated using a numerical scheme. In this thesis I use a grid equispaced in X (i.e. in $R^{1/2}$) with spacing ΔX. Using a first-order

explicit scheme to integrate Equation 1.84, at timestep t and in grid cell i

$$\left(\frac{\partial^2 Y}{\partial X^2}\right)_i^t = \frac{1}{\Delta X^2}\left(Y_{i-1}^t + Y_{i+1}^t - 2Y_i^t\right) \tag{1.85}$$

so that

$$S_i^{t+1} = S_i^t + \Delta t\left[\frac{1}{\Delta X^2}\left(Y_{i-1}^t + Y_{i+1}^t - 2Y_i^t\right)\right]. \tag{1.86}$$

where Δt is the length of the current timestep.

The results of this numerical scheme are compared to the exact solution (taken from the source code to the SPLASH software package; Price, 2007) in Figure 1.7. On the basis of the excellent agreement, I use the numerical rather than the exact solution, purely for convenience, to calibrate the SPH Navier-Stokes viscosity described in Section 1.4.1.2.

2

Planetary eccentricity growth from planet-disc interactions

In this chapter, I present SPH simulations of giant planets embedded in protoplanetary discs. The aim of these simulations is to determine how interactions with the disc affect the eccentricity of the planet. It has been suggested that the origin of highly eccentric exoplanet orbits may be attributable to this mechanism (e.g. Papaloizou et al., 2001, hereafter PNM01). Before discussing the simulations and the results thereof, I will briefly discuss the history of exoplanet observations that motivate this work.

2.1 Introduction

2.1.1 Exoplanet observations & trends

Observations of planets of course start with our own Solar System. The orbits of the 8 planets are co-planar to within $\sim 10°$ and predominantly circular (Mercury, with its eccentricity $e = 0.2$ is the only planet with $e > 0.1$). As it was the only known system of planets until very recently, models of planet formation were of course tailored to reproduce these characteristics. It is still unclear just how typical the Solar System is of planetary systems in the wider Galaxy, as observational surveys are only just beginning to probe systems of a comparable scale.

The most obvious method to use to look for exoplanets is direct imaging (DI). At optical wavelengths Solar System planets reflect light from the Sun towards the Earth and are thus visible, rather than shining with their own intrinsic luminosity (this is not the case in the infrared, see below). This process does not allow us to observe exoplanets, as they are outshone by their parent star by many orders of magnitude at interstellar distances. That this is true can be seen by considering the maximum fraction f_r of the star's light that will be reflected by a planet of radius r_p orbiting with semimajor axis a. This is given by the surface area of the planet's hemisphere divided by that of a sphere at the planet's orbit

$$f_r = \frac{\pi r_p^2}{4\pi a^2}.$$ (2.1)

For an Earth-like planet orbiting at 1 au, $f_r \sim 5\times10^{-10}$, while for Jupiter (the largest planet in the Solar System, and therefore likely the easiest to detect) $f_r \sim 2 \times 10^{-9}$. Indeed, not even this fraction of the star's light will be reflected, as the planet's albedo (how much light it absorbs rather than reflects) must be taken into account. Evidently if we are to directly image exoplanets, looking directly at them in the optical will not yield success. The problem becomes harder when one considers the necessity to resolve the planet's orbit.

Solutions to this problem have been found by masking the light from the central star to decrease the contrast problem, and observing in the infrared (e.g. Marois et al., 2008). At these wavelengths, it is in fact the planets' own thermal emissions rather than reflection of stellar light that is observed. This provides the bonus that it gives us information about the planet's thermal structure. However, DI is still limited to very massive planets in wide orbits, and requires 8m-class telescopes to achieve from the ground even after removing the contrast problem, but as a method is responsible for discovering some of the most famous exoplanets (Fomalhaut b, the HR8799 system and β Pic b; Kalas et al., 2008; Marois et al., 2008; Lagrange et al., 2009).

If most exoplanets are impossible to directly image using current telescopes, other more circuitous methods must be used. Long the most popular method, and the first to successfully detect an exoplanet around a main-sequence star (51 Peg b; Mayor and Queloz, 1995) is radial velocity (RV). This method uses high-resolution spectroscopy to track doppler shifting of spectral lines in the stellar light to infer the existence of a planet, which for a non-eccentric orbit traces a simple sine curve. For a planet of mass M_p in a circular orbit around a star of mass $M_\star$, the amplitude K of the observed doppler shift is related to the radial velocity of the star $v_\star$ and the inclination i of the system with respect to the plane of the sky by

$$K = v_\star \sin i \qquad (2.2)$$

and conservation of momentum ensures that $v_\star$ is related to the radial velocity of the planet v_p by $v_\star M_\star = v_p M_p$. In the limit of $M_p \ll M_\star$, a circular Keplerian orbit with semimajor axis a has orbital velocity $v = \sqrt{GM_\star/a}$. So for a planet with semimajor axis a_p the observed doppler shift amplitude K gives the planet mass via

$$K = \left(\frac{M_p}{M_\star}\right)\sqrt{\frac{GM_\star}{a_p}} \sin i. \qquad (2.3)$$

Therefore if we know $M_\star$, we know the *minimum* possible mass of the planet, which is $M_p \sin i$. For eccentric orbits, the doppler shifting signal is not sinusoidal, but instead changes as a function of the planet's orbit. This can mean that poorly sampled light-curves can produce spurious eccentricity, but given that RV surveys have been going since the mid-1990s, this is rarely a problem with current statistics (Armitage, 2010).

While the RV method is not without its problems (prime amongst them is that stellar activity can mimic a planetary signal; e.g. Desidera et al., 2004), it was long the workhorse of the exoplanet community and these are by now well understood and characterized.

More recently however, large transit surveys such as the ground-based SuperWASP and the satellites *CoRoT* and *Kepler* have overtaken RV as the prime contributor of newly discovered planets. The transit method detects planets that pass directly in front of their star and block out some of the light. The fraction f_t of light blocked out is simply the ratio of areas

$$f_t = \frac{\pi r_p^2}{\pi r_\star^2}.$$ (2.4)

where $r_\star$ is the radius of the star[1]. Transits therefore probe a different property of the planet than do RV signals, as they depend on the planet's radius rather than mass. This permits the planet's mean density to be found in cases where both transit and RV data are available, and explaining the relationship (if indeed one exists) between them is a field of very active research.

Taken together, RV- and transit-discovered planets make up the vast bulk of currently known exoplanets, which currently stands at around 700 (Wright et al., 2011), with *Kepler* contributing some 3000 additional unconfirmed candidate planets (Batalha et al., 2013). Additional techniques such as microlensing and astrometry have been used to look for exoplanets, but the numbers found pale in comparison.

The first major trend to emerge in the few years following the first exoplanet discoveries was the discovery of a large number of so-called 'hot Jupiters'. As the name suggest, these are Jovian-mass (or higher) and in very close orbits ($a_p \lesssim 0.1$ au) around their host stars. While RV surveys are indeed biased towards finding close and massive planets such as these, which have a much stronger doppler signal, their very existence at all was something of a surprise. As the time baseline of these surveys has increased and number statistics have improved greatly, they seem to be less common than initially thought (around ~ 1 percent of sun-like stars are thought to host them; Wright et al., 2012) but are still an interesting phenomenon.

However, more relevant for the content of this chapter is the discovery that giant planets inhabit predominantly eccentric orbits. The distribution of measured eccentricities e (which are usually only obtainable from RV measurements) approaches uniformity for $0 < e \lesssim 0.4$, but stretches all the way up to $e \simeq 1$ (e.g. Wright et al., 2011; Kane et al., 2012). There is a trend towards moderate eccentricity ($e \sim 0.2 - 0.4$) for higher mass planets, and this is especially pronounced for $M_p \gtrsim 5\,M_{Jup}$ (Marcy et al., 2005; Udry and Santos, 2007).

In contrast, planets with small semimajor axis (such as the hot Jupiters) show smaller

---

[1] This is only correct to first order, and neglects effects such as grazing transits and limb darkening.
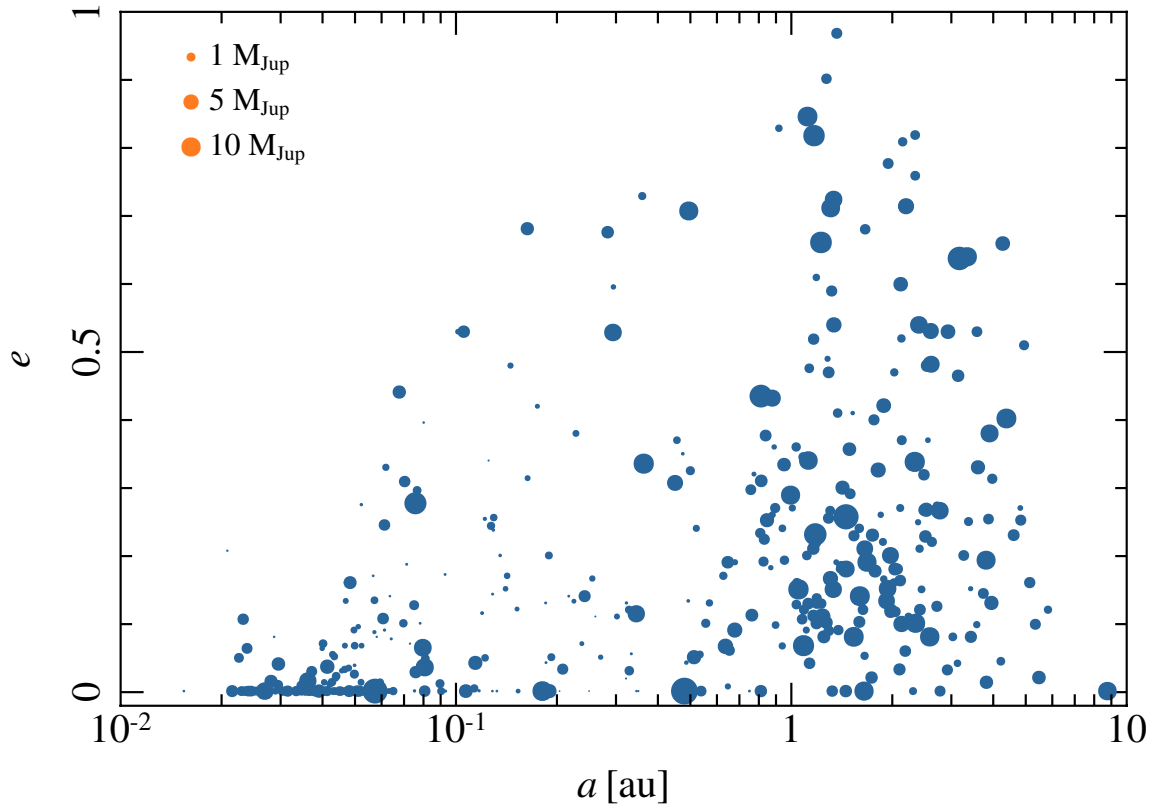
Figure 2.1: Observed eccentricity against semimajor axis for 627 known exoplanets (data taken from the Exoplanet Data Explorer; Wright et al., 2011). Planet mass is encoded in the circle radius, which follows $r \propto M_\mathrm{p}^{1/3}$. Orange circles are example sizes for 1, 5 and 10 M_Jup planets. The trend towards low e at small a is clearly visible, as is the preference for $e \gtrsim 0.1$ for planets $M_\mathrm{p} \gtrsim 5 M_\mathrm{Jup}$ at $a \sim 1$ au. These planets are particularly hard to explain with planet-planet scattering events (e.g. Ford and Rasio, 2008). It is also noteworthy that the range of e extends all the way up to $e \sim 1$.

---

eccentricities. This is well explained by tidal circularisation due to proximity to their host star (Rasio et al., 1996). These trends are clearly visible in Figure 2.1, which plots observed eccentricities against semimajor axis for 627 planets for which e_p, a_p and M_p (or $M_\mathrm{p} \sin i$ if i is not known) have been measured. An interesting sidenote is that *Kepler* data shows a different population of hot Jupiters than does the RV data. A long-observed pile-up of hot Jupiters in orbits with periods of ~ 3 days (e.g. Cumming et al., 2008) is not observed by *Kepler*, and this is thought to be due to differences in the metallicities of the target systems (e.g. Dawson et al., 2012; Dawson and Murray-Clay, 2013). A number of scenarios have also been proposed to to explain the unexpectedly large eccentricities of giant exoplanets. These fall into three broad categories: dynamical interactions of planets in multi-planet systems; secular interactions with companion stars; and tidal interactions

between planets and their parent protoplanetary discs.

2.1.2   Explaining eccentric exoplanet orbits

Large regions of the observed eccentricity distribution can be populated by invoking in-
teractions between multiple planets, be it via direct close encounters (Ford et al., 2001;
Jurić and Tremaine, 2008; Chatterjee et al., 2008) or through mean-motion resonances
over longer time-scales (Chiang et al., 2002). While simulations of such encounters are
able to reproduce the observed distribution to a reasonable accuracy, it is unclear if such
close interactions are frequent enough in nature to provide a universal source of planetary
eccentricity. There are also problems reproducing the observed rates of giant planets in
eccentric orbits ($e \gtrsim 0.2$) at $a \sim 1$ au (see Figure 2.1 Ford and Rasio, 2008).

Another method for growing eccentricity is through secular interactions with inclined
companion stars, which lead to a resonant exchange of angular momentum between the
planets and the external body (Kozai, 1962; Lidov, 1962). This results in long-period
changes in inclination and eccentricity and although this mechanism seems inviting as
an alternative explanation for the observed eccentricity distribution, numerical work has
shown that it does not produce the correct eccentricity distribtion (Takeda and Rasio,
2005), although recent work has found that it may explain some eccentric misaligned Hot
Jupiters (Naoz et al., 2012). Similar interactions between planets in the same system have
also been suggested as a chaotic formation mechanism for highly eccentric planets (Wu
and Lithwick, 2011).

As described in Section 1.3, the interaction between an embedded planet and its parent
gas disc can have a strong effect on the planet, and this has also been suggested as a
mechanism for driving eccentricity growth. For companions with masses comparable
to the central body it has long been known that tidal interactions with the disc lead to
eccentricity excitation. This result has applications to stellar binaries (e.g. Artymowicz
et al., 1991) and binary super-massive black holes (e.g Cuadra et al., 2009), but how it
extends to the more extreme mass ratios of star-planet systems is still not clear.

Semi-analytic calculations combining prescriptions from Goldreich and Tremaine (1980)
and Goldreich and Sari (2003) have been somewhat inconclusive. Moorhead and Adams
(2008) found eccentric damping rather than growth in most cases, although in the cases
where they did find growth it was extremely strong, leading to $e \sim 1$ after only a few
thousand orbits. However, as such highly eccentric planets would be unable to maintain
an equally eccentric gap their orbits would be circularised as they interact with coorbital
disc material (e.g. Bitsch and Kley, 2010).

By contrast, numerical simulations looking specifically at eccentricity growth have so far shown more consistent and positive results. PNM01 found that relatively massive embedded companions (in the brown dwarf regime) undergo eccentricity growth. Lower-mass planets were not found to experience this growth, although it seems likely that this was for numerical rather than physical reasons (Masset and Ogilvie, 2004). Later simulations have indeed found eccentricity growth, albeit at modest levels, down to $M_{\rm p} \sim M_{\rm Jup}$(D'Angelo et al., 2006). Extensive analysis of the behaviour and morphology of the disc by PNM01 and Kley and Dirksen (2006) attributed this eccentricity excitation to an instability launched at the 3:1 outer Lindblad resonance, which drives a large eccentricity at the inner edge of the disc. For the large companion masses considered by PNM01 a wide gap is opened in the disc, so coorbital co-rotation resonances are not present, and non-coorbital ones only operate once the planet's orbit is already eccentric. Kley and Dirksen (2006) also extended this analysis down to planets of a few $M_{\rm Jup}$, and found that this mechanism still operates down to planets of mass $M_{\rm p} \sim 3\ M_{\rm Jup}$, although the magnitude of the eccentricity induced depends strongly on the disc viscosity and temperature.

To date the majority of the numerical simulations of this problem have been performed in only two dimensions (2D), and all have used Eulerian (grid-based) methods. However, it has been suggested that a full three-dimensional (3D) treatment weakens the effect of resonant torques (Tanaka et al., 2002). Each study has also typically only considered a single disc model, with little consistency in the choice of parameters, and we have already seen in Section 1.3.2 that the parameters of the disc strongly affect how it interacts with an embedded planet. Moreover, Eulerian methods are not always ideal for following the dynamics of gas on non-circular orbits; in general, one expects Lagrangian methods to track eccentric orbits with greater accuracy.

In this chapter I present results of high-resolution 3D SPH simulations of eccentricity growth due to planet-disc interactions. I first explain in Section 2.2 how the Navier-Stokes viscosity described in Section 1.4.1.2 was calibrated against solutions to the spreading ring problem (Section 1.4.2).

2.2   Navier-Stokes viscosity calibration

To test and calibrate the Navier-Stokes viscosity in the SPH code, I conducted test simulations which model the viscous spreading of a gas ring around a point mass. In these tests a thin ring of initially uniform surface density Σ_0 was allowed to evolve under the action of viscous torques. The rings had a thickness of $0.2\ R_0$ centred on radius R_0. The
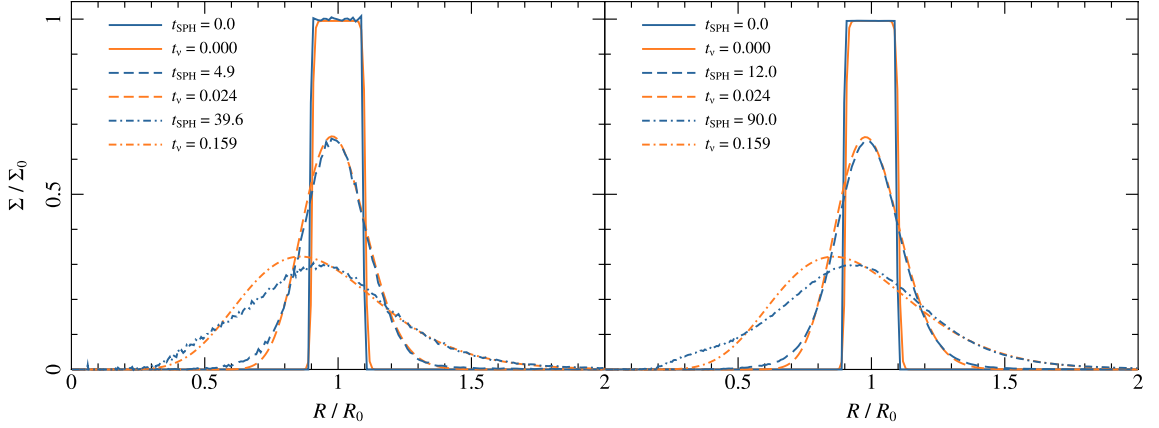
Figure 2.2: Radial surface density evolution of a viscously spreading ring. The left panel shows a ring with 10^5 particles in its initial configuration (solid blue line), after 4.9 orbits (dashed blue line) and after 39.6 orbits (dash-dotted blue line). The corresponding orange lines show the best-fitting profile from an explicit 1D ring code, plotted as a fraction of the viscous spreading time. The right panel is as the left, but for a ring with 10^6 particles, in its initial configuration (solid line), after 12 orbits (dashed line) and after 90 orbits (dash-dotted line). The Navier-Stokes viscosity approximates a uniform viscosity well except at very low resolution, where the 'best fit' to the 1D ring is rather poor in both cases. The initial configuration in the right hand panel is representative of the resolution obtained in the full disc models used in the later simulations. The SPH rings shown are from the runs with $\nu_{in} \gtrsim 10^{-5}$ (see Table 2.1).

---

ring is expect the ring to spread, as described in Sections 1.4.2 (see also Pringle, 1981; Frank et al., 2002). These spreading rings were then used to test the accuracy of the SPH viscosity prescription by comparing them against results from the 1D explicit scheme described in Section 1.4.2. I modelled the ring using 10^5 and 10^6 SPH particles and four different levels of viscosity (see Table 2.1). The ring was set up orbiting a single point mass, and was allowed to evolve for 200 orbits.

To compare the SPH ring spreading to that of the 1D code, I performed a simple least-squares fit for $\Sigma(R)$ at several different times in each run, and measure the effective fraction of the viscous time t_ν. The ring spreading is initially linear with the 1D model as expected, but later the approximation of constant viscosity becomes invalid as the artificial viscosity becomes strong at the ring edges. This is because the artificial viscosity scales with the SPH smoothing length (Equation 1.74) and thus is more viscid in low-density regions. I have only fit spreading times during this initial phase where the linear relationship exists. Comparisons between the surface density profiles of the SPH and 1D rings at various times at the different resolutions are shown in Figure 2.2.

The fits between the SPH and 1D rings allow a comparison between the imposed viscosity ν_{in} and the measured rate of viscous angular momentum transport ν_{out}, in order

N	ν_{in}	ν_{out}	Corresponding α
10^5	0	3.01×10^{-4}	0.019
10^5	10^{-5}	3.06×10^{-4}	0.020
10^5	10^{-4}	3.45×10^{-4}	0.022
10^5	10^{-3}	6.92×10^{-4}	0.045
10^6	0	1.29×10^{-4}	0.008
10^6	10^{-5}	1.38×10^{-4}	0.009
10^6	10^{-4}	2.25×10^{-4}	0.015
10^6	10^{-3}	1.05×10^{-3}	0.068

Table 2.1: Summary of ring spreading tests. ν_{in} denotes the magnitude of the imposed kinematic shear viscosity, and is related to the kinematic viscosity η in Equation 1.80 via the density ρ, with $\nu = \eta/\rho$. ν_{out} is the measured viscosity in the SPH runs, calculated by fitting the viscous time (Equation 1.22) to the time in the SPH runs. The effective α values are calculated by assuming that $\nu_{out} = \alpha c_{s0}H_0$, where subscript 0 denotes values at $R = R_0$. For very small values of ν_{in} the artificial viscosity is the dominant source of angular momentum transport, but for $\nu_{in} \gtrsim 10^{-5}$ the measured viscosity increases as expected.

to determine the accuracy of the SPH viscosity. I further parametrize the viscosity in terms of an effective Shakura and Sunyaev (1973) α parameter by assuming that $\nu = \alpha c_{s0}H_0$, where 0 indicates values at R_0, and the measured values are given in Table 2.1.

From these results I was able to estimate the true level of angular momentum transport present in a full disc simulation. Table 2.1 shows that for both the 10^5- and 10^6-particle runs, there is essentially no difference in the measured viscosity between the runs with $\nu_{in} = 0$ (i.e., artificial viscosity only) and $\nu_{in} = 10^{-5}$. The viscosity in the higher-resolution runs is smaller than that in the lower-resolution runs by a factor of approximately $10^{1/3} \simeq 2.15$, suggesting that in this regime artificial transport of angular momentum is dominant. For larger values of ν_{in}, however, the angular momentum transport increases as expected, showing that the imposed Navier-Stokes viscosity is the dominant source of angular momentum transport for $\nu_{in} \gtrsim 10^{-5}$ (or, equivalently, $\alpha \gtrsim 0.008$). An input α of 0.01 gives a value of ν at R_0 of 1.5×10^{-4}.

The SPH smoothing lengths throughout the simulated discs described in Section 2.3 are comparable to those in the 10^6-particle spreading rings at radius R_0, being of order 0.01 in code units, and so I use this set of rings for comparison. This suggests that the artificial viscosity sets a floor to the effective viscosity in the SPH simulations, approximately at or slightly below a canonical imposed value of $\alpha = 0.01$. I am therefore satisfied that artificial transport of angular momentum does not dominate the viscosity in the disc

models[2]. Moreover, it is known that in the case of a shearing disc, the SPH artificial viscosity behaves similarly to a Shakura-Sunyaev α viscosity (Murray, 1996). Consequently, although the SPH artificial viscosity prevents me from running simulations with very low disc viscosities, the imposition of an explicit Navier-Stokes viscosity means that the angular momentum transport in the disc can be explicitly controlled. Therefore spurious effects from the artificial viscosity should not dominate the results.

2.3 Simulations of giant planets embedded in discs

2.3.1 Initial conditions & code units

Using the SPH code described in Section 1.4.1 and calibrated as described above, I have performed a suite of simulations of planets embedded in protoplanetary discs. I model the star and planet as point masses, and the disc with 10^7 SPH particles. The main simulations and the differences in parameters between them are given in Table 2.2 The number of nearest neighbours used for the SPH was set to 50, and smoothing lengths adjusted accordingly when this changes by ± 2. The Courant parameter used to determine the maximum permitted time-step for SPH particles was set to 0.1. The gravitational softening length for the point mass particles was in each case set to be the same as the sink radius for the planet particle. For the star particle, the sink radius was set to 0.4 times the initial semimajor axis of the planet, and for the planet this was set to 0.4 of its Hill radius, given in Equation 1.59.

I use a a system of units (mass M_0, distance R_0 and time T_0) such that for a planet mass M_p and a stellar mass $M_\star$, $M_p + M_\star = M_0$. The unit of time T_0 is the Keplerian orbital period for a semimajor axis R_0. This choice of units fixes the gravitational constant to be $G = 4\pi^2$. This is particularly convenient as for a mass $M_0 = M_\odot$ and radius $R_0 = 1$au it gives $T_0 = 1$ year.

The initial conditions consist of a gas disc which is axisymmetric about the centre of mass and extends radially from 0.4 to 6 R_0. It has a power-law surface density such that

$$\Sigma(R) = \Sigma_0 \left(\frac{R}{R_0}\right)^{-\gamma} \tag{2.5}$$

where $\Sigma_0 = \Sigma(R_0)$ is a reference surface density used for normalisation. The viscosity

---

[2] The exception is run PNM (see Table 2.2), which uses a much lower explicit viscosity. In this case the artificial viscosity is expected to dominate the angular momentum transport.

Σ_0 [Code Units] [a]	Σ_0 [g cm^{-2}] [b]	γ	q	Model Name
1.12×10^{-5}	10^2	1	0.005	LOW5
1.10×10^{-5}	10^2	1	0.025	LOW25
1.12×10^{-4}	10^3	1	0.005	HIGH5
1.10×10^{-4}	10^3	1	0.025	HIGH25
1.10×10^{-4}	10^3	0	0.025	FLAT
7.03×10^{-4}	6.4×10^3	0	0.025	PNM [c]
7.03×10^{-4}	6.4×10^3	1	0.025	PNMSLOPE

Table 2.2: Summary of simulation runs and the parameters used.

[a] As the unit of mass depends on the planet mass M_P, different values of q give a different Σ_0 for the same physical model.

[b] These values correspond to a $M_\star = M_\odot$ and an initial semimajor axis of 1 au.

[c] Corresponds to the disc model used in PNM01. For this model the viscosity ν was 1.59×10^{-6} in dimensionless units, far less than the artificial viscosity (Sections 2.2). This was set in error, and should have been 2.68×10^{-5} in our units to match that used in PNM01.

ν which goes in to equation 1.80 for each disc model (barring that used in run PNM, see Table 2.2) was chosen to make $\nu\Sigma$ constant (i.e., so that the disc is a steady-state accretion disc), and normalized such that $\nu_0 = 0.01 c_{s0} H_0$ (where the subscript 0 indicates values at R_0). This treatment reduces to a Shakura and Sunyaev (1973) alpha-prescription, with $\alpha = 0.01$, in the canonical case of a $\Sigma \propto R^{-1}$ surface density profile.

The vertical scale-height H is determined by imposing a locally isothermal equation of state, where the temperature $T_{\mathrm{iso}} \propto R^{-1/2}$. This gives the disc an aspect ratio that goes as $H/R \propto R^{1/4}$, normalized so that $H/R = 0.05$ at R_0.

The initial particle positions were created by randomly distributing particles in the radial and azimuthal directions according to Equation 2.5. Vertical positions were generated by randomly sampling a Gaussian density profile with scale height H. Radial and vertical velocities were set to zero, and azimuthal velocities set according to Equation 1.20.

The planet masses used were $M_{\mathrm{p}} = 5\,M_{\mathrm{Jup}}$ and $25\,M_{\mathrm{Jup}}$. Those in the full simulation runs described in Table 2.2 were started on initially circular orbits. Two additional runs, described in Section 2.3.2, used the $25\,M_{\mathrm{Jup}}$ planet with an initial eccentricity $e_0 = 0.05$, in order to test whether the artificial viscosity scheme causes spurious numerical eccentricity damping.
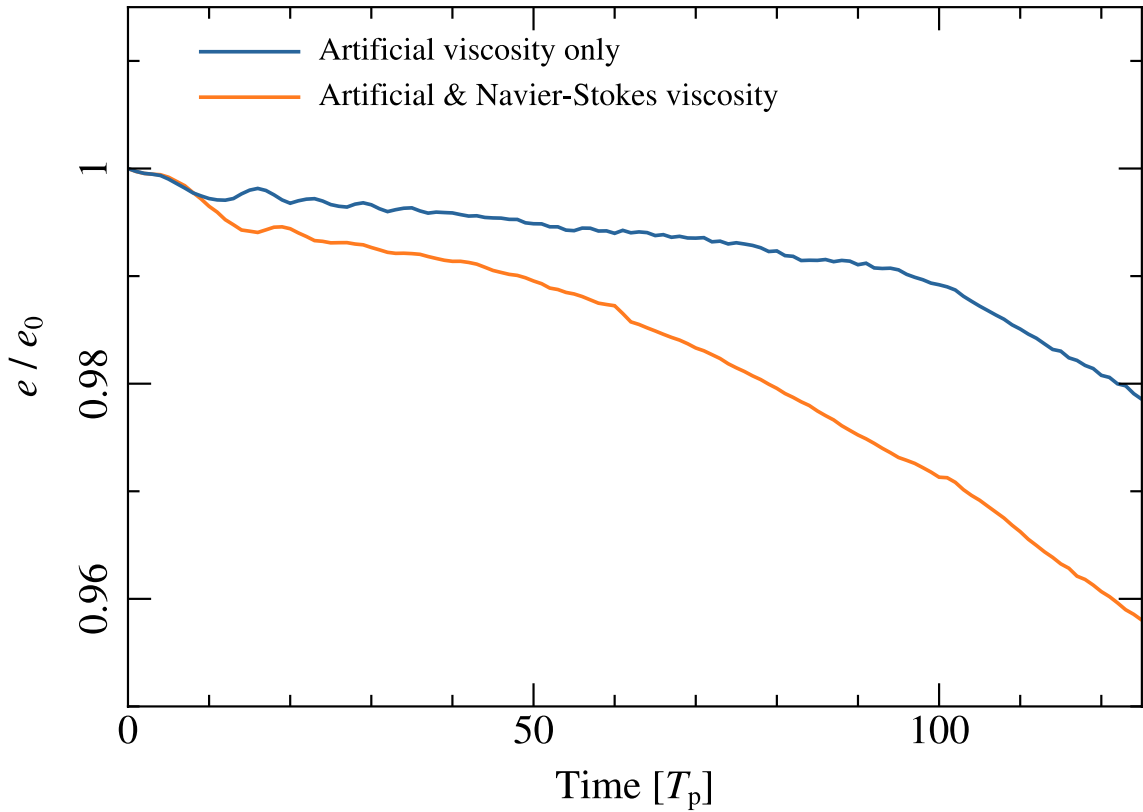
Figure 2.3: Comparison between initially eccentric models with just artificial viscosity (blue line) and with both artificial viscosity and a Navier-Stokes viscosity (orange line). Both used disc LOW25 (see Table 2.2) and were given an initial eccentricity $e_0 = 0.05$. The model without the Navier-Stokes viscosity shows small initial damping of eccentricity which soon flattens off, while the model with the full viscosity scheme implemented sees continued damping. This shows that the SPH artificial viscosity is not causing spurious eccentricity damping. In the case of the full Navier-Stokes viscosity model, at later times the eccentricity decay reversed and began to grow again. This is due to the eccentric planet causing stronger disc eccentricity, and is in agreement with the findings of D'Angelo et al. (2006).

2.3.2   Numerical eccentricity damping

As an initial test, and to ensure that the results of the simulations are physical, I first verified that a planet on an eccentric orbit does not undergo spurious eccentricity damping due to the SPH artificial viscosity (or other numerical effects). To this end I ran 2 realisations the disc model LOW25, where in each case the planet was given an initial eccentricity $e_0 = 0.05$. In one version of this model the Navier-Stokes prescription described in Section 1.4.1.2 was switched off, so that in this case the only source of angular momentum transport was from the artificial viscosity. These were allowed to evolve for 125 orbits, and the eccentricity evolution is shown in Figure 2.3.

The model with no physical (Navier-Stokes) viscosity sees very little eccentricity decay during the initial period while the disc settles into an equilibrium state and the planet opens its full gap. A loss of some eccentricity during this phase is fully expected and is in agreement with simulations by Bitsch and Kley (2010), who found that for non-gap opening planets eccentricity is damped by the surrounding gas. With the physical viscosity switched on, additional damping of eccentricity occurs, at a much more pronounced level. Both cases display exponential damping after the initial phase (beyond ∼ 75 orbits), after the planet has fully cleared its gap. The rate of eccentricity damping during this phase is similar between the two models. This is because once the gap has fully formed, the planet is not directly interacting with the gas to any great extent so the angular momentum exchange here is due to gravitational resonances. At later times, the eccentricity began to rise again in the case of the full viscosity model. This is expected, as D'Angelo et al. (2006) found that even for very low planet masses an initially eccentric planet can undergo far stronger eccentricity growth than one on an initially circular orbit. Consequently I conclude that the SPH artificial viscosity is not causing significant spurious eccentricity damping in our disc models.

2.3.3   Reproducing PNM01 results

As a further test, I have also attempted to reproduce the results of PNM01. To this end I have created a disc model that is as near as possible in form to that used in their calculations[3]. Using the parameters for run PNM given in Table 2.2, this approximates run N4 from that paper with the obvious caveat that these simulations are in 3D and use SPH. Unlike the other models described in Table 2.2, the normalisation for the Navier-Stokes viscosity was taken to be 1.59×10^{-6} in dimensionless code units[4]. This is constant across the disc, fulfilling the steady-state accretion requirement that $\nu\Sigma$ be constant. Note, however, that with this setup the angular momentum transport due to the explicit viscosity is smaller than that due to the SPH artificial viscosity (see Section 2.2), so in practice this test calculation is somewhat more viscous than that of PNM01 (by a factor of $2-3$). This model was allowed to evolve for 340 orbital periods. A series of surface density maps as the disc evolves are shown in Figure 2.4, and the evolution of the planet's eccentricity is shown in Figure 2.5.

The evolution of the disc structure is broadly in line with that found by PNM01, with

---

[3] PNM01 used a 2D fixed-grid code for their simulations, so it is not possible to run a completely identical simulation.

[4] Equivalent to 2.50×10^{-6} in the units of PNM01.
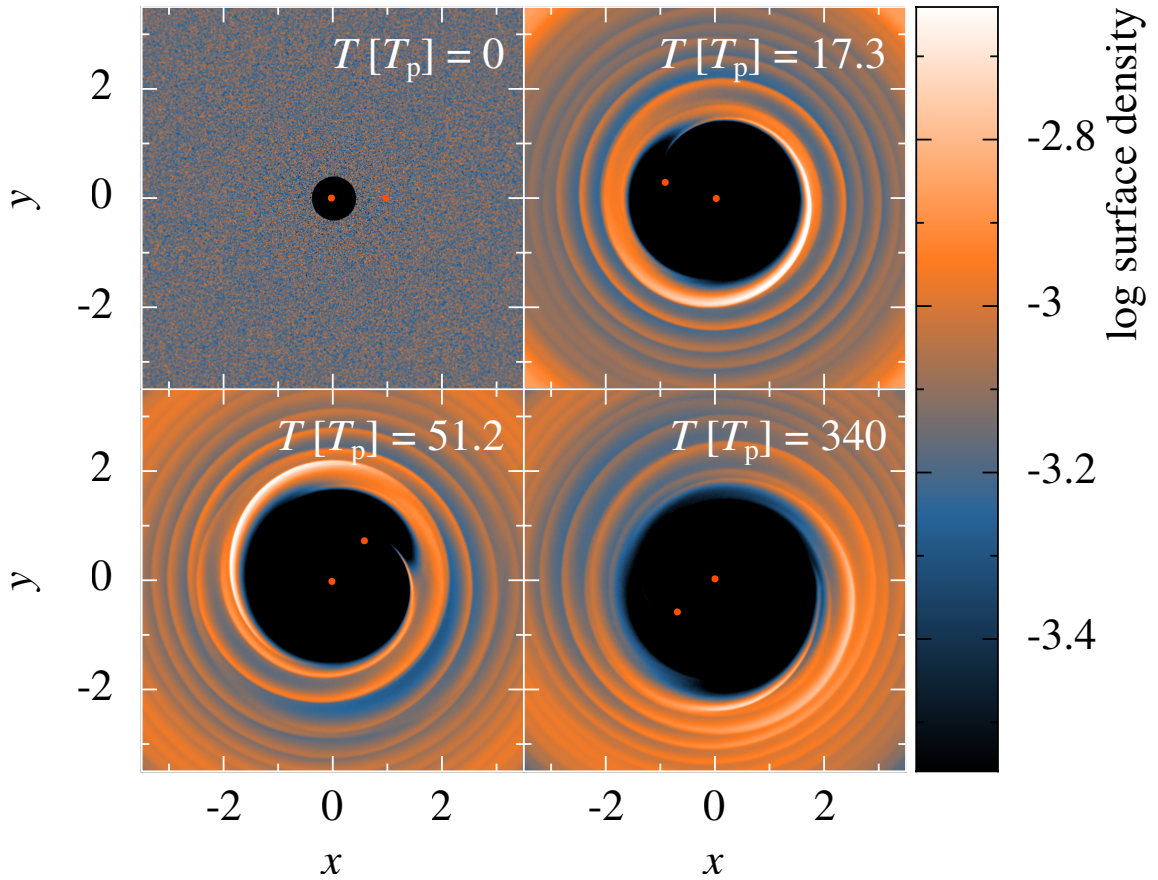
Figure 2.4: Surface density evolution of the central region of my disc model PNM (see Table 2.2), roughly equivalent to run N4 from PNM01. The distance unit is equal to the initial separation between the star and planet. Times shown are in units of the initial orbital period of the planet. The eccentricity evolution for this model is shown in Figure 2.5. After the inner disc clears (upper panels), the presence of the planet drives the inner edge of the disc eccentric (bottom panels). As the disc evolves it exerts torques back upon the planet, causing the planet's eccentricity to grow.

---

the planet rapidly opening a wide gap in the disc, and the inner part of the disc quickly accreting onto the central star. As the system evolves the planet begins to drive eccentricity in the disc at its inner edge, while its own orbit remains essentially circular. At later times this is no longer the case and the planet's orbit becomes significantly eccentric [above the ~ 0.01-0.05 level required by Ogilvie and Lubow (2003) for non-coorbital corotation resonances to saturate, at which point further eccentricity growth is expected]. The long-period oscillations in eccentricity seen in Figure 2.5 are due to the relative precession of the planet and the eccentric disc inner edge of the disc.

The level of eccentricity growth seen in our simulations is somewhat less than that seen by PNM01, but it is still comparable. Moreover, given the larger effective disc
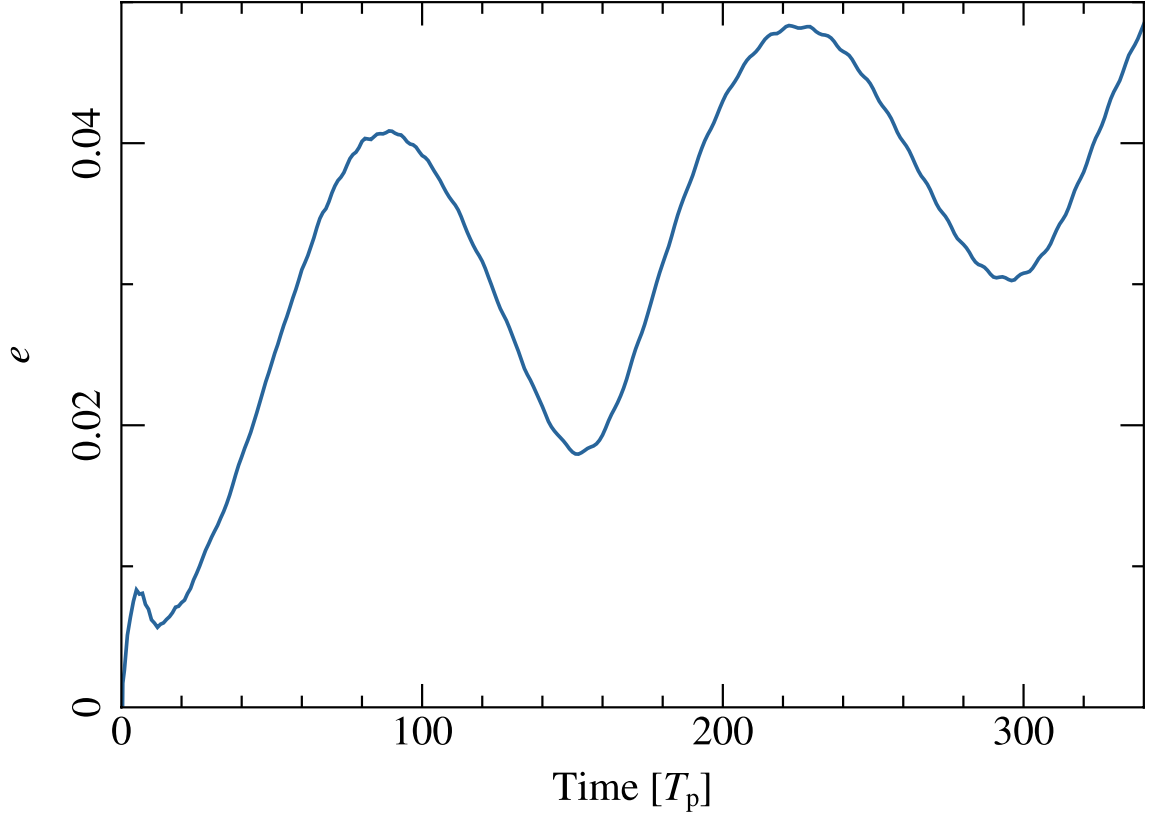
Figure 2.5: Evolution of planet eccentricity for the disc model PNM, which is approximately equivalent to run N4 from PNM01 (see Table 2.2). I find growth of eccentricity in general agreement with that paper. Surface density plots from this run are shown in Figure 2.4. The ~ 100 orbital period oscillations are due to the relative precession of the planet and the eccentric inner edge of the disc.

viscosity, and the inherent differences between the methods (2D fixed-grid versus full 3D SPH calculation), exact agreement is not to be expected. I also note that these simulations are extremely computationally expensive (using up to approximately 150,000 CPU hours per run), so the length for which the runs could be evolved for is limited. Consequently I have been unable to find a level of eccentricity at which growth saturates (the eccentricity was still growing at the end of this simulation), but otherwise there is good agreement between this 3D result and the 2D simulations of PNM01.
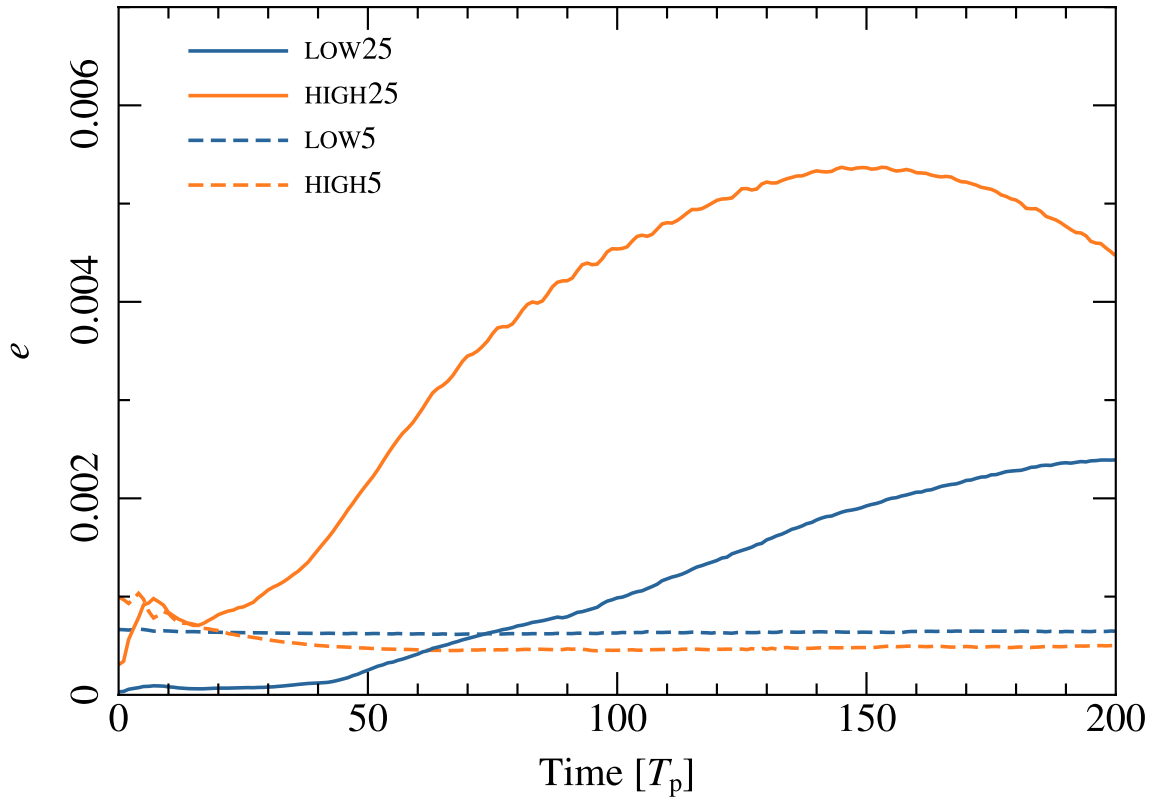
Figure 2.6: Eccentricity evolution for planets of different masses in disc models with different surface densities. LOW5 and HIGH5 (dashed lines, blue and orange respectively) have a planet-star mass ratio of $q = 0.005$, while LOW25 and HIGH25 (solid lines, blue and orange) have $q = 0.025$. LOW and HIGH refer to the choice of disc surface density – see Table 2.2 for the values – but all have a power-law index of $\gamma = 1$.

2.4   Main results

2.4.1   The effect of the planet mass

To test the effect of different planet masses, I ran models with both LOW and HIGH surface densities (see Table 2.2) with planet-star mass ratios of $q = 0.005$ & 0.025. For a $1M_\odot$ star this corresponds to planet masses of approximately 5 and $25M_{\mathrm{Jup}}$ respectively. Figure 2.6 shows the time evolution of the orbital eccentricity in each case. For both of the models with $q = 0.005$ (Low5 and High5), no eccentricity growth was seen at any level. For the models with $q = 0.025$ (LOW25 and HIGH25) some modest growth was seen, but none above $e = 0.005$.

This is consistent with the conclusions of both PNM01 and D'Angelo et al. (2006), who found that eccentricity is first induced in the disc, driven by the outer 3:1 Lindblad

resonance. In these simulations this disc eccentricity clearly develops in both runs with $q = 0.025$, but not in either of the lower mass planet cases. A comparison of the surface density maps for both planet masses in the LOW model is shown in Figure 2.7. The reason for the lack of eccentricity growth in both of the $q = 0.005$ models is that the 3:1 outer Lindblad resonance induced by the planet is too weak to affect the disc structure significantly on the time-scale of the simulations. The difference in eccentricity evolution between models LOW25 and HIGH25 arises because the higher surface density disc is simply more massive, and consequently is able to exert stronger torques upon the planet, resulting in more (but still extremely limited) eccentricity growth.

In the case of both planet masses, eccentric co-rotation resonances (that is, CRs which are not co-rotational with the planet but which co-rotate with some component of the potential; see Section 1.3) are resolved where present. Using the width formula provided by Masset and Ogilvie (2004), the smoothing lengths are smaller than the resonant widths at the nominal resonant locations by a factor of at least two, for resonances that are not fully in the open gap. For the case of the 5 M_{Jup} planet, this is assuming an eccentricity of 10^{-2}, as for the eccentricity measured from the simulations the prescribed widths are vanishing.

2.4.2   The effect of the radial surface density profile

I now consider the effect of the disc surface density profile on the evolution of the embedded planet. To this end I have run two additional models with the same star-planet mass ratio of $q = 0.025$, and different radial surface density profiles ($\gamma = 1$ & 0). The disc models were FLAT and PNMSLOPE (see table 2.2), and each was evolved for 200 orbits. The evolution of the orbital eccentricity for these runs, along with two previously described (HIGH25 and PNM, for the purposes of comparison), are shown in Figure 2.8. Again, due to computational limitations these models did not run for long enough to determine at what level of eccentricity growth saturates; I am instead more concerned here with determining the conditions under which growth will occur. An eccentric inner disc is again driven by the presence of the massive planet, as described above, and its effect on the eccentricity of the planet depends strongly on the precise disc model used.

The results from these models show two things. First, the magnitude of the disc surface density has a strong effect on the eccentricity evolution of an embedded planet. Figure 2.8 clearly shows that the two models with $\Sigma_0 = 7.03 \times 10^{-4}$ (PNM and PNMSLOPE) see significant eccentricity growth, while the models with $\Sigma_0 = 1.10 \times 10^{-4}$ (HIGH25 and FLAT) do not. This suggests that eccentricity can only be excited above a threshold surface
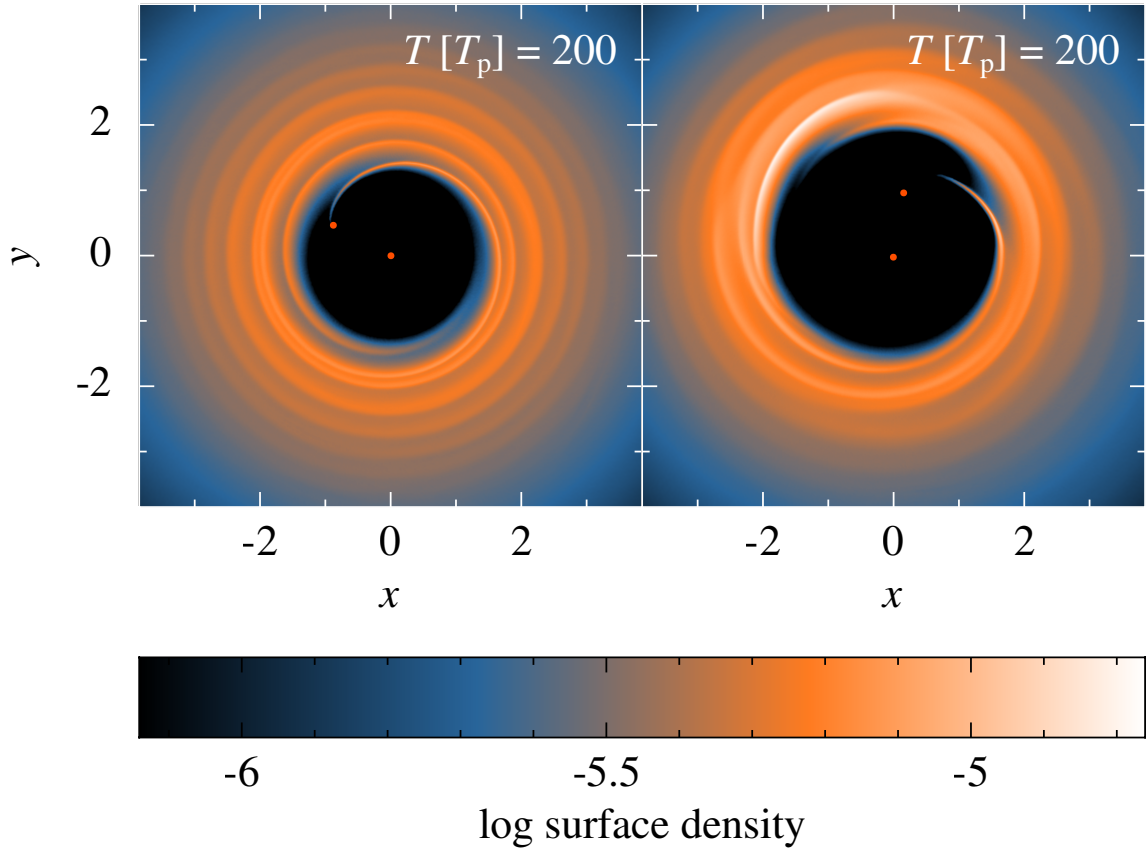
Figure 2.7: Surface density maps after 200 planetary orbits:. The left panel shows the central region of model LOW5 ($q = 0.005$) and the right panel shows LOW25 ($q = 0.025$; see table 2.2). It is clear from the left panel that the lower mass planet has a far smaller perturbing effect on its host disc, driving spiral waves but not further disrupting the disc shape or structure. This is in stark contrast to the right hand panel, where the higher mass planet has driven the inner edge of the disc quite eccentric.

---

density. This behaviour was suggested by PNM01 but not investigated in detail. Comparing the ratio between the planet mass and the local disc mass (approximated by $\Sigma \pi a^2$ in the unperturbed disc) suggests that values between $\sim 1 - 13.5$ may result in eccentricity growth (see Figure 2.9). I also note in passing that this is a surprisingly strong effect for a relatively modest (factor of 6.4) change in the disc surface density; the surface densities in real protoplanetary discs are thought to change by factors $\gtrsim 10^3$ over their lifetimes (e.g. Hartmann et al., 1998; Alexander et al., 2006b).

The power-law index γ also has quite a strong effect on the level of eccentricity growth seen. The two models with $\gamma = 1$ (HIGH25 and PNMSLOPE) show slower, weaker growth of eccentricity than their counterparts with the same normalisation value of Σ_0 but a flat $\gamma = 0$ radial power-law dependance. There are two primary reasons for this. Firstly,
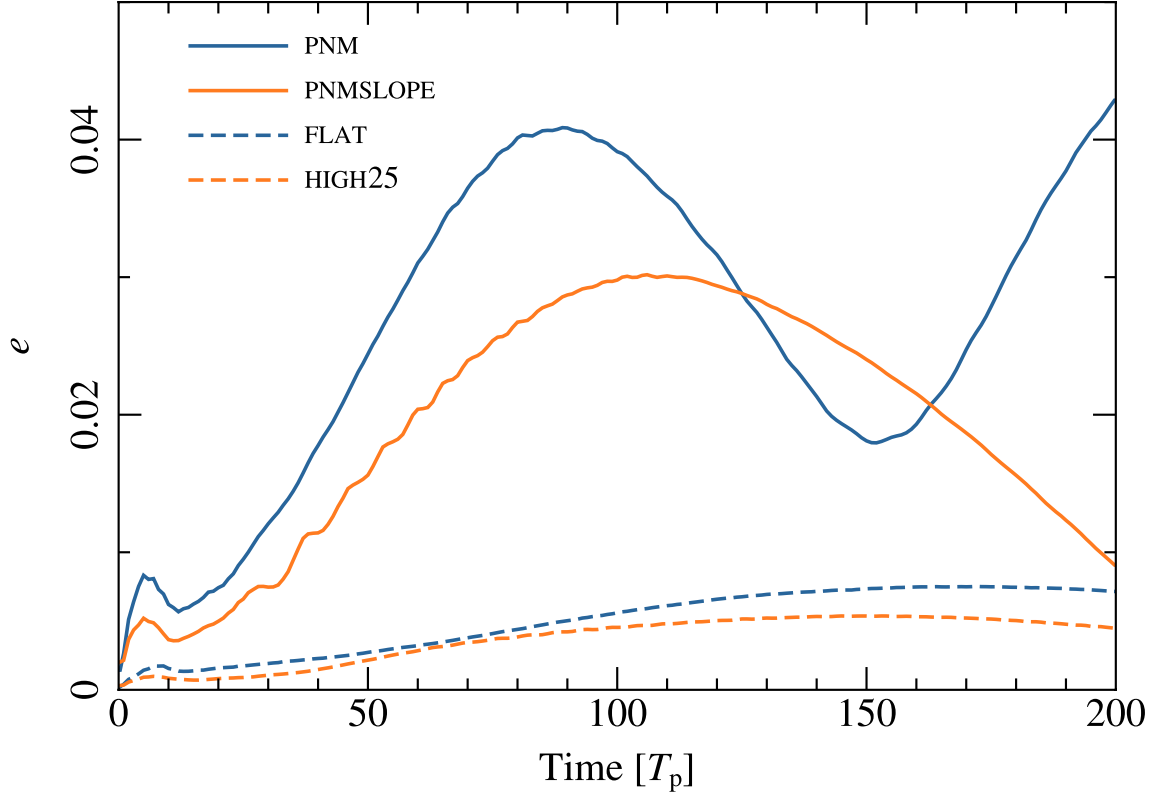
Figure 2.8: Eccentricity evolution for disc models with different surface density profiles. The solid blue line is the same as in Figure 2.5, truncated at $T = 200$ for reference. The dashed orange line is the same as the solid orange line in Figure 2.6. The solid orange line traces the eccentricity of model PNMSLOPE, and the dashed blue line is run FLAT. These models show that while higher values of Σ_0 give consistently stronger eccentricity growth, shallower radial profiles (lower values of γ, see table 2.2) also show stronger growth. The oscillations in the curves are again due to the relative precession of the planet and the eccentric disc. Note also that model PNMSLOPE was run for a further 100 orbits to verify that the eccentricity does grow indeed continue to grow; the declining eccentricity at $T = 200$ is simply due to this precession effect.

---

a flatter radial profile puts less mass inside the planet's orbit. The inner disc therefore accretes on to the star more rapidly for flatter surface density profiles (i.e. lower values of γ), and as eccentricity only begins to grow considerably after the inner disc has accreted, this takes place sooner for flatter surface density profiles. A flatter radial profile also puts more mass into the outer Lindblad resonances, which are responsible for the torques that drive eccentricity growth. This changes the torque balance on the planet, and gives rise to more rapid eccentricity growth. In addition, in the special case when Lindblad torques cancel (as proposed by Goldreich and Sari 2003), the resultant net torque is a function of the surface density gradient, with a steeper power-law dependance damping eccentricity. However, this effect is not expected to act in these models, as the planets
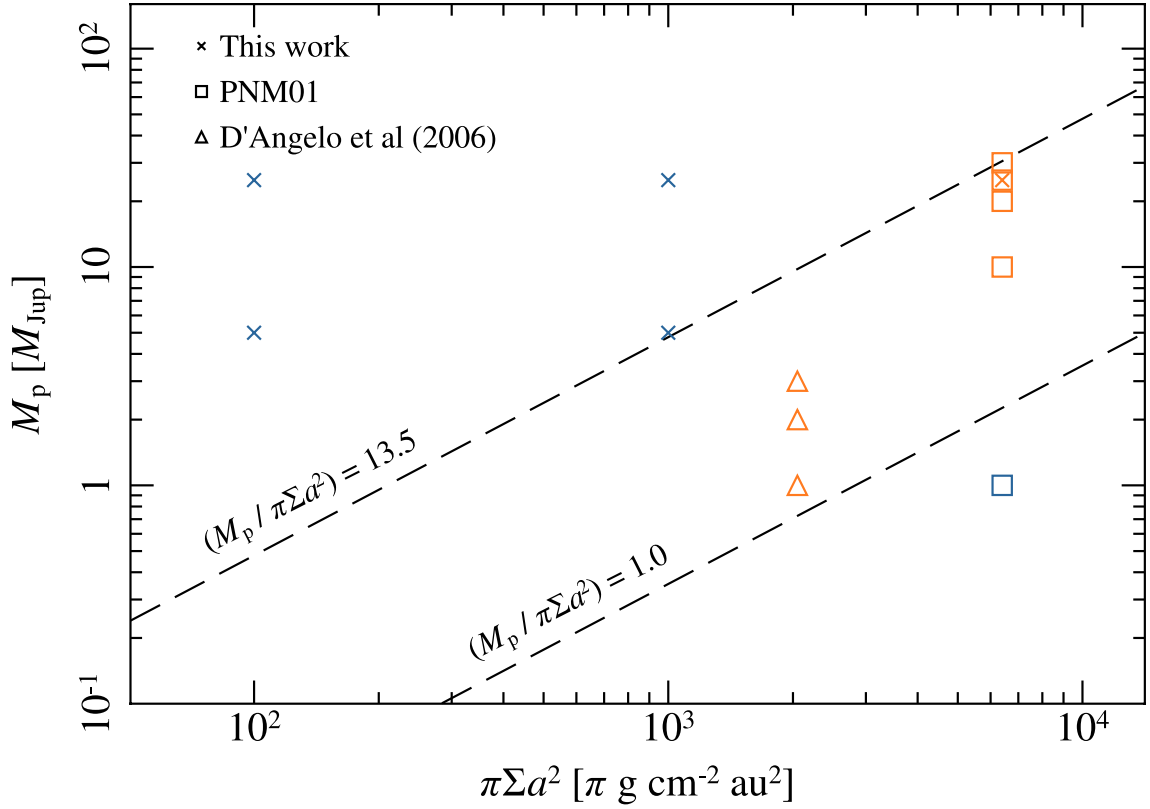
Figure 2.9: Comparison between the local unperturbed disc mass (approximated by $\pi\Sigma a^2$) and planet mass for simulations presented here and in two other papers. Crosses indicate simulations from this chapter, squares from PNM01 and triangles from D'Angelo et al. (2006). Orange symbols indicate that eccentricity growth was seen, while blue indicates that it was not. The choice of units on the x-axis gives the reference surface density at the semimajor axis of the planet for that model (except for the models of D'Angelo et al. (2006), where it is 5.2^2 times that value). The dashed lines show where the ratio of $M_p/\pi\Sigma a^2$ cross values of 1.0 and 13.5. I suggest that these values represent cases where the planet is massive enough to significantly perturb the inner edge of the disc, but where the disc is also massive enough that it can exert sufficiently strong torques upon the planet, although one should note that this ratio is not the lone deciding factor. I also draw attention to the fact that the single point below the threshold $M_p/\pi\Sigma a^2 = 1.0$ has been questioned by Masset and Ogilvie (2004) as being affected by spurious numerical factors, and thus the lower limit may not be real.

---

open gaps sufficiently wide that co-rotation resonances are ineffective.

To investigate this effect further I follow the method of Artymowicz et al. (1991) to calculate values of $\dot{e}$, time averaged over several orbits at the end of each simulation. The radial contributions to this, normalized against the magnitude of the surface density in each model, are shown in figure 2.10. For all runs with $q = 0.025$, in the low-surface density limit (models LOW25, HIGH25 and FLAT), the eccentricity is being damped by a peak at $R \sim 1.8$, and the magnitude of the surface density acts as a scaling factor with only a

very weak dependence on the radial slope γ. In these three disc cases, the net value of $\dot{e}$ summed over the disc radius is negative. By contrast, In the models with very high surface densities, where eccentricity is growing (models PNM and pnmslope), the sense of the contribution reverses, and the total $\dot{e}$ is positive. Here there is also a more pronounced effect of the radial gradient in the surface density. For the runs with $q = 0.005$ (LOW5 and HIGH5), the net $\dot{e}$ is always negative. This is unsurprising, as neither of these models saw any growth of eccentricity (Figure 2.6.

It is unclear exactly what mechanism causes the change in the sense of $\dot{e}$ for high surface densities, but it appears that there is a threshold surface density above which the analysis performed in the low disc mass limit (e.g. Goldreich and Tremaine, 1980) no longer applies. These results, and those of previous simulations (see Figure 2.9), suggest that this threshold is given by

$$\pi \Sigma a^2 > M_{\mathrm{p}}/C \tag{2.6}$$

where C is a constant with a value $C \sim 10$. Discs with surface densities below this threshold are unable to excite significant eccentricity in the planet's orbit. This behaviour was suggested by PNM01, and these results support their tentative prediction. The threshold must presumably depend on several other factors (disc viscosity, H/R, etc.), but a complete exploration of this parameter space is beyond the scope of this chapter. Nevertheless, the results of these simulations show that eccentricity is only excited in discs with very high surface densities.

2.5 Discussion

2.5.1 Numerical limitations

The biggest potential numerical problem with this work is that the SPH artificial viscosity may cause spurious eccentricity damping. Artificial viscosity can be especially problematic for shearing-disc type problems such as this, as the differential rotation is often mistaken by the algorithm for a shock. However, given the high resolution of the simulations and results of the tests presented in Sections 2.2 and 2.3.2 I am confident that the SPH artificial viscosity has not had a significant influence on the results.

I further note that the approximation of a locally isothermal equation of state is somewhat idealised, and in particular may not give an accurate description of the spiral density waves induced in the disc. Bitsch and Kley (2010) looked at the evolution of initially eccentric planets in fully radiative discs, and made comparisons to models which used an
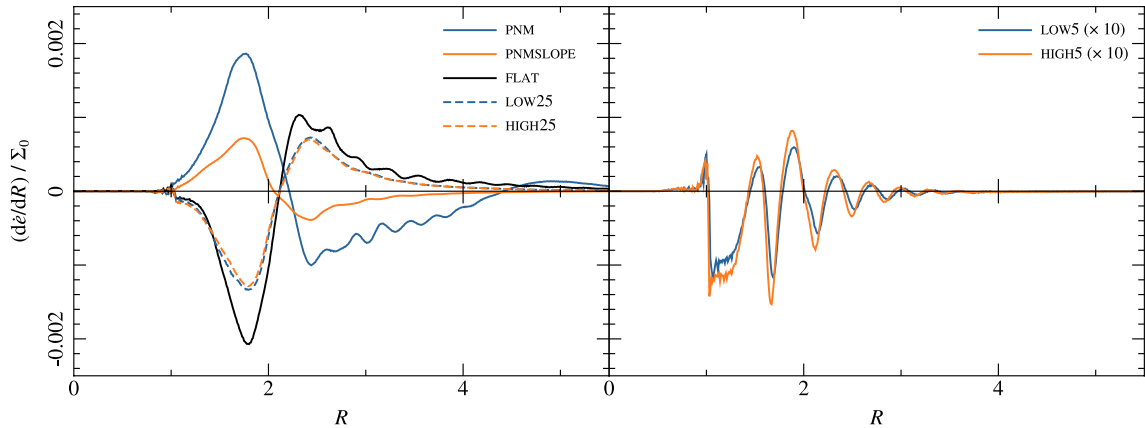
Figure 2.10: Radial contributions to $\dot{e}$, calculated using a Gaussian perturbation approach following the method of Artymowicz et al. (1991), and normalized to the disc surface density. These values are time-averaged over 5 orbital periods of the planet at the end of each simulation. The left hand panel shows all models with high-mass planets ($q = 0.025$). Models LOW25, HIGH25 and FLAT clearly show that in the low-surface density limit, $\dot{e}$ is approximately linear with the magnitude of the surface density, with a weak dependance on the radial profile γ. For models PNM and PNMSLOPE which have very high surface densities that lie above the threshold for eccentricity growth, the sign of the curves flip and the net $\dot{e}$ becomes positive. In this limit the radial profile of the surface density becomes more important, and the effect is no longer linear with its magnitude. The right panel shows the runs with low-mass planets ($q = 0.005$), LOW5 and HIGH5, with the amplitude on the y-axis multiplied by a factor of 10. In both of these cases, the net values of $\dot{e}$ are negative, as neither run saw any level of eccentricity growth.

---

isothermal approximation. They only found the results to be inconsistent for relatively low planet masses ($M_\mathrm{p} \lesssim 0.6\,M_\mathrm{Jup}$). The planet masses considered here are far above this, well into the regime where the isothermal approximation holds, and so I do not expect the use of an isothermal equation of state to introduce significant uncertainties in the results.

It must also be borne in mind that the Navier-Stokes viscosity used in the models is merely a first-order approximation, attempting to mimic the effect globally of a process that occurs far below the scales I am able to resolve – namely the MRI thought to drive angular momentum transport in protoplanetary discs (see Section 1.2.2; Balbus and Hawley, 1991). By adopting a Shakura and Sunyaev (1973) alpha-prescription I implicitly assume that the small-scale effects of turbulence in the disc behave like a viscosity on large scales, but it is not clear whether this approximation holds for length-scales $\lesssim H$. In the simulations in this chapter the planets open gaps on length-scales $\gtrsim H$, so turbulent fluctuations on smaller scales are unlikely to have a strong effect. However, there is some overlap between the length-scales considered here and the typical scales of MRI turbulence, so I note that these results may not hold if the MRI drives significant power on moderate or large scales (as suggested by recent simulations; Simon et al., 2012)). Detailed investiga-

tion of this issue is beyond the scope of this thesis, but if the viscous approximation does break down at scales $\sim H$ then it seems likely that eccentricity growth could be affected, particularly for low-mass planets.

2.5.2 Applications to real systems

The major result from these simulations is that resonant torques are not generally an efficient means of exciting eccentricity, and that planet-disc interactions are unlikely to be responsible for the eccentricities seen in the majority of (giant) exoplanet systems. While I do find eccentricity growth in agreement with PNM01, I note that their calculations considered a very massive planet in a disc with a very high surface density. Their reference surface density of $6.4 \times 10^3 \mathrm{g/cm}^2$ at 1 au is a factor of several larger than predicted by the Minimum Mass Solar Nebula (Weidenschilling, 1977) or more realistic accretion disc models (e.g., Hartmann et al., 1998). I have found that a modest reduction in the disc surface density results in no significant eccentricity growth for similarly massive planets, in agreement with behaviour suggested by PNM01. Consequently it is unlikely that this mechanism will be able to excite eccentricity in real protoplanetary discs.

I have also failed to find eccentricity growth for lower planet masses, but in this case the results must be with more caution. Using 2D simulations D'Angelo et al. (2006) found eccentricity growth up to $e \sim 0.1$ for planets between 1-3 M_{Jup}, but typically this occurred on time-scales of thousands of orbits. Given the high computational cost of the 3D simulations presented here I am not able to follow their evolution for such long time-scales, and consequently cannot rule out this slower mode of growth. However, I note that the disc model used by D'Angelo et al. (2006) also uses a very large surface density: if one re-scales their model to match the units used here, their surface density normalisation (Σ_0) becomes 2.0×10^3 g/cm^2 at 1 au, larger than in the HIGH models from this chapter. I also note that their relatively flat choice surface density profile ($\gamma = 1/2$) is likely to promote eccentricity growth compared with steeper values. Although the SPH simulations in this chapter do not rule out growth on very long time-scales, it seems likely that differences in the choice of disc model are responsible for the apparent discrepancy between these results and those of D'Angelo et al. (2006).

As noted in Section 2.4.2, there are two reasons why the radial surface density profile plays a role in exciting eccentricity growth. At a basic level, a lower value of γ puts more mass into the outer Lindblad resonances, which exert a net torque that causes eccentricity to grow, and conversely less mass into the inner resonances which damp eccentricity (Tanaka et al., 2002). In the case of lower mass planets, where a gap is not fully opened,

there is another effect which takes place that brings the radial surface density gradient into play. Goldreich and Sari (2003) suggest that in a near-Keplerian disc where an outer and inner Lindblad resonance cancel to a reasonable approximation, the resulting net torque scales with $d\Sigma/dr$ rather than with Σ. This implies that not only the Σ_0 level but also the radial profile may become as important as the planet mass in discerning the physical contribution of resonant torques for low mass planets. This latter effect is unlikely to play a role in the simulations presented here as the gaps opened by the planets are very wide. It is therefore more likely the former effect at work with both the surface density profile and its normalisation level both having a strong effect on if, and how, the orbital eccentricity of an embedded planet will grow (Figure 2.8).

This analysis of the effect of both the magnitude and radial profile of the surface density is supported by looking at the radial contributions to the change in eccentricity in different models, shown in figure 2.10. Below the threshold for eccentricity growth the surface density profile has only a small effect, and its magnitude is an approximately linear scaling factor, which is in agreement with the findings of Artymowicz et al. (1991). In contrast, above the threshold for growth the sense of the contributions reverse and the difference between models with the same Σ_0 but different γ becomes more pronounced. I am confident that this effect is real, and warrants further study, but detailed investigation is beyond the scope of this chapter. Other disc parameters including viscosity and temperature structure must also have an effect on this result, but I do not investigate them here.

One can however extend this analysis by taking Equation 2.6 to be a necessary condition for eccentricity growth:

$$C\pi\Sigma a^2 \gtrsim M_\mathrm{p} \tag{2.7}$$

where $C \sim 10$ is a constant. However, giant planets continue to accrete via tidal streams even after opening a gap in the disc, and in discs which are sufficiently massive to meet this condition they are likely to accrete rapidly. This accretion increases M_p, making it less likely that the threshold for eccentricity growth will be met. The critical issue is therefore one of time-scales: does eccentricity grow on a shorter time-scale than the planet's mass? For massive giant planets ($\gtrsim 5M_\mathrm{Jup}$) tidal torques strongly suppress accretion from the disc on to the planet (e.g. Lubow et al., 1999), and therefore eccentricity would be expected to grow more rapidly than the planet mass. However, Figure 2.9 suggest that eccentricity growth in this regime requires very high disc surface densities, $\Sigma_0 > 10^3 \mathrm{g\,cm}^{-2}$, and such massive discs are not commonly observed.

By contrast, lower-mass giant planets ($M_\mathrm{p} \sim 1M_\mathrm{Jup}$) accrete very efficiently from their

parent discs (e.g. D'Angelo et al., 2002). The planetary accretion rate can be parametrized as $\dot{M}_p = \epsilon \dot{M}_d$, where $\dot{M}_d = 3\pi \nu \Sigma$ is the disc accretion rate in the absence of the planet. ϵ represents the efficiency of accretion on to the planet; simulations show that this efficiency has a peak value of $\epsilon \sim 1$ for planets of approximately Jupiter mass, and declines to higher planet masses (Lubow et al., 1999; D'Angelo et al., 2002, see also Veras and Armitage 2004). Substituting $\nu = \alpha \Omega H^2$ and rearranging gives

$$\frac{M_p}{\dot{M}_p} \lesssim \frac{C}{3\epsilon} \alpha^{-1} \Omega^{-1} \left(\frac{H}{a}\right)^{-2} . \tag{2.8}$$

The quantity on the left hand side is the time-scale for planet growth through accretion of gas from the disc, τ_{acc}, so that meeting the condition for eccentricity growth (Equation 2.6) also sets an upper limit to the time-scale for planet growth by accretion. Taking standard values of $\alpha = 0.01$ and $H/a \simeq H/R = 0.1$, and assuming that $\epsilon \sim 1$ gives

$$\tau_{acc} \lesssim 3 \times 10^4 \, \Omega^{-1}, \tag{2.9}$$

for planets of approximately Jupiter mass. This is comparable to the time-scale for eccentricity growth seen in previous studies of Jupiter-mass planets (D'Angelo et al., 2006). This therefore suggests that in this case the planet would accrete rapidly, and 'migrate' out of the region of allowed eccentricity growth highlighted in Figure 2.9 before it attains a significant eccentricity.

I have shown that in moderately viscous discs (with $\alpha \sim 0.01$), eccentricity growth only occurs if the disc surface density is high. Unfortunately, as touched upon in Chapter 1, observational constraints on the surface densities of discs are extremely weak, and at au scales such as those considered here, almost non-existent. Andrews and Williams (2007) and Andrews et al. (2009, 2010) made systematic studies of protoplanetary discs in Taurus and Ophiuchus, fitting surface density profiles to submillimeter observations of thermal dust emission. However, the angular resolution of such observations means that they can only resolve scales a few tens of au (in the very best cases), and are not sensitive to the region of most interest for planet formation. Extrapolating their results down to the unresolved inner disc, their fits yield values of Σ_0 at au radii between ~ 30 and ~ 1700 g/cm^2, with most values lying at $\sim 700 - 800$ g/cm^2. The radial power-law indices (γ) they fit to their observations range between 0.4 and 1.1, with a clear preference towards the upper end of this range. These correspond approximately to the ʜɪɢʜ disc model from the simulations here, and have both lower surface densities and steeper power-law indices

than the disc models used by PNM01 and D'Angelo et al. (2006). However it must be stressed that extrapolating these sub-mm observations to smaller radii is by no means justified. It has also been suggested that substantial mass reservoirs may exist in 'dead zones' close to the star (e.g. Gammie, 1996; Hartmann et al., 2006; Zhu et al., 2010), but there are currently no useful constraints on disc surface densities at au radii. ALMA may soon provide real constraints on protoplanetary discs with far higher angular resolution, but it will be some time before it is able to probe radii of a few au (e.g. Cossins et al., 2010).

Despite this, I argue that significant eccentricity growth due to the planet-disc interaction is unlikely given realistic protoplanetary disc conditions. There is an increasing consensus in the literature that this is the case – while D'Angelo et al. (2006) did see eccentricity growth by this method, it did not rise above ~ 0.15, lower than the observed values of 0.2–0.3 that this effect is invoked to explain. Similarly, the semi-analytic models Moorhead and Adams (2008) were unable to reproduce the observed low-eccentricity distribution. Coupled with our findings, it seems that the planet-disc interaction alone is incapable of reproducing the eccentricities seen in exoplanet observations. More recently, Bitsch et al. (2013) have performed 3D grid-based simulations of eccentric Jovian planets, and in general also find eccentricity damping for $M_{\rm p} \lesssim 5\,M_{\rm Jup}$.

This negative result of course begs the question of what the true origin of exoplanet eccentricities is. An emerging consensus seems to be that scattering events are responsible for most if not all of the distribution (e.g. Chatterjee et al., 2008). This is backed up by a vast number of tightly-packed multi-planet systems observed by Kepler (Batalha et al., 2013) and by a plethora of planetary systems that have clearly undergone strong interactions (e.g. highly inclined and retrograde planets; Wright et al., 2011). However, there is some suggestion that the low-eccentricity end of the distribution perhaps cannot be explained by this mechanism (Goldreich and Sari, 2003; Jurić and Tremaine, 2008), and further work in this area is still required.

2.6   Conclusions

I have performed high-resolution 3D SPH simulations of giant planets embedded in proto-planetary discs, using the SPH code described in Section 1.4.1.2 and calibrated using 1D viscously spreading rings. For high disc surface densities and planet masses I have found that the planet-disc interaction leads to eccentricity growth, in agreement with previous studies. However, I have also shown that for realistic planet masses and disc properties,

the planet-disc interaction is incapable of exciting significant orbital eccentricity growth.

I have identified a threshold surface density for eccentricity growth, and it is noteworthy that this threshold is rarely, if ever, met in real systems, except in cases where the time-scale for eccentricity growth is comparable to the time-scale for mass accretion by the planet. I conclude from these simulations that in the case of a real giant planet, the interaction with its parent disc is unlikely to yield growth of its orbital eccentricity at measurable levels. Therefore I suggest that the low but non-zero exoplanet eccentricities observed, not accounted for by simulations of planet-planet scattering events, must have some other origin. I have identified a number of possible areas for future work, including exploring how the viscosity and temperature of the disc affects the eccentricity evolution of the planet.

3

Kepler-16b: Evidence for eccentricity damping in a circumbinary disc

T HE previous chapter concerned the eccentricities of giant planets orbiting single stars. I now turn in this chapter to consider circumbinary planets – those orbiting both components of a stellar binary, and how disc interactions may have affected their orbits, with specific application to the planet Kepler-16b.

3.1 Introduction

3.1.1 Planets in binary systems

While planets around single stars have been known to exist for almost 20 years (or considerably longer if one includes our Solar System), planets in binary star systems have only been found much more recently. These planets come in two flavours of orbit, S-type and P-type. S-type planets orbit either the primary or the secondary star in the system, while P-type ('circumbinary') planets orbit both stars. Planets of the former type have been known to exist for some time – for example, υ And A is known to host a multi-planet system, although this was known before its stellar companion was discovered (Butler et al., 1997; Lowrance et al., 2002). Although the number statistics are comparatively low, there seems to be little difference between the eccentricity distribution of S-type stars and those orbiting single stars (Eggenberger et al., 2004).

It has recently been shown that S-type planets can exist in stable orbits previously thought impossible (e.g. the planet orbiting the nearby α Centauri B, part of a binary with eccentricity $e \sim 0.5$; Dumusque et al., 2012, although the significance of the signal in that detection has recently been challenged by Hatzes 2013). Much theoretical work has been put into discovering how such systems form and evolve (e.g. Kley and Nelson, 2008; Müller and Kley, 2012), and they are fascinating systems in their own right. However, for this chapter I shall concentrate purely on circumbinary planets in P-type orbits around both of their host stars.

Even prior to the discovery of a single planet in a circumbinary orbit around a main-sequence star[1], a large body of theoretical work on how they might form and evolve had been built up (e.g. Paardekooper et al., 2008; Pierens and Nelson, 2008; Marzari et al., 2008; Fragner et al., 2011). Largely based on numerical simulations, the broad

---

[1] Planets around non-main sequence binaries have been known for some time. Thorsett et al. (1993) suggested that timing variations in the PSR B1620-26 system, a millisecond pulsar with a white dwarf binary companion, could be explained by a planet in a $\sim$ 10 au orbit. A number of other planets are known to inhabit P-type orbits around evolved (post-main sequence) binaries, such as HW Virginis and NN Serpentis (Lee et al., 2009; Beuermann et al., 2010).

consensus of this work was that planet formation in a circumbinary disc sees significant challenges due to interaction with the binary that are not present for single stars. In particular, planetesimal growth at small semimajor axis (within a factor of a few of the binary semimajor axis) is strongly inhibited by the large velocity dispersion induced by the central binary (Paardekooper et al., 2008; Marzari et al., 2008). Additionally, the binary causes strong perturbations to the disc that can affect the orbit of a fully-formed planet and cause unusual migration patterns not seen for single-star planets (e.g. Pierens and Nelson, 2008).

The recent detection of a number of planets orbiting main-sequence binary stars by the *Kepler* mission (Doyle et al., 2011; Welsh et al., 2012; Orosz et al., 2012a,b; Schwamb et al., 2013) challenges this carefully built theoretical understanding. All of the six systems reported to date contain planets in relatively short-period orbits around close eclipsing binaries (binary semimajor axes $a_b < 0.25$ au, planet semimajor axes $a_p \lesssim 1$ au). While transit surveys such as *Kepler* are naturally biased towards finding short-period systems first, it is noteworthy that these planets are predominantly found just on the edge of dynamical stability (Holman and Wiegert, 1999). Moreover, the fraction of close binaries hosting circumbinary planets is estimated to be $\gtrsim$ 1–10 per cent (Welsh et al., 2012), implying that planets form readily in circumbinary discs.

Four of the seven of the circumbinary planets discovered by *Kepler* to date have low eccentricities, $e_p < 0.05$. Of the exceptions, Kepler-34b and PH1 both orbit eccentric binaries ($e_b \gtrsim 0.2$), while Kepler-47c is the outer planet in a two-planet system and is therefore a special case; the remainder are in near-circular orbits around low- or moderate-eccentricity binaries. Figure 3.1 shows the measured values of e_p plotted against the binary eccentricity e_b for the known *Kepler* circumbinary planets.

At close separations circumbinary orbits are highly non-Keplerian, so directly measuring their eccentricities does not necessarily give a clear picture of the orbits. The planet's measured eccentricity e_p actually comprises the sum of two components, referred to as the forced and free eccentricities (e_{forced} and e_{free}). The forced eccentricity is driven by the potential of the central binary, and only the free eccentricity is a parameter of the planet's orbit alone (Murray and Dermott, 1999). However, an analysis of the orbits of Kepler-16b, Kepler-34b and Kepler-35b by Leung and Lee (2013) found that e_{free} is an order of magnitude higher for Kepler-34b than the others, consistent with its higher measured eccentricity.
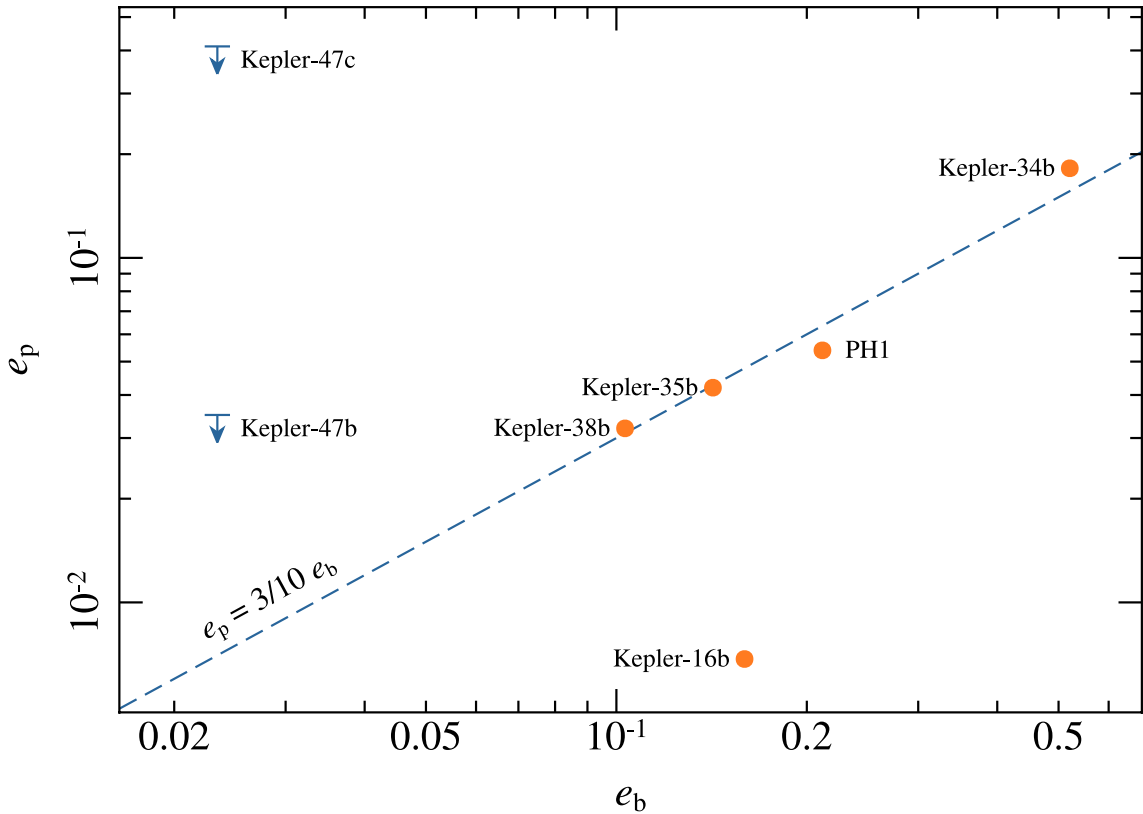
Figure 3.1: Measured planet eccentricities e_p against binary eccentricity e_b for all the reported *Kepler* circumbinary eccentricities. Orange circles indicate quoted values while blue arrows are upper limits. While e_p is in fact a function of the e_b due to gravitational forcing from the binary (Murray and Dermott, 1999), there seems to be a broad trend of increasing e_p with increasing e_b. For reference, the blue dashed line plots the relation $e_p = 3/10\,e_b$. Kepler-16b is noteworthy in that it has e_p almost an order of magnitude less than the median value.

3.1.2   Kepler-16b

Of particular interest is Kepler-16, the first of these systems discovered. It is a closely packed system and is aligned to a very high degree: the planes of the binary and planetary orbits, are aligned to within $0.3°$ (Doyle et al., 2011) and the binary orbit is aligned to the spin of the primary star to within $3°$ (Winn et al., 2011). The total stellar mass is approximately $0.9M_\odot$, with a mass ratio of 0.3, and the binary has eccentricity $e_b = 0.16$. The planet Kepler-16b is approximately the mass of Saturn ($0.333M_{\mathrm{Jup}}$ Doyle et al., 2011; Bender et al., 2012), with a measured eccentricity of $e_p = 0.0069$; Leung and Lee (2013) find that $e_{\mathrm{free}} = 0.03$. Figure 3.1 clearly shows that Kepler-16b's measured eccentricity is much smaller than that of its compatriots.

Since its discovery, there has been much discussion in the literature of how these

planets can have formed, focussing primarily on Kepler-16b. Paardekooper et al. (2012) argued that it is unlikely to have formed in situ, due to the high planetesimal velocity dispersion induced by the binary. Meschiari (2012) found similar results, and while Rafikov (2013) showed analytically that a massive disc could damp the velocity dispersion, this effect is probably not sufficient to allow Kepler-16b to form in situ. Indeed, numerical simulations by Marzari et al. (2013) suggest that in practice disc self-gravity may have the opposite effect, again inhibiting planetesimal growth.

The broad consensus is emerging that Kepler-16b cannot have formed in its current orbit, and must presumably have formed at larger radius and migrated inwards. This picture is consistent with the highly aligned nature of the system, as well as the fact that it is at the very edge of stability – the planet is located at the natural inner edge for a circumbinary disc (Holman and Wiegert, 1999; Pelupessy and Portegies Zwart, 2013). However, this picture of gentle disc-driven migration is complicated by the work of Pierens and Nelson (2008), who found that a Saturn-mass planet migrating through such a disc is likely to attain a significant eccentricity. Although the measured orbital parameters are expected to osculate under the influence of the binary, Leung and Lee (2013) have shown that the maximum eccentricity is still low ($e_\mathrm{p} < 0.07$). Moreover, Popova and Shevchenko (2013) find that the planet's eccentricity cannot ever have been much greater than $e_\mathrm{p} = 0.05$ without its orbit becoming unstable.

These arguments point very strongly towards a picture where Kepler-16b's orbit has never been significantly eccentric, which in turn suggests that its free eccentricity was damped to near zero due to the interaction with its parent protoplanetary disc. In order to investigate this further, I have carried out high resolution SPH simulations of an analogue to the Kepler-16 system embedded in a circumbinary disc, seeking to characterize how the disc would have affected the angular momentum of the planet (and thus its eccentricity) as it migrated. In Section 3.2 I describe the method used, and present the results of the simulations in Section 3.3. I discuss the implications of these results in Section 3.4 and outline the conclusions in Section 3.5.

3.2 Simulations

I follow the same numerical method as in Chapter 2, using SPH particles to simulate a gas disc and 3 *N*-body particles to represent the planet and binary components, with their orbital elements free to evolve throughout the simulation. These act as sinks for gas particles, and the sink radii were set to 0.25 times the binary semimajor axis a_b for the

binary components and 0.4 times its Hill radius for the planet (which is given by Equation 1.59, with $M_\star$ replaced by the total binary mass M_b).

The Navier-Stokes viscosity used for the SPH followed the standard Shakura-Sunyaev α-viscosity prescription, with $\alpha = 0.01$. The disc thickness was again defined by using an imposed locally-isothermal equation of state, with temperature $T_{iso} \propto R^{-1/2}$ as in Chapter 2, so that the disc aspect ratio scales as $H/R \propto R^{1/4}$. This was normalized so that the $H/R = 0.05$ at $R = a_p$.

The positions of the disc particles were distributed by random sampling in a similar manner to that described in Chapter 2, this time following

$$\Sigma(R) = \Sigma_p \left(\frac{R}{a_p} \right)^{-\gamma} \tag{3.1}$$

where $\Sigma_p = 100$ g cm^{-2} is the surface density at the radius of the planet's semimajor axis.

In order to reduce the impact of initial numerical transients, the disc was relaxed for 100 binary orbits before the planet was inserted into the simulation. Although the initial surface density profile has $\gamma = 1$, by the time the disc had relaxed this had become $\gamma \sim 3/2$, as expected for a circumbinary decretion[2] disc (Pringle, 1991). In addition, in order to prevent the planet undergoing a transient burst of accretion immediately following its insertion, I modified the (relaxed) surface density profile to include a gap at the initial orbital semimajor axis of the planet, using the parametrization of Lubow and D'Angelo (2006). The surface density profile $\Sigma(R)$ given in Equation 3.1 is multiplied by a factor f_Σ

$$f_\Sigma = \exp\left[-\frac{1}{9} \frac{q^2 a_p^2 \Omega_p}{\nu_p} \left(\frac{a_p}{\Delta_p} \right)^3 \right] \tag{3.2}$$

where $q = m_p/m_b$ is the mass ration between the planet and the binary, Ω_p and ν_p are respectively the orbital frequency and viscosity in the disc at $R = a_p$. The factor $\Delta_p = \max[H, |R - a_p|]$ ensures that length-scales shorter than H do not dominate (i.e. the gap width will be of order H). Note that this means that although the disc surface density was normalized to 100 g cm^{-2} in Equation 3.1 the imposition of the gap means that the actual value is much smaller than this after the planet is inserted.

For these simulations I adopted a system of code units such that the unit of distance is 0.22431 au (equal to the binary semimajor axis; Doyle et al., 2011). The unit of mass

---

[2] A decretion disc is one where the net transport of angular momentum and mass are both in the outward direction. For a circumbinary disc, the resonant torques from the binary hold the disc out and prevent it from accreting a significant amount of mass, turning the initial accretion disc into a decretion disc.

Model name	M_p [M_{Jup}]	e_0
REFERENCE	0.333	0.0
MASSIVE	1.0	0.0
ECCENTRIC	1.0	0.05

Table 3.1: Table of model parameters. M_p is the planet mass in units of the mass of Jupiter, M_{Jup}, and e_0 is the initial eccentricity of the planet's orbit.

is equal to the sum of the masses of the the binary components and the planet, and the time unit is set to 41.08 days, the orbital period of the binary. This sets the gravitational constant $G = 4\pi^2$ in code units.

I ran a total of 3 such simulations for a period of 400 orbits of the planet (approximately 2200 binary orbits). In each run the physical parameters of the binary are those reported by Doyle et al. (2011), and both the planet and disc were coplanar with the binary. The initial semimajor axis of the planet was $a_p = 0.7048$ au (3.142 in code units) as reported in the discovery paper, and the initial radial extent of the disc (before relaxation) was $1.5 < R < 10$ in code units. The disc scale height was resolved into approximately 6 SPH smoothing lengths at the radius of the planet's orbit, sufficient to avoid numerically underestimating the midplane density (Nelson, 2006).

The REFERENCE model used a planet mass $M_p = 0.333\,M_{Jup}$, as reported by Doyle et al. (2011). To test the effect of a higher planetary eccentricity, I also ran a model with initial eccentricity $e_0 = 0.05$ (ECCENTRIC). In this case I used a higher planet mass of 1 M_{Jup} to ensure that unphysical numerical eccentricity damping did not occur (e.g. Masset and Ogilvie, 2004). In order to compare this run with the REFERENCE model, I also ran a model identical to the REFERENCE model but with this higher planet mass (MASSIVE). The model parameters which were varied between the simulations are summarised in Table 3.1.

3.3 Results

A representative snapshot of the central region of the MASSIVE simulation after 200 orbits of the planet (T_p) is shown in Figure 3.2. The eccentricity trend in all of the models (shown in Figure 3.3) follows the pattern found by Leung and Lee (2013), where the osculating eccentricity e_p consists of forced and free components. They found that this occurs between $e_{min} = |e_{forced} - e_{free}| \simeq 0.006$, and $e_{max} = e_{forced} + e_{free} \simeq 0.066$. There is excellent agreement with this in the simulations with no initial eccentricity, for both low
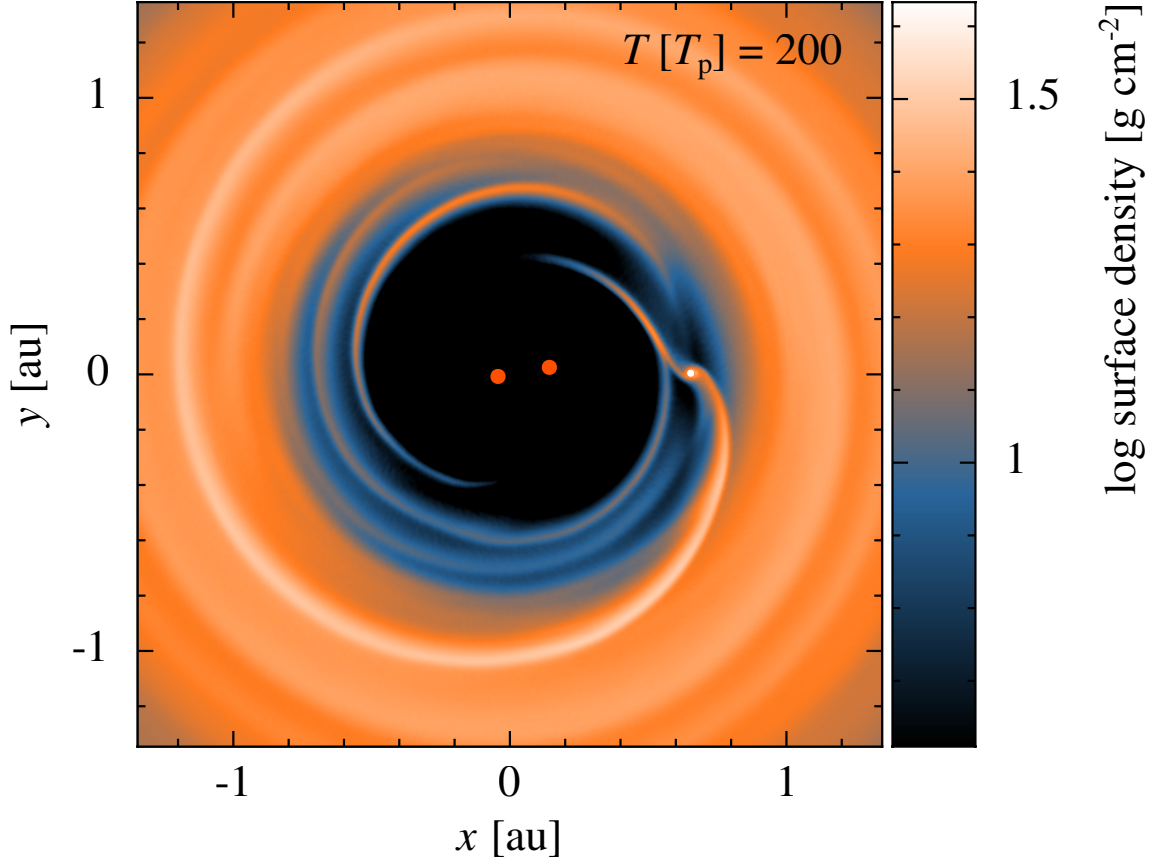
Figure 3.2: Snapshot of the central region of the MASSIVE model (see Table 3.1) after 200 planetary orbits. The positions of the (stellar) binary components are shown as orange points, while the white point denotes the position of the planet; the colour-scale shows the gas surface density. The planet orbits very close to the inner disc edge, and the spiral density waves induced by the planet are clearly visible.

———————————————————————————————————

and high planet masses, suggesting that torques from the disc do not alter the planet's orbit significantly on dynamical time-scales. In the initially eccentric model there is a similar osculating eccentricity, but with a larger amplitude (as expected), giving $e_{max} \simeq 0.12$ for $e_0 = 0.05$.

As the planet's eccentricity evolution is dominated by the binary forcing pattern, it is difficult to determine if there is any underlying longer-time-scale change in the planet's eccentricity due to the interaction with the disc. In order to test this, I compare the eccentricity in the REFERENCE simulation with that from a pure N-body run (that is, an identical binary set-up but with no disc, using the same code and integrator as in the hydrodynamic runs) in Figure 3.4. The orbit-averaged REFERENCE planet eccentricity (e_{hydro}) is divided by the orbit-averaged N-body planet eccentricity ($e_{N\text{-}body}$) as a function of time. Relative to
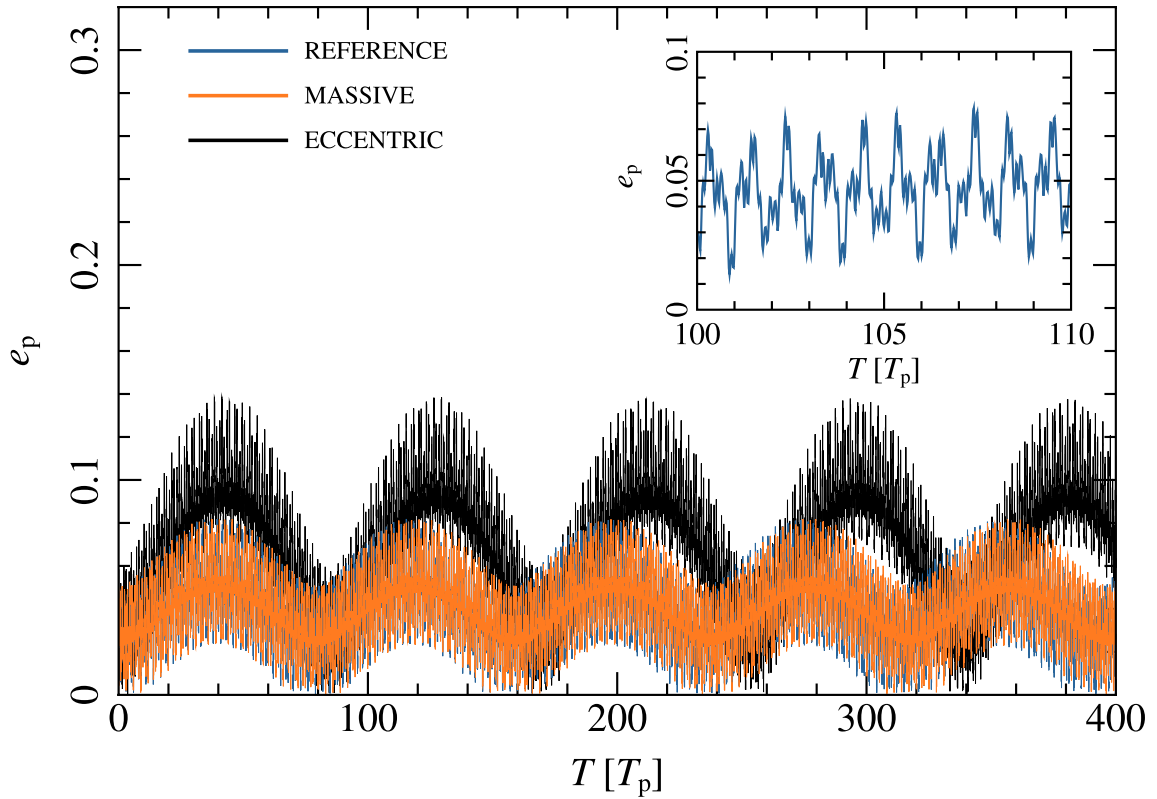
Figure 3.3: Planet eccentricity as a function of time for the three models described in Table 3.1. The osculation of eccentricity for the initially circular models (REFERENCE and MASSIVE) agrees with those found by Doyle et al. (2011) and with the analytic theory of Leung and Lee (2013), independent of planet mass. The inset shows the eccentricity over 10 orbits in the REFERENCE model, illustrating the intricate forcing pattern driven by the central binary.

the N-body model (which closely matches the analytic solution of Leung and Lee, 2013), the REFERENCE model shows initial eccentricity damping (at the per cent level) due to the relaxation of the disc initial conditions, followed by periodic oscillations. This periodicity matches the osculation period shown in Figure 3.3, and consequently I do not attach any significance to the minima seen in Figure 3.4 (as these represent ratios of pairs of very small numbers). The maxima, however, are more reliable, and after the decay of initial transients these eccentricity peaks show a shallow decline with time. Due to the orbital precession induced by the disc, the eccentricity cycles from the REFERENCE model and the N-body run are slightly out of phase by the end of the simulations. This causes the 'sharpening' of the maxima seen in Figure 3.4, but the phase difference is sufficiently small that the peak values are not strongly affected. I therefore conclude that the disc damps the planet eccentricity (relative to the N-body run) at the 1 per cent level over the duration of
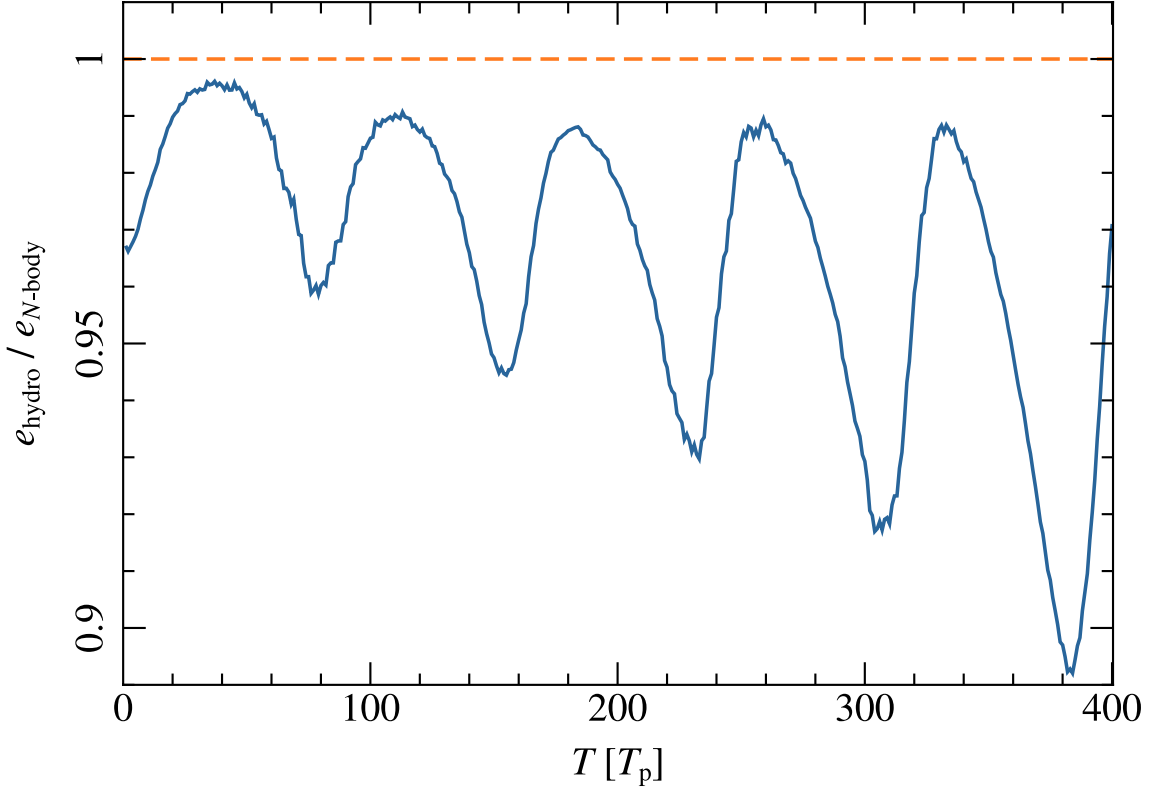
Figure 3.4: Comparison between the orbit-averaged planet eccentricity from the REFERENCE run (e_{hydro}) with that from an identical simulation with no gas disc ($e_{N\text{-body}}$). For reference, the orange dashed line denotes $e_{\text{hydro}}/e_{N\text{-body}} = 1$. The strong oscillations are due to the binary forcing and matches the periodicity in Figure 3.3, but after initial transients have dissipated the peak eccentricity settles into a broadly constant rate of decay.

the simulations.

3.3.1 Torque analysis

The planet's orbit is dominated by the binary forcing throughout the SPH simulations, as discussed above, but additional insight into its behaviour on longer time-scales can be gained by looking at the torques exerted on the planet by the disc. Figure 3.5 shows the orbit-averaged disc torques on the planet for each of our simulations. The torque Γ is defined such that a negative torque corresponds to eccentricity growth (Armitage, 2010); positive torques damp the planet's eccentricity. Although the details vary between models, the same broad trend is seen in all of the simulations. The torques undergo periodic oscillations on the time-scale of the eccentricity osculation shown in Figure 3.3, with the maximum torque corresponding to the peak eccentricity. The underlying trend is
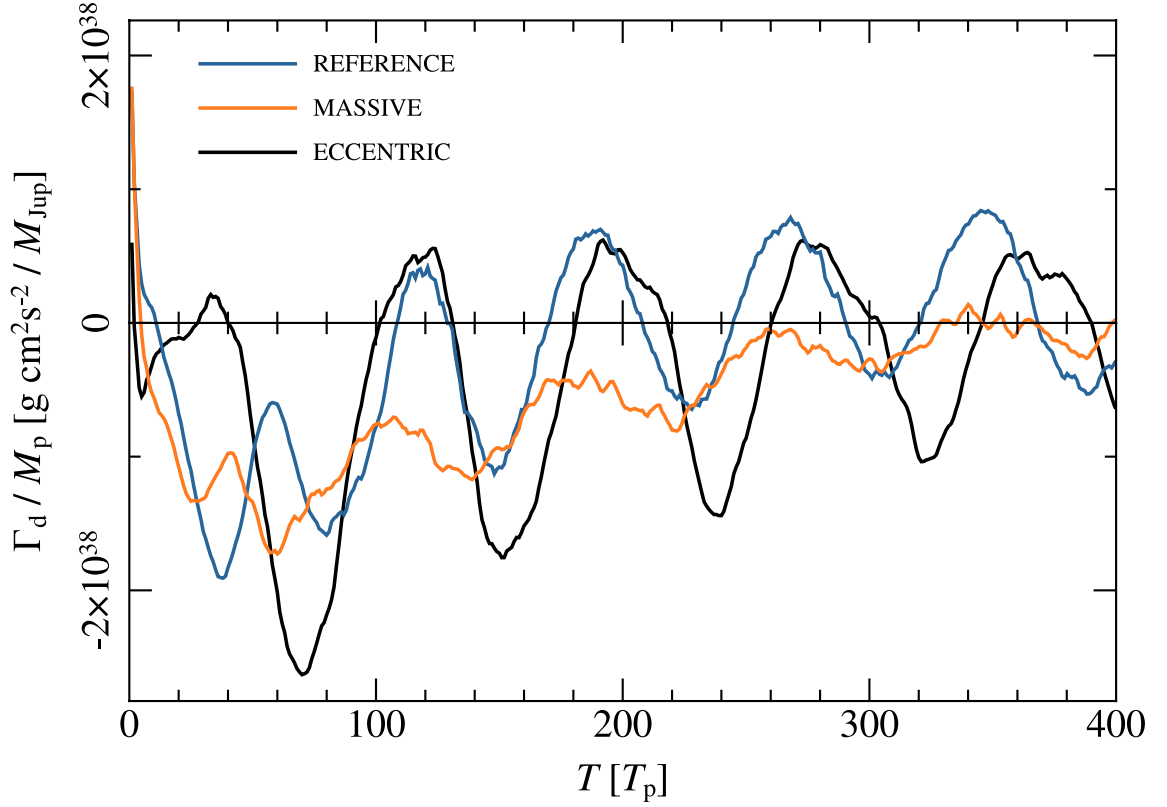
Figure 3.5: Orbit-averaged disc torques, normalized to the planet mass, as a function of time (in planetary orbital periods) for the models described in Table 3.1. The torques oscillate on the same time-scale as the eccentricity, and after initial transients have decayed settle into a pattern with a net positive torque of 6×10^{36} g cm^{2}s^{-2} (for the REFERENCE model). The torques on the planet from the binary are $\sim 10^4$ times stronger, but additional runs show that the disc torques scale linearly with the surface density over a wide dynamic range in Σ. Lower-resolution test runs indicate that the torques in the REFERENCE model are approximately converged, and the MASSIVE and ECCENTRIC models are almost converged by the end of the simulation.

---

that the torques are initially negative, but increase with time and eventually settle into a quasi-steady (oscillating) state. In the case of the REFERENCE model, this has a net positive torque, with a time-averaged value of $\langle \Gamma_{\mathrm{disc}} \rangle \simeq 6 \times 10^{36}$g cm^{2}s^{-2}. In a sense this is to be expected: the planet orbits near the inner edge of the disc, and the presence of more material exterior to the planet's orbit favours positive torques. These provide angular momentum to the planet, in principle allowing it either to reduce its eccentricity or to migrate outwards. However, the model setup here assumes that the planet has previously migrated from a formation point further out in the disc, and this configuration disfavours outward migration, so the long-term effect of these positive torques is to damp the planet's eccentricity (as seen in Figure 3.4).

The MASSIVE and ECCENTRIC models show the same basic trend of torque growth but

here the net torque at the end of the simulation runs is negative. This is consistent with the results of Pierens and Nelson (2008), who found generally negative torques in their 2D disc models that resulted in the growth of eccentricity. Both MASSIVE and ECCENTRIC models and the Saturn-mass run from Pierens and Nelson (2008) were in the gap-opening regime due to high-mass planets and low disc viscosity respectively (see Figure 3.6 for a comparison between the REFERENCE and MASSIVE models with respect to their gap opening). In contrast, the REFERENCE model is not fully in the gap-opening regime, and so feels additional torques from co-orbital gas not present in the MASSIVE, ECCENTRIC or Pierens and Nelson's runs. I address numerical concerns with simulating non-gap opening planets in isothermal discs in Section 3.4.1.

In all of the simulations presented here, the absolute values of Γ are smaller than the tidal torques from the binary by a factor $\sim 10^4$, and additional calculations (run for shorter durations) show that the disc torques scale linearly with Σ over a wide range (at least 3 orders of magnitude) in disc surface density. I therefore conclude that if the planet is not in the gap-opening regime, the disc-planet interaction in the Kepler-16 system during its disc phase leads to damping of the planet's free eccentricity, with the damping time-scale determined primarily by the local disc surface density.

3.3.2 Disc properties

I now use this result to infer a limit on the gas surface density of the real disc in which Kepler-16b formed. Previous simulations of circumbinary planets have have found that eccentricity grows on a time-scale comparable to (or shorter than) the migration time-scale τ_{mig} (e.g. Pierens and Nelson, 2008). If the planet's orbit never became significantly eccentric during its migration, then that the eccentricity must be damped on a time-scale $\tau_e \lesssim \tau_{\mathrm{mig}}$. In order to connect this to the disc torque, the angular momentum ΔJ gained by the planet from the damping torque must be estimated. To first-order the potential is Keplerian, so I approximate the planet's angular momentum as $J_{\mathrm{p}} = M_{\mathrm{p}} \sqrt{GM_{\mathrm{b}}\, a_{\mathrm{p}}(1 - e^2)}$. Differentiating with respect to eccentricity gives

$$\left(\frac{\partial J_{\mathrm{p}}}{\partial e}\right)_{a_{\mathrm{p}}} = -M_{\mathrm{p}} \sqrt{GM_{\mathrm{b}}\, a_{\mathrm{p}}} \left(1 - e^2\right)^{-1/2} e, \tag{3.3}$$

so to first-order in e

$$\frac{\mathrm{d}J}{\mathrm{d}e} \simeq -e M_{\mathrm{p}} \sqrt{GM_{\mathrm{b}}\, a_{\mathrm{p}}} \tag{3.4}$$

and

$$\Delta J = \left|\frac{\mathrm{d}J}{\mathrm{d}e}\right| e = e^2 M_{\mathrm{p}} \sqrt{GM_{\mathrm{b}}\, a_{\mathrm{p}}}\,. \tag{3.5}$$

Maintaining a low eccentricity requires a damping torque that fulfils

$$\Gamma_{\mathrm{d}} \gtrsim \frac{\Delta J}{\tau_e}. \tag{3.6}$$

As noted above, Γ_{d} scales linearly with the disc surface density (in the limit $M_{\mathrm{d}} \ll M_{\mathrm{b}}$) so this allows a lower limit on Σ. The planet is unlikely to have attained an eccentricity $e > 0.1$ (e.g. Popova and Shevchenko, 2013), which gives $\Delta J = 2.3 \times 10^{47}$ g cm^{2}s^{-1}. For the REFERENCE model, which has $\Sigma_{\mathrm{p}} = 100$ g cm^{-2}, the net damping torque is $\Gamma_{\mathrm{d}} \simeq 6 \times 10^{36}$ g cm^{2}s^{-2} (see Figure 3.5), and substituting these values into Equation 3.6 and re-arranging (assuming $\Gamma_{\mathrm{d}} \propto \Sigma_{\mathrm{p}}$) yields

$$\Sigma_{\mathrm{p}} \gtrsim 120 \left(\frac{\tau_e}{10^4\,\Omega_{\mathrm{p}}^{-1}}\right)^{-1} \mathrm{g\,cm}^{-2}. \tag{3.7}$$

This is a surprisingly large surface density, and implies that Kepler-16b formed in a comparatively massive circumbinary disc. Pierens and Nelson (2008) found that the migration and eccentricity growth time-scale for a Saturn-mass planet is $\sim 10^4\,\Omega_{\mathrm{p}}^{-1}$, which I adopt as a reference value above. The true migration time-scale is not known, and the planet may well have migrated more slowly. However, even a conservative assumption of $\tau_e \sim 10^5\,\Omega_{\mathrm{p}}^{-1}$ requires $\Sigma_{\mathrm{p}} \gtrsim 10$ g cm^{-2}, and I regard this as a robust lower limit to the surface density of Kepler-16b's parent disc.

3.4 Discussion

3.4.1 Limitations of the model

A major source of uncertainty in the simulations is the treatment of the disc thermodynamics, where I adopted a locally (vertically) isothermal equation of state. In the single-star case it is well known that a full radiative hydrodynamical treatment can give different results for low-mass planets below the gap-opening limit (e.g. Bitsch and Kley, 2010, 2011a,b). For Saturn-mass planets, it seems that migration behaviour is broadly unaffected by the disc equation of state (Bitsch and Kley, 2011a), but the disc structure can still be strongly affected. Marzari et al. (2013) compared their fully radiative treatment in the circumbinary disc case with the locally isothermal approximation of Pelupessy
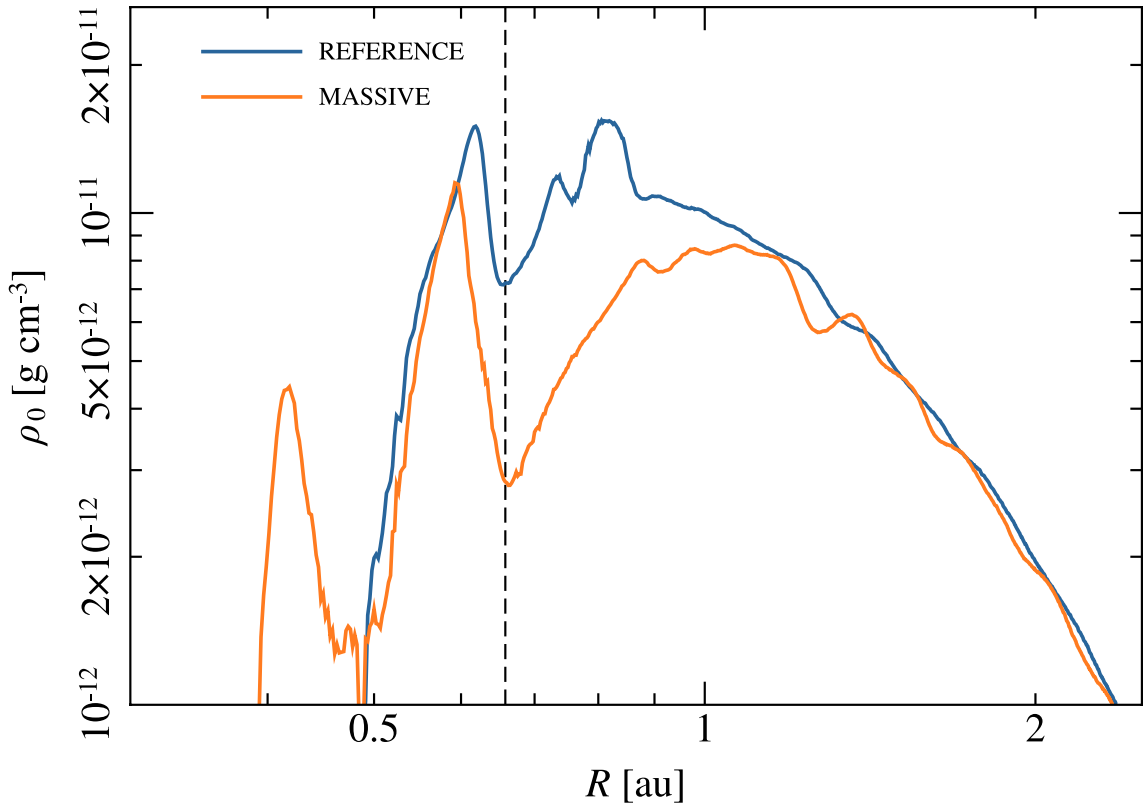
Figure 3.6: Midplane gas densities for the REFERENCE and MASSIVE models (see Table 3.1) after 200 planetary orbits, as a function of disc radius R along an axis offset from that of the planet by 0.1 rad (to avoid the density enhancement of material being accreted by the planet). A dashed line indicates the semimajor axis of the planet in both models. While a Jupiter-mass planet is able to maintain a strong gap in the disc, the lower-mass planet in the REFERENCE model only opens a partial gap, with a significant mass of gas remaining in the corotation region.

———————————————————————————————

and Portegies Zwart (2013) and found significant differences in the temperature profile, though the different initial surface density profiles between their models must also play a strong role. Marzari et al. (2012) also performed a comparison between a fully radiative treatment with an an imposed temperature profile, and found that disc eccentricity is strongly damped in the radiative case. They also found that spiral waves are smoothed somewhat by the radiative treatment. In the simulations presented here, these two factors are mitigated somewhat by the (relatively) high level of disc viscosity, which independently damps disc eccentricity and smooths spiral waves.

Unfortunately the parameter space for radiative circumbinary disc models in 3D is vast, and exploring even a modest sub-set of this space is not feasible. However, in single-star models the critical uncertainty where a gap is not opened is usually the (thermo)dynamics

of gas in the co-rotation region of the planet, but it is unclear if these simulations are in this regime. Comparing the models with different planet masses, shown in Figure 3.6, I plot the midplane density for the REFERENCE and MASSIVE models after 200 planetary orbits. The REFERENCE model, with a planet of approximately Saturn mass, only opens a partial gap in the disc, and only the more massive planet is really in the gap-opening regime. The results of our REFERENCE model must therefore be taken with some caution. The torque analysis in Figure 3.5 shows that the torque oscillates on the same time-scale for all planet masses, but the net (time-averaged) torque is negative for the runs with higher planet mass, as seen in previous studies (e.g., Pierens and Nelson, 2008). Crucially however, Bitsch and Kley (2011b) showed in the single-star case that an accurate treatment of the disc thermodynamics generally increases the disc torques for low-mass non gap-opening planets, suggesting that the net positive Γ_d seen in the REFERENCE simulation is a robust result. Clearly, further investigation of this issue is still required.

A further limitation of the high-resolution 3D treatment is that as in Chapter 2, due to computational expense I am only able to simulate time-scales of hundreds of planetary orbits. In contrast, Pierens and Nelson (2008) ran their 2D simulations for tens of thousands of planetary orbits. Although I cannot follow changes in the planet's orbit on such long time-scales, the torque analysis in Section 3.3.1 provides a clean way to probe longer time-scale behaviour in our models. However, it is notable that this result, that disc torques damp eccentricity, disagrees with Pierens and Nelson (2008), who found eccentricity growth and outward migration in comparable models. It is well known that a 2D approach can lead to artificially higher resonant torques (Tanaka et al., 2002), but this is unlikely to be the origin of such a qualitative change in behaviour.

Instead, I attribute this apparent discrepancy to the different disc viscosities used. I chose $\alpha = 10^{-2}$, while Pierens and Nelson (2008) used $\alpha = 10^{-3}$, which is below the limit set by the artificial viscosity in my simulations (see Section 2.2). The different dimensionality (3D versus 2D) means that the gap-opening criteria in these simulations cannot easily be compared directly, and it is computationally unfeasible to simulate such low disc viscosities (see Section 2.2). The critical planet mass for gap-opening typically decreases with lower viscosities, as the disc-planet torques more easily overcome the viscous torques in the disc (as shown in Equation 1.58). Comparing Figure 3.2 with Figures 5 and 8 in Pierens and Nelson (2008), their planets open much deeper and wider gaps in their discs, and therefore qualitatively different behaviour is not unexpected. Measurements of real disc viscosities are uncertain at best, and are almost certainly not constant (particularly in the case of "dead zones" in disc midplanes). In general, however, the adopted

value of $\alpha = 10^{-2}$ is consistent with the values inferred from observations of protoplanetary discs (e.g. Hartmann et al., 1998; King et al., 2007). I note also that the simulations presented here are specifically tailored to the observed parameters of the Kepler-16 system, and therefore differ from those of Pierens and Nelson (2008, who used $q = 0.1$, and initial eccentricities of $e_b = 0.08$ and $e_p = 0.02$). Their more extreme binary mass ratio in particular may change the system's dynamics significantly (as the secondary orbits much closer to the inner disc edge); further exploration of the binary parameter space is necessary to understand these effects in more detail.

The relatively short duration of the simulations means that I cannot rule out the possibility that the positive torques on the planet drive some outward migration in addition to damping the planet's eccentricity. However, as noted in Section 3.3.1 this is unlikely because these simulations consider a planet that has already migrated to the orbit it is observed in today. Figure 3.4 shows that there is indeed some small level of eccentricity damping in the simulations, although again I cannot rule out the possibility of outward migration at later times. If this is considered as a possible outcome, it in fact strengthens the minimum surface density requirement. This is because if some fraction of the total disc torque is 'diverted' towards driving outward migration, the remaining torque which acts to damp eccentricity is smaller, requiring a higher surface density to maintain the same level of eccentricity damping.

This analysis also neglects the possibility that the binary can drive eccentricity in the disc, which in turn can alter the angular momentum of the planet. No significant disc eccentricity is seen in any of the simulations, but Pierens and Nelson (2008) showed that a binary with eccentricity similar to Kepler-16 can excite a disc eccentricity of $e_d \sim 0.05$ on time-scales only slightly longer than the migration time-scale of the planet (i.e., several thousand planetary orbits). To see what effect this may have upon the planet, I consider the maximal case where for the radial region $R_p - H < R < R_p + H$, all the angular momentum from the disc's eccentricity is transferred to the planet. Starting by analogy with Equation 3.5, the change in angular momentum due to the disc eccentricity e_d is

$$\Delta J_d = e_d^2 M_d \sqrt{GM_b R_p} \tag{3.8}$$

where I have assumed the planet's eccentricity to be small so that its orbital radius $R_p \simeq a_p$. Assuming that the disc eccentricity is also small so that it may be approximated as a series

of concentric annuli, the mass is given by

$$M_{\rm d} = \int_{R_{\rm p}-H}^{R_{\rm p}+H} 2\pi R \Sigma \, dR. \tag{3.9}$$

For a disc with surface density profile of the form given in Equation 3.1, this becomes

$$\Delta J_{\rm d} = 2\pi e_{\rm d}^2 \sqrt{GM_{\rm b}} \Sigma_{\rm p} R_{\rm p}^{5/2} f \tag{3.10}$$

where f is the factor

$$f = \frac{\left(1 + \frac{H}{R}\right)^{5/2-\gamma} - \left(1 - \frac{H}{R}\right)^{5/2-\gamma}}{5/2 - \gamma}. \tag{3.11}$$

Pierens and Nelson (2008) found the time-scales for migration and disc eccentricity driving to be similar, so I can compare this with the angular momentum change the planet experiences from damping torques given by Equation 3.5:

$$\frac{\Delta J_{\rm d}}{\Delta J} = 2\pi \left(\frac{e_{\rm d}}{e_{\rm p}}\right)^2 \frac{\Sigma_{\rm p} R_{\rm p}^2}{M_{\rm p}} f. \tag{3.12}$$

Taking $e_{\rm d} \sim 0.1$ as before and $e_{\rm d} =\sim 0.05$ from Pierens and Nelson (2008), for a disc with $H/R = 0.05$ and a surface density profile with $\gamma = 3/2$ this becomes

$$\frac{\Delta J_{\rm d}}{\Delta J} = 2.7 \times 10^{-4} \, \Sigma_{\rm p}. \tag{3.13}$$

For this to be a significant effect, $\Delta J_{\rm d}$ must be some significant fraction of ΔJ. Arbitrarily setting this to be 10 percent requires that

$$\Sigma_{\rm p} \gtrsim 370 \, {\rm g \, cm}^{-2}. \tag{3.14}$$

A more detailed consideration of this effect is beyond the scope of these simulations and this thesis, but this suggests that disc eccentricity is unlikely to alter our conclusions significantly unless the disc is very massive. Instead, as the time-scale for disc eccentricity growth is relatively long, the most likely outcome is that a modest disc eccentricity will slow, but not reverse, the damping of the planet's eccentricity. While Pierens and Nelson (2013) recently performed simulations of the Kepler-16 system using a similar method to their previous work (Pierens and Nelson, 2008), their neglect of the influence of the disc on the binary does not conserve momentum to any reasonable degree. The large disc eccentricities in their simulations ($e_{\rm d} \sim 0.2 - 0.4$ in the inner regions) are likely due to a

convolution this non-conservation and the dimensionality of the simulations.

Finally, I note that these simulations and analysis are predicated on the assumption that Kepler-16b's eccentricity was damped during the disc phase, rather than at later times in the systems evolution. This is supported by stability modelling of the system, which suggests that the planet's orbit would be unstable if its eccentricity increased above $e_{\mathrm{p}} \simeq$ 0.05 in the post-disc phase (Popova and Shevchenko, 2013), but it is possible that an additional body (or bodies) in the system could influence the planet's orbit. The simplest dynamical explanation for the low eccentricity is that the planet is at the low-e point of a Kozai cycle (Kozai, 1962; Lidov, 1962). This scenario is rather contrived, however, and requires the presence of a massive perturber on an inclined orbit in addition to the system being observed at a very specific time in its evolution. There is some evidence for an additional planetary-mass body in the system (Bender et al., 2012), but detailed dynamical modelling of its potential effects is not yet possible. I therefore cannot exclude a dynamical origin, but given the level of fine-tuning required I do not consider this a likely explanation for the low eccentricity of Kepler-16b.

3.4.2 Implications for circumbinary planet formation

These results suggest that Kepler-16b formed in, and migrated through, a circumbinary disc with a relatively high surface density. Observations currently tell us little about the surface densities of circumbinary discs, particularly for young binaries with separations as close as Kepler-16. At larger separations ($a \gtrsim 10$ au), where young binaries are more readily observed, most circumbinary discs appear to be rather low mass: the majority of the circumbinary discs identified by Kraus et al. (2011, 2012) have masses $M_{\mathrm{d}} \leqslant 10^{-3}$ $M_{\odot}$ (Andrews and Williams, 2005). At the other end of the binary period distribution, Rosenfeld et al. (2012) measured the mass of CO gas in the disc surrounding the very close (2.4-day period) binary V4046 Sgr to be $M_{\mathrm{CO}} \sim 3 \times 10^{-6} \, M_{\odot}$, implying a total gas mass of $\sim 0.01 - 0.1 \, M_{\odot}$. This system is particularly noteworthy because it is old (~ 12 Myr; Rosenfeld et al., 2012), yet it still retains a very massive disc.

Once again however, all of these measurements are based on (sub-) millimetre observations, which only probe the outer regions of the disc. Estimates of disc surface densities at au radii can only be made by extrapolating inwards. Identifying young binaries at au separations remains challenging (e.g., Kraus et al., 2012) and, while some subset of the sample of known circumstellar discs are presumably circumbinary, observations of discs in known close binary systems remain scarce. The limit on the surface density of Kepler-16b's parent disc derived here therefore provides a useful new insight into circumbinary
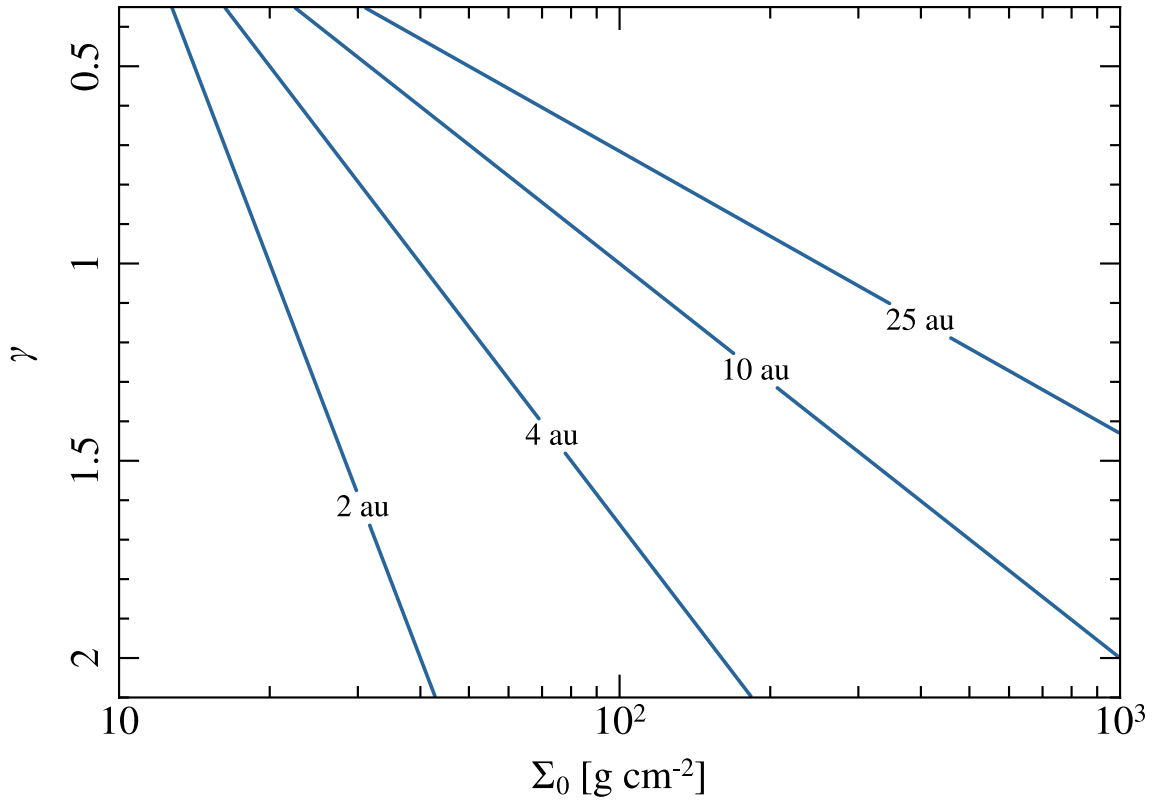
Figure 3.7: Contour plot for disc models the form of equation 3.15. Contours are labelled with the radius R at which the surface density $\Sigma(R) = \Sigma_{min}$ for given values of the normalization factor Σ_0 and power-law index γ.

---

disc physics in this poorly-explored regime.

I extend this analysis further by considering the radial distribution of mass in such a disc. Figure 3.7 shows a contour plot for different disc models with the form

$$\Sigma(R) = \Sigma_0 \left(\frac{R}{1 \text{ au}}\right)^{-\gamma}, \tag{3.15}$$

(note that this is of the same form as but subtly different to Equation 3.1) where contours indicate the radius R at which the surface density equals $\Sigma_{min} = 10 \text{ g cm}^{-2}$,. The true radius at which Kepler-16b formed is unknown, although there is general agreement that it must have formed beyond approximately 10 au (Meschiari, 2012; Rafikov, 2013; Marzari et al., 2013). Figure 3.7 shows that to satisfy $\Sigma(R \gtrsim 10 \text{ au}) > \Sigma_{min}$ requires discs with $\Sigma_0 \gtrsim 10^2$ g cm^{-2}, even for shallow surface density gradients ($\gamma < 1$).

I have focused our attention on Kepler-16b because of its extremely low measured eccentricity (see Figure 3.1, but it in fact shares a number of characteristics with other

circumbinary systems. Kepler-35b, -38b and -47b have low eccentricities ($e_{\rm p} < 0.05$) and similar semimajor axis ratios $a_{\rm p}/a_{\rm b} \sim$ 3–4. Moreover, the secular disc evolution models of Alexander (2012) find that photoevaporation rapidly erodes the inner regions of circumbinary discs, and high surface densities persist only for a short fraction of the discs' lifetimes ($\lesssim$ 1Myr). Taken together with the results of the simulations presented here, this suggests that these circumbinary planets must have formed early in the evolution of their parent discs, when the discs were still massive enough to damp their eccentricities. At present the poor observational statistics for both circumbinary discs and circumbinary planets do not allow me to draw detailed conclusions regarding the frequency of such planets, but as the census of both discs and planets improves I expect circumbinary systems to continue to provide crucial insights into the physics of planet formation.

3.5   Conclusions

I have performed high resolution 3D SPH simulations of a Kepler-16-like system embedded in a circumbinary disc using the code described in Sections 1.4.1 and 2.2. Analysing the disc torques from these simulations and using observational constraints on the system's history, I conclude that the planet must have maintained its low eccentricity by migrating through this disc without opening a significant gap, leading to the damping of eccentricity by co-orbital material.

As this damping is directly proportional to the disc mass, I have derived a limit on the local disc surface density in which Kepler-16b was once embedded, $\Sigma_{\rm min} \sim 10 \, {\rm g\,cm^{-2}}$. Applying this result to similar circumbinary systems also discovered by the *Kepler* mission, I conclude that this low limit may be used to provide constraints on the route planet formation takes in the circumbinary environment. I have argued that the process requires relatively massive circumbinary discs, and that their secular evolution implies that these planets form far out in the disc at early times, and migrated inwards before the disc mass was significantly depleted.

I have identified a number of areas requiring more work, including a more detailed consideration of the equation of state. Similarly, I have neglected to consider that the binary orbit must have evolved, and if the planet is understood to have formed and migrated early in the lifetime of the system this evolution may play an important role. The possibility that Kepler-16b's eccentricity may have been damped through dynamical interactions with a fourth body in the system is unlikely but cannot currently be ruled out, so this is another area where further work may yield greater understanding.

4

Binary supermassive black hole accretion & the last parsec problem

I N this chapter I turn to the problem of SMBH binaries, to investigate how gas accretion discs can help to coalesce them at a high enough rate to reconcile theory with observational statistics.

4.1   Introduction

4.1.1   Coalescing SMBH binaries

As detailed in Chapter 1, there is a mismatch between the number of SMBH binaries observed in other galaxies and current theoretical understanding of the galaxy merger process (Begelman et al., 1980; Milosavljević and Merritt, 2001). In the aftermath of a major galaxy merger, it is expected the resultant merged galaxy will contain two SMBHs, one from each progenitor galaxy. Dynamical friction between stars and the two SMBHs is an efficient process at migrating the SMBHs towards the centre of the galactic potential on short time-scales (Milosavljević and Merritt, 2001), where the SMBHs will form a binary (Mayer et al., 2007).

Dynamical interactions with stars on orbits that allow them to interact strongly with the binary can continue to shrink the binary once it has formed. Here, three-body encounters eject the least massive body (i.e. the star) from the system, taking some angular momentum and energy with it. This process can allow the binary to shrink to separations of ~ 1 pc (Begelman et al., 1980).

However, once the binary has shrunk to this scale, the ejection of stars will have cleared a gap in phase space (the six-dimensional space of stellar positions and velocities) called the 'loss cone', leaving no stars able to strongly interact with the binary and affect its evolution. While the loss cone can be refilled somewhat by scattering events between stars, the rates at which this is expected to occur are far too long to continue driving the binary evolution (of the order $\sim 10^9$ years; Merritt and Wang, 2005).

If the binary semimajor axis shrinks to some critical value a_{crit} then gravitational wave emission from the binary will remove sufficient energy to allow it to coalesce on very short time-scales. This critical value depends on the mass ratio, total binary mass and binary eccentricity but is typically of the order of $10^{-3} - 10^{-2}$ pc (Armitage and Natarajan, 2002, 2005; Merritt and Milosavljević, 2005; Lodato et al., 2009). While simulations show that for low binary masses ($\sim 10^6 - 10^7 \, M_\odot$), torques from a prograde un-inclined accretion disc can take the binary to a_{crit} on short time-scales (Armitage and Natarajan, 2002; Lodato et al., 2009) this process is very inefficient for higher binary masses and

typically takes longer than a Hubble time before coalescence (Lodato et al., 2009).

Observations of galaxies show signs for multiple SMBHs in only some very small fractions of galaxies. Some large separation ($\sim$ kpc) dual SMBHs are known to exist (e.g. NGC6240; Fabbiano et al., 2011), where both components show typical AGN signatures. These are seen in recent merger remnants and are therefore currently in the dynamical friction regime. Despite expectations that binary coalescence stalls at these distances, very few objects are known at $\sim$ pc separations. Exceptions include the radio galaxy 0402+379, which shows evidence for a SMBH binary with a projected separation of 7.3 pc (Rodriguez et al., 2006). OJ 287, perhaps the most famous candidate SMBH binary, is thought to be in an orbit with semimajor axis $a_{\rm b} \sim 0.1$ pc (Sillanpaa et al., 1988; Pursimo et al., 2000). There is some evidence from periodic AGN variability that binaries at extremely small separations (with periods of fractions to tens of years) may be more common (e.g. Merritt and Milosavljević, 2005), but if these are indeed due to binary orbits the semimajor axes would have to be on the order of au, so they are well below the 0.01 pc needed for gravitational waves to become an efficient emission mechanism.

This lack of evidence for binary stalling at parsec separations, known as the 'final parsec problem', requires that some mechanism be found to bridge the gap from separations of ~ 1 pc to ~ 0.01 pc. Circumbinary accretion discs could play an important role in this process, as it has been long known that they can strongly affect the dynamics of the bodies they form around (e.g. Chapter 1, Goldreich and Tremaine, 1979, 1980). However, the time-scales to drive direct binary coalescence by the usual mechanism of disc resonances are far too long to explain the observations (e.g. Cuadra et al., 2009; Lodato et al., 2009). Instead, Cuadra et al. (2009) showed that the rate of SMBH eccentricity growth due to a circumbinary disc is $\dot{e} \sim 10^{-4} \, \Omega$. For a binary of mass $10^7 \, M_\odot$ and semimajor axis of 1 pc, this gives a time-scale for eccentricity growth of $t_e \sim 10^7$ yr.

Recent work by Nixon et al. (2011a,b, 2013) has shown that misaligned or counter-rotating discs may be a possible avenue towards coalescing binaries in this way. They find that a circumbinary disc rotating in a retrograde direction with respect to the orbit of the binary components interacts much more strongly with the binary. This is due to the absence of resonant torques in the retrograde case. For prograde discs, these torques act to hold disc material away from the binary, stopping it from interacting directly by accreting large amounts of disc material. In contrast, retrograde discs can directly accrete disc material and by conservation of linear and angular momentum their orbits show much greater evolution (Nixon et al., 2011a).

Another interesting aspect of disc misalignment is the discovery of the phenomenon

of disc tearing (Nixon et al., 2012, 2013). By including some precession effect in a disc with a strong warp (in the circumbinary case, the precession is caused simply by the non-axisymmetry of the binary; the warp is a manifestation the short alignment time at the very inner disc radius compared with the rest of the disc), if the viscosity in the disc is not strong enough to communicate the precession through the warp the disc can 'tear' into distinct planes. The intersection points between the different planes allow efficient angular momentum cancellation between the two flows so some gas can simply fall straight into the centre and be accreted by one of the SMBHs (Nixon et al., 2013).

4.1.2 Chaotic accretion

However, even if counter- or mis-aligned discs can drive the binary across the final parsec, we must first answer the question of why one might expect the discs to be out of the plane of the binary in the first place. One way to justify this is to assume that for any given accretion event, the initial gas that forms the disc does not 'know' the orientation of the binary plane, and therefore the angular momenta of the disc and binary will be uncorrelated. This assumption can in turn be justified by invoking the chaotic accretion paradigm developed by King and Pringle (2006).

In order to explain very massive SMBHs at very high redshifts ($M_{bh} \sim 10^{8-9} M_{\odot}$ at $z \sim 6 - 7$; e.g. Mortlock et al., 2011; Willott et al., 2013), this paradigm invokes a large number of small, randomly orientated accretion events. This is motivated by a requirement that the black hole spin be kept small. A small spin allows a lower value for the black hole's radiative efficiency (King and Pringle, 2006), which describes how much of the accreted material's mass-energy is radiated away as accretion luminosity compared to how much is actually accreted onto the black hole. The accretion efficiency is set by the radius of the black hole's ISCO, which is larger for a black hole with high (positive) spin – a larger ISCO gives lower accretion efficiency, so a black hole with low spin accretes more efficiently (King and Pringle, 2006).

The spin of the black hole is affected by the orientation of the angular momentum vector of the accreted material. Prograde (conversely, retrograde) accretion events are defined as those with angular momenta closer to being aligned than counter-aligned (conversely, closer to being counter-aligned than aligned) with the black hole spin. King and Pringle (2006) used analytical arguments to show that prograde and retrograde accretion events cause the black hole to respectively increase or decrease its spin. Notably, an otherwise identical event must cause more spin-down if it occurs in the retrograde direction than it would cause spin-up if it was prograde. This is due to the difference in angular momen-

tum of material at the ISCO, as the radius of the ISCO is larger for a retrograde disc. The process of many small events with random orientations allows the spin to remain small, permitting the black hole to accrete large amounts of mass in short times. King et al. (2008) showed that the net effect of disc accretion is a small (prograde) black hole spin, and this is largely unaffected by mergers of pairs of black holes. Crucially, they also found that the spin axis of the black hole changes by large values over a number of accretion events, making misalignment inevitable.

We therefore have a reason to expect misalignment between binary SMBHs and an infalling cloud of gas, arrived at independently of the desire to drive binary coalescence. There is also observational evidence that the angular momentum of the central object in a galaxy and the angular momentum of the galaxy itself indeed uncorrelated. This comes from observations of AGN jets, which are indeed misaligned from the axis of the host galaxy in nearly all cases (e.g. Kinney et al., 2000). It is unclear if the jet should be aligned to the spin axis of the black hole or of its accretion disc (e.g. Nixon and King, 2013), but one of these components must be misaligned to the plane of the galaxy. This is strong evidence that by the time they (either the black hole or infalling gas) reach the inner $\sim$ pc of the galaxy, any alignment with the larger galaxy is lost. Misaligned discs have also been invoked to explain the circumnuclear stellar rings at the centre of the Milky Way (e.g. Levin and Beloborodov, 2003; Alexander et al., 2012; Lucas et al., 2013).

4.1.3   Dynamical effect of misaligned discs

The exact effect of an individual accretion event onto a SMBH (binary or single) depends strongly upon the cloud's mass and angular momentum, and especially how they are related to the mass and angular momenta of the black holes. For a disc counter-aligned to the spin of a SMBH binary, the binary eccentricity can grow to very high values if the secondary (the least massive binary component) captures[1] approximately its own mass in disc gas (Nixon et al., 2011a). The exact (mis-) alignment between the binary and disc angular momentum vectors becomes important here because the change in eccentricity is caused by the angular moment cancelling – so an exactly counter-aligned can grow a greater eccentricity than a nearly counter-aligned one.

While this requirement is not going to be fulfilled for any single accretion event (which as we have seen must be small), chaotic accretion allows a large number of such events to be invoked – indeed, it requires them to grow an SMBH from a stellar mass black hole in

---

[1] Note that it does not have to actually accrete this gas, merely capture it into a bound orbit.

Gyr time-scales (King and Pringle, 2006). Therefore while one single event may not be able to cause significant eccentricity growth or semimajor axis evolution, small changes summed over many events can cause huge changes in the dynamics of the system.

Torques from the binary on the disc will cause a misaligned disc to co- or counter-align on some (radius dependent) alignment time-scale. The exact outcome depends on the relationship between the angular momentum of the binary j_b and that of the disc j_d with angle θ between them. If counter-alignment is to occur, King et al. (2005) showed[2] that

$$\cos \theta < -\frac{|j_d|}{2\,|j_b|} \tag{4.1}$$

is required, leading to the conditions for counter-alignment being that

$$\theta > \frac{\pi}{2}, \qquad |j_d| < 2\,|j_b| \tag{4.2}$$

We expect in the chaotic accretion scenario that $|j_d|$ would be comparatively low for an infalling gas cloud, both due to the fact that it is falling to the centre of the galaxy (i.e. on a low angular momentum orbit) and that its mass is expected to be low. Therefore the second condition for stable counter-alignment is likely to be unfulfilled in a number of cases, leading to somewhat less than half of the events resulting in a counter-aligned disc, assuming an isotropic distribution of events.

Numerical tests of alignment and retrograde accretion discs have to date started with an *a priori* assumption than an infalling cloud will form a smooth axisymmetric accretion disc (e.g. Nixon et al., 2011b; Nixon, 2012), while analytical arguments typically consider only simple angular momentum vectors (e.g. King and Pringle, 2006). While this is clearly expected from accretion disc theory, in this chapter I will test this assumption by simulating the infall of a cloud within the chaotic accretion paradigm with some initial misalignment and see how this translates into the final accretion disc, and also to investigate the differences between how the system evolves given different initial configurations.

4.1.4   Gravitationally unstable accretion discs

A large amount of work has been put into characterisation and understanding of gravitational instability in discs in recent years. While much of this work has been in the context of understanding planet formation in protoplanetary discs by gravitational fragmentation (e.g. Cossins et al., 2009; Harsono et al., 2011), a large amount of effort has been put into

---

[2] Although the original calculation was for alignment between black hole spin and the disc, the result is the same when considering alignment due to binary torques (Nixon, 2012)

understanding the fragmentation of accretion discs in the centres of galaxies (e.g. Levin, 2007; Alexander et al., 2008; Bonnell and Rice, 2008). A strong driver of this work has been the discovery of rotating stellar rings around Sgr A⋆ (Levin and Beloborodov, 2003; Genzel et al., 2003; Paumard et al., 2006; Lu et al., 2009).

There are two aspects to understanding the process of gravitational instability: how it manifests initially, and how it evolves from there. The first part is relatively easy to understand in general terms – a disc is gravitationally unstable simply if it is cold and massive enough that it cannot support itself against its own self-gravity.

The three parameters that matter the most are therefore the disc surface density, temperature and rate of shear. The first is controlled by the disc hydrodynamics, the second by the thermodynamics of the gas (e.g. Levin, 2007) and the latter by a convolution of the thermodynamics and mass of the central object (a pressure-dominated disc has a lower orbital speed than a purely Keplerian disc, and therefore has less rotational support against gravitational collapse – see Section 1.2.1). For numerical simulations such as those presented in this chapter, the last two terms are primarily controlled by the cooling prescription used.

Once a disc has become gravitationally unstable, the situation becomes less trivial to understand. If the disc is unstable over a broad radial range, one expects large spiral arms to develop. These allow the disc to regulate itself against its self-gravity, with the shocks produced allowing heating (Cossins et al., 2009), and they also act to transport angular momentum. Lodato and Rice (2004) and Forgan et al. (2011) found that for discs less massive than 25 percent of the mass of the central body[3], this transport is well approximated as a local effect and can therefore be parametrized as an α-type viscosity. In this case a steady-state is reached where the disc is able to regulate the instability by balancing heating from spiral shocks with cooling. However, for more massive discs the effect is less easily parametrized, as there is no steady state. Accretion outbursts and short-lived low-m spiral modes manifest, and the transport cannot be well approximated as a local α over any reasonable radial or temporal range (Lodato and Rice, 2005; Forgan et al., 2011; Balbus and Papaloizou, 1999).

An alternative possible outcome for an unstable disc is that it may fragment into bound clumps. Gammie (2001) found that in a shearing box this occurs where the condition

$$t_{\text{cool}} \lesssim \frac{3}{\Omega} \tag{4.3}$$

---

[3] This was done in the context of circumstellar discs, but the process is nominally scale-free provided that one assumes that the cooling times assumed also scale similarly. While this is not true in general it is only likely to affect the timescales on which the instability manifests.

is satisfied, where t_{cool} is the cooling time of the gas. This has broadly been confirmed in global disc simulations although the exact criterion depends on various details of the simulations (Rice et al., 2003, 2005; Forgan et al., 2011)[4].

Toomre (1964)'s parameter Q_{T} is an effective way of measuring how and where the disc becomes unstable, and is given by

$$Q_{\mathrm{T}} = \frac{c_{\mathrm{s}}\kappa}{\pi G \Sigma} \simeq \frac{c_{\mathrm{s}}\Omega}{\pi G \Sigma} \tag{4.4}$$

where κ and Ω are the epicyclic and orbital frequencies, with the right hand side only being strictly true for a Keplerian disc. This parameter measures how stable the disc is against the self-gravity caused by some perturbation. In the original derivation (which considered an axisymmetric perturbation to a ring of stars), $Q_{\mathrm{T}} \leqslant 1$ is the condition for gravitational instability to occur. In the case of non-axisymmetric perturbations, the value depends on the dominant Fourier mode m of the perturbation, but becomes $Q_{\mathrm{T}} \lesssim 2$ in the limit of high-m (Lodato, 2007; Durisen et al., 2007).

4.2   Simulations

To investigate how the initial configuration of an infalling cloud affects the final disc formed around a SMBH binary system, and how the process affects the binary itself, I again use high-resolution SPH simulations. I have run four simulations of a turbulent gas cloud falling on to a binary with mass ratios $q = M_2/M_1 = 1/3$ and $q = 1/10$ for primary and secondary masses M_1 and M_2 respectively, and the total mass of the binary being $M_{\mathrm{b}} = 10^7 \, M_\odot$. For each mass ratio, I have run one simulation of a cloud moving close to prograde with respect to the orbit of the binary, and one close to retrograde. The cooling prescription used is described in Section 4.2.1, and the initial setup of the clouds is described in Section 4.2.2. The outcomes of the simulations and an analysis & discussion are presented in Sections 4.3 and 4.4, and I draw conclusions in Section 4.5.

4.2.1   Cooling

As described in Section 1.4.1, in this Chapter I abandon the isothermal equation of state used in Chapters 2 and 3, as this approximation is no longer valid for the situation under

---

[4] Most of these studies use the traditional β parameterisation of cooling. Although problems of numerical non-convergence have recently been identified with this scheme by Meru and Bate (2011), it still seems that this threshold approximately holds for realistic discs (Rice et al., 2012).

consideration. In the former cases, the disc saw (approximately) constant heating from the central star(s), but in the case of a gas cloud falling into the central region of a galaxy no such heating source exists. It is therefore necessary to consider the time-dependent temperature of the gas.

The exact thermal behaviour of the gas requires detailed radiative-transfer calculations, which are difficult to carry out even when the exact geometry of the problem is known (e.g. the calculations of Bitsch and Kley, 2011a, where the geometry of the initial conditions do not evolve significantly during the simulations). In the situation I consider here, the very nature of the simulations performed is to investigate what disc geometry forms and how it evolves – rendering the problem even harder.

While approximations of radiative cooling have been implemented in SPH in the past (Stamatellos et al., 2007; Forgan et al., 2009), the accuracy of these schemes is disputed (Wilkins and Clarke, 2012, although it is unclear how valid the arguments are). I avoid the potential pitfalls of these methods by adopting an explicit cooling prescription, chosen to mimic optically thin cooling.

The rationalisation behind such a cooling law is that the infalling cloud is expected to be of relatively low mass, as noted in Section 4.1.2. At low temperatures ($T \sim 100$ K), the dominant cooling mode is simple thermal radiation, and the opacity is approximately constant with temperature and density in this regime (e.g. Semenov et al., 2003).

In the optically thin regime, the mean free path of an emitted photon is greater than the physical size of the cloud, and so the cooling rate is independent of the geometry of the problem. In this case the cooling rate is expected to scale inversely with the density. This is because the cooling time depends only on the rate of photon emission, which scales with the mass (density) of the radiating gas.

The exact parameterisation follows the form of the widely-used 'beta-cooling', but modified to follow this $\propto \rho^{-1}$ scaling. The internal energy ϵ of the gas

$$\frac{d\epsilon}{dt} = -\frac{u}{t_{\text{cool}}} \tag{4.5}$$

where the cooling time t_{cool} of the gas is set to be

$$t_{\text{cool}} = \frac{\beta}{\Omega_{\text{ref}}} D \tag{4.6}$$

with

$$D = \max\left[1, \frac{\rho_0}{\rho}\right]. \tag{4.7}$$

Ω_{ref} is equal to the Keplerian orbital frequency at a reference radius ($R = 5$), taken to be the approximate radius of maximum surface density in the discs formed in test runs where no cooling was included. ρ_0 is set to the density at one scale-height H at that same radius in the same test runs.

The decision to cap D at $\rho = \rho_0$ is to allow very low density gas to cool more efficiently. While somewhat unphysical, it prevents a situation where the bulk of the computational time is spent integrating orbits for a large number of super-heated particles at very large radii, and test runs indicate that it does not significantly affect the results beyond slightly increasing the amount of gas that ends up bound to the binary.

This implementation results in gas which cools exponentially, with the term $d\epsilon_a/dt$ used to evolve the entropy function A_a of particle a (Equation 1.70) given by

$$\frac{d\epsilon_a}{dt} = \begin{cases} -\epsilon_a \frac{\Omega_{\text{ref}}}{\beta} & \text{for } \rho \leqslant \rho_0, \\ -\epsilon_a \frac{\Omega_{\text{ref}}}{\beta} \frac{\rho}{\rho_0} & \text{for } \rho > \rho_0. \end{cases} \tag{4.8}$$

Although this implementation is by no means exactly representative of reality, it is a close approximation of the expected scaling relation in the regime (where the gas is optically thin to the cooling radiation) broadly expected to hold under real conditions, although in reality the disc may be optically thick in the hotter regions. The simplicity of the parameterisation has the added benefit that its behaviour under given conditions are easily calculable, and so there should be no unexpected effects. This makes interpreting the results of the simulations a far simpler affair, and makes understanding how they differ from reality possible.

The free parameters in the cooling used are β and ρ_0, which are unchanged between the simulations, with $\beta = 20$ and $\rho_0 = 8 \times 10^{-4}$ in code units the values used. While an important aspect in interpreting the results is how they are affected by these variables, I chose to restrict the parameter space explored by the simulations to the initial cloud orbits and binary mass ratios due to limited computational resources.

The parametrization and values chosen effectively ensure that the discs will eventually fragment into stars, as it sets the cooling to be fastest in areas of high density – areas of high surface density are automatically going to cool the quickest and become the most unstable to self-gravity. While this is not a physically unreasonable expectation (especially given the observations of stellar rings in the galactic centre and the findings of similar simulation work, e.g. Bonnell and Rice, 2008; Lucas et al., 2013) it must be borne in mind that other choices of parameters may alter the outcomes entirely. To demonstrate this, consider the effect of the parameter ρ_0: as it is set to a value below the expected mid-

plane density, the midplane of the disc must cool and become denser, thus cooling faster. Eventually ρ_0/ρ becomes small enough that t_{cool} fulfils Gammie (2001)'s condition for fragmentation (Equation4.3). Once fragmentation occurs, the cooling prescription again acts to ensure that the fragments cool further and collapse. However, this is not an unphysical process. We know that bound, dense gaseous clouds do indeed become unstable to self-gravity in their densest regions – this is how stars are born.

4.2.2   Turbulent cloud

The cloud was set up so that it initially had constant density, and a turbulent velocity field imposed upon it using a method similar to Bate et al. (2003) and Bonnell and Rice (2008). The imposed field is divergence free (i.e. $\nabla \cdot \boldsymbol{u} = 0$). The turbulent field is constructed in Fourier space on a 64^3 grid, and real values for the velocity are found by taking fast inverse Fourier transforms interpolated from the grid. The power spectrum of the turbulence is $P(k) \propto k^{-4}$ for wavenumber k, to match observed Larson (1981) turbulent scaling relations.

The kinetic energy of the turbulence was set to half of its potential energy (i.e. it has virial energies), and the cloud was given a bulk velocity that put it on an eccentric orbit around the barycentre of the binary. Initially at apocentre with a distance of 10 times the binary semimajor axis, the eccentricity of the orbit was set to $e_c = 0.667$ so that its pericentre distance would be equal to the initial cloud diameter (although evolution of the cloud during its passage means that this is no longer strictly the case by the time the cloud reaches the binary). The initial plane of the orbit was offset by $15°$ to the plane of the binary for the 'prograde' cloud, and $165°$ for the 'retrograde' cloud.

The turbulence acts to support the cloud against initial catastrophic collapse. Indeed, real GMCs in the Galaxy are known to be highly turbulent (e.g. Larson, 1981), so it also adds some semblance of realistic initial conditions for the simulation. However, by the time the cloud has reached the binary much of the initial turbulent energy has been dissipated and tests show that the exact details of the setup do not affect the results.

4.2.3   Units and other parameters

The system of units adopted in these simulations set the unit of distance equal to the initial binary semimajor axis a_b. The unit of mass is the total mass of the binary components M_b, and the unit of time to be the initial binary orbital period T_b, setting $G = 4\pi^2$.

As in the previous chapters, the binary components are modelled as N-body partciles

q	Direction	T_{end} [T_{b}]	e_{end}
1/3	Prograde	393	0.0016
1/3	Retrograde	511	0.0033
1/10	Prograde	405	0.0035
1/10	Retrograde	515	0.011

Table 4.1: Summary of simulation outcomes.

with sink radii equal to 0.15 a_{b}. This is a large value, but is required to ensure that the simulation is not slowed by a few particles on very short timesteps inside the binary. SPH particles are also removed from the simulation if they are unbound from the binary and are at a radius greater than 500 a_{b} from the binary centre of mass. As the binary components are 'live', their orbital elements are allowed to evolve throughout the simulations. The mass of each cloud M_{c} is set to 10^{-2} of the total binary mass. For $a_{\mathrm{b}} = 1$ pc and $M_{\mathrm{b}} = 10^7\,M_{\odot}$, this gives $M_{\mathrm{c}} = 10^5\,M_{\odot}$ and sets $T_{\mathrm{b}} = 2.96 \times 10^4$ years.

4.3   Results

Table 4.1 shows a summary of the state of each simulation at the end of the run. Figure 4.1 shows a series of surface density renderings of the simulation at 4 times during the simulation for the prograde cloud and $q = 1/3$, and Figure 4.2 shows the same for the retrograde case. The major difference between the simulations with different q values is the level of eccentricity gained by the binary, but in qualitative terms at least the simulations with the same q proceed identically regardless of the direction of the cloud's initial orbit.

The cloud becomes smeared out into a stream upon first passage past the binary (top right panels in both Figures 4.1 and 4.2), and over repeated passages this forms into a disc (bottom left panels). Eventually the initial turbulent energy dissipates through shocks and the imposed cooling scheme, and the disc becomes laminar (bottom right panels). In both cases the eccentricity of the final disc is approximately $e_{\mathrm{d}} = 0.1$. The vertical structure of the disc is resolved into $4 - 6$ SPH smoothing lengths at the midplane, and have aspect ratios in the region $0.02 \lesssim H/R \lesssim 0.05$. This is somwhat thicker than is thought typical for AGN discs (where it is thought to be of the order 10^{-3}, although for self-gravitating discs $\sim 10^{-2}$ may be possible; Lodato, 2007) but is necessary to meet the resolution requirements of Nelson (2006). As shown in Table 4.1, for a given q the final eccentricity of the binary is higher in the retrograde case than the prograde case by a factor
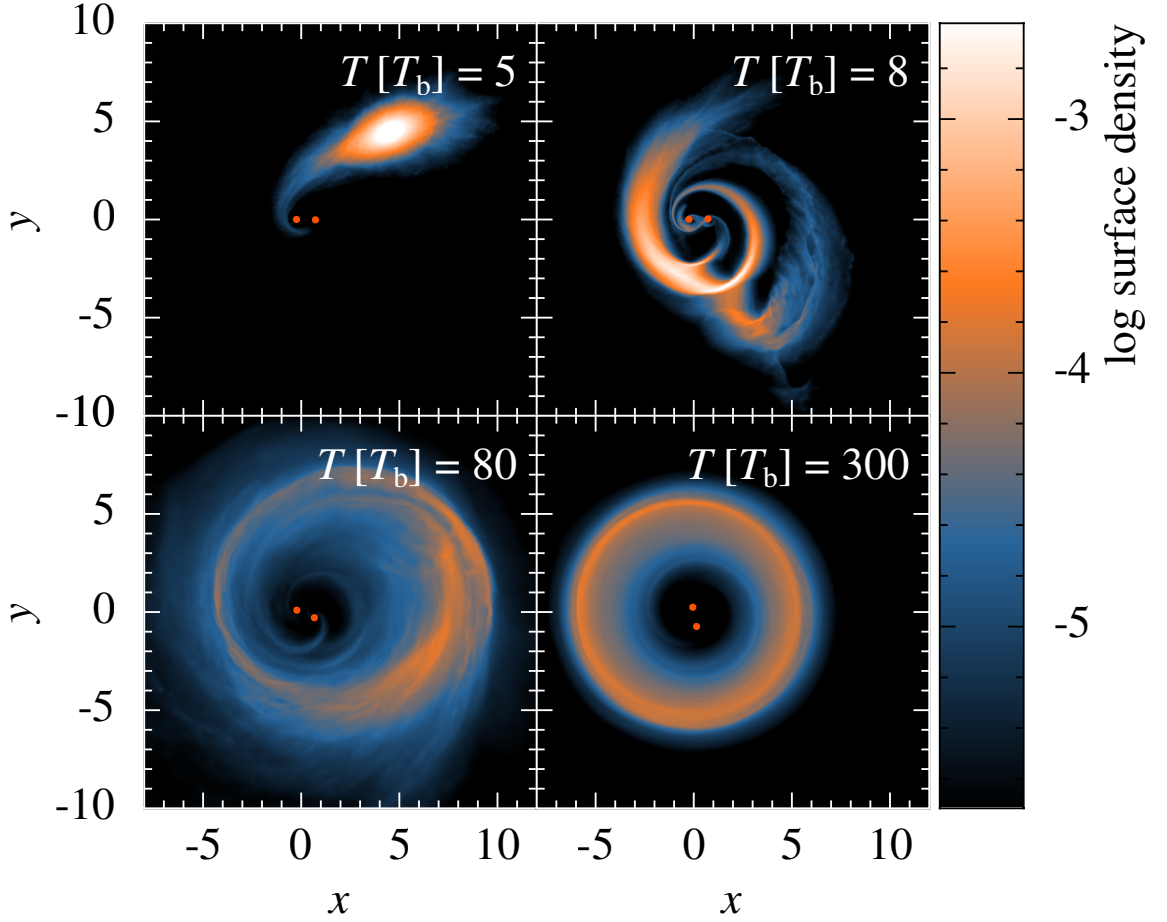
Figure 4.1: Surface density renderings for the simulation of an infalling cloud onto a SMBH binary with $q = 1/3$. The initial orbit of the cloud is prograde with respect to the binary orbit. Time in binary orbits is given in each panel.

———————————————————————————————

of $\sim 2 - 3$.

As the disc continues to evolve and cool, it eventually becomes unstable to its own self-gravity as outlined in Section 4.1.4. As the disc cools, the sound speed c_s decreases and so in the regions where surface density Σ is high, Q_T (Equation 4.4) will be at a minimum. Indeed, given the inverse relation between the cooling time and density in our cooling prescription, the highest density regions will continue to cool at a faster rate.

Simply by comparing the bottom right panels of Figures 4.1 and 4.2 it is obvious that this instability will proceed differently in the prograde and retrograde cases, and this is shown more explicitly in Figure 4.3, which shows the azimuthally-averaged surface density of the discs in each case (again, for the simulations with $q = 1/3$) at $T = 350\,T_b$.

While the distributions peak at the same radius, it is obvious that the prograde peak
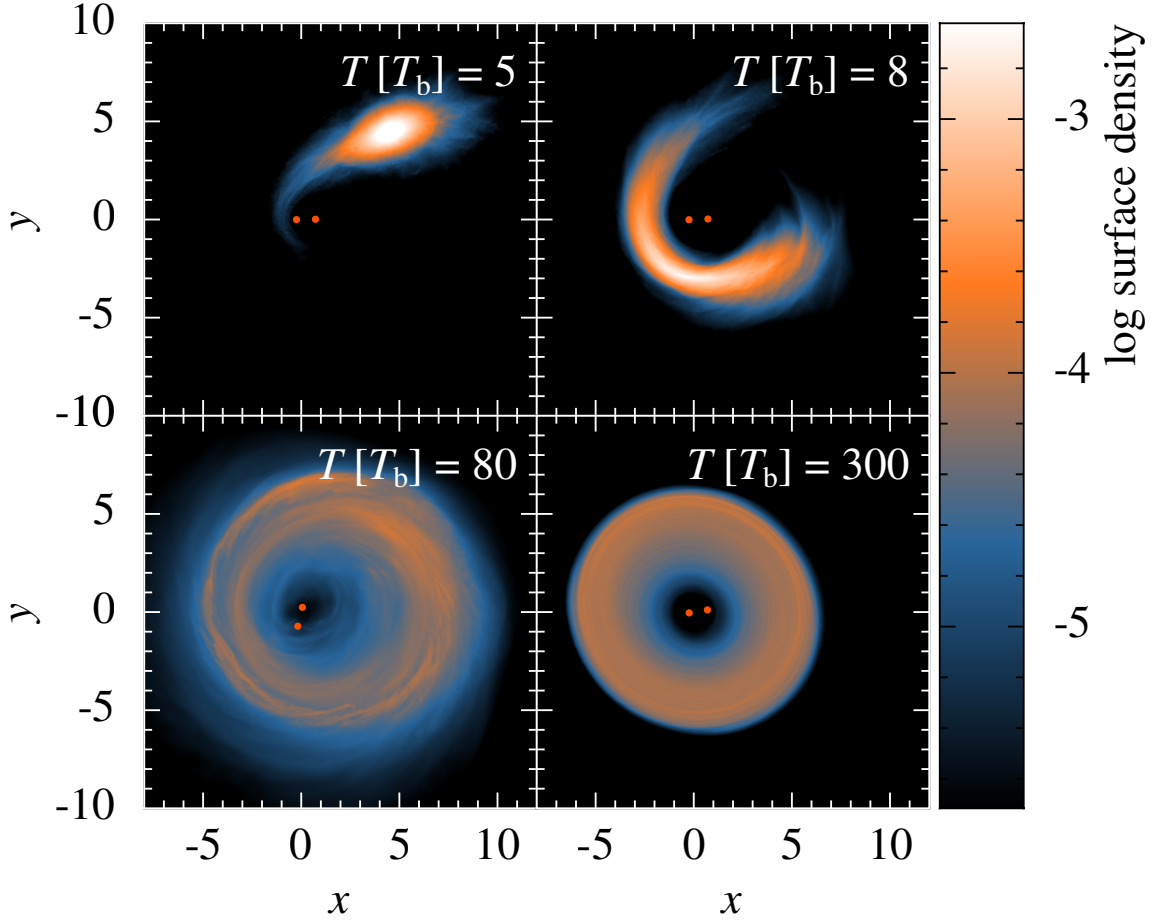
Figure 4.2: As Figure 4.1, but for a cloud with initial orbit retrograde with respect to the binary orbit. Comparing this sequence with that shown in Figure 4.1 shows that while the qualitative behaviour is identical, the disc is able to spread closer to the binary, as it is not held out by Lindblad resonances as in the prograde case.

---

is higher and narrower than in the retrograde case, which has two consequences for the evolution of the discs. Firstly, if the disc cools sufficiently that gravitational instability ensues, its onset will be sooner in the prograde case than in the retrograde case. Secondly, if this occurs the disc will be unstable over a larger area in the retrograde than in the prograde case. I discuss the physical implications of both of these facts for the evolution of real SMBH binaries in Section 4.4.

The simulations were stopped at the point where the disc began to form bound clumps that will continue to collapse under their own gravity and according to the prescribed cooling rate. Physically this process represents the onset of star formation, and I do not attempt to simulate this process. The time T_{end} in the simulations at which this occuredf
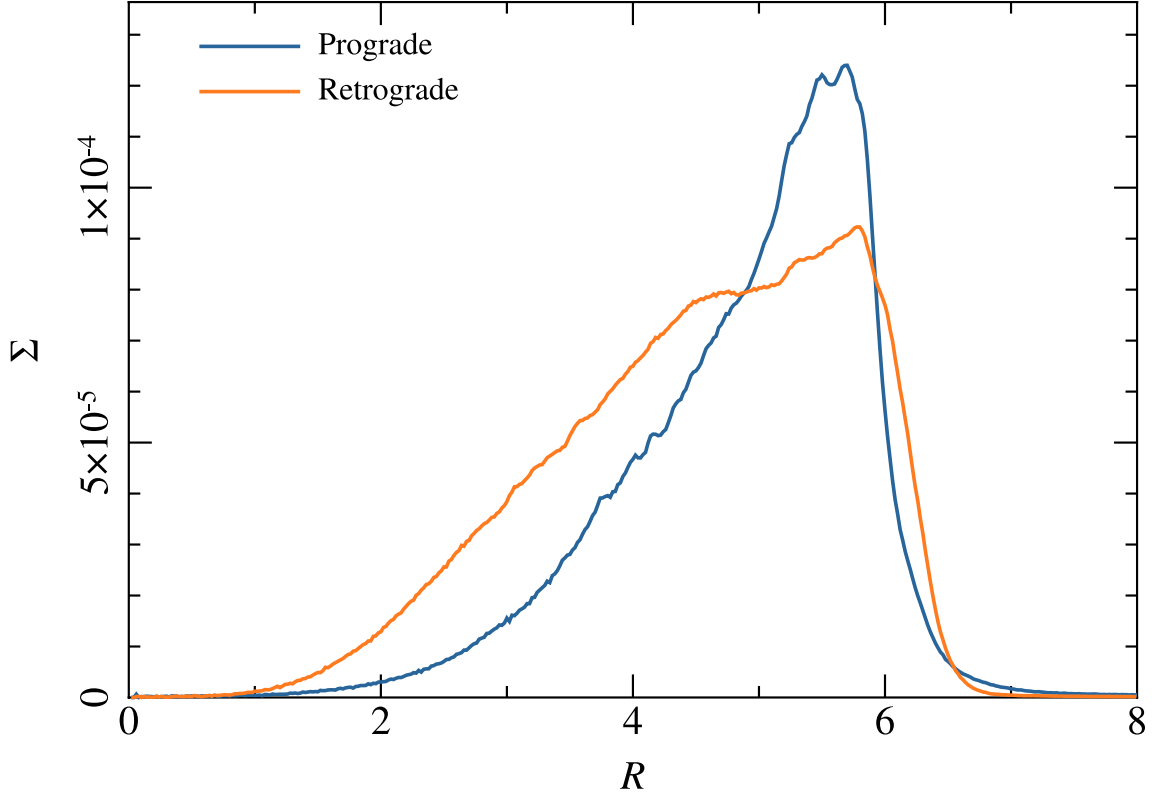
Figure 4.3: Azimuthally averaged disc surface density distributions for prograde and retrograde cases where $q = 1/3$, at time $T = 350\,T_{\mathrm{b}}$.

is given in Table 4.1. Again, I provide a physical interpretation of the differences here in Section 4.4.

None of the four simulations show significant alignment between the disc and the plane of the binary orbit. This is not unexpected, as the alignment time-scale is thousands of dynamical times for a disc of negligible mass (Nixon, 2012). The dynamical time[5] at $R = 5$ in the discs here is $T_{\mathrm{dyn}} = 1.8\,T_{\mathrm{b}}$, so they only live for $\sim 700 - 900$ dynamical times before the simulations end.

The only major quantitative effect of changing q between the two sets of simulations (besides minor differences in the times at which the simulations end) is the extent to which the binary eccentricity is affected. Nixon et al. (2011a) showed that a retrograde disc can drive the binary eccentricity to near unity if the secondary is able to capture its own mass in gas from the disc. The total mass of the clouds simulated here is only 10^{-1} of the

---

[5] The dynamical time is defined as the time for a particle to crosss a system. This is given by $T_{\mathrm{dyn}} = T_{\mathrm{orb}}/2\pi$.

secondary mass for the most extreme mass ratio simulated, so eccentricity growth is not expected to be large for any single, low mass infalling cloud.

While it is notable that the eccentricity change is smaller than a percent in most cases (the exception being for a more extreme mass ratio and a retrograde cloud, and then only just above one percent), the chaotic accretion paradigm predicts a large number of these small events over Myr to Gyr time-scales in order to drive rapid growth of black hole masses ($\sim$ 100 accretion events in somewhat less than a Gyr; King and Pringle, 2006). This would therefore lead to an eccentricity in excess of $e = 0.5$ and aid the onset of rapid gravitational wave-driven binary coalescence.

4.4 Discussion

4.4.1 Disc evolution

As described above, the infalling cloud forms a disc around the SMBH binary in each case regardless of the initial orbital sense of the cloud with respect to the orbit of the binary. Due to the length of the simulations, the discs do not see any significant alignment with respect to the plane of the binary, but lower-resolution versions which ran for longer confirm that this indeed would occur at later times[6].

As noted by Nixon et al. (2011a), a prograde disc is effectively held out at a distance from the binary by Lindblad resonances, but an otherwise identical retrograde disc experiences no such resonances. That this is true is shown simply by considering the condition for resonances exterior to the orbit of the binary (i.e. OLRs, where $r > a_\text{b}$, Equation 1.34):

$$\Omega_\text{d}(R) = m\left(\Omega_\text{d}(R) - \Omega_\text{b}\right)$$

or more simply that

$$\Omega_\text{b} = \Omega_\text{d}(R)\left(1 - \frac{1}{m}\right).$$

where $\Omega_\text{d}(r)$ is the orbital frequency at radius r in the disc, Ω_b is the binary orbital frequency and m is an integer value. For a retrograde disc, Ω_b has the opposite sign to Ω_d, so the condition is never satisfied for $m > 0$ OLRs. While ILRs can satisfy the criterion for resonance in Equation 1.34, there is no gas interior to the orbit of the secondary in these

---

[6] These lower resolution simulations ran for longer times because the effect of reduced resolution is the smoothing out of high density regions – therefore the formation of dense bound clumps occurs at much later times at low resolution.

simulations so this has no effect.

The effect of this is that, as shown in Figure 4.3, the retrograde disc is able to spread much closer to the binary. The implications of this for the evolution of the system are twofold. The first of these is that while the disc survives, the retrograde disc is able to feed more gas to the binary. (Nixon et al., 2011a) showed that an approximately equal-mass binary (where $q \gtrsim 1/10$) both components will accrete an even amount of gas, but for more extreme mass ratios the secondary will accrete the majority. Regardless, while the disc persists a retrograde disc is a more efficient route to feeding (and growing) the SMBHs. Furthermore, as noted above the onset of gravitational instability is delayed significantly in the retrograde case. Therefore not only is a retrograde disc more capable of feeding gas to the binary, it is able to do so for longer before fragmenting into stars.

4.4.2   Star formation

In Section 4.3 I noted that the simulations evolve to the point where dense bound clumps form and end there. These bound clumps represent the onset of star formation, and are a natural result of the cooling prescription employed. Figure 4.4 shows Toomre's Q_{T} parameter (Equation 4.4) as a function of radius at the end point of the prograde and retrograde simulations with binary mass ratio $q = 1/3$. The details of what exact value of Q_{T} represents gravitational instability are somewhat uncertain and depend on the specific problem, but $Q_{\mathrm{T}} \lesssim 2$ is the quoted value for high-m perturbations (i.e. those which do not manifest as large spiral arms but instead are local and tightly-wound, such as those clearly shown in Figure 4.5; Lodato, 2007).

Q_{T} as calculated using the approximate version in Equation 4.4 is clearly not an exact value, as the disc is not perfectly Keplerian. However it is clear from Figure 4.4 that in this case instability is present at $Q_{\mathrm{T}} \sim 2$, as expected. Given the sharpness of the prograde surface density profile in Figure 4.3, it is unsurprising that the Q_{T} profile is also strongly peaked, and the converse statement is true for the retrograde case.

The physical implication of the Q_{T} profiles shown in Figure 4.4 is that if star formation is to occur, it will produce very a narrow stellar ring in the prograde case, and a wider ring of stars in the retrograde case. Figure 4.5 shows a surface density rendering of each case at the end of the simulations, and indeed it is plain that the annulus in which high density clumps have formed is far thinner radially in the prograde case (left hand panel) that that which forms in the retrograde disc (right hand panel). Such stellar rings are a common outcome of simulations of gas discs at the centre of galaxies (e.g. Bonnell and Rice, 2008; Mapelli et al., 2012; Amaro-Seoane et al., 2013).
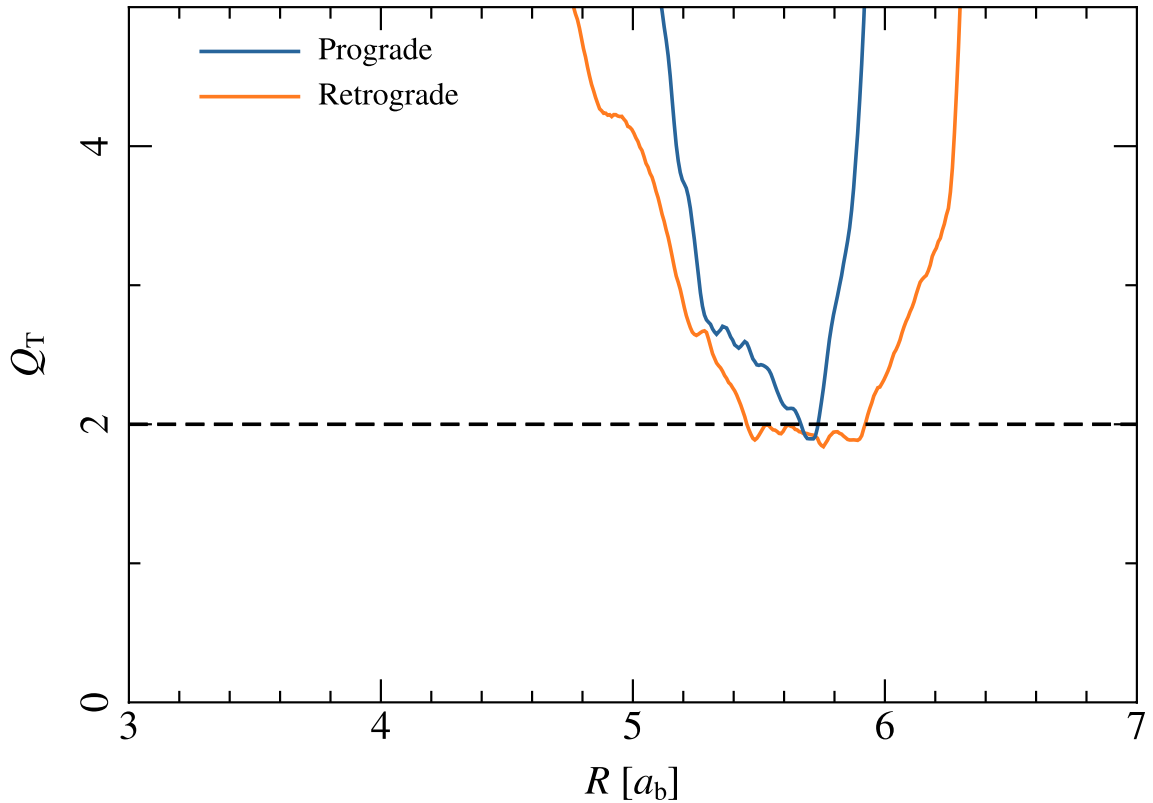
Figure 4.4: Toomre Qs (calculated using Equation 4.4) as a function of radius at the end of the simulation for prograde and retrograde discs, from runs with binary mass ratio $q = 1/3$. The horizontal dashed line denotes $Q_T = 2$.

_____

It is therefore interesting to consider the future evolution of the resultant stellar ring, in particular if and how they may further influence the binary. Recall that gravitational scattering of stars is able to drive the binary down to parsec separations. The gap in phase space opened by this process (due to stars receiving gravitational kicks from the binary and moving out of that region of phase space) is called the 'loss cone'. A broad definition of the loss cone is that it is the region of phase space inhabited by stars on orbits which take them close enough to the binary to interact strongly with it (Merritt, 2013).

While it is evident that stars forming in an approximately circular ring around the binary will not immediately repopulate the loss cone, it is still possible that individual stars may be scattered away from their initial orbits onto orbits that do put them in the loss cone. Gualandris et al. (2012) took the stellar rings formed in the SPH simulations of Mapelli et al. (2012) and performed N-body integrations to examine the long term behaviour of the initially thin disc of stars formed therein. They found that the discs thicken in both
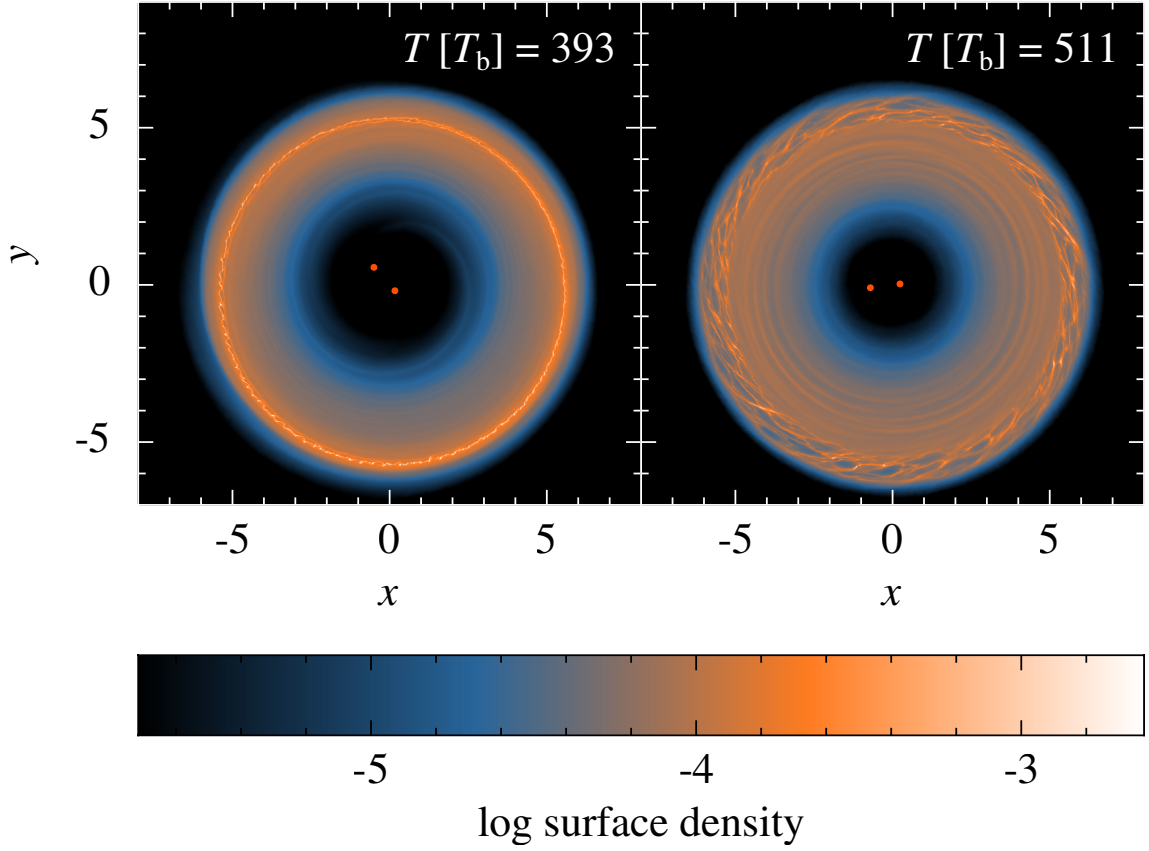
Figure 4.5: Surface density rendering of prograde (left panel) and retrograde (right panel) simulations for a binary mass ratio $q = 1/3$ at the end of each simulation. Each disc is plainly gravitationally unstable in the outer regions, but the radial range of the instability is much smaller in the prograde case. This has strong implications for the evolution of the system after star formation occurs.

_____

the R and z directions due to the growth of orbital eccentricity and inclination on time-scales of a Myr. The radial extent of these rings is a factor of 10 smaller than the ones in the simulations presented here, but the smaller mass of Sgr A* further complicates the comparison somewhat. The same time-scales in the discs described here would be longer by a factor F simply due to the different scales involved, without considering other differences (e.g. initial mass, eccentricity and thickness of the stellar rings). F is given by

simply comparing the orbital frequencies of the two discs, giving

$$
\begin{aligned}
F &\sim \frac{\Omega_{\text{Gualandris et al. (2012)}}}{\Omega_{\text{This work}}} \\
&= \sqrt{\frac{GM_{\text{Sgr A}^\star}}{(R = 0.5\,\text{pc})^3}} \bigg/ \sqrt{\frac{GM_\bullet}{(R = 5\,\text{pc})^3}} \\
&= 20.
\end{aligned}
\tag{4.9}
$$

This at least gives an order of magnitude of a few tens of Myrs as the time-scale for the stellar rings produced in the simulations here to spread in phase space and begin to repopulate the loss cone, driving further evolution of the SMBH binary. Further work in the form of N-body simulations or the inclusion of a star-formation prescription into the code is necessary to fully quantify how the process proceeds, but broad conclusions can be drawn without doing so.

Alexander et al. (2007) found that the relaxation time for a stellar ring with scale height H and radial thickness ΔR centred at radius R with velocity dispersion σ follows the relation

$$
t_{\text{relax}} \propto R_0 \Delta R \, \sigma^4.
\tag{4.10}
$$

Using this relation and measuring R_0, ΔR and σ for the SPH particles in the unstable region in each case, the relaxation time for a prograde stellar disc is at least an order of magnitude shorter than for the retrograde case.

However, Amaro-Seoane et al. (2013) found that the binary evolution from stars formed in a circumbinary disc is an order of magnitude or more slower than that caused by the circumbinary disc of Cuadra et al. (2009). It is therefore not clear how important this difference in relaxation time is, and further work is necessary to address this uncertainty. It is possible that binary stars may be able to help the situation somewhat ejecting stars from the ring, both through three-body interactions (e.g. Perets et al., 2008; Cuadra et al., 2008) and by supernovae (e.g. Zubovas et al., 2013). Massive stars predominantly form in binaries (e.g. Kobulnicky and Fryer, 2007), and are short-lived. As they go supernova, the companion star will sometimes become unbound and move onto a new orbit. This process could help to repopulate the loss cone, although the effect would strongly depend on the rate of supernovae in the stellar ring. Zubovas et al. (2013) estimate supernovae rates of approximately 10^4 per $10^6\,M_\odot$ of stars formed on time-scales of 100 Myr – this would be even lower for the rings formed in the simulations here, as the initial cloud mass is only $10^5\,M_\odot$. It seems unlikely that this would contribute significantly, but further work is again required.

This leads to the interesting conclusion that the net effect of a single accretion event comprising an infalling gas cloud on a central SMBH binary may be stronger in either the prograde and the retrograde case, depending on the details of how the physical system evolves. If it is true that the gas disc is more capable of driving the binary evolution, then the retrograde cloud has the dominant effect; if stellar scattering from a ring of stars is dominant, then the prograde case must have the stronger effect.

While it is perfectly possible to tune physical parameters of a simulation in favour of either outcome, it is not trivial to see which is naturally favoured in the universe. Uncertainties in the cooling rate for a real gas disc also hinder a full understanding of which process dominates. Further work to explore the parameter space is therefore required to break the degeneracies.

4.5 Conclusions

In this chapter I have described SPH simulations of a low-mass gas cloud infalling on to an initially circular SMBH binary, varying both the binary mass ratio and the initial direction of the cloud with respect to the binary orbit. In all cases, a slightly eccentric disc forms ($e_d \sim 0.1$), but does not align with the plane of the binary on the time-scales of the simulations ($\sim 400 - 500$ binary orbits). The retrograde simulations result in higher binary eccentricities, with $e \sim 0.01$ in the case of a retrograde initial cloud orbit and a mass ratio of 1/10. There is significant scope for future work most importantly involving a more vigorous exploration of the cooling parameters, and following the future evolution of the stellar rings formed at the end of the simulations. Another interesting aspect would be to vary the parameters of the initial cloud, both its mass and initial orbit, to see how strongly these affect the results.

The major result of this work is that for a cooling law of the form $t_{cool} \propto \rho^{-1}$, the prograde discs fragment and form stars both sooner and across a narrower radial range than retrograde discs. This has strong implications for how the binary evolution progresses. If we assume that this evolution is driven primarily by the gas disc, then a retrograde disc will be able to drive stronger evolution than a prograde one with the same parameters. They live longer and are also able to extend closer to the binary allowing a more direct interaction (see Figures 4.3 and 4.5).

While it seems to be disfavoured on time-scale arguments (e.g. Amaro-Seoane et al., 2013), it may be that the converse is true and evolution of the binary may be dominated by interactions with the stars that form at the end of the simulations presented here. If

this is the case, then the prograde discs will result in stronger binary evolution, as they form stars sooner (and the stars which form may be more massive, as the surface density is higher). It remains to be seen which is true in real systems.

5
Conclusions

CHAPTERS 2−4 present the results of SPH simulations of disc interactions and their influence on the eccentricity of planetary orbits around single stars (Chapter 2) and in circumbinary cases (Chapter 3), and also how they can drive evolution of SMBH binaries (Chapter 4). In this chapter I summarise the results of each of these chapters and of the thesis as a body of work, and I discuss possible directions for future research.

5.1   Chapter summaries

5.1.1   Chapter 1: Introduction

Chapter 1 introduced many of the key concepts used throughout the thesis. I began by summarising the contexts relevant to the science in the later chapters, and what observational constraints we have on the accretion discs in those situations. I then derived the classical 1D accretion disc equations, and discussed current theoretical understanding and constraints on the viscosity and surface density, which are key parameters considered in Chapters 2 and 3.

I reviewed the classical theory of how satellites embedded in a disc interact with it, with particular attention to gap-opening by giant planets, before describing the main numerical method used in the simulations in the thesis. I gave an overview of how SPH works and the ways in which I have modified my version of GADGET-2 that differ from the standard methods of the field, particularly the implementation of a Navier-Stokes viscosity. Finally, I discussed the 1D viscous code used in Chapter 2 to calibrate this viscosity, which I preferred to solving the 1D viscous diffusion equation directly with modified Bessel functions. I showed that the agreement between this exact solution and the 1D code is almost exact, so its use is very well justified.

5.1.2   Chapter 2: Planetary eccentricity growth

In Chapter 2 I introduced the basic methods by which most exoplanets have been discovered to date, radial velocity measurements and transit surveys, and identified the observed trend of eccentric orbits for giant exoplanets. I then used a comparison between rings of spreading gas modelled using SPH and 1D viscous diffusion described in Chapter 1 to calibrate and characterize the behaviour of the Navier-Stokes viscosity implemented in the SPH code and also described in Chapter 1.

I then proceeded to describe simulations using this code of a giant planet embedded in a circumstellar disc, in order to test how the interaction with the disc affected the planet's orbit. The main parameters in the models which I varied between simulations were the planet mass and the disc surface density profile, in contrast to previous simulation work where in general the latter was kept constant between simulations (e.g. PNM01; D'Angelo et al., 2006). By plotting the contribution to the change in eccentricity as a function of radius in the disc, I have been able to show that there is a threshold surface density above which eccentricity can be excited.

Combining my results with those of previously published simulations, I showed that above this threshold, the question of whether a planets eccentricity will grow is in fact dependant upon the time-scale on which it accretes mass from the disc. If the mass growth time-scale is shorter than the eccentricity growth time-scale, then the planet will move out of the region of parameter space where eccentricity growth is permitted before becoming significantly eccentric. I conclude therefore that disc interactions are unlikely to be responsible for a large number of the eccentric giant planets observed in the universe, but that the process is not necessarily ruled out in individual cases.

5.1.3 Chapter 3: Kepler-16b and eccentricity damping

In Chapter 3 I moved on to consider the case of a circumbinary planet embedded in its parent disc. The system I chose to model was Kepler-16, the first planet discovered orbiting a main-sequence binary by *Kepler* (Doyle et al., 2011). In particular, Kepler-16b has a very low measured eccentricity, and Leung and Lee (2013) showed that its inherent (free) eccentricity is also very low, independent of the osculation pattern driven by its proximity to the binary.

I presented results of SPH simulations of this system when it was embedded in a circumbinary disc, based on the notion that the planet must have migrated through such a disc to reach its current orbit, given the dynamical coldness of the system and its other observed properties. The results of these simulations indicate that the planet would have had its eccentricity damped by the disc if it was unable to maintain a significant gap as it migrated. Using the disc torques measured in the simulations and the fact that they scale linearly with the disc surface density, I derived a lower limit on the surface density of the disc Kepler-16b formed in of $\Sigma_{min} \sim 10 \, \mathrm{g\,cm^{-2}}$.

The implication of this limit is that Kepler-16b must have formed in a relatively massive disc, and therefore must have formed early in the disc's lifetime before it lost significant mass to accretion, photoevaporation or other processes. As observations of discs

around very close binaries such as Kepler-16 are very difficult and provide few if any constraints, this limit has important implications for circumbinary planet formation theory.

5.1.4   Chapter 4: Binary supermassive black hole accretion

I then turned in Chapter 4 to simulations of a low-mass gas cloud infalling on to a SMBH binary. I first described the motivation behind these simulations, and how the last parsec problem requires some mechanism to drive these massive binaries from separations of ~ 1 down to ~ 0.01 pc in order to satisfy observations. I introduced the concept of chaotic accretion and described why theoretical considerations lead to the expectation that gas clouds must arrive in the central region of a galaxy with no correlation between the angular momentum vector of the cloud and the binary. I then briefly discussed the idea that mis- or counter-aligned gas circumbinary discs may drive faster binary evolution than aligned prograde discs. As the discs are expected to be gravitationally unstable, I discussed how the instability is understood to manifest in discs of given parameters.

Before presenting the simulations in the chapter I also described the modifications to the code that I made which differed from those used in Chapters 2 and 3. As the equation of state was no longer isothermal I described a density-dependant cooling prescription which set the cooling time for an SPH particle $t_{\text{cool}} \propto \rho^{-1}$ for density ρ above some threshold limit. This prescription was chosen to mimic optically thin cooling, and while this is not necessarily physical for the discs simulated it is a well behaved prescription and allows the results to be interpreted with confidence.

The simulations in this chapter consist of a turbulent gas cloud on an initially eccentric orbit around a binary, offset from the plane of the binary by $15°$. For two different binary mass ratios, I simulated one cloud in an orbit initially prograde with respect to the binary orbit, and one retrograde leading to four runs in total. The mass of each cloud was a hundredth of the total binary mass, and so only very small binary eccentricities were excited (the binary orbits were initially circular).

The major result of the simulations is that (for the cooling law imposed) the prograde disc fragments and forms stars at an early time than the retrograde disc, and does so across in a narrower radial region. As I did not simulate this process, but stopped the simulations when dense bound clumps formed in unstable regions of the disc, it is something of an open question how the system would subsequently evolve. I discuss the possibilities, and conclude that if the disc phase drives stronger binary evolution than the stars are able to do after they form (as seems likely), then retrograde circumbinary discs are indeed able to dominate the evolution of SMBH binaries over prograde discs. Applying these results to

the last parsec problem indicates that misaligned and retrograde discs are indeed capable of overcoming the problem of coalescing SMBH binaries, by driving the eccentricity of the binary to high values and allowing gravitational wave emission to dominate the binary evolution at earlier times, as suggested by Nixon et al. (2011a).

5.2   Discussion & avenues for further research

In this thesis I have simulated accretion discs interacting with bodies across different astrophysical contexts using SPH. One interesting outcome from the work presented in Chapter 2 is that the effect of a disc upon the dynamics of a planet are incredibly sensitive to the parameters of the disc. In that chapter I performed a modest exploration of the parameters space where the disc surface density profile is concerned, but this is by no means exhaustive. Other parameters I have neglected to consider include how accretion into the star and planet is handled. In this thesis I use only open sink boundaries, where any SPH particle found within the radius is swallowed. This is computationally cheap and I have been hesitant to add further complexity to what is already a somewhat inefficient code. Other possibilites are however possible, including the original sink formulation of Bate et al. (1995). In particular, Ayliffe and Bate (2010, 2011) have found that modelling planets as solid bodies with surfaces rather than sinks affects how their migration proceeds. While it is less clear that a solid body is a good model for a giant planet, it may be a better approximation than an open sink, and how this approach would affect the dynamics of the planetary orbit is an interesting question.

The treatment of the disc surface density is less important in the circumbinary case, as the binary torques tend to drive the disc towards an approximate $\Sigma(R) \propto R^{-3/2}$ solution as the disc relaxes, although the binary mass ratio and eccentricity certainly affect how this progresses. In both the single-star and binary cases though the treatment of the gas as locally isothermal is certainly unphysical. Again this was done for reasons of simplicity and efficiency, and there has been a significant amount of work done investigating how using fully radiative discs affects planet and planetesimal dynamics (e.g. Ayliffe and Bate, 2010; Bitsch and Kley, 2010, 2011a,b; Marzari et al., 2013), the details of what treatment is exactly correct have yet to be fully calculated.

Initially eccentric discs or inclined planets, and how the change in total angular momentum in the system alters the evolution of the system, would be another interesting parameter to explore. This is unlikely to make a large difference in binary systems, which are dominated by the angular momentum of the binary, but for single star systems it is pos-

sible that a Kozai-like mechanism will act and how this influences the other parameters is only just beginning to be considered in the literature (e.g. Xiang-Gruess and Papaloizou, 2013; Bitsch et al., 2013).

The science presented in Chapter 4 also has many viable options for further work. Once again, the treatment of the thermodynamics is somewhat unphysical, and it is important to quantify if and how the results change under more physical assumptions, or even different normalisations of the same cooling law. While current implementations of radiative transfer in SPH are not necessarily well calibrated for general problems (e.g. Wilkins and Clarke, 2012), it would be worthwhile comparing how these schemes change the evolution of the cloud simulations.

It would also be instructive to perform the same or similar simulations with an initially eccentric binary. Nixon et al. (2011a) suggest, and the results of the work in Chapter 4 agree, that retrograde accretion events should only increase the eccentricity of the binary (provided that the disc does not have sufficient angular momentum to cause it to co-align with the binary). Recently, Roedig and Sesana (2013) suggested that retrograde discs are only stable for $e_b = 0$, becoming co-aligned and causing the eccentricity to saturate at $e_b \simeq 0.6$ otherwise. This seems counter to previous results (e.g. Nixon et al., 2011a; Nixon, 2012), and further simulations would be a convenient clarification of the issue.

The recent discovery by Nixon et al. (2012, 2013) of 'tearing' of highly misaligned discs allowing rapid accretion of large amounts of gas directly onto a spinning SMBH or a binary SMBH provides another fascinating avenue for further work. Their simulations are all isothermal, and the possibility that cooling such as that implemented here could allow stars to form in a tearing disc could result in high rates of tidal disruption events, among other phenomena.

5.3   Final conclusions

I leave the reader with the following concise conclusions from the work presented in this thesis. Giant planets will not grow eccentricity via disc-planet interactions unless the disc is very massive and they accrete mass from the disc very slowly. Circumbinary planets (such as Kepler-16b) have their eccentricity damped by their parent disc, which must have been relatively massive when they formed. Finally, a prograde disc falling onto a SMBH binary will fragment and form stars sooner and in a narrower ring than will an identical retrograde disc, allowing many small randomly-aligned accretion events to drive the binary eccentricity to high values and cause coalescence.

A

SPH derivations

THE equations of motion for SPH as used in this thesis are given in Chapter 1 by equations 1.61, 1.67 and 1.70. The SPH implementation of a physical Navier-Stokes shear viscosity used in the simulations described in Chapters 2 and 3 is given in equation 1.79. In this appendix I derive the former from the basic SPH estimator, the Lagrangian, and by considering the equation of state (Section A.1), and the latter by combining the discretised Navier-Stokes equation with an improved SPH divergence estimator (Section A.2).

A.1 SPH equations of motion

A.1.1 Continuity equation

The continuity equation in SPH is arrived at simply and immediately from the density estimator (Equation 1.61), as it naturally enforces conservation of mass.

A.1.2 Momentum equation

To derive the SPH momentum equation, we start with the Lagrangian

$$\mathcal{L}(\boldsymbol{r}, \boldsymbol{u}) = \int_V \rho \left(\frac{\boldsymbol{u}^2}{2} - \epsilon \right) d\boldsymbol{r} \tag{A.1}$$

with ϵ the specific internal energy defined by the equation of state, which will be discussed below. A simple discretisation of this integral gives the SPH lagrangian,

$$\mathcal{L}_{\text{SPH}}(\boldsymbol{r}, \boldsymbol{u}) = \sum_a^N m_a \left(\frac{\boldsymbol{u}_a^2}{2} - \epsilon_a \right) \tag{A.2}$$

where the sum is taken over all particles.

This form of the Langragian allows us to use the Euler-Lagrange equation

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{u}_b} \right) - \frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{r}_b} = 0 \tag{A.3}$$

to derive the momentum equation for SPH. Considering this term-by-term, the partial

differential with respect to the velocity of the bth particle is

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{u}_b} &= \frac{\partial}{\partial \boldsymbol{u}_b} \sum_a^N m_b \left(\frac{\boldsymbol{u}_a^2}{2} - \epsilon_a \right) \\
&= m_b \boldsymbol{u}_b,
\end{aligned}
\tag{A.4}
$$

the second line being true because the differential is zero unless $a = b$ as the velocities of the particles are independent of each other.

The next term is the spatial differential with respect to the position of particle b, given by

$$
\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{r}_b} = \frac{\partial}{\partial \boldsymbol{r}_b} \sum_i^N m_a \left(\frac{\boldsymbol{u}_a^2}{2} - \epsilon_a \right).
\tag{A.5}
$$

In order to proceed we must know the form of the internal energy ϵ. Typically one uses the ideal gas equation of state, which gives the pressure P of particle a as

$$
P_a = A_a \rho_a^\gamma,
\tag{A.6}
$$

where A_a is some entropic function given by $A_a = (\gamma - 1)\epsilon_a / \rho_a^{(\gamma-1)}$ and γ is the adiabatic index of the gas, taken to be $\gamma = 5/3$ throughout. ϵ_a is then simply

$$
\epsilon_a = \frac{P_a}{(\gamma - 1)\rho_a} = \frac{A_a \rho_a^{(\gamma-1)}}{\gamma - 1}.
\tag{A.7}
$$

Equation A.5 now becomes

$$
\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{r}_b} = -\sum_a^N m_a \frac{\partial \epsilon_a}{\partial \rho_a} \frac{\partial \rho_a}{\partial \boldsymbol{r}_b}.
\tag{A.8}
$$

We arrive at this by assuming that entropy (and therefore the entropic function A_a) is constant, as there should be no inherent dissipation – and therefore the only factor in ϵ_a that is a function of position is the density ρ_a.

We can evaluate $\partial \epsilon / \partial \rho$ by considering the first law of thermodynamics, which states

$$
dU = TdS - PdV
\tag{A.9}
$$

where dU is the internal energy per unit volume and TdS is heat per unit volume added to the system (S being entropy and T being temperature), which disappears in the absence of dissipation. Using the SPH estimate for the volume given by $V = m/\rho$, dV becomes

$-(m/\rho^2)d\rho$. By changing from dU to $d\epsilon$ (i.e. going from internal energy per unit volume to per unit mass), we obtain

$$d\epsilon = \frac{P}{\rho^2}d\rho \tag{A.10}$$

and so

$$\frac{\partial \epsilon_a}{\partial \rho_a} = \frac{P_a}{\rho_a^2}. \tag{A.11}$$

The evaluation of $\partial \rho_a/\partial r_a$ is more tricky as we must account for the fact that smoothing lengths in SPH are variable. The process is simpler if we first proceed as if they are not (i.e. keeping h_a constant for all a). This gives

$$\frac{\partial \rho_a}{\partial r_b} = \sum_c^{N_{\mathrm{neigh}}} m_c \frac{\partial W_{ac}(h_a)}{\partial r_b}$$

$$= \sum_c^{N_{\mathrm{neigh}}} m_c \nabla_b W_{ac}(h_a)[\delta_{ab} - \delta_{cb}] \tag{A.12}$$

where $W_{ac}(h_a)$ represents the smoothing kernel $W(|r_a - r_c|, h_a)$ and δ_{ab} is the Dirac delta function at $a = b$. If we now consider smoothing lengths to be variable, Equation A.12 now becomes

$$\frac{\partial \rho_a}{\partial r_b} = \sum_c^{N_{\mathrm{neigh}}} m_c \left(\nabla_b W_{ac}(h_a)[\delta_{ab} - \delta_{cb}] + \frac{\partial W_{ac}}{\partial h_a}\frac{dh_a}{d\rho_a}\frac{\partial \rho_a}{\partial r_b} \right). \tag{A.13}$$

Collecting like terms and after some rearranging, this becomes

$$\frac{\partial \rho_a}{\partial r_b} = \frac{1}{\Lambda_a} \sum_c^{N_{\mathrm{neigh}}} m_c \nabla_b W_{ac}(h_a)[\delta_{ab} - \delta_{cb}] \tag{A.14}$$

where

$$\Lambda_a = 1 - \frac{dh_a}{d\rho_a} \sum_a^{N_{\mathrm{neigh}}} m_a \frac{\partial W_{ab}}{\partial h_a}$$

$$= 1 + \frac{h_a}{3\rho_a} \sum_a^{N_{\mathrm{neigh}}} m_a \frac{\partial W_{ab}}{\partial h_a}, \tag{A.15}$$

the second line coming from differentiation of equation 1.63, which gives the relationship between density and smoothing length directly.

Combining Equations A.8, A.11 and A.14, the second term in the Euler-Lagrange

equation becomes

$$\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{r}_b} = -\sum_a^{N_{\text{neigh}}} m_a \frac{P_a}{\Lambda_a \rho_a^2} \sum_c^{N_{\text{neigh}}} m_c \nabla_b W_{ab}(h_a)[\delta_{ab} - \delta_{cb}]$$

$$= -m_b \frac{P_b}{\Lambda_b \rho_b^2} \sum_c^{N_{\text{neigh}}} m_c \nabla_b W_{bc}(h_b) + \sum_a^{N_{\text{neigh}}} m_a \frac{P_a}{\Lambda_a \rho_a^2} m_b \nabla_b W_{ab}(h_a). \qquad \text{(A.16)}$$

Using the fact that the gradient of the Kernel is antisymmetric in r (i.e. $\nabla_b W_{ab}(h_a) = -\nabla_b W_{ba}(h_a)$) and by changing the summation index from c to a in the first term, this can be expressed as

$$\frac{\partial \mathcal{L}_{\text{SPH}}}{\partial \boldsymbol{r}_b} = -m_b \sum_a^{N_{\text{neigh}}} m_a \left[\frac{P_b}{\Lambda_b \rho_b^2} \nabla_b W_{ba}(h_b) + \frac{P_a}{\Lambda_a \rho_a^2} \nabla_b W_{ba}(h_a) \right]. \qquad \text{(A.17)}$$

We can now substitute Equations A.4 and A.17 back into the Euler-Lagrange equation (Equation A.3) to find the full momentum equation for SPH:

$$\frac{d\boldsymbol{u}_a}{dt} = -\sum_b^{N_{\text{neigh}}} m_b \left[\frac{P_a}{\Lambda_a \rho_a^2} \nabla_a W_{ab}(h_a) + \frac{P_b}{\Lambda_b \rho_b^2} \nabla_a W_{ab}(h_b) \right]. \qquad \text{(A.18)}$$

A.1.3   Energy equation

In the formulation of SPH used in this thesis, it is the entropic function A_i in Equation A.6 that is directly evolved, rather than the internal energy ϵ_a as in some other formulations. This has the advantage of directly controlling the sources of entropy, but in the limit of a smooth and well-resolved flow the approaches are identically conservative. In the absence of dissipation

$$\frac{dA_a}{dt} = 0. \qquad \text{(A.19)}$$

In the simulations described in Chapters 2 and 3, this is explicitly enforced by the isothermal temperature structure applied. However, in Chapter 4 dissipation is required due both to cooling and shocks. In these cases this becomes

$$\frac{dA_a}{dt} = \frac{\gamma - 1}{\rho^{\gamma-1}} \left(\frac{d\epsilon_a}{dt} \right) \qquad \text{(A.20)}$$

where the contributions to the term $d\epsilon_a/dt$ are described explicitly in Section 1.4.1.1 and Chapter 4.

A.2 Navier-Stokes viscosity

As described in Section 1.4.1.2, I use a non-standard modification to the SPH momentum equation (Equation 1.67) to explicitly include a Navier-Stokes viscosity term in the simulations described in Chapters 2 and 3. This is done to accurately model an Shakura and Sunyaev (1973) α-viscosity in the disc. The parameterisation is taken from Lodato and Price (2010), using the 'two first-derivatives' method described therein.

Before deriving the SPH implementation of the Navier-Stokes equation, it is worth noting that the standard way of calculating gradients and divergence in SPH can be improved. This can be shown by considering a constant function $f(\boldsymbol{r}) = 1$, which obviously has a gradient of $\nabla f(\boldsymbol{r}) = 0$. Using standard SPH estimates, these become

$$1 \approx \sum_a^{N_{\text{neigh}}} \frac{m_a}{\rho_a} W(\boldsymbol{r} - \boldsymbol{r}_a, h) \tag{A.21}$$

and

$$0 \approx \sum_a^{N_{\text{neigh}}} \frac{m_a}{\rho_a} \nabla W(\boldsymbol{r} - \boldsymbol{r}_a, h). \tag{A.22}$$

While neither of these is necessarily exactly true, in most cases the effect is small and is therefore acceptable. It is however possible to construct estimators that are exactly correct for constant functions.

For any real number n, the gradient of a function $f\rho^n$ can be found from

$$\nabla(f\rho^n) = nf\rho^{n-1}\nabla\rho + \rho\nabla f. \tag{A.23}$$

Rearranging, this gives

$$\nabla f = \frac{1}{\rho^n} \left[\nabla(f\rho^n) - nf\rho^{n-1}\nabla\rho \right] \tag{A.24}$$

which we can combine with the usual discretisation and SPH estimates for $\nabla(f\rho^n)$ and $\nabla\rho$ to find

$$\nabla f(\boldsymbol{r_a}) = \frac{1}{\rho_a^n} \sum_b^{N_{\text{neigh}}} \left[f(\boldsymbol{r_b})\rho_b^{n-1} - nf(\boldsymbol{r_a})\rho_a^{n-1} \right] \nabla W_{ab}. \tag{A.25}$$

Evaluating this at $n = 1$ gives an SPH gradient estimate that is exact for constant functions,

$$\nabla f(\boldsymbol{r}_a) = \frac{1}{\rho_a} \sum_b^{N_{\text{neigh}}} m_b \left[f(\boldsymbol{r}_b) - f(\boldsymbol{r}_a) \right] \nabla_a W_{ab}, \tag{A.26}$$

while $n = -1$ gives the expression

$$\nabla f(\boldsymbol{r}_a) = \rho_a \sum_b^{N_{\text{neigh}}} m_b \left[\frac{f(\boldsymbol{r}_b)}{\rho_b^2} + \frac{f(\boldsymbol{r}_a)}{\rho_a^2} \right] \nabla_a W_{ab}. \tag{A.27}$$

Taking variable smoothing lengths into consideration, this becomes

$$\nabla f(\boldsymbol{r}_a) = \rho_a \sum_b^{N_{\text{neigh}}} m_b \left[\frac{f(\boldsymbol{r}_b)}{\Lambda_b \rho_b^2} \nabla_a W_{ab}(h_b) + \frac{f(\boldsymbol{r}_a)}{\Lambda_a \rho_a^2} \nabla_a W_{ab}(h_a) \right]. \tag{A.28}$$

Interestingly, this provides an alternative derivation of Equation 1.67, the SPH momentum equation.

The discretised Navier-Stokes equation is given by (e.g. Springel, 2010)

$$\frac{du_a^i}{dt} = -\frac{\nabla_a P_a}{\rho_a} + \frac{\nabla_a^j S_a^{ij}}{\rho_a} \tag{A.29}$$

where vectors in summation notation i, j, k have been introduced (so that $\nabla_a^j = \partial/\partial r_a^j$). S is the stress tensor, which in the absence of any bulk viscosity (i.e. only considering a shear viscosity) is given by

$$S_a^{ij} = \eta_a \left[\frac{\partial u_a^i}{\partial r_a^j} + \frac{\partial u_a^j}{\partial r_a^i} \right] - \frac{2}{3} \delta^{ij} \eta_a \left(\frac{\partial u_a^k}{\partial r_a^k} \right) \tag{A.30}$$

where η_a is the shear viscosity at particle a. To evaluate S_a^{ij}, I use an expanded version of Equation A.26 where variable smoothing lengths are again accounted for,

$$\frac{\partial f^i(r_a^i)}{\partial r_a^j} = \frac{1}{\rho_a \Lambda_a} \sum_b^{N_{\text{neigh}}} \left[f^i(r_b^i) - f^i(r_a^i) \right] \frac{\partial W_{ab}(h_a)}{\partial r_a^j}. \tag{A.31}$$

This is sometimes referred to as the standard variable-smoothing-length gradient operator (e.g. Lodato and Price, 2010), and when applied to the velocity u in Equation A.30 it becomes

$$\frac{\partial u_a^i}{\partial r_a^j} = \frac{1}{\rho_a \Lambda_a} \sum_b^{N_{\text{neigh}}} \left[u_b^i - u_a^i \right] \frac{\partial W_{ab}(h_a)}{\partial r_a^j}. \tag{A.32}$$

Equation A.28 allows the viscous component of the Navier-Stokes equation to be

written as

$$\frac{du_a^i}{dt}\bigg|_{\text{visc}} = -\sum_b^{N_{\text{neigh}}} m_b \left[\frac{\mathcal{S}_a^{ij}}{\Lambda_a \rho_a^2} \nabla_a^j W_{ab}(h_a) + \frac{\mathcal{S}_b^{ij}}{\Lambda_b \rho_b^2} \nabla_a^j W_{ab}(h_b) \right] \quad \text{(A.33)}$$

and now the complete momentum equation, including viscous terms from the Navier-Stokes viscosity and the artificial viscosity, is

$$\frac{du_a^i}{dt} = -\sum_b^{N_{\text{neigh}}} m_b \left[\frac{P_a + \mathcal{S}_a^{ij}}{\Lambda_a \rho_a^2} \nabla_a^j W_{ab}(h_a) + \frac{P_b + \mathcal{S}_b^{ij}}{\Lambda_b \rho_b^2} \nabla_a^j W_{ab}(h_b) + \Pi_{ab} \, \nabla_a^j \overline{W}_{ab} \right]. \quad \text{(A.34)}$$

References

Adams, F. C., Lada, C. J., and Shu, F. H.: 1987, *ApJ* **312**, 788

Alexander, R.: 2012, *ApJ* **757**, L29

Alexander, R. D., Armitage, P. J., Cuadra, J., and Begelman, M. C.: 2008, *ApJ* **674**, 927

Alexander, R. D., Begelman, M. C., and Armitage, P. J.: 2007, *ApJ* **654**, 907

Alexander, R. D., Clarke, C. J., and Pringle, J. E.: 2006a, *MNRAS* **369**, 216

Alexander, R. D., Clarke, C. J., and Pringle, J. E.: 2006b, *MNRAS* **369**, 229

Alexander, R. D., Smedley, S. L., Nayakshin, S., and King, A. R.: 2012, *MNRAS* **419**, 1970

Amaro-Seoane, P., Brem, P., and Cuadra, J.: 2013, *ApJ* **764**, 14

André, P., Ward-Thompson, D., and Barsony, M.: 1993, *ApJ* **406**, 122

Andrews, S. M., Rosenfeld, K. A., Kraus, A. L., and Wilner, D. J.: 2013, *ApJ* **771**, 129

Andrews, S. M. and Williams, J. P.: 2005, *ApJ* **631**, 1134

Andrews, S. M. and Williams, J. P.: 2007, *ApJ* **659**, 705

Andrews, S. M., Wilner, D. J., Espaillat, C., et al.: 2011, *ApJ* **732**, 42

Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., and Dullemond, C. P.: 2009, *ApJ* **700**, 1502

Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., and Dullemond, C. P.: 2010, *ApJ* **723**, 1241

Antonucci, R.: 1993, *ARA&A* **31**, 473

Armitage, P. J.: 2007, *ArXiv:0701485*

Armitage, P. J.: 2010, *Astrophysics of Planet Formation*, Cambridge Univ. Press, Cambridge

Armitage, P. J.: 2011, *ARA&A* **49**, 195

Armitage, P. J. and Natarajan, P.: 2002, *ApJ* **567**, L9

Armitage, P. J. and Natarajan, P.: 2005, *ApJ* **634**, 921

Artymowicz, P., Clarke, C. J., Lubow, S. H., and Pringle, J. E.: 1991, *ApJ* **370**, L35

Ayliffe, B. A. and Bate, M. R.: 2010, *MNRAS* **408**, 876

Ayliffe, B. A. and Bate, M. R.: 2011, *MNRAS* **415**, 576

Balbus, S. A.: 2011, in P. J. V. Garcia (ed.), *Physical Processes in Circumstellar Disks around Young Stars*, University of Chicago Press, Chicago

Balbus, S. A. and Hawley, J. F.: 1991, *ApJ* **376**, 214

Balbus, S. A. and Hawley, J. F.: 1998, *Rev. Modern Phys.* **70**, 1

Balbus, S. A. and Papaloizou, J. C. B.: 1999, *ApJ* **521**, 650

Balsara, D. S.: 1995, *J. Comput. Phys.* **121**, 357

Barrado y Navascués, D. and Martín, E. L.: 2003, *AJ* **126**, 2997

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al.: 2013, *ApJS* **204**, 24

Bate, M. R., Bonnell, I. A., and Bromm, V.: 2003, *MNRAS* **339**, 577

Bate, M. R., Bonnell, I. A., and Price, N. M.: 1995, *MNRAS* **277**, 362

Begelman, M. C., Blandford, R. D., and Rees, M. J.: 1980, *Nature* **287**, 307

Bender, C. F., Mahadevan, S., Deshpande, R., et al.: 2012, *ApJ* **751**, L31

Bertout, C.: 1989, *ARA&A* **27**, 351

Beuermann, K., Hessman, F. V., Dreizler, S., et al.: 2010, *A&A* **521**, L60

Bird, J., Martini, P., and Kaiser, C.: 2008, *ApJ* **676**, 147

Bitsch, B., Crida, A., Libert, A.-S., and Lega, E.: 2013, *A&A* **555**, A124

Bitsch, B. and Kley, W.: 2010, *A & A* **523**, A30

Bitsch, B. and Kley, W.: 2011a, *A&A* **530**, A41

Bitsch, B. and Kley, W.: 2011b, *A&A* **536**, A77

Blandford, R. D. and Payne, D. G.: 1982, *MNRAS* **199**, 883

Bonnell, I. A. and Rice, W. K. M.: 2008, *Science* **321**, 1060

Butler, R. P., Marcy, G. W., Williams, E., Hauser, H., and Shirts, P.: 1997, *ApJ* **474**, L115

Calvet, N. and Gullbring, E.: 1998, *ApJ* **509**, 802

Chatterjee, S., Ford, E. B., Matsumura, S., and Rasio, F. A.: 2008, *ApJ* **686**, 580

Chiang, E. I., Fischer, D., and Thommes, E.: 2002, *ApJ* **564**, L105

Chiang, E. I. and Goldreich, P.: 1997, *ApJ* **490**, 368

Cossins, P., Lodato, G., and Clarke, C. J.: 2009, *MNRAS* **393**, 1157

Cossins, P., Lodato, G., and Testi, L.: 2010, *MNRAS* **407**, 181

Cossins, P. J.: 2010, *ArXiv:1007.1245v2*

Cuadra, J., Armitage, P. J., and Alexander, R. D.: 2008, *MNRAS* **388**, L64

Cuadra, J., Armitage, P. J., Alexander, R. D., and Begelman, M. C.: 2009, *MNRAS* **393**, 1423

Cuadra, J., Nayakshin, S., Springel, V., and Di Matteo, T.: 2006, *MNRAS* **366**, 358

Cumming, A., Butler, R. P., Marcy, G. W., et al.: 2008, *PASP* **120**, 531

D'Alessio, P., Calvet, N., Hartmann, L., Franco-Hernández, R., and Servín, H.: 2006, *ApJ* **638**, 314

D'Angelo, G., Henning, T., and Kley, W.: 2002, *A&A* **385**, 647

D'Angelo, G., Lubow, S. H., and Bate, M. R.: 2006, *ApJ* **652**, 1698

Dawson, R. I. and Murray-Clay, R. A.: 2013, *ApJ* **767**, L24

Dawson, R. I., Murray-Clay, R. A., and Johnson, J. A.: 2012, *ArXiv:1211.0554*

de Val-Borro, M., Edgar, R. G., Artymowicz, P., et al.: 2006, *MNRAS* **370**, 529

Desidera, S., Gratton, R. G., Endl, M., et al.: 2004, *A&A* **420**, L27

Doyle, L. R., Carter, J. A., Fabrycky, D. C., et al.: 2011, *Science* **333**, 1602

Dumusque, X., Pepe, F., Lovis, C., et al.: 2012, *Nature* **491**, 207

Durisen, R. H., Boss, A. P., Mayer, L., et al.: 2007, *Protostars and Planets V* pp 607–622

Eggenberger, A., Udry, S., and Mayor, M.: 2004, *A&A* **417**, 353

Fabbiano, G., Wang, J., Elvis, M., and Risaliti, G.: 2011, *Nature* **477**, 431

Feigelson, E. D. and Montmerle, T.: 1999, *ARA&A* **37**, 363

Ford, E. B., Havlickova, M., and Rasio, F. A.: 2001, *Icarus* **150**, 303

Ford, E. B. and Rasio, F. A.: 2008, *ApJ* **686**, 621

Forgan, D., Rice, K., Cossins, P., and Lodato, G.: 2011, *MNRAS* **410**, 994

Forgan, D., Rice, K., Stamatellos, D., and Whitworth, A.: 2009, *MNRAS* **394**, 882

Fragner, M. M., Nelson, R. P., and Kley, W.: 2011, *A&A* **528**, A40

Frank, J., King, A., and Raine, D. J.: 2002, *Accretion Power in Astrophysics: Third Edition*, Cambridge Univ. Press, Cambridge

Furlan, E., Hartmann, L., Calvet, N., et al.: 2006, *ApJS* **165**, 568

Gammie, C. F.: 1996, *ApJ* **457**, 355

Gammie, C. F.: 2001, *ApJ* **553**, 174

Genzel, R., Schödel, R., Ott, T., et al.: 2003, *ApJ* **594**, 812

Ghez, A. M., Salim, S., Weinberg, N. N., et al.: 2008, *ApJ* **689**, 1044

Gingold, R. A. and Monaghan, J. J.: 1977, *MNRAS* **181**, 375

Glassgold, A. E., Feigelson, E. D., and Montmerle, T.: 2000, *Protostars and Planets IV* pp 429–455

Goldreich, P. and Sari, R.: 2003, *ApJ* **585**, 1024

Goldreich, P. and Tremaine, S.: 1979, *ApJ* **233**, 857

Goldreich, P. and Tremaine, S.: 1980, *ApJ* **241**, 425

Goodman, A. A., Benson, P. J., Fuller, G. A., and Myers, P. C.: 1993, *ApJ* **406**, 528

Greene, J. E., Seth, A., den Brok, M., et al.: 2013, *ApJ* **771**, 121

Greene, T.: 2001, *American Scientist* **89**, 316

Gressel, O., Nelson, R. P., and Turner, N. J.: 2012, *MNRAS* **422**, 1140

Gualandris, A., Mapelli, M., and Perets, H. B.: 2012, *MNRAS* **427**, 1793

Gullbring, E., Calvet, N., Muzerolle, J., and Hartmann, L.: 2000, *ApJ* **544**, 927

Gullbring, E., Hartmann, L., Briceno, C., and Calvet, N.: 1998, *ApJ* **492**, 323

Hacar, A., Tafalla, M., Kauffmann, J., and Kovács, A.: 2013, *A&A* **554**, A55

Haisch, Jr., K. E., Lada, E. A., and Lada, C. J.: 2001, *ApJ* **553**, L153

Harris, R. J., Andrews, S. M., Wilner, D. J., and Kraus, A. L.: 2012, *ApJ* **751**, 115

Harsono, D., Alexander, R. D., and Levin, Y.: 2011, *MNRAS* **413**, 423

Hartmann, L.: 2009, *Accretion Processes in Star Formation: Second Edition*, Cambridge Univ. Press, Cambridge

Hartmann, L., Calvet, N., Gullbring, E., and D'Alessio, P.: 1998, *ApJ* **495**, 385

Hartmann, L., D'Alessio, P., Calvet, N., and Muzerolle, J.: 2006, *ApJ* **648**, 484

Hatzes, A. P.: 2013, *ApJ* **770**, 133

Hawley, J. F., Gammie, C. F., and Balbus, S. A.: 1995, *APJ* **440**, 742

Hayashi, C.: 1981, *Prog. Theor. Phys. Suppl.* **70**, 35

Hillenbrand, L. A.: 2009, in E. E. Mamajek, D. R. Soderblom, and R. F. G. Wyse (eds.), *The Ages of Stars*, Vol. 258 of *IAU symposium proceedings*, pp 81–94

Hirose, S., Blaes, O., and Krolik, J. H.: 2009, *ApJ* **704**, 781

Holman, M. J. and Wiegert, P. A.: 1999, *AJ* **117**, 621

Jurić, M. and Tremaine, S.: 2008, *ApJ* **686**, 603

Kalas, P., Graham, J. R., Chiang, E., et al.: 2008, *Science* **322**, 1345

Kane, S. R., Ciardi, D. R., Gelino, D. M., and von Braun, K.: 2012, *MNRAS* **425**, 757

Kenyon, S. J. and Hartmann, L.: 1987, *ApJ* **323**, 714

Kenyon, S. J. and Hartmann, L.: 1995, *ApJS* **101**, 117

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

King, A. R., Pringle, J. E., and Livio, M.: 2007, *MNRAS* **376**, 1740

Kinney, A. L., Schmitt, H. R., Clarke, C. J., et al.: 2000, *ApJ* **537**, 152

Klaassen, P. D., Testi, L., and Beuther, H.: 2012, *A&A* **538**, A140

Kley, W. and Dirksen, G.: 2006, *A&A* **447**, 369

Kley, W. and Nelson, R. P.: 2008, *A&A* **486**, 617

Kobulnicky, H. A. and Fryer, C. L.: 2007, *ApJ* **670**, 747

Königl, A. and Salmeron, R.: 2011, in P. J. V. Garcia (ed.), *Physical Processes in Circumstellar Disks around Young Stars*, pp 283–352

Kozai, Y.: 1962, *AJ* **67**, 591

Kraus, A. L., Ireland, M. J., Hillenbrand, L. A., and Martinache, F.: 2012, *ApJ* **745**, 19

Kraus, A. L., Ireland, M. J., Martinache, F., and Hillenbrand, L. A.: 2011, *ApJ* **731**, 8

Lada, C. J.: 1987, in M. Peimbert and J. Jugaku (eds.), *Star Forming Regions*, Vol. 115 of *IAU symposium proceedings*, pp 1–17

Lagrange, A.-M., Gratadour, D., Chauvin, G., et al.: 2009, *A&A* **493**, L21

Larson, R. B.: 1981, *MNRAS* **194**, 809

Lee, J. W., Kim, S.-L., Kim, C.-H., et al.: 2009, *AJ* **137**, 3181

Leung, G. C. K. and Lee, M. H.: 2013, *ApJ* **763**, 107

Levin, Y.: 2007, *MNRAS* **374**, 515

Levin, Y. and Beloborodov, A. M.: 2003, *ApJ* **590**, L33

Lidov, M. L.: 1962, *Planet. Space Sci.* **9**, 719

Lin, D. N. C. and Papaloizou, J.: 1979a, *MNRAS* **186**, 799

Lin, D. N. C. and Papaloizou, J.: 1979b, *MNRAS* **188**, 191

Lin, M.-K., Krumholz, M. R., and Kratter, K. M.: 2011, *MNRAS* **416**, 580

Livio, M., Ogilvie, G. I., and Pringle, J. E.: 1999, *ApJ* **512**, 100

Lodato, G.: 2007, *Riv. Nuovo Cimento* **30**, 293

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 Rice, W. K. M.: 2004, *MNRAS* **351**, 630

Lodato, G. and Rice, W. K. M.: 2005, *MNRAS* **358**, 1489

Lowrance, P. J., Kirkpatrick, J. D., and Beichman, C. A.: 2002, *ApJ* **572**, L79

Lu, J. R., Ghez, A. M., Hornstein, S. D., et al.: 2009, *ApJ* **690**, 1463

Lubow, S. H. and D'Angelo, G.: 2006, *ApJ* **641**, 526

Lubow, S. H., Seibert, M., and Artymowicz, P.: 1999, *ApJ* **526**, 1001

Lucas, W. E., Bonnell, I. A., Davies, M. B., and Rice, W. K. M.: 2013, *MNRAS* **433**, 353

Lucy, L. B.: 1977, *AJ* **82**, 1013

Lynden-Bell, D. and Pringle, J. E.: 1974, *MNRAS* **168**, 603

Mapelli, M., Hayfield, T., Mayer, L., and Wadsley, J.: 2012, *ApJ* **749**, 168

Marcy, G., Butler, R. P., Fischer, D., et al.: 2005, *Prog. Theor. Phys. Suppl.* **158**, 24

Marois, C., Macintosh, B., Barman, T., et al.: 2008, *Science* **322**, 1348

Martin, R. G., Armitage, P. J., and Alexander, R. D.: 2013, *ApJ* **773**, 74

Marzari, F., Baruteau, C., Scholl, H., and Thebault, P.: 2012, *A&A* **539**, A98

Marzari, F., Thébault, P., and Scholl, H.: 2008, *ApJ* **681**, 1599

Marzari, F., Thébault, P., Scholl, H., Picogna, G., and Baruteau, C.: 2013, *A&A* **553**, A71

Masset, F. S. and Ogilvie, G. I.: 2004, *ApJ* **615**, 1000

Mayer, L., Kazantzidis, S., Madau, P., et al.: 2007, *Science* **316**, 1874

Mayor, M. and Queloz, D.: 1995, *Nature* **378**, 355

Merritt, D.: 2013, *ArXiv:1307.3268*

Merritt, D. and Milosavljević, M.: 2005, *Living Reviews in Relativity* **8**, 8

Merritt, D. and Wang, J.: 2005, *ApJ* **621**, L101

Meru, F. and Bate, M. R.: 2011, *MNRAS* **411**, L1

Meschiari, S.: 2012, *ApJ* **752**, 71

Mihos, J. C. and Hernquist, L.: 1994, *ApJ* **431**, L9

Milosavljević, M. and Merritt, D.: 2001, *ApJ* **563**, 34

Moorhead, A. V. and Adams, F. C.: 2008, *Icarus* **193**, 475

Morris, J. P. and Monaghan, J. J.: 1997, *J. Comput. Phys.* **136**, 41

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al.: 2011, *Nature* **474**, 616

Müller, T. W. A. and Kley, W.: 2012, *A&A* **539**, A18

Murray, C. D. and Dermott, S. F.: 1999, *Solar System Dynamics*, Cambridge Univ. Press, Cambridge

Murray, J. R.: 1996, *MNRAS* **279**, 402

Naoz, S., Farr, W. M., and Rasio, F. A.: 2012, *ApJ* **754**, L36

Nelson, A. F.: 2006, *MNRAS* **373**, 1039

Nixon, C. and King, A.: 2013, *ApJ* **765**, L7

Nixon, C., King, A., and Price, D.: 2013, *ArXiv:1307:0010*

Nixon, C., King, A., Price, D., and Frank, J.: 2012, *ApJ* **757**, L24

Nixon, C. J.: 2012, *MNRAS* **423**, 2597

Nixon, C. J., Cossins, P. J., King, A. R., and Pringle, J. E.: 2011a, *MNRAS* **412**, 1591

Nixon, C. J., King, A. R., and Pringle, J. E.: 2011b, *MNRAS* **417**, L66

Ogilvie, G. I. and Lubow, S. H.: 2003, *ApJ* **587**, 398

Orosz, J. A., Welsh, W. F., Carter, J. A., et al.: 2012a, *ApJ* **758**, 87

Orosz, J. A., Welsh, W. F., Carter, J. A., et al.: 2012b, *Science* **337**, 1511

Owen, J. E., Ercolano, B., Clarke, C. J., and Alexander, R. D.: 2010, *MNRAS* **401**, 1415

Paardekooper, S.-J., Leinhardt, Z. M., Thébault, P., and Baruteau, C.: 2012, *ApJ* **754**, L16

Paardekooper, S.-J., Thébault, P., and Mellema, G.: 2008, *MNRAS* **386**, 973

Papaloizou, J. C. B., Nelson, R. P., and Masset, F.: 2001, *A&A* **366**, 263

Paumard, T., Genzel, R., Martins, F., et al.: 2006, *ApJ* **643**, 1011

Pelupessy, F. I. and Portegies Zwart, S.: 2013, *MNRAS* **429**, 895

Perets, H. B., Kupi, G., and Alexander, T.: 2008, in E. Vesperini, M. Giersz, and A. Sills (eds.), *Dynamical Evolution of Dense Stellar Systems*, Vol. 246 of *IAU symposium proceedings*, pp 275–276

Pierens, A. and Nelson, R. P.: 2008, *A&A* **483**, 633

Pierens, A. and Nelson, R. P.: 2013, *ArXiv:1307.0713*

Popova, E. A. and Shevchenko, I. I.: 2013, *ApJ* **769**, 152

Price, D. J.: 2004, *Ph.D. thesis*, Univ. Cambridge

Price, D. J.: 2007, *Proc. Astron. Soc. Aust.* **24**, 159

Price, D. J.: 2012, *J. Comput. Phys.* **231**, 759

Pringle, J. E.: 1981, *ARA&A* **19**, 137

Pringle, J. E.: 1989, *MNRAS* **239**, 361

Pringle, J. E.: 1991, *MNRAS* **248**, 754

Pursimo, T., Takalo, L. O., Sillanpää, A., et al.: 2000, *A&AS* **146**, 141

Rafikov, R. R.: 2013, *ApJ* **764**, L16

Rasio, F. A., Tout, C. A., Lubow, S. H., and Livio, M.: 1996, *ApJ* **470**, 1187

Rice, W. K. M., Armitage, P. J., Bate, M. R., and Bonnell, I. A.: 2003, *MNRAS* **339**, 1025

Rice, W. K. M., Forgan, D. H., and Armitage, P. J.: 2012, *MNRAS* **420**, 1640

Rice, W. K. M., Lodato, G., and Armitage, P. J.: 2005, *MNRAS* **364**, L56

Rodriguez, C., Taylor, G. B., Zavala, R. T., et al.: 2006, *ApJ* **646**, 49

Roedig, C. and Sesana, A.: 2013, *ArXiv:1307.6283*

Rosenfeld, K. A., Andrews, S. M., Wilner, D. J., and Stempels, H. C.: 2012, *ApJ* **759**, 119

Salmeron, R. and Wardle, M.: 2008, *MNRAS* **388**, 1223

Schneider, N., André, P., Könyves, V., et al.: 2013, *ApJ* **766**, L17

Schwamb, M. E., Orosz, J. A., Carter, J. A., et al.: 2013, *ApJ* **768**, 127

Searle, L. and Zinn, R.: 1978, *ApJ* **225**, 357

Semenov, D., Henning, T., Helling, C., Ilgner, M., and Sedlmayr, E.: 2003, *A&A* **410**, 611

Shakura, N. I. and Sunyaev, R. A.: 1973, *A&A* **24**, 337

Shu, F. H., Adams, F. C., and Lizano, S.: 1987, *ARA&A* **25**, 23

Sillanpaa, A., Haarala, S., Valtonen, M. J., Sundelius, B., and Byrd, G. G.: 1988, *ApJ* **325**, 628

Simon, J. B., Beckwith, K., and Armitage, P. J.: 2012, *MNRAS* **422**, 2685

Springel, V.: 2005, *MNRAS* **364**, 1105

Springel, V.: 2010, *ARA&A* **48**, 391

Stamatellos, D., Whitworth, A. P., Bisbas, T., and Goodwin, S.: 2007, *A&A* **475**, 37

Stone, J. M., Hawley, J. F., Gammie, C. F., and Balbus, S. A.: 1996, *ApJ* **463**, 656

Takeda, G. and Rasio, F. A.: 2005, *ApJ* **627**, 1001

Tanaka, H., Takeuchi, T., and Ward, W. R.: 2002, *ApJ* **565**, 1257

Thorsett, S. E., Arzoumanian, Z., and Taylor, J. H.: 1993, *ApJ* **412**, L33

Toomre, A.: 1964, *ApJ* **139**, 1217

Tsamis, Y. G., Rawlings, J. M. C., Yates, J. A., and Viti, S.: 2008, *MNRAS* **388**, 898

Udry, S. and Santos, N. C.: 2007, *ARA&A* **45**, 397

Urry, C. M. and Padovani, P.: 1995, *PASP* **107**, 803

van den Bosch, R. C. E., Gebhardt, K., Gültekin, K., et al.: 2012, *Nature* **491**, 729

Veras, D. and Armitage, P. J.: 2004, *MNRAS* **347**, 613

Walter, F. M., Brown, A., Mathieu, R. D., Myers, P. C., and Vrba, F. J.: 1988, *AJ* **96**, 297

Weidenschilling, S. J.: 1977, *Ap&SS* **51**, 153

Welsh, W. F., Orosz, J. A., Carter, J. A., et al.: 2012, *Nature* **481**, 475

Whitworth, A. P., Bhattal, A. S., Francis, N., and Watkins, S. J.: 1996, *MNRAS* **283**, 1061

Wilkins, D. R. and Clarke, C. J.: 2012, *MNRAS* **419**, 3368

Williams, J. P. and Cieza, L. A.: 2011, *ARA&A* **49**, 67

Willott, C. J., Omont, A., and Bergeron, J.: 2013, *ApJ* **770**, 13

Winn, J. N., Albrecht, S., Johnson, J. A., et al.: 2011, *ApJ* **741**, L1

Wolf, S. and D'Angelo, G.: 2005, *ApJ* **619**, 1114

Wright, J. T., Fakhouri, O., Marcy, G. W., et al.: 2011, *PASP* **123**, 412

Wright, J. T., Marcy, G. W., Howard, A. W., et al.: 2012, *ApJ* **753**, 160

Wu, Y. and Lithwick, Y.: 2011, *ApJ* **735**, 109

Xiang-Gruess, M. and Papaloizou, J. C. B.: 2013, *MNRAS* **431**, 1320

Zhu, Z., Hartmann, L., and Gammie, C.: 2010, *ApJ* **713**, 1143

Zubovas, K., Wynn, G. A., and Gualandris, A.: 2013, *ApJ* **771**, 118