Astronomy Unit
School of Physics and Astronomy
Queen Mary, University of London

Gap Formation in Protoplanetary Discs

Paul Derek Hallam

Supervised by S.-J. Paardekooper

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

Declaration

I, Paul Derek Hallam, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Details of collaboration and publications: Part of this work has been done in collaboration with my supervisor, Sijme-Jan Paardekooper. I have made a major contribution to the original research presented within this thesis. This thesis is based on the following papers, all of which have been published or submitted for publication:

- The role of gap edge instabilities in setting the depth of planet gaps in protoplanetary discs
  P. D. Hallam and S.-J. Paardekooper, *2017, MNRAS, 469, 3813*
  Arxiv:1705.07528 [astro-ph.EP]

- Investigating the possibility of reversing giant planet migration via gap edge illumination
  P. D. Hallam and S.-J. Paardekooper, *2018, MNRAS, 481, 1667*
  Arxiv:1808.09554 [astro-ph.EP]

- Constraining the masses of planets in protoplanetary discs from the presence or absence of vortices - Comparison with ALMA observations
  P. D. Hallam and S.-J. Paardekooper, *2019, MNRAS, preprint*

This work was supported by an STFC PhD studentship.

Signature: Paul Derek Hallam
Date: 22/07/2019

Abstract

In this thesis, we address a wide range of topics revolving around the interactions between a protoplanetary disc and a planet situated within. Initially, we present an overview of the background to protoplanetary disc physics. We then discuss our work, which is the body of this thesis and addresses three different problems that exist regarding planet-disc interactions.

The first of these problems is the significant difference between the depth of gaps formed by planets in one dimensional and two dimensional numerical simulations. We investigate this by applying one dimensional gap forming potentials axisymmetrically across a two dimensional disc containing no planet and observe the results. We find that the discrepancy between gap depths is reduced and the reason for this is the Rossby Wave Instability in two dimensions, which is not accounted for in one dimension.

The second problem we address is that Type II planetary migration is too fast to explain the population of gas giant planets at larger orbital radii. Hence, we investigate a method which can slow down Type II migration, the heating of the outer edge of a giant planet gap by incident radiation from the central star. We find this can reduce the net torque on the planet, potentially slowing or even reversing Type II migration.

The third problem we investigate is more observational in nature. Almost any observed gap can have a planetary explanation if the viscosity of the disc and mass of the planet are unknown. Hence, we present a way to break this degeneracy, using the presence or absence of vortices to help constrain these parameters. We find that often high mass planets can be ruled out in axisymmetric discs and that if discs are low viscosity, planets may not need to be particularly massive to form vortices.

Acknowledgements

This thesis would not have been possible without the help and support of many people over the last four years. Firstly, I would like to thank my supervisor, Sijme-Jan Paardekooper, not only for his help and guidance but also his patience and understanding. Also, for always having time to listen, which set my mind at rest on many occasions.

I would like to thank all of my fellow PhD students for the relaxed and pleasant atmosphere of our office and department, which has made the whole PhD experience a lot of fun! I would like to thank my old deskmates, Matthew Mutter and Gavin Coleman, for their advice in my early PhD years, both regarding my subject area and PhD life in general. I would also like to thank my current deskmates, Francesco Lovascio and Callum Boocock for always listening when I have something completely random to say and for all of our unusual conversations.

Also, I would like to thank John Strachan for joining me on our daily lunch time journeys and Clark Baker for his ever 'interesting' emails and presentations.

But honestly, I would really like to thank all of my colleagues, Pedro Carrilho, Shaille Imrith, Charalambos Pittordis, Domenico Trotta, John Ronayne, Jessie Durk, Louis Coates, Kit Gallagher, Rebeca Carrillo, Sanson Poon, Jack Skinner, Jorge Fuentes, Sandy Zeng, Eline DeWeerd, Ahdab Althukair, Ali Barlas, Jesse Coburn, Alice Giroul and Paula Soares for making this department exactly what it is!

Additionally, I would like to thank my long time friends Christopher Haynes, Murray Maynard, Tommaso Covella, Emrys Jenkins, Talha Ahmed, Rustam Majainah, Jessica Dendura, Nivedha Radhakrishnan, Darren Maskell and Issie Barrett for all their support and distractions outside of physics.

I would especially like to thank my good friend Jean Duffy, for being there when I needed to talk.

Finally, I would like to give special thanks to my parents and my brother. Their constant support and belief in me really made the completion of this PhD possible.

Contents

1 Introduction

1.1 Exoplanets

The idea of planets existing around stars other than our own has always been an area of great interest to researchers, even if for a long period of time the Solar System was our only point of reference (Lissauer, 1993). It was only recently that the first confirmed detection was made of a planet orbiting a main sequence star other than our sun (Mayor & Queloz, 1995). Since then the number of detected exoplanets has soared, with many different detection methods (Fischer et al., 2014) and observational instruments designed to push the boundaries of exoplanet detection (e.g. Mayor et al., 2003; Pollacco et al., 2006; Baglin et al., 2009; Ricker et al., 2015). Perhaps the most well known of these missions is the Kepler space telescope (Borucki et al., 2010), which aimed to detect small Earth-like planets that are desirable to understand due to their similarity to the Earth. The interest in the field is ever increasing, with new and exciting discoveries being made even very recently (Anglada-Escudé et al., 2016; Gillon et al., 2017; Ribas et al., 2018).

Exoplanet research has provided us with a great deal of insight into the structure and distribution of planets and planetary systems in our galaxy. Some of the discoveries made have come as a complete surprise to us, based on our prior limited frame of reference, our solar system. Since Mayor & Queloz (1995), we have learned that planetary systems are extremely common around other stars, with currently 3846 exoplanets detected across 2867 planetary systems, 636 of which are multiple planet systems [1]. Studies suggest that there is, on average, one planet orbiting at $0.5 - 10$AU around every star in our galaxy and that low mass planets are very common (Cassan et al., 2012).

1.1.1 Detection Methods

The two most prominent methods for detecting exoplanets are the radial velocity and transit methods, although exoplanets have also been detected using microlensing

[1] http://exoplanet.eu/catalog/, retrieved 25/09/18.

(Bond et al., 2004), astrometry (Pravdo et al., 2005; Dahn et al., 2008) and by direct imaging (Marois et al., 2010). Here we will give a brief overview of the radial velocity and transit method.

Radial velocity method

The radial velocity, or Doppler method, relies on the gravitational interaction between the planet and the star it is orbiting. Realistically, the star and planet will both orbit around a common barycentre, the centre of mass of the star-planet system. This is significantly closer to the star than the planet itself, however, it causes the star to move slightly over the course of one orbital period. This small change in the star's position causes a Doppler shift in the emitted light. If this Doppler shift is periodic, it can be inferred that a planetary companion is present. This method is biased towards the detection of larger planets with smaller semi-major axis, as these will exert a stronger gravitational force on the star and hence provide a larger Doppler shift (Fischer et al., 2014). The radial velocity method has provided the second largest number of exoplanet detections so far, including the very first detected exoplanet around a main sequence star (Mayor & Queloz, 1995).

Transit method

The transit method is perhaps the most well known of the exoplanet detection methods, maybe due to the success of the Kepler mission (Borucki et al., 2010). This method relies on the exoplanet being in roughly the same plane as the Earth and the star we are observing, such that as the exoplanet passes in between the observer and the star it blocks out a portion of the star's light. This provides a dip in the stellar light curve which, if periodic, can imply the presence of a planetary companion. Again this method is biased towards larger planets at a smaller semi-major axis, as these will block out more light and provide a deeper transit. This method has produced the most confirmed exoplanet detections to date.

1.1.2 Exoplanet Diversity

A surprising result from our exploration into the detection of exoplanets was the diversity of these systems, many of which do not resemble the layout of our solar system at all. For example, the compact system orbiting the star TRAPPIST-1 (Gillon et al., 2017) or short period giant planets such as 51 Pegasi b (Mayor & Queloz, 1995). In general, despite the broad variety in architecture of planetary

systems there are two main classifications of planet, terrestrial and giant.

Terrestrial planets

Terrestrial planets are planets that are similar to those found in our inner solar system, for example Venus, Mars and the Earth itself. Such planets are likely to have a rocky composition and masses $M_p < 10M_\oplus$ (Raymond et al., 2014). Terrestrial planets can be particularly interesting as these are planets that are the most similar to the Earth. Planets sub-Neptune in size are known to be common on orbits $R_p \ll 1$AU around Sun-like stars (Howard et al., 2010; Mayor et al., 2011; Howard et al., 2012). Additionally, approximately half of all stars in the galaxy are expected to have planets between the size of the Earth and the size of Neptune, orbiting with a semi-major axis smaller than Mercury's (Dressing & Charbonneau, 2015; Burke et al., 2015). Hence, it is clear that terrestrial planets are very common and it is likely that some of these will orbit in the habitable zone of their star, the region at which liquid water can exist on the planet's surface.

It is very desirable to detect Earth-like exoplanets orbiting in the habitable zone of their star, as it could be the first step to detecting life outside our own planet. The most commonly detected Earth-like exoplanets are called Super-Earths, as they are more massive than the Earth and so are often easier to detect. Unfortunately the detection of planets this size is still very difficult, as most methods of detecting exoplanets rely on the perturbation of light from the star by the planet, be this due to gravitational influence or the blocking of emitted light. Hence, small planets have a weaker effect and a smaller observational signature. Despite this we know that small planets are more common than massive planets (Mayor et al., 2011). Earth mass planets have been detected in habitable zones around M-dwarf stars (Anglada-Escudé et al., 2016; Bonfils et al., 2018) with an occurrence rate of one planet per four stars (Dressing & Charbonneau, 2015). For sun-like stars it is more difficult to detect terrestrial planets residing in their habitable zones, due to their longer periods and weaker transit signal. However, it is expected that there is a habitable zone Earth-like planet per ten stars (Burke et al., 2015). While these results may be due to extrapolation, they provide evidence that habitable zone planets similar to our own are not hugely uncommon throughout the galaxy.

Giant planets

Giant planets, on the other hand, are planets similar to those that can be found in our outer solar system, for example Jupiter and Saturn. These planets are usually

characterised by a rocky core surrounded by a gaseous envelope and as such are often known as 'gas giants'. Gas giant planets are of particular importance to the work we present in this thesis due to their interactions with the protoplanetary disc that formed them, which shall be discussed in future sections (e.g. Section 1.5).

Observationally, gas giant planets provided one of the most unexpected discoveries to come out of the study of exoplanets. This was the existence of a population of short-period giant planets, orbiting very close to their parent stars (Mayor et al., 2011; Wright et al., 2012). These 'hot Jupiters' have no analogue in our solar system and were not predicted by planet formation models, so discovering these was a vital step towards better understanding the dynamics and formation of planetary systems.

Hot Jupiters make up a large portion of the detected exoplanets, as observational bias favours their detection via the radial velocity and transit methods and these are the methods that have provided the majority of exoplanet detections. As such they were among the only exoplanets detectable using early ground based instruments, indeed the very first detected exoplanet around a main sequence star was a hot Jupiter (Mayor & Queloz, 1995). Despite the bias towards detecting hot Jupiters, there still exists a large population of 'cold Jupiters' which are a better analogue to the gas giant planets found in our solar system. In fact, we now know that hot Jupiters are less common than our familiar cold Jupiters (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016), but the existence of both still presents problems for planet formation models. Again, despite the observational bias towards them, giant planets are also less common than smaller, terrestrial planets (Mayor et al., 2011).

The bias towards hot Jupiters can be highlighted using Figure 1.1. This shows the mass (or minimum mass) of detected exoplanets against their semi-major axis. We can see there is a large number of detected massive planets at small semi-major axis, with the hot Jupiter population at $\ll$ 1AU clearly visible. We know that the radial velocity and transit methods have a higher probability of detecting short period giant planets. Therefore, correcting for this bias shows that there must be a greater number of giant planets at larger semi-major axis. Figure 1.2 shows the distribution of currently detected transiting exoplanets across the sky, clearly highlighting the increased detection rate of giant planets using this method.

1.1.3 Motivation

Overall, the motivation to study exoplanets arises from two simple questions, which prove to be very difficult to answer. Firstly, we want to understand the dynamics,

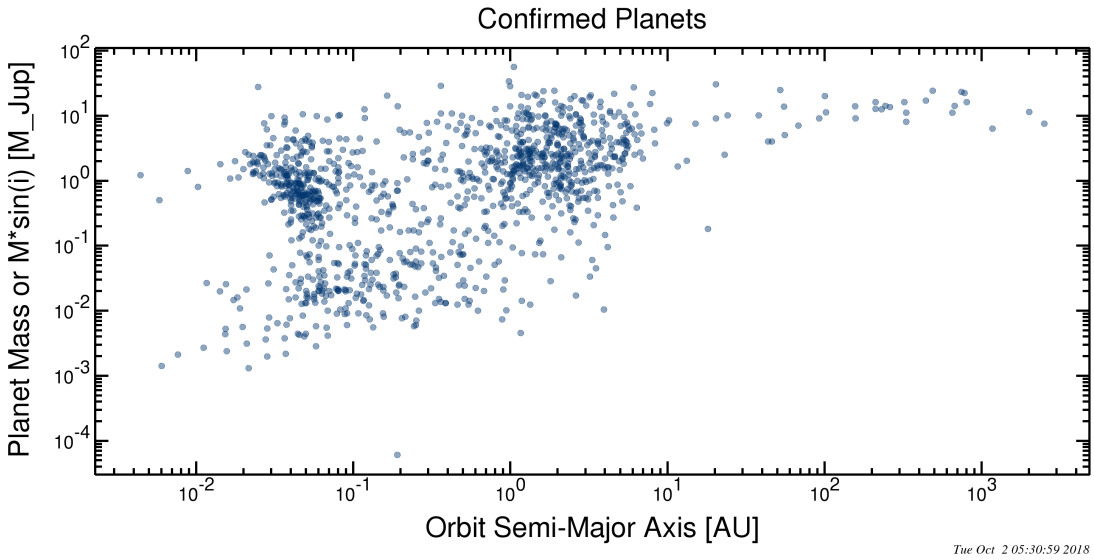

Figure 1.1: Mass distribution of confirmed exoplanets compared to their semi-major axis. Approximately 90% of these exoplanets have been detected using the transit or radial velocity methods. This figure is courtesy of https://exoplanetarchive.ipac.caltech.edu/.

evolution and formation of planetary systems, such that we can explain the diverse distributions of exoplanetary systems that we detect. Secondly, we want to know if life exists around stars other than our own. Hence, the focus on Earth-like worlds and habitable zones.

We focus our work on explaining the first of these two challenging problems. The dynamics, evolution and formation of planetary systems proves difficult to understand, as our models are purely the result of theory and time-constrained numerical simulations. In addition our understanding of these systems is hampered not only by our current observational equipment but also our static position in their evolutionary timeline. Nevertheless, great progress has been made in this field of research in order to explain our detections. It is this drive to understand these systems that motivates this work, at a very base level. To understand why these systems exist the way they do, we must first take a step away from exoplanets and understand the processes that drive their formation, back to the very early stages of the system's lifetime.

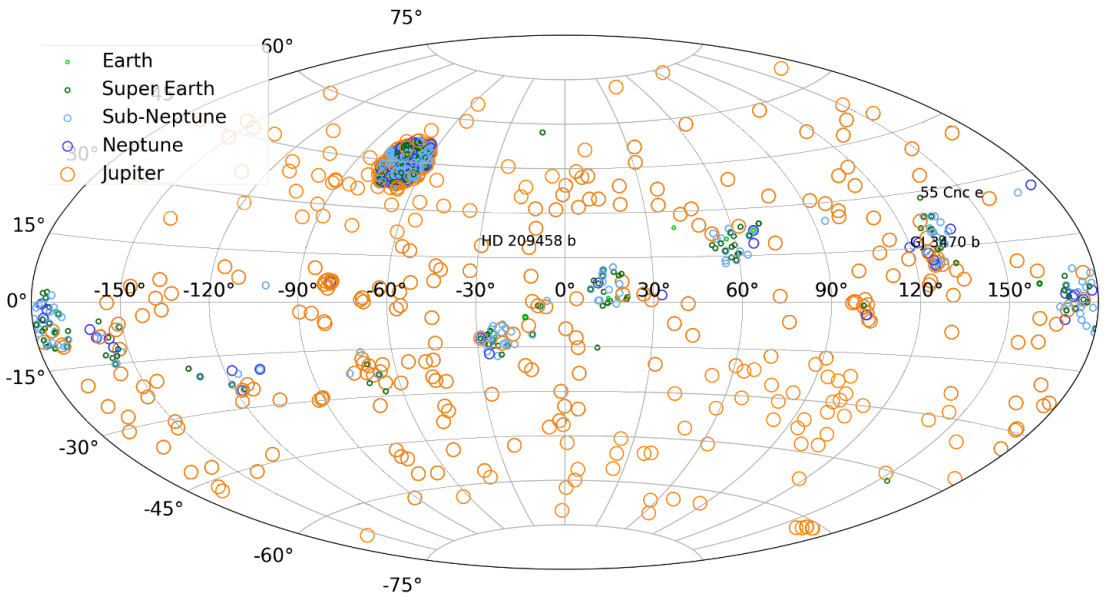

Figure 1.2: Distribution of currently detected transiting exoplanets across the sky. Clearly this shows the increased detection rate of Jupiter mass planets, most often hot Jupiters. The transit method has currently produced the largest number of confirmed exoplanet detections. Also visible in this plot is a clustered region showing the field of view of the Kepler space telescope. This figure has been adapted from Edwards et al. (2018).

1.2 Protoplanetary Discs

To understand the final state of a planetary system, we are required to look back at its earliest evolution. Initially young stars are surrounded by discs comprised of gas and dust. This is a result of angular momentum conservation during the gravitational collapse of the interstellar cloud that formed the star. Protoplanetary discs form rapidly after this gravitational collapse begins and will contain all the materials that can eventually form a planetary system.

Material at the inner edge of this disc will rapidly accrete onto the star, however, over time this accretion rate will decrease as the material is used up or disperses. This rapid accretion means that the disc mass does not increase during the gravitational collapse that forms the star. Following this, material accretion onto the star is driven by angular momentum loss of material within the disc. To conserve angular momentum this means the disc will spread out over time (Williams & Cieza, 2011). How this angular momentum transfer radially outwards occurs is one of the major

questions in understanding protoplanetary disc dynamics (Armitage, 2011). Naively one would expect friction due to disc viscosity to dissipate the angular momentum of disc material and drive accretion onto the star, however, the disc's molecular viscosity is not large enough to result in such angular momentum loss (Lynden-Bell & Pringle, 1974). A number of theoretical solutions to this problem have been posed, which largely revolve around the existence of a 'turbulent viscosity' (Shakura & Sunyaev, 1973) that mimics the effect of the kinematic viscosity in the disc. However, recent results imply that this turbulent viscosity may also be low (Flaherty et al., 2018). This means that the accretion flow may be laminar, rather than turbulent.

A prominent process that can provide turbulent viscosity is the Magnetorotational Instability (MRI). However, as we shall see this instability may not be applicable in all protoplanetary discs. As such we need to invoke angular momentum transport by hydrodynamic instabilities to account for this. These will be discussed in more detail in Section 1.3, as a number of these are particularly relevant to the work presented in this thesis. First we shall present the properties and basic equations that describe the evolution of a protoplanetary disc.

1.2.1 Properties

The primary source of heating for a protoplanetary disc is radiation emitted from the central star. As a result, there is a temperature gradient across the disc, with the inner regions being hotter and the outer disc being cooler. This means that protoplanetary discs emit a range of wavelengths in the sub-millimetre bands and so are best observed using infra-red and radio telescopes. One of the most well known of such telescopes is the Atacama Large Millimetre Array (ALMA). Similarly to exoplanet detection, the detection of protoplanetary discs has flourished over recent years, with ALMA providing some of the best high resolution images to date. Perhaps the most well known of these, HL Tau and TW Hydrae, can be seen in Figure 1.3 (ALMA Partnership et al., 2015; Andrews et al., 2016).

Observational evidence dictates that protoplanetary discs can have a wide range of properties, but we shall give a summary of these here for an average protoplanetary disc.

Mass

Protoplanetary discs often have masses of a few times a Jupiter mass (M_J) but can be significantly more massive, even around $50 M_\mathrm{J}$ (Williams & Cieza, 2011; van der Marel et al., 2019). Figure 1.4 shows the masses for protoplanetary discs

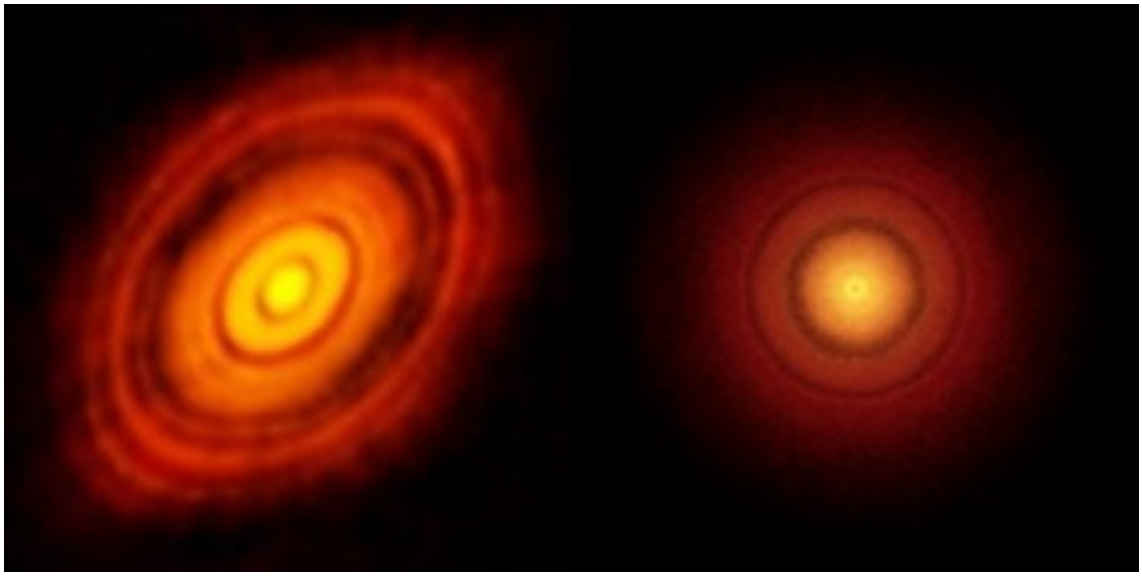

Figure 1.3: High resolution ALMA images of the protoplanetary discs surrounding the young ($\approx 10^6$yr) star HL Tau (left) and the older ($\approx 10^7$yr) star TW Hydrae (right). This shows the dust emission from these discs. There are clear similarities between these discs, despite their separation in age. This figure is courtesy of https://www.almaobservatory.org/en/home/.

in the Taurus, Ophiuchus and Orion regions (Andrews & Williams, 2005, 2007; Mann & Williams, 2010). This illustrates the range of masses generally expected for protoplanetary discs and shows that discs of extremely high mass are less common. These masses show reasonable agreement with the minimum mass solar nebulae model (MMSN), which predicts that the protoplanetary disc that formed our solar system must have a mass of $0.01 - 0.07 M_\odot$, based on the masses of the planets in our solar system (Kusaka et al., 1970; Weidenschilling, 1977). For reference, $1 M_\mathrm{J} \approx 10^{-3} M_\odot$. Inferring the masses of protoplanetary discs from observational results requires some simplifying assumptions to be made, which will of course result in some uncertainty on these values. For example, often only dust emitting in millimetre wavelengths is considered in observations. This leaves out possible large, more massive dust grains, resulting in an underestimate of the mass (Williams & Cieza, 2011). The total disc mass is dependent on the stellar mass and for low mass stars generally follows the relation $M_\mathrm{disc}/M_* \approx 0.01$. This breaks down for more massive stars, such as O stars, potentially due to their stronger photoevaporation effects on the disc (Williams & Cieza, 2011).

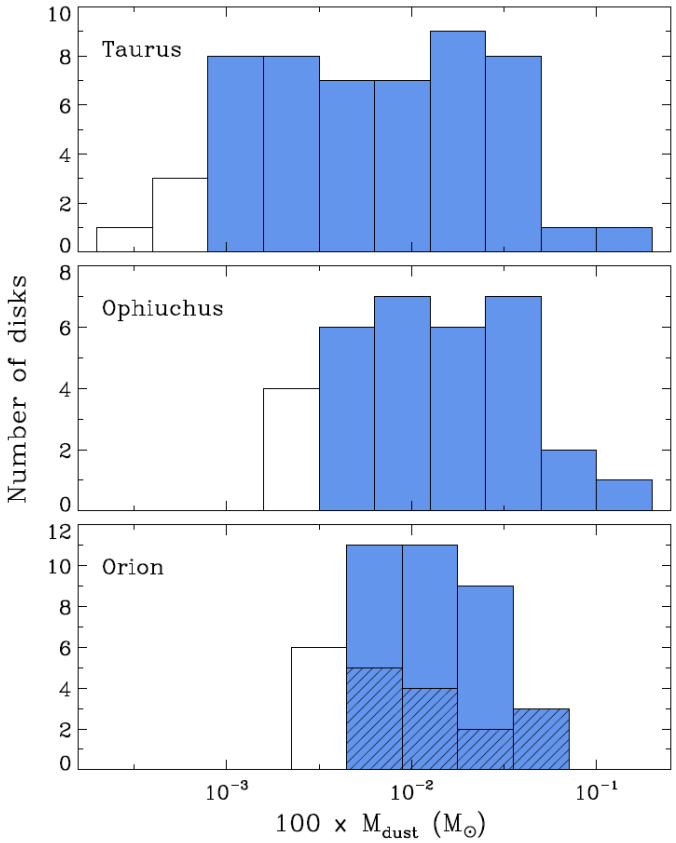

Figure 1.4: Mass distribution for protoplanetary discs in the Taurus, Ophiuchus and Orion regions (Andrews & Williams, 2005, 2007; Mann & Williams, 2010). Dust masses are derived from millimetre fluxes. The filled blue bars indicate ranges in which the millimetre flux data was complete. The dashed bars show discs at large distance from a highly photoionising star, where external photoevaporation may be less important (See Section 1.2.5). This figure is courtesy of Williams & Cieza (2011).

Radius

Another property of protoplanetary discs that provides some difficulties in measuring is their radial size. This arises from the temperature gradient across the disc, meaning that the outer, cool regions emit less strongly and so are more difficult to detect observationally. Hence, it can be difficult to say at what point the disc ends. In addition, we observe sub-millimetre emission from the dust in protoplanetary discs, while our observation of gas in the disc is significantly more limited. Often a tracer molecule such as ^{12}CO is used to determine the extent of the gas disc. Despite these difficulties, we know that protoplanetary discs can extend out to a few hundreds of AU from their dust emission (Vicente & Alves, 2005; Huang et al., 2018). Additionally, we know that the gas disc can extend out significantly further than the

dust disc (e.g. van der Marel et al., 2019). Finally, older discs are generally smaller than younger discs, a result of accretion onto the central star. Protoplanetary discs will spread due to angular momentum transport radially outwards. However, less material is moved outwards than inwards, resulting in the accretion onto the central star (Pringle, 1981). This also causes the outer regions to be less dense which, in combination with their lower temperature, makes them difficult to see observationally. Potentially, the reduction in disc radius could also be caused by phenomena such as photoevaporation blowing off the discs outer regions or material in the outer disc accreting to planetesimal size (van der Marel et al., 2019).

Lifetimes

The lifetimes and ages of protoplanetary discs are parameters that are very closely tied to the age of the star around which they are orbiting. If a protoplanetary disc is observed around a star, a good estimate of the age of that disc will be the same as the age of the star itself. Similarly, if a star is observed with no protoplanetary disc, that disc must have dissipated and the age of the star can act as an upper limit on the lifetime of the disc that once surrounded it. Unfortunately, stellar ages are not such an easy parameter to predict and are very model dependent (Siess et al., 2000; Baraffe et al., 2015). This can lead to large uncertainties in the predicted ages of discs, although we can generally determine whether a disc is young or old at worst (van der Marel et al., 2019). Nevertheless, we know that the large majority of protoplanetary discs have ages on the order of $1-10$Myr with some particularly old discs having ages of around 15Myr (Mamajek, 2009; Williams & Cieza, 2011; Long et al., 2018; Andrews et al., 2018; van der Marel et al., 2019). This is highlighted in Figure 1.5, which shows the age of a number of stellar samples against the fraction of those that have protoplanetary discs.

Composition

As we previously mentioned, protoplanetary discs are comprised of a combination of gas and dust. This is initially the composition of the interstellar medium (ISM), from which the interstellar cloud is formed that eventually gives rise to the star-disc system. The dust in the disc dominates the disc opacity and is what we see in sub-millimetre observations. Dust is comprised largely of grains with $r_{\rm dust} \leq 0.1\mu m$ with a mixture of graphite grains and polycyclic aromatic hydrocarbons (Draine, 2003). Gas, on the other hand, is very difficult to see as it emits only at specific wavelengths and hence requires high resolution spectroscopy to detect. Additionally,

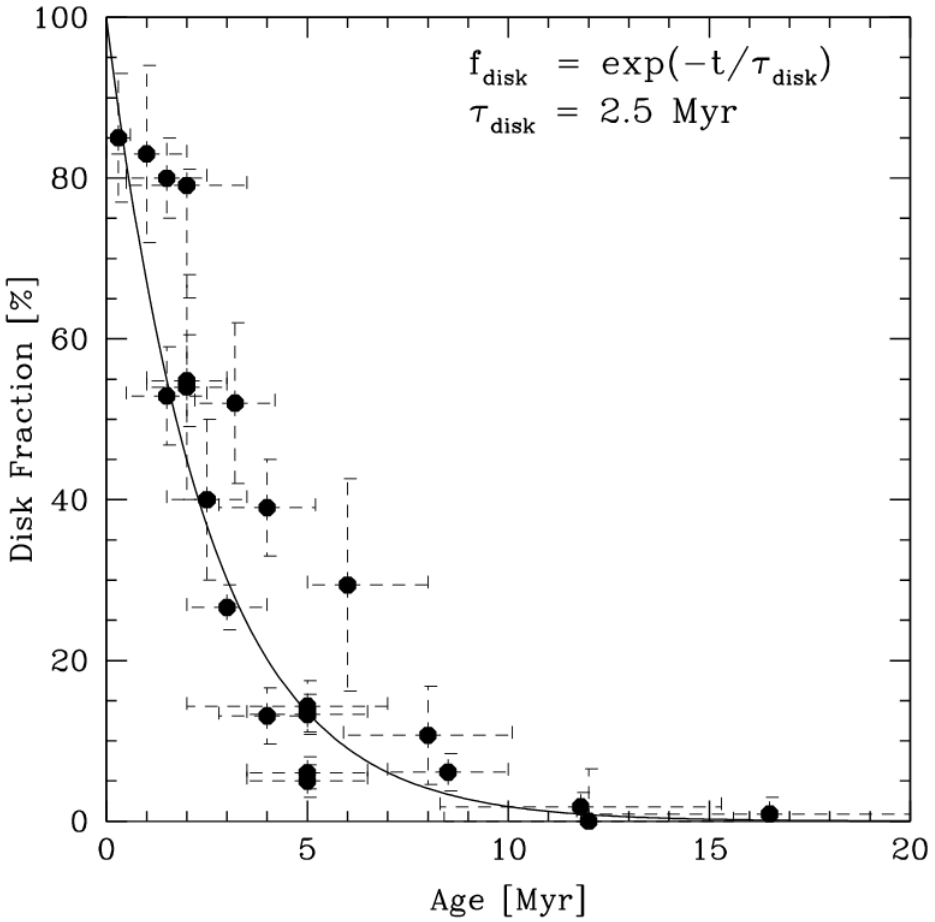

Figure 1.5: Age of a number of stellar samples vs the fraction of stars with discs visible in Hα emission or infrared excess. An exponential best fit curve has been applied which clearly shows that older stars are significantly less likely to host protoplanetary discs. This figure is courtesy of Mamajek (2009).

line emission can be hidden if the excitation temperature coincides with that of the dust (Williams & Cieza, 2011). Gas in protoplanetary discs exists largely in the form of H_2 but small amounts of atomic hydrogen can also exist due to photodissociation (Kamp et al., 2007). More complex molecules (e.g. C_2H_2, HCN, CO_2, OH, H_2O) have also been detected from absorption and/or emission lines (Lahuis et al., 2006; Carr & Najita, 2008). Similarly to the ISM, gas is more prevalent than dust in the disc, comprising 99% of the disc mass. Despite this, the behaviour of dust in the disc is still very important, not only because dust dominates the disc's opacity but also because it provides the building blocks for planet formation (see Section 1.4).

Structure

Protoplanetary discs are generally more dense in the inner regions and less dense in the outer. The gas density distribution usually follows a radial power law of the form $\rho \propto R^{-p}$, where ρ is the density of the disc, R is the radial position within the disc and p is a positive constant. However, protoplanetary discs are three dimensional structures. We can eliminate variation in the azimuthal direction by assuming axisymmetry, but disc parameters will often still vary in the vertical direction. The vertical gas density distribution of a vertically isothermal disc is set by the balance of the vertical components of stellar gravity and the disc pressure gradient. Hence, the density distribution of the gas in a disc can be written as,

$$\rho(R, z) = \rho(R)\exp\left(-\frac{z^2}{2H^2}\right). \tag{1.1}$$

Here z is the vertical direction and $\rho(R)$ is often taken to be a radial power law as discussed. $H = c_s/\Omega_{\rm K}$ is the disc scale height, where c_s is the sound speed in the disc and $\Omega_{\rm K}$ is the Keplerian angular velocity. Both of these quantities are also expected to follow a radial power law, $H \propto R^{(-q+3)/2}$ and $c_s^2 \propto R^{-q}$ where q is a positive constant (Takeuchi & Lin, 2002). The radial dependence of the scale height results in flaring of the disc. Observational evidence for flaring can be seen in discs silhouetted against scattered light from the central star, or background nebula (Burrows et al., 1996; Stapelfeldt et al., 1998; Padgett et al., 1999; Smith et al., 2005).

Despite this flaring, protoplanetary discs are thin such that $H \ll R$. This is a result of angular momentum conservation as the initial interstellar cloud collapses to form the star-disc system. This also means that an important quantity often of interest is the surface density of the disc, which is given by

$$\Sigma = \int_{-\infty}^{\infty} \rho dz. \tag{1.2}$$

This parameter is often important in two dimensional models, which can assume vertical equilibrium based on the thin nature of protoplanetary discs.

As alluded to in the discussion of disc masses, protoplanetary discs are significantly less massive than the central star. Hence, we would expect material in them to orbit at the Keplerian angular velocity,

$$\Omega_{\rm K} = \left(\frac{GM_*}{R^3}\right)^{\frac{1}{2}}, \tag{1.3}$$

where M_* is the stellar mass. However, due to the pressure force on the gas, the gas motion will differ, orbiting at a slightly sub-Keplerian velocity. Additionally, in a disc that is not in vertical equilibrium, there will be vertical variation in the angular velocity of the gas. The angular velocity distribution of the disc is then given by

$$\Omega = \Omega_{\mathrm{K}} \left(1 - \frac{1}{2} \left(\frac{H}{R} \right)^2 \left(p + q + \frac{q}{2} \frac{z^2}{H^2} \right) \right), \tag{1.4}$$

where p and q are the respective constants from the radial power laws. Here we can see that as z increases and we move away from the disc midplane, the angular velocity of the gas is reduced. At the disc midplane, the deviation from Keplerian velocity due to the disc pressure gradient is of the order of the aspect ratio squared, $h^2 = (H/R)^2$ (Takeuchi & Lin, 2002).

The dust in a protoplanetary disc behaves in a slightly different manner to the gas. Small dust grains will be well coupled to the gas, following its motion closely. However, as dust collides and grows in the disc it becomes more decoupled from the gas and begins to behave differently. As the dust does not feel the disc's pressure gradient, we expect this to orbit at the aforementioned Keplerian angular velocity. This means that the dust at a given radius will orbit faster than the gas. Therefore, the dust will feel a drag force from the gas ahead of it, which reduces the speed of the dust and causes the dust particles to drift radially inwards (Takeuchi & Lin, 2002). The strength of this drag force is dependent on dust size, increasing with size until the dust reaches cm to m sizes, then decreasing for dust larger than this. Additionally, as dust collides and grows it will begin to strongly feel the vertical component of the star's gravity. This acts to concentrate dust towards the disc midplane, which further increases the rate of collisions and dust growth. However, due to disc turbulence, the disc does not become completely vertically stratified. These processes are known as grain growth and dust settling and are the first steps towards planet formation, although growing the dust many orders of magnitude to terrestrial planet size is still an area of much debate (Dullemond & Dominik, 2005; Williams & Cieza, 2011).

Temperature

As we know, the temperature of a disc is dependent largely on the temperature of the star it surrounds. In addition if the disc is turbulent the temperature can also be set by viscous dissipation, although this is only effective in the inner disc. The temperature and sound speed, c_s, are related by

$$P = \frac{\Sigma R_{\rm m} T}{\mu} = c_s^2 \Sigma, \tag{1.5}$$

where P is the pressure, Σ is the disc surface density, R_m is the molar gas constant and μ is the mean molecular weight. For geometrically flat discs, which are non-flaring, we can relate the effective temperature of the disc to that of the star (Chiang & Goldreich, 1997),

$$T_e \approx \left(\frac{2}{3\pi}\right)^{\frac{1}{4}} \left(\frac{R_*}{R}\right)^{\frac{3}{4}} T_*, \tag{1.6}$$

where R_* and T_* are the stellar radius and temperature respectively.

The scale height of the disc is set by the disc temperature. For a blackbody disc in hydrostatic equilibrium (Chiang & Goldreich, 1997),

$$\frac{H}{R} = \left(\frac{T_e}{T_c}\right)^{\frac{1}{2}} \left(\frac{R}{R_*}\right)^{\frac{1}{2}} \tag{1.7}$$

where,

$$T_c = \frac{G M_* \mu}{k R_*} \tag{1.8}$$

is a measure of the gravitational potential at the surface of the star. Here M_* is the stellar mass. As protoplanetary discs are cold the scale height will be small and therefore the disc will be thin.

1.2.2 Disc Evolution

The evolution of a protoplanetary disc can be determined by solving the continuity equation for the disc's surface density, Σ,

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0 \tag{1.9}$$

and the two dimensional Navier-Stokes equation,

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}\right) = -\nabla P - \nabla \cdot T - \Sigma \nabla \Phi, \tag{1.10}$$

where T is the Newtonian viscous stress tensor, Φ is the gravitational potential of the star and $\mathbf{v}$ is the velocity vector. We use the two dimensional equations as we expect discs to be thin and for their vertical structure to be independent from the equations describing their radial evolution. This allows us to integrate vertically

across a thin disc and work in terms of surface density rather than material density. We use a cylindrical coordinate system so that $\mathbf{v} = (v_R, R\Omega)$, where v_R and Ω are the radial and angular velocities respectively. Now, splitting Equation 1.10 into its component forms we find,

$$\frac{\partial v_R}{\partial t} + v_R\frac{\partial v_R}{\partial R} + \Omega\frac{\partial v_R}{\partial \phi} - R\Omega^2 = -\frac{1}{\Sigma}\frac{\partial P}{\partial R} - \frac{2}{\Sigma R}\frac{\partial}{\partial R}\left(\nu\Sigma R\frac{\partial v_R}{\partial R}\right) - \frac{1}{\Sigma R}\frac{\partial}{\partial \phi}\left[\nu\Sigma\left(R\frac{\partial \Omega}{\partial R} + \frac{1}{R}\frac{\partial v_R}{\partial \phi}\right)\right] - \frac{\partial \Phi}{\partial R} \quad (1.11)$$

and, noting that the star's potential is axisymmetric,

$$\frac{\partial (R\Omega)}{\partial t} + v_R R\frac{\partial \Omega}{\partial R} + \Omega\frac{\partial (R\Omega)}{\partial \phi} + 2v_R\Omega = -\frac{1}{R\Sigma}\frac{\partial P}{\partial \phi} - \frac{2}{\Sigma R}\frac{\partial}{\partial \phi}\left(\nu\Sigma\frac{\partial \Omega}{\partial \phi}\right) + \frac{1}{\Sigma R^2}\frac{\partial}{\partial R}\left[\nu\Sigma R^2\left(R\frac{\partial \Omega}{\partial R} + \frac{1}{R}\frac{\partial v_R}{\partial \phi}\right)\right], \quad (1.12)$$

where ν is the kinematic viscosity of the disc. We assume axisymmetry, this means we can take the azimuthal average of Equation 1.12,

$$\Sigma R\frac{\partial (R\Omega)}{\partial t} + \Sigma v_R\frac{\partial (R^2\Omega)}{\partial R} = \frac{1}{R}\frac{\partial}{\partial R}\left(\nu\Sigma R^3\frac{\partial \Omega}{\partial R}\right). \quad (1.13)$$

Assuming Keplerian rotation, Equation 1.13 can be written as,

$$\Sigma v_R R = -3R^{\frac{1}{2}}\frac{\partial}{\partial R}(\nu\Sigma R^{\frac{1}{2}}). \quad (1.14)$$

Using the azimuthally averaged form of Equation 1.9,

$$R\frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial R}(\Sigma v_R R) = 0, \quad (1.15)$$

we eliminate the $\Sigma v_R R$ term and find the surface density evolution of an axisymmetric two dimensional disc will be given by (Pringle, 1981)

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{R}\frac{\partial}{\partial R}\left[3R^{\frac{1}{2}}\frac{\partial}{\partial R}\left(\nu\Sigma R^{\frac{1}{2}}\right)\right]. \quad (1.16)$$

The time evolution of an axisymmetric protoplanetary disc's surface density has one main outcome, the accretion of disc material onto the central star. Solving Equa-

tion 1.16 shows us that the disc will spread out over time, with the majority of disc material losing angular momentum and drifting radially inwards. This angular momentum is transported outwards such that a smaller amount of disc material will move radially outwards, causing the spreading of the disc. This angular momentum transport arises due to the viscosity acting on material in the disc. However, it is not currently clear what gives rise to viscosity in protoplanetary discs, as the molecular viscosity is not large enough to provide angular momentum loss due to friction. Hence, it is very desirable to find methods in which a turbulent viscosity (Shakura & Sunyaev, 1973) can be introduced to the disc, to provide angular momentum transport. A turbulent viscosity can provide a much larger effective viscosity than the molecular viscosity, due to the increase in the length scale involved. This could potentially explain the observed accretion rates of protoplanetary discs, which demand a viscosity orders of magnitude larger than the molecular viscosity. One such method of providing a turbulent viscosity is the Magnetorotational Instability.

1.2.3 Magnetorotational Instability

As previously discussed, a major problem in the understanding of protoplanetary discs is finding a satisfying explanation for the transport of angular momentum outwards across the disc. In order to solve this problem a turbulent viscosity was invoked (Shakura & Sunyaev, 1973) as the disc's molecular viscosity alone is not large enough to explain the transport of angular momentum (Lynden-Bell & Pringle, 1974). One process that can provide such a turbulent viscosity is the Magnetorotational Instability (MRI, Balbus & Hawley, 1991). This instability was the first widely accepted explanation for the turbulent viscosity in protoplanetary discs and indeed was commonly believed to be the solution to the angular momentum transport problem. As we will see, this may not necessarily be correct.

The MRI occurs due to the impact of the magnetic field of the star on the material orbiting in the accretion disc. The action of this instability is as follows. Consider a differentially rotating disc of material with $d\Omega^2/dR < 0$, (i.e a protoplanetary disc) threaded by a weak vertical magnetic field. An outwardly displaced fluid element in such a disc is orbiting at a smaller angular velocity than its corresponding inner fluid element. However, the magnetic field is resisting the shearing and attempting to enforce rigid rotation. This means that the magnetic field is trying to force the fluid element to rotate too quickly for its radial position. The excess centrifugal force pushes the fluid element even further radially outwards. The magnetic field also applies a stabilising force which resists the stretching and attempts to return the

fluid element to its original position. However, at long wavelengths this stabilising force is weak, hence the destabilisation wins and the instability is excited (Balbus & Hawley, 1991).

The requirements and procedure listed above does need the assumption of ideal Magnetohydrodynamics (MHD) to be made. These conditions make the MRI a likely source of angular momentum transport in discs surrounding objects such as white dwarves, neutron stars and black holes. However, this is not the case in protoplanetary discs, where the gas is cool, dense and mostly neutral and therefore we cannot make the assumption of ideal MHD. In the non-ideal regime, we now have to contend with a number of extra effects. If we consider just a single effect, the Ohmic resistivity, we now introduce a magnetic resistivity η into the disc that will attempt to smooth out small scale structure in the magnetic field (Gammie, 1996; Armitage, 2011). Considering a length equal to the disc scale height, H, the timescale for the growth of the MRI in an ideal disc is $\tau_{\rm MRI} \approx H/v_A$ where v_A is the Alfven velocity. Similarly the timescale for resistive damping is $\tau_\eta \approx H^2/\eta$. As $\tau_\eta \leq \tau_{\rm MRI}$ to suppress the MRI, we find that $\eta \geq H v_A$. Defining the magnetic Reynolds number Re_M, this condition can be written as,

$$\mathrm{Re}_M = \frac{H v_A}{\eta} \leq 1 \tag{1.17}$$

for the MRI to not be excited in a non-ideal MHD protoplanetary disc containing solely Ohmic diffusion. The next question is, is it possible for $\mathrm{Re}_M > 1$ in a protoplanetary disc? This relies on the fraction of electrons to neutral particles in the disc, n_e/n. Hence, ionisation becomes particularly important. Thermal ionisation is prominent in the inner regions of the disc due to proximity to the star, however, as we move radially outwards we must rely on non-thermal sources to ionise the disc sufficiently to excite the MRI.

Summarising, there are three main factors that must be considered if the MRI can be excited in protoplanetary discs,

- The impact of the non-ideal terms on the MRI.
- The distribution of ionisation across the disc due to non-thermal processes.
- The recombination rate due to gas phase and surface processes on dust grains.

These have been discussed in detail in Armitage (2011). Gammie (1996) combines potential answers to these questions to form a disc model in which angular momentum is transported due to the MRI. It is proposed that the protoplanetary

disc develops a layered structure with three distinct radial regions. Firstly, the region of the disc closest to the star, which as we have already discussed is sufficiently thermally ionised by stellar radiation. Secondly, an outer disc region in which ionisation occurs due to non-thermal effects, for example cosmic rays and x-ray emission from the star. Finally, the region in between these inner and outer regions is potentially the most interesting. Here it is predicted that a magnetic dead zone will form around the mid plane of the disc, where the disc is cool enough and well shielded enough from ionising radiation that the MRI will not be excited. Here the disc has a layered vertical structure, as closer to the surface (top and bottom) of the disc the MRI can still be active (Gammie, 1996). This structure can be seen in Figure 1.6.

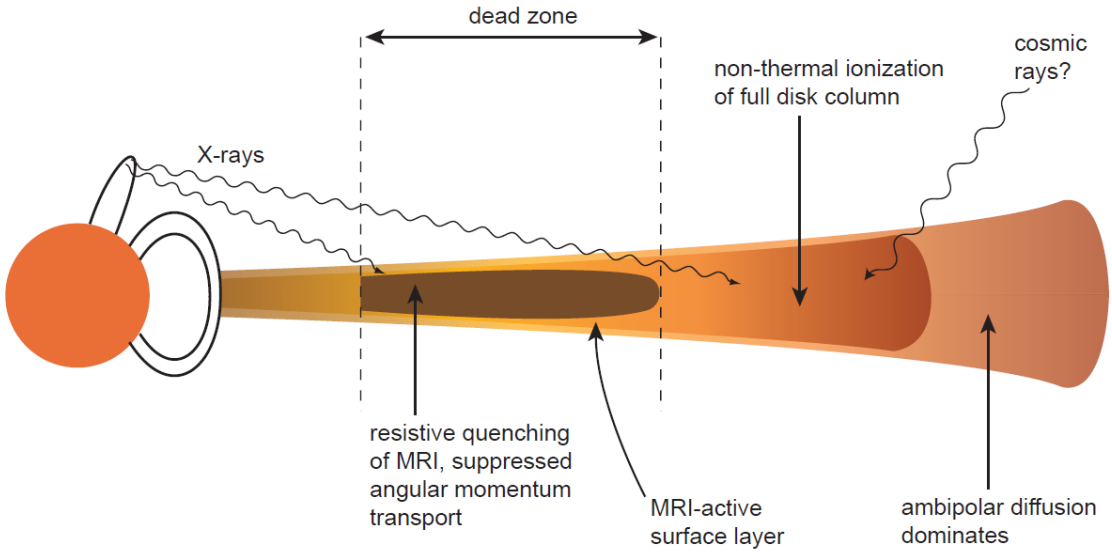

Figure 1.6: Vertical and radial structure of an MRI active protoplanetary disc following the prescription of Gammie (1996). Visible here are the three distinct radial regions. The inner thermally ionised region, the outer non-thermally ionised region and the vertically layered region in between, in which the MRI is inactive around the midplane. This figure has been adapted from Armitage (2011).

Theoretically, it is possible that these dead zones can exist in the range $R = 1-10$AU (Salmeron & Wardle, 2008; Terquem, 2008; Bai & Goodman, 2009; Turner et al., 2010), although there is no conclusive observational evidence for or against their existence. The fact that dead zones may fall in this radial range makes the dynamics in these regions particularly worth investigating, as this is the region that we are most interested in when it comes to planet formation (see Section 1.4). A particularly important consequence of magnetic dead zones is that the MRI cannot transport angular momentum in these regions. Hence, we must invoke other methods

by which angular momentum is transported.

1.2.4 Other Angular Momentum Transport Methods

As mentioned previously, the existence of substantial magnetic dead zones means there must be other methods of transporting angular momentum across the disc. In these dead zones we are in a regime where purely hydrodynamical instabilities can be excited. These largely fall into two categories, instabilities driven by the shear in the disc and those driven by the entropy gradient across the disc. In MHD, the MRI is a shear driven instability arising from the restoring force of the magnetic field. Generally hydrodynamic shear driven instabilities arise due to a deviation from the Keplerian rotation profile of the disc. Examples include the Rossby Wave and Rayleigh instabilities, discussed in Section 1.3, that can arise due to over or under dense regions perturbing the orbital velocity of material. An example of an entropy driven instability is the Subcritical Baroclinic Instability (SBI), which relies on a radially unstable entropy gradient and a suitable thermal diffusion timescale. This, again, will be discussed in more detail in Section 1.3.

Angular momentum transport may also arise, in some part, due to the formation of planets within the disc. The above listed instabilities (gravitational, MRI, SBI etc.) all act to concentrate disc material into vortices, which could potentially be the birthplace of planets. Planets could contribute to angular momentum transport via gravitational interaction with the disc material, exciting density waves at the inner and outer Lindblad resonances. If this method can provide significant angular momentum transport, it is expected that the main contributors would be low mass planets that have not opened gaps in the disc (Goodman & Rafikov, 2001). This process will be discussed in more detail in Section 1.5. However, it is very unlikely that this is the driving method of transporting angular momentum in protoplanetary discs. Nevertheless planet-disc interactions can have a large effect on disc evolution. Specifically giant planets can heavily modify the density structure of the disc, as we shall see in Section 1.5.2.

1.2.5 Photoevaporation

Despite not acting as a method for transporting angular momentum across the disc, the process known as photoevaporation can be very important for driving the evolution of the disc. In this process ionising photons are incident upon the outer layers (top and bottom) of the protoplanetary disc. This heats a thin outer layer of the

disc material, causing the gas in this layer to become unbound and flow freely away from the disc. Combining the ideas of photoevaporation and angular momentum transport we can picture the evolution of a disc in which photoevaporation is important. Initially, the accretion rate of material onto the central star is high. The disc spreads due to angular momentum transfer radially outwards and the outer regions of the disc supply material to the accreting inner regions. As time passes this accretion rate decreases, until it becomes equal to the rate of photoevaporation. At this point, the inner disc is accreted over the viscous timescale but is not resupplied by the outer disc. This leaves an inner cavity and exposes the disc's edge to direct radiation from the central star. This causes the disc to rapidly photoevaporate from the inside out, until no disc remains (Clarke et al., 2001; Alexander et al., 2006a,b). As a result, we can see that the disc lifetime and evolution is closely linked to photoevaporation, if it occurs in the disc (Williams & Cieza, 2011).

Photoevaporation can be split into two types, dependent on the source of the ionising radiation. These are internal and external photoevaporation. Internal photoevaporation occurs due to radiation from the central star and can be important for both massive and low mass stars. Internal photoevaporation is impacted by the far ultraviolet, extreme ultraviolet and x ray emission from the star (Armitage, 2011). External photoevaporation occurs due to radiation from external stellar sources (Johnstone et al., 1998; Adams et al., 2004). Hence, it is at its strongest for discs in compact stellar clusters and the strength of this type of photoevaporation increases with the mass and density of stars in such a cluster (Armitage, 2000; Fatuzzo & Adams, 2008). This is apparent observationally, as discs in high density stellar clusters tend to be more compact (Eisner et al., 2018) than those in less dense clusters (Long et al., 2018). External photoevaporation is mostly dependent on the far ultraviolet and extreme ultraviolet emission from these stars. Despite the strength of the ionising radiation from compact clusters, it is possible that internal photoevaporation could still be the more dominant of these processes. This is because the length scales involved across the protoplanetary disc are significantly shorter than those across interstellar space (Armitage, 2011).

1.3 Hydrodynamic Instabilities in Protoplanetary Discs

We have discussed disc instabilities in Section 1.2, as a method of transporting angular momentum and providing a turbulent viscosity. Specifically the MRI, while

briefly touching upon hydrodynamic instabilities such as the Rayleigh and Rossby wave. Here we will discuss in more detail the hydrodynamic instabilities that can occur in protoplanetary discs.

Hydrodynamic instabilities and subsequent vortex formation have been proposed as a method of providing turbulent viscosity in protoplanetary discs (Johnson & Gammie, 2005). Interest in these has been renewed following the realisation that many discs may not be sufficiently ionised for the MRI to be active, especially in the regions in which planet formation is expected to occur. This means that there are significantly large regions within protoplanetary discs in which turbulence may arise mainly from hydrodynamic instabilities.

There is some evidence for the presence of vortices in observations, as non-axisymmetric structure has been identified in a number of discs (van der Marel et al., 2013; Fukagawa et al., 2013; Pérez et al., 2014; Marino et al., 2015; Cazzoletti et al., 2018). As vortices are the non-linear outcome of many hydrodynamic instabilities, this provides strong evidence that vortex forming instabilities occur in protoplanetary discs. Examples of such observations are given in Figure 1.7, which clearly show the presence of non-axisymmetric structure. Vortices are also well known in thin two-dimensional planetary atmosphere models, with a common example being Jupiter's Great Red Spot (Sommeria et al., 1988; Marcus, 1988, 1990).

There are many candidates for providing hydrodynamic turbulence in protoplanetary discs. Here we shall discuss a number of hydrodynamic instabilities that can arise in protoplanetary discs. Some of which, such as the aforementioned Rossby wave and Rayleigh instabilities, are very relevant to the work in this thesis. Others, such as the gravitational instability, are less important to our work but remain important in understanding the behaviour of protoplanetary discs.

1.3.1 Convection

One method that can possibly transport angular momentum across a protoplanetary disc is convection. Convection was a primary method of producing a turbulent viscosity in early disc models (Lin & Papaloizou, 1980), however, have since been abandoned due to its inefficiency. More recently it was shown that convection should again be considered as a method of transporting angular momentum, although it is more efficient at moving heat vertically in the disc than it is at moving angular momentum radially outwards (Lesur & Ogilvie, 2010).

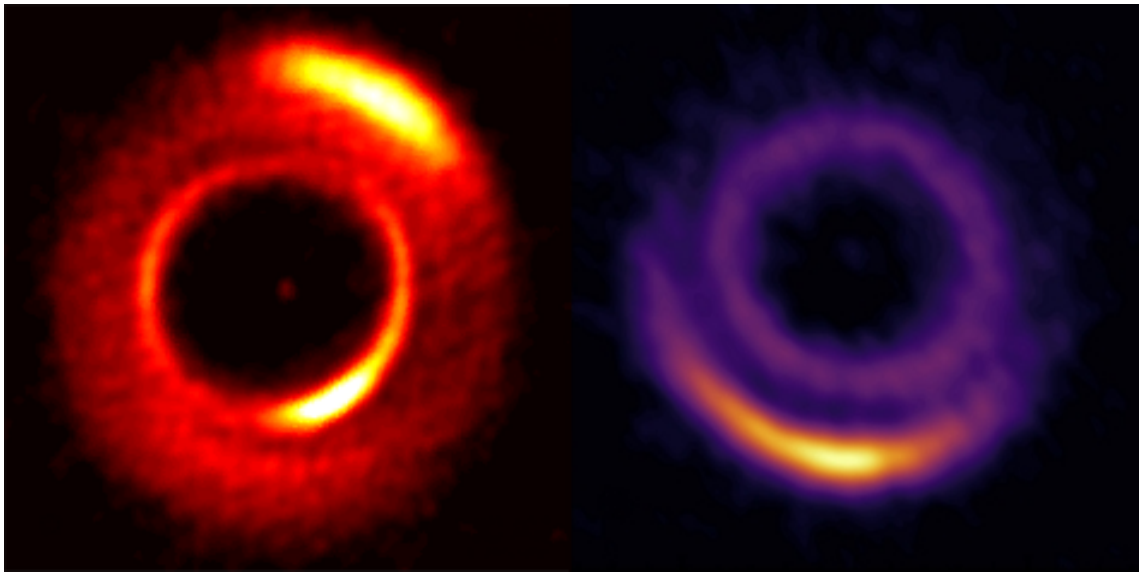

Figure 1.7: High resolution ALMA images of non-axisymmetric protoplanetary discs surrounding the star MWC758 (left, adapted from Dong et al., 2018) and the star HD135344B (right, adapted from Cazzoletti et al., 2018). Clearly these show the presence of potential vortices, which could be the result of the excitation of an instability, such as the Rossby Wave Instability.

1.3.2 Gravitational Instability

The gravitational instability has been considered an alternative to the MRI when it comes to providing a turbulent viscosity in the disc. However, as we shall see it too has some limitations that prevent it from being applicable to all discs. Hence, it is not an all encompassing solution to the angular momentum transport problem.

The gravitational instability arises due to the self-gravity of material within the disc. The stability of a thin, axisymmetric disc to the gravitational instability is set by the Toomre Q parameter (Toomre, 1964),

$$Q = \frac{c_s \kappa}{\pi G \Sigma}, \tag{1.18}$$

where c_s is the sound speed, $\kappa \approx \Omega_K$ is the epicyclic frequency and Σ is the surface density. Axisymmetric perturbations to the equilibrium disc structure begin to grow exponentially for $Q \leq 1$. However, often protoplanetary discs will not enter this regime, as self-gravity becomes important and impacts disc structure when Q is of the order 1. As the mass enclosed within a radius R is $M_d \approx \pi R^2 \Sigma$ and the disc scale height $H = c_s/\Omega_K$, we can rewrite $Q \leq 1$ as,

$$\frac{M_d}{M_*} \geq \frac{H}{R}. \tag{1.19}$$

From which we can see that for the gravitational instability to be excited, the disc must be massive (Armitage, 2011). This means that the gravitational instability is likely to be more common in younger discs, as accretion onto the star will reduce the disc mass.

Once the gravitational instability is excited the system becomes non-linear and the outcome cannot be predicted via analytical solutions to evolution equations, rather numerical simulations are required to explore this regime. There are two broad consequences to the disc becoming gravitationally unstable. The disc may enter a quasi-steady state of self-gravitating turbulence with spiral arms transporting angular momentum outwards via gravitational torques (Lynden-Bell & Kalnajs, 1972) or the disc may fragment (Boss, 1997; Armitage, 2011).

These outcomes are largely differentiated by the thermodynamics of the disc (Durisen et al., 2007). As the gravitational instability is excited, spiral waves begin to grow in the disc which act to transport angular momentum outwards. This is the first outcome and is a direct consequence of the excitement of the gravitational instability. As these waves grow they can give rise to shocks which result in strong heating at locations in the disc. If the cooling is too fast this can result in fragmentation of the disc. This is due to cool material 'clumping' to form a dense region in which the pressure due to gravity exceeds the opposing thermal pressure of the material. However, if the heating and cooling are balanced the disc can remain in its quasi-steady state with no fragmentation occurring.

For fragmentation to occur the cooling timescale must be less than a critical value (Gammie, 2001; Meru & Bate, 2010),

$$\beta = t_{\rm cool}\Omega \lesssim \beta_{\rm c}, \tag{1.20}$$

in which,

$$t_{\rm cool} = u\left(\frac{du_{\rm cool}}{dt}\right)^{-1}, \tag{1.21}$$

where u is the internal energy and $du_{\rm cool}/dt$ is the total cooling rate. For a viscous disc in thermal equilibrium, this cooling timescale is related to the turbulent, α, viscosity (Shakura & Sunyaev, 1973) by (Pringle, 1981; Gammie, 2001; Rice et al., 2005)

$$\alpha = \left(\gamma(\gamma-1)\frac{9}{4}\beta\right)^{-1}. \tag{1.22}$$

Here γ is the adiabatic index (the ratio of specific heats) and $\nu = \alpha c_s H$. Hence, the parameter β_c is dependent on the ratio of specific heats. It has been shown in two dimensions that for $\gamma = 2$, $\beta_c \approx 3$ (Gammie, 2001). Additionally in three dimensions it has been shown that for $\gamma = 5/3$ and $7/5$, $\beta_c \approx 7$ and 13 respectively (Rice et al., 2005). However, as the cooling rate in protoplanetary discs is difficult to estimate Equation 1.22 does not provide us with a definitive value for the turbulent viscosity (Armitage, 2011).

Disc fragmentation is particularly relevant to planet formation, as will be discussed in Section 1.4.2. However, while the conditions for disc fragmentation are well understood, whether or not these are realistic and can result in planets is not (Durisen et al., 2007). Otherwise, in non-fragmenting discs, the excitement of the gravitational instability provides a means of angular momentum transport outwards across the disc (Armitage, 2011).

The existence of the gravitational instability does not mean that other hydrodynamic instabilities are of no importance. As can be seen, for the gravitational instability to be excited the protoplanetary disc must be very massive and potentially, very young. As the excitement of the gravitational instability has quite a distinct effect on the structure of the disc, we also have observational evidence that many discs are not gravitationally unstable. For example, compare the simulated image of a non-fragmenting gravitationally unstable protoplanetary disc shown in Figure 1.8 to the observational images of non-axisymmetric protoplanetary discs shown in Figure 1.7. The difference in disc structure between these images is extremely clear. Hence, there is a significant number of discs in which we expect disc self-gravity does not drive angular momentum transport and are not sufficiently ionised for the MRI to be excited. It is in these discs that other hydrodynamic instabilities become the primary method for providing a turbulent viscosity and angular momentum transport.

1.3.3 Subcritical Baroclinic Instability

The Subcritical Baroclinic Instability (SBI, Lesur & Papaloizou, 2009, 2010) is an entropy driven instability that can occur in protoplanetary discs. This instability has an interesting history. It was originally observed by (Klahr & Bodenheimer, 2003) in two and three dimensional protoplanetary disc simulations containing a negative

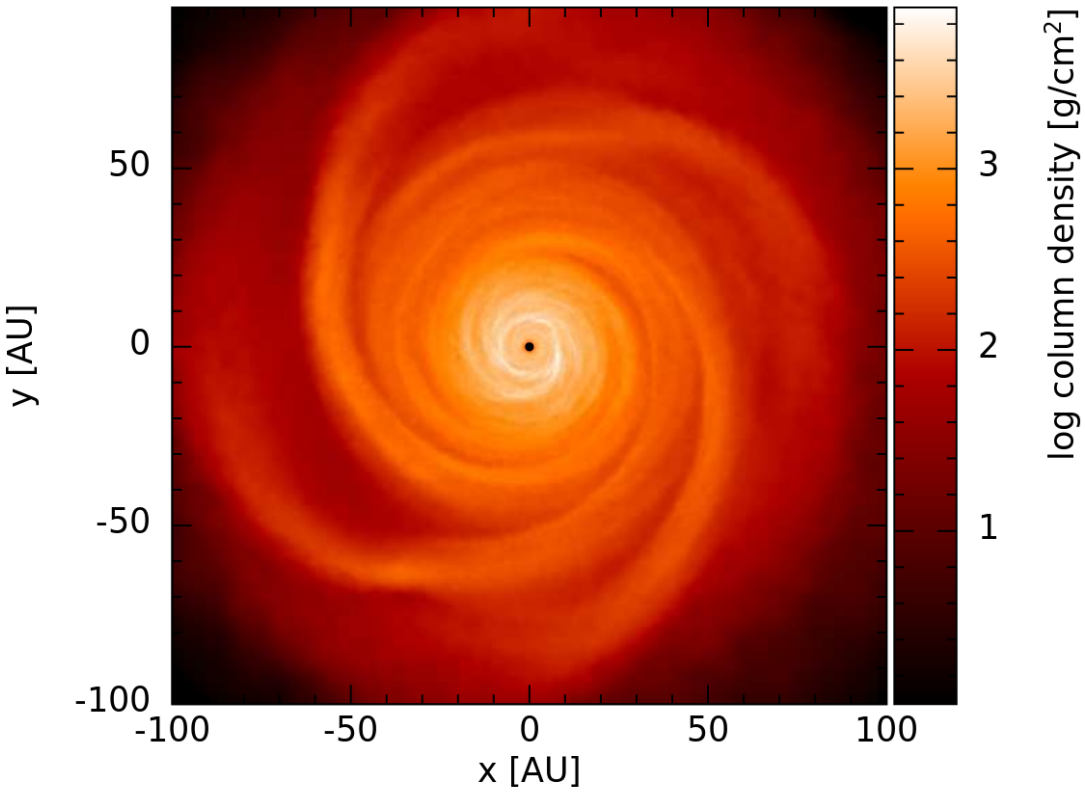

Figure 1.8: A simulated image of a disc in which the gravitational instability is active. This highlights the formation of spiral waves, but the image is taken before fragmentation occurs. This particular protoplanetary disc will undergo fragmentation at a later time. If we are to compare this disc to those shown in Figure 1.7 we can see the outcome of the gravitational instability looks extremely different to the standard vortex formation. This figure has been adapted from Meru (2015).

entropy gradient. However, the existence of this instability was not confirmed until much later, in two dimensional global and local simulations (Petersen et al., 2007a,b; Lesur & Papaloizou, 2010). This instability became known as the SBI due to its subcritical nature, requiring a finite amplitude initial perturbation to the disc for it to be excited (Lesur & Papaloizou, 2010). The excitement of this instability requires a radially unstable entropy gradient, which is dependent on the stability of the system against convection. The presence of this depends on two instability criteria, the Schwarzschild criterion and the Solberg-Hoiland criterion for the onset of convection in non-shearing and shearing flows respectively. Both of these depend on the Brunt-Väisälä frequency which describes the buoyancy in the disc,

$$N_R^2 = -\frac{1}{\gamma\rho}\frac{dP}{dR}\frac{d}{dR}\ln\left(\frac{P}{\rho^\gamma}\right), \tag{1.23}$$

where γ is the adiabatic index and ρ is the density. Then, the Schwarzschild criterion states,

$$N_R^2 < 0, \tag{1.24}$$

for the system to be unstable. Similarly, for shearing flows the Solberg-Hoiland criterion states,

$$N_R^2 + \Omega_{\rm K}^2 < 0, \tag{1.25}$$

for the growth of perturbations in the system. For the SBI to exist, the disc must be Schwarzschild unstable, but Solberg-Hoiland stable (Petersen et al., 2007a,b; Lesur & Papaloizou, 2010; Armitage, 2011).

However, this is not the only factor that is important in the existence of the SBI. While the above describes the excitement of the SBI, for its continued survival the thermal diffusion timescale in the disc becomes very important. In an SBI unstable disc, vortices will exist with a given aspect ratio. As material circulates these vortices it will travel in two directions, radially inwards/outwards and azimuthally. On the radially inwards/outwards journey material is subject to buoyancy forces, whereas on the azimuthal journeys thermal diffusion is felt by the material. If thermal diffusion occurs on a certain timescale, not too fast to reduce the buoyancy effects and not too slow to become close to adiabatic, then this vortex can become self-sustaining. To better describe this process we will refer to Figure 1.9, which shows the motion of a fluid particle circulating a vortex. Consider a fluid particle starting at point A, moving radially outwards towards point B. Assuming the particle moves quickly from A to B, it retains the temperature and density it had in thermal equilibrium with the disc at point A. Hence, it is hotter and lighter than the surrounding material and experiences a buoyancy force directed radially outwards. At B, the particle moves azimuthally towards C. On the route from B to C, the temperature and density is constant. Hence, if cooling happens fast enough, the fluid particle will reach thermal equilibrium with its surroundings. When the fluid particle reaches C, it moves radially inwards towards D. Similarly to the previous radial displacement, the fluid particle is now cooler and heavier than its surroundings on its way to D, so experiences a buoyancy force directed radially inwards. At D, the particle moves azimuthally again back towards A and will reach thermal equilibrium with

its surroundings on this route, as before. Overall, this causes an amplification of the vortex due to the acceleration from the bouyancy forces. This process highlights the importance of the thermal relaxation timescale, as if this was too fast the material would reach equilibrium with the surroundings while on the AB and CD routes and if it was too slow, would never reach thermal equilibrium with the surrounding on the BC and DA routes (Lesur & Papaloizou, 2010).

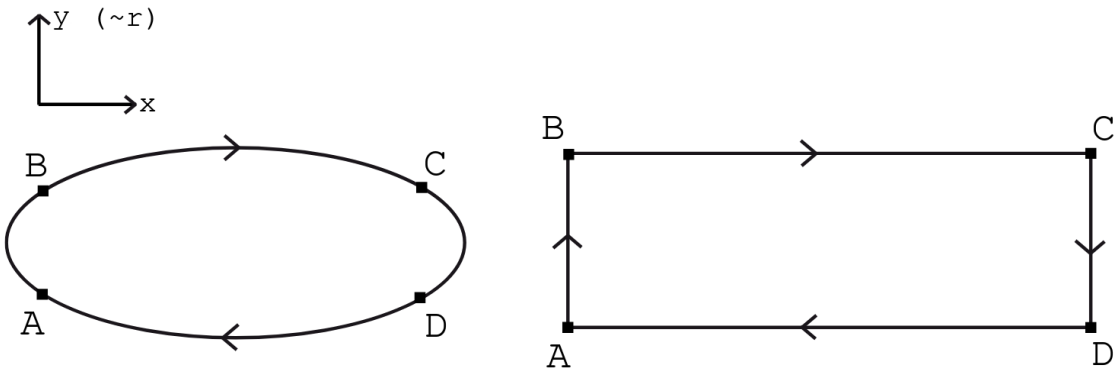

Figure 1.9: A streamline for a vortex undergoing the SBI in a protoplanetary disc. For certain thermal relaxation timescales, a vortex such as this can be self-sustaining. The process behind this is described in Section 1.3.3. This figure has been adapted from Lesur & Papaloizou (2010).

1.3.4 Convective Overstability

The Convective Overstability (Klahr & Hubbard, 2014; Lyra, 2014) is an instability with a similar driving mechanism to the SBI. Crucially, the Convective Overstability drives fluid elements to undergo epicyclic motion, rather than circulating a vortex. This means that there are no vortices associated with the Convective Overstability. This instability occurs when a fluid element performs epicyclic motion in a disc containing a weak entropy gradient and hence is cooler at larger radii. As the fluid element moves radially outwards it transfers some of its heat to the cooler surroundings, such that when it returns to the initial radial position it is now cooler than it originally was. This means the fluid element has a smaller entropy than the equilibrium at this radius and so feels a buoyancy force that pushes it radially inwards. While radially inwards the fluid element raises in temperature, such that when it returns to its initial position it now has larger entropy than the surrounding material. The buoyancy force on the fluid element now pushes it radially outwards. This causes the oscillations to grow larger and clearly the process begins to run away.

However, this instability can be suppressed based on the thermal diffusion of the system. If the thermal diffusion is too fast the temperature of the fluid element would rapidly reach equilibrium with its surroundings. Therefore, its entropy would never diverge from that of the surrounding material and would never feel the buoyancy force. If the thermal diffusion was too slow the material would never change in temperature across its epicycle. Hence, the entropy would remain constant and the buoyancy force would adjust the epicycles (Latter, 2016).

1.3.5 Zombie Vortex Instability

Another instability that may exist in three dimensional protoplanetary discs is the Zombie Vortex Instability (Marcus et al., 2013, 2015). This instability occurs in the magnetic dead zones of the disc and requires the vertical thermodynamical structure to be stably stratified, introducing baroclinic critical layers to the flow. Anticyclones in the disc can perturb critical layers above and below them, forming dipolar vortex layers, oppositely signed layers of potential vorticity close together, at these critical layers. Vortex layers that have the same rotation as the background shear form vortices and eventually create anticyclones at the critical layers. These anticylcones then repeat this process and the vortices spread across the disc. Whether this instability is realistic or not is a matter of debate, as it is easy to produce in simulations using hyper-diffusion operators but not when using classical viscous diffusion operators. This is due to the dependence of the critical layer's characteristic lengthscale on the diffusion. Therefore, it is very important that realistic dissipation operators are used, not hyper-diffusion or numerical diffusion. It was also found that the Zombie Vortex Instability is unlikely to occur in discs around typical T-Tauri stars (Lesur & Latter, 2016).

1.3.6 Vertical Shear Instability

The Vertical Shear Instability (Nelson et al., 2013) can occur in three dimensional protoplanetary discs in which the stability is impacted by thermal and viscous diffusion. This instability occurs when viscous diffusion is much less efficient than thermal diffusion and therefore was originally observed in locally isothermal, inviscid three dimensional protoplanetary disc simulations. A shear in the vertical structure of the disc occurs when there is a radial temperature gradient across the disc. This results in a deviation from cylindrical rotation as the angular velocity profiles now depend on both radius and height. This can be seen from Equation

1.4, in which the vertical distribution of the angular velocity depends on the temperature gradient across the disc. The Vertical Shear Instability acts to transport angular momentum vertically in order to eliminate the vertical shear, in addition to radially across the disc. In order for the Vertical Shear Instability to continue transporting angular momentum radially, the disc must be heated by an external source. This is so that the vertical angular momentum transport is not successful in eliminating the vertical shear and hence does not eliminate the instability driving the radial angular momentum transport (Stoll & Kley, 2014; Barker & Latter, 2015). The Vertical Shear Instability shows similarities to the Goldreich-Schubert-Fricke Instability (Goldreich & Schubert, 1967; Fricke, 1968), which occurs in the radiative zones of differentially rotating stars.

1.3.7 Rossby Wave Instability

The Rossby Wave Instability (RWI) and subsequent formation of vortices is another process that can occur in protoplanetary discs. The RWI in protoplanetary discs was originally investigated in two dimensional discs using linear analysis (Lovelace et al., 1999; Li et al., 2000). It was found that for a disc to be unstable to the RWI there must be a local maxima or minima in the key function,

$$\mathcal{L}(R) = \mathcal{F}S^{\frac{2}{\gamma}}, \tag{1.26}$$

where $S = P/\Sigma^{\gamma}$ is the entropy, $\mathcal{F} = 1/2\xi$ and $\xi = \hat{\mathbf{z}} \cdot (\nabla \times \mathbf{v})/\Sigma$ is the potential vorticity. In addition to a local maxima or minima an instability only occurs if a threshold in the variation of the key function is reached and exceeded. This threshold is dependent on the background of the disc, such as the disc surface density and temperature profiles (Lovelace et al., 1999; Ono et al., 2016). This condition could be fulfilled in protoplanetary discs as a result of material accretion trapping heat in regions of the disc, allowing temperature and entropy to increase in these locations and forming vortices there. Hence, it is likely necessary for the disc to be optically thick for the instability to occur this way (Li et al., 2000). These vortices then facilitate the angular momentum transfer outwards. The linear results were followed up by two dimensional numerical simulations (Li et al., 2001) which found good agreement in the linear growth regime and showed that these vortices are long lived. It has also been shown that the RWI can exist in MHD dominated discs at the edge of dead zones, where the viscosity is not enhanced by the effects of the MRI (Varnière & Tagger, 2006). The RWI has since been investigated in three dimensions both analytically (Lin, 2012b, 2013; Meheut et al., 2012a) and numerically (Meheut

et al., 2010, 2012b; Richard et al., 2013), including three dimensional numerical simulations of specific excitation scenarios, such as the shear in the disc at the edge of MHD dead zones (Lyra & Mac Low, 2012). It was found from these studies that there are minimal differences between the RWI in two and three dimensions, further cementing the idea that the RWI arises as a result of two dimensional phenomena.

Another method of exciting the RWI in a protoplanetary disc arises when we consider a planet perturbing the disc (See Section 1.5). It is well known that a giant planet in a protoplanetary disc will open a gap in the disc material. This gap is a region of comparably low density with respect to the inner and outer disc regions. Hence, there is a steep gradient in the velocity of disc material across this region, as material in the empty gap region deviates from the expected Keplerian rotation (Kanagawa et al., 2015). This can cause vortices to form at the gap edge (de Val-Borro et al., 2006, 2007). It was found that for moderate mass planets $q = M_p/M_* = 10^{-4}$ the RWI could form vortices in the coorbital region of the planet, a result of the shocks created by the planet's inner and outer wakes (Koller et al., 2003; Li et al., 2005). The RWI has also been observed in numerical simulations of discs containing giant gap opening planets with $q = M_p/M_* = 10^{-3}$ or greater, roughly equivalent to one Jupiter mass around a solar type star. This has been observed in two dimensional simulations (de Val-Borro et al., 2006, 2007; Fu et al., 2014a; Hammer et al., 2016; Hallam & Paardekooper, 2017) and three dimensional simulations (Lin, 2012a) where once again similar results were found due to the two dimensional nature of the RWI. The excitation of the RWI by gap opening giant planets is one of the most important factors to the investigations we present in this thesis.

1.3.8 Rayleigh Instability

Another instability that can be excited by the steep gap edges is the Rayleigh Instability. The Rayleigh Instability is a more violent instability than the RWI and can strongly shape the structure of the disc. The Rayleigh Instability occurs when the Rayleigh stable condition is violated. This states that angular momentum in the disc must increase with radius (Chandrasekhar, 1961), therefore the Rayleigh stability criterion is

$$\frac{d}{dR}(R^2\Omega) > 0. \tag{1.27}$$

This is clearly the case in discs containing no planet, or a small mass planet, but if a massive planet opens a gap in the disc this condition may be violated. This is due

to the deviation from Keplerian velocity of material in a deep gap. The excitation of this instability promotes angular momentum transfer radially outwards and thus a reduction in the steepness of the surface density gradient at the gap edge as the velocity deviation of the material is reduced (Kanagawa et al., 2015). This instability is not particularly well studied in protoplanetary disc physics, for the simple reason that protoplanetary discs always satisfy its stability criteria, by definition. For a disc to stabilise against it, it must lead to angular momentum transport, however, this is not a method that we expect to provide a turbulent viscosity. Nevertheless, the onset of this instability can become important, for example in the results presented in Chapter 4 (Hallam & Paardekooper, 2018).

1.4 Planet Formation

As alluded to previously, when a disc is young planets may begin to form via accretion of dust. These planets may impact the disc geometry and eventually form the planetary system surrounding the star, that we can detect observationally. Using our Solar system as a reference, there are two main classifications of planet that may form. These are terrestrial and gas giant planets. In this work we are specifically interested in gas giant planets, for reasons that will become clear. However, here we will discuss the formation of both.

1.4.1 Terrestrial Planets

Terrestrial planets are planets similar those that populate our inner solar system, such as Venus, Mars and the Earth. The definition of terrestrial planets has expanded somewhat with observations of exoplanetary systems, however, generally these are planets with densities that are consistent with a rocky composition. This means that terrestrial planets have a mass $M_p < 10 M_\oplus$, as larger than this we begin to enter the regime of giant planet cores (Raymond et al., 2014).

Terrestrial planets progress through a number of different growth regimes during formation. These are orderly, runaway and oligarchic growth. In the orderly growth regime, planetesimal growth via accretion occurs at a constant rate regardless of the size of the planetesimal. This phase of growth allows the formation of the first small planetesimals. When these planetesimals become large enough such that the velocity dispersion of the material is less than the escape velocity of the planetesimals they undergo an enhanced accretion due to gravitational focussing and enter the next regime, the runaway growth phase. In this phase the larger masses accrete faster

than the smaller masses. Hence, it is dominated by the largest planetesimals that formed during orderly growth, rather than the small planetesimals or formation of new planetesimals. This means that as time goes on, the mass difference between an initially large and small planet will increase (Raymond et al., 2014). As the larger planetesimals accrete faster they will empty their feeding zones and the rate of accretion will slow down.

The next phase of planetary growth occurs at the end of the runaway growth regime, this is oligarchic growth. In this regime a small number of 'oligarchs' dominate the dynamics of the system (Kokubo & Ida, 1998, 2000; Thommes et al., 2003). These oligarchs have similar masses, a result of the self-limiting nature of runaway growth. These oligarchs are now referred to as planetary embryos rather than planetesimals, as they are approaching the mass to form a terrestrial planet. The emptying of the surrounding feeding zones slows the accretion rate, ending the runaway growth regime and growth becomes orderly again. The planetary embryos interact while accreting and remain roughly 10 mutual Hill radii apart through a process known as orbital repulsion. This is a combined effect of the competition between gravitational scattering increasing a planetary embryos orbital radii and eccentricity and friction from smaller planetesimals decreasing it (Kokubo & Ida, 1995). When the friction becomes too weak to oppose the gravitational scattering, due to depletion of smaller planetesimals, the oligarchic growth phase ends and the final regime of terrestrial planet formation begins, the late stage accretion. This regime is dominated by planetary embryo collisions and sets many of the important characteristics of the resultant planetary system, such as the planet's masses and orbital paths, their feeding zones, composition and spin. After this stage the disc will eventually dissipate leaving the newly formed terrestrial planets at their existing locations.

1.4.2 Giant Planets

Giant planets are those similar to the planets that populate our outer solar system, such as Jupiter and Saturn, and are often characterised by a rocky core surrounded by a gaseous envelope. There are two major models for the formation of giant planets, these are the core accretion model (Pollack et al., 1996) and the gravitational instability model (Boss, 1997).

The core accretion model begins similarly to terrestrial planet formation. The giant planet's core begins to form via planetesimal accretion as discussed above, until the later stages of runaway growth in which the planetary embryo has been

formed. This planetary embryo is $< 1 M_{\oplus}$ in mass and can begin to capture gas from within the disc, as the local thermal velocity of the gas is smaller than the embryo's escape velocity, while continuing to accrete solids as usual. Gas accretion can occur very rapidly when the mass of the gaseous envelope becomes approximately equal to the mass of the core. This mechanism can be the driving factor for producing giant planets with masses similar to Jupiter's, with a timescale of approximately 10^6 years (Pollack et al., 1996). At some point gas dissipation or gap formation (Section 1.5.2) will clear the area surrounding the giant planet, halting its envelope accretion and ending the giant planet formation process (Helled et al., 2014).

More recently a new model for the growth of planetary embryos has been proposed, known as the 'pebble accretion' model (Lambrechts & Johansen, 2012). This helps to explain the formation of giant planets over significantly shorter timescales than planetesimal accretion. Rather than planetesimal-planetesimal collisions, the largest planetesimals can very efficiently accrete pebbles of a few tenths of a metre in size. These pebbles are strongly coupled to the gas and feel a deflection to their orbit as they encounter the planetesimal. If the encounter is close enough, the pebbles spiral inwards and are accreted onto the planetesimal. Therefore, if these pebbles dominate a disc's mass planetesimals can grow to a few Earth masses in size very rapidly (Raymond et al., 2014).

There is observational evidence for the prescence of pebbles in protoplanetary discs (Wilner et al., 2005; Rodmann et al., 2006; Pérez et al., 2012), however, the ratio of pebbles to planetesimals is still unknown. Crucially, pebble accretion is a faster method of forming a giant planet core, which is its major advantage over the planetesimal accretion model of core accretion. Despite this, as the pebbles within the disc are coupled to the gas pebble accretion will end when the gas dissipates from the disc, whereas runaway and oligarchic planetesimal accretion can continue. Nevertheless, pebble accretion may better explain why outer terrestrial planets are more massive than inner terrestrial planets and can drive planetesimals to accrete enough mass to become rocky cores for giant planets.

In the gravitational instability model the driving mechanism behind giant planet formation is the fragmentation of the protoplanetary disc, previously discussed in Section 1.3.2. For fragmentation to occur the disc must be particularly massive so that effects arising from self-gravity cause destabilisation and form density perturbations across the disc. These density perturbations will occur for Q of the order 1 (See Equation 1.19, Toomre, 1964). This means that non-axisymmetric disc perturbations will occur as $Q \rightarrow 1$ resulting in shocks and torques. These redistribute angular momentum and mass in order to stabilise the disc. This can result in the

formation of spiral arms in the disc structure which act as a stabilising factor by spreading the disc mass, competing against radiative cooling which acts to destabilise the disc. If these two factors are balanced the disc will retain spiral structure and not become unstable. However, if the effects of cooling outweigh the mass redistribution it can cause these spiral arms to collapse into self-gravitating clumps, which can be the birthplace of giant planets (Helled et al., 2014). This is more likely to occur at large distances from the star, where the cooling timescale is short. However, it is currently not known if realistic discs cool fast enough to fragment or if the fragments can survive long enough to indeed form giant planets (Durisen et al., 2007). Although, it has been shown that clumps may be able to survive up to several thousand years (e.g. Ilee et al., 2017). If the disc fragments can form giant planets, this method has the advantage of forming them at large orbital radii and quickly, while the disc is still young. Figure 1.10 shows a simulated disc that has undergone fragmentation.

Both the core accretion and gravitational instability models for giant planet formation predict the existence of a circumplanetary disc, a disc of solids and gas that resides within the planet's Hill sphere. The Hill sphere is the region in which the planet's gravity dominates over that of external bodies, defined by the Hill radius,

$$R_H \approx a \left(\frac{M_p}{3M_*} \right)^{\frac{1}{3}}, \tag{1.28}$$

where a is the planet's semi-major axis and we assume a circular orbit with eccentricity $e = 0$. As a giant planet core undergoes runaway gas accretion it will fill up its Hill sphere while gravitationally contracting. When the gas accretion rate cannot match the gravitational contraction rate due to the planet's increasing mass, the planet will contract within its Hill sphere. There are two processes subsequent to this that can result in the formation of a circumplanetary disc. Firstly, as the planet accretes material angular momentum is transferred to the planet, increasing the planet's rotation rate. When this rotation rate exceeds a critical rate a rotational instability is excited and the planet sheds its outer layers. This forms a 'spin-out' circumplanetary disc. Secondly, if the gaseous disc is still present while the planet is contracting gas accretion onto the planet can continue. At some point the gas inflowing into the planet's Hill sphere has too much angular momentum to accrete directly onto the now contracted planet so instead flows into orbit around the planet. This forms a circumplanetary 'accretion disc' (Ward & Canup, 2010). Circumplanetary discs have been postulated to explain the existence of the Gallilean satellites of Jupiter (Canup & Ward, 2002).

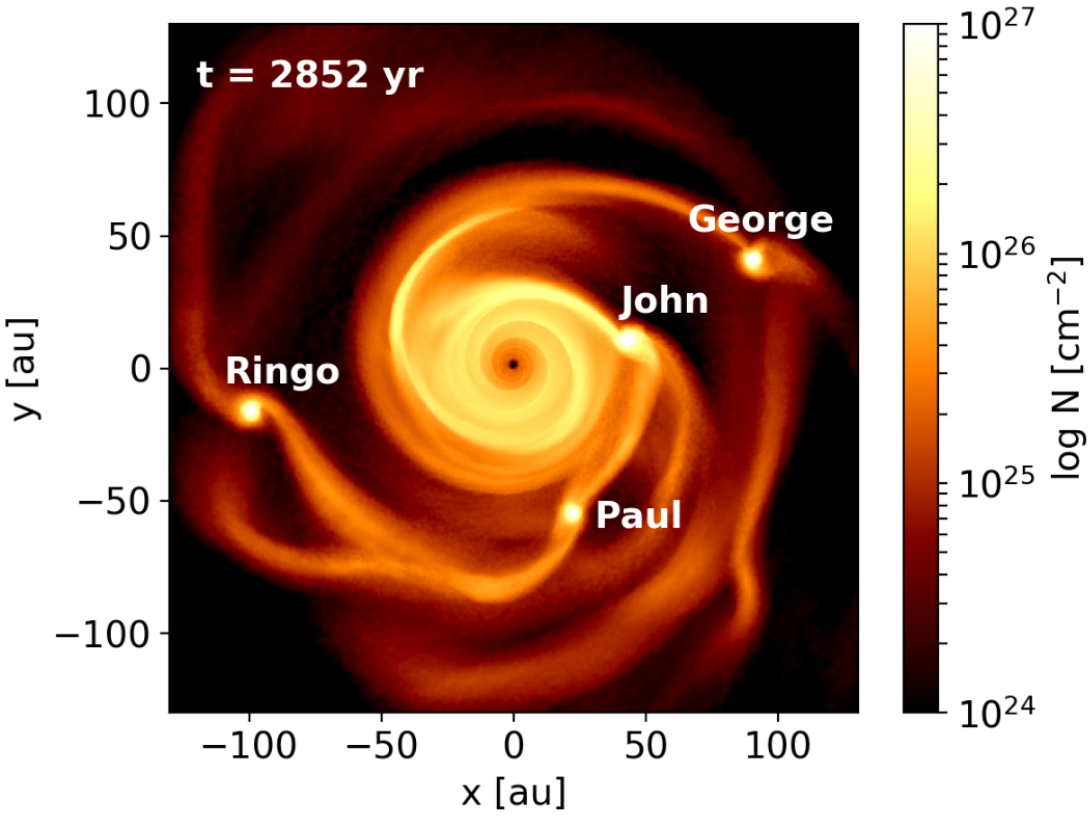

Figure 1.10: A simulated image of a protoplanetary disc in which the gravitational instability has been excited and fragmentation has occurred. The fragments are clearly labelled as John, Paul, George and Ringo. This image was taken at $t = 2852$yrs, showing that these fragments can survive for several thousand years. This figure has been adapted from Ilee et al. (2017).

1.5 Planet-Disc Interactions

Both during their formation and after forming, planets within a protoplanetary disc will gravitationally interact and perturb the surrounding disc material. This can impact the shape of the disc, especially for giant planets. The presence of a planet in a protoplanetary disc modifies the gravitational potential term of the two dimensional Navier-Stokes equation. The two dimensional Navier-Stokes equation is given by

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P - \nabla \cdot T - \Sigma \nabla \Phi. \tag{1.29}$$

We follow similar steps to those taken in Section 1.2.2 in order to determine the one dimensional surface density evolution equation for the evolution of a protoplanetary disc containing a planet. Initially, we again assume axisymmetry. However, once axisymmetry is assumed the potential term due to the planet in Equation 1.29 disappears. Therefore, we must add a one dimensional approximation to the torque density distribution at this step. We use an impulse approximation to obtain a torque density distribution, Λ (Lin & Papaloizou, 1986), to reintroduce the effect of the planet. The impulse approximation determines the angular momentum exchange between the planet and the disc by the interaction of the planet and the individual disc particles shearing past the planet. Assuming axisymmetry we find,

$$\Sigma R\frac{\partial(R\Omega)}{\partial t}+\Sigma v_R\frac{\partial(R^2\Omega)}{\partial R}=\frac{1}{R}\frac{\partial}{\partial R}\left(\nu\Sigma R^3\frac{\partial\Omega}{\partial R}\right)+\Sigma\Lambda. \tag{1.30}$$

This then modifies the simple axisymmetric picture of an unperturbed disc described in Section 1.2 to

$$\frac{\partial\Sigma}{\partial t}=\frac{1}{R}\frac{\partial}{\partial R}\left[3R^{\frac{1}{2}}\frac{\partial}{\partial R}(\nu\Sigma R^{\frac{1}{2}})-\frac{2\Lambda\Sigma R^{\frac{3}{2}}}{(GM_*)^{\frac{1}{2}}}\right]. \tag{1.31}$$

Comparing this with Equation 1.16 we can clearly see the impact of the planet in the second term on the right hand side of Equation 1.31. Of course Equation 1.31 does not include two dimensional features that can become important when considering a planet in a protoplanetary disc.

Planet-disc interactions also drive the migration of the planet, which is important in determining the final orbital positions of planets in planetary systems. The interaction between a giant planet and the disc, its migration and the features it can create are the central topic of this thesis. In this section we shall discuss planet-disc interactions for low and high mass planets.

1.5.1 Low Mass Planets

Low mass planets are those that do not strongly perturb the disc material and so disc structure and geometry are not greatly affected by their presence. This means much of the analysis regarding low mass planet-disc interactions can be done in the linear regime. A planet, low or high mass, will exert a torque on the disc, which in turn will then exert a torque on the planet. This gives rise to planet-disc interactions due to angular momentum exchange. There are two torques that are important to low mass planet-disc interactions (and indeed planet-disc interactions as a whole),

these are the Lindblad torques and the corotation torques. The Lindblad torques arise from spiral density waves that are excited at the inner and outer Lindblad resonance locations (Goldreich & Tremaine, 1979, 1980), which correspond to the locations at which the gas azimuthal velocity (relative to the planet) matches the phase velocity of acoustic waves in the azimuthal direction (Baruteau et al., 2014). The superposition of these density waves can be referred to as the inner and outer wake of the planet. These wakes are visible in Figure 1.13, although it is worth noting that this figure shows a high mass planet, that we shall discuss in the next section. The wakes carry angular momentum from the planet and deposit it in the disc, which exerts a torque on the disc and the planet. The torque on the planet is positive as a result of the inner wake and negative as a result of the outer wake and it is the sum of these 'one sided' torques that produces the overall net Lindblad torque on the planet. In an unperturbed disc, for a small mass planet, the one sided torques are given by (Ward, 1997)

$$\mathcal{T}_{\rm os} = C\left(\frac{M_p}{M_*}\right)^2\left(\frac{1}{h}\right)^3\Sigma_0 R_0^4\Omega_0^2, \tag{1.32}$$

where C is a constant that is positive for the inner disc and negative for the outer disc, M_p is the mass of the planet and the subscript 0 denotes values at the location of the planet. In an unperturbed disc the one sided net torques will mostly cancel, hence the resultant Lindblad torque will be small. However, the gas pressure gradient in the disc acts to shift these locations of the Lindblad resonances slightly radially inwards, meaning that the outer Lindblad torque will usually dominate over the inner Lindblad torque as it is now closer to the planet (Ward, 1997; Papaloizou et al., 2007). In a self-gravitating disc the positions of the Lindblad resonances are shifted closer to the planet and hence the one sided torques on the planet are stronger (Pierens & Huré, 2005; Baruteau & Masset, 2008b).

The second torque acting upon a low mass planet in a disc is the corotation torque. This arises from material on horseshoe orbits. Picturing an axisymmetric unperturbed disc without a planet, we can see that the angular velocity distribution is continuous and for a given radius R, will be close to the Keplerian angular velocity at that point, Ω_K. Now putting a planet into the disc, we separate the disc into three distinct regions, the outer disc in which material is moving slower than the planet, the inner disc in which material is moving faster than the planet and the planet's coorbital region separating the inner and outer disc. Hence, in the reference frame of the planet, the material in the outer disc has negative angular velocity, while the material in the inner disc has positive angular velocity and material in the coorbital

region has, on average, the same angular velocity as the planet, $\Omega_0 = 0$. Material in this coorbital region librates on horseshoe orbits. This occurs as material within the coorbital region, inside or outside of the planet's orbit, is gravitationally perturbed by the presence of the planet. This perturbation causes the material to accelerate if outside the planet's orbit, or decelerate if inside the planet's orbit. Conservation of angular momentum causes the accelerated material to move radially inwards and the decelerated material to move radially outwards such that it crosses the orbit of the planet. Now the material is faster or slower than the planet, so as the material orbits the star the separation between them will increase and the material will be too far away for the planet to perturb its orbit. However, as the angular velocity is now different to the planet's, the material will now catch up to the planet (if faster) or be caught by the planet (if slower) and again be perturbed. This will cause the material to cross the planet's orbit again, but in the opposite way. For example if the material had originally been accelerated to cross the planet's orbit and move radially inwards it will then be decelerated and move radially back outwards. The angular momentum exchange between the material on this horseshoe orbit and the planet exerts a torque on the planet (and of course, a torque on the disc). This process is outlined in Figure 1.11. The net torque from material making horseshoe turns upstream and downstream of the planet is called the corotation torque.

While we have implied that low mass planets can be studied in the linear regime due to their weak perturbation of the disc material, it has in fact been shown that the corotation torque is non-linear for all planet masses (Paardekooper & Papaloizou, 2008). As the total corotation torque on the planet is the sum of each contribution from material on horseshoe orbits, for the net torque to be non-zero there must be an asymmetry across the horseshoe region. The strength of this asymmetry is dependent on the magnitude of the potential vorticity and entropy gradients across the horseshoe region (Ward, 1991; Baruteau & Masset, 2008a; Paardekooper & Papaloizou, 2008).

As there are no waves involved in the horseshoe region, in order to sustain the corotation torque a flow of angular momentum into the horseshoe region is required. Therefore, unlike the Lindblad torque the corotation torque is prone to saturation if viscous effects are not included (Baruteau et al., 2014). If the corotation torque saturates the strength of the torque can be greatly reduced, or dissapear completely. This occurs as material on horseshoe orbits have different libration timescales, which leads to phase mixing and a flattening of the aforementioned vortensity and entropy gradients across this region. Viscosity and heat diffusion act to re-establish these gradients and hence are required for the corotation torque to remain unsaturated

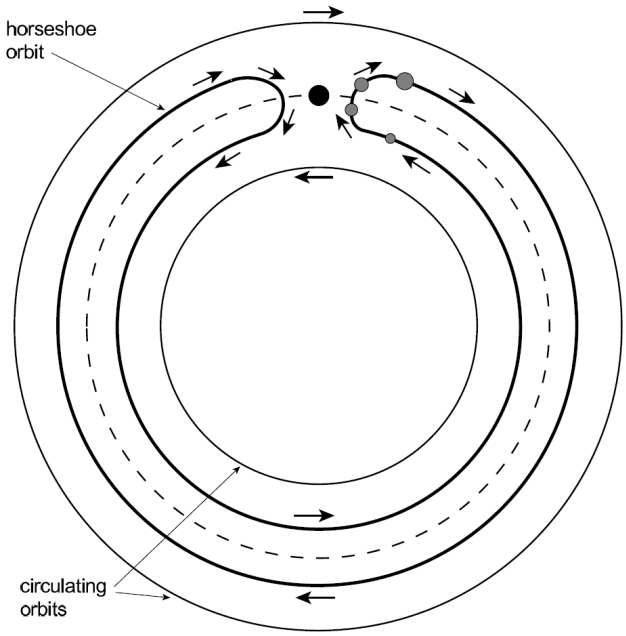

Figure 1.11: Streamlines for the flow of material in the inner disc, outer disc and close to the planet in the coorbital region. This is shown in the planet's orbital frame. Clearly the horseshoe behaviour of material in the coorbital region can be seen. This figure has been adapted from Kley & Nelson (2012).

(Kley & Nelson, 2012). Specifically, in order for the horseshoe torque to remain unsaturated, the viscous diffusion timescale across the horseshoe region must be shorter than the libration timescale (Masset, 2001, 2002), which can be estimated by (Kley & Nelson, 2012)

$$\tau_{\rm lib} = \frac{8\pi R_0}{3\Omega_0 x_s}, \tag{1.33}$$

where $x_s = \mathcal{C} R_0 \sqrt{q/h}$ is the horseshoe region half-width for two dimensional discs, in which $\mathcal{C}$ is a constant of order unity (Masset et al., 2006; Baruteau & Masset, 2008a; Paardekooper & Papaloizou, 2008). Conversely, in extremely viscous discs the corotation torque can be reduced as the viscosity disrupts the horseshoe streamlines. This occurs as the viscous diffusion timescale becomes comparable to the timescale in which the gas performs a horseshoe turn. Similar arguments can be made for the desaturation and reduction of the entropy related corotation torques, this time regarding the thermal relaxation timescale rather than the viscous diffusion timescale (Paardekooper et al., 2010, 2011; Kley & Nelson, 2012). Hence, we can see that the corotation torque can also be impacted by the magnitude of thermal and viscous diffusion within the disc (Masset, 2001; Paardekooper et al., 2010, 2011).

Overall, the Lindblad and corotation torques sum to make the total torque on the planet. This total torque drives the planet's migration, which in the case of small mass planets, is called Type I migration.

1.5.2 High Mass Planets

Unlike low mass planets, high mass planets will strongly perturb the disc structure and modify the surface density distribution of the disc. Hence, we cannot make the assumption that we are in a linear regime, as we can with low mass planets. The single most defining feature of high mass planet-disc interactions is the opening of a gap in the disc material. As a planet's mass increases, so does the strength of the torque exerted on the disc by the planet and as such the amount of angular momentum transferred. This angular momentum is transferred from the planet and deposited in the disc due to exited spiral density waves in the disc material, similarly to the small mass case, the only difference being the amount of angular momentum and strength of the resultant torque exerted. If the angular momentum is deposited close to the planet (Lin & Papaloizou, 1993; Bryden et al., 1999) then that material will gain (or lose) angular momentum and be pushed away from the planet.

There are two methods of determining the planet's transfer of angular momentum to the disc. These are summing over the Lindblad resonances (Goldreich & Tremaine, 1980) or using an impulse approximation (Lin & Papaloizou, 1979). In the impulse approximation the angular momentum exchange is determined from the interaction between the planet and individual particles in the disc shearing past the planet. Both of these approaches provide similar results. The impulse approximation gives us a rate of angular momentum transfer from the planet to the disc of (Lin & Papaloizou, 1986)

$$\frac{dJ_{\rm tid}}{dt} = fq^2\Sigma_0 R_0^4\Omega_0^2\left(\frac{R_0}{\Delta}\right)^3, \tag{1.34}$$

where $f = 0.23$ and $\Delta = |R_0 - R| \approx R_{\rm H}$ is the impact parameter between the unperturbed path of the particle and the planet. The disc material will then attempt to viscously diffuse back into the now empty region surrounding the planet, with a viscous torque of (Lynden-Bell & Pringle, 1974)

$$\frac{dJ_\nu}{dt} = 3\pi\Sigma_0\nu R_0^2\Omega_0. \tag{1.35}$$

Now for the gap to form, the torque due to the tidal effect of the planet must exceed the viscous diffusion torque of the disc material, otherwise the gap will not remain

empty as the material viscously diffuses back into the now empty region surrounding the planet (Lin & Papaloizou, 1979; Goldreich & Tremaine, 1980; Takeuchi et al., 1996; Crida et al., 2006). This is known as the viscous condition for gap opening and is given by (Lin & Papaloizou, 1993)

$$q \gtrsim \frac{40\nu}{R_0^2 \Omega_0}. \tag{1.36}$$

This process is illustrated in Figure 1.12, which shows a two dimensional azimuthal slice of a disc. This figure highlights the competition between the viscous diffusion of the disc material and the tidal torque of the planet, the balance of which acts to set the locations of the gap edges (Takeuchi et al., 1996).

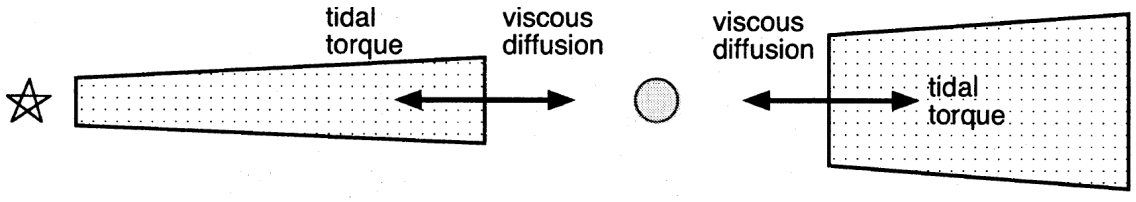

Figure 1.12: A two dimensional, radial slice of a disc containing a high mass planet. This shows the competition between the tidal torque of the planet and the viscous diffusion of the disc material in setting the edges of the gap formed by the planet. This figure is courtesy of Takeuchi et al. (1996).

As previously stated, angular momentum must be deposited close to the planet in order for a gap to open. This is because the disc pressure also opposes gap opening. For this to occur, the planet's Hill sphere must be larger than the scale height of the disc, $R_{\rm H} \gtrsim H$. This is known as the thermal criterion for gap opening (Ward, 1997, as an aside, it is this reason that protoplanetary discs containing gap forming planets can often be safely modelled in two dimensions, as the disc appears two dimensional with respect to the planet). The mass of the planet, the pressure and viscosity of material in the disc all contribute to setting the parameters of the gap formed, such as its depth in surface density and its width. In general, the half-width of a planetary gap is often of the order of twice the planet's Hill radius, Equation 1.28 (Baruteau et al., 2014). As a result of this, a giant planet in a disc will be surrounded by a region of low density, mostly devoid of disc material. An example of this can be seen in Figure 1.13, which shows a simulated result in which a Jupiter mass planet has opened a gap in a disc.

Considering the thermal and viscous criteria for gap formation, Crida et al. (2006) derived a unified gap opening criteria. This shows that a planet will open a gap for

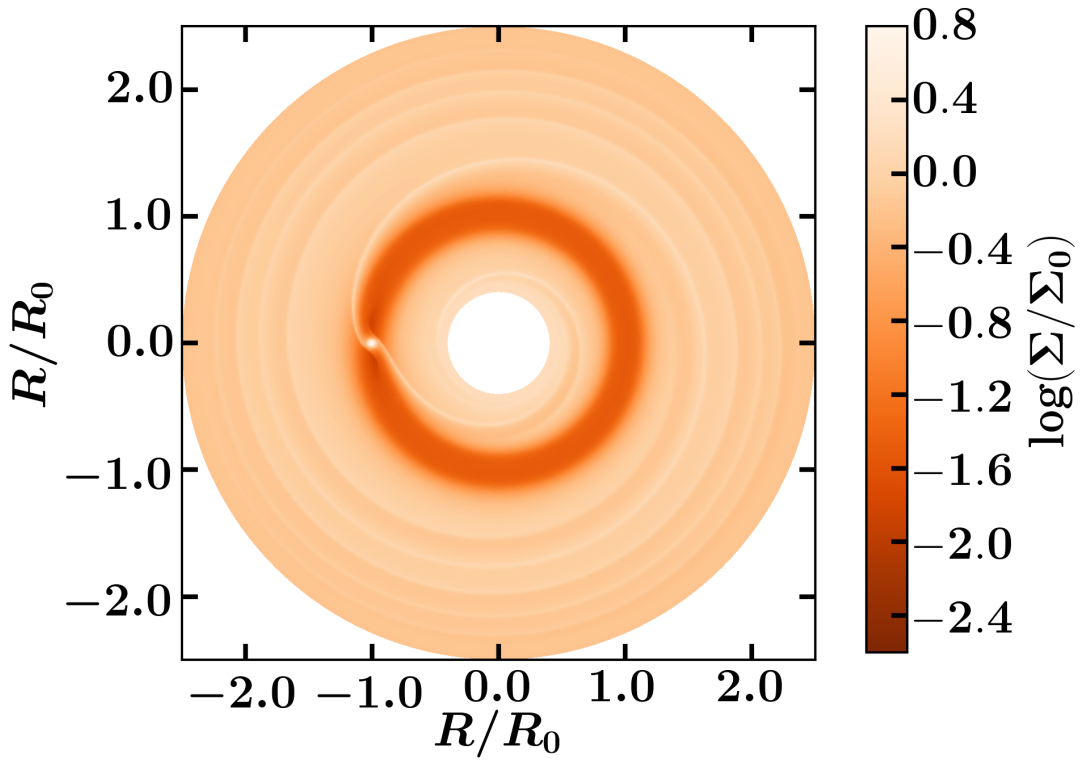

Figure 1.13: A simulated result of a Jupiter mass planet opening a gap in a proto-planetary disc. It can be seen from this simulation that there is a dark region surrounding the planet's orbital path, corresponding to a low density region in the disc. The spiral waves excited by the planet are also clearly visible. This figure is courtesy of Hallam & Paardekooper (2017).

which $\Sigma_{\rm Gap}/\Sigma_0 \leq 0.1$ if the parameter $\mathcal{P} < 1.0$, where

$$\mathcal{P} = \frac{3H}{4R_{\rm H}} + \frac{50}{q\mathcal{R}}. \tag{1.37}$$

Here $q = M_p/M_*$ and $\mathcal{R} = R_0^2\Omega_0^2/\nu$ is the Reynolds number.

Gap formation has important implications for planet formation, as a giant planet that is still accreting can reach a gap opening mass and subsequently empty its accretion zone. Hence, opening a gap can affect the final mass of the planet. These factors play an important role in population synthesis models (Ida & Lin, 2008; Mordasini et al., 2009) and also more recent pebble accretion models (Bitsch et al., 2015; Ali-Dib, 2016). Giant planets are not necessarily the only planets that can open gaps, as it has been shown that small mass planets can also open gaps at greater distance from the planet, where the excited density wave shocks (Goodman & Rafikov, 2001; Rafikov, 2002). Also in low viscosity discs, it may be possible for single planets to open multiple gaps (Zhang et al., 2018; Guzmán et al., 2018).

Finally one of the most important results of giant planets opening gaps in the disc is the impact this has on the migration of the planet. This migration regime is different to that of small mass planets and is called Type II migration.

1.5.3 Migration

It is clear from our understanding of planet-disc interactions, planet formation and observations that many planets are unlikely to have formed in the location we observe them today. Hence, planetary migration is a very important part of planet-disc interactions. It is the main factor that determines where a planet will reside when the disc dissipates and the planet settles into its orbital position. This means that by understanding planet migration, we can much better understand the layout of planetary systems and as a result it has been the subject of much research.

Overall, a planet will migrate based on the magnitude and direction of the total torque it feels. This is the resultant torque when considering the Lindblad and corotation torques acting upon a planet. The total torque on the planet is given by

$$\mathcal{T} = \int_{R_{\rm Min}}^{R_{\rm Max}} \int_{-\pi}^{\pi} \Sigma(R,\phi) \frac{d\Phi_p}{d\phi} d\phi R dR, \tag{1.38}$$

where $R_{\rm min}$ and $R_{\rm max}$ are the radial locations of the inner and outer edges of the disc and Φ_p is the gravitational potential of the planet, given by

$$\Phi_p = -\frac{GM_p}{\sqrt{R^2 + R_0^2 - 2RR_0\cos(\phi - \phi_0)}}, \tag{1.39}$$

in which the subscript 0 denotes the location of the planet. There are three types of planet migration, which each pertain to a different mass regime of planet. These are Type I, Type II and Type III migration.

Type I migration

Type I migration occurs for low mass planets that do not open a gap in the disc material. These migrate due to the Lindblad and corotation torques acting upon them. The net torque due to these drives the direction and rate of the Type I migration. As the Lindblad torque due to the planet's outer wake is stronger than the Lindblad torque due to the planet's inner wake, this migration is most often inwards. The outer Lindbad torque is stronger as the outer Lindblad resonance is closer to the planet than the inner Lindblad resonance, a result of the gas pressure in the disc causing the angular velocity of disc material to be slightly sub-Keplerian.

This moves the locations of the Lindblad resonances slightly inwards (Ward, 1997; Papaloizou et al., 2007). Type I migration is a very rapid form of migration, with a timescale often shorter than the disc lifetime. This problem has been the focus of much research over recent years (Tanaka et al., 2002; Paardekooper & Mellema, 2006b; Paardekooper et al., 2010, 2011; Baruteau et al., 2014).

Type II migration

Type II migration occurs for giant planets that have opened a gap in the disc material. In the classical picture of Type II migration, the presence of the gap means that the region surrounding the planet is empty of material. This means that the planet no longer feels the corotation and Lindblad torques that drove the Type I migration. Now, the planet migrates on the viscous timescale of the disc, $\tau_\nu = R_0^2/\nu$ (Lin & Papaloizou, 1986). The disc material accretes onto the star over this timescale, causing the gap and planet move with it. As we shall see in Section 1.6.4, classical Type II migration is only an idealised case and the rate may vary from the viscous evolution timescale. Despite this, Type II migration is still slower than Type I migration.

The opening of a gap marks the transition between Type I and Type II planet migration, another important result of giant planet-disc interactions. Even though the migration rate has been reduced, it has been found that the predicted Type II migration rate is still shorter than the lifetime of the disc. This suggests that most planets undergoing Type II migration will migrate into and be absorbed by the star (Hasegawa & Ida, 2013).

Type III migration

The final type of planetary migration is Type III migration, which occurs for intermediate (roughly Saturn) mass planets in discs with masses a few times that of the minimum mass solar nebulae (MMSN). This type of migration depends on the fact that the planet is already migrating and can lead to periods of very rapid migration. Hence, this is also known as runaway migration. As usual, this method of migration could potentially be either inwards or outwards, but is most often inwards due to the Lindblad torques acting on the planet.

As a planet migrates inwards it will only undergo one interaction with material at the edge of the horseshoe region in the inner disc. This material will execute a single horseshoe turn, moving to the outer disc before being left behind by the migrating planet. This exerts a total corotation torque on the planet due to the mass flowing

past the planet,

$$\Gamma_{\rm F} = \delta J \frac{dM}{dt} = 2\pi \Sigma_{\rm in} \Omega_0 \frac{dR_0}{dt} R_0^3 x_s, \tag{1.40}$$

where $\Sigma_{\rm in}$ is the surface density of material at the edge of the horseshoe region in the inner disc and the subscript 0 denotes the location of the planet, which varies with time due to the planet migration. This is known as the flow through corotation torque and the positive correlation between migration rate and corotation torque can lead to runaway migration (Masset & Papaloizou, 2003).

During planet migration, material on horseshoe orbits and within the planet's Hill sphere will be carried along by the planet. This trapped material has masses of $m_{\rm hs}$ and $m_{\rm Hill}$ respectively. This material will exert a positive torque on the planet, acting as a drag force opposing the planet's inwards migration. The migration rate of the planet plus this trapped material is then given by

$$(m_p + m_{\rm hs} + m_{\rm Hill})\Omega_0 \frac{dR_0}{dt} R_0 = 2(\Gamma_{\rm F} + \Gamma_L) = (4\pi R_0^2 x_s \Sigma_{\rm in})\Omega_0 \frac{dR_0}{dt} R_0 + 2\Gamma_L, \tag{1.41}$$

where Γ_L is the magnitude of the Lindblad torque on the planet. The quantity $(4\pi R_0^2 x_s \Sigma_{\rm in})$ is the mass of the material that would be contained within the horseshoe region if the disc was unperturbed. Therefore, we can define the parameter $\delta m = 4\pi R_0^2 x_s \Sigma_{\rm in} - m_{\rm hs}$, which is known as the coorbital mass deficit (Masset & Papaloizou, 2003). This parameter increases for more massive planets that begin to open a gap in the disc. Now if we redefine the mass of the planet to consist of both the planet mass and the mass of the material trapped in the planet's Hill sphere, $m_p' = m_p + m_{\rm Hill}$, we can find an expression for the migration rate,

$$\frac{dR_0}{dt} = \frac{2\Gamma_L}{\Omega_0 R_0 (m_p' - \delta m)}. \tag{1.42}$$

From this, we can see that if the coorbital mass deficit is 0 and hence the planet does not open a gap, then the only relevant torques driving the planet's migration are the Lindblad torques. This means that the flow through corotation torque is balanced by the drag from the material trapped in the horseshoe region. Also we can clearly see that the migration rate drastically increases when $\delta m \approx m_p'$. This is the regime for which Type III migration occurs, when the planet is massive enough to partially open a gap in the disc for which the coorbital mass deficit is approximately equal to the mass of the planet. As mentioned previously, this occurs for approximately Saturn mass planets in typical disc models (Kley & Nelson, 2012).

1.5.4 Observing planet-disc interactions

As mentioned previously, telescopes such as ALMA are providing high resolution images of protoplanetary discs in the sub-millimetre wavelength. Present in many of these images, particularly those shown in Figure 1.3, are dark bands in the otherwise bright disc. These imply regions of low density compared to their surroundings. Figure 1.3 shows the ALMA images of two protoplanetary discs, HL Tau and TW Hydrae, a young and old disc respectively, as examples of this. Hence, this shows these features can be present across the lifetime of a disc. Given the information presented in Section 1.5.2, one might conclude that these dark bands are gaps created by high mass planets forming in the disc, however, resolution is not high enough to directly image the planet. Recent work has been done regarding the detectability of circumplanetary discs surrounding potential planets in these discs, but the resolution required would still need to be extremely high (Szulágyi et al., 2017; Zhu et al., 2018). This means that, while some evidence for circumplanetary discs has been detected (Kraus & Ireland, 2012; Sallum et al., 2015), there are currently no unambiguous detections. Hence, we cannot outright attribute this structure to gaps carved by high mass planets. There are a number of other phenomena that could be the cause of this structure, for example a 'clumping' instability that organises dust into narrow rings (Lyra & Kuchner, 2013), dust trapping due to magnetic or Rossby Wave instabilities (Pinilla et al., 2012, see also Section 1.3) or rapid pebble growth at condensation fronts within the disc (Zhang et al., 2015).

This does not mean that it is any more or less likely that these dark bands are not tracers for high mass planets, as there is no conclusive evidence for any of the proposed phenomena that can explain them. Therefore, there has been much work into predicting the characteristics of planets that could explain the observed structures (Dipierro et al., 2015; Bae et al., 2016; Tsukagoshi et al., 2016; Dipierro et al., 2018).

Another important point to consider is that these images of protoplanetary discs show only the dust emission and tell us nothing about the gas within the disc. Hence, the structure we are seeing in these discs may only be present in the dust, as it is easier for planets to form gaps in the dust than in the gas (Paardekooper & Mellema, 2004).

1.6 Problems with Planet Formation

Unfortunately, despite understanding the various steps necessary for planet formation, the problems forming planets in practice are abundant, arising almost immediately when concerning realistic discs, at many length scales.

1.6.1 Radial Drift

Consider the most common picture of the formation of a terrestrial planet or rocky giant planet core. To do this it is necessary to grow dust of size approximately 10^{-6}m into planetesimals of size approximately 10^{3}m before these collide and accrete to become the aforementioned terrestrial or giant planet core. It is at this early stage that we encounter one of the first problems in planet formation theories. While it has previously been stated that material in protoplanetary discs is expected to orbit at the Keplerian angular velocity, this is only strictly true of the dust in the disc. In most protoplanetary discs there is a negative radial gradient in the gas pressure, which causes the gas to orbit at slightly sub-Keplerian velocities. The dust particles do not feel the gas pressure and hence orbit at the expected Keplerian velocity. The problem now arises from the slight difference in the velocities of these two species, at a given radius. As there is a continuous stream of gas and dust orbiting at a given radius in the disc there will always be gas ahead of the dust, even though the gas orbits at a lower velocity. Hence, the dust will feel a headwind from the gas. This acts as a drag force on the dust, reducing its velocity and causing the dust to lose angular momentum. This promotes radial drift inwards towards the star, that is dependent on the size of the dust particle. The rate of this radial drift increases with particle size until particles reach roughly cm to m sizes, at which point the rate begins to decrease with size. The rate of radial drift is also dependent on disc properties, but the maximum drift velocity can be as fast as 100ms^{-1} (Weidenschilling, 1977). Hence, it is very possible that dust particles will be accreted onto the star before they can grow in mass to decouple from the gas and long before they reach the sizes necessary to form planetesimals. This is known as the metre-size barrier to planet formation (Morbidelli & Raymond, 2016).

As a result, it is clear that there must be some method of slowing or halting the inwards radial drift of dust particles in order to allow them to grow to planetesimal size. Such a method can arise from perturbations to the background gradient in the gas pressure. If regions of the disc contain local enhancements in gas pressure this can modify the gas velocity and hence the drag force on the dust particles.

This can result in reduced inwards drift speed or even outwards radial drift of dust (Haghighipour & Boss, 2003a,b). Such local pressure maxima may arise following the excitement of instabilities (see Section 1.3), resulting in vortex formation that traps dust and prevents radial drift.

Another method of solving the radial drift problem is to force dust particles together, such that they grow in size fast enough to overcome the metre-size barrier. An instability that can potentially do this is the streaming instability (Youdin & Goodman, 2005). This instability arises from the interaction between the dust and gas in the disc, unlike the hydrodynamical instabilities discussed in Section 1.3. Consider the standard picture of radial dust drift in a protoplanetary disc outlined above. Additionally, the dust that feels a headwind from the gas will exert a back reaction on the gas, slightly increasing the velocity of the gas. This pushes the velocity of the gas closer to that of the dust, reducing the drag force on the dust and hence its rate of inwards radial drift. This can result in an overdensity of dust at a certain radius, where the strength of this back reaction is increased, further reducing the radial drift rate of the dust. With significant enough back reaction, arising from a large pile up of dust, the gas velocity will be pushed to approximately the Keplerian velocity, the same as the dust velocity at this radius. This will significantly reduce the drag force and hence the radial drift of the dust particles. Dust particles will then drift to this radius faster than they drift away, resulting in a large pile up of dust. Therefore, vortices can form at these locations and the trapping of dust may act as a catalyst for planet formation. However, getting this instability to be excited does pose some problems. For dust particles of centimetre and larger size in the inner disc a super-solar solid/gas surface density ratio is required to excite this instability. This ratio is known to decrease over the disc lifetime (Lambrechts & Johansen, 2014; Morbidelli & Raymond, 2016).

Disc self-gravity can also provide a potential solution to the radial drift problem. In a gravitationally unstable disc, spiral waves will be excited that can act to provide local pressure maxima. The drag force the dust particles feel then directs their radial drift towards the centre of the spiral waves. It has been shown in three dimensional simulations that small particles that would be strongly affected by the gas drag become concentrated at the centre of the spiral waves (Rice et al., 2004; Meru & Bate, 2010; Meru, 2015). This implies that the disc entering a self-gravitating phase may be important in the formation of giant planets, even if they do not form directly by the gravitational instability (Durisen et al., 2007). However, it has been shown that it may only be possible to concentrate dust particles at the centre of spiral waves at large radial distances from the central star (Booth & Clarke, 2016).

1.6.2 Growth of Planetesimals

Even if we can prevent dust particles from rapidly accreting onto the central star, another problem arises in the actual growing of planetesimals. In the standard picture of planet formation, dust particles collide and accrete in order to grow in size. However, when dust grains grow to approximately a millimetre in size, these collisions result in dust particles bouncing off of each other, rather than accreting (Zsom & Dullemond, 2008). Additionally, at far distances from the star (beyond the ice line), this is partially suppressed and dust can grow to approximately 0.1m before bouncing. As mentioned previously these particles will rapidly migrate inwards, resulting in disruptive collisions with dust of different sizes. This is known as the bouncing barrier to planet formation (Morbidelli & Raymond, 2016).

A number of ideas have been presented in an attempt to solve this problem. It has been shown that beyond the ice line in the disc there is a tendency for dust particles to form highly porous aggregates when accreting (Okuzumi et al., 2012; Kataoka et al., 2013). These are resistant to being disrupted by collisions and have a reduced radial drift for a particle of a given size, due to their reduced density. This helps to overcome the bouncing barrier, however, the radial drift of dust very much remains a problem.

It is also known that small dust particles will be collected at pressure maxima, which can arise due to turbulence in the disc. A large collection of dust particles will increase the self-gravity of the clump, which will hold the dust together until it is large enough to form a planetesimal. This, however, shifts the problem to requiring turbulence in the disc, which has been discussed in detail in Sections 1.2 and 1.3. As turbulence is the leading explanation for angular momentum transport in discs, it would be desirable for it to also neatly solve the problems of planet formation. However, it should not be assumed that discs are turbulent by default. Therefore, if the dust particles themselves generate turbulence that drives them to clump and form planetesimals, we can detach this solution from the problems arising from transporting angular momentum across the disc. One such method of generating turbulence is the aforementioned streaming instability (Youdin & Goodman, 2005). Another method follows a similar idea to the streaming instability, however, regards the settling of dust particles to the midplane of the disc. In this case the overdensity of dust at the midplane causes the midplane to rotate faster than the upper layers, a result of the back reaction on the gas, again similar to the streaming instability. The resultant shear in velocity can excite a Kelvin-Helmholtz instability that can act to collect dust particles into self-gravitating clumps (Weidenschilling, 1995; Johansen

et al., 2006). Unfortunately, to collect dust particles in turbulent clumps the particles themselves must be larger than is allowed by the bouncing barrier in the inner disc, approximately 20cm (Johansen et al., 2007).

1.6.3 Formation Timescales

Even assuming the problems with planetesimal formation are overcome, there are still additional barriers to forming giant planets. A prominent problem, affecting the core accretion explanation for giant planets, is the timescale problem. At large radii, the core accretion timescale to form a giant planet is extremely long (Dodson-Robinson et al., 2009; Rafikov, 2011). This can be longer than the lifetime of the disc at radii even as small as 5AU (Goldreich et al., 2004). The pebble accretion model of giant planet formation goes some way to alleviating the timescale problem, as this method can form giant planets orders of magnitude faster than the planetesimal accretion model of core accretion. Hence, using pebble accretion, giant planets can form in the lifetime of the disc, even at large radii (Lambrechts & Johansen, 2012). This does depend on the availability of pebbles in the disc, which is time dependent and needs further investigation. Otherwise, forming a giant planet through gravitational instability is difficult as well. Close to the central star the cooling timescale is often too slow, relative to the orbital timescale, for fragmentation to occur (Rafikov, 2005; Matzner & Levin, 2005). Conversely, at distances very far from the central star it is difficult to collect enough material to form a planet without exceeding the limit for star formation (Kratter et al., 2010).

1.6.4 Migration Timescales

Finally, even if we are able to successfully form planets of terrestrial or giant size, there are problems with planet migration to consider. The foremost of these is the problem with the timescales in which planets undergo their respective type of migration. This is a problem that impacts both Type I and Type II migration and so is common to planets of a range of sizes. As migration is extremely important in setting the final architecture of planetary systems, such as those we observe, this problem has been the subject of much research. Indeed, attempting to address the speed at which Type II migration occurs is the topic of Chapter 4 of this thesis.

Type I migration

As mentioned in Section 1.5.3, Type I migration is very rapid with a migration timescale often shorter than the disc lifetime. For example, a planet on a circular orbit has a Type I migration timescale given by (Baruteau et al., 2014)

$$\tau_{\rm I} = \frac{h^2 M_*}{2q\Sigma_0 R_0^2 \Omega_0}, \tag{1.43}$$

where the subscript 0 denotes the location of the planet. Assuming typical values, $\Sigma_0 = 2000(R_0/1\text{AU})^{-3/2}\text{gcm}^{-2}$, $M_* = M_\odot$, $h = 0.05$, and $R_0 = 1\text{AU}$ then,

$$\tau_{\rm I} \approx \frac{1}{q} \text{ years}. \tag{1.44}$$

Hence, an Earth mass planet would migrate inwards on a timescale $\tau_{\rm I} \approx 3 \times 10^5$ years, which is significantly less than the expected disc lifetimes of $1 - 10$Myr (Baruteau et al., 2014). However, the processes discussed in Section 1.5.1 which prevent the corotation torque from becoming saturated can impact the migration of the planet and has even been shown to potentially reverse the direction of migration (Paardekooper & Mellema, 2006b).

As smaller, terrestrial, planets require less material to accrete than giant planets it is more believable that these planets could have formed further out in the disc, where the formation timescale is longer. Additionally, the abundance of super-Earth planets close to their parent star (Dressing & Charbonneau, 2015; Burke et al., 2015) is consistent with inwards migration of terrestrial planets.

Type II migration

As mentioned in Section 1.5.3, Type II migration is slower than Type I migration. However, this does little to alleviate the timescale problem, as it has been found that the predicted Type II migration rate is still shorter than the lifetime of the disc (Hasegawa & Ida, 2013). Hence, we expect that most planets undergoing Type II migration will migrate into and be absorbed by the star. Nelson et al. (2000) has shown that the Type II migration timescale for a Jupiter mass planet at 5AU is approximately 10^5 years. Observational evidence dictates that gas giant planets are more common at radii $R > 1\text{AU}$ (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016) which would not be the case if the predicted Type II migration timescale was correct. Current models estimate that for gas giant planets to survive migration they must form at $R > 20\text{AU}$ (Coleman & Nelson, 2014). At this radius the core accretion timescale exceeds the lifetime of the disc,

hence giant planets cannot form. This implies that there must be a process that limits the rate of Type II migration (Nelson et al., 2000). Understanding this process could go a long way to improving our understanding of the formation of planetary systems.

One possible method of slowing Type II migration can occur by reducing the viscous accretion rate onto the star. If a planet is massive enough, such that $M_p > 4\pi\Sigma_{\rm out}R_0^2$ the viscous accretion rate can be slowed. Here $\Sigma_{\rm out}$ is the surface density of the outer disc just outside the gap region. In this case the inner disc accretes onto the central star, however, the outer disc is held by the planet, resulting in partial or total depletion of the inner disc and a reduced migration rate for the planet. The Type II migration timescale is then set by a balance between the viscous torque and the planet's inertia (Ivanov et al., 1999; Baruteau et al., 2014) and so is increased to

$$\tau_{\rm II} = \frac{\tau_\nu M_p}{4\pi\Sigma_{\rm out}R_0^2}. \tag{1.45}$$

The above arguments only hold for massive planets that open very deep gaps in the disc. For planets that only partially satisfy the gap opening criterion of Crida et al. (2006), such that $\mathcal{P} \geq 1.5$ (corresponding to $\Sigma_{\rm Gap}/\Sigma_0 \geq 0.2$) the gas in the gap is no longer decoupled from the planet. This means the planet can feel a positive corotation torque which will also slow down Type II migration, in these marginal cases (Baruteau et al., 2014).

While in certain cases it may be possible to reduce the viscous accretion rate of the disc, this may not be the overall factor that sets the rate of Type II migration. The argument for classical Type II migration relies on the absence of material surrounding the planet, a result of the torque exerted on the disc by the massive planet clearing the area around it and forming a gap. This is clearly the case in one dimensional simulations such as those in Lin & Papaloizou (1986), however, extending this to higher dimensions it is found not to be the case (Kanagawa et al., 2015; Hallam & Paardekooper, 2017). It can easily be seen from two or three dimensional simulations that disc material can be present within the gap and can cross the gap on horseshoe orbits, which means that the assumption that there is no corotation torque acting on the planet is not correct. Hence, we can only view classical Type II migration as an idealised case, whereas realistically it is closer to a slowed form of Type I migration. This results in a deviation from the classical Type II migration rate, the viscous evolution timescale of the disc. The migration rate can now be significantly faster or slower than the classical rate, depending on the disc and

planet parameters (Duffell et al., 2014). It was found by Dürmann & Kley (2015) that only for small disc masses $M_d/M_J < 0.2$ is the Type II migration rate slower than the classical rate.

More recently it has been shown that the rate of Type II migration is indeed proportional to the disc viscosity, as was classically believed. However, it is still the net torque on the planet, primarily the inner and outer Lindblad torques, that drives the planet's migration (Dürmann & Kley, 2015; Robert et al., 2018). This does not mean that the net torque on the planet is proportional to the disc viscosity. Then, the behaviour of Type II migration is as follows. A massive planet opens a gap in the disc and reaches equilibrium with the disc material. For low viscosity discs, the amount of material that can cross the gap is negligable. The planet still feels a net torque due to the inner and outer Linblad torques, which drives the planet to migrate (usually inwards). Now, the density distribution of the disc material is no longer in equilibrium and so must adjust, over the viscous timescale. Once it does the net torque on the planet will cause it to continue its migration and the process repeats. Hence, the migration rate is proportional to the viscous timescale, while still being driven by the torques on the planet. As the torques driving migration are in no way related to the viscous evolution timescale of the unperturbed disc, this rate may still be faster than the classical Type II migration. Regardless, the migration rate must be proportional to the disc viscosity, as the amount of material that can cross the gap is negligible. If this was not the case the outer disc would deplete and the net torque driving the planet inwards would disappear (Robert et al., 2018).

Overall, the rate of Type II migration is one of the largest problems facing anyone attempting to explain the structure of observed planetary systems. As such, it has been the subject of much research, including our own presented in Chapter 4. Despite this, there is currently no conclusive result that can solve this problem.

Gap formation and migration

While Type II migration provides some means of slowing down planet migration, it may also be difficult for a planet to enter this regime. As has been previously stated, the opening of a gap in the disc material acts as a transition between the faster Type I migration and the slower (although still too fast) Type II migration. In a simple, non-migrating picture opening a gap is quite straightforward. The planet will accrete disc material until it becomes massive and exerts a strong enough torque on the disc, opening a gap. However, in a realistic disc we cannot make the assumption that the planet is stationary. Until the gap is opened, we would expect the planet to be in

the Type I migration regime, so the planet will be moving rapidly while trying to form a gap. This means that if the planet cannot open a gap fast enough, it will be unable to open a gap and will not enter the Type II migration regime. Hence, if a planet's gap formation timescale is longer than the migration timescale, a gap will not be formed (Malik et al., 2015). Nevertheless, it has been shown that it is possible for migrating planets to open gaps in the disc material (Crida & Bitsch, 2017).

1.7 Numerical Methods

As implied in previous sections, numerical simulations are extremely important in furthering our understanding of protoplanetary discs. Therefore, a large number of numerical codes have been developed over the years in order to solve the equations that govern disc dynamics. As a number of different codes existed it was necessary to perform a consistancy check between them, to ensure their outputs converged (de Val-Borro et al., 2006). This was also partially motivated by the discovery of instabilities within these numerical simulations, to ensure they arose as a result of physical processes rather than numerical error. There are two main types of code that can be used to simulate the evolution of protoplanetary discs, grid based codes and particle based codes. In this section we will give a brief overview of these two types of code. Additionally we will discuss the behaviour of the code we use to perform our simulations, FARGO3D (Benítez-Llambay & Masset, 2016) and our reasons for choosing it.

1.7.1 Grid based codes

Grid based codes use a mesh of grid cells, usually square or rectangular, to cover the computational domain. At each of these cells the relevant quantities are calculated. For example, the velocities and density in the case of a protoplanetary disc. In grid based codes there are generally two different methods of solving the Navier-Stokes equation for the evolution of fluids. These are upwind methods (Ziegler & Yorke, 1997) and shock capturing methods (Paardekooper & Mellema, 2006a). Both of these use finite difference and finite volume methods when solving the relevant evolution equations.

In both of these methods there are a number of factors that must be taken into consideration. Firstly the Courant-Friedrichs-Levy (CFL) condition, which means that information (i.e. particle velocity, bulk motion) can only travel over a single cell

per timestep. This means that in hydrodynamical codes the timestep is limited by the sound speed and velocity. If this condition is not met, causality is violated and this can lead to non-physical effects, as finite difference and finite volume methods only consider neighbouring cells (de Val-Borro et al., 2006).

Another factor grid based codes must consider is that most algorithms for solving hydrodynamical equations are designed only for a one dimensional flow. Hence, most codes use a method known as Strang splitting to extend the one dimensional algorithm to higher dimensions (Strang, 1968). This method comprises of a partial solution of the equations in one direction (say, x) followed by a full update in another (say, y), using the updated values from the original direction. This is then followed by another partial update in the original direction to complete. At the end of each of these updates the CFL condition is calculated. This process is very important, as grid based codes only consider directly adjacent cells as neighbours, however, the flow in a cell will not necessarily be perpendicular to a cell facing. This means that Strang splitting ensures the flow correctly handles diagonally adjacent cells which otherwise would not be impacted by the vertical component of the flow from that cell. However, most codes do not do exactly this and rather perform a full step in each direction, alternating which is taken first (de Val-Borro et al., 2006).

Finally, in a grid based code using a polar grid, such as those that deal with protoplanetary discs, care must be taken when considering the effects of the Coriolis force. Treating the Coriolis force in the simplest manner, as an extra force in the Navier-Stokes equation, leads to incorrect conservation of angular momentum. Hence, we need to use a more conservative treatment (Kley, 1998).

Upwind methods

Upwind codes solve the Navier-Stokes equations using an operator splitting technique. In this manner some operators are solved for using a finite difference method, while others are solved for using a finite volume method. Generally, in an operator splitting scheme each timestep is split into two steps, the source step and the transport step. In the source step, the velocities and source terms from the Navier-Stokes equations, such as the pressure, are updated. In the transport step, these velocities are then used to advect the other quantities. This is usually done using the integral forms of the equations in order to ensure conservation of mass through the cell boundaries. Artificial viscosity is also introduced to damp post shock oscillations and ensure the stability of the code. These codes generally use a staggered mesh, in which scalar quantities such as the density are cell centred, whereas vector

quantities such as the velocity are stored at cell faces (de Val-Borro et al., 2006).

The code we have chosen to use for the simulations presented in this thesis is a grid based, upwind method code, FARGO3D (Benítez-Llambay & Masset, 2016). This code generally follows the setup discussed here, but we will go into greater detail in Section 1.7.3. This code was chosen for a number of reasons, such as its GPU compatibility allowing for moderate resolution simulations to be run for extended periods of time. However, a major factor that makes FARGO3D suited for protoplanetary disc simulations is its implementation of the FARGO algorithm (Masset, 2000). This allows FARGO3D to circumvent the restrictions the CFL condition places upon the timestep by separating a single timestep into hydrodynamic and advection steps. Once solved, the results are then recombined at the end of the timestep. In the advection scheme, both the material in the disc and the cells rotate at the Keplerian angular velocity and so no material can cross a cell boundary. Hence, the CFL condition will never apply here. Therefore, in the hydrodynamic scheme only the perturbation to the background Keplerian angular velocity is considered and this is the only way material can cross cell boundaries. This means that only the perturbations to the Keplerian rotation will impact the calculation of the CFL condition. This is especially important in protoplanetary discs, where the inner disc is rotating considerably faster than the outer, otherwise the rotation of the inner disc would impose a very small timestep. Using the FARGO algorithm allows the timestep to be roughly an order of magnitude larger than it would be otherwise, so it is clear that this procedure will significantly reduce computational time (de Val-Borro et al., 2006).

Shock capturing methods

Shock capturing codes rely on the fact that there is an analytical solution to the Riemann problem (one dimensional shock tube). At the boundaries between the cells in these codes there is a sudden change in the variables within the cell, a shock. Solving the Riemann problem allows transport of quantities between cells, across these shocks. Initially these codes treated variables within a cell as constant across the whole cell, for example each cell would have no gradient in density across it (Godunov, 1959). This was improved upon using parabolic interpolation, assuming all quantities are cell centred, which allowed a gradient of a quantity across a cell (Colella & Woodward, 1984). However, to satisfy Godunov's theorem, slope limiters must be applied to ensure the method of solving the Riemann problem is not greater than first order around the shock. Otherwise this will add additional oscillations

to the simulation. Hence, this method does not require an explicit artificial viscosity, but does automatically add numerical viscosity to handle shocks, due to the aforementioned slope limiters. Additionally, many shock-capturing codes use an approximate solution to the Riemann problem, as the full problem is computationally expensive (de Val-Borro et al., 2006).

1.7.2 Particle based codes

Particle based codes are the other type of code that can be used to model fluids. These codes differ from grid based codes in that they do not attempt to solve the equations of hydrodynamics on a grid, instead they treat the fluid as a collection of small particles and follow the evolution of these. The most prominent method of doing this is known as Smoothed Particle Hydrodynamics (SPH, Gingold & Monaghan, 1977; Lucy, 1977). In SPH, particles are treated as points with their influence spread over a small region of space, the extent of which is given by a smoothing length. The smoothing length is also used to limit the timestep in these simulations, similarly to the CFL condition in the grid based codes (de Val-Borro et al., 2006).

The main advantage of SPH over grid based codes is that it does not require advection in the equations of motion for the system, making the code easier both to understand and to write. SPH codes also naturally enhance resolution in high density areas, as more dense regions directly correlate to the presence of more particles. Unfortunately, when it comes to investigating gap formation in protoplanetary discs this can mean a low resolution in the gap itself, as this is a low density region. However, the biggest problem with SPH codes is the necessity of an artificial viscosity, in order to avoid particles passing through each other (de Val-Borro et al., 2006). This can make it difficult to perform simulations of protoplanetary discs, which are expected to have low viscosity. It is for these reasons that we have chosen to use a grid based, upwind method code for our simulations, FARGO3D, rather than a particle based code.

1.7.3 FARGO3D

FARGO3D (Fast Advection in Rotating Gaseous Objects) is a magnetohydrodynamic code designed to simulate disc evolution in one to three dimensions by solving the hydrodynamic equations of motion. This code is written in C, and uses a C to CUDA translator to allow the code to be run on graphics processing units (GPUs) in addition to central processing units (CPUs). GPUs are desirable for

running simulations as they can tackle computationally expensive problems such as three dimensional disc simulations significantly faster than a CPU. In general, a three dimensional GPU simulation will run roughly as fast as the equivelent simulation in two dimensions on a CPU (Benítez-Llambay & Masset, 2016). In addition FARGO3D features a C to CUDA translator, meaning that no code has to be written in CUDA. This means that in order to modify the code it is not necesarry to learn the CUDA language and makes modifications much simpler, faster and more reliable. This does, however, mean the format of the code follows strict rules that allow this conversion process to occur.

FARGO3D is not the only code capable of running on GPUs, however, this capability does separate it from the vast number of numerical codes available, into a more select pool (Schive et al., 2010; Fung et al., 2014; Bryan et al., 2014; Schneider & Robertson, 2015). In addition to the C to CUDA translator, what sets FARGO3D apart from the existing GPU codes is its small memory-footprint per cell. One problem with the use of a GPU to reduce computational time is that, compared to a standard cluster of CPUs, the GPU has significantly less RAM available. This can be circumvented using CPU to GPU communications, allowing the GPU to make use of the whole RAM available to the CPU, however, this comes at the cost of reduced computational speed when using the GPU/CPU system. FARGO3D does not suffer from these issues, as in the context of running protoplanetary disc simulations computational speed is more important than memory. Thus, FARGO3D is written in such a way that it has the smallest possible number of temporary arrays, meaning that its memory imprint is as small as it can be. This now means that the whole simulation will run on the GPU, circumventing the slow down due to CPU to GPU communications, which now only occur when data is being output. This fact makes FARGO3D a very good choice for running moderate resolution simulations for extended periods of time, which makes it very desirable for two and three dimensional protoplanetary disc simulations.

FARGO3D is an upwind code, similar the code ZEUS (Stone & Norman, 1992). In these codes the velocities are a staggered quantity rather than a cell centred quantity, which means that they are evaluated at the edges of the cells allowing for easier calculation of mass, momentum and specific energy fluxes at the cell edge via upwind methods. This separates FARGO3D from shock capturing codes which use Godunov's method to solve the Riemann problem and determine the fluxes at the boundaries between cells. Unlike shock capturing codes, FARGO3D evolves the pressure, rather than the total energy of the system. This means that the total energy in the system is not conserved to machine accuracy. Considering only the

kinetic (E_K) and internal (E_i) energies,

$$E_{\text{tot}} = E_K(\rho, v) + E_i(p). \tag{1.46}$$

Due to the high Mach number in protoplanetary discs $E_K \gg E_i$. Therefore, if total energy was conserved, errors in the total energy due to time and space resolution could result in large changes to the pressure, p, including potentially a change in sign. Instead evolving the pressure means that errors in pressure arising from the same quantities will have a much smaller impact on the total energy. This is known as the high Mach number problem in hydrodynamics (Ryu et al., 1993; Trac & Pen, 2004).

In the same manner as ZEUS, FARGO3D updates its fields step by step using an operator splitting technique (Stone & Norman, 1992). This separates the evolution equation for a quantity, Q,

$$\frac{\partial Q}{\partial t} + \nabla \cdot (Q\mathbf{v}) = S(Q, \mathbf{v}, t), \tag{1.47}$$

into two partial equations,

$$\frac{\partial Q}{\partial t} = S(Q, \mathbf{v}, t) \tag{1.48}$$

and

$$\frac{\partial Q}{\partial t} + \nabla \cdot (Q\mathbf{v}) = 0. \tag{1.49}$$

Here Equation 1.48 is the source step and Equation 1.49 is the transport step. A full timestep is now comprised of a partial update from the source term and a secondary update by the transport term. The updates due to the source term are split into three sub-steps which consider the impact of the centrifugal force, the pressure gradient, the body forces and the work done by the pressure forces. In sub-step 1 the velocity field is updated due to the pressure gradients and gravitational forces. In sub-step 2 an artificial von Neumann-Richtmyer viscosity and corresponding heating terms are added, to smooth shocks over a small number of cells. In sub-step 3 the work done by the pressure forces is considered. In the transport step the momentum and energy is considered and the orbital advection scheme is introduced.

Like its predecessor, FARGO (Masset, 2000) and as implied by its name, the FARGO3D code includes orbital advection. Unlike FARGO, the orbital advection scheme has been extended to both hydrodynamic and magnetohydrodynamic schemes. The orbital advection method allows the azimuthal velocity of a ring of cells at a given radius and colatitude to be split into two parts, such that,

$$v = v_0 + \delta v, \tag{1.50}$$

where v_0 is the Keplerian velocity and δv is a smaller residual velocity. In FARGO3D, the maximum timestep is set by the Courant-Friedrichs-Levy (CFL) condition,

$$\Delta t = 0.44 \times \min \left[\left(\sum_i \Delta t_i^{-2} \right)^{-\frac{1}{2}} \right], \tag{1.51}$$

which means that information can only travel over a single cell per timestep. Here Δt_i are the processes which limit the timestep, which can be due to the speed of sound waves, fluid motion, artificial viscosity, viscosity and resistivity (Stone & Norman, 1992; Benítez-Llambay & Masset, 2016). Using orbital advection, the gas moving at v_0 is corotating with the cells, so this material will never cross a cell boundary. Hence, the only velocity that sets the maximum CFL timestep is the deviation from the Keplerian velocity, δv. As $v \approx v_0 \gg \delta v$, this means that the timestep can be much larger than if v was considered as a whole. This reduces computational time, so the addition of this orbital advection algorithm is another factor that makes FARGO3D a very good choice for simulating the evolution of protoplanetary discs.

2 Overview of Thesis

In this thesis we have investigated a variety of problems pertaining to the interaction between giant, gap forming planets and the protoplanetary disc from which they have formed. Key concepts to this work have been highlighted in Chapter 1, specifically the evolution of protoplanetary discs (Section 1.2.2) and the interaction between the planet and the disc (Section 1.5). Also very important is the understanding of how planets form in the disc (Section 1.4) and the understanding of certain instabilities (Section 1.3.7, 1.3.8). Each of the following chapters is self-contained in such a way that many of the basics and any specifics required to understand the work presented within the chapter are included in the chapter itself. Here we shall provide a brief summary of the problems we are investigating in each of the subsequent chapters.

In Chapter 3 (Hallam & Paardekooper, 2017) we address a long standing problem in protoplanetary disc simulations, namely the difference between results of one and two dimensional disc simulations. It is well known that gap formation in one dimensional simulations results in significantly deeper gaps than their two dimensional counterparts (Lin & Papaloizou, 1986; Crida et al., 2006; Kanagawa et al., 2015). We find that this is due to a two dimensional instability, the Rossby Wave Instability (RWI, Lovelace et al., 1999, Section 1.3.7), which is unaccounted for in one dimensional simulations. We also find evidence that this means gaps in protoplanetary discs are maintained at marginal stability, with the RWI limiting their depth.

In Chapter 4 (Hallam & Paardekooper, 2018) we propose a potential solution to another well known problem in protoplanetary disc physics. This problem is simply that Type II planet migration is too fast to explain the observed distributions of gas giant planets at $R > 1$AU in exoplanetary systems (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016). Hence, there must be something slowing down their migration (Nelson et al., 2000). We propose that heating of the outer gap edge by radiation from the central star can act to slow down Type II migration. In some extreme cases, we also show that this method could result in reversing the direction of planet migration.

In Chapter 5 (Hallam & Paardekooper, 2019) we investigate a problem that has

become very prevalent in interpreting observations of protoplanetary discs over the recent years. This problem is the degeneracy in planetary explanations for observed structure, based on our lack of knowledge of the parameters that describe these discs (e.g. van Boekel et al., 2017; Long et al., 2018; Huang et al., 2018). Hence, we propose constraints on the parameters of planets and discs, based on the observed structure, in an attempt to limit the degeneracy. We use the presence or absence of vortices in observational results to determine these constraints, as it is well known from numerical simulations that vortices will form when a steep gap edge becomes hydrodynamically unstable.

In Chapter 6 we investigate the effect of the non-isothermal regime, with a thermal relaxation timescale, on gap edge vortex formation. Our thermal relaxation timescale is implemented in a similar manner to that Chapter 4 (Hallam & Paardekooper, 2018). We observe the behaviour of vortices across a range of thermal relaxation timescales and use this to make predictions about the presence or absence of vortices as a tracer for planets in observations of protoplanetary discs. This chapter is linked quite closely to Chapter 5 and our predictions act as a theoretical extension to the conclusions drawn there. We also find interesting gap edge behaviour at sufficiently long cooling timescales that could potentially link to the conclusions presented in Chapter 3 (Hallam & Paardekooper, 2017), that two dimensional planet gaps are maintained at marginal stability.

3 The role of gap edge instabilities in setting the depth of planet gaps in protoplanetary discs

Abstract

It is known that an embedded massive planet will open a gap in a protoplanetary disc via angular momentum exchange with the disc material. The resulting surface density profile of the disc is investigated for one dimensional and two dimensional disc models and, in agreement with previous work, it is found that one dimensional gaps are significantly deeper than their two dimensional counterparts for the same initial conditions. We find, by applying one dimensional torque density distributions to two dimensional discs containing no planet, that the excitement of the Rossby Wave Instability (RWI) and the formation of vortices play a critical role in setting the equilibrium depth of the gap. Being a two dimensional instability, this is absent from one dimensional simulations and does not limit the equilibrium gap depth there. We find similar gap depths between two dimensional gaps formed by torque density distributions, in which the RWI is present, and two dimensional planet gaps, in which no RWI is present. This can be understood if the planet gap is maintained at marginal stability, even when there is no obvious RWI present. Further investigation shows the final equilibrium gap depth is very sensitive to the form of the applied torque density distribution, and using improved one dimensional approximations from three dimensional simulations can go even further to reducing the discrepancy between one and two dimensional models, especially for lower mass planets. This behaviour is found to be consistent across discs with varying parameters.

3.1 Introduction

A planet situated in a protoplanetary disc will excite density waves in the disc material, which transport angular momentum away from the planet. The angular

momentum is then deposited in the disc as these waves dissipate. Therefore, the planet exerts a torque on the surrounding disc material, forming a gap when the torque is strong enough.

For a planet to open a gap it must fulfil two conditions. Firstly, it must be massive enough (Rafikov, 2002; Crida et al., 2006) such that this torque overcomes the viscous torque from the disc material attempting to diffuse back into the low density gap region (e.g. Kley & Nelson, 2012). This is known as the viscous condition and it is the balance of these torques that sets the equilibrium profile of the gap (Lin & Papaloizou, 1979; Goldreich & Tremaine, 1980; Takeuchi et al., 1996; Crida et al., 2006). Secondly, the planet's Hill radius must be larger than the disc scale height, such that the excited density waves deposit angular momentum close to the planet (Lin & Papaloizou, 1993; Bryden et al., 1999). This is known as the thermal criterion for gap opening. It has also been shown that for small viscosities low mass planets can open gaps further away in the disc, where the excited density wave shocks (Goodman & Rafikov, 2001). Gap opening has an important impact on both theoretical and observational studies of protoplanetary discs.

Recent results from the Atacama Large Millimetre Array (ALMA) have provided our first high resolution images of protoplanetary discs and the features present across the course of their lifetime (ALMA Partnership et al., 2015; Andrews et al., 2016). Of particular note is the presence of a number of gaps visible in the disc. We cannot currently say with certainty whether or not these gaps are planetary in origin, due to lack of resolution and understanding of possible gap opening processes. Despite this it is possible that a number of the gaps we see could be the result of planetary interaction and this furthers the desire to understand this phenomenon (Dipierro et al., 2015; Gonzalez et al., 2015; Zhang et al., 2015; Rapson et al., 2015).

In the context of planet formation theory, gap formation plays a vital role in the transition between two regimes of planet migration, Type I and Type II. Type I migration occurs when the planet is not massive enough to open a gap in the disc, hence the disc is almost unperturbed by the presence of a planet. Type II migration occurs when a more massive planet opens a gap in the disc. The migration timescale for the planet will then depend on the viscous evolution timescale of the disc (Lin & Papaloizou, 1986). Type I migration is considerably faster than Type II migration (Ward, 1997). These types of migration are important to theoretical models used to explain planetary system formation as they dictate the final position of a planet and can help to explain the features of the numerous exoplanetary systems discovered. Gap formation is also a very important factor when considering how massive a planet can grow. When an accreting planet becomes massive enough to open a gap in the

disc its accretion rate will fall significantly, as the planet has emptied its accretion zone. Hence, opening a gap can act as a limit on the final mass of a planet. These factors play an important role in population synthesis models (Ida & Lin, 2008; Mordasini et al., 2009) and also the more recent pebble accretion models (Bitsch et al., 2015; Ali-Dib, 2016).

Gap formation as a result of planet-disc interactions is an extensively studied subject area, and has been explored in one, two and three dimensions. The results of one dimensional models show thin deep gaps (Lin & Papaloizou, 1986; Kanagawa et al., 2015), comparable to their higher dimensional analogues. In two dimensions gap formation criteria and shape have been studied (Crida et al., 2006; Duffell & MacFadyen, 2013) and a power law scaling relation for gap depth has been deduced (Fung et al., 2014; Kanagawa et al., 2015). It is well known that the depth and shape of two dimensional gaps from numerical simulations are inconsistent with expected results from one dimensional analytical and numerical models (Crida et al., 2006; Kanagawa et al., 2015). Crida et al. (2006), however, show that including the pressure torque in the balance of torques on the discs returns good agreement between semi-analytical and numerical gap profiles. The differences between two and three dimensional simulations are a less explored area, but nonetheless have received some investigation (Fung & Chiang, 2016) and improved one dimensional models of torque distributions have been determined from three dimensional hydrodynamic simulations (D'Angelo & Lubow, 2010).

The evolution of gaps in two and three dimensional discs can be heavily influenced by the presence of instabilities and the resultant formation of an unstable gap edge. Two relevant instabilities are the Rayleigh Instability and the Rossby Wave Instability (RWI). The Rayleigh Instability occurs due to violation of the Rayleigh stable condition, $d(R^2\Omega)/dR \geq 0$ (Chandrasekhar, 1961), which can be due to the deviation from Keplerian velocity of material within a deep gap formed by a massive planet. If this condition is violated the gap would become unstable and promote angular momentum transfer from the gap edge, lowering the surface density gradient. The effect of this on the aforementioned discrepancy has been investigated by Kanagawa et al. (2015). The RWI occurs due to the shear in velocity of the material at the edge of the gap. This can form vortices when the velocity shear is significant, which corresponds to a gap edge with a steep surface density gradient (Lovelace et al., 1999; Li et al., 2000).

We focus our efforts on investigating the discrepancy between the results of one dimensional and two dimensional simulations, specifically the significant difference in gap depth. Here we explore a new approach to explaining this discrepancy, we

remove the gap forming planet and instead apply a gap forming one dimensional torque density distribution radially across the two dimensional disc. This now becomes the mechanism for gap formation, and is the same distribution that forms the gap in one dimensional simulations. We proceed to investigate the resultant gap depth for a number of different forms of this one dimensional torque density distribution. We also extend our investigation to both lower viscosity and higher aspect ratio discs.

This chapter is arranged as follows. In Section 3.2 we derive the relevant equations solved to simulate disc evolution. In Section 3.3 we discuss the code used and the numerical setup of our simulations. In Section 3.4 we present the results of our one and two dimensional simulations. In Section 3.5 we present the results of varying the disc parameters and compare these to our previous results. In Section 3.6 we discuss our findings in the context of prior work, while justifying assumptions made and postulating any impact they may have. Finally, in Section 3.7 we present our conclusions.

3.2 Basic Equations

3.2.1 Two dimensional protoplanetary disc

The continuity equation for the evolution of a protoplanetary discs surface density, Σ, is

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0, \tag{3.1}$$

where $\mathbf{v}$ is the velocity field. We simulate the evolution of a protoplanetary disc's surface density due to the presence of a planet by solving the two dimensional Navier-Stokes equation for the motion of the disc-planet system,

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P - \nabla \cdot \mathbf{T} - \Sigma \nabla \Phi, \tag{3.2}$$

where $\mathbf{T}$ is the Newtonian viscous stress tensor, P is the pressure and Φ is the gravitational potential of the planet and star system. We assume the disc is locally isothermal, with equation of state $P = c_s^2 \Sigma$, where $c_s(R)$ is the sound speed at a radius R. We impose a constant aspect ratio $h = H/R = 0.05$ by selecting a profile for the sound speed, where H is the disc scale height. A cylindrical coordinate system is used, such that $\mathbf{v} = (v_R, R\Omega)$ where $v_R(R)$ and $\Omega(R)$ are the radial and angular velocities respectively, at a given radius. Now splitting Equation 3.2 into

its component forms we find,

$$\frac{\partial v_R}{\partial t} + v_R\frac{\partial v_R}{\partial R} + \Omega\frac{\partial v_R}{\partial \phi} - R\Omega^2 = -\frac{1}{\Sigma}\frac{\partial P}{\partial R} - \frac{2}{\Sigma R}\frac{\partial}{\partial R}\left(\nu\Sigma R\frac{\partial v_R}{\partial R}\right) - \frac{1}{\Sigma R}\frac{\partial}{\partial \phi}\left[\nu\Sigma\left(R\frac{\partial \Omega}{\partial R} + \frac{1}{R}\frac{\partial v_R}{\partial \phi}\right)\right] - \frac{\partial \Phi}{\partial R} \quad (3.3)$$

and

$$\frac{\partial (R\Omega)}{\partial t} + v_R R\frac{\partial \Omega}{\partial R} + \Omega\frac{\partial (R\Omega)}{\partial \phi} + 2v_R\Omega = -\frac{1}{R\Sigma}\frac{\partial P}{\partial \phi} - \frac{2}{\Sigma R}\frac{\partial}{\partial \phi}\left(\nu\Sigma\frac{\partial \Omega}{\partial \phi}\right) + \frac{1}{\Sigma R^2}\frac{\partial}{\partial R}\left[\nu\Sigma R^2\left(R\frac{\partial \Omega}{\partial R} + \frac{1}{R}\frac{\partial v_R}{\partial \phi}\right)\right] - \frac{1}{R}\frac{\partial \Phi}{\partial \phi}, \quad (3.4)$$

where ν is the kinematic viscosity of the disc.

3.2.2 One dimensional protoplanetary disc

We reduce the two dimensional equations to standard one dimensional equations for disc evolution outlined in Pringle (1981), modified to account for the presence of a planet. We begin by averaging azimuthally Equation 3.4 so that only the radial dimension remains. This removes the potential term due to the planet, so we reintroduce an approximation to this as a torque density distribution, Λ,

$$\Sigma R\frac{\partial (R\Omega)}{\partial t} + \Sigma v_R\frac{\partial (R^2\Omega)}{\partial R} = \frac{1}{R}\frac{\partial}{\partial R}\left(\nu\Sigma R^3\frac{\partial \Omega}{\partial R}\right) + \Sigma\Lambda. \quad (3.5)$$

Assuming Keplerian rotation, $\Omega = (GM/R^3)^{1/2}$, where M is the stellar mass, Equation 3.5 can be written as,

$$\Sigma v_R R = -3R^{\frac{1}{2}}\frac{\partial}{\partial R}(\nu\Sigma R^{\frac{1}{2}}) + \frac{2\Sigma\Lambda R^{\frac{3}{2}}}{(GM)^{\frac{1}{2}}}. \quad (3.6)$$

Using the azimuthally averaged form of Equation 3.1,

$$R\frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial R}(\Sigma v_R R) = 0, \quad (3.7)$$

we eliminate the $\Sigma v_R R$ term and find,

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{R}\frac{\partial}{\partial R}\left[3R^{\frac{1}{2}}\frac{\partial}{\partial R}\left(\nu\Sigma R^{\frac{1}{2}}\right) - \frac{2\Lambda\Sigma R^{\frac{3}{2}}}{(GM)^{\frac{1}{2}}}\right]. \tag{3.8}$$

We then solve this form of the continuity equation using standard finite difference techniques (Richtmyer & Morton, 1967).

3.2.3 Torque density distributions

We investigate three major forms of the torque density distribution, Λ. These are an impulse approximation (Equation 3.9), an improved model of the one dimensional torque density distribution from D'Angelo & Lubow (2010) (Equation 3.11) and an azimuthally averaged torque density distribution calculated from the planet itself in two dimensional simulations (Equation 3.13).

The impulse approximation for Λ is in the manner of Lin & Papaloizou (1986),

$$\Lambda = \text{sign}(R - R_0)\frac{fq^2GM}{2R}\left(\frac{R}{|\Delta_0|}\right)^4, \tag{3.9}$$

where $q = M_p/M$ is the ratio of planet mass to stellar mass and R_0 is the location of the planet. We take the dimensionless constant $f = 0.23$, and Δ_0 to be the greater of the disk scale height H or $|R - R_0|$.

We split the disk into three sections, defined as inner disk, outer disk and inside gap. Discontinuities are removed by ensuring equality at the boundaries of these regions. To this end the $\text{sign}(R - R_0)$ term is approximated to

$$\text{sign}(R - R_0) = \begin{cases} -1, & \text{if } R - R_0 < -H \\ \frac{R-R_0}{H}, & \text{if } |R - R_0| \leq H \\ 1, & \text{if } R - R_0 > H \end{cases} \tag{3.10}$$

providing a continuous form for Λ. In Equation 3.10 and the definition of Δ_0 we replace H with the Hill Radius, R_{Hill}, for planet masses of which $R_{\text{Hill}} > H$.

The improved model of the one dimensional torque density distribution from D'Angelo & Lubow (2010) is of the form

$$\Lambda = \mathcal{F}(x, \beta, \zeta)\,\Omega^2 R_0^2 q^2 \left(\frac{R_0}{H}\right)^4, \tag{3.11}$$

where $x = (R - R_0)/H$, $\mathcal{F}$ is a dimensionless function and $\beta = 0.5$ and $\zeta = 1.0$ are the surface density and temperature radial gradients respectively. $\mathcal{F}$ is determined

by D'Angelo & Lubow (2010) to be

$$\mathcal{F}(x,\beta,\zeta) = \left(p_1 e^{\left(-\frac{(x+p_2)^2}{p_3^2}\right)} + p_4 e^{\left(-\frac{(x-p_5)^2}{p_6^2}\right)} \right) \tanh(p_7 - p_8 x). \tag{3.12}$$

$p_1 - p_7$ are constant for a given β, ζ and can be found in Table 3.1 of this chapter and table 1 of D'Angelo & Lubow (2010). The function $\mathcal{F}$ is found by fitting the results of three dimensional simulations.

Table 3.1: Values of the parameter p in Equation 3.12.

p_n	Value
p_1	0.0297597
p_2	1.09770
p_3	0.938567
p_4	0.0421186
p_5	0.902328
p_6	1.03579
p_7	0.0981183
p_8	4.68108

We then consider the torque density distributions presented in the upper panel of figure 15 of D'Angelo & Lubow (2010). These distributions are determined from three dimensional discs in which the tidal torques are stronger than the viscous torques, such that a gap is formed. In this regime Equation 3.11 requires some modification to match the distributions shown in D'Angelo & Lubow (2010). From their figure 15 we can see a non-linear progression in the magnitude of the local maxima and minima of the torque density distributions, despite the linear mass progression. To account for this deviation from linearity we introduce a scaling factor such that our applied torque follows the trend present in their results. This factor is a cubic polynomial in planet mass and is applied in the region immediately before transition from using the disc aspect ratio to the Hill radius of the planet, smoothing the transition between the low mass region (with no scaling factor) and the high mass region (in which a scaling factor of 1/2 is applied as discussed in D'Angelo & Lubow, 2010).

The last torque density distribution we investigate is calculated from the torque applied to the disc from the planet in the two dimensional planet simulations. We average this torque azimuthally to form a one dimensional torque density distribution. The torque density distribution is given by

$$\Lambda = \frac{\int_{-\pi}^{\pi} \Sigma(R,\phi)\frac{\partial \Phi_p}{\partial \phi} d\phi}{2\pi\Sigma_{Av}(R)}, \tag{3.13}$$

where $\Sigma_{Av}(R)$ is the average surface density at a given radius and the gravitational potential of the planet Φ_p is given by

$$\Phi_p = -\frac{GM_p}{\sqrt{R^2 + R_0^2 - 2RR_0\cos(\phi - \phi_0) + s^2}}, \tag{3.14}$$

where $s = 0.6H_0$ is the smoothing length, H_0 is the scale height at the location of the planet and ϕ_0 is the azimuthal position of the planet. This provides the average torque the disc feels at a given radius due to the presence of the planet. A comparison of these two distributions can be seen in Figure 3.1 for a $q = 10^{-3}$ planet.

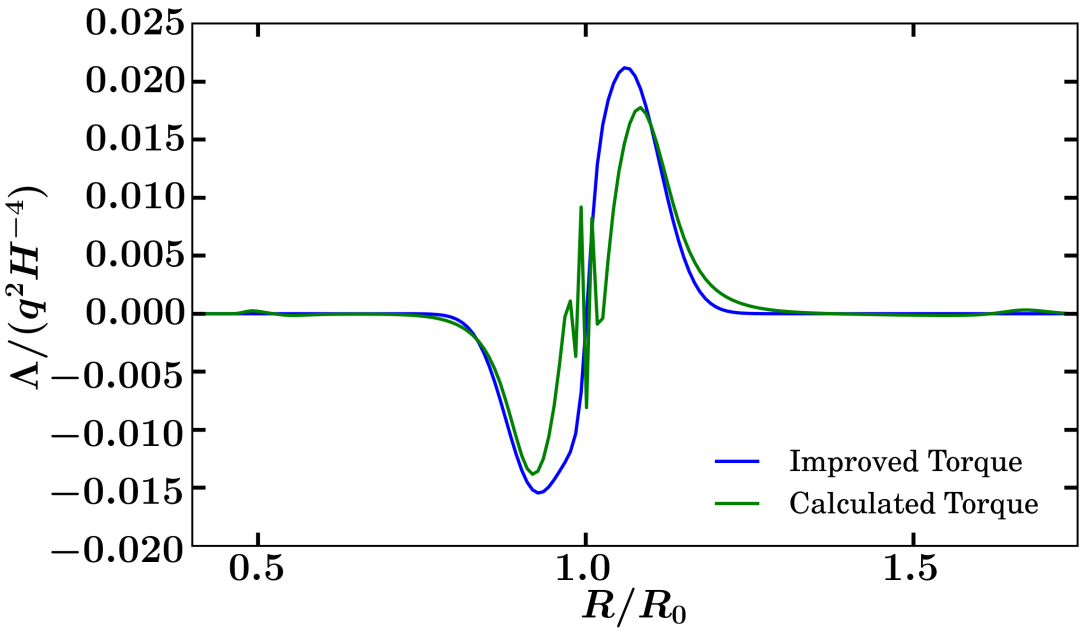

Figure 3.1: Torque density distributions from Equations 3.11 (Improved Torque) and 3.13 (Calculated Torque) in Section 3.2.3 for a $q = 10^{-3}$ planet acting on the disc.

3.3 Numerical Setup

3.3.1 One dimensional protoplanetary disc

One dimensional simulations run over a radial domain of $0.4 \leq R/R_0 \leq 2.5$ using 2100 radial elements, providing a high resolution of $\Delta R_{\text{Cell}} = 10^{-3}$ necessary for re-

solving the extremely steep gap edges, while maintaining computational efficiency. This is possibly an unnecessarily large domain for a one dimensional simulation, however, it allows for more accurate and direct comparisons with our two dimensional simulations discussed in Section 3.3.2. The initial surface density distribution was such that $\Sigma_{\rm int} \propto R^{-1/2}$ and a constant viscosity $\nu = 10^{-5}$ was used. This corresponds to a dimensionless viscosity $\alpha = 4 \times 10^{-3}$ at $R = R_0$, using the alpha prescription $\nu = \alpha c_s H$ of Shakura & Sunyaev (1973). We impose boundary conditions such that the surface density at the edge of the simulation is constant in time and hence feels no perturbation due to gap evolution. Simulations are run until equilibrium is attained.

3.3.2 Two dimensional protoplanetary disc

Two dimensional simulations were run using the FARGO3D code. This is a magnetohydrodynamic code developed with specific emphasis on simulating disc evolution for one dimensional to three dimensional models. The hydrodynamics equations are solved using a finite differences method. Importantly, FARGO3D uses a C to CUDA translator to allow simulations to run on GPUs. GPUs have limited memory, however, decrease computational time significantly, making them a good choice for running moderate resolution simulations for extended periods of time. This makes FARGO3D a very good choice for investigating gap forming planets, as these are often computationally expensive and need to be run until equilibrium is attained. For further detail see Benítez-Llambay & Masset (2016).

The two dimensional simulation uses, wherever possible, the same parameters as the one dimensional simulation to draw accurate comparisons between them. The two dimensional simulation extends from $0.4 \leq R/R_0 \leq 2.5$ radially and $-\pi \leq \phi \leq \pi$ azimuthally with 256 by 768 cells respectively, with an initial surface density distribution of $\Sigma_{\rm int} \propto R^{-1/2}$. This corresponds to a radial resolution of $\Delta R_{\rm Cell} = 8.203 \times 10^{-3}$. The radial range was chosen to allow density waves excited by the embedded planet to propagate and potentially impact gap evolution, while the azimuthal range provides a view of the entire disk. This resolution was found to give consistent results with higher resolution runs, whilst still providing computational efficiency. Reflecting boundary conditions were used with parabolic damping wave killing zones in the regions $0.4 \leq R/R_0 \leq 0.505$ and $2.29 \leq R/R_0 \leq 2.5$, as per de Val-Borro et al. (2006). The combination of this closed boundary and wave killing zones acts as a good approximation to setting the boundaries at infinity, as excited density waves are significantly damped before reaching the boundaries and no wave

reflections are observed. The planet is held on a fixed circular orbit ($e = 0$) with no migration and no disc self-gravity. We use this setup as the basis for simulating disc evolution in both planet and applied torque density distribution simulations.

We now consider the further modifications to this setup for our new approach to investigating the prior observed discrepancy in gap depth between one and two dimensional simulations. We apply a gap forming one dimensional axisymmetric torque density distribution radially across a two dimensional disc devoid of embedded planet. As a result the formation of the gap will be due to a torque applied to every cell, with magnitude dependent on the radial position of the cell. We attempt to closely mimic the torque due to the presence of the planet and measure resultant gap depth. To this end we use the parameters of the simulation as described here, with the addition that random noise of magnitude 10^{-5} was applied to the disc surface density, to remove axisymmetry. For the simulations containing planets the noise would be swiftly damped. We investigate the formation of a gap due to the one dimensional torque density distributions listed in Section 3.2.3.

To calculate gap depth from both planet and torque density distribution simulations two dimensional results were averaged azimuthally to form a one dimensional profile and the gap depth was taken to be the minimum of this. A small region of 10 cells either side of the planet's position was removed when taking the azimuthal average so the presence of the planet does not artificially decrease gap depth, similarly to Fung et al. (2014). See Figure 3.2 and Section 3.4.2. Gap depth was always calculated at equilibrium, defined after 1000 orbits had elapsed for a $\nu = 10^{-5}$ simulation, and always calculated from the average of the gap depth per orbit over 100 orbits. This eliminates any short term variance in equilibrium gap depth due to gap turbulence. Figure 3.3 shows that beyond 1000 orbits gap depth is, on average, approximately constant and so provides good reasoning for this definition of equilibrium.

3.4 Results

3.4.1 One dimensional results

The equilibrium surface density profiles for a number of sample runs of the one dimensional simulation can be seen in Figure 3.4. This shows the variation in shape of the gap profile with planet mass, namely the wider and deeper gaps are resultant of higher mass planets. The $q = 10^{-3}$ planet case shows how sensitive the gap depth is to planet mass, forming a gap of depth $\Sigma/\Sigma_0 \approx 10^{-24}$, considerably deeper than

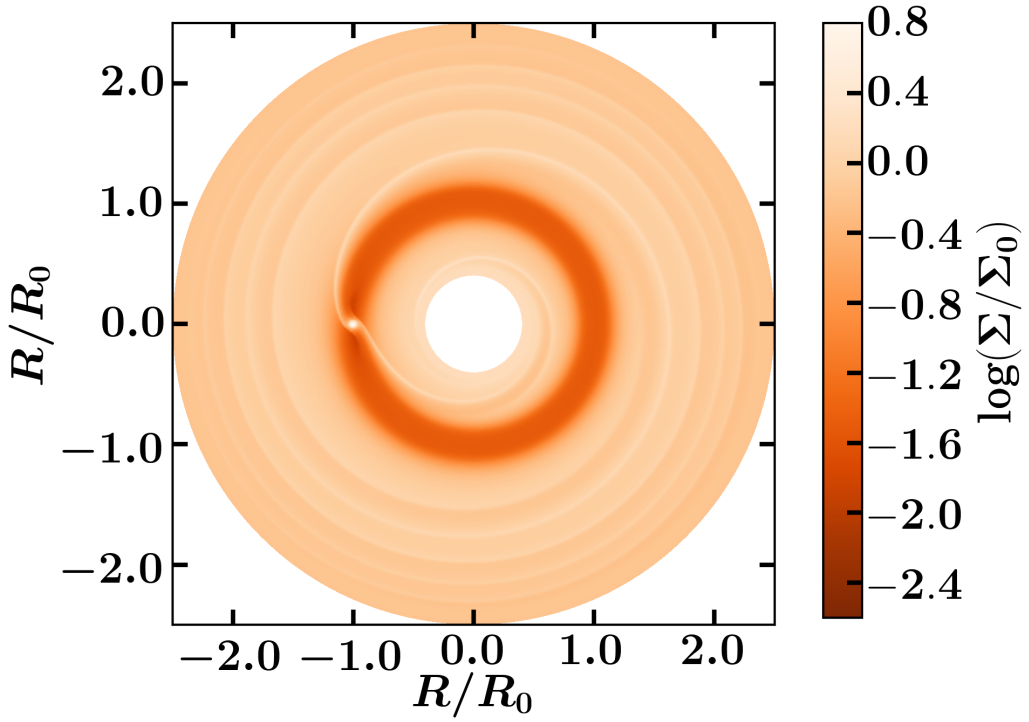

Figure 3.2: A two dimensional simulation using FARGO3D after a $q = 10^{-3}$ planet reaches equilibrium at 1000 orbits in a $\nu = 10^{-5}$, $h = 0.05$ disc.

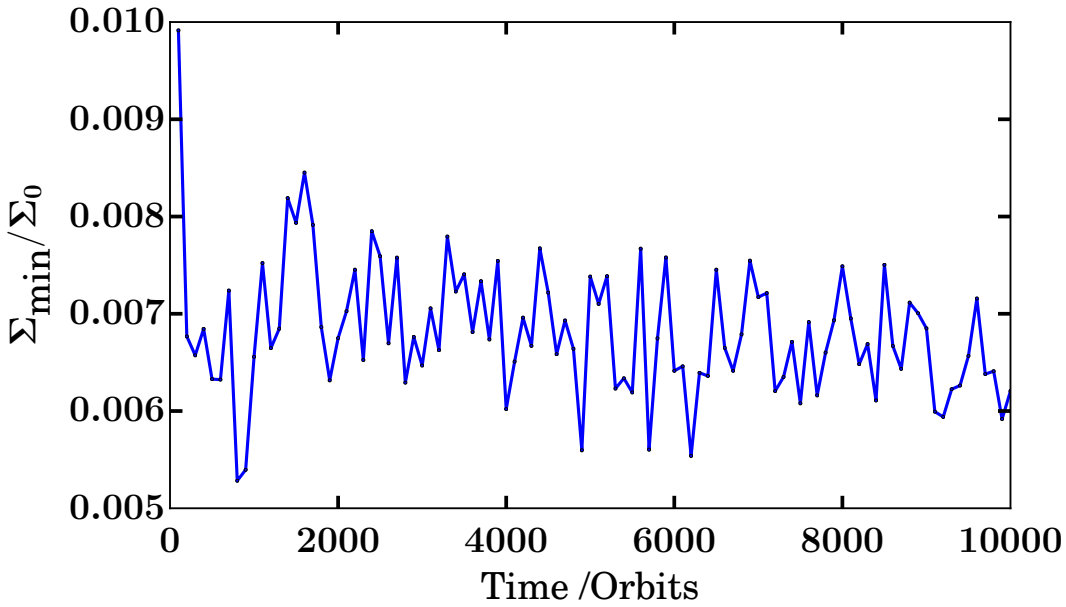

Figure 3.3: Gap depths from a sample two dimensional simulation of a turbulent gap profile evolving beyond the defined point of equilibrium at 1000 orbits. The gap was formed using the torque density distribution in Equation 3.11, discussed in greater detail in Section 3.2.3. If the gaps had been formed by a planet the result would be similar, but less noisy. Gap depth calculation is discussed in Section 3.3.2.

the other profiles shown in Figure 3.4. It was also found that higher mass planets approach equilibrium faster.

For confirmation of accurate definition of equilibrium, Equation 3.8 was solved analytically for $\partial\Sigma/\partial t = 0$. These results can be seen in Figure 3.5, showing good agreement over a large range of planet masses. This prompts the conclusion that the gap depths from the one dimensional simulations are at an acceptable degree of accuracy.

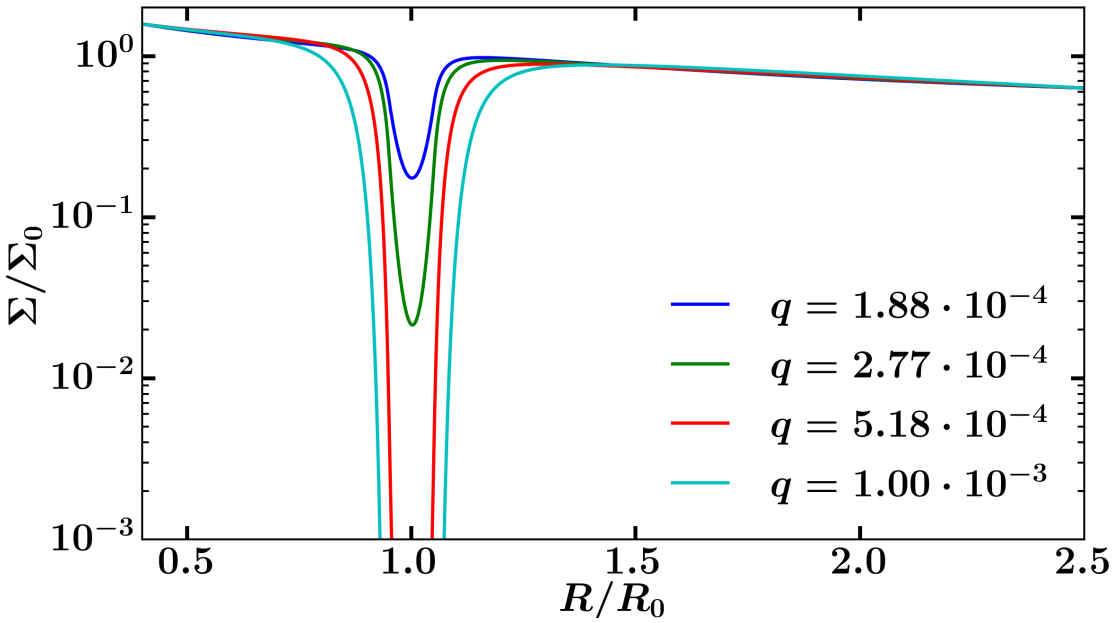

Figure 3.4: One dimensional radial surface density profiles for gaps formed by planets of $q \approx 10^{-4} - 10^{-3}$ after reaching equilibrium. Equation 3.8 was solved numerically to determine these profiles. Disc parameters are discussed in Section 3.3.1.

3.4.2 Two dimensional planet results

Gap depths in two dimensions show the same expected mass-depth trend as they do in one dimension. The most interesting result of the two dimensional simulations arises from the comparison between one and two dimensional results. A direct comparison of one and two dimensional gap profiles for the same planet and disc can be seen in Figure 3.6. The gap profiles are distinctly different as the one dimensional case is thinner and significantly deeper, with $\Sigma/\Sigma_0 \approx 10^{-24}$. This discrepancy is also illustrated in Figure 3.7, which shows us it is present for any mass planet that can open a significant gap in the disc.

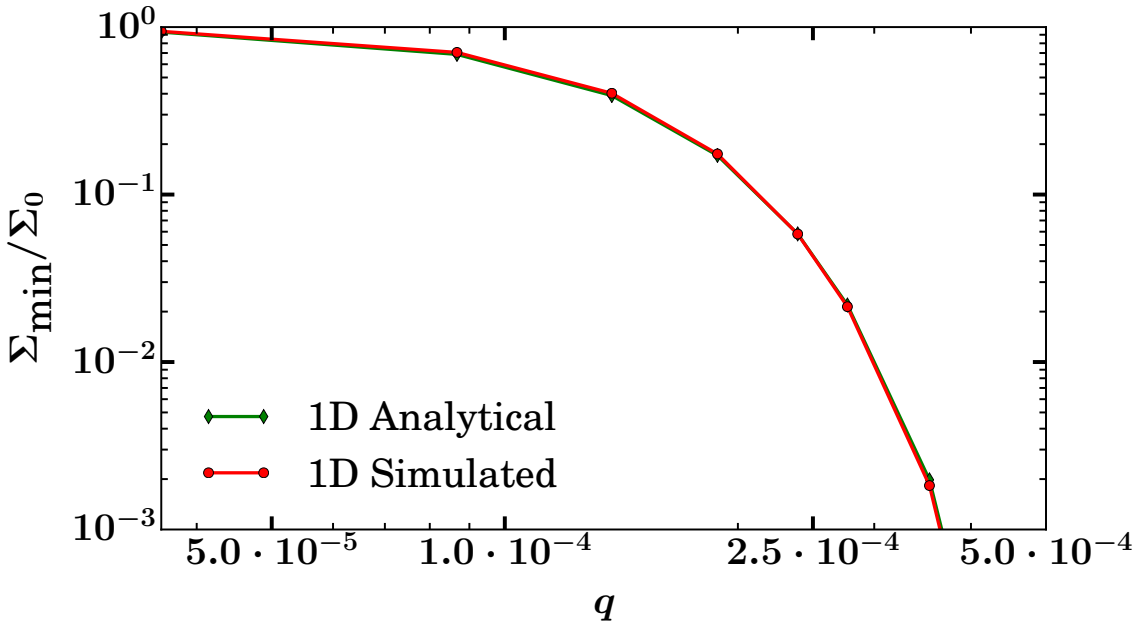

Figure 3.5: One dimensional gap depths from Equation 3.8 for a number of planet masses. Equation 3.8 was solved analytically and numerically via one dimensional simulation. The close agreement between these two curves is an indicator that the numerical simulation has reached equilibrium.

We check the consistency of our two dimensional gap depth results against the gap depth scaling law of Fung et al. (2014), valid for $10^{-4} \leq q \leq 5 \times 10^{-3}$,

$$\frac{\Sigma_{\text{Gap}}}{\Sigma_0} = 0.14 \left(\frac{q}{10^{-3}}\right)^{-2.16} \left(\frac{\alpha}{10^{-2}}\right)^{1.41} \left(\frac{h}{0.05}\right)^{6.61}, \tag{3.15}$$

where Σ_{Gap} is the gap depth. This scaling was calculated from two dimensional disc-planet models. From Figure 3.7 we can see that, for masses of roughly $q \geq 3 \times 10^{-4}$, the results of our two dimensional simulations follow closely the scaling law.

3.4.3 Two dimensional impulse approximation

We now apply the impulse approximation, described by Equation 3.9, directly from the one dimensional simulations radially across a two dimensional disc devoid of any planet. We do this for a range of masses comparable to both prior simulations. The results can be seen in Figure 3.8, which compares the results of this simulation to the prior two simulations. A sample simulated output can be seen in Figures 3.9 and 3.10. Figure 3.9 is a parallel to Figure 3.2, however, using the one dimensional torque density distribution rather than the gap forming planet. Particularly noticeable is the presence of a turbulent gap edge, absent in Figure 3.2 and unaccounted for in

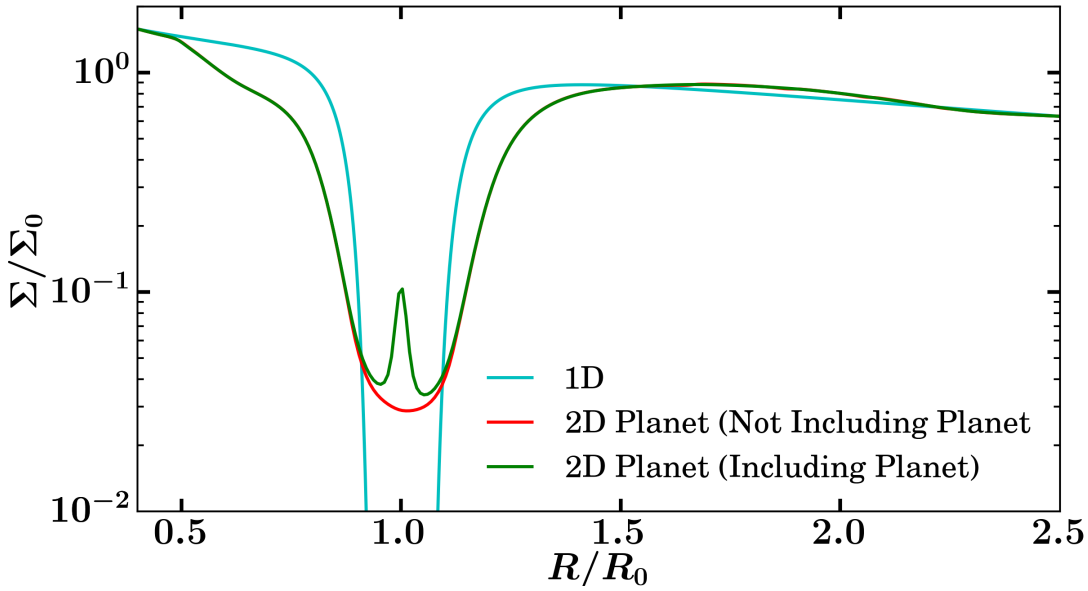

Figure 3.6: Comparison of one dimensional surface density profile and two dimensional azimuthally averaged surface density profiles at equilibrium for a $q = 10^{-3}$ planet. The two dimensional profiles correspond to the result shown in Figure 3.2. The two dimensional profiles here show the necessity of removing the planet's surface density perturbation when averaging surface density, as discussed in Section 3.3.2, however, the most important comparison is the large discrepancy between the one dimensional and two dimensional gap depths.

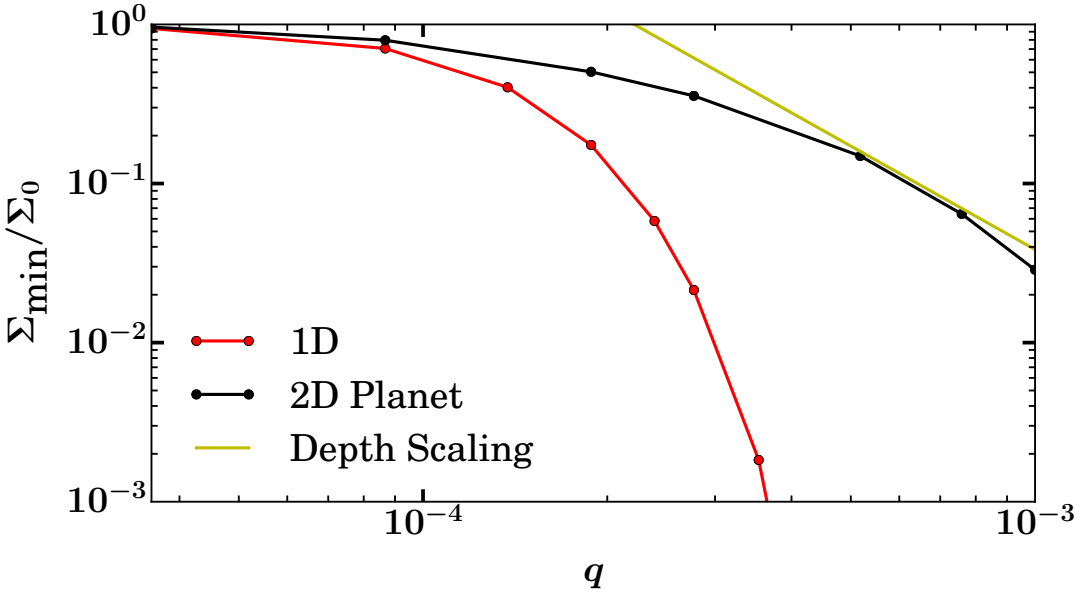

Figure 3.7: Comparison of one dimensional and two dimensional results for a range of planet masses. The large discrepancy in gap depth between these simulations is clearly visible. The gap depth scaling law given by Equation 3.15 is also shown here and our good agreement provides confirmation that our two dimensional simulations are returning expected results.

Section 3.4.1. The ramifications of this can be seen in Figure 3.10, which shows the significantly reduced depth of the two dimensional impulse approximation gap compared to the one dimensional simulation.

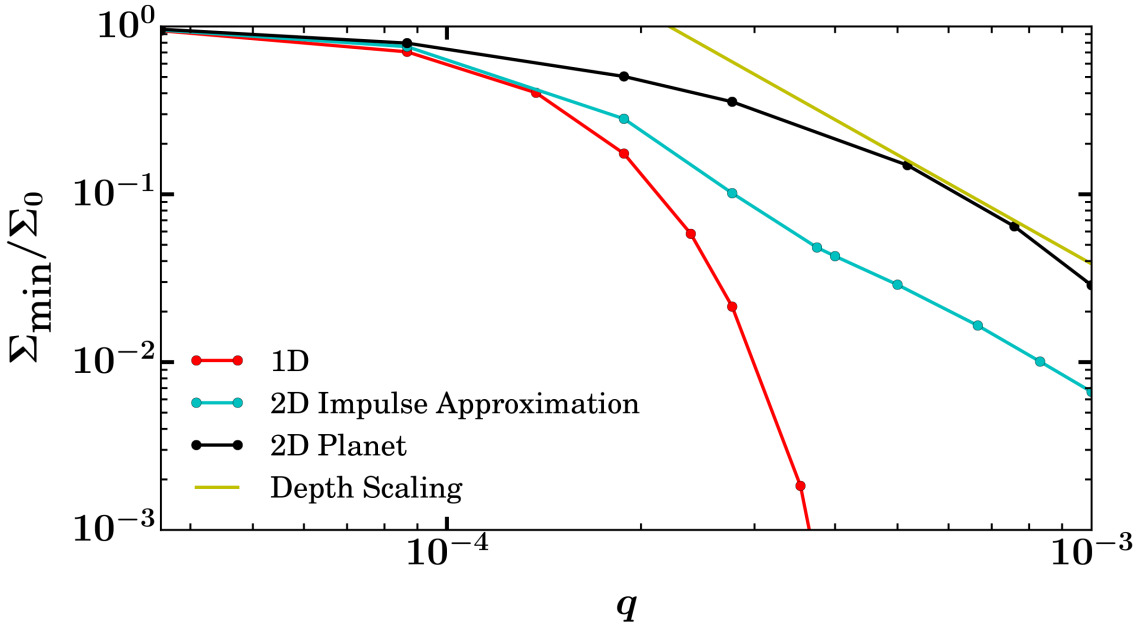

Figure 3.8: Comparison of one dimensional, two dimensional planet and two dimensional impulse approximation (Equation 3.9) results for a range of planet masses. We discuss the reason for this significant improvement over the one dimensional model in Section 3.4.3.

Figure 3.8 shows the significant difference in gap depth resulting from using the impulse approximation in two dimensions. Applying this approximation to the torque of the planet accounts for the majority of the discrepancy, with only roughly an order of magnitude in difference between the two dimensional planet simulations and the two dimensional impulse approximation simulations.

It is not expected that the applied torque density distribution would return a gap depth of equal magnitude, as it is only an approximation to the planet, however, it is important to understand why the formed gap is significantly shallower in two dimensions than in one. The disc material is forced by the torque density distribution to attempt to form a gap of similar shape to those in Figure 3.4, namely deep, thin and axisymmetric, characteristic of one dimensional simulations. This results in a steep surface density gradient at the gap edge, which causes vortices to form and the gap edge to become unstable (Lovelace et al., 1999; Li et al., 2000). This results in the excitation of density waves which redistribute angular momentum. This lessens the surface density gradient at the gap edge forming a more stable gap and reducing

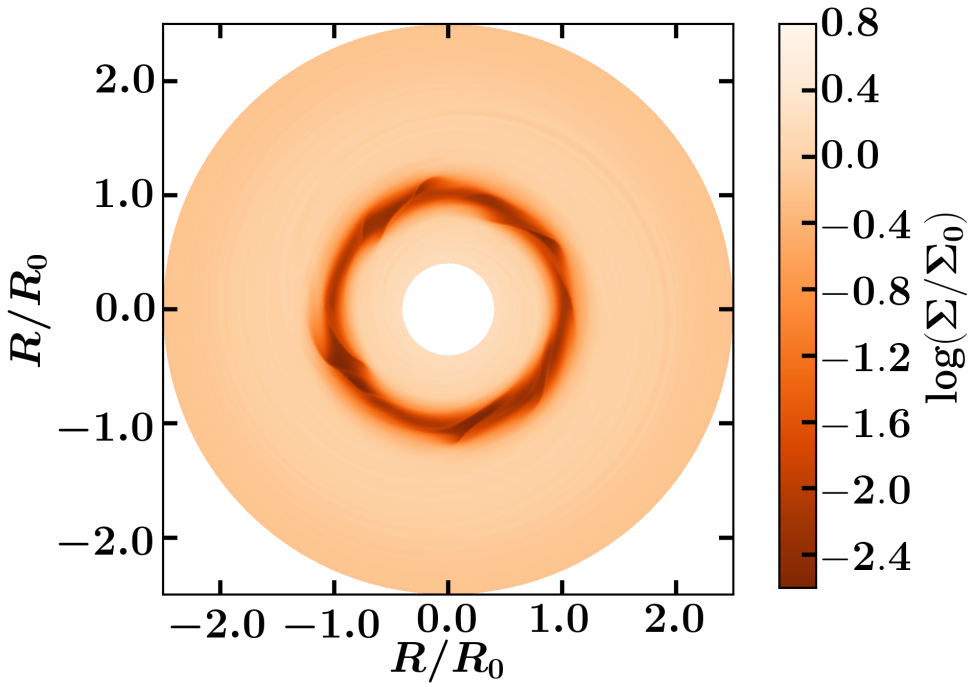

Figure 3.9: A two dimensional simulation using FARGO3D, after evolving to equilibrium at 1000 orbits under an applied one dimensional torque density distribution of the impulse approximation form. The magnitude of the torque density distribution was as if it was caused by a $q = 10^{-3}$ planet, and was applied to a $\nu = 10^{-5}$, $h = 0.05$ disc.

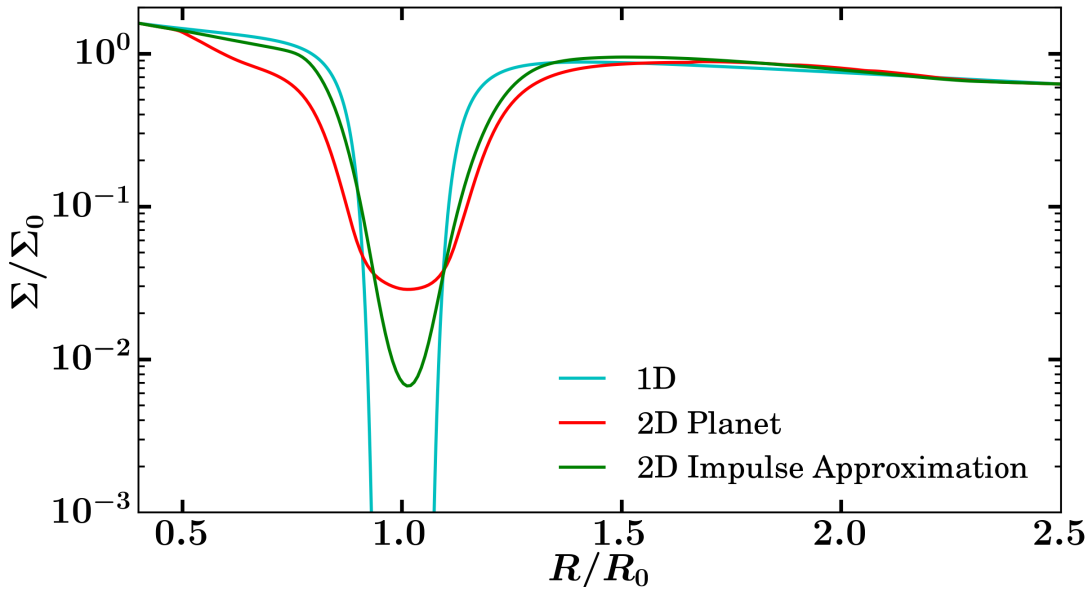

Figure 3.10: Comparison of one dimensional surface density profile and two dimensional azimuthally averaged surface density profiles from the planet and impulse approximation (Equation 3.9) cases at equilibrium for a $q = 10^{-3}$ planet. This also shows the extent to which azimuthally averaging the turbulent two dimensional surface density distribution from Figure 3.9 returns a smooth one dimensional surface density profile.

the gap depth, providing the shallower gap we see. The vortices outside the gap edge will often merge and weaken before equilibrium is reached. They are, however, present throughout the discs lifetime after excitation to ensure the disc does not return to an axisymmetric state. The profile of this shallower gap can be seen in Figure 3.10, not dissimilar to that of the two dimensional planet case.

Despite the angular momentum transfer creating a shallower gap, a quick comparison of Figures 3.2 and 3.9 shows that the one dimensional torque density distribution forms a significantly more turbulent gap. The presence of this turbulence gives rise to short term variability in equilibrium gap depth, which makes averaging gap depth over 100 orbits an important factor in determining true equilibrium gap depth.

While this rectifies a large portion of the discrepancy between these simulations, the gap depths are still close to an order of magnitude lower than in the two dimensional planet simulations. We proceed to make further amendments to the applied torque density distribution in order to investigate the remaining difference.

3.4.4 Improved one dimensional torque density distribution

Using the results of D'Angelo & Lubow (2010) we proceed to modify the torque density distribution to the form given in Equation 3.11 and measure the impact on gap depth. The variation of the minimum surface density with mass under the influence of this torque density distribution can be seen in Figure 3.11. Again we can see better agreement between the model and the two dimensional planet model to a larger mass, but still a large discrepancy in the higher mass regime.

Returning to the torque density distributions presented by D'Angelo & Lubow (2010) in their figure 15, we now consider the lower panel. The torque density distributions here correspond to high mass planets, the regime in which our simulations are inconsistent with the two dimensional planet results. Comparing the shapes of these distributions we see interesting phenomena present in the larger mass distributions that are otherwise absent in the lower mass distributions. As the mass of the planet increases, the peak/trough of the torque density distribution becomes narrower, it approaches zero at a greater distance from the planet, and remains approximately zero for longer. This is a significant deviation from the torque density distribution currently being applied at these high masses. Therefore, we crudely reproduce this phenomenon in our model by setting our torque density distribution to zero for $|R - R_0|/R_{\rm Hill} < 0.95$ with a sharp transition region from $0.95 \leq |R - R_0|/R_{\rm Hill} \leq 1.05$. The difference between these two torque models is clearly visible in Figure 3.12. This new torque model is applied to the high mass

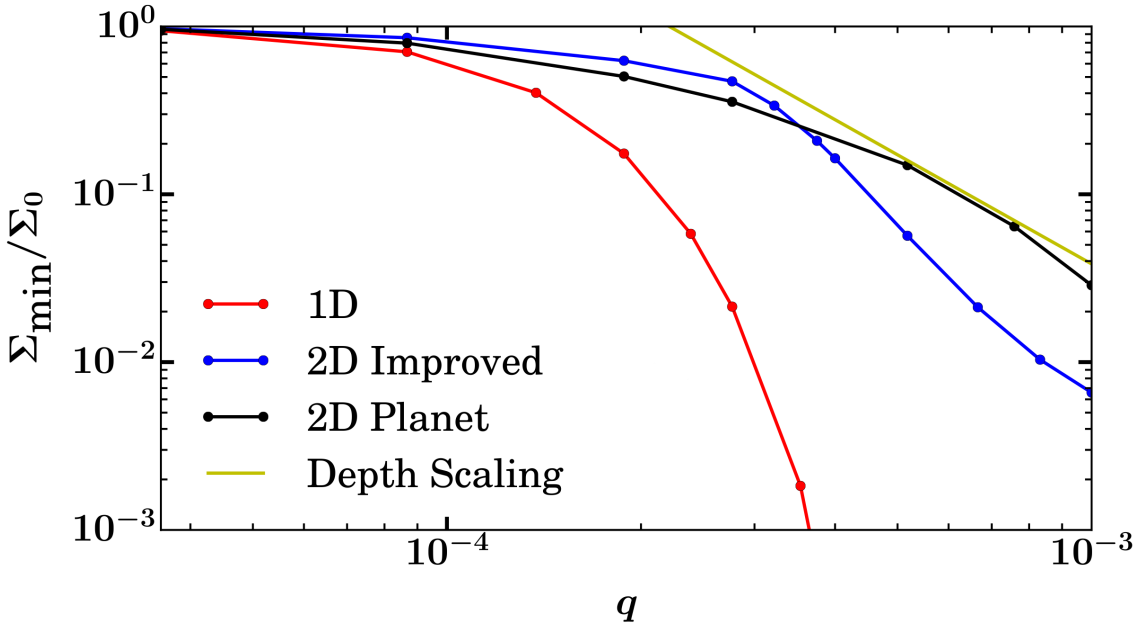

Figure 3.11: Comparison of one dimensional, two dimensional planet and the improved torque density distribution (Equation 3.11) results, as discussed in Section 3.4.4. This gives a similar improvement over the one dimensional model as the impulse approximation, but also shows consistency with the two dimensional planet simulation to a larger mass.

simulations, where the previous model begins to diverge from the two dimensional planet simulation. The results of this can be seen in Figure 3.13. This shows greater consistency with the two dimensional planet simulation to even larger masses, compared to previous simulations, however, we still observe a trend away from the two dimensional planet simulation as the mass increases. The discontinuous region present at roughly $q = 3 \times 10^{-4}$ is a result of the regime change from the low mass to the high mass torque density distribution shown in Figure 3.12.

A conclusion we can draw from this is that the shape of the applied torque density distribution has a significant impact on the equilibrium depth of the resultant gap. Additionally it appears that the shape of the distribution at close proximity to the planet plays an important role in determining gap depth. Further investigation shows that while the shape of the distribution in this region does impact the gap depth, it is currently unknown how important and what other factors are contributing to the gap depth. This is discussed in greater detail in Section 3.6. Nevertheless, we can clearly see the form of the torque density distribution is important, and worth further investigation.

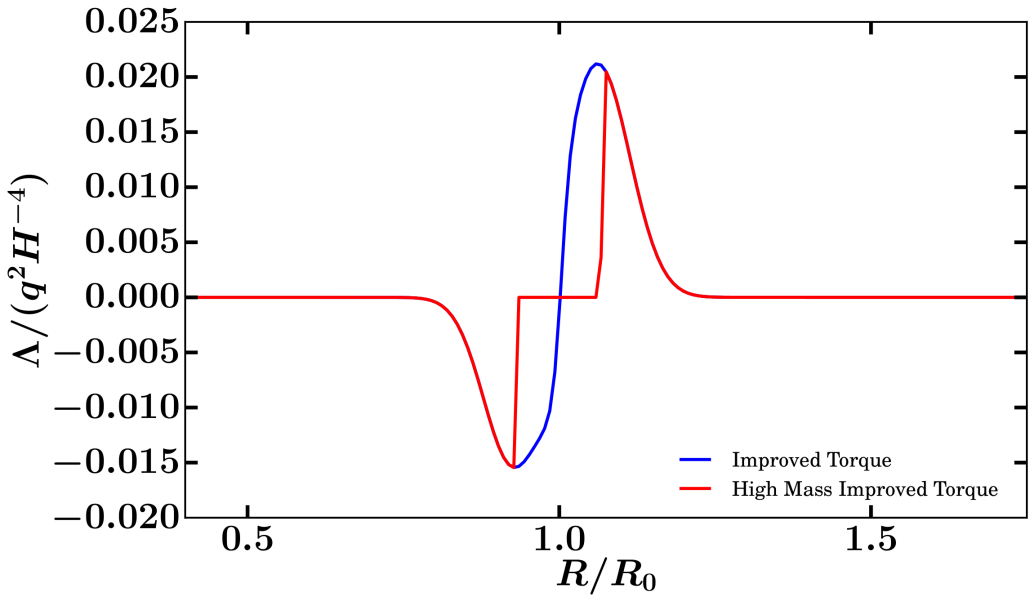

Figure 3.12: Comparison of the torque density distribution given by Equation 3.11 (Improved Torque), and the torque density distribution for high mass planets detailed in Section 3.4.4, for which the torque in the immediate vicinity of $R/R_0 = 1.0$ was set to 0. The example here is shown for a $q = 10^{-3}$ planet. The significant difference between these two distributions is both the absence of torque close to the planet, and the steep gradient transitioning into this area. Again, we show the torque from the planet acting on the disc.

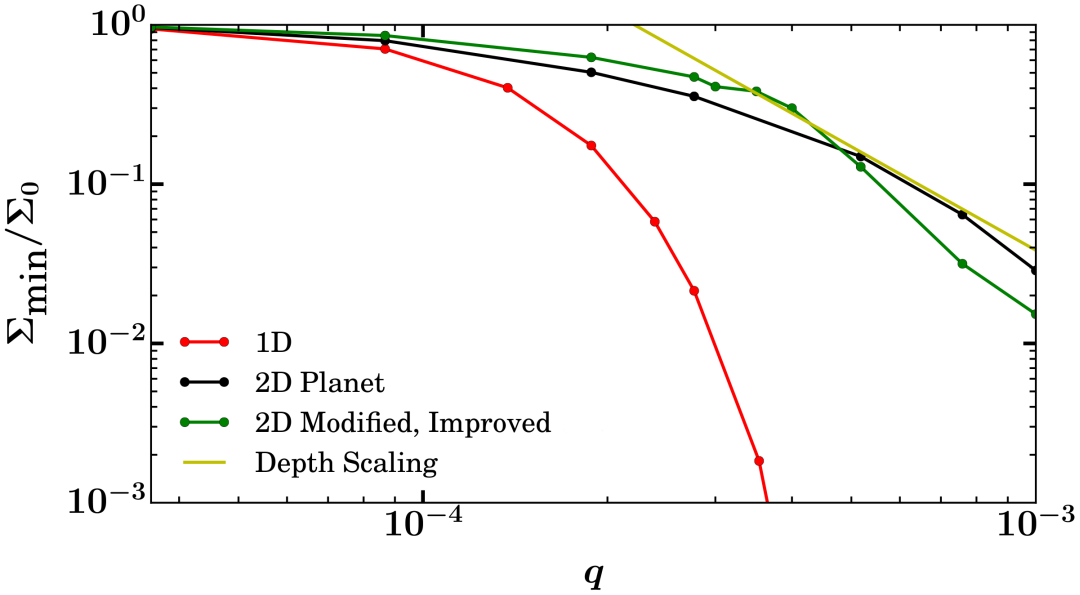

Figure 3.13: Comparison of one dimensional, two dimensional planet and the modified high mass torque density distribution (Section 3.4.4) results. This shows even greater consistency with the two dimensional planet simulation to larger masses when compared to previous simulations.

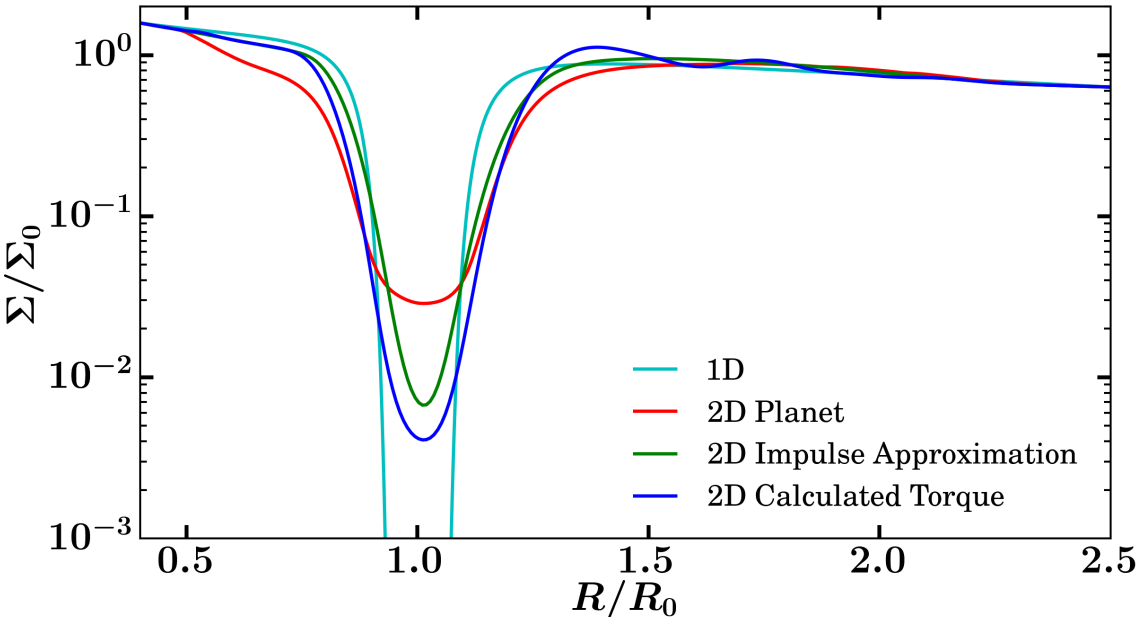

Figure 3.14: Comparison of azimuthally averaged gap profiles for a number of different gap forming mechanisms in an $h = 0.05$, $\nu = 10^{-5}$, $q = 10^{-3}$ simulation.

3.4.5 One dimensional form of the planet's torque density distribution

Concluding that the torque density distribution is an important factor, we now apply the azimuthally averaged shape of this distribution calculated from our two dimensional planet simulations, as presented in Equation 3.13. For a sample case of a $q = 10^{-3}$ mass planet, this torque distribution can be seen in Figure 3.1. This was calculated after equilibrium had been attained, resulting in a constant torque distribution orbit to orbit. This distribution is comparable to the previous distributions used, although shows interesting detail at very close proximity to the planet. We now apply this torque to a two dimensional disc in the same manner as our previous distributions. In Figure 3.14 we compare the azimuthally averaged gap profile of this with the corresponding profiles from prior simulations. In Figure 3.15 we compare the calculated one dimensionally averaged torque distributions for a range of planet masses with our prior simulations.

This clearly shows that using the exact torque distribution is an improvement over the results of the one dimensional model, but does not match the previous accuracy of the modified improved torque density distribution. As a result, we have to conclude here that the torque density distribution is significant to the depth of the gap and using the correct form in two dimensions can greatly reduce the discrepancy

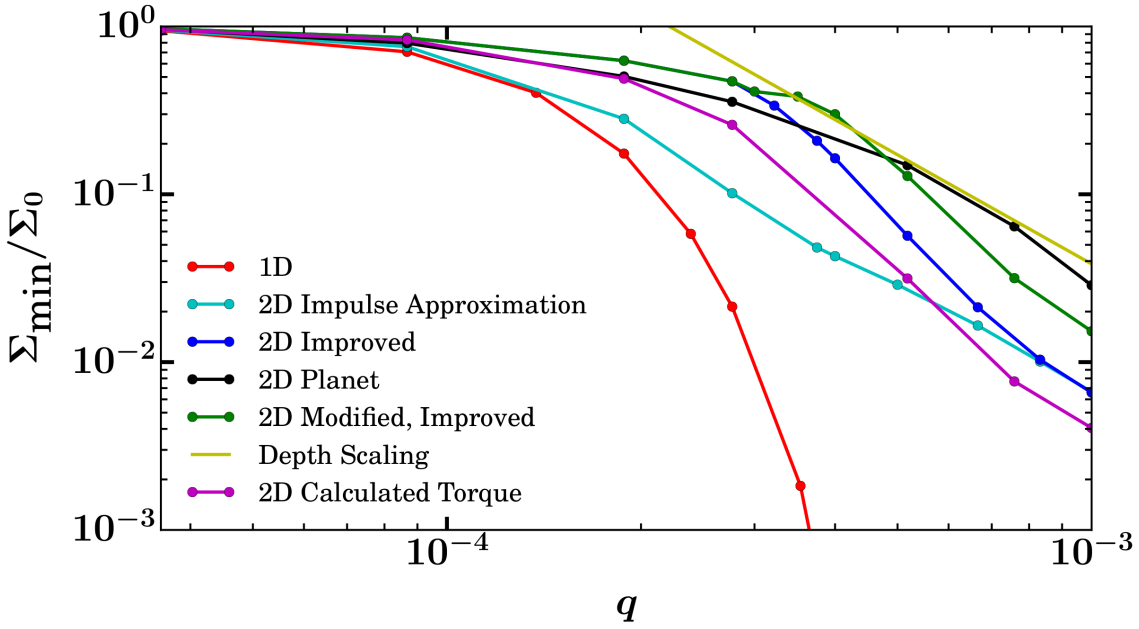

Figure 3.15: Comparison of all simulations discussed in Section 3.4. This shows that, to high mass, the modified high mass torque density distribution provides the closest fit to the two dimensional planet gap depths. At low mass ($q \leq 1.88203 \times 10^{-4}$), however, the one dimensionally averaged calculated torque density distribution provides the best fit.

between one and two dimensional models. However, there appears to be a limit to which the torque density distribution can improve over the one dimensional model, as the torque density distribution is only an approximation to the planet. We now investigate the behaviour of the gap under applied torque density distributions in discs with different parameters.

3.5 Results of varying Disc Parameters

3.5.1 Low viscosity disc

We now proceed to investigate the behaviour of a disc with $\nu = 10^{-6}$, corresponding to an $\alpha = 4 \times 10^{-4}$ at $R = R_0$. We initialise the disc as discussed in Section 3.3.2 with the only difference being the lower viscosity and follow a procedure largely the same as that described in Section 3.4, however, we do not investigate the improved torque density distribution that we discuss in Section 3.4.4 as unfortunately this disc is outside the range of discs studied in D'Angelo & Lubow (2010) and therefore they do not have fit parameters pertaining to it. We investigate this disc for one planet mass, which we choose to return similar gap depths to our previous $q = 10^{-3}$ planet

simulations. This corresponds to $q = 2.77 \times 10^{-4}$ in this case. In two dimensions, simulations now run to 10000 orbits before reaching equilibrium due to this low viscosity.

The cases investigated were the one dimensional, two dimensional planet, two dimensional impulse approximation and two dimensional azimuthally averaged calculated torque distribution. The azimuthally averaged gap profiles for these can be seen in Figure 3.16. This shows a similar result to those presented in Figure 3.14, the one dimensional gap is significantly deeper than its two dimensional counterparts, and the application of one dimensional torques to the two dimensional disc make up a sizeable amount of the discrepancy, but still remain roughly an order of magnitude lower than the two dimensional planet case. A deviation from the results of Figure 3.14 is that here the azimuthally averaged calculated torque distribution provides better agreement than the impulse approximation, whereas in the higher viscosity disc this is not the case. Despite this we see good agreement between the low viscosity disc and our standard disc model, in that the gaps formed by one dimensional torque density distributions in two dimensions are significantly shallower than in one dimension, however, some discrepancy still remains between these gap depths and the two dimensional planet case.

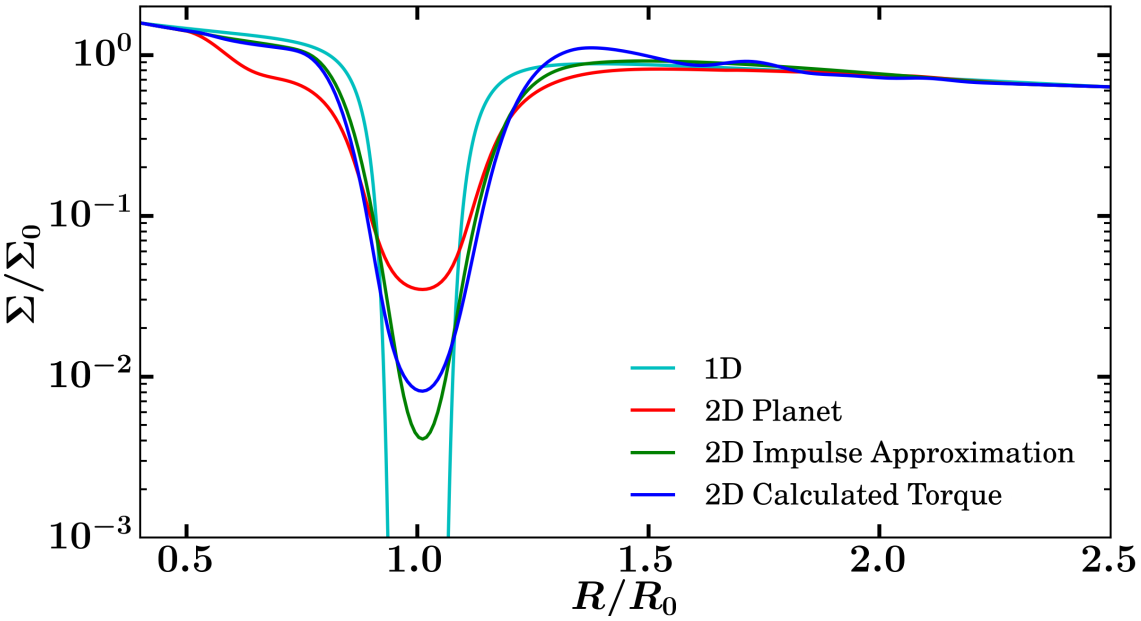

Figure 3.16: Comparison of azimuthally averaged gap profiles for a number of different gap forming mechanisms in an $h = 0.05$, $\nu = 10^{-6}$, $q = 2.77 \times 10^{-4}$ simulation. The one dimensional gap depth is of the order $\Sigma/\Sigma_0 \approx 10^{-18}$.

3.5.2 High aspect ratio disc

We now proceed to investigate the behaviour of a disc with $h = H/R = 0.063$ and $\nu = 10^{-5}$, corresponding to an $\alpha = 2.52 \times 10^{-3}$ at $R = R_0$. We initialise the disc as discussed in Section 3.3.2 with the only difference being the higher aspect ratio. We investigate this disc for the same cases as the low viscosity disc discussed in Section 3.5.1. Again we investigate for one planet mass, $q = 2 \times 10^{-3}$, selected with the same reasoning as in Section 3.5.1.

The azimuthally averaged gap profiles for the cases investigated can be seen in Figure 3.17. Again we see a similar result to that presented in Figure 3.14, with the one dimensional gap still significantly deeper than its two dimensional counterparts, and the application of one dimensional torque density distributions to the two dimensional disc make up a sizeable amount of the discrepancy, but still remain roughly an order of magnitude lower than the two dimensional planet case. However, in contrast to Section 3.5.1, here we see the impulse approximation providing a better agreement to the two dimensional planet case than the azimuthally averaged calculated torque distribution. This is the exact result found in Figure 3.14 and hence we see the two dimensional gaps formed by one dimensional torque density distributions are significantly closer in depth to the two dimensional planet gap than their one dimensional counterparts.

3.6 Discussion

In this chapter we find that the discrepancy between equilibrium gap depths from one and two dimensional simulations can be explained by the absence of the RWI in one dimension. We determine this using one dimensional torque density distributions as the gap forming mechanism in two dimensional simulations. We find good agreement with prior work regarding our one and two dimensional planet simulations (Lin & Papaloizou, 1986; Crida et al., 2006; Kanagawa et al., 2015), specifically our two dimensional planet gap depths closely follow the established gap depth scaling law (Fung et al., 2014).

Our results show the excitement of the RWI in two dimensional simulations when a one dimensional torque density distribution is used as a gap forming mechanism. Prior work (de Val-Borro et al., 2006, 2007) shows that the RWI is present in two dimensional planet simulations for low viscosity discs or high mass planets. It is also shown that these instabilities die off before equilibrium is reached (Fu et al., 2014a; Hammer et al., 2016), suggesting that the gap is maintained at marginal stability

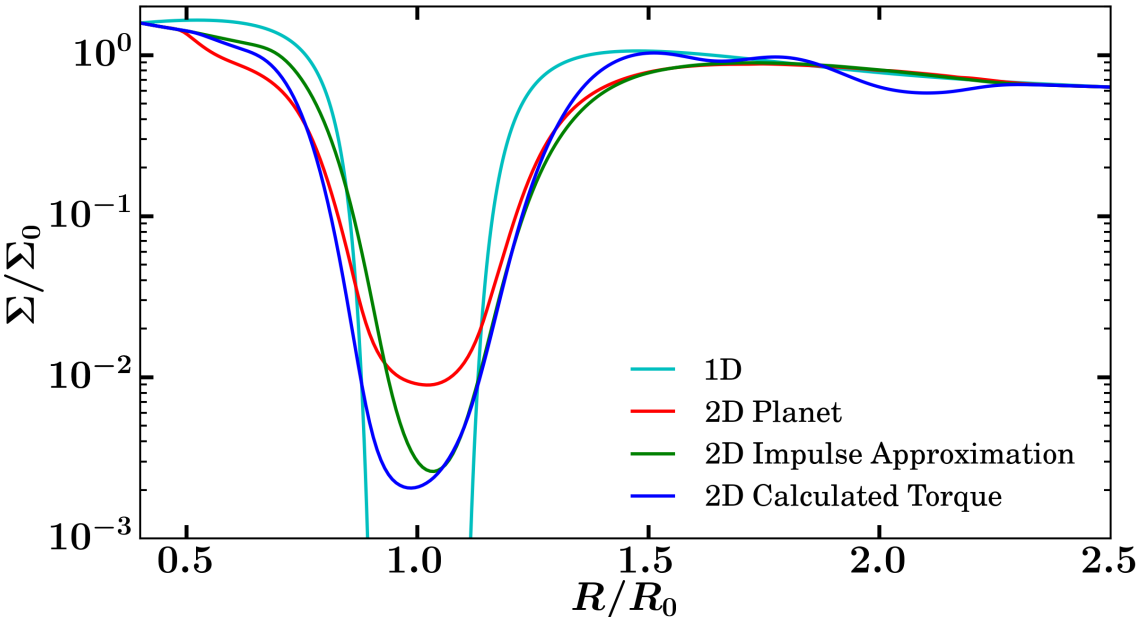

Figure 3.17: Comparison of azimuthally averaged gap profiles for a number of different gap forming mechanisms in an $h = 0.063$, $\nu = 10^{-5}$, $q = 2 \times 10^{-3}$ simulation. The one dimensional gap depth is of the order $\Sigma/\Sigma_0 < 10^{-30}$.

set by the point at which the excitement of the RWI occurs.

We also find that when the planet is not massive enough to open a gap in the disc there is good agreement between one and two dimensional simulations as shown in Figure 3.15. This is because there is no RWI present in the two dimensional simulations that do not open a gap. This result could be important for modelling the behaviour of low mass planets in one and two dimensions.

For population synthesis models (e.g. Mordasini et al., 2009), especially those regarding the formation of giant planets (Ida & Lin, 2008), the depth of the gap formed and the feedback of the planet onto the disc structure is very important. If the depth of the gap is set by the instabilities observed, such as the Rayleigh Instability investigated in Kanagawa et al. (2015) and the RWI presented here, this could lead to an improved recipe for gap formation, which would be very important in future population synthesis models. The fact that our results hold for differing disc parameters (Figure 3.16, Figure 3.17) reinforces the relevance of our results to such a recipe for gap formation.

We can find some similarity to our investigation of the gap depth discrepancy in Kanagawa et al. (2015). They examine gap depth in viscous, one dimensional discs accounting for deviation from Keplerian disc rotation and include the Rayleigh

stable condition. They find a shallowing of the gap due to angular momentum transfer from the gap edge, arising as a result of the deviation from Keplerian disc rotation, or violation of the Rayleigh stable condition. We see a similar situation in our two dimensional simulations, however, the shallowing of the gap is a result of the excitement of the RWI, which is unaccounted for in one dimension. Kanagawa et al. (2015) also show that a significantly large wave propagation length before angular momentum deposition occurs can make the one dimensional gap shallower and wider. They find a wave propagation length that gives a comparable depth to the two dimensional model for high mass planets, however, the increase in mass corresponds to a long propagation length and a wider gap. The correct width of the two dimensional gap is unknown, with prior work both agreeing (Varnière et al., 2004) and disagreeing (Duffell & MacFadyen, 2013) with their result.

Closely related to this is the work of Crida et al. (2006). They find that the equilibrium gap profile of two dimensional simulations can be accurately reproduced semi-analytically by balancing the viscous, gravitational and pressure torques on the disc. The pressure torque here arises from the fraction of the gravitational torque that is carried away by density waves instead of being locally damped and is important to increasing the width of the analytical gaps to match their numerical counterparts. An ansatz for the pressure torque is determined from reference numerical simulations, hence the result is not fully analytic.

The results of both Crida et al. (2006) and Kanagawa et al. (2015) imply that wave propagation can account for the discrepancy between one and two dimensional simulations, however, both approaches introduce parameters that must be fitted in order to determine the wave propagation length, or the magnitude of the pressure torque. From our results we find that two dimensional gaps formed by one dimensional torque density distributions have depths significantly closer to those of two dimensional planet gaps than their one dimensional counterparts. This is a consequence of the excitement of the RWI. Considering this together with the result that high mass planets appear to have gaps maintained at marginal stability (Fu et al., 2014a; Hammer et al., 2016), we suggest that the pressure torque required to form the characteristic two dimensional gaps is such that the gap edge is maintained at marginal stability.

Our results from exploring different discs provide good agreement between themselves. We find in each case that using a one dimensional torque density distribution as an approximation to the torque on the disc due to the planet provides significantly better agreement with the planet case in two dimensions than in one. We also find that in the $\nu = 10^{-6}$ case, the averaged one dimensional torque array proves to

be a better approximation than the impulse approximation, however, in both the $\nu = 10^{-5}$ discs we find the opposite to be true.

The results of Section 3.4.4 give the impression that the treatment of the torque at close proximity to the location of the planet has a major impact on the depth of the resultant gap. This was investigated using the azimuthally averaged calculated torque density distribution, largely due to the detailed shape of the distribution at close proximity to the planet (see Figure 3.1). It was found that setting this detailed region to zero, similarly to the modification made in the second part of Section 3.4.4, had very minimal impact on the gap depth. Widening this area, such that the transition to this region is steeper, moves the gap depth in the correct direction, but not significantly. This was unexpected, as the shape of the azimuthally averaged calculated torque density distribution was crudely forced until it was very close to the modified improved torque density distribution from Section 3.4.4, yet it still does not return a similar gap depth. In order to determine where the difference between these two torque density distributions arises, we split these torque density distributions into two regions, the inner disc ranging from the inner edge of the simulation to the beginning of the zero torque region and the outer disc ranging from the end of the zero torque region to the outer edge of the simulation. We then create two more torque density distributions, consisting of the inner part of one torque density distribution and the outer part of another. In each of these distributions one region is modelled by the improved torque density distribution while the other is the azimuthally averaged calculated torque density distribution. We find that the Inner Disc Modelled simulation has minimal effect in changing the gap depth, whereas the Outer Disc Modelled returns a gap depth extremely close to the original, modified improved torque density distribution. At this point we can conclude that it is not only the torque close to the planet that has a large impact on the gap depth, there are a variety of other factors. Currently, however, we are unable to deduce exactly what all of these factors are or specifically determine how these impact the depth of the formed gap.

One such factor that could be important to consider is the presence of material on horseshoe orbits and its impact on gap depth. Horseshoe orbits are present in simulations containing planets, as the planet perturbs the orbit of material at similar orbital radius to itself as usual. However, this is absent in the case of a torque density distribution forming the gap. The interaction between the planet and material on a horseshoe orbit gives rise to a corotation torque, due to the vortensity (Ward, 1991; Masset, 2001) and entropy (Baruteau & Masset, 2008a; Paardekooper & Papaloizou, 2008; Paardekooper et al., 2010) gradients across this region. The

presence of this torque in the planet case may provide an explanation as to why there is consistently a discrepancy between gap depths from planet formed gaps and torque density distribution formed gaps.

Recent studies into the excitement of the RWI have shown that using more realistic (longer) planetary growth times induce shorter lived and lower amplitude vortices (Hammer et al., 2016). We do not increase the intensity of our applied torque distributions over any number of orbits, instead we apply our torque density distributions from the beginning of each simulation. We do not believe the results of Hammer et al. (2016) will affect our findings, as in the torque density distribution simulations the instabilities are present throughout the whole simulation. If the instabilities did die off the disc would return to an axisymmetric state and the gap would again try to become deep and narrow, which would in turn re-excite the RWI. Hence, the RWI will be present throughout the discs lifetime.

We make a number of simplifying assumptions during this investigation. Planets were held on a constant, circular orbit ($e = 0$) and we do not include planetary migration. We use a locally isothermal equation of state, as radiative transfer and cooling is difficult to accurately model and is very computationally expensive. Hence, disc thickness H is linear in R and gives us a constant aspect ratio $h = H/R$, for $H \ll R$. We use a constant kinematic viscosity across the disc and neglect disk self-gravity. The effect of disc self-gravity on vortex instabilities has been investigated (Lin & Papaloizou, 2011a) and it was found that larger disc masses form more vortices. With sufficiently large self-gravity, however, vortex formation was suppressed. It was also found that self-gravity acts to delay vortex merging. The fact that vortex formation is still present with self-gravity gives confidence that our results would still be applicable in that regime.

We also neglect magnetohydrodynamic (MHD) effects on the disc, which would otherwise cause the disc to become turbulent. Gap formation in MHD turbulent discs has been investigated (Papaloizou et al., 2004) and has been found to return wider and shallower gaps than their purely hydrodynamic counterparts (Winters et al., 2003). The impact of turbulence would be difficult to account for in the applied torque distribution cases and would be computationally expensive. Papaloizou et al. (2004) show that in the vicinity of the planet, local shearing box simulations are a good approximation to global disc models and are less computationally expensive, which would be useful if accounting for MHD was pursued. These factors all constitute a very idealised case of gap formation and removing some of these simplifying assumptions could also impact our results.

We compare results of two dimensional and one dimensional simulations for one

dimensional gap forming mechanisms such as the impulse approximation. Currently it is unknown as to how a gap formed by a one dimensional torque density distribution such as the impulse approximation would behave in three dimensions and how this would compare to both its one and two dimensional counterparts. Recent studies into three dimensional gap formation have found three dimensional gaps have similar depth and shape to their two dimensional counterparts (Fung & Chiang, 2016). With this in mind, and considering the gap depth is set by the excitement of the RWI, a two dimensional instability, we can predict with some confidence that there would be minimal change in our findings if it was taken to three dimensions. Of course this area is far from fully explored and a number of assumptions are still involved, so further investigation would be required before any conclusions could be drawn with certainty.

3.7 Conclusions

We have investigated gap formation in viscous protoplanetary discs in both one and two dimensions, with intent to measure the depth in surface density of the gap formed. We encountered a well known discrepancy (Crida et al., 2006; Kanagawa et al., 2015) in depth between one and two dimensional gaps. We proceeded to investigate this discrepancy using a new method, the application of a one dimensional gap forming torque density distribution to a two dimensional disc devoid of a planet. We studied this for a number of forms of the torque density distribution;

- An impulse approximation (Equation 3.9), as of Lin & Papaloizou (1986).
- An improved one dimensional torque density distribution using results from three dimensional models (Equation 3.11), as of D'Angelo & Lubow (2010).
- A modified version of the above torque density distribution from D'Angelo & Lubow (2010), in which $\Lambda = 0$ at close proximity to the planet.
- An azimuthally averaged one dimensional torque distribution calculated due to the presence of the planet in two dimensional simulations (Equation 3.13).

We found that applying a one dimensional torque density distribution across the disc as a gap forming mechanism results in the formation of a gap significantly shallower than the one dimensional gap, however, still roughly an order of magnitude deeper than the equivalent two dimensional planet simulation. This occurs as a deep thin gap is attempted to be formed, much like in the one dimensional case, however,

as the two dimensional gap edge becomes too steep vortices begin to form and the gap edge becomes unstable. This excites density waves that redistribute angular momentum across the disc, which acts to reduce the steepness of the gap edge and as a result the gap becomes shallower. de Val-Borro et al. (2006, 2007); Fu et al. (2014a); Hammer et al. (2016) show that the RWI is present in low viscosity discs or high mass planet simulations, however, the instability dies off before equilibrium is reached. This suggests that the gap is maintained at marginal stability, set by the point at which the excitement of the RWI occurs. Our results show a similar gap depth between two dimensional gaps formed by torque density distributions, in which the RWI is present, and two dimensional planet gaps, in which no RWI is present. This can be understood if the planet gap is maintained at marginal stability, even when there is no obvious RWI present.

We observe very similar behaviour across all applied torque density distributions. We see that while the choice of torque density distribution does affect the resultant gap depth, the change is often minimal and cannot make up for the order of magnitude discrepancy remaining. The only torque density distribution that makes any significant progression towards the two dimensional planet model is the modified version of the improved torque density distribution from D'Angelo & Lubow (2010), in which the torque at close proximity to the planet is set to zero. This would propose that the torque close to the planet is the dominant factor in the resultant gap depth, however, while it does impact the gap depth, this does not appear to be the case. Therefore, while we can say the gap depth is sensitive to the torque density distribution, the exact dependencies are still unknown. We are considering here only the gap depths in the higher mass regime, where the discrepancy still exists. For lower mass planets, roughly $q \leq 3 \times 10^{-4}$, the improved torque density distribution and the azimuthally averaged calculated torque density distribution are significant improvements over the impulse approximation, and almost entirely eliminate the discrepancy. Hence, in this mass range, these are very good approximations to the effect of a two dimensional planet.

When we extend this study into discs with different parameters we find very good agreement with our prior results. While we only study two additional discs, changing the aspect ratio and viscosity, and study each case for a single planet mass, they return the expected results. With this we can confidently say our results are not constrained solely to the disc we have been investigating.

4 Investigating the possibility of reversing giant planet migration via gap edge illumination

Abstract

A massive planet in a protoplanetary disc will open a gap in the disc material which acts as a transition between Type I and Type II planetary migration. Type II migration is slower than Type I migration, however, it is still desirable to slow down Type II migration to allow gas giant planets with semi-major axis in the range $5-10$AU to exist, similarly to our Solar system. We investigate a method of slowing down and reversing Type II migration by heating the outer gap edge due to incident radiation from the central star. Using an approximate vertically averaged heating method we find that Type II migration can be slowed or in extreme cases reversed if we assume near maximum allowed irradiation from the central star. Therefore, we believe this is a very interesting phenomenon that should be investigated in greater detail using three dimensional hydrodynamic and radiative transfer simulations.

4.1 Introduction

Planets form in protoplanetary discs by accretion of gas and dust particles and will excite density waves in the disc material that transport angular momentum away from the planet. This angular momentum is deposited in the disc, exerting a torque on the disc material. If the angular momentum is deposited close to the planet (Lin & Papaloizou, 1993; Bryden et al., 1999) and the resultant torque is stronger than the viscous diffusion torque of the disc material (Lin & Papaloizou, 1979; Goldreich & Tremaine, 1980; Takeuchi et al., 1996; Crida et al., 2006) then a gap is formed around the location of the planet. These two conditions are known as the thermal and viscous criteria for gap opening. It has also been shown that small mass planets

can open gaps at a greater distance from the planet, where the excited density wave shocks (Goodman & Rafikov, 2001; Rafikov, 2002).

The interaction between the planet and the disc plays an important role in the evolution of the planet's orbital radius. There are two main regimes of planetary migration, Type I and Type II, which are differentiated between by the mass of the planet and as a result, the presence of a gap in the disc material. Type I migration occurs for planets that are not massive enough to open a gap in the disc, and migrate due to the Lindblad and corotation torques acting upon them (Goldreich & Tremaine, 1980). The corotation torque depends on the magnitude of the thermal and viscous diffusion in the disc (Masset, 2001; Paardekooper et al., 2010, 2011) and these can alter the speed of migration. Type II migration occurs when a planet is massive enough to open a gap in the disc. The low density area surrounding the planet slows down the Type I migration, and the planet now migrates with the gap on the viscous evolution timescale of the disc (Lin & Papaloizou, 1986). Hence, Type I migration is considerably faster than Type II migration (Ward, 1997). Despite this reduction in migration rate, it has been found that the predicted timescale for Type II migration is shorter than the lifetime of the disc, suggesting that most planets undergoing Type II migration will migrate into and be absorbed by the central star (Hasegawa & Ida, 2013). Observational evidence dictates that gas giant planets are more common at orbital distances $R > 1$AU (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016), which would not be the case if this predicted timescale was correct. Current models estimate that for giant planets to survive migration they must form at distances $R > 20$AU (Coleman & Nelson, 2014). Hence, it is implied that something must be limiting the migration speed of a planet in the Type II migration regime (Nelson et al., 2000).

The argument for classical Type II migration relies on the absence of material surrounding the planet, a result of the torque exerted on the disc by the massive planet clearing the area around it and forming a gap. This is clearly the case in one dimensional simulations, such as those in Lin & Papaloizou (1986), however, extending this to higher dimensions this is found not to be the case (Kanagawa et al., 2015; Hallam & Paardekooper, 2017). It can easily be seen from two or three dimensional simulations that disc material can be present within the gap, and can cross the gap on horseshoe orbits. Hence, we can only view classical Type II migration as an idealised case. The result is a deviation from the classical Type II migration rate, the viscous evolution timescale of the disc. The migration rate can now be significantly faster or slower than the viscous evolution timescale of the disc and is dependent on the planet and disc parameters (Duffell et al., 2014). Dürmann

& Kley (2015) found that only for small disc masses $M_d/M_J < 0.2$ is the migration rate slower than the classical rate. Therefore, we investigate a new method of slowing the rate of planet migration in the Type II regime. We propose that radiation from the central star is incident on the outer edge of gap formed by the giant planet, visible as a result of the flaring of the disc. This incident radiation heats the outer disc, increasing the scale height in this region. This process is outlined in Figure 4.1. As the one-sided torques scale with the aspect ratio as h^{-3}, the result of this may be a lowered contribution to the net torque on the planet from the planet's outer wake, which in turn causes the net torque on the planet to become more positive. As the rate of migration is proportional to the magnitude of the torque on the migrating planet and the direction given by the sign of the torque, this process could slow the planet's migration. Observational signatures of irradiated gap edges have been studied by Jang-Condell & Turner (2013).

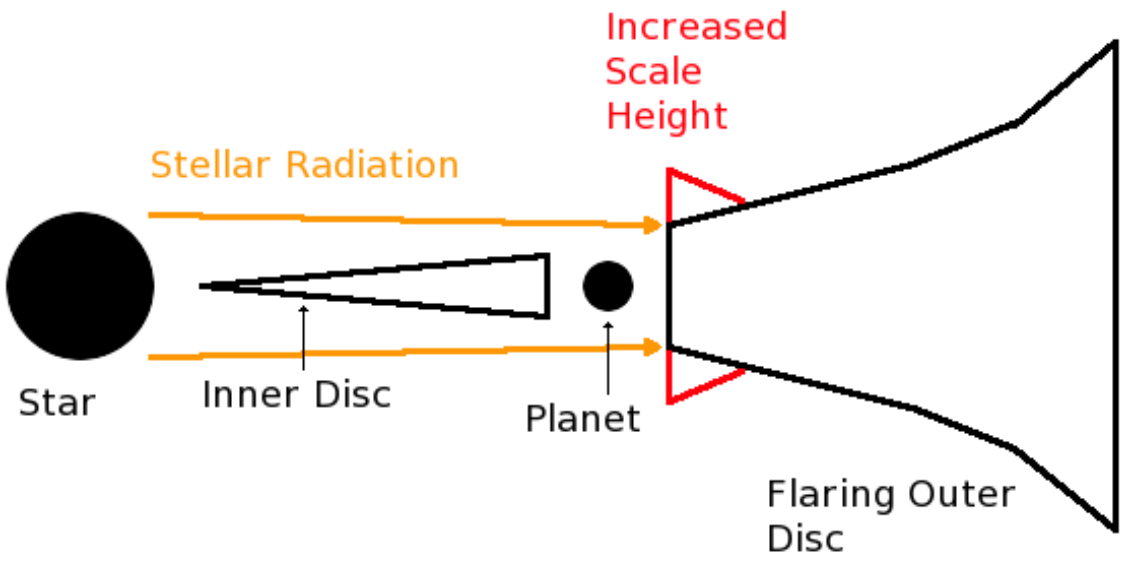

Figure 4.1: A diagram showing the process by which the outer gap edge is heated by the star. This is the method in which the torque contribution from the outer disc may be reduced and hence the net torque on the planet may become more positive.

This chapter is arranged as follows. In Section 4.2 we derive the relevant equations solved to simulate disc evolution. In Section 4.3 we discuss the code used and the numerical setup for our simulations. In Section 4.4 we justify the choice of parameters used in our simulations. In Section 4.5 we discuss the onset of instabilities in our simulations and how we go about avoiding them. In Section 4.6 we present our

results of heating the gap edge and the resultant dependence of the net torque on previously justified parameters. In Section 4.7 we discuss our results in the context of prior work while justifying assumptions made and postulating any impact they may have. Finally, in Section 4.8 we present our conclusions.

4.2 Basic Equations

In this chapter we perform a parameter study in two dimensions to outline a region for which there is the possibility of reversing the direction of Type II migration. While the heating of the outer disc is a radiative hydrodynamic process which is clearly a three dimensional phenomenon, we choose to represent this phenomenon in a vertically averaged approach outlined in Section 4.2.2. A full radiation-hydrodynamic parameter study in three dimensions is beyond current computational resources and an approximate study such as presented here can serve as a guide for future more realistic simulations.

4.2.1 Two dimensional protoplanetary disc

The continuity equation for the evolution of a protoplanetary disc's surface density, Σ, is

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0, \tag{4.1}$$

where $\mathbf{v}$ is the velocity field. We simulate the evolution of a protoplanetary disc's surface density due to the presence of a planet by solving the two dimensional Navier-Stokes equation for the motion of the disc-planet system,

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P - \nabla \cdot \mathbf{T} - \Sigma \nabla \Phi, \tag{4.2}$$

where $\mathbf{T}$ is the Newtonian viscous stress tensor, P is the pressure and Φ is the gravitational potential of the planet and star system. We solve the energy equation,

$$\frac{\partial e}{\partial t} + \nabla \cdot e\mathbf{v} = -P\nabla \cdot (\mathbf{v}) + \mathcal{C}, \tag{4.3}$$

with an equation of state $P = (\gamma - 1)e$, where e is the volumetric internal energy and γ is the ratio of specific heats. $\gamma = c_P/c_V = 1.4$ for a diatomic gas. $\mathcal{C}$ is a cooling term, in which we introduce a thermal relaxation timescale, τ, as discussed in Section 4.2.2. A cylindrical coordinate system is used, such that $\mathbf{v} = (v_R, R\Omega)$

where v_R and Ω are the radial and angular velocities at a given radius. The planet is held on a fixed circular orbit.

4.2.2 Heating the outer gap edge

In order to attempt to reduce the rate of Type II migration we consider the impact on the net torque on the planet if the outer gap edge was directly heated by radiation from the central star. This may reduce the torque contribution from the outer wake of the planet, while leaving the contribution from the inner wake largely unchanged. To represent this in our simulations we impose a Gaussian distribution in the sound speed profile of the disc, c_s, of the form

$$c_s = \left(1.0 + (A - 1.0)e^{\left(-\frac{((R-R_G)^2}{2\sigma^2}\right)}\right) h \left(\frac{R}{R_0}\right)^f \sqrt{\frac{GM_*}{R}}, \tag{4.4}$$

where A is the factor by which the sound speed increases at $R = R_G$, R is the radial position within the disc, R_G is the location of the Gaussian peak, σ is the standard deviation of the Gaussian, R_0 is the location of the planet, f is the flaring index and M_* is the mass of the star. This represents the stellar heating of the outer gap edge, resulting in an increase in sound speed at that location. This changes both the sound speed and aspect ratio of the disc at this location. The impact a Gaussian has on the sound speed profile of the disc is shown in Figure 4.2. In this chapter we will be investigating the effect of the shape of this Gaussian on the net torque on the planet.

We deviate from a purely adiabatic simulation by applying a thermal relaxation function at constant density, given by

$$\mathcal{C} = -\frac{e - e_i\frac{\Sigma}{\Sigma_i}}{\tau}, \tag{4.5}$$

where e is the volumetric internal energy, $\tau = 2\pi/\Omega_0$ is the thermal relaxation timescale, Ω_0 is the angular velocity at the location of the planet and the subscript i denotes initial and therefore equilibrium values. As τ is a constant this means the rate of cooling in the inner disc will be slower than cooling in the outer disc when compared with the orbital timescale at that location. This does not affect our results, as we are interested in cooling at and around the location of the planet. This is then applied using implicit time stepping to cool the disc in the form

$$e^{n+1} = e_i\frac{\Sigma^n}{\Sigma_i} + \frac{e^n - e_i\frac{\Sigma^n}{\Sigma_i}}{1 + \frac{\Delta t}{\tau}}, \tag{4.6}$$

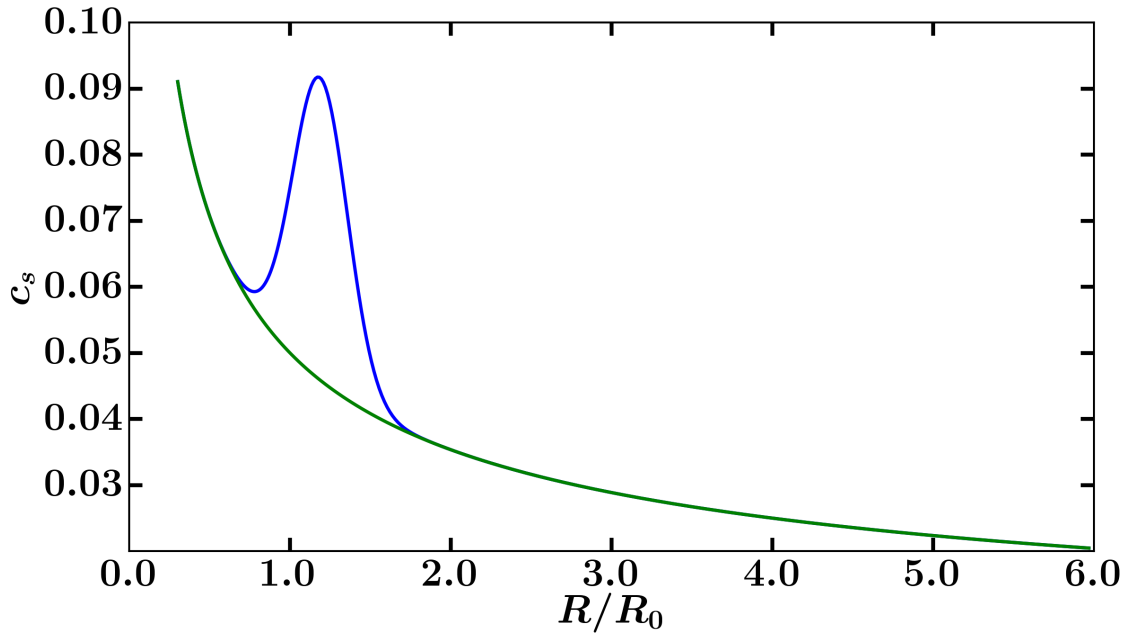

Figure 4.2: Azimuthally averaged initial sound speed profiles for a disc containing a Gaussian with $A = 2.0$ and $\sigma = 0.17$, described by Equation 4.4 and a disc containing no Gaussian in the sound speed profile. Disc setups are described in more detail in Section 4.3 and the parameters of this Gaussian are discussed in more detail in Section 4.6.1.

where the superscript n denotes the timestep at which the relevant field is evaluated.

4.2.3 Total Torque

The total torque per unit mass on the planet was calculated using

$$\mathcal{T}_m = \frac{1}{M_p} \int_{R_{\rm Min}}^{R_{\rm Max}} \int_{-\pi}^{\pi} \Sigma(R,\phi) \frac{d\Phi_p}{d\phi} d\phi R dR, \tag{4.7}$$

where $R_{\rm Min}$ and $R_{\rm Max}$ are the minima and maxima of the radial domain respectively, and Φ_p is the gravitational potential of the planet, given by

$$\Phi_p = -\frac{GM_p}{\sqrt{R^2 + R_0^2 - 2RR_0\cos(\phi - \phi_0) + s^2}}, \tag{4.8}$$

in which $s = 0.6H_0$ is the smoothing length, R_0 is the location of the planet, H_0 is the scale height at the location of the planet and ϕ_0 is the azimuthal location of the planet. The contribution of the planet envelope within 0.8 Hill Radii is not included in the calculation of torque on the planet, similarly to Crida et al. (2008); Dürmann & Kley (2015). Our cut off is applied before force calculation, and ramps sinusoidally back from 0 to the default value at that radii,

$$\mathcal{H}_{\rm c} = \begin{cases} 0 & \text{if } r/r_{\rm Hill} < 0.8 \\ 1 & \text{if } r/r_{\rm Hill} > 1.0 \\ \sin^2\left(\frac{\pi}{0.4}\left(\frac{r}{r_{\rm Hill}} - 0.8\right)\right) & \text{otherwise,} \end{cases} \tag{4.9}$$

where r is the distance from the planet. Then,

$$F_{\text{cut-off}} = F\mathcal{H}_{\rm c}. \tag{4.10}$$

We expect a change in sound speed to impact the net torque on the planet. For a low mass planet, the one sided torques in the disc are given by (Ward, 1997)

$$\mathcal{T}_{\rm os} = C\left(\frac{M_p}{M_*}\right)^2\left(\frac{1}{h}\right)^3 \Sigma_0 R_0^4 \Omega_0^2, \tag{4.11}$$

where C is a constant that is positive for the inner disc and negative for the outer disc and Σ_0 is the unperturbed surface density at $R = R_0$. In an unperturbed disc the one sided torques will mostly cancel, resulting in a small net torque on the planet. Tanaka et al. (2002) has shown that for a two dimensional disc containing a low mass planet the overall net torque on the planet is given by

$$\mathcal{T} = -(1.160 + 2.828\alpha_{\rm sl})\left(\frac{M_p}{M_*}\frac{1}{h}\right)^2 \Sigma_0 R_0^4 \Omega_0^2, \tag{4.12}$$

where $\alpha_{\rm sl}$ is the negative of the radial surface density gradient. As $h \ll 1$, therefore $\mathcal{T} \ll \mathcal{T}_{\rm os}$. If the temperature and therefore sound speed in the outer disc is greater than that of the inner disc Equation 4.11 implies that the inner torque will be favoured. Despite Equations 4.11 and 4.12 pertaining to small planets, this is expected to hold for higher mass planets.

4.3 Numerical Setup

Our simulations are run in two dimensions using the FARGO3D code. This is a magnetohydrodynamic code designed to simulate disc evolution in one to three dimensions by solving the hydrodynamic equations of motion. FARGO3D is a good choice for investigating gap opening planets as it uses a C to CUDA translator to allow simulations to be run on Graphics Processing Units (GPUs). GPUs have limited memory, but decrease computational time significantly, allowing moderate resolution simulations to be run for extended periods of time. For more details see Benítez-Llambay & Masset (2016).

Our disc and planet parameters remain constant between simulations, the variables modified pertain to the applied Gaussian in the sound speed profile only. Hence, the setup described below is the standard disc-planet setup used in all simulations. We use a $q = M_p/M_* = 0.001$ planet, which is equivalent to one Jupiter mass around a Solar type star. This is large enough to undergo gap formation and enter the Type II migration regime. Our disc extends over a radial domain of $0.3 \leq R/R_0 \leq 6.0$ and azimuthal domain $-\pi \leq \phi \leq \pi$ with 285 by 596 cells respectively, using logarithmic cell spacing in the radial direction. This radial range was chosen to ensure no torque contribution from the outer disc was suppressed when calculating the net torque on the planet. The initial surface density gradient is $\Sigma_{\text{int}} \propto R^{-1/2}$ with $\Sigma_0 = 2.67 \times 10^{-3}$ at $R_0 = 1.0$ and we use a constant viscosity $\nu = 7.5 \times 10^{-6}$ corresponding to a dimensionless viscosity $\alpha = 3 \times 10^{-3}$ using the alpha prescription $\nu = \alpha c_s H$ of Shakura & Sunyaev (1973). We use a constant aspect ratio $h = H/R = 0.05$. Reflecting boundary conditions were used with wave killing zones in the regions $0.3 \leq R/R_0 \leq 0.585$ and $5.43 \leq R/R_0 \leq 6.0$ similarly to de Val-Borro et al. (2006), meaning that excited waves are damped before reaching the edge of the simulation. The planet is held on a fixed circular orbit with no migration or disc self-gravity accounted for. The migration of the planet is predicted from the net torque on the planet, which is determined by allowing the simulation to run for 600 orbits, and averaging the net torque on the planet over the last 100 orbits. Gap edge heating from the star is represented by a Gaussian in the sound speed profile of the disc, centred at the outer edge of the planet gap. Deviation from a purely adiabatic simulation is a result of the applied thermal relaxation function, Equation 4.5.

4.4 Justification of Gaussian Parameters

While the parameters pertaining to the disc and the planet have been discussed in Section 4.3, in this section we will justify the selection of the Gaussian parameters used in our simulations. Here we must reiterate what the presence of the Gaussian in the sound speed profile is intending to represent, the heating of the gap edge due to the incident radiation from the star. The important parameters are all contained within Equation 4.4, namely the amplitude of the Gaussian, A, the location of the Gaussian peak R_G and the full width at half maximum of the Gaussian, $\mathcal{W}$, from which σ is determined.

The easiest to identify of these parameters is R_G, as this will obviously occur at

the gap edge and can be roughly determined from running a very basic simulation of a disc plus a planet (no Gaussian applied yet). The specific location of the gap edge, however, is more difficult to determine, as its exact definition would be dependent on the point at which the disc material becomes opaque to the stellar radiation. As exact calculations involving disc opacity are beyond the scope of this chapter, we instead consider the outer slope of the gap in the surface density profile, from which we decide that $R_G = 1.20 - 1.40$ are a reasonable range within which the disc is heated. We also find that the change in c_s due to the Gaussian does not alter the location of the gap edge, hence these values of R_G are always located on the gap edge.

As the Gaussian acts as a local increase in the sound speed profile by a factor A, in order to predict a rough value of A we must estimate the sound speed increase due to the heating of the gap edge. This means we must first determine how much the temperature of the gap edge increases due to stellar heating, and then how this corresponds to an increase in sound speed. Therefore, we must calculate the unperturbed equilibrium temperature of the disc, and then the increased temperature due to the gap. Using the results of Chiang & Goldreich (1997) for geometrically flat discs, we use the equation

$$T_{\rm e} \approx \left(\frac{2}{3\pi}\right)^{\frac{1}{4}} \left(\frac{R_*}{R}\right)^{\frac{3}{4}} T_*, \tag{4.13}$$

where $T_{\rm e}$ is the effective temperature of the disc and R_* and T_* are the stellar radius and temperature respectively.

To calculate the temperature of the heated gap edge we make the further assumption that there is no planet and no inner disc. This allows us to treat the gap edge as the inner edge of the disc from which we can apply the results of Dullemond et al. (2001). While clearly this is not the exact case we are investigating, the presence of an inner disc could be represented by an efficiency factor, blocking some of the radiation incident on the gap edge in our simplified case. This means that the case we present here represents the maximum possible gap edge irradiation for a given set of stellar parameters. Therefore, to calculate the temperature of the heated gap edge we can use

$$T_G \approx \left(\frac{L_*}{4\pi R_G^2 \sigma_{sb}}\right)^{\frac{1}{4}}, \tag{4.14}$$

where T_G is the effective temperature of the gap edge, L_* is the luminosity of the star and $\sigma_{\rm sb}$ is the Stefan-Boltzmann constant. In determining Equation 4.14 we

also make the very reasonable assumption that $R_G \gg H_G$, namely the scale height at the edge of the gap is much smaller than the radial location of the gap edge.

Now from these we have a ratio of temperatures, T_G/T_e which we need to convert into a ratio of sound speeds. The temperature and sound speed are related by

$$P = \frac{\Sigma R_\mathrm{m} T}{\mu} = c_s^2 \Sigma, \tag{4.15}$$

where P is the pressure, R_m is the molar gas constant and μ is the mean molecular weight. Therefore,

$$A = \frac{c_G}{c_\mathrm{e}} = \left(\frac{T_G}{T_\mathrm{e}}\right)^{\frac{1}{2}}, \tag{4.16}$$

so the ratio of sound speeds and therefore the parameter A, depends on the root of the ratio of the temperatures. Using Equation 4.16 we can see that the A has a weak dependence of R_G, such that $A \propto R_G^{1/8}$, however, this is strongly dependent on the assumed disc geometry. Hence, we estimate the value of A at $R_G = 1.0$AU for a number of different stellar classifications. We include a T-Tauri star in this range of stellar classifications, using the stellar parameters found in Dullemond et al. (2001) for our T-Tauri calculation.

This is presented in Table 4.1, but the general conclusion is that A ranges from around $1.8-2.5$ and increases as the star becomes smaller, cooler and less luminous. When we choose a value of A for our Gaussian, we can consider it the maximum irradiation from the star, in which case the values presented in Table 4.1 would be the A we select, or we could consider it as a reduced magnitude of irradiation from a star that would provide a greater value of A at maximum irradiation. For example for a T Tauri star the maximum gap edge irradiation from Table 4.1 is 2.18, while an $A = 2.00$ or 1.75 could still represent a T Tauri star at near maximum gap edge irradiation, or a different classification of star at maximum irradiation.

Recalling that this is evaluated at $R_G = 1.0$AU rather than the actual radial location of the gap edge, the exact location of heating depends greatly on the opacity of the disc material and the inner disc could act as an efficiency factor to block some of the incident radiation, we can only call this a rough estimate. Despite this it gives us a good range in the parameter A for us to explore.

The next parameter, $\mathcal{W}$, is more difficult to select. Theoretically $\mathcal{W}$ represents the redistribution of temperature from the primary heating location (R_G), however, this will be dependent once again on the opacity of the material and also the density of material, which will be modified by the presence of a gap opening planet. We select

Table 4.1: Values of the parameter A for different stellar classifications.

Classification	T_*/K	$R_*/R_\odot$	L_*/W	A
O	30000	6.60	1.2×10^{31}	1.87
B	20000	4.20	9.8×10^{29}	1.99
A	8750	1.60	5.2×10^{27}	2.24
F	6750	1.30	1.2×10^{27}	2.30
G (Sun)	5780	1.00	3.8×10^{26}	2.37
K	4500	0.83	9.8×10^{25}	2.43
M	3700	0.70	3.2×10^{25}	2.49
T-Tauri	3800	2.54	2.9×10^{26}	2.18

$\mathcal{W}$ based on the constraint discussed in Section 4.5, that a too narrow Gaussian leads to the excitement of instabilities. This sets a lower limit in our selection of $\mathcal{W}$. We explore a range of Gaussians with $\mathcal{W}$ above this, however, we know that if we make the Gaussian too wide the heating will spread into the corotation region and the inner wake of the planet. If the torque contribution from the outer wake of the planet is also reduced by the Gaussian then the overall impact on the net torque due to the gap edge heating will be reduced. Hence, we believe this will act to reduce the parameter space for which we can successfully slow Type II migration. The region covered by the Gaussian corresponds to a 'puffed up' gap edge in the disc and hence regions behind this will be self-shadowed from stellar radiation. As a result the temperature in these regions will be reduced. This effect is an extra complication, in addition to not knowing the exact heating profile.

4.5 Instability Study

Implementing a Gaussian distribution in the sound speed profile of the disc often results in the excitation of Rayleigh or Rossby Wave instabilities, and subsequent destabilisation of the Gaussian, resulting in a very different shape from the input equilibrium profile. We study the parameter space for which an input Gaussian remains stable and use this to ensure stable Gaussians are implemented when investigating the net torque on the planet. We do this by investigating the behaviour of a Gaussian in the sound speed profile of a disc with setup described by Sections 4.2 and 4.3, however, these simulations contain no planet. We investigate these setups by both simulating the time evolution of the disc and by numerically analysing instability criteria. There are two potential instabilities excited by the implementation of a Gaussian in the sound speed profile, these are the Rayleigh Instability and the Rossby Wave Instability (RWI), which are addressed in Section 4.5.1 and Section

4.5.2 respectively.

4.5.1 Rayleigh Instability

For a disc to be Rayleigh stable the Rayleigh stability condition,

$$\frac{d}{dR}\left(R^2\Omega\right) > 0, \tag{4.17}$$

must be satisfied (Chandrasekhar, 1961). This is satisfied for a disc containing no planet undergoing Keplerian rotation, however, modifying the sound speed profile opens up the possibility for this criterion to be violated. The angular velocity is given by

$$\Omega^2 = \Omega_K^2 + \frac{1}{R\Sigma}\frac{\partial P(c_s)}{\partial R}, \tag{4.18}$$

so clearly a change in the sound speed profile will impact the stability of the system. For large enough deviations from Keplerian angular velocity, Ω_K, we would expect to find the disc Rayleigh unstable. We test this stability criterion for the sound speed profile given in Equation 4.4, with $R_G = 1.20$, and find a range of Gaussians for which the Rayleigh criterion is satisfied. It is found that for a given A, a $\mathcal{W}$ exists beyond which Gaussians will be Rayleigh stable. This information is presented in the second column of Table 4.2.

Table 4.2: $\mathcal{W}$ limits for Rossby and Rayleigh stable Gaussians of height A.

	Rayleigh Stable	Rossby and Rayleigh Stable
A	$\mathcal{W}$	$\mathcal{W}$
1.50	0.20	0.25
2.00	0.30	0.40
2.15	0.35	0.45
2.25	0.35	0.50
2.50	0.40	0.60

Running simulations of discs with Gaussians of these parameters we find that the Gaussians still go unstable even when the Rayleigh criterion is satisfied, however, as expected an increase in $\mathcal{W}$ keeps the Gaussian stable for longer. This leads us to conclude that the Rayleigh Instability is not the only instability we must be concerned with.

4.5.2 Rossby Wave Instability

In addition to the Rayleigh criterion from Equation 4.17 a Gaussian must also satisfy the Rossby Wave criterion to be stable to the RWI. Similarly to the Rayleigh criterion there is a critical value of $\mathcal{W}$ for a given A beyond which the Gaussian will be Rossby stable. This is in all cases a larger $\mathcal{W}$ than that which satisfies the Rayleigh criterion, meaning that Gaussians with very small $\mathcal{W}$ below those listed in the second column of Table 4.2 will be both Rossby and Rayleigh unstable.

A local maximum or minimum in the radial profile of the key function,

$$\mathcal{L}(R) = \mathcal{F}S^{\frac{2}{\gamma}}, \tag{4.19}$$

is required for the RWI to occur (Lovelace et al., 1999). Here $\mathcal{F}$ is approximately the inverse of the potential vorticity and $S = P/\Sigma^{\gamma}$ is the entropy. As both potential vorticity and entropy have dependence on the sound speed, again we can see that modifying the sound speed profile will impact the Rossby stability of the system. In addition to the local maxima/minima, an instability only occurs if a threshold in the variation of the key function is reached and exceeded. This threshold is dependent on the input parameters of the simulation, such as the surface density profile of the disc (Lovelace et al., 1999; Ono et al., 2016). We can study the Rossby stability of our Gaussians simply by simulating their time evolution in a disc and observing the results in the two dimensional sound speed distribution, the azimuthally averaged sound speed distribution and the radial velocity distribution as the presence of the RWI is clearly visible from these. Hence, we can easily find a $\mathcal{W}$ for a given A at which the Gaussian becomes Rossby stable, as at this point there will be no non-axisymmetric structure in these distributions and the sound speed profile will be unchanged from the initial condition to the final output. This information is presented in the third column of Table 4.2.

4.5.3 Instabilities excited by the presence of the planet

Even ensuring that the Gaussians themselves are stable in a disc containing no planet, we find that it is very difficult to stop them from going unstable once a planet has been introduced. Keeping our Gaussians in the Rossby stable regime is still important as it confirms that any instability present is the result of excitation by the planet, and not an intrinsic property of the Gaussian itself. The magnitude of the excited instability can greatly vary with the parameters of the Gaussian, similarly to the Rossby and Rayleigh instabilities, and in some cases, such as for very low A,

the instability is almost non-existent or is not even excited. Unfortunately it seems impossible to completely remove these instabilities from our simulations. These instabilities, in addition to the presence of the planet, act to perturb the Gaussian profile such that the equilibrium profile is different from the initial profile. However, as long as the instabilities are weak the difference between these profiles is very small. This can be seen in Figure 4.3. Accepting that these instabilities will always be present, the problem now is do they impact the net torque on the planet? This is investigated in more depth when we discuss the variation of $\mathcal{W}$ in Section 4.6.2, but we find that the average net torque appears to be unaffected by the presence of the instabilities. However, these instabilities do impact the amplitude of the net torque variation, meaning that particularly violent instabilities produce a much greater amplitude. This can increase the error in the net torque averaged over the last 100 orbits and can also impact the shape of the net torque distribution over these orbits, resulting in a state that is further from equilibrium than a more stable case. As a result we must be careful to not use Gaussians that excite violent instabilities, to preserve the validity of our average net torque result. An example of a result in which instabilities are excited but the net torque remains largely unchanged is given in Figure 4.4. Here we can see the variability in net torque is larger in the Gaussian case than in the no Gaussian case, however, the locally smoothed average is not impacted by this variability. It is important to note that the strength of the instability decreases with time, such that many of the instabilities that are weak at 500 orbits are much stronger at 100 orbits.

Finally we decide to average the net torque from $500 - 600$ orbits as it is clear that the simulation has reached an approximate equilibrium. To ensure this is a valid assumption we have run a number of our simulations to 3000 orbits (including that shown in Figure 4.4) in order to check the behaviour of the torque over large timescales and we find no significant deviation from the approximate equilibrium.

Between our findings that 500 orbits is a good approximation to equilibrium and that the instabilities do not have a great impact on the net torque, we feel confident that averaging the net torque from $500 - 600$ orbits provides an accurate description of the overall net torque.

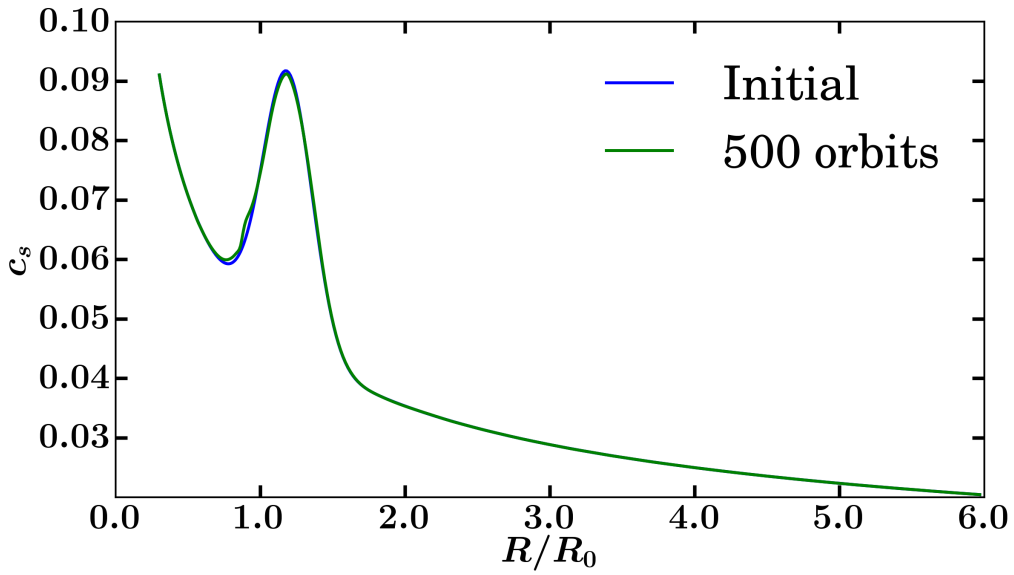

Figure 4.3: An example showing two Gaussians in the sound speed distribution of a disc containing a planet, 500 orbits apart. One Gaussian is the initial Gaussian, while the other shows the same Gaussian after 500 orbits have elapsed. The disc setup is described in Section 4.3 and the Gaussian parameters are described in Section 4.6.1. Here we can see that there is some change in the Gaussian across these 500 orbits, and while it is only a small difference, it is not unnoticable.

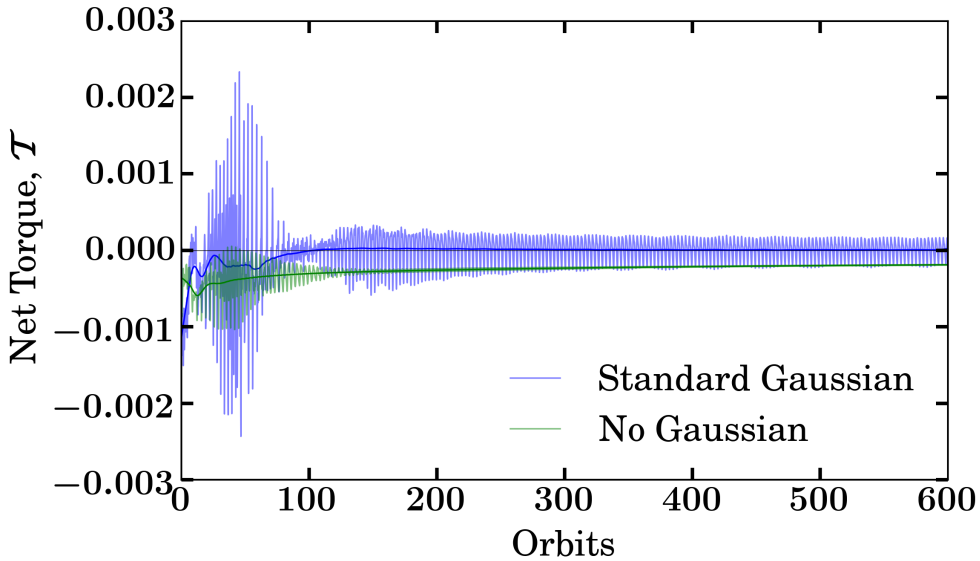

Figure 4.4: Net torque on the planet for a disc containing a Gaussian in the sound speed profile and a disc containing no Gaussian in the sound speed profile. This is an example of how instabilities excited by the presence of the planet and Gaussian affect the net torque on the planet, for a Gaussian setup described in more detail in Section 4.6.1. Clearly we can see the system reaches an approximate equilibrium, from which averaging over the last 100 orbits provides us with a good description of the net torque. The bold lines represent the locally smoothed average of the net torque.

4.6 Results

4.6.1 Reversing the Direction of Migration

Using the setup described in Section 4.3 and results of Section 4.4 as a guideline for parameter space, we investigated the net torque on a gap opening planet with a number of different Gaussians in the sound speed profile acting as stellar heating at the gap edge. We aimed to investigate whether this could change the sign of the net torque on the planet, as a negative net torque would result in inwards migration, and a positive net torque would result in outwards migration. We found that this method of heating can reduce the rate at which the planet migrates inwards, and can even change the sign of the net torque on the planet and produce outwards migration, as can be seen in Figure 4.4. Not only that, but we found the net torque on the planet can be positive even for the limited parameter space provided from Section 4.4, in which the Gaussians are stable and the amplitude of the heating is in good agreement with the expected heating from a number of classifications of stars, assuming the gap edge is exposed to maximum or near maximum stellar irradiation. From this we settle on a standard case in which the direction of migration is reversed, for which $A = 2.0$, $\mathcal{W} = 0.40$ and $R_G = 1.20$. This returns a net torque of $\mathcal{T} = 7.395 \times 10^{-6}$, compared to the net torque for a planet and no Gaussian heating, $\mathcal{T} = -1.939 \times 10^{-4}$. This positive torque is two orders of magnitude smaller than the normal case for a planet and no Gaussian, hence we can deduce that although the direction of migration has been reversed, the rate of migration will be significantly slower. Figure 4.5 clearly shows the impact of this Gaussian in reducing the net torque contribution from the planet's outer wake. Using this setup as a standard case, we can now investigate the sensitivity of the net torque to the parameters of the Gaussian.

4.6.2 Torque Dependence on $\mathcal{W}$

The first parameter we investigate is $\mathcal{W}$. As this parameter is heavily constrained by the onset of instabilities for $\mathcal{W} < 0.40$, this gives us little parameter space to work with. As a result we still investigate the impact on the torque for $\mathcal{W} < 0.40$, but consider that they are unstable and as such the resultant Gaussian is very likely to be different to that which we describe. The reasoning is to get a larger picture of how $\mathcal{W}$ affects the net torque, however, outwards migration from unstable Gaussians will not be regarded as a viable solution to reversing the direction of migration.

The net torques for different $\mathcal{W}$ are listed in Table 4.3. From this we can see that net torque does not have a simple dependence on $\mathcal{W}$, rather the net torque peaks for

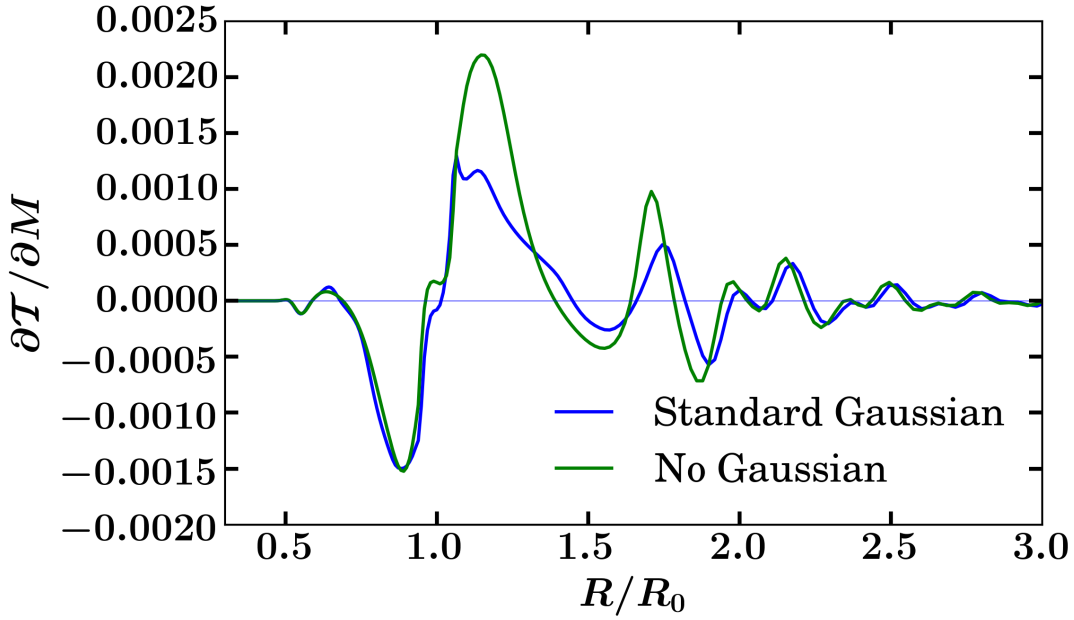

Figure 4.5: Azimuthally averaged torque profiles for a disc containing a Gaussian in the sound speed profile and a disc containing no Gaussian in the sound speed profile. This Gaussian's parameters are discussed in detail in Section 4.6.1. Torque is calculated as described in Section 4.2.3 and averaged over 100 orbits after the simulation reaches approximate equilibrium at orbit 500. This clearly shows the reduction in the net torque contribution from the planet's outer wake due to the irradiation of the outer gap edge.

a certain $\mathcal{W}$, and increasing $\mathcal{W}$ beyond this reduces the net torque. Fortunately for this setup the net torque remains positive for large $\mathcal{W}$, in which the Gaussians are stable, however, it is clear that increasing $\mathcal{W}$ further will continue to reduce the net torque. This is likely due to the Gaussian now infringing on the inner wake excited by the planet, which also contributes to the net torque on the planet. In a similar regard having $\mathcal{W}$ too low also lowers the net torque on the planet, which is likely due to the Gaussian only acting on a small radial strip of the disc, meaning that a large portion of the torque profile is unaffected by its presence. It is also important to reiterate here that for unstable Gaussians, especially those in the Rayleigh and Rossby unstable regime, it is unclear how much the instability affects the net torque and as such that could be producing the interesting $\mathcal{W}$ torque dependence we see here. Despite this Table 4.3 shows us that, at least in the Rayleigh stable regime, increasing $\mathcal{W}$ lowers the value of the net torque, which is a very important conclusion as stable Gaussians demand a large $\mathcal{W}$.

Figure 4.6 also shows the results presented in Table 4.3. Here it is easier to see the

Table 4.3: Net Torque on the planet for stable/unstable Gaussians with different $\mathcal{W}$ at $A = 2.00$, $R_G = 1.20$.

$\mathcal{W}$	$\mathcal{T}$	Stability
0.10	1.653×10^{-5}	Rayleigh and Rossby unstable
0.20	3.712×10^{-5}	Rayleigh and Rossby unstable
0.30	4.115×10^{-5}	Rossby unstable
0.35	2.585×10^{-5}	Rossby unstable
0.40	7.395×10^{-6}	Stable
0.45	-1.730×10^{-5}	Stable
0.50	-3.942×10^{-5}	Stable

impact of changing $\mathcal{W}$ on the net torque and also the behaviour of the torque as it crosses the regions of Rayleigh and Rossby stability. From this we can see that the torque increases with $\mathcal{W}$ in the Rayleigh unstable regime, but outside this regime it falls with increasing $\mathcal{W}$. The most important thing to note from Figure 4.6 is that the trend in the net torque does not change as $\mathcal{W}$ crosses the Rossby stability limit, hence leading us to believe that the RWI does not influence the net torque on the planet.

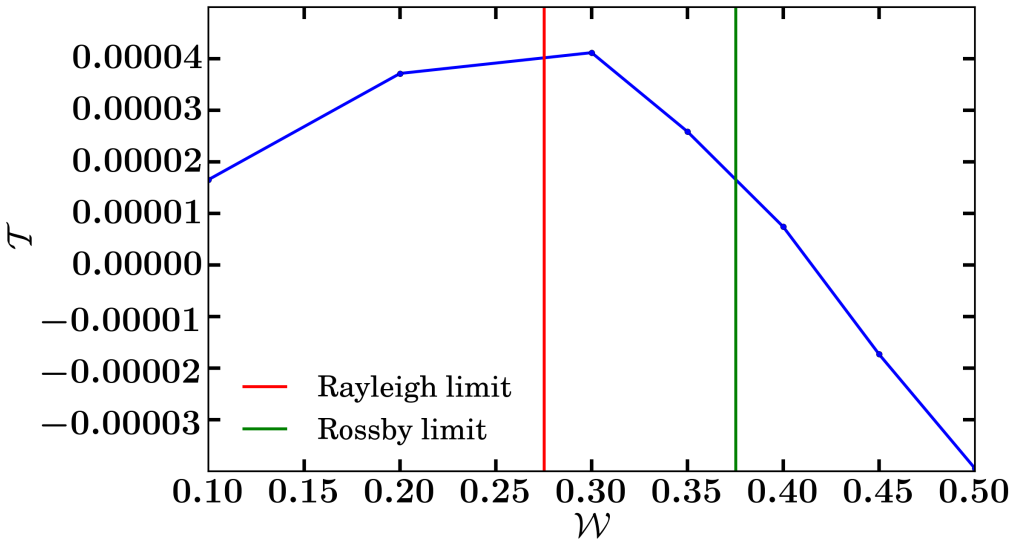

Figure 4.6: The change in net torque on the planet as $\mathcal{W}$ is increased. Here we can see that the increase in $\mathcal{W}$ causes an increase in torque in the Rayleigh unstable regime and a decrease in torque in the Rayleigh stable regime. Also it is apparent that the excitement of the RWI does not influence the net torque on the planet as the trend is unaffected by the crossing of the Rossby stability limit. The Rossby and Rayleigh stability limits shown here are at an estimate of the $\mathcal{W}$ at which they occur, as we have only investigated their location to a $\mathcal{W}$ resolution of 0.05.

4.6.3 Torque Dependence on A

We investigate modifying the parameter A for the range $A = 1.5 - 2.5$, consistent with our findings from Section 4.4. Here the value of A could represent the heating of the gap edge at maximum irradiation from the star or a reduced irradiation from a star that at maximum would give a higher A value. Reasonable A values for a variety of stellar classifications are given in Table 4.1. However, we already know from Sections 4.4 and 4.5 that $\mathcal{W}$ is closely linked to the A of the Gaussian, to ensure we do not enter Rossby or Rayleigh unstable regimes. Therefore, in a deviation from modifying only one parameter, we must also increase $\mathcal{W}$ to the appropriate value for stability for each value of A. These $\mathcal{W}$ are given in Table 4.2.

We find that modifying the $\mathcal{W}$ and A of the Gaussian each have the opposite effect on the net torque on the planet. Section 4.6.2 already shows that outside of the Rayleigh unstable regime increasing $\mathcal{W}$ acts to move the net torque on the planet towards the negative. We find that increasing A acts to move the net torque on the planet towards the positive and hence the effect of modifying both of these parameters compete with each other. However, $\mathcal{W}$ must be increased so much for larger values of A that the effect of $\mathcal{W}$ on the net torque is stronger than the effect of the parameter A. This results in stable Gaussians with a larger A actually making the net torque on the planet more negative and being unable to reverse the direction of migration. We then find that, at $R_G = 1.20$, there is a 'sweet spot' where the effect of increasing A on the torque overcomes the effect of increasing $\mathcal{W}$, and we get outwards migration. This information is presented in Table 4.4, which shows that in the range $1.75 \leq A \leq 2.15$ we can return outwards migration for stable Gaussians.

Table 4.4: Net Torque on the planet for stable Gaussians with different values of A at $R_G = 1.20$.

A	$\mathcal{W}$	$\mathcal{T}$
1.50	0.25	-2.742×10^{-5}
1.75	0.35	1.022×10^{-6}
2.00	0.40	7.395×10^{-6}
2.15	0.45	2.775×10^{-6}
2.25	0.50	-1.149×10^{-5}
2.50	0.60	-2.499×10^{-5}

4.6.4 Torque Dependence on R_G

So far all of the variations made to parameters of the standard Gaussian have held the location of the Gaussian constant, $R_G = 1.20$. We now move the Gaussian radially outwards across the disc to $R_G = 1.30 - 1.40$, still located on the gap edge but corresponding to a lower opacity disc, as radiation from the star now penetrates further into the gap edge. The immediate change that results from moving the Gaussian is that the $\mathcal{W}$ required to stabilise the Gaussian increases. This is because the Gaussian now resides at a location with a larger aspect ratio. Therefore, we can no longer use the results presented in Table 4.2 and instead recalculate the region for which the Gaussians are stable. This information is presented in Table 4.5.

Table 4.5: $\mathcal{W}$ limits for Rossby and Rayleigh stable Gaussians of height A.

		Rayleigh Stable	Rossby and Rayleigh Stable
R_G	A	$\mathcal{W}$	$\mathcal{W}$
1.30	1.50	0.20	0.30
1.30	1.75	0.30	0.35
1.30	2.00	0.35	0.45
1.30	2.15	0.35	0.50
1.30	2.25	0.40	0.55
1.30	2.50	0.45	0.65
1.35	1.50	0.20	0.30
1.35	1.75	0.30	0.35
1.35	2.00	0.35	0.45
1.40	1.50	0.25	0.30
1.40	1.75	0.30	0.40
1.40	2.00	0.35	0.45
1.40	2.25	0.40	0.55

A concern with further increasing $\mathcal{W}$ was outlined in Section 4.4, namely that the wider the Gaussian the more the heating infringes onto the gap region and subsequently could reduce the torque contribution from the inner wake. In this case it appears that the increase in $\mathcal{W}$ is significantly smaller than the distance the Gaussian was moved, $\Delta R_G = 0.10 - 0.20$, hence the inner wake will be unaffected by the increase in $\mathcal{W}$. We now use these $\mathcal{W}$ as the limit for stable Gaussians, similarly to the way we used $\mathcal{W}$ in Table 4.2 as the limit for stable Gaussians at $R_G = 1.20$. With this in mind we replicate Table 4.4 but for stable Gaussians at $R_G = 1.30 - 1.40$. This information is presented in Table 4.6. We also present a number of these results in Figure 4.7, which shows the smoothed net torque for each simulation, similar to that shown in Figure 4.4. The deviation in the last < 10

orbits in each curve is an artefact of the smoothing mechanism we use, and should be disregarded.

Table 4.6: Net Torque on the planet for stable Gaussians with different input parameters.

R_G	A	$\mathcal{W}$	$\mathcal{T}$
1.30	1.50	0.30	-4.395×10^{-5}
1.30	1.75	0.35	3.123×10^{-5}
1.30	1.75	0.40	2.432×10^{-5}
1.30	2.00	0.45	5.270×10^{-5}
1.30	2.00	0.50	3.431×10^{-5}
1.30	2.15	0.50	4.845×10^{-5}
1.30	2.25	0.55	3.378×10^{-5}
1.30	2.50	0.65	1.449×10^{-5}
1.35	1.50	0.30	-9.006×10^{-5}
1.35	1.75	0.35	4.177×10^{-5}
1.35	1.75	0.40	-4.802×10^{-7}
1.35	2.00	0.45	6.710×10^{-5}
1.35	2.00	0.50	4.908×10^{-5}
1.40	1.50	0.30	-1.323×10^{-4}
1.40	1.75	0.40	-4.661×10^{-5}
1.40	2.00	0.45	6.008×10^{-5}
1.40	2.00	0.50	3.178×10^{-5}
1.40	2.25	0.55	3.870×10^{-5}
1.40	2.25	0.60	6.140×10^{-5}

The results shown in Table 4.6 are very interesting as they show that there is much larger potential for outward migration for heating at these larger radii. The net torque at these locations is also often an order of magnitude larger than at $R_G = 1.20$, meaning that the error in the net torque calculation will be small compared to the actual value and hence we can be confident in our result. This increase in magnitude of net torque also means that outward migration will be more rapid for heating at this location, or in the case of a negative torque, inwards migration will be slower.

We briefly investigated modifying $\mathcal{W}$ while keeping A and R_G constant, in a similar manner to Section 4.6.2. It was found that the net torque behaves in the same way as it does at $R_G = 1.20$ and increasing the $\mathcal{W}$ of the Gaussian too much begins to reduce the net torque. This was not unexpected, as only the location of the Gaussian has changed, but the overall more positive net torque at $R_G = 1.30-1.40$ means that $\mathcal{W}$ can be increased more than it can at $R_G = 1.20$ while still retaining outwards migration. This is another example of the increased parameter space available at

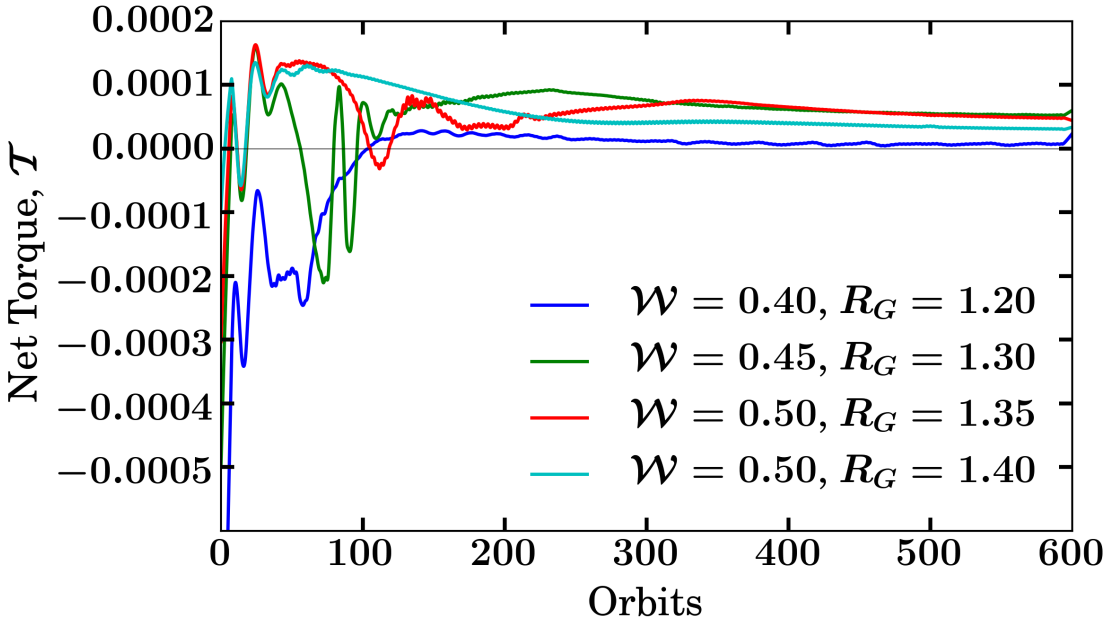

Figure 4.7: Smoothed net torque curves for a number of the simulations presented in Table 4.6. This shows the improvement made in returning a positive net torque when we increase the radius at which the Gaussian in the sound speed profile is centred. This increased net torque corresponds to more parameter space for outward migration at these larger radii.

this location for stable Gaussians with outward migration, which gives us greater confidence in saying that outwards migration could potentially occur here.

4.6.5 Exploring Asymmetric Distributions

As discussed in Section 4.4 we intend our Gaussian distribution to only reduce the net torque contribution from the planet's outer wake. From our results in Section 4.6.2 we can see that should $\mathcal{W}$ be increased too much the net torque from the planet's inner wake is also reduced. Therefore, in order to investigate the effect of increasing $\mathcal{W}$ without allowing the Gaussian to further reduce the torque contribution from the planet's inner wake, we modify our description of the Gaussian. We name our new distribution an 'Asymmetric Gaussian distribution' as our heating method remains based around a Gaussian of height A at radius R_G, however, we now have two different $\mathcal{W}$'s, one for the inner half of the Gaussian ($\mathcal{W}_{\text{in}}$, heating the gap edge and gap region) and one for the outer half of the Gaussian ($\mathcal{W}_{\text{out}}$, heating the gap edge and outer disc). Therefore, we can now hold $\mathcal{W}_{\text{in}}$ constant at the smallest stable Gaussian $\mathcal{W}$ while extending $\mathcal{W}_{\text{out}}$ further into the outer disc. The results of running a number of asymmetric Gaussians with different initial parameters are

presented in Table 4.7.

Table 4.7: Net Torque on the planet for stable asymmetric Gaussians with different input parameters.

R_G	A	$\mathcal{W}_{\text{in}}$	$\mathcal{W}_{\text{out}}$	$\mathcal{T}$
1.20	1.75	0.35	0.35	1.022×10^{-6}
1.20	1.75	0.35	0.45	3.527×10^{-6}
1.20	2.00	0.40	0.40	7.395×10^{-6}
1.20	2.00	0.40	0.50	1.205×10^{-5}
1.20	2.00	0.40	0.60	1.430×10^{-5}
1.30	1.75	0.35	0.35	3.123×10^{-5}
1.30	1.75	0.35	0.45	2.227×10^{-5}
1.30	1.75	0.40	0.40	2.432×10^{-5}
1.30	1.75	0.40	0.50	1.488×10^{-5}
1.30	2.00	0.45	0.45	5.270×10^{-5}
1.30	2.00	0.45	0.55	5.147×10^{-5}
1.30	2.00	0.45	0.65	5.270×10^{-5}
1.30	2.00	0.50	0.50	3.431×10^{-5}
1.30	2.00	0.50	0.60	3.471×10^{-5}
1.30	2.00	0.50	0.70	3.506×10^{-5}
1.35	1.75	0.35	0.35	4.177×10^{-5}
1.35	1.75	0.35	0.45	-1.412×10^{-5}
1.35	1.75	0.40	0.40	-4.802×10^{-7}
1.35	1.75	0.40	0.50	-6.146×10^{-6}
1.35	2.00	0.45	0.45	6.710×10^{-5}
1.35	2.00	0.45	0.55	6.552×10^{-5}
1.35	2.00	0.50	0.50	4.908×10^{-5}
1.35	2.00	0.50	0.60	5.323×10^{-5}
1.40	1.75	0.40	0.40	-4.661×10^{-5}
1.40	1.75	0.40	0.50	-4.709×10^{-5}
1.40	2.00	0.45	0.45	6.008×10^{-5}
1.40	2.00	0.45	0.55	2.183×10^{-5}
1.40	2.00	0.50	0.50	3.178×10^{-5}
1.40	2.00	0.50	0.60	2.416×10^{-5}

From Table 4.7 we can see there is minimal difference in the net torque when we increase $\mathcal{W}_{\text{out}}$. This is especially true for the largest radii results, at $R_G = 1.40$, however, for the smallest radii $R_G = 1.20$ the net torque increases by a significant amount. This shows that at $R_G = 1.20$ the net torque contribution from the outer wake is not fully covered by the Gaussian, whereas at larger radii the Gaussian already covers the wake, and the Gaussian now encroaches on the less perturbed outer disc region. This result further increases the parameter space for which outward migration is allowed, as although torque is often reduced by extending $\mathcal{W}_{\text{out}}$ there is

still room to extend $\mathcal{W}_{\rm out}$ before the torque becomes negative again. Also extending $\mathcal{W}_{\rm out}$ appears to increase the net torque at $R_G = 1.20$, which could potentially solve the problem of the limited parameter space for outward migration at this location.

4.7 Discussion

In this chapter we have found that it is possible to reduce the magnitude of a planet's inwards Type II migration rate by heating the outer gap edge due to stellar irradiation. We have also found for a range of parameters that it is possible to reverse the direction of migration of the planet by causing the disc to exert a positive net torque on the planet, if the gap edge irradiation is the maximum or near maximum possible from the star. We represent the irradiation of the outer gap edge by initialising a Gaussian distribution in the sound speed profile of the disc at the gap edge, which reduces the contribution of the planet's outer wake on the net torque on the planet. We then measure the resultant net torque on the planet and use this to infer the relative rate and direction of migration.

We make a number of assumptions and approximations in our simulations which allow us to investigate a simplified model of the system we describe. Primarily, we are investigating a three dimensional problem using two dimensional simulations. This is a result of the increased computational time required for three dimensional simulations meaning it would be impossible to perform a parameter study of this magnitude within a reasonable time frame. Therefore, we present these results as an exploration into the effect of heating the outer gap edge and should be followed up in three dimensions. We also use a very simplified model for the heating of the gap edge, namely an increase in the sound speed profile of the disc at this location following a Gaussian distribution. Realistically the three dimensional location and radial dependence of this heating would be a function of the disc's opacity, which is not considered here. The magnitude of our heating of the outer gap edge is also calculated ignoring the presence of the planet and the inner disc (see Section 4.4). Therefore, we present our results at maximum to near maximum possible gap edge irradiation from the star, however, the inner disc and planet should be accounted for in more realistic estimates of the gap edge heating. As such, the heating of the outer gap edge should also be explored numerically using three dimensional radiative transfer simulations.

An important concept in our work is the reduction of the net torque contribution from the planet's outer wake, while leaving the inner wake unchanged. This is not

necessarily always the case in our simulations. We expect some torque difference at the inner wake due to the presence of the Gaussian in the sound speed profile, an unfortunate consequence of the required $\mathcal{W}$ to avoid the Gaussians becoming inherently unstable. However, we find even in cases in which $\mathcal{W}_{\rm in}$ of an asymmetric Gaussian is held constant and $\mathcal{W}_{\rm out}$ is modified the torque contribution from the inner wake can still change, albeit a small amount. This is because the torque does not depend solely on the sound speed, it also has a surface density dependence. As the gap profile is modified by the change in sound speed, this can have an impact on the torque across the gap region. It is also known that the gap edge steepness, gap width and gap depth can all impact the torque (Petrovich & Rafikov, 2012). Overall, the change in the torque at the inner wake is negligible, especially compared to the change in the torque at the outer wake, therefore this is largely irrelevant to our results.

The selected $\mathcal{W}$s of our Gaussians are also an important point for discussion in our work. The reasons for selecting the values we do has already been discussed in Sections 4.4 and 4.5, however, here we shall briefly mention further difficulties surrounding our selection of $\mathcal{W}$. The $\mathcal{W}$ we have selected are entirely based on stability arguments from Section 4.5. Hence, we keep our $\mathcal{W}$ close to the stability cut off. However, $\mathcal{W}$ realistically would represent a mixture of the absorption of stellar radiation by the disc at locations outside the peak in absorption efficiency (R_G) and the self-heating of the disc by re-emission of radiation from the location of maximum absorption efficiency. This means that $\mathcal{W}$ would be closely linked to the opacity of the disc and its rate of heat transport. By selecting a $\mathcal{W}$, and an R_G, we are indirectly making statements about these parameters. Of course these would depend on more than just these Gaussian parameters, such as the composition of the disc and the gas to dust ratio, however, these are non-trivial to gain exact values for and so we do not directly consider them in this work. Taking these into account will likely modify the range of parameter space for which the Gaussian is realistic heating for a disc and therefore further constrain the outwards migration regime.

We make further simplifying assumptions within the model we have outlined above, which we do not believe will affect the overall conclusion, that outwards migration is possible, however, will undoubtedly affect the exact numerical results. We use a non-flaring disc, while claiming that, theoretically, the flaring of the disc is what allows the stellar radiation to impact the gap edge and thus the heating to occur. This is a simplifying assumption that may affect the structure of the outer disc and the heat transport outwards, but will largely not affect the net torque on the planet.

We consider a simplified prediction of the magnitude of heating from a number of stellar sources provided in Table 4.1. These predictions agree with the A parameters we use within our simulations, with parameter space large enough that should a more complex heating model provide slightly different heating magnitudes (most likely lower) we would still expect outwards migration in a number of cases. As stated above we neglect the presence of the inner disc and planet when determining the magnitude of heating of the gap edge, such that we can follow the example of Dullemond et al. (2001). Realistically the inner disc would block some of the radiation from the star and hence this would likely further reduce the magnitude of heating. This effect would be difficult to quantify in this simple approximation, as the planet impacts both the inner and outer disc structure, hence radiative transfer simulations would be required to adequately investigate. In addition, it has been shown that reflection, absorption and re-radiation from the outer gap edge can cause the inner gap edge to also become heated, however, to a lesser extent (Turner et al., 2012). This can cause a reduction in the one sided net torque at this location, and a 'puffing up' of the inner gap edge, blocking more of the stellar radiation that otherwise would be incident on the outer gap edge. If the heating is significant, this could affect our results by causing the net torque on the planet to become more negative.

We address the usage of a Gaussian for the shape of the heating distribution in Section 4.6.5, and find that the result changes minimally in most cases for an asymmetric heating distribution. In fact in some cases we find this improves the likelihood of outwards migration, rather than reducing the net torque on the planet. This shows that our results are not specific to Gaussian distributions.

Our Gaussian heating setup does not capture the effect of shadowing the outer disc by the 'puffed up' heated gap edge. This would reduce the temperature in the outer disc and as a result move the net torque on the planet towards the negative. As has been shown in Section 4.6.2 we require a few scale heights of heating to reduce the torque contribution from the planet's outer wake, so if the cooling effect from self-shadowing does not allow for this then this could cause the net torque on the planet to become more negative. The results of Jang-Condell & Turner (2013) imply that gap edge heating may occur over a region a few scale heights wide.

We heat the location of the gap edge before the gap has formed, as our planet is stationary and we know the location at which the gap will occur. This may cause non-physical effects at early times, as the gap edge cannot be visible to the star if it has not formed yet, but we are confident that by 500 orbits the simulation has reached equilibrium and the result is independent of early time effects.

We use only stable Gaussians in our simulations, however, it is clear that the presence of the planet excites instabilities in the Gaussian which is unavoidable. We are confident the instabilities have no effect on the net torque once we reach an approximate equilibrium, as while there is short term variability in the net torque on the planet the magnitude is approximately constant such that the average over 100 orbits is a good indicator of the overall net torque on the planet. We are also confident that the RWI has no effect on the net torque on the planet, as shown in Figure 4.6. Despite arguing that these instabilities have little to no effect on the overall net torque and that instabilities are unavoidable in simulations containing planets and Gaussians, we endeavour to avoid inputting instabilities into our simulations where we can control their excitement. Hence, we do not thoroughly investigate the net torque on the planet for the regime in which the input Gaussian is Rossby unstable, despite Section 4.6.2 showing us that this region can return outwards migration. From our results, it is not unbelievable that should a system similar to those we describe exist in nature, it could be unstable. In this case Rossby instabilities excited due to the gap edge irradiation could potentially occur and so this region would be important. From our results in Section 4.6.2 we believe this would only further open up parameter space for which outward migration is a possibility.

As the planet is stationary in the disc and our measure of migration comes from the net torque on the planet, the gap edge and therefore the Gaussian representing heating, is also stationary. Realistically in Type II migration the gap edge would move with the planet as it migrates, and this would change the magnitude of heating on the gap edge. Hence, the predicted net torque on the planet will not be constant as the planet moves and the shape of the input Gaussian should change as the gap edge moves. This means it is very possible that a planet will undergo both inwards and outwards migration as the system evolves, however, investigating this is beyond the scope of this chapter.

We use a fixed value for the thermal relaxation timescale, $\tau = 2\pi$, in our simulations. This acts as a compromise between an adiabatic and an isothermal simulation, for which $\tau = 0$. This is because a longer thermal relaxation timescale will cause the simulation to damp back slower to the initial Gaussian in the sound speed profile. Hence, the Gaussian heating acting on the disc will be less accurate to the initial distribution for larger τ. Richert et al. (2015) have shown that for inefficient cooling timescales a planet in a two dimensional adiabatic disc can excite instabilities that significantly alter the disc structure. In some cases it is found that a massive planet will not even open a gap. Therefore, we select our thermal relaxation timescale to be much shorter than the gap formation timescale. Despite this, during our inves-

tigation we briefly experimented with increasing the thermal relaxation timescale, for the standard case of a stable Gaussian we describe in Section 4.6.1. We found that increasing the thermal relaxation timescale to $\tau = 10\pi$ and $\tau = 20\pi$ greatly deformed the shape of the Gaussian and excited instabilities in the disc similar to those described in Section 4.5, even before a planet was introduced. From this and the results of Richert et al. (2015), we decided that the thermal relaxation timescale was not a parameter to be investigated in our parameter study. It should be made clear though that in a realistic disc the local thermal relaxation timescale will vary with radius as, for example, in an optically thin outer disc heat will be redistributed faster than for an optically thick inner disc. As such, the results we present are valid for a thermal relaxation timescale of $\tau = 2\pi$ at the location of heating. Far from the location of heating the sound speed profile of the disc is unperturbed and as a result the shape of the profile is independent of the thermal relaxation timescale here, regardless of its realistic value.

As we hold our planets on a fixed circular orbit there is no accounting for the eccentricity evolution of the planet's orbit. It is known that the eccentricity of a giant planet is excited by its Lindblad resonances and damped by its non-coorbital corotation torques. Unless the gap formed is very large, the damping via corotation torques dominates and the eccentricity is not excited. However, it has been shown that the illumination of the outer gap edge by stellar irradiation can modify the planet's corotation torque to allow the eccentricity excitation via Lindblad resonances to dominate (Tsang, 2014; Tsang et al., 2014). Hence, this could change the planets orbital path from the circular case we have investigated, which could have implications on the net torque on the planet.

Our results act as proof that this method of modifying the classical Type II planetary migration may affect the migration of giant planets. Therefore, this method of reducing or reversing migration rate is important to consider when studying the evolution of planetary systems, either as a sole method of modifying planetary migration, or in tandem with other previously proposed methods, i.e. Coleman & Nelson (2016).

4.8 Conclusions

We have investigated the heating of the outer gap edge by stellar radiation using a Gaussian distribution in the sound speed profile of a disc. Using this method it is possible to lower the contribution to the net torque on the planet from the planet's

outer wake, while leaving the contribution from the inner wake mostly unchanged. The goal was to address the problem that observational evidence dictates that giant planets are more common at orbits $R > 1$AU (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016) and that for a gas giant planet to survive throughout the disc's lifetime it must begin Type II migration at an orbital radius $R > 20$AU (Coleman & Nelson, 2014). At this radius the core accretion timescale exceeds the lifetime of the disc, therefore there must be a process that limits Type II migration speed (Nelson et al., 2000). We found heating the gap edge via stellar radiation is a method of reducing the net torque on the planet and by extension the rate of inwards migration. We also found that for the extreme case of maximum to near maximum gap edge irradiation it is possible to return a positive net torque on the planet, which corresponds to outwards migration for the planet. We have investigated the impact on the net torque of modifying the following Gaussian parameters:

- A, the amplitude of the Gaussian.
- $\mathcal{W}$, the full width at half maximum of the Gaussian.
- R_G, the radial location of the Gaussian peak.
- The impact of asymmetric Gaussians.

We found that the range of applicable Gaussians was limited by $\mathcal{W}$ constraints to avoid Rossby and Rayleigh unstable regimes. Our test case of $R_G = 1.20$ was studied extensively, and was found to provide very low magnitude positive torque for a small range of A and $\mathcal{W}$, providing the potential for weak outward migration, or a large reduction in the rate of inwards migration. At larger radii, $R_G = 1.30 - 1.40$, we found that there is a significantly larger parameter space for which outward migration is possible, both in A and $\mathcal{W}$ and that the magnitude of the net torque is larger, so the planet is further from the edge of the outwards migration regime and has a higher rate of migration. We found that in general the height of the Gaussian must be at least $A = 1.75$ to achieve outwards migration and that increasing the $\mathcal{W}$ of the Gaussian lowers the net torque on the planet, a result of the Gaussian impacting the torque contribution from both the inner and outer wakes, rather than just the outer wake. We find that for a number of sample stellar classifications the predicted magnitude of heating at the gap edge corresponds to $A \approx 2.0$ at maximum or near maximum stellar irradiation with a weak radial dependence, $A \propto R_G^{1/8}$, which would make outwards migration a possibility. For more modest A, which could account

for the presence of an inner disc, re-radiation heating the inner gap edge and self-shadowing of the outer disc, we still find significant effects on the net torque on the planet, meaning that even if outwards migration is not a possibility there is still the potential for reduction in Type II migration rate. We also investigate asymmetric Gaussian distributions, in which $\mathcal{W}$ is greater on the outer Gaussian edge, and at the limit of stability on the inner Gaussian edge. We found that increasing $\mathcal{W}_{\rm out}$ can both increase the net torque on the planet at smaller radii ($R_G = 1.20$) and decrease the net torque at larger radii, such as the reduction at $R_G = 1.40$. This could potentially solve the problem of the small parameter space for outwards migration at $R_G = 1.20$, while the weak reduction in net torque for ranges $R_G = 1.30 - 1.35$ still leaves a large parameter space for outwards migration there.

Overall we believe our results show that there is a significant possibility for the irradiation of the gap edge to severely slow the rate of Type II planetary migration. The results from our simplified approach can act as a good starting point for future three dimensional radiation-hydrodynamical simulations, which will reveal how important this effect is for allowing gas giant planets to remain at larger orbital radii.

5 Constraining the masses of planets in protoplanetary discs from the presence or absence of vortices - Comparison with ALMA observations

Abstract

A massive planet in a protoplanetary disc will open a gap in the disc material. A steep gap edge can be hydrodynamically unstable, which results in the formation of vortices that can act as tracers for the presence of planets in observational results. However, in a viscous disc, the potential formation of these vortices is dependent on the timescale over which the massive planet accretes mass and with a sufficiently long timescale it is possible for no vortices to form. Hence, there is a connection between the presence of vortices and the growth timescale of the planet and it may therefore be possible to exclude a planetary interpretation of observed structure from the absence of vortices. We have investigated the effect of the planet growth timescale on vortex formation for a range of planet masses and viscosities and have found an approximate relation between the planet mass, viscosity and planet growth timescale for which vortices are not formed within the disc. We then interpret these results in the light of recent observations. We have also found that planets do not need to be close to a Jupiter mass to form vortices in the disc if these discs have low viscosity, as these can be caused by planets as small as a few Neptune masses.

5.1 Introduction

Planets form in protoplanetary discs via the accretion of dust and gas. These planets will excite density waves in the disc material that transport angular momentum from

the planet and deposit it in the disc, exerting a torque on the disc material. If the deposition of torque is close enough to the planet (Lin & Papaloizou, 1993; Bryden et al., 1999) and the resultant torque is stronger than the viscous diffusion of the disc material (Lin & Papaloizou, 1979; Goldreich & Tremaine, 1980; Takeuchi et al., 1996; Crida et al., 2006) then a low density region, a gap, is formed around the orbital location of the planet. These two criteria are known as the thermal and viscous criteria for gap opening.

Observational results over the past few years (e.g. ALMA Partnership et al., 2015; Andrews et al., 2016; van der Marel et al., 2019) have provided high resolution images and information about the features that exist across the lifetime of protoplanetary discs. One important feature we can see from these images is the presence of dark bands in the dust emission of the disc. A clear example of this can be seen in images of the protoplanetary disc HL Tau (ALMA Partnership et al., 2015). While we do not currently have a definitive answer as to what causes these, there is a possibility they could be planetary in origin, gaps formed by massive planets within the disc. Unfortunately, even assuming these dark bands are the result of planet-disc interaction, there is a degeneracy in the planet masses and disc viscosities that can adequately explain them (e.g. Mulders et al., 2013). If both the disc viscosity and the planet mass are unknown, almost any gap can be explained by a planet. Hence, we would benefit greatly from a method that would allow us to rule out a planet of a given mass from having formed a gap.

When a planet opens a gap in the disc, there is the potential for the excitement of instabilities in the disc material. The steep edge of a gap carved by a planet is prone to hydrodynamic instabilities such as the Rossby Wave Instability (RWI). The non-linear evolution of this instability can result in the formation of vortices at the gap edge (Lovelace et al., 1999; Li et al., 2000).

Vortices are potential explanations for brightness asymmetries seen at the edge of a gap or cavity in observational results (van der Marel et al., 2013; Fukagawa et al., 2013; Pérez et al., 2014; Marino et al., 2015; Kraus et al., 2017; Dong et al., 2018; Cazzoletti et al., 2018). Numerical simulations show that indeed massive planets can excite the RWI (de Val-Borro et al., 2006, 2007; Fu et al., 2014a). However, vortices can also be formed by processes that do not involve a planet (e.g. Lovelace et al., 1999; Li et al., 2000; Klahr & Bodenheimer, 2003; Lyra & Klahr, 2011; Raettig et al., 2013; Lyra, 2014; Bae et al., 2015). Despite this, it is not unbelievable that the observed vortices in the discs were formed as a result of planets we cannot see. In many of the simulations in which these vortices occur the formation of the planet is not accounted for. In other words, the planet is already at its maximum

mass before it begins to open a gap in the disc. As the strength of the deposited torque is dependent on the mass of the planet, the gap will open slower for a planet growing in mass than for a planet at maximum mass, giving the disc time to adjust viscously. It has been shown by Hammer et al. (2017) that over sufficient growth timescales it is possible for a massive planet to open a gap without exciting the RWI and therefore without producing vortices. Hence, it may be possible to rule out a planet of a certain mass as the creator of the gap based on the presence or absence of vortices in the disc. Additionally, recent investigation into the shape of planet induced vortices have accounted for the timescale over which the planet grows and have found that vortices can become elongated (Hammer et al., 2017). As a result elongated vortices have their own specific observational features (Hammer et al., 2019) that have potentially been observed (Pérez et al., 2014; Cazzoletti et al., 2018). However, the majority of observational results do not show the features that imply elongated vortices, although this may be due to the resolution of the observations.

In this chapter we are investigating the required planet growth timescales for which no vortices are formed in a protoplanetary disc. We study this for a range of masses and viscosities, with the aim to find an approximate relation between these parameters. This approximate relation could then be used to estimate limits for parameters of observed protoplanetary discs, or to determine whether or not it is likely that a planet exists in these discs. This acts to constrain the mass of a planet that can be formed in a disc of a certain age and viscosity, based on the presence or absence of vortices in the disc. Therefore, we aim for our result to aid in breaking the degeneracy that exists in explaining observed disc structure using planets.

This chapter is arranged as follows. In Section 5.2, we discuss the relevant equations solved to simulate disc evolution. In Section 5.3, we discuss the code and numerical setup for our simulations. In Section 5.4, we present our results. In Section 5.5, we compare our findings to prior observational work. In Section 5.6 we discuss our results and any assumptions made in our simulations. Finally, in Section 5.7, we present our conclusions.

5.2 Basic Equations

In this chapter we perform a parameter study to find the growth timescales for which the RWI is not excited by gap formation in a protoplanetary disc. We investigate a wide range of viscosities, mass ratios and growth timescales and so perform this

study in two spatial dimensions, making the approximation of a locally isothermal disc. As the Hill sphere of the planet is larger than the scale height for gap opening, the disc appears two dimensional with respect to the planet. Hence, it is safe to assume that the approximation of two dimensions is sufficient to draw meaningful conclusions from.

5.2.1 Two dimensional protoplanetary disc

The continuity equation for the evolution of a protoplanetary disc's surface density, Σ, is given by

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0, \tag{5.1}$$

where $\mathbf{v}$ is the velocity field. We simulate the evolution of a protoplanetary disc's surface density due to the presence of a planet by solving the two dimensional Navier-Stokes equation for the motion of the disc,

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P - \nabla \cdot \mathbf{T} - \Sigma \nabla \Phi, \tag{5.2}$$

where $\mathbf{T}$ is the Newtonian viscous stress tensor, P is the pressure and Φ is the gravitational potential of the planet and star system. A cylindrical coordinate system is used, such that $\mathbf{v} = (v_R, R\Omega)$ where v_R and Ω are the radial and angular velocities at a given radius. We assume the disc is locally isothermal, with an equation of state $P = c_s^2 \Sigma$ where $c_s(R)$ is the sound speed at cylindrical radius R. We impose a constant aspect ratio $h = H/R$ where H is the disc scale height. The sound speed in the disc is given by

$$c_s = h \sqrt{\frac{GM_*}{R}}, \tag{5.3}$$

where M_* is the mass of the star. The planet is held on a fixed circular orbit at $R = R_0$, with angular velocity $\Omega = \Omega_0$ and orbital period P_0.

5.2.2 Growing the planet mass

In our simulations it is important that the planet mass is not constant, it should grow with time over a number of orbits. This is in an attempt to eliminate the excitation of instabilities and formation of vortices that otherwise would be present if the planet is initialised at its maximum mass. Therefore, we grow the planet mass over a timescale $t_{\rm G}$ for $t \leq t_{\rm G}$ using the following relation,

$$q(t) = q_f \frac{1}{2}\left(1 - \cos\left(\pi \frac{t}{t_{\rm G}}\right)\right), \tag{5.4}$$

where $q(t) = M_p(t)/M_*$, $q_f = q(t_{\rm G})$.

5.2.3 Rossby Wave Instability

The RWI can occur due to a steep gradient in the radial profile of the angular velocity of the disc. In a disc with a gap opening planet, this location corresponds to the gap edge in the surface density of the disc. Using a long enough planet growth timescale, we expect the gap to open slower and hence the disc viscosity has more time to smooth out the steep gap edges. This will result in a smaller gradient in the velocity profile of the disc at this location.

For the RWI to be excited within an adiabatic disc there must be a local maximum or minimum in the key function (Lovelace et al., 1999),

$$\mathcal{L}(R) = \mathcal{F} S^{\frac{2}{\gamma}}, \tag{5.5}$$

where $\mathcal{F}$ is approximately the inverse of the potential vorticity, $S = P/\Sigma^\gamma$ is the entropy and γ is the ratio of specific heats. As this function clearly has surface density dependence, the opening of a gap in the disc will have an effect on whether the RWI is excited and we would expect to see a local maximum or minimum occurring near a steep gap edge in the surface density of the disc. The presence of a maximum or minimum in this function is only one prerequisite for the excitation of the RWI, it must also exceed a threshold value before the RWI is excited (Lovelace et al., 1999; Ono et al., 2016). Hence, from this function alone we cannot predict the Rossby stability of a system. In addition, our simulations are not adiabatic, they are locally isothermal and as such are in the limit of $\gamma \to 1$. It is therefore difficult to predict whether the RWI will be present a priori and we therefore determine from simulations whether the RWI has led to the formation of vortices by calculating the potential vorticity across the disc. The potential vorticity is given by

$$\xi = \hat{\mathbf{z}} \cdot \frac{\nabla \times \mathbf{v}}{\Sigma}. \tag{5.6}$$

A local minimum in the potential vorticity of the disc indicates the presence of a vortex, which is the non-linear outcome of the RWI.

In addition to observing the two dimensional potential vorticity distribution, we also investigate the orbit-to-orbit difference in the potential vorticity of the disc.

This method is helpful when instabilities are marginal and more difficult to determine from the two dimensional distribution. As the vortices have a different orbital velocity to that of the planet, we will see them move on an orbit-to-orbit basis, therefore if we see a sudden increase in the orbit-to-orbit maximum potential vorticity difference we can imply the presence of the RWI.

5.3 Numerical Setup

Our simulations are run in two dimensions using the FARGO3D code. This is a magnetohydrodynamic code designed to simulate disc evolution in one to three dimensions by solving the hydrodynamic equations of motion. FARGO3D is a good choice for investigating gap opening planets as it uses a C to CUDA translator to allow simulations to be run on Graphics Processing Units (GPUs). GPUs have limited memory, but decrease computational time significantly, allowing moderate resolution simulations to be run for extended periods of time. For more details see Benítez-Llambay & Masset (2016).

Our simulations use a disc model that extends over a radial domain of $0.3 \leq R/R_0 \leq 6.0$ and an azimuthal domain of $-\pi \leq \phi \leq \pi$ with 285 by 596 cells respectively, using logarithmic cell spacing in the radial direction and the planet located at $R = R_0$. The initial surface density is $\Sigma_{\text{int}} = \Sigma_0(R/R_0)^{-1/2}$ with $\Sigma_0 = 2.67 \times 10^{-3}$. The indirect potential due to the planet is accounted for and the mass of the disc is arbitrary, as the planet does not feel the disc material. Reflecting boundary conditions were used with wave killing zones in the regions $0.3 \leq R/R_0 \leq 0.585$ and $5.43 \leq R/R_0 \leq 6.0$ similarly to de Val-Borro et al. (2006), meaning that excited waves are damped before reaching the edge of the simulation. The planet is held on a fixed circular orbit with no migration or disc self-gravity accounted for. We use a constant kinematic viscosity ν and a fixed aspect ratio h. We vary both the final mass ratio, q_f and ν between simulations.

We run our simulations for a number of orbits beyond the chosen planet growth timescale, t_{G}, in order to ensure the RWI is not excited shortly after the planet has reached its maximum mass. This depends on the gap formation timescale and viscous timescale of the disc and as a result will vary depending on the plant mass and viscosity, however, is often of the order of a few hundred orbits.

5.4 Results

5.4.1 Method

We run our simulations for a range of mass ratios, $1.5 \times 10^{-4} \leq q \leq 10^{-3}$ and disc viscosities $10^{-6} \leq \nu \leq 10^{-5}$. For each combination of planet mass and viscosity, we investigate the planet growth timescale for which we do not observe the formation of vortices in the disc. We use an aspect ratio $h = 0.05$ for the majority of our simulations. We focus on this one aspect ratio to reduce the size of the parameter space we investigate, in order to lower the computational time necessary. However, we do investigate changing the aspect ratio in Section 5.4.4.

In order to determine the timescale at which vortices are not formed we calculate the potential vorticity across the disc. From this we determine the orbit-to-orbit maximum potential vorticity difference. As the vortices orbit at a different speed to the planet they move on an orbit-to-orbit basis. An extreme example of these vortices can be seen in Figure 5.1, which shows the formation of vortices in a $t_{\rm G} = 0$ setup. This is an extreme case for the purpose of showing the vortices and using a growth timescale $t_{\rm G} > 0$ will reduce the magnitude of these vortices. Hence, approaching the values of $t_{\rm G}$ that stabilise the disc it will become more and more difficult to identify the presence of vortices from these plots.

Such marginal cases are still important to distinguish, as it is known that vortices are efficient at trapping dust and this could enhance the visibility of a vortex observationally (Barge & Sommeria, 1995). If these vortices form a sharp increase in the orbit-to-orbit potential vorticity difference this should imply the excitation of the RWI. An example of using the potential vorticity difference to highlight the presence of vortices can be seen in Figure 5.2. In this case we can see the comparison between a stable simulation with $t_G = 500$ and an unstable simulation with $t_G = 300$. In the stable case the potential vorticity difference has some small variations but is overall roughly constant, while in the unstable case there is a clear sudden sharp increase in the potential vorticity difference indicating the formation of vortices. At the potential vorticity difference peak, the two dimensional potential vorticity distributions for both of these cases would look similar to the stable and unstable plots shown in Figure 5.3. However, in the $t_{\rm G} = 300$ case vortices would be even more marginal. This highlights how useful this method is to identifying these vortices, especially given that this case is only marginally unstable. For comparison, the vortices in Figure 5.1 produce a maximum $\Delta\xi \approx 200 - 300$.

In particularly marginal cases, in which it becomes difficult to determine stability

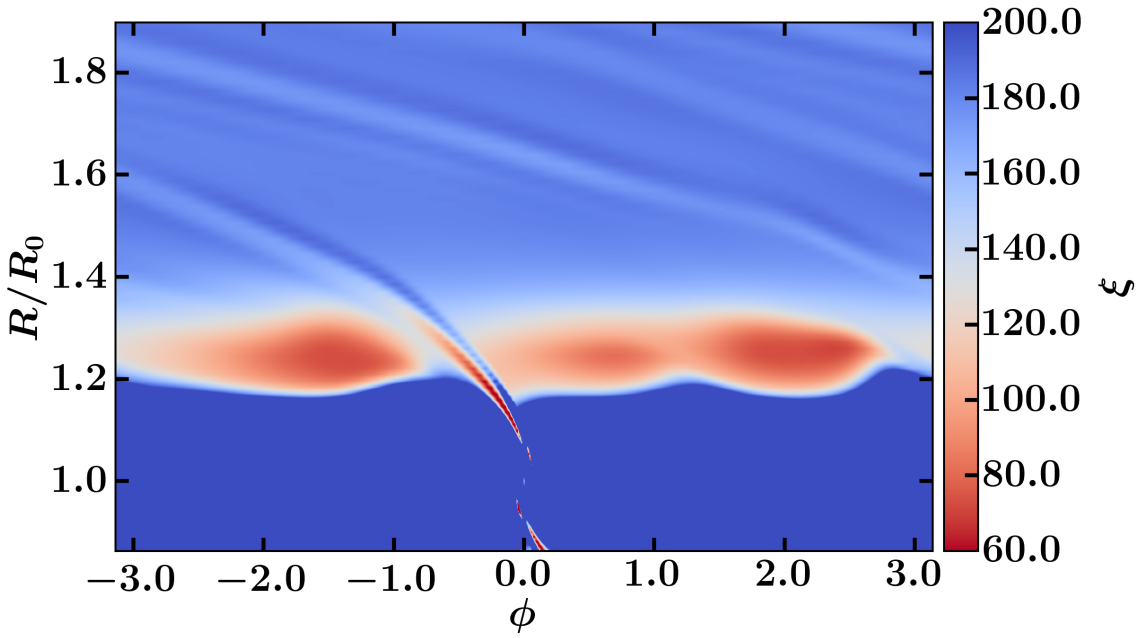

Figure 5.1: An example of vortices clearly visible in the two dimensional potential vorticity distribution of the disc. In this case the disc has $\nu = 8 \times 10^{-6}$ with a $q_f = 6 \times 10^{-4}$ planet and a $t_G = 0$, meaning that the planet is initialised in the disc at $t = 0$ at its final mass. Vortices began to form within 100 orbits of initialising the simulation.

solely from the potential vorticity difference, we also use the two dimensional potential vorticity distribution to aid in determining whether or not the system is stable. An example of using the two dimensional potential vorticity distribution to determine if a vortex is present can be seen in Figure 5.3, which shows two simulations with $t_G = 1400$ and $t_G = 1000$ both at $t = 1500$. While this is a considerably clearer case of vortex formation than those that produce the potential vorticity difference shown in Figure 5.2, this highlights the difference between a stable and unstable system.

We limit our potential vorticity difference calculations to the region close to the gap edge where the vortices are formed, excluding a region $\Delta\phi \approx 0.95$ either side of the planet. We do this because even in cases where there is little to no vortices small fluctuations in the location of the gap edge or the planet's wake, on the order of one to two radial cells can cause very large potential vorticity differences for steep gap edges. This can cause a false result implying the presence of vortices when there are none, hence we must take care to exclude this from our sampling region.

In addition, in some cases we find that weak features can be present in the potential vorticity distribution of the disc, even in simulations that are Rossby stable. These may be due to the interaction between orbiting material and the planet's

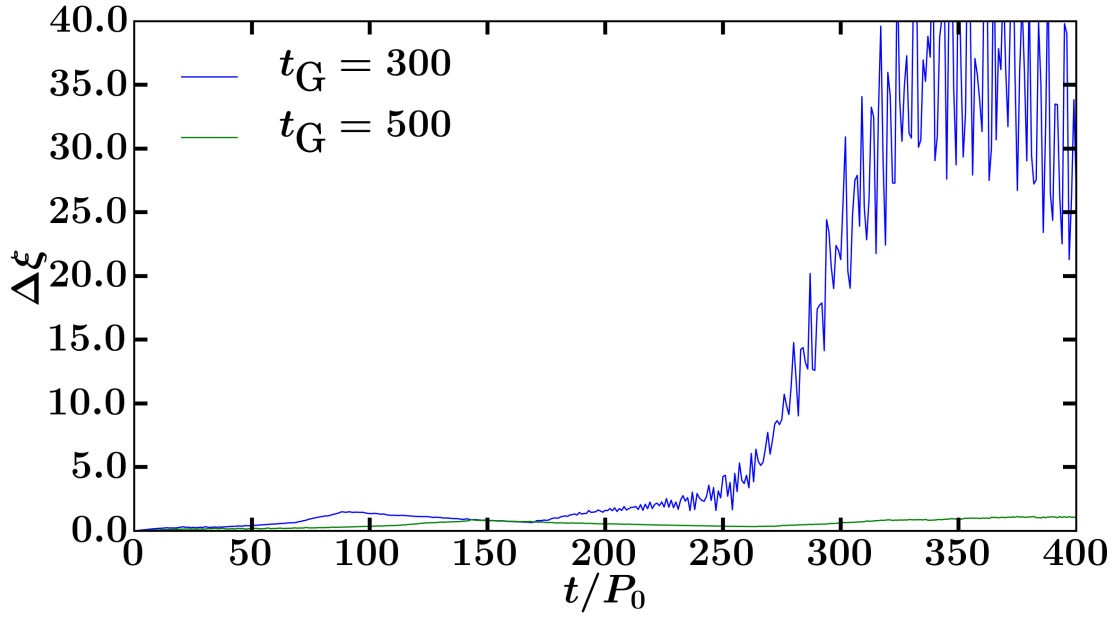

Figure 5.2: Orbit-to-orbit maximum potential vorticity difference for a given growth timescale against the number of orbits elapsed. Illustrated here is the difference between a stable case, with $t_G = 500$ and an unstable case, with $t_G = 300$. The parameters for these two simulations are $\nu = 10^{-5}$ and $q_f = 9\times10^{-4}$. For comparison, a $t_G = 0$ results in a $\Delta\xi \approx 400-1000$ at its most unstable.

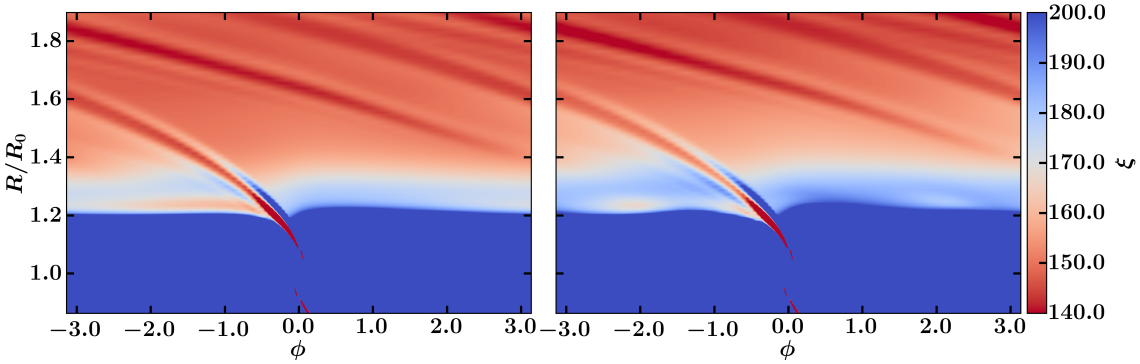

Figure 5.3: A comparison showing the two dimensional potential vorticity distributions for a stable disc (left) and a disc containing vortices (right). Both of these discs show a $q_f = 9\times10^{-4}$ planet in a $\nu = 8\times10^{-6}$ disc at $t = 1500$. The difference between them is the stable case has a $t_G = 1400$ while the unstable case has $t_G = 1000$ and hence has excited an instability, which can be clearly seen from the presence of the vortex.

outer wake, in which material orbiting past the planet's wake receives a decrease in potential vorticity flowing past the shock. This can be seen in the stable system in the left panel of Figure 5.3. This mechanism was first identified by Koller

et al. (2003). These weak features often have no impact on the potential vorticity difference, even in the rare cases in which they stray out of the excluded zones.

5.4.2 Error Estimates

As we discretely sample the growth timescale, a higher sampling resolution will provide a more accurate value of the growth timescale at which vortices do not appear. Therefore, we choose errors on our selected growth timescale based on our sampling resolution. This should account for the possibility that the exact value we are looking for lies between our sampled growth timescales. In selecting whether a growth timescale results in the formation of vortices or not, we first make an educated guess at the growth timescale that results in no vortices and perform a simulation. We analyse the result of this, and increase or decrease the growth timescale by an amount, usually on the order of a few hundred orbits, but heavily dependent on the accuracy of the initial guess and the computational time required to complete the simulation. This is an iterative process, until we find the smallest growth timescale for which the system has no vortices. Hence, we have a discrete resolution in growth timescale, which is necessary to reduce computational time. The error due to this gets larger as the growth timescale increases, as these are more computationally expensive. Based on these considerations, for the purposes of this study we select errors as presented in Table 5.1.

Table 5.1: Errors for different ranges of t_G.

t_G	Δt_G
$0 \leq t_G \leq 500$	50
$500 < t_G \leq 1000$	100
$1000 < t_G \leq 2000$	200

5.4.3 Minimum planet growth timescale

Table 5.2 shows the minimum planet growth timescales for which vortices are not formed, over the range of mass ratios and viscosities. This information is also presented in Figure 5.4 which fits exponential curves of the form

$$t = t_0 e^{\frac{q_f}{q_0}}, \tag{5.7}$$

where t_0 and q_0 are constants dependent on the fit, given in Table 5.3. From this figure we can see that the best fit exponential curves are a reasonable fit to the data

and that as we increase the viscosity the curves become less steep and q_0 increases. From this we find a relation between q_0 and the viscosity,

$$\nu = 9.1884 \times 10^{-4} q_0^{0.5588}. \tag{5.8}$$

This relation can be seen in Figure 5.5. This shows a reasonable fit to the data in the mid to high viscosity regime, however, deviation begins to become significant in the low viscosity regions.

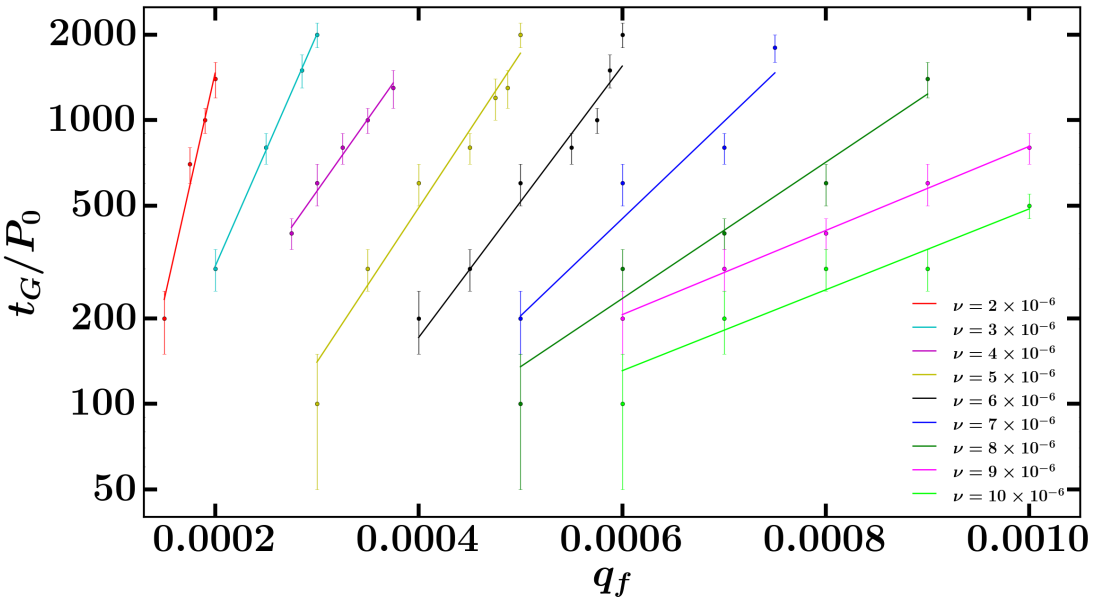

Figure 5.4: Exponential best fit curves for the data presented in Table 5.2 for each value of ν, including errors decided upon as discussed in Section 5.4, presented using a logarithmic scale in orbits. A planet of mass q_f in a disc of viscosity ν will create vortices while forming if the growth timescale t_G is below the best fit curve for that viscosity, and will not create vortices if it is above. This clearly shows the trend of decreasing gradient with increasing viscosity.

As t_0 does not follow a simple law (see Table 5.3) we cannot predict how t_0 evolves with viscosity. Hence, we cannot determine a continuous relation between the viscosity and planet growth timescale. This, in combination with Equation 5.8's poor fitting at low viscosities means that we use Equation 5.7 at our discrete viscosity intervals to form a relation between mass and viscosity. We can then use Equation 5.7 evaluated at these viscosities in order to make statements about the potential masses of gap opening planets and viscosities of the disc based on the age of the system in question, using observational results regarding the detection or absence of vortices within these discs.

Table 5.2: Planet growth timescales for which no vortices are formed for different mass ratios and disc viscosities.

ν	q_f	$t_{\rm G}$	ν	q_f	$t_{\rm G}$
1.0×10^{-6}	1.500×10^{-4}	2000	6.0×10^{-6}	5.500×10^{-4}	800
2.0×10^{-6}	1.500×10^{-4}	200	6.0×10^{-6}	5.750×10^{-4}	1000
2.0×10^{-6}	1.750×10^{-4}	700	6.0×10^{-6}	5.875×10^{-4}	1500
2.0×10^{-6}	1.900×10^{-4}	1000	6.0×10^{-6}	6.000×10^{-4}	2000
2.0×10^{-6}	2.000×10^{-4}	1400	7.0×10^{-6}	5.000×10^{-4}	200
3.0×10^{-6}	2.000×10^{-4}	300	7.0×10^{-6}	6.000×10^{-4}	600
3.0×10^{-6}	2.500×10^{-4}	800	7.0×10^{-6}	7.000×10^{-4}	800
3.0×10^{-6}	2.850×10^{-4}	1500	7.0×10^{-6}	7.500×10^{-4}	1800
3.0×10^{-6}	3.000×10^{-4}	2000	8.0×10^{-6}	5.000×10^{-4}	100
4.0×10^{-6}	2.750×10^{-4}	400	8.0×10^{-6}	6.000×10^{-4}	300
4.0×10^{-6}	3.000×10^{-4}	600	8.0×10^{-6}	7.000×10^{-4}	400
4.0×10^{-6}	3.250×10^{-4}	800	8.0×10^{-6}	8.000×10^{-4}	600
4.0×10^{-6}	3.500×10^{-4}	1000	8.0×10^{-6}	9.000×10^{-4}	1400
4.0×10^{-6}	3.750×10^{-4}	1300	9.0×10^{-6}	6.000×10^{-4}	200
5.0×10^{-6}	3.000×10^{-4}	100	9.0×10^{-6}	7.000×10^{-4}	300
5.0×10^{-6}	3.500×10^{-4}	300	9.0×10^{-6}	8.000×10^{-4}	400
5.0×10^{-6}	4.000×10^{-4}	600	9.0×10^{-6}	9.000×10^{-4}	600
5.0×10^{-6}	4.500×10^{-4}	800	9.0×10^{-6}	1.000×10^{-3}	800
5.0×10^{-6}	4.750×10^{-4}	1200	1.0×10^{-5}	6.000×10^{-4}	100
5.0×10^{-6}	4.875×10^{-4}	1300	1.0×10^{-5}	7.000×10^{-4}	200
5.0×10^{-6}	5.000×10^{-4}	2000	1.0×10^{-5}	8.000×10^{-4}	300
6.0×10^{-6}	4.000×10^{-4}	200	1.0×10^{-5}	9.000×10^{-4}	300
6.0×10^{-6}	4.500×10^{-4}	300	1.0×10^{-5}	1.000×10^{-3}	500
6.0×10^{-6}	5.000×10^{-4}	600			

Table 5.3: Values of the constants t_0 and q_0 for different disc viscosities.

Disc viscosity, ν	t_0	q_0
2.0×10^{-6}	0.9469	2.7238×10^{-5}
3.0×10^{-6}	7.0582	5.3100×10^{-5}
4.0×10^{-6}	16.7741	8.5449×10^{-5}
5.0×10^{-6}	3.2684	7.9798×10^{-5}
6.0×10^{-6}	2.1016	9.0870×10^{-5}
7.0×10^{-6}	3.9502	1.2675×10^{-4}
8.0×10^{-6}	8.5080	1.8079×10^{-4}
9.0×10^{-6}	26.5984	2.9289×10^{-4}
1.0×10^{-5}	18.2717	3.0466×10^{-4}

From Figure 5.4 we can see that for higher viscosity discs it is possible to form more massive planets over shorter timescales, without creating vortices. Consider the planets formed with $q_f = 6 \times 10^{-4}$. At $\nu = 10^{-5}$, it is possible to form this planet in only 100 orbits with no vortices created. However, as we lower the viscosity the growth timescale to form this planet without creating vortices increases. Hence, at $\nu = 6 \times 10^{-6}$ the same planet takes 2000 orbits to form without creating vortices.

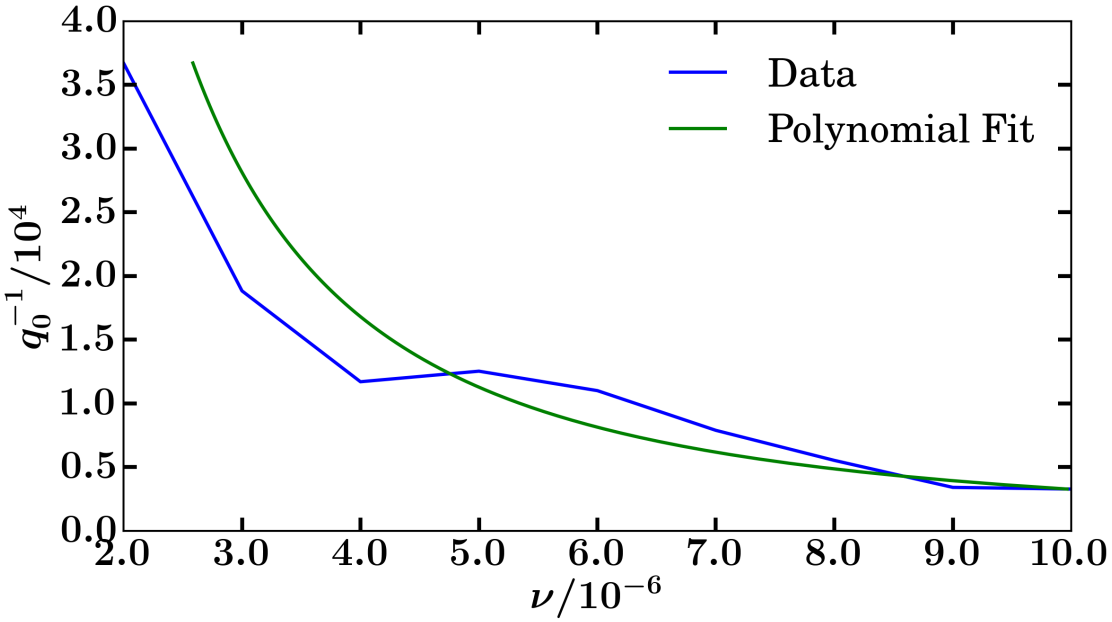

Figure 5.5: Polynomial best fit curve for the gradients of the curves in Figure 5.4 given by Equation 5.8. This shows relatively good agreement between the fit and the data in the mid to high viscosity range, however, this gets worse at low viscosities and begins to significantly overestimate the gradient of the mass/growth timescale curve.

Another important result we can see in Figure 5.4 pertains to the masses of planets that can form vortices in the disc. The exponential best fit curves in Figure 5.4 show the minimum growth timescale required for a planet of a given mass to form without creating vortices. As we can see there is large parameter space at lower viscosities in which planets can create vortices while forming. Indeed, it is not even necessary for such planets to be particularly massive. For example consider a disc with $\nu = 2 \times 10^{-6}$. In this disc, if a planet of $q_f = 2 \times 10^{-4}$ is to form over a timescale $t_{\mathrm{G}} < 1400$, it will create vortices in the disc. This planet is a similar mass to Saturn, and significantly less massive than Jupiter. Hence, if discs have low viscosities we may not need to predict planets on the order of Jupiter's mass to explain potential vortices in observational results.

5.4.4 Larger aspect ratio

So far we have only discussed the formation of vortices in a disc with $h = 0.05$. It is expected that as the disc aspect ratio changes, so will the planet growth timescale required for a planet to be formed without creating vortices, for a given viscosity and planet mass. We have investigated this behaviour, increasing the aspect ratio

of the disc from $h = 0.05$ to $h = 0.06$. We only investigate one different aspect ratio and only for a small number of setups. Adding this additional parameter to our investigation into q_f and ν would be infeasible due to increased computational time required. However, we can make a physically motivated guess at how the results vary by changing h and then perform a number of simulations to check this.

We know that the gap depth is set by the K parameter (Kanagawa et al., 2015),

$$K = q_f^2 \left(\frac{R_0}{h}\right)^5 \alpha^{-1}, \tag{5.9}$$

where $\nu = \alpha c_s H$. Hence, we may expect that holding the K parameter constant while increasing the aspect ratio gives similar results. However, a constant K at a given h is degenerate in q_f and ν and therefore we need a method of selecting values for these parameters. To this end we use the relation q_f/h^3 to set q_f at the increased aspect ratio. This relation governs the non-linearity of the flow in the inviscid case (Korycansky & Papaloizou, 1996). Using this we keep the non-linearity of the flow constant, while now having specific values for the parameters q_f and ν at our selected value of $h = 0.06$. With these parameters we find a similar result to the $h = 0.05$ case, namely an unstable run remains unstable and a stable run remains stable, however, using these parameters does seem to slightly improve the overall stability of the system compared to the $h = 0.05$ case. Therefore, we investigate using the relation q_f/h^2 to select q_f and ν instead, still with $h = 0.06$. Using these parameters we found had the opposite effect, driving the system more towards instability and vortex formation compared to the respective $h = 0.05$ case. These results can be seen in Figure 5.6 which shows the orbit-to-orbit potential vorticity difference that we use as a tracer for vortex formation for both $h = 0.06$ cases and the $h = 0.05$ case. The $h = 0.05$ setup has $\nu = 9.0 \times 10^{-6}$ and $q_f = 9.0 \times 10^{-3}$, while all setups have $t_G = 600$. Clearly it can be seen that the q_f/h^2 case is driven more towards vortex formation, while the q_f/h^3 case is closer to reproducing the same results as $h = 0.05$. The earlier vortex formation occurs as the gap opens faster in the lower viscosity cases. From these results we can see that using the relation q_f/h^3 gives reasonably consistent results with $h = 0.05$, despite the increase to $h = 0.06$, whereas q_f/h^2 is less consistent. However, we do not know if this result is applicable in every setup.

5.5 Comparison with Observations

In the previous section we have found an approximate relation between the mass of a planet and the viscosity of the disc that predicts the growth timescale of the

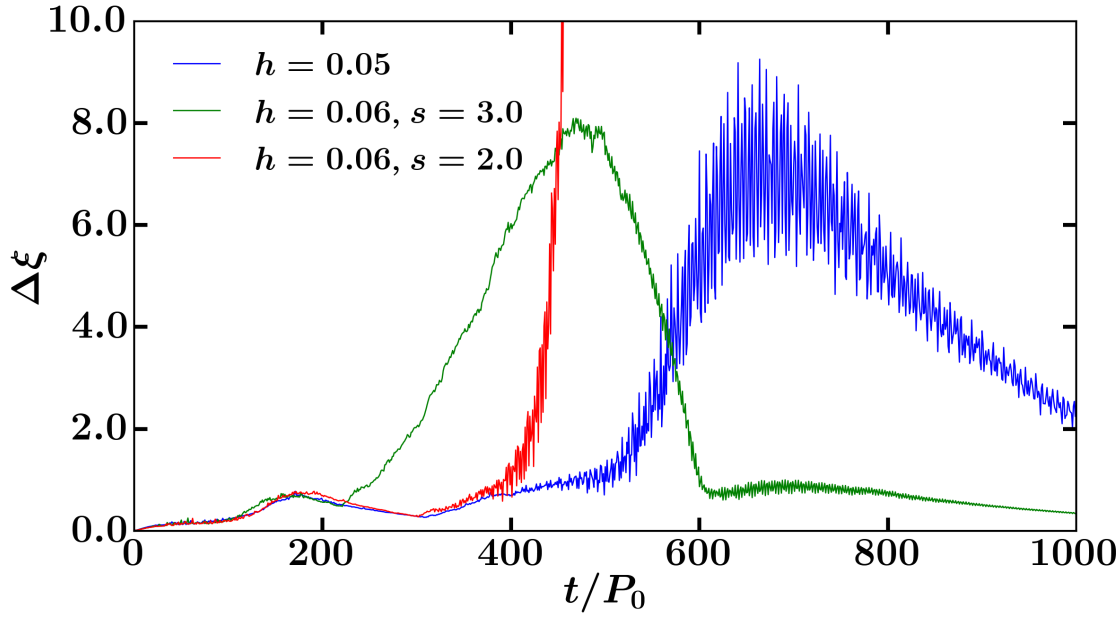

Figure 5.6: Orbit-to-orbit potential vorticity difference for the parameters discussed in Section 5.4.4, where s represents the power of h in the parameter setting relation. Here we can see that it is possible to get similar results despite increasing the aspect ratio by holding K constant, however, depending on how the parameters are selected the system may be forced more towards stability or vortex formation.

planet required in order for the planet to form without creating vortices within the disc. This is a result which allows us to make comparisons with observational data. Specifically we can use observational evidence for both the presence and absence of vortices in protoplanetary discs that show structure indicative of embedded planets. From these we can make tentative statements about the parameters of the discs and planets in question.

An important consideration in our analysis here is the choice of growth timescale for planets within the discs. Whereas planet masses and disc viscosities are often predicted parameters from observational results, which we can use in our comparison, planet growth timescale is not often considered. As a result, for the comparisons in this section we use the age of the system as a limit on the planet growth timescale. For a planet in a disc that has reached maximum mass, the growth timescale of the planet cannot exceed the lifetime of the disc. Hence, the growth timescales considered here represent the maximum possible growth timescale for a planet in that disc. As we use ages considerably longer than the growth timescales fit by Equation 5.7, we have to assume that our fits are still applicable in this regime.

Additionally, we have generally endeavoured to make statements using the units

of mass presented in the papers to which we are comparing. However, for ease of comparison we also provide the conversion to mass ratio q alongside each. With regards to viscosity, many of the papers to which we are comparing use a turbulent α viscosity (Shakura & Sunyaev, 1973). We provide the conversion to the kinematic viscosity given the disc parameters used in our setup, at the location of the planet $R = R_0$.

5.5.1 Axisymmetric discs

HL Tau

One of the most well documented protoplanetary discs of the last few years is HL Tau (ALMA Partnership et al., 2015), for good reason as it is one of the highest resolution images of a protoplanetary disc we currently have, in addition to showing very interesting structure. The origin of this structure is still debated, as there are a number of different proposed factors that can explain the gap structure observed, besides planets. Nevertheless, there are some factors that support the hypothesis that planets have carved these gaps, as the locations of these dark bands are located close to the resonances between planets orbiting at these radial locations (Wolf et al., 2002; ALMA Partnership et al., 2015). Dipierro et al. (2015) performed smoothed particle hydrodynamics simulations to estimate the masses of planets that could potentially explain this structure. Their results imply that the observed structure could be the result of three planets of masses $q = 2.0 \times 10^{-4}$, 2.7×10^{-4} and 5.5×10^{-4} located at 13.2, 32.3 and 68.8AU respectively. This prediction is very interesting to us, as crucially the HL Tau system shows no sign of vortices. Using Equation 5.7 and assuming $t_{\rm G} = 1$Myr (an upper limit on the age of the HL Tau system, Testi et al., 2015), we can determine if our predictions regarding vortex formation agree with their planet masses. The results of this can be seen in Figure 5.7, which shows the solution to Equation 5.7 for a range of viscosities. Dipierro et al. (2015) use a turbulent viscosity (Shakura & Sunyaev, 1973) of $\alpha \approx 0.005$ ($\nu = 1.25 \times 10^{-5}$), however, more recent results from dust settling models imply that for HL Tau $\alpha \approx 10^{-4}$ ($\nu = 2.5 \times 10^{-7}$) is an upper limit (Pinte et al., 2016). Dipierro et al. (2015) use a flared disc with aspect ratio $h_{\rm in} = 0.04$ and $h_{\rm out} = 0.08$.

From Figure 5.7 we can see that our results are consistent with this prediction for the very large viscosity cases. For $\nu > 8 \times 10^{-6}$ we predict it is possible to form planets up to $\approx 1 M_J$ ($q = 1.0 \times 10^{-3}$) at all of these radii with no vortices created. Using $\nu = 2.5 \times 10^{-7}$ it appears that the planets predicted by Dipierro et al. (2015) are unable to form without creating vortices. Our results predict that planets

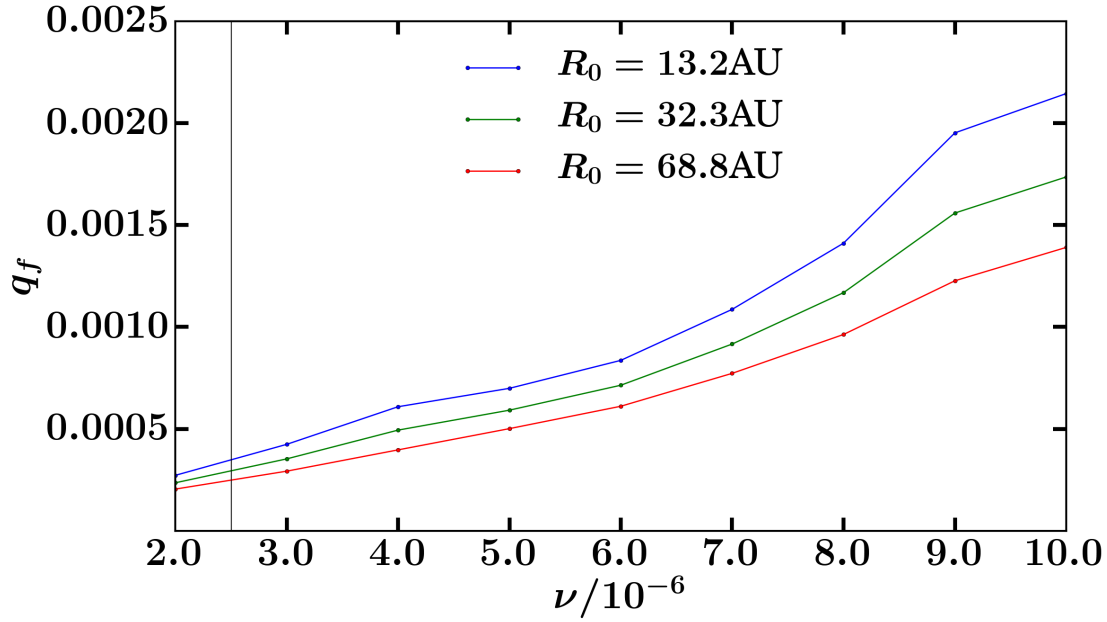

Figure 5.7: Maximum masses of planets that can be formed over a timescale t_G in a disc of a given viscosity. t_G is set by the number of orbits able to be completed by a planet orbiting at a radius R over the lifetime of the disc. This means that masses above this line will not form at this radius without creating vortices, while masses below will, according to our result in Equation 5.7. In this case the three radii chosen correspond to the predicted locations of planets in the disc HL Tau (Dipierro et al., 2015). The vertical black line corresponds to the viscosity used by Jin et al. (2016).

with masses of roughly $q = 2.5 \times 10^{-5}$ are the largest that could form here without creating vortices. Using the viscosity from Dipierro et al. (2015), however, it is very clear from our results that their planets could form without creating vortices.

Jin et al. (2016) make predictions in a lower viscosity regime than Dipierro et al. (2015). They imply that the observed structure can be explained by planets of masses $0.35M_J$, $0.17M_J$ and $0.26M_J$ ($q = 3.5, 1.7$ and 2.6×10^{-4} respectively) at approximately the same radii as above. Jin et al. (2016) use a disc viscosity of $\alpha = 10^{-3}$ ($\nu = 2.5 \times 10^{-6}$) which lies in between the above viscosities. Fortunately, this viscosity also lies within our explored parameter space. From Figure 5.7 we can see that at this viscosity we can expect planets of $q \approx 2.5 - 3.5 \times 10^{-4}$ to form without creating vortices. Therefore, the planets of Jin et al. (2016) are roughly the maximum mass planets our results predict could be present in the HL Tau disc at this viscosity. Hence, our results show consistency with their predictions.

TW Hydrae

Another well documented disc that shows similar features to HL Tau is TW Hydrae, also the result of high resolution ALMA observations (Andrews et al., 2016). This disc also shows axisymmetric dark radial rings that could potentially be the result of gaps formed by planet-disc interactions. An interesting difference between TW Hydrae and HL Tau is that while the latter is a young disc, TW Hydrae is particularly old, with an age of $3-15$Myr (Sokal et al., 2018). Another difference that separates these discs is the dark bands themselves. In TW Hydrae these dark bands are significantly brighter (compared to the surrounding disc) and narrower than those in HL Tau. The masses of planets that could potentially explain these gaps has been investigated by van Boekel et al. (2017), using the results of Duffell (2015) and was found to be in the range of several earth masses. This corresponds to roughly $4-10M_{\oplus}$, $10-20M_{\oplus}$ and $25-50M_{\oplus}$ ($q(1M_{\oplus}) \approx 3.0\times10^{-6}$) at 6.0, 21.0 and 85.0AU respectively, for viscosities between $\alpha = 1.0-4.0\times10^{-4}$ ($\nu = 2.5\times10^{-7}-1.0\times10^{-6}$). From our results we predict that the gaps at 6 and 21AU are capable of being opened by these planets without forming vortices (except for the gap at 21AU in the lowest viscosity case). However, our results imply that the larger mass planet at 85AU could not form there without creating vortices. These predictions are based on extrapolation, as we have not investigated at viscosities this low. Tsukagoshi et al. (2016) also imply that the most prominent, central gap could be caused by an approximately 1.5 Neptune mass planet ($q = 7.72\times10^{-5}$) in an $h = 0.05$, $\alpha = 10^{-3}$ ($\nu = 2.5\times10^{-6}$) disc. As this planet is low mass and TW Hydrae is an older star, our results predict that this planet could form without causing the creation of vortices. Hence, we display some consistency here with previous studies.

Taurus star forming region

More recently, Long et al. (2018) performed an analysis on 32 discs in the Taurus star forming region using ALMA. From this sample 12 discs containing axisymmetric structure were identified. This structure takes the form of dark band bright ring pairs, emission bumps and cavities. Overall 19 gap ring pairs were identified, indicating that a number of these systems contain multiple gaps. These gaps range in location from $R = 10-120$AU with no preferred distance. The majority of these gaps are narrow, but the weak correlation between gap location and gap width potentially implies formation via planet-disc interactions. In addition a significant number of these gaps can not be explained by condensation fronts. Long et al. (2018) perform an analysis on the width of these gaps in order to determine possible

masses of planets that could form them. Assuming the width of the gaps corresponds to $4R_{\rm Hill}$ (Dodson-Robinson & Salyk, 2011) they estimate the masses of the planets to be in the $0.1 - 0.5M_J$ ($q = 1.0 \times 10^{-4} - 5.0 \times 10^{-4}$) range, however, they stress that these masses have large uncertainties and that the gap widths may be as large as $7 - 10R_{\rm Hill}$ (Pinilla et al., 2012). Alternatively they compare the distance between the minimum of the gaps and the maximum of the rings to the results of hydrodynamic simulations (Rosotti et al., 2016) with an $\alpha = 10^{-4}$ ($\nu = 2.5 \times 10^{-7}$), resulting in a predicted planet mass of $0.05M_J$ ($q = 5.0 \times 10^{-5}$) and an $\alpha = 10^{-2}$ ($\nu = 2.5 \times 10^{-5}$), resulting in a predicted planet mass of $0.3M_J$ ($q = 3.0 \times 10^{-4}$). Using Equation 5.7 and an estimated disc lifetime of 2Myr (based on stellar age estimates for the spectral type of the target stars, Baraffe et al., 2015; Feiden, 2016) we can see that in the higher viscosity case planets of $0.3M_J$ ($q = 3.0 \times 10^{-4}$) can form at any of these radii without creating vortices. The lower viscosity case is more difficult to predict as it is significantly lower than the viscosities we have explored, however, for this planet mass it seems unlikely from our results that the predicted planet could form without creating vortices. Instead, we expect the maximum mass planets that could be found here to be roughly half the mass they predict. In addition, from our results planets of mass $0.5M_J$ ($q = 5.0 \times 10^{-4}$) would require the disc viscosity to be $\nu \gtrapprox 3.0 - 5.0 \times 10^{-6}$ to form without creating vortices, depending on the radius at which they are located. Conversely we find planets as small as $0.1M_J$ ($q = 1.0 \times 10^{-4}$) can even form at $\nu = 2.0 \times 10^{-6}$ without creating vortices. This can be seen in Figure 5.8. Hence, our results show that these gaps could be opened by planets of the masses predicted by Long et al. (2018) without forming vortices.

DSHARP

Recently results from DSHARP (Disc Substructures at High Angular Resolution Project) have been published (Andrews et al., 2018). This is one of the first large programs carried out using the ALMA telescope. In this project millimetre emission from a sample of 20 protoplanetary discs was observed with intent to find and characterise substructures, such as dark gaps and bright rings. It was found that these substructures are present at almost any radius across the whole sample of discs. They are significantly more common than spiral arm structure, occurring in 3 of the discs and crescent asymmetries, occurring in 2 of the discs. Both of these additional structures coexist with the dark band substructures present across the sample. Similarly to Long et al. (2018) it is found that condensation fronts are not a reliable method of forming these potential gaps (Huang et al., 2018).

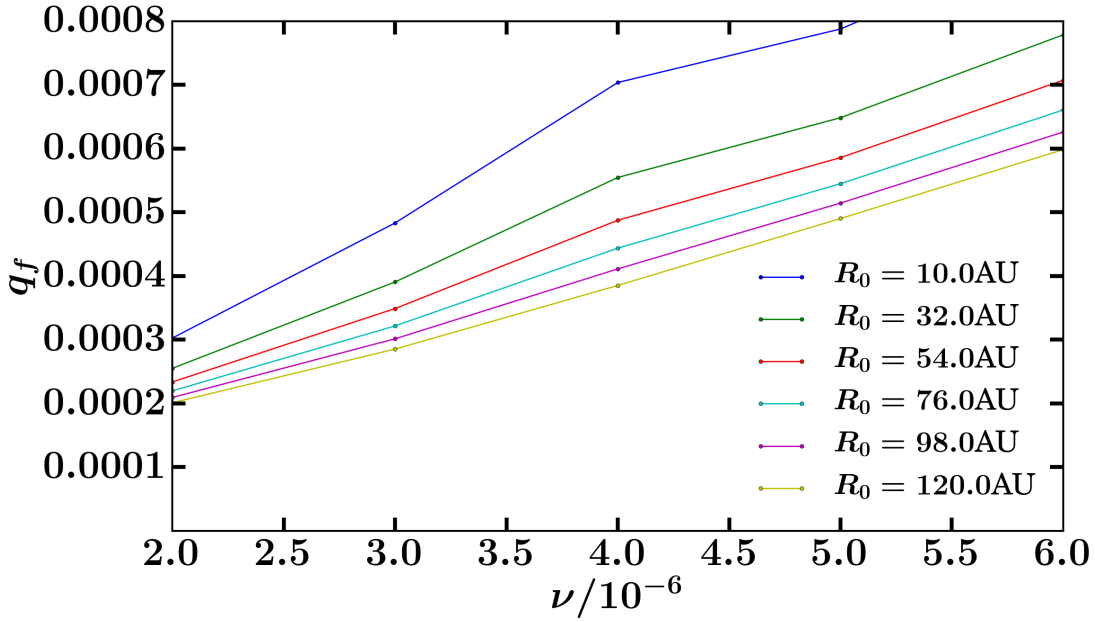

Figure 5.8: Maximum masses that can be formed over a timescale t_G in a disc of a given viscosity, similarly to Figure 5.7. In this case the radii chosen correspond to the predicted locations of planets in the sample of 12 discs analysed by Long et al. (2018). Specifically this plot focusses on the region for which planets of masses in the range $0.1 - 0.5 M_J$ ($q = 1.0 \times 10^{-4} - 5.0 \times 10^{-4}$) can form.

Zhang et al. (2018) analyse 14 discs from this sample, in which a total of 19 gaps are present. Relations between gap width/depth and planet/disc properties are found using two dimensional hydrodynamic simulations including dust particles for a range of disc aspect ratios, turbulent viscosities and planet masses appropriate to the sample. These relations are then used to infer the planet mass from the intensity profiles of Huang et al. (2018). However, only 12 of these gaps are wide enough to be fit using their fitting formula, whereas the remainder are fitted by eye. Hence, we will focus our comparison here on the more robustly fit sample of 12.

The predicted planet masses listed in Zhang et al. (2018) clearly highlight the degeneracy we are attempting to address. Not only does the planet mass differ depending on which turbulent viscosity is assumed but also the assumed dust size distribution. This results in between 3 and 9 possible planet masses that can explain each gap. While our results can do nothing to address the degeneracy arising from the assumed dust size distribution, we can make some predictions based on the turbulent viscosity. The turbulent viscosities assumed in their analysis are $\alpha = 10^{-4}, 10^{-3}$ and 10^{-2} ($\nu = 2.5 \times 10^{-7}, \times 10^{-6}$ and $\times 10^{-5}$ respectively). For each of these viscosities the predicted masses range from $0.01 - 1.06 M_J$, $0.03 - 2.16 M_J$

and $0.06 - 4.41 M_J$ ($q(1 M_J) = 1.0 \times 10^{-3}$) respectively in the axisymmetric discs. The aspect ratios of their discs range from $h = 0.04 - 0.09$. We perform a general overall comparison using a disc lifetime of 1Myr (Andrews et al., 2018), $h = 0.05$ and $R = 10 - 100$AU as the locations of the gaps for our comparison. Using this our results show that in the lowest viscosity case, planets of up to $0.025 M_J$ ($q = 2.5 \times 10^{-5}$) can form without creating vortices and in the middle viscosity case planets of up to $0.36 M_J$ ($q = 3.6 \times 10^{-4}$) can form without creating vortices. In this case the lowest viscosity prediction requires some extrapolation to make as this viscosity lies outside the domain of our results. Similarly, the highest viscosity case is significantly outside the domain of our results, such that it is difficult to make meaningful extrapolations towards. At the edge of our results domain, at $\nu = 10^{-5}$, planets of up to $2.0 M_J$ ($q = 2.0 \times 10^{-3}$) can form without creating vortices. These are the maximum masses in each case, which corresponds to the smallest radial location of the gap and increasing the radial location will reduce the mass of a planet that can form. In general for the lower viscosities this mass reduction is very small. Hence, our results provide a limit which can rule out the low viscosity, high planet mass cases. This information is presented concisely in Table 5.4.

Table 5.4: Predicted planet masses from DSHARP (Zhang et al., 2018) and our results ($q(M_J) = 10^{-3}$).

Viscosity, ν	DSHARP masses, q/M_J	Predicted maximum mass q/M_J
2.5×10^{-7}	$0.01 - 1.06$	0.025
2.5×10^{-6}	$0.03 - 2.16$	0.36
2.5×10^{-5}	$0.06 - 4.41$	> 2.0

Finally, we must address the point that two of the discs in this sample (HD143006 and HD163296) show asymmetries and hence require a slightly different analysis. In HD143006 the asymmetry lies beyond the continuum disc edge and such is not a direct analogue to our gap edge vortices. However, this disc is predicted to contain a planet of mass $\approx 2 - 40 M_J$ ($q = 2 - 40 \times 10^{-3}$) dependent on the degeneracies listed above, which would be impossible to form without creating vortices according to our results. The asymmetry in HD163296 is more consistent with those we see in our simulations, occurring at the $R = 48$AU gap in this disc. Using this gap location and an age of 10Myr (Andrews et al., 2018) we find that all of the predicted masses $0.54 - 4.45 M_J$ ($q = 0.54 - 4.45 \times 10^{-3}$) across the above degeneracies will not be able to form without creating vortices. The only potential exception to this is in the high viscosity case, which suffers from the same extrapolation problem as in the comparison above. At $\nu = 10^{-5}$ it is already possible for their smallest predicted planet to form while

retaining axisymmetry, for their highest viscosity. This parameter space will very likely become larger as the viscosity increases. Nevertheless, the overall agreement with this disc is very good.

Another very interesting result from the analysis of Zhang et al. (2018) is for the specific case of the disc AS209 which contains axisymmetric gaps at 9, 24, 35, 62, 90, 105 and 137AU. They find that a planet of mass $q = 1.0 \times 10^{-4}$ at $R = 100$AU in a disc with $\alpha = 10^{-5}$ ($\nu = 2.5 \times 10^{-8}$) and $h = 0.05$ can very well explain both position and amplitude of the 24, 35, 62, 90 and 105AU gaps. The age of this disc is 1Myr (Andrews et al., 2018). Again we must extrapolate as the viscosity is so small, but unfortunately from our results it is very clear that a planet of this mass would not be able to form in these conditions without also creating vortices. Interestingly, despite reproducing well the location and amplitude their results also show some asymmetry in the disc substructure. This is overcome by using a radially varying turbulent viscosity, which also provides a better fit to the observations. Hence, this is something that may need to be considered.

5.5.2 Non-axisymmetric discs

Oph IRS 48

There are many observational examples of discs containing vortices that could potentially be the result of planet-disc interactions (e.g. van der Marel et al., 2013; Isella et al., 2013; Fukagawa et al., 2013; Pérez et al., 2014; Marino et al., 2015; Kraus et al., 2017; Dong et al., 2018; Cazzoletti et al., 2018). These are also particularly interesting cases to us as, similarly to the HL Tau and TW Hydrae cases, they allow us to make tentative predictions about the planets in these systems. van der Marel et al. (2013) investigate the disc surrounding the young A type star Oph IRS 48, in which a dust trap is located between 45 and 80AU at the edge of an inner cavity. They suggest a planet of mass $\approx 10 M_J$ ($q = 1.0 \times 10^{-2}$) located at 20AU could explain this and that the dust trap at the edge of the cavity could be a vortex arising from the excitation of the RWI. In their simulations, they use an $\alpha = 10^{-4}$ ($\nu = 2.5 \times 10^{-7}$) and an $h = 0.05$ at the location of the planet. This aspect ratio is consistent with our investigation, however, their disc is flaring. From Equation 5.7 we can see that attempting to form such a massive planet over standard young disc lifetimes ($1 - 5$Myr) will always result in vortices.

HD135344B

The disc surrounding the Herbig Ae/Be star HD135344B (Pérez et al., 2014; Cazzoletti et al., 2018) shows interesting phenomena. This disc contains a cavity with a radius of $\approx$ 50AU at the edge of which is a bright ring of material. At 80AU there is a crescent asymmetry. The disc also contains two bright spiral arms extending out to $\approx$ 75AU. This disc is also particularly interesting as it is currently the only observational result we have that shows the features identified by Hammer et al. (2019), pertaining to an elongated vortex forming as a result of a planet growing in the disc. Fung & Dong (2015); Dong & Fung (2017); Cazzoletti et al. (2018) show that these phenomena could possibly be explained by a planet of $5.0 M_{\rm J}$ ($q = 5.0 \times 10^{-3}$) situated between the bright ring and the asymmetry. For standard disc lifetimes $1 - 10$Myr and viscosity ν on the order of a few 10^{-6} we can confirm from Equation 5.7 that forming such a massive planet at this location will create vortices. Also particularly interestingly Cazzoletti et al. (2018) show that, by using the results of Rosotti et al. (2016); Facchini et al. (2018), the gap between these features can be explained by a $0.2 M_{\rm J}$ ($q = 2.0 \times 10^{-4}$) planet at 68AU with a viscosity of $\alpha = 10^{-3}$ ($\nu = 2.5 \times 10^{-6}$). In this situation the planet is not massive enough to explain the spiral arms, hence is not a result that explains all of the features in the disc. Our results show the formation of this planet should not create vortices for ages $1 - 10$Myr. This is interesting, as while it has been predicted that a planet of this mass could form the gap we see, both our results and the analysis by Cazzoletti et al. (2018) predict that this planet could not form the additional features present in this disc. This is an example of using our results to rule out a planet of a certain mass as an explanation for a gap in a disc, despite this planet being a reasonable candidate when a large number of disc parameters are unknowns.

MWC758

Another system of interest surrounds the Herbig Ae/Be star MWC758 (Marino et al., 2015; Dong et al., 2018). This star is 3.7 ± 2Myr old and has a cavity of radius 50AU in which there is the possibility of a low mass companion. The outer disc extends to $\approx$ 100AU with two clumps at 50 and 85AU. Dong et al. (2018) postulate that a $5 - 10 M_{\rm J}$ ($q = 5 - 10 \times 10^{-3}$) planet at 100AU could be the cause of this (Dong et al., 2015). Once again for a reasonable range of disc viscosities, using the age of the star as an upper limit we find from Equation 5.7 that attempting to form a planet this massive at this radius will create vortices.

5.5.3 Concluding Remarks

The majority of our investigation has been undertaken assuming a constant disc aspect ratio $h = 0.05$, however, as discussed in Section 5.4.4 we have investigated the effect of varying this parameter. As the results discussed here are observational, it is very difficult to know the aspect ratios of the discs in question. As a result predictions made regarding masses of gap opening planets will often assume the value of h. Our investigation into the aspect ratio shows that a thicker disc will tend to be more stable, for the same viscosity and planet mass. Therefore, should the discs in these observations have a larger aspect ratio than has been assumed (or the assumed aspect ratio is greater than $h = 0.05$) the planet masses and viscosities could be larger before vortices are created. This also means that if we predict a planet of a given mass can form in a disc without causing vortices to appear, then that same planet can still form without creating vortices if the disc's aspect ratio was larger than we assumed.

The findings from our results and the predictions regarding the above systems is twofold. Firstly, it is apparent that given our relation and for reasonable ranges of disc viscosity and age, several $M_{\rm J}$ planets will always create vortices in the disc during their formation. Hence, in these discs that we have observed, if we expect to see a planet of multiple $M_{\rm J}$ we should also expect to find asymmetries. Secondly, our results show that vortices are formed for planets significantly smaller than multiple $M_{\rm J}$ for low viscosities. This implies that, in systems observationally showing signs of vortices, the planets may not be as giant as we previously expected. Hence, vortices in discs may not be a tracer of multiple $M_{\rm J}$ mass planets, instead they could show the presence of planets with masses in the sub-Saturn range, down to even only a few Neptune masses. In addition, the absence of vortices in low viscosity discs implies that if any planets exist in the disc, they must be very low mass.

5.6 Discussion

Our results and the above conclusion are still very tentative. We have made a number of assumptions in our simulations that could have some impact on the validity of our main result, in addition to the aforementioned errors. We will justify and discuss these assumptions here.

We make no statements regarding how these planets could form at these locations within the disc. We use a very simplified method of growing the planet mass over a timescale, discussed in Section 5.2.2, that does not take into account any particular

formation mechanism. There are two main theories as to how giant planets form in protoplanetary discs, the core accretion and gravitational instability theories (Pollack et al., 1996; Boss, 1997; Helled et al., 2014), both of which contain their own problems that can be difficult to solve. In neither case do we expect the growth of the planet to follow Equation 5.4. However, we expect that it is the growth timescale that is the most important parameter in determining whether the gap edge becomes unstable and any method of planet formation will have a corresponding growth timescale.

We also assume our planet does not migrate in the disc. Planet migration would clearly impact the planet formation process, as it is more difficult to form giant planets at large radial distances in the disc. As we expect planets to migrate inwards this means that it will be easier to form giant planets as the migrating planet approaches the warmer, more dense regions of the disc. Again, similarly to planet formation there are different regimes of planet migration through which a planet will progress as its mass increases. These are Type I migration for low mass planets (Goldreich & Tremaine, 1980) and Type II migration for more massive, gap opening planets (Lin & Papaloizou, 1986). As a planet grows it will initially migrate at high speed via Type I migration until it begins to open a gap in the disc and transfers to Type II migration, at which this speed will be reduced (Ward, 1997). Despite this reduction in migration rate the predicted Type II migration rate is still shorter than the lifetime of the disc, implying that most planets undergoing Type II migration will be absorbed by the central star (Hasegawa & Ida, 2013). This is problematic for planet formation models, as observational evidence dictates that gas giant planets are more common at orbital distances $R > 1$AU (Mayor et al., 2011; Cassan et al., 2012; Fressin et al., 2013; Santerne et al., 2016), which could not occur if this predicted timescale is correct. Current models predict that giant planets must form at $R > 20$AU to survive this migration (Coleman & Nelson, 2014). Hence, it is expected that something is limiting the rate of Type II migration (Nelson et al., 2000). Due to its complex nature, we neglect planet migration in this initial study and leave accounting for it to future work.

The impact of the RWI on migrating planets is still largely unexplored, however, there has been some recent investigation into its effects (McNally et al., 2019). As the planet's wake migrates with the planet, the effect of the shock will be more spread out over the disc, possibly leading to a gap edge that is less steep. Hence, it is possible that the migration of the planet may act to suppress the RWI. Additionally, the relation between gap opening and planet migration is not well understood. It has been shown that for a migrating planet to open a gap, the gap formation

timescale must be shorter than the migration timescale, otherwise a gap will not be formed (Malik et al., 2015). This can be difficult considering the rapid rate of planet migration. Also, growing a planet in the disc will cause the gap formation timescale to be longer than if the planet was initialised at its final mass. If the migrating planet cannot open a gap, then the RWI will not be excited as no steep gap edge is formed. Nevertheless, it has been shown that migrating giant planets can still open gaps in the disc material (Crida & Bitsch, 2017).

In addition to our simplistic approach to planet growth and migration, the systems we simulate are significantly less complex than those we are comparing our results to. We use a constant viscosity across our disc and hold our planets on a circular orbit with zero eccentricity. Most importantly we only consider a single planet within the disc, whereas observationally we often see tracers of multiple planets, such as the multiple dark bands in both HL Tau and TW Hydrae. However, in discs such as HL Tau the dark bands are far apart, so that if they were gaps formed by planets we can safely treat them as seperate entities. Also, the vortices formed as a result of instability excitation in our simulations always occurs on the outer edge of the gap formed by the planet. This is often the case in observational results, but not always as we can also see vortices formed at the edge of cavities and also interior to the planet. This should have little implication for comparison with our results, but shows that there are differences between our simulations and the systems we observe.

We perform our simulations using a simplified, locally isothermal equation of state. In further work we intend to extend this investigation to a non-isothermal equation of state, using a cooling timescale. However, when considering non-isothermal discs, the cooling timescale is a parameter that is difficult to estimate as it is dependent on disc opacity. The cooling timescale can be chosen to ensure initial equilibrium between (viscous) heating and cooling (e.g. Pierens & Lin, 2018). More realistic cooling prescriptions can be constructed under the same assumption (e.g. Faure et al., 2015).

Another factor that may impact the validity of our comparisons with observational results is that we simulate only the gas in the disc, while observationally we can only view the emission from the dust in the disc. This means that unless the dust and gas behave very similarly there may be some difference between our predictions and observations. It is already known that it is easier to form gaps in dust than in the gas (Paardekooper & Mellema, 2004), therefore it is possible that gaps we see from observational results may still contain gas.

We make no predictions concerning the lifetime of the vortices arising as the result

of instability excitation in our simulations. We find that some vortices formed during planet growth will mostly disperse before the simulation has elapsed, but in other cases vortices will still be present when the simulation has been finished. We note that the lifetime of a vortex is difficult to predict. There are a number of factors that can reduce the lifetime of vortices, such as the viscosity (de Val-Borro et al., 2007; Ataiee et al., 2013; Fu et al., 2014a; Regály et al., 2017), dust feedback (Johansen et al., 2004; Inaba & Barge, 2006; Lyra et al., 2009; Fu et al., 2014b), disc self-gravity (Regály & Vorobyov, 2017a; Pierens & Lin, 2018) and the streaming instability (Raettig et al., 2015). The lifetimes of vortices formed by giant planets has been investigated and were found to be strongly dependent on planet mass and viscosity, capable of sustaining vortices up to 10^4 orbits in some cases (Fu et al., 2014a). It has also been shown that dust build up in vortices can act to speed up their dispersion (Railton & Papaloizou, 2014). Vortices in protoplanetary discs are very efficient at trapping dust (Barge & Sommeria, 1995), however, the dust feedback can destroy the vortex if the dust density becomes too large. This occurs due to the excitement of a dynamical instability which destroys the potential vorticity minimum, a prerequisite for sustaining the vortex. Hence, when this is removed the vortex quickly dissipates (Fu et al., 2014b). It has also been shown that a dust vortex will live longer than a gas vortex, hence after the gas vortex dissipates we expect to still see asymmetries in the dust. The lifetime of asymmetries in the dust can be four times longer than those in the gas (Fu et al., 2014b). As our simulations only consider the gas in the disc, it is possible that the vortices in the dust would be observable for a longer period. However, it is also possible that the vortices would dissipate faster if we considered their tendency to trap dust. If a brightness asymmetry exists in an observation, we know there is the potential for a vortex to exist and can use this to constrain the mass of any potential planets in the disc. If we do not see a potential vortex, then we can use their absence to constrain the mass of any potential planets in the disc. However, we must be aware that it is possible for a vortex to have existed earlier in the disc's lifetime and has since dissipated. This is an unfortunate consequence of only being able to observe a disc at a single point during its evolution. Additionally, dust enhancement of vortices can make these vortices more visible in the dust than in the gas. As we have discussed, observational results show the dust in the disc, whereas our simulations represent the gas in the disc. Therefore, it is possible that weak vortices in our simulations could correspond to more visible vortices after accounting for dust enhancement. Hence, we have been particularly careful analysing marginal cases with weak vortices in establishing the planet growth timescale at which vortices disappear.

Finally, we do not include any effects of disc self-gravity in our model. However, it is known that disc self-gravity can have major effects on vortex formation and lifetime (Regály & Vorobyov, 2017a). It is also known that the indirect potential caused by a vortex can aid the onset of the RWI (Regály & Vorobyov, 2017b). Hence, our results are not applicable to massive discs for which self-gravity becomes important, but this would be an interesting direction to take this study.

5.7 Conclusions

In this chapter we have derived a relation between the viscosity of a disc and the mass of planets forming in this disc as a function of the growth timescale of the planet. We intend this relation to help break the degeneracy in planet masses and viscosities that can adequately explain observed gaps in protoplanetary discs. If both of these parameters are unknown then almost any gap can have a planetary explanation. Using this relation we can make predictions regarding the mass of planets in protoplanetary discs in which we observationally detect the presence or absence of vortices, assuming standard disc ages and viscosities. We find that our results are consistent with predictions made regarding a variety of observations, both for discs containing and devoid of vortices, including well documented discs such as HL Tau (ALMA Partnership et al., 2015) and TW Hydrae (Andrews et al., 2016). However, we must stress that our relation is very tentative and is by no means a definitive prediction, due to both the inherent inaccuracies in the method from which it was derived and the assumptions made in our simulations.

Furthermore, using our relation we can make the prediction that it is not necessary for planets to have masses on the order of one Jupiter mass to form vortices within low viscosity discs. While massive planets do indeed form these vortices, the onset of them occurs for planets that are significantly less massive, for acceptable disc viscosities and ages. We find that planets of sub-Saturn mass can form vortices, down to a few Neptune masses. Whether or not planets of these low masses can explain the dark bands we see in many of these observations is a different question. We are in no way precluding giant Jupiter mass planets from being the cause of these, however, this prediction must be something to keep in mind, especially as we delve deeper into attempting to directly image these planets within their gaps. We also hope that our results are a step on the way to breaking the degeneracy in planetary explanations for observational gaps, by imposing limits on the mass of the planet and/or the viscosity of the disc based on the presence or absence of vortices

in the disc.

6 The generation of planet induced vortices in non-isothermal protoplanetary discs

Abstract

A massive planet in a protoplanetary disc will open a gap in the disc material. If the gap edge is steep it can be hydrodynamically unstable, which results in the formation of vortices. Therefore, vortices can act as tracers for the presence of planets in observational results. In order to predict whether a planet can exist in a disc, based on the presence or absence of vortices, we must understand well the processes in the disc that can impact the formation of vortices. One such process is the temperature evolution of the disc material. Often it is assumed that discs are isothermal, however, a more realistic interpretation is of a disc that is non-isothermal, with a finite thermal relaxation timescale. Realistic values of the thermal relaxation timescale are often difficult to obtain. We have investigated the impact of the thermal relaxation timescale on vortex formation in a disc containing a gap opening planet. We have found that as the thermal relaxation timescale is increased the disc goes through a number of different regimes which impact the behaviour of the vortices formed at the gap edge. We then predict how each of these behaviours may impact the formation of vortices in a disc with a growing planet, such as those we can potentially see in observational results.

6.1 Introduction

In a protoplanetary disc, planets can form due to accretion of dust and gas. These planets excite density waves in the disc material that transport angular momentum from the planet and deposit it in the disc. Hence, planets exert a torque on the disc

material. The more massive the planet, the stronger the torque exerted. Should this torque be deposited locally (Lin & Papaloizou, 1993; Bryden et al., 1999) and the resultant torque be stronger than the viscous diffusion of the disc material (Lin & Papaloizou, 1979; Goldreich & Tremaine, 1980; Takeuchi et al., 1996; Crida et al., 2006) then a gap will be formed around the orbit of the planet. These are the thermal and viscous criteria for opening a gap.

Much of our understanding of protoplanetary disc dynamics and planet-disc interaction comes from numerical simulations (Lin & Papaloizou, 1986; Crida et al., 2006; Armitage, 2011; Duffell & MacFadyen, 2013; Kanagawa et al., 2015). However, in recent years observational results have begun to probe the structure of protoplanetary discs at high resolution. While these results provide some support to previous theoretical predictions, they also open up many new questions. The most well known observations of protoplanetary discs are those of the systems HL Tau (ALMA Partnership et al., 2015) and TW Hydrae (Andrews et al., 2016), observed using the Atacama Large Millimetre Array (ALMA). Both of these discs show similar substructure, despite their difference in age. Even more recently, surveys of protoplanetary discs have shown the existence of substructure across a variety of disc ages and masses (Long et al., 2018; Andrews et al., 2018; van der Marel et al., 2019). A substructure of particular interest present in many of these discs are dark bands, implying the presence of a low density region. There are a number of potential explanations for these regions, but a commonly cited theory is the aforementioned gap opening by massive planets within the disc. Often, a planetary interpretation of this structure is assumed (e.g. Huang et al., 2018) and parameters for these planets, such as their mass, are predicted. Unfortunately, there is a degeneracy in the disc-planet parameters that can explain the substructures we can see, as many of these parameters are not well constrained (e.g. Mulders et al., 2013). For example if both the disc viscosity and the planet mass is unknown, almost any gap can be explained by a planet. We have attempted to address this degeneracy and provide some constraints on planet masses and viscosities in Chapter 5 (Hallam & Paardekooper, 2019).

An important factor in our previous work on this topic (Chapter 5, Hallam & Paardekooper, 2019) is the presence of vortices induced by gap opening planets in protoplanetary discs. A prominent instability in protoplanetary discs is the Rossby Wave Instability (RWI), which can occur due to a steep density gradient in the disc, such as those at the gap edge. This can result in the formation of vortices at the gap edge (Lovelace et al., 1999; Li et al., 2000). Hence, these vortices can act as a tracer for massive gap opening planets. Indeed, we can see potential vortices existing at the edge of a gap or cavity in many observations (van der Marel et al., 2013; Fukagawa

et al., 2013; Pérez et al., 2014; Marino et al., 2015; Kraus et al., 2017; Dong et al., 2018; Cazzoletti et al., 2018), which could very easily be due to the existence of planets within the disc. However, when considering vortex formation in discs, we must be aware that many simulations do not grow planets over a particular timescale, rather start planets in the disc at their maximum mass. It has been shown that by increasing the growth timescale of a planet, we can eliminate vortex formation by allowing the disc to adjust viscously to a slower opening gap (Hammer et al., 2017). Hence, it may be possible to rule out planets of a certain mass from existing in the disc based on the presence or absence of vortices (Hallam & Paardekooper, 2019).

In the previous chapter (Chapter 5, Hallam & Paardekooper, 2019), we presented a relation between disc viscosity and planet mass for which vortices were not formed in the disc. We used this to predict the maximum masses of planets that could exist in certain observed discs, based on the presence or absence of vortices in these discs. We then compared these predictions with the predictions of others, in which planetary explanations were often degenerate. We found that in general, our results ruled out the more massive, $\geq 1 M_J$, planets from axisymmetric discs containing no vortices. We also found that for discs with low viscosity (as is expected of protoplanetary discs) planets significantly smaller than a Jupiter mass can cause vortices to form. These planets could even be as small as a few Neptune masses.

In this chapter we extend the results of the previous chapter. Our studies in the previous chapter use a locally isothermal equation of state. Here, we investigate the behaviour of the disc using a non-isothermal equation of state with cooling timescales implemented similarly to our setup in Chapter 4. However, in this chapter we investigate a number of cooling timescales significantly longer than the one investigated in Chapter 4. We observe interesting behaviour across the range of cooling timescales investigated. From this we predict the impact moving to the non-isothermal regime has on planet growth timescales and vortex formation as an observational tracer of planets. We also present an interesting phenomenon found when using extremely long cooling timescales, which potentially links to the results presented in Chapter 3, that gaps are maintained at marginal stability.

This chapter is arranged as follows. In Section 6.2 we discuss the relevant equations solved to simulate disc evolution. In Section 6.3 we discuss the numerical setup for our simulations. In Section 6.4 we discuss relevant instabilities to our work in this chapter. In Section 6.5 we present our results. In Section 6.6 we discuss our results and any assumptions we have made during our simulations. Finally in Section 6.7 we conclude.

6.2 Basic Equations

In this chapter we perform a parameter study to find the impact of a thermal relaxation timescale on the RWI in a protoplanetary disc containing a gap forming planet. This work has some similarities to that of Chapter 5 (Hallam & Paardekooper, 2019). However, in this chapter we implement a thermal relaxation timescale and a non-isothermal equation of state. We investigate a number of thermal relaxation timescales to compare to the previous work and as such our simulations remain in two dimensions. Additionally protoplanetary discs are expected to be thin with respect to the Hill radius of the gap forming planet. Therefore, we believe two dimensions is sufficient to draw meaningful conclusions from.

6.2.1 Two dimensional protoplanetary disc

The continuity equation for the surface density, Σ, evolution of a protoplanetary disc is given by

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0, \tag{6.1}$$

where $\mathbf{v}$ is the velocity field. We simulate the evolution of a protoplanetary disc's surface density by solving the continuity equation and the two dimensional Navier-Stokes equation for the motion of the planet-disc system,

$$\Sigma \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P - \nabla \cdot \mathbf{T} - \Sigma \nabla \Phi, \tag{6.2}$$

where $\mathbf{T}$ is the Newtonian viscous stress tensor, P is the pressure and Φ is the gravitational potential of the planet and star system. We solve the energy equation,

$$\frac{\partial e}{\partial t} + \nabla \cdot e\mathbf{v} = -P\nabla \cdot (\mathbf{v}) + \mathcal{C}, \tag{6.3}$$

with an equation of state $P = (\gamma - 1)e$ where e is the volumetric internal energy and $\gamma = 1.4$ is the ratio of specific heats. $\mathcal{C}$ is a cooling term, in which we introduce a thermal relaxation timescale, τ, as discussed in Section 6.2.2.

6.2.2 Thermal Relaxation Function

We deviate from a purely adiabatic simulation by applying a thermal relaxation function at constant density, similarly to Chapter 4. This is given by

$$\mathcal{C} = -\frac{e - e_i \frac{\Sigma}{\Sigma_i}}{\tau}, \tag{6.4}$$

where e is the volumetric internal energy, $\tau = 2\pi/\Omega_0$ is the thermal relaxation timescale and the subscript i denotes initial and therefore equilibrium values. As τ is constant and set at the location of the planet the rate of cooling will vary radially across the disc when compared to the orbital timescale at that location. In general we are only interested in cooling around the location of the planet and at the edges of the gap region. This is applied, again similarly to Chapter 4, using implicit time stepping to cool the disc in the form

$$e^{n+1} = e_i \frac{\Sigma^n}{\Sigma_i} + \frac{e^n - e_i \frac{\Sigma^n}{\Sigma_i}}{1 + \frac{\Delta t}{\tau}}, \tag{6.5}$$

where the superscript n denotes the timestep at which the relevant field is evaluated. This is a similar method of applying a thermal relaxation function to that in Chapter 4, however, in this setup we do not include gap edge illumination to heat the disc.

6.3 Numerical Setup

We run our simulations in two dimensions using the code FARGO3D. This is a magnetohydrodynamic code which is designed to simulate disc evolution in one to three dimensions by solving the hydrodynamic equations of motion. FARGO3D's C to CUDA translator makes it a good choice for investigating gap opening planets, as this allows it to be run on Graphics Processing Units (GPUs). GPUs have limited memory, but decrease computational time significantly, allowing moderate resolution simulations to be run for extended periods of time. For more details see Benítez-Llambay & Masset (2016).

In our simulations we use a disc model with radial domain $0.3 < R/R_0 < 6.0$ and an azimuthal domain $-\pi < \phi < \pi$ with resolution 285 by 596 cells respectively. We use logarithmic cell spacing in the radial direction. We set an initial surface density distribution $\Sigma_{\text{int}} = \Sigma_0 (R/R_0)^{-1/2}$ where $\Sigma_0 = 2.67 \times 10^{-3}$. We investigate a single setup containing a planet of mass $q = M_p/M_* = 2.5 \times 10^{-4}$ and constant viscosity $\nu = 3.0 \times 10^{-6}$. The aspect ratio of the disc is initially set to $h = H/R = 0.05$, where H is the disc scale height. This corresponds to an $\alpha = 1.2 \times 10^{-3}$ at the location of the planet using the alpha prescription $\nu = \alpha c_s H$ of Shakura & Sunyaev (1973), where c_s is the sound speed. Reflecting boundary conditions were used with wave killing zones in the regions $0.3 \leq R/R_0 \leq 0.585$ and $5.43 \leq R/R_0 \leq 6.0$ similarly to de

Val-Borro et al. (2006), meaning that excited waves are damped before reaching the edge of the simulation. The planet is held on a fixed circular orbit with no migration or self-gravity accounted for. Deviation from a purely adiabatic simulation arises from the applied thermal relaxation function as described in Section 6.2.2. We use a variable thermal relaxation timescale τ between simulations. We let our simulations run for a number of orbits, usually in the range $500 - 1000$.

6.4 Instabilities

Before presenting our results we will first briefly discuss two particularly relevant instabilities. These are the RWI and the Subcritical Baroclinic Instability (SBI). Using a particularly long cooling timescale makes the disc more prone to instabilities (Richert et al., 2015). Hence, we must be aware that we are more likely to encounter such instabilities when leaving the isothermal regime.

6.4.1 Subcritical Baroclinic Instability

The SBI is an entropy driven instability (Lesur & Papaloizou, 2010) that can occur in protoplanetary discs that have a radially unstable entropy gradient and a carefully selected thermal diffusion timescale. As this instability is subcritical, it also requires an initial finite amplitude perturbation, such as a vortex. The SBI can cause the vortex to be self-sufficient, depending on the interaction between buoyancy forces and rate of cooling on material within the vortex. Consider a particle of material circulating the vortex. It begins at a point with a given temperature. As it is quickly displaced radially outwards it retains the temperature and density of its original position, so is hotter and lighter than its new surroundings. Therefore, it feels a buoyancy force directed radially outwards. Next, the particle circulates azimuthally around the vortex at this new radius and on this journey it reaches thermal equilibrium with its surroundings. Now it has crossed the vortex azimuthally it will move radially back inwards while retaining its new temperature and density, similarly to its last radial displacement, feeling a buoyancy force directed inwards. Now it has returned to its original radial position, however, it is displaced azimuthally and is cooler and heavier than its surroundings. It now circulates azimuthally back to its initial position, both radial and azimuthal, while reaching thermal equilibrium with its surroundings. This overall process causes and amplification of the vortex due to acceleration by buoyancy forces (Lesur & Papaloizou, 2010).

Clearly from this circulation the thermal relaxation timescale is very important.

Should the thermal relaxation timescale be too large the material would never reach equilibrium with its surrounding on the azimuthal parts of its journey. If the thermal relaxation timescale is too small, the material would reach equilibrium as it is displaced radially and hence would not feel the buoyancy forces. Therefore, it is clear that a carefully selected thermal relaxation timescale can result in the excitement of the SBI (Lesur & Papaloizou, 2010).

6.4.2 Rossby Wave Instability

The RWI can occur in protoplanetary discs due to a steep gradient in the radial profile of the azimuthal velocity of the disc. In a disc containing a gap opening planet, this corresponds to the location of the gap edge. For a non-isothermal disc to be unstable to the RWI there must be a local maximum or minimum in the key function (Lovelace et al., 1999),

$$\mathcal{L}(R) = \mathcal{F}S^{\frac{2}{\gamma}}, \tag{6.6}$$

where $\mathcal{F}$ is approximately the inverse of the potential vorticity, $S = P/\Sigma^{\gamma}$ is the entropy and γ is the ratio of specific heats. In addition to a maxima or minima in this function, it must also exceed a threshold value before the RWI will occur (Lovelace et al., 1999; Ono et al., 2016). Nevertheless we can see that a gap opening planet will impact the key function, as it has surface density dependence. However, it is difficult to predict the stability of the system from this function alone.

We determine if our discs are unstable to the RWI as we have previously (Chapter 5), using local minima in the potential vorticity of the disc as a tracer for vortex formation, which is the non-linear outcome of the RWI. The potential voricity is given by

$$\xi = \hat{\mathbf{z}} \cdot \frac{\nabla \times \mathbf{v}}{\Sigma}. \tag{6.7}$$

Again, we also investigate the orbit-to-orbit potential vorticity difference in the disc. As vortices orbit at a different speed to the planet, sudden differences in the maximum orbit-to-orbit potential vorticity difference of the disc can be a good indicator of the presence of vortices.

We limit our potential vorticity calculations to the area close to the gap edge, in which the vortices are formed. We also exclude a region $\Delta\phi \approx 0.95$ either side of the planet. This is because even in cases where there is little to no vortices small fluctuations in the location of the gap edge or the planet's wake can cause very

large potential vorticity differences for steep gap edges. This can cause a false result implying the presence of vortices when there are none, therefore we take care to exclude this from our sampling region.

6.5 Results

We investigate the behaviour of a disc-planet system for a wide range of thermal relaxation timescales, from near isothermal to near adiabatic. We focus on observing the maximum of the orbit-to-orbit potential vorticity difference for each of these thermal relaxation timescales, as we use this as a tracer for the formation of vortices. Using the results we find here, we can predict the effect of these longer thermal relaxation timescales on the formation of vortices in discs containing growing planets. For the purposes of this investigation we use only one disc-planet system, with $h = 0.05$, $\nu = 3.0 \times 10^{-6}$ and $q = 2.5 \times 10^{-4}$. In the isothermal regime, this system shows signs of vortex formation within a few 100 orbits of initialising the simulation.

We show the results of varying the thermal relaxation timescale in Figure 6.1. We find that, unsurprisingly, for extremely low thermal relaxation timescales our results do not differ hugely from the isothermal result. Indeed, in the limit $\gamma \to 1$ and $\tau \to 0$ we reproduce exactly the isothermal result. As we move further into the non-isothermal regime it begins to get more interesting. Increasing the thermal relaxation timescale appears to move the system more towards stability, which is unexpected as systems with long thermal relaxation timescales are more prone to becoming unstable (Richert et al., 2015). However, this relation does not remain linear and seems to reach a maximum stability (which is stable and absent of vortices) when the thermal relaxation timescale is of the order of 1 orbit at the location of the planet. Beyond this, increasing the thermal relaxation timescale trends the system more towards instability, as we would expect from a system with less efficient cooling. From Equation 6.6 we know that the stability of the system against the RWI is dependent on the entropy. As a result, we investigate the azimuthally averaged entropy profiles of the disc. We find that the entropy profiles of the systems with $\tau \leq 1$ orbit show inversion in the gap region, whereas the entropy profiles of those of the order $\tau \geq 10$ orbits have minima in the gap region, surrounded by maxima at the gap edges. The entropy profiles for a number of thermal relaxation timescales can be seen in Figure 6.2. These profiles are at $t = 75$, within the initial peak shown in Figure 6.1, however, the main features present in Figure 6.2 are consistent regardless of sampling time. Figure 6.2 shows that for $\tau > 1.0$ or $\tau < 1.0$ there is a

steeper gradient in the entropy profiles. In these regimes we are moving towards a more unstable system. This implies that the entropy variation stabilises against the excitement of the RWI when τ is the on the order of 1 orbit. We also find that we can reproduce the result of (Richert et al., 2015), using the same planet mass and viscosity, in which a purely adiabatic disc becomes entirely unstable.

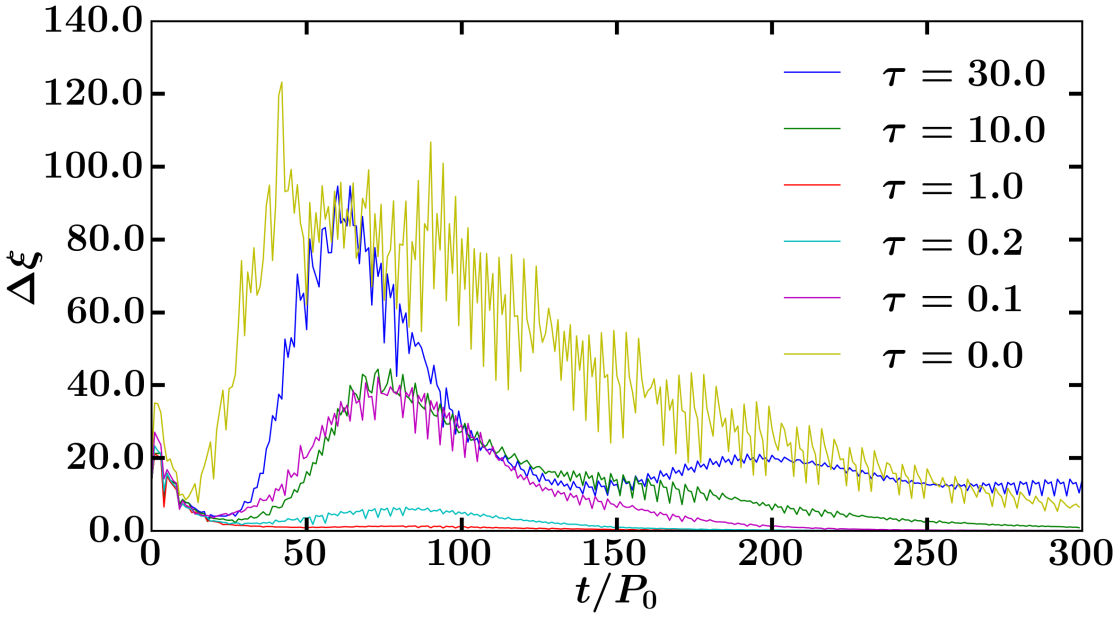

Figure 6.1: Orbit-to-orbit potential vorticity difference for a number of thermal relaxation timescales, including the isothermal case. These timescales correspond to the near isothermal and efficient cooling regimes. Here we can see that the system peaks in stability at $\tau = 1.0$ and increasing or decreasing τ from this point causes the disc to tend more towards instability.

Increasing the thermal relaxation timescale from the order of 1 orbit to the order of 10s of orbits we see the vortices return, similarly to in the isothermal case. The magnitude of the orbit-to-orbit potential vorticity difference and hence the strength of the vortices, increases with thermal relaxation timescale until reaching $\tau \approx 80$. At this point the increase in orbit-to-orbit potential vorticity difference with thermal relaxation timescale is negligible at best, even as $\tau \to \infty$. For smaller thermal relaxation timescales in this range, $\tau < 40$ orbits, the system rapidly approaches stability after the initial vortices disperse. For $\tau = 40$ orbits, we find that weaker vortices are longer lived in the disc, but the system tends towards stability over 1000 orbits. This appears to be a transition case, as for longer thermal relaxation timescales, in the range of $50-60$ orbits, we see long lived weak vortices that appear to survive as the system reaches equilibrium over 1000 orbits. These vortices are

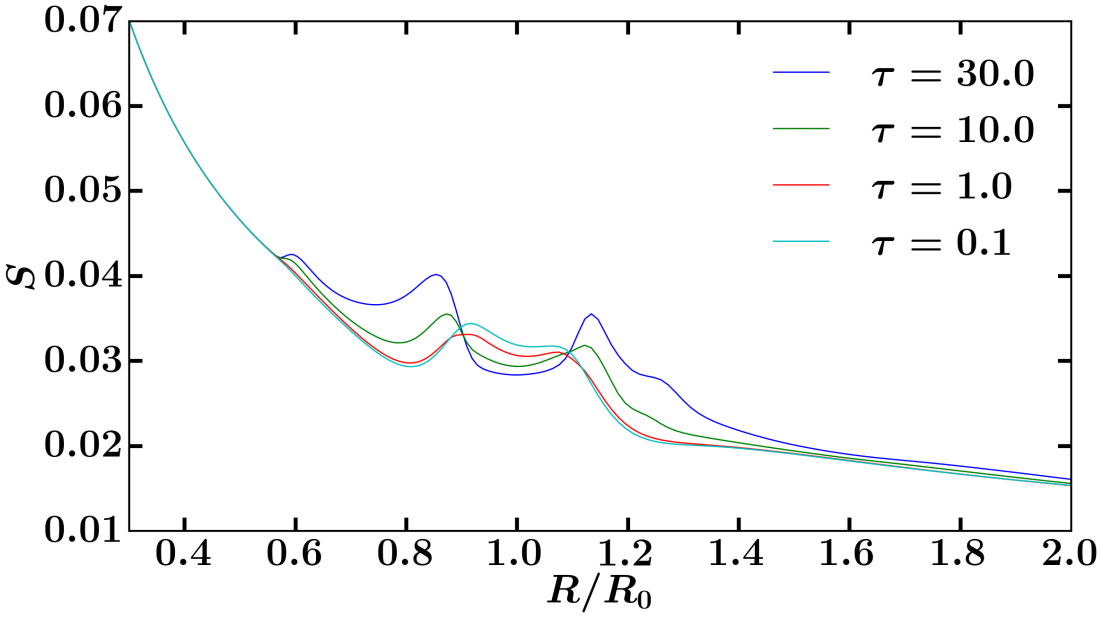

Figure 6.2: Entropy profiles for a number of cases shown in Figure 6.1, at $t = 75$ orbits. Here we can see the steeper gradients in the $\tau \neq 1.0$ cases.

elongated and weak and do not fulfil the criteria for the SBI, regarding the vortex crossing time and thermal relaxation timescale. These are shown in Figure 6.3.

Further increasing the thermal relaxation timescale we now enter the regime in which the thermal relaxation timescale is of the order of 100 orbits. Here we observe another interesting phenomenon. Starting at $\tau \approx 80$ orbits we find that after the initial peak in the orbit-to-orbit potential vorticity the system tends towards stability, before a secondary peak occurs. After this peak the system tends towards stability for a number of orbits, before the process occurs again. We observe this for two thermal relaxation timescales $\tau = 80$ and $\tau = 120$ orbits. Interestingly, we begin to see this phenomenon at thermal relaxation timescales as low as $\tau = 30$ orbits, where the orbit-to-orbit potential vorticity difference oscillates before levelling out to an equilibrium with or without vortices. We see these oscillations grow larger with increasing thermal relaxation timescale, until we observe the additional peaks we can see in $\tau = 80$ and $\tau = 120$ orbit systems, shown in Figure 6.4. We find that these peaks occur at approximately every 2τ after the secondary peak occurs and are stronger for the larger thermal relaxation timescale. However, we note we have not investigated beyond $\tau = 120$ orbits, outside of our comparison study with (Richert et al., 2015). Hence, we postulate that this secondary formation of gap edge structure arises from the similarity between the gap formation timescale and thermal relaxation timescale in these setups, as these are both on the order of 100

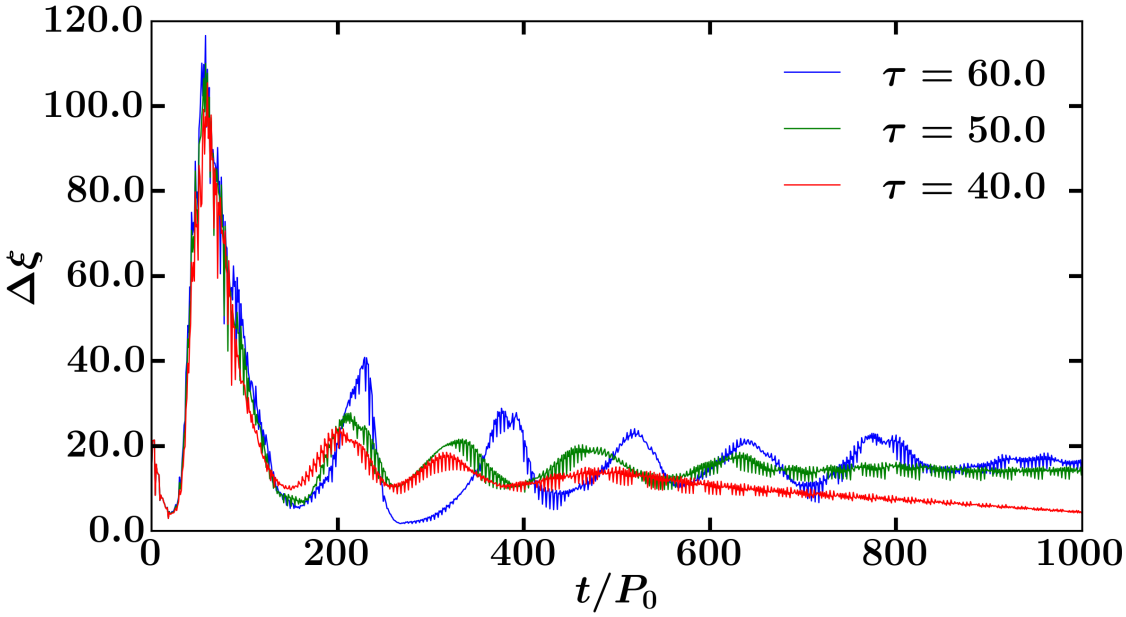

Figure 6.3: Orbit-to-orbit potential vorticity differences for longer thermal relaxation times. It can be seen that the $\tau = 50$ and 60 cases show signs of weak vortices even at equilibrium, whereas the $\tau = 40$ case eventually heads towards stability. Also present are stability oscillations beginning at $t \approx 200$, which eventually die off.

orbits. If these timescales are similar, we may expect that the effect of one will become dominant, then the other. This may result in a cycle which can smooth out and reform structure at the gap edge, resulting in the behaviour shown in Figure 6.4. The presence of this gap edge structure being formed repeatedly could potentially link with the ideas of Chapter 3 (Hallam & Paardekooper, 2017), that gaps are maintained at marginal stability.

6.6 Discussion

We have studied the impact of a non-isothermal equation of state and a variety of thermal relaxation timescales on vortex formation in protoplanetary discs containing gap forming planets. Here we shall predict how these effects could impact vortex formation in protoplanetary discs for which the planet grows in mass over a planet growth timescale, as described in Chapter 5. We will also justify any assumptions made in our simulations. Many of the assumptions in the previous chapter still apply to our predictions here, as we make predictions for discs with growing planets based on our previous results.

Observationally, the presence or absence of vortices in a radially structured pro-

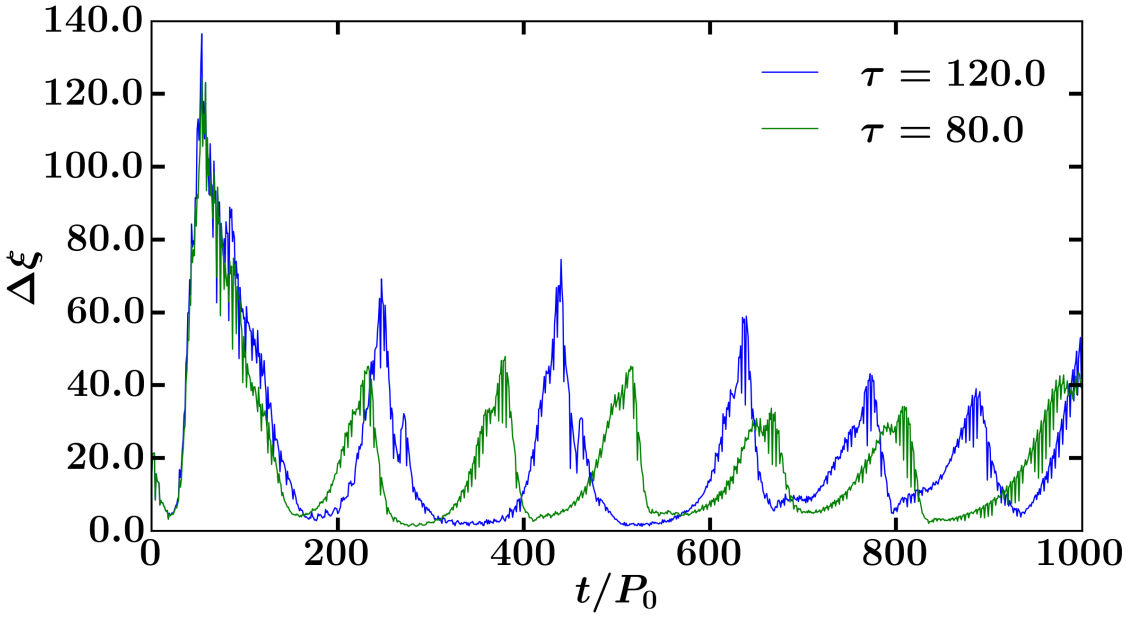

Figure 6.4: Orbit-to-orbit potential vorticity difference for the longest investigated cooling timescales, outside of the purely adiabatic regime. Here it can be seen that the oscillations present at lower cooling timescales (Figure 6.3) have overtaken these systems and are showing little to no indication of damping.

toplanetary disc can give us a great deal of information regarding the potential presence of an unseen planet. As shown in Chapter 5, it may be possible to rule out a planetary explanation for disc substructure if the disc is axisymmetric and vortices are absent. Hence, it is important to understand vortex formation in realistic discs. In Chapter 5, we investigated a locally isothermal disc setup, however, it is more realistic to assume the disc is non-isothermal, with the cooling dictated by a thermal relaxation timescale. Realistic values of this thermal relaxation timescale are often difficult to obtain due to its dependence on disc opacity. The opacity of a protoplanetary disc is set by the smallest size dust grains and during the planet formation process small grains grow into larger grains. Hence, it is difficult to determine the opacity of a disc observationally. Therefore, we have explored a wide range of thermal relaxation timescales, from near isothermal to 100s of orbits.

We have observed a number of distinct regimes over the investigated thermal relaxation timescales. Initially, for near-isothermal thermal relaxation timescales, the isothermal result is almost exactly reproduced. At $\tau \approx 1$ orbit, we find that the system is driven towards stability. This is particularly interesting, as $\tau \approx 1$ is very rapid cooling and so we would expect similarity to the isothermal result. As this is not the case, our results imply that even very short cooling timescales are important.

Increasing the cooling timescale so that $\tau \approx 20 - 40$ orbits we find that the system moves back towards vortex formation. The magnitude of these vortices can be lower than in the isothermal regime and the system will return to equilibrium absent of vortices. For $\tau \approx 50 - 60$ orbits we find a similar result, however, weak vortices remain as the system reaches equilibrium. At larger thermal relaxation timescales, $\tau \approx 80 - 120$ orbits we see an interesting feature in which, following the primary vortex formation, the system oscillates between a stable gap edge and the formation of gap edge structure. Finally, at extremely long thermal relaxation timescales such that the system is adiabatic, the whole disc becomes unstable and any discernible gap edge structure is lost.

Based on these results we can predict the effect these non-isothermal conditions have on vortex formation for a growing planet. Clearly, if a disc has a thermal relaxation timescale low enough that it is close to the isothermal regime there will be no additional effects to consider beyond those in Chapter 5. Should the cooling timescale be on the order of $\tau \approx 1$ we would expect the magnitude of the vortices to be reduced. As our results show, this does not necessarily mean there will be no vortices, as $\tau < 1$ and $\tau > 1$ still shows a sudden increase in orbit-to-orbit potential vorticity, just of a reduced magnitude compared to the isothermal result. This implies that efficient, non-isothermal cooling will result in weaker vortices created by the planet as it forms. However, this does also present the possibility that systems that would show vortices in the isothermal regime may not if the cooling is slightly less efficient. Compared to the isothermal regime, it also appears that such vortices will not be as long lived, however, this is difficult to predict as by growing the planet mass the vortices will be formed over a longer period of time.

Moving to considerably less efficient cooling, $\tau > 40$, we find that weak vortices exist in the disc even when the system has reached equilibrium. As we have discussed in Chapter 5, even weak vortices are important to consider as dust enhancement of gaseous vortices could considerably improve their visibility (Barge & Sommeria, 1995). With this in mind, the longevity of these vortices means that in systems with less efficient cooling the vortices created during planet formation may be visible even significantly after the planet has formed. Hence, we may not need to be concerned with 'missing' a vortex signature of a planet due to the vortex dissipating over time after its excitation.

In the last interesting case, for $\tau \approx 80 - 120$ orbits, even less efficient cooling, we observe the system oscillating between stability and a structured gap edge. We have already postulated that this is a result of the disc evolution and temperature evolution occurring on similar timescales. When looking for vortices acting as a

planetary tracer in protoplanetary discs, we expect these vortices to appear at the edge of the planet gap. The structure formed in this regime also occurs at the gap edge and may be less detectable observationally. We also expect that this phenomenon will impact vortex formation as the planet grows in mass and forms a gap. Hence, we predict that the planet growth timescale would need to be longer than the thermal relaxation timescale, otherwise vortex formation may be masked by the temperature evolution and surface density evolution alternating in impacting the formation of structure at the gap edge.

Finally, in the purely adiabatic regime all discernible gap edge structure is lost and vortices form across the disc. In particularly extreme cases (Richert et al., 2015) all structure across the whole disc is lost. This makes it very difficult to compare these results to observations. We will stay clear of making predictions in this regime, other than to say that protoplanetary discs observed are more likely to be better explained by non-isothermal models with a finite thermal relaxation timescale. The adiabatic regime has been studied in more detail by Richert et al. (2015).

In addition to the fact that the above are only predictions based on simulations of non-isothermal systems, we must also stress that only a single disc setup has been investigated. Hence, it is possible that the behaviour witnessed is not consistent across all non-isothermal discs. For example, it may depend on planet mass or disc viscosity. Therefore, the predictions above are tentative due to both not knowing how the systems would actually behave with a growing planet and not knowing if the phenomenon observed is consistent across all discs.

One major assumption made here that is not present in Chapter 5 is that of a constant thermal relaxation timescale. This is measured in orbits at $R = R_0$, hence the cooling will be slower in the inner disc and faster in the outer disc relative to the orbital timescale in that region. As we are only particularly interested in the region surrounding the disc and gap, this is not an unreasonable assumption. However, this does become important when the disc structure becomes heavily modified, such as by the opening of a gap and the build up of material and energy deposited at the gap edge. This approximation will get worse as cooling becomes more inefficient. Hence, particularly long thermal relaxation timescales may be unrealistic approximations to the actual behaviour of the system.

Many of the assumptions made in Chapter 5 apply to our predictions here, as we base these predictions off of the results from the previous chapter. However, assumptions regarding planet growth do not apply to the results we present here, only the predictions we make. One assumption that is still relevant to our results here is that we only simulate the gas in the disc. As we only observe the dust

distribution of the disc we must bear this in mind as we make comparisons between observations and simulations. It is known that vortices can be enhanced by trapping dust within them (Barge & Sommeria, 1995), meaning that it is possible that weak vortices in our simulations can be more visible in the dust. It has also been shown that dust build up in vortices can act to speed up their dispersion (Railton & Papaloizou, 2014). This means that potentially long lived vortices such as those in the $\tau \approx 50 - 60$ orbit range may have shorter lifetimes than we observe and also may not actually be present as the disc reaches equilibrium.

6.7 Conclusions

In this chapter we have investigated gap formation in a non-isothermal regime including a thermal relaxation timescale. We have studied a range of thermal relaxation timescales, from almost isothermal to adiabatic, in order to determine the impact of this timescale on the behaviour of the system. Specifically we focus on the formation of vortices at the gap edge. Hence, we used the orbit-to-orbit potential vorticity difference in this region to act as a tracer for vortex formation, in a similar manner to Chapter 5. We choose to focus on vortex formation as we use these results to predict the impact of the non-isothermal regime on the results of Chapter 5. Namely, we predict the impact of the non-isothermal regime on the presence of vortices created by growing, gap forming planets in a protoplanetary disc. Our results here are very preliminary, in that we have only investigated a single disc setup and have not grown the planet mass over a planet growth timescale. Rather, we assume the findings from this single setup apply across all disc setups and predict how they may impact the system for a growing planet.

We have found that as we increase the thermal relaxation timescale from 0 to ∞ there are a number of distinct regions in which disc structure is modified, in addition to the near isothermal and near adiabatic regimes. In the near isothermal region we find that the disc behaves very similarly to the isothermal case. Similarly, in the near adiabatic regime the disc becomes largely unstable and forms vortices across it and in some cases all discernible disc structure is lost (Richert et al., 2015).

The remaining regions are more interesting and perhaps, more surprising. In the low thermal relaxation timescale regime, $\tau \approx 1$, we find that the disc is driven more towards stability and no vortices are formed. This means that the vortices become weaker as we increase the thermal relaxation timescale for $\tau < 1$. At slightly larger cooling timescales $\tau \approx 20 - 40$ we find that the disc is driven more towards vortex

formation again, although perhaps not as strongly as in the isothermal regime. Hence, increasing the thermal relaxation timescale for $\tau > 1$ results in stronger vortices. As we reach a critical point, $\tau \approx 50 - 60$, we find that it is possible to sustain very weak gap edge vortices even at equilibrium, meaning these vortices will be very long lived or even permanent. However, we find that these vortices do not satisfy the criteria for the SBI. At the largest investigated thermal relaxation timescales, $\tau \approx 80 - 120$, we find that gap edge structure is continuously formed and decays back to stability, resulting in secondary peaks in the orbit-to-orbit potential vorticity. This phenomenon appears to be present across the disc's lifetime, however, we have not investigated for extremely long timescales > 1000 orbits.

Applying these results to those of a previous chapter we find there are a few effects here that must be accounted for. Firstly, it seems unlikely that vortices will be particularly enhanced by the presence of less efficient cooling, at least not significantly enough to consider. Secondly, we find that vortex formation may actually be suppressed if the disc is non-isothermal and its cooling efficiency falls in the 'stability valley' provided around low to mid thermal relaxation timescales. For longer thermal relaxation timescales we found that vortices can survive at disc equilibrium. Hence, we may not need to consider vortex dissipation in protoplanetary discs with inefficient enough cooling. Finally, with extremely long thermal relaxation timescales we find that gap edge structure is continually formed and suppressed. This is possibly due to the surface density and temperature evolving on roughly equal timescales. Hence, this result may imply that for observable vortices to be present in the disc the planet growth timescale may need to be longer than the thermal relaxation timescale. Otherwise, vortex formation may be masked as the temperature evolution and surface density evolution alternate in impacting the formation of structure at the gap edge.

Overall, the effects of the non-isothermal regime are very interesting and complex. This means that it is difficult to predict with some certainty how the systems with growing planets will behave, especially when vortex formation is the main consideration. Protoplanetary discs are complex systems and identifying planets in these from their structure is no easy task. Hence, at least a basic understanding of the behaviours of these systems under different conditions is necessary to make informed predictions. Therefore, we believe we have shown the importance of understanding the effects of this regime and have provided evidence that this is an area that should be investigated in much greater detail.

7 Conclusions

In this thesis we have investigated a variety of problems in the area of planet-disc interactions, specifically those related to the interactions between a giant, gap forming planet and its surrounding disc. We have looked into theoretical problems, such as the discrepancy between one and two dimensional numerical models, in addition to problems that exist in observational results, such as the degeneracy in planetary explanations for observed disc structure. Hence, in this thesis we have covered a wide range of areas in which planet-disc interactions are important and while there are many similarities across chapters, each pertains to a separate, distinct problem.

7.1 Chapter 3

Chapter 3 concerns the role that gap edge instabilities have in setting the depth of gaps in protoplanetary discs. In this chapter we investigated the well known discrepancy between the depth of gaps in one and two dimensional simulations. We studied this by applying a one dimensional gap forming torque density distribution radially across a two dimensional disc that contains no planets. This mimics the effect of gap formation in one dimension, in two dimensions. We used a variety of different one dimensional torque density distributions to form the gap in two dimensions (Lin & Papaloizou, 1986; D'Angelo & Lubow, 2010). Using these torque distributions we can make significant progress to eliminating the gap depth discrepancy. However, an order of magnitude difference still remains at best. We found that the discrepancy is due to the Rossby Wave Instability (RWI) in two dimensions, promoting angular momentum transport from the steep gap edge. As the gap edge becomes less steep because of this, the depth of the gap is reduced. The RWI is a two dimensional instability and so this explains why one dimensional gaps are much deeper, as it does not exist in one dimension to limit the gap depth. This also implies that gaps are maintained at marginal stability, with the RWI limiting their depth.

7.2 Chapter 4

In Chapter 4 we investigated a very well known problem in the area of planetary astrophysics. This is the problem that we cannot explain the observed distribution of gas giant planets with current models of planet formation and evolution. Observational evidence dictates that cold Jupiters far from their parent star are more common than hot Jupiters. However, the current Type II migration rate is so fast that these Jupiters would have to have formed at locations in which the core accretion timescale is longer than the lifetime of the disc. Hence, something must slow Type II migration (Nelson et al., 2000). In this chapter we have investigated one such process. In this idea, the central star emits radiation that is incident on the outer gap edge, visible due to the flaring of the disc. This heats the outer gap edge and increases the scale height in this region, reducing the net torque on the planet due to the planet's outer wake and consequently, the inwards migration rate of the planet.

We approximated the heating of the outer gap edge in two dimensions and investigated a variety of combinations of parameters that describe this heating. We found that using this heating method we could reduce the rate of inwards migration and in extreme cases even reverse the direction of planetary migration. Therefore, we believe we have shown that this is a potential method of slowing Type II migration that should be investigated in greater detail using three dimensional radiation-hydrodynamical simulations.

7.3 Chapter 5

In Chapter 5 we investigated another problem which has become very prevalent in recent literature. This is the problem of predicting whether or not a planet exists in an observed protoplanetary disc, based on the gap structure present. However, if the mass of the planet and the disc viscosity are both unknowns almost any gap can have a planetary explanation. Unfortunately these parameters are often not well constrained. Hence, there is a degeneracy in planet masses and disc viscosities that can adequately explain these substructures (e.g. Mulders et al., 2013).

We attempted to break this degeneracy by predicting whether or not a planet can exist in a protoplanetary disc based on the presence or absence of vortices in the disc. As it is well known that vortices can be the non-linear outcome of hydrodynamically unstable steep gap edges, it is possible that these could act as tracers for gap forming planets. It is also known that growing a planet over a sufficient timescale

can eliminate vortex formation by giving the disc time to viscously respond to the opening of the gap (Hammer et al., 2017). Hence, we found a relation between the viscosity of a disc and the mass of a planet forming in the disc as a function of the growth timescale of the planet. This allows us to place some constraints on the masses of planets in observed discs, or the viscosities of the discs themselves. One main result that falls out of our prediction is that we observe the onset of vortices in low viscosity discs for planets significantly smaller than a Jupiter mass. Hence, planets opening gaps in these discs may not need to be as 'giant' as is often thought.

7.4 Chapter 6

Finally, in our last chapter, Chapter 6, we investigated the non-isothermal regime of disc-planet interactions. We intend the results of this chapter to act as an extension of those in Chapter 5 by focussing on the behaviour of gap edge vortices in this regime. Here we investigated a wide range of thermal relaxation timescales, from near isothermal to near adiabatic. Our goal was to determine how the non-isothermal regime impacts the formation of the gap edge vortices that were the focus of Chapter 5.

We found that there are a variety of regimes that the system will pass through as the thermal relaxation timescale is increased from near-isothermal to near-adiabatic. Initially we found that if cooling is efficient vortex formation may be suppressed and hence a planet that is more massive may be able to form without creating vortices. Secondly, if the cooling is less efficient the vortices created by the planet may not die off across the disc's lifetime. Hence, we may not need to be concerned that vortices which were present in the disc at an earlier time are now absent in observations. Finally, we may require the growth timescale of the planet to be longer than the thermal relaxation timescale of the system, as for longer thermal relaxation timescales gap edge structure forms and disappears periodically. This may make the observation of gap edge vortices difficult in this regime.

7.5 Discussion

The field of protoplanetary disc research is a diverse one, with a broad range of problems providing a large number of distinct areas of research. Even within the smaller branch of planet-disc interactions that we have focussed on, there is still great diversity in the problems to address. This can clearly be seen from the chapters

presented within this thesis, each of which is centred around a topic distinct from the others. Despite each of these being separate, individual problems, they all contribute towards the overall understanding of protoplanetary disc physics.

The central topic, that links all the chapters of this thesis together, is the opening of gaps in protoplanetary discs by massive planets. Understanding this interaction contributes to our knowledge of protoplanetary discs in a variety of ways, which have been explored separately in each chapter. Gap opening is a very important process in planet-disc interactions, as it can act to limit the mass of a growing giant planet and also sets the migration regime of the planet. The opening of a gap also perturbs the gas pressure gradient across the disc and can cause dust and gas to pile up outside the gap edge. Additionally, a particularly prevalent topic in this thesis is the excitement of instabilities and subsequent formation of vortices that can be the result of the presence of a steep gap edge. These vortices can trap dust, provide a turbulent viscosity and limit the depth of the gap that can form. The non-linear outcome of such instabilities are also potentially visible in observations and can be used to constrain the masses of potential planets within these discs.

In Chapter 3, we have shown that the reason two dimensional gaps are shallower than those in one dimension is due to the RWI. This implies that gap depths are maintained at marginal stability, at the onset of the RWI. This has important implications for planet formation, as if a planet is allowed to open a gap with no limit on its depth it may be unable to accrete material. Conversely, if a planet is unable to completely clear its surroundings then it may still be allowed to grow in size. This could also impact the migration rate of giant planets, as classically Type II migration assumes that no material can cross the gap. Clearly this is not correct in higher dimensional simulations, but the more material existing within the gap region the further away we get from classical Type II migration. The impact of the planet-disc interactions studied in this chapter on planet migration and formation show the links between this chapter and Chapters 4 and 5 respectively.

It may not be an understatement to say that the rate of Type II migration is one of the largest problems facing anyone attempting to explain the variety of protoplanetary systems we observe. As such, this is one of the problems we attempted to address within this thesis, in Chapter 4. Whereas the problem discussed in Chapter 3 has links to planet formation and migration, this problem, while centred around planet migration, is very directly linked to observations of exoplanetary systems. As we discussed at the very beginning of Chapter 1, one of the primary motivations of exoplanetary science as a whole is attempting to understand and explain the diversity of systems we observe. Hence, the results presented in this chapter

have implications beyond even the broad area of protoplanetary disc physics. When considering problems with migration, we must always also be aware of the link with planet formation. While changing the rate of Type II migration has implications on the location at which a planet can begin forming, we must also remember that a planet will migrate during its formation, not just after. This can have implications on planet formation, such as the rate of formation, but also on composition. For example if a planet crosses the ice line while forming, the composition of material it accretes will change. As such our findings in Chapter 4 touch on a very broad range of topics.

We will consider Chapters 5 and 6 together, as one is largely an extension of the results of the other. These chapters focus around another observational problem, the degeneracy in planetary explanations for observed disc structure. While based more in protoplanetary disc physics, this problem has large implications for the understanding of exoplanetary systems. After all, if we cannot predict the masses of planets at the early stages of a systems life, how can we predict their evolution to explain how they will appear in later life? Chapter 5 is also closely linked to planet formation, as we grew our planets over a planet growth timescale. The rate at which a planet grows may depend on its position within the disc and this may also change as the planet migrates. Hence, despite these chapters addressing an observational problem that regards the layout of exoplanetary systems, these clearly show the link between observations and the protoplanetary disc physics.

Overall, the goal of every problem we have addressed in this thesis is to take a step towards understanding observed exoplanetary systems. Specifically why and how they look the way they do. This is not an easy goal, as exoplanetary systems are considerably diverse and their evolution spans a variety of time and length scales. Additionally, their diversity in architecture and conditions means it is very likely a single explanation will not be sufficient for all discs. Despite this a great deal of progress has been made towards this goal, across many areas of research. Chapters 4, 5 and 6 really highlight how protoplanetary disc physics is a step towards this goal. While observations of exoplanetary systems and simulations of protoplanetary discs may seem far apart, these chapters show that by looking at early evolution of a system we can make predictions about its final state. In such a way, we can see that by understanding the behaviour of protoplanetary discs we can build up to understanding exoplanetary systems.

7.6 Future work

The work described in this thesis pertains to many of the current major problems surrounding protoplanetary disc physics. Despite going some way to providing solutions to these, there is still a great deal of work that must be done to provide satisfying answers. Hence, there are many directions in which this work can be taken, in order to go even further to solving these problems. One such direction would be to extend the results in this thesis into three dimensions, as all of our results pertain only to two dimensions. This would be extremely important for Chapter 4, as in this chapter we approximated a three dimensional phenomena, the heating of the outer gap edge, in two dimensions. We would expect modelling this in three dimensions may result in some differences, so the extension to three dimensions and modelling the heating via radiative transfer codes would be the next step in understanding these results. This would be the most important direction to take this in order to determine if the outer gap edge heating is a reliable method of slowing Type II migration. Regarding Chapters 3 and 5, the two dimensional nature of our simulations may not be as much of a limitation as one may think. As we have discussed in Chapter 3, the RWI is a two dimensional instability and hence we believe the outcome would be similar in three dimensions. In Chapter 5 we have discussed that the disc scale height is small compared to the Hill sphere of the planet and hence to the planet, the disc appears two dimensional. Therefore, extending the results of these to three dimensions may not be the most pressing of directions to take them in the future.

A common theme across a number of chapters in this thesis is planet migration. This is important in Chapter 4, as we have investigated a method of slowing migration and Chapter 5, as the formation of the planet is dependent on its surroundings and planets will migrate during their formation. However, in no simulations run across this thesis do we allow the planet to migrate, instead it is always held on a fixed, circular orbit. In Chapter 4 we determined the reduction in speed of migration by investigating the net torque on the planet. In Chapter 5 we assumed that the planet growth timescale is the only important factor and that any impact on planet formation that migration may have can be modelled by merely changing this timescale. These are reasonable approximations, however, realistically it is unknown exactly how including migration would impact these results. For example, in Chapter 4 the planet migrating inwards or outwards would impact the intensity of the radiation incident on the outer gap edge, resulting in a modification to the net torque on the planet that could either be positive or negative, depending on the

radial position. With regards to Chapter 5 including migration without an accurate description of planet formation will tell us very little. However, the RWI may be affected by the inclusion of migration. The impact of migration on the RWI needs further investigation, although there has been some recent studies into this (McNally et al., 2019). This could also provide an interesting extension to the results presented in Chapter 3, as these focus on the impact of the RWI excited at the gap edge of a stationary planet. Additionally, the relation between gap opening and migration is not well understood. Any impact of this would be relevant across all of our simulations, as gap opening is the primary planet-disc interaction studied in each. It has been shown that for a migrating planet to open a gap, the gap formation timescale must be faster than the migration timescale, otherwise a gap will not be formed (Malik et al., 2015). This can be difficult considering the rapid rate of all three types of migration. Nevertheless, migrating giant planets can still open gaps in the disc material (Crida & Bitsch, 2017). Overall, including planet migration in our simulations would be an interesting direction to take these and could have important effects very worth investigating.

In our simulations we have ignored disc self-gravity. As we have discussed in Chapter 1 self-gravity can lead to the excitement of the gravitational instability, however, only in massive discs with a small enough Toomre Q parameter. This means that the results presented in this thesis only apply to discs that are not massive enough to excite the gravitiational instability. The inclusion of self-gravity is known to introduce interesting effects in massive discs. It has been shown that vortex formation at the gap edge can behave differently in self-gravitating discs and can modify the torque exerted on the planet (Lin & Papaloizou, 2011a,b). It has also been implied that a gravitationally unstable gap in an otherwise stable disc can significantly impact the migration of the planet (Lin & Papaloizou, 2012). As vortex formation and migration are important topics to the work presented in this thesis, it would be interesting to investigate how our results behave in massive, self-gravitating discs.

In all the simulations presented in this thesis, we have observed a planet (or potential) impacting the surface density structure of the disc. However, these discs are comprised only of gas, with no dust. As we are aware, this is not perfectly analogous to a real disc, which is composed of both gas and dust. This problem becomes even more apparent when we compared our results to observations, such as those in Chapter 5, as often we only observe the emission from the dust in the disc and not the gas. Clearly the dust and gas will behave similarly, for example both will be excavated by a gap forming planet, but not necessarily in exactly the same

way. We also know that vortices can be enhanced by the presence of dust, but dust can also reduce the lifetime of vortices. Hence, it is very clear that including dust in our simulations would impact the results we presented in Chapter 5. Therefore, for better comparisons with observations and a deeper insight into the behaviour of vortices, extending our simulations to include dust would be an important direction to take this work.

Part of the work presented in this thesis has been completed in the locally isothermal regime. In a more realistic picture of a protoplanetary disc, we would expect the disc to be approximated by a non-isothermal equation of state with a finite cooling timescale, such as we have used in Chapters 4 and 6. It is well known that the efficiency of heat redistribution in the system will affect the opening of gaps, as significant energy is deposited in the disc where the excited density waves shock. Indeed, this is seen in our non-isothermal simulations. Therefore, it is clear that moving into the non-isothermal regime will impact our results. We have already begun the process of extending the results presented in Chapter 5 into the non-isothermal in Chapter 6. However, the results presented in Chapter 6 are very much a first step in this direction. This area requires significantly more exploration before the impact of the non-isothermal equation of state on our results is fully understood. Furthermore, in Chapter 6 we still need to explore the behaviour of the system when including planet growth timescales, similarly to those in Chapter 5. For future work, these two directions would be the next step in investigating the results presented in Chapters 5 and 6.

Finally, it is notable that in all of the simulations performed across this thesis, none of the systems we investigate contain more than one planet, which orbits on a circular orbit. The fixing of the orbit has already been discussed with respect to migration, however, some thought should be given to multi-planet systems. In the majority of discs showing multiple ring like structures we can approximate each gap to have been formed independent of others, based on the length scales involved. As such, our simulations approximate the system well. Despite this we cannot always assume this will be the case, especially if we consider the effects of migration. Hence, investigating gap formation in multi-planet systems could be an interesting direction to extend some of this work.

Overall, while our understanding of protoplanetary disc physics has increased dramatically over the past $30 - 40$ years it is not an overstatement to say we still have a great deal to learn. This covers such a wide range of physics, from better understanding of fluid dynamics and the processes within accretion discs, to planet formation, migration, gap formation all the way to the formation of the solar system

itself. It is my hope that the results presented in this thesis contribute to a tiny part of the huge puzzle that is understanding protoplanetary disc physics.

Bibliography

ALMA Partnership et al., 2015, ApJL, 808, L3

Adams F. C., Hollenbach D., Laughlin G., Gorti U., 2004, ApJ, 611, 360

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

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

Ali-Dib M., 2016, preprint, (arXiv:1611.03128)

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

Andrews S. M., Williams J. P., 2007, ApJ, 671, 1800

Andrews S. M., et al., 2016, ApJL, 820, L40

Andrews S. M., et al., 2018, ApJL, 869, L41

Anglada-Escudé G., et al., 2016, Nature, 536, 437

Armitage P. J., 2000, A&A, 362, 968

Armitage P. J., 2011, ARAA, 49, 195

Ataiee S., Pinilla P., Zsom A., Dullemond C. P., Dominik C., Ghanbari J., 2013, A&A, 553, L3

Bae J., Hartmann L., Zhu Z., 2015, ApJ, 805, 15

Bae J., Zhu Z., Hartmann L., 2016, ApJ, 819, 134

Baglin A., Auvergne M., Barge P., Deleuil M., Michel E., CoRoT Exoplanet Science Team 2009, in Pont F., Sasselov D., Holman M. J., eds, IAU Symposium Vol. 253, Transiting Planets. pp 71–81, doi:10.1017/S1743921308026252

Bai X.-N., Goodman J., 2009, ApJ, 701, 737

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

Baraffe I., Homeier D., Allard F., Chabrier G., 2015, A&A, 577, A42

Barge P., Sommeria J., 1995, A&A, 295, L1

Barker A. J., Latter H. N., 2015, MNRAS, 450, 21

Baruteau C., Masset F., 2008a, ApJ, 672, 1054

Baruteau C., Masset F., 2008b, ApJ, 678, 483

Baruteau C., et al., 2014, Protostars and Planets VI, pp 667–689

Benítez-Llambay P., Masset F., 2016, preprint, (arXiv:1602.02359)

Bitsch B., Lambrechts M., Johansen A., 2015, A&A, 582, A112

Bond I. A., et al., 2004, ApJL, 606, L155

Bonfils X., et al., 2018, A&A, 613, A25

Booth R. A., Clarke C. J., 2016, MNRAS, 458, 2676

Borucki W. J., et al., 2010, Science, 327, 977

Boss A. P., 1997, Science, 276, 1836

Bryan G. L., et al., 2014, ApJS, 211, 19

Bryden G., Chen X., Lin D. N. C., Nelson R. P., Papaloizou J. C. B., 1999, ApJ, 514, 344

Burke C. J., et al., 2015, ApJ, 809, 8

Burrows C. J., et al., 1996, ApJ, 473, 437

Canup R. M., Ward W. R., 2002, AJ, 124, 3404

Carr J. S., Najita J. R., 2008, Science, 319, 1504

Cassan A., et al., 2012, Nature, 481, 167

Cazzoletti P., et al., 2018, A&A, 619, A161

Chandrasekhar S., 1961, Hydrodynamic and hydromagnetic stability

Chiang E. I., Goldreich P., 1997, ApJ, 490, 368

Clarke C. J., Gendrin A., Sotomayor M., 2001, MNRAS, 328, 485

Colella P., Woodward P. R., 1984, Journal of Computational Physics, 54, 174

Coleman G. A. L., Nelson R. P., 2014, MNRAS, 445, 479

Coleman G. A. L., Nelson R. P., 2016, MNRAS, 460, 2779

Crida A., Bitsch B., 2017, Icarus, 285, 145

Crida A., Morbidelli A., Masset F., 2006, Icarus, 181, 587

Crida A., Sándor Z., Kley W., 2008, A&A, 483, 325

D'Angelo G., Lubow S. H., 2010, ApJ, 724, 730

Dahn C. C., et al., 2008, ApJ, 686, 548

Dipierro G., Price D., Laibe G., Hirsh K., Cerioli A., Lodato G., 2015, MNRAS, 453, L73

Dipierro G., et al., 2018, MNRAS, 475, 5296

Dodson-Robinson S. E., Salyk C., 2011, ApJ, 738, 131

Dodson-Robinson S. E., Veras D., Ford E. B., Beichman C. A., 2009, ApJ, 707, 79

Dong R., Fung J., 2017, ApJ, 835, 38

Dong R., Zhu Z., Rafikov R. R., Stone J. M., 2015, ApJL, 809, L5

Dong R., et al., 2018, ApJ, 860, 124

Draine B. T., 2003, ARAA, 41, 241

Dressing C. D., Charbonneau D., 2015, ApJ, 807, 45

Duffell P. C., 2015, ApJL, 807, L11

Duffell P. C., MacFadyen A. I., 2013, ApJ, 769, 41

Duffell P. C., Haiman Z., MacFadyen A. I., D'Orazio D. J., Farris B. D., 2014, ApJL, 792, L10

Dullemond C. P., Dominik C., 2005, A&A, 434, 971

Dullemond C. P., Dominik C., Natta A., 2001, ApJ, 560, 957

Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, Protostars and Planets V, pp 607–622

Dürmann C., Kley W., 2015, A&A, 574, A52

Edwards B., et al., 2018, Experimental Astronomy,

Eisner J. A., et al., 2018, ApJ, 860, 77

Facchini S., Pinilla P., van Dishoeck E. F., de Juan Ovelar M., 2018, A&A, 612, A104

Fatuzzo M., Adams F. C., 2008, ApJ, 675, 1361

Faure J., Fromang S., Latter H., Meheut H., 2015, A&A, 573, A132

Feiden G. A., 2016, A&A, 593, A99

Fischer D. A., Howard A. W., Laughlin G. P., Macintosh B., Mahadevan S., Sahlmann J., Yee J. C., 2014, Protostars and Planets VI, pp 715–737

Flaherty K. M., Hughes A. M., Teague R., Simon J. B., Andrews S. M., Wilner D. J., 2018, ApJ, 856, 117

Fressin F., et al., 2013, ApJ, 766, 81

Fricke K., 1968, Zeitschrift für Astrophysik, 68, 317

Fu W., Li H., Lubow S., Li S., 2014a, ApJL, 788, L41

Fu W., Li H., Lubow S., Li S., Liang E., 2014b, ApJL, 795, L39

Fukagawa M., et al., 2013, PASJ, 65, L14

Fung J., Chiang E., 2016, preprint, (arXiv:1606.02299)

Fung J., Dong R., 2015, ApJL, 815, L21

Fung J., Shi J.-M., Chiang E., 2014, ApJ, 782, 88

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

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

Gillon M., et al., 2017, Nature, 542, 456

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Godunov S. K., 1959, Mat. Sb., 47, 271

Goldreich P., Schubert G., 1967, ApJ, 150, 571

Goldreich P., Tremaine S., 1979, ApJ, 233, 857

Goldreich P., Tremaine S., 1980, ApJ, 241, 425

Goldreich P., Lithwick Y., Sari R., 2004, ARAA, 42, 549

Gonzalez J.-F., Laibe G., Maddison S. T., Pinte C., Ménard F., 2015, MNRAS, 454, L36

Goodman J., Rafikov R. R., 2001, ApJ, 552, 793

Guzmán V. V., et al., 2018, ApJL, 869, L48

Haghighipour N., Boss A. P., 2003a, ApJ, 583, 996

Haghighipour N., Boss A. P., 2003b, ApJ, 598, 1301

Hallam P. D., Paardekooper S.-J., 2017, MNRAS, 469, 3813

Hallam P. D., Paardekooper S.-J., 2018, MNRAS, 481, 1667

Hallam P. D., Paardekooper S.-J., 2019, preprint

Hammer M., Kratter K. M., Lin M.-K., 2016, preprint, (arXiv:1610.01606)

Hammer M., Kratter K. M., Lin M.-K., 2017, MNRAS, 466, 3533

Hammer M., Pinilla P., Kratter K. M., Lin M.-K., 2019, MNRAS, 482, 3609

Hasegawa Y., Ida S., 2013, ApJ, 774, 146

Helled R., et al., 2014, Protostars and Planets VI, pp 643–665

Howard A. W., et al., 2010, Science, 330, 653

Howard A. W., et al., 2012, ApJS, 201, 15

Huang J., et al., 2018, ApJL, 869, L42

Ida S., Lin D. N. C., 2008, ApJ, 685, 584

Ilee J. D., et al., 2017, MNRAS, 472, 189

Inaba S., Barge P., 2006, ApJ, 649, 415

Isella A., Pérez L. M., Carpenter J. M., Ricci L., Andrews S., Rosenfeld K., 2013, ApJ, 775, 30

Ivanov P. B., Papaloizou J. C. B., Polnarev A. G., 1999, MNRAS, 307, 79

Jang-Condell H., Turner N. J., 2013, ApJ, 772, 34

Jin S., Li S., Isella A., Li H., Ji J., 2016, ApJ, 818, 76

Johansen A., Andersen A. C., Brandenburg A., 2004, A&A, 417, 361

Johansen A., Henning T., Klahr H., 2006, ApJ, 643, 1219

Johansen A., Oishi J. S., Mac Low M.-M., Klahr H., Henning T., Youdin A., 2007, Nature, 448, 1022

Johnson B. M., Gammie C. F., 2005, ApJ, 635, 149

Johnstone D., Hollenbach D., Bally J., 1998, ApJ, 499, 758

Kamp I., Freudling W., Chengalur J. N., 2007, ApJ, 660, 469

Kanagawa K. D., Tanaka H., Muto T., Tanigawa T., Takeuchi T., 2015, MNRAS, 448, 994

Kataoka A., Tanaka H., Okuzumi S., Wada K., 2013, A&A, 557, L4

Klahr H. H., Bodenheimer P., 2003, ApJ, 582, 869

Klahr H., Hubbard A., 2014, ApJ, 788, 21

Kley W., 1998, A&A, 338, L37

Kley W., Nelson R. P., 2012, ARAA, 50, 211

Kokubo E., Ida S., 1995, Icarus, 114, 247

Kokubo E., Ida S., 1998, Icarus, 131, 171

Kokubo E., Ida S., 2000, Icarus, 143, 15

Koller J., Li H., Lin D. N. C., 2003, ApJL, 596, L91

Korycansky D. G., Papaloizou J. C. B., 1996, ApJS, 105, 181

Kratter K. M., Murray-Clay R. A., Youdin A. N., 2010, ApJ, 710, 1375

Kraus A. L., Ireland M. J., 2012, ApJ, 745, 5

Kraus S., et al., 2017, ApJL, 848, L11

Kusaka T., Nakano T., Hayashi C., 1970, Progress of Theoretical Physics, 44, 1580

Lahuis F., et al., 2006, ApJL, 636, L145

Lambrechts M., Johansen A., 2012, A&A, 544, A32

Lambrechts M., Johansen A., 2014, A&A, 572, A107

Latter H. N., 2016, MNRAS, 455, 2608

Lesur G. R. J., Latter H., 2016, MNRAS, 462, 4549

Lesur G., Ogilvie G. I., 2010, MNRAS, 404, L64

Lesur G., Papaloizou J. C. B., 2009, A&A, 498, 1

Lesur G., Papaloizou J. C. B., 2010, A&A, 513, A60

Li H., Finn J. M., Lovelace R. V. E., Colgate S. A., 2000, ApJ, 533, 1023

Li H., Colgate S. A., Wendroff B., Liska R., 2001, ApJ, 551, 874

Li H., Li S., Koller J., Wendroff B. B., Liska R., Orban C. M., Liang E. P. T., Lin D. N. C., 2005, ApJ, 624, 1003

Lin M.-K., 2012a, MNRAS, 426, 3211

Lin M.-K., 2012b, ApJ, 754, 21

Lin M.-K., 2013, ApJ, 765, 84

Lin D. N. C., Papaloizou J., 1979, MNRAS, 186, 799

Lin D. N. C., Papaloizou J., 1980, MNRAS, 191, 37

Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846

Lin D. N. C., Papaloizou J. C. B., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. pp 749–835

Lin M.-K., Papaloizou J. C. B., 2011a, MNRAS, 415, 1426

Lin M.-K., Papaloizou J. C. B., 2011b, MNRAS, 415, 1445

Lin M.-K., Papaloizou J. C. B., 2012, MNRAS, 421, 780

Lissauer J. J., 1993, ARAA, 31, 129

Long F., et al., 2018, ApJ, 869, 17

Lovelace R. V. E., Li H., Colgate S. A., Nelson A. F., 1999, ApJ, 513, 805

Lucy L. B., 1977, AJ, 82, 1013

Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1

Lynden-Bell D., Pringle J. E., 1974, MNRAS, 168, 603

Lyra W., 2014, ApJ, 789, 77

Lyra W., Klahr H., 2011, A&A, 527, A138

Lyra W., Kuchner M., 2013, Nature, 499, 184

Lyra W., Mac Low M.-M., 2012, ApJ, 756, 62

Lyra W., Johansen A., Klahr H., Piskunov N., 2009, A&A, 493, 1125

Malik M., Meru F., Mayer L., Meyer M., 2015, ApJ, 802, 56

Mamajek E. E., 2009, in Usuda T., Tamura M., Ishii M., eds, American Institute of Physics Conference Series Vol. 1158, American Institute of Physics Conference Series. pp 3–10 (arXiv:0906.5011), doi:10.1063/1.3215910

Mann R. K., Williams J. P., 2010, ApJ, 725, 430

Marcus P. S., 1988, Nature, 331, 693

Marcus P. S., 1990, Journal of Fluid Mechanics, 215, 393

Marcus P. S., Pei S., Jiang C.-H., Hassanzadeh P., 2013, Physical Review Letters, 111, 084501

Marcus P. S., Pei S., Jiang C.-H., Barranco J. A., Hassanzadeh P., Lecoanet D., 2015, ApJ, 808, 87

Marino S., Casassus S., Perez S., Lyra W., Roman P. E., Avenhaus H., Wright C. M., Maddison S. T., 2015, ApJ, 813, 76

Marois C., Zuckerman B., Konopacky Q. M., Macintosh B., Barman T., 2010, Nature, 468, 1080

Masset F., 2000, A&AS, 141, 165

Masset F. S., 2001, ApJ, 558, 453

Masset F. S., 2002, A&A, 387, 605

Masset F. S., Papaloizou J. C. B., 2003, ApJ, 588, 494

Masset F. S., D'Angelo G., Kley W., 2006, ApJ, 652, 730

Matzner C. D., Levin Y., 2005, ApJ, 628, 817

Mayor M., Queloz D., 1995, Nature, 378, 355

Mayor M., et al., 2003, The Messenger, 114, 20

Mayor M., et al., 2011, preprint, (arXiv:1109.2497)

McNally C. P., Nelson R. P., Paardekooper S.-J., Benítez-Llambay P., 2019, MNRAS, 484, 728

Meheut H., Casse F., Varniere P., Tagger M., 2010, A&A, 516, A31

Meheut H., Yu C., Lai D., 2012a, MNRAS, 422, 2399

Meheut H., Keppens R., Casse F., Benz W., 2012b, A&A, 542, A9

Meru F., 2015, MNRAS, 454, 2529

Meru F., Bate M. R., 2010, MNRAS, 406, 2279

Morbidelli A., Raymond S. N., 2016, Journal of Geophysical Research (Planets), 121, 1962

Mordasini C., Alibert Y., Benz W., 2009, A&A, 501, 1139

Mulders G. D., Paardekooper S.-J., Panić O., Dominik C., van Boekel R., Ratzka T., 2013, A&A, 557, A68

Nelson R. P., Papaloizou J. C. B., Masset F., Kley W., 2000, MNRAS, 318, 18

Nelson R. P., Gressel O., Umurhan O. M., 2013, MNRAS, 435, 2610

Okuzumi S., Tanaka H., Kobayashi H., Wada K., 2012, ApJ, 752, 106

Ono T., Muto T., Takeuchi T., Nomura H., 2016, ApJ, 823, 84

Paardekooper S.-J., Mellema G., 2004, A&A, 425, L9

Paardekooper S.-J., Mellema G., 2006a, A&A, 450, 1203

Paardekooper S.-J., Mellema G., 2006b, A&A, 459, L17

Paardekooper S.-J., Papaloizou J. C. B., 2008, A&A, 485, 877

Paardekooper S.-J., Baruteau C., Crida A., Kley W., 2010, MNRAS, 401, 1950

Paardekooper S.-J., Baruteau C., Kley W., 2011, MNRAS, 410, 293

Padgett D. L., Brandner W., Stapelfeldt K. R., Strom S. E., Terebey S., Koerner D., 1999, AJ, 117, 1490

Papaloizou J. C. B., Nelson R. P., Snellgrove M. D., 2004, MNRAS, 350, 829

Papaloizou J. C. B., Nelson R. P., Kley W., Masset F. S., Artymowicz P., 2007, Protostars and Planets V, pp 655–668

Pérez L. M., et al., 2012, ApJL, 760, L17

Pérez L. M., Isella A., Carpenter J. M., Chandler C. J., 2014, ApJL, 783, L13

Petersen M. R., Julien K., Stewart G. R., 2007a, ApJ, 658, 1236

Petersen M. R., Stewart G. R., Julien K., 2007b, ApJ, 658, 1252

Petrovich C., Rafikov R. R., 2012, ApJ, 758, 33

Pierens A., Huré J.-M., 2005, A&A, 433, L37

Pierens A., Lin M.-K., 2018, MNRAS, 479, 4878

Pinilla P., Birnstiel T., Ricci L., Dullemond C. P., Uribe A. L., Testi L., Natta A., 2012, A&A, 538, A114

Pinte C., Dent W. R. F., Ménard F., Hales A., Hill T., Cortes P., de Gregorio-Monsalvo I., 2016, ApJ, 816, 25

Pollacco D. L., et al., 2006, PASP, 118, 1407

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62

Pravdo S. H., Shaklan S. B., Lloyd J., 2005, ApJ, 630, 528

Pringle J. E., 1981, ARAA, 19, 137

Raettig N., Lyra W., Klahr H., 2013, ApJ, 765, 115

Raettig N., Klahr H., Lyra W., 2015, ApJ, 804, 35

Rafikov R. R., 2002, ApJ, 572, 566

Rafikov R. R., 2005, ApJL, 621, L69

Rafikov R. R., 2011, ApJ, 727, 86

Railton A. D., Papaloizou J. C. B., 2014, MNRAS, 445, 4409

Rapson V. A., Kastner J. H., Millar-Blanchaer M. A., Dong R., 2015, ApJL, 815, L26

Raymond S. N., Kokubo E., Morbidelli A., Morishima R., Walsh K. J., 2014, Protostars and Planets VI, pp 595–618

Regály Z., Vorobyov E., 2017a, MNRAS, 471, 2204

Regály Z., Vorobyov E., 2017b, A&A, 601, A24

Regály Z., Juhász A., Nehéz D., 2017, ApJ, 851, 89

Ribas I., et al., 2018, Nature, 563, 365

Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2004, MNRAS, 355, 543

Rice W. K. M., Lodato G., Armitage P. J., 2005, MNRAS, 364, L56

Richard S., Barge P., Le Dizès S., 2013, A&A, 559, A30

Richert A. J. W., Lyra W., Boley A., Mac Low M.-M., Turner N., 2015, ApJ, 804, 95

Richtmyer R. D., Morton K. W., 1967, Difference methods for initial-value problems

Ricker G. R., et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Robert C. M. T., Crida A., Lega E., Méheut H., Morbidelli A., 2018, A&A, 617, A98

Rodmann J., Henning T., Chandler C. J., Mundy L. G., Wilner D. J., 2006, A&A, 446, 211

Rosotti G. P., Juhasz A., Booth R. A., Clarke C. J., 2016, MNRAS, 459, 2790

Ryu D., Ostriker J. P., Kang H., Cen R., 1993, ApJ, 414, 1

Sallum S., et al., 2015, Nature, 527, 342

Salmeron R., Wardle M., 2008, MNRAS, 388, 1223

Santerne A., et al., 2016, A&A, 587, A64

Schive H.-Y., Tsai Y.-C., Chiueh T., 2010, ApJS, 186, 457

Schneider E. E., Robertson B. E., 2015, ApJS, 217, 24

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Siess L., Dufour E., Forestini M., 2000, A&A, 358, 593

Smith N., Bally J., Licht D., Walawender J., 2005, AJ, 129, 382

Sokal K. R., Deen C. P., Mace G. N., Lee J.-J., Oh H., Kim H., Kidder B. T., Jaffe D. T., 2018, ApJ, 853, 120

Sommeria J., Meyers S. D., Swinney H. L., 1988, Nature, 331, 689

Stapelfeldt K. R., Krist J. E., Ménard F., Bouvier J., Padgett D. L., Burrows C. J., 1998, ApJL, 502, L65

Stoll M. H. R., Kley W., 2014, A&A, 572, A77

Stone J. M., Norman M. L., 1992, ApJS, 80, 753

Strang G., 1968, SIAM Journal on Numerical Analysis, 5, 506

Szulágyi J., van der Plas G., Meyer M. R., Pohl A., Quanz S. P., Mayer L., Daemgen S., Tamburello V., 2017, preprint, (arXiv:1709.04438)

Takeuchi T., Lin D. N. C., 2002, ApJ, 581, 1344

Takeuchi T., Miyama S. M., Lin D. N. C., 1996, ApJ, 460, 832

Tanaka H., Takeuchi T., Ward W. R., 2002, ApJ, 565, 1257

Terquem C. E. J. M. L. J., 2008, ApJ, 689, 532

Testi L., et al., 2015, ApJL, 812, L38

Thommes E. W., Duncan M. J., Levison H. F., 2003, Icarus, 161, 431

Toomre A., 1964, ApJ, 139, 1217

Trac H., Pen U.-L., 2004, New Astronomy, 9, 443

Tsang D., 2014, ApJ, 782, 112

Tsang D., Turner N. J., Cumming A., 2014, ApJ, 782, 113

Tsukagoshi T., et al., 2016, ApJL, 829, L35

Turner N. J., Carballido A., Sano T., 2010, ApJ, 708, 188

Turner N. J., Choukroun M., Castillo-Rogez J., Bryden G., 2012, ApJ, 748, 92

Varnière P., Tagger M., 2006, A&A, 446, L13

Varnière P., Quillen A. C., Frank A., 2004, ApJ, 612, 1152

Vicente S. M., Alves J., 2005, A&A, 441, 195

Ward W. R., 1991, in Lunar and Planetary Science Conference.

Ward W. R., 1997, Icarus, 126, 261

Ward W. R., Canup R. M., 2010, AJ, 140, 1168

Weidenschilling S. J., 1977, MNRAS, 180, 57

Weidenschilling S. J., 1995, Icarus, 116, 433

Williams J. P., Cieza L. A., 2011, ARAA, 49, 67

Wilner D. J., D'Alessio P., Calvet N., Claussen M. J., Hartmann L., 2005, ApJL, 626, L109

Winters W. F., Balbus S. A., Hawley J. F., 2003, ApJ, 589, 543

Wolf S., Gueth F., Henning T., Kley W., 2002, ApJL, 566, L97

Wright J. T., Marcy G. W., Howard A. W., Johnson J. A., Morton T. D., Fischer D. A., 2012, ApJ, 753, 160

Youdin A. N., Goodman J., 2005, ApJ, 620, 459

Zhang K., Blake G. A., Bergin E. A., 2015, ApJL, 806, L7

Zhang S., et al., 2018, ApJL, 869, L47

Zhu Z., Andrews S. M., Isella A., 2018, MNRAS, 479, 1850

Ziegler U., Yorke H. W., 1997, Computer Physics Communications, 101, 54

Zsom A., Dullemond C. P., 2008, A&A, 489, 931

de Val-Borro M., et al., 2006, MNRAS, 370, 529

de Val-Borro M., Artymowicz P., D'Angelo G., Peplinski A., 2007, A&A, 471, 1043

van Boekel R., et al., 2017, ApJ, 837, 132

van der Marel N., et al., 2013, Science, 340, 1199

van der Marel N., Dong R., di Francesco J., Williams J. P., Tobin J., 2019, ApJ, 872, 112

List of Figures

List of Tables