The Evolution of Gravitationally Unstable Protoplanetary Discs: From Instability to Stability

by

Sahl Rowther

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Physics

Department of Physics

September 2022

Contents

List of Figures

Acknowledgments

Completing this PhD would have been a lot harder without the support and guidance of a couple important people. Firstly, I would like to thank my supervisor, Farzana Meru for giving me the chance, and resources to develop as a scientist (and for funding my hobby of wildlife photography). Secondly, I am also grateful to Bec Nealon, who alongside with Farzana, never tired of whatever questions and thoughts my brain came up with it.

The work in this thesis has benefitted from valuable conversations with Christophe Pinte, Ken Rice, Dmitri Veras, Grant Kennedy, Cathie Clarke, and Colin McNally. Additionally, I would like to thank Hossam Aly, Sean Andrews, Richard Teague, and Cassandra Hall for some very interesting conversations in the final weeks of my PhD which will no doubt be of great help during the next stage of my academic career.

Finally, I would like to thank my parents for their constant support throughout my life.

The research presented in thesis was funded by the Royal Society Enhancement Award, and utilised TINIS, ORAC, and AVON, the HPC clusters at the University of Warwick.

Declarations

I submit this thesis to the University of Warwick Doctoral College for the degree of Doctor of Philosophy, and declare that all work presented is my own research. The contents of this thesis have not been submitted for any other degree at any university.

All work in this thesis was carried out by me and includes material from either published papers or papers in preparation:

- Chapter 3 has been published as *"Planet migration in self-gravitating discs: survival of planets", Rowther, S., and Meru, F.*, 2020, MNRAS, 496, 1598.
- Chapter 4 has been published as *"Continuing to Hide Signatures of Gravitational Instability in Protoplanetary Discs", Rowther, S., Meru, F., Kennedy, G., Nealon, R., and Pinte, C.*, 2020, ApJ, 904, L18.
- Chapter 5 has been published as *"Continuing to Hide Signatures of Gravitational Instability in Protoplanetary Discs", Rowther, S., Nealon, R., and Meru, F.*, 2023, MNRAS, 518, 763
- Chapter 6 has been published as *"Warping Away Gravitational Instabilities in Protoplanetary Discs", Rowther, S., Nealon, R., and Meru, F.*, 2022, ApJ, 925, 163.

Abstract

In their youth, protoplanetary discs are expected to be massive and self-gravitating. A characteristic feature of such discs are non-axisymmetric spiral structures. However recent observations of young protoplanetary discs with the Atacama Large Millimeter/submillimeter Array (ALMA) have revealed that discs with large-scale spiral structure in the mid-plane are rarely observed. Instead, axisymmetic discs with ring & gap structures are more commonly observed. Using 3D smoothed particle hydrodynamic simulations, the aim of this thesis is to explore how additional physical processes occurring in a gravitationally unstable protoplanetary disc can explain this phenomenon.

The first process considered is planet-disc interactions. I first show the importance of modelling the disc thermodynamics for studying the fate of planet migration. In contrast to previous work using a simpler cooling model, I show that planets are able to slow their migration and survive in the gravitationally stable inner regions of the disc. I then show that if the planet is sufficiently massive, the spiral wakes of the planet can heat up the disc causing it to become gravitationally stable, suppressing spiral structures due to gravitational instability.

The second process considered is warps. An idealised warp is used to isolate the effect of the warp on the gravitational instabilities present in the disc. As with planet-disc interactions, I show that the warp is able to heat up the disc rendering it gravitational stable.

The results of this thesis provide a plausible explanation for why gravitationally unstable discs are rarely seen – physical processes that are likely to occur in these discs will alter its evolution from an unstable to a stable disc.

Abbreviations

GI	Gravitational Instability
SED	Spectral Energy Distribution
SPH	Smoothed Particle Hydrodynamics
ALMA	Atacama Large Millimeter/submillimeter Array
IR	Infrared
AU	Astronomical Unit
MHD	Magnetohydrodynamics
DSHARP	Disk Substructures at High Angular Resolution Project

1

Introduction

Accretion discs are ubiquitous in astronomy. At the smallest scale, they can be found as rings around planets like Saturn. At the largest scale, they are found around supermassive black holes. Somewhere between these massive scales lies the subject of this thesis, protoplanetary discs around stars.

1.1 Formation of protoplanetary discs

The story of a protoplanetary disc begins in the early stages of star formation in a molecular cloud made up of gas and small dust grains. As the cloud cools, a protostar begins to form as gravitational forces seeking to collapse the cloud eventually overcomes thermal pressure preventing collapse. The initial cloud is neither homogeneous nor static. Hence, there is always some angular momentum present which prevents the entire infalling cloud collapsing onto the protostar. As the protostar accretes material, the rotation of the cloud increases due to conservation of angular momentum, preventing further infall onto the protostar. The gas loses energy through collisions as it orbits, while maintaining its angular momentum. Since a circular orbit is favoured for the lowest energy state, the infalling material adopts a circular orbit (Pringle, 1981). The gas also has a range of angular momenta and energy and so these orbits also occur at a range of radii from the central star. This structure that forms around the protostar is a protoplanetary disc (see simulations by Bate, 1998, 2011). The evolutionary phases of forming a planetary system from a giant molecular cloud is illustrated in Figure 1.1. While a useful cartoon, it is limited as stars do not form in isolation. They form in clusters where dynamic

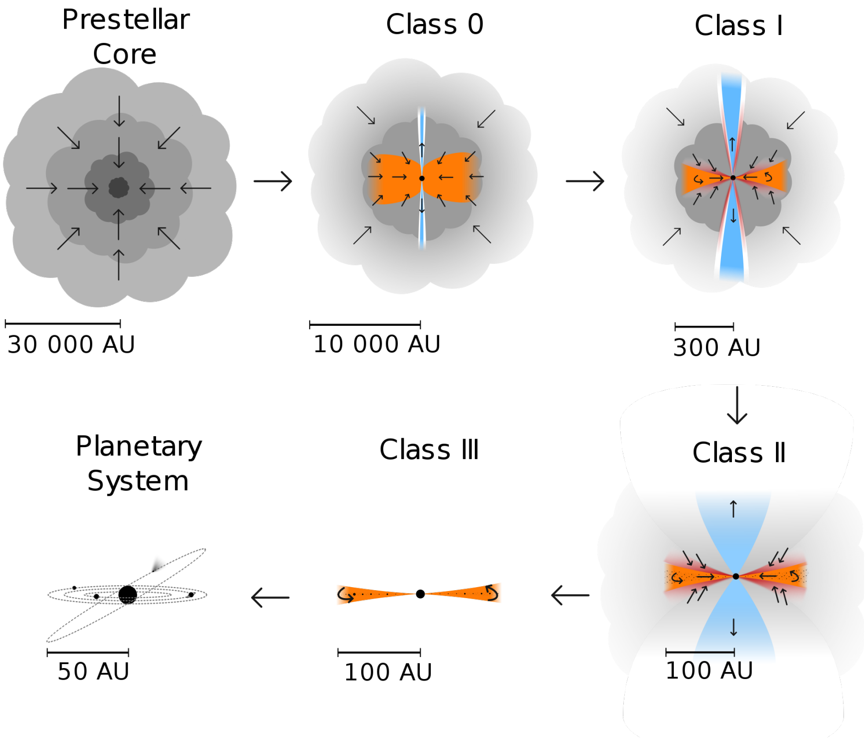


Figure 1.1: Cartoon of the evolution from a cloud to a disc to a planetary system. Figure by Magnus Vilhelm Persson.

interactions influence disc evolution, e.g. Bate (2018), such that most stars form in pairs called binaries.

Protoplanetary discs are typically classified by the shape of their infra-red (IR) spectral energy distribution (SED) as originally proposed by Lada (1987) and Adams et al. (1987) which categorised protoplanetary discs into three classes, I-II-III respectively based on the slope of their SED. This classification scheme was extended by Andre et al. (1993) to include Class 0 discs, the earliest evolutionary phase of a protoplanetary disc. The four classes and their corresponding SEDs are illustrated in Figure 1.2.

In the Class 0 phase, the disc is beginning to form and the protostar is still collapsing with the emission being dominated by its envelope. The SED peaks in the far IR or sub-mm without any flux measured in the near and mid IR due to light being absorbed by circumstellar dust and re-emitted at longer wavelength.

Next, we enter Class I, the accretion phase of protostellar evolution. During this phase, the envelope accretes onto the forming disc which is then in turn accreted onto the star. The disc mass increases substantially and emission from both the star

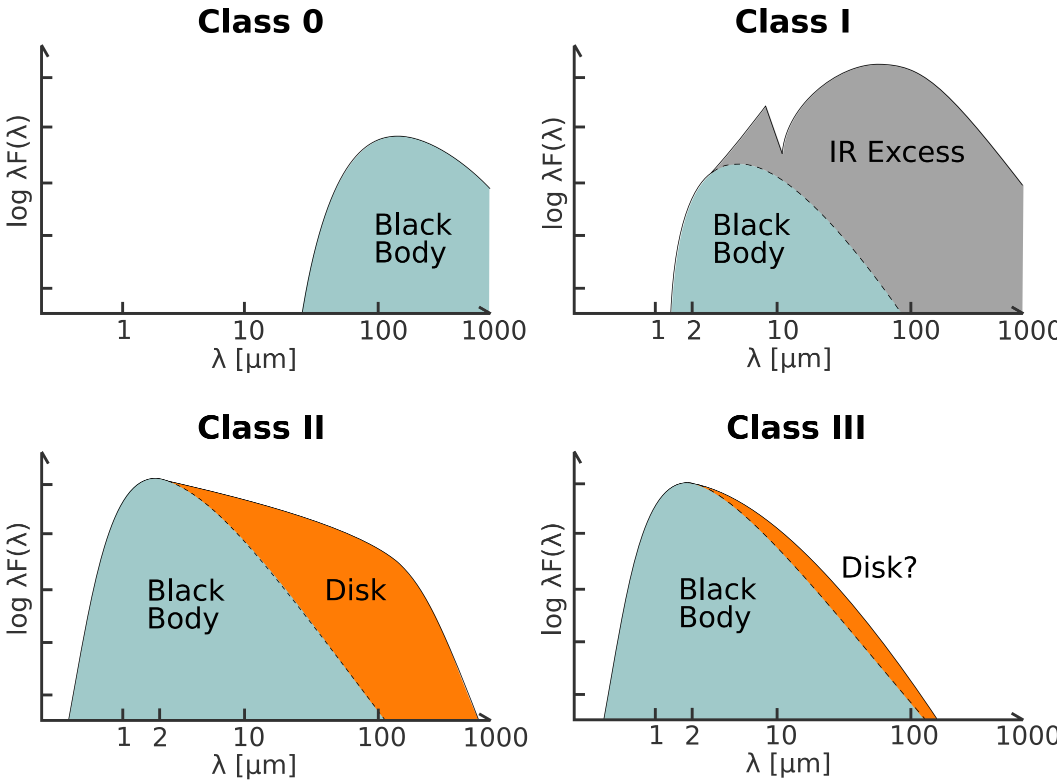


Figure 1.2: Cartoon of the SED associated at each stage of protoplanetary disc formation. Figure by Magnus Vilhelm Persson.

and disc is detectable. The star heats up the inner regions of the disc which will emit at a wavelength of $\sim 1\mu$m. Most of this light will be absorbed by the now optically thinner cloud and re-emitted at longer wavelengths. Hence, the SED peaks in the mid or far-IR with a rising slope in the near-IR.

Most of the observed protoplanetary discs are in the Class II phase. The envelope has now completely disappeared and the protostar, now classified as a T Tauri or a Herbig Ae/Be star, continues to accrete material from the surrounding protoplanetary disc. The SED now peaks in the near-IR, though excess IR emission from the disc is still present, but it has declined from the Class I phase (Andrews & Williams, 2005). Recent observations of substructures in Class I (eg. Segura-Cox et al., 2020) and Class II discs (eg. Andrews et al., 2018; Long et al., 2018) strongly implies that planet formation is connected to disc formation.

Although total disc masses are difficult to obtain, an insight to the mass loss at each stage can be gained by comparing dust mass estimates. Using ALMA Tychoniec et al. (2020) measured a median dust mass of $47M_{\oplus}$ for a population of Class 0 discs in the Perseus star forming region. While for Class I, the mass dropped to $12M_{\oplus}$. However, these discs are still a lot more massive compared to the more evolved Class II discs in the Lupus star forming region which has a median dust mass of only $3M_{\oplus}$ (Ansdell et al., 2016). The work in this thesis focuses on the earlier phases of the disc's evolution when they are still massive enough to be gravitationally unstable.

The final phase (Class III) is when only remnants of the protoplanetary disc is left in the form of solids, usually referred to as debris disc. There is little to none of the gaseous disc left as most of it has been accreted onto the central star (see review by Matthews et al., 2014). The SED peaks in the near-IR with little to no IR excess due to the lack of disc material. However, discs do not necessarily evolve in a steady evolutionary sequence as described above. Dynamical interactions can result in Class III discs in the earliest stages of star formation (Kurosawa et al., 2004).

1.2 Observations of protoplanetary discs

The dust and gas dynamics in a protoplanetary disc provides valuable insight on the observed structures and their origins.

1.2.1 Dust substructures

Assuming the composition of protoplanetary discs is similar to that of the Milky Way, which has a dust-to-gas mass ratio of 1:100 (Bohlin et al., 1978), then it follows that dust makes up only a tiny portion of the disc. Despite the disc mass — and hence its evolution — being dominated by gas, the more readily observed dust is used to reveal the processes occurring in protoplanetary discs. While all the work in this thesis involves gas only simulations, connections to observations are made by using the above dust-to-gas mass ratio, and by assuming the dust is perfectly coupled to the gas. The latter assumption is valid as long as the Stokes number, the ratio of the stopping time to the orbital timescale (Weidenschilling, 1977), is low. This occurs when the gas mass is sufficiently high compared to the the dust size (Kwok, 1975; Laibe & Price, 2012), which is the case for the massive discs in this thesis. However, the grain size is still important for decoupling. Observations of young discs such as HL Tau (ALMA Partnership et al., 2015; Dipierro et al., 2015) already show evidence of decoupling from a mm-emitting dust disc with $H/R \lesssim 0.01$.

Figure 1.3 shows how different dust grain sizes distribute themselves in the disc (see review by Testi et al., 2014). Observations of the dust in protoplanetary discs either trace scattered light emitted by the host star, or the thermal continuum

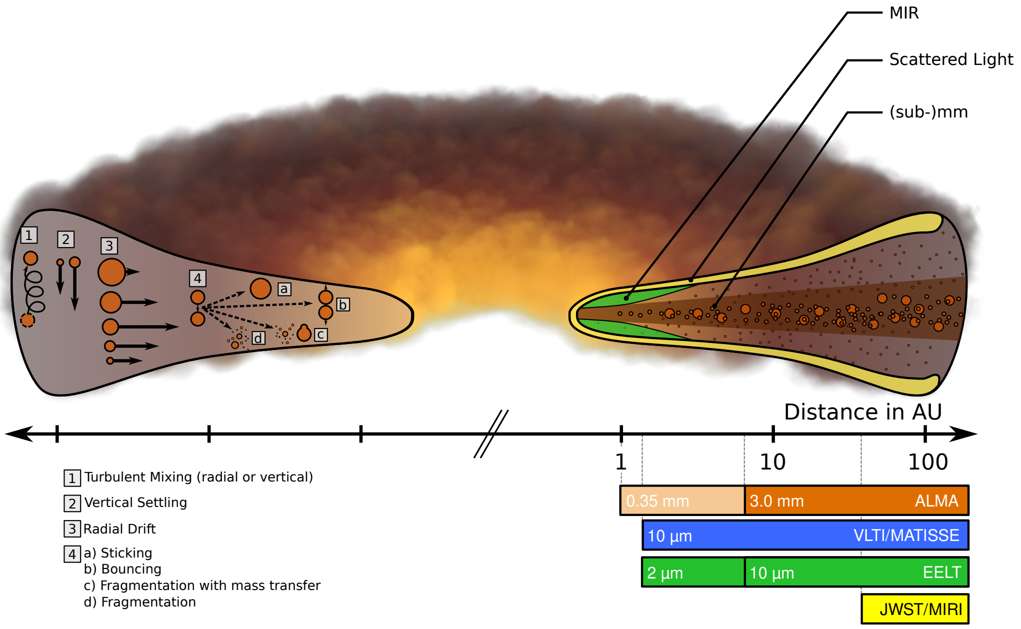


Figure 1.3: Illustration of dust dynamics and structure in a protoplanetary disc. Various processes governing the dust dynamics is shown on the left. The distribution of different grain sizes is shown on the right. The left edge of the horizontal bars represent the highest angular resolution achieved by different instruments at the typical distance (∼140pc) to a protoplanetary disc. Image from Testi et al. (2014).

emitted by the dust. Scattered light observations probe the upper layers of the disc atmosphere, which is dominated by small μm-sized grains. Although the dust thermal continuum covers wavelengths from 1μm to cm, at short wavelengths the stellar radiation is absorbed by the dust. Therefore, it traces the upper layers and is a temperature diagnostic. The density structure is revealed by observations of the dust continuum at sub-mm or mm wavelengths where the disc is expected to be optically thin, since optical depth decreases with wavelength. At these wavelengths, the observations trace the larger grains which are well-settled (Weidenschilling, 1980; Nakagawa et al., 1981) and probe the disc mid-plane, where planet formation is thought to occur.

Rings & gaps

One of the more common type of substructures seen in observations of protoplanetary discs are axisymmetric rings & gaps. They are routinely observed in both the mm continuum and scattered light as shown by a collection of discs in Figure 1.4. Even amongst just these subset of discs, their morphology is varied. The rings

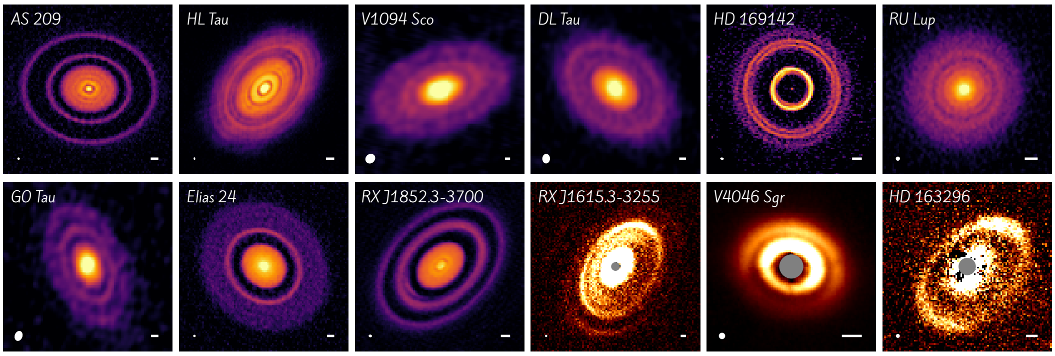


Figure 1.4: A collection of protoplanetary discs exhibiting ring & gap structure (adapted from Andrews, 2020). AS 209 (Guzmán et al., 2018), HL Tau (ALMA Partnership et al., 2015), V1094 Sco (van Terwisga et al., 2018), DL Tau (Long et al., 2018), HD 169142 (Pérez et al., 2019), RU Lup (Andrews et al., 2018), GO Tau (Long et al., 2018), Elias 24 (Andrews et al., 2018), RX J1852.3-3700 (Villenave et al., 2019), RX J1615.3-3255 (Avenhaus et al., 2018), V4046 Sgr (Avenhaus et al., 2018), and HD 163296 (Monnier et al., 2017).

can be narrow in some discs (eg. AS 209, Guzmán et al., 2018) or wide in others (eg. GO Tau, Long et al., 2018). Likewise, the gaps appear deep and fully depleted in some discs (eg. Elias 24, Andrews et al., 2018) or faint in others (eg. RU Lup, Andrews et al., 2018).

There are a few mechanisms to explain the origin of the rings & gaps. As planets must form in protoplanetary discs, they are a natural choice as a mechanism which can alter the disc structure to form rings & gaps. Once a planet becomes massive enough, it begins interacting with disc through the spiral waves generated by the planet through constructive interference of waves launched at Lindblad resonances (Ogilvie & Lubow, 2002). The exchange of angular momentum between the planet and the disc results in disc material being repelled away from the planet's orbit, thus carving open a gap in the disc (Lin & Papaloizou, 1979, 1986). The width and depth of the gap carved open by the planet depends on its mass, and the disc properties (Kley & Nelson, 2012). The resulting pressure maxima produced outside the gap can trap inwardly drifting solids resulting in bright rings (Rice et al., 2006; Zhu et al., 2012b; Kanagawa et al., 2021). Planet-disc interactions are discussed in more detail in §1.4. If the disc viscosity is low enough, multiple rings & gaps can be formed by a single planet (Zhang et al., 2018; Miranda & Rafikov, 2019). The migration of planets can also complicate the signatures of ring & gap structure (Nazari et al., 2019; Meru et al., 2019). Alternatively, MHD turbulence has been found to produce narrow rings & gaps (Dittrich et al., 2013; Suriano et al., 2019), as well as wider gaps (Flock et al., 2015). Concentric concentrations of magnetic flux modify the gas dynamics by repelling gas from regions of peak magnetic stress resulting in rings & gaps. Dust growth near condensation fronts or snowlines is yet another mechanism which can explain the ring & gap structures (van der Marel et al., 2018b; Zhang et al., 2015). However, by measuring the radii of the predicted main freeze-out temperatures, no correlation was found between the location of dust gaps and the snowlines (van der Marel et al., 2019).

Ring/Cavity

Figure 1.5 shows a collection of discs characterised by a primary ring around a central cavity. These cavities tend to be completely or partially depleted, resulting in a deficit in the infra-red SED emission (Espaillat et al., 2007; van der Marel et al., 2016), and thus these discs are often considered to be transition discs (Espaillat et al., 2014). Numerical simulations have shown outflows from MHD-driven winds (Suzuki et al., 2016; Takahashi & Muto, 2018) or photoevaporative flows (Alexander et al., 2014; Ercolano & Pascucci, 2017) are possible mechanisms for creating a cavity with a ring exterior to it. Alternatively, as with rings & gaps, dynamical interactions with planetary (Zhu et al., 2011) or stellar companions (HD 142527, Price et al., 2018b) can deplete the inner regions of the disc. However, dynamical clearing (by planets Zhu et al., 2011) is unable to explain the high mass accretion rates observed (Manara et al., 2014).

Spirals

Figure 1.6 shows that large scale spiral structures are seen in both scattered light and in the mm continuum, although far less frequently in the latter. Planets are, as usual, a popular choice as an explanation for the spiral structures. Dong et al. (2015a) and Zhu et al. (2015) have shown that a giant planet at large radii can generate spiral arms interior and exterior to the planet's orbit. The inner arms are able to match the morphology of the spirals seen in observations.

If the disc is sufficiently massive and cold enough, the self-gravity of the gas drives the evolution of the disc resulting in gravitational instabilities in the form of large-scale spiral structures (Boss, 1997; Dong et al., 2015b; Meru et al., 2017; Forgan et al., 2018b). These two mechanisms can be distinguished by the shape of the spiral structures. Planet-induced spirals are expected to be asymmetric with variable pitch angles (Bae & Zhu, 2018), whereas gravitational instability forms symmetric spirals with a constant pitch angle (Forgan et al., 2018b). Contrast

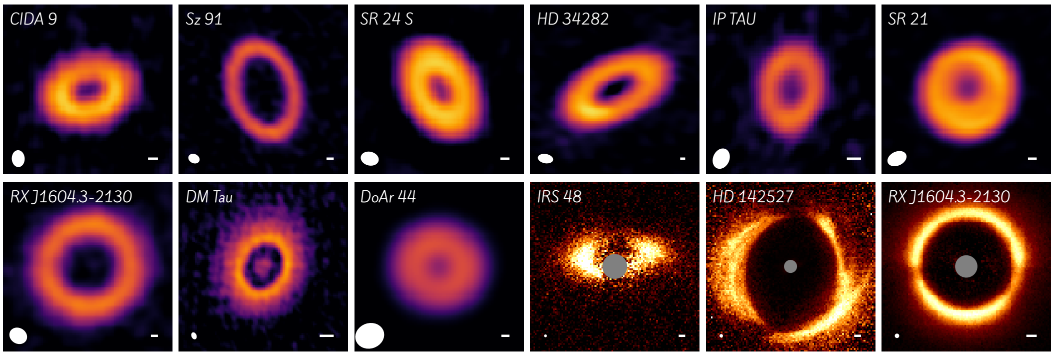


Figure 1.5: A collection of protoplanetary discs exhibiting a depleted cavity surrounded by a ring (adapted from Andrews, 2020). CIDA 9 (Long et al., 2018), Sz 91 (Tsukagoshi et al., 2019), SR 24 S (Pinilla et al., 2017), HD 34282 (van der Plas et al., 2017), IP Tau (Long et al., 2018), SR 21 (van der Marel et al., 2018a), RX J1604.3-2130 (Pinilla et al., 2018a), DM Tau (Kudo et al. 2018), DoAr 44 (Pinilla et al., 2018a), IRS 48 (Follette et al. 2015), HD 142527 (Avenhaus et al., 2014), and RX J1604.3-2130 (Pinilla et al., 2018b).

curves can also be used to rule out giant planets at large radii as proposed for MWC758 (Ren et al., 2018). The morphology of the spiral structures has been used to propose gravitational instabilities as the origin of the spiral structures in Elias 2-27 (Paneque-Carreño et al., 2021).

Arcs

A few protoplanetary discs exhibit asymmetries either as a lone incomplete ring around a central cavity or as an additional brightness asymmetry in ring & gap structures as seen in Figure 1.7. A potential explanation are vortices resulting from Rossby wave instabilities driven – once again – by a planet. At the edge of a gap opened by a planet, material can pile up resulting in a localised over-density. Eventually, the accumulation of gas triggers the Rossby wave instability (Lovelace et al., 1999; Lyra et al., 2009). However, this model is not favoured by observations (van der Marel et al., 2021; Garg et al., 2021). Instead, the preferred model is a planetary or stellar companion creating a pressure maxima where grains concentrate (Ragusa et al., 2017; Price et al., 2018b; Calcino et al., 2019; Ragusa et al., 2020, 2021).

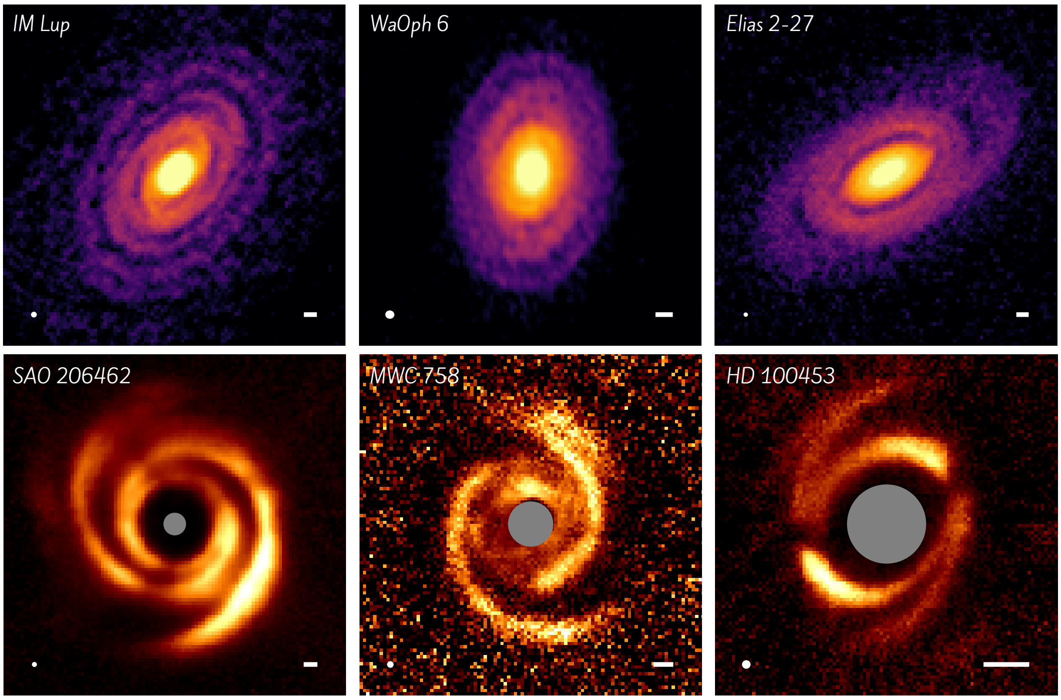


Figure 1.6: A collection of discs featuring large-scale spiral structures (adapted from Andrews, 2020). IM Lup, WaOph 6, Elias 2-27 (all from Huang et al., 2018b), SAO 206462 (Stolker et al., 2017), MWC 758 (Benisty et al., 2015), and HD 100453 (Benisty et al., 2017).

Shadows

Recent observations suggest that not all discs have coplanar geometry with respect to their central star. The complex geometry is revealed by scattered light observations where shadowing effects are visible as shown in Figure 1.8. The shadows cast either appear as narrow lanes spanning a few degrees in azimuth (Avenhaus et al., 2014; Benisty et al., 2017; Ginski et al., 2021), or as broad shadows spanning around half the disc (Benisty et al., 2018). The shadowing effects are often associated with misaligned or warped discs (Facchini et al., 2018). The narrow shadows are cast by an inner disc that is strongly misaligned (and perhaps *broken*) relative to the outer disc (Marino et al., 2015). Broad extended shadows are in general caused by small misalignments or warps (Benisty et al., 2018). Planets or stellar companions on misaligned orbits are a popular explanation for the misalignment of the inner and outer disc (Larwood et al., 1996; Facchini et al., 2013; Nixon et al., 2013; Nealon et al., 2018, 2020b; Ballabio et al., 2021). A recent observation highlighting this is the triple star system GW Ori (Kraus et al., 2020). Alternatively, infalling material

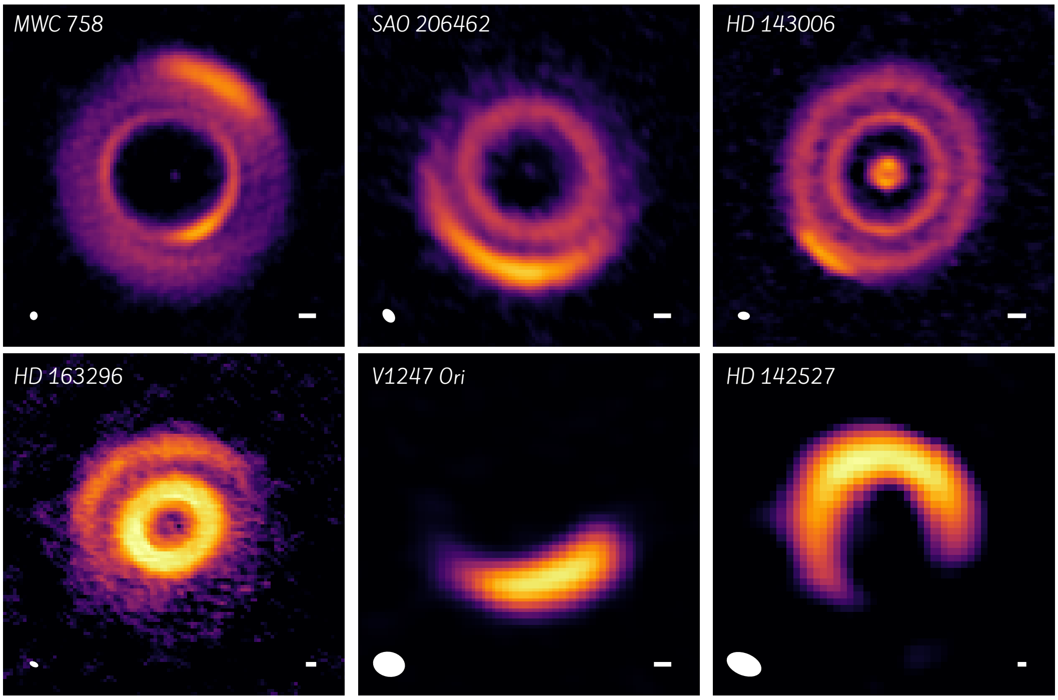


Figure 1.7: A collection of protoplanetary discs exhibiting extended aximuthal asymmetries (adapted from Andrews, 2020). MWC 758 (Dong et al., 2018), SAO 206462 (Cazzoletti et al., 2018), HD 143006 (Pérez et al., 2018), HD 163296 (Isella et al., 2018), V1247 Ori (Kraus et al., 2017), HD 142527 (Casassus et al., 2013).

with independent angular momentum is also a plausible mechanism which has been seen in star forming simulations (Bate, 2018), and at late stages as in the Class II disc SU Aur (Ginski et al., 2021).

1.2.2 Gas substructures

While the dust reveals a diverse variety of substructures with many possible origins, we now turn to the gas. Although $\sim 99\%$ of the disc mass is in gas, most of it is molecular hydrogen which is unobservable due to it not having a permanent dipole moment and not emitting efficiently over most of the disc volume. Instead, observations rely on rarer molecules which have rotational transitions at sub-mm wavelengths. As with different sized dust grains, different molecules trace different parts of the disc. Optically thick isotopologues of molecules such as CO_2, the second most abundant molecule, trace the upper layers of the disc which underestimates the disc mass (Bergin & Williams, 2017). Optically thin molecules such as $^{13}C^{17}O$ probe the disc mid-plane and are a robust tracer of the disc mass if the abundance

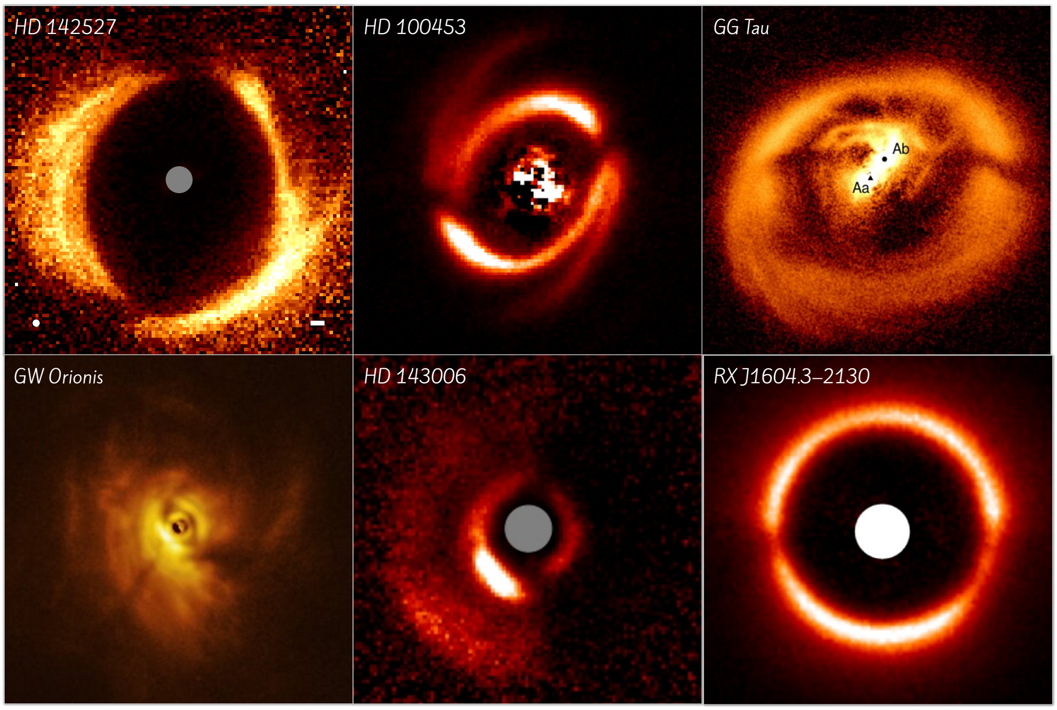


Figure 1.8: A collection of protoplanetary discs exhibiting shadowing effects. HD 142527 (Avenhaus et al., 2014), HD 100453 (Benisty et al., 2017), GG Tau, GW Orionis, HD 143006 (Benisty et al., 2018), RX J1604.3-2130 (Pinilla et al., 2018b).

ratio relative to H_2 is known (Booth et al., 2019; Booth & Ilee, 2020).

The gas emission lines also allow the disc kinematics to be measured (see review by Pinte et al., 2022). The line-of-sight velocities are determined by calculating the frequency shift of a line due to the Doppler effect caused by the rotation of the disc. Since material moving towards/away from the observer is blue/red-shifted, the velocities at every region of the disc can be determined. Data cubes are then produced composed of channel maps showing the disc emission at a specific velocity (or frequency). Figure 1.9 illustrates how the molecular line emission can be viewed. The individual channel maps are shown in the centre. Summing over all the channel maps shows the overall morphology of the line emission (top right of Fig 1.9). A velocity weighted summation of the channel maps reveals the overall velocity structure of the disc (middle right of Fig 1.9).

In the absence of any processes altering the disc structure, the rotation of the disc is expected to show a Keplerian ("butterfly") pattern, i.e. opposite sides of the disc at a specific velocity are symmetric and smooth (Horne & Marsh, 1986). Any deviations from Keplerian rotation, as will be discussed below, can then be at-

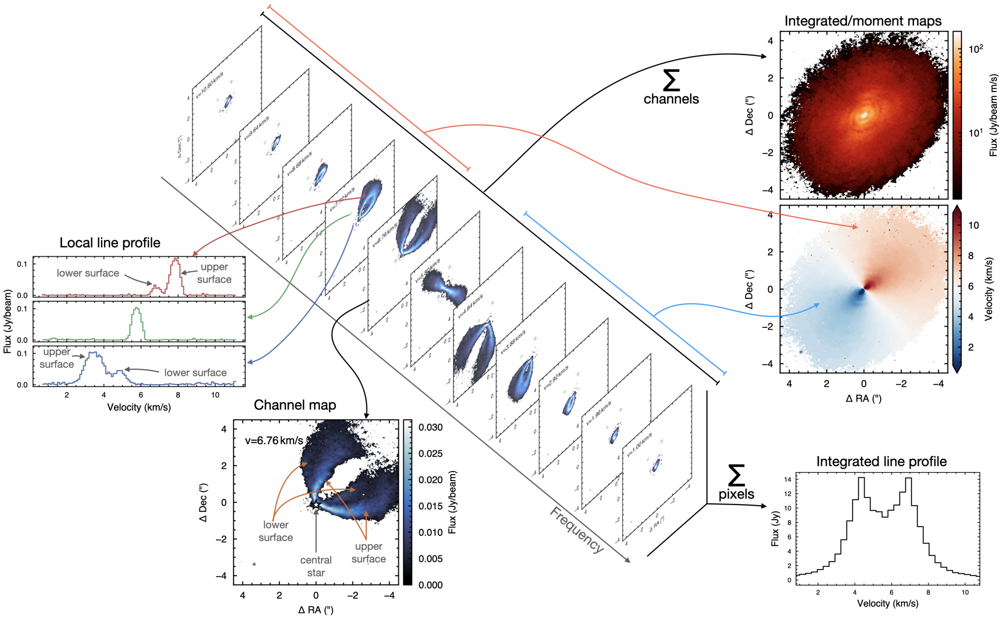


Figure 1.9: This figure shows the multiple ways to view a molecular data cube using ^{12}CO J=2-1 ALMA observations of HD 163296 (Isella et al., 2016; Andrews et al., 2018) as an example. From the individual channel maps at a given velocity shown in the center of the figure, a single channel map can be extracted, the local line profile can be obtained by integrated the flux from a given region, moment maps can be computed by integrating along the velocity axis, and a global spectrum can be obtained by spatially integrating over all the pixels. Image from Pinte et al. (2022).

tributed to some of the processes that are the origin of the dust structures mentioned in §1.2.1.

Kinks due to a planet

As mentioned earlier in §1.2.1, planets are often used to explain many of the observed substructures. However, they are not a unique explanation, and often cannot be distinguished from non-planet explanations based on the continuum data alone. If the planet is massive enough it will generate spiral wakes perturbing the velocity structure of the disc. The presence of the planet can then be detected in the gas emission by localised deviations from the Keplerian rotation of the disc in specific channel maps (Perez et al., 2015) as shown recently by Pinte et al. (2018b, 2019, 2020). These deviations are referred to as *kinks*.

An example of a kink due to a planet in HD 97048 is shown in Figure 1.10

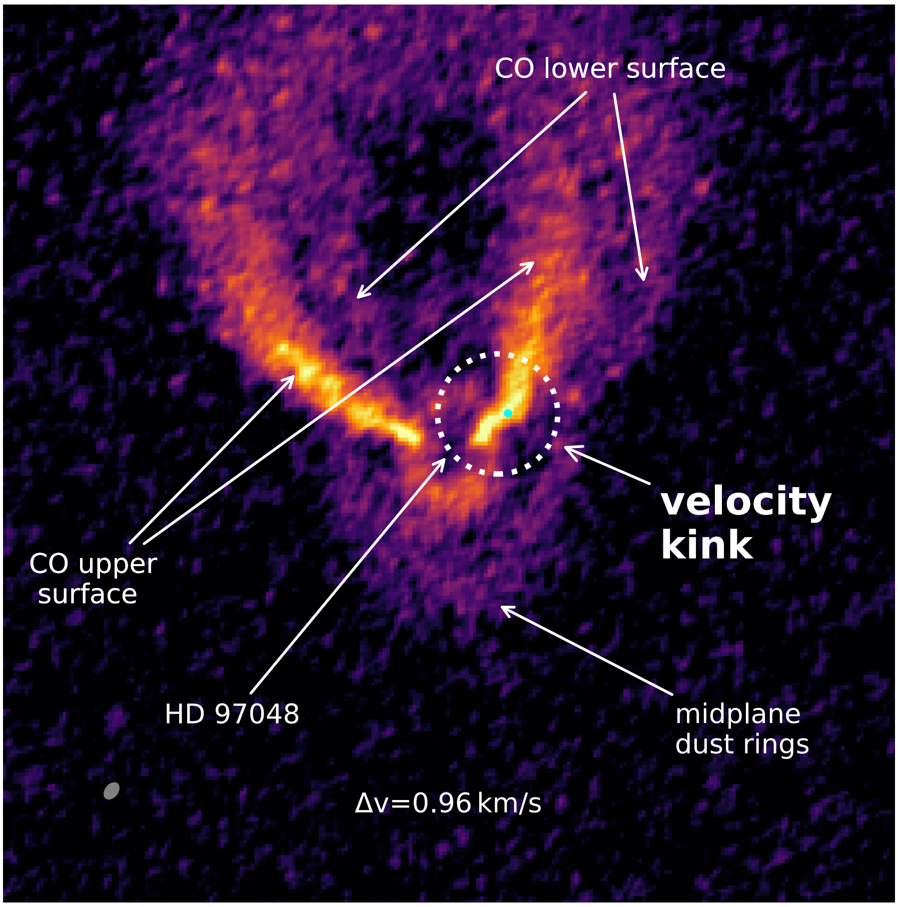


Figure 1.10: ALMA observations of ^{13}CO in HD 97048 at velocity +0.96km/s from the systemic velocity. The presence of a planet is inferred from kink which is marked by a dotted circle. The location of the kink is coincident with the gap detected in the continuum, strengthening the evidence of planet-disc interactions (Pinte et al., 2019).

(Pinte et al., 2019). Here, both the continuum emission and channel maps showing iso-velocity curves are visible. If the disc was perfectly Keplerian, then the butterfly pattern of the channel maps would be symmetric and smooth. However in Fig 1.10, only the left side of the disc shows a smooth iso-velocity curve. On the right side, a kink is seen in the iso-velocity curve which lines up with the gap seen in the continuum data. Since the kink is asymmetric and localised to a small region, alternative mechanisms such as snowlines or zonal flows could be ruled out. The combined signatures strengthen the evidence of planet-disc interactions being responsible for ring & gap structure in HD 97048.

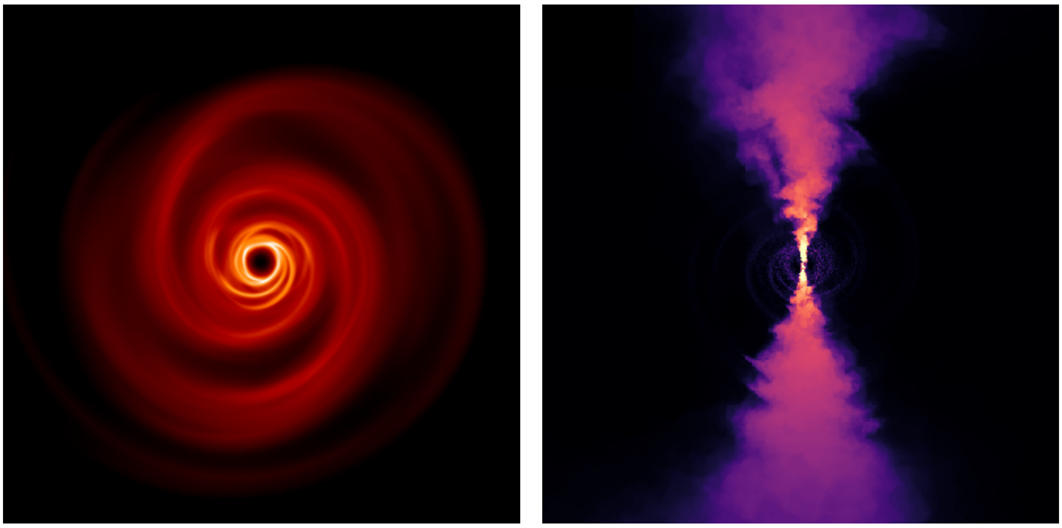


Figure 1.11: Left – A simulation of a $0.25M_\odot$ disc around a $1M_\odot$ star where spirals due to gravitational instability are clearly present. Right – The channel map at systemic velocity showing the GI wiggle, the kinematic signature of gravitational instability, which is present at all radii. Image adapted from Terry et al. (2022).

GI wiggles

A characteristic signature of massive discs are large scale spiral structures due to gravitational instabilities. However, spiral structures can also be explained by planets and can appear morphologically similar (Dong et al., 2015a,b). While the pitch angles of the spiral arms allow the two scenarios to be distinguished from each other, it requires high spatial resolution, and can be made more difficult by deprojection effects or low contrast ratios between spiral arms (Forgan et al., 2018b). The gas kinematics offer another method to distinguish these two scenarios. Kinks due to the planet follow the planet-induced spiral wake (Bollati et al., 2021; Calcino et al., 2021; Verrios et al., 2022). However, spirals due to GI are global and continuously perturb the disc velocity. Inside/outside the spiral arms the velocity relative to average background velocity is increased/decreased. Hence, the emission is shifted to an adjacent channel map with a larger/smaller velocity. This results in a zigzag pattern, or a *GI wiggle*, that is present at all azimuths and radii as shown in Figure 1.11 (Hall et al., 2020). The global nature of the deviations from Keplerian velocity reveals the presence of GI in a disc. Recently, the kinematics of Elias 2-27 have revealed large global deviations (Paneque-Carreño et al., 2021) hinting at the presence of GI consistent with previous theoretical predictions (Meru et al., 2017; Forgan et al., 2018b; Hall et al., 2018).

1.2.3 Disc mass estimates

Accurate measurements of the disc mass is necessary in understanding both disc evolution and planet formation. However, the estimates from observations are often poorly constrained as the majority of the disc is composed of unobservable molecular hydrogen. Hence, the disc mass can only be inferred indirectly based on the small fraction of the disc that is observable.

A common method to infer the total mass is to measure the dust mass from the integrated continuum emission which is more easily observable. Assuming the continuum emission is from an optically thin, isothermal part of the disc, the first measures of the disc mass, M_{disc} was directly calculated from the flux density F_ν using (Beckwith et al., 1990; Andre & Montmerle, 1994)

$$M_{\mathrm{disc}} = \frac{d^2 F_\nu}{\kappa_\nu B_\nu(T_{\mathrm{dust}})}, \tag{1.1}$$

where d, is the distance, κ_ν is the opacity which assumes a dust-to-gas mass ratio, $B_\nu(T_{\mathrm{dust}})$ is the Planck function at a constant dust temperature, T_{dust}. The typical dust-to-gas mass ratio used for this calculation is 0.01, which is the same as that of the Milky Way (Bohlin et al., 1978). Furthermore, this often assumes the dust distribution traces the gas, and does not take into account dust evolution. Hence, the mass estimates remain uncertain due to the many assumptions that go into the disc radius, dust-to-gas mass ratio, dust opacities, distribution of grain sizes, or local over/under-densities (Testi et al., 2014; Andrews, 2015, and references therein).

While a more direct tracer is an option with hydrogen deuteride (HD), it has only been detected in three protoplanetary discs (Bergin et al., 2013; McClure et al., 2016). Although HD is an independent tracer of the gas mass, the vertical structure of the disc influences the HD $J = 1{-}0$ flux, which leads to increasingly uncertain disc mass for higher mass discs. This uncertainty can be decreased by complementary observations of HD $J = 2 - 1$, which is less influenced by disc mass, and with knowledge of the vertical structure (Trapman et al., 2017). An alternative is to use carbon monoxide (CO) as a tracer for the disc mass. It's the second most abundant molecule, and its rotational transitions are bright at millimeter wavelengths. The total mass is found by assuming a constant CO/H_2 abundance ratio. However, it's not without its issues: CO and some of its more common isotopologues are optically thick which can result in an underestimated disc mass (Bergin & Williams,

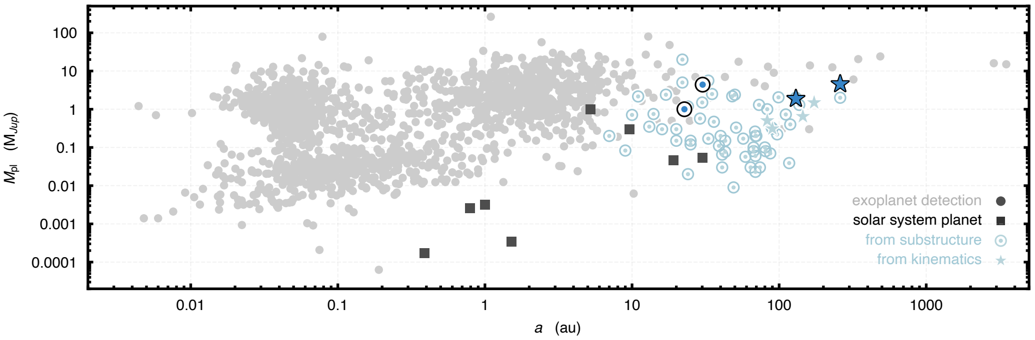


Figure 1.12: The exoplanet population showing both detected exoplanets (grey dots) and the planets inferred (blue markers) from disc substructures (encircled dots) and kinematics (stars). Larger symbols represent the confirmed planets in protoplanetary discs: PDS 70b and PDS 70c, HD 97048b and HD 163296b. Image from Pinte et al. (2022).

2017). Additionally, the abundance of CO and its isotopologues can be depleted by freeze-out onto icy grains in the cold mid-plane (van Zadelhoff et al., 2001), and photodissociation by far UV radiation in the upper parts of the disc (Reboussin et al., 2015). These issues can result in large discrepancies of mass estimates from CO. Using ^{13}CO and C^{18}O Ansdell et al. (2016) found gas masses to be low relative to the dust mass, and from estimates using HD, indicating a large amount of CO depletion has occurred. However, using the rarest stable CO isotopologue, ^{13}C^{17}O, Booth et al. (2019); Booth & Ilee (2020) found disc masses to be much higher compared to optically thicker CO isotopologues. By using N_2H, a tracer of CO poor gas, combined with C^{18}O to measure the CO/H_2 ratio, Trapman et al. (2022) measured disc masses more in line with estimates from HD. In general, the disc masses tend to be higher for more direct and robust tracers. However, disc masses remain a major source of uncertainty.

1.3 Planet formation

If planets are assumed to be cause of the substructures observed in protoplanetary discs, their masses can be estimated. Using a combination of simulations and analytics Zhang et al. (2018); Lodato et al. (2019) obtained planet mass estimates from analysis of gap widths and depths in the DSHARP discs (Andrews et al., 2018). A majority of the planet masses were found to be in the giant planet regime, i.e. around a Saturn mass or larger. Similarly, planet mass estimates in some of the DSHARP discs were also obtained from the analysis of the kinematics (Pinte et al., 2020) which were found to be 4-10 times larger, i.e. Jupiter masses or larger. Figure 1.12 shows the inferred planet mass from disc substructures and kinematics compared to the detected exoplanets.

If the hypothesis of planets is correct, how they formed is vital for understanding the evolution of the disc, and the planet especially if they ultimately have to evolve from the coloured dots to the grey dots in Figure 1.12. However, it is also possible that the population of observed exoplanets are not the same as inferred planets in protoplanetary discs. There are two main mechanisms for planet formation: core accretion (Pollack et al., 1996); and gravitational instability (Boss, 1997).

1.3.1 Core accretion

The core accretion model describes a *bottom up* method of building giant planets. It begins with small micron sized dust grains in the disc. By colliding, coalescing, and growing larger, the small grains become planetesimals. Gravitational interactions then produce a planetary embryo of a fraction of an Earth mass. The process of building a giant planet can then be described in three phases (Pollack et al., 1996; Alibert et al., 2005; Helled et al., 2014). In phase 1, the core is formed by accreting planetesimals within a few Hill radii (the feeding zone) of the growing planet. The accretion rate of solids during this phase is much higher than the gas accretion rate. This phase ends when there are no more planetesimals available for the planet to accrete. Once in phase 2, the mass increase is much slower due to the lack of planetesimals. Instead the planet slowly gains mass by accreting gas in the envelope. While gaining mass, the feeding zone of the planet also increases which allows it to accrete further planetesimals. This phase ends when the mass of the accreted gas is comparable to mass of the accreted planetesimals. At this point the escape velocity of the gas around the core is large enough to prevent gas from escaping. Thus phase 3 begins with runaway gas accretion during which time the mass also increases substantially.

Recent observations of protoplanetary discs pose a challenge to traditional core accretion models. One issue is the timescale required to form giant planets, which requires $\sim 10^6 - 10^8$ years (Lissauer et al., 2009; Goldreich et al., 2004). As giant planets are often used to explain the observed substructures, the formation timescale is problematic as it exceeds protoplanetary disc lifetimes which is only a few million years (Williams & Cieza, 2011; Kimura et al., 2016). There is an increasing amount of evidence indicating that planet formation begins much earlier in the disc's lifetime. Ring & gap structures have been found in Class I discs, such as HL Tau (ALMA Partnership et al., 2015), which are still surrounded by an external envelope. Simulations have suggested multiple giant planets as the cause of the structures in HL Tau (Dipierro et al., 2015). Even discs that are less than half a million years old reveal ring & gap structure (IRS 63, Segura-Cox et al., 2020). Given the youth of these discs, and the masses and radial location of the proposed planets, core accretion struggles to form these planets fast enough. This is referred to as the *timescale* problem.

A second issue involves the growth of small dust grains to the kilometer sized planetesimals which are the starting point in typical simulations of core accretion (Helled et al., 2014). While dust grains can grow to millimeter sized grains by sticking, it is less straightforward for the grains to continue to grow and form kilometer sized planetesimals. This is due to the *bouncing barrier*. Once grains reach a certain size (which depend on the disc properties), they no longer stick to similar sized grains. Instead they bounce off each other. Hence further growth becomes difficult (Windmark et al., 2012). However, the large grains can still grow by sticking to smaller grains.

A popular mechanism to bypass this barrier is the streaming instability (Youdin & Goodman, 2005; Johansen & Youdin, 2007; Youdin & Johansen, 2007; Simon et al., 2016) which is an instability of aerodynamically coupled system of gas and dust particles. The instability arises from speed difference of gas and dust. As gas is pressure supported against the radial component of stellar gravity, it will orbit at sub-Keplerian velocities. However, the dust orbits at Keplerian velocity as it is not pressure supported. A consequence of different velocities is the dust experiences a headwind and loses angular momentum resulting in inward migration of the dust (Adachi et al., 1976; Weidenschilling, 1977). Migration accelerates as dust grains grow larger. However, if there are any density enhancements (dust clumps), back-reaction from the particle onto the gas will result in the clumps migrating slower. Dust further out will continue drifting inwards and eventually pile on top of the clump, increasing its mass and further slowing its migration. This positive feedback can result in planetesimals growing quickly and is the basis of the streaming instability.

1.3.2 Gravitational instability

When protoplanetary discs are massive enough the self-gravity of the disc plays a key role in the disc's evolution resulting in gravitational instabilities. A key difference in relation to the core accretion model is that it does not require building a solid core *bottom up* from small grains. Instead, gravitational instabilities in the disc can form clumps of gas and dust, which then contract and collapse to form giant planets (Boss, 1997). Traditionally, gravitational instability has been primarily used to explain the formation of giant planets, however recent work in MHD simulations have shown that this model can also form intermediate mass planets, ranging from Super-Earth to Neptune-sized planets (Deng et al., 2021). A major advantage over core accretion is that it can form planets much more rapidly on dynamical timescales (Gammie, 2001).

To analytically understand the conditions required for gravitational instability, some assumptions must first be made. Any perturbations must be small enough to be described by linear perturbation theory. The spiral density waves resulting from these small perturbations must be tightly wound (the WKB approximation), allowing the long-range effects of gravity to be negligible. Assuming an infinitesimally thin disc, the dispersion relation $D(\omega, k, m)$, can be described by the wave frequency ω, the azimuthal wavenumber m, and the radial wavenumber $k = 2\pi/\lambda$,

$$D(\omega, k, m) = (m\Omega - \omega)^2 - \kappa_{\mathrm{ep}}^2 + 2\pi G\Sigma|k| - k^2 c_s^2 = 0 \tag{1.2}$$

or

$$(m\Omega - \omega)^2 = \kappa_{\mathrm{ep}}^2 - 2\pi G\Sigma|k| + k^2 c_s^2 \tag{1.3}$$

where Ω is the angular velocity, κ_{ep} is the epicyclic frequency, c_s is the sound speed, Σ is the surface mass density, and G is the gravitational constant. The middle term on the right side of Eq. 1.3 is the destabilising gravitational forces, and the first and last term are the stabilising rotational and pressure forces. For axisymmetric disturbances $m = 0$, giving

$$\omega^2 = \kappa_{\mathrm{ep}}^2 - 2\pi G\Sigma|k| + k^2 c_s^2. \tag{1.4}$$

If $\omega^2 > 0$, the solution is real and the disc is stable. For an unstable disc, the most unstable wavelength is when $\mathrm{d}\omega^2/\mathrm{d}k = 0$. Hence, by directly differentiating, this is when

$$|k| = \frac{\pi G\Sigma}{c_s^2}. \tag{1.5}$$

Substituting this back into Eq. 1.4 and requiring $\omega^2 > 0$ then gives

$$\begin{aligned} \kappa_{\rm ep}^2 - k^2 c_s^2 &> 0 \\ \kappa_{\rm ep} &> k c_s \\ Q = \frac{\kappa_{\rm ep} c_s}{\pi G \Sigma} &> 1 \end{aligned} \tag{1.6}$$

where Q is known as Toomre stability parameter (Toomre, 1964). For a marginally stable disc $Q \approx 1$. For a Keplerian disc, $\kappa_{\rm ep} = \Omega$ and hence Q can be rewritten as

$$Q = \frac{c_s \Omega}{\pi G \Sigma} \tag{1.7}$$

where $\Omega = \sqrt{GM/R^3}$ is the orbital Keplerian frequency of the disc. The WKB approximation is valid when the radial wavelength is much smaller than the radius ($|kR| \gg 1$ or $\lambda/R \ll 1$). Typical protoplanetary discs are thin with aspect ratios of $H/R \approx 0.1$. The largest wavelength possible in a disc is $\lambda = \zeta H$, where the constant $\zeta = \mathcal{O}(1)$. Hence, the stability criterion is still valid for protoplanetary discs which have a finite thickness, despite the derivation assuming a razor-thin disc.

This model requires two conditions to be fulfilled in order to work. First, the self-gravity of the disc needs to be comparable to the gravitational effects of the central star. The second condition is that the cooling timescale $t_{\rm cool}$ must be comparable to the orbital timescale Ω^{-1}. A simple prescription satisfying this condition is (Gammie, 2001),

$$t_{\rm cool} = \beta \Omega^{-1}, \tag{1.8}$$

where β is of order unity. Cooling must be efficient enough for instabilities to be able to further collapse into fragments. If cooling is too slow, increasing temperatures in the instabilities provide strong enough pressures to prevent further collapse.

1.4 Planet-disc interactions

Planets being hypothesised as the origin of many of the substructures in §1.2.1 necessitates understanding how planets interact with disc. Any planet that forms in a disc will result in the exchange of angular momentum between the disc and the planet via the spiral wakes generated by the planet (Lin & Papaloizou 1979, also see review by Kley & Nelson 2012). Exactly how the planet interacts with the disc depends on both the planet and disc properties, and can be divided into several regimes as outlined below (see review by Paardekooper et al., 2022).

1.4.1 Type I migration

The regime of Type I migration is the realm of low mass planets in discs with larger turbulent viscosities. For planetary sized objects the dominant source of torques comes from gravitational interaction with the disc – the wave torque (Goldreich & Tremaine, 1980; Artymowicz, 1993), and the corotation torque (Ward, 1991). Type I migration is defined by an unsaturated corotation torque due to turbulent viscosity (Masset, 2001). If the planet mass is not high enough or the disc viscosity is large, then coorbital material lost by the disc to the planet is refilled by the disc viscosity, i.e. the planet is unable to carve open a gap in the disc.

1.4.2 Type II migration

Type II migration is characterised by the ability of the planet to open up a gap. In this regime, the disc viscosity is unable to replenish the coorbital material lost by the disc to the planet. Traditional Type II migration assumed that the gap carved by the planet was devoid of gas. Thus the planet migrates at the viscous evolution time of the disc (Lin & Papaloizou, 1986). However, more recent hydrodynamic simulations have shown that gas flowing across the gap is not negligible (Artymowicz & Lubow, 1996; Bryden et al., 1999; Kley, 1999; Lubow et al., 1999; Masset & Snellgrove, 2001). Additionally, gap-crossing flow has been shown to cause the planet to migrate independently of disc accretion (Duffell et al., 2014; Dürmann & Kley, 2015, 2017). Type II migration does not necessarily require a massive planet. If instead the disc viscosity is sufficiently low ($\alpha < 10^{-4}$), even a Neptune mass planet can open up a gap in a disc (Duffell & MacFadyen, 2013). An updated gap-opening criteria (Kanagawa et al., 2018) can be derived by requiring the angular momentum flux due to the spiral wakes of the planet to exceed the angular momentum flux of the viscous accretion disc (see Duffell & MacFadyen, 2013; Kanagawa et al., 2015). A planet is able to open up a gap if its mass ratio planet/star, q exceeds

$$q \gtrsim 5h^{3/2}\left(\frac{\nu}{r_p^2\Omega_p}\right)^{1/2} = 5h^{5/2}\alpha^{1/2}, \tag{1.9}$$

where $h = H/R$ is the disc aspect ratio, $\nu = \alpha c_s H$ is the kinematic viscosity, and r_p is the planet's radial location.

1.4.3 Type III migration

In the Type III regime, the flow of material in the coorbital region plays a dominant role in the migration of the planet. As a planet migrates, an asymmetry can develop in the coorbital region between the leading and trailing side of the planet, resulting in a non-cancellation of torques. This can lead to very rapid migration in either direction depending on the gas flow near the planet (Masset & Papaloizou, 2003; Pepliński et al., 2008a,b,c). The fate of the planet in this regime depends on whether it's on the slow or rapid Type III rate (both types are faster than Type I and II). If it's slow Type III, then once migration is slow enough that the planet can begin to carve open a gap, it can transition to Type II. If it's fast Type III, then the planet migrates too quickly for it to be able to open up a gap.

1.4.4 In the context of GI discs

In their youth, discs can be massive enough that they become gravitationally unstable developing large scale spiral structures throughout the disc. Studies of planet migration usually assume a GI-stable disc (Paardekooper et al., 2022, and references therein), which tend to be smooth and axisymmetric. In contrast, the spiral structures due to gravitationally instabilities create a far more turbulent environment where planets simply migrate rapidly inward (Baruteau et al., 2011; Malik et al., 2015). Even giant Jupiter sized planets were unable to open up a gap.

The treatment of the thermodynamics changes how the disc evolves, and hence how a planet migrates. In Baruteau et al. (2011) and Malik et al. (2015) the thermodynamics was treated using a simple constant β model (see §2.2.5). This resulted in a globally turbulent gravitationally unstable disc. However, with more realistic thermodynamics, the inner regions of the disc remains gravitationally stable and smooth. Stamatellos & Inutsuka (2018) found that the planets are able to slow down and open up a gap in the inner regions of the disc after an initial phase of rapid inward migration in the outer unstable parts of the disc. However, in those simulations the planet accreted a significant amount of material growing to beyond the brown dwarf limit. Thus, it's unclear which had a bigger role in the planet's ability to open up a gap, the planet's mass or the thermodynamics. This aspect is investigated in detail in Chapter 3.

1.5 Warp theory

Simulations of gravitationally unstable protoplanetary discs are often done with a disc around a single star isolated from the effects of its star forming environment. Additionally, the angular momentum vectors of the star and disc are assumed to be parallel. While this simplified approach allows for an easier analysis of the physics of interest, it neglects the chaotic star forming environment which can alter the disc's evolution.

1.5.1 Star forming environment

Contrary to the simplified picture presented earlier, observational and theoretical studies strongly suggest that protoplanetary discs do not evolve in isolation (Lada & Lada, 2003). From observations we see that a majority of stars are part of multi-star systems (Fischer & Marcy, 1992; Frankowski et al., 2007; Raghavan et al., 2010; Luo et al., 2022). These systems can either be formed by disc fragmentation (Adams et al., 1989; Kratter et al., 2010a; Backus & Quinn, 2016) or from turbulent fragmentation (Bate, 2018). Disc fragmentation primarily results in close in multiples (Tobin et al., 2016a; Ilee et al., 2018) at ∼100s of au in scale, while turbulent fragmentation initially produces multiples at wider separations, which can also migrate in (Bate, 2018).

Theoretical studies have shown that if a companion is misaligned relative to the orbital plane of the disc, the disc can become warped. The companion can either be internal (Facchini et al., 2013) or external (Xiang-Gruess & Papaloizou, 2014). A disc is considered to be warped if it's angular momentum vector varies with radius. There are a few mechanisms that can result in warped discs which are more relevant when discs are young, and hence more likely when they are gravitationally unstable. Multiples formed at wide separation as a result of turbulent fragmentation accrete gas where the net angular momentum vectors are in different directions. A consequence of this is misaligned stellar spins and accretion discs (Bate, 2018; Offner et al., 2022).

While discs are young enough to still be influenced by their star forming environment, stellar encounters are also more prevalent (Vincke & Pfalzner, 2016; Bate, 2018). The rate of encounters diminishes over time (Pfalzner, 2013). Recent estimates indicate a Sun like star has roughly a 20% probability of encountering a flyby within 3Myrs in star clusters of various densities (Winter et al., 2018). Flybys can approach the disc from any direction and if misaligned relative to the disc, the flyby provides a misaligned torque which warps the disc (Clarke & Pringle, 1993; Nealon et al., 2020a).

In these early times, infalling material can also play a role in the disc's evolution (Bate, 2018; Sakai et al., 2019). If the angular momentum vector of the infalling material is different to that of the disc, i.e. if it's misaligned relative to the disc, then interaction with the infalling material will result in the angular momentum vector no longer being constant with radial location. Figure 1.13 from Bate (2018) shows the outcome of a disc after it has interacted with both misaligned

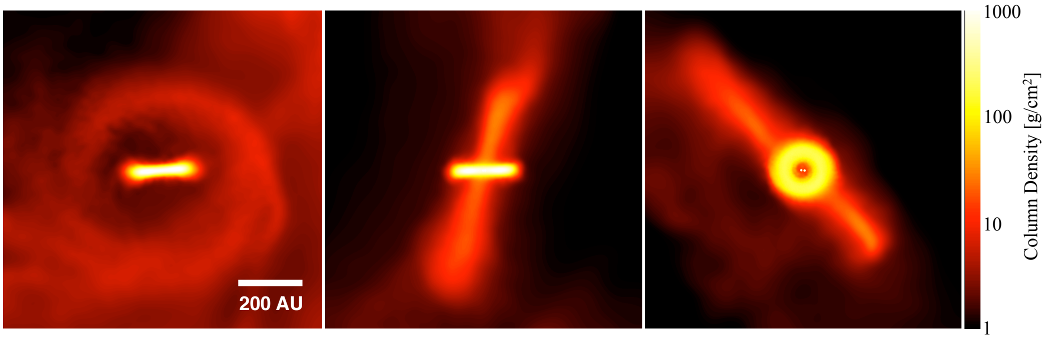


Figure 1.13: The formation of a strongly misaligned circumbinary disc due to the chaotic star forming environment. Initially, the two protostars had their own circumstellar disc around them. After interacting with each other, a larger circumbinary disc was formed. Later on, the disc captured cloud material with angular momentum completely misaligned relative to binary resulting in a strongly misaligned disc (Bate, 2018).

cloud material and a stellar companion. Initially the two protostars had their own separate discs of around $0.1M_{\odot}$ around them. After they encountered each other, the product was a larger circumbinary disc. Later, the binary captured additional cloud material with angular momentum completely misaligned relative to the binary. This resulted in a strongly misaligned disc as shown in Fig 1.13.

1.5.2 Important timescales

In aligned discs there are a couple of important timescales. The first is the dynamical timescale which is defined by

$$t_{\mathrm{dynamical}} = \frac{1}{\Omega(R)}, \tag{1.10}$$

where Ω is the angular frequency at radius R. It is defined as the time taken for material to complete an orbit. This is useful for understanding the timescale of various physics such as the cooling timescale in self-gravitating discs in §2.2.5. The second is the sound-crossing timescale which describes how long it takes for material to communicate over distance R. It is defined using the sound speed c_{s} as

$$t_{\mathrm{s}} = \frac{R}{c_{\mathrm{s}}(R)}. \tag{1.11}$$

For warped discs an important timescale is how long it takes for a warp to be damped. The timescale depends on whether it is in the diffusive or wave-like regime.

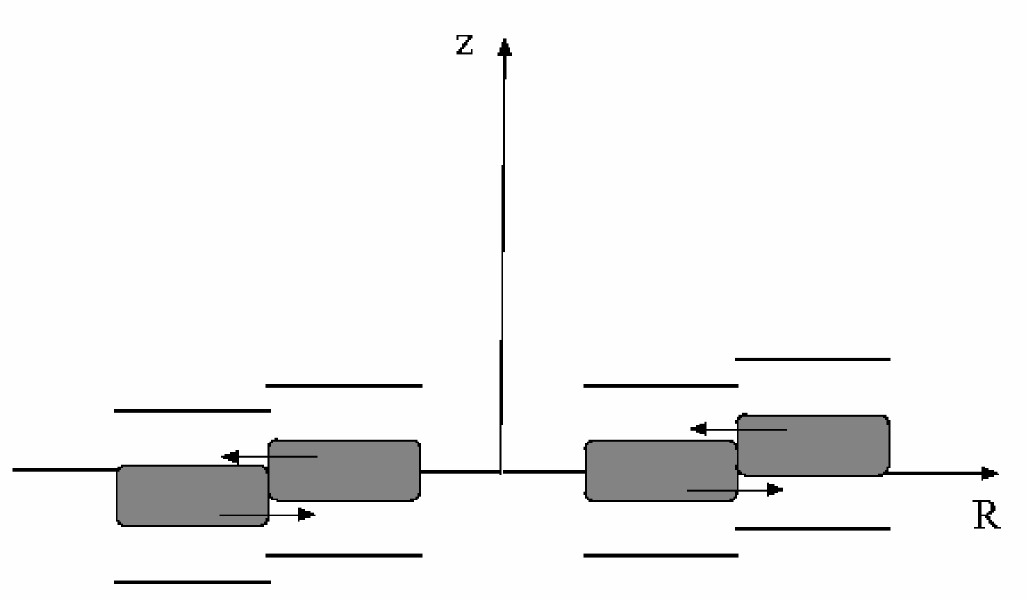


Figure 1.14: A schematic view of a warped disc (Lodato & Pringle, 2007). The grey boxes represent the pressure between neighbouring rings of gas. The arrows show the direction of the radial pressure gradient induced by the warp. As gas orbits, it will feel an oscillating pressure gradient that varies with the vertical height.

In gravitationally unstable discs, α can be larger than H/R implying the warp propagates diffusively. However, in the regions of interest α is comparable to H/R. Hence, the warp is in the intermediate regime where the warp propagates diffusively in the inner regions, and wave-like in the outer regions. Hence, as the outer gravitationally unstable parts of the disc are the regions of interest in this thesis, the warp is considered to be in the wave-like regime (Martin et al., 2019). The damping timescale is defined by the disc viscosity α, and is proportional to the orbital timescale,

$$t_{\rm damp} \sim \frac{1}{\alpha\Omega(R)}. \tag{1.12}$$

With these timescales, we can understand how warps can affect the behaviour of protoplanetary discs, as done in Chapter 6.

1.5.3 Warp propagation

To visualise how a warp propagates in a disc, consider Figure 1.14 from Lodato & Pringle (2007) which shows a schematic view of a warped disc. The grey boxes represent the pressure between neighbouring annuli of gas. At this moment in the orbit, adjacent rings of gas are strongly misaligned. As a result a radial pressure gradient is induced by the warp which depends on the vertical height, as indicated by the arrows. After the material completes another quarter of an orbit the neighbouring annuli of gas will be almost aligned. Hence, the radial pressure gradient induced by the warp will be weak. Thus, as material orbits in a warped disc, it will feel an oscillating radial pressure gradient that depends on the amplitude of the warp. This oscillating radial pressure gradient results in epicyclic motion in the gas orbits, which in turn launches a wave that propagates the warp. Since the warps in this thesis are located in the outer parts of the disc, they are considered to be in the wavelike regime and so propagates through the disc as a wave which is damped non-locally. However, as the warps are in the intermediate regime, the further in the warp is located, the more they act in a diffusive manner where the warp is locally damped.

1.6 Thesis focus

The focus of this thesis explores how the evolution of young, massive, gravitationally unstable discs is influenced by physical processes occurring in the disc. Recent continuum observations of protoplanetary discs (eg. Andrews et al., 2018) reveal that large scale spiral structures are rarely seen. From these observations, one could fairly say that discs in general may not be massive enough to be gravitationally unstable. In this thesis, we instead explore whether evidence of gravitational instability can be hidden. To investigate this question, two processes that are plausible when discs are young are considered: planet-disc interactions; and warps.

Before the impact of planet-disc interactions on the structure of the disc could be investigated, the fate of planet migration in a gravitationally unstable disc had to be studied. In Chapter 3, I investigate whether planets can survive in a gravitationally unstable disc if the disc thermodynamics is modelled in a way that mimicked a realistic gravitationally unstable disc. The aim is to see whether the gravitational stable inner disc was a region where the planet would be able to slow its migration.

In Chapter 4, I introduce the concept of how a planet interacts with a gravitationally unstable disc. Mock observations were made to study the resulting implications on the observability of evidence of gravitational instability. This is initially shown for a single planet and disc mass, but is extended to a suite of simulations in Chapter 5. By carrying out a grid of simulations with different planet and disc masses, I explore when planet-disc interactions dominate the disc's evolution over gravitational instability and vice-versa.

In Chapter 6, I explore how a warp alters the evolution of a gravitationally unstable disc. As warps had not been previously considered in discs in this mass regime, I start with a warp that was numerically introduced to isolate the impact of the warp on the disc structure. As with planet-disc interactions, the aim is to explore whether a warp could hide evidence of gravitational instability.

In Chapter 7, I describe how the results in this thesis could be further extended. Finally, I conclude the work in Chapter 8.

2

Numerical Methods

Studies of hydrodynamic problems in astrophysics are investigated using numerical simulations. There are two methods that are typically used to perform these simulations: grid-based, and particle-based hydrodynamics codes. Grid-based codes use an Eulerian grid which is generally fixed. In the Eulerian approach, individual fluid particles are not followed. Instead the evolution of the fluid properties is considered at all grid cells. In contrast, particle based codes use a Lagrangian approach that follows individual particles of fixed mass which describe the fluid properties. Due to their fundamental difference in how they solve the fluid equations, both methods have their advantages and disadvantages depending on the specific problem at hand.

This thesis focuses on the evolution of gravitationally unstable protoplanetary discs where gravitational instabilities result in regions of high density contrasts. While grid-based codes can use adaptive meshes or nested grids to resolve the high density regions, particle-based codes by their nature have higher resolution in the denser regions of the disc. Grid-based codes require all cells to have non-zero mass, wasting computational time in low density regions where nothing of interest is occurring. Whereas particle-based codes can save computational time in these regions to instead focus on the high density areas. This thesis also features warped gravitationally unstable protoplanetary discs where the complex geometry of the system is naturally suited to particle-based codes as by its formulation, conservation of properties such as linear and total angular momentum is as accurate as time-stepping methods allow and is independent of the disc orientation. Thus given the nature of the problems studied in this thesis, I have used PHANTOM (Price et al., 2018a), a smoothed particle hydrodynamics (SPH) code for the advantages it provides over grid-based codes.

2.1 Smoothed Particle Hydrodynamics

2.1.1 Representing a fluid with particles

SPH was originally independently developed by Lucy (1977) and Gingold & Monaghan (1977) for simulations of astrophysical fluids (also see reviews by Monaghan, 1992, 2005). It is a Lagrangian particle method that solves the equations of hydrodynamics. Although SPH is a particle based method that computes fluid properties at a particular point, the particles do not represent individual point masses. Instead, the concept behind SPH is that each particle represent a volume of fluid with properties representative of the surrounding region.

There are two important concepts that determine how the properties of the fluid is calculated. The first is the kernel W, which is a weighting function that determines how much each neighbouring particle should contribute based on the distance $\boldsymbol{r}_a - \boldsymbol{r}_b$, where $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$ are the positions of particle a and b, respectively. The second is the smoothing length h, which defines the length-scale over which contribution of particles are weighted.

To illustrate these concepts, consider calculating the density of example particles in Figure 2.1. As we shall see, measuring the density at the location of a particle is crucial to our calculations. The density is simply mass divided by volume, hence the relative positions and masses of the particles is used to determine the local density. In Fig 2.1, assuming equal mass particles the more closely packed red region represents an area of higher density compared to the loosely packed particles in the blue region. The closer a neighbouring particle is to a particular particle, the more influence it has on the calculations. Therefore a weighting function is required so that the closest neighbours contribute the most while the further away particles have little influence. This weighting is represented by the colour gradient in Fig 2.1. The size of the shaded regions represents the smoothing length. Bringing these concepts together, the density ρ at the location of particle a can be calculated in SPH form as

$$\rho_a = \sum_b m_b W(\boldsymbol{r}_a - \boldsymbol{r}_b, h_a), \tag{2.1}$$

where the subscript b refers to all the neighbour particles. A good choice of smoothing length should ensure that the size of region being considered, over which the fluid properties are calculated, must be

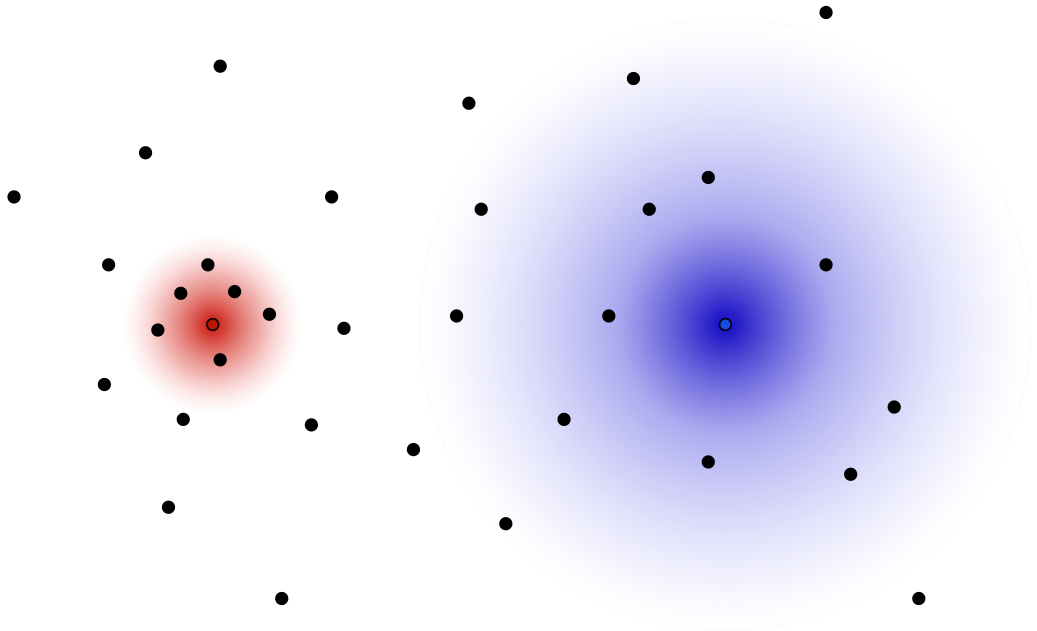


Figure 2.1: A set of particles representing the density distribution of a fluid based on Figure 1c of Price (2012). A high and low density region is indicated at the red and blue particle respectively. The shaded regions shows the particles that contribute to the density calculation and how the influence of each particle is considered. The size of the shaded region represents the size of the region considered for the calculation and is dependent on the smoothing length. The gradient represents the weighting function which weights the particle according to their distance from the indicated particle.

1. Smaller than the length-scale over which a variable evolves.
2. Large enough that the number of neighbours involved in the calculations represent a fluid.
3. Dynamic so that it can resolve both dense and sparse regions.

For the simulations in this thesis, the smoothing length is set as

$$h_a = h_{\text{fact}} \left(\frac{m_a}{\rho_a} \right)^{1/3}, \tag{2.2}$$

where m_a is the mass of the particle, and $h_{\text{fact}} = 1.2$ is a dimensionless parameter such that, for kernels extending to $2h$, the average number of neighbours for any particle is ~ 60 (e.g. Price, 2012). The mass of the particle is derived from the properties of the fluid being modelled, for equal mass particles this is the total physical mass divided by the number of particles. For equal mass particles, the smoothing length is then dependent on the density. Thus, it's evident that Eq 2.2 satisfies the criteria outlined above. A good choice for the kernel (hereafter referred to as $W_{ab}(h_a) \equiv W(\boldsymbol{r}_a - \boldsymbol{r}_b,\, h_a))$ must satisfy the following criteria:

1. The kernel must be strictly positive, smoothly decrease towards zero with increasing distance, and have smooth derivatives.
2. The kernel must not be influenced by the particle position. It must depend on the distance between two particles rather than whether the kernel is calculated on particle a or b. Thus a symmetric kernel is required, $W_{ab} = W_{ba}$.
3. At close distances, the gradient of the kernel must be shallow so that small movements by the closest neighbours do not greatly change the density estimate.
4. Conservation of mass requires the magnitude of the kernel function to be normalised using
$$\int_V W_{ab}(h_a)\mathrm{d}V' = 1. \tag{2.3}$$

A Gaussian kernel is a natural choice that satisfies all the above criteria. However, it is impractical as even at very large distances the kernel will never exactly equal zero. Hence, the density calculation becomes expensive as every particle will have to be considered regardless of how little the furthest away particles contribute. A more useful kernel is one that has a similar profile to a Gaussian, but truncated to zero beyond a couple of smoothing lengths. The kernel is set as

$$W_{ab}(h_a) = \frac{C_{\mathrm{norm}}}{h_a^d} f(q) \tag{2.4}$$

where the normalisation constant C_{norm} depends on the number of dimensions d. For three dimensions (as we will use in all our calculations), $C_{\mathrm{norm}} = 1/\pi$. The parameter q is defined as

$$q \equiv \frac{|\boldsymbol{r}_a - \boldsymbol{r}_b|}{h_a}. \tag{2.5}$$

The kernel function used for the simulations in this thesis is the M_4 cubic spline from the Schoenberg (1946) B-spline family given by

$$f(q) = \begin{cases} 1 - \frac{3}{2}q^2 + \frac{3}{4}q^3, & 0 \leq q < 1 \\ \frac{1}{4}(2-q)^3, & 1 \leq q < 2 \\ 0, & q \geq 2. \end{cases} \tag{2.6}$$

By only considering particles within a couple of smoothing lengths, only the nearby particles that influence the density the most are considered, decreasing the computational cost for this kernel to $\mathcal{O}(N_{\mathrm{neighbours}}N)$ compared to the $\mathcal{O}(N^2)$ of a Gaussian kernel.

2.1.2 General expressions in SPH

The calculation of the density and the concepts behind it in §2.1.1 are the core of SPH since other fluid properties such as the pressure, energy, and velocity depend on the density. However, the concept of smoothing can also be generalised for any property. For a scalar variable A, the approximate value at position $\boldsymbol{r}$ can be interpolated using

$$A(\boldsymbol{r}) = \int_V A(\boldsymbol{r}')\,\delta(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3\boldsymbol{r}' \tag{2.7}$$

where δ is the Dirac delta function and V is the volume of the region. The delta function can be replaced by the kernel function,

$$A(\boldsymbol{r}) = \int_V A(\boldsymbol{r}')\,W(\boldsymbol{r}-\boldsymbol{r}',\,h)\,\mathrm{d}^3\boldsymbol{r}' + \mathcal{O}(h^2), \tag{2.8}$$

if $W \to \delta$ as $h \to 0$. This expression is accurate to second order for positive kernels, where the last term is the error due to a finite kernel. For any finite density within the volume element, the above equation is equivalent to

$$A(\boldsymbol{r}) = \int_V \frac{A(\boldsymbol{r}')}{\rho(\boldsymbol{r}')}\,W(\boldsymbol{r}-\boldsymbol{r}',\,h)\,\rho(\boldsymbol{r}')\mathrm{d}^3\boldsymbol{r}' + \mathcal{O}(h^2), \tag{2.9}$$

which can be discretised to a set of particles of mass $m = \rho(\boldsymbol{r}')\mathrm{d}^3\boldsymbol{r}'$ allowing the scalar field to be calculated using

$$A(\boldsymbol{r}_a) \approx \sum_b^N m_b \frac{A_b}{\rho_b} W_{ab}(h_a) \tag{2.10}$$

where the summation is done for all b particles, and $A_b = A(\boldsymbol{r}_b)$, m_b, and $\rho_b = \rho(\boldsymbol{r}_b)$ are the scalar value, mass, and density of the b^{th} particle at its position $\boldsymbol{r}_b$. In this form, the only variable that depends on $\boldsymbol{r}_a$ is the smoothing kernel. Thus, the gradient of the scalar field is given by (Monaghan, 1992; Price, 2012)

$$\nabla A(\boldsymbol{r}_a) \approx \sum_b^N m_b \frac{A_b}{\rho_b} \nabla W_{ab}(h_a). \tag{2.11}$$

The usefulness of the general form of eq. 2.10 is further emphasised by being easily extended to vector quantities resulting in the following identities,

$$\boldsymbol{A}(\boldsymbol{r}_a) \approx \sum_b^N m_b \frac{\boldsymbol{A}_b}{\rho_b} W_{ab}(h_a), \tag{2.12}$$

$$\nabla \cdot \boldsymbol{A}(\boldsymbol{r}_a) \approx \sum_b^N m_b \frac{\boldsymbol{A}_b}{\rho_b} \cdot \nabla W_{ab}(h_a) \tag{2.13}$$

Using Eq 2.1 and the generalised equations above, many of the other physical properties required to model a fluid can be calculated.

2.1.3 Equation of state

All simulations use an ideal gas equation of state,

$$P = (\gamma - 1)\rho u \tag{2.14}$$

which relates the pressure P, to the density and internal energy, u. Here $\gamma = 5/3$ is the adiabatic index and the speed of sound, c_s is given by

$$c_s = \sqrt{\frac{\gamma P}{\rho}}. \tag{2.15}$$

The internal energy can be related to the gas temperature, T using

$$P = \frac{\rho k_B T}{\mu m_H}, \tag{2.16}$$

and hence,

$$T = \frac{\mu m_H}{k_B}(\gamma - 1)u, \tag{2.17}$$

where k_B is the Boltzmann constant, μ is the mean molecular weight, and m_H is the mass of hydrogen.

2.1.4 Hydrodynamic equations

For simplicity, I shall first consider the equations of compressible hydrodynamics for an inviscid fluid in the absence of any gravitational forces and external heating or cooling. The smoothing lengths will be considered constant, and equal for all particles which further simplify the equations. In Lagrangian form, the three main equations that need to be solved for hydrodynamic simulations at their simplest are the continuity equation,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\rho(\nabla \cdot \boldsymbol{v}), \tag{2.18}$$

the momentum equation,

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\frac{\nabla P}{\rho}, \tag{2.19}$$

and the energy equation,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\nabla \cdot \boldsymbol{v}. \tag{2.20}$$

Starting with the continuity equation for a particular particle a,

$$\begin{aligned} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} &= -\rho_a(\nabla \cdot \boldsymbol{v})_a \\ &= \boldsymbol{v}_a \cdot \nabla\rho_a - \nabla_a \cdot (\rho_a \boldsymbol{v}_a), \end{aligned} \tag{2.21}$$

which using the general identities in §2.1.2 gives the discretised version of the continuity equation as

$$\begin{aligned} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} &= \boldsymbol{v}_a \cdot \sum_b m_b \nabla_a W_{ab}(h) - \sum_b m_b \frac{(\rho_b \boldsymbol{v}_b)}{\rho_b} \cdot \nabla_a W_{ab}(h) \\ &= \sum_b m_b (\boldsymbol{v}_a - \boldsymbol{v}_b) \cdot \nabla_a W_{ab}(h). \end{aligned} \tag{2.22}$$

Similarly, the momentum equation for a particle a,

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{v}_a}{\mathrm{d}t} &= -\frac{(\nabla P)_a}{\rho_a}, \\ &= -\nabla_a \left(\frac{P_a}{\rho_a}\right) - \frac{P_a}{\rho_a^2}\nabla_a \rho_a \end{aligned} \tag{2.23}$$

where the identity, $\nabla(\psi/\phi) = (\phi\nabla\psi - \psi\nabla\phi)/\phi^2$ was used. The above expression is then similarly discretised as

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{v}_a}{\mathrm{d}t} &= -\sum_b m_b \frac{P_b}{\rho_b^2} \nabla_a W_{ab}(h) - \frac{P_a}{\rho_a^2} \sum_b m_b \nabla_a W_{ab}(h) \\ &= -\sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2}\right) \nabla_a W_{ab}(h) \end{aligned} \tag{2.24}$$

Finally, using Eq 2.18 the energy equation for a particle a can be written as,

$$\begin{aligned}\frac{\mathrm{d}u_a}{\mathrm{d}t} &= -\frac{P_a}{\rho_a}(\nabla\cdot\boldsymbol{v})_a \\ &= \frac{P_a}{\rho_a^2}\frac{\mathrm{d}\rho_a}{\mathrm{d}t}\end{aligned} \tag{2.25}$$

and can be discretised into SPH form using Eq 2.21 as

$$\frac{\mathrm{d}u_a}{\mathrm{d}t} = -\frac{P_a}{\rho_a^2}\sum_b m_b(\boldsymbol{v}_a - \boldsymbol{v}_b)\cdot\nabla_a W_{ab}(h). \tag{2.26}$$

2.1.5 Errors and accuracy

The discretisation of the governing equations of SPH in §2.1.4 include errors arising from the integral (Eq 2.9), and from incomplete sampling of the smoothing kernel. While accurate to second order in h^2, they are not the equations that would naturally arise from kernel interpolation theory (Eq 2.10 – 2.13). The errors can be illustrated by considering a constant scalar field $f(\boldsymbol{r}) = 1$. For such a field, the gradient must clearly be zero, $\nabla f(\boldsymbol{r}) = 0$. However, using the basic SPH approximation for the gradient (Eq 2.11),

$$\nabla f(\boldsymbol{r}_a) = 0 \approx \sum_b^N m_b\frac{1}{\rho_b}\nabla W_{ab}(h). \tag{2.27}$$

As there are no identities that result in the the right hand side of the above equation to equal zero, it's clear that using the basic gradient estimate directly results in errors. However, using vector calculus identities, a better gradient estimate can be found,

$$\nabla(1f(\boldsymbol{r}_a)) = 1\nabla f(\boldsymbol{r}_a) + f(\boldsymbol{r}_a)\nabla 1 \tag{2.28}$$

and hence

$$\nabla f(\boldsymbol{r}_a) = \nabla(1f(\boldsymbol{r}_a)) - f(\boldsymbol{r}_a)\nabla 1. \tag{2.29}$$

Since the SPH approximation for $\nabla 1 \neq 0$, we can define a different estimate for $\nabla f(\boldsymbol{r})$ as

$$\begin{aligned}\nabla f(\boldsymbol{r}_a) &= \sum_b \frac{m_b}{\rho_b} f(\boldsymbol{r}_b)\nabla W_{ab}(h) - f(\boldsymbol{r}_a)\sum_b \frac{m_b}{\rho_b}\nabla W_{ab}(h) \\ &= \sum_b m_b\frac{f(\boldsymbol{r}_b) - f(\boldsymbol{r}_a)}{\rho_b}\nabla W_{ab}(h),\end{aligned} \tag{2.30}$$

which has the advantage of being exactly zero for a constant $f(\boldsymbol{r})$. More generally, the improved gradient estimate comes in two forms,

$$\nabla A = \frac{1}{\phi}[\nabla(\phi A) - A\nabla\phi] \approx \sum_b \frac{m_b}{\rho_b}\frac{\phi_b}{\phi_a}(A_b - A_a)\nabla_a W_{ab}, \tag{2.31}$$

and

$$\nabla A = \phi\left[\frac{A}{\phi^2}\nabla\phi + \nabla\left(\frac{A}{\phi}\right)\right] \approx \sum_b \frac{m_b}{\rho_b}\left(\frac{\phi_b}{\phi_a}A_a + \frac{\phi_a}{\phi_b}A_b\right)\nabla_a W_{ab}, \tag{2.32}$$

where ϕ is any arbitrary, differential scalar field defined on the particles. Using $\phi = 1$, Eq 2.31 reduces to Eq 2.30, which is exact. Identities such as above can be used to get the equations of SPH into more useful forms. For example, using $\phi = \rho$ in the gradient estimate given by Eq 2.32 for $(\nabla P)/\rho$ results in the momentum equation (Eq 2.24). Although Eq 2.32 no longer vanishes for constant A, it does have the advantage of being pairwise symmetric, ensuring conservation of momentum. Similar expressions can be derived for the other vector derivatives.

2.1.6 Variable smoothing length

While a constant smoothing length is useful for deriving the equations of motion in SPH form, it is not ideal for simulations with a large dynamic density range. If the smoothing length was held constant such that the high density regions were accurately modelled, the small smoothing length required would result in the low density regions being poorly resolved with few particles within a couple of smoothing lengths. Similarly, if the constant smoothing length was chosen to model low density regions accurately, then the simulation becomes extremely expensive due to particles in high density regions having extremely large number of neighbours. Hence, a smoothing length for each individual particle that varies with density (Hernquist & Katz, 1989; Benz, 1990), as in Eq 2.2, is a natural choice for more accurate modelling of both high and low density regions as it ensures the number of neighbours for all particles are roughly the same. However, the temporal and spatial gradients of the smoothing length are no longer zero and cannot be ignored. Hence, the equations in §2.1.4 must be written in a form that self-consistently accounts for the variation in smoothing length, ensuring conservation of energy.

Continuity equation

Taking the full time derivative for the density (Eq 2.1),

$$\frac{\mathrm{d}\rho_a}{\mathrm{d}t} = \sum_b m_b \frac{\mathrm{d}W_{ab}(h_a)}{\mathrm{d}t}. \tag{2.33}$$

In 3 dimensions, $W_{ab}(h_a) = C_{\mathrm{norm}}\, f(q)/h_a^3$. Substituting into the above gives,

$$\begin{aligned} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} &= \sum_b m_b\, C_{\mathrm{norm}} \left[\frac{h_a^3 \frac{\mathrm{d}f(q)}{\mathrm{d}t} - f(q) \frac{\mathrm{d}h_a^3}{\mathrm{d}t}}{h_a^6} \right] \\ &= \sum_b m_b \frac{C_{\mathrm{norm}}}{h_a^6} \left[h_a^3\, f'(q) \frac{\mathrm{d}q}{\mathrm{d}t} - f(q) \frac{\mathrm{d}h_a^3}{\mathrm{d}\rho_a} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} \right]. \end{aligned} \tag{2.34}$$

Recalling that $q = |\boldsymbol{r}_{ab}|/h_a$, which gives

$$\begin{aligned} \frac{\mathrm{d}q}{\mathrm{d}t} &= \frac{h_a \frac{\mathrm{d}|\boldsymbol{r}_{ab}|}{\mathrm{d}t} - |\boldsymbol{r}_{ab}| \frac{\mathrm{d}h_a}{\mathrm{d}t}}{h_a^2} \\ &= \frac{\hat{\boldsymbol{r}}_{ab} \cdot \boldsymbol{v}_{ab}}{h_a} - \frac{q}{h_a} \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} \frac{\mathrm{d}\rho_a}{\mathrm{d}t}, \end{aligned} \tag{2.35}$$

where $\boldsymbol{v}_{ab} \equiv \boldsymbol{v}_a - \boldsymbol{v}_b$. Substituting the above back into Eq 2.34 gives

$$\begin{aligned} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} &= \sum_b m_b \frac{C_{\mathrm{norm}}}{h_a^4} f'(q)\; \hat{\boldsymbol{r}}_{ab} \cdot \boldsymbol{v}_{ab} - \sum_b m_b \frac{C_{\mathrm{norm}}}{h_a^6} \left[f'(q)\, h_a^3 \frac{q}{h_a} \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} + f(q) \frac{\mathrm{d}h_a^3}{\mathrm{d}\rho_a} \right] \frac{\mathrm{d}\rho_a}{\mathrm{d}t} \\ &= \sum_b m_b\, \boldsymbol{v}_{ab} \cdot \nabla_a W_{ab}(h_a) - \sum_b m_b \frac{C_{\mathrm{norm}}}{h_a^6} \left[h_a^2\, q f'(q) \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} + 3h_a^2\, f(q) \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} \right] \frac{\mathrm{d}\rho_a}{\mathrm{d}t} \\ &= \sum_b m_b\, \boldsymbol{v}_{ab} \cdot \nabla_a W_{ab}(h_a) - \sum_b m_b \frac{C_{\mathrm{norm}}}{h_a^4} \left[q f'(q) + 3 f(q) \right] \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} \\ &= \sum_b m_b\, \boldsymbol{v}_{ab} \cdot \nabla_a W_{ab}(h_a) + \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} \frac{\mathrm{d}\rho_a}{\mathrm{d}t} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a}, \end{aligned} \tag{2.36}$$

where the following identities have been used (Price, 2012)

$$\nabla_a W_{ab}(h_a) = \frac{C_{\mathrm{norm}}}{h_a^4} f'(q)\; \hat{\boldsymbol{r}}_{ab}, \tag{2.37}$$

and

$$\frac{\partial W_{ab}(h_a)}{\partial h_a} = -\frac{C_{\mathrm{norm}}}{h_a^4} \left[3 f(q) + q f'(q) \right]. \tag{2.38}$$

Finally, grouping like terms in Eq 2.36 gives the discretised form of the continuity equation as

$$\frac{\mathrm{d}\rho_a}{\mathrm{d}t} = \frac{1}{\Omega_a} \sum_b m_b \boldsymbol{v}_{ab} \cdot \nabla_a W_{ab}(h_a), \tag{2.39}$$

where the Ω_a is a convenient term for variable smoothing lengths relating to the temporal gradient of the smoothing length, and is given by

$$\begin{aligned} \Omega_a &= 1 - \frac{\mathrm{d}h_a}{\mathrm{d}\rho_a} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a} \\ &= 1 + \frac{h_a}{3\rho_a} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a}, \end{aligned} \tag{2.40}$$

using Eq 2.2 for $\mathrm{d}h_a/\mathrm{d}\rho_a$. For a constant smoothing length $\Omega_a = 1$ giving the standard expression for the continuity equation in Eq 2.21.

Momentum equation

The Lagrangian of a fluid, defined as the difference total kinetic energy and total internal energy, is given by (Morrison, 1998)

$$\mathcal{L}(\boldsymbol{r}, \boldsymbol{v}) = \int_V \left(\frac{1}{2} \rho \boldsymbol{v} \cdot \boldsymbol{v} - \rho u(\rho, s) \right) \mathrm{d}^3\boldsymbol{r}, \tag{2.41}$$

which is discretised to a set of particles of mass $m = \rho(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r}$ as

$$\mathcal{L}(\boldsymbol{r}, \boldsymbol{v}) = \sum_b m_b \left(\frac{1}{2} \boldsymbol{v}_b \cdot \boldsymbol{v}_b - u_b(\rho_b, s_b) \right). \tag{2.42}$$

The momentum equation is derived from the equations of motion through the Euler-Lagrange equations, written for particle a as

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}_a} \right) - \frac{\partial \mathcal{L}}{\partial \boldsymbol{r}_a} = 0 \tag{2.43}$$

Now, all that is needed is to determine the derivatives as a function of the particle velocity and position coordinates. The first term in Eq 2.43 is found by differentiating the Lagrangian with respect to the velocity as

$$\begin{aligned} \frac{\partial \mathcal{L}_a}{\partial \boldsymbol{v}_a} &= \frac{\partial}{\partial \boldsymbol{v}_a} \sum_b m_b \left(\frac{1}{2} \boldsymbol{v}_b \cdot \boldsymbol{v}_b - u_b(\rho_b, s_b) \right) \\ &= m_a \boldsymbol{v}_a, \end{aligned} \tag{2.44}$$

since u does not depend on the velocities and because the velocities are independent, resulting in the differential becoming zero unless $b = a$. Similarly, the second term in Eq 2.43 is given by

$$\begin{aligned}\frac{\partial \mathcal{L}_a}{\partial \boldsymbol{r}_a} &= -\frac{\partial}{\partial \boldsymbol{r}_a}\sum_b m_b u_b(\rho_b, s_b)\\ &= -\sum_b m_b \frac{\partial u_b}{\partial \rho_b}\frac{\partial \rho_b}{\partial \boldsymbol{r}_a},\end{aligned} \tag{2.45}$$

where the assumption of a dissipationless fluid means the entropy is constant, and hence only the partial derivative with respect to the density is needed. From the first law of thermodynamics we have,

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V, \tag{2.46}$$

where, $\delta Q \equiv T\mathrm{d}S$ is the energy added to the system as heat (per unit volume), and $\delta W \equiv P\mathrm{d}V$ is the work done by expansion and compression of the fluid. As before, the volume is estimated as $V = m/\rho$, and thus the change in volume is $\mathrm{d}V = -m/\rho^2\mathrm{d}\rho$. Dividing by mass gives Eq 2.46 in quantities per unit mass as

$$\mathrm{d}u = T\mathrm{d}s + \frac{P}{\rho^2}\mathrm{d}\rho. \tag{2.47}$$

Hence, as entropy is constant, we have

$$\frac{\partial u_b}{\partial \rho_b} = \frac{P}{\rho^2}, \tag{2.48}$$

giving

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{r}_a} = -\sum_b m_b \frac{P_b}{\rho_b^2}\frac{\partial \rho_b}{\partial \boldsymbol{r}_a}. \tag{2.49}$$

From Eq 2.1, the gradient of the density at particle b with respect to the coordinates of particle a is

$$\begin{aligned}\frac{\partial \rho_b}{\partial \boldsymbol{r}_a} &= \sum_c m_c \frac{\partial W_{bc}(h_b)}{\partial \boldsymbol{r}_a}\\ &= \sum_c m_c \left[\frac{\partial W_{bc}}{\partial r_{bc}}\frac{\partial r_{bc}}{\partial \boldsymbol{r}_a} + \frac{\partial W_{bc}(h_b)}{\partial h_b}\frac{\partial h_b}{\partial \boldsymbol{r}_a}\right],\end{aligned} \tag{2.50}$$

where $r_{bc} = |\boldsymbol{r}_{bc}|$, and using the fact that the kernel is spherically symmetric. Directly differentiating r_{bc} with respect to $\boldsymbol{r}_a$ is only non-zero when b or c is equal to

a,

$$\frac{\partial r_{bc}}{\partial \boldsymbol{r}_a} = (\delta_{ba} - \delta_{ca})\hat{\boldsymbol{r}}_{bc}. \tag{2.51}$$

Substituting this back into the previous equation gives

$$\begin{aligned} \frac{\partial \rho_b}{\partial \boldsymbol{r}_a} &= \sum_c m_c \left[\frac{\partial W_{bc}(h_b)}{\partial r_{bc}}(\delta_{ba} - \delta_{ca})\hat{\boldsymbol{r}}_{bc} + \frac{\partial W_{bc}(h_b)}{\partial h_b}\frac{\mathrm{d}h_b}{\mathrm{d}\rho_b}\frac{\partial \rho_b}{\partial \boldsymbol{r}_a}\right] \\ &= \sum_c m_c \left[\nabla_b W_{bc}(h_b)(\delta_{ba} - \delta_{ca}) + \frac{\partial W_{bc}(h_b)}{\partial h_b}\frac{\mathrm{d}h_b}{\mathrm{d}\rho_b}\frac{\partial \rho_b}{\partial \boldsymbol{r}_a}\right]. \end{aligned} \tag{2.52}$$

Finally grouping like terms we find that

$$\frac{\partial \rho_b}{\partial \boldsymbol{r}_a} = \nabla_a \rho_b = \frac{1}{\Omega_b}\sum_c m_c \nabla_b W_{bc}(h_b)(\delta_{ba} - \delta_{ca}) \tag{2.53}$$

where Ω_b, the term for the spatial gradient of the smoothing length, is identical to that for the temporal gradient. Now, Eq 2.49 can written as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \boldsymbol{r}_a} &= -\sum_b m_b \frac{P_b}{\rho_b^2}\frac{1}{\Omega_b}\sum_c m_c \nabla_b W_{bc}(h_b)(\delta_{ba} - \delta_{ca}) \\ &= -m_a \frac{P_a}{\rho_a^2}\frac{1}{\Omega_a}\sum_c m_c \nabla_a W_{ac}(h_a) + \sum_b m_b \frac{P_b}{\rho_b^2}\frac{1}{\Omega_b} m_a \nabla_b W_{ba}(h_b) \\ &= -m_a \sum_b m_b \left[\frac{P_a}{\rho_a^2 \Omega_a}\nabla_a W_{ab}(h_a) + \frac{P_b}{\rho_b^2 \Omega_b}\nabla_a W_{ab}(h_b)\right] \end{aligned} \tag{2.54}$$

where the final expression is obtained by changing the summation index of the first term from c to b, and by noting that the gradient of the kernel is antisymmetric, $\nabla_a W_{ab} = -\nabla_b W_{ba}$. Finally, substituting Eq 2.44 and 2.54 back into Eq 2.43, and dividing by the common factor m_a gives the discretised form of the momentum equation as

$$\frac{\mathrm{d}\boldsymbol{v}_a}{\mathrm{d}t} = -\sum_b m_b \left[\frac{P_a}{\rho_a^2 \Omega_a}\nabla_a W_{ab}(h_a) + \frac{P_b}{\rho_b^2 \Omega_b}\nabla_a W_{ab}(h_b)\right]. \tag{2.55}$$

For a constant smoothing length $h_a = h_b \equiv h$ the above expression reduces to Eq 2.24, since $\Omega_a = \Omega_b = 1$, and $W_{ab}(h_a) = W_{ab}(h_b) \equiv W_{ab}(h)$.

Energy equation

All that is required to derive the energy equation 2.20 is the time derivative of the density, which has already been derived in Eq 2.39. Thus, the discretised form of the energy equation accounting for a variable smoothing length is

$$\frac{\mathrm{d}u_a}{\mathrm{d}t} = \frac{P_a}{\rho_a^2\Omega_a}\sum_b m_b \boldsymbol{v}_{ab}\cdot\nabla_a W_{ab}(h_a). \tag{2.56}$$

Conservation properties

As shown above, the equations that govern SPH can be derived from just the density and the Lagrangian. The time, translational, and rotational invariance of the Lagrangian gives conservation of total energy, total linear momentum, and total angular momentum, respectively. Thus allowing the SPH algorithm to be fully conservative in the absence of dissipation terms.

2.2 Phantom

We now turn to methods and implementations that are specific to our SPH code of choice, PHANTOM. We choose PHANTOM because because it has been extensively used for studies of protoplanetary discs (for e.g., but not limited to Lodato & Price, 2010; Price et al., 2018b; Hall et al., 2020; Nealon et al., 2020a; Terry et al., 2022), and accretion discs in general.

In particular, PHANTOM has tried and tested methods for sink particles, external torques, thermodynamics and time-stepping that are advantageous for our study. We introduce each of these in turn. A more detailed introduction to PHANTOM can be found in Price et al. (2018a).

2.2.1 Self-gravity

Self-gravity between particles can be be included with an additional acceleration term in the equation of motion

$$\boldsymbol{a}_{\mathrm{selfgrav}} = -\nabla\Phi, \tag{2.57}$$

which is done by solving Poisson's equation,

$$\nabla^2\Phi = 4\pi G\rho(\boldsymbol{r}) \tag{2.58}$$

where Φ is the gravitational potential and ρ represents a continuous fluid density. Care must be taken when solving Eq 2.58 to prevent unphysically large accelerations when particles get too close to each other. The problem arises due to SPH particles representing interpolation points in a continuous density field rather than individual point masses. Thus, to avoid this issue gravitational softening is used for SPH particles that are too close to each other. The proximity at which gravitational softening is applied is defined later (see equations 2.66 and 2.67). The gravitational potential with softening is given by Price & Monaghan (2007)

$$\Phi_a = -G \sum_b m_b \, \phi(|\boldsymbol{r}_a - \boldsymbol{r}_b|, \, \epsilon_a), \tag{2.59}$$

where ϕ is the gravitational softening kernel, and ϵ_a is the softening length, which is typically set to the smoothing length. As the smoothing length is variable, an average of the kernel has to be taken to satisfy energy conservation,

$$\Phi_a = -G \sum_b m_b \left[\frac{\phi_{ab}(\epsilon_a) + \phi_{ab}(\epsilon_b)}{2} \right]. \tag{2.60}$$

The gravitational softening kernel is related to the density kernel by

$$W(r, \epsilon) = \frac{1}{4\pi r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) \tag{2.61}$$

where $r \equiv |\boldsymbol{r} - \boldsymbol{r}'|$. Integrating Eq 2.61 once gives the gravitational force softening kernel, ϕ'. Integrating a second time gives the gravitational potential softening kernel, ϕ. The resulting additional term in the equations of motion due to self-gravity can now be written as

$$\begin{aligned} \boldsymbol{a}^a_{\mathrm{selfgrav}} &= -\nabla \Phi \\ &= -G \sum_b m_b \left[\frac{\phi'_{ab}(\epsilon_a) + \phi'_{ab}(\epsilon_b)}{2} \right] \hat{\boldsymbol{r}}_{ab} \\ &\quad - \frac{G}{2} \sum_b m_b \left[\frac{\zeta_a}{\Omega^\epsilon_a} \nabla_a W_{ab}(\epsilon_a) + \frac{\zeta_b}{\Omega^\epsilon_b} \nabla_a W_{ab}(\epsilon_b) \right], \end{aligned} \tag{2.62}$$

where $\Omega^\epsilon_{a,b}$ and $\zeta_{a,b}$ are correction terms required to satisfy energy conservation for a variable smoothing length and are defined by

$$\Omega^\epsilon_{a,b} = 1 - \frac{\partial \epsilon_a}{\partial \rho_a} \sum_c m_c \frac{\partial W_{ac}(\epsilon_a)}{\partial \epsilon_a} \tag{2.63}$$

and

$$\zeta_{a,b} = \frac{\partial \epsilon_a}{\partial \rho_a} \sum_c m_c \frac{\partial \phi_{ac}(\epsilon_a)}{\partial \epsilon_a}. \tag{2.64}$$

In PHANTOM, the calculation of Eq 2.62 is split into 'short-range' and 'long-range' contributions,

$$\boldsymbol{a}^a_{\text{selfgrav}} = \boldsymbol{a}^a_{\text{short}} + \boldsymbol{a}^a_{\text{long}}. \tag{2.65}$$

PHANTOM uses a kd-tree to efficiently include self-gravity. The tree is built by recursively splitting the particles at the centre of mass and bisecting the longest axis at each level, grouping particles into cells. This process is repeated until every cell contains less than $N_{\text{min}} = 10$ particles, referred to as a 'leaf node' (see Price et al., 2018a). The short-range acceleration is evaluated by direct summation of Eq 2.62 for all particles in the leaf node n over the trial neighbour list obtained by node-node pairs, $n - m$, which satisfy either the tree opening criteria,

$$\theta^2 < \left(\frac{s_m}{r_{nm}}\right)^2, \tag{2.66}$$

or if the smoothing spheres of the nodes intersect,

$$r^2_{nm} < [s_n + s_m + \max(R_{\text{kern}} h^n_{\text{max}},\, R_{\text{kern}} h^m_{\text{max}})]^2, \tag{2.67}$$

where r_{nm} is the separation between the nodes, s is the node size which is the smallest radius about the center of mass containing every particle in the node, $R_{\text{kern}} = 2$ is the dimensionless cutoff radius of the smoothing kernel, and h_{max} is the maximum smoothing length of the particles in the node. The node size is defined as the minimum radius about the center of mass which contains all the particles. For any particles pairs that are outside each other's softening radius, gravitational softening is unnecessary. In this case, the short range interaction reduces to

$$\boldsymbol{a}^a_{\text{short}} = -G \sum_b m_b \frac{\boldsymbol{r}_a - \boldsymbol{r}_b}{|\boldsymbol{r}_a - \boldsymbol{r}_b|^3}, \tag{2.68}$$

avoiding unnecessary calculations of the softening kernel. If neither of Eq 2.66 or 2.67 is satisfied, the particles are considered to be distant from each other where $\zeta = 0$, while the first term in Eq 2.62 reduces to $1/r^2$.

2.2.2 Artificial viscosity

So far, the equations introduced that govern SPH do not include any dissipation terms. In a real fluid, as particles approach a shock front discontinuities can arise in the fluid properties (such as pressure, density, and energy) due to bulk kinetic energy being instantaneously converted to heat. In numerical schemes, discontinuities are problematic as the flow becomes multivalued. They are fundamentally unresolved as a discontinuity will always occur on a length-scale smaller than the resolution length of the particles. Hence to accurately capture shocks in any numerical method including SPH, artificial viscosity is essential to smooth shocks, allowing it to be resolved and preventing multivalued flows. Additionally, artificial viscosity in SPH can damp particle motion and prevent particle interpenetration. Artificial viscosity introduces two more dissipative terms, Π_{shock} and Λ_{shock} in the momentum and energy equation respectively, to give a correct increase in entropy at the shock front.

The artificial viscosity term q_a^{AV} is defined using two artificial viscosity parameters α^{AV} and β^{AV} and is described as

$$q_a^{\text{AV}} = \begin{cases} -\frac{1}{2}\rho_a v_{\text{sig},\,a}\, \boldsymbol{v}_{ab}\cdot\hat{\boldsymbol{r}}_{ab}, & \boldsymbol{v}_{ab}\cdot\hat{\boldsymbol{r}}_{ab} < 0, \\ 0, & \boldsymbol{v}_{ab}\cdot\hat{\boldsymbol{r}}_{ab} \geq 0, \end{cases} \tag{2.69}$$

where $v_{\text{sig},\,a}$ is the signal velocity, the maximum velocity at which particles can exchange information and is given by

$$v_{\text{sig},\,a} = \alpha_a^{\text{AV}} c_{s,\,a} + \beta^{\text{AV}} |\boldsymbol{v}_{ab}\cdot\hat{\boldsymbol{r}}_{ab}|. \tag{2.70}$$

In the early days of SPH, artificial viscosity was modelled using a fixed $\alpha^{\text{AV}} \sim 1$ and $\beta^{\text{AV}} = 2$ (see review by Monaghan, 1992). Capturing shocks using a fixed artificial viscosity is far from ideal as all shocks are treated equally regardless of their strength leading to poor modelling of strong shocks, overdamping of convergent flows, or both. To overcome this, Morris & Monaghan (1997) introduced a time dependent switch based on the local convergence of the flow which evolved α^{AV} according to

$$\frac{\mathrm{d}\alpha_a^{\text{AV}}}{\mathrm{d}t} = S_a - \frac{\alpha_a^{\text{AV}} - \alpha_{\text{min}}^{\text{AV}}}{\tau_a}. \tag{2.71}$$

Here, the first term is a source term

$$S_a = \max(-\nabla\cdot\boldsymbol{v}_a,\, 0) \tag{2.72}$$

that increases α_a^{AV} as the particle approaches a shock. The second term exponentially decays α_a^{AV} to $\alpha_{\text{min}}^{\text{AV}}$ beyond the shock over the timescale $\tau \equiv h/(\sigma_{\text{decay}} v_{\text{sig}})$, with $\sigma_{\text{decay}} = 0.1$. Typically, an $\alpha_{\text{min}}^{\text{AV}} = 0.1$ is used to correctly model shocks. Compared to the fixed artificial viscosity approach, this method reduces α_a^{AV} by an order of magnitude to $\alpha_{\text{min}}^{\text{AV}}$ away from shocks where it's not needed, whilst still modelling shocks properly. Despite the improvements of the Morris & Monaghan (1997) method, a background $\alpha_{\text{min}}^{\text{AV}} = 0.1$ was still required to remove the ripples bedind the shock arising due to the switch on of α_a^{AV} being slightly delayed with respect to the shock front.

To eliminate artificial viscosity completely ($\alpha_a^{\mathrm{AV}} \rightarrow 0$) except within the shocked regions, Cullen & Dehnen (2010) introduced a new shock indicator using the time derivative of the velocity divergence, $\mathrm{d}(\nabla \cdot \boldsymbol{v})/\mathrm{d}t \equiv \dot{\nabla} \cdot \boldsymbol{v}$. An advantage to using $\dot{\nabla} \cdot \boldsymbol{v}$ as a shock indicator is that it distinguishes between pre- and post-shock regions, as well as velocity discontinuities (shocks) from convergent flows, which is evident from differentiating the continuity equation,

$$-\dot{\nabla} \cdot \boldsymbol{v} = \frac{\mathrm{d}^2 \ln \rho}{\mathrm{d}t^2}. \tag{2.73}$$

When $\dot{\nabla} \cdot \boldsymbol{v} < 0$, it indicates density is increasing non-linearly and a steepening of flow convergence, typical of pre-shock regions. Conversely, $\dot{\nabla} \cdot \boldsymbol{v} > 0$ for the post-shock regions. Thus, the shock indicator need only to consider the negative values and takes an analogous form to Eq 2.72,

$$A_a = \xi_a \max\left[-\frac{\mathrm{d}}{\mathrm{d}t}(\nabla \cdot \boldsymbol{v}_a),\, 0\right] \tag{2.74}$$

where

$$\xi = \frac{|\nabla \cdot \boldsymbol{v}|^2}{|\nabla \cdot \boldsymbol{v}|^2 + |\nabla \times \boldsymbol{v}|^2} \tag{2.75}$$

is a modified version of the Balsara (1995) viscosity limiter for shear flows. This is used to set α^{AV} according to

$$\alpha^{\mathrm{AV}}_{\mathrm{loc},\,a} = \min\left(\frac{10 h_a^2 A_a}{c_{s,\,a}^2},\, \alpha^{\mathrm{AV}}_{\max}\right), \tag{2.76}$$

where $\alpha^{\mathrm{AV}}_{\max} = 1$. If $\alpha^{\mathrm{AV}}_{\mathrm{loc},\,a} > \alpha_a^{\mathrm{AV}}$, then $\alpha_a^{\mathrm{AV}} = \alpha^{\mathrm{AV}}_{\mathrm{loc},\,a}$. Otherwise, α_a^{AV} is evolved according to

$$\frac{\mathrm{d}\alpha_a^{\mathrm{AV}}}{\mathrm{d}t} = -\frac{\alpha_a^{\mathrm{AV}} - \alpha^{\mathrm{AV}}_{\mathrm{loc},\,a}}{\tau_a}, \tag{2.77}$$

where τ is defined identically as in the Morris & Monaghan (1997) method. In PHANTOM, the second artificial viscosity parameter, $\beta^{\mathrm{AV}} = 2$ by default. It is used to prevent interpenetration of approaching particles and is independent of α^{AV} (Price & Federrath, 2010; Meru & Bate, 2012).

Finally, the additional dissipative term in the momentum equation used to capture shocks correctly can be written as

$$\Pi^a_{\text{shock}} = -\sum_b m_b \left[\frac{q^{\text{AV}}_a}{\rho^2_a \Omega_a} \nabla_a W_{ab}(h_a) + \frac{q^{\text{AV}}_b}{\rho^2_b \Omega_b} \nabla_b W_{ab}(h_b) \right]. \tag{2.78}$$

A similar dissipative term is added to the energy equation and is given by

$$\begin{aligned} \Lambda^a_{\text{shock}} = &- \frac{1}{\Omega_a} \sum_b m_b v_{\text{sig},\,a} \frac{1}{2} (\boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab})^2 F_{ab}(h_a) \\ &+ \sum_b m_b \alpha_u v^u_{\text{sig}} (u_a - u_b) \frac{1}{2} \left[\frac{F_{ab}(h_a)}{\Omega_a \rho_a} + \frac{F_{ab}(h_b)}{\Omega_b \rho_b} \right], \end{aligned} \tag{2.79}$$

where the first term provides viscous shock heating, and the second term is the thermal conductivity necessary for dissipating the jump in energy at a shock. For the thermal conductivity, $\alpha_u = 1$, and it uses a different signal velocity from Eq 2.70. As all the problems in this thesis include self-gravity, the signal velocity is given by

$$v^u_{\text{sig}} = |\boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab}|. \tag{2.80}$$

All simulations in this thesis use the Cullen & Dehnen (2010) method for artificial viscosity with $\alpha^{\text{AV}}_a \in [0, 1]$ and $\beta^{\text{AV}} = 2$ to correctly capture shocks without excess dissipation away from shocks.

2.2.3 Sink particles

Sink particles were introduced by Bate et al. (1995) to follow the evolution of star forming molecular clouds after fragmentation. Once a fragment forms, calculations grind to a halt due to the extremely small timesteps required to accurately model particles in this high density region. To circumvent this issue, a sink particle is used to represent all the SPH particles within this high density region. The only properties of a sink particle are its mass, and spatial and velocity coordinates. Thus, it only interacts with other particles gravitationally.

The acceleration on any particular sink particle i of mass M_i is given by

$$\frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = -\sum_{j=1}^{N_{\text{sink}}} G M_j \phi'_{ij}(\epsilon)\, \hat{\boldsymbol{r}}_{ij} - \sum_{b=1}^{N_{\text{part}}} G m_b \phi'_{ib}(\epsilon_{ib})\, \hat{\boldsymbol{r}}_{ib} \tag{2.81}$$

where ϕ' is the gravitational force softening kernel defined in §2.2.1, N_{part} and N_{sink} are the total number of gas and sink particles respectively. The sink-gas softening length is defined as $\epsilon_{ib} \equiv \max(\epsilon, \epsilon_b)$, the maximum of the softening length for the sink particles ϵ, and the gas particle ϵ_b. The acceleration on a particular gas particle from a sink particle is given by

$$\boldsymbol{a}^a_{\text{sink}-\text{gas}} = -\sum_{j=1}^{N_\text{sink}} GM_j \phi'_{aj}(\epsilon_{aj})\,\hat{\boldsymbol{r}}_{aj}. \tag{2.82}$$

If the softening length of the sink particles is set to zero, as is the default behaviour, then sink-gas interactions are not softened. In which case, the above expression reduces to

$$\boldsymbol{a}^a_{\text{sink}-\text{gas}} = -\sum_{j=1}^{N_\text{sink}} \frac{GM_j}{|\boldsymbol{r}_a - \boldsymbol{r}_j|^3}\,\boldsymbol{r}_{aj}. \tag{2.83}$$

For a sink particle interacting with other sink particles, interactions are unsoftened by default, giving the acceleration on a sink particle from other sink particles as

$$\boldsymbol{a}^i_{\text{sink}-\text{sink}} = -\sum_{j=1}^{N_\text{sink}} \frac{GM_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3}\,\boldsymbol{r}_{ij}. \tag{2.84}$$

Accretion onto sink particles

Sink particles are able to accrete gas particles under certain circumstances. In Phantom, gas particles are accreted if they approach too close to a sink radius or if they are close and bound to the sink. The numerical implementation is as follows: the criteria for accretion depends on two parameters, the accretion radius of the sink particle r_acc, and f_acc which determines when further checks are required, where $0 \leq f_\text{acc} \leq 1$. If the gas particle is within $f_\text{acc} r_\text{acc}$ of the sink particle, then it gets accreted without any further checks. If the particle is instead within $f_\text{acc} r_\text{acc} \leq r \leq r_\text{acc}$, then a particle is only accreted if:

1. Its specific angular momentum is less than that of a Keplerian orbit at r_acc,

$$|\boldsymbol{L}_{ai}| < |\boldsymbol{L}_\text{acc}|, \tag{2.85}$$

   where the relative specific angular momentum of the gas-sink pair, $\boldsymbol{L}_{ai}$ is given by

$$|\boldsymbol{L}_{ai}|^2 \equiv |\boldsymbol{r}_{ai} \times \boldsymbol{v}_{ai}|^2 = r_{ai}^2 v_{ai}^2 - (\boldsymbol{r}_{ai} \cdot \boldsymbol{v}_{ai})^2, \tag{2.86}$$

   while $|\boldsymbol{L}_\text{acc}| = r_\text{acc}^2 \Omega_\text{acc}$ and $\Omega_\text{acc} = \sqrt{GM_i / r_{ai}^3}$ are the specific angular momentum and Keplerian angular speed at r_acc, respectively, and r_{ai} and v_{ai} are the relative position and velocity, respectively.

2. It is gravitationally bound to the sink particle, which is when the total energy

is negative,

$$e = \frac{v_{ai}^2}{2} - \frac{GM_i}{r_{ai}} < 0 \tag{2.87}$$

3. It has the smallest value of e when compared with all the other sink particles, i.e., it is most gravitationally bound to this specific sink particle.

When a particle is accreted onto a sink, the position, velocity, acceleration, spin angular momentum, and mass of the sink particle is updated to

$$\begin{aligned}
\boldsymbol{r}_i &= \frac{\boldsymbol{r}_a m_a + \boldsymbol{r}_i M_i}{M_i + m_a} \\
\boldsymbol{v}_i &= \frac{\boldsymbol{v}_a m_a + \boldsymbol{v}_i M_i}{M_i + m_a} \\
\boldsymbol{a}_i &= \frac{\boldsymbol{a}_a m_a + \boldsymbol{a}_i M_i}{M_i + m_a} \\
\boldsymbol{S}_i &= \boldsymbol{S}_i + \frac{m_a M_i}{M_i + m_a} \left[(\boldsymbol{r}_a - \boldsymbol{r}_i) \times (\boldsymbol{v}_a - \boldsymbol{v}_i) \right] \\
M_i &= M_i + M_a.
\end{aligned} \tag{2.88}$$

During this process, total mass, total linear and angular momentum is conserved, however total energy is not. In PHANTOM, accreted particles are tagged to be ignored in all subsequent calculations by setting their smoothing length to be negative.

2.2.4 External forces

An alternative method to model the central star instead of a sink particle is to use an external force describing a point mass, M at the origin, where the gravitational potential and acceleration is given by

$$\Phi_a = \frac{GM}{r_a}; \qquad \boldsymbol{a}_{\rm ext} = -\nabla \Phi_a = -\frac{GM}{|\boldsymbol{r}_a|^3} \boldsymbol{r}_a. \tag{2.89}$$

As with a sink particle, gas particles that are within $R_{\rm acc}$ from the origin can be accreted. However, unlike a sink particle, the properties of the accreted particles are not added to the central mass when an external potential is used. While this formally breaks the conservation laws, the error can be considered negligible as long the mass of the central potential is the dominant mass in the system. An external potential can be favourable to use instead of sink particles in certain scenarios as in the warped GI discs in Chapter 6. In those simulations, the artificial introduction of the warp in a disc with non-axisymmetric structure due to gravitational instabilities would influence the sink particle, which would in turn influence the disc making analysis more complicated. However, using an external potential avoids such complications.

2.2.5 Thermodynamics

In §2.2.2 and 2.1.6, the heating effects due to viscous heating, and work done by expansion and compression of the fluid were included as additional terms in the energy equation. Similarly, cooling the disc requires an additional cooling term $\Lambda_{\rm cool}$. The simulations in this thesis use the straightforward 'β-cooling' prescription of Gammie (2001),

$$\Lambda_{\rm cool} = \frac{\rho u}{t_{\rm cool}}. \tag{2.90}$$

Where in the standard implementation, the cooling timescale is proportional to the orbital time by a constant factor of β,

$$t_{\rm cool} = \beta \Omega^{-1}, \tag{2.91}$$

where Ω is defined using the spherical distance r from the central star. However, β can also be defined to vary with radial location. The constant β implementation is used in Chapters 3 and 6, while the variable β implementation is introduced and used in Chapters 3 – 5.

2.2.6 Timestepping

To evolve the fluid properties PHANTOM uses a Leapfrog integrator in the 'Velocity Verlet' form. It is time-reversible, and conserves both total linear and angular momentum. The particle positions and velocities are updated from t^n to t^{n+1} based on the previous half time-step,

$$\boldsymbol{v}^{n+\frac{1}{2}} = \boldsymbol{v}^n + \frac{1}{2}\Delta t \boldsymbol{a}^n, \tag{2.92}$$

$$\boldsymbol{r}^{n+1} = \boldsymbol{r}^n + \frac{1}{2}\Delta t \boldsymbol{v}^{n+\frac{1}{2}}, \tag{2.93}$$

$$\boldsymbol{a}^{n+1} = \boldsymbol{a}\left(\boldsymbol{r}^{n+1}\right), \tag{2.94}$$

$$\boldsymbol{v}^{n+1} = \boldsymbol{v}^{n+\frac{1}{2}} + \frac{1}{2}\Delta t \boldsymbol{a}^{n+1}, \tag{2.95}$$

where $\Delta t \equiv t^{n+1} - t^n$. However, a complication arises from the accelerations due to the shock capturing dissipation terms which depend on the velocity, making Eq 2.95 implicit. To resolve this, a first order prediction of the velocity is calculated as follows,

$$\boldsymbol{v}^{n+\frac{1}{2}} = \boldsymbol{v}^n + \frac{1}{2}\Delta t \boldsymbol{a}^n, \tag{2.96}$$

$$\boldsymbol{r}^{n+1} = \boldsymbol{r}^n + \frac{1}{2}\Delta t \boldsymbol{v}^{n+\frac{1}{2}}, \tag{2.97}$$

$$\boldsymbol{v}^* = \boldsymbol{v}^{n+\frac{1}{2}} + \frac{1}{2}\Delta t \boldsymbol{a}^n, \tag{2.98}$$

$$\boldsymbol{a}^{n+1} = \boldsymbol{a}\left(\boldsymbol{r}^{n+1}, \boldsymbol{v}^*\right), \tag{2.99}$$

$$\boldsymbol{v}^{n+1} = \boldsymbol{v}^* + \frac{1}{2}\Delta t \left[\boldsymbol{a}^{n+1} - \boldsymbol{a}^n\right] \tag{2.100}$$

Timestep constraints

To determine how frequently the fluid properties have to be calculated for numerical accuracy and stability, there are various constraints that have to be considered. In addition to being accurate, the code must also be numerically stable, which is achieved by constraining the timestep to be smaller than the maximum stable timestep. For a particular particle a, this is given by (Monaghan & Lattanzio, 1986; Monaghan, 1992)

$$\Delta t_{C,a} = C_{\rm Cour} \frac{h_a}{v_{{\rm sig},a}^{\rm dt}}, \tag{2.101}$$

where $C_{\rm Cour} = 0.3$, and $v_{{\rm sig},a}^{\rm dt}$ is taken as the maximum of Eq 2.70 over particle neighbours assuming $\alpha^{\rm AV} = \max(\alpha^{\rm AV}, 1)$. For particles that suffer large accelerations, a constraint is applied based on the accelerations (the 'force condition') where

$$\Delta t_{f,a} = C_{\rm force} \sqrt{\frac{h_a}{|\boldsymbol{a}_a|}}, \tag{2.102}$$

with $C_{\rm force} = 0.25$. Separate constraints are also applied for external forces,

$$\Delta t_{\rm ext} = C_{\rm force} \sqrt{\frac{h}{|\boldsymbol{a}_{{\rm ext},a}|}}, \tag{2.103}$$

and for accelerations due to sink-gas interactions,

$$\Delta t_{{\rm sink-gas},a} = C_{\rm force} \sqrt{\frac{h_a}{|\boldsymbol{a}_{{\rm sink-gas},a}|}}. \tag{2.104}$$

For external forces defined as $\Phi \to 0$ as $r \to 0$, such as for a point mass in §2.2.4, an additional constraint is

$$\Delta t_{\Phi,\,a} = C_{\rm force}\, \eta_\Phi \sqrt{\frac{|\Phi_a|}{|\nabla \Phi|_a^2}}, \tag{2.105}$$

where $\eta = 0.05$. A final constraint is applied to ensure the cooling does does not decouple from the other equations, and to avoid cooling instabilities from numerical errors,

$$\Delta t_{{\rm cool},\,a} = C_{\rm cool} \left| \frac{u}{({\rm d}u/{\rm d}t)_{\rm cool}} \right|, \tag{2.106}$$

where $C_{\rm cool} = 0.3$ (Glover & Mac Low, 2007). In practice, for a given particle all applicable constraints from above are checked and the smallest time-step from these is adopted. All the constants stated above are standard values (see Price et al., 2018a).

2.2.7 Summary of the complete governing equations of SPH

A summary of the equations that control the movement of particles and the evolution of their properties as implemented in PHANTOM are described here. The equation describing the density of a particle at any position, and its smoothing length,

$$\rho_a = \sum_b m_b W_{ab}(h_a) \qquad \text{and} \qquad h_a = h_{\rm fact} \left(\frac{m_a}{\rho_a}\right)^{1/3}. \tag{2.107}$$

The continuity equation,

$$\frac{{\rm d}\rho_a}{{\rm d}t} = \frac{1}{\Omega_a} \sum_b m_b (\boldsymbol{v}_a - \boldsymbol{v}_b) \cdot \nabla_a W_{ab}(h_a). \tag{2.108}$$

The equation of motion,

$$\begin{aligned} \frac{{\rm d}\boldsymbol{v}_a}{{\rm d}t} = & -\sum_b m_b \left[\frac{P_a}{\rho_a^2 \Omega_a} \nabla_a W_{ab}(h_a) + \frac{P_b}{\rho_b^2 \Omega_b} \nabla_a W_{ab}(h_b) \right] \\ & + \Pi_{\rm shock} + \boldsymbol{a}_{\rm ext}(\boldsymbol{r}, t) + \boldsymbol{a}_{\rm sink-gas} + \boldsymbol{a}_{\rm selfgrav}, \end{aligned} \tag{2.109}$$

and the energy equation,

$$\frac{{\rm d}u_a}{{\rm d}t} = \frac{P_a}{\rho_a^2 \Omega_a} \sum_b m_b \boldsymbol{v}_{ab} \cdot \nabla_a W_{ab}(h_a) + \Lambda_{\rm shock} - \frac{\Lambda_{\rm cool}}{\rho}. \tag{2.110}$$

The accelerations from self-gravity (§2.2.1), sink particles (§2.2.3), and external forces (§2.2.4) are given by $\boldsymbol{a}_{\rm selfgrav}$, $\boldsymbol{a}_{\rm sink-gas}$, and $\boldsymbol{a}_{\rm ext}$, respectively. The dissipative terms introduced by artificial viscosity (§2.2.2) to give a correct increase in entropy at the shock front are given by $\Pi_{\rm shock}$ and $\Lambda_{\rm shock}$. Lastly, $\Lambda_{\rm cool}$ is a cooling term (§2.2.5).

2.3 Post-processing SPH simulations

While hydrodynamic simulations enable the testing of various theories, they cannot be compared to observations directly. To make proper comparisons to understand whether the theory matches reality, radiative transfer modelling is necessary to produce synthetic observations. Examples of using this approach can be found in Terry et al. (2022); Ballabio et al. (2021); Hall et al. (2020). Here we introduce the fundamentals of the radiative transfer code MCFOST (Pinte et al., 2006, 2009) that is used to generate the synthetic observations presented in this work.

2.3.1 The radiative transfer problem

In a dusty protoplanetary disc the radiative transfer equation involves solving for the (polarised) specific intensity, $\mathbf{I}_\lambda(\boldsymbol{r}, \boldsymbol{n})$ at every point $\boldsymbol{r}$, and direction $\boldsymbol{n}$, of the volume at all wavelengths λ. Assuming time independent radiation transport, randomly orientated dust particles, and adopting Stokes formalism we can write the radiative transfer equation as (e.g. Pinte et al., 2009)

$$\begin{aligned}\frac{\mathrm{d}\mathbf{I}_\lambda(\boldsymbol{r}, \boldsymbol{n})}{\mathrm{d}s} =& -\kappa_\lambda^{\rm ext}(\boldsymbol{r})\mathbf{I}_\lambda(\boldsymbol{r}, \boldsymbol{n}) + \kappa_\lambda^{\rm abs}(\boldsymbol{r})B_\lambda(T(\boldsymbol{r}))\mathbf{I}_0 \\ &+ \kappa_\lambda^{\rm sca}(\boldsymbol{r})\frac{1}{4\pi}\iint_\Omega \mathbf{S}_\lambda(\boldsymbol{r}, \boldsymbol{n}', \boldsymbol{n})\,\mathbf{I}_\lambda(\boldsymbol{r}, \boldsymbol{n}')\,\mathrm{d}\Omega',\end{aligned} \tag{2.111}$$

where $\mathbf{I}_\lambda(\boldsymbol{r}, \boldsymbol{n}) = (I, Q, U, V)$ is the Stokes vector, and s is distance along the direction of propagation. The total intensity is represented by I, the linearly polarised intensities by Q and U, and the circularly polarised intensity by V. The absorption, scattering, and extinction opacities are given by $\kappa_\lambda^{\rm abs}(\boldsymbol{r})$, $\kappa_\lambda^{\rm sca}(\boldsymbol{r})$, and $\kappa_\lambda^{\rm ext}(\boldsymbol{r}) = \kappa_\lambda^{\rm abs}(\boldsymbol{r}) + \kappa_\lambda^{\rm sca}(\boldsymbol{r})$, respectively. The 4×4 Mueller matrix, $\mathbf{S}_\lambda(\boldsymbol{r}, \boldsymbol{n}', \boldsymbol{n})$ describes how the Stokes vector changes when light is scattered from the direction $\boldsymbol{n}' \rightarrow \boldsymbol{n}$. The Planck function is $B_\lambda(T(\boldsymbol{r}))$, and unpolarised emission is represented by $\mathbf{I}_0 = (1, 0, 0, 0)$. By assuming the dust is at local thermal equilibrium, and the radiation field is the only source of energy, the temperature, $T(\boldsymbol{r})$, can be calculated

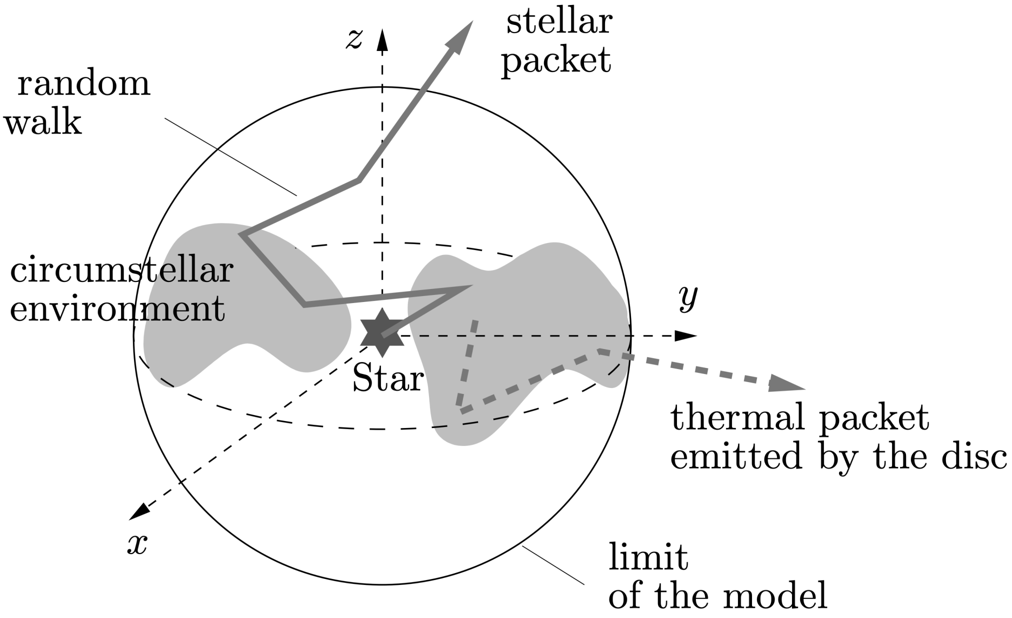


Figure 2.2: A schematic diagram of how photon packets propagate (image from Pinte, 2015). Photons undergo a random walk due to absorption, scattering, and re-emission events until it has escaped the circumstellar environment.

by solving

$$\int_0^\infty \kappa_\lambda^{\mathrm{abs}}(\boldsymbol{r}) B_\lambda(T(\boldsymbol{r}))\,\mathrm{d}\lambda = \int_0^\infty \kappa_\lambda^{\mathrm{abs}}(\boldsymbol{r}) J_\lambda\,\mathrm{d}\lambda, \tag{2.112}$$

where J_λ is the specific intensity averaged over all solid angles.

2.3.2 The Monte-Carlo method

In MCFOST, the radiative transfer equations are solved numerically using a stochastic, statistical approach. The technique relies on numerous packets of energy, hereafter referred to as photon packets, emitted by the star that interact with the dust. The temperature distribution and optical properties of the medium dictate the journey of the photons through absorption, scattering, and re-emission events. The random journey of a photon is illustrated in Figure 2.2. As the interactions are statistical in nature, the results can become noisy if the number of photons used is low. This can be overcome by increasing the amount of photons used at the cost of computational expense.

2.3.3 Computing the temperature

The temperature is determined as follows. Photon packets are generated by the stellar photosphere and the dust grains. The interactions of each photon is tracked individually until it exits the protoplanetary disc. The distance a photon travels until an interaction event with a dust grain depends on the optical depth τ, and thus the probability of interaction is given by

$$P(\tau_\lambda) = e^{-\tau_\lambda}. \tag{2.113}$$

Integrating the optical depth gives the distance between two interactions l,

$$\tau_\lambda = \int_0^l \kappa^{\mathrm{ext}}(\lambda, \boldsymbol{r})\rho(\boldsymbol{r})\mathrm{d}s. \tag{2.114}$$

Once the position of interaction is found, the interaction is determined probabilistically by the local albedo to either be a scattering or absorption event. If a photon is scattered, the propagation vector is altered, but its wavelength is unchanged. If a photon is absorbed, it is immediately re-emitted at a frequency wavelength in-situ and isotropically. The new frequency is chosen by the frequency redistribution approach of Bjorkman & Wood (2001)

$$\frac{\mathrm{d}P_i}{\mathrm{d}\nu} = \frac{\kappa_\nu}{K}\left(\frac{\mathrm{d}B_\nu}{\mathrm{d}T}\right)_{T=T_i}, \tag{2.115}$$

where $\mathrm{d}P_i/\mathrm{d}\nu$ is the probability of re-emitting a photon between frequencies ν and $\nu + \mathrm{d}\nu$, and $K = \int_0^\infty \kappa_\nu(\mathrm{d}B_\nu/\mathrm{d}T)\,\mathrm{d}\nu$ is the normalisation constant. Once all the photons have been propagated, the mean intensity J_λ, can be determined as (Lucy, 1999).

$$J_\lambda = \frac{1}{4\pi V_i}\sum_\gamma \epsilon_\gamma \Delta I_\gamma. \tag{2.116}$$

where γ indicates a photon packet, ϵ_γ is the packet luminosity, V_i is the volume of the cell, and ΔI_γ is the distance the photon propagated through the cell i. The final dust temperature can then be calculated using Eq 2.112.

2.3.4 Making synthetic images

To produce an image (to compare with observations), the basic Monte Carlo method places detectors in a given direction. A major disadvantage is the low probability of a photon to reach this detector, which is compensated for by using a large number of photons to ensure enough reach the detector. However, for certain disc and wavelength configurations (like an edge-on disc in the mid infra-red), the probability of a photon reaching the observer is so small that the computational time required becomes impractical.

An alternative method is to use ray-tracing to produce the images. Only the mean specific intensity and temperature structure is required to build the image. Both of these quantities can be estimated from the Monte Carlo method. First there is an initial calculation with a multi-wavelength Monte Carlo run to get the temperature structure. Then monochromatic Monte Carlo runs estimate the specific intensity at every point in the model for a particular wavelength. Ray-tracing is then used to produce the image at the required wavelength by numerically integrating the solution to radiative transfer equation along the observer's line of sight.

2.3.5 Line radiative transfer

The method of determining the mean intensity J_ν for the gas differs to the method used for the dust. This is because for gas line emission the absorption and emission coefficients, α_ν^{ul} and j_ν^{ul}(gas), are determined from transitions between energy levels u and l with number densities n_u and n_l. The Einstein coefficients, A_{ul}, B_{lu}, and B_{ul} determine the transition probabilities for spontaneous emission, absorption, and stimulated emission respectively between two adjacent energy levels with an energy difference of $\Delta E = E_u - E_l = h\nu_0$, where h is Planck's constant. The absorption and emission coefficients are given by

$$\alpha_\nu^{ul}(\text{gas}) = \frac{h\nu_0}{4\pi}\left(n_l B_{lu} - n_u B_{ul}\right)\phi(\nu) \tag{2.117}$$

$$j_\nu^{ul}(\text{gas}) = \frac{h\nu_0}{4\pi}\, n_u A_{ul}\, \phi(\nu). \tag{2.118}$$

Here, ν_0 is the rest frequency of the transition, and $\phi(\nu)$ is a line-broadening function that depends on frequency. The level populations are determined through collisions and radiation with the equation of statistical equilibrium,

$$n_l\left[\sum_{k<l} A_{lk} + \sum_{k\neq l}\left(B_{lk}J_\nu + C_{lk}\right)\right] = \sum_{k>l} n_k A_{kl} + \sum_{k\neq l} n_k\left(B_{kl}J_\nu + C_{kl}\right) \tag{2.119}$$

where C_{kl} is the collision rates. However, this equation cannot be solved directly due to the mutual dependence of J_ν and n_l. Hence, it must be solved iteratively starting by guessing the initial level populations. Then J_ν is calculated, and statistical equilibrium is solved resulting in new level populations. This is repeated through Monte Carlo integrations until the radiation field and populations converge on a solution (see Hogerheijde & van der Tak, 2000, for more detail on solving radiative transfer and molecular excitation).

2.3.6 MCFOST

The synthetic continuum images and gas kinematics for all the work in this thesis are produced by post-processing the simulations using the 3D radiative transfer code MCFOST. It is optimised for PHANTOM as it uses the location of the SPH particles to create a Voronoi mesh where each MCFOST cell corresponds to an individual SPH particle. For gas only simulations, like the ones presented in this thesis, MCFOST generates the synthetic continuum images using the radiative transfer schemes outlined above and by assuming the dust is perfectly coupled to the gas. The gas kinematics for specific transitions are created using the velocity information from the SPH simulations along with an abundance ratio of the chosen gas species. In addition to studies such as Terry et al. (2022); Ballabio et al. (2021); Hall et al. (2020) which post-process PHANTOM simulations with MCFOST to make observational predictions and comparisons, MCFOST can also be coupled directly with PHANTOM to calculate the temperature structure on the go (see Nealon et al., 2020b).

3

Planet Migration in Self-Gravitating Discs: Survival of Planets

3.1 Introduction

During the last couple of decades the number of exoplanets discovered has increased drastically thanks to advances in technology. As of current writing 4154 exoplanets have been confirmed according to the NASA Exoplanet Archive[1], and this number will continue to increase significantly with ongoing missions such as TESS and future missions such as PLATO

Over the last few years, advances in imaging have resulted in surveys of high-resolution images of young protoplanetary discs. A large number of these discs have substructures, such as rings and gaps (ALMA Partnership et al., 2015; Andrews et al., 2016, 2018; Huang et al., 2018a; Dipierro et al., 2018; Booth & Ilee, 2020). A common explanation for their presence is planets (Dipierro et al., 2015, 2016; Long et al., 2018; Zhang et al., 2018; Dullemond et al., 2018). However, many of these discs appear to be quite young, which would require giant planets to form very early in the lifetime of the disc. Ansdell et al. (2017) also find that even in the younger discs, there it not enough material to form the core of a giant planet. This suggests that if planets are the source of gaps and rings, then planet formation must occur

---

[1] https://exoplanetarchive.ipac.caltech.edu/

at a much faster timescale. In the early stages of disc evolution, the disc is expected to be quite massive and large in size. However, there is a growing body of evidence that such discs can exist (Tobin et al., 2016b; Pérez et al., 2016).

In the GI model, giant planet formation is caused by disc fragmentation. This first involves the breakup of the massive protoplanetary disc into self-gravitating clumps at distances of tens to hundreds of AU inside $10^4 - 10^5$ years. The initial masses of protoplanets formed by these clumps are expected to be in the range of $1 M_{\rm Jup} - 10 M_{\rm Jup}$ (Boley et al., 2010). Disc fragmentation is dependent on two criteria. The first is the Toomre criterion,

$$Q = \frac{c_s \Omega}{\pi G \Sigma} \lesssim 1, \tag{3.1}$$

where G is the gravitational constant, c_s and Σ are the disc sound speed and surface mass density respectively, and Ω is the angular frequency. This criteria requires the disc to be massive enough so that self gravity plays a role in its evolution. The second criteria is the cooling timescale. This requires that disc must cool fast on a timescale comparable to the dynamical timescale to avoid clumps being disrupted.

An issue with this model is whether planets that form early on in these discs, when the discs are potentially gravitationally unstable, can survive. Previous studies such as Baruteau et al. (2011) have found that giant planets in self-gravitating discs rapidly migrate inwards with no signs of gap opening. It is much harder to open up gaps in self-gravitating region of a disc due to the high levels of turbulence. Malik et al. (2015) also found it difficult to open up gaps, even with $30 M_{\rm Jup}$ companions. However, these studies used a simple dimensionless globally constant cooling parameter β, to cool the disc, where β is defined as the ratio between the cooling time and the local orbital time. Although this method is quick as it gives a simple expression for the cooling time, it does have its issues. The main one being that as the disc evolves, spiral structure is formed throughout the disc. This behaviour is not expected from realistic discs where they should only be gravitationally unstable in the outer regions (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009).

Other studies such as Stamatellos & Inutsuka (2018); Stamatellos (2015) have found that when realistic thermodynamics are included, giant planets can open up gaps and slow their inward migration. However, in their simulations the planets were also allowed to accrete material which could aid gap opening. It is therefore unclear if their planet migration slowed down due to the mass or once they reached the gravitationally stable inner disc. The gravitational instability of a disc can be defined by Q, the Toomre parameter (Toomre, 1964)

$$Q = \frac{c_s \Omega}{\pi G \Sigma}, \tag{3.2}$$

where c_s is the sound speed of the disc, Σ is the disc surface density, Ω is the angular frequency, and G is the gravitational constant.

Cossins et al. (2009) showed that the amplitude of density fluctuations decreases at a rate inversely proportional to β. This means that in constant β models as in Baruteau et al. (2011) and Malik et al. (2015), there will be density fluctuations triggered by gravitoturbulence throughout the disc. These fluctuations can result in mass deficits around the planet and induce stochastic kicks or even Type III migration.

The motivation behind this work is to determine whether planets – irrespective of how they form – can slow down their migration in the inner gravitationally stable part of the disc. To accomplish this, we mimic the effects of more realistic thermodynamics by using the aforementioned simple cooling method but we vary it with radius such that the cooling timescale in the outer regions is much smaller than the inner regions so that the disc is only gravitationally unstable in the outer part. A larger β (and hence a longer cooling timescale) in the inner gravitationally stable region resulting in smaller density fluctuations could provide a safe haven for the planet. The main advantage of this method is to avoid having to use the more complicated and computationally expensive method of radiative transfer. We study the migration of planets in these more realistic discs to determine whether they can slow their inward migration and survive.

The outline of this Chapter is as follows. In Section 3.2, we describe how the simulations are set up and how the disc thermodynamics is treated. In Section 3.3, we present the comparison of the new implementation of the cooling parameter with the standard implementation. We compare the effect it has on the evolution of the disc, and also how planet migration is affected. In Section 3.4, we investigate the impact of numerics on the results. In Sections 5.4, we put the work done here in context of past work and observations. Finally, we make conclusions in Section 5.5.

3.2 Model

3.2.1 Numerical Method

We perform 3D hyrdrodynamic simulations using Phantom, a Smoothed Particle Hydrodynamics (SPH) code developed by Price et al. (2018a). Most of the simulations presented here are run with 2 million gas particles. To ensure the results are not affected by resolution, a subset of the simulations are also done with 1 and 4 million particles. The relative change in error is ∼20% with each step up/down in resolution. The star and planets are modelled as sink particles (Bate et al., 1995). The accretion radius of the star is set to be equal to the disc inner boundary, $R_{\rm in}$. The accretion radius of each planet is set to 0.001, which is much smaller than their Hill radius. This is to prevent the planets from accreting too much material in order to perform as close as a comparison to Baruteau et al. (2011) and Malik et al. (2015).

To model shocks, we use an artificial viscosity switch that utilises the time derivative of the velocity divergence introduced by Cullen & Dehnen (2010). The artificial viscosity parameter α_v, is varied depending on how far away the shock is. Close to the shock, it is a maximum, $\alpha_{\rm max} = 1$. Far from the shocks, it is a minimum, $\alpha_{\rm min} = 0$. The artificial viscosity coefficient β_v is set to 2.

3.2.2 Initial Conditions

We model a disc between $R_{\rm in} = 1$ and $R_{\rm out} = 25$, with disc to star mass ratio of 0.1. It should be noted that these simulations are scale free. The initial surface mass density is set as a smoothed power law and is given by

$$\Sigma = \Sigma_0 \left(\frac{R}{R_0}\right)^{-1} f_s, \tag{3.3}$$

where Σ_0 is the surface mass density at $R = R_0 = 1$ and $f_s = 1 - \sqrt{R_{\rm in}/R}$ is the factor used to smooth the surface density at the inner boundary of the disc.

The energy equation is given by

$$\frac{{\rm d}u}{{\rm d}t} = -\frac{P}{\rho}(\nabla \cdot \mathbf{v}) + \Lambda_{\rm shock} - \frac{\Lambda_{\rm cool}}{\rho} \tag{3.4}$$

where u is the specific internal energy, the first term on the RHS is the $P{\rm d}V$ work, $\Lambda_{\rm shock}$ is a heating term that is due to the artificial viscosity used to correctly deal with shock fronts, and $\Lambda_{\rm cool}$ is a cooling term to cool the disc.

We assume an adiabatic equation of state. For an ideal gas this relates the pressure P, to the density ρ, and specific internal energy by

$$P = (\gamma - 1)\rho u = \frac{c_s^2 \rho}{\gamma}, \tag{3.5}$$

where the adiabatic index is $\gamma = 5/3$ and the sound speed is $c_s = \sqrt{\gamma k_{\rm B} T/\mu m_{\rm H}}$. The mean molecular weight is assumed to be $\mu = 2.381$. The initial temperature profile is expressed using a power law as

$$T = T_0 \left(\frac{R}{R_0}\right)^{-0.5}. \tag{3.6}$$

The disc aspect ratio H/R, is set to 0.05 at $R = R_0$.

3.2.3 The cooling timescales

In this Chapter, we simply set the cooling term as

$$\Lambda_{\rm cool} = \frac{\rho u}{t_{\rm cool}} \tag{3.7}$$

where the cooling time is given by

$$t_{\rm cool} = \beta \Omega^{-1}. \tag{3.8}$$

Here, β is the cooling parameter. Under the assumption that the transfer of angular momentum driven by gravitoturbulence occurs locally, we can relate the cooling parameter β with the α viscosity parameter by (Gammie, 2001)

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma - 1)}\frac{1}{\beta}. \tag{3.9}$$

We perform two sets of simulations with different cooling implementations to model the disc thermodynamics. Firstly, we define β as

$$\beta = \beta_0 \left(\frac{R}{R_0}\right)^{-2}, \tag{3.10}$$

where $\beta_0 = 5500$. Since β varies with radius, so does α. In these discs α ranges from $\sim 10^{-4}$ in the inner regions to $\sim 10^{-2}$ in the outer regions. Secondly, to compare with previous studies, we run a set of simulations where the cooling parameter is globally constant and set to $\beta = 15$. Using equation 3.9, this is equivalent to $\alpha = 0.027$. This value and the normalisation constant in equation 3.10 is chosen to

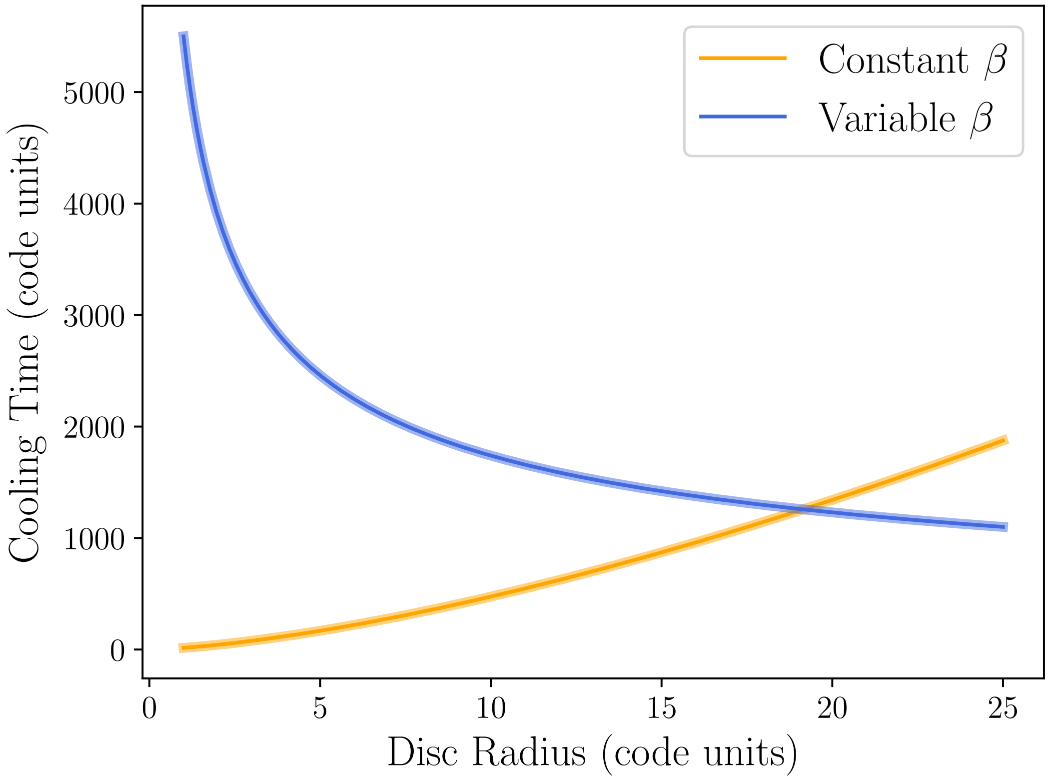


Figure 3.1: Comparison between a constant ($\beta = 15$) and a variable ($\beta = \beta_0(R/R_0)^{-2}$) cooling parameter. The cooling time in the inner regions is significantly longer using a variable β.

ensure that while the disc cools fast enough to develop spiral structure, it does not fragment into clumps. This is done so that the migration of the planet is not affected by the interaction with the clumps. The difference between these cooling timescales on disc evolution and planet migration is shown in Figure 3.1 where it can be seen that using the variable β defined in equation 3.10 results in the cooling timescale in the inner regions being significantly longer than with constant β. Hence, the inner regions now cool slowly enough to avoid becoming gravitationally unstable, while the outer regions still cool fast enough to become gravitationally unstable and form spiral structure.

3.2.4 Embedding the planets

After the disc has evolved long enough to form spiral structure (around 10 orbits at the outer edge, $R = 25$), a planet is added at $R = 20$. To explore the effect the variable cooling parameter has on planet migration, the evolution of both giant and low mass planets is followed. The giant planets have planet to star mass ratios of $q = 2.857 \times 10^{-4},\ 9.543 \times 10^{-4}$ and 9.543×10^{-3}. The low mass planets have planet to star mass ratios of $q = 3 \times 10^{-6},\ 3 \times 10^{-5}$ and 5.149×10^{-5}. This is equivalent to a $1M_{\text{Sat}}$, $1M_{\text{Jup}}$, and a $5M_{\text{Jup}}$ planet and a $1M_{\oplus}$, $10M_{\oplus}$, and a $1M_{\text{Nep}}$ planet in a $0.1M\odot$ disc around a $1M_{\odot}$ star.

3.3 Results

The simulations presented here are done in two steps. First the disc is set up as described in Section 3.2.2 and is allowed to evolve until spiral structure develops. The evolution of the disc is described in Section 3.3.1 Then after the inclusion of planets of various masses, the simulations are resumed and the planet's migration is shown in Section 3.3.2. These are done with both a constant and variable β.

3.3.1 Disc evolution

Both discs start with identical surface density and temperature profiles, so that the evolution of the disc is determined by how it cools. The difference is seen in Figures 3.2 and 3.3. In Fig 3.2, the density rendered plots of both discs is shown at two different times. The top row shows the disc early in its evolution just as the spiral structure begins to form. The disc that evolves with constant β develops spiral structure from the inside first (top left) and develops spiral structure throught the disc (bottom left), which is consistent with past studies. Whereas the disc that evolves with variable β only develops spiral structure in the outermost regions (right). Similarly, the bottom row shows how the discs differ shortly before a planet is added. This behaviour in the bottom right panel is more in line with what is expected from realistic discs, where discs are expected to be gravitationally unstable only in the outer regions (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009).

The difference between the two discs is due to the different expressions for β used. Since the variable β increases with decreasing radius, it results in the cooling time in the inner regions being much larger. This effect of the cooling time on the disc evolution is shown in Figure 3.3 by comparing the temperature and the the Toomre parameter of both discs at the same moment in time. Using equation 3.5 and the expression for the sound speed, the temperature as the disc evolves can be calculated as

$$T = \frac{\mu m_{\text{p}}}{k_{\text{B}}} (\gamma - 1)\, u. \tag{3.11}$$

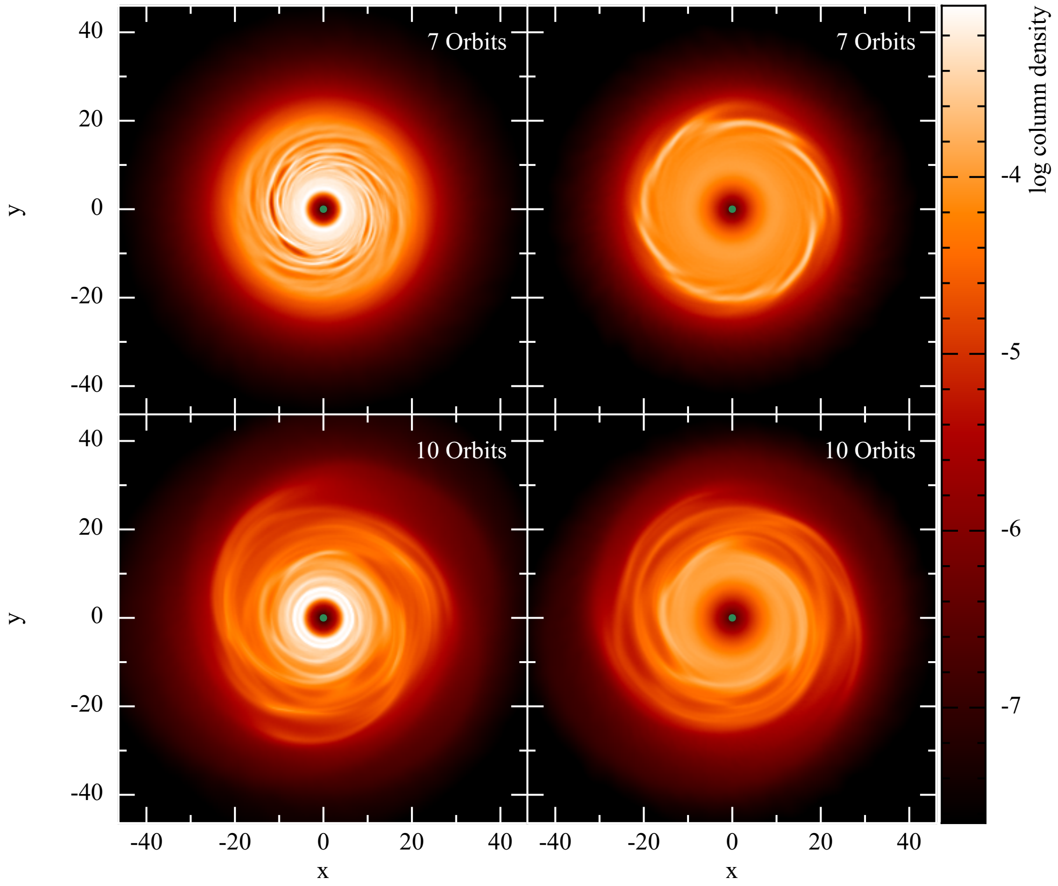
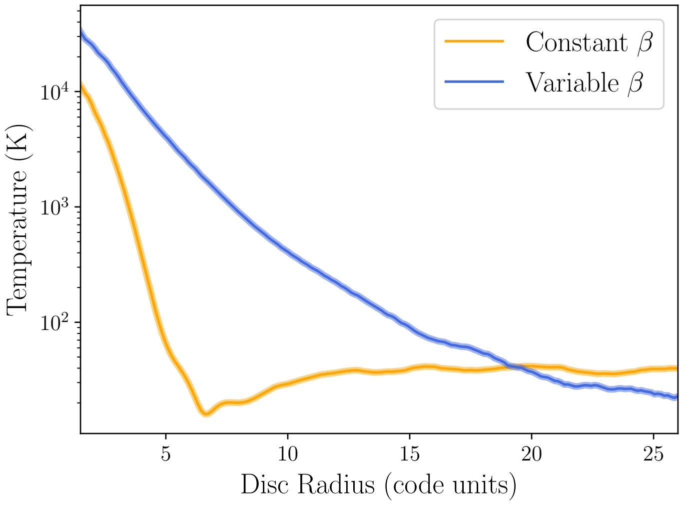


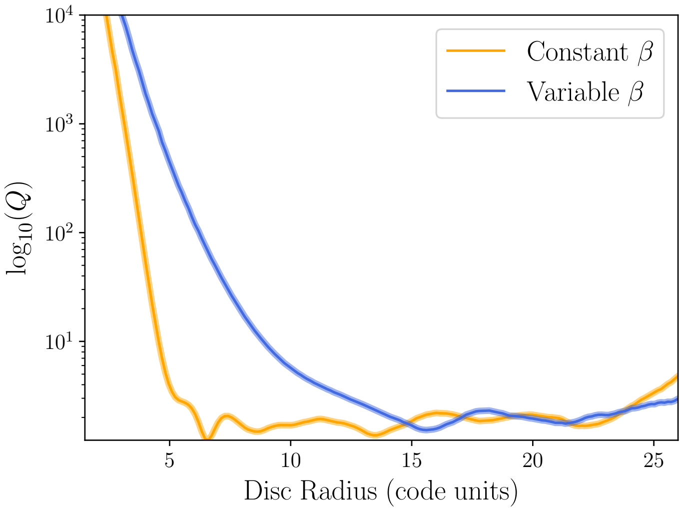
Figure 3.2: Comparison of the evolution of two discs evolved with a constant cooling parameter β given by equation 3.10 (left), and a variable β (right), after 7 and 10 orbits (at $R = 25$). The top row shows the two discs at an early stage where the spiral structure has just begun to form. The bottom row shows both discs at a later stage just before a planet is added. This figure highlights that a variable β mimics a more realistic disc where the disc is only gravitationally unstable in the outer regions.

A consequence of the variable β model is the disc cools extremely inefficiently in the innermost regions of the disc, resulting in very high temperatures as seen in Figure 3.3a. However, the simulations present here are scale free. Hence, although the temperatures are quite high, they would be more in line with expected values if the disc was scaled up to sizes of typical gravitationally unstable disc, while also scaling the initial Q profile to be identical. Furthermore, the high temperatures at $R < 10$ is not a major concern for the purposes of this Chapter since the planet reaches this part of the disc.

Figure 3.3b compares the azimuthally averaged Toomre parameter, Q. Since



(a) Comparison of the disc temperature.



(b) Comparison of the Toomre parameter, Q.

Figure 3.3: Azimuthally-averaged temperature (left) and Toomre parameter (right) for the variable and constant β simulations after 10 orbits at $R = 25$. This figure shows why the disc cooled with a variable β is only gravitationally unstable in the outer regions. Due to the significantly longer cooling times in the inner regions, the inner disc cannot cool fast enough, which results in the Toomre parameter not reaching the required values for gravitational instability. Although the temperatures are quite high, it should be noted that if these discs were scaled up to sizes of typical gravitationally unstable discs, while scaling the initial Q profile to be identical, the temperatures would be more sensible. The discs at this time are shown in the bottom row in Figure 3.2.

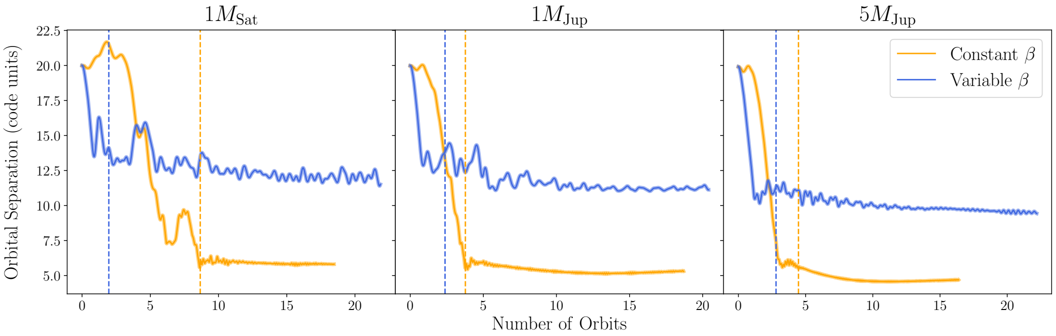


Figure 3.4: Migration tracks of a $1M_{\text{Sat}}$ (left), $1M_{\text{Jup}}$ (middle), and a $5M_{\text{Jup}}$ (right) planet in a disc cooled with a variable β (blue) and constant β (orange). The planet slows down in the inner disc modelled with a variable β, whereas with a constant β, the inward migration does not slow down and the planets reach the inner edge of the disc. The dashed lines are defined as the time when migration has slowed down (or reached the inner edge).

Q is dependent on the temperature via the sound speed, this results in $Q \gg 1.7$ in the inner regions. Hence, the conditions for spiral structure to form is only satisfied in the outer parts (Durisen et al., 2007).

3.3.2 Planet migration

Giant Planets

To investigate the effect the variable β has on planet migration, a single planet is added at $R = 20$ in the simulations in Section 3.3.1 for both the constant and variable β discs. The value of β at the initial location of the planet is roughly similar in both discs (see Figure 3.1).

To compare with previous studies that also used β-cooling to study planet migration in self-gravitating discs, we first evolve the planets in a constant β disc. From the migration tracks, shown in orange in Figure 3.4, and in the top row of Figure 3.5, it can be seen that the planets reach the inner edge of the disc in a few orbits. These compare well with previous studies such as Baruteau et al. (2011). However, in a disc cooled with a variable β, the migration tracks are no longer the same. Although both initially migrate inwards very rapidly, with a variable β, the inward migration slows down in the inner gravitationally stable part of the disc as seen in the bottom row of Figure 3.5. This is more clearly seen in Figure 3.6 which shows the surface density (top row) and a 2D map of the Toomre stability parameter Q (bottom row) of a constant β disc (left column) when a $1M_{\text{Jup}}$ planet reaches the

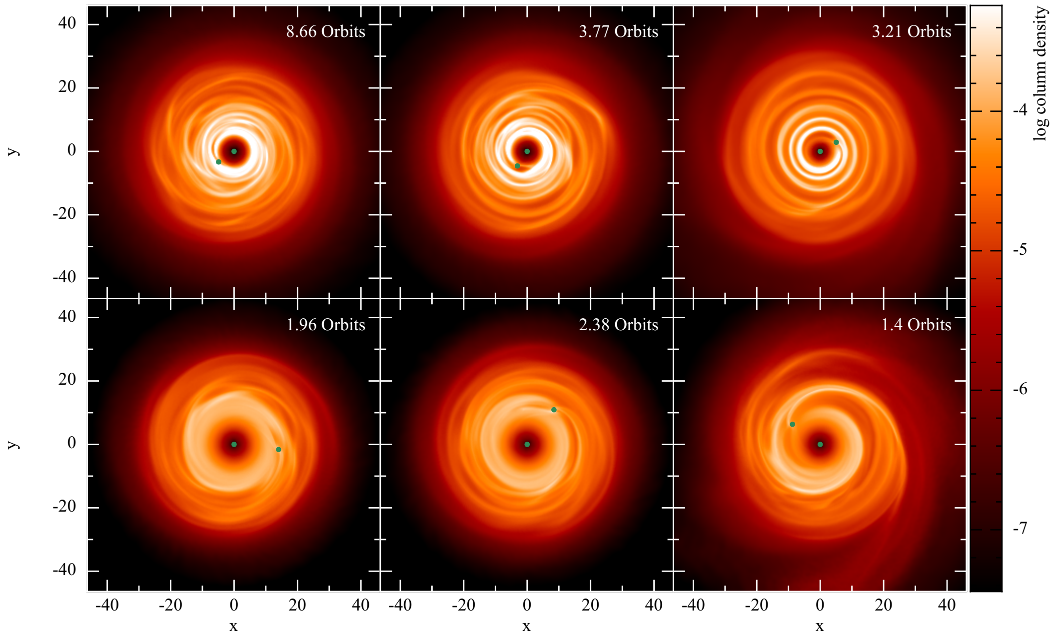


Figure 3.5: This figure shows the location of a $1M_{\rm Sat}$ (left), a $1M_{\rm Jup}$ (middle), and a $5M_{\rm Jup}$ planet (right) soon after their migration has slowed down. This is shown for a disc modelled with a constant β (top row) and a variable β (bottom row). The planets in a constant β disc only stop migrating when they reach the inner edge of the disc. Whereas, with a variable β, migration slows down before reaching the disc's inner edge. The times in each panel correspond to the dashed lines in Figure 3.4.

disc's inner edge, and of a variable β disc (right column) when the planet's migration slows down. It can be seen that unlike in a constant β disc, the inner region of a variable β disc is gravitationally stable, and it is in this region where the planets are able to slow down their inward migration.

Stochastic kicks

As seen from Figure 3.4, planet migration in self-gravitating discs is not a smooth process. The planets can experience random kicks that push them either inwards or outwards. To ensure that these random kicks due to the turbulent nature of the gravitationally unstable disc do not impact the results of the giant planet simulations, the simulations in Section 3.3.2 are repeated another 3 times. In each new run the planets are embedded at the same distance, but started 0.2 orbits earlier. Starting at a different time means that the turbulent structure around the planet is different in each simulation. Hence, it is a good test to determine whether the

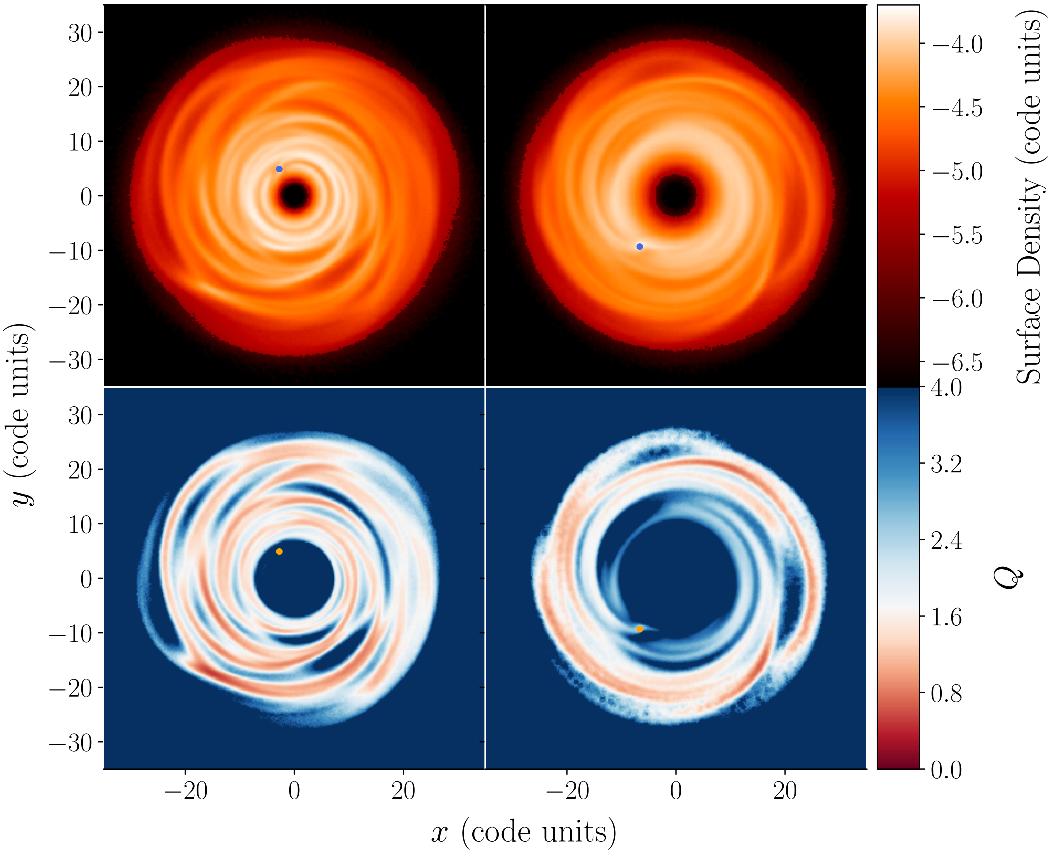


Figure 3.6: Surface density (top) and 2D Toomre parameter (bottom) plots of the disc soon after a $1M_{\text{Jup}}$ planet reaches the inner edge of a constant β (left) and slows its migration in a variable β (right) disc. The 2D Q plots show how gravitationally unstable the disc is. It can be seen that the migration of a $1M_{\text{Jup}}$ planet slows down in the gravitationally stable part of the disc.

turbulence has a significant impact on planet migration.

The results from the various restarts are shown in Figure 3.7. The migration of the giant planets are very similar to each other, regardless of when they start. In all cases, the initial rapid inward migration slows down in the gravitationally stable inner disc at roughly the same location and time. There is one exception, shown by the lightest line in Figure 3.7, where the planets do not immediately migrate inwards. This is due to random chance, where the planet is initially embedded inside one of the spiral arms of the disc. In other words, the planet is in a region of much higher density fluctuation, increasing the stochasticity. This is seen in the migration tracks for all three planets, which fluctuates much more than the other runs. Nevertheless, even in this random scenario, the planet eventually slows down

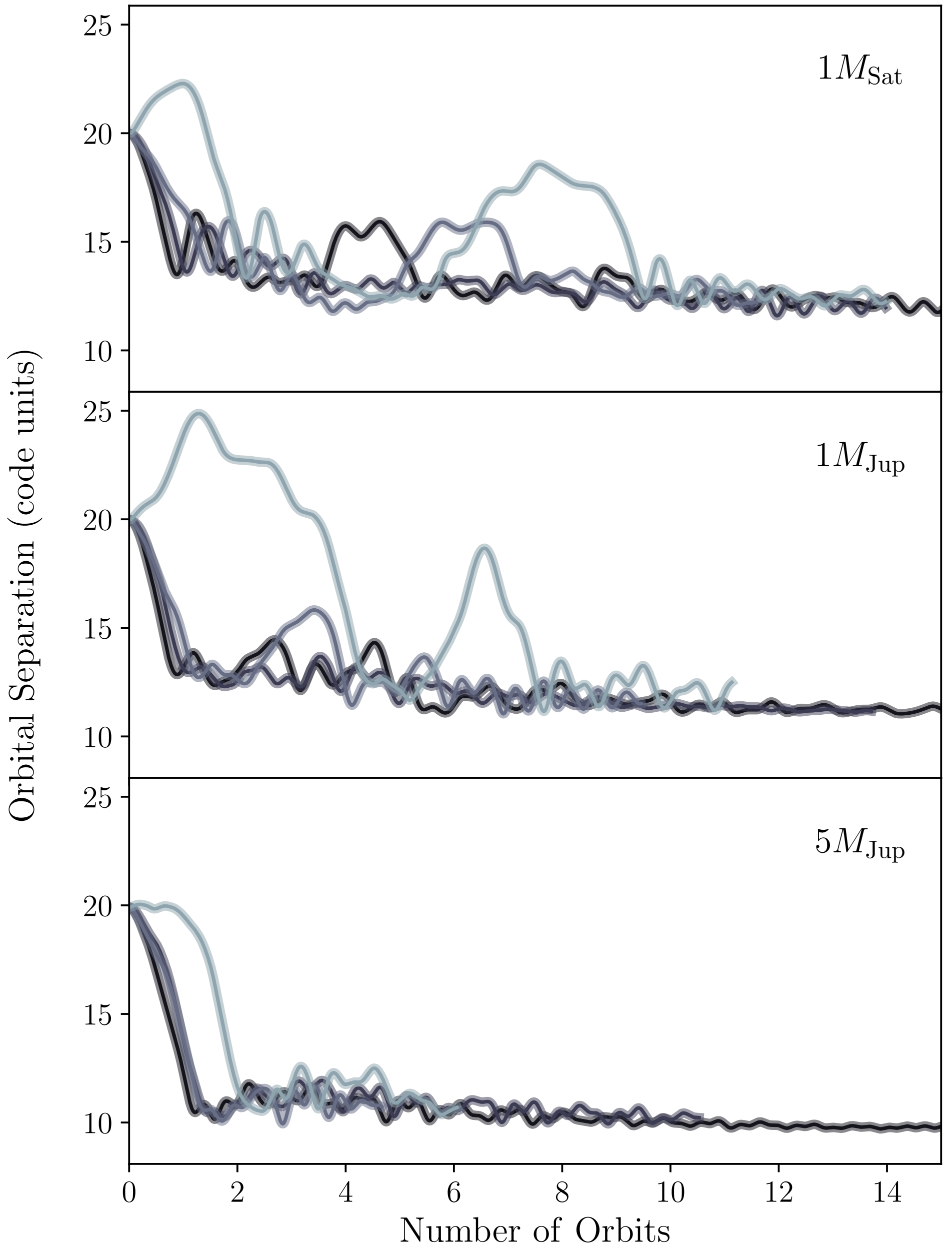


Figure 3.7: Migration tracks of a $1M_{\rm Sat}$ (top), $1M_{\rm Jup}$ (middle), and a $5M_{\rm Jup}$ (right) planet started at different times to understand the effect of different turbulent environments. Despite the random turbulent structure around the planet at each start time, the planets always end up migrating inwards rapidly and slow down when they reach the gravitationally stable inner disc.

its migration in the inner disc.

Low mass planets

As seen in Figure 3.7 (and as noted by Baruteau et al. 2011) the influence of stochastic kicks increases with decreasing planet mass. Hence, for low mass planets it is especially important to run multiple simulations as done so previously to observe the trend in the migration.

We begin by looking at the migration in the disc with a constant β, shown in the left column of Figure 3.8. Although the low mass planets migrate inwards on a longer timescale compared to the giant planets, they show no signs of slowing down. A couple of simulations of the $10M_{\oplus}$ and $1M_{\rm Nep}$ planet even reach the inner edge. It is expected that with enough computational time, all simulations of the low mass planets would reach the inner edge.

As with the giant planets, the migration of the low mass planets in the disc with a variable β (shown in the right column of Figure 3.8) compared to a constant β is quite different. In some cases, the $1M_{\oplus}$ and $10M_{\oplus}$ hardly show any inward migration. In the cases where the planets migrate inwards, the migration is significantly slowed down once the planet reaches the inner regions of the disc which is gravitationally stable. The slow down is best seen with the $1M_{\rm Nep}$ planet where after the initial rapid inward motion, the migration rate is much slower. With a variable β, there are some instances where the $1M_{\oplus}$ and $10M_{\oplus}$ planet do not migrate in at all. This is likely to be due to stochastic kicks, and will be discussed further in Section 3.5.1.

3.4 Impact of numerics

3.4.1 Resolution tests

To ensure that the above results are not affected by the resolution, we repeat the simulations with the giant planets, but with 1 and 4 million particles. However, it should be noted that the relative change in error is only $\sim$20% with each step up/down in resolution here, which may not be enough to determine convergence.

The first check for any resolution effect is done by comparing the disc evolution. Both the low and high resolution broadly compared very well with the initial simulations; all 3 simulations are only gravitationally unstable in the outer regions of the disc. A minor difference is that the gravitationally unstable regime extends further in with higher resolution. However, the higher resolution also resolves more of the inner disc. Thus the extent of the gravitational stable region in all 3 discs is similar in size, but shifted further in with increasing resolution. The planets are

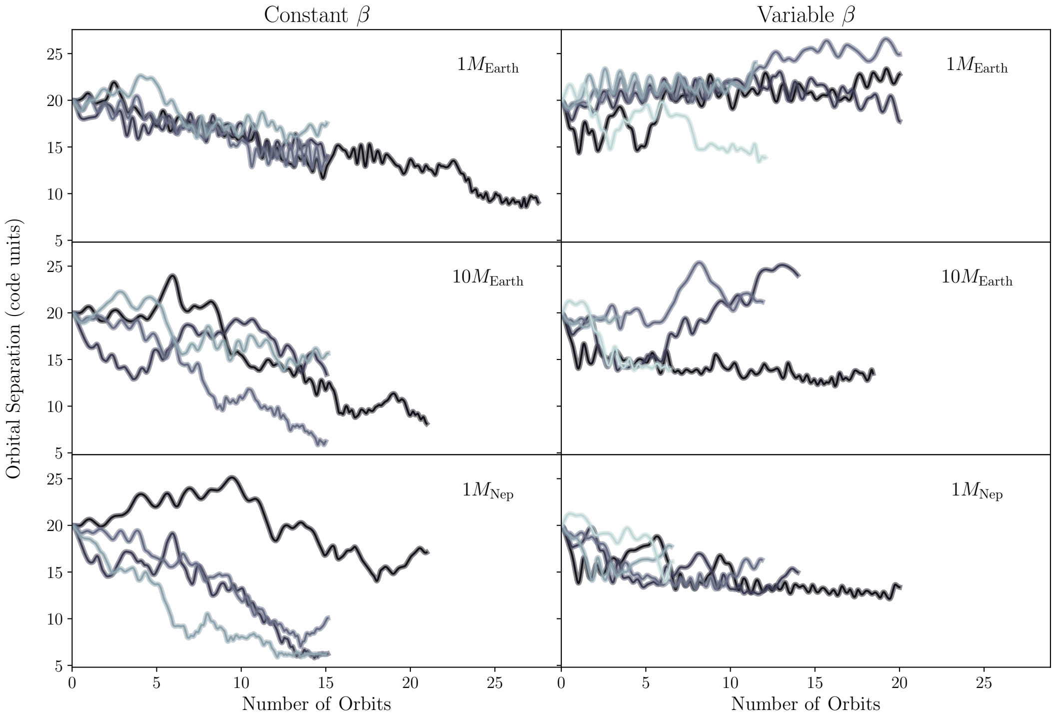


Figure 3.8: Migration tracks of the $1M_{\oplus}$, $10M_{\oplus}$, and a $1M_{\text{Nep}}$ planet started at different times for a constant (left) and a variable (right) β disc. The general trend is that even low mass planets migrate to the inner edge of a constant β disc. However, in a variable β disc, like the giant planets, they slow down in the gravitationally stable inner disc. The sporadic outward migration is due to the opposite effect described later in Figure 3.12; there are under-dense regions behind the planet resulting in a positive corotation torque.

introduced at the same time in all three discs. The migration of each planet with different resolutions is shown in Figure 3.9.

Although each planet does travel further in with increasing resolution, the conclusion that planet migration slows down in the inner gravitationally stable region of the disc remains unchanged. The further inward migration is due to the self-gravitating nature being resolved further in. This is shown in Figure 3.10 where the 2D map of the Toomre parameter Q is shown for all three resolutions just before the planet is added. In red are the regions which are gravitationally unstable. To compare how much of the inner disc is gravitationally stable, a grey circle is overlaid to represent the inner gravitationally stable disc of the 4 million particle simulation. From this it can be seen that as the resolution decreases, the gravitationally stable inner disc becomes larger. However, regardless of resolution, the migration of a

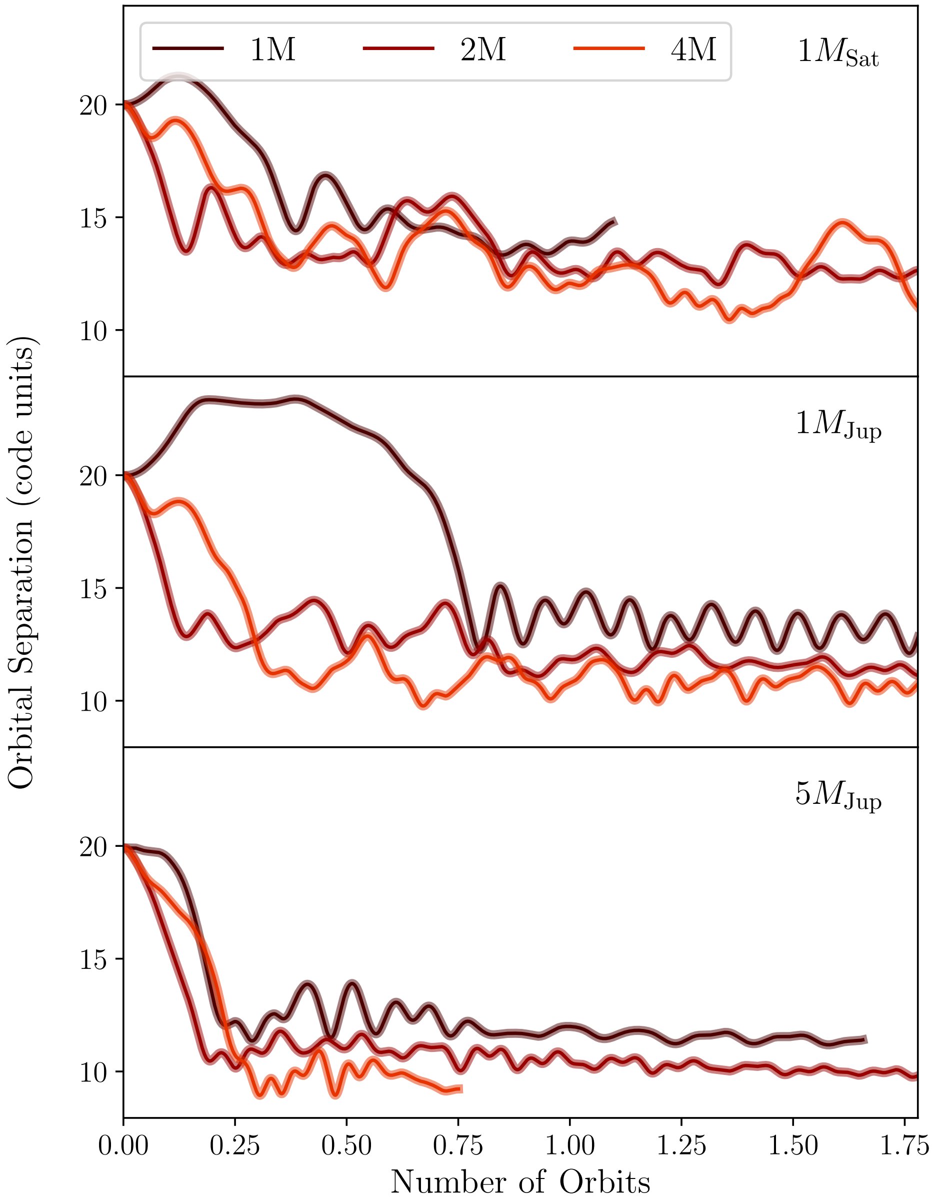


Figure 3.9: Comparison of the planet migration in a disc with variable β modelled using 1, 2, and 4 million SPH particles. The resolution increases with lighter shades. Inward migration slows down in the gravitationally stable inner disc in all cases. Although the exact location where the planet slows down varies with resolution, the relative difference is less than the expected error of the different resolutions, showing the results are approaching convergence.

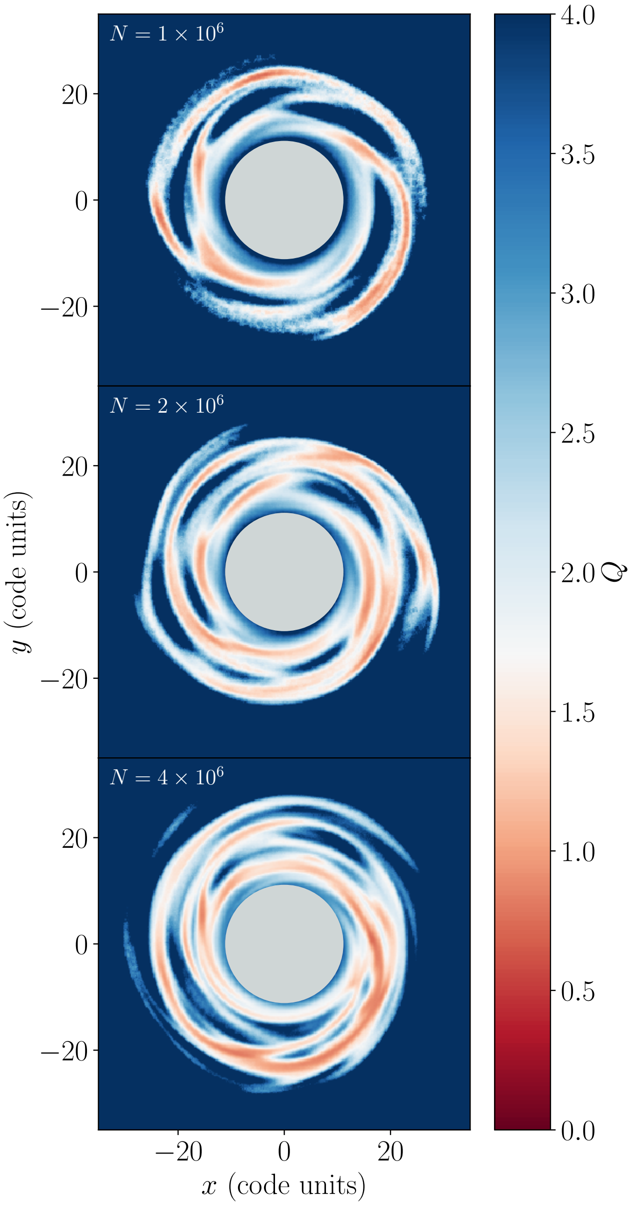


Figure 3.10: 2D maps of the Toomre parameter Q of the variable β disc just before the planets are added for a disc modelled with 1 million (top), 2 million (middle), and 4 million (bottom) particles. In red are the regions where the disc is gravitationally unstable. The size of the grey circle in each plot is the size of the gravitationally stable inner disc in the highest resolution simulation. It can be seen that in the lowest resolution simulation, more of the inner disc is gravitationally stable.

planet in a disc with variable β always slows down in the gravitationally stable part of the disc and before reaching the inner disc edge.

3.4.2 Different β profiles

Since the choice of how β should vary with radius was arbitrarily chosen, we generalise it to

$$\beta(R) = \beta_0 \left(\frac{R}{R_0}\right)^{-\delta} \tag{3.12}$$

and investigate the impact of different values of δ. We repeat the previous simulations with $\beta = 27500 \ \ (R/R_0)^{-2.5}$ and $\beta = 137500 \ \ (R/R_0)^{-3}$, where β_0 is chosen such that $\beta\,(R = 25) = 8.8$ in all three cases. We do not investigate $\delta \leqslant 1.5$ since we require the cooling time, $t_{\rm cool}$ to decrease with radius so that only the outer regions become gravitationally unstable. With $\delta = 1.5$, the cooling time is constant throughout the disc.

A $1M_{\rm Sat}$, $1M_{\rm Jup}$, and $5M_{\rm Jup}$ planet is introduced at $R = 20$ after the discs have evolved for 10 outer orbital periods. The different β profiles do not change the key result as we find that the inward migration is slowed down when the planet reaches the inner gravitationally stable part of the disc.

3.5 Discussion

3.5.1 Torque

To investigate why the rapid inward migration of the planets slows down in a variable β disc, we evaluate the tidal torque between the disc material and the planet. The total z-component of the torque, which drives the radial motion of the planet, is calculated by summing the individual torque contributions from each fluid element on the planet and is given by

$$T_z = \sum_i^N \frac{GM_{\rm p} m_i}{d_i^3}\ \mathbf{r}_{\rm p} \times \mathbf{d}_i\ , \tag{3.13}$$

where $\mathbf{d}_i$ is the separation between each fluid element and the planet, m_i is the mass of the fluid element, and $\mathbf{r}_{\rm p}$ and $M_{\rm p}$ are the position and mass of the planet respectively. To decrease numerical noise, all material inside half the Hill radius of the planet is considered to be circumplanetary material and excluded from calculations. This exclusion zone varies with the stellar mass which increases as disc material is accreted by the star, and with the mass and location of the planet as it migrates.

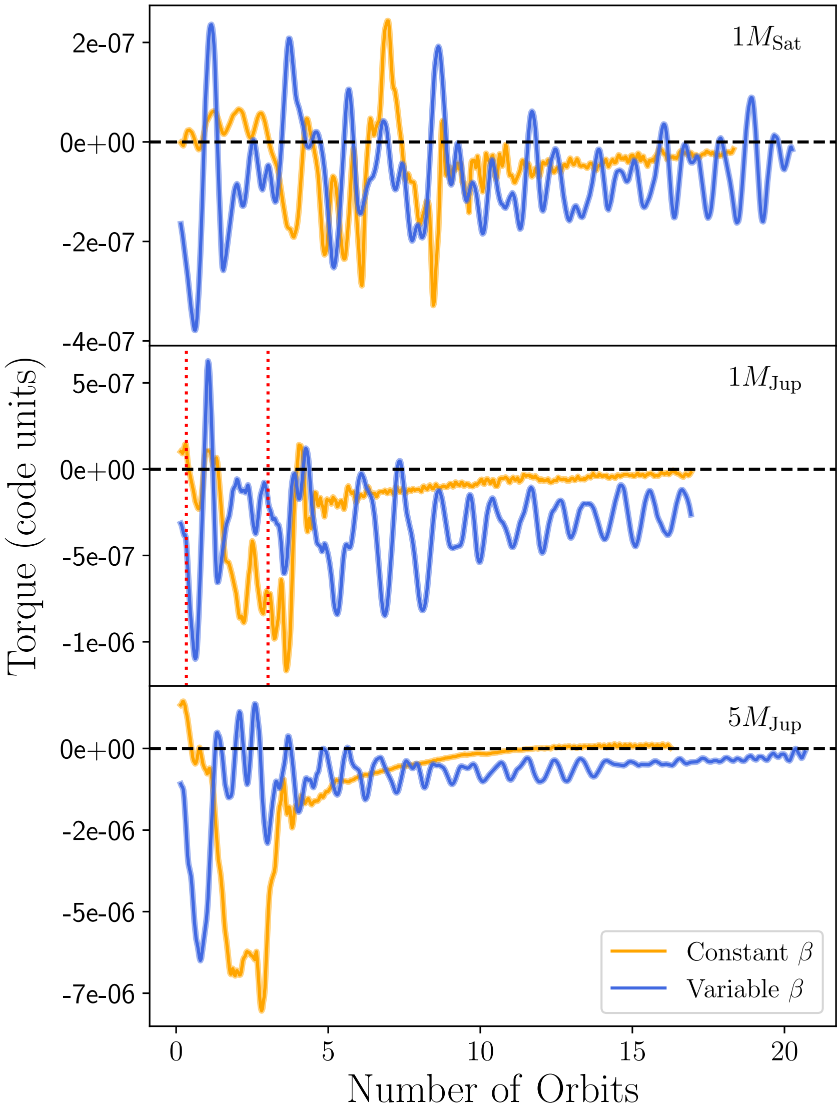


Figure 3.11: Total z-component of the torque exerted by the disc material on a $1M_{\rm Sat}$ (top), a $1M_{\rm Jup}$ (middle), and a $5M_{\rm Jup}$ planet (bottom) in a variable β (blue) and a constant β (orange) disc. In a variable β disc, the torque is most negative in the first orbit when the planet is migrating inwards rapidly. After which the torque is slightly negative, and thus the planet continues to migrate inwards but at a slower rate. In a constant β disc, the torque is just as negative during the rapid inward migration, however it stays negative for longer and only decreases to near zero when the inner edge of the disc is reached. The red dashed lines indicate the times shown in Figures 3.12 and 3.13.

The torques are also smoothed by taking the average over 0.2 orbits.

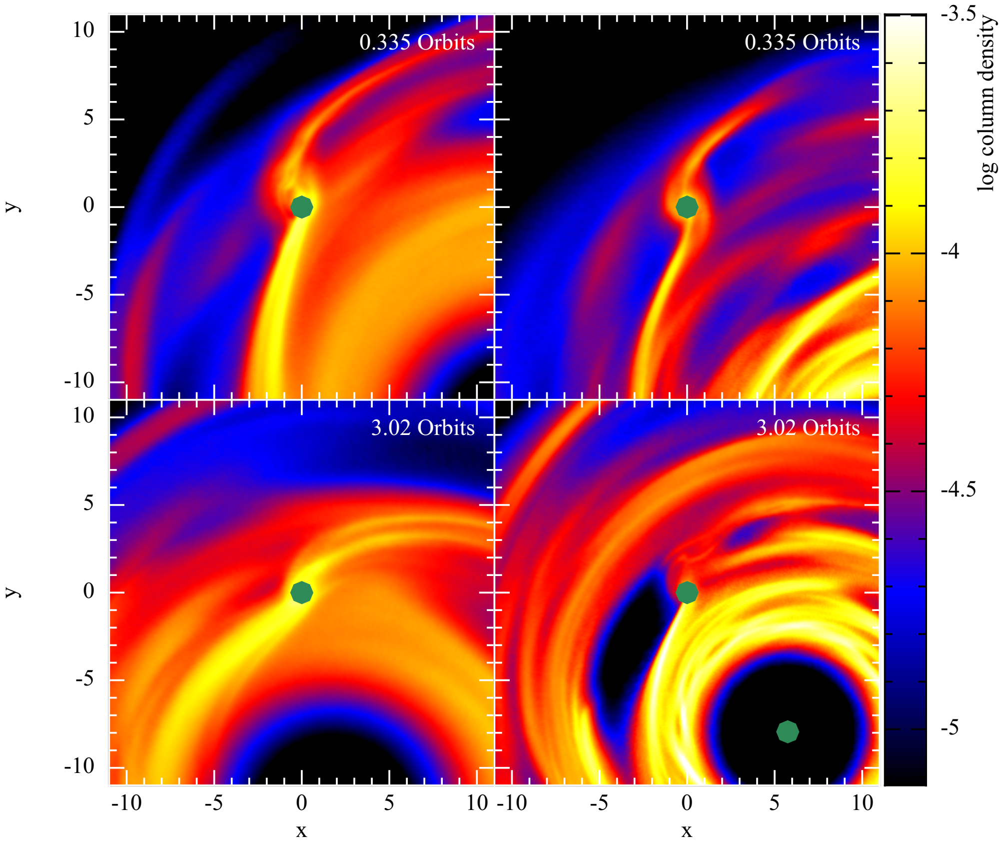


Figure 3.12: This figure shows how the surface density of the disc near a $1M_{\mathrm{Jup}}$ planet changes as it migrates in a variable β (left) and a constant β (right) disc. Top left panel is when the planet is rapidly migrating inward in a variable β disc. Bottom left is when the migration of the planet has slowed down. Top right is just before the planet starts rapidly migrating inwards in a constant β disc. Bottom right is when the planet is rapidly migrating inwards. When the surface density of the region in front and behind the planet is comparable, there is little migration (bottom left and top right). However when there is an under-dense region in front and and over-dense region behind, the planet migrates inwards due to the azimuthal density gradients (top left and bottom right, also see middle panel of Figure 3.4).

Figure 3.11 compares how T_z varies as the planet migrates in both a constant and variable β disc. The behaviour of the torque is quite different in each disc. With a constant β, the $1M_{\mathrm{Jup}}$ and $5M_{\mathrm{Jup}}$ planet experience a large negative torque until they reach the inner edge, at which point a lack of disc material results in the torque flattening to near zero very quickly. Due to random outward kicks while the $1M_{\mathrm{Sat}}$ planet is migrating inwards, the torque profile is not continuously negative.

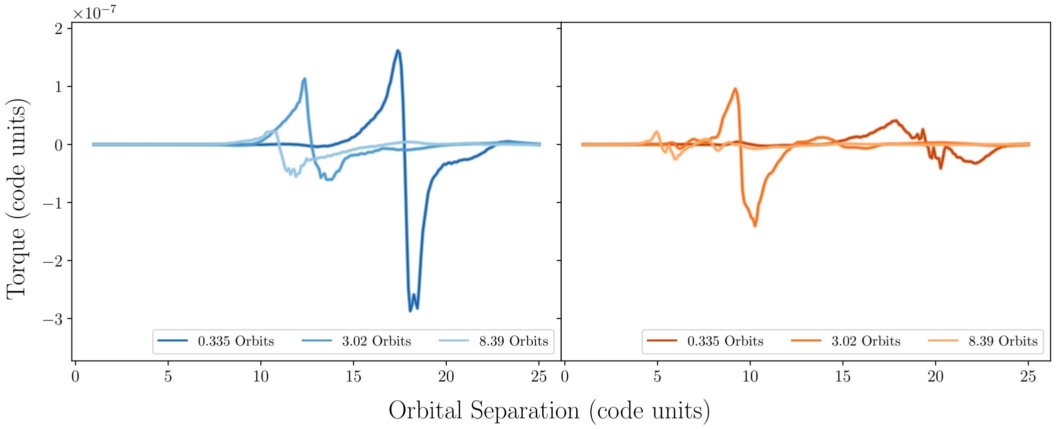


Figure 3.13: Azimuthally-averaged torque acting on a $1M_{\mathrm{Jup}}$ as it migrates in a variable (left) and constant (right) β disc. When the corotation torque is mostly negative, the planet is rapidly migrating inwards (compare times in Figure 3.5) due to an underdense region in front of the planet (see Figure 3.12).

Despite this, it is still clear that like the $1M_{\mathrm{Jup}}$ and $5M_{\mathrm{Jup}}$ planet, the magnitude of the torque only significantly decreases when it reaches the inner edge of the disc. Whereas with a variable β, even though the torque once again quickly becomes just as negative, once the planet reaches the gravitationally stable inner disc, the magnitude of the torque decreases but remains negative resulting in more inward migration at a much slower pace. Unlike with a constant β, the torque does not flatten indicating it is still migrating inwards, albeit very slowly.

The difference in the torque evolution as the planet migrates can be explained by considering the disc structure near the planet as it migrates. The corotation torque, exerted by the gas within the horseshoe region plays a significant role in the magnitude and direction of the total torque. Asymmetries in the disc structure in this region can result in either inward or outward migration. Figure 3.12 shows the disc structure at important times during the planet's migration in a variable (left) and constant β (right) disc. Comparing with Figure 3.11 shows that the torque is most negative when the region in front of the planet is underdense relative to the region behind the planet (top left and bottom right). This density contrast results in a large negative co-oribital corotation torque which drives the planet inwards. This is seen in azimuthally-averaged torques in Figure 3.13 which shows that when the planet is rapidly migrating inwards, the corotation torque is mostly negative. Whereas when there is little density contrast around the planet (top right and bottom left in Fig 3.12), the torque profile in Fig 3.13 is more symmetric resulting in slower migration.

As expected, comparing the torque on the $1M_{\rm Sat}$ to the $1M_{\rm Jup}$ and $5M_{\rm Jup}$ planets show that lower mass planets are more affected by stochastic kicks. Unlike the latter two, the torque on a $1M_{\rm Sat}$ frequently becomes positive resulting in the sporadic moments of small outward migration as seen in Figure 3.4. However, it is still clear the torque on average is negative and the planet is still migrating inwards.

However, the increased effect of stochastic kicks as planet mass decreases could explain why some of the low mass planets do not migrate inwards. Figure 3.14 shows the migration (right column) of a $1M_{\oplus}$ planet, and how T_z (left column) varies as it migrates. The first five rows are the different runs in a variable β disc. In nearly all cases, the migration is dominated by stochastic kicks as seen by the torque fluctuating about zero, resulting in small inward and outward migration. The bottom row is in a constant β disc, where it is clear that despite stochastic kicks being present, the torque on average is slightly negative resulting in inward migration.

3.5.2 Comparison with previous work

We compare our results to those of Baruteau et al. (2011) and Malik et al. (2015) who considered giant planet migration in self-gravitating discs. There are a couple of differences in this work. They perform their simulations using a 2D grid based code, whereas the simulations in this work are performed using a 3D SPH code. The initial conditions of the disc are also slightly different. Their disc surface density and temperature decreased as R^{-2} and R^{-1} respectively, whereas in this work, the disc surface density and temperature decreases as R^{-1} and $R^{-0.5}$ respectively. To ensure that the different initial conditions and codes used do not influence the results, we initially performed a set of simulations with a constant value of $\beta = 15$ to compare with Baruteau et al. (2011) and Malik et al. (2015). Despite the differences, we agree with their findings that using a constant β results in the giant planets rapidly migrating towards the inner edge of the disc.

However, the crucial development in our work is how the disc is cooled: whilst, we also utilise the β-cooling approach, we allow β to vary with radius to mimic a more realistic disc such that our discs remain gravitationally stable in the inner regions, and spiral structure only forms in the outer regions. The planets are able to slow down their rapid inward migration before reaching the inner disc edge.

Using radiative transfer, Stamatellos & Inutsuka (2018) migrate $1M_{\rm Jup}$ planets in self-gravitating discs using a 3D SPH code. The initial conditions in their discs are more similar to ours; the same initial disc mass and surface density profile. Their

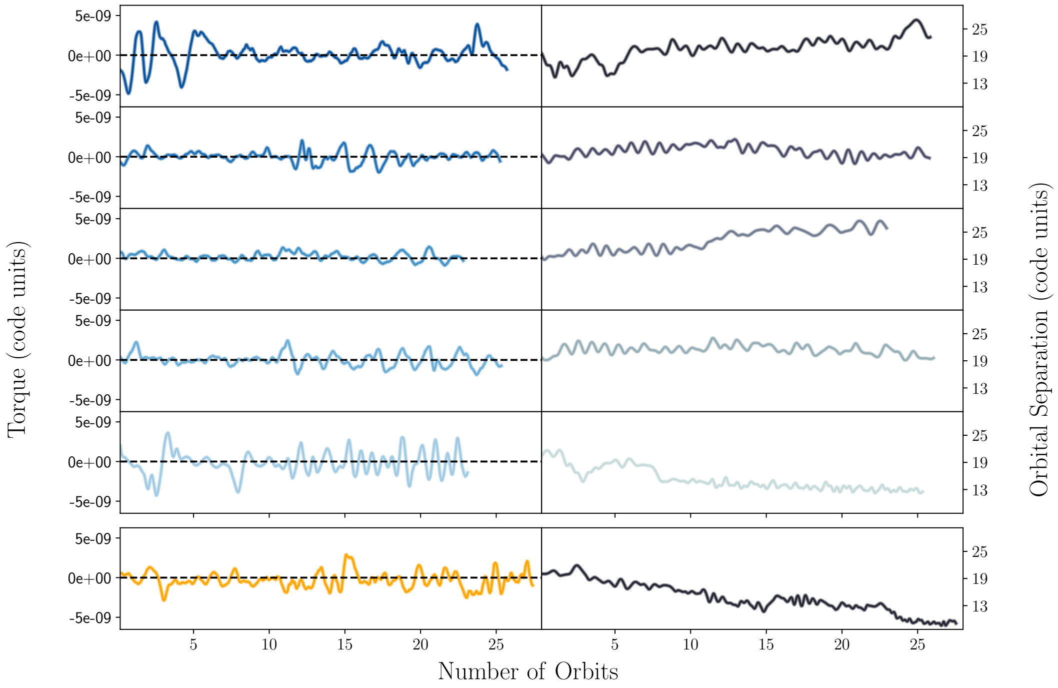


Figure 3.14: Total z-component of the torque exerted by the disc material on a $1M_\oplus$ planet is shown on the left panels. The migration tracks are shown on the right. The top 5 panels are simulations modelled with a variable β, whilst the bottom panel is one of the simulations modelled with a constant β. The migration is dominated by stochastic kicks which results in random outward or inward motion, as seen in the torques oscillating about zero.

usage of a realistic treatment of the disc thermodynamics meant that like in the simulations presented here, the planets are evolved in discs that are only gravitationally unstable in the outer regions. Using realistic thermodynamics to allow the gas to control its cooling/heating based on its properties, the thermodynamics of the circumstellar and circumplanetary disc is better captured. This enhances the mass growth of the planet causing the planets to grow to beyond the brown dwarf limit. This resulted in their planets opening up gaps and slowing their inward migration. Thus it is unclear whether the mass growth or its presence in the more gravitationally stable inner disc was the dominant factor in slowing the inward migration. We do not allow our planets to grow ($R_{\mathrm{acc}} = 0.001$) and hence show that modelling the cooling in the discs such that only the outer disc is gravitationally unstable, enables a planet to slow its migration in the inner gravitationally stable disc. Therefore, while we expect the growth of a planet would help slow its migration, our work shows that the location of the planet in the disc has an important effect too.

Recently Mercer et al. (2018), using their radiative transfer simultions calculated an effective value of beta, β_{eff}. Using this, they showed that a constant β does not accurately reflect a realistic gravitationally unstable protoplanetary disc. Their calculated β_{eff} was found to vary both spatially and with time as the disc evolves. Since the variable β model in this Chapter does not vary with time, to achieve a similar comparison, we compare the variable β profile here with their β_{eff} of a high mass disc that has not yet developed spiral arms. In general, the two compare reasonably well with both decreasing as roughly R^{-2} in the inner regions of the disc. The main difference between the two is in the outer regions where the β_{eff} of Mercer et al. (2018) decreased more slowly tending towards a non-zero β. However this difference could be enhanced by the differences in the simulation setup, where their simulations have a ambient radiative field of 10K and they simulate a higher disc-to-star mass ratio. Furthermore, their discs eventually fragment, whereas the β profile chosen here is such that disc fragmentation is avoided.

3.5.3 In the context of observations

The viability of early planet formation is important given recent observations of young protoplanetary discs. Some of these discs are a few million years old or younger (Andrews et al., 2018) and already show substructure such as rings and gaps. Assuming that the gaps were carved out by planets, Lodato et al. (2019) determined that a wide range of planet masses from super Earths to a few Jupiters at wide orbits ($10 - 100$ AU) were typically required. There is also an increasing body of evidence (Clarke et al., 2018; Flagg et al., 2019) that show that planets have formed in young discs ($\sim$ 1Myr).

Although the discs in the aforementioned studies may not be expected to be gravitationally unstable given their mass estimates, they still indicate that planets can form very early in a disc's lifetime. However, measuring the disc mass can be quite challenging. Recently Booth & Ilee (2020) using the rarer CO isotopologue, ^{13}C^{17}O measured the mass of HL Tau to be at least $0.2M_{\odot}$, up from the previous gas mass measurements of $2 - 40 \times 10^{-3} M_{\odot}$ (Wu et al., 2018). This puts HL Tau in the gravitationally unstable regime beyond 50AU, where a potential planet might be carving the gaps seen.

If planets can form in young discs, the next most natural question to ask is whether they can survive their migration. Our work shows that planets formed – by any process – in young discs can survive and evolve in the disc.

3.5.4 In the context of population synthesis models

The work presented here is also important for studying population synthesis models in self-gravitating discs, such as the ones in Galvagni & Mayer (2013), Forgan (2013), and Forgan et al. (2018a). A common problem in these models is using the rapid migration times. This results in very low survivability of any clumps formed via gravitationally stability, unless the clumps could open up a gap to slow their migration and survive. The results here show that with more realistically modelled thermodynamics, the inward migration of planets in gravitationally unstable discs can be suppressed without requiring gap-opening. Further work will be required to study the impact that a slow down of a planet's migration has on population synthesis studies.

3.5.5 Caveats

The main goal of this Chapter is to study the impact of implement a more realistic treatment of the thermodynamics on planet migration. Thus, the main difference compared to past studies with β-cooling models is how β varies with radius. Several other possible important parameters are kept fixed.

We only consider one disc model. The temperature and surface density profiles are the same for all simulations. The disc mass is also kept fixed. Although these would change how the Toomre parameter evolves, and thus how the gravitational stability of the disc evolves, it is not expected to change the result that planet migration can be slowed down in the gravitationally stable inner disc.

We have also limited accretion onto the planets. Studies have shown that disc fragments ($\sim 1 - 10 M_J$ Boley et al. 2010) can easily accrete enough material to become brown dwarfs (Stamatellos & Whitworth, 2009; Kratter et al., 2010b; Zhu et al., 2012a). To keep things simple, the accretion radii of the planets are small enough to ensure the planet's mass stays roughly constant. This is obviously not very realistic, but is done firstly to directly compare with Baruteau et al. (2011) and Malik et al. (2015) and secondly to understand the impact that different locations of gravitationally unstable discs have on planet migration and survivability.

Despite using a variable β to mimic the varying cooling in a real self-gravitating disc, the cooling in a disc is likely to be much more complex. Further work will be done in the future comparing the results here with a disc modelled using radiative transfer.

3.6 Conclusion

We perform 3D SPH simulations to investigate the migration of both giant planets with masses of $1M_{\mathrm{Sat}}$, $1M_{\mathrm{Jup}}$, and $5M_{\mathrm{Jup}}$, and low mass planets with masses of $1M_{\oplus}$, $10M_{\oplus}$, and $1M_{\mathrm{Nep}}$ in self-gravitating discs. Our study shows that by implementing a radially decreasing cooling time to more accurately cool the disc such that only the outer region of the disc is gravitationally unstable, the inward migration of planets is slowed down in the inner gravitationally stable part of the disc without requiring a gap to open up. These results show, most importantly that planets may be able to survive their inwards migration in real self-gravitating discs.

4

Hiding Signatures of Gravitational Instability in Protoplanetary Discs

4.1 Introduction

In the last few years, a large number of discs have revealed substructure in the form of rings and gaps when observed at millimeter wavelengths with ALMA (ALMA Partnership et al., 2015; Andrews et al., 2016; Fedele et al., 2018; Andrews et al., 2018; Huang et al., 2018a; Dipierro et al., 2018; Booth & Ilee, 2020). A few of these discs are thought to be very young (< 1Myr) (ALMA Partnership et al., 2015; Fedele et al., 2018; Dipierro et al., 2018). Ring and gap structures have also been observed in even younger ($\lesssim 0.5\,$Myrs) Class 1 discs (Sheehan & Eisner, 2018; Segura-Cox et al., 2020). Young discs are thought to be massive and could potentially be gravitationally unstable. Such discs are expected to harbour spiral arms. There is evidence that discs with spiral arms in the midplane exist (Pérez et al., 2016; Huang et al., 2018a), although they seem to be quite rare. Is the observed rarity due to young discs being less massive or can spiral structures in massive discs be hidden by other processes?

The goal of this Chapter is to investigate the impact that a migrating giant planet, irrespective of how it formed, has on the structure of a gravitationally unstable disc and the resulting implications for observations of such discs. We find that a migrating giant planet is able to suppress spiral structures yielding discs that appear axisymmetric with rings and gaps.

4.2 Method

4.2.1 Hydrodynamic simulations & initial conditions

We perform scale free 3D gas hydrodynamic simulations using PHANTOM, a smoothed particle hydrodynamics (SPH) code developed by Price et al. (2018a).

The disc setup is identical to that in Rowther & Meru (2020). We model a disc using 2 million particles between $R_{\text{in}} = 1$ and $R_{\text{out}} = 25$ in code units with a disc-to-star mass ratio of 0.1. The central star, and the planet are modelled as sink particles (Bate et al., 1995). The accretion radius of the central star is set to be equal to the disc inner boundary, R_{in}. To maintain a roughly constant planet mass throughout the simulation, the accretion radius of the planet is limited by setting it to 0.001 in code units. This is ~10 times smaller than the minimum Hill radius for a $3M_{Jup}$ planet at R_{in}. The initial surface mass density is set as a smoothed power law and is given by

$$\Sigma = \Sigma_0 \left(\frac{R}{R_0}\right)^{-1} f_s, \tag{4.1}$$

where Σ_0 is the surface mass density at $R = R_0 = 1$ and $f_s = 1 - \sqrt{R_{\text{in}}/R}$ is the factor used to smooth the surface density at the inner boundary of the disc. The initial temperature profile is expressed as a power law

$$T = T_0 \left(\frac{R}{R_0}\right)^{-0.5}, \tag{4.2}$$

where T_0 is set such that the disc aspect ratio $H/R = 0.05$ at $R = R_0$. The energy equation is

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\left(\nabla \cdot \mathbf{v}\right) + \Lambda_{\text{shock}} - \frac{\Lambda_{\text{cool}}}{\rho} \tag{4.3}$$

where we assume an adiabatic equation of state, u is the specific internal energy, the first term on the RHS is the $P\mathrm{d}V$ work, Λ_{shock} is a heating term that is due to the artificial viscosity used to correctly deal with shock fronts, and

$$\Lambda_{\text{cool}} = \frac{\rho u}{t_{\text{cool}}} \tag{4.4}$$

controls the cooling in the disc. Here we use a simple implementation of the cooling time which is proportional to the dynamical time by a factor of β,

$$t_{\text{cool}} = \beta(R)\Omega^{-1} = \beta_0 \left(\frac{R}{R_0}\right)^{-2} \Omega^{-1}, \tag{4.5}$$

where Ω is the orbital frequency and we have varied β with radius (Rowther & Meru, 2020). This allows us to mimic a realistic self-gravitating disc that is only gravitationally unstable in the outer regions (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009).

To model shocks, we use an artificial viscosity switch that utilises the time derivative of the velocity divergence introduced by Cullen & Dehnen (2010). The artificial viscosity parameter α_v has a maximum of $\alpha_{\text{max}} = 1$ near the shock and a minimum of $\alpha_{\text{min}} = 0$ far away from the shock. The artificial viscosity coefficient β_v is set to 2 (see Price et al. 2018a).

Embedding the planet

The simulation is first run for 10 orbits to allow spiral structure in the disc to develop. We then split it into two simulations, one with an embedded planet and one which continues to evolve without a planet. The latter is a control simulation to compare the impact a migrating giant planet has on the spiral structure in a gravitationally unstable disc. A planet with a planet-to-star mass ratio of $q = 2.9 \times 10^{-3}$, equivalent to $3M_{\text{Jup}}$ in a $0.1M_\odot$ disc around a $1M_\odot$ central star is added at $R_p = 20$.

The simulations are run for another 8 orbits, with a total simulation time of 18 orbits.

4.2.2 Post processing of simulations

The raw synthetic continuum images at 1.3mm are created using MCFOST (Pinte et al., 2006, 2009). We first scale the simulations such that $R_{\text{out}} = 200$AU, and initial $R_p = 160$AU. We use 10^8 photon packets on a Voronoi tesselation where each MCFOST cell corresponds to an SPH particle. The luminosity of the star is calculated using an assumed mass of $1M_\odot$ and a 1Myr isochrone from Siess et al. (2000) which corresponds to a temperature of $T_\star = 4286$K. Since the gas surface density of the disc is quite large, the Stokes numbers in the disc are small enough (< 0.1 for mm sized grains) that we assume that the dust is well coupled to gas. Therefore we assume that the dust distribution is identical to the gas distribution and a constant dust-to-gas ratio of 0.01. The dust sizes vary between 0.3 and 1000 μm and are distributed across 100 different sizes with a power-law exponent of -3.5. We assume all dust grains are made of astronomical silicates, and are spherical and homogeneous. We compute the dust properties using Mie theory. The disc is assumed to be at a distance of 140pc.

To create mock millimeter continuum observations, we use the ALMA Observation Support Tool (Heywood et al., 2011). We use integration times of 12, 30, 60, and 120 minutes at 230GHz (1.3mm) in Band 6 with the ALMA Cycle 8 C43-7 configuration. We assume a bandwidth of 7.5GHz and a precipitable water vapour level of 0.913mm. `CLEAN` images are created using natural weights resulting in a beam size of $0.107'' \times 0.124''$, or equivalently 15.0AU ×17.3AU at 140pc. As shown by Mayer et al. (2016), different ALMA configurations can alter the detectability of features. Hence a similar set mock observations are created with the C43-6 configuration ($0.156'' \times 0.196''$) and with a disc inclined at 40° using the C43-7 configuration.

Assuming a disc inclination and position angle of 40°, optically thick ^{12}CO, and optically thinner ^{13}C^{16}O CO-isotopologue channel maps are generated for the $J = 3 - 2$ transition using a velocity resolution of 0.1 km/s. We choose this transition as it can be observed with a good compromise between observation time and signal-to-noise ratio. The abundances for ^{12}CO and ^{13}C^{16}O are assumed to be a fraction 1×10^{-4} and 2×10^{-7} of the total disc mass (i.e. relative to H_2) respectively. We account for CO freeze-out at $T < 20$K, and photo-dissociation and photo-desorption in regions of high UV radiation (see Appendix B of Pinte et al. 2018a). We then convolve the resulting images with a beam size of $0.05'' \times 0.05''$; the maximum resolution expected for CO line observations.

4.3 Results

4.3.1 Impact on spiral structure

As shown in Rowther & Meru (2020), planets migrate inwards rapidly until they reach the gravitationally stable inner disc. For the simulations presented here, the gravitationally stable region is inside $R \approx 120$AU. The impact of the planet on the disc as it migrates is compared to the control simulation using the Toomre parameter (Toomre, 1964),

$$Q = \frac{c_s \Omega}{\pi G \Sigma}, \tag{4.6}$$

which gives a measure of how gravitationally unstable the disc is, where Σ and c_s are the disc surface density and sound speed, respectively.

Figure 4.1 shows the surface density (left) and the 2D Toomre parameter Q

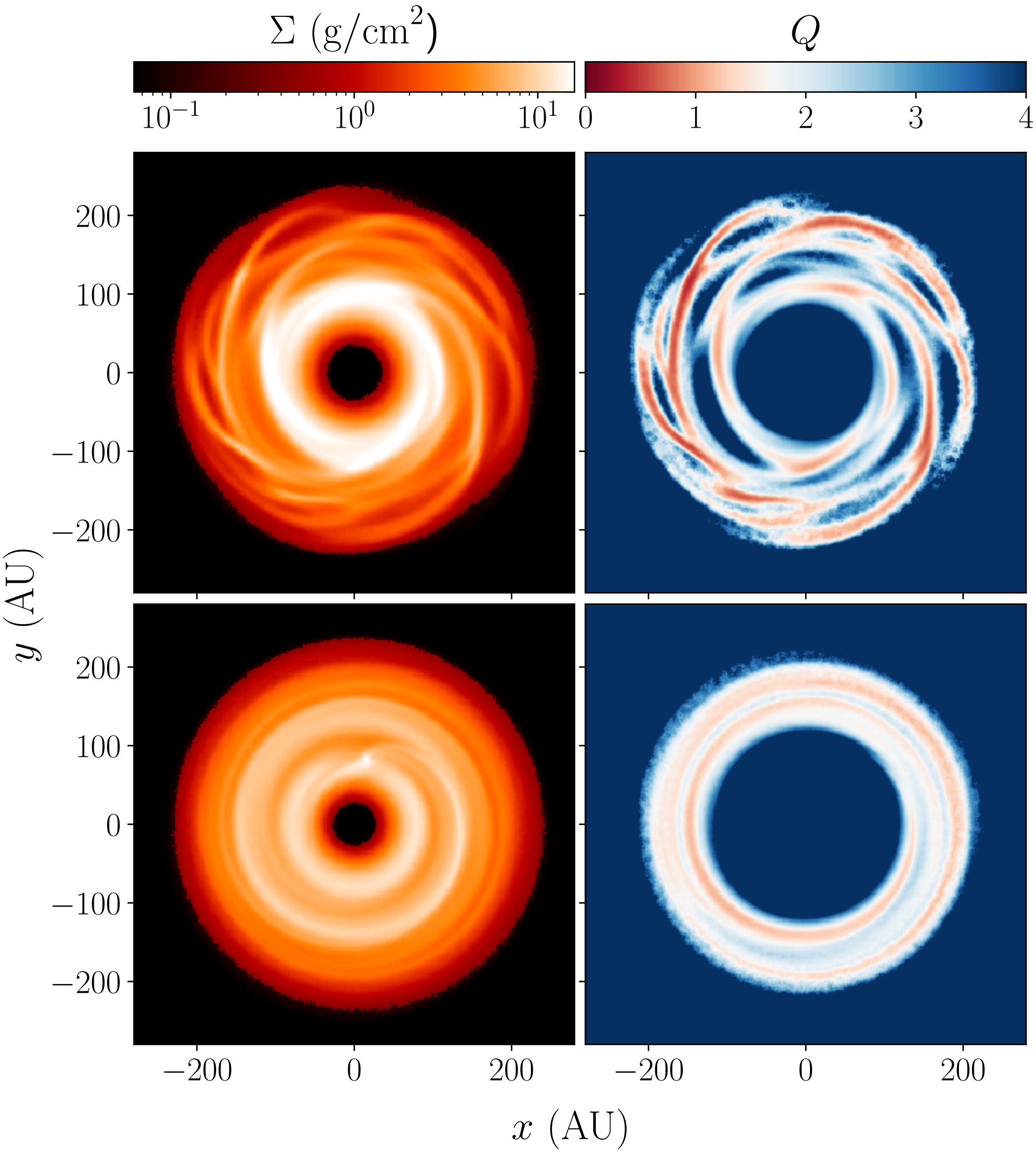


Figure 4.1: Surface density (left) and 2D Toomre parameter (right) plots of a disc without a planet (top) and with a $3M_{\rm Jup}$ planet (bottom) at the end of the simulation. The 2D Toomre plots show how gravitationally unstable the disc is. The critical value for non-axisymmetric instabilities is when $Q \lesssim 1.7$ (Durisen et al., 2007), shown in red. The planet erases the spiral structure and is massive enough to carve out a gap. This results in a gravitationally stable axisymmetric disc.

(right) of the simulations without a planet (top) and with a $3M_{\rm Jup}$ planet (bottom) at the end of the simulation. It is clear that the presence of the planet significantly impacts the structure. The planet suppresses the spiral structure and opens up a

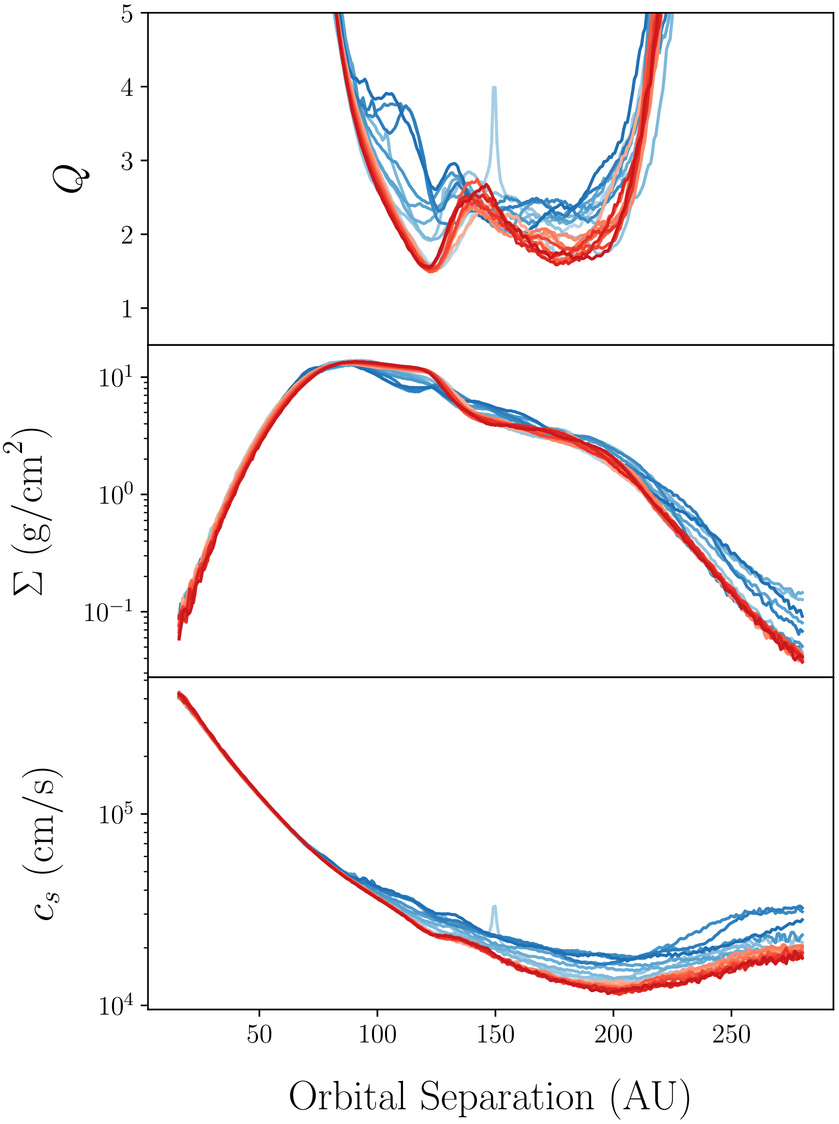


Figure 4.2: Azimuthally averaged Toomre parameter, Q (top), surface density, Σ (middle), and the sound speed, c_s (bottom) for both the disc without a planet (red lines) and with a $3M_{\text{Jup}}$ planet (blue lines) for the first 2 orbits every 0.2 orbits after the planet is added. The time evolution during the first 2 orbits is shown by the darker shades representing later times. The increase in c_s due to the planet is the dominant reason for the increase in Q in the outer regions of the disc (beyond $R \approx 120$AU), causing the disc to become gravitationally stable.

gap. Despite both discs retaining the same disc mass ($\sim$0.09$M_\odot$) within 200AU, the 2D Q plots show that the presence of the planet results in the disc becoming mostly gravitationally stable with $Q > 1.7$ throughout the disc. Without a planet, strong gravitationally unstable ($Q < 1.7$) arms are present.

The dominant reason for the increase in Q is due to the planet's influence on the disc temperature, or equivalently the sound speed c_s, in the disc. The spiral wakes generated by the planet as it migrates are regions of relative overdensities with respect to the disc background. The radially propagating spiral wakes can evolve into shocks. The exchange of momentum and energy between the density wakes and the disc at the shocks influence the global properties of the disc, heating it up (Goodman & Rafikov, 2001; Rafikov, 2016; Ziampras et al., 2020).

The azimuthally averaged Toomre parameter (top), surface density (middle), and sound speed (bottom) is plotted in Figure 4.2 for the first 2 orbits after the $3M_{\rm Jup}$ planet (blue lines) is embedded. The red lines represent the disc without the planet at the same times. It can be seen, particularly in the regions of the disc beyond $R \approx 120$AU, that there is a quick and significant increase in the sound speed after the planet has been added, whereas, the change in Σ is relatively small. This increase in temperature causes the disc to become gravitationally stable, resulting in a higher value of the Toomre parameter.

To ensure the sudden inclusion of the $3M_{\rm Jup}$ planet did not trigger artificial excess heating, we compared the artificial viscosity to the gravitational stress parameter immediately after the planet was embedded. The increase in artificial viscosity due to the planet is negligible compared to the magnitude of the gravitational stress parameter.

4.3.2 Continuum images

The left column of Figure 4.3 compares the mock observations of a disc with a $3M_{\rm Jup}$ planet (bottom) and without a planet (top) with an integration time of 30 minutes using the C43-7 configuration. An axisymmetric flux map is produced as the azimuthally averaged flux of the mock observations and is shown in the middle column. Finally to highlight non-axisymmetric features, the residual flux is plotted in the right column by subtracting the axisymmetric map from the mock observation. The top subset of panels in Figure 4.4 show the mock observations (top half) and residuals (bottom half) using the higher resolution C43-7 configuration. The bottom left subset of panels show the same with the lower resolution C43-6 configuration. The bottom right subset of panels are for a disc inclined by 40° using the C43-7 configuration.

From Figure 4.3, it can be seen that without a planet, spiral arms due to gravitational instability would be readily apparent. Whereas with the planet, the only major non-axisymmetric feature that remains are the spiral arms caused by the

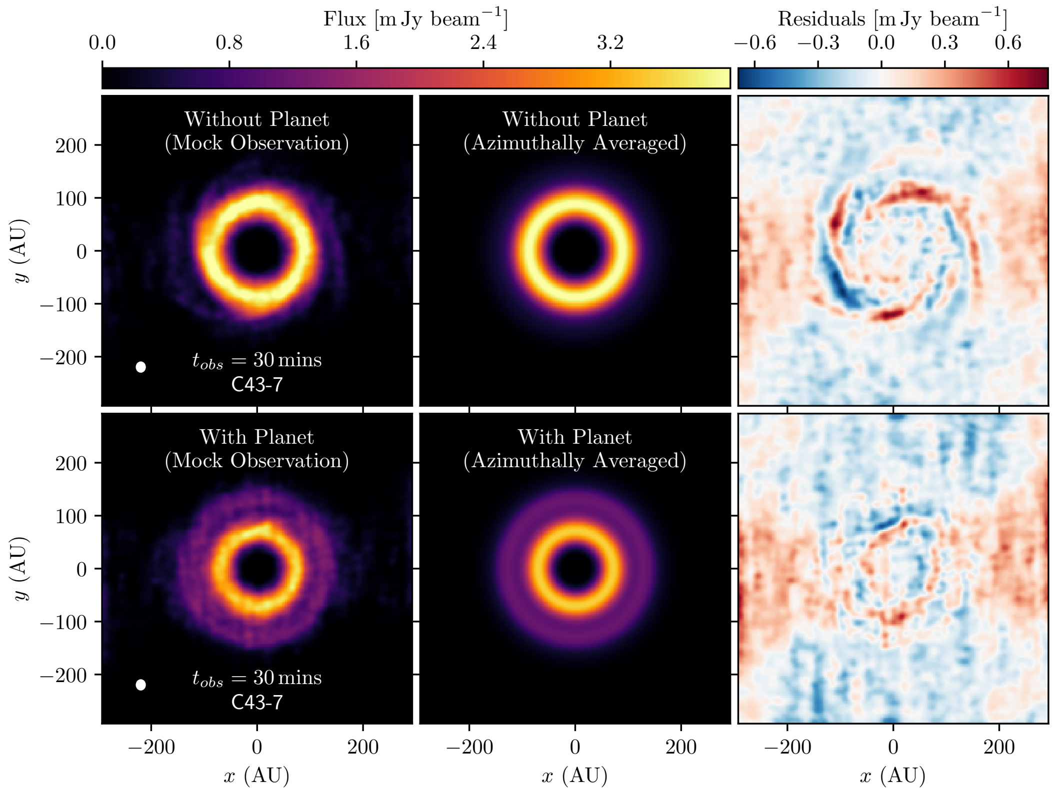


Figure 4.3: Mock observation (left) using the C43-7 configuration with an integration time of 30 minutes, axisymmetric flux map (middle), and residual flux (right) of a disc without a planet (top) and with a $3M_{\mathrm{Jup}}$ planet (bottom). The white ellipse in the bottom left of each mock observation represents the beam size. The presence of the planet results in a disc that no longer consists of spiral structure due to gravitational instability and appears to be more axisymmetric.

planet as shown in Figures 4.3 and 4.4. The higher resolution mock observations are able to better resolve the gap, however the planet induced spiral arms are only visible with integration times > 1hr. This is true for both the non-inclined and moderately inclined disc. Using a lower resolution, which increases the signal-to-noise, allows the spiral arms to be easily seen in the residuals at lower integration times. The tradeoff is a gap that is less resolved and shallower in the continuum image. Despite the lower resolution images showing the spiral arms of the planet more easily, higher resolution images would still be favourable as a less resolved and shallower gap can lead to underestimates of the planet's mass. This could in turn lead to predictions with other methods of estimating the planet's mass, such as with the CO kinematics, that are inconsistent.

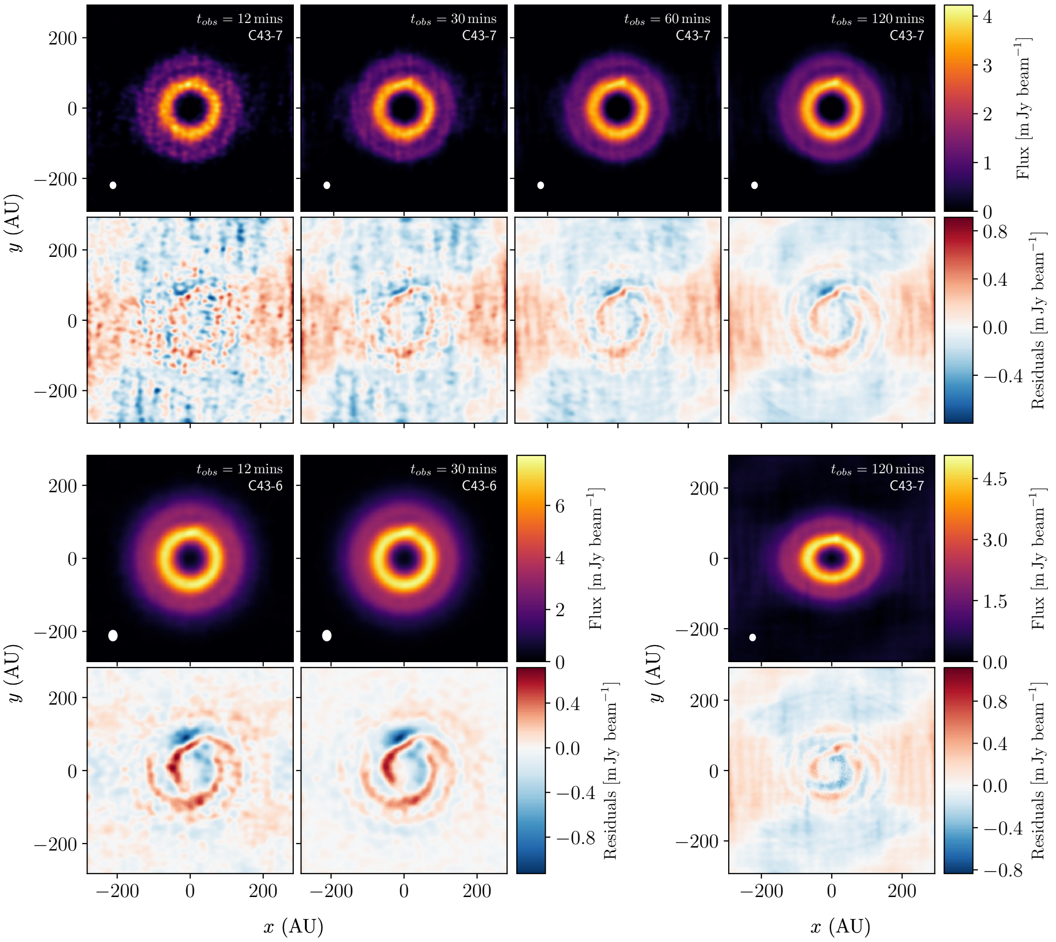
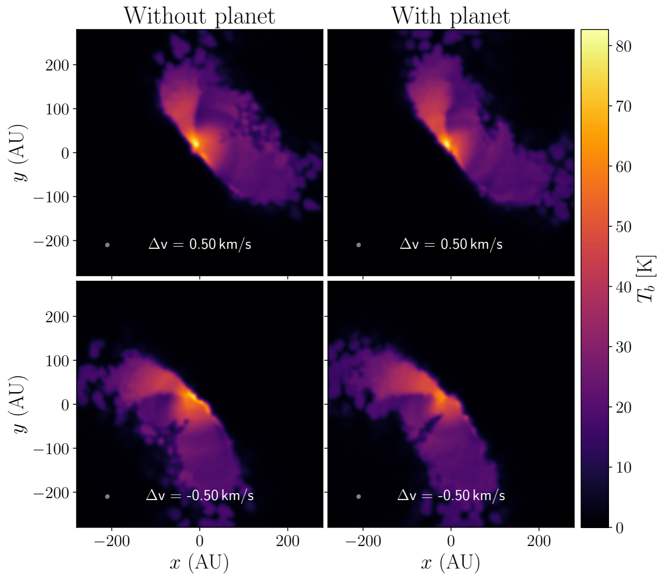


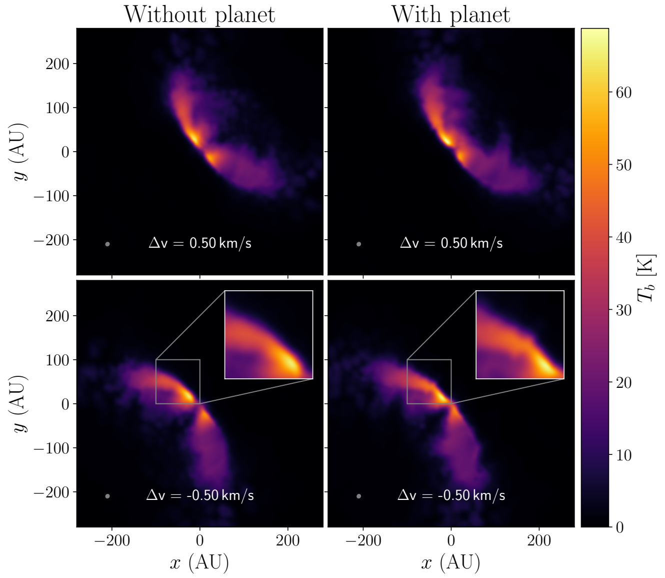
Figure 4.4: Top subset of panels show the mock observations (top half) and residual plots (bottom half) of a disc with a $3M_{\mathrm{Jup}}$ planet with integration times of 12, 30, 60, and 120 minutes using the higher resolution C43-7 configuration. The bottom left subset of panels is similar but only for integrations times of 12 and 30 minutes using the C43-6 configuration. At high resolution the disc appears completely axisymmetric with short integration times. Whereas with longer integration times (over 60 minutes), spiral arms caused by the planet can be seen in the residual plots. At lower resolutions, the gap carved by the planet appears shallower, but the planet's spiral arms are easier to detect even with short integration times. The bottom right subset of panels show that the planet's spiral arms are also detectable for moderately inclined discs. This is shown for a disc inclined by 40°.

4.3.3 CO kinematics

The spiral waves generated by the planet can cause localised deviations in the Keplerian flow of the disc as shown recently by Pinte et al. (2018b, 2019, 2020). These



(a) ^{12}CO



(b) ^{13}C^{16}O

Figure 4.5: Synthetic continuum subtracted channel maps (J = 3 − 2 transitions) for a) ^{12}CO and b) ^{13}C^{16}O at $\Delta v = \pm 0.5\,\mathrm{km\ s^{-1}}$ from the systemic velocity of the disc with and without the planet. A kink is not detected in the optically thick ^{12}CO, but is visible in the optically thin ^{13}C^{16}O in the negative velocity channel within the vicinity of the planet, see bottom right inset in Fig 4.5b. The bend in the velocity profile in the bottom right inset is referred to as the kink. This is contrasted with the smooth velocity profile in the bottom left inset. The disc inclination and position angle are both 40°.

deviations can be detected as *kinks* in the gas channel maps.

Figure 4.5 shows the channel maps for ^{12}CO and ^{13}C^{16}O at $\Delta v = \pm 0.5\,\mathrm{km\ s^{-1}}$ from the systemic velocity with the continuum subtracted. The optically thick ^{12}CO does not reveal any signs of a planet, but a kink is noticeable in the optically thinner ^{13}C^{16}O in the $-0.5\,\mathrm{km\ s^{-1}}$ channel within the vicinity of the planet. We can exclude large scale perturbations or any azimuthally symmetric mechanisms since the kink is not seen in the opposite velocity channel where the profile remains smooth. The kink is shown in the bottom right panel of Fig 4.5b in the inset. The kink is also seen in the channels from $v = -0.3$ to $-1.2\,\mathrm{km\ s^{-1}}$.

Although the channel maps without the planet are not perfectly smooth as would be expected from a Keplerian disc, this is likely to be due to the gravitational instabilities. Hall et al. (2020) show that for a gravitationally unstable disc the perturbations due to gravitational instabilities are seen in all azimuths and in both the positive and negative velocity channels. However in our simulations with a planet, the kink is more localised appearing in only the negative velocity channel. To ensure that the observed kink is due to the planet, we re-calculated the channel maps with the planet at slightly different times with the planet being in different azimuthal locations. Although the kink is not detectable at all times, whenever the kink is detected it is in a different location determined by the position of the planet. Additionally, in some cases the kink was only detected for a narrower range of Δv.

4.4 Discussion

4.4.1 Implications for dust-to-gas mass ratios

In the early stages of evolution, discs are expected to be massive and gravitationally unstable (e.g. Bate, 2018). A characteristic feature of such discs are their spiral arms. However, from observations so far (Pérez et al., 2016; Huang et al., 2018a), these discs appear to be quite rare. Due to the difficulty in directly measuring the gas mass of the disc, the disc mass is often inferred via the dust mass using a fixed dust-to-gas mass ratio. In some scenarios, such as with MWC 480 (Liu et al., 2019), the observed dust mass can be high enough such that inferring the gas mass via the canonical dust-to-gas mass ratio of 0.01 results in a disc that is massive enough to be gravitationally unstable. Thus in their models, they assume a higher dust-to-gas mass ratio to ensure that the disc is gravitationally stable. Facchini et al. (2020) show that LkCa-15 retains a significant amount of dust mass in its rings. Consequently, when modelling LkCa-15, they assume an upper limit on the disc gas mass (and hence a higher dust-to-gas mass ratio) to ensure the disc is not locally gravitationally unstable.

In both cases this assumption was deemed necessary to explain the lack of spiral features in the observations, which would be expected from more massive discs. However, we show that a migrating $3M_{\mathrm{Jup}}$ planet can cause a massive disc to become gravitationally stable and suppress any spiral structure that would otherwise be present. Hence, large dust-to-gas mass ratios are not necessarily required to explain a lack of spiral features.

4.4.2 Caveats

A caveat to the results presented here is the long term gravitational stability of the discs due to the cooling mechanism used. Although it mimics the characteristics of a realistic self-gravitating disc, i.e. a disc which is only gravitationally unstable in the outer regions, it is still a straight-forward implementation where the cooling time is a simple expression determined by the location in the disc.

The actual cooling in the disc is likely to be more complex and evolve over time. Recently Mercer et al. (2018) calculated an effective β from their radiative transfer calculations, which was found to vary both spatially and with time. Hence, it is unknown what impact using more complex cooling methods such as radiative transfer to cool the disc will have on the long term gravitational stability of the disc with a migrating planet. Whilst we show that a migrating giant planet can erase the spiral structure in a self-gravitating disc, it remains to be investigated whether the loss of spiral structure remains for a significant amount of time.

The results presented here initially appear contrary to Meru (2015) where it was shown that a fragment formed by gravitational instability in the outer disc could trigger subsequent fragmentation in the inner disc. The main difference is that the Toomre profiles are quite different. In Meru (2015), the inner disc was on the edge of fragmentation ($Q \sim 1$). Hence the increase in the surface density as a result of the fragment driving material inwards towards the disc caused the disc to fragment. This behaviour is not seen or expected in this Chapter as the inner disc remains well above the gravitationally stable regime ($Q \gtrsim 1.7$).

In this Chapter, we have only presented results for one planet and disc mass. Lower mass planets would be less likely to impact self-gravitating spiral structures. On the other hand, lower mass discs, which have weaker gravitational instabilities, would be more susceptible to having their spiral structures suppressed by planets. Although we find that a planet that is large enough to open up a gap also suppresses the spiral structure, we cannot conclude that non-gap-opening planets are unable to affect the disc structure. We will present a follow up study that considers a variety of planet and disc properties, as well as ALMA configurations, to investigate the conditions at which spiral suppression is likely to be observed.

4.5 Conclusion

We perform 3D SPH simulations to investigate the impact a migrating giant planet has on the structure of a gravitationally unstable disc. Our work shows that the presence of the planet suppresses the spiral structure in the disc and causes the disc to become gravitationally stable because it alters the temperature structure. This interaction between the planet and the disc causes the self-gravitating phase of the disc to be shortened, while retaining the same disc mass. The planet is able to carve open a gap resulting in an axisymmetric disc.

The mock ALMA observations of the continuum presented here show that the disc can appear completely axisymmetric for higher resolution ALMA configurations. However, with longer integration times or by sacrificing resolution, spiral arms from the planet become observable. In the latter case, the gap carved by the planet will be less resolved and appear shallower. We also show that the planet can be detected with high resolution kinematics using optically thin CO-isotopologues like ^{13}C^{16}O. Our results show it is possible to explain a lack of spiral structure in high mass discs without requiring high dust-to-gas mass ratios to limit the gas mass.

5

Continuing to Hide Signatures of Gravitational Instability in Protoplanetary Discs

5.1 Introduction

In the last few years the Atacama Large Millimeter/submillimeter Array (ALMA) has provided us with a large number of well resolved discs at millimeter wavelengths. These observations have revealed the ubiquity of disc substructures. The most common of these substructures being axisymmetric rings & gaps. (ALMA Partnership et al., 2015; Andrews et al., 2016; Fedele et al., 2018; Andrews et al., 2018; Huang et al., 2018a; Dipierro et al., 2018; Long et al., 2018; Booth & Ilee, 2020). A few of these discs are thought to be very young ($<$ 1Myr) (ALMA Partnership et al., 2015; Fedele et al., 2018; Dipierro et al., 2018). In their youth discs are expected to be more massive and in the regime where the disc self-gravity drives its evolution. In this phase of their lives, young massive discs develop gravitationally instabilities in the form of large-scale spiral structures. These discs are often referred to as GI discs. However, even young Class I discs show evidence of rings & gaps (Segura-Cox et al., 2020; Sheehan & Eisner, 2018). Although rare, there is evidence that discs with spiral features in the midplane exist (Pérez et al., 2016; Huang et al., 2018a). *Is the rarity of spiral structures in the disc midplane due to young discs being less massive than we expect? Or can gravitational instabilities in massive discs be hidden* *by other processes?*

Protoplanetary discs are primarily made up of circumstellar gas H_2, which is difficult to observe and makes measuring the mass of the disc challenging. Thus, the total mass of the disc has to be inferred from the other constituents of the disc, the dust or molecular gas (usually CO). By assuming a globally constant dust-gas mass ratio, one can infer the gas mass from the more easily observed dust. However, the inferred gas mass remains very uncertain due to the assumptions that go into the disc radius, dust-to-gas mass ratio, dust opacities, distribution of grain sizes, or local over/under-densities (Testi et al., 2014; Andrews, 2015, and references therein). Inferring the disc mass using CO has other problems. The disc masses measured can be underestimated if the CO line emission is optically thick, and hence doesn't trace the disc midplane (Bergin & Williams, 2017). Additionally, measurements of the disc mass using CO can also be affected by freeze-out onto icy grains (van Zadelhoff et al., 2001) and photodissociation (Reboussin et al., 2015). However, Booth et al. (2019) showed that disc masses can be larger when using $^{13}C^{17}O$, an optically thin isotopologue which is a more robust tracer of the disc mass, consistent with early mass estimates by Greaves et al. (2008). In HL Tau, Booth & Ilee (2020) found that the disc mass estimates pushed the outer regions of the disc into the gravitationally unstable regime. Interestingly, the continuum data for HL Tau shows axisymmetric ring & gap structure (ALMA Partnership et al., 2015) instead of spiral structures; in contrast to what is expected from massive gravitationally unstable discs.

The physical interaction between a planet and the disc can alter the disc structure through the exchange of angular momentum (Kley & Nelson, 2012). Thus it is natural to invoke planets as an explanation for the disc substructures such as rings & gaps. Comparisons of hydrodynamic and radiative transfer simulations to observations from ALMA (Clarke et al., 2018; Zhang et al., 2018) have validated this interpretation. The evidence for planet-disc interactions became stronger by **?**Pinte et al. (2020) where it was shown that the planet could be detected by the localised kink it produces in the gas kinematics. Given the young ages of some of these discs with rings & gaps, if they were formed by planets then it's reasonable to assume the planet may have formed when the disc was younger and perhaps gravitationally unstable. Hence, it's necessary to understand how planet-disc interactions impact gravitationally unstable discs and their observability.

Recent work has shown that the evolution of a gravitationally unstable disc can be altered by the same mechanisms that are often used in lower mass discs, as shown in Rowther et al. (2020) with a gap-opening planet, and by a warp in Rowther et al. (2022). In the case of a planet, its spiral wake influences global properties of the disc. While in a warped disc, the non-coplanar geometry results in an oscillating radial pressure gradient, which alters the velocities in the disc. In both, the disc was heated up rendering it gravitationally stable with an axisymmetric ring & gap structure.

In this Chapter we perform three-dimensional global numerical simulations to extend the study in Rowther et al. (2020) by exploring a wider set of planet and disc masses to fully understand the interplay between planet-disc interactions and gravitational instabilities, and its implications on observations. This paper is organised as follows. In §5.2 we describe the simulations presented in this Chapter. In §5.3 we present our results on how planet-disc interactions impact gravitationally unstable discs. The limitations and observational implications of this Chapter are discussed in §5.4. We conclude our work in §5.5.

5.2 Model

We use Phantom, a smoothed particle hydrodynamics (SPH) code developed by Price et al. (2018a) to perform the suite of simulations presented here.

5.2.1 Disc setup

The disc setup is identical to that in Rowther & Meru (2020) & Rowther et al. (2020), and is summarised below. All discs are modelled using 2 million particles between $R_{\text{in}} = 1$ and $R_{\text{out}} = 25$ in code units where the fiducial disc has a disc-to-star mass ratio of 0.1. This is equivalent to a $0.1M_{\odot}$ disc around a $1M_{\odot}$ star. Sink particles (Bate et al., 1995) are used to model both the central star and the planet. The accretion radius of the central star is set to be equal to the disc inner boundary, R_{in}. The surface density profile Σ is given by

$$\Sigma = \Sigma_0 \left(\frac{R}{R_0}\right)^{-1} f_s, \tag{5.1}$$

where Σ_0 is the surface mass density at $R = R_0 = 1$ and $f_s = 1 - \sqrt{R_{\text{in}}/R}$ is the factor used to smooth the surface density at the inner boundary of the disc. The initial temperature profile is expressed as a power law

$$T = T_0 \left(\frac{R}{R_0}\right)^{-0.5}, \tag{5.2}$$

where T_0 is set such that the disc aspect ratio $H/R = 0.05$ at $R = R_0$. The energy equation is

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}\left(\nabla \cdot \boldsymbol{v}\right) + \Lambda_{\mathrm{shock}} - \frac{\Lambda_{\mathrm{cool}}}{\rho} \tag{5.3}$$

where we assume an adiabatic equation of state, and u is the specific internal energy, P is the pressure, ρ is the density and $\boldsymbol{v}$ is the velocity. The first term on the RHS is the $P\mathrm{d}V$ work, and Λ_{shock} is a heating term that is due to the artificial viscosity used to correctly deal with shock fronts. The final term

$$\Lambda_{\mathrm{cool}} = \frac{\rho u}{t_{\mathrm{cool}}} \tag{5.4}$$

controls the cooling in the disc. Here the cooling time is straightforwardly implemented to be proportional to the dynamical time by a factor of $\beta(R)$,

$$t_{\mathrm{cool}} = \beta(R)\Omega^{-1} = \beta_0 \left(\frac{R}{R_0}\right)^{-2} \Omega^{-1}, \tag{5.5}$$

where Ω is the orbital frequency and we have varied β with radius (Rowther & Meru, 2020). In addition to the disc mass, the strength of gravitational instability also depends on the cooling factor (Cossins et al., 2009). Hence, we set $\beta_0 = 5500$. The cooling factor is $\beta = 13.75$ and 8.8 at the planet's location, $R_p = 20$ and at R_{out}, respectively. This allows us to mimic a realistic self-gravitating disc that is only gravitationally unstable in the outer regions (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009), while being comparable to constant β simulations in the outer regions.

Cullen & Dehnen (2010) introduced an artificial viscosity switch that utilises the time derivative of the velocity divergence, which we use here to model shocks. The artificial viscosity parameter α_{AV} varies depending on the proximity to a shock. Close to the shock, it takes a maximum of $\alpha_{\mathrm{max}} = 1$, and a minimum of $\alpha_{\mathrm{min}} = 0$ far away. The artificial viscosity coefficient β_{AV} is set to 2 (see Price et al. 2018a; Nealon et al. 2015).

5.2.2 The Suite of Simulations

In our suite we consider three different disc masses and three planet masses. We begin by simulating three discs – that differ only in their mass – for 10 orbits at $R = R_{\mathrm{out}}$ to allow spiral structure in the disc to develop. The final snapshot from each of these three simulations are then used as the initial condition for four subsequent simulations; three of these containing embedded planets of differing masses and

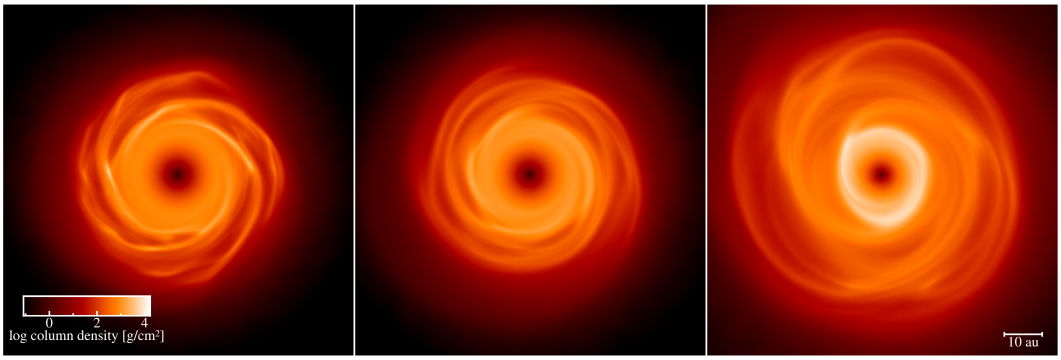


Figure 5.1: Surface density rendered plots of a $0.08M_\odot$, $0.1M_\odot$, and a $0.25M_\odot$ disc after 10 orbits at $R = R_{\rm out}$, right before a planet is embedded. Spiral structures due to gravitational instabilities are seen in all three discs to different extents. The higher the disc mass, the more gravitationally unstable the disc becomes as highlighted by the decreasing size of the gravitationally stable inner disc as the disc mass increases.

one without a planet. The latter is a control simulation to compare the impact a migrating giant planet has on the spiral structure in a gravitationally unstable disc.

To investigate the importance of the planet mass on suppressing spiral structures in a GI disc, we choose $1M_{\rm Sat}$, $1M_{\rm Jup}$, and a $3M_{\rm Jup}$ for our planet masses. Each planet is embedded (in code units) at $R_p = 20$ with an accretion radius of 0.001, and the simulation is run for another 8 orbits, with a total simulation time of 18 orbits. To investigate the importance of the disc mass on planet-disc interactions in a GI disc, for our disc masses we choose a less gravitationally unstable ($0.08M_\odot$) and a more gravitationally unstable ($0.25M_\odot$) disc in addition to our fiducial ($0.1M_\odot$) disc. Figure 5.1 shows each of the three discs right before the planet is embedded.

5.2.3 Post processing of simulations

The raw synthetic continuum images at 1.3mm are created using MCFOST (Pinte et al., 2006, 2009). The simulations are scaled such that the disc size is 200AU, and initial $R_p = 160$AU. We use 10^8 photon packets on a Voronoi tesselation where each MCFOST cell corresponds to an individual SPH particle. The luminosity of the star is calculated assuming a mass of $1M_\odot$ and a 1Myr isochrone from Siess et al. (2000) which corresponds to a temperature of $T_\star = 4286$K. We assume that the dust is perfectly coupled to the gas and a constant dust-to-gas ratio of 0.01. As the Stokes numbers in all the simulations in this Chapter are less than unity, this assumption is valid. Additionally, in a GI disc, the dust is stil able to couple quite well to the gas even at Stokes numbers of ∼10 (Baehr & Zhu, 2021). The dust sizes vary between

0.3 and 1000 μm and are distributed across 100 different sizes with a power-law exponent of -3.5. We assume all dust grains are made of astronomical silicates, and are spherical and homogeneous. We compute the dust properties using Mie theory. The disc is assumed to be at a distance of 140pc.

Mock millimeter continuum observations are created using the ALMA Observation Support Tool (Heywood et al., 2011). We use an integration time of 60 minutes in the ALMA Cycle 9 C-7 configuration. We assume a bandwidth of 7.5GHz and a precipitable water vapour level of 0.913mm. `CLEAN` images are created using natural weights resulting in a beam size of $0.107'' \times 0.124''$, or equivalently 15.0AU $\times$17.3AU at 140pc.

In Rowther et al. (2020), the kink caused by a $3M_{\mathrm{Jup}}$ planet in a $0.1M_{\odot}$ disc was not visible in the channel maps of the optically thick ^{12}CO. Therefore in this Chapter, the kinematics is investigated using just the optically thinner ^{13}C^{16}O. Assuming a disc inclination and position angle of 40°, channel maps are generated for the $J = 3 - 2$ transition using a velocity resolution of 0.1 km/s. We choose this transition as it can be observed with a good compromise between observation time and signal-to-noise ratio. The abundance of ^{13}C^{16}O is assumed to be a fraction 7×10^{-7} of the total disc mass (i.e. relative to H_2) respectively. Unless specified, we do not account for CO freeze-out at $T < 20$K, and photo-dissociation and photo-desorption in regions of high UV radiation (see Appendix B of Pinte et al. 2018a).

5.3 Results

Once the disc develops spiral structures, two different scenarios are followed. In the first the simulation is continued as normal to show how the disc would evolve in the absence of any planets. In the second, a planet is embedded. This is done for all combinations of a $1M_{\mathrm{Sat}}$, $1M_{\mathrm{Jup}}$, and a $3M_{\mathrm{Jup}}$ planet in a $0.08M_{\odot}$, $0.1M_{\odot}$, and a $0.25M_{\odot}$ disc, totalling 9 simulations with planets. In all cases, the planet starts in the outer gravitationally unstable parts of the disc and migrates rapidly inwards. As the planet migrates, the spiral wakes generated by the planet begin influencing the global disc properties (Goodman & Rafikov, 2001; Rafikov, 2016; Ziampras et al., 2020). The final fate of the migrating planets and their impact on the disc structure depend on both the planet's mass and the strength of gravitational instabilities (i.e. the disc mass).

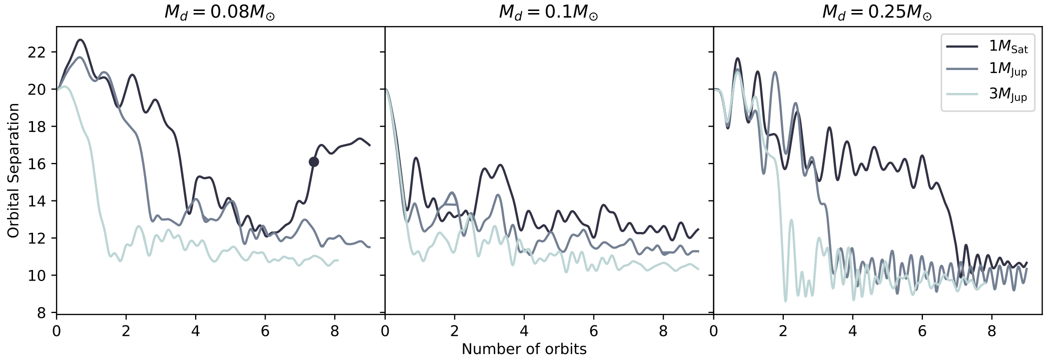


Figure 5.2: The migration tracks of a $1M_{\rm Sat}$, $1M_{\rm Jup}$, and a $3M_{\rm Jup}$ in a $0.08M_\odot$ (left), $0.1M_\odot$ (middle), and $0.25M_\odot$ (right) disc (as shown in Figure 5.1). The planets are able to slow their migration in the $0.08M_\odot$ and $0.1M_\odot$ disc as the inner regions are gravitationally stable. However, the planets are unable to slow down in the $0.25M_\odot$ disc as it is completely gravitationally unstable. The dot in the left panel marks the location of a $1M_{\rm Sat}$ planet as it is undergoing a brief period of outward migration in a $0.08M_\odot$ disc.

5.3.1 Planet Migration

In the $0.08M_\odot$ and $0.1M_\odot$ discs, the variable β model results in a disc with a gravitationally stable inner region as shown in Rowther & Meru (2020). Whereas, the $0.25M_\odot$ disc is massive enough to still be gravitationally unstable everywhere despite the variable β model. The migration of a $1M_{\rm Sat}$, $1M_{\rm Jup}$, and a $3M_{\rm Jup}$ planet is shown in Figure 5.2 in each of the three discs. Fig 5.2 demonstrates how the strength of gravitational instabilities, and thus the mass of the disc (increasing from left to right), determines the fate of the planet.

In the left and middle panels (the $0.08M_\odot$ and $0.1M_\odot$ disc), the migration of each planet slows down when it reaches the stable inner disc, but at different locations. The lighter the planet, the further out the planet slows down. After which, the planet continues to migrate inwards at a slower pace, consistent with the results in Rowther & Meru (2020). However, as the entire $0.25M_\odot$ disc is gravitationally unstable, the migration of the planets (right panel) is much more similar with simulations using a constant β where the planet rapidly migrates inwards towards the inner boundary of the disc (Baruteau et al., 2011; Malik et al., 2015; Rowther & Meru, 2020). After which, migration ceased.

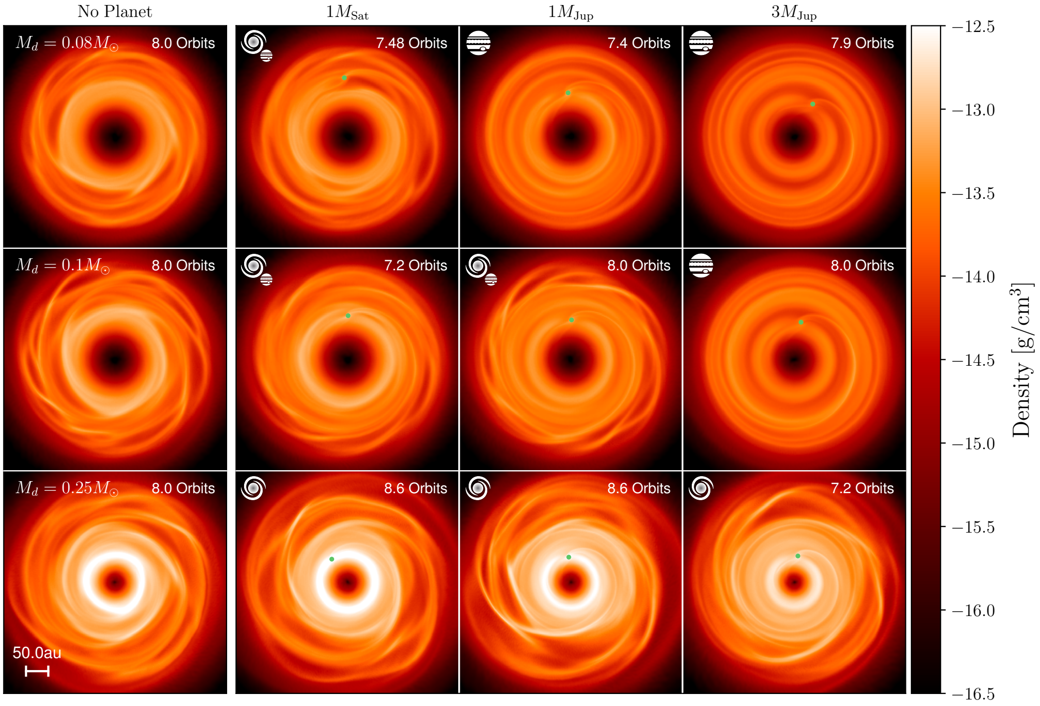


Figure 5.3: Cross section slices of density in the $z = 0$ plane of various disc and planet masses. From top to bottom, the disc masses are $0.08M_\odot$, $0.1M_\odot$, and $0.25M_\odot$. The left-most, detached column shows how the disc evolves without a planet. The rest from left to right contain a $1M_{\mathrm{Sat}}$, $1M_{\mathrm{Jup}}$, and a $3M_{\mathrm{Jup}}$ planet. These show that if a planet is massive enough it can completely suppress spiral structures due to GI, e.g. a $0.08M_\odot$ or $0.1M_\odot$ disc with a $3M_{\mathrm{Jup}}$ planet. If its mass is too low, the disc remains gravitationally unstable, e.g. with any planet in a $0.25M_\odot$ disc. When the planet mass lies somewhere between those two extremes, spiral structures due to GI are weakened, e.g. a $0.1M_\odot$ disc with a $1M_{\mathrm{Jup}}$ planet.

5.3.2 Gravitational Instability vs Planet's Spiral Wakes

Figures 5.3 and 5.4 show the density and velocity divergence $\nabla \cdot \boldsymbol{v}$ rendered plots for all the simulations in this Chapter. In both figures, each row corresponds to a different disc mass. From top to bottom, the disc masses are $0.08M_\odot$, $0.1M_\odot$, and $0.25M_\odot$. Similarly, each column represents a different planet mass. The left-most detached column shows how the disc would have evolved in the absence of a planet. The rest from left to right contain a $1M_{\mathrm{Sat}}$, $1M_{\mathrm{Jup}}$, and a $3M_{\mathrm{Jup}}$ planet. The time at which the simulation is shown is chosen such that planets are on the North side of the disc, hence the times are not identical. Having the planet at roughly the same location makes for easier comparisons for mock observations in Figures 5.7 and 5.8,

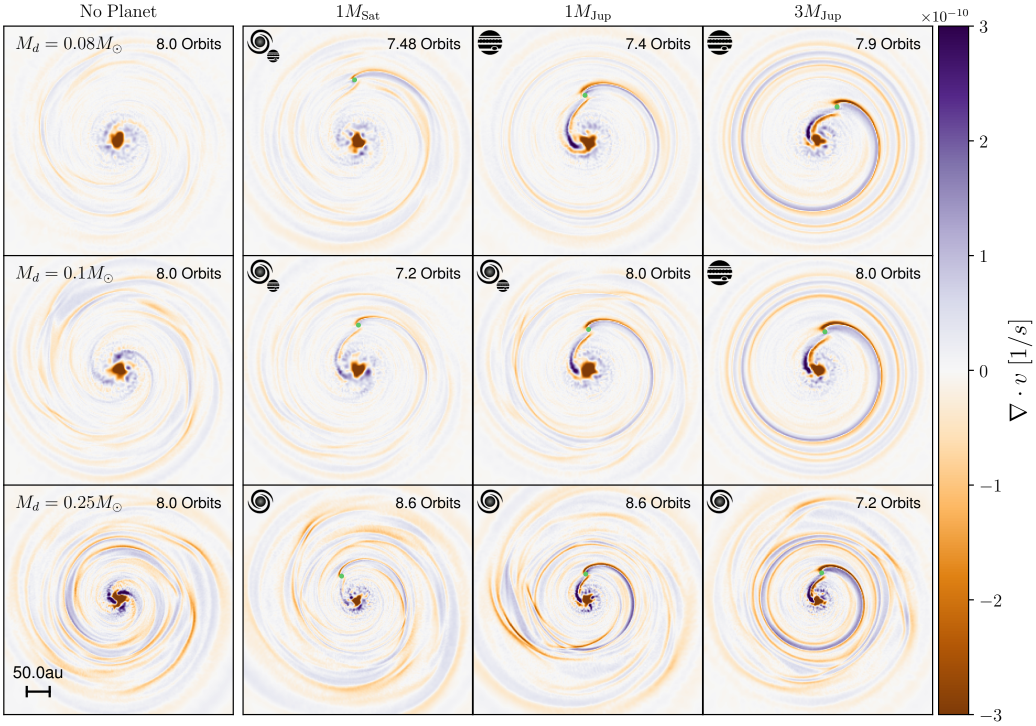


Figure 5.4: Cross section slices of velocity divergence $\nabla \cdot \boldsymbol{v}$ in the $z = 0$ plane of various disc and planet masses. The simulations are arranged as in Figure 5.3. When the magnitude of $\nabla \cdot \boldsymbol{v}$ due to the planet's spiral wake is greater relative to the $\nabla \cdot \boldsymbol{v}$ due to GI, the evolution of the disc is driven by the planet resulting in a gravitationally stable disc with ring & gap structure, e.g. a $0.08M_\odot$ or $0.1M_\odot$ disc with a $3M_{\rm Jup}$ planet. When the magnitude of $\nabla \cdot \boldsymbol{v}$ due to the planet and GI is comparable, GI structures are weakened, but not suppressed, e.g. a $0.1M_\odot$ disc with a $1M_{\rm Jup}$ planet. If the magnitude of $\nabla \cdot \boldsymbol{v}$ due to GI is dominant, then GI structures are unaffected, e.g. with any planet in a $0.25M_\odot$ disc.

but the general results described here are applicable even at the same simulation time. Similarly to Rowther et al. (2022), we use the $\nabla \cdot \boldsymbol{v}$ plots to understand how the spiral wakes of the planet are responsible for heating up the disc as from eq 5.3, changes in $\nabla \cdot \boldsymbol{v}$ directly contribute to the energy equation, and thus the temperature of the disc.

From the various disc and planet masses explored in this Chapter, we find that the planet's impact on the spiral structures due to GI can be quite different. For sufficiently massive discs, the gravitational instabilities are too strong to be suppressed by the planet. This is the case with the $0.25M_\odot$ disc in the bottom row of Fig 5.3. Likewise, if the planet is massive enough, it can alter the evolution of

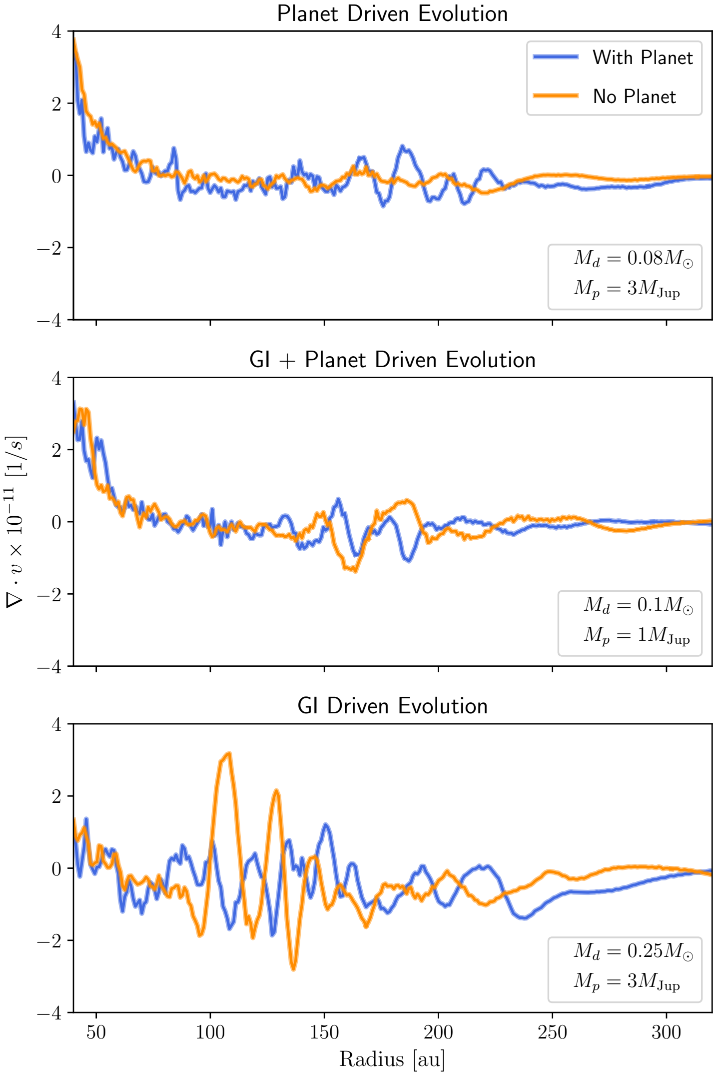


Figure 5.5: Azimuthally averaged $\nabla \cdot v$ plots showing the impact of a planet on representative simulations from Figure 5.4 that highlight the three scenarios that may occur. In this view, spiral structures in the disc are represented by radial variations in $\nabla \cdot v$. In the top panel, the spiral wakes generated by a $3M_\mathrm{Jup}$ planet in a $0.08M_\odot$ disc dominate over GI, thus controlling the evolution of the disc. In the middle panel, the magnitude of $\nabla \cdot \boldsymbol{v}$ due to GI is comparable to a $1M_\mathrm{Jup}$ planet in a $0.1M_\odot$ disc. Hence, both the planet and GI play a role in the evolution of the disc. In the bottom panel, a $3M_\mathrm{Jup}$ planet in a $0.25M_\odot$ disc has a lesser impact on $\nabla \cdot \boldsymbol{v}$ compared to GI. Thus the evolution of the disc is unchanged.

the disc by heating it up enough to push it into the gravitationally stable regime, and thus suppressing spiral structures (Rowther et al., 2020). The suppression of

GI is most apparent with the $3M_{\rm Jup}$ planet in the $0.08M_\odot$ and $0.1M_\odot$ discs. An in-between scenario is also seen where the planet mass is just large enough to weaken GI structures, but not enough to completely suppress it. The weakening of GI can be seen with both $1M_{\rm Sat}$ and $1M_{\rm Jup}$ planets in a $0.1M_\odot$ disc in the middle row of Fig 5.3.

By expanding our parameter space beyond Rowther et al. (2020), we can explore in more detail how the spiral wakes generated by the planet interacts with the gravitational instabilities present in the disc, seen in Figure 4. In general, the magnitude of $\nabla\cdot\boldsymbol{v}$ due to GI is larger with increasing disc mass. This is expected as more massive discs are expected to be more gravitationally unstable. Similarly, $\nabla\cdot\boldsymbol{v}$ due to the spiral wakes generated by the planet is also larger with increasing planet mass which is consistent with expectations of larger planets being more easily able to influence the disc's evolution (Kley & Nelson, 2012). However, the impact of the planet on $\nabla\cdot\boldsymbol{v}$ is broadly similar regardless of disc mass.

Figure 5.5 shows the azimuthally averaged $\nabla\cdot\boldsymbol{v}$ of a subset of the simulations in Fig 5.4. In each panel the blue lines represent a simulation with a planet that is representative of one of the aforementioned scenarios, while the orange lines are for discs that continued to evolve without a planet. They highlight the increasing dominating effect of GI on the magnitude of $\nabla\cdot\boldsymbol{v}$ as the disc mass increases. Figures 5.4 and 5.5 now reveal why none of the planets have a significant impact on the spiral structures due to GI in the most massive disc. The magnitude of $\nabla\cdot\boldsymbol{v}$ due to GI is larger comparable to $\nabla\cdot\boldsymbol{v}$ due to the planet's wake. As a result, the planet does not contribute much to the energy equation (Eq. 5.3) and the disc remains gravitationally unstable. Only the $3M_{\rm Jup}$ planet has a minor impact on the disc structure. However, the planet's wake is restricted to the inner disc. Thus, the outer disc remains mostly unaffected. Similarly, Figs 5.4 and 5.5 also reveal that gravitational instabilities are completely suppressed when $\nabla\cdot\boldsymbol{v}$ due to the planet's wake is dominant, as shown in the simulations with a $3M_{\rm Jup}$ planet in a $0.08M_\odot$ and $0.1M_\odot$ discs, and a $1M_{\rm Jup}$ planet in the $0.08M_\odot$ disc. In the in-between scenario where GI is weakened, this occurs when both GI and the planet have a comparable influence on $\nabla\cdot\boldsymbol{v}$ as seen when there is a $1M_{\rm Sat}$ or $1M_{\rm Jup}$ planet in the $0.1M_\odot$ disc, and a $1M_{\rm Sat}$ planet in the $0.08M_\odot$ disc. The Toomre Q parameter gives a measure of how gravitationally unstable a disc is, and is defined by (Toomre, 1964)

$$Q = \frac{c_s\Omega}{\pi G\Sigma}. \tag{5.6}$$

To best highlight the three scenarios that may occur, the azimuthally averaged

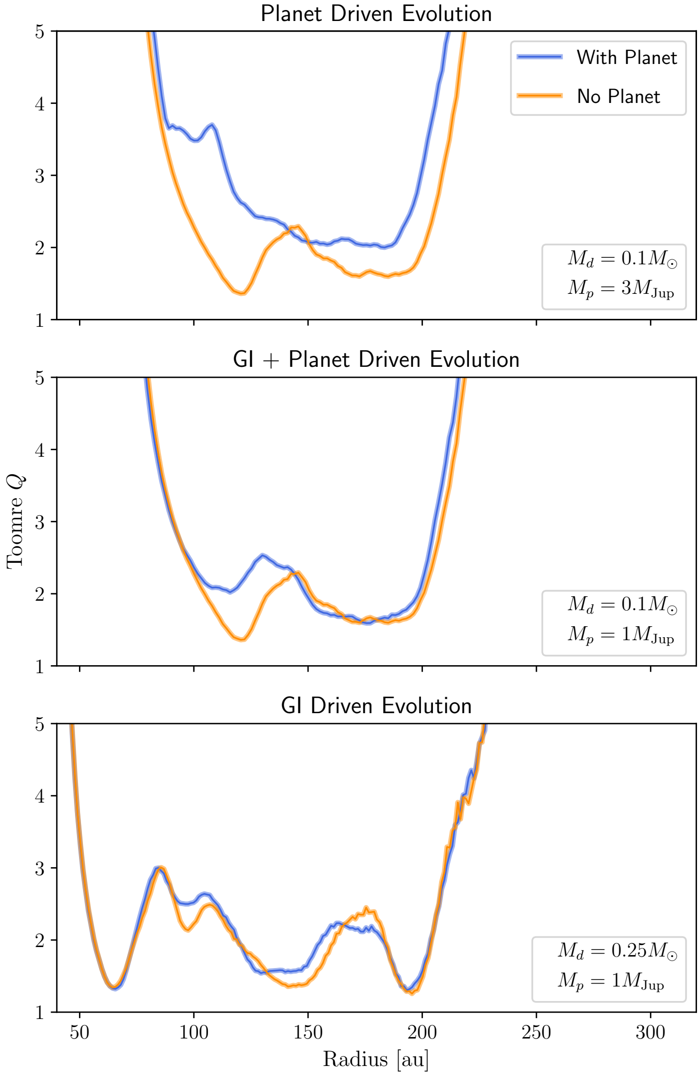


Figure 5.6: Azimuthally averaged Toomre Q plots after two orbits showing the impact of a planet on three representative simulations that highlight the three scenarios that may occur. In the top panel, a $3M_{\text{Jup}}$ planet in a $0.1M_\odot$ disc increases Q globally, thus controlling the evolution of the disc and causing it to become gravitationally stable. In the middle panel, a $1M_{\text{Jup}}$ planet in a $0.1M_\odot$ disc is unable to affect Q in the outermost regions of the disc, but has a minor impact on the inner regions. Hence, both the planet and GI play a role in the evolution of the disc. In the bottom panel, a $1M_{\text{Jup}}$ planet in a $0.25M_\odot$ disc has little impact on Q, thus the evolution of the disc is unchanged.

Toomre Q plots in Figure 5.6 are shown two orbits after the planet has been embedded for three representative simulations. To summarise;

i) **Planet Driven Evolution** 🪐 – The spiral wakes generated by a $3M_{\mathrm{Jup}}$ planet in a $0.1M_{\odot}$ or $0.08M_{\odot}$ have a large effect on $\nabla \cdot \boldsymbol{v}$ causing the disc to heat up as seen by the global increase in the Q profile. Thus, GI structures are completely suppressed as the disc is pushed into the gravitationally stable regime.

ii) **GI + Planet Driven Evolution** 🌀🪐 – The spiral wakes of a $1M_{\mathrm{Jup}}$ planet in a $0.1M_{\odot}$ disc are less extended compared to the $3M_{\mathrm{Jup}}$ planet. As the spiral wakes do not extend far out, the the Q profile of outer parts of the disc remain in the gravitationally unstable regime. Thus the simulations still show GI structures in the outer regions.

iii) **GI Driven Evolution** 🌀 – The spiral wakes generated by any of the three planets in a $0.25M_{\odot}$ disc are unable to have a major impact on $\nabla \cdot \boldsymbol{v}$, and thus they have very little effect on the Q profile. Hence the disc remains just as gravitationally unstable as if it had evolved without a planet.

5.3.3 Mock ALMA observations

The top half of Figure 5.7 shows mock observations for all simulations in this Chapter at the same times as shown in Figures 5.3 and 5.4 at 1.3mm with an integration time of 60 minutes using the C-7 configuration. The bottom half highlights non-axisymmetric features by plotting the residuals of subtracting the mock observations from an axisymmetric model. From top to bottom, the disc masses are $0.08M_{\odot}$, $0.1M_{\odot}$, and $0.25M_{\odot}$. The left-most detached column shows how the disc would be observed in the absence of a planet. The rest from left to right show how the disc would appear if it contained a $1M_{\mathrm{Sat}}$, $1M_{\mathrm{Jup}}$, and a $3M_{\mathrm{Jup}}$ planet.

From Figure 5.3, in a $0.25M_{\odot}$ disc the planets do not have much, if any, impact on gravitational instabilities in the disc. This is reflected in the mock observations in Figure 5.7, where both the flux and residual plots are similar regardless of whether a planet is present in the disc.

In contrast, Figure 5.3 shows that the $0.08M_{\odot}$ and $0.1M_{\odot}$ discs are much more affected by the presence of a planet. If the planet is massive enough and drives the disc's evolution, as is the case with a $3M_{\mathrm{Jup}}$ planet in a $0.08M_{\odot}$ and $0.1M_{\odot}$ discs, ring & gap structure is seen in the mock observations. The weakening of GI by the lower mass planets is also reflected in the mock observations where evidence of GI in the form of large-scale spiral structures is far less clear. The lack of obvious GI structures is highlighted by the plots in the left-most detached column which show that in the absence of a planet, evidence of GI is apparent in the mock observations, and especially the residuals in Fig 5.7.

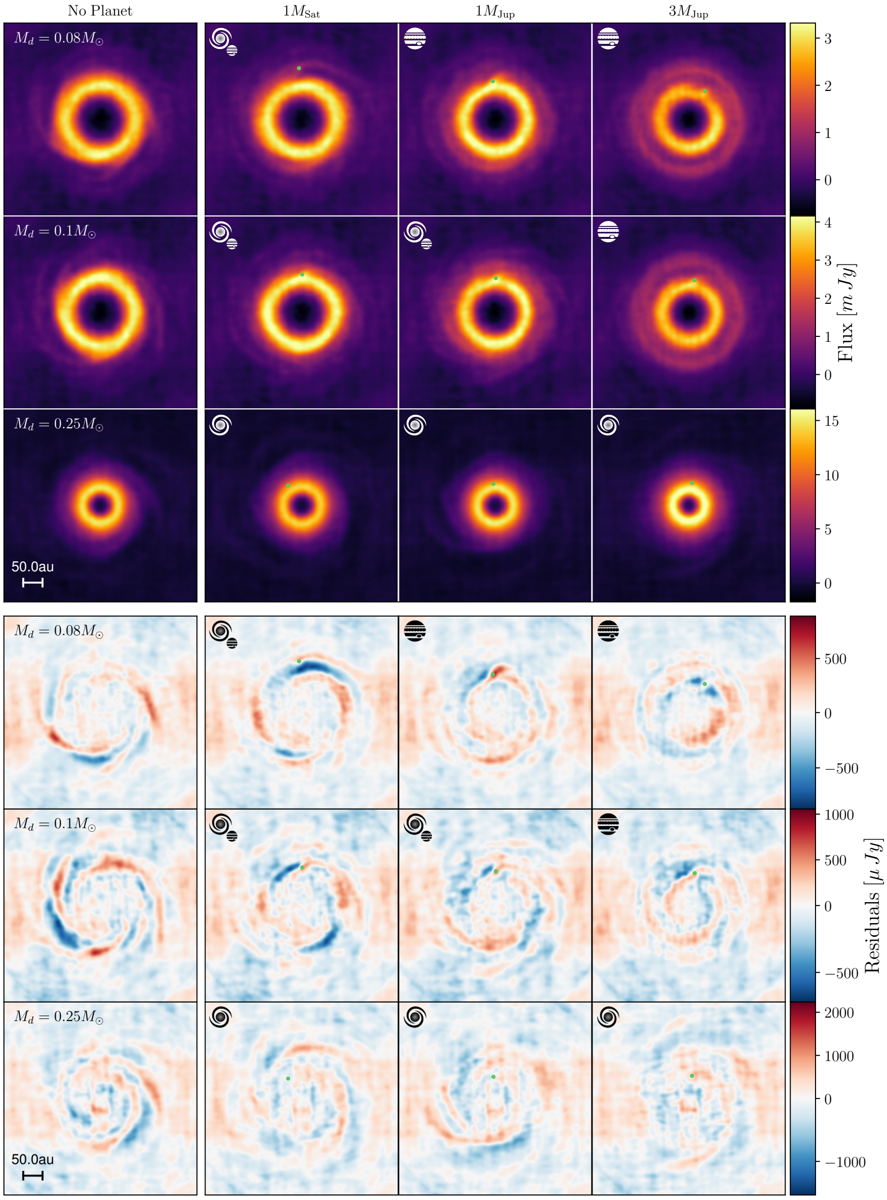


Figure 5.7: Upper panel: mock observations at 1.3mm using the C43-7 configuration with an integration time of 60 minutes. Lower panel: the residuals of subtracting the mock observations from an axisymmetric model which highlights non-axisymmetric features. The simulations are arranged as in Figure 5.3. The planet can completely suppress spiral structures if it's massive enough, or have no affect on GI if its mass is too low, or weaken GI and making it harder to observe if the planet mass lies between the two extremes.

The planet leaves a few hints of its presence in Fig 5.7. For clarity, this will be described using the mock observations of a $3M_{\text{Jup}}$ planet in either the $0.08M_{\odot}$ or $0.1M_{\odot}$ disc as the planet dominates over GI, where all the residual features can be confidently attributed to the planet. The largest of the features in the residuals show that the rings are asymmetric in brightness, highlighted by the red regions in the vicinity of the planet. Comparison with Fig 5.3 reveals that this asymmetry can be attributed to overdense regions resulting from the spiral wakes of the planet. As mentioned earlier, the gravitationally stable inner disc slows down planet migration in the $0.08M_{\odot}$ or $0.1M_{\odot}$ discs. The inward migration is due to a large negative coorbital torque resulting from an underdense region in front of the planet (Rowther & Meru, 2020; Malik et al., 2015; Baruteau et al., 2011), which is revealed in the residuals as a blue (underdense) region in front of the planet. The opposite is seen for an outwardly migrating $1M_{\text{Sat}}$ planet in a $0.08M_{\odot}$ disc where the blue (underdense) region is behind the planet. The time at which the mock observation is created is marked in the migration track in Figure 5.2 highlighting the outward migration. The planet also appears as a bright spot in the mock continuum data. It is likely that this bright spot is artificially enhanced by our choice of accretion radius, see §5.4.2.

5.3.4 Impact on the Kinematics

The gas channel maps can also be used to detect GI or the presence of giant planets (Pinte et al., 2018b, 2019, 2020; Hall et al., 2020). Figure 5.8 shows the synthetic channel maps at $\Delta v = v_{\text{obs}} - v_{\text{systemic}} = 0$ km/s for all the simulations in this Chapter. In real observations, the channel maps would be affected by CO freeze-out, and photodissociation and photodesorption in regions of high UV radiation. However, the channel maps in Fig 5.8 are created excluding these effects for easier comparison with Hall et al. (2020) and Terry et al. (2022). Additionally, these have not been convolved with a beam size. Both are discounted for clarity in comparing the kinematic signatures of GI and a planet.

Primary Kink

In the case of a planet, its presence can be detected by a localised kink near the location of the planet (Pinte et al., 2018b, 2019, 2020). This is most apparent for a $3M_{\text{Jup}}$ planet in a $0.1M_{\odot}$ or $0.08M_{\odot}$ disc where a kink is visible near the vicinity of the planet and is not seen at all radii. A weaker kink is also visible for a $1M_{\text{Jup}}$ planet in both the $0.1M_{\odot}$ and $0.08M_{\odot}$ discs, but this is likely to be detectable only

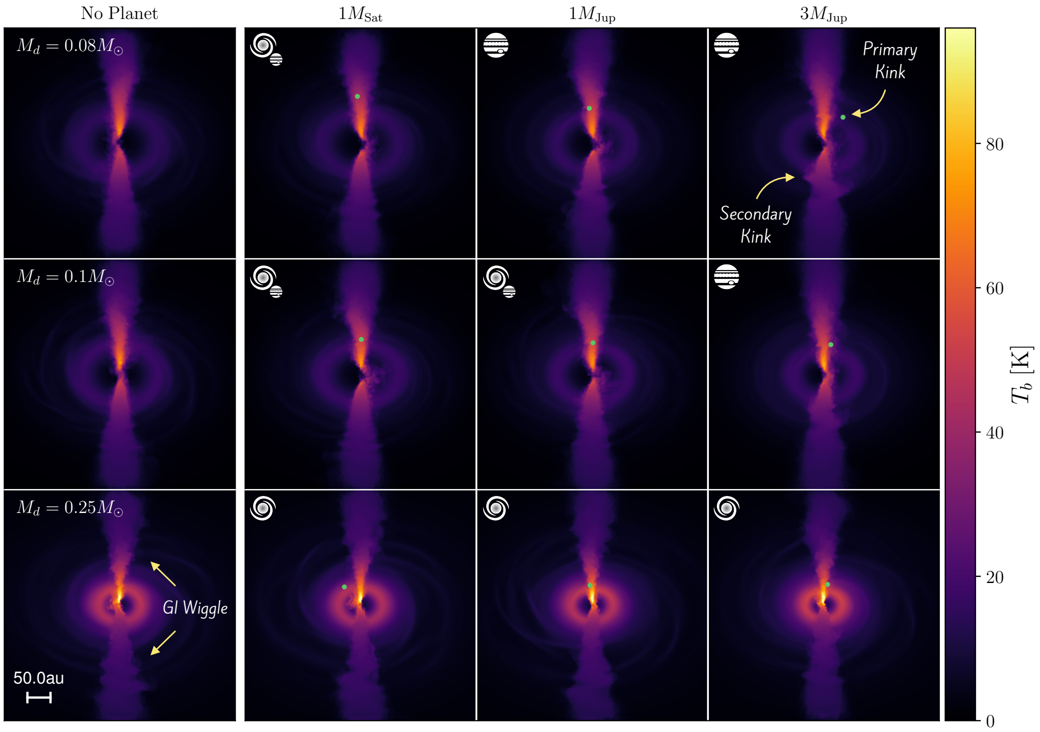


Figure 5.8: Synthetic channel maps (J = 3 – 2 transitions) for ^{13}C^{16}O at $\Delta v = 0$ km s^{-1} from the systemic velocity of the disk. The simulations are arranged as in Figure 5.3. When the planet is able to open up a gap in the disc, a primary kink is seen at the planet's location along with a secondary kink on the opposite side due to the planet's spiral wake. GI wiggles are only clearly seen for the $0.25M_\odot$ disc.

in synthetic observations which are not affected by noise and not smoothed by the beam size. However, if the planet's influence does not dominate over GI, the planet is undetectable from the kinematics as evident by the lack of any localised kink in the $0.25M_\odot$ disc.

Secondary Kink

The presence of a planet in the kinematics can be detected by any extended features caused by the spiral wake of the planet (Bollati et al., 2021; Calcino et al., 2021; Verrios et al., 2022). In Fig 5.8, this extended feature is visible as a secondary kink on the opposite side of the planet. The imprint of the planet's spiral wake is most easily seen for a 1 and $3M_{\rm Jup}$ in a $0.08M_\odot$ disc, and a $3M_{\rm Jup}$ in a $0.1M_\odot$ disc. The imprint left behind by the wake of the planet as seen in these simulations have also been detected in the disc around HD 163296 (Calcino et al., 2021).

GI Wiggle

In the case of GI, its presence can be detected in the form of *'GI wiggles'* (Hall et al., 2020). The global nature of GI results in these wiggles being non-localised appearing at all velocities and radii in the disc. In Fig 5.8, a clear GI wiggle is only seen for a $0.25M_{\odot}$ disc where strong gravitational instabilities are present. The amplitude of the wiggle is larger further out in the disc. Since β decreases with radius, this behaviour is consistent with Longarini et al. (2021) which showed the amplitude of the wiggle is stronger for lower values of β. As mentioned earlier, when GI is stronger compared to the planet's spiral wake, the disc structure is unaffected. This is reflected in the kinematics by the lack of any features associated with the planet visible in the $0.25M_{\odot}$ disc.

Despite being gravitationally unstable and harbouring large scale spiral features, the $0.08M_{\odot}$ and $0.1M_{\odot}$ discs only show tentative GI wiggles, which is likely to be hidden when observational effects (such as a beam size) are included. However, this is expected from previous work which showed GI wiggles are weaker for lower disc masses (Longarini et al., 2021; Terry et al., 2022).

Including CO Freeze-out, Photodissociation and Photodesorption

To determine if the kinematic features due to a planet or GI are impacted by observational effects, Figure 5.9 compares the channel maps with (right panels) and without (left panels) the inclusion of CO freeze-out below 20K, and photodissociation and photodesorption in regions of high UV radiation. The comparison is only presented for two simulations; a $3M_{\mathrm{Jup}}$ planet in a $0.1M_{\odot}$ disc and a $0.25M_{\odot}$ disc without a planet. The former is chosen to investigate the observability of the kinematic signatures due to a planet which shows that both the primary and secondary kink are unaffected. The latter is chosen to determine observability of GI structures in the channel maps which show that the GI wiggles are far less visible due to the inclusion of the aforementioned observation effects.

5.4 Discussion

5.4.1 Observational Implications

Simulations of star cluster formation (Bate, 2018) show that discs are expected to be massive and potentially gravitationally unstable in the earliest stages of their lives. A characteristic feature of such massive discs are their large-scale spiral arms. Despite the increasing number of resolved protoplanetary discs, these type of discs

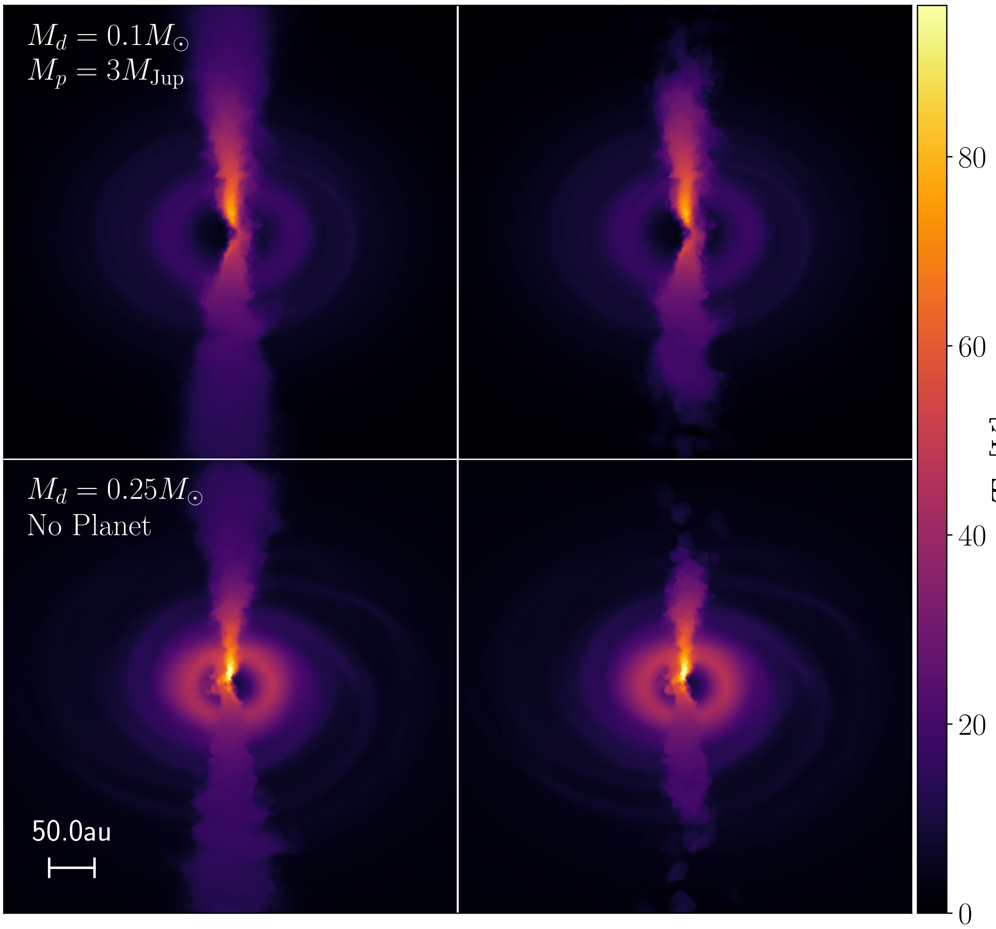


Figure 5.9: Synthetic channel maps (J = 3 − 2 transitions) for $^{13}C^{16}O$ at $\Delta v = 0$ km s^{-1} from the systemic velocity of the disk. The top panels are for a $3M_{\rm Jup}$ in a $0.1M_\odot$ disc, while the bottom panels are for a $0.25M_\odot$ disc without a planet. The right panels includes CO freeze-out below 20K, and photodissociation and photodesorption in regions of high UV radiation, while the left panels do not. Inclusion of these observational effects do not affect the visibility of the kinematic features of the planet. However, as GI is strongest in the outer colder regions of the disc, they are more affected and are less visible.

remain quite rare in observations (Pérez et al., 2016; Huang et al., 2018a). Additionally, more complex methods have measured disc masses to be higher than previous estimates using a constant fixed dust-to-gas mass ratio to infer the total mass of the disc. Using the rare optically thinner $^{13}C^{17}O$, Booth & Ilee (2020) measured the mass of HL Tau to be large enough to enter the gravitationally unstable regime. Utilising the aerodynamic properties of dust grains Powell et al. (2019) measured new disc masses to be 9-27% of their stellar hosts. Such high masses push the discs close to the gravitationally unstable regime where the disc would be expected to show evidence of gravitational instabilities, but instead they show axisymmetric ring & gap structures. Our results show that weak and moderate gravitational instabilities can be suppressed by planet-disc interactions, thus highlighting that massive discs cannot be simply dismissed by the lack of large-scale spiral structures in the disc midplane. Veronesi et al. (2021) showed that the self-gravity of the Elias 2-27 protoplanetary disc can have a noticeable impact on its rotation curve. Hence, our results provide additional motivation for obtaining high resolution molecular data for axisymmetric discs so that dynamical mass estimates can be calculated to determine if their spiral structures due to GI are hidden by planets or other processes.

5.4.2 Observability of Continuum and Kinematic Features

The results presented here have made a few assumptions and simplifications which affect the observability of the continuum and kinematic substructures described in §5.3.

The bright spot at the planet's location in our synthetic observations is most likely to be impacted due to the simplification of using an accretion radius of 0.001 code units. We choose a small accretion radius to ensure the planet does not accrete too much material. If the planet was allowed to freely accrete, it would quickly reach the threshold required to open up a gap at which point all the planets would behave identically, which defeats the purpose of investigating the impact of different planet masses. The consequence of such a tiny accretion radius is a lot of gas swirling around the planet. Although a circumplanetary disc is an expected physical feature, resolving it is not the purpose of this study and it is unlikely that we do so in these simulations.

In this Chapter, the dust has been assumed to be perfectly coupled to the gas. Consequently, any asymmetric structure due to the planet present in the simulation is also evident in the mock continuum observation. If the dust is decoupled from the gas, the asymmetric features are less likely to be observable (Dipierro et al., 2015) as the dust will be able to form substructures that do not trace the gas substructures perfectly. However, given the high masses of the disc involved the Stokes numbers of millimeter sized grain is smaller than one and thus are more likely to be coupled to the gas. Nevertheless, simulations with a gas and dust mixture are required to test the validity of this assumption further. Additionally, the observability of these substructures also depend on ALMA configuration and integration time used. Recently, Speedie et al. (2022) also found that planet driven spirals could be observable with ALMA with longer integration times. This is also true for the observability of GI structures when they are weakened by the planet like with a $1M_{\rm Sat}$ or $1M_{\rm Jup}$ planet in the $0.1M_{\odot}$ disc.

The cavity seen in all simulations is due to the inner-most regions of the disc not being resolved. Consequently, the inner boundary of the disc shows up as a ring in the contiuum. Higher resolution and more realistic thermodynamics could both resolve these issues.

In §5.3.4 it was shown that the inclusion of CO freeze-out, and photodissociation and photodesorption had more of an effect on the kinematic signatures of GI rather than a planet. Gravitational instabilities are expected to be strongest in the outer colder regions of the disc. Hence, their signature GI wiggle is more likely to be affected by CO freeze-out. This is seen in the bottom right panel of Fig 5.9 by the lack of emission (and thus wiggles) in the outer parts of the disc. However, as the planet is located in the inner warmer parts of the disc, the kinematic features associated with the planet are still visible with the extra physics included.

5.4.3 Cooling prescription

A caveat to the simulations presented in this study is the cooling prescription used. Although it is able to mimic a realistic gravitationally unstable disc, i.e. one that only has GI structures in the outer regions of the disc, it does not account for the evolution of the disc. Using radiative transfer simulations, Mercer et al. (2018) calculated an effective β which was found to vary both spatially and with time. While this does not change our conclusions qualitatively, the long term evolution of planet-disc interactions in a GI disc requires studies with more realistic thermodynamics.

5.5 Conclusions

We perform 3D SPH simulations to investigate the impact of planet-disc interactions on the structure of a gravitationally unstable disc. Our work shows that the more gravitationally unstable a disc is, the harder it is for a planet to influence the disc's evolution. However, the planet can affect the disc's evolution when the strength of the planet's spiral wake is comparable to or larger than gravitational instabilities. When the planet is massive enough, it is able to open up a gap in the disc.

The mock ALMA observations of the continuum show that the observability of GI structures is either diminished or completely suppressed when the planet is massive enough to impact the disc structure. The analysis of the kinematics reveals that if either the kinematic signature of a planet or GI is present, then the other is not detected. If the planet drives the disc's evolution, then the planet's imprint is visible in the kinematics. The GI wiggle is only clearly visible for the highest mass disc, and as the planet is not massive enough to impact the disc structure, the GI wiggles are not affected by the presence of the planet.

Our results show that it is possible for high mass discs to appear axisymmetric, and lacking large-scale spiral structures in the presence of a giant planet. The implications on the origins of disc substructures, planet formation, and disc evolution necessitate these findings to be taken into account when considering the many possible discrepancies between the disc structure and inferred mass in observations.

6

Warping Away Gravitational Instabilities in Protoplanetary Discs

6.1 Introduction

In recent years, a large number of protoplanetary discs have been observed at millimeter wavelengths with ALMA. Most of these discs are axisymmetric, with some also containing rings and gaps (ALMA Partnership et al., 2015; Andrews et al., 2016; Fedele et al., 2018; Andrews et al., 2018; Huang et al., 2018a; Dipierro et al., 2018; Booth & Ilee, 2020), even though some of them are quite young (< 1Myr). Ring and gap structures have also been observed in even younger ($\lesssim 0.5$ Myrs) Class 1 discs (Sheehan & Eisner, 2018; Segura-Cox et al., 2020). Young discs are thought to be massive and could potentially be gravitationally unstable. A characteristic feature of such discs are large scale spiral features. There is evidence that discs with spiral arms in the midplane exist (Pérez et al., 2016; Huang et al., 2018b), although they seem to be quite rare.

Does the lack of observed large scale spiral structures imply that young discs are not as massive as expected? Or can signatures of gravitational instabilities be hidden? It is reasonable to assume that gravitationally unstable discs do not evolve in isolation. The physical processes that are often used to explain observed substructures such as rings and gaps, along with the mechanisms that can warp a disc will also influence the evolution of young self-gravitating discs. A common explanation for rings and gaps are planet-disc interactions which can also suppress spiral structures in gravitationally unstable discs (Rowther et al., 2020).

There are several mechanisms that can warp a protoplanetary disc. These include a misaligned internal (planetary or stellar) companion, a misaligned unbound stellar companion (flyby), or misaligned infalling material from chaotic accretion episodes. The rate at which these discs are observed also suggests that warps and misalignments are common in protoplanetary discs (Benisty et al., 2017; Walsh et al., 2017; Casassus et al., 2018; Ballabio et al., 2021). The latter two mechanisms are more relevant for gravitationally unstable discs. Flybys are much more common earlier in a disc's lifetime and become less frequent over time (Pfalzner, 2013; Vincke & Pfalzner, 2016; Bate, 2018). Misaligned infall is also expected to occur very early in a disc's lifetime when there is still plenty of material around the protostar (Bate, 2018; Sakai et al., 2019). Both of these mechanisms are more likely to occur precisely when discs are more likely to be gravitationally unstable and can alter the evolution of the disc.

There have been some simulations of flybys interacting with a gravitationally unstable disc, but in these studies the flybys were coplanar or head-on. Additionally, the distance of closest approach occurred inside the disc causing major disruption to the disc structure and completely suppressing gravitational instabilities (Lodato et al., 2007; Forgan & Rice, 2009). These works demonstrate that a flyby altering the disc structure can affect the heating and thus the self-gravitating structures. However, they do not consider warps or misalignments. Thies et al. (2010) on the other hand did study the impact of inclined flybys and found that stellar encounters induced fragmentation. However, their disc masses were much more massive and hence more prone to fragmentation.

In this Chapter we use three-dimensional global numerical simulations to consider the evolution of an isolated self-gravitating disc subjected to a warp. To our knowledge, this is the first study of a self-gravitating disc subjected to a warp. This Chapter is organised as follows. In §6.2 we recall the relevant warp theory and its context in self-gravitating discs. In §6.3 we describe the simulations presented in this Chapter. In §6.4 we present our results of the impact that an artificially introduced warp has on the structure and evolution of gravitationally unstable protoplanetary discs. We discuss our work in the context of its limitations and observations in §6.5. We conclude our work in §6.6.

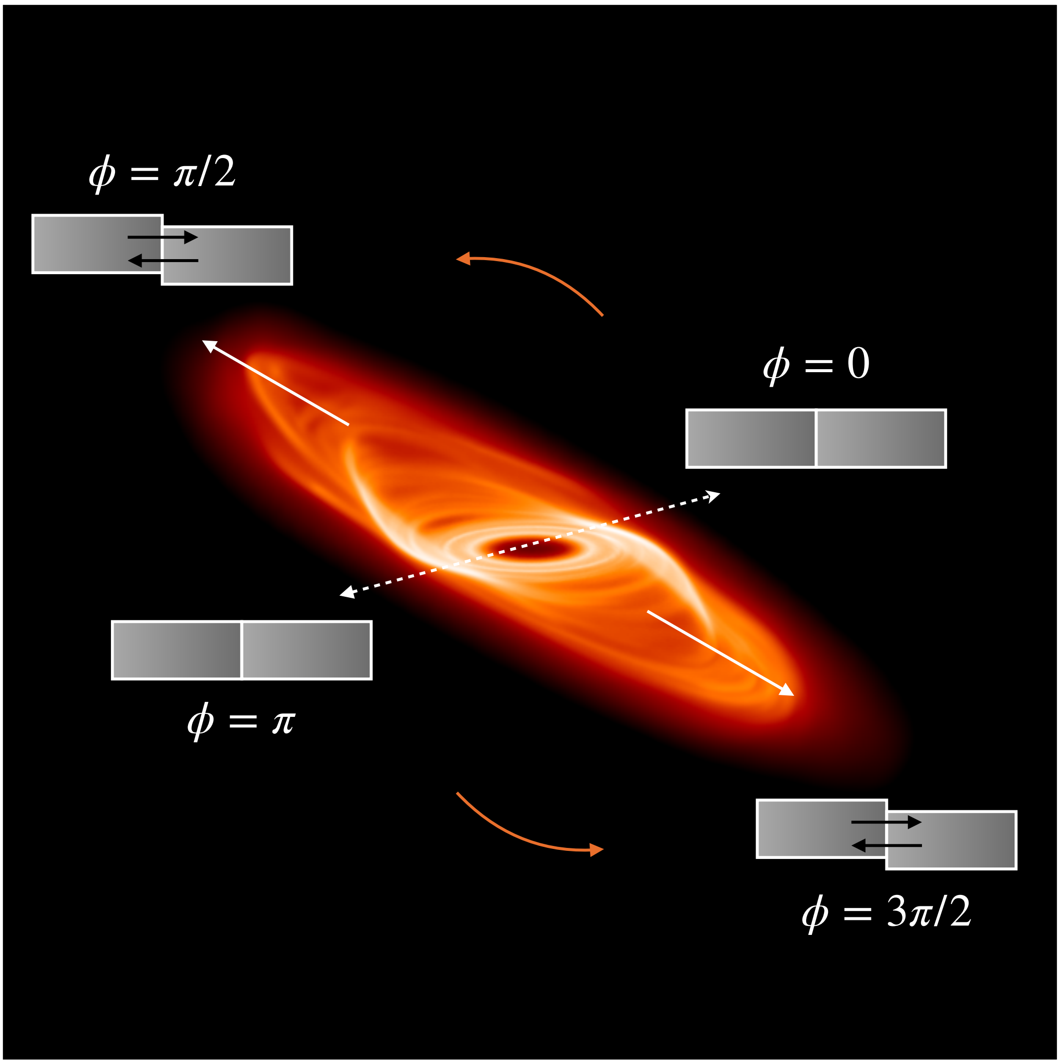


Figure 6.1: A 3D projection of the disc after it has been warped. The annotations (see Fig 10 of Lodato & Pringle, 2007) show the radial pressure gradients induced by the warp at different azimuths as a fluid element orbits the disc. The orange arrows show the rotation of the disc. The shaded grey boxes represent the pressure between adjacent annuli of gas. At $\phi = 0$ and $\phi = \pi$ the annuli are aligned, hence there is no pressure gradient. At all other azimuths, the heights of the annuli are offset resulting in an oscillating radial pressure gradient as the gas traverses an orbit. The direction of the resulting pressure gradient is shown by the black arrows. The colour scale represents the surface density of the disc with brighter colours correspond to regions of higher density.

6.2 Warp Theory

Warped discs are characterised by a radial dependence on the angular momentum vector. The steepness of the warp is described by the warp amplitude ψ which is given by (eg. Ogilvie, 1999; Lodato & Price, 2010)

$$\psi = R\left|\frac{\partial \boldsymbol{l}}{\partial R}\right|, \tag{6.1}$$

where $\boldsymbol{l}$ is the unit angular momentum vector, and R is the radial location spherical coordinates.

To demonstrate the effect that the warp has on the disc, in Figure 6.1 we show a 3D projection of a warped gravitationally unstable disc with annotations to illustrate how the warp propagates through the disc. Consider a fluid element as it orbits around the disc with a position given by ϕ. Here the disc may be described as concentric rings or annuli of gas and the warp causes adjacent rings to be vertically offset. The rotation of the disc is represented by the orange arrows. The shaded grey boxes represent the direction of the pressure between adjacent annuli of gas. At $\phi = 0$ and π, adjacent rings will be perfectly aligned with no variation in the vertical height z, hence at these locations there is no pressure gradient. At all other azimuths, adjacent rings will have variations in z, resulting in pressure gradients due to overpressure regions above or below the local midplane. The magnitude of the induced pressure gradient is given by (Lodato & Pringle, 2007)

$$\frac{\partial P}{\partial R} \sim \frac{\partial P}{\partial z}\psi \sim \frac{P\psi}{H}, \tag{6.2}$$

where P is the pressure, z is the vertical height, and H is disc scale height. The direction of the resulting pressure gradient, represented by the black arrows, changes based on whether the fluid element is flowing towards or away from misaligned regions. As it moves towards vertically offset regions, away from $\phi = 0$ and π, the resulting radial pressure gradient points inwards. The direction of the pressure gradient is reversed when flowing away from the vertically offset regions, towards $\phi = 0$ and π. Hence the fluid element feels an oscillating pressure gradient as it orbits in the warp which will drive the disc's evolution. This oscillating radial pressure gradient can trigger a strong response in the velocity flow in the disc. The response for example in the radial velocity field from Lodato & Pringle (2007)

$$v_R \propto \psi \cos\phi \tag{6.3}$$

depends on the warp amplitude of the disc and the azimuthal angle ϕ.

While there have been extensive analytic efforts studying warps in accretion discs (Hatchett et al., 1981; Papaloizou & Terquem, 1995; Ogilvie, 1999, 2000; Lodato & Price, 2010), simplifications are often required, or the regime studied is restricted to either wave-like or diffusive. However, in the work presented here where we consider warps in gravitationally unstable discs, which contains large scale spiral structures that cause shocks, $P\mathrm{d}V$ work is naturally important. Hence, if a warp can excite strong responses in the velocity flow, the spiral structures in the disc are expected to be impacted due to additional $P\mathrm{d}V$ heating. Furthermore, in gravitationally unstable discs, the disc viscosity can be comparable to the disc aspect ratio H/R, placing it in the intermediate regime where the warp propagation is neither fully wave-like or diffusive. Martin et al. (2019) (and later Dullemond et al. 2022) showed that in the intermediate regime, the warp propagation is wave-like in the outer regions of the disc, and diffusive in the inner regions. As gravitational instability is more relevant in the outer parts of the disc, we consider the warp propagation to be wave-like in this chapter.

6.3 Method

6.3.1 Hydrodynamic simulations & initial conditions

We use PHANTOM, a smoothed particle hydrodynamics (SPH) code developed by Price et al. (2018a) to perform the suite of simulations presented here.

The disc is modelled using 2 million particles between $R_{\mathrm{in}} = 3$ and $R_{\mathrm{out}} = 150$ in code units with a disc-to-star mass ratio of 0.1. The central star is modelled using a fixed external potential. The initial surface mass density is set as a smoothed power law and is given by

$$\Sigma = \Sigma_{\mathrm{in}} \left(\frac{R}{R_{\mathrm{in}}} \right)^{-1} f_s, \tag{6.4}$$

where Σ_{in} is the surface mass density at $R = R_{\mathrm{in}}$ and $f_s = 1 - \sqrt{R_{\mathrm{in}}/R}$ is the factor used to smooth the surface density at the inner boundary of the disc. The initial temperature profile is expressed as a power law

$$T = T_{\mathrm{in}} \left(\frac{R}{R_{\mathrm{in}}} \right)^{-0.5}, \tag{6.5}$$

where T_{in} is set such that the disc aspect ratio $H/R = 0.05$ at $R = R_{\mathrm{in}}$. The internal energy equation is

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -\frac{P}{\rho}(\nabla \cdot \boldsymbol{v}) + \Lambda_{\mathrm{shock}} - \frac{\Lambda_{\mathrm{cool}}}{\rho} \tag{6.6}$$

where we assume an adiabatic equation of state, and u is the specific internal energy. The first term on the RHS is the $P\mathrm{d}V$ work, Λ_{shock} is a heating term that is due to the artificial viscosity used to correctly deal with shock fronts, and

$$\Lambda_{\mathrm{cool}} = \frac{\rho u}{t_{\mathrm{cool}}} \tag{6.7}$$

controls the cooling in the disc. Here, the cooling time is modelled using a simple prescription such that it is proportional to the dynamic time by a constant factor β_{cool} (Gammie, 2001),

$$t_{\mathrm{cool}} = \beta_{\mathrm{cool}}\,\Omega^{-1}, \tag{6.8}$$

where Ω is the orbital frequency. Assuming the transfer of angular momentum is locally driven by gravitoturbulence (Gammie, 2001), β_{cool} can be related to the α viscosity by (Shakura & Sunyaev, 1973)

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma-1)}\frac{1}{\beta_{\mathrm{cool}}}. \tag{6.9}$$

Here $\beta_{\mathrm{cool}} = 15$, which using the above equation gives a theoretical $\alpha = 0.027$. To model shocks, we use an artificial viscosity switch that utilises the time derivative of the velocity divergence introduced by Cullen & Dehnen (2010). The artificial viscosity parameter α_v varies depending on the proximity to a shock. It takes a maximum of $\alpha_{\mathrm{max}} = 1$ close to the shock and a minimum of $\alpha_{\mathrm{min}} = 0$ far away. The artificial viscosity coefficient β_v is set to 2 (see Price et al. 2018a; Nealon et al. 2015).

6.3.2 Warping the disc

Before the disc is warped, it is evolved for 10 orbits to allow it to become gravitationally unstable and develop spiral structures. A warped disc can be described by two angles, its tilt β, and its twist γ. Using this, the unit angular momentum vector can be written as (Pringle, 1996)

$$\boldsymbol{l} = (\cos\gamma\sin\beta,\ \sin\gamma\sin\beta,\ \cos\beta), \tag{6.10}$$

where the disc is considered warped if $\boldsymbol{l}$ varies with radius. To introduce a warp in our simulation, the position of each particle is rotated such that the unit vector of the angular momentum is given by (Lodato & Price, 2010)

$$
\begin{aligned}
l_x &= \begin{cases} 0 & \text{for } R < R_1 \\ \frac{A}{2}\left[1 + \sin\left(\pi \frac{R-R_0}{R_2-R_1}\right)\right] & \text{for } R_1 < R < R_2 \\ A & \text{for } R > R_2 \end{cases} \\
l_y &= 0, \\
l_z &= \sqrt{1 - l_x^2},
\end{aligned}
\tag{6.11}
$$

where in our fiducial setup the warp is centered at $R_0 = 75$ and extends from $R_1 = 45$ to $R_2 = 105$ in code units. The initial misalignment between the outer and inner disc, $i = 30°$ defines $A = \sin(i)$. All discs in this work are initially untwisted since $l_y = 0$.

Figure 6.1 also helps to visualise how the warp profile affects the disc's evolution, as also described in equations 6.2 and 6.3. For a steeper warp, i.e. a larger warp amplitude ψ, a fluid element will experience a larger pressure gradient from adjacent rings as it orbits around the disc. Hence to investigate the importance of the warp amplitude ψ on the disc evolution, we consider multiple simulations. First, we consider various inclinations where simulations are performed with $i = 5°, 15°,$ and $60°$. By changing only the initial disc misalignment, the only difference to the fiducial setup is the maximum value of ψ. Second, the impact of the warp location is also investigated with two additional simulations with the same warp width ($R_2 - R_1 = 60$) and initial disc misalignment ($i = 30°$), but at $R_0 = 50$ and 100.

6.4 Results

6.4.1 Evolution into a gravitationally stable disc

As the disc in our fiducial simulation evolves, the warp propagates radially both inwards and outwards. The influence of the warp quickly suppresses the spiral structures yielding an axisymmetric gravitationally stable disc. Since the warp is not maintained by external forces, as the disc continues to evolve, the warp smooths out and the disc becomes coplanar. The top panels of Figure 6.2 shows the surface density of the disc at three snapshots at 0.5, 4, and 14 orbits after the disc is warped. The leftmost panels shows the disc in the early stages when it is still warped. The middle panels show the disc while the warp is dissipating at $t = 4$ orbits. Here, the disc is smooth with no signs of any spiral structures left. The rightmost panels show the disc much later in its evolution after the warp has completely dissipated at $t = 14$ orbits. Due to the constant $\beta_{\rm cool}$ used to cool the disc, the disc cools back down and becomes gravitationally unstable once more after the warp has dissipated and the disc is back to being flat.

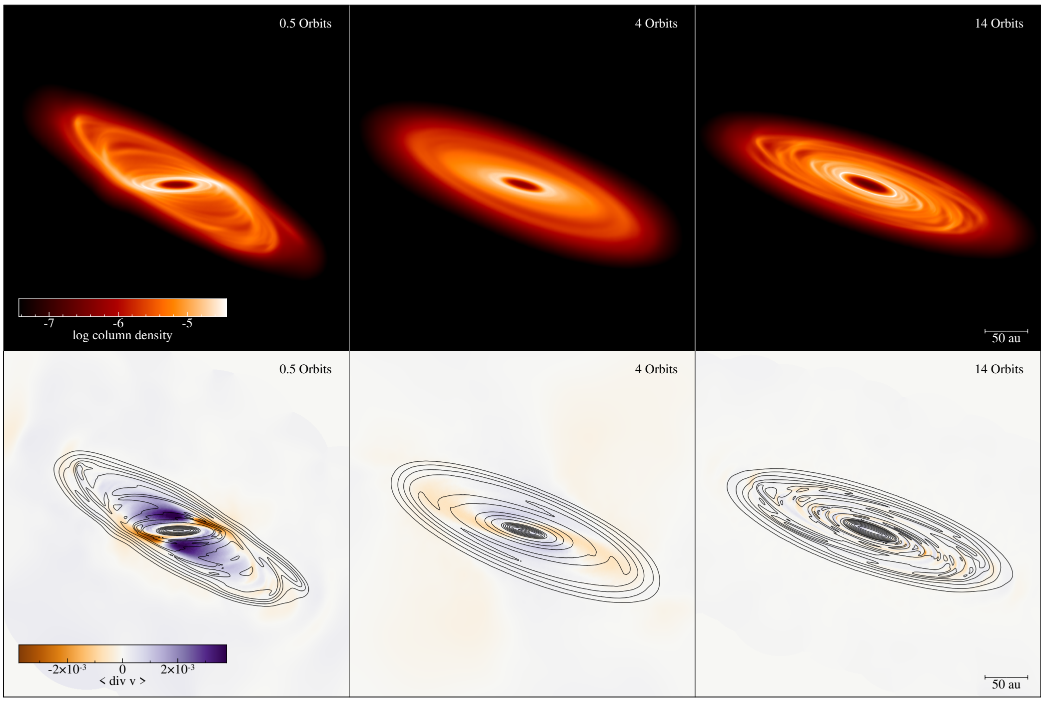


Figure 6.2: Plots of the surface density, Σ (top) and divergence of velocity, $\nabla \cdot \boldsymbol{v}$ (bottom) in code units where the angled brackets represent a density weighted average, showing the evolution of a $0.1 M_\odot$ disc 0.5, 4, and 14 orbits (from left to right) after a warp is introduced. The warp induces a strong response in the velocity flow of the disc (left panels), which heats it up resulting in an axisymmetric gravitationally stable disc (middle panels). After the disc has realigned, cooling takes over with the disc eventually recovering its spiral features (right panels).

To demonstrate what drives the disc to become gravitationally stable, the azimuthally averaged surface density Σ, sound speed c_s, and Toomre Q parameter (Toomre, 1964) are plotted in Figure 6.3, where the darker lines represent later times. The Toomre parameter is given by

$$Q = \frac{c_s \Omega}{\pi G \Sigma}. \tag{6.12}$$

Protoplanetary discs are considered to be gravitationally unstable when spiral structures form. The critical value to form non-axisymmetric instabilities such as spiral arms is when $Q \lesssim 1.7$ (Durisen et al., 2007). We see that the surface density of the disc is mostly unchanged once the warp is introduced. However, as seen by the increase in c_s the influence of the warp heats up the disc by enough to put it in the gravitationally stable regime. During the first couple of orbits when the disc is still warped, Figure 6.3 shows that the disc continues to heat up. As the warp dissipates as the disc evolves, cooling takes over and the disc starts to cool back down. Due to the simple cooling prescription, the inner regions of the disc begin to cool down first. This is expected in a disc cooled with a constant $\beta_{\rm cool}$, where the the disc cools faster in the inner disc. Eventually the warp has dissipated such that the disc can cool down enough to reform its spiral structures. The rightmost panels in Figure 6.2 show the disc at 14 orbits when it is gravitationally unstable again.

To determine why the disc initially heats up, we investigate the role of $P\mathrm{d}V$ work. The bottom panel in Figure 6.2 shows $\nabla \cdot \boldsymbol{v}$, the divergence of the velocity. Recalling from §6.2 that the warp induces a response in the radial velocity, we examine the consequence of this on the heating in the disc. As the divergence of the velocity directly contributes to the energy (and thus temperature) of the disc, see equation 6.6, we can use $\nabla \cdot \boldsymbol{v}$ as a proxy for the heating due to $P\mathrm{d}V$ work that is occuring in the disc. In the early stages of evolution at 0.5 orbits, the disc is still warped. As mentioned in §6.2, strong radial pressure gradient are induced due to adjacent annuli of gas varying with vertical height z. This triggers a response in the induced radial velocity of the disc which heats it up as seen by the large magnitude of $\nabla \cdot \boldsymbol{v}$ in the bottom left panel of Figure 6.2. As the disc evolves, the warp dissipates as there is no external torque sustaining the warp. At 4 orbits, the disc is nearly flat and hence there is little variance in the pressure gradient. Thus, the magnitude of $\nabla \cdot \boldsymbol{v}$ has greatly decreased, as expected, leading to less heating from PdV work. At 14 orbits, the only contributions to $P\mathrm{d}V$ heating is due to the gravitational instabilities in the disc, which is much weaker compared to the $P\mathrm{d}V$ work due to the warp.

6.4.2 Impact of warp inclination

Figure 6.4 shows the initial warp profile for three additional simulations, each with a different initial disc misalignment of $i = 5°, 15°, 60°$ to compare with the fiducial simulation with $i = 30°$. Recalling eq 6.2, the pressure gradient induced by the warp depends on the pressure p, disc scale height H, and warp amplitude ψ. By fixing R_0 and $R_2 - R_1$, the location and extent of the warp respectively, p and H are unchanged in the warped regions for all four discs. Thus, only changing the maximum value of ψ allows for easier comparison of how the warp affects the evolution of the disc.

Figures 6.5 and 6.6 show the surface density and the divergence of the velocity

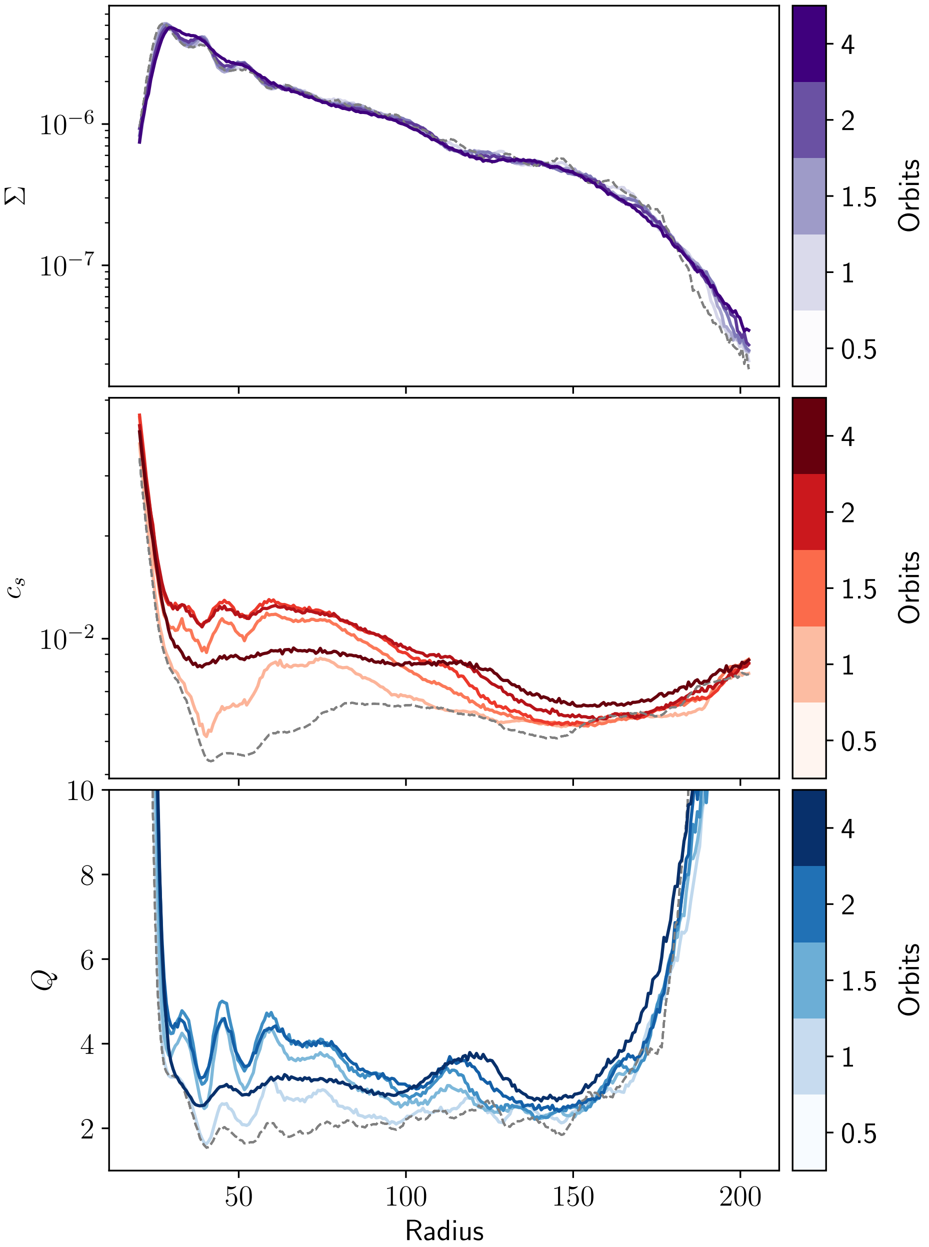


Figure 6.3: Azimuthally averaged surface density, Σ (top), the sound speed, c_s (middle), and the Toomre parameter, Q (top) in code units at 0.5, 1, 1.5, 2, and 4 orbits after the warp has been introduced. The darker shades represent later times. The dashed gray line in each panel is the initial profile at the moment the warp is introduced. The increase in c_s due to the warp is the reason for the increase in Q causing the disc to become gravitationally stable.

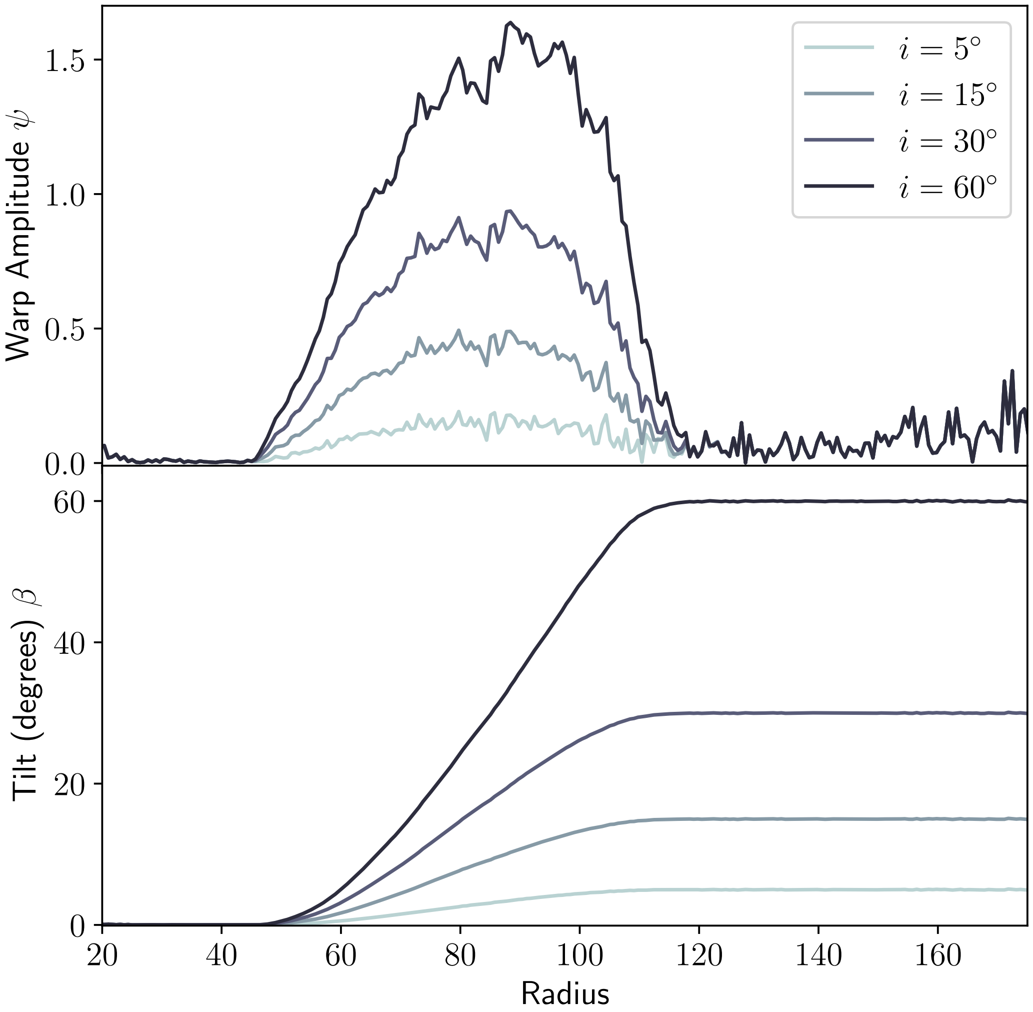


Figure 6.4: The warp amplitude, ψ (top) and the tilt, β (bottom) for the simulations where the initial disc misalignment, i varies from 5° to 60°. The darker lines represent larger i. A larger initial disc misalignment results in a steeper tilt profile, and thus a larger warp amplitude.

at 0.5 and 4 orbits respectively after the warp has been introduced. From the plots of $\nabla \cdot \boldsymbol{v}$ in Fig 6.5 it is clear that, all else being equal, the discs with the stronger warp profiles trigger greater responses in the induced radial velocity of the disc. This is easiest seen comparing the two extremes where $i = 5°$ and $i = 60°$ (top left and bottom right sub panels). For the disc where $i = 5°$, there is little change to the $\nabla \cdot \boldsymbol{v}$ of the disc. Whereas with $i = 60°$, there is a large impact on $\nabla \cdot \boldsymbol{v}$ throughout most of the disc. Fig 6.6 shows the consequence of the warp on the disc structure. It can be seen that the discs with stronger misalignments ($i = 30°$ and $i = 60°$

in particular), which trigger stronger responses in the induced radial velocity of the disc, experience greater heating resulting in the discs becoming gravitationally stable with no signs of spiral structures after a few orbits.

This is reflected in Figure 6.7 which shows the azimuthally averaged surface density Σ (top panels), sound speed c_s (middle panels), and Toomre Q parameter (bottom panels). In all cases, the surface density is mostly unchanged as with the fiducial case. The change in the sound speed on the other hand has a strong dependence on the initial disc misalignment. In general, a larger initial disc misalignment, and hence a larger warp amplitude ψ, results in more heating. However, simply having a misaligned disc does not necessarily result in the disc becoming gravitationally stable. With a slight misalignment of $i = 5^\circ$, since there is very little change to the velocity flow of the disc, the contribution to the $P\mathrm{d}V$ work due to the warp is negligible and the disc simply continues to cool as expected due to the constant β_{cool} used in these simulations. Although there is some heating in the $i = 15^\circ$ case, it is not enough to push the disc into the gravitationally stable regime as it quickly cools back down to pre-warp levels. It is only in the $i = 30^\circ$ and 60° cases that there is enough heating to result in the disc becoming completely gravitationally stable.

6.4.3 Impact of warp location

Unlike in §6.4.2 where varying the initial disc misalignment i, only changes the warp profile, varying the warp location R_0, adds additional complexities to the evolution of a warped disc. Figure 6.8 shows the warp profiles (first two panels), the initial pressure p (third panel), and disc scale height H (bottom panel), for the simulations with $R_0 = 50, 75$ and 100. The bottom two panels are important to understanding the pressure gradients induced by the warp, recalling eq 6.2.

Since the warp amplitude, ψ is larger for increasing R_0, based on the results in §6.4.2 we might naïvely expect the magnitude of $\nabla \cdot \boldsymbol{v}$ to be largest for the disc warped at $R_0 = 100$. However, from Figure 6.9 this is clearly not the case. Instead, the largest change to $\nabla \cdot \boldsymbol{v}$ is for the disc warped at $R_0 = 50$ which has the smallest ψ. To reconcile this discrepancy, we have to also consider the pressure and scale height at each R_0. From Fig 6.8, the pressure increases while the disc scale height decreases for smaller R_0. Both of these work towards increasing the pressure gradients induced by the warp (see equation 6.2), thus triggering a stronger response in $\nabla \cdot \boldsymbol{v}$ for smaller R_0, which is consistent with our results.

Based on the results in §6.4.2 we should expect that the spiral structures to be most weakened for the discs warped at smaller R_0, since they have the largest change in $\nabla \cdot \boldsymbol{v}$. However, from Figure 6.10, this is clearly not the case. Instead, the

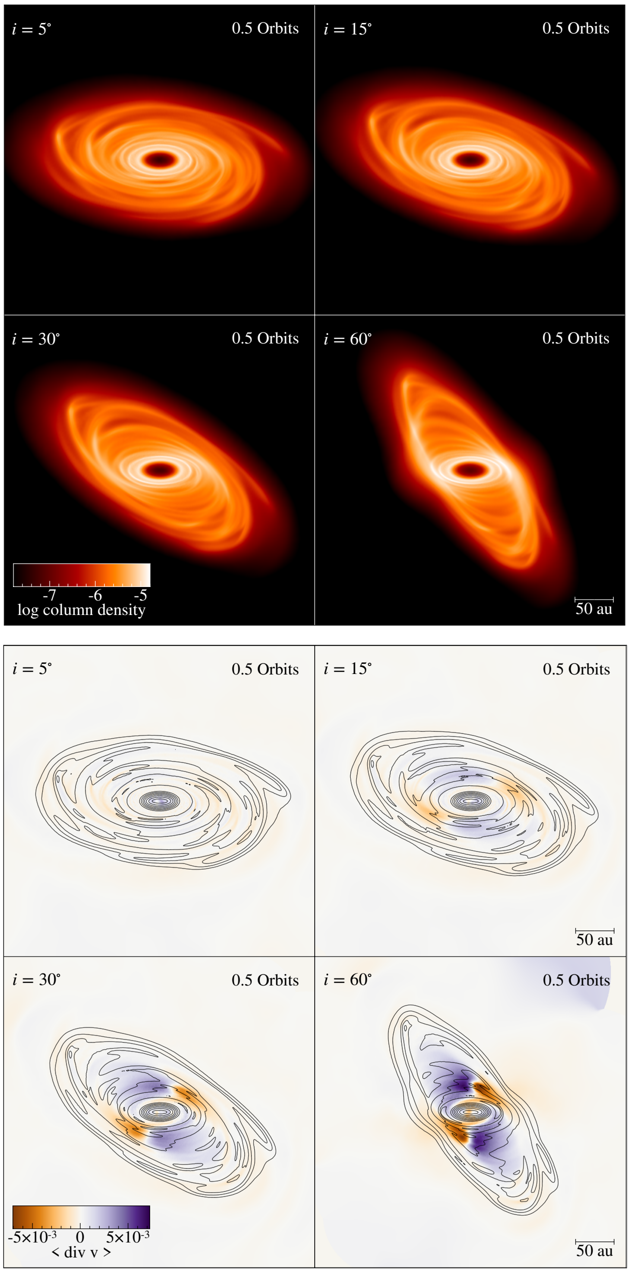


Figure 6.5: The surface density, Σ (top) and divergence of velocity, $\nabla \cdot \boldsymbol{v}$ (bottom) in code units where the angled brackets represent a density weighted average, showing the evolution of a $0.1M_{\odot}$ disc 0.5 orbits after a warp has been introduced. The subplots are for a disc with initial misalignments of $i = 5^{\circ}, 15^{\circ}, 30^{\circ}$ and 60°. The $\nabla \cdot \boldsymbol{v}$ plots show that the strength of the response to the velocity flow of the disc due to the warp is greater for larger i.

impact to the disc structure is greater at larger R_0. To understand why, we have to

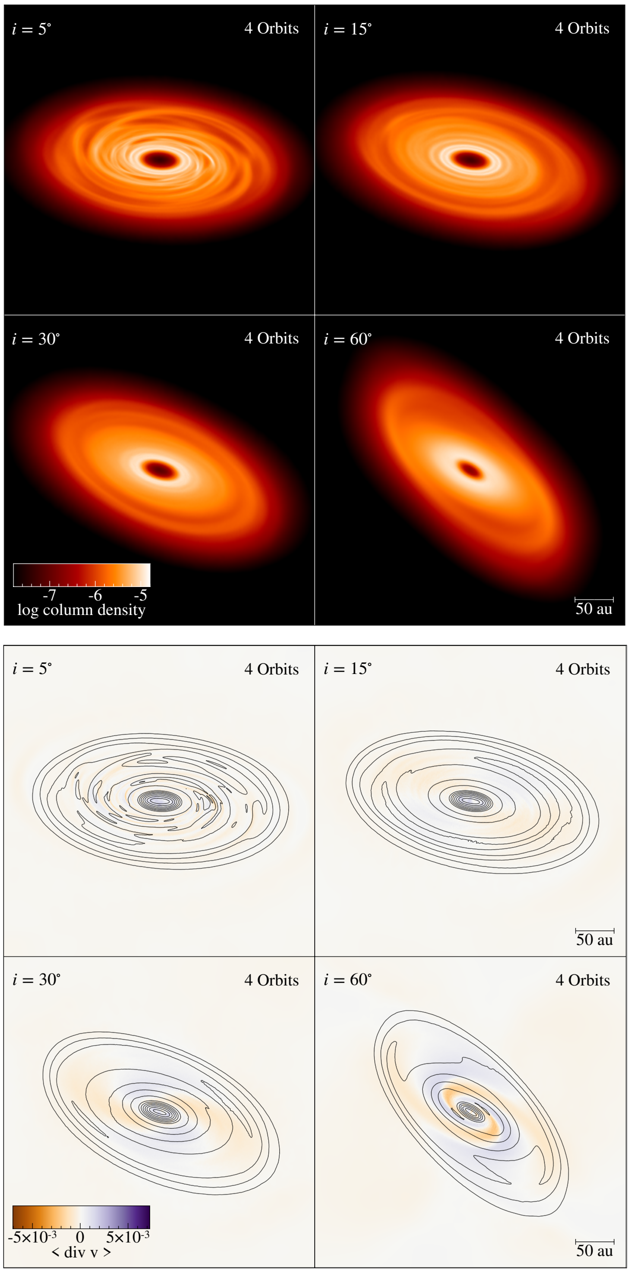


Figure 6.6: Same as Figure 6.5, but at 4 orbits after the warp has been introduced. Due to the strong response to the velocity flow, the discs in the bottom row ($i = 30°$ and $60°$) have heated up enough to become gravitationally stable and lose their spiral structures. Whereas for a slight misalignment ($i = 5°$), the spiral structures are unaffected as the disc has experienced negligible heating.

now consider the timescale at which warp propagation is damped, $t_{\rm damp} = (\alpha\Omega)^{-1}$ (Lubow & Ogilvie, 2000), and the cooling timescale, $t_{\rm cool} = \beta_{\rm cool}\Omega^{-1}$ (Gammie,

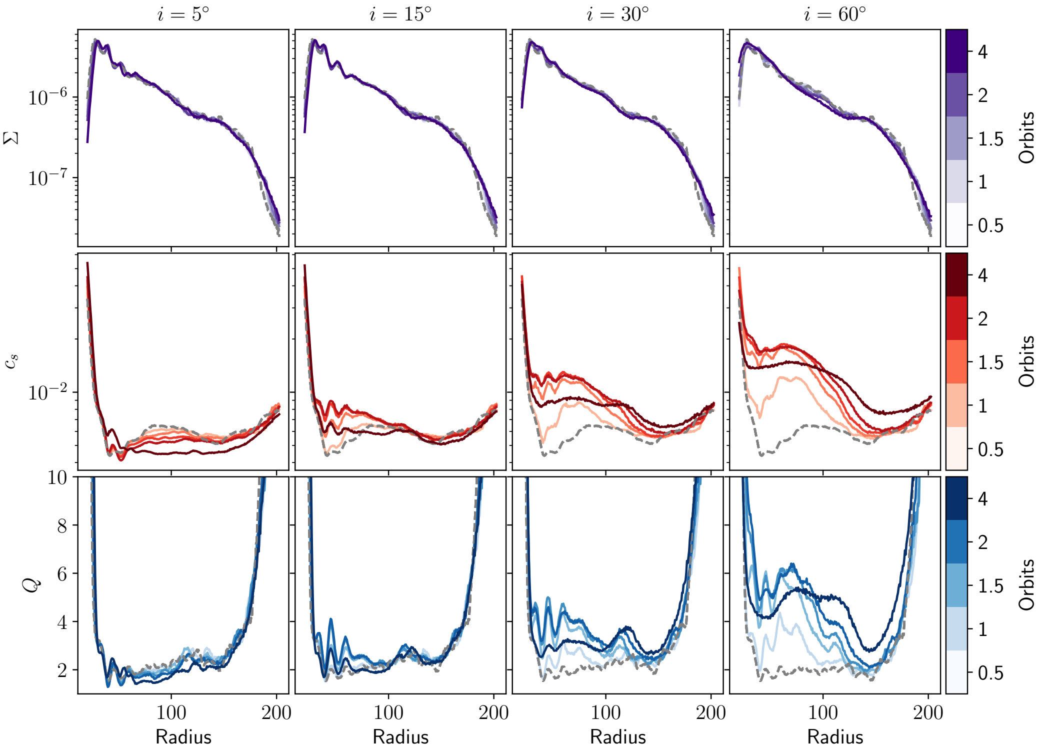


Figure 6.7: The azimuthally averaged surface density Σ (top panels), sound speed c_s (middle panels), and Toomre Q parameter (bottom panels) in code units for the discs with initial disc misalignment $i = 5°, 15°, 30°$ and $60°$ (from left to right). The darker shades represent later times. The dashed line represents the conditions at $t = 0$. The disc only experiences sufficient heating to become gravitational stable when $i = 30°$ and $60°$.

2001). For an $\alpha \sim 0.1$ (a typical value for GI discs), the damping timescale is ∼10 orbital periods at $R = 75$AU, which is consistent with our simulations. Both of these timescales are longer at larger R_0. A longer $t_{\rm damp}$ results in the disc taking longer to realign. Since the disc is in a warped state for an increased duration, the heating due to the warp is prolonged. Comparing the right panels of Figs 6.9 and 6.10 shows that the magnitude of $\nabla \cdot \boldsymbol{v}$ has decreased the most for $R_0 = 50$. Whereas for $R_0 = 100$ the magnitude of $\nabla \cdot \boldsymbol{v}$ has not changed too much. This is consistent with the disc realigning much faster for a warp at $R_0 = 50$. Since $t_{\rm cool}$ is also shorter at smaller R_0, cooling is faster in the inner regions of the disc due to the simple cooling model used. Hence, the heating due to the warp is less efficient for smaller R_0.

Figure 6.11 shows the azimuthally averaged surface density Σ (top panels), sound speed c_s (middle panels), and Toomre Q parameter (bottom panels) for these

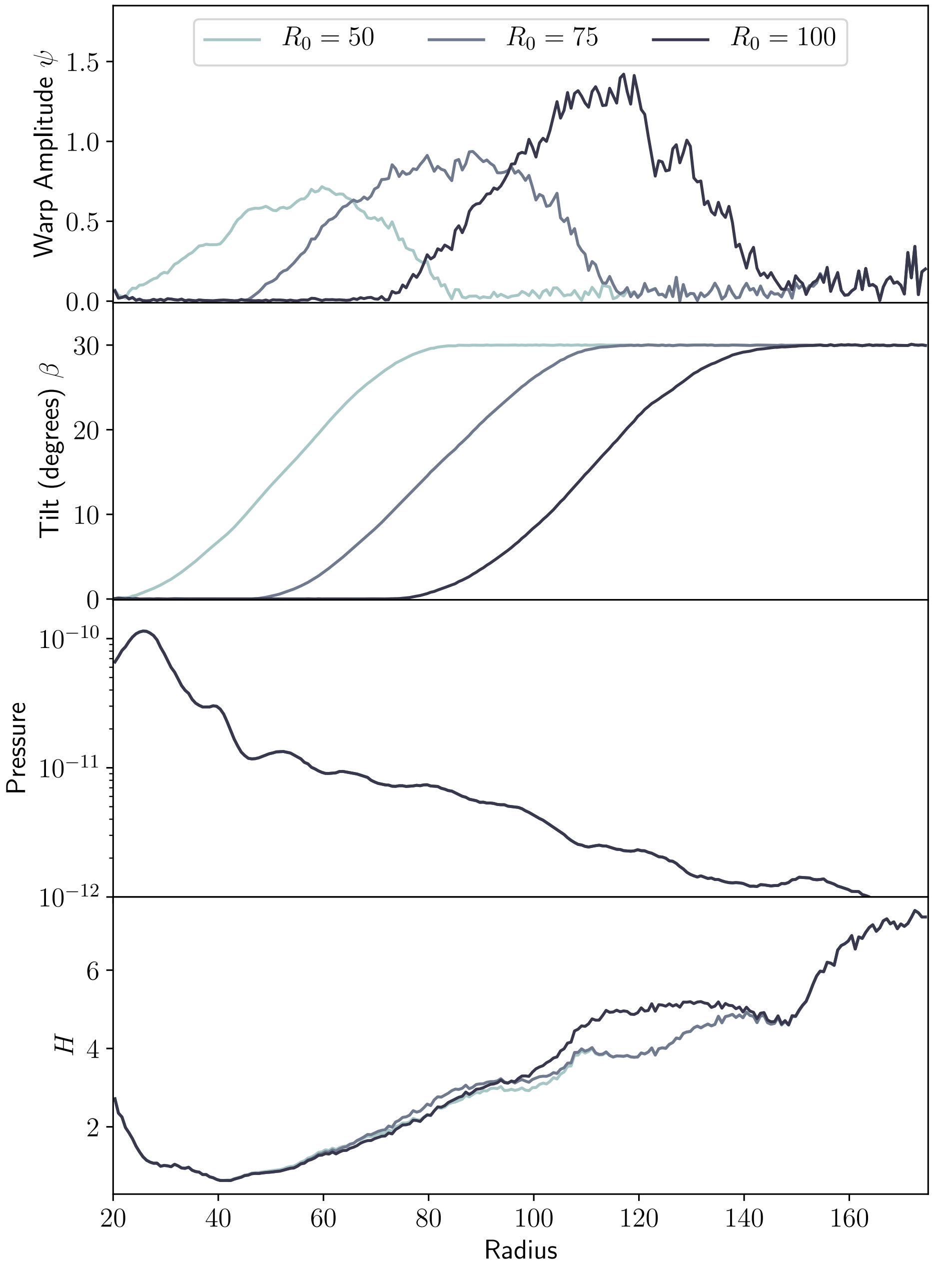


Figure 6.8: The warp amplitude, ψ (first panel) and the tilt, β (second panel) for the simulations where the warp location, $R_0 = 50, 75$, and 100. The darker lines represent larger R_0. The maximum value of ψ increases at higher R_0. The pressure and disc scale height in code units are shown in the third and fourth panels respectively. The pressure profile is the same for all simulations, hence the lines are coincident.

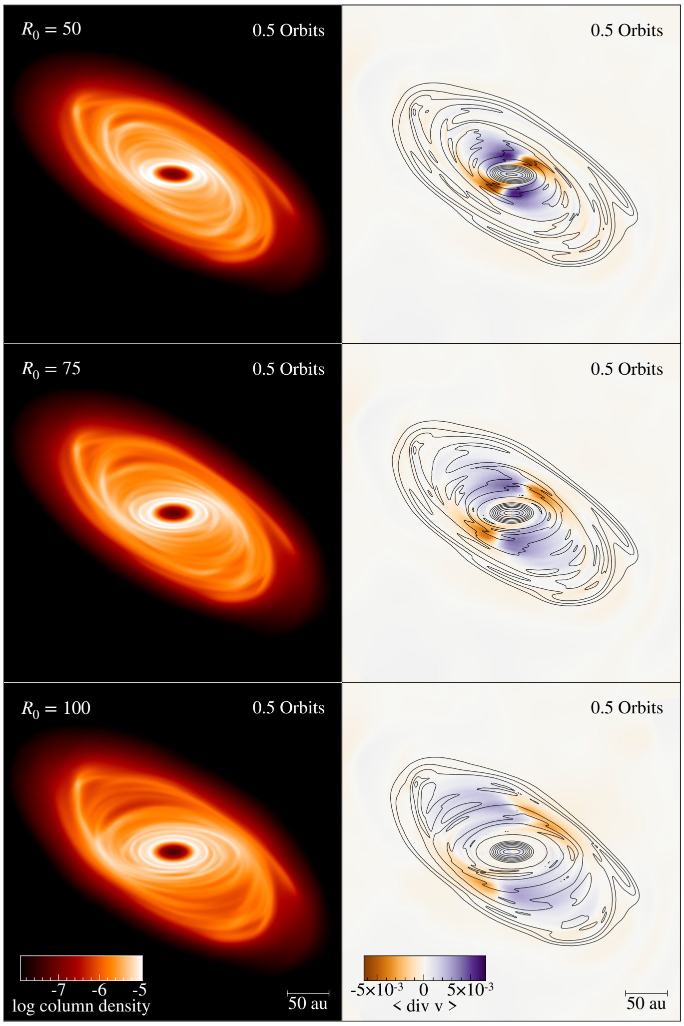


Figure 6.9: The surface density, Σ (left) and divergence of velocity, $\nabla \cdot \boldsymbol{v}$ (right) in code units showing the evolution of a $0.1M_\odot$ disc 0.5 orbits after a warp has been introduced. The subplots are for a disc with the warp located at $R_0 = 50, 75$ and 100. Although the warp amplitude ψ increases with higher R_0, the magnitude of $\nabla \cdot \boldsymbol{v}$ is largest at lower R_0, which is at odds with the results in Figure 6.5. However, once the differences in pressure p, and disc scale height H at different R_0 are considered, this discrepancy is reconciled. Higher pressures and and a smaller scale height at lower R_0 counter the lower warp amplitude giving rise to greater response in the induced radial velocity.

simulations. The impact of the shorter cooling and realignment timescales at smaller R_0 is more apparent in this figure. The disc with $R_0 = 50$ cools the quickest. Additionally, since the disc also realigns the fastest for $R_0 = 50$, the heating due to the warp does not last long enough to spread throughout the disc. This is in contrast

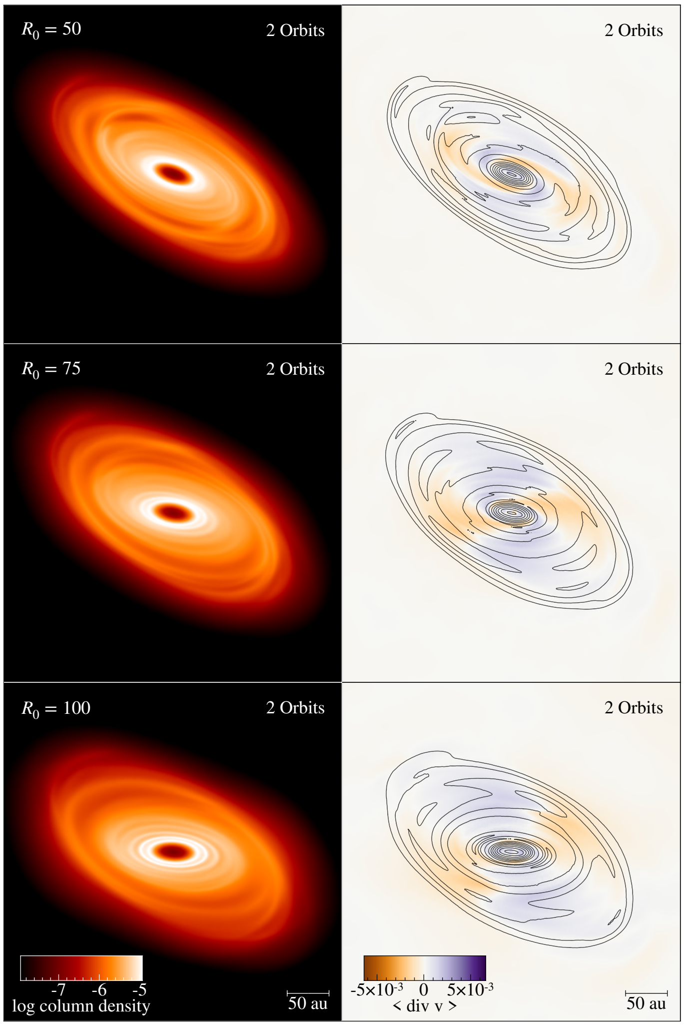


Figure 6.10: Same as Figure 6.9 but at 2 orbits after the warp has been introduced. Despite the magnitude of $\nabla \cdot \boldsymbol{v}$ being initially larger at lower R_0, the impact to the spiral structures is greater at larger R_0. This result initially appears at odds with Figure 6.6 where the spiral structures are most suppressed with increasing magnitudes of $\nabla \cdot \boldsymbol{v}$. However consideration of the cooling and realignment timescales reconciles this discrepancy. Both timescales are shorter at smaller R_0. Hence, we expect the magnitude of $\nabla \cdot \boldsymbol{v}$, and hence the amount of heating, to decrease faster for discs warped at lower R_0, which can be seen by comparing with the right panels of Figure 6.9 to see how much the magnitude of $\nabla \cdot \boldsymbol{v}$ has changed by. Therefore, the results here are consistent with our expectations of spiral structures being most impacted at larger R_0 since the heating due to the warp is prolonged and more efficient at larger R_0.

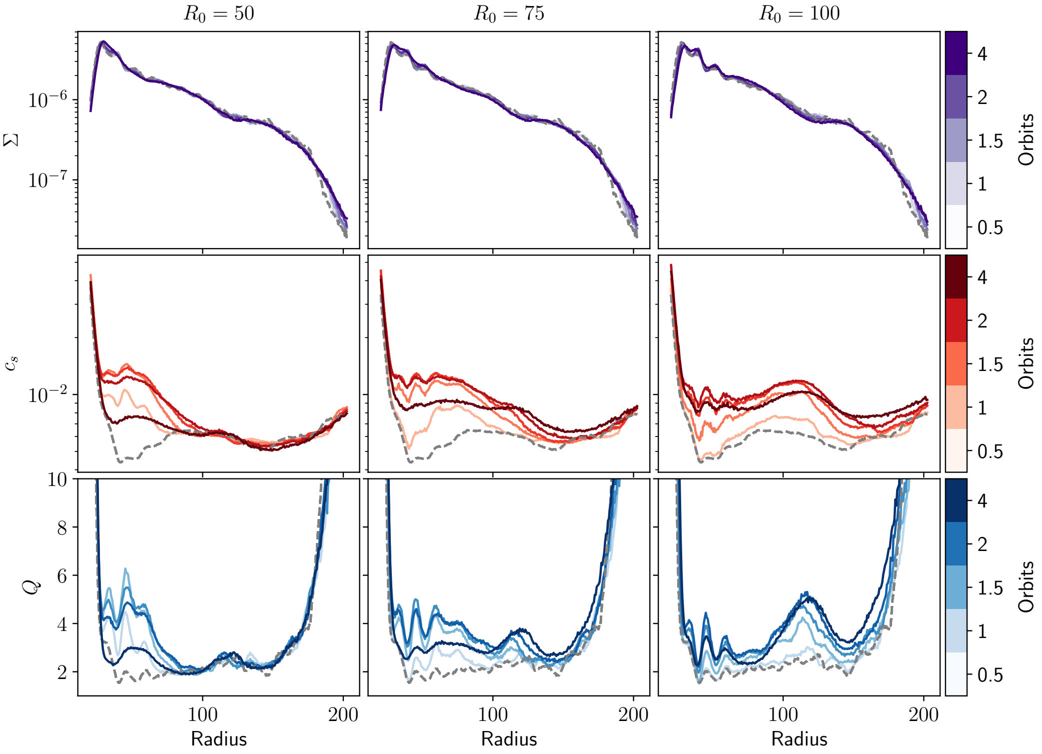


Figure 6.11: The azimuthally averaged surface density Σ (top panels), sound speed c_s (middle panels), and Toomre Q parameter (bottom panels) in code units for the discs with the warp located at $R_0 = 50, 75$, and 100 (from left to right). The darker shades represent later times. The dashed line represents the conditions at $t = 0$. The disc experiences greater heating, and thus is more impacted, at larger R_0 due to longer cooling and realignment timescales at larger radii.

to the disc with $R_0 = 100$ which has experienced the most heating globally due to a longer realignment timescale. In general, we find that the spiral structures are more affected by a warp located further out which results in the disc being heated up for a longer duration.

6.5 Discussion

6.5.1 Disc cooling & long term evolution

The cooling parameter, β_{cool} is constant both spatially and temporally. Although commonly used to model self-gravitating discs, comparison with radiative transfer models show that β_{cool} should vary in both time and space (Mercer et al., 2018). For the purposes of this Chapter, a constant β_{cool} is justified as we are primarily interested in how the evolution of the disc structure responds to a warp. Since PdV heating dominates over cooling while the disc is warped, we expect the qualitative results to hold; a warp acts to move a self-gravitating disc towards stability.

However, the long term evolution of the disc will be better understood when taking into account the non-coplanar geometry with more sophisticated disc thermodynamics. In all simulations where the disc becomes gravitationally stable due to the warp, the disc eventually cools back down to become gravitationally unstable and recover its spirals throughout the disc. This is simply a result of the cooling model which is time-independent and thus does not take into account the structural evolution of the disc. Additionally, the shorter cooling time in the inner disc naturally results in spiral structures forming more quickly in the inner regions, contrary to expectation of realistic self-gravitating discs (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009). Hence, more realistic thermodynamics are needed to better understand the longer term evolution of the suppression of spiral structures due to the warp.

6.5.2 Formation of rings & gaps

An interesting result which could have important implications on observations is the formation of rings and gaps from the evolution of a warped gravitationally unstable disc. We find tentative evidence of ring and gap structure, most easily seen in the bottom row of Figures 6.6 ($i = 30°$ and $60°$) and 6.10 ($R_0 = 100$). The evolution of the warp amplitude which has a continuous decline in magnitude combined with the surface density profiles suggests that the gap is not a result of disc breaking. We also do not expect the disc to break due to the high α viscosities (Doğan et al., 2018) associated with gravitationally unstable discs generally and in our simulations ($\alpha \sim 0.1$). Our results instead suggest that the spiral structures initially present in our disc evolved into ring and gap structure as a result of the warp propagation. However, the morphology of the spiral structures depend on the disc thermodynamics. Thus, more detailed simulations are required to determine if warps in gravitationally unstable discs can explain some of the sub-structures seen in observations. This may explain features in existing discs such as Elias 2-27, which contains a gap, spiral structures and evidence of a warp (Paneque-Carreño et al., 2021). If the warp in Elias 2-27 is located in the inner parts of the disc, the impact of the warp could be localised to the inner disc allowing the spiral structures in the outer regions of the disc to survive.

6.6 Conclusion

We perform 3D SPH simulations to investigate the impact of a warp on the structure of a gravitationally unstable disc. Our work shows that if the warp is strong enough, it can suppress spiral structures due to gravitational instability. This is due to the oscillating radial pressure gradient induced by the warp which triggers a response in the velocity flow of the disc. This causes the disc to heat up and become gravitationally stable. In some cases, the disc evolves to form ring & gap structure, which could have important observational implications.

We find that the structure is more impacted for warps with larger initial disc misalignments and for warps located further out. In the former case, a larger initial misalignment simply results in larger pressure gradients due to a larger warp amplitude, and thus a stronger response in the velocity flow of the disc leading to increased heating. The latter case is more complicated due to physical properties of the disc varying with radius, but a warp located further out has a greater impact on the spiral structure due to a longer realignment timescale, which results is prolonged heating. Finally, we note that a more detailed treatment of the disc thermodynamics is required to understand the long term evolution of the tentative ring and gap structure we identify in some of our simulations.

7

Future Work

7.1 Star-disc interactions

The results of Chapter 6 demonstrated the concept of how a warp could alter the evolution of a gravitationally unstable disc. In those simulations, I used an ideal warp such that the effect of the warp was isolated from whatever mechanisms might induce the warp. Previous works that study mechanisms of warping a disc neglect self-gravity as the disc masses are much lower compared to the discs in this thesis. Hence, my future studies will explore whether external companions are able to warp gravitationally unstable discs.

The simulations will primarily feature bound companions on retrograde orbits as they provide more favourable conditions for warping a disc (Xiang-Gruess, 2016; Cuello et al., 2019). In addition to companions on inclined orbits, there will be two control simulations. One will be a standard GI disc around a single star to compare the evolution of the disc on its own. The other will contain a companion on a coplanar retrograde orbit. The latter will allow the impact of the companion to be differentiated from the impact of the warp. Figure 7.1 shows the preliminary results of a simulation showing a $0.1M_{\odot}$ disc warped by a $1M_{\odot}$ companion inclined at 135° at a pericenter distance of 350au. It shows that the concept introduced in Chapter 6 can be physically motivated.

Star formation simulations (Bate, 2018) have shown that during their formation discs do encounter neighbouring stars. Additionally, the star systems formed can be made up of multiples, also confirmed by numerous examples of binary and triple systems (e.g. Tobin et al., 2016a). However, due to the large scale of the

Figure 7.1: A simulation of a $0.1M_\odot$ disc around a $1M_\odot$ star warped by a $1M_\odot$ companion inclined at 135° at a pericenter distance of 350au.

simulation, the discs were poorly resolved. These controlled simulations aim to understand how disc-star interactions affect the evolution of gravitationally unstable discs and the resulting implications on planet formation.

7.2 Misaligned infall

An alternative scenario that can result in misaligned discs is by infalling material. The disc can become warped or even broken if the angular momentum vector of the infalling material is misaligned relative to that of the disc. This has been demonstrated in Bate (2018); Kuffmeier et al. (2021), but with unresolved discs and in a much more chaotic environment; the misaligned disc that had formed in that star formation simulation was subsequently destroyed by stellar encounters. My future studies based on Chapter 6 aims to study how infalling material affects the evolution of gravitationally unstable discs in a more controlled environment. It will involve setting up a cloud of material on the orbit of a flyby. This will allow for control of the shape, mass, rotational speed, and distance of closest approach of the infalling material resulting in a stronger understanding of the conditions required for infalling material to warp gravitationally unstable discs. Another possible study is to replicate the conditions in figure 1.13 by studying the impact of misaligned infall on a circumbinary gravitationally unstable disc.

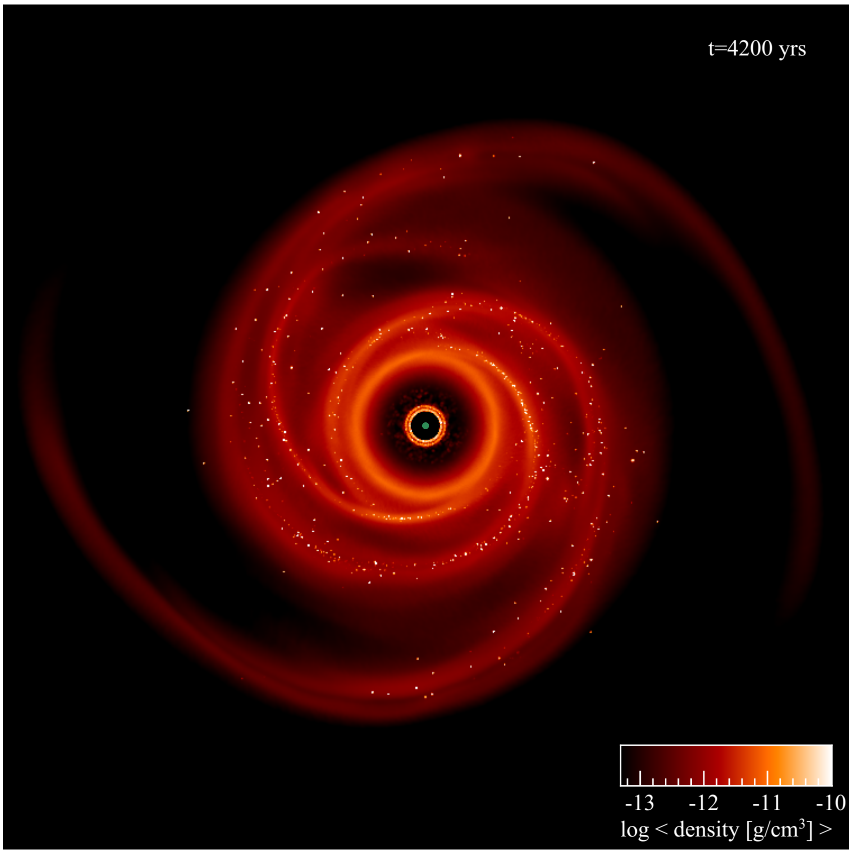


Figure 7.2: Dust+gas density rendered image of a $0.25 M_\odot$ disc. Dust with Stokes numbers of approximately 4 are efficiently trapped in the spiral arms forming numerous clumps.

7.3 Dust concentration in the spiral arms of GI discs

Chapter 3 showed that planets that form in gravitationally unstable disc can survive their inward migration. These results will motivate future work studying planet formation in self-gravitating discs. While planets forming through gravitational instability is well studied, the dust is often neglected. The spiral arms due to gravitational instabilities are regions of pressure maxima where dust can get trapped (Rice et al., 2004, 2006) forming planetary embryos (Baehr et al., 2022).

My initial simulations will explore the role of self-gravity and drag in trapping dust of different sizes in the spiral arms, and the formation of clumps. Later, this will be expanded to investigate dust growth in self-gravitating discs. These simulations will be the first to explore how fast dust can grow in gravitationally unstable discs and whether they can grow to the sizes required to fragment into planetary sized embryos. Figure 7.2 shows a preliminary simulation where dust with Stokes numbers of approximately 4 are efficiently trapped in the spiral arms of a $0.25M_\odot$ disc and have formed numerous clumps.

7.4 Role of thermodynamics on the signatures of planet-disc interactions

The results of Chapters 3 - 5 highlight the importance of the disc thermodynamics on the fate of planet migration, and how planet-disc interactions shape the disc. However, a vast majority of non-self-gravitating discs typically assume a locally isothermal equation of state. As shown in Chapter 4 and 5 the heating due to spiral wakes generated by the planet can play a role in the disc's evolution, which is missed by a vertically isothermal equation of state. Recently Speedie et al. (2022) showed that the observational signatures of planets can be missed if the simpler thermodynamic model is used.

My future work will focus on giant gap-opening planets. It will involve comparing the observational signatures using three thermodynamic models: vertically isothermal; adiabatic with cooling; and radiative transfer. One of the main areas of investigation will be the discrepancy of planet mass estimates from gap widths and depths (Zhang et al., 2018; Lodato et al., 2019), and the kinematics (Pinte et al., 2020). A change in how the disc thermodynamics is modelled will change planet mass estimates from either the gap profile or the kinematics, or both. Hence, the aim is to see whether more accurate modelling of the disc thermodynamics resolves this discrepancy.

8

Conclusion

I have carried out three-dimension numerical simulations of gravitationally unstable protoplanetary discs to investigate whether evidence of gravitational instability can by hidden by physical processes such as planet-disc interactions and warps. To do this I have used a combination of three-dimensional hydrodynamic simulations and radiative transfer calculations, producing both models with demonstrative physics and synthetic observations.

In Chapter 3, by modelling the thermodynamics using variable β-cooling to mimic a realistic gravitationally unstable discs where the inner regions are gravitationally stable (Rafikov, 2005; Stamatellos & Whitworth, 2009; Rice & Armitage, 2009; Clarke, 2009), I find that planets are able to slow down their migration in the inner parts of the disc in contrast to previous studies (Baruteau et al., 2011; Malik et al., 2015) which used a constant β to cool the disc. A consequence of the simpler thermodynamics is a globally gravitationally unstable disc which is inconsistent with theoretical expectations. My result is more consistent with work that included realistic thermodynamics which showed planets could open up gaps and slow their migration (Stamatellos, 2015; Stamatellos & Inutsuka, 2018). However, an important difference between the work in this thesis with the latter is accretion onto the planet. In their simulations, the planets grew in mass beyond the brown dwarf limit allowing the planet to easily open up a gap. Hence, it was unclear whether the thermodynamics or the gap in the disc was the dominant reason for slowing down the migration. Whereas in my simulations the planet's mass is fixed and most of the planet masses chosen in my simulations were insufficient to open a gap. Thus it was clear that the treatment of the disc thermodynamics was the relevant factor in slowing down the migration of the planet. Once the planet was beyond the turbulent unstable parts of the disc, the corotation torques greatly decreased allowing the planet to slow down. The survivability of planets in GI discs is important as it shows that planets that form in these discs could influence the disc's evolution if it is massive enough.

In Chapter 4, two simulations were compared. One where the disc evolved with a $3M_{\rm Jup}$ planet, and the other being a control simulation to see how the disc would have evolved in the absence of the planet. I showed that the planet was able to alter the evolution of the disc, completely suppressed spiral structures by heating up the disc and thus pushing it into the gravitationally stable regime. On the other hand, the control simulation simply continued to be gravitationally unstable. I made mock observations to show that in the absence of a planet, spiral structures due to GI is easily observable. Whereas with a planet, even a massive disc that would otherwise be gravitationally unstable could look like many of the observed discs with ring & gap structure.

In Chapter 5, I considered a range of disc and planet masses to study in further detail how planet-disc interactions can influence the evolution of a GI disc. Using plots of the velocity divergence ($\nabla \cdot v$) and Toomre Q, I show three scenarios that can occur when a planet is migrating in a GI disc. If the disc is massive enough, then from analysis of $\nabla \cdot v$ it's clear that GI is too strong relative to the planet's spiral wake. Hence, the planet has little impact on the velocity structure and is thus unable to suppress the spiral structures due to GI. Conversely, if the planet's spiral wakes are too strong relative to GI, the disc is pushed into the gravitationally stable regime resulting in a disc with ring & gap structure. Between these two extremes, both the planet and GI play a role in the disc's evolution. While spiral structures are not completely suppressed in this scenario, they are weakened. Even though the planet cannot completely hide evidence of GI in the simulations, evidence of GI is less obvious in the mock observations. Finally, analysis of the kinematics showed that in this set of simulations the signature of a planet cannot be detected simultaneously with the signature of GI.

In Chapter 6, I introduced a warp in a GI disc to study how the evolution of a GI disc might be affected. The warp induces an oscillating radial pressure gradient which triggers a response in the velocity flow of the disc. A consequence of the warp is the loss of spiral structures as the disc is then pushed into the gravitationally stable regime as it heats up while the warp is still present. A range of warp profiles was explored to determine how the disc structure was impacted for different disc misalignments and by the location of the warp. I showed that larger misalignments results in stronger pressure gradients due to a larger warp amplitude. Hence, the larger the misalignment, the more the disc structure was impacted. I also showed that since the realignment timescale is longer further out in the disc, the more the disc structure is impacted. This is because the lifetime of the warp is prolonged, and hence the disc structure is heated up for a longer duration making it easier for the disc to become gravitationally stable and lose its spiral structures. In some simulations, the disc also formed ring & gap structure, which again shows that massive discs can appear in observations without their characteristic large scale spiral structures.

While recent observations of protoplanetary discs (eg. Andrews et al., 2018) suggest that gravitationally unstable discs may be intrinsically rare, the results here demonstrate that the observed rarity could be due to neglecting many of the complex processes that can shape the evolution of a GI disc and suppress its spiral structures. Furthermore, the results also show that upper limits on the disc mass are not always necessary if there is a lack of visible spiral structure when the disc is expected to be gravitationally unstable. This is particularly relevant for mass estimates of protoplanetary discs due to the many uncertainties that go into the various methods. While estimates from the readily measured dust mass indicate discs are not expected to be gravitationally unstable, more robust tracers of the disc mass can push the disc into the gravitationally unstable regime despite the disc having ring & gap structure, as was the case with HL Tau (Booth & Ilee, 2020). Hence, if planets are the cause, they must have formed when the disc was younger and potentially gravitationally unstable. Planets being able to survive the gravitationally unstable phase of a disc also has important implications as many of the substructures seen in discs can be explained by planets.

References

ALMA Partnership et al., 2015, ApJ, 808, L3 [pp. 4, 6, 18, 57, 83, 96, 97, and 117]

Adachi I., Hayashi C., Nakazawa K., 1976, Progress of Theoretical Physics, 56, 1756 [p. 18]

Adams F. C., Lada C. J., Shu F. H., 1987, ApJ, 312, 788 [p. 2]

Adams F. C., Ruden S. P., Shu F. H., 1989, ApJ, 347, 959 [p. 23]

Alexander R., Pascucci I., Andrews S., Armitage P., Cieza L., 2014, Protostars and Planets VI, p. 475 [p. 7]

Alibert Y., Mordasini C., Benz W., Winisdoerffer C., 2005, A&A, 434, 343 [p. 17]

Andre P., Montmerle T., 1994, ApJ, 420, 837 [p. 15]

Andre P., Ward-Thompson D., Barsony M., 1993, ApJ, 406, 122 [p. 2]

Andrews S. M., 2015, PASP, 127, 961 [pp. 15, 97]

Andrews S. M., 2020, ARA&A, 58, 483 [pp. 6, 8, 9, and 10]

Andrews S. M., Williams J. P., 2005, ApJ, 631, 1134 [p. 3]

Andrews S. M., et al., 2016, ApJ, 820, L40 [pp. 57, 83, 96, and 117]

Andrews S. M., et al., 2018, ApJ, 869, L41 [pp. 3, 6, 12, 16, 26, 57, 80, 83, 96, 117, and 144]

Ansdell M., et al., 2016, ApJ, 828, 46 [pp. 4, 16]

Ansdell M., Williams J. P., Manara C. F., Miotello A., Facchini S., van der Marel N., Testi L., van Dishoeck E. F., 2017, AJ, 153, 240 [p. 57]

Artymowicz P., 1993, ApJ, 419, 155 [p. 21]

Artymowicz P., Lubow S. H., 1996, ApJ, 467, L77 [p. 21]

Avenhaus H., Quanz S. P., Schmid H. M., Meyer M. R., Garufi A., Wolf S., Dominik C., 2014, ApJ, 781, 87 [pp. 8, 9, and 11]

Avenhaus H., et al., 2018, ApJ, 863, 44 [p. 6]

Backus I., Quinn T., 2016, MNRAS, 463, 2480 [p. 23]

Bae J., Zhu Z., 2018, ApJ, 859, 119 [p. 7]

Baehr H., Zhu Z., 2021, ApJ, 909, 135 [p. 100]

Baehr H., Zhu Z., Yang C.-C., 2022, ApJ, 933, 100 [p. 140]

Ballabio G., Nealon R., Alexander R. D., Cuello N., Pinte C., Price D. J., 2021, MNRAS, 504, 888 [pp. 9, 52, 56, and 118]

Balsara D. S., 1995, Journal of Computational Physics, 121, 357 [p. 45]

Baruteau C., Meru F., Paardekooper S.-J., 2011, MNRAS, 416, 1971 [pp. 22, 58, 59, 60, 66, 70, 78, 81, 102, 110, and 142]

Bate M. R., 1998, ApJ, 508, L95 [p. 1]

Bate M. R., 2011, MNRAS, 417, 2036 [p. 1]

Bate M. R., 2018, MNRAS, 475, 5618 [pp. 2, 10, 23, 24, 93, 112, 118, 138, and 139]

Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362 [pp. 46, 60, 84, and 98]

Beckwith S. V. W., Sargent A. I., Chini R. S., Guesten R., 1990, AJ, 99, 924 [p. 15]

Benisty M., et al., 2015, A&A, 578, L6 [p. 9]

Benisty M., et al., 2017, A&A, 597, A42 [pp. 9, 11, and 118]

Benisty M., et al., 2018, A&A, 619, A171 [pp. 9, 11]

Benz W., 1990, Numerical Modelling of Nonlinear Stellar Pulsations Problems and Prospects, p. 269 [p. 36]

Bergin E. A., Williams J. P., 2017, Formation, Evolution, and Dynamics of Young Solar Systems, 445, 1 [pp. 10, 15, and 97]

Bergin E. A., et al., 2013, Nature, 493, 644 [p. 15]

Bjorkman J. E., Wood K., 2001, ApJ, 554, 615 [p. 54]

Bohlin R. C., Savage B. D., Drake J. F., 1978, ApJ, 224, 132 [pp. 4, 15]

Boley A. C., Hayfield T., Mayer L., Durisen R. H., 2010, Icarus, 207, 509 [pp. 58, 81]

Bollati F., Lodato G., Price D. J., Pinte C., 2021, MNRAS, 504, 5444 [pp. 14, 111]

Booth A. S., Ilee J. D., 2020, MNRAS, p. L14 [pp. 11, 16, 57, 80, 83, 96, 97, 113, 117, and 144]

Booth A. S., Walsh C., Ilee J. D., Notsu S., Qi C., Nomura H., Akiyama E., 2019, ApJ, 882, L31 [pp. 11, 16, and 97]

Boss A. P., 1997, Science, 276, 1836 [pp. 7, 17, and 19]

Bryden G., Chen X., Lin D. N. C., Nelson R. P., Papaloizou J. C. B., 1999, ApJ, 514, 344 [p. 21]

Calcino J., Price D. J., Pinte C., van der Marel N., Ragusa E., Dipierro G., Cuello N., Christiaens V., 2019, MNRAS, 490, 2579 [p. 8]

Calcino J., Hilder T., Price D. J., Pinte C., Bollati F., Lodato G., Norfolk B. J., 2021, arXiv e-prints, p. arXiv:2111.07416 [pp. 14, 111]

Casassus S., et al., 2013, Nature, 493, 191 [p. 10]

Casassus S., et al., 2018, MNRAS, 477, 5104 [p. 118]

Cazzoletti P., et al., 2018, A&A, 619, A161 [p. 10]

Clarke C. J., 2009, MNRAS, 396, 1066 [pp. 58, 63, 85, 99, 136, and 142]

Clarke C. J., Pringle J. E., 1993, MNRAS, 261, 190 [p. 23]

Clarke C. J., et al., 2018, ApJ, 866, L6 [pp. 80, 97]

Cossins P., Lodato G., Clarke C. J., 2009, MNRAS, 393, 1157 [pp. 59, 99]

Cuello N., et al., 2019, MNRAS, 483, 4114 [p. 138]

Cullen L., Dehnen W., 2010, MNRAS, 408, 669 [pp. 45, 46, 60, 85, 99, and 122]

Deng H., Mayer L., Helled R., 2021, Nature Astronomy, 5, 440 [p. 19]

Dipierro G., Price D., Laibe G., Hirsh K., Cerioli A., Lodato G., 2015, MNRAS, 453, L73 [pp. 4, 18, 57, and 114]

Dipierro G., Laibe G., Price D. J., Lodato G., 2016, MNRAS, 459, L1 [p. 57]

Dipierro G., et al., 2018, MNRAS, 475, 5296 [pp. 57, 83, 96, and 117]

Dittrich K., Klahr H., Johansen A., 2013, ApJ, 763, 117 [p. 6]

Dong R., Zhu Z., Rafikov R. R., Stone J. M., 2015a, ApJ, 809, L5 [pp. 7, 14]

Dong R., Hall C., Rice K., Chiang E., 2015b, ApJ, 812, L32 [pp. 7, 14]

Dong R., et al., 2018, ApJ, 860, 124 [p. 10]

Doğan S., Nixon C. J., King A. R., Pringle J. E., 2018, MNRAS, 476, 1519 [p. 136]

Duffell P. C., MacFadyen A. I., 2013, ApJ, 769, 41 [p. 21]

Duffell P. C., Haiman Z., MacFadyen A. I., D'Orazio D. J., Farris B. D., 2014, ApJ, 792, L10 [p. 21]

Dullemond C. P., et al., 2018, ApJ, 869, L46 [p. 57]

Dullemond C. P., Kimmig C. N., Zanazzi J. J., 2022, MNRAS, 511, 2925 [p. 121]

Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, Protostars and Planets V, p. 607 [pp. 66, 87, and 124]

Dürmann C., Kley W., 2015, A&A, 574, A52 [p. 21]

Dürmann C., Kley W., 2017, A&A, 598, A80 [p. 21]

Ercolano B., Pascucci I., 2017, Royal Society Open Science, 4, 170114 [p. 7]

Espaillat C., Calvet N., D'Alessio P., Hernández J., Qi C., Hartmann L., Furlan E., Watson D. M., 2007, ApJ, 670, L135 [p. 7]

Espaillat C., et al., 2014, Protostars and Planets VI, p. 497 [p. 7]

Facchini S., Lodato G., Price D. J., 2013, MNRAS, 433, 2142 [pp. 9, 23]

Facchini S., Juhász A., Lodato G., 2018, MNRAS, 473, 4459 [p. 9]

Facchini S., et al., 2020, A&A, 639, A121 [p. 93]

Fedele D., et al., 2018, A&A, 610, A24 [pp. 83, 96, and 117]

Fischer D. A., Marcy G. W., 1992, ApJ, 396, 178 [p. 23]

Flagg L., Johns-Krull C. M., Nofi L., Llama J., Prato L., Sullivan K., Jaffe D. T., Mace G., 2019, ApJ, 878, L37 [p. 80]

Flock M., Ruge J. P., Dzyurkevich N., Henning T., Klahr H., Wolf S., 2015, A&A, 574, A68 [p. 7]

Forgan Duncan & Rice K., 2013, MNRAS, 432, 3168 [p. 81]

Forgan D., Rice K., 2009, MNRAS, 400, 2022 [p. 118]

Forgan D. H., Hall C., Meru F., Rice W. K. M., 2018a, MNRAS, 474, 5036 [p. 81]

Forgan D. H., Ilee J. D., Meru F., 2018b, ApJ, 860, L5 [pp. 7, 14, and 15]

Frankowski A., Jancart S., Jorissen A., 2007, A&A, 464, 377 [p. 23]

Galvagni M., Mayer L., 2013, MNRAS, 437, 2909 [p. 81]

Gammie C. F., 2001, ApJ, 553, 174 [pp. 19, 20, 49, 61, 122, and 130]

Garg H., et al., 2021, MNRAS, 504, 782 [p. 8]

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375 [p. 29]

Ginski C., et al., 2021, ApJ, 908, L25 [pp. 9, 10]

Glover S. C. O., Mac Low M.-M., 2007, ApJS, 169, 239 [p. 51]

Goldreich P., Tremaine S., 1980, ApJ, 241, 425 [p. 21]

Goldreich P., Lithwick Y., Sari R., 2004, ApJ, 614, 497 [p. 17]

Goodman J., Rafikov R. R., 2001, ApJ, 552, 793 [pp. 89, 101]

Greaves J. S., Richards A. M. S., Rice W. K. M., Muxlow T. W. B., 2008, MNRAS, 391, L74 [p. 97]

Guzmán V. V., et al., 2018, ApJ, 869, L48 [p. 6]

Hall C., Rice K., Dipierro G., Forgan D., Harries T., Alexander R., 2018, MNRAS, 477, 1004 [p. 15]

Hall C., et al., 2020, ApJ, 904, 148 [pp. 14, 41, 52, 56, 93, 110, and 112]

Hatchett S. P., Begelman M. C., Sarazin C. L., 1981, ApJ, 247, 677 [p. 121]

Helled R., et al., 2014, Protostars and Planets VI, p. 643 [pp. 17, 18]

Hernquist L., Katz N., 1989, ApJS, 70, 419 [p. 36]

Heywood I., Avison A., Williams C. J., 2011, arXiv e-prints, p. arXiv:1106.3516 [pp. 86, 101]

Hogerheijde M. R., van der Tak F. F. S., 2000, A&A, 362, 697 [p. 56]

Horne K., Marsh T. R., 1986, MNRAS, 218, 761 [p. 11]

Huang J., et al., 2018a, ApJ, 869, L42 [pp. 57, 83, 93, 96, 113, and 117]

Huang J., et al., 2018b, ApJ, 869, L43 [pp. 9, 117]

Ilee J. D., Cyganowski C. J., Brogan C. L., Hunter T. R., Forgan D. H., Haworth T. J., Clarke C. J., Harries T. J., 2018, ApJ, 869, L24 [p. 23]

Isella A., et al., 2016, Phys. Rev. Lett., 117, 251101 [p. 12]

Isella A., et al., 2018, ApJ, 869, L49 [p. 10]

Johansen A., Youdin A., 2007, ApJ, 662, 627 [p. 18]

Kanagawa K. D., Muto T., Tanaka H., Tanigawa T., Takeuchi T., Tsukagoshi T., Momose M., 2015, ApJ, 806, L15 [p. 21]

Kanagawa K. D., Tanaka H., Szuszkiewicz E., 2018, ApJ, 861, 140 [p. 21]

Kanagawa K. D., Muto T., Tanaka H., 2021, ApJ, 921, 169 [p. 6]

Kimura S. S., Kunitomo M., Takahashi S. Z., 2016, MNRAS, 461, 2257 [p. 17]

Kley W., 1999, MNRAS, 303, 696 [p. 21]

Kley W., Nelson R. P., 2012, ARA&A, 50, 211 [pp. 6, 20, 97, and 106]

Kratter K. M., Matzner C. D., Krumholz M. R., Klein R. I., 2010a, ApJ, 708, 1585 [p. 23]

Kratter K. M., Murray-Clay R. A., Youdin A. N., 2010b, ApJ, 710, 1375 [p. 81]

Kraus S., et al., 2017, ApJ, 848, L11 [p. 10]

Kraus S., et al., 2020, Science, 369, 1233 [p. 9]

Kuffmeier M., Dullemond C. P., Reissl S., Goicovic F. G., 2021, A&A, 656, A161 [p. 139]

Kurosawa R., Harries T. J., Bate M. R., Symington N. H., 2004, MNRAS, 351, 1134 [p. 4]

Kwok S., 1975, ApJ, 198, 583 [p. 4]

Lada C. J., 1987, Symposium - International Astronomical Union, 115, 1–18 [p. 2]

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57 [p. 23]

Laibe G., Price D. J., 2012, MNRAS, 420, 2365 [p. 4]

Larwood J. D., Nelson R. P., Papaloizou J. C. B., Terquem C., 1996, MNRAS, 282, 597 [p. 9]

Lin D. N. C., Papaloizou J., 1979, MNRAS, 186, 799 [pp. 6, 20]

Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846 [pp. 6, 21]

Lissauer J. J., Hubickyj O., D'Angelo G., Bodenheimer P., 2009, Icarus, 199, 338 [p. 17]

Liu Y., et al., 2019, A&A, 622, A75 [p. 93]

Lodato G., Price D. J., 2010, MNRAS, 405, 1212 [pp. 41, 120, 121, and 123]

Lodato G., Pringle J. E., 2007, MNRAS, 381, 1287 [pp. 25, 119, and 120]

Lodato G., Meru F., Clarke C. J., Rice W. K. M., 2007, MNRAS, 374, 590 [p. 118]

Lodato G., et al., 2019, MNRAS, 486, 453 [pp. 16, 80, and 141]

Long F., et al., 2018, ApJ, 869, 17 [pp. 3, 6, 8, 57, and 96]

Longarini C., Lodato G., Toci C., Veronesi B., Hall C., Dong R., Patrick Terry J., 2021, ApJ, 920, L41 [p. 112]

Lovelace R. V. E., Li H., Colgate S. A., Nelson A. F., 1999, ApJ, 513, 805 [p. 8]

Lubow S. H., Ogilvie G. I., 2000, ApJ, 538, 326 [p. 130]

Lubow S. H., Seibert M., Artymowicz P., 1999, ApJ, 526, 1001 [p. 21]

Lucy L. B., 1977, AJ, 82, 1013 [p. 29]

Lucy L. B., 1999, A&A, 344, 282 [p. 54]

Luo Q.-y., et al., 2022, ApJ, 931, 158 [p. 23]

Lyra W., Johansen A., Klahr H., Piskunov N., 2009, A&A, 493, 1125 [p. 8]

Malik M., Meru F., Mayer L., Meyer M., 2015, ApJ, 802, 56 [pp. 22, 58, 59, 60, 78, 81, 102, 110, and 142]

Manara C. F., Testi L., Natta A., Rosotti G., Benisty M., Ercolano B., Ricci L., 2014, A&A, 568, A18 [p. 7]

van der Marel N., Verhaar B. W., van Terwisga S., Merín B., Herczeg G., Ligterink N. F. W., van Dishoeck E. F., 2016, A&A, 592, A126 [p. 7]

van der Marel N., et al., 2018a, ApJ, 854, 177 [p. 8]

van der Marel N., Williams J. P., Bruderer S., 2018b, ApJ, 867, L14 [p. 7]

van der Marel N., Dong R., di Francesco J., Williams J. P., Tobin J., 2019, ApJ, 872, 112 [p. 7]

van der Marel N., et al., 2021, AJ, 161, 33 [p. 8]

Marino S., Perez S., Casassus S., 2015, ApJ, 798, L44 [p. 9]

Martin R. G., et al., 2019, ApJ, 875, 5 [pp. 25, 121]

Masset F. S., 2001, ApJ, 558, 453 [p. 21]

Masset F. S., Papaloizou J. C. B., 2003, ApJ, 588, 494 [p. 22]

Masset F., Snellgrove M., 2001, MNRAS, 320, L55 [p. 21]

Matthews B. C., Krivov A. V., Wyatt M. C., Bryden G., Eiroa C., 2014, Protostars and Planets VI, p. 521 [p. 4]

Mayer L., Peters T., Pineda J. E., Wadsley J., Rogers P., 2016, ApJ, 823, L36 [p. 86]

McClure M. K., et al., 2016, ApJ, 831, 167 [p. 15]

Mercer A., Stamatellos D., Dunhill A., 2018, MNRAS, 478, 3478 [pp. 80, 94, 115, and 135]

Meru F., 2015, MNRAS, 454, 2529 [p. 94]

Meru F., Bate M. R., 2012, MNRAS, 427, 2022 [p. 45]

Meru F., Juhász A., Ilee J. D., Clarke C. J., Rosotti G. P., Booth R. A., 2017, ApJ, 839, L24 [pp. 7, 15]

Meru F., Rosotti G. P., Booth R. A., Nazari P., Clarke C. J., 2019, MNRAS, 482, 3678 [p. 6]

Miranda R., Rafikov R. R., 2019, ApJ, 878, L9 [p. 6]

Monaghan J. J., 1992, ARA&A, 30, 543 [pp. 29, 32, 44, and 50]

Monaghan J. J., 2005, Reports on Progress in Physics, 68, 1703 [p. 29]

Monaghan J. J., Lattanzio J. C., 1986, A&A, 158, 207 [p. 50]

Monnier J. D., et al., 2017, ApJ, 838, 20 [p. 6]

Morris J. P., Monaghan J. J., 1997, Journal of Computational Physics, 136, 41 [pp. 44, 45]

Morrison P. J., 1998, Reviews of Modern Physics, 70, 467 [p. 38]

Nakagawa Y., Nakazawa K., Hayashi C., 1981, Icarus, 45, 517 [p. 5]

Nazari P., Booth R. A., Clarke C. J., Rosotti G. P., Tazzari M., Juhasz A., Meru F., 2019, MNRAS, 485, 5914 [p. 6]

Nealon R., Price D. J., Nixon C. J., 2015, MNRAS, 448, 1526 [pp. 99, 122]

Nealon R., Dipierro G., Alexander R., Martin R. G., Nixon C., 2018, MNRAS, 481, 20 [p. 9]

Nealon R., Cuello N., Alexander R., 2020a, MNRAS, 491, 4108 [pp. 23, 41]

Nealon R., Price D. J., Pinte C., 2020b, MNRAS, 493, L143 [pp. 9, 56]

Nixon C., King A., Price D., 2013, MNRAS, 434, 1946 [p. 9]

Offner S. S. R., Moe M., Kratter K. M., Sadavoy S. I., Jensen E. L. N., Tobin J. J., 2022, arXiv e-prints, p. arXiv:2203.10066 [p. 23]

Ogilvie G. I., 1999, MNRAS, 304, 557 [pp. 120, 121]

Ogilvie G. I., 2000, MNRAS, 317, 607 [p. 121]

Ogilvie G. I., Lubow S. H., 2002, MNRAS, 330, 950 [p. 6]

Paardekooper S.-J., Dong R., Duffell P., Fung J., Masset F. S., Ogilvie G., Tanaka H., 2022, arXiv e-prints, p. arXiv:2203.09595 [pp. 20, 22]

Paneque-Carreño T., et al., 2021, ApJ, 914, 88 [pp. 8, 14, and 136]

Papaloizou J. C. B., Terquem C., 1995, MNRAS, 274, 987 [p. 121]

Pepliński A., Artymowicz P., Mellema G., 2008a, MNRAS, 386, 164 [p. 22]

Pepliński A., Artymowicz P., Mellema G., 2008b, MNRAS, 386, 179 [p. 22]

Pepliński A., Artymowicz P., Mellema G., 2008c, MNRAS, 387, 1063 [p. 22]

Perez S., Dunhill A., Casassus S., Roman P., Szulágyi J., Flores C., Marino S., Montesinos M., 2015, ApJ, 811, L5 [p. 12]

Pérez L. M., et al., 2016, Science, 353, 1519 [pp. 58, 83, 93, 96, 113, and 117]

Pérez L. M., et al., 2018, ApJ, 869, L50 [p. 10]

Pérez S., Casassus S., Baruteau C., Dong R., Hales A., Cieza L., 2019, AJ, 158, 15 [p. 6]

Persson M. V., 2014a, doi:10.6084/m9.figshare.654555.v7 [p. 2]

Persson M. V., 2014b, doi:10.6084/m9.figshare.1121574.v2 [p. 3]

Pfalzner S., 2013, A&A, 549, A82 [pp. 23, 118]

Pinilla P., et al., 2017, ApJ, 839, 99 [p. 8]

Pinilla P., et al., 2018a, ApJ, 859, 32 [p. 8]

Pinilla P., et al., 2018b, ApJ, 868, 85 [pp. 8, 11]

Pinte C., 2015, EPJ Web of Conferences, 102, 00006 [p. 53]

Pinte C., Ménard F., Duchêne G., Bastien P., 2006, A&A, 459, 797 [pp. 52, 85, and 100]

Pinte C., Harries T. J., Min M., Watson A. M., Dullemond C. P., Woitke P., Ménard F., Durán-Rojas M. C., 2009, A&A, 498, 967 [pp. 52, 85, and 100]

Pinte C., et al., 2018a, A&A, 609, A47 [pp. 86, 101]

Pinte C., et al., 2018b, ApJ, 860, L13 [pp. 12, 91, and 110]

Pinte C., et al., 2019, Nature Astronomy, 3, 1109 [pp. 12, 13, 91, and 110]

Pinte C., et al., 2020, ApJ, 890, L9 [pp. 12, 17, 91, 97, 110, and 141]

Pinte C., Teague R., Flaherty K., Hall C., Facchini S., Casassus S., 2022, arXiv e-prints, p. arXiv:2203.09528 [pp. 11, 12, and 16] van der Plas G., Ménard F., Canovas H., Avenhaus H., Casassus S., Pinte C., Caceres C., Cieza L., 2017, A&A, 607, A55 [p. 8]

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62 [p. 17]

Powell D., Murray-Clay R., Pérez L. M., Schlichting H. E., Rosenthal M., 2019, ApJ, 878, 116 [p. 113]

Price D. J., 2012, Journal of Computational Physics, 231, 759 [pp. 30, 32, and 37]

Price D. J., Federrath C., 2010, MNRAS, 406, 1659 [p. 45]

Price D. J., Monaghan J. J., 2007, MNRAS, 374, 1347 [p. 42]

Price D. J., et al., 2018a, PASA, 35, e031 [pp. 28, 41, 43, 51, 60, 84, 85, 98, 99, 121, and 122]

Price D. J., et al., 2018b, MNRAS, 477, 1270 [pp. 7, 8, and 41]

Pringle J. E., 1981, ARA&A, 19, 137 [p. 1]

Pringle J. E., 1996, MNRAS, 281, 357 [p. 122]

Rafikov R. R., 2005, ApJ, 621, L69 [pp. 58, 63, 85, 99, 136, and 142]

Rafikov R. R., 2016, ApJ, 831, 122 [pp. 89, 101]

Raghavan D., et al., 2010, ApJS, 190, 1 [p. 23]

Ragusa E., Dipierro G., Lodato G., Laibe G., Price D. J., 2017, MNRAS, 464, 1449 [p. 8]

Ragusa E., Alexander R., Calcino J., Hirsh K., Price D. J., 2020, MNRAS, 499, 3362 [p. 8]

Ragusa E., et al., 2021, MNRAS, 507, 1157 [p. 8]

Reboussin L., Wakelam V., Guilloteau S., Hersant F., Dutrey A., 2015, A&A, 579, A82 [pp. 16, 97]

Ren B., et al., 2018, ApJ, 857, L9 [p. 8]

Rice W. K. M., Armitage P. J., 2009, MNRAS, 396, 2228 [pp. 58, 63, 85, 99, 136, and 142]

Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2004, MNRAS, 355, 543 [p. 140]

Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2006, MNRAS, 372, L9 [pp. 6, 140]

Rowther S., Meru F., 2020, MNRAS, 496, 1598 [pp. ix, 84, 85, 86, 98, 99, 102, and 110]

Rowther S., Meru F., Kennedy G. M., Nealon R., Pinte C., 2020, ApJ, 904, L18 [pp. ix, 97, 98, 101, 105, 106, and 118]

Rowther S., Nealon R., Meru F., 2022, ApJ, 925, 163 [pp. ix, 97, and 104]

Rowther S., Nealon R., Meru F., 2023, MNRAS, 518, 763 [p. ix]

Sakai N., Hanawa T., Zhang Y., Higuchi A. E., Ohashi S., Oya Y., Yamamoto S., 2019, Nature, 565, 206 [pp. 23, 118]

Schoenberg I. J., 1946, Quarterly of Applied Mathematics, 4, 45 [p. 31]

Segura-Cox D. M., et al., 2020, Nature, 586, 228 [pp. 3, 18, 83, 96, and 117]

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337 [p. 122]

Sheehan P. D., Eisner J. A., 2018, ApJ, 857, 18 [pp. 83, 96, and 117]

Siess L., Dufour E., Forestini M., 2000, A&A, 358, 593 [pp. 85, 100]

Simon J. B., Armitage P. J., Li R., Youdin A. N., 2016, ApJ, 822, 55 [p. 18]

Speedie J., Booth R. A., Dong R., 2022, ApJ, 930, 40 [pp. 115, 141]

Stamatellos D., 2015, ApJ, 810, L11 [pp. 58, 142]

Stamatellos D., Inutsuka S.-i., 2018, MNRAS, 477, 3110 [pp. 22, 58, 78, and 142]

Stamatellos D., Whitworth A. P., 2009, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 1094, 557 [pp. 58, 63, 81, 85, 99, 136, and 142]

Stolker T., et al., 2017, ApJ, 849, 143 [p. 9]

Suriano S. S., Li Z.-Y., Krasnopolsky R., Suzuki T. K., Shang H., 2019, MNRAS, 484, 107 [p. 6]

Suzuki T. K., Ogihara M., Morbidelli A., Crida A., Guillot T., 2016, A&A, 596, A74 [p. 7]

Takahashi S. Z., Muto T., 2018, ApJ, 865, 102 [p. 7]

Terry J. P., Hall C., Longarini C., Lodato G., Toci C., Veronesi B., Paneque-Carreño T., Pinte C., 2022, MNRAS, 510, 1671 [pp. 14, 41, 52, 56, 110, and 112]

van Terwisga S. E., et al., 2018, A&A, 616, A88 [p. 6]

Testi L., et al., 2014, Protostars and Planets VI, p. 339 [pp. 4, 5, 15, and 97]

Thies I., Kroupa P., Goodwin S. P., Stamatellos D., Whitworth A. P., 2010, ApJ, 717, 577 [p. 118]

Tobin J. J., et al., 2016a, Nature, 538, 483 [pp. 23, 138]

Tobin J. J., et al., 2016b, ApJ, 818, 73 [p. 58]

Toomre A., 1964, ApJ, 139, 1217 [pp. 20, 59, 86, 106, and 124]

Trapman L., Miotello A., Kama M., van Dishoeck E. F., Bruderer S., 2017, A&A, 605, A69 [p. 15]

Trapman L., Zhang K., van't Hoff M. L. R., Hogerheijde M. R., Bergin E. A., 2022, ApJ, 926, L2 [p. 16]

Tsukagoshi T., et al., 2019, ApJ, 871, 5 [p. 8]

Tychoniec Ł., et al., 2020, A&A, 640, A19 [p. 3]

Veronesi B., Paneque-Carreño T., Lodato G., Testi L., Pérez L. M., Bertin G., Hall C., 2021, ApJ, 914, L27 [p. 114]

Verrios H. J., Price D. J., Pinte C., Hilder T., Calcino J., 2022, ApJ, 934, L11 [pp. 14, 111]

Villenave M., et al., 2019, A&A, 624, A7 [p. 6]

Vincke K., Pfalzner S., 2016, ApJ, 828, 48 [pp. 23, 118]

Walsh C., Daley C., Facchini S., Juhász A., 2017, A&A, 607, A114 [p. 118]

Ward W. R., 1991, Lunar and Planetary Science Conference, 22, 1463 [p. 21]

Weidenschilling S. J., 1977, MNRAS, 180, 57 [pp. 4, 18]

Weidenschilling S. J., 1980, Icarus, 44, 172 [p. 5]

Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67 [p. 17]

Windmark F., Birnstiel T., Güttler C., Blum J., Dullemond C. P., Henning T., 2012, A&A, 540, A73 [p. 18]

Winter A. J., Clarke C. J., Rosotti G., Ih J., Facchini S., Haworth T. J., 2018, MNRAS, 478, 2700 [p. 23]

Wu C.-J., Hirano N., Takakuwa S., Yen H.-W., Aso Y., 2018, ApJ, 869, 59 [p. 80]

Xiang-Gruess M., 2016, MNRAS, 455, 3086 [p. 138]

Xiang-Gruess M., Papaloizou J. C. B., 2014, MNRAS, 440, 1179 [p. 23]

Youdin A. N., Goodman J., 2005, ApJ, 620, 459 [p. 18]

Youdin A., Johansen A., 2007, ApJ, 662, 613 [p. 18]

van Zadelhoff G. J., van Dishoeck E. F., Thi W. F., Blake G. A., 2001, A&A, 377, 566 [pp. 16, 97]

Zhang K., Blake G. A., Bergin E. A., 2015, ApJ, 806, L7 [p. 7]

Zhang S., et al., 2018, ApJ, 869, L47 [pp. 6, 16, 57, 97, and 141]

Zhu Z., Nelson R. P., Hartmann L., Espaillat C., Calvet N., 2011, ApJ, 729, 47 [p. 7]

Zhu Z., Hartmann L., Nelson R. P., Gammie C. F., 2012a, ApJ, 746, 110 [p. 81]

Zhu Z., Nelson R. P., Dong R., Espaillat C., Hartmann L., 2012b, ApJ, 755, 6 [p. 6]

Zhu Z., Dong R., Stone J. M., Rafikov R. R., 2015, ApJ, 813, 88 [p. 7]

Ziampras A., Ataiee S., Kley W., Dullemond C. P., Baruteau C., 2020, A&A, 633, A29 [pp. 89, 101]