THE UNIVERSITY *of* EDINBURGH

Planet formation and the early evolution of self-gravitating protoplanetary discs

James Cadman

Doctor of Philosophy
The University of Edinburgh
June 2022

Abstract

When a Giant Molecular Cloud (GMC) collapses to form a stellar core, conservation of angular momentum will lead to the formation of a protoplanetary disc, with an initial mass potentially of the order of its stellar host. If a massive disc forms, then the disc's self-gravity will play a crucial role in the earliest stages of its evolution; driving its viscous evolution, and potentially leading to the formation of wide orbit, giant planets and brown dwarfs through disc fragmentation.

I begin this thesis by placing improved constraints on the conditions required for disc fragmentation, specifically focusing on how the disc's environment may influence its evolution and eventual fate.

Recent results from direct imaging surveys suggest that wide orbit giant planets and brown dwarfs are found more frequently around higher mass stars. I use Smoothed Particle Hydrodynamics (SPH) simulations to show that a disc's susceptibility to fragmentation is dependent on the mass of its host star. I demonstrate that discs around higher mass stars may fragment for lower disc-to-star mass ratios, making them favourable sites for the formation of wide orbit, massive objects, such as those found in direct imaging surveys. Low mass stars may support high mass discs, in principle providing large reservoirs of material for core accretion planet formation.

Results from direct imaging surveys also find that stars hosting close in giant planets or brown dwarfs display an excess of outer binary companions, with indications that some of these objects may have formed through the gravitational instability (GI). I use SPH to simulate a suite of self-gravitating discs with a binary companion, and show that there is a narrow region of parameter space where intermediate separation companions may trigger fragmentation. Short separation encounters are destructive, whilst wide orbit companions have little

effect. The range of binary separations found to favour the formation of short period, giant planets is consistent with results from direct imaging surveys.

Although numerical models suggest that GI may dominate a disc's early evolution, it is still unclear from observations whether massive, self-gravitating discs exist in nature. Recent high-resolution infrared imaging of protoplanetary discs have given rise to unparalleled observations of their substructure, including rings, gaps and spirals, providing us with crucial insights to the earliest stages of planet formation.

Observations of the protoplanetary disc surrounding AB Aurigae have revealed the possible presence of two massive planets in the process of forming. The young measured age for the system places strict time constraints on the planet's formation histories. I use analytic core accretion models to show that their expected core accretion formation timescales are longer than the system's current age. Using SPH and viscous evolution models of self-gravitating discs, I show that a proto-AB Aurigae disc could have been massive enough to fragment in the past, with typical fragment masses consistent with the masses of the protoplanets which have been observed in the disc.

Finally, I use Monte Carlo radiative transfer models to generate observational predictions of self-gravitating discs using ALMA. I develop an existing 3D semi-analytic model to include a prescription for dust trapping in the disc's spirals. I make predictions about the disc properties which may drive spirals that could be visible to ALMA, in particular focusing on the impact of dust trapping. I also use these models to analyse 3 discs from the DSHARP survey, and discuss the plausibility of their observed spirals being the result of GI.

Lay Summary

Throughout human history we have attempted to understand the Universe around us. Some of the earliest astronomical theories focused on making sense of the objects which are nearest to us and the most easily visible; those being the Sun, the Moon and the nearby Solar System planets. Initially, motivated by the observation that these objects move across the night sky in an elliptical motion, astronomers believed that all of the Solar System objects orbited around the Earth, and that the Earth was at the center of the Universe. It wasn't until after the 16^{th} century it became widely accepted that, in fact, all of the Solar System planets were in orbit around the Sun. As it was also true that the planets orbit the Sun in the same plane and in the same direction as each other, it was hypothesised they must have all formed in some sort of disc-like structure around the young, newly-formed Sun.

These *protoplanetary discs* are what we now believe to be the formation sites of all of the known planets. The most widely accepted theory of how planets form, known as the *core accretion theory*, suggests that grains of dust within these discs will collide with each other, stick together and slowly grow over millions of years to form planets potentially more massive than Jupiter. Core accretion theory is generally successful in explaining the formation of all of the planets in our Solar System, as well as most of the known planets which we have discovered around other stars, known as *exoplanets*. However, to date we have discovered over 5,000 exoplanets, lots of which have very different properties to the planets in our Solar System, and whose origins can be challenging to explain through core accretion theory.

Of particular importance to the work in this thesis is the formation of giant planets (with masses greater than that of Jupiter) at large distances from their stars (often much further than the outer reaches of our Solar System). The formation of these wide-orbit, giant planets is difficult to explain through core

accretion theory. The time that it would take for them to form would be much longer than the typical lifetimes of protoplanetary discs, as discs are only expected to survive for around 10 million years before all of the material has either been accreted onto the star or has been blown away by the star's radiation.

The gravitational instability theory is a complementary model of planet formation to core accretion theory, and it may be able to explain the origin of these wide-orbit, giant planets. It suggests that the outer regions of a protoplanetary disc may collapse (or *fragment*) to form gravitationally bound clumps of gas, which may go on to become gas giant planets. This gravitational collapse of the disc may only occur if the disc has a mass comparable to the mass of its host star, such that the gravitational potential of the disc is comparable to that of the star. Hence it is only likely to occur whilst the disc is very young and before its mass has been depleted.

However, this mode of planet formation is still very poorly understood. Firstly, it is uncertain whether discs will ever be massive enough to be gravitationally unstable, as we are still yet to find concrete evidence of gravitationally unstable discs existing in nature. Even if sufficiently massive discs do exist, it is uncertain how often, if at all, fragmentation can occur and the specific disc conditions which would be required, such as their masses, sizes and temperatures. Further, if the disc does fragment, it is uncertain whether the clumps which form will go on to become planets, or whether instead they can only ever become objects too massive to be planets but too low mass to be stars, known as *Brown Dwarfs*.

Throughout this thesis I will use computer simulations to study the role of the gravitational instability in the early evolution of protoplanetary discs. In Chapters 1 and 2 I will describe in detail our current understanding of star and planet formation, derive some key equations and outline some important numerical methods which will be used throughout the research here. In Chapters 3 and 4 I will place improved constraints on the disc's environmental factors which may provide favourable conditions for the disc to undergo fragmentation. In Chapter 5 I will analyze the AB Aurigae system, focusing on the formation histories of the two protoplanets which have recently been observed in the disc and evaluating the possibility that they may have formed through the gravitational instability. Finally, in Chapter 6 I will generate simulated observations of gravitationally unstable discs with the aim of providing improved predictions of the disc conditions required if we are to successfully observe these systems with today's telescopes.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of the work presented in this thesis have been published or will be submitted for publication in:

1. Cadman J., Rice K., Hall C., Haworth T. J., Biller B., 2020, **Fragmentation favoured in discs around higher mass stars**, MNRAS, 492, 5041 (Chapter 3)
2. Haworth T. J., Cadman J., Meru F., Hall C., Albertini E., Forgan D., Rice K.,Owen J. E., 2020, **Massive discs around low mass stars**, MNRAS, 494, 413017 (Chapter 3)
3. Cadman J., Hall C., Fontanive C., Rice K., 2022, **Binary companions triggering fragmentation in self-gravitating discs**, MNRAS, 511, 457, (Chapter 4)
4. Cadman J., Rice K., Hall C., 2021, **AB Aurigae: Possible evidence of planet formation through the gravitational instability**, MNRAS, 504, 2877 (Chapter 5)
5. Cadman J., Hall C., Rice K., Harries T. J., Klaassen P. D., 2020, **The observational impact of dust trapping in self-gravitating discs**, MNRAS, 498, 4256 (Chapter 6)

.

(*James Cadman, June 2022*)

Acknowledgements

The work presented in this thesis would not have been remotely possible without the incredible supervision of Professor Ken Rice. His inspired thoughts and research ideas, combined with his supervision style which has allowed me to work with freedom yet careful guidance, have enabled me to generate this body of work largely without hindrance. I am eternally grateful for the hours spent explaining things to me (both related to physics and how to hit a golf ball properly), the conferences he encouraged me to attend and the collaborators he enabled me to work with.

Of the many collaborators I have had the pleasure of working with, there are a few I would like to give a special mention to. I am grateful to Thomas Haworth who assisted me a lot with my very first publication (also allowing me to contribute to his own work) and has since publicised our work at various conferences and meetings over the past few years. I am also grateful to Cleménce Fontanive, without whom there would have been no Chapter 4. Clem provided the research question for arguably the best chapter in this thesis, whilst also contributing to the writing of the paper as well as many hours of discussion on the topic. Finally, I would of course like to give an extra special mention to Cassandra Hall. Cass has contributed research ideas, code, proof reading of papers, and lots of advice at every stage throughout my PhD. She was kind enough to take me under her wing on the very first day when I met her at a conference in Edinburgh, and has since become a good friend who has helped me immensely over the past 4 years (as can be seen from her feature on the author list of every one of my papers).

I would also like to thank my family, in particular my parents. As well as being generally faultless parents, they have contributed significantly to my pursuit of science. I would like to thank my Mum for forcing me to revise and laboriously drawing up revision timetables for me throughout school no matter how much I stubbornly refused. This, as well as her knowledge and love of maths, have undoubtedly driven me to reach this point. My Dad for his excitement and enthusiasm towards science, particularly towards astronomy. He has relentlessly sent me articles and pushed his bonkers theories since I first showed an interest in astronomy. As well as this, he has instilled in me a strong work ethic, and encouraged me since a child to always chase after my own carrots.

Contents

List of Figures

List of Tables

Chapter 1

An overview of star and planet formation

1.1 Introduction

Astronomy is generally believed to be the oldest of the sciences, and can be traced back to most ancient civilisations. The earliest astronomical records date back to around 5000 years ago as humans attempted to use the motions of the stars, planets and the moon to keep track of things such as the yearly seasonal changes (of crucial importance for farming and agriculture) as well as for navigational purposes. Clear evidence of ancient civilisations utilising the motions of the heavenly bodies can still be seen today. Monolithic structures such as Stonehenge, dating back to $\sim 3000\,\mathrm{BC}$, demonstrate how the motion of the Sun was potentially used as an astronomical clock, whilst the ancient Egyptian pyramids, also dating to around the same time, are aligned to point directly north - a feat presumably only achievable through utilisation of the positions of the stars.

Fast forward a few millenia past huge astronomical milestones such as Copernicus' development of the heliocentric model of the Solar System in the 16^{th} century, the invention of the telescope in the early 17^{th} century, and Newton's theory of gravity in the late 17^{th} century, and astronomers were beginning to piece together the origins of our Solar System and how we got here. The earliest modern theories of star and planet formation can be traced back to the works of Kant and Laplace in the 18^{th} century, in what is now known as *the nebular hypothesis*. In 1755 Immanuuel Kant proposed that the Solar System formed from a dispersed cloud

of particles which would slowly coalesce and grow as they collided with each other. In 1756 Pierre-Simon Laplace took this further. Realising that all of the planets orbit the Sun in the same direction and in the same plane, he proposed that their formation must have occurred in a rotating disc-like structure. The nebular hypothesis suggested that the Sun was initially much larger, with an atmosphere extending beyond the outer Solar System, and that it was rotating. As the Sun cooled and contracted, conservation of angular momentum would result in the formation of a heliocentric disc, within which the collisional accumulation of particles proposed by Kant would ensue.

Since then, modern theories of star and planet formation have built on this. We now know that star formation occurs in extremely large, gravitationally bound clouds of gas and dust, know as *Giant Molecular Clouds* (GMCs). These structures can span anywhere between a few and a few hundred parsecs, with masses up to a few million solar masses. The most massive clouds may potentially form thousands of stars which may later exist as constituents of dense stellar clusters. Famous examples include the Perseus and Orion molecular clouds, which are pictured in Figure 1.1.

GMCs are typically cold, with temperatures $\sim 10\,\mathrm{K}$, hence their constituent gas will exist in its molecular form, of which the vast majority will be molecular hydrogen, H_2. The distribution of this gas within each cloud will be non-homogenous, consisting of complex filametary structures as can be seen in the clouds in Figure 1.1. At the knots of these filaments exist the densest regions of GMCs and the actual sites of star formation, *molecular cores*, with typical sizes $\sim 0.1\,\mathrm{pc}$. If a molecular core becomes massive enough it may begin to contract under the influence of its self-gravity to form a dense, central protostar, as illustrated in Figure 1.2. As it does so, and the core becomes progressively smaller, conservation of angular momentum dictates that the forming protostar's rotational velocity must increase. As the rotational velocity of the protostar continues to increase, centrifugal acceleration will drive the formation of a large, extended disc structure which rotates coincident with the axis of rotation of the collapsing region.

These *protostellar* or *protoplanetary discs*[1] are what will go on to become the

[1]Note: throughout this thesis I will use these terms somewhat interchangeably. *Protostellar disc* will generally refer to discs in their earliest stages of formation, whilst existing around a *protostar*. *Protoplanetary disc* will generally be used to refer to slightly more evolved systems, within which the planet formation process has begun to take place.

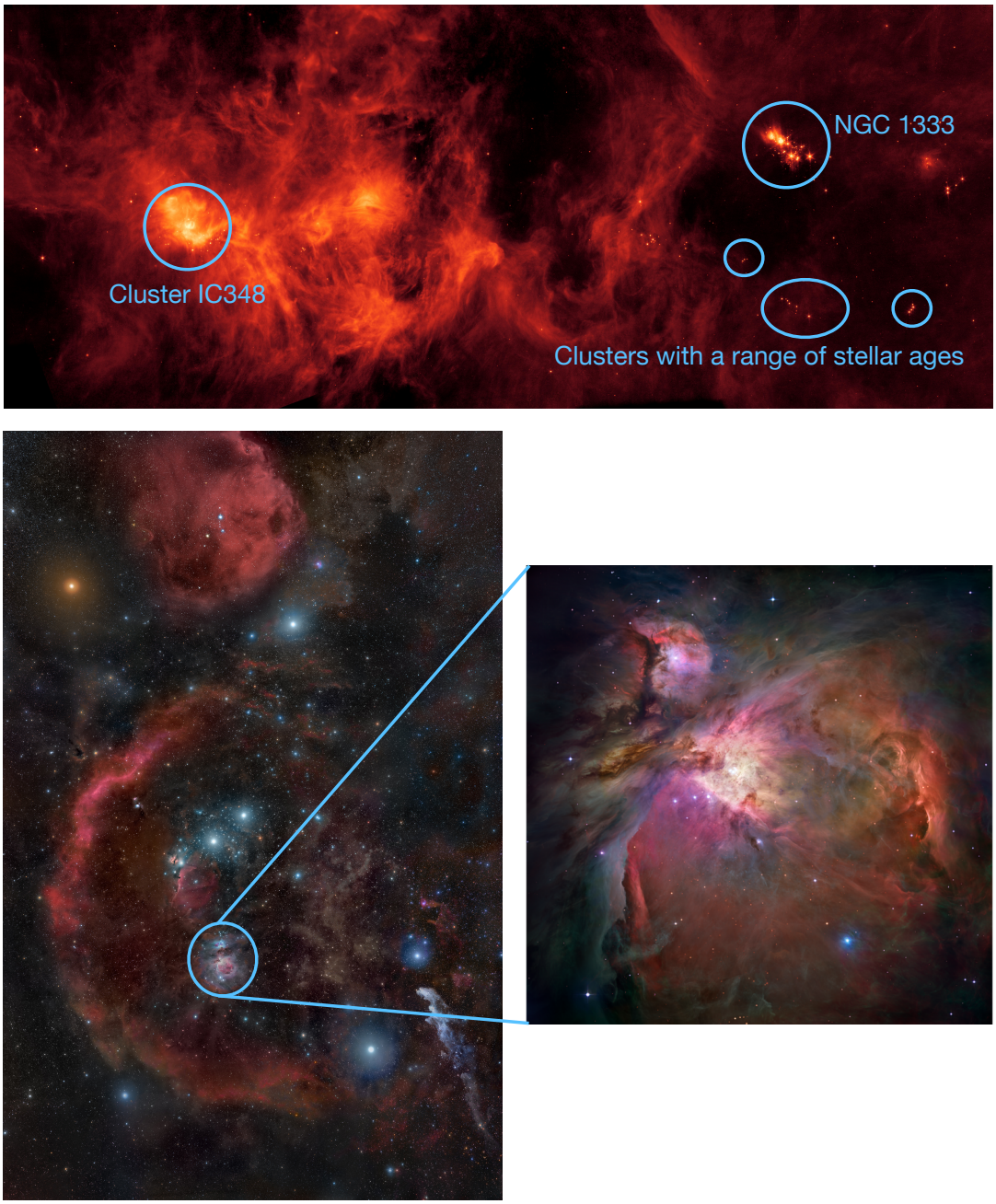

Figure 1.1 *Top: Perseus molecular cloud with some notable features highlighted. The cloud is* $\sim$ 150 *pc wide and contains* $\sim$ $10^4\,M_\odot$ *of gas and dust (Credit: NASA/JPL-Caltech). Bottom left: Orion molecular cloud, located in the constellation of Orion, with the location of the famous Orion nebula highlighted (Image credit: Rogelio B. Andreo, RBA Premium Astrophotography). Bottom right: Orion nebula. The image spans* $\sim$ 4 *pc and contains* $\sim$ 3000 *stars in different stages of formation (Image credit: NASA,ESA, M. Robberto (Space Telescope Science Institute/ESA) and the Hubble Space Telescope Orion Treasury Project Team).*

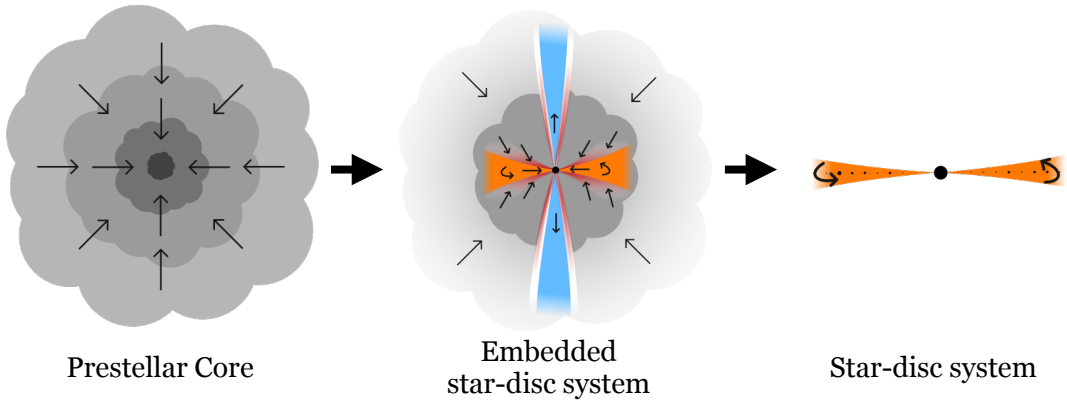

Figure 1.2 *An illustration of star-disc formation. Dense regions of a cloud may collapse to form the first prestellar cores. Conservation of angular momentum during the collapse results in the formation of extended protostellar discs (Image credit: M.V. Persson).*

sites of planet formation, and are consequently the primary research focus of this thesis. The first observational evidence for the existence of circumstellar discs came in 1984 when Aumann et al. (1984) found a large Infrared excess when observing Vega, indicative of there being a shell of solids surrounding the star. Shortly after, the first published image of a disc also came in 1984, when Smith & Terrile (1984) observed an edge-on disc-like structure around the star β-Pictoris, shown in Figure 1.3. Since then, our capabilities of observing these discs has vastly improved, with collaborations such as the Disc Substructures at High Angular Resolution Project (DSHARP) (Andrews et al., 2018b) having recently published some of the most spectacular images of protoplanetary discs to date (see Figure 1.4), revealing complex substructures indicative of the presence of planets and various instabilities.

Throughout this chapter I will outline some of the key processes which take place throughout the course of star and planet formation, beginning with the collapse of a molecular cloud to form a star-disc system, and concluding with our current understanding of how planets form within the resultant protoplanetary discs.

1.2 Gravitational collapse of a GMC and the formation of prestellar cores.

Over 50 years ago, Larson (1969) performed the first numerical simulations of a collapsing cloud, outlining the key phases of star formation. An initial phase of

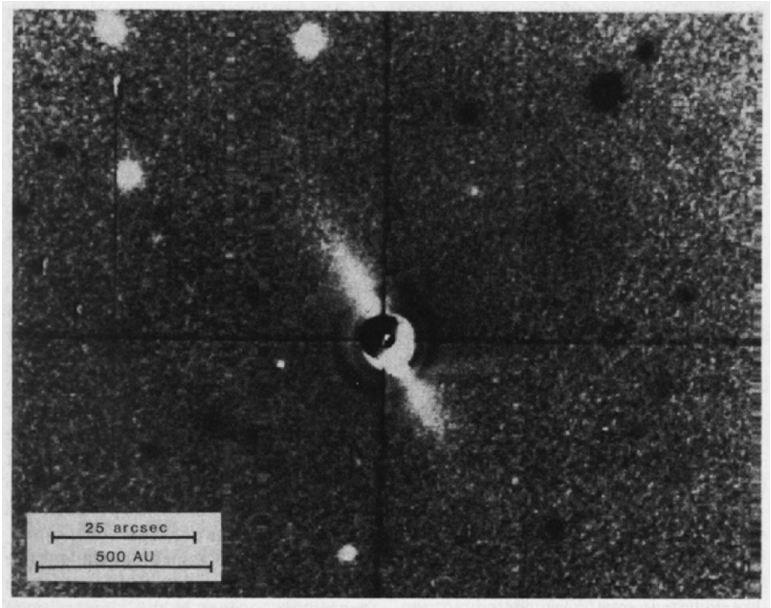

Figure 1.3 *First published image of a circumstellar disc obtained in 1984. Image shows $\beta-$Pictoris at the centre surrounded by an edge-on disc-like structure extending $\sim 400\,AU$ from the star. Image credit: Smith & Terrile (1984)*

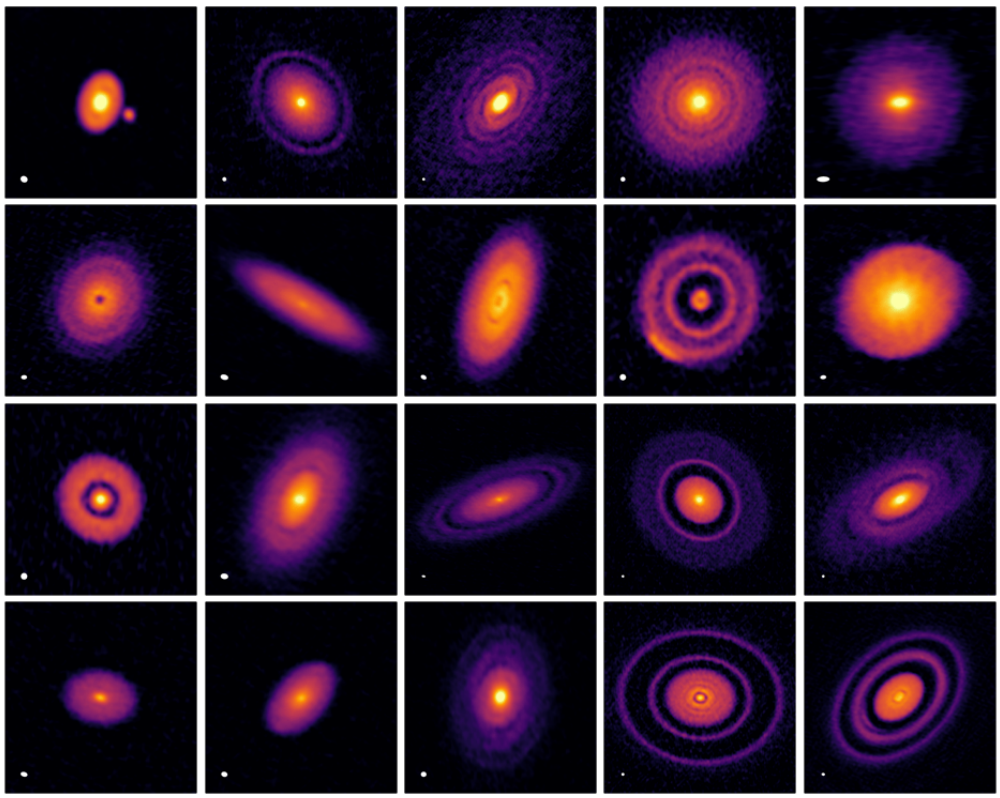

Figure 1.4 *The 20 discs imaged in the DSHARP data release (Andrews et al., 2018b). Rings, gaps and spirals can be seen in many of the discs, some of which are indicative of the presence of young planets which may have carved out such substructures.*

isothermal collapse will occur as the cloud contracts under its own gravitational potential. Collapse will continue until the density at the centre of the collapsing region increases such that the gas becomes optically thick, heats up, and halts any further contraction. These dense, optically thick central regions are known as the *first cores*. This process typically proceeds in a hierarchical manner, and the initial collapsing region will fragment to form many cores. The first cores will continue to accrete mass from the surrounding gas, and a phase of second collapse may occur once the first core heats up sufficiently to dissociate molecular hydrogen, thus proceeds the formation of the *second cores*. Collapse will continue in a manner similar to before, until most of the molecular hydrogen has been dissociated and collapse is halted again. The dense second cores will continue to accrete material, and their formation is followed by a long period of accretion and contraction as they evolve toward stellar densities.

The initial collapse of a cloud may occur if the mass within a given region exceeds some threshold value, known as the *Jeans mass*, M_{J} (Jeans, 1902). As a region of the cloud contracts, the gas density will increase thus providing pressure support acting to stabilise the region. However, if the mass within the collapsing region exceeds the Jeans mass, the timescale on which the cloud collapses (known as the free fall timescale, t_{ff}) will be shorter than the timescale on which pressure support is able to prevent it (given by the sound crossing timescale, t_{s}) hence collapse may continue to form the first prestellar cores.

The Jeans mass can be derived by considering the ratio of the free fall timescale and the sound crossing timescale. Beginning with Newton's second law for a uniform density cloud of mass, M, size, R, and density, ρ, we have,

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = -\frac{\mathrm{G}M}{R^2} = -\frac{\mathrm{G}}{R^2}\frac{4\pi R_0^3 \rho}{3}. \tag{1.1}$$

Substituting,

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}R}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}R} = v\frac{\mathrm{d}v}{\mathrm{d}R}, \tag{1.2}$$

and rearranging, we have,

$$\int v\,\mathrm{d}v = -\frac{4\pi \mathrm{G} R_0^3 \rho}{3}\int \frac{\mathrm{d}R}{R^2}, \tag{1.3}$$

$$\frac{1}{2}v^2 = \frac{4\pi \mathrm{G} R_0^3 \rho}{3R} + C. \tag{1.4}$$

Using the boundary conditions that $v = 0$ when $R = R_0$, we find that,

$$C = -\frac{4\pi \mathrm{G} R_0^3 \rho}{3}, \tag{1.5}$$

hence, rearranging gives,

$$|v| = \sqrt{\frac{8\pi \mathrm{G} R_0^2 \rho^2}{3}\left(\frac{R_0}{R} - 1\right)}. \tag{1.6}$$

This velocity represents the velocity with which the cloud would collapse if in free fall and subject to no opposing forces. The *free fall timescale* on which this collapse would happen can be calculated as,

$$t_{\mathrm{ff}} = \int_0^{R_0} \frac{\mathrm{d}r}{|v|}, \tag{1.7}$$

which gives,

$$t_{\mathrm{ff}} = \sqrt{\frac{3\pi}{32\mathrm{G}\rho}}. \tag{1.8}$$

However, as the cloud collapses its density and pressure will increase, and sound waves will propagate through the cloud to re-establish pressure balance. These will travel at the gas sound speed, c_{s}, on a timescale given by the sound crossing time, $t_{\mathrm{s}} = R/c_s$. Collapse will only be able to continue if $t_{\mathrm{ff}} < t_{\mathrm{s}}$, hence when,

$$\frac{R}{c_{\mathrm{s}}} < \sqrt{\frac{3\pi}{32\mathrm{G}\rho}}, \tag{1.9}$$

which gives,

$$R_{\mathrm{J}} = c_{\mathrm{s}}\sqrt{\frac{3\pi}{32\mathrm{G}\rho}}, \tag{1.10}$$

Here, we have arrived at the *Jeans length*, which represents the maximum size

of a density perturbation within a cloud of uniform density, ρ, which can be maintained by pressure support. Any region of the cloud smaller than this will continue to collapse to form the first prestellar cores. From the Jeans length, we can also calculate the Jeans mass as the mass of gas contained within the volume defined by the Jeans length,

$$M_{\rm J} = \frac{4}{3}\pi R_{\rm J}^3 \rho. \tag{1.11}$$

1.3 The formation of protoplanetary discs

Observations of GMCs find that their internal velocity dispersions, σ, follow approximate scaling relations with their masses and radii. These are known as *Larson's scaling relations* (Larson, 1981), which approximately give $\sigma \propto M^{0.25}$ and $\sigma \propto R^{0.5}$. The velocity dispersions within GMCs are dominated by turbulence, with turbulent eddys which operate on a large range of size scales. Turbulence is inherently rotational, so any turbulent region will posses some angular momentum which must be conserved as the region collapses to form stellar cores.

For a collapsing, rotating spherical cloud, its pre-collapse angular momentum will be,

$$J_{\rm c} = M_{\rm c} R_{\rm c}^2 \omega_{\rm c}, \tag{1.12}$$

where $M_{\rm c}$ and $R_{\rm c}$ are the masses and radii of the cloud and $\omega_{\rm c}$ is the cloud's angular velocity. The cloud will also have a rotational energy equal to,

$$E_{\rm rot} = \frac{1}{2} I_{\rm c} \omega_{\rm c}^2 = \frac{1}{5} M_{\rm c} R_{\rm c}^2 \omega_{\rm c}^2, \tag{1.13}$$

and a gravitational potential energy,

$$E_{\rm grav} = -\frac{3}{5}\frac{{\rm G} M_{\rm c}^2}{R_{\rm c}}. \tag{1.14}$$

Measurements of molecular cores reveal that the ratio of their rotational to

gravitational potential energies, commonly characterised as β, have typical values $\beta \sim 0.02$ (Goodman et al., 1993). Hence, rearranging Equations 1.12, 1.13 and 1.14, and substituting $\beta = E_{\rm rot}/E_{\rm grav}$ gives,

$$J_{\rm c} = \sqrt{3\beta {\rm G} M_{\rm c}^3 R_{\rm c}}. \tag{1.15}$$

Using typical values for a collapsing region of mass $1\,{\rm M}_\odot$ and size $0.1\,{\rm pc}$, we find that a collapsing cloud will have angular momentum, $J_{\rm c} \approx 10^{54}\,{\rm g\,cm^2\,s^{-1}}$ - 4 orders of magnitude higher than the total angular momentum of the solar system, which has $J \approx 10^{50}\,{\rm g\,cm^2\,s^{-1}}$.

In what is commonly referred to as *the angular momentum problem*, the question then remains - how is this excess angular momentum lost? Assuming that a cloud collapses at constant angular momentum, the spin required to form an isolated stellar core would be sufficient to tear it apart. Instead, an extended circular disc of material will form around the protostar acting as a vessel for angular momentum transport outwards and the transport of material inwards for further stellar accretion. We can equate the angular momentum of the collapsing cloud to that of the eventual disc which forms. Assuming that a disc forms with radius, $R_{\rm disc}$, and the gas within the disc orbits at Keplerian velocities with orbital frequency,

$$\Omega_{\rm K} = \sqrt{\frac{G M_{\rm c}}{{R_{\rm disc}}^3}}, \tag{1.16}$$

then,

$$J_{\rm disc} = M_{\rm c} R_{\rm disc}^2 \Omega_{\rm K} = \sqrt{{\rm G} M_{\rm c}^3 R_{\rm disc}}. \tag{1.17}$$

Equating this to the angular momentum obtained above, we find the when $M_{\rm c} = 1\,{\rm M}_\odot$, $R_{\rm disc} \approx 100\,{\rm AU}$. The formation of large rotating gaseous discs around newly formed stars is therefore a natural outcome of cloud collapse and angular momentum conservation.

Of course, the solution to the angular momentum problem is a combination of other processes, as cloud collapse will occur in a dynamic environment of magnetic fields, turbulence, interactions with other forming stars and tidal forces, hence

the values derived above are approximate. Observations of discs through their mm continuum emission suggest they may have sizes anywhere between $\approx 10^1 - 10^3$ AU (e.g. Andrews et al., 2018a,b; Eisner et al., 2018; Hendler et al., 2020), although these are typically for evolved systems, as young, newly formed discs will be embedded within their optically thick natal environments, making them challenging to observe. Equally, or potentially more challenging to measure is the typical masses of these newly formed systems. Again we are only able to observe evolved systems whose masses will have been significantly depleted since formation, and in addition we rely on uncertain conversion factors between a disc's mm dust emission and its gas mass (e.g. Dutrey et al., 2014; Andrews, 2020, and references therein). As will be discussed in detail in the next section, the masses of these newly formed discs will be of crucial importance to their very early evolution, and their potential susceptibility to gravitational instabilities.

1.4 Young self-gravitating discs

One-dimensional models (Lin & Pringle, 1990; Rice et al., 2010) and numerical simulations (Machida et al., 2010) suggest that when a star-disc system forms, the disc will have a mass comparable to that of the newly formed star. When this is the case, the disc's self-gravity will be of crucial importance to its early evolution. Such systems are said to be prone to *the gravitational instability* (GI), which will act to drive spiral density waves through the disc (as can be seen in Figure 1.5), acting as a key source of viscosity for driving the disc's early angular momentum transport and potentially leading to the direct formation of gas giant planets and brown dwarf stars through disc fragmentation. In this section, I will derive some of the key quantities associated with GI, and describe how the disc's self-gravity may influence its evolution.

1.4.1 The Toomre parameter

We can understand the stability of a self-gravitating, rotating disc by beginning with the WKB dispersion relation for a razor-thin disc, originally derived in Lin & Shu (1964) in the context of galactic dynamics. Considering for now axisymmetric perturbations, and assuming that the radial wavelength of a perturbation is much less than the radial extent of the disc, the dispersion relation is given by,

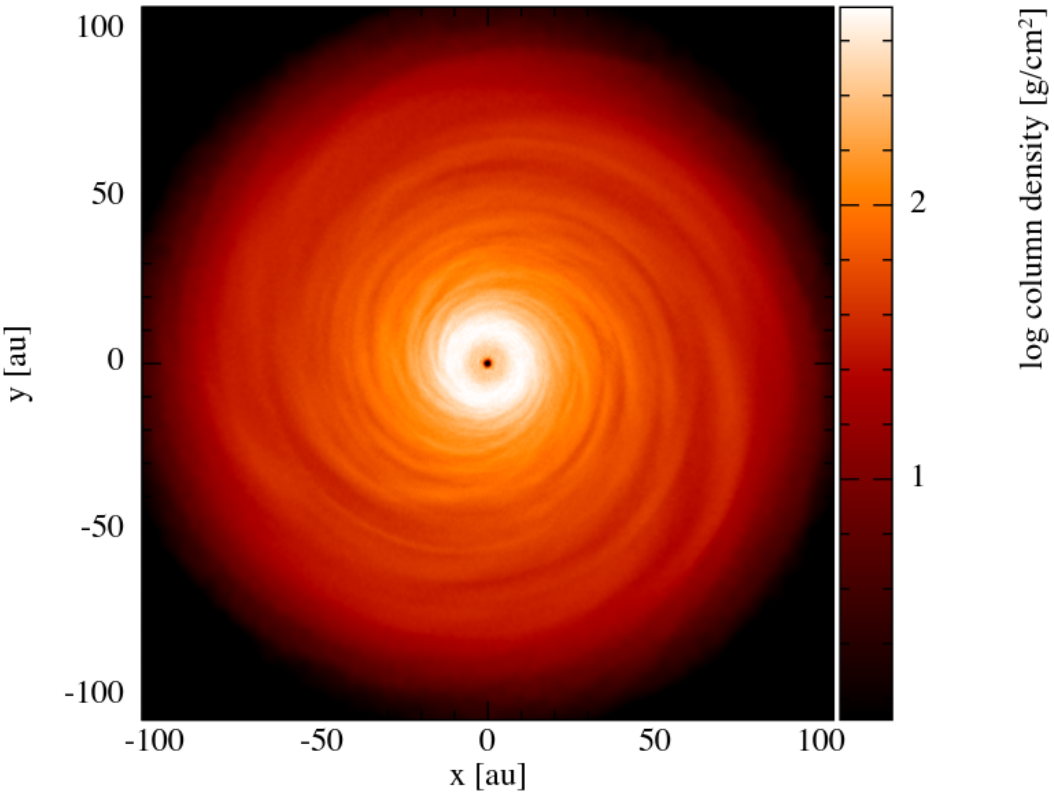

Figure 1.5 *Snapshot of a self-gravitating protoplanetary disc obtained from a Smoothed Particle Hydrodynamics simulation. The disc was set up with a mass of 0.2 $M_\odot$, an outer radius of 100 AU, and a central star mass of 1 $M_\odot$. Spiral substructure can be seen propagating throughout the disc as consequence of the gravitational instability.*

$$\omega^2 = c_s^2 k^2 - 2\pi G\Sigma|k| + \kappa^2, \tag{1.18}$$

which will have a solution of the form $e^{i(kx-\omega t)}$, where ω is the wave frequency, k is the radial wavenumber, Σ is the disc's surface density, c_s is the sound speed and κ is the epicyclic frequency, which will be equal to the Keplerian frequency in a rotationally supported disc. On the right hand side (RHS) of the equation, the first and third terms, which are positive, represent pressure and rotation respectively, whilst the central, negative term represents the disc's self-gravity.

When the RHS is positive we get $\omega^2 > 0$, hence ω will be real and the disc will be stable. However when the RHS is negative we get $\omega^2 < 0$, hence ω will be imaginary giving a solution to the wave equation, $e^{\pm|\omega|t}$, meaning any instability will grow exponentially. It then follows that the pressure and rotation terms in Equation 1.18 act to stabilize the disc as they drive the RHS up, whilst self-gravity will push the disc towards instability, and Equation 1.18 embodies the balance between these forces.

Defining the limit for stability as being when $\omega = 0$, we are left with a quadratic equation in terms of k,

$$c_s^2 k^2 - 2\pi G\Sigma|k| + \kappa^2 = 0. \tag{1.19}$$

If we consider Equation 1.18 to be a parabola in ω^2 and k, then the most unstable wavelength, $k_{\min}$, is the value of k at the parabola minimum which can be obtained by setting $\mathrm{d}\omega^2/\mathrm{d}k = 0$, giving,

$$k_{\min} = \frac{\pi G\Sigma}{c_s^2}. \tag{1.20}$$

It then follows that the disc is stable when $\omega^2(k_{\min}) > 0$ and unstable when $\omega^2(k_{\min}) < 0$. Substituting $k_{\min}$ into Equation 1.18 gives the criterion for stability as,

$$Q = \frac{c_s \kappa}{\pi G\Sigma} > 1. \tag{1.21}$$

Hence we have arrived at the *Toomre stability criterion*, where Q is commonly

known as the *Toomre parameter* (Toomre, 1964). The limit for stability derived in Equation 1.21 represents that for axisymmetric perturbations, such as rings. For non-axisymmetric perturbations, such as the development of spiral density waves which are characteristic of GI, it has been shown through numerical simulations that the limit for stability is $Q < 1.5 - 1.7$ (Durisen et al., 2007). Equation 1.21 again demonstrates the balance between rotational support, thermal support, and self-gravity. Assuming that the orbital frequency remains fixed at the Keplerian frequency, we see that hotter discs (higher c_s) will increase Q acting to stabilize the system, whilst more massive discs (higher Σ) will drive Q down toward the threshold for instability.

We can take this derivation further to obtain a more intuitive expression for instability in terms of the disc-to-star mass ratio. If we substitute,

$$M_{\rm disc} = \int_0^R 2\pi r \Sigma(r) dr, \tag{1.22}$$

and assume that the surface density follows some radially negative power-law distribution,

$$\Sigma(r) = \Sigma_0 \left(\frac{r}{R_0} \right)^{-q}, \tag{1.23}$$

then we get that,

$$M_{\rm disc} = \frac{2\pi\Sigma_0}{(2-q)R_0^{-q}} R^{2-q}. \tag{1.24}$$

Substituting Equations 1.23 and 1.24 back into Equation 1.21, noting that $c_s = H\Omega$ (will be derived later in Chapter 2), and κ is the Keplerian orbital frequency where $\kappa = \Omega = \sqrt{GM_*/R^3}$, we get the limit for stability as,

$$Q = f \frac{M_*}{M_{\rm disc}} \frac{H}{R}, \tag{1.25}$$

where H is the disc's vertical *scale height*, and $f(= 2/(2-q))$ is a numerical factor of order unity which depends on the specific surface density profile. From Equation 1.25 we see that, for a fixed disc-to-star mass ratio, thinner discs will

be more unstable. Observations of discs find that they are generally thin, with $H/R \approx 0.05 - 0.2$ (Andrews et al., 2010), hence we expect discs with $\gtrsim 5 - 20\%$ the mass of their parent star to be gravitationally unstable. As mentioned at the beginning of the section, models of star and disc formation predict initial disc-to-star mass ratios which easily exceed this (Lin & Pringle, 1990; Rice et al., 2010; Machida et al., 2010), hence we may reasonably expect discs to be gravitationally unstable when they are young.

1.4.2 Marginal stability and self-regulation

Through the Toomre parameter and the demonstrated balance between the disc mass and temperature, we can understand how a disc close to the limit for instability will evolve. A massive disc at a high temperature will be stable, and will radiatively cool. As the disc temperature decreases so does its sound speed, hence Q will decrease. If the disc is massive enough and able to cool efficiently enough, Q may approach unity, at which point the disc will become susceptible to the growth of spiral density waves. The presence of spirals will cause disc heating as they drive shock waves, hence the disc temperature will increase, Q will increase, and the disc will move back toward a state of stability. The spirals then become weakened as the instability is quenched, and the disc will begin to cool again, driving Q back down toward instability. Paczynski (1978) originally showed that GI discs can reach this state of self-regulation, where heating is balanced by cooling and the disc remains close to the limit for instability. Such a disc is said to be in a *marginally stable* state, where the spirals act to thermally regulate the system. In a marginally stable disc, spirals will propagate with a range of length scales and timescales which acts to generate turbulence in the disc. This *gravito-turbulence* will provide the dominant source of viscosity at early times when the disc is massive enough to be gravitationally unstable, and too cold to be significantly magnetized such that magnetohydrodynamic (MHD) turbulence is not important. The importance of viscosity to angular momentum transport, and the effective viscosity associated with the gravitational instability will be formalised in more detail in Chapter 2.

1.5 Observational properties of protostars and protostellar discs

Formation of stars and discs and the subsequent evolution of star-disc systems can be understood through observations of star forming regions. When these objects are young they will be heavily embedded within envelopes of optically thick gas, hence emission will mostly be at far-IR and mm wavelengths. To gain a good understanding of the different phases of star and disc formation requires observations of many systems, such that we can derive a strong statistical sample. However, spatially resolving discs at these wavelengths and at distances which can be in the 100s of parsecs is challenging, and has only recently become possible with the advent of long baseline interferometeres such as the Atacama Large Millimeter/submillimeter Array (ALMA) and the Square Kilometer Array (SKA).

Prior to this being possible it has been convention to classify Young Stellar Objects (YSOs) through their unresolved spectral energy distribution (SED) (Lada, 1987; Andre et al., 1993; Andre & Montmerle, 1994). More specifically, YSOs are classified by the slope of their SED in the near- to mid-IR, between wavelengths $\sim 2 - 20\,\mu$m, where the slope is calculated as,

$$\alpha_{\rm IR} = \frac{\Delta \log(\lambda F_\lambda)}{\Delta \lambda}. \tag{1.26}$$

This classification scheme is generally made up of 4 key stages, which are outlined below and illustrated in Figure 1.6

- **Class 0:** A compact, dense object has formed within the prestellar core. The protostar is heavily embedded in an optically thick circumstellar envelope, where the envelope mass is comparable to that of the protostar. This phase is likely prior to any disc formation. The SED peaks in the far-IR/mm, with very little/no emission in the near-IR at wavelengths $\lambda \lesssim 10\,\mu$m. Typical ages of Class 0 protostars are $\sim 10^4 - 10^5$ yrs.
- **Class I:** The protostar is less embedded within the now less massive envelope, and a protostellar disc is now present. The SED consists of protostellar blackbody emission, with a large amount of excess emission in the far-IR, representing reprocessed stellar radiation from the envelope

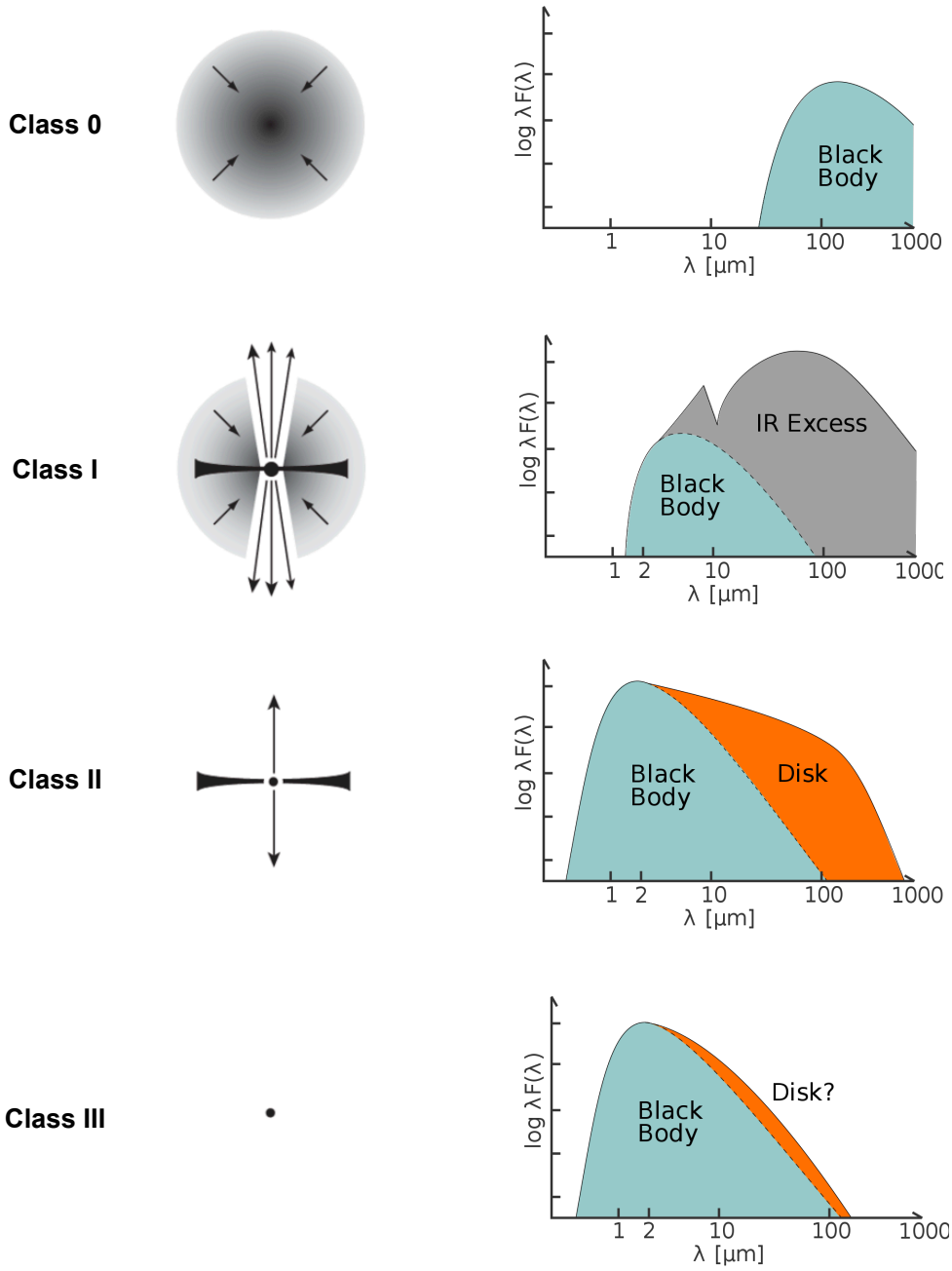

Figure 1.6 *An illustration of the classification scheme for Young Stellar Objects. Image credit: Adapted from an image originally found in Armitage (2010) and another image by Magnus Vilhelm Persson.*

and disc. The SED is now flat, or slightly rising, between near and mid-IR wavelengths, with $\alpha_{\rm IR} > 0$. The protostar is highly accreting, hence a large amount of its luminosity comes from accretion, and outflows and jets are often detected. These objects typically represent ages of a few 10^5 yrs.

- **Class II:** The circumstellar envelope is now mostly depleted, and a protostar surrounded by a protostellar disc remains. The SED now consists of stellar blackbody combined with IR excess due to reprocessing in the disc, with $-1.5 < \alpha_{\rm IR} < 0$. These objects are often referred to as Classical T Tauri stars, and have ages $\sim 3 \times 10^6$ yrs. The protostar is still strongly accreting by means of the protostellar disc.

- **Class III:** The final phase of star formation, before the protostar evolves onto the main sequence. The disc is now mostly depleted, hence the protostar is no longer strongly accreting, and the SED is mostly that of stellar blackbody. These are often referred to as Weak-lined T Tauri stars, with ages $\sim 10^7$ yrs. The spectral index is now $\alpha_{\rm IR} < -1.5$.

These classes broadly represent the key stages of star and disc formation. To date, observations of discs typically represent older, Class I and Class II sources, when the disc is no longer embedded within the envelope. Famous examples include the images of the Class I disc HL-Tau (ALMA Partnership et al., 2015a) and the Class II disc Elias 2-27 (Pérez et al., 2016; Huang et al., 2018b) from which substructure in clearly observable, possibly indicative of the presence of planets, or the gravitational instability in the case of Elias 2-27 (see Figure 1.7). However younger, possibly Class 0, sources have also been imaged, such as L1448 IRS 3B (Tobin et al., 2016) which shows a triple protostar system where the disc is possibly in the process of *fragmentation* (also see Figure 1.7). Disc self-gravity is likely only important during the Class 0 and Class I phases whilst the disc is still massive and being replenished through accretion from the envelope.

Crucially, and of central importance to this thesis, throughout these stages of star and disc formation the formation of planets is also slowly taking place. I will now outline our current understanding of planet formation, and how it fits into the picture of star formation laid out thus far.

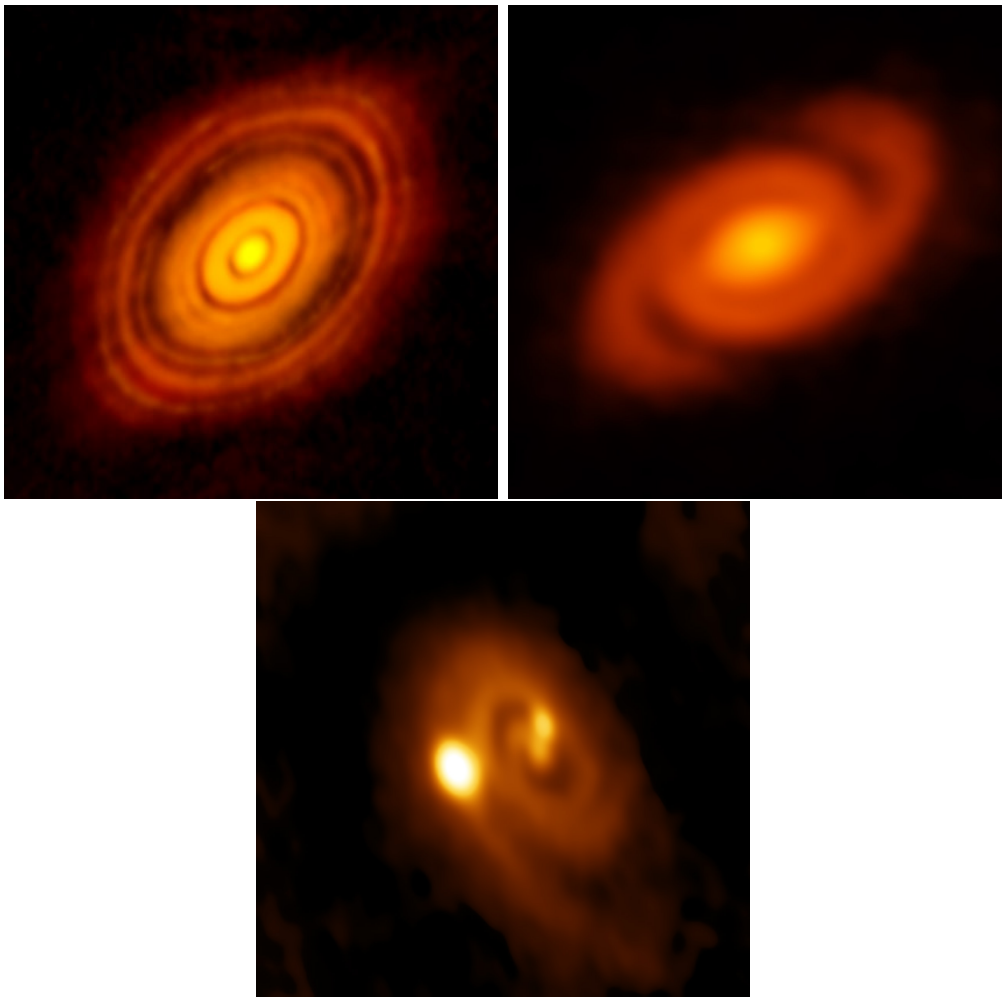

Figure 1.7 *Images of protoplanetary discs obtained with ALMA. Top left: HL Tau. Top right: Elias 2-27. Bottom: L1448 IRS 3B. Image credit for all 3 images: B. Saxton (NRAO/AUI/NSF); ALMA (ESO/NAOJ/NRAO)*

1.6 Planet formation

Early theories of planet formation (e.g. Safronov, 1969) were developed long before the first exoplanets were discovered in 1992, when two terrestrial mass planets were found in orbit around a pulsar (Wolszczan & Frail, 1992). Hence, the theories being developed were primarily concerned with explaining the architecture of our own Solar System, where we have rocky, terrestrial planets in the inner Solar System and gas/ice giant planets in the outer Solar System. As we began to discover more systems through large exoplanet surveys (e.g. Borucki et al., 2010; Howell et al., 2014; Ricker et al., 2015) we found that, perhaps unsurprisingly, there exists a rich diversity of exoplanets and exoplanetary architectures different to what we have in the Solar System. For example, radial velocity surveys such as the *Kepler* mission (Borucki et al., 2010) have so far revealed hundreds of planets classified as *hot-Jupiters*; planets of masses comparable to that of Jupiter, on very short orbits with periods less than 10 days. We are constantly discovering systems which challenge our understanding of planet formation, such as the Trappist-1 system (Gillon et al., 2017) where 7 terrestrial planets were found orbiting a $0.08\,\mathrm{M}_{\odot}$ star all within 0.062 AU, which would require an extremely high planet formation efficiency if the planets were to have formed at their current locations.

As it stands, there are two complementary theories of planet formation, namely the core accretion theory (Pollack et al., 1996; Mordasini et al., 2010) and the gravitational instability theory (Kuiper, 1951; Boss, 1997) of planet formation (see also Papaloizou & Terquem, 2005; Brandner, 2006; Armitage, 2020, for comprehensive overviews of both). It is widely accepted that the terrestrial planets, and probably most of the gas giant planets, formed through core accretion; where smaller bodies steadily grow through collisions and accumulation of other smaller bodies to form progressively larger bodies. However core accretion theory struggles to explain the existence of the most massive planets which are often found on wide orbits from their host stars. The gravitational instability model of planet formation can potentially help with this, as it predicts that giant gaseous planets may form directly through the gravitational collapse of dense regions of the disc. The research undertaken throughout this thesis will focus primarily on the gravitational instability model, however a discussion of planet formation would not be complete without first discussing core accretion theory.

1.6.1 Core Accretion theory of planet formation

Core accretion theory consists of 3 general phases; the formation of planetesimals of $\sim$km scales from dust grains of $\sim\mu$m scales, the formation of terrestrial mass protoplanets from planetesimals, and the eventual possible formation of gas giant planets.

In the first stage it is thought that micron sized dust grains are able to grow through random collisional accumulation of other dust grains, which will steadily coagulate to form planetesimals (for an overview of dust evolution in discs see Testi et al., 2014). The initial phase of growth up to particles of meter sizes is thought to occur fairly rapidly, on timescales 10^3-10^4 yrs for particles at distances of a few AU from their parent star (Weidenschilling, 1980; Nakagawa et al., 1981). However growth beyond meter scales to form planetesimals presents a number of challenges, and is not currently well understood, in what is commonly referred to as the *meter-barrier*.

A protoplanetary disc will generally have a gas pressure gradient which decreases between regions of high gas density in the inner disc and low gas density in the outer disc. Thus the gas component of the disc is subject to an additional outward pressure force, such that gas at a given radius is able to orbit with slightly sub-Keplerian velocities whilst remaining in a stable orbit (with $v \approx 0.996\, v_{\rm kep}$). Smaller, micron-sized dust grains will be strongly coupled to the gas in the disc and will also orbit with sub-Keplerian velocities. However, as the dust grains grow toward meter-sizes they will become less strongly coupled and try to orbit at Keplerian velocities which generates a resultant gas-dust drag. This drag results in deceleration of the dust and inward radial migration. It becomes most effective when dust grains reach meter-sizes, and may generate large radial velocities resulting in either accretion onto the star or destruction of the solids as they rapidly migrate through the disc (Weidenschilling, 1977). This is commonly referred to as the *radial drift barrier*.

Equally, as smaller particles will have small relative velocities, when they collide they will likely coalesce in a hit-and-stick process. However as they grow toward meter sizes and their *Stokes number*, which can be used to quantify the gas-dust coupling, approaches unity, particle's relative velocities will increase. Hence as particles grow, their impact velocities will increase accordingly and collisions will result in particles either bouncing off each other, compacting their densities in the process (Güttler et al., 2010; Zsom et al., 2010), or shattering into several smaller

fragments (Brauer et al., 2008; Birnstiel et al., 2010). These two growth barriers are known as the *bouncing barrier* and the *fragmentation barrier* respectively.

How particles overcome the meter-barrier is still not well understood, and an active area of research. Solutions often invoke various instabilites which can generate regions of high particle densities and accelerated planetesimal growth, such as the streaming instability (Youdin & Goodman, 2005; Youdin & Johansen, 2007), or dust-trapping in the spirals of self-gravitating discs (Rice et al., 2004). It has also been proposed that these instabilities may be capable of generating local regions of extremely high particle densities, which may then undergo gravitational collapse to form the first 100 – 1000 km planetesimals directly (Rice et al., 2006; Johansen et al., 2007).

Sidestepping this issue for now, and assuming that growing dust grains are able to overcome the meter barrier to form planetesimals, we now enter the next stage of core accretion as the planetesimals go on to form terrestrial mass protoplanets. Once planetesimals of km sizes have formed, their subsequent evolution is largely dictated by their gravitational influence on the dust in the disc and other nearby planetesimals. Mutual collisions between planetesimals may result in them sticking together or shattering into several fragments, with either outcome dependent on the energy of the collision. For massive collisions, shattering may result in re-accretion of the fragments if they remain gravitationally bound. As they grow, the most massive planetesimals will enter a phase of *runaway growth*, where their gravitational cross sections exceed their geometric cross sections, and their escape velocities exceed their relative velocities, hence gravitational focussing of other nearby planetesimals results in a phase of rapid growth. During this phase it is the most massive planetesimals which will grow fastest, and usually one dominant mass will form locally, hence it is often referred to as *oligarchic growth*. Many oligarchs will form across the disc, each feeding on their own local material.

This phase will generally continue until the protoplanet core, or oligarch, has depleted the solid material from its local feeding zone, or it becomes massive enough to accrete a massive gaseous envelope. If the protoplanetary core has depleted all the material from its local region of the disc, growth may still continue in a chaotic manner as oligarchs interact with each other, collide, and scatter other objects through large radial distances.

If a protoplanetary core is able to grow to a mass $\sim 10\,\mathrm{M}_{\oplus}$, then it may enter the

phase of giant planet formation as the core's high mass will trigger the onset of rapid gas accretion (Mizuno, 1980; Bodenheimer & Pollack, 1986; Pollack et al., 1996). The core will accrete all the gas within its local feeding zone, referred to as its *Hill sphere*, which is defined as the local region of the disc where the core's gravitational potential exceeds the global potential generated by the star. Gas accretion will halt when either the protoplanet has depleted all of its local supply of gas, at which point it will have opened a gap in the disc, or when the global disc has dissipated through photoevaporation.

Models of core accretion are generally able to successfully explain the formation of Jupiter-mass planets at Jupiter-like distances of a few AU from their parent star (e.g. Rice & Armitage, 2003; Lissauer et al., 2009). However the formation of giant planets on wider orbits than this at a few 10s of AU, which are often found through direct imaging surveys, present challenges to core accretion theory. At such distances the disc surface density will be low, thus the mass reservoir available to the forming protoplanet will be low, and formation timescales will be long. The observed lifetimes of protoplanetary discs are typically found to be a few Myrs (e.g. Haisch, Lada & Lada, 2001), whilst core accretion formation timescales of giant planets beyond 10 AU are generally predicted to far exceed these timescales (e.g. Pollack et al., 1996). It is possible that there exists mechanisms which may allow for such planets to form through core accretion, whether they act to speed up formation timescales (e.g. pebble accretion, Johansen & Lambrechts (2017), streaming instabilities Youdin & Goodman (2005)), or allow for planets to form elsewhere in the disc and migrate outwards later. However, for systems as challenging to explain as HR 8799, where four wide-orbit, massive planets (with semi-major axes in the range 15 AU and 70 AU and masses in the range 5.7 $M_{Jupiter}$ and 9.1 $M_{Jupiter}$) have been directly imaged (Marois et al., 2008, 2010), it is tempting to look toward alternative planet formation theories, such as the gravitational instability model.

1.6.2 Gravitational Instability theory of planet formation

In an attempt to explain the origin of the Solar System's gas giant planets, Kuiper (1951) first theorised that planet formation may proceed in a manner similar to how stars form in GMCs, where a dense region of the disc may *fragment* to form gravitationally bound gaseous clumps, which may then go on to become gas giant planets. With the advent of the first exoplanet discoveries, including observations

of a number of giant planets, Boss (1997) later revived this idea, evaluating the possibility that GI may be a plausible mechanism for the formation of these obejcts. An illustration of disc undergoing fragmentation is shown in Figure 1.8.

This work sparked interest and research into the conditions necessary for a disc to undergo fragmentation. In Section 1.4 I discussed the conditions required for the onset of the gravitational instability. If a disc is sufficiently massive, such that Q (Equation 1.21) is of order unity, a phase of linear instability may set in which manifests in the formation of spiral density waves. A disc may enter a phase of marginal stability if it is unable to rapidly radiate away the heat generated by the spirals, such that they have the effect of thermally regulating the disc and maintaining Q close to unity. However, if the disc is able to cool efficiently then another outcome may occur, whereby a phase of exponential growth of the perturbations sets in and the spirals may break up to form gravitationally bound fragments.

This leads us to our second requirement for a disc to undergo fragmentation: the critical cooling rate. Consider a gravitationally unstable disc where the thermal timescale can be approximated by the disc dynamical timescale, $t_{\rm dyn} \sim 1/\Omega$ (e.g. Pringle, 1981). This is the timescale on which shock waves driven by the spirals will act to heat the disc, and the disc will radiatively cool. If the cooling rate is balanced by the heating rate then the disc will enter a phase of marginal stability. However if the disc is able to cool at a rate faster than the heating rate, such that $t_{\rm cool} \lesssim t_{\rm dyn}$, then the instability will grow and the disc may fragment. Early work considered this condition by introducing a dimensionless cooling parameter, β, such that the critical cooling rate can be described by some critical value of β (Gammie, 2001; Rice et al., 2003),

$$t_{\rm cool}\Omega = \beta < \beta_{\rm crit}. \tag{1.27}$$

This was later taken further, where the local equation of state was also taken into account, and the critical cooling rate was described in terms of a critical rate of angular momentum transport (Rice, Lodato & Armitage, 2005),

$$\alpha = \frac{4}{9}\frac{1}{\gamma(\gamma-1)\beta}. \tag{1.28}$$

Here, α is the Shakura-Sunyaev viscous-α (Shakura & Sunyaev, 1973) which is

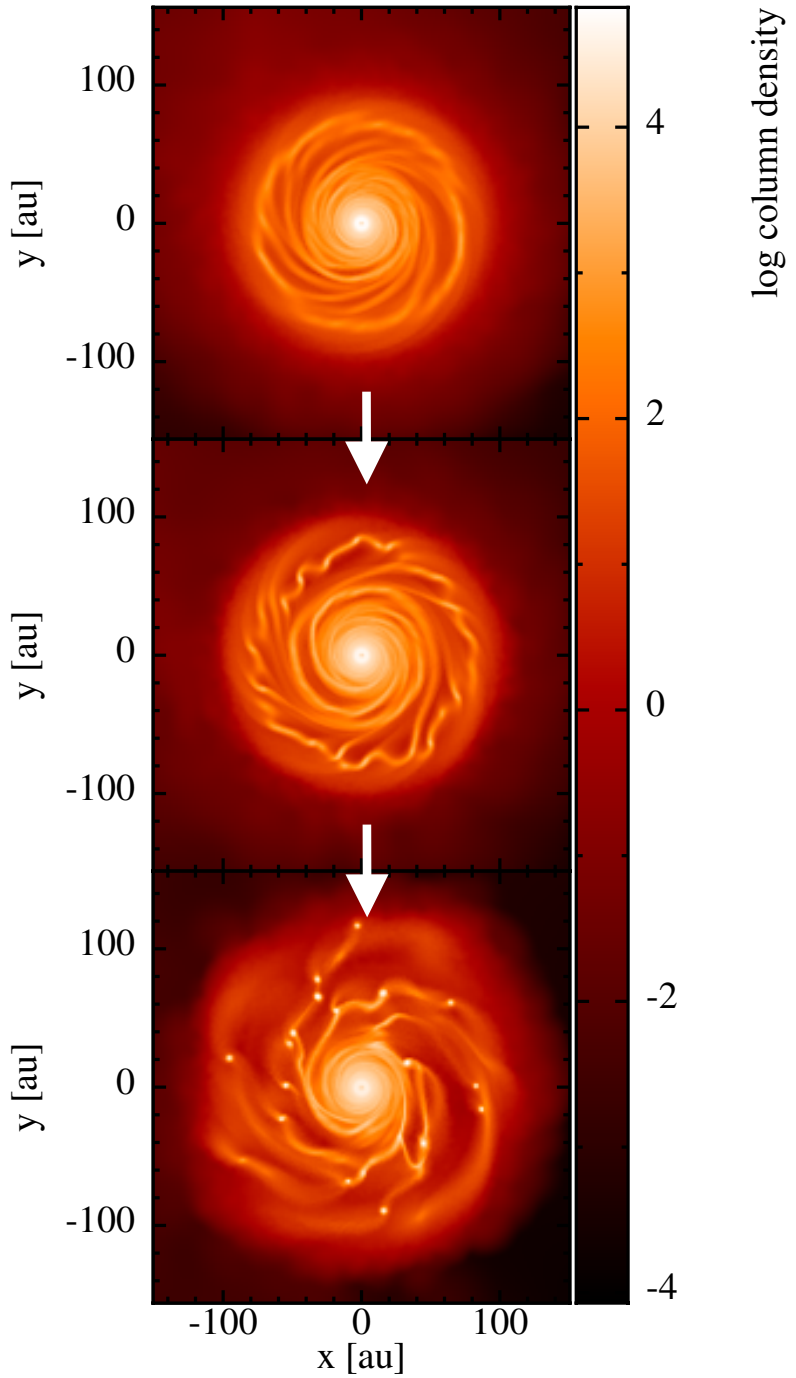

Figure 1.8 *Evolution of a Smoothed Particle Hydrodynamics simulation showing a self-gravitating disc undergoing fragmentation. The disc was set up with a mass of 0.3 $M_\odot$, an outer radius of 100 AU, and a central star mass of 1 $M_\odot$. Strong spiral features initially form (top panel) which later become unstable and start to form high density clumps (middle panel). In the final state of the disc, many fragments have formed which may go on to become giant planets or brown dwarf stars (bottom panel).*

used here to parameterise the effective viscosity resulting from GI (which will be introduced in more detail in Chapter 2) and γ is the ratio of specific heats. Typically it is thought that a disc may be thermally regulated if $\beta \approx 10$–20 (i.e. if the cooling time is 10–20$\times$ the dynamical timescale), and that fragmentation may occur when $\beta \lesssim 3$, or when $\alpha \sim 0.06$ (Gammie, 2001; Rice et al., 2003, 2005).

Hence we have arrived at the two main conditions necessary for fragmentation to occur. If a disc is massive enough such that $Q \approx 1$ and able to cool fast enough, gravitationally bound gaseous clumps will begin to form and condense at a rate faster than the disc is able to disperse them through heating or dynamical shear.

Having derived the conditions necessary for fragmentation we can now also ask: what type of objects might form through fragmentation? We can use the critical cooling rate criterion to understand where in the disc fragmentation is likely to occur. In the dense, inner regions of disc the gas will typically be optically thick thus rapid cooling isn't possible. Additionally, stellar irradiation will generate high temperatures, hence any instability will be quenched (Rafikov, 2005). We would therefore only expect fragmentation to be possible in the outer disc ($R \gtrsim 50 - 100\,\mathrm{AU}$), where the gas is cooler and surface densities are lower such that rapid radiative cooling may occur (Clarke, 2009; Rice & Armitage, 2009; Rice et al., 2010).

To understand the typical fragment masses which may form, we begin by noting that the most unstable wavelength of GI (when $Q = 1$) is of the order of the disc scale height, when $\lambda = 2\pi H$. We can then take the Jeans mass of a fragment as being the mass enclosed within this wavelength, such that,

$$M_{\mathrm{frag}} = \Sigma\lambda^2 = \Sigma(2\pi H)^2. \tag{1.29}$$

If we assume that fragmentation may occur when $Q = 1$, then we can rearrange Equation 1.21 for Σ. Also substituting that $H = c_{\mathrm{s}}/\Omega$ (will be derived later in Chapter 2), we get that,

$$M_{\mathrm{frag}} = 4\pi M_* \left(\frac{H}{r}\right)^3. \tag{1.30}$$

For a disc around a solar mass star, with $H/r = 0.1$ we therefore expect typical

fragments to have masses $M_{\rm frag} = 13.2\,{\rm M_{Jup}}$.

To conclude this chapter, we have derived that planet formation through the gravitational instability will likely only form the most massive planets on wide orbits from their parent star. It has been argued that subsequent growth of fragments may lead to the majority of GI-born objects going on to exceed the deuterium burning limit, thus planets formed through GI will be rare as most fragments will become brown dwarfs (Kratter et al., 2010). It has also been argued that fragments will enter a phase of *tidal downsizing* after formation, where mass stripping, thermal contraction, sedimentation of solids at the fragment's core and migration can lead to the formation of all planetary types, including terrestrial planets (Nayakshin, 2010a). Throughout this thesis my research will primarily focus on the conditions necessary for discs to undergo fragmentation, with less focus on the subsequent evolution of fragments.

Chapter 2

Fundamental physics and useful numerical methods

2.1 Fundamental physics

2.1.1 Hydrodynamics

I will begin this section by deriving the fluid equations, which are fundamental to the work undertaken here. Protoplanetary discs consist primarily of large reservoirs of gas which can be treated as a *compressible, Newtonian*[1] fluid, hence the equations derived below will be essential in both understanding and in modelling the evolution of these systems. The derivations outlined below are all standard results, but are primarily based on the derivations presented in Chorin & Marsden (1993).

These derivations are based on three basic principles, namely the conservation of mass, the conservation of momentum and the conservation of energy.

2.1.1.1 The continuity equation

We can derive the equation for the conservation of mass by first considering a volume element, V, which contains some fluid of density, ρ, such that the mass contained inside the volume is given by $\int_V \rho \, \mathrm{d}V$. If we define S as the surface of

[1] where the viscous stresses are linearly proportional to the strain

the volume element and let $\mathrm{d}A$ denote the surface area of the volume element we can state that the net flow of mass, m, in the volume element is,

$$m = \oint_S \rho \mathbf{v} \cdot \mathrm{d}\mathbf{A}\,, \tag{2.1}$$

where $\mathbf{v}$ represents the fluid's velocity. Defining the rate of change of mass in the volume element as,

$$m = -\frac{\partial}{\partial t} \int_V \rho \,\mathrm{d}V\,, \tag{2.2}$$

conservation of mass then dictates that these two equations must be equal,

$$\oint_S \rho \mathbf{v} \cdot \mathrm{d}\mathbf{A} = -\frac{\partial}{\partial t} \int_V \rho \,\mathrm{d}V\,. \tag{2.3}$$

Using the divergence theorem we can substitute the surface integral for a volume integral as,

$$\int_V \nabla \cdot \rho \mathbf{v} \,\mathrm{d}V = \oint_S \rho \mathbf{v} \,\mathrm{d}A\,. \tag{2.4}$$

Equation 2.3 then becomes,

$$\int_V \nabla \cdot \rho \mathbf{v} \,\mathrm{d}V = -\frac{\partial}{\partial t} \int_V \rho \,\mathrm{d}V\,, \tag{2.5}$$

hence,

$$\int_V \left(\nabla \cdot \rho \mathbf{v} + \frac{\partial \rho}{\partial t} \right) \mathrm{d}V = 0. \tag{2.6}$$

Recognising that this must hold true for all V, this gives,

$$\boxed{\nabla \cdot \rho \mathbf{v} + \frac{\partial \rho}{\partial t} = 0.} \tag{2.7}$$

Hence we have arrived at the *continuity equation* which embodies the conservation

of mass, stating that the rate of change of mass within the volume element must be equal to the mass flowing in or out of the volume element.

2.1.1.2 The momentum equation

We can use the same volume element to construct a similar equation which represents the conservation of momentum within a fluid. The force, $\mathbf{F}$, exerted on the surface of the volume element is given by,

$$\mathbf{F} = -\oint_S P \, \mathrm{d}A = -\int_V \nabla P \, \mathrm{d}V \,, \tag{2.8}$$

where we have again used the divergence theorem to convert the surface integral to a volume integral. For now, we will assume that we are dealing with an *ideal fluid*[2] where the pressure force acts perpendicular to the surface of the volume element only. We can then appeal to Newton's second law to construct an equation for the force balance, where the force per unit area on a mass element is given by,

$$\mathbf{F} = \rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}, \tag{2.9}$$

hence,

$$\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\nabla P. \tag{2.10}$$

The above derivative is *Lagrangian*, meaning that $\mathbf{v}$ represents the path followed by an individual particle within the fluid. The particle's velocity is therefore a function of position and time, such that $\mathbf{v} = \mathbf{v}(x(t), y(t), z(t), t)$, and,

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\partial \mathbf{v}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathbf{v}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathbf{v}}{\partial z}\frac{\partial z}{\partial t} + \frac{\partial \mathbf{v}}{\partial t}, \tag{2.11}$$

which can be rewritten as,

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}. \tag{2.12}$$

[2] an incompressible, inviscid, non-turbulent fluid. A simplification for the purposes of our derivations here. Ideal fluids do not exist in nature.

This is known as the *material derivative*, which in its most general form can be expressed as,

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla. \tag{2.13}$$

Substituting Equation 2.12 back into Equation 2.10 and rearranging, we arrive at the *momentum equation*,

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho} \nabla P. \tag{2.14}$$

In the presence of some external force, $\mathbf{F}$, (such as gravity) this becomes,

$$\boxed{\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho}(\nabla P + \mathbf{F}).} \tag{2.15}$$

2.1.1.3 The Navier-Stokes equation

Until now, in our derivation of Equation 2.15 we have assumed that we are dealing with an ideal fluid where all the forces acting on the surface are being applied perpendicular to the surface plane. More generally, when considering a viscous fluid, the surface will be subject to additional stress forces which act both perpendicular and parallel to the surface plane.

A schematic diagram which illustrates these additional stress forces is shown in Figure 2.1, where the forces can be expressed in the form of a viscous stress tensor,

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \tag{2.16}$$

This can be broken down into two parts. The volumetric stress tensor which consists of forces which act to change the volume of the body (i.e. the pressure forces), and the stress deviator tensor which consists of forces which act to change the shape of the body (i.e. the viscous shear stresses). The stress tensor then becomes,

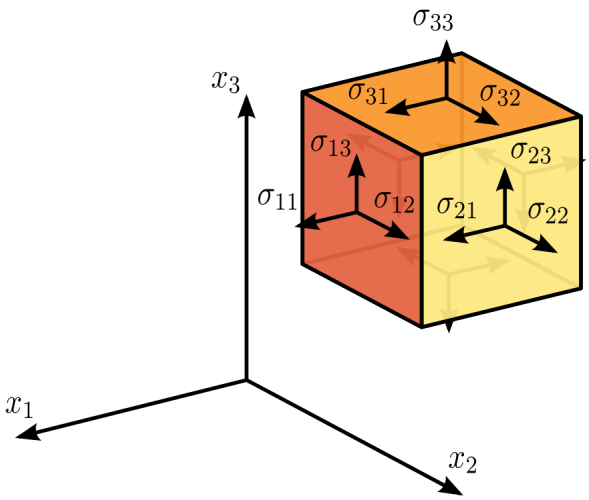

Figure 2.1 *Illustration of the viscous stress tensor from Equation 2.16 acting on a volume element. Image credit: Sanpaz*

$$\sigma = -\begin{bmatrix} P & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & P \end{bmatrix} + \begin{bmatrix} \sigma_{11} + P & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} + P & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} + P \end{bmatrix} \tag{2.17}$$

equivalent to,

$$\sigma = -P\mathbb{I} + \mathbb{T}. \tag{2.18}$$

Substituting this into our form of the Navier-Stokes equation gives,

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho}\nabla P + \frac{1}{\rho}\nabla \cdot \mathbb{T} + \mathbf{F}. \tag{2.19}$$

The specific form of $\mathbb{T}$ will depend on the type of fluid we are considering. For an *incompressible, Newtonian* fluid,

$$\nabla \cdot \mathbb{T} = \nu \nabla^2 \mathbf{v}, \tag{2.20}$$

where ν is the fluid viscosity. Hence, Equation 2.19 becomes,

$$\boxed{\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho}\nabla P + \frac{1}{\rho}\nu \nabla^2 \mathbf{v} + \mathbf{F},} \tag{2.21}$$

2.1.1.4 The energy equation

Finally, we can derive the equation for the conservation of energy by first stating that the total energy of the fluid within a volume element is the sum of its kinetic and internal energies,

$$E_{\text{total}} = E_{\text{kinetic}} + E_{\text{internal}}. \tag{2.22}$$

Here we will consider an *isentropic fluid*, whereby both the kinetic and internal energies are affected by work done on the system, which may be a result of pressure or other body forces such as gravity or magnetic fields. If we define u as the internal energy per unit mass and e as the total energy per unit mass, then Equation 2.22 can be rewritten as,

$$E = \rho e = \frac{1}{2}\rho |\mathbf{v}|^2 + \rho u. \tag{2.23}$$

The rate of change of energy is then,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{1}{2}\rho \frac{D}{Dt}|\mathbf{v}|^2 + \rho \frac{\mathrm{d}u}{\mathrm{d}t}, \tag{2.24}$$

where we have again used the material derivative from Equation 2.13 when differentiating $\mathbf{v}$ with respect to time. This gives,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \rho \left[\mathbf{v} \cdot \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right) \right] + \rho \frac{\mathrm{d}u}{\mathrm{d}t}. \tag{2.25}$$

We can substitute for the rate of change of internal energy using the First Law of Thermodynamics, which states,

$$\mathrm{d}U = T\,\mathrm{d}S - P\,\mathrm{d}V\,. \tag{2.26}$$

where U is internal energy, T is temperature and S is entropy. Converting these to quantities per unit mass, where u and s are the internal energy and entropy per unit mass, we get,

$$\mathrm{d}u = T\,\mathrm{d}s - P\,\mathrm{d}\left(\frac{1}{\rho}\right) = T\,\mathrm{d}s + \frac{P}{\rho^2}\,\mathrm{d}\rho\,, \tag{2.27}$$

where $1/\rho$ is equivalent to the volume per unit mass. Taking the derivative with respect to time then gives,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = T\frac{\mathrm{d}s}{\mathrm{d}t} + \frac{P}{\rho^2}\frac{\mathrm{d}\rho}{\mathrm{d}t}. \tag{2.28}$$

As stated earlier, we are considering an isentropic flow, which by definition will have constant entropy, hence $\frac{\mathrm{d}s}{\mathrm{d}t} = 0$, and we are left with,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{P}{\rho^2}\frac{\mathrm{d}\rho}{\mathrm{d}t}. \tag{2.29}$$

Substituting this back into Equation 2.25 we get,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \rho\left[\mathbf{v}\cdot\left(\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right)\right] + \frac{P}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}. \tag{2.30}$$

Using the continuity equation (Equation 2.7) this can be rewritten as the *energy equation* for an isentropic flow,

$$\boxed{\frac{\mathrm{d}E}{\mathrm{d}t} = \rho\left[\mathbf{v}\cdot\left(\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right)\right] - P(\nabla\cdot\mathbf{v}).} \tag{2.31}$$

2.1.2 Gravity

Of the external forces mentioned in Equation 2.15, the influence of gravity will dominate the early evolution of protoplanetary discs, with contributions primarily from the mass of the central star, but also from the mass of the disc material itself when the system is young.

Newton's law of gravitation gives that the gravitational force between two masses,

M_1 and M_2, is inversely proportional to square of their separation, R, as,

$$F_{\rm g} = \frac{{\rm G}M_1 M_2}{R^2}, \tag{2.32}$$

where the force is related to the gravitational potential, Φ, as,

$$F_{\rm g} = -\nabla\Phi. \tag{2.33}$$

In a system where the gravitational potential is dominated by the mass of the star,

$$\Phi = -\frac{{\rm G}M_*}{R}. \tag{2.34}$$

Hence, including this in Equation 2.21 the Navier-Stokes equation becomes

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} = -\frac{1}{\rho}\nabla P + \frac{1}{\rho}\nu\nabla^2\mathbf{v} - \nabla\Phi. \tag{2.35}$$

2.1.3 The equation of state

The sound speed is defined as the rate at which a medium can respond to disturbances, and can be expressed as,

$$c_s^2 = \frac{{\rm d}P}{{\rm d}\rho}. \tag{2.36}$$

In an adiabatic, ideal gas we can express the equation of state as,

$$P = K\rho^\gamma, \tag{2.37}$$

where K is some constant of proportionality and γ is the ratio of specific heats. We then obtain that the sound speed in a medium is,

$$c_s^2 = \frac{\gamma P}{\rho}. \tag{2.38}$$

2.1.4 Radiative transfer

Another process fundamental to the work undertaken in this thesis is that of radiative transfer. Radiative transfer describes the transport of energy through a medium, such as the radiation emitted from a star which then travels through the gas and dust in a protoplanetary disc. The star's radiation will contribute thermal energy to the disc, thus understanding how this energy is transported allows us to understand and model the disc's thermodynamics. The following equations are again standard results, but have been primarily derived following Rybicki & Lightman (1986).

Consider a beam of radiation travelling through a cylinder of material with surface area, $\mathrm{d}A$, in time, $\mathrm{d}t$, at an angle of incidence, θ, and solid angle, $\mathrm{d}\Omega$. The beam (or *photon packet*, consisting of photons with frequencies in the range $\nu \rightarrow \nu + \mathrm{d}\nu$) will carry energy,

$$\mathrm{d}E_\nu = I_\nu \cos(\theta)\, \mathrm{d}A\, \mathrm{d}t\, \mathrm{d}\nu\, \mathrm{d}\Omega\,. \tag{2.39}$$

Here, I_ν is the *specific intensity* of the photon packet which describes the rate of radiation transport through a medium. As a photon packet travels through a medium, such as some gas, it will be subject to absorption and scattering as it interacts with the atoms in the gas. Equally, the medium itself may contribute to the radiation field through emission as excited atoms de-excite. Following how the specific intensity of a photon packet changes along its trajectory is a highly complex, high dimensional problem and challenging to model in full. Picture the number of photons emitted from a radiation source, such as a star, per unit frequency per unit time, and trying to model the individual interactions of those photons with all of the atoms in a 3D volume of gas in, say, a protoplanetary disc. The problem quickly grows out of hand. Instead, useful numerical approximations to radiative transfer are usually employed when modelling systems such as a star-disc system, some of which will be discussed in Sections 2.4.1.4 and 2.4.2.

Throughout this section I will outline some of the fundamental processes of radiative transfer, namely emission, absorption and scattering, and derive the *equation of radiative transfer* which encapsulates it all.

Emission:

The energy emitted by the medium can be described by defining an emission

coefficient, j_ν, which represents the energy emitted per unit volume per unit solid angle per unit time per unit frequency. The change in energy of the radiation field due to emission is then,

$$\mathrm{d}E_\nu = j_\nu \,\mathrm{d}V \,\mathrm{d}\Omega \,\mathrm{d}t \,\mathrm{d}\nu \,. \tag{2.40}$$

Equating this to Equation 2.39 gives,

$$\mathrm{d}I_\nu \,\mathrm{d}A = j_\nu \,\mathrm{d}V \,. \tag{2.41}$$

If we again consider radiation travelling through a cylinder of material with surface area, $\mathrm{d}A$, length, $\mathrm{d}s$ and volume, $\mathrm{d}V$, we can substitute that $\mathrm{d}V = \mathrm{d}A\,\mathrm{d}s$ which gives,

$$\mathrm{d}I_\nu = j_\nu \,\mathrm{d}s \,. \tag{2.42}$$

Hence, emission contributes to the specific intensity of a beam by an amount equal to the emission coefficient multiplied by the path length.

Absorption:

As a beam interacts with the atoms in a medium it will be subject to absorption which acts to remove energy from the beam. Absorption is generally formalised by introducing an absorption coefficient, α_ν. This is defined as the amount of absorption which occurs per unit length, and will be equal to $\alpha_\nu = n_a \sigma_\nu$, where n_a is the number of absorbers per unit volume and σ is their cross sectional areas. For astrophysical problems it is common to write this in terms of the medium's density, ρ, and its *opacity*, κ_ν, as,

$$\alpha_\nu = \rho \kappa_\nu. \tag{2.43}$$

The change in the specific intensity of a beam due to absorption is then,

$$\mathrm{d}I_\nu = -\alpha_\nu I_\nu \,\mathrm{d}s \,. \tag{2.44}$$

Scattering:

As well as being absorbed, photons may bounce off atoms in the medium, changing their direction of travel but not affecting their frequencies. Whether a photon packet is scattered or absorbed can be determined by the medium's *albedo*,

$$a = \frac{n_s \sigma_s}{n_s \sigma_s + n_a \sigma_a}, \tag{2.45}$$

where n_s and n_a are the number densities of scatterers and absorbers respectively, and σ_s and σ_a are their respective cross sections. From Equation 2.45 we can see that photons will be scattered with a probability a and absorbed with a probability $1 - a$ when interacting with atoms in the medium. When a photon is scattered, their new directions will follow some angular probability distribution function, $P(\cos\theta)$, which is normalised such that,

$$\int P(\cos\theta) \frac{\mathrm{d}\Omega}{4\pi} = 1. \tag{2.46}$$

Scattering will act to deflect photons out of the cylinder of material we are considering, thus reducing the specific intensity in that region.

Equation of radiative transfer:

Putting this all together into the *equation of radiative transfer*, we can say that the change in specific intensity of a photon packet as it travels through a medium is,

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = j_\nu - \alpha_\nu I_\nu, \tag{2.47}$$

where for simplicity we have considered absorption and emission only and have neglected the effects of scattering. We can rewrite this by introducing a new term, the *source function*, defined as the ratio of the emission and absorption coefficients, $S_\nu = j_\nu / \alpha_\nu$, which gives,

$$\frac{\mathrm{d}I_\nu}{\alpha_\nu \, \mathrm{d}s} = S_\nu - I_\nu. \tag{2.48}$$

It is also useful to introduce another new term here, the *optical depth* of a medium, defined as,

$$\tau_\nu = \alpha_\nu \, \mathrm{d}s = \rho \kappa_\nu \, \mathrm{d}s \, . \tag{2.49}$$

Hence, we can rewrite Equation 2.48 as,

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = S_\nu - I_\nu, \tag{2.50}$$

which has the solution,

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} S(\tau_\nu)e^{-\tau_\nu} \, \mathrm{d}\tau_\nu \, . \tag{2.51}$$

In the limit where the medium is optically thick such that $\tau_\nu \to \infty$ we obtain that $I_\nu(\tau_\nu) = S_\nu$. In other words, the specific intensity is equal to the ratio of emission and absorption coefficients, with no dependence on the initial intensity of the beam. When the medium is optically thin, such that $\tau_\nu \to 0$ and $e^{-\tau_\nu} \approx 1 - \tau_\nu$, we obtain that $I_\nu(\tau_\nu) = I_\nu(0) - (I_\nu(0) - S_\nu)\tau_\nu$. Hence when $\tau_\nu = 0$ the intensity of the beam will remain unchanged.

2.2 Protoplanetary disc structure

From the equations derived above we can derive analytic equations for the structure and evolution of discs, allowing us to build semi-analytic models as well as use these equations in fully hydrodynamical models. The equations of disc structure and disc physics outlined below have primarily been derived following (Pringle, 1981).

2.2.1 The thin disc approximation

A common simplifying approximation invoked when modelling accretion discs is to assume they are *razor-thin*, such that their radial extents, R, are much greater than their vertical extents,

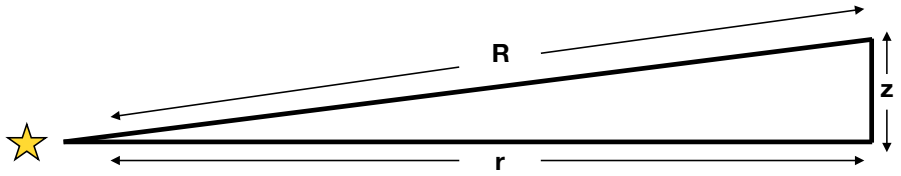

Figure 2.2 *Schematic of the geometry of a protoplanetary disc, highlighting various values for use in the derivations in Section 2.2.2.*

$$\frac{H}{R} \ll 1, \tag{2.52}$$

where H is the disc *scale height*. This *thin disc approximation* will prove very useful in the subsequent derivations of a disc's vertical and radial structure. In addition to being useful, observations of young protoplanetary discs suggest that it is likely a reasonable assumption, with typical measured values close to $H/R \sim 0.1$ (e.g. Andrews et al., 2010).

2.2.2 Equations of vertical disc structure

Consider a thin, non-self-gravitating disc which is in hydrostatic equilibrium such that,

$$\frac{\partial P}{\partial z} = -\rho \frac{\partial \Phi}{\partial z}. \tag{2.53}$$

The above equation expresses the balance in the z-direction between pressure (LHS) and gravitational forces (RHS). Substituting for Φ, and using the disc geometry illustrated in Figure 2.2 gives,

$$\frac{\partial P}{\partial z} = -\rho \frac{\mathrm{G}M_*}{R^2}\hat{z} = -\rho \frac{\mathrm{G}M_* z}{R^3}, \tag{2.54}$$

where $R = \sqrt{r^2 + z^2}$, and we are assuming a thin disc where $z \ll r$ such that $R \approx r$. Using Equation 2.36 we can substitute,

$$\frac{\partial P}{\partial z} = \frac{\partial P}{\partial \rho}\frac{\partial \rho}{\partial z} = c_s^2 \frac{\partial \rho}{\partial z}. \tag{2.55}$$

hence,

$$\frac{1}{\rho}\,\mathrm{d}\rho = -\frac{\mathrm{G}M_* z}{c_s^2 R^3}\,\mathrm{d}z\,, \tag{2.56}$$

which has the solution,

$$\rho = \rho_0 \exp\left[\frac{-\mathrm{G}M_* z^2}{2c_s^2 R^3}\right]. \tag{2.57}$$

Here, ρ_0 is a constant which represents the midplane density of the disc. This can be rewritten by substituting $\Omega^2 = GM_*/R^3$, and introducing a variable called the *disc scale height*, H, where,

$$H = \frac{c_s}{\Omega}, \tag{2.58}$$

such that,

$$\rho = \rho_0 \exp\left(\frac{-z^2}{2H^2}\right). \tag{2.59}$$

Thus we have derived that the vertical density profile of a disc can be approximated by a Gaussian function, where H represents the length scale over which the disc's density will decrease by a factor e. In the case of a self-gravitating disc, where we now include the disc's mass in the calculation of the gravitational potential, the vertical density profile is given by (Spitzer, 1942),

$$\rho = \frac{\rho_0}{\cosh^2(\frac{\mathrm{z}}{\mathrm{H_{sg}}})}, \tag{2.60}$$

where the self-gravitating scale height is,

$$H_{sg} = \frac{c_s^2}{\pi \mathrm{G}\Sigma}. \tag{2.61}$$

2.2.3 Equations of radial disc structure

We can understand the radial structure of a protoplanetary disc by first considering the radial component of the Navier-Stokes equation for an inviscid fluid,

$$\frac{\partial \mathrm{v_r}}{\partial t} + v_r \frac{\partial v_r}{\partial t} + \frac{v_\phi}{r}\left(\frac{\partial v_r}{\partial \phi} - v_\phi\right) + v_z \frac{\partial v_r}{\partial z} = -\frac{1}{\rho}\frac{\partial P}{\partial r} - \frac{\partial \mathbf{\Phi}}{\partial r}, \tag{2.62}$$

where we have made use of the material derivative of a vector in cylindrical polar coordinates.

In a radially stable disc we can assume that the radial drift velocity will be highly subsonic ($v_r \ll c_s$). Using the equation for the disc scale height (Equation 2.58), substituting $\Omega = v_\phi/R$, and making use of the thin disc approximation ($H/R \ll 1$) then gives that $v_r \ll c_s \ll v_\phi$. Thus we can discount the v_r terms from Equation 2.62 as being negligible. In doing so we arrive at,

$$\frac{v_\phi^2}{r} = \frac{\mathrm{G}M_*}{r^2} + \frac{1}{\rho}\frac{\partial P}{\partial r}. \tag{2.63}$$

If we assume that the radial pressure gradient is negligible, then we arrive at the standard equations for the Keplerian orbital velocity and orbital frequency,

$$v_\phi = v_\mathrm{K} = \sqrt{\frac{\mathrm{G}M_*}{r}}, \tag{2.64}$$

$$\Omega = \frac{v_\mathrm{K}}{r} = \sqrt{\frac{\mathrm{G}M_*}{r^3}}. \tag{2.65}$$

More accurately, the pressure gradient term in Equation 2.63 will be small, but not negligible. As the gas pressure will generally be a decreasing function of radius, the pressure gradient term will be negative, hence the gas in the disc will orbit at slightly sub-Keplerian velocities. This will become important when considering the gas' interaction with the dusty component of the disc in later chapters, as aerodynamic drag between the gas and dust results in radial migration of the solids.

Assuming that the pressure gradient follows some radial power-law,

$$P = P_0 \left(\frac{r}{r_0} \right)^{-n}, \tag{2.66}$$

and substituting $P_0 = \rho_0 c_s^2$, Equation 2.63 can be rearranged to give,

$$v_\phi = v_K \left(1 - n \frac{c_s^2}{v_K^2} \right)^{1/2}. \tag{2.67}$$

For a thin disc, where $H/R \ll 1$, our definition of the disc scale height from Equation 2.58 gives that $c_s/v_K \ll 1$, hence any deviation from Keplerian velocities will be small, typically with $v_\phi \gtrsim 0.99 v_\mathrm{K}$.

2.3 Protoplanetary disc evolution

2.3.1 Surface density evolution of a disc

The evolution of a protoplanetary disc's surface density can be derived from the continuity and momentum equations derived in Section 2.1.1. Starting with the conservation of mass within a radial annulus of the disc of width, Δr, such that the mass within the annulus is, $m = 2\pi r \Delta r \Sigma$, we can rewrite Equation 2.7 as,

$$\frac{\partial}{\partial t}(2\pi r \Delta r \Sigma) = 2\pi r \Sigma(r) v_r(r) - 2\pi (r + \Delta r) \Sigma(r + \Delta r) v_r(r + \Delta r). \tag{2.68}$$

Dividing by $2\pi \Delta r$, noting that neither r nor Δr are varying with time, and taking the limit $\Delta r \to 0$, we get,

$$r \frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial r}(r \Sigma v_r) = 0. \tag{2.69}$$

Now considering the conservation of angular momentum within the annulus, where the angular momentum is $L = m v_\phi r$ and the azimuthal velocity of the gas is $v_\phi = r\Omega$, we get that $L = 2\pi r \Delta r \Sigma r^2 \Omega$. Hence the rate of change of angular momentum within the annulus is,

$$\frac{\partial L}{\partial t} = 2\pi r \Delta r \frac{\partial}{\partial t}(\Sigma r^2 \Omega). \tag{2.70}$$

Similarly, we can also state that the rate of change of angular momentum within an annulus is equal to rate of change of surface density between the inner and outer edges of the annulus, *plus* the difference in the viscous torque, Γ, exerted at the inner and outer edges, such that,

$$\begin{aligned}\frac{\partial L}{\partial t} = 2\pi r \Sigma(r) r^2 \Omega(r) v_r(r) - \\ 2\pi(r+\Delta r)\Sigma(r+\Delta r)(r+\Delta r)^2 \Omega(r+\Delta r) v_r(r+\Delta r) + \\ \Gamma(r) - \Gamma(r+\Delta r).\end{aligned} \tag{2.71}$$

Again, dividing by $2\pi\Delta r$ and taking the limit $\Delta r \to 0$ we get,

$$\frac{1}{2\pi\Delta r}\frac{\partial L}{\partial t} = -\frac{\partial}{\partial r}(r\Sigma r^2 \Omega v_r) + \frac{1}{2\pi}\frac{\partial}{\partial r}\Gamma. \tag{2.72}$$

For a viscous fluid with viscosity, ν, the viscous torque exerted on a radial annulus can be calculated from the circumference of the annulus ($2\pi r$) the viscous force per unit length ($\nu\Sigma r\frac{\mathrm{d}\Omega}{\mathrm{d}r}$) and the length of the moment arm, r, as,

$$\Gamma = 2\pi r \cdot \nu\Sigma r \frac{\mathrm{d}\Omega}{\mathrm{d}r} \cdot r. \tag{2.73}$$

Substituting this into Equation 2.72 and equating Equations 2.70 and 2.72 gives,

$$\frac{\partial}{\partial t}(\Sigma r^2 \Omega) + \frac{1}{r}\frac{\partial}{\partial r}(r\Sigma r^2 \Omega v_r) = \frac{1}{r}\frac{\partial}{\partial r}\left(\nu\Sigma r^3 \frac{\partial \Omega}{\partial r}\right). \tag{2.74}$$

We can then use Equations 2.69 and 2.74 to eliminate v_r (where we note that $\frac{\partial}{\partial t}(\Sigma r^2\Omega) = r^2\Omega\frac{\partial\Sigma}{\partial t}$ as r and Ω don't vary with time). After some rearranging we find,

$$v_r = \frac{2}{r^2 \Omega \Sigma}\frac{\partial}{\partial r}\left(\nu\Sigma r^3 \frac{\partial \Omega}{\partial r}\right). \tag{2.75}$$

Substituting this equation for v_r into Equation 2.69, and assuming Keplerian rotation with $\Omega = \sqrt{GM/r^3}$, we arrive at the final form of our equation for the surface density evolution of a viscous disc,

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left(r^{1/2}\frac{\partial}{\partial r}(\nu \Sigma r^{1/2}) \right). \tag{2.76}$$

2.3.2 Steady state solution for mass accretion

Solving the equations of surface density evolution for a disc which has settled into a state with a constant mass accretion rate will prove useful in later chapters when modelling the instantaneous state of a disc. This is known as the *steady state solution*, and can be derived beginning with Equation 2.74 noting that in the steady state solution the $\partial/\partial t$ terms go to zero. Hence we have,

$$\frac{\partial}{\partial r}(\Sigma r^3 \Omega v_r) = \frac{\partial}{\partial r}\left(r^3 \nu \Sigma \frac{\partial \Omega}{\partial r} \right). \tag{2.77}$$

Taking the mass within a radial annulus as, $M = 2\pi r \Delta r \Sigma$, hence the rate of change of mass travelling out from the inner edge of the annulus is,

$$\dot{M} = -2\pi r \Sigma v_r = \text{constant}. \tag{2.78}$$

Rearranging for Σv_r and substituting into Equation 2.77 gives,

$$\frac{\partial}{\partial r}\left(\frac{-\dot{M}}{2\pi} r^2 \Omega \right) = \frac{\partial}{\partial r}\left(r^3 \nu \Sigma \frac{\partial \Omega}{\partial r} \right). \tag{2.79}$$

Integrating from the inner edge of the disc, r_{in}, to an arbitrary location in the outer disc, r, and assuming that there is no torque exerted on the inner edge of the disc such that $\frac{\partial \Omega}{\partial r_{\text{in}}} = 0$, we get,

$$\frac{-\dot{M}}{2\pi} r^2 \Omega + \frac{\dot{M}}{2\pi} r_{\text{in}}^2 \Omega = r^3 \nu \Sigma \frac{\partial \Omega}{\partial r}. \tag{2.80}$$

Substituting $\Omega = \sqrt{GM/r^3}$, hence $\frac{\partial \Omega}{\partial r} = \frac{-3}{2}\sqrt{GM/r^5}$, and rearranging for $\dot{M}$

gives,

$$\dot{M} = 3\pi\Sigma\nu\left(1 - \sqrt{\frac{r_{\rm in}}{r}}\right). \tag{2.81}$$

When $r \gg r_{\rm in}$ this can be simplified to the more common expression for the steady state mass accretion rate of a protoplanetary disc,

$$\dot{M} = 3\pi\Sigma\nu. \tag{2.82}$$

2.3.3 Viscosity and $\alpha-$discs

The equations derived in Section 2.3.1 and Section 2.3.2 demonstrate the importance of viscosity in carrying away angular momentum whilst allowing material to move inwards and accrete onto the central star. Hence to understand the evolution of a disc and solve the above equations requires consideration of the nature of the viscosity within a disc, which we have so far alluded to.

Kinetic viscosity is the simplest and immediately obvious choice to consider. This arises from particle-particle collisions, and can be characterised by the mean free path of particles, λ, and their typical collisional velocities, which may be approximated by the local sound speed, c_s. Hence we can state that,

$$\nu_{\rm kin} \sim c_s\lambda = \frac{c_s}{n\sigma}, \tag{2.83}$$

where n is the number density of particles and σ is their collisional cross section.

More generally, viscosity can be written as $\nu = l \cdot v = l^2/t$, where l is the characteristic length. Hence we can state that the viscous timescale of a disc, which represents the timescale on which angular momentum will be redistributed under the effect of viscous torques, is,

$$t_\nu = \frac{l^2}{\nu}. \tag{2.84}$$

Assuming typical values for a protoplanetary disc, consisting mostly of molecular Hydrogen with $\sigma \approx 2 \times 10^{-15}\,{\rm cm}^2$, $n = 10^{12}\,{\rm cm}^{-3}$, and a sound speed at 10 AU of $c_s \approx 500\,{\rm ms}^{-1}$, we obtain that $t_\nu \approx 10^{13}\,{\rm yrs}$ - longer than the age of the Universe,

and around 6 orders of magnitude larger than typical disc lifetimes. Consequently we must look elsewhere for an appropriate source of viscosity in discs.

It is interesting to consider the *Reynolds number* implied if molecular collisions are the dominant source of disc viscosity. The Reynolds number is defined as the ratio of inertial forces to viscous forces within a fluid,

$$Re = \frac{|\mathbf{F}_{\text{interial}}|}{|\mathbf{F}_{\text{viscous}}|} = \frac{|\mathrm{d}\mathbf{v}/\mathrm{d}t_{\text{dyn}}|}{|\mathrm{d}\mathbf{v}/\mathrm{d}t_{\text{visc}}|} = \frac{vl}{\nu}. \tag{2.85}$$

Fluids with low viscosities imply a large Reynolds number, where any perturbation to a flow will be able to persist with little viscous resistance. Hence the flow may become unstable and *turbulent*. Substituting typical values for a protoplanetary disc at 10 AU into Equation 2.85, approximating that $v = c_s$ and $l = H = 0.1R$, gives $Re \sim 10^{10}$. At such high Reynolds numbers it is thought that a fluid will be highly turbulent, hence it is commonly thought that protoplanetary discs are turbulent.

The presence of a turbulent flow may provide an effective *turbulent viscosity*, whereby turbulent eddies give rise to mixing of the fluid which generates friction in a manner similar to kinetic viscosity except on a macroscale. We can approximate the magnitude of the turbulent viscosity within a disc by appealing to dimensional constraints. Assuming that the length scale of any turbulent eddies must be of the order of the disc scale height, whilst the characteristic velocity of any turbulent eddies must not exceed the sound speed, we can state that,

$$\nu = \alpha c_s H, \tag{2.86}$$

where α is a dimensionless coefficient of value less than 1. This was first proposed by Shakura & Sunyaev (1973), and has since become commonly known as the Shakura-Sunyaev $\alpha-$disc model, characterised by the viscous$-\alpha$ parameter. This model is convenient, as it allows us to contain all of our uncertainty about the source of the turbulence within the α coefficient, where we now need to only approximate a reasonable value for the effective α. The value of α can be measured through observations of discs, which has traditionally been done through measuring how disc accretions rates decrease with disc age, allowing us to measure the disc viscous timescale hence also allowing us to estimate α (Hartmann et al., 1998). Recently using high resolution spectroscropy with ALMA, it has also been possible to measure turbulence in discs through measurements of molecular

emission line widths, again allowing us to estimate α (Teague et al., 2016; Flaherty et al., 2015, 2017, 2018, 2020). Typical values are found to be in the range $\alpha = 10^{-1} - 10^{-3}$, with a strong dependence on the phase of disc evolution being considered. In the early phases of disc evolution whilst the disc-to-star mass ratio is high, GI may generate an effective gravito-turbulent viscosity. In the later stages of disc evolution when the disc density is lower and more optically thin to stellar radiation, the gas may become highly ionised resulting in magnetohydrodynamic (MHD) turbulence (for a review see Lesur et al., 2022). Throughout this thesis we will generally be concerned with young, self-gravitating discs which aren't highly ionised, hence when gravito-turbulence is effective.

2.4 Useful numerical methods

2.4.1 Smoothed Particle Hydrodynamics

In order to understand the 3-dimensional evolution of protoplanetary discs we require sophisticated numerical methods which model the system's fluid dynamics. Historically two main methods have been employed to approach this, namely grid-based and particle-based methods.

In grid-based methods the simulation domain is set up within a 3 dimensional box which is further sub-divided into many smaller grid cells. The system is evolved by solving the equations of hydrodynamics at grid cell interfaces and calculating the flux of mass density moving between neighbouring cells. When grid-based methods were first developed, the grids were set up at the start of the simulation and remained fixed throughout the remainder of the system's evolution. However this approach was subject to resolution issues, which led to the development of *Adaptive Mesh Refinement* (AMR) grid-based approaches where the grid cells are dynamically refined based on some mass density resolution criteria.

Particle-based methods were developed later. Instead of splitting the computational domain into a set of grid cells, the fluid being simulated (which is continuous in nature) is discretised into a set of N pseudo-particles which move with the flow of the fluid, and have properties equal to the fluid properties at each particle's location. In order for the particles to more accurately represent a continuous fluid the properties of each particle are *smoothed* over a finite volume centered on the particle's location. Each particle's mass, density, velocity and internal

energy (and any other fluid properties desired to be determined) are calculated as a weighted average over the properties of the nearest neighbouring particles, defined as any other particles located within the particle's *smoothing kernel*. Such a method has become known as *Smoothed Particle Hydrodynamics* (SPH) and was initially developed by Lucy (1977) and Gingold & Monaghan (1977) for use in astrophysical problems (also see Hernquist & Katz, 1989; Benz, 1990; Monaghan, 1992). It has since been adopted by a wide range of fields involving the modelling of fluid mechanics, such as biology and engineering.

It would be incorrect to say that either grid-based or particle-based approaches are better than the other. Instead, the two methods are better suited to different astrophysical problems. Grid-based approaches aren't optimised for problems with a large dynamical range. They can be subject to resolution issues when, for example, simulating star formation and a parsec-sized cloud collapses to form an object of scale $\sim 10^5\,\mathrm{km}$, or region of a protoplanetary disc collapses to form a protoplanet as we consider in this thesis. Although this has become less problematic since the introduction of AMR grids, grid-based codes would still require extremely large parent cells to be defined in order to capture the full computational domain in such scenarios. SPH naturally deals with this by following the motions of particles, hence the computational domain and spatial resolution are naturally adaptable.

Throughout this thesis we use the SPH method when considering 3-dimensional hydrodynamical models. In the following section I will outline some of the fundamental principles of SPH as well as some of the specific SPH algorithms utilised here.

2.4.1.1 The SPH smoothing kernel

For the research throughout this thesis, I use the PHANTOM SPH code (Price et al., 2018), hence the following discussion will be with reference to the specific prescriptions which PHANTOM uses.

Each SPH pseudo-particle has an associated mass, velocity, position and internal energy, from which all other necessary quantities can be calculated and subsequently evolved using the equations of hydrodynamics derived in Section 2.1.1. The density of particle a is equal to the fluid's density at the location, $\mathbf{r}_a$, and is calculated by summing over the masses of the nearest neighbouring

particles as,

$$\rho(\mathbf{r}_a) = \sum_b m_b W(\mathbf{r}_a - \mathbf{r}_b, h_a), \tag{2.87}$$

where W is the SPH smoothing kernel, and h is the *smoothing length*. The smoothing kernel defines the volume in which the sum is performed, while also weighting the density sum so that neighbouring particles at the edge of the kernel volume contribute less than those near the centre. An illustration of this is shown in Figure 2.3. Typically the functional form of the kernel will be spherically symmetric, monotonically decreasing and continuous so that derivatives can be calculated. An immediately obvious choice would be a Gaussian function, however a Gaussian would be impractical in practise as at no point is it equal to zero. In the case of PHANTOM, we use an M_4 cubic B-spline of the form,

$$W(\mathbf{r}_a - \mathbf{r}_b, h_a) = \frac{1}{\pi h_a^3} \begin{cases} 1 - \frac{3}{2}q^2 + \frac{3}{4}q^3, & 0 \leq q < 1; \\ \frac{1}{4}(2-q)^3, & 1 \leq q < 2; \\ 0, & q \geq 2, \end{cases} \tag{2.88}$$

,

where $q = (\mathbf{r}_a - \mathbf{r}_b)/h_a$. From this, it is clear that any particles at $q \geq 2$ won't contribute to the local fluid properties, i.e.

$$W = 0 \text{ for } |\mathbf{r}_a - \mathbf{r}_b| > 2h_a. \tag{2.89}$$

The smoothing length is set to be dynamic and a function of the local density so that the kernel volume always contains the same number of neighbouring particles. This is desirable to ensure that the simulation remains resolved at both high and low particle densities. When considering a 3D simulation domain we use a smoothing length of the form,

$$h_a = h_{\text{fact}} \left(\frac{m_a}{\rho_a} \right)^{1/3}, \tag{2.90}$$

where h_{fact} is a normalisation constant.

For the smoothing kernel to make physical sense it must be normalised as,

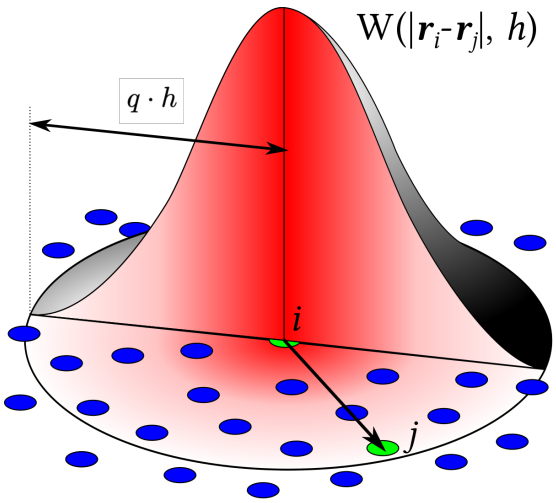

Figure 2.3 *An illustration of how the SPH smoothing kernel is used to determine the state variables of particle i as a weighted sum over the properties of its nearest neighbouring particles, defined as those which lie within the extent of the kernel's radius. Image credit: Jlcercos*

$$\int W(\mathbf{r}_a - \mathbf{r}_b, h_a)\, \mathrm{d}r_b = 1, \tag{2.91}$$

and converge to a delta-function in the limit $h \to 0$,

$$\lim_{h\to 0} W(\mathbf{r}_a - \mathbf{r}_b, h_a) = \delta(\mathbf{r}_a - \mathbf{r}_b). \tag{2.92}$$

A more general form of Equation 2.87 for calculating any arbitrary function, $f(\mathbf{r}_a)$, in terms of the smoothing kernel is,

$$f(\mathbf{r}_a) = \sum_b \frac{m_b}{\rho_b} f_b W(\mathbf{r}_a - \mathbf{r}_b, h_a), \tag{2.93}$$

where m_b/ρ_b represents the effective volume of particle b. From this we can also calculate derivatives of any function in terms of the smoothing kernel straightforwardly. As $W(\mathbf{r}_a - \mathbf{r}_b, h_a)$ is the only term in Equation 2.93 which explicitly depends on r we get that,

$$\nabla f(\mathbf{r}_a) = \sum_b \frac{m_b}{\rho_b} f_b \nabla W(\mathbf{r}_a - \mathbf{r}_b, h_a). \tag{2.94}$$

2.4.1.2 Equations of hydrodynamics in SPH

Having introduced the smoothing kernel we can now write the equations of hydrodynamics in the form that they take in SPH.

The continuity equation: The equation of mass continuity can be derived straightforwardly from Equation 2.87. Abbreviating $W(\mathbf{r}_a - \mathbf{r}_b, h_a)$ to W_{ab}, we can state,

$$\frac{\mathrm{d}\rho_a}{\mathrm{d}t} = \frac{\partial}{\partial t}\left(\sum_b m_b W_{ab}\right). \tag{2.95}$$

As the mass of each pseudo-particle remains constant we need to only consider the time derivative of the smoothing kernel, which is equal to,

$$\frac{\mathrm{d}W_{ab}}{\mathrm{d}t} = \frac{\partial W_{ab}}{\partial r_{ab}}\frac{\partial r_{ab}}{\partial t} = \mathbf{v}_{ab}\cdot\nabla_a W_{ab}, \tag{2.96}$$

where v_{ab} is the relative velocity of particles a and b, and ∇_a is the gradient of the kernel at the location of particle a. Hence we have that,

$$\boxed{\frac{\mathrm{d}\rho_a}{\mathrm{d}t} = \sum_b m_b \mathbf{v_{ab}}\cdot\nabla_a W_{ab}.} \tag{2.97}$$

Equation 2.97 is not directly solved in SPH when calculating the particle density. As the particle masses don't change, mass is conserved naturally in SPH. In the case of PHANTOM, Equation 2.97 is only really used to reduce the number of iterations required when predicting the smoothing length at the next timestep, as h and ρ have a mutual dependence so they must be determined simultaneously through an iterative, rootfinding procedure (described in more detail in Price et al. (2018)).

The momentum equation: Starting with Equation 2.10 we can state that in the absence of any external forces,

$$\frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t} = -\frac{\nabla P_a}{\rho_a} = \frac{-1}{\rho_a}\sum_b \frac{m_b}{\rho_b} P_b \nabla_a W_{ab}, \tag{2.98}$$

where we have substituted for ∇P_a using Equation 2.94. However if we consider the pairwise force between particles a and b,

$$\mathbf{F}_{ab} = \left(m_a \frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t}\right)_b = -\frac{m_a}{\rho_a}\frac{m_b}{\rho_b} P_b \nabla_a W_{ab} \tag{2.99}$$

$$\mathbf{F}_{ba} = \left(m_b \frac{\mathrm{d}\mathbf{v}_b}{\mathrm{d}t}\right)_a = -\frac{m_b}{\rho_b}\frac{m_a}{\rho_a} P_a \nabla_b W_{ba} = \frac{m_b}{\rho_b}\frac{m_a}{\rho_a} P_a \nabla_a W_{ab}, \tag{2.100}$$

we can see that when $P_a \neq P_b$ (which in general will be the case) we have that $F_a \neq -F_b$, hence momentum will not be conserved. Instead, we can derive the *symmetrized* version of the momentum equation. Starting by using the product rule as,

$$\frac{\nabla P}{\rho} = \nabla\left(\frac{P}{\rho}\right) + \frac{P}{\rho^2}\nabla\rho. \tag{2.101}$$

We can substitute this into Equation 2.94 to obtain that,

$$\frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t} = -\frac{\nabla P_a}{\rho_a} = -\sum_b \frac{m_b}{\rho_b}\frac{P_b}{\rho_b}\nabla_a W_{ab} - \frac{P_a}{\rho_a^2}\sum_b \frac{m_b}{\rho_b}\rho_b \nabla_a W_{ab}. \tag{2.102}$$

Hence,

$$\boxed{\frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t} = -\sum_b m_b\left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2}\right)\nabla_a W_{ab},} \tag{2.103}$$

which conserves momentum as the pressure term is symmetric in a and b, so that $\frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t} = -\frac{\mathrm{d}\mathbf{v}_b}{\mathrm{d}t}$.

The energy equation: The internal energy equation can be derived straightforwardly, starting with Equation 2.29,

$$\frac{\mathrm{d}u_a}{\mathrm{d}t} = \frac{P_a}{\rho_a^2}\frac{\mathrm{d}\rho_a}{\mathrm{d}t} = \frac{P_a}{\rho_a^2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_b m_b W_{ab}\right), \tag{2.104}$$

Substituting the time derivative for Equation 2.96 then gives the energy equation in SPH,

$$\boxed{\frac{\mathrm{d}u_a}{\mathrm{d}t} = \frac{P_a}{\rho_a^2}\sum_b m_b \mathbf{v}_{ab}\cdot\nabla_a W_{ab}.} \tag{2.105}$$

2.4.1.3 Including artificial viscosity to capture shocks

Regions of steep pressure or density gradients in the fluid, known as *shocks*, present a problem in SPH. They appear as a discontinuity in the flow on scales smaller than the smoothing length, meaning that the differential equations which govern the particle motions cannot be solved. On very small scales (i.e. the mean free path of the gas particles) shock fronts will be continuous, but these scales will be *much* smaller than any realistically achievable resolution in our simulations. Equally, if the shock front's length scale is smaller than our particle resolution, there is no way of knowing where the shock should be inserted between neighbouring particles.

A common solution for dealing with shocks in SPH is to introduce an additional term which acts to smooth the discontinuity over a length scale which can be resolved by the smoothing kernel. This is achieved by including an *artificial viscosity* term into the momentum and energy equations, which are modified to be:

The momentum equation

$$\frac{\mathrm{d}\mathbf{v}_a}{\mathrm{d}t} = -\sum_b m_b\left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2} + \Pi_{ab}\right)\nabla_a W_{ab}. \tag{2.106}$$

The energy equation

$$\frac{\mathrm{d}E_a}{\mathrm{d}t} = \frac{P_a}{\rho_a^2}\sum_b m_b \mathbf{v}_{ab}\cdot\nabla_a W_{ab} + \frac{1}{2}\sum_b m_b \Pi_{ab}\mathbf{v}_{ab}\cdot\nabla_a W_{ab}, \tag{2.107}$$

where,

$$\Pi_{ab} = \begin{cases} \frac{-\alpha \overline{c}_{ab} \mu_{ab} + \beta \mu_{ab}^2}{\overline{\rho}_{ab}}, & \mathbf{v}_{ab} \cdot r_{ab} \leq 0; \\ 0, & \mathbf{v}_{ab} \cdot r_{ab} > 0, \end{cases} \tag{2.108}$$

where $\overline{c}_{ab}$ is the average sound speed, $\overline{\rho}_{ab}$ is the average density, and,

$$\mu_{ab} = \frac{h \mathbf{v}_{ab} \cdot \mathbf{r}_{ab}}{r_{ab}^2 + \epsilon \overline{h}_{ab}^2}. \tag{2.109}$$

α and β in Equation 2.108 are free parameters which characterise the artificial viscosity. The α term is linear in the particles' velocity differences so contributes a bulk and shear viscosity. The β term is quadratic in $\mathbf{v}_{ab}$ so dominates in cases of large velocity differences, thus preventing particle-particle interpenetration. It is generally recommended that $\beta = 2\alpha$, although the exact values for α and β vary between different works. Throughout the work in this thesis we use $\alpha = 0.1$ and $\beta = 0.2$.

2.4.1.4 Radiative transfer in SPH

Accurate modelling of radiation transport within our simulations is essential to the work undertaken here. As discussed in Section 1.4 the growth of gravitational instabilities, and ultimately whether the disc undergoes fragmentation, is sensitive to the balance of heating and cooling within the disc. We therefore want to model the disc's thermodynamics as accurately as possible. However, as always, there is a compromise to be struck between the accuracy of our model and the computational expense we can feasibly afford. Full treatment of radiation transport is currently beyond our capabilities. Highly accurate methods exist, such as Monte Carlo radiative transfer (which will be discussed in Section 2.4.2), but would take far too long to compute at each SPH timestep to be a viable option.

Early work studying gravitationally unstable discs considered the cooling prescription mentioned in Section 1.6.2, where the fluid's radiative cooling rate is proportional to the disc's dynamical timescale. Whilst this is a convenient and computationally efficient method to model disc evolution, it doesn't really capture any of the actual thermodynamics of the disc.

Throughout the work undertaken in this thesis we generally model radiative transfer in SPH using the hybrid method developed in Forgan et al. (2009), where the disc thermodynamics are modelled through a combination of hydrodynamics, $\dot{u}_{i,\mathrm{hydro}}$, (such as $P\,\mathrm{d}V$ work) radiative cooling, $\dot{u}_{i,\mathrm{cool}}$, and diffusive energy exchange between neighbouring particles, $\dot{u}_{i,\mathrm{diff}}$, as,

$$\dot{u}_{i,\mathrm{total}} = \dot{u}_{i,\mathrm{hydro}} + \dot{u}_{i,\mathrm{cool}} + \dot{u}_{i,\mathrm{diff}}. \tag{2.110}$$

In this formalism, we calculate the radiative cooling term using the polytropic cooling approximation from Stamatellos et al. (2007), where each SPH particle is treated as a spherically symmetric, polytropic pseudo-cloud, of polytropic index $n = 2$. Each particle's density, ρ_i, temperature, T_i and gravitational potential, Φ_i, are used to calculate the mean column density, $\overline{\Sigma}_i$, opacity, κ_i, and mass weighted opacity, $\overline{\kappa}_i$, from which the cooling term can be calculated as,

$$\dot{u}_{i,\mathrm{cool}} = -\frac{4\sigma(T_i^4 - T_0^4)}{\overline{\Sigma}_i^2\overline{\kappa}_i(\rho_i, T_i) + \kappa_i^{-1}(\rho_i, T_i)}, \tag{2.111}$$

where T_0 represents the minimum disc temperature set by background irradiation. From Equation 2.111 we see that there are two limiting cases:

1. When we are in a region of large column density, hence also a high optical depth so that $\overline{\Sigma}_i^2\overline{\kappa}_i(\rho_i, T_i) \gg \kappa_i^{-1}(\rho_i, T_i)$, and the cooling term reduces to,

$$\dot{u}_{i,\mathrm{cool}} = -\frac{4\sigma(T_i^4 - T_0^4)}{\overline{\Sigma}_i^2\overline{\kappa}_i(\rho_i, T_i)} \tag{2.112}$$

2. When we are in a region of low optical depth so that $\overline{\Sigma}_i^2\overline{\kappa}_i(\rho_i, T_i) \ll \kappa_i^{-1}(\rho_i, T_i)$, the cooling term reduces to,

$$\dot{u}_{i,\mathrm{cool}} = -4\sigma(T_i^4 - T_0^4)\kappa_i(\rho_i, T_i) \tag{2.113}$$

From 1 and 2 it is clear that cooling becomes inefficient in both the optically thick and optically thin regimes, as $\dot{u}_{\mathrm{cool}}$ becomes small in both scenarios.

Stamatellos et al. (2007) outlines two main limitations to this polytropic approximation, the first being that it assumes each particle is a spherically symmetric pseudo-cloud, which may not always be accurate for regions of the

simulation where the geometry is not so. The other main limitation is that the model doesn't deal with energy exchange between neighbouring particles.

This second point is addressed by the diffusion term in Equation 2.110. Forgan et al. (2009) use the flux limited diffusion approximation (e.g. Bodenheimer et al., 1990; Cleary & Monaghan, 1999; Mayer et al., 2007b) to calculate $\dot{u}_{i,\mathrm{diff}}$, as,

$$\dot{u}_{i,\mathrm{diff}} = \sum_b \frac{4m_b}{\rho_i \rho_b} \frac{k_i k_b}{k_i + k_b} (T_i - T_b) \frac{\mathbf{r}_{ib} \cdot \nabla W}{|r_{ib}|^2}. \tag{2.114}$$

Here, the subscript b refers the nearest neighbours of particle i, hence the diffusion term is calculated by summing over the energy exchange from all neighbouring particles. The thermal conductivity, k, is calculated as,

$$k_i = \frac{16\sigma}{\rho_i k_i} \lambda_i T_i^3, \tag{2.115}$$

where λ_i is the *flux limiter*, described by the local radiation field,

$$\lambda_i(R_i) = \frac{2 + R_i}{6 + 3R_i + R_i^2} \tag{2.116}$$

where,

$$R_i = \frac{|\nabla u_i(\mathbf{r}_i)|}{u_i(\mathbf{r}_i)\rho_i \kappa_i}. \tag{2.117}$$

Again, we find two limiting cases for $\dot{u}_{i,\mathrm{diff}}$:

1. In optically thick regions of high density, ρ_i and κ_i are large, hence R_i is small, λ_i reduces to $1/3$ and the thermal conductivity reduces to the Rosseland diffusion approximation, $k_i = \frac{16\sigma T_i^3}{3\rho_i k_i}$.

2. In optically thin regions where ρ_i and κ_i are small, R_i becomes large and $\lambda_i \to 0$. Therefore the conductivity terms and $\dot{u}_{i,\mathrm{diff}}$ also go to zero, hence we have no diffusive energy exchange between neighbouring particles.

On its own, the flux limited diffusion approximation models energy exchange between particles well in optically thick regions, but breaks down in regions of

low optical depths where the equation reduces to zero. It also doesn't allow the system to radiatively lose energy, as it only accounts for diffusive energy exchange. The hybrid method in Equation 2.110 combines both the polytropic cooling approximation and the flux limited diffusion approximation to form a complimentary method which is able to model the disc thermodynamics well at both high and low optical depths.

In the chapters of this thesis which represent earlier work (Chapters 3 and 6) we note that the full hybrid method had not yet been implemented into PHANTOM, hence radiation transport in SPH was modelled using only the polytropic approximation from Stamatellos et al. (2007). By the time of the later chapters (Chapters 4 and 5) we had implemented the full hybrid method into the code.

2.4.2 Monte Carlo radiation transport

In Chapter 6 we generate synthetic observations of self-gravitating discs, which requires accurate calculation of the dust temperatures throughout the system so that continuum images can be produced. In situations such as this, where we consider an instantaneous snapshot of a disc, we can afford a more expensive treatment of radiation transport than was discussed in Section 2.4.1.4. *Monte Carlo radiation transport* (MCRT) (Lucy, 1999) is a probabilistic approach to radiative transfer, where a fixed number of photon packets are iteratively emitted from a radiation source, such as the star at the center of the disc. Photons are then followed along their trajectories from the star to the edge of the computational domain. As they interact with the gas they will either be absorbed and re-emitted, or scattered through some angle. This will continue until all of the photons which were initially emitted from the source have reached the edge of the computational domain, at which point dust temperatures can be calculated. Another set of photon packets will then be emitted from the star with the updated dust temperatures, and the process is repeated until the dust temperatures converge.

To do this, we begin by constructing a disc within an AMR grid. The grid initially consists of a singular parent cell which is centred on the disc centre and encloses the full computational domain. It is then repeatedly subdivided from parent cells into 2^D child cells based on some mass resolution criteria, where D represents the dimensions of the computational domain (3 dimensional here). If the mass within

a cell exceeds the resolution criteria then a parent cell will be divided into 2^D child cells such that child cells then become parent cells. This continues until the mass in each cell is less than or equal to our predefined mass resolution criteria.

Once the grid has been constructed, N photon packets are emitted instantaneously and isotropically from the star, and their trajectories are followed through the grid, as illustrated in Figure 2.4. Once a photon is emitted it will travel a random optical depth through the disc, calculated using a random number, r,

$$\tau_\nu = -\ln(1 - r), \tag{2.118}$$

which, combined with the medium's density and opacity can be translated to a physical distance (see Equation 2.49). After travelling through a depth, τ_ν, the photon will either be absorbed (and re-emitted) or scattered through some angle. Which of these outcomes occurs is determined probabilistically using the albedo of the material, which is simply the ratio of the scattering opacity to the scattering plus absorption opacities. If the photon is absorbed and re-emitted then its new frequency and direction will also be randomly sampled from the disc emissivity, and if it is scattered then the new direction of travel will also be randomly sampled from some angular probability distribution function.

Once all of the photons have reached the edge of the computational domain dust temperatures can be calculated. A photon will contribute to the energy of each grid cell by an amount proportional to the time which it spent within the cell, $\epsilon\delta t/\Delta t$, where $\delta t = l/c$ (the distance travelled over the photon velocity) and Δt is the total duration of the iteration. At the end of each iteration the energy density in a cell of volume, V, contributed to by photons in the frequency range ν and $\nu + \mathrm{d}\nu$ is given by,

$$U_\nu \,\mathrm{d}\nu = \frac{\epsilon}{c\Delta t}\frac{1}{V}\sum l. \tag{2.119}$$

Dust temperatures can be calculated assuming radiative equilibrium (where the absorption rate equals the emission rate) as,

$$T_d = \left(\frac{\dot{A}}{4\sigma\alpha_P}\right)^{1/4}, \tag{2.120}$$

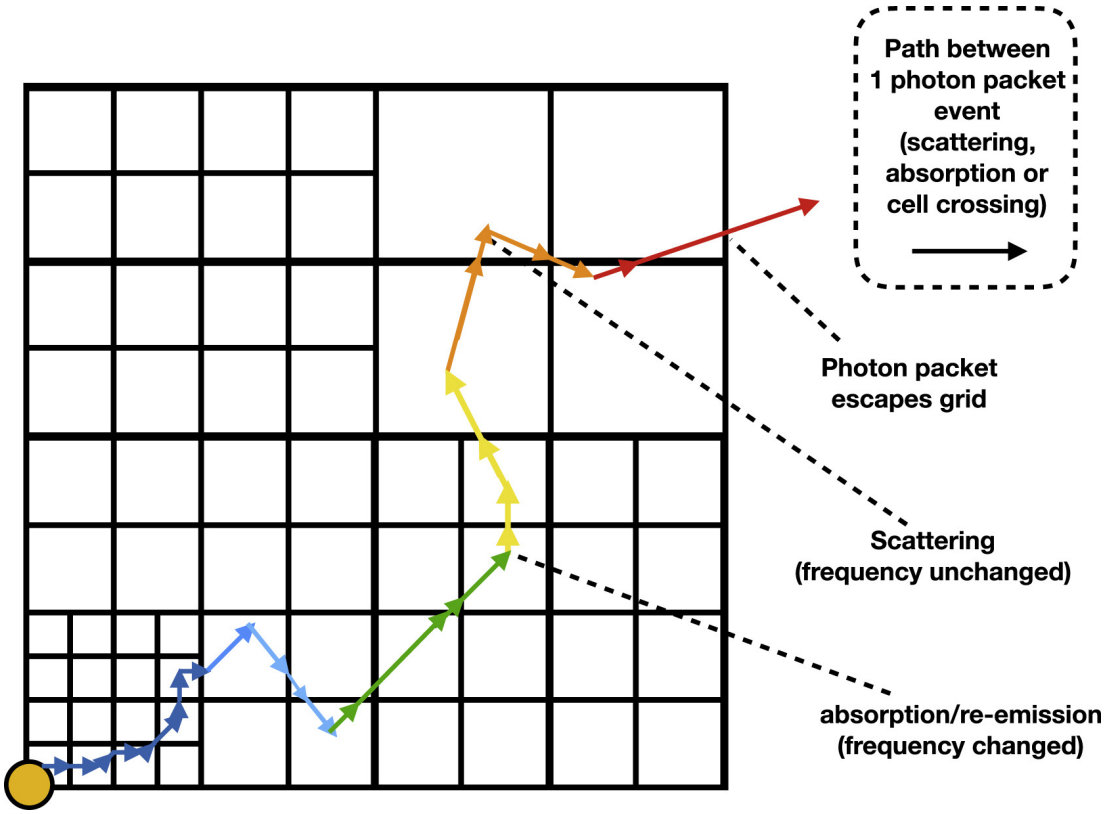

Figure 2.4 *Illustration of a photon packet's trajectory through an AMR grid during MCRT, highlighting the processes of absorption/re-emission and scattering. Image credit: Harries et al. (2019)*

where α_P is the Planck mean absorption coefficient and $\dot{A}$ is the absorption rate given by,

$$\dot{A} = \frac{\epsilon}{\Delta t}\frac{1}{V}\sum \alpha_\nu l. \tag{2.121}$$

The temperatures calculated at the end of each iteration are then used as a starting point for the next iteration, when another set of N photon packets are emitted from the star. The process is repeated until the temperature within each grid cell changes by an amount smaller than some user-defined tolerance between subsequent iterations. The final dust temperatures can then be used to generate continuum images of the disc or spectral energy distributions.

2.5 Concluding remarks

I have introduced some of the key equations, physical processes and numerical methods which will be used and referred back to throughout this thesis when modelling self-gravitating protoplanetary discs. In the following 4 chapters I will detail the research undertaken throughout the past 3 and a half years of my PhD. In Chapters 3 and 4 I focus on using SPH simulations to identify trends in the parameter space of star-disc systems which may provide favourable conditions for disc fragmentation. In Chapter 5 I use SPH and various 1D models to try and understand the formation histories of the protoplanets recently identified in the AB Aurigae disc. Finally, in Chapter 6 I use MCRT models to generate observational predictions of self-gravitating discs, focusing on the impact of dust-trapping in the discs' spiral arms.

Chapter 3

Fragmentation favoured in discs around higher mass stars

This chapter primarily contains work from Cadman et al. (2020a), with some references to work from a companion paper, Haworth et al. (2020). Whilst Cadman et al. (2020a) was led by myself, the project was undertaken in collaboration with various listed co-authors, hence some parts of the text may contain contributions from these co-authors.

Specifically, Section 3.7 was led by K. Rice.

3.1 Motivation

Early results from radial velocity surveys suggested that giant planets are found more frequently around higher-mass hosts (Johnson et al., 2007; Bowler et al., 2010), although Lloyd (2011) express some concerns regarding how accurately the mass of these host stars can be measured. These results stimulated large-scale searches for directly imaged exoplanet companions around high-mass hosts (primarily A stars, Janson et al., 2011; Vigan et al., 2012; Nielsen et al., 2013), even though intrinsically higher contrasts are needed to detect companions around these bright stars relative to solar analogues. Recently, considering the first 300 stars observed during the Gemini GPIES survey, Nielsen et al. (2019) found a higher frequency of wide-orbit ($R = 10-100\,\mathrm{AU}$) giant planets ($M = 5-13\,\mathrm{M_{Jup}}$) around higher mass stars ($M > 1.5\,\mathrm{M_\odot}$) than around lower mass stars ($M < 1.5\,\mathrm{M_\odot}$), while direct imaging surveys of low mass stars (M stars) have not yielded

any companion detections (Lannier et al., 2016).

As already discussed in Section 1.6 giant planet formation through GI will likely only operate in the outer disc, whilst core accretion struggles to explain the formation of giant planets on wide orbits. Hence the preference for finding more wide-orbit, giant planets around higher mass stars may suggest that discs around higher mass stars are favourable sites for planet formation through GI.

It has previously been shown that GI in discs around low-mass stars is quenched by a combination of viscous heating and stellar irradiation, making planet formation through fragmentation unlikely (Matzner & Levin, 2005). Kratter & Matzner (2006) found a critical disc outer radius of $R \approx 150\,\mathrm{AU}$, above which discs around massive stars may become prone to fragmentation. This critical radius is set by two competing factors; increased stellar irradiation with increasing stellar mass pushing the radius out, whilst the increased rate of accretion around the more massive stars favours fragmentation. Kratter & Lodato (2016) used the scaling of Q with disc-to-star mass ratio to suggest analytically that we may expect some scaling of instability with stellar mass. Recently this relation has been further explored by Haworth et al. (2020).

Haworth et al. (2020) demonstrated that low-mass stars are able to maintain discs with high disc-to-star mass ratios, with masses comparable to that of the central protostar, without becoming gravitationally unstable and fragmenting. The large mass reservoirs which these discs could in principle support may have important consequences for planet formation through core accretion, and may help to explain the origin of multi-planet systems around very low-mass stars, such as Trappist-1 (Gillon et al., 2017) whose formation would require extremely high planet formation efficiencies given the canonical disc-to-star mass ratio of $q = 0.1$.

The work we present here builds on the work from Haworth et al. (2020), but conversely aims to investigate how the critical disc-to-star mass ratio for fragmentation varies with increasing stellar mass around higher mass stars. To approach this, 1D disc models for various stellar masses have been used to calculate the effective viscous-α values (Shakura & Sunyaev, 1973; Lodato & Rice, 2004) for a range of disc radii and accretion rates. These models are used to estimate the critical disc-to-star mass ratios for fragmentation given a range of disc parameter space, which are then used to inform the setup parameters for a suite of 3D SPH simulations where we determine the critical disc-to-star mass

ratios more reliably.

This chapter is organised as follows. In Sections 3.2 and 3.3 we describe the setup of our 1D disc models and 3D SPH simulations respectively, and present the results of these in Sections 3.4 and 3.5. In Section 3.6 we analyse the Jeans masses in self-gravitating discs, allowing us to predict the planet masses we might expect to form through disc fragmentation. In Section 3.7 we discuss the timescales over which we might expect the conditions for fragmentation to be satisfied. Finally, in Sections 3.8 and 3.9 we summarise our results and discuss the implications for planet formation through disc fragmentation.

3.2 Methods - 1D disc models

To investigate how disc stability against fragmentation varies with stellar mass, we have implemented the 1D disc models first presented by Clarke (2009) and then further developed by Forgan & Rice (2011). Specifically, we use the formalism in which external irradiation is also included (Forgan & Rice, 2013a). We consider two cases; one in which irradiation leads to a constant background temperature of $T_{\rm irr} = 10\,{\rm K}$ and another in which the stellar irradiation is based on the MIST stellar models for 0.5 Myr stars (Dotter, 2016; Choi et al., 2016).

The four stellar masses considered in this analysis are $0.25\,{\rm M}_\odot$, $0.5\,{\rm M}_\odot$, $1.0\,{\rm M}_\odot$ and $2.0\,{\rm M}_\odot$. For host-star masses greater than $2\,{\rm M}_\odot$, these models become complicated as the outer disc becomes optically thick and dynamical heating may become important. We therefore choose not to model stellar masses greater than this. For each stellar mass we have generated a suite of 1D disc models and investigated the conditions necessary for fragmentation to occur, assuming that fragmentation is possible for $\alpha \gtrsim 0.1$ (Rice, Lodato & Armitage, 2005).

A self-gravitating disc is constructed by assuming that it settles into a state with a steady mass accretion rate given by (Pringle, 1981),

$$\dot{M} = \frac{3\pi\alpha c_{\rm s}^2 \Sigma}{\Omega}, \tag{3.1}$$

where α is the Shakura-Sunyaev viscous$-\alpha$. Note that this equation is the same as the one derived in Equation 2.82, where we have used Equations 2.86 and 2.58 in place of ν. We use Equation 1.28 to calculate α, where we use that $\beta_c = (u/\dot{u})\Omega$

where,

$$u = \frac{c_s^2}{\gamma(\gamma - 1)}, \tag{3.2}$$

and the cooling function is given by,

$$\dot{u} = \frac{\sigma_{\rm SB} T^4}{\tau + 1/\tau}, \tag{3.3}$$

where $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant, T is the disc temperature and τ is the optical depth. We can estimate the optical depth using $\tau = \Sigma\kappa(\rho, c_{\rm s})$, where $\Sigma = 2H\rho$ and κ is the opacity. Values of γ, T and κ are obtained from ρ and $c_{\rm s}$ using a look up table and the equation of state from Stamatellos et al. (2007). The scaling of this cooling function with optical depth as $\tau + 1/\tau$ allows us to account for both optically thin regimes, where the $1/\tau$ term will dominate, and optically thick regimes, where the τ term will dominate (Levin, 2003, 2007).

Assuming that the disc is gravitationally unstable at all radii, with $Q = 2$, and specifying a mass accretion rate (which is assumed to be constant at all radii) Equations 1.21, 1.28 and 3.1 allow us to calculate the values of three unknown quantities: α, Σ and $c_{\rm s}$.

A disc is constructed by solving these equations at successive radial intervals. By summing the mass within each radial bin we can determine the total disc mass for a given $\dot{M}$ and $R_{\rm out}$. In this way we can construct discs which span a wide range of parameter space in $\dot{M}$ and $R_{\rm out}$, where here we use values in the range $10^{-10} - 10^{-1}\,{\rm M}_\odot\,{\rm yr}^{-1}$ and $1 - 200\,{\rm AU}$ respectively.

In the case where irradiation is modelled using a constant background temperature, the disc is prevented from dropping below a floor temperature of $T_{\rm irr} = 10\,{\rm K}$. In the case of stellar irradiated discs we model the irradiation temperature as,

$$T_{\rm irr} = \left(\frac{L_*}{4\pi\sigma R^2}\right)^{1/4}, \tag{3.4}$$

where L_* is obtained from the MIST stellar evolution tracks at 0.5 Myr (Dotter, 2016; Choi et al., 2016). These tracks are plotted in Fig. 3.1 and the values of L_* used here for the cases of $0.25\,{\rm M}_\odot$, $0.5\,{\rm M}_\odot$, $1.0\,{\rm M}_\odot$ and $2.0\,{\rm M}_\odot$ stellar masses are $0.44\,{\rm L}_\odot$, $1.19\,{\rm L}_\odot$, $3.40\,{\rm L}_\odot$ and $10.11\,{\rm L}_\odot$ respectively.

This modelling of stellar irradiation assumes the disc to be optically thin and thus passively irradiated. In reality there will be significant self-shielding in the inner disc and the true disc heating will lie somewhere in between these two cases. An initial assessment on the impact of self-shielding finds the mid-plane dust radiative equilibrium temperature to be a factor $\approx 3-4$ smaller than that from Equation 3.4. We therefore might expect the true critical disc-to-star mass ratios to be closer to the predictions of the $T_{\rm irr} = 10\,{\rm K}$ discs than the stellar irradiated discs.

When calculating the resultant temperatures due to stellar irradiation in the outer regions of a $R_{\rm out} = 150\,{\rm AU}$ disc, we find that the systems with a $0.25\,{\rm M}_{\odot}$, $0.5\,{\rm M}_{\odot}$, $1\,{\rm M}_{\odot}$ and $2\,{\rm M}_{\odot}$ stellar host have temperatures 26.1K, 33.5 K, 43.6 K and 57.2 K respectively. These higher disc temperatures in the presence of stellar irradiation will further suppress fragmentation as it will both provide greater pressure support against gravitational collapse and also reduce the disc effective-α.

The 1D disc models presented here assume local angular momentum transport in which the disc viscosity can be represented by a local $\alpha-$parameter (e.g. Equation 1.28). This assumption may be violated in some cases where global effects become important. Forgan et al. (2011) found the local approximation to be valid up to disc-to-star mass ratios of $q \approx 0.5$, above which global effects become important and the effective viscosity is not well represented by this local parameterisation. We should therefore proceed with caution when interpreting the results of these models at high disc-to-star mass ratios. However, they do provide useful information that informs the 3D SPH simulations which follow.

3.3 Methods - SPH simulations

To extend the results from the 1D disc models we have produced a suite of 3D SPH simulations using the PHANTOM SPH code (Price et al., 2018). We use the polytropic cooling approximation introduced in Stamatellos et al. (2007) to model radiation transport in each disc, as outlined in Section 2.4.1.4.

The gas discs are represented by 500,000 SPH particles, allowing us to simulate a large number of discs spanning a wide range of parameter space. The stellar masses are the same as those from the 1D models: $M_* = 0.25\,{\rm M}_{\odot}$, $0.5\,{\rm M}_{\odot}$, $1.0\,{\rm M}_{\odot}$

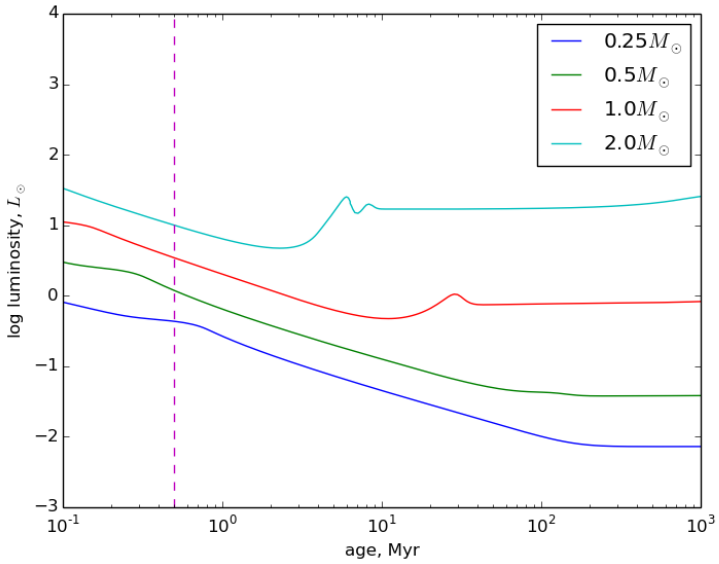

Figure 3.1 *MIST stellar evolution tracks (see Dotter, 2016; Choi et al., 2016). The 0.5Myr luminosities extracted from these plots have been used in these analyses.*

and 2.0 $M_\odot$. Each disc has an initial surface density profile of $\Sigma \propto R^{-1.5}$, and an initial sound speed profile of $c_s \propto R^{-0.5}$, with an inner disc radius of 1 AU in each case. Any gas particles falling within this inner radius will be accreted onto the central protostar, represented here as a point mass. These surface density and sound speed profiles were chosen to be consistent with those resulting from the 1D models. The steep surface density profile also avoids artificially inducing fragmentation by initially putting too much mass in the outer disc. In Haworth et al. (2020) shallower surface density and sound speed profiles have been used, with the authors finding consistent results to what we find here. Artificial viscosity is modelled using the standard $\alpha - \beta$ viscosity parameters, where we use $\alpha_{\rm SPH} = 0.1$ and $\beta_{\rm SPH} = 0.2$.

Again, we assume two cases of disc irradiation in line with the 1D models: one where irradiation leads to a constant background temperature of 10 K, and another where the disc is being passively irradiated by the star.

For the cases of 10 K and stellar irradiation, a total of 192 and 58 discs have been simulated respectively. The specific disc masses and radii were selected from inspection of the 1D model results, considering disc parameters which lie close to the $\alpha = 0.1$ contour.

Each disc has been allowed to evolve for 5 outer orbital periods, assuming that if

it has not fragmented by this point then it will not fragment in the future. Discs are considered to have not fragmented if they initially appear to form clumps, but these clumps are then rapidly destroyed by dynamical effects within the 5 orbital periods.

3.4 Results - 1D disc models

Throughout this subsection I will refer only to the contours in Figures 3.2 and 3.3, which represent the results of the 1D models. I will then go on to discuss the overplotted dots and crosses in Section 3.5, which represent the results of the SPH simulations.

In Figures 3.2 and 3.3, the blue contours show how the disc-to-star mass ratio, q, varies with accretion rate as a function of disc outer radius. For example, in Figure 3.2 for a $0.25\,\mathrm{M}_\odot$ stellar host, a disc with an accretion rate $\dot{M} = 10^{-7}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ and a radius $R_{\mathrm{out}} = 80\,\mathrm{AU}$ will have a disc-to-star mass ratio, $q = 0.520$.

The black contours show the Shakura-Sunyaev effective viscous-α values from Equation 1.28. We show contours for $\alpha = 0.01$ and $\alpha = 0.1$. As discussed in section 1.6.2, the canonical fragmentation boundary is typically taken to be $\alpha = 0.06$. There is, however, some uncertainty in this exact value, partly due to convergence issues in the simulations (Meru & Bate, 2011), partly due to possible stochasticity (Paardekooper, 2012), and partly because there is some evidence for an alternative mode of fragmentation (Young & Clarke, 2015). It seems likely, though, that fragmentation will occur somewhere in the region between the $\alpha = 0.01$ and $\alpha = 0.1$ contours. Figures 3.2 and 3.3 illustrate that this will require discs with masses that are a significant fraction of the mass of the central protostar.

3.4.1 $T_{\mathrm{irr}} = 10\,\mathrm{K}$

Figure 3.2 shows the scenario in which we assume that background irradiation prevents the disc temperature from dropping below $T = 10\,\mathrm{K}$. It shows that as we increase the host star mass from $0.25\,\mathrm{M}_\odot$ to $2\,\mathrm{M}_\odot$ the critical mass ratio for the discs to become unstable against fragmentation generally decreases. If we consider the $\alpha = 0.1$ contour in Figure 3.2, for a $0.25\,\mathrm{M}_\odot$ stellar host the disc-to-

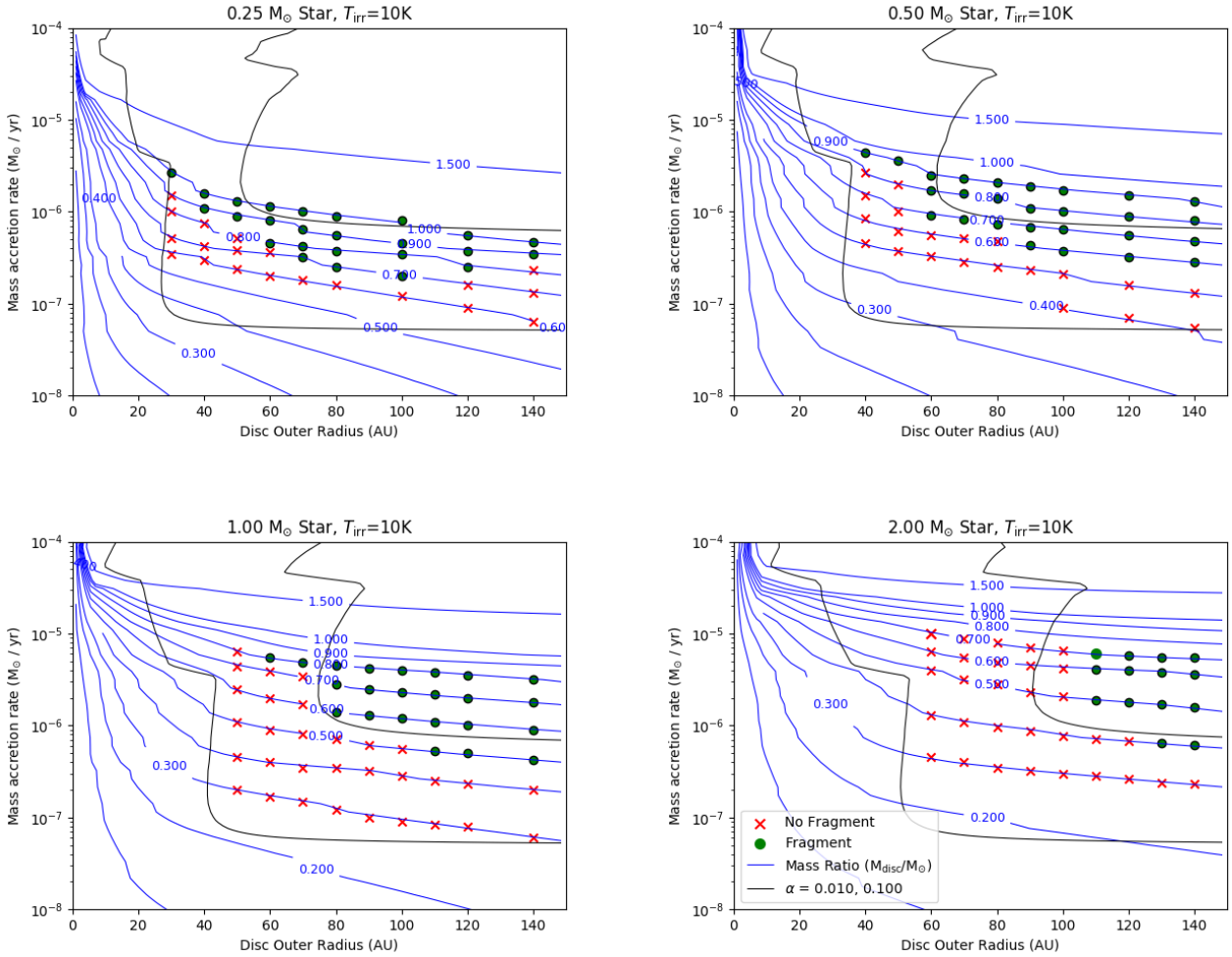

Figure 3.2 *Results of the 1D models (contours) and 3D SPH simulations (dots and crosses) for the case of 10 K irradiated discs. The 2D contour plots show how the disc-to-star mass ratio (blue contours) varies as a function of accretion rate and disc outer radius for the cases of* $0.25\,M_{\odot}$ *(top left),* $0.5\,M_{\odot}$ *(top right),* $1.0\,M_{\odot}$ *(bottom left) and* $2.0\,M_{\odot}$ *(bottom right) host star masses. The effective Shakura-Sunyaev viscous-*α *values from the 1D models are shown as black contours. The results of the 3D SPH simulations are shown by the dots and crosses, representing fragmenting and non-fragmenting discs respectively.*

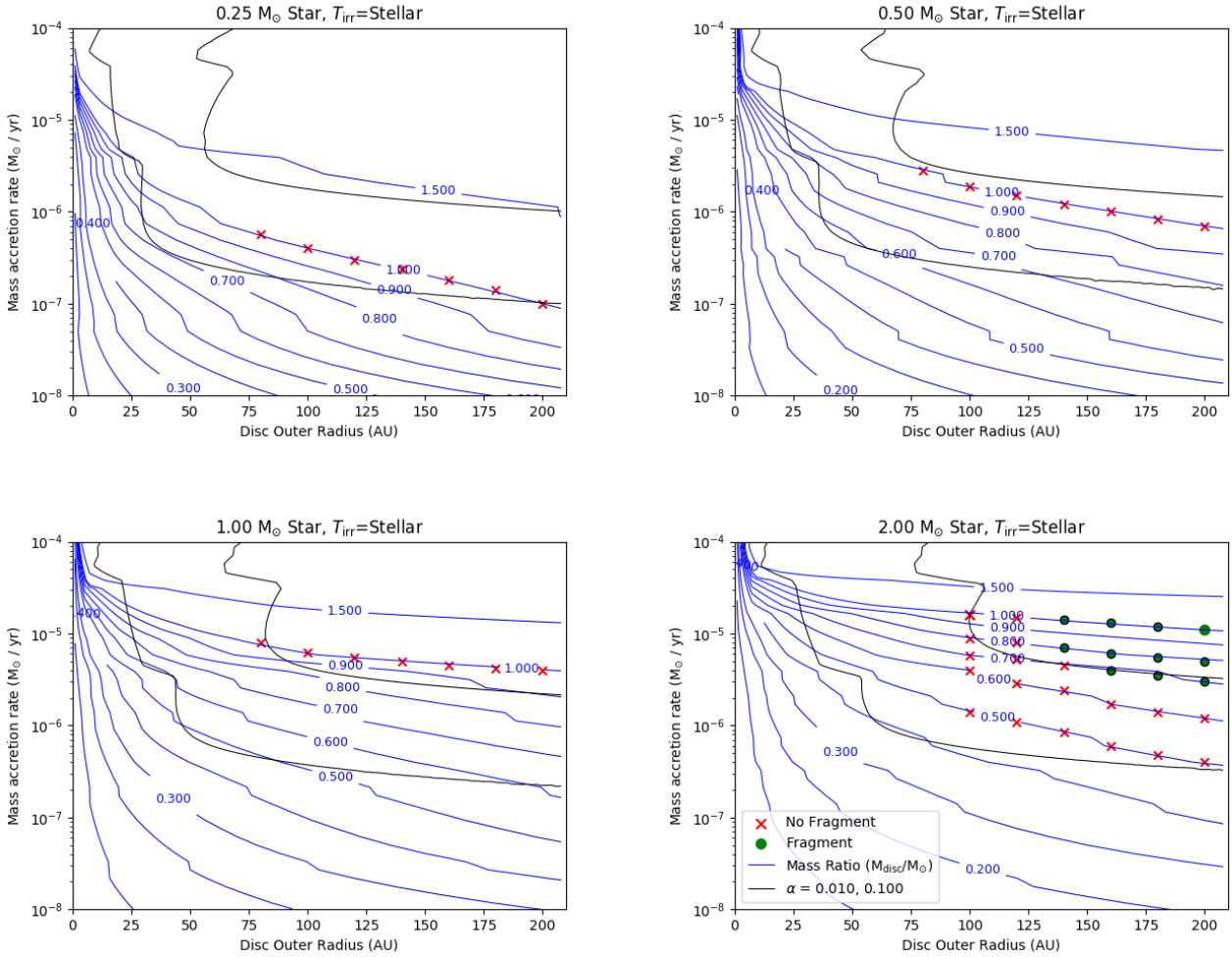

Figure 3.3 *Results of the 1D models (contours) and 3D SPH simulations (dots and crosses) for the case of 0.5 Myr MIST Stellar irradiated discs. The 2D contour plots show how the disc-to-star mass ratio (blue contours) varies as a function of accretion rate and disc outer radius for the cases of* $0.25\,M_{\odot}$ *(top left),* $0.5\,M_{\odot}$ *(top right),* $1.0\,M_{\odot}$ *(bottom left) and* $2.0\,M_{\odot}$ *(bottom right) host star masses. The effective Shakura-Sunyaev viscous-*α *values from the 1D models are shown as black contours. The results of the 3D SPH simulations are shown by the dots and crosses, representing fragmenting and non-fragmenting discs respectively.*

star mass ratio needs to exceed $q = 1$ before the disc's viscous-α values exceed $\alpha = 0.1$. We would therefore expect these discs to avoid fragmenting even for very large disc-to-star mass ratios. As stellar mass is increased to $2\,M_\odot$, the disc's viscous$-\alpha$ exceeds $\alpha = 0.1$ for mass ratios of around $q = 0.4 - 0.5$. The minimum radius for fragmentation also tends to shift outwards with increasing stellar mass. Fragmentation is only expected in discs larger than $R \approx 90\,\mathrm{AU}$ in the case of a $2\,M_\odot$ stellar host, compared to $R \approx 50\,\mathrm{AU}$ in the case of a $0.25\,M_\odot$ stellar host.

3.4.2 T_{irr} = Stellar

When considering the case of stellar irradiation, shown in Figure 3.3, the critical mass ratios are now shifted to even higher masses compared to when $T_{\mathrm{irr}} = 10\,\mathrm{K}$. This is due to the now higher disc temperatures suppressing GI. For a $0.25\,M_\odot$ stellar host we now require $q \gtrsim 1.4$ before the disc's viscous-α values exceed $\alpha = 0.1$. Increasing the stellar mass to $2\,M_\odot$ reduces the required disc-to-star mass ratio to $q \gtrsim 0.7$. The minimum radii at which fragmentation is likely to occur has also been pushed outward compared to the 10 K irradiated discs. Fragmentation will now only occur in discs larger than $R \approx 100\,\mathrm{AU}$ for a $2\,M_\odot$ stellar host, and $R \approx 60\,\mathrm{AU}$ for a $0.25\,M_\odot$ stellar host.

These models suggest that fragmentation is favourable in discs around higher mass stars compared to those around lower mass stars, as the disc-to-star mass ratios required for fragmentation may be lower. When including the effects of stellar irradiation we find that discs become less prone to fragmentation, as we now require far higher disc-to-star mass ratios before the discs' viscous$-\alpha$ values exceed $\alpha = 0.1$. We note again that above $q \approx 0.5$, global effects may become important which are not accounted for in these 1D models. However, we do not expect this to invalidate the general trends demonstrated by these results.

3.5 Results - SPH simulations

In Figures 3.2 and 3.3, we also show the results of the 3D SPH simulations. These are represented by the markers over-plotted on the mass-ratio contours. Each marker represents an individual simulation, which has been set up as described in Section 3.3. Red crosses show discs that have not fragmented after 5 outer orbital periods and green circles show discs in which a bound fragment has formed.

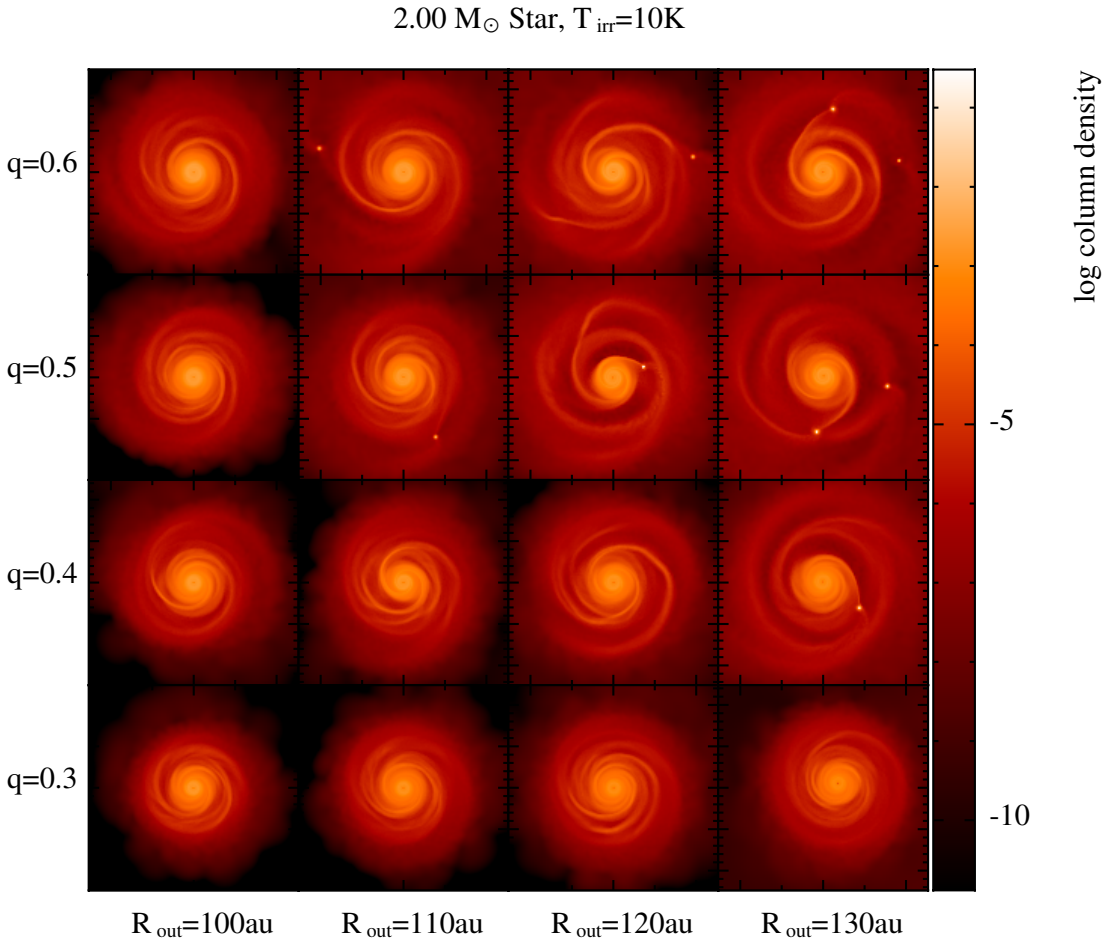

Figure 3.4 *3D SPH results demonstrating how discs become more gravitationally unstable and prone to fragmentation as we increase the disc-to-star mass ratio and the disc outer radius. The discs shown here are for a* $2\,M_{\odot}$ *host star and* $T_{\rm irr} = 10\,K$*. Each disc has been allowed to evolve for 5 outer orbital periods, with only the largest and most massive discs having formed bound fragments.*

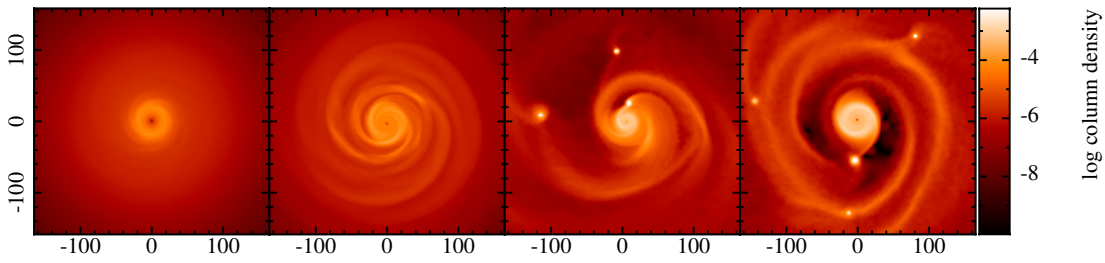

Figure 3.5 *3D SPH results showing how the final states of the discs vary with stellar mass in the case of* $T_{\rm irr} = 10\,K$*. The discs shown have mass ratios,* $q = 0.5$*, and outer radii,* $R_{\rm out} = 140\,AU$*, with stellar masses, from left to right, of* $0.25\,M_{\odot}$*,* $0.5\,M_{\odot}$*,* $1.0\,M_{\odot}$ *and* $2.0\,M_{\odot}$*.*

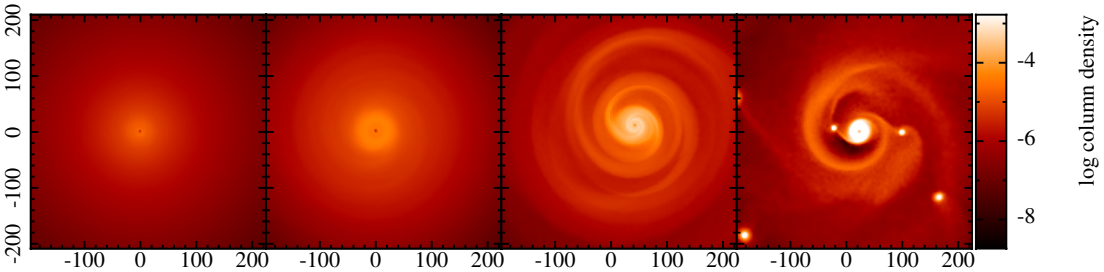

Figure 3.6 *Results of the 3D SPH simulations in the case of 0.5 Myr MIST Stellar irradiated discs. The discs shown have mass ratios, $q = 1.0$, and outer radii, $R_{\rm out} = 200\,AU$, with stellar masses, from left to right, of $M_* = 0.25\,M_\odot$, $0.5\,M_\odot$, $1.0\,M_\odot$ and $2.0\,M_\odot$.*

Example plots of the final states of these simulated discs are displayed in Figure. 3.4. The discs shown are for a $2\,M_\odot$ host star when $T_{\rm irr} = 10\,K$, and demonstrate how discs become increasingly gravitationally unstable and prone to fragmentation as we increase the disc's outer radius and mass. Bound fragments have clearly formed in the largest and most massive discs, whilst the smaller and less massive discs display spiral arm structure only.

As we mentioned previously, our 1D models assume local angular momentum transport which may not be valid at high disc-to-star mass ratios. The effect of this can be clearly seen in Figure 3.2 when comparing the 1D predictions to the 3D results at the highest disc-to-star mass ratios ($q \geq 0.5$). In the $0.25\,M_\odot$ case for example, the SPH simulations find that fragmentation can occur for much lower q values than initially predicted by the 1D models. When comparing the 1D predictions to the 3D results for slightly lower mass ratio discs, for example in the calculations involving the $2.0\,M_\odot$ stars, we find the results to be far more consistent with the 1D models as they are now more reliable. Despite this, the SPH results shown in Figures 3.2 and 3.3 display the same general trend as suggested by the 1D disc models; the critical disc-to-star mass ratio for fragmentation generally decreases with increasing stellar mass, and the critical radius steadily shifts outward.

In Figure 3.2, for a $0.25\,M_\odot$ host star with $T_{\rm irr} = 10\,K$, discs are able to fragment for mass ratios $q \geq 0.7$. This is lower than suggested by the $\alpha = 0.1$ contour but still broadly consistent with the 1D models. For the discs around the $2\,M_\odot$ host star, we find that fragmentation occurs for mass ratios $q \geq 0.4$. Discs as small as $R_{\rm out} = 30\,AU$ are able to fragment around a $0.25\,M_\odot$ host star, with this value increasing to $R_{\rm out} = 110\,AU$ for a $2\,M_\odot$ stellar host.

In Figure 3.3, when considering stellar irradiated discs, only the $2\,M_\odot$ stellar hosts produce fragments after 5 orbital periods. Fragmentation can occur for $q \geq 0.7$ in these systems. All other stellar masses show no fragmentation for discs with masses up to $q = 1.0$, with these being the highest mass discs modelled in our SPH simulations. We have chosen to not model discs with mass ratios greater than this as it is unclear whether these would exist as disc-star systems at all, or whether the system would instead be deeply embedded in an envelope. The critical radii at which we expect discs to fragment has again shifted outward with respect to the 10 K irradiated discs, with discs around a $2\,M_\odot$ star only being able to fragment when $R_{\text{out}} \gtrsim 140\,\text{AU}$.

Figures 3.5 and 3.6 further illustrate the effects of increasing stellar mass on disc instability. It can be seen that as we increase the star mass for discs with constant R_{out} and q, they become increasingly gravitationally unstable. In Figure 3.5 the discs with a $1\,M_\odot$ and a $2\,M_\odot$ host star have formed bound fragments, whilst in the disc with a $0.5\,M_\odot$ host star we observe spiral structure, and for the $0.25\,M_\odot$ host star we observe almost no spiral structure at all. A similar trend can be seen when considering stellar irradiated discs in Figure 3.6, with only the $2\,M_\odot$ host star case forming bound fragments here.

3.6 Jeans mass in a spiral wave perturbation

When a region of a gravitationally unstable disc fragments, it will collapse to form bound clumps of mass comparable to the local Jeans mass. Hence calculation of the Jeans mass allows us to place constraints on the type of objects which may be produced through disc fragmentation. To derive an analytical expression for the Jeans mass in a spiral density perturbation of a self-gravitating disc, we begin with the equation derived in Section 1.2,

$$M_J = \frac{4}{3}\pi R_J^3 \rho_{\text{pert}} = \frac{4}{3}\left(\frac{3}{32}\right)^{3/2} \pi^{5/2} \frac{c_s^3}{G^{3/2}\rho_{\text{pert}}^{1/2}} \tag{3.5}$$

Where ρ_{pert} represents the density of the spiral perturbation. The scale height, H, and local surface density of the perturbation, Σ_{pert} are related to ρ_{pert} as,

$$\rho_{\text{pert}} = \Sigma_{\text{pert}}/2H, \tag{3.6}$$

where,

$$\Sigma_{\rm pert} = \Sigma\Big(1 + \frac{\Delta\Sigma}{\Sigma}\Big). \tag{3.7}$$

Here, $\frac{\Delta\Sigma}{\Sigma}$ represents the fractional amplitude of the spiral wave perturbation. Rearranging Equation 1.21 in terms of G and Σ gives,

$$(G\Sigma)^{1/2} = \Big(\frac{c_s\Omega}{\pi Q}\Big)^{1/2}. \tag{3.8}$$

Substituting Equations 3.7, 3.8 and $H = c_s/\Omega$ into Equation 3.5 gives,

$$M_J = \frac{4\sqrt{2}}{3}\Big(\frac{3}{32}\Big)^{3/2}\frac{\pi^3}{G}\frac{Q^{1/2}c_s^2 H}{(1+\frac{\Delta\Sigma}{\Sigma})^{1/2}} \tag{3.9}$$

In the presence of external irradiation Rice et al. (2011) showed that,

$$\langle\frac{\Sigma_{\rm RMS}}{\Sigma}\rangle = 4.47\sqrt{\alpha}, \tag{3.10}$$

where α is the self-gravitating viscous-α. Substituting this into Equation 3.9 gives the expression for the Jeans mass in an irradiated self-gravitating disc,

$$M_J = \frac{\sqrt{3}}{32G}\frac{\pi^3 Q^{1/2}{c_{\rm s}}^2 H}{(1+4.47\sqrt{\alpha})^{1/2}}. \tag{3.11}$$

Note that this equation differs slightly from the equation derived previously in Forgan & Rice (2013a). We now have a different prefactor and the $1 + 4.47\sqrt{\alpha}$ term is now square-rooted.

Figure 3.7 shows the calculated Jeans masses from the 1D disc models for a $2\,{\rm M}_\odot$ stellar host, considering both 10 K and stellar irradiation. We consider here the case of a $2\,{\rm M}_\odot$ host star as we are primarily concerned with fragmentation around the more massive stellar hosts. The Jeans mass for each value of $\dot{M}$ and $R_{\rm out}$ has been calculated using Equation 3.11 and plotted as the green contours in Figure 3.7.

The minimum Jeans masses in the 10 K and stellar irradiated discs are $1.10\,{\rm M}_{\rm Jup}$ and $6.18\,{\rm M}_{\rm Jup}$ respectively, assuming fragmentation is only possible above the

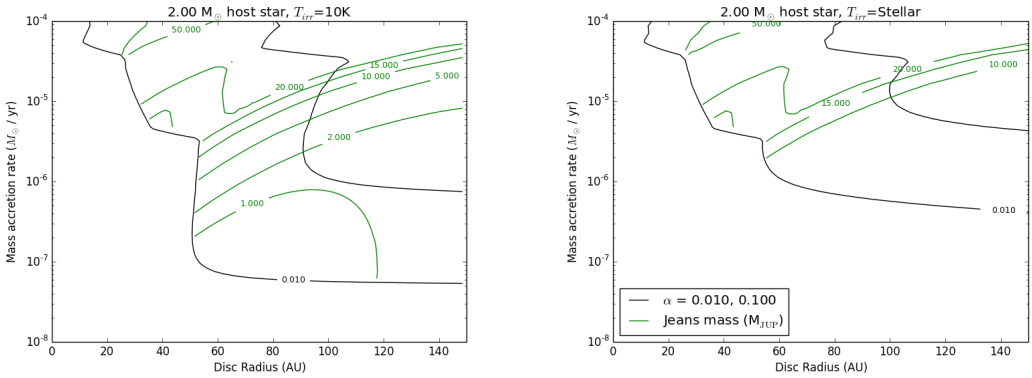

Figure 3.7 *Predicted Jeans masses from Equation 3.11 in discs around a* $2\,M_{\odot}$ *host star for the cases of* $T_{\rm irr} = 10\,K$ *(left) and stellar irradiation (right).*

$\alpha = 0.1$ contour. The tendency for the Jeans mass to increase with the level of irradiation is a consequence of higher disc temperatures reducing the effective-α thus causing discs to be more massive for a given $\dot{M}$ and $R_{\rm out}$. Equally, the higher temperatures provide greater pressure support against gravitational collapse, as previously discussed in Forgan & Rice (2013a).

The analysis in Forgan & Rice (2013a) considered the case of a $1\,{\rm M}_{\odot}$ stellar host, and the values found here remain similar to those found previously despite the changes made to Equation 3.11. For the case of a $1\,{\rm M}_{\odot}$ stellar host, we find minimum Jeans masses of $1.10\,{\rm M}_{\rm Jup}$ and $4.60\,{\rm M}_{\rm Jup}$ when using $10\,{\rm K}$ and stellar irradiation respectively, compared to values of $4.1\,{\rm M}_{\rm Jup}$ and $11.2\,{\rm M}_{\rm Jup}$ found in Forgan & Rice (2013a) previously.

3.7 Timescale for fragmentation

Our results indicate that fragmentation is preferred in discs around higher mass stars, and could potentially be completely suppressed in very-low-mass stars if the level of irradiation is sufficient. However, another factor to consider is the timescale over which a disc may sustain the conditions that are suitable for fragmentation. This is not possible to assess using the results from the 1D model and the 3D SPH simulations, since the 1D models are not time-dependent and the 3D SPH simulations are simply sampling regions of parameter space.

To consider this, we use the time-dependent models presented in Rice & Armitage

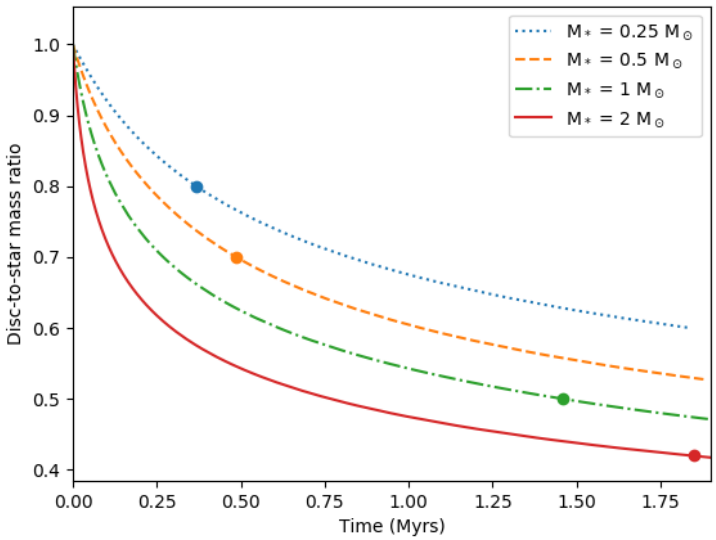

Figure 3.8 *Figure showing the evolution of disc-to-star mass ratio, q, in discs in which the gravitational instability is the dominant angular momentum transport mechanism, for host star masses of $M_* = 0.25, 0.5, 1$ and 2 $M_\odot$. The markers show the disc-to-star mass ratios above which disc fragmentation is possible, based on the results presented in Figure 3.2.*

(2009), which assume that angular momentum transport is predominantly driven by GI. Given that we don't actually know what the initial conditions will be, we assume that all discs start with an outer radius of $R_{\rm out} = 100\,{\rm AU}$ and with a disc-to-star mass ratio of $q = 1$. We also only consider the case where $T_{\rm irr} = 10\,{\rm K}$.

Figure 3.8 shows the time evolution of the disc-to-star mass ratio for the same host star masses as considered before ($M_* = 0.25, 0.5, 1$ and 2 $\rm M_\odot$). The markers show, for each host star, the disc-to-star mass ratio above which fragmentation is possible, based on the results presented in Figure 3.2. What Figure 3.8 illustrates is that, in conjunction with the required disc-to-star mass ratio decreasing with increasing stellar mass, the timescale over which fragmentation could occur also increases.

Of course, Figure 3.8 does assume that sufficiently massive discs can indeed exist, but - if they can - the conditions for fragmentation would only persist around a $0.25\,{\rm M_\odot}$ host star for a few $100\,{\rm kyr}$. Around a 2 $\rm M_\odot$ host star, however, the timescale for fragmentation could be much longer, potentially a Myr, or longer. However, this does assume that GI is the dominant mass transport mechanism, which may not be the case once the disc mass, and mass accretion rate, have

become low enough for other mechanisms to become more important (Rice et al., 2010).

3.8 Discussion

The results presented here illustrate that disc fragmentation is potentially favoured around higher-mass stars. If we consider the case where $T_{\rm irr} = 10\,{\rm K}$, and assume that the fragmentation boundary is at $\alpha = 0.1$, fragmentation requires a disc-to-star mass ratio of close to unity for a $0.25\,{\rm M}_{\odot}$ host star, but requires $q \gtrsim 0.4$ around a $2\,{\rm M}_{\odot}$ host. If we then consider stellar irradiation (e.g. Figure 3.3), fragmentation around a $0.25\,{\rm M}_{\odot}$ host star would then require disc masses that exceed the mass of the central protostar, while fragmentation around a $2\,{\rm M}_{\odot}$ host could still occur for mass ratios of $q \gtrsim 0.6$. This might suggest that stellar irradiation could completely suppress fragmentation around lower-mass host stars.

However, the simple modelling of stellar irradiation used in these models does not account for these being young, massive discs and there likely being a large amount of material in the inner disc regions. We therefore neglect factors such as self-shielding by material in the inner disc that could lead to stellar irradiation having less of an impact at large radii than we've assumed here. We expect that the true heating to be somewhere in between the two irradiation cases we've considered, (possibly being closer to the $T_{\rm irr} = 10\,{\rm K}$ as already mentioned in section 3.2) and that the critical mass ratio where fragmentation can occur is probably somewhere within the range we've presented. We don't expect, however, that this will influence the trend that fragmentation is preferred around higher-mass host stars.

In general, discs with mass ratios of order unity, or above, are probably unrealistic. For Class II sources we expect the disc mass to be small compared to the stellar mass, usually no more than 10%. Higher mass-ratio systems would likely be in the Class I phase whilst there is still a large amount of material in the envelope. For even higher mass ratio systems, with q approaching unity, we would expect them to still be in the Class 0 phase where the source is deeply embedded. In this phase it is uncertain if there would be a star-disc system at all, or if instead there would be a massive envelope or torus. Additionally, even if such a system could exist, it would probably evolve very rapidly. It's, therefore, unclear if there

would be sufficient time for fragmentation to actually occur in a disc with $q > 1$.

A full discussion for the implications of lower mass stars being capable of hosting high-mass-ratio discs before becoming susceptible to gravitational instabilities can be found in Haworth et al. (2020). The key points to note are that the results suggest these systems may potentially have very large mass reservoirs available to them for planet formation through core accretion, thus loosening the constraint that any formation scenario (e.g. for the Trappist-1 system, Gillon et al. 2017) must involve highly efficient dust growth. They also find that the high mass ratio discs ($q \gtrsim 0.3$) required from photoevaporation models of the formation of the Trappist-1 system (Haworth et al., 2018) to be entirely plausible, with our models also finding discs to be gravitationally stable with masses greater than this.

It is intriguing that Morales et al. (2019) recently discovered a $0.46\,\mathrm{M_{Jup}}$ planet orbiting a very low mass, $0.12\,\mathrm{M_\odot}$, M dwarf on a 204 day period, with the authors proposing GI as the likely formation scenario. The results presented here suggest that only very massive discs around these very low mass stars may be permitted to be gravitationally unstable, thus indicating that such massive discs may indeed exist. We also require that these discs be optically thick to stellar irradiation, which would likely be the case for such a massive disc.

Several direct imaging surveys for companions around $M > 1.5\,\mathrm{M_\odot}$ stars have tentatively pointed to a higher fraction of exoplanet and brown dwarf companions to higher mass stars relative to solar analogues or very-low-mass stars (Janson et al., 2011; Nielsen et al., 2013; Vigan et al., 2012). Recently, considering the first 300 stars observed during the Gemini GPIES survey, Nielsen et al. (2019) demonstrates this more conclusively, finding a higher frequency of wide-orbit ($R = 10 - 100\,\mathrm{AU}$) giant planets ($M = 5 - 13\,\mathrm{M_{Jup}}$) around higher mass stars ($M > 1.5\,\mathrm{M_\odot}$) than around lower mass stars ($M < 1.5\,\mathrm{M_\odot}$). Nielsen et al. (2019) find an occurrence rate for wide-orbit ($10 - 100\,\mathrm{AU}$) giant planets ($M = 5 - 13\,\mathrm{M_{Jup}}$) of $9^{+5}_{-4}\%$ for their high mass stellar sample, vs. a brown dwarf occurrence rate ($M = 13 - 80\,\mathrm{M_{Jup}}$, $10 - 100\,\mathrm{AU}$) of $0.8^{+0.8}_{-0.5}\%$ around all survey stars. The mass divisions adopted in (Nielsen et al., 2019) do not straightforwardly map to a specific formation mechanism – the brown dwarfs they detect could likely have formed via gravitational instability, whereas some of the planets in their cohort (e.g. 51 Eri b, for instance) are likely lower than the Jeans masses we have calculated here, and thus not as likely to be disc instability objects. However, these results imply that the total companion frequency ($M = 5 - 80\,\mathrm{M_{Jup}}$, $10 - 100\,\mathrm{AU}$) must be higher for their high mass vs.

low mass stellar sample, qualitatively consistent with the work presented here.

Although our analysis does indicate that disc fragmentation is more likely around higher-mass stars, it also suggests that it will probably occur at radii $\gtrsim 100\,\mathrm{AU}$, with this critical radius moving outward with increased levels of irradiation. We might also expect the fragments to have initial masses above $5\,\mathrm{M_{Jup}}$ and to undergo further growth. However, since we expect GI to act when the disc is young ($< 0.1\,\mathrm{Myr}$) and massive, we would expect these fragments to undergo significant inward radial migration and potentially tidal downsizing after they form (e.g. see Nayakshin, 2010a; Forgan & Rice, 2013a).

3.9 Conclusions

In this chapter we have used a set of 1D disc models followed by a suite of 3D SPH simulations to investigate how the conditions necessary for gravitational instability in protoplanetary discs vary with host star mass. In these models we have varied the disc masses and radii, focusing on determining the critical disc-to-star mass ratio at which fragmentation is able to occur for stellar masses $M_* = 0.25\,\mathrm{M_\odot}$, $0.5\,\mathrm{M_\odot}$, $1\,\mathrm{M_\odot}$ and $2\,\mathrm{M_\odot}$. We have run models for both $T_{\mathrm{irr}} = 10\,\mathrm{K}$ and for stellar irradiation, with the true disc irradiation likely lying somewhere in between these two cases.

The primary conclusions drawn from this work are that,

1. Discs become more susceptible to GI as we increase the host star mass, with discs around higher mass stars being prone to fragmentation which will tend to produce wide-orbit giant planets and brown dwarfs.

2. Discs around lower mass stars ($M \leq 1.0\,\mathrm{M_\odot}$) are able to host very high mass-ratio discs whilst still remaining gravitationally stable. When considering discs which are passively irradiated by their stellar host we find fragmentation to be completely suppressed in discs up to mass ratios of order unity. This may have important implications for core accretion, since it may be possible for these discs to have larger mass reservoirs available for planet formation than previously thought.

3. Discs around higher-mass stars $M \geq 2\,\mathrm{M_\odot}$ are more susceptible to GI and fragmentation. For a $2\,\mathrm{M_\odot}$ host star, we find that discs may fragment for

mass ratios $q \geq 0.4$ and $q \geq 0.7$ in the cases of $T_{\rm irr} = 10\,{\rm K}$ and stellar irradiated discs respectively. We find that fragmentation will only likely occur at radii $\gtrsim 100\,{\rm AU}$, with this critical radius increasing with increased irradiation and with increasing host star mass. Fragment masses are found to be strongly dependent on disc irradiation with hotter discs producing more massive planets, as inferred through analytical calculations of their Jeans masses. Fragmentation in discs around $2\,{\rm M}_{\odot}$ stars will produce objects of masses $\geq 1.10\,{\rm M}_{\rm Jup}$ and $\geq 6.18\,{\rm M}_{\rm Jup}$ in discs with $T_{\rm irr} = 10\,{\rm K}$ and stellar irradiation respectively, thus preferentially producing wide-orbit, giant planets and brown dwarfs.

4. Discs around $2\,{\rm M}_{\odot}$ stars are able to sustain the conditions necessary for fragmentation for far longer timescales than discs around lower mass stars are. This is due to these discs becoming unstable against fragmentation for lower disc-to-star mass ratios, thus the conditions necessary for discs to be unstable against fragmentation will be satisfied for longer.

Chapter 4

Binary companions triggering fragmentation in self-gravitating discs

This chapter contains work from Cadman et al. (2022). Whilst Cadman et al. (2022) was led by myself, the project was undertaken in collaboration with various listed co-authors, hence some parts of the text may contain contributions from these co-authors.

Specifically, Section 4.4.3 was led by C. Fontanive.

4.1 Motivation

It is generally accepted that if GI is a viable mode of planet formation it will likely only operate in the outer parts of extended protoplanetary discs (Stamatellos & Whitworth, 2009; Vorobyov & Basu, 2010), potentially explaining the origin of some directly-imaged, wide-orbit planetary-mass and brown dwarf companions (Nero & Bjorkman, 2009; Kratter et al., 2010; Cadman et al., 2021; Humphries et al., 2021). It has also been suggested that objects that form via GI on wide orbits could migrate inwards rapidly (Baruteau et al., 2011) and potentially undergo tidal stripping (Nayakshin, 2010a; Boley et al., 2010) to produce close-in planets with a wide range of masses (Nayakshin & Fletcher, 2015). However, hydrodynamics simulations of such systems show that these objects either stay

on wide orbits or are destroyed during the migration process (Hall et al., 2017). Population synthesis models also suggest that such an outcome is relatively rare, and that most objects that form via GI will remain on wide-orbits as giant planets, or brown dwarfs (Forgan & Rice, 2013b; Forgan et al., 2018).

Such objects could, though, still be scattered onto highly eccentric orbits that can then tidally circularise onto close-in orbits (Rice et al., 2015). This will tend to form gas giant planets, or brown dwarfs, with very close-in circular orbits (with orbital properties similar to those of "hot" Jupiters) or eccentric orbits that are still undergoing tidal circularisation.

Given that the scatterer is likely to be a companion to the host star, this motivated a search for companions to systems with close-in ($< 1\,\mathrm{AU}$) massive planets or brown dwarfs ($M > 7\,\mathrm{M_{Jup}}$) (Fontanive et al., 2019). The results of this search did indeed indicate a binary fraction twice as high as for field stars on projected separations between 20–10,000 AU. However, only about half of these systems were consistent with high eccentricity migration through secular interactions with the outer stellar companion, the others being on orbits where the tidal circularisation timescale was far too long to explain their origin (Fontanive et al., 2019).

Nonetheless, even if the close-in objects were not scattered onto their current orbits, the high binary fraction for these systems suggests that the existence of a companion may still influence their formation. There are also indications that some of these objects may have formed via GI. The sample of stars studied in Fontanive et al. (2019), hosting close-in companions with masses between 7–60 $\mathrm{M_{Jup}}$, has a mean metallicity of $\langle[\mathrm{Fe/H}]\rangle = -0.12$, consistent with the mean field metallicity (Moe et al., 2019). This is substantially lower than the mean metallicity for hosts to genuine hot Jupiters (0.2–4 $\mathrm{M_{Jup}}$) of $\langle[\mathrm{Fe/H}]\rangle = 0.23$ (Santos et al., 2004; Fischer & Valenti, 2005), which also do not show the same excess in multiplicity frequency (Ngo et al., 2016; Moe & Kratter, 2021).

This lower-mass planetary population is thought to have formed via the alternative scenario for planet formation, core accretion (Pollack et al., 1996). This formation mechanism shows a strong metallicity dependence in the formation of giant planets with masses above a few Jupiter masses (Mordasini et al., 2012; Jenkins et al., 2017). In contrast, the GI formation process has no metallicity dependence (Meru & Bate, 2010), and preferentially forms massive planets or brown dwarfs (Kratter et al., 2010; Forgan & Rice, 2011), with a transition at

around ∼4–10 M_{Jup} between the two mechanisms (Schlaufman, 2018).

This suggests that only the most massive planetary and brown dwarf companions, likely forming via GI, are effected by wide-orbit stellar binarity. Fontanive & Bardalez Gagliuffi (2021) recently confirmed this idea, finding that close-in exoplanets and substellar companions with masses of several Jupiter masses and above are almost exclusively observed in binary star systems with separations of a few hundred AU or less. In contrast, sub-Jovian and wide giant planets are less frequently seen in multiple-star systems, mostly observed in binaries with wider separations, and show similar planet properties when compared to the population of planets orbiting single and binary stars (Fontanive & Bardalez Gagliuffi, 2021). We therefore investigate how the presence of a companion at a few hundred AU can influence the likelihood of a disc undergoing fragmentation and forming such high-mass planetary systems.

There is little agreement in the literature as to whether binary companions or stellar flyby events can trigger fragmentation in a disc which would be marginally stable in isolation. Early work considering isothermal discs suggested that encounters during a flyby event could trigger fragmentation (Boffin et al., 1998; Watkins et al., 1998a,b). Boss (2006) also found that a binary star will act to promote fragmentation, as the spiral arms driven by the companion will typically go on to form self-gravitating clumps. Other authors, however, found that tidal heating during the binary orbit generally acts to stabilise the disc against fragmentation (Nelson, 2000; Mayer et al., 2005; Lodato et al., 2007; Forgan & Rice, 2009). Whilst none of their simulations resulted in fragmentation, and the majority of their results suggest that the effect of encounters is to prohibit fragmentation, Forgan & Rice (2009) find that, for some orbital parameters, their discs become more unstable over a larger range of radii, suggesting that there may be some region of parameter space which is favourable to fragmentation. It has also been shown that once a fragment forms in a GI disc, further fragmentation may be triggered as material is channelled inward causing the inner spirals to become sufficiently dense to fragment (Meru, 2015).

In this chapter we present a suite of SPH simulations of binary star systems. We extensively test the parameter space of binary orbital properties for configurations which may trigger fragmentation in discs which would be marginally stable in isolation. We evolve a total of 62 discs which, to our knowledge, represents the most thorough search of this parameter space to date, and in each simulation we model realistic cooling through the Forgan et al. (2009) hybrid radiative transfer

method. Section 4.2 details the various disc setups explored in our simulations. Section 4.3 presents the results obtained, which we discuss in Section 4.4. Our conclusions are presented in Section 4.5.

4.2 Methods – SPH simulations

We simulate a three-dimensional gaseous disc using the PHANTOM SPH code (Price et al., 2018), which has been modified to include the hybrid radiative transfer method introduced in Forgan et al. (2009) and outlined in Section 2.4.1.4. We also include the standard SPH artificial viscosity, with parameters $\alpha_{\rm SPH} = 0.1$ and $\beta_{\rm SPH} = 0.2$.

Each disc is initialised with $N = 1 \times 10^6$ SPH particles, distributed such that the initial surface density profile of the disc is $\Sigma(R) = \Sigma_0 (R/R_0)^{-1.5}$ and the temperature profile is $T(R) = T_0 (R/R_0)^{-1.0}$ between $R_0 = 1\,\mathrm{AU}$ and $R_{\rm out} = 100\,\mathrm{AU}$. In each disc T_0 and Σ_0 are determined self-consistently, with $T_0 = 374\,\mathrm{K}$ for all discs set up here, and Σ_0 varying with the disc mass being considered. Any particles that fall within R_0 are accreted onto the central star, which is represented as a point mass particle. When considering binary star systems, we set up circumprimary discs only, and the companion star behaves as a gravitationally bound point mass, modelled using a sink particle.

4.2.1 Suite of SPH models

To effectively explore the parameter space in binary star separation, a, orbital eccentricity, e, orbital inclination, i, and companion star mass, $M_{*,\rm companion}$, we set up 4 suites of discs – one for studying each variable individually. In each case, the disc setup parameters are selected to be close to where we find the limit for disc fragmentation to be, identified during our reference run of discs which are detailed below.

4.2.1.1 Reference run of discs with no companion

We initially set up a reference run of discs with no companion. This allows us to understand how our discs would evolve in isolation, whilst also being able

N_{SPH}	$R_{\mathrm{out,disc}}$	M_*	M_{disc}
1×10^6	100 AU	$1\,\mathrm{M}_\odot$	$[0.1, 0.2, 0.3, 0.4]\,\mathrm{M}_\odot$

Table 4.1 *SPH disc setup parameters for the reference run of discs with no companion star. Final states of these discs are shown in Figure 4.1*

$M_{*,\mathrm{primary}}$	$M_{*,\mathrm{companion}}$	a	M_{disc}	e	i
$1\,\mathrm{M}_\odot$	$0.2\,\mathrm{M}_\odot$	[100, 250, 500, 1000] AU	$[0.1, 0.2, 0.3, 0.4]\,\mathrm{M}_\odot$	0	0°

Table 4.2 *SPH disc setup parameters for the initial suite of discs, where we explore the parameter space in binary semi-major axis and disc mass. Final states of these discs are shown in Figure 4.2.*

$M_{*,\mathrm{primary}}$	$M_{*,\mathrm{companion}}$	a	M_{disc}	e	i
$1\,\mathrm{M}_\odot$	$0.2\,\mathrm{M}_\odot$	[150, 200, 325, 400] AU	$0.2\,\mathrm{M}_\odot$	0	0°

Table 4.3 *SPH disc setup parameters where we probe the parameter space in binary semi-major axis further, considering small changes in binary semi-major axis and keeping the disc mass constant. Final states of these discs are shown in Figure 4.4.*

$M_{*,\mathrm{primary}}$	$M_{*,\mathrm{companion}}$	a	M_{disc}	e	i
$1\,\mathrm{M}_\odot$	$0.2\,\mathrm{M}_\odot$	[150, 200, 250, 325, 400, 500] AU	$0.2\,\mathrm{M}_\odot$	[0.25, 0.5, 0.75]	0°

Table 4.4 *SPH disc setup parameters where we explore the parameter space in binary semi-major axis and orbital eccentricity. Final states of these discs are shown in Figure 4.5.*

$M_{*,\mathrm{primary}}$	$M_{*,\mathrm{companion}}$	a	M_{disc}	e	i
$1\,\mathrm{M}_\odot$	$0.2\,\mathrm{M}_\odot$	[100, 150, 200, 250] AU	$0.2\,\mathrm{M}_\odot$	0	[30°, 60°, 90°]

Table 4.5 *SPH disc setup parameters where we explore tha parameter space in binary semi-major axis and binary inclination. Final states of these discs are shown in Figure 4.6.*

$M_{*,\mathrm{primary}}$	$M_{*,\mathrm{companion}}$	a	M_{disc}	e	i
$1\,\mathrm{M}_\odot$	$[0.1, 0.5]\,\mathrm{M}_\odot$	[150, 250, 325, 400] AU	$0.2\,\mathrm{M}_\odot$	0	0°

Table 4.6 *Additional SPH disc setup parameters where we explore the parameter space in binary semi-major axis and companion mass. Final states of these discs are shown in Figure 4.7.*

to identify the region of parameter space where our discs are near to the limit for fragmentation. We set up 4 discs here with masses $M_{\rm disc} = 0.1, 0.2, 0.3$ and $0.4\,{\rm M}_{\odot}$, and a parent star of mass $M_{*} = 1.0\,{\rm M}_{\odot}$. The value of Σ_0 for each of these discs, and all subsequent discs of the same mass, are 322, 644, 966, $1288\,{\rm g\,cm^{-2}}$, respectively. A summary of these setup parameters can be found in Table 4.1.

4.2.1.2 Varying binary star separation

In the first of our suites which include a companion star, we aim to determine how binary separation affects a disc's susceptibility to fragmentation. We set up a grid of 16 systems in which we vary a between 100–1000 AU and $M_{\rm disc}$ between 0.1–$0.4\,{\rm M}_{\odot}$. We then explore 4 additional cases of a with finer steps in binary separation between $a = 150$–$400\,{\rm AU}$ and a fixed disc mass $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$. In all the setups we consider a primary star with mass $M_{*,\rm primary} = 1.0\,{\rm M}_{\odot}$ and a companion star with mass $M_{*,\rm companion} = 0.2\,{\rm M}_{\odot}$, hence a stellar mass ratio $q_{\rm binary} = 0.2$. For this initial suite we set up binaries on circular orbits in the plane of the disc, with $e = 0$ and $i = 0$. These disc setups are summarised in Tables 4.2 and 4.3.

Results from this initial suite of discs, where we vary the companion's semi-major axis, were then used to inform the range of semi-major axes considered for the subsequent suites of disc simulations.

4.2.1.3 Varying orbital eccentricity

We then wish to study the effect of varying the orbital eccentricity of the binary orbit. 18 new discs are setup where we introduce eccentricities, $e = 0.25$, 0.5 and 0.75. We vary the binary separation between $a = 150$–$500\,{\rm AU}$ whilst keeping the disc mass constant at $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$. In all cases we consider a primary star mass, $M_{*,\rm primary} = 1.0\,{\rm M}_{\odot}$, a companion star mass, $M_{*,\rm companion} = 0.2\,{\rm M}_{\odot}$, and binary orbits in the plane of the disc, with $i = 0$. A summary of these disc setups is outlined in Table 4.4.

4.2.1.4 Varying orbital inclination

A third suite of discs is set up where we study the effect of varying the companion star's orbital inclination relative to the plane of the disc. We set up 12 new discs which include inclinations $i = 30°$, $60°$ and $90°$. We consider binary separations in the range $a = 100$–$250\,\mathrm{AU}$, whilst keeping the disc mass constant at $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$. In each case we consider a primary star mass, $M_{*,\mathrm{primary}} = 1.0\,\mathrm{M}_{\odot}$, a companion star mass, $M_{*,\mathrm{companion}} = 0.2\,\mathrm{M}_{\odot}$, and circular binary orbits with $e = 0$. A summary of these disc setups is outlined in Table 4.5.

4.2.1.5 Varying companion star mass

Finally, we set up a suite of discs to explore the effect of varying the mass of the companion star. We set up a grid of discs which includes 8 new setups, where we introduce companion star masses $M_{*,\mathrm{companion}} = 0.1\,\mathrm{M}_{\odot}$ and $0.5\,\mathrm{M}_{\odot}$. We vary the binary separation between $a = 150$–$400\,\mathrm{AU}$, whilst keeping the disc mass constant at $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$. For all disc setups we consider a primary star mass, $M_{*,\mathrm{primary}} = 1.0\,\mathrm{M}_{\odot}$, and circular binary orbits in the plane of the disc, with $e = 0$ and $i = 0$. A summary of these disc setups is outlined in Table 4.6.

4.3 Results

Results from the final states of all the discs simulated here are summarised in Table 4.7. In all cases we allow the discs to evolve for at least 5 orbital periods at the disc outer edge ($R_{\mathrm{out}} = 100\,\mathrm{AU}$), equivalent to $t = 5000\,\mathrm{yrs}$, or until fragmentation occurs. In the case of the wide orbit binary systems, for which the binary orbit is longer than 5 orbital periods at $R_{\mathrm{out}} = 100\,\mathrm{AU}$, we allow the simulations to evolve for at least a full binary orbit. We define a simulation as having fragmented when a local clump forms where the density is significantly higher than the surrounding disc gas. Typically, these clumps have densities that are a few orders of magnitude greater than the surrounding gas.

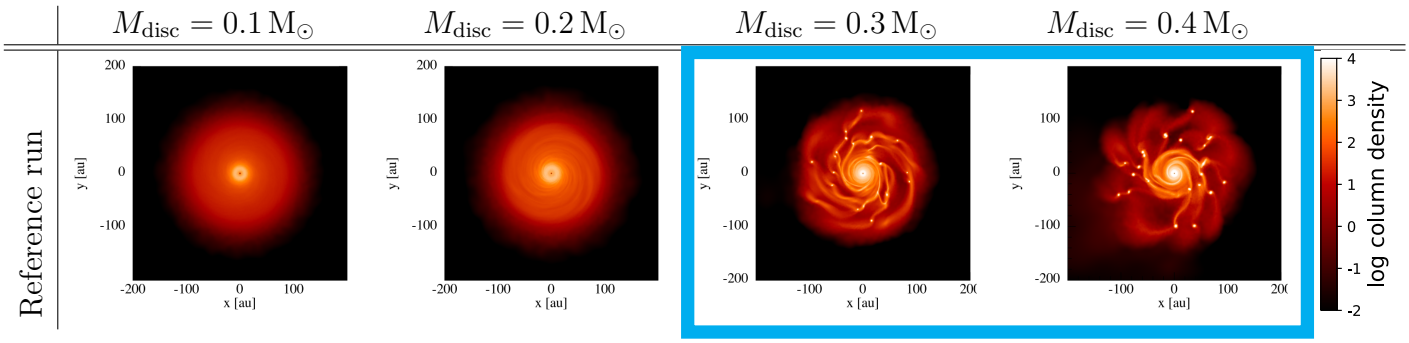

Figure 4.1 *Final states of the reference run of disc setups with no companion star included. A summary of the disc setup parameters laid out in Table 4.1, and outlined in detail in Section 4.2.1.1. Blue boxes are included to highlight the reference run discs which resulted in fragmentation.*

4.3.1 Reference run of discs with no companion

Figure 4.1 shows the final states of the reference run of discs with no companion star included. Setup parameters for these discs are outlined in Section 4.2.1.1 and summarised in Table 4.1. We find the lower mass limit for disc fragmentation to be in the range $0.2 < M_{\rm disc} < 0.3\,{\rm M}_\odot$. The disc with $M_{\rm disc.} = 0.2\,{\rm M}_\odot$ is able to evolve for the full simulation time without fragmenting, whilst the disc with $M_{\rm disc} = 0.3\,{\rm M}_\odot$ fragments quickly, after 700 yrs of evolution.

In Figure 4.3 we plot the azimuthally averaged midplane disc properties from some of the systems simulated here. We include plots of the Toomre parameter, Q, the disc cooling time, $t_{\rm cool} (= u/\dot{u})$, from Forgan et al. (2009), and the dimensionless cooling parameter, $\beta_{\rm cool} = t_{\rm cool}\Omega$, where $\Omega = \sqrt{{\rm G}M_*/r^3}$ is the Keplerian frequency at a given radius. Early work has shown that fragmentation may occur if the disc is able to cool on dynamical timescales, with $\beta_{\rm cool} \lesssim 3$ (Gammie, 2001; Rice et al., 2003).

From Figure 4.3, we find that the system with $M_{\rm disc} = 0.2\,{\rm M}_\odot$ reaches a marginally stable final state with $Q_{\rm min} = 1.06$ at $R = 74\,{\rm AU}$, and $\beta_{\rm cool,min} = 9.75$ at $R = 100\,{\rm AU}$.

In order to ensure that the fragmentation threshold found here is independent of the discs' initial setup, we also ran each disc such that $Q \approx 1$ at $R = R_{\rm out}$, by adjusting the value of the disc aspect ratio (H/R) at $R = R_{\rm out}$. Again, we found the limit for fragmentation to be in the range $0.2 < M_{\rm disc} < 0.3\,{\rm M}_\odot$.

4.3.2 Varying binary separation

4.3.2.1 Initial suite of discs

Results from the initial suite of discs where we vary binary separation and disc mass are displayed in Figure 4.2. Setup parameters for these discs are outlined in Section 4.2.1.2 and summarised in Table 4.2. As with the results in Section 4.3.1, we also ran each of these discs with slightly different initial $Q-$profiles to ensure our conclusions remain consistent.

When comparing the results in Figure 4.2 to those from the reference run in Figure 4.1, we find one disc configuration, with $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$ and $a = 250\,{\rm AU}$, where the simulation results in fragmentation, and its analog from the reference run, with $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$ and no companion star, did not. The companion star's initial eccentricity is $e = 0$, however energy exchange with disc material throughout the companion's orbit results in a periastron binary separation of $r_{\rm peri,actual} = 186\,{\rm AU}$. As the companion approaches and passes through periastron, an $m = 2$ spiral mode propagates through the disc generating a bump in the surface density at $R \approx 60\,{\rm AU}$.

This can be seen in the plots of azimuthally averaged disc properties shown in Figure 4.3. When comparing the disc properties for the system where $a = 250\,{\rm AU}$ and $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$ immediately before fragmentation occurs (at $t \approx 2700\,{\rm yrs}$), and the properties from the final state of the analog disc from the reference run (where $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$), we can see how the surface density increases in the disc of the $a = 250\,{\rm AU}$ system, consistent with the location of the spirals driven by the companion. Efficient cooling, evident from the drop in $t_{\rm cool}$ between $\approx$60–90 AU, is able to prevent the disc temperature from increasing significantly at the spiral location. Hence Q can drop to below $Q = 1$ and fragmentation ensues. Immediately before this disc fragments we find $\beta_{\rm cool,min} = 2.92$ at $R = 75\,{\rm AU}$.

Binaries on wide orbits ($a = 500, 1000\,{\rm AU}$) converge to the single star solution, where the mass limit for fragmentation is $0.2\,{\rm M}_{\odot} < M_{\rm disc} < 0.3\,{\rm M}_{\odot}$. In Figure 4.3 we also plot the azimuthally averaged midplane disc properties for the setup where $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$ and $a = 500\,{\rm AU}$ at the time of periastron passage. The temperature profile and $Q-$profile are almost identical to the analog disc from the reference run, with $Q_{\rm min} = 1.05$ at $R = 75\,{\rm AU}$. Similarly, we find no significant bump in the surface density profile which would be consistent with a spiral being driven by the companion star. The $M_{\rm disc} = 0.3\,{\rm M}_{\odot}$ system with a companion at

$a = 500\,\mathrm{AU}$ fragments quickly, as it did around a single star.

Binaries with small semi-major axes, whose orbits result in the companion star passing through the disc ($a = 100\,\mathrm{AU}$), are found to be prohibitive to fragmentation. In Figure 4.3 we plot the azimuthally averaged midplane properties for the disc with $a = 100\,\mathrm{AU}$ and $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$, at the point of periastron passage (where $r_{\mathrm{peri,actual}} = 78\,\mathrm{AU}$). As the companion moves through the disc the midplane temperature increases, whilst material is simultaneously ejected from the outer disc and channeled toward the inner disc. The final surface density profiles for the $a = 100\,\mathrm{AU}$ discs, set up with $M_{\mathrm{disc}} = 0.1\,\mathrm{M}_{\odot}$ and $0.2\,\mathrm{M}_{\odot}$, are consequently truncated at $a \approx 40\,\mathrm{AU}$, with final disc masses $M_{\mathrm{disc}} = 0.06\,\mathrm{M}_{\odot}$ and $0.12\,\mathrm{M}_{\odot}$ respectively. Hence no fragmentation can occur. The massive discs with $M_{\mathrm{disc}} \geq 0.3\,\mathrm{M}_{\odot}$ are still able to fragment quickly, before completing a full binary orbital period, in a spiral arm which trails the path of the companion star.

4.3.2.2 Further probing the parameter space in binary separation

As the results in Figure 4.2 indicate that there may be a sweet spot in binary separation at $a \approx 250\,\mathrm{AU}$ which can trigger fragmentation, we ran an additional set of discs where we probe this region of parameter space with greater granularity. This consists of 4 additional discs with $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$ and $a = 150, 200, 325$ and $400\,\mathrm{AU}$. Setup parameters for these are outlined in Section 4.2.1.2 and summarised in Table 4.3. Final states of these discs are shown in Figure 4.4.

We find that the disc setups with binary semi-major axes between $150\,\mathrm{AU} \leq a \leq 250\,\mathrm{AU}$ all result in fragmentation. In the configuration with $a = 150\,\mathrm{AU}$, the companion narrowly avoids passing through the outer extent of the disc, with $r_{\mathrm{peri,actual}} = 116\,\mathrm{AU}$. As the companion approaches and passes through periastron, an $m = 2$ spiral mode propagates through the disc causing a significant drop in Q at the inner regions of one of these spirals, between $30\,\mathrm{AU} \leq R \leq 50\,\mathrm{AU}$, and 4 fragments initially form. All 4 of these fragments survive as the companion travels back towards apastron, as can be seen in the final state of the disc in Figure 4.4. A similar process occurs in the system with $a = 200\,\mathrm{AU}$.

As we increase the binary separation beyond $a = 250\,\mathrm{AU}$ the influence of the companion star becomes progressively weaker. Once $a \geq 325\,\mathrm{AU}$ the companion star can no longer trigger fragmentation in the disc. The spiral mode induced by the $a = 325\,\mathrm{AU}$ companion is much weaker than in the $a = 250\,\mathrm{AU}$ disc. We find

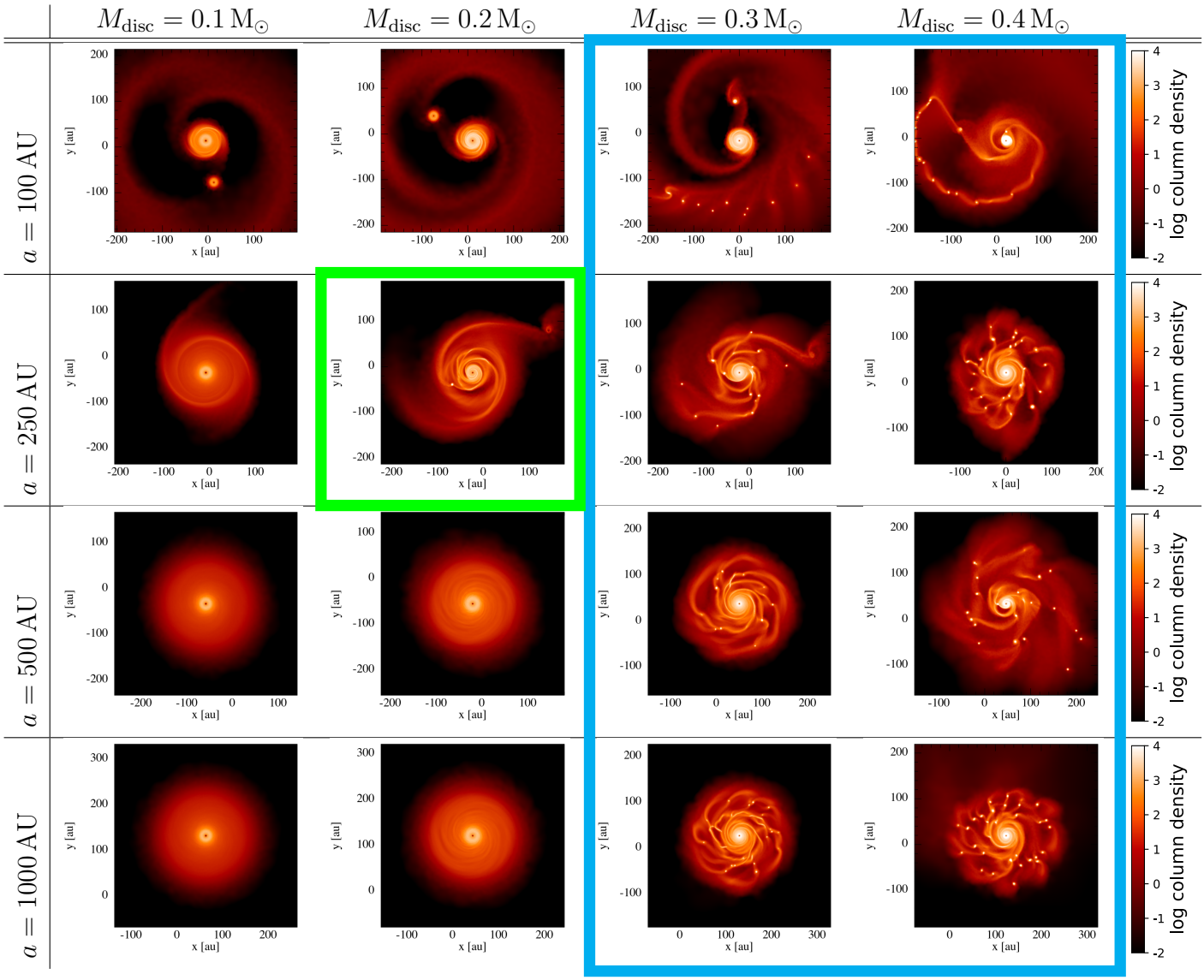

Figure 4.2 *Final states of the discs where we vary the disc mass and the semi-major axis of the companion star. Disc setup parameters are summarised in Table 4.2 and outlined in detail in Section 4.2.1.2. Blue boxes are included to highlight discs which resulted in fragmentation when their reference run analogs also fragmented. Green boxes are included to highlight disc configurations which resulted in fragmentation when their reference run analog did not fragment.*

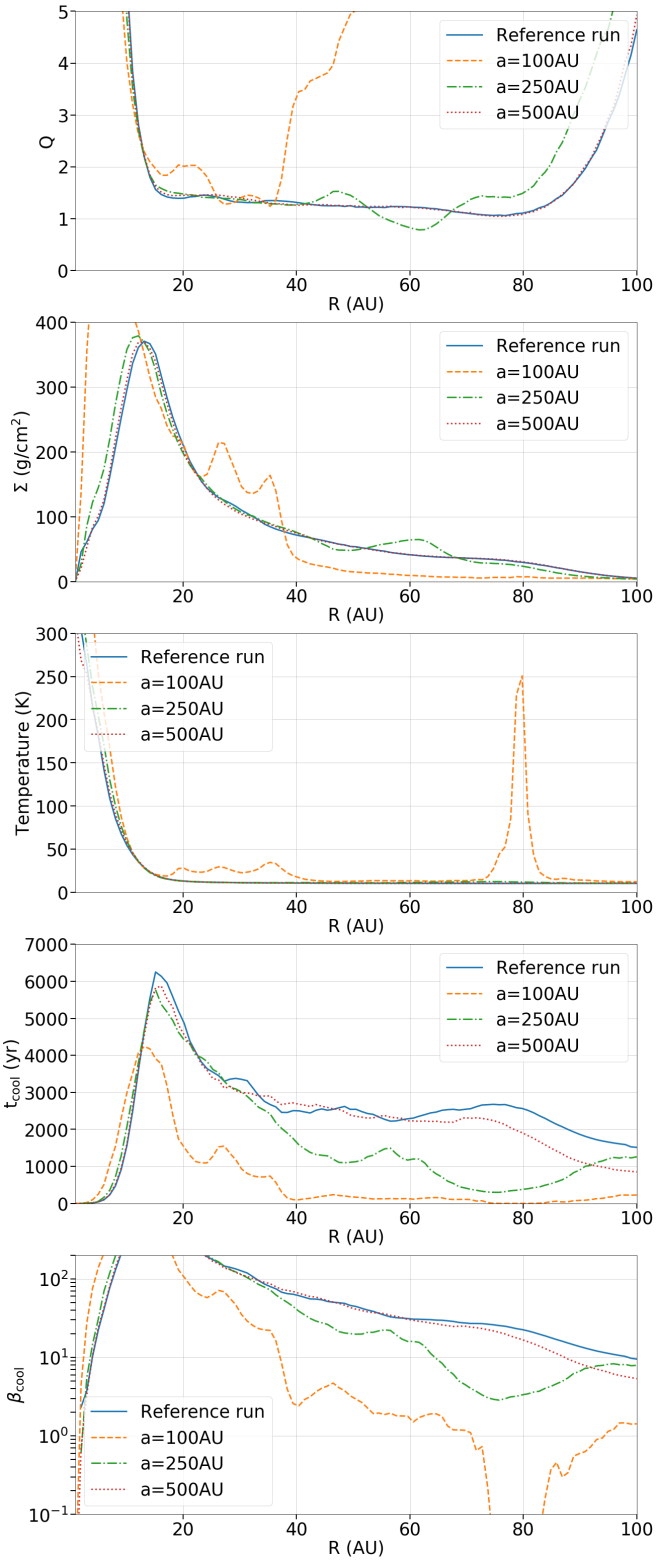

Figure 4.3 *Azimuthally averaged midplane disc properties calculated from the* $M_{\rm disc} = 0.2\,M_{\odot}$ *discs from the initial suite which includes a binary companion (setup parameters in Table 4.2, final states in Figure 4.2). We plot the reference run final state, the* $a = 100\,AU$ *run at periastron, the* $a = 250\,AU$ *run immediately before fragmentation, and the* $a = 500\,AU$ *run at periastron.*

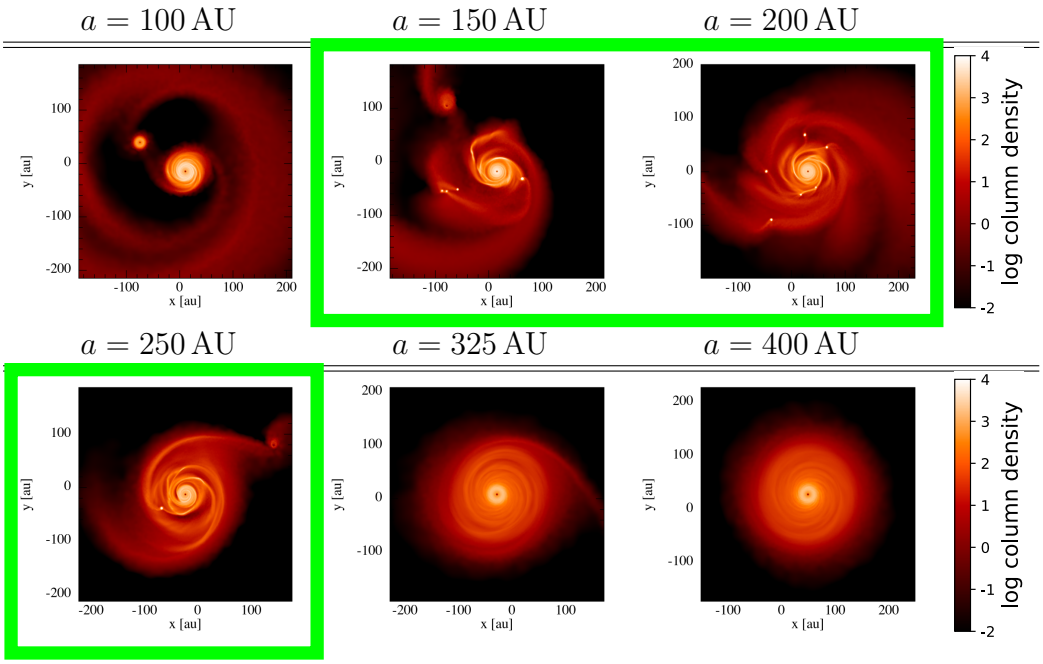

Figure 4.4 *Final states of the discs where we keep the disc mass constant and consider small changes in the semi-major axis of the companion star. Disc setup parameters are summarised in Table 4.3 and outlined in detail in Section 4.2.1.2. Green boxes are included to highlight disc configurations which resulted in fragmentation, when their reference run analog did not.*

that Q drops slightly in the outer disc of the $a = 325\,\mathrm{AU}$ system when compared to its analog from the reference run (with $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$ and no companion), but not enough to push the disc over the fragmentation threshold. Once we increase the binary semi-major axis to $a = 400\,\mathrm{AU}$ we find a similar result as with the $a = 500\,\mathrm{AU}$ system in the previous section. The surface density profile, $Q-$profile and temperature profile at periastron passage become increasingly similar to the final state of the reference run disc with $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$.

4.3.3 Varying orbital eccentricity

So far we have only considered companion stars on circular orbits, with $e = 0$. In reality it is likely that there will be some orbital eccentricity. Hence in this section we simulate 18 new discs, introducing eccentricities, $e = 0.25, 0.5$ and 0.75. We consider setup parameters found to be near to the limit for fragmentation, keeping the disc mass constant at $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$, and varying a and e only. These setups are outlined in Section 4.2.1.3 and summarised in Table 4.4. Final states of these discs are shown in Figure 4.5.

In the previous section we found that companion stars with semi-major axes $150\,\mathrm{AU} \leq a \leq 250\,\mathrm{AU}$ ($116\,\mathrm{AU} \leq r_{\mathrm{peri,actual}} \leq 186\,\mathrm{AU}$) may induce fragmentation in a disc which would not fragment in isolation. When including eccentricity we find a wider range of semi-major axes are capable of inducing fragmentation, given that $r_{\mathrm{peri,actual}}$ falls roughly within the same range as found previously.

When including an eccentricity of $e = 0.5$ we find that disc setups with $a = 325\,\mathrm{AU}$ and $a = 400\,\mathrm{AU}$ now also result in fragmentation. The periastron distances observed in these simulations are $r_{\mathrm{peri,actual}} = 134\,\mathrm{AU}$ and $r_{\mathrm{peri,actual}} = 163\,\mathrm{AU}$ respectively. Fragmentation occurs in a similar manner here to what was found in the previous section; an $m = 2$ spiral which is generated as the companion approaches and passes through periastron becomes unstable and forms bound, self-gravitating clumps.

We find that fragmentation occurs in this suite for a slightly lower $r_{\mathrm{peri,actual}}$ than was found in the previous section. In the configuration with $a = 150\,\mathrm{AU}$ and $e = 0.25$, corresponding to $r_{\mathrm{peri,actual}} = 92\,\mathrm{AU}$, the companion passes through the very outer edge of the disc and a single fragment forms in the spiral 180° from the companion's location. Disc heating at periastron is much less here than was the case when $a = 100\,\mathrm{AU}$ and $e = 0$ ($r_{\mathrm{peri,actual}} = 78\,\mathrm{AU}$), which was found to inhibit fragmentation. The amount of mass ejected is also significantly less here, with a final disc mass $M_{\mathrm{disc}} = 0.18\,\mathrm{M}_{\odot}$. What remains is a slightly more compact disc, truncated at $R \approx 75\,\mathrm{AU}$.

For any companion which passes closer than $r_{\mathrm{peri,actual}} = 92\,\mathrm{AU}$, the disc-star interaction becomes destructive. Disc material is dispersed as the companion passes through the disc, ejecting a significant amount of mass, leaving a compact, lower mass disc, thus entirely preventing fragmentation. All configurations here with $e = 0.75$ suffer this fate.

An interesting case is the disc setup with $a = 500\,\mathrm{AU}$ and $e = 0.75$. Despite this system's actual periastron distance of $r_{\mathrm{peri,actual}} = 104\,\mathrm{AU}$ falling within the sweet spot found which may induce fragmentation, no fragmentation occurs here. As the companion passes through periastron a self-gravitating clump begins to form at the inner edge of one of the spirals but immediately disperses as the companion quickly moves away from periastron. As the companion moves back toward apastron the disc stabilizes again.

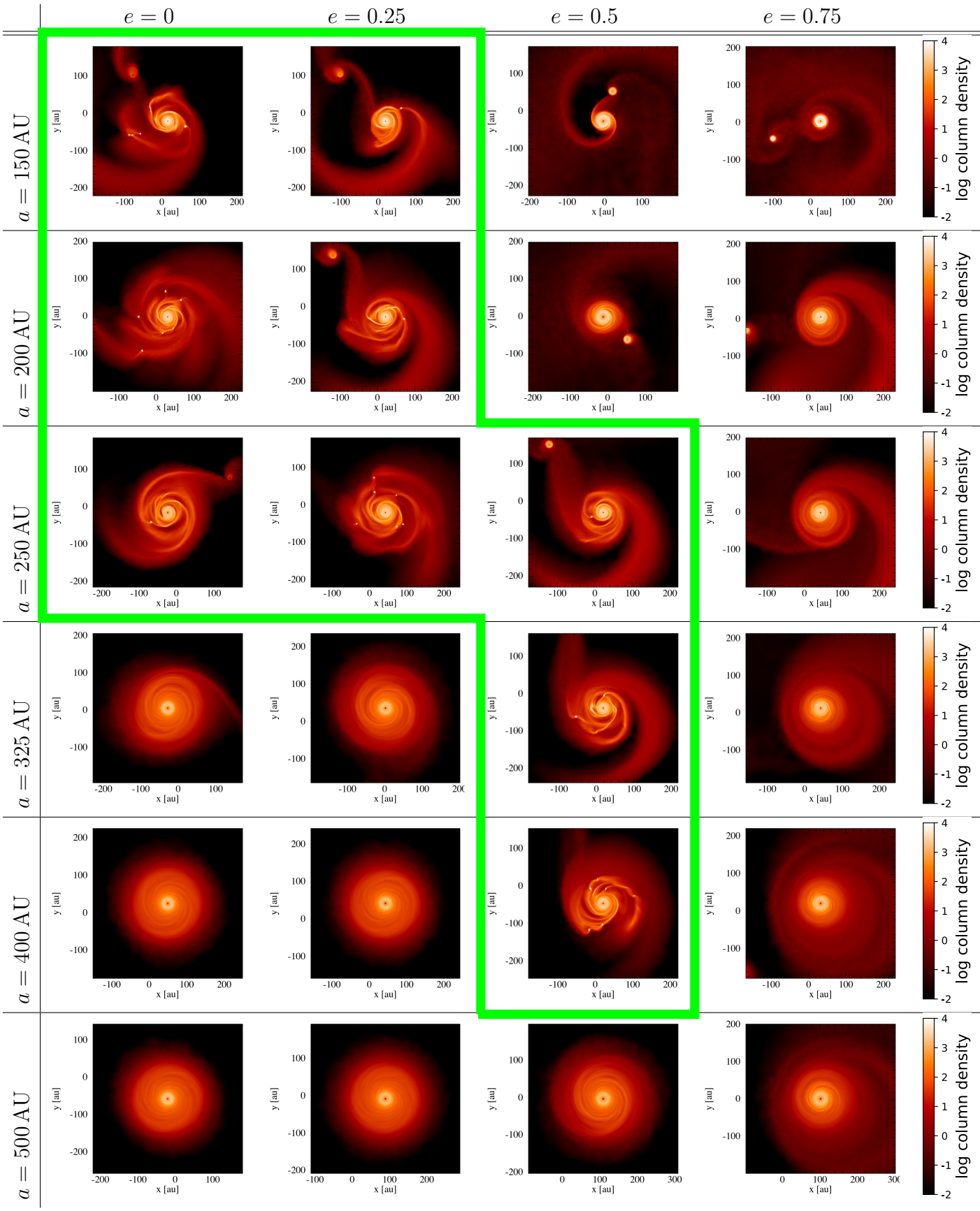

Figure 4.5 *Final states of the discs where we vary the orbital eccentricity and the semi-major axis of the companion star. Disc setup parameters are summarised in Table 4.4 and outlined in detail in Section 4.2.1.3. Green boxes are included to highlight disc configurations which resulted in fragmentation, when their reference run analog did not.*

4.3.4 Varying orbital inclination

We now consider systems with some orbital inclination relative to the plane of the disc. We introduce inclinations $i = 30°, 60°$ and $90°$, once again keeping the disc mass constant at $M_{\rm disc} = 0.2\,{\rm M}_{\odot}$ and varying the binary separation close to the limit for fragmentation. Disc setup parameters are outlined in Section 4.2.1.4 and summarised in Table 4.5. Final states of these discs are shown in Figure 4.6.

In the short orbit system, with $a = 100\,{\rm AU}$, we find that including an inclination $i \geq 60°$ results in a less destructive disc-star interaction than when the companion orbits in the plane of the disc, hence fragmentation can occur. Despite their $r_{\rm peri,actual}$ being similar, the resulting surface density profile after the companion has passed through the plane of the disc is much less truncated when $i = 60°$ than was the case when $i = 0°$, extending to $R_{\rm out} \approx 80\,{\rm AU}$ after several binary orbital periods. An $m = 2$ spiral mode forms quickly, and a fragment forms at the inner edge of one the spirals, at $R = 55\,{\rm AU}$. The fragment's orbit is initially slightly inclined relative to the disc, with $i \approx 8°$, but it quickly settles into the plane of the disc after an orbital period. The final state of the $i = 60°$ disc is more flared compared to when $i = 0°$, with $H/R = 0.15$ at $R = 100\,{\rm AU}$ compared to $H/R = 0.10$ at $R = 100\,{\rm AU}$.

In the wider orbit systems, with $a = 200\,{\rm AU}$ and $250\,{\rm AU}$, we find that including an inclination can weaken the disc-star interaction compared to when $i = 0°$ such that disc fragmentation no longer occurs. Considering the discs with $a = 250\,{\rm AU}$, as we increase the companion's inclination from $i = 0°$ to $i = 90°$ we find that the $m = 2$ spiral mode generated by the companion becomes progressively weaker. When $i = 30°$, we observe a much smaller bump in the surface density profile when compared to the system with $i = 0°$. Hence the $Q-$profile remains relatively flat, and no fragmentation occurs. For the discs with $a = 200\,{\rm AU}$, only a single fragment forms when $i = 60°$, and increasing the inclination to $i = 90°$ prevents fragmentation from happening altogether.

4.3.5 Varying companion mass

The gravitational influence of the companion star on the disc will vary with the star's mass. Hence we setup 8 new discs where we explore the parameter space in $M_{*,\rm companion}$ and a for configurations which result in fragmentation. In particular, we focus on how varying the companion star's mass affects the binary separations

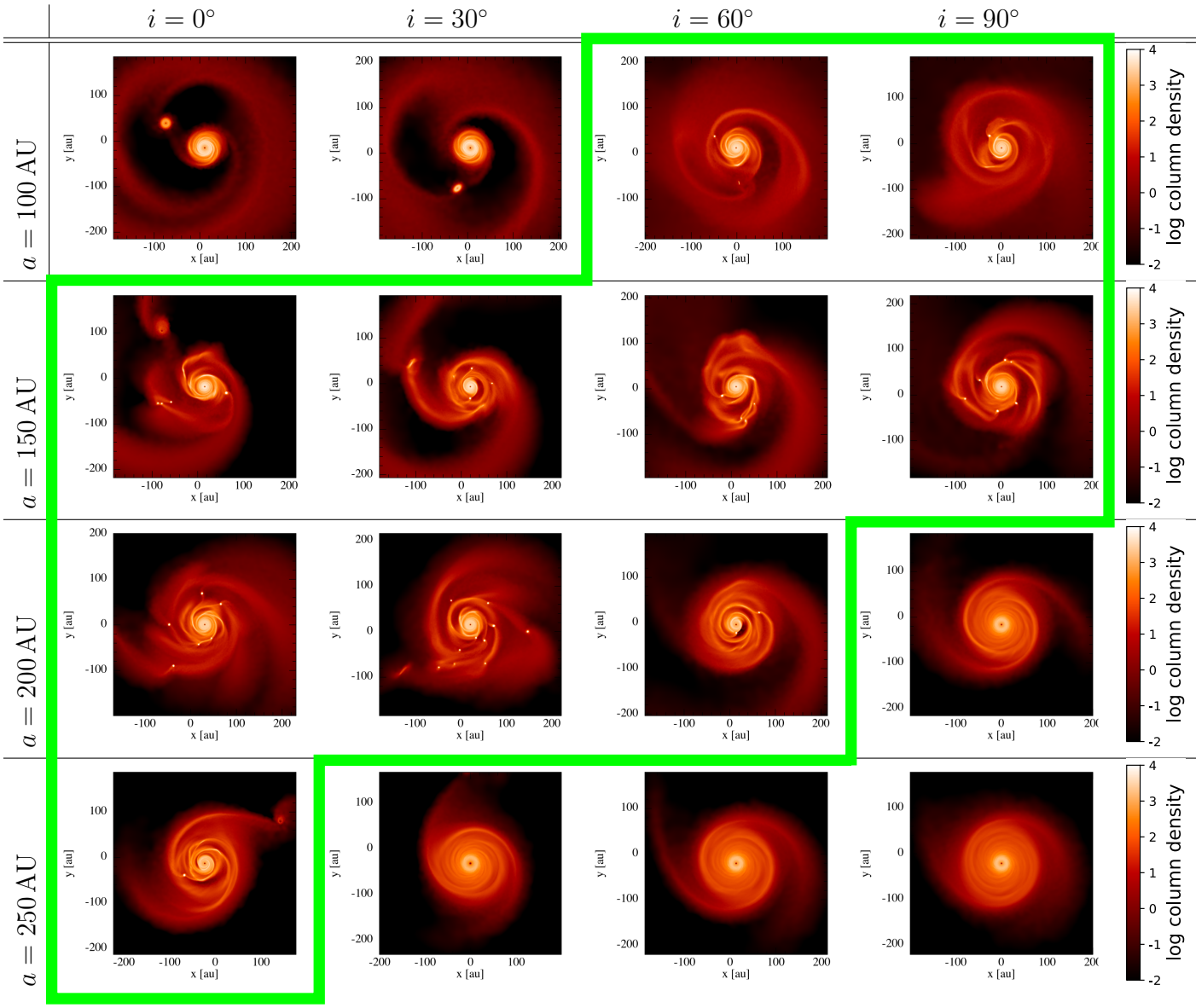

Figure 4.6 *Final states of the discs where we vary the orbital inclination and the semi-major axis of the companion star. Disc setup parameters are summarised in Table 4.5 and outlined in detail in Section 4.2.1.4. Green boxes are included to highlight disc configurations which resulted in fragmentation, when their reference run analog did not.*

which are capable of inducing fragmentation.

Discs are setup with parameters close to the fragmentation limit found until now, keeping the disc mass constant at $M_{\rm disc} = 0.2\,{\rm M}_\odot$ and varying $M_{*,\rm companion}$ and a only. We introduce new companion masses $M_{*,\rm companion} = 0.1\,{\rm M}_\odot$ and $M_{*,\rm companion} = 0.5\,{\rm M}_\odot$. These setups are outlined in Section 4.2.1.5 and summarised in Table 4.6. The final states of these discs are shown in Figure 4.7.

Considering the discs with $a = 150\,{\rm AU}$, we find that both of the new setups, with $M_{*,\rm companion} = 0.1\,{\rm M}_\odot$ and $M_{*,\rm companion} = 0.5\,{\rm M}_\odot$, result in fragmentation, with a trend for the discs to fragment faster with increasing $M_{*,\rm companion}$. We find $t_{\rm frag} = 900, 750$ and $550\,{\rm yrs}$ when $M_{*,\rm companion} = 0.1, 0.2$ and $0.5\,{\rm M}_\odot$ respectively.

When $a = 250\,{\rm AU}$ we find that fragmentation no longer occurs after decreasing $M_{*,\rm companion}$ from $0.2\,{\rm M}_\odot$ to $0.1\,{\rm M}_\odot$. A weaker $m = 2$ spiral mode is driven by the $0.1\,{\rm M}_\odot$ companion, which results in an increase in Σ and a decrease in Q at the spiral location, but the change is not significant enough to induce fragmentation. Instead, the spiral mode persists for two full binary orbits until the simulation ends. For the discs which do fragment, we again find a trend for discs to fragment earlier with increasing companion mass, finding $t_{\rm frag} = 3050$ and $1550\,{\rm yrs}$ for $M_{*,\rm companion} = 0.2$ and $0.5\,{\rm M}_\odot$ respectively.

Considering the discs with $a = 325\,{\rm AU}$ we find that increasing $M_{*,\rm companion}$ from $0.2\,{\rm M}_\odot$ to $0.5\,{\rm M}_\odot$ causes the disc to fragment, as the more massive star generates a stronger spiral mode. A single fragment forms after $4000\,{\rm yrs}$ in the disc with $a = 325\,{\rm AU}$ and $M_{*,\rm companion} = 0.5\,{\rm M}_\odot$. Of all the discs simulated here, this is the widest periastron separation ($r_{\rm peri,actual} = 256\,{\rm AU}$) for which we find fragmentation can be triggered.

None of the configurations where $a = 400\,{\rm AU}$ result in fragmentation, for any of the new companion star masses considered here.

4.4 Discussion

4.4.1 Summary of results

We have identified a "sweet spot" in the orbital parameter space of binary stars which may trigger fragmentation in a disc which does not fragment in isolation.

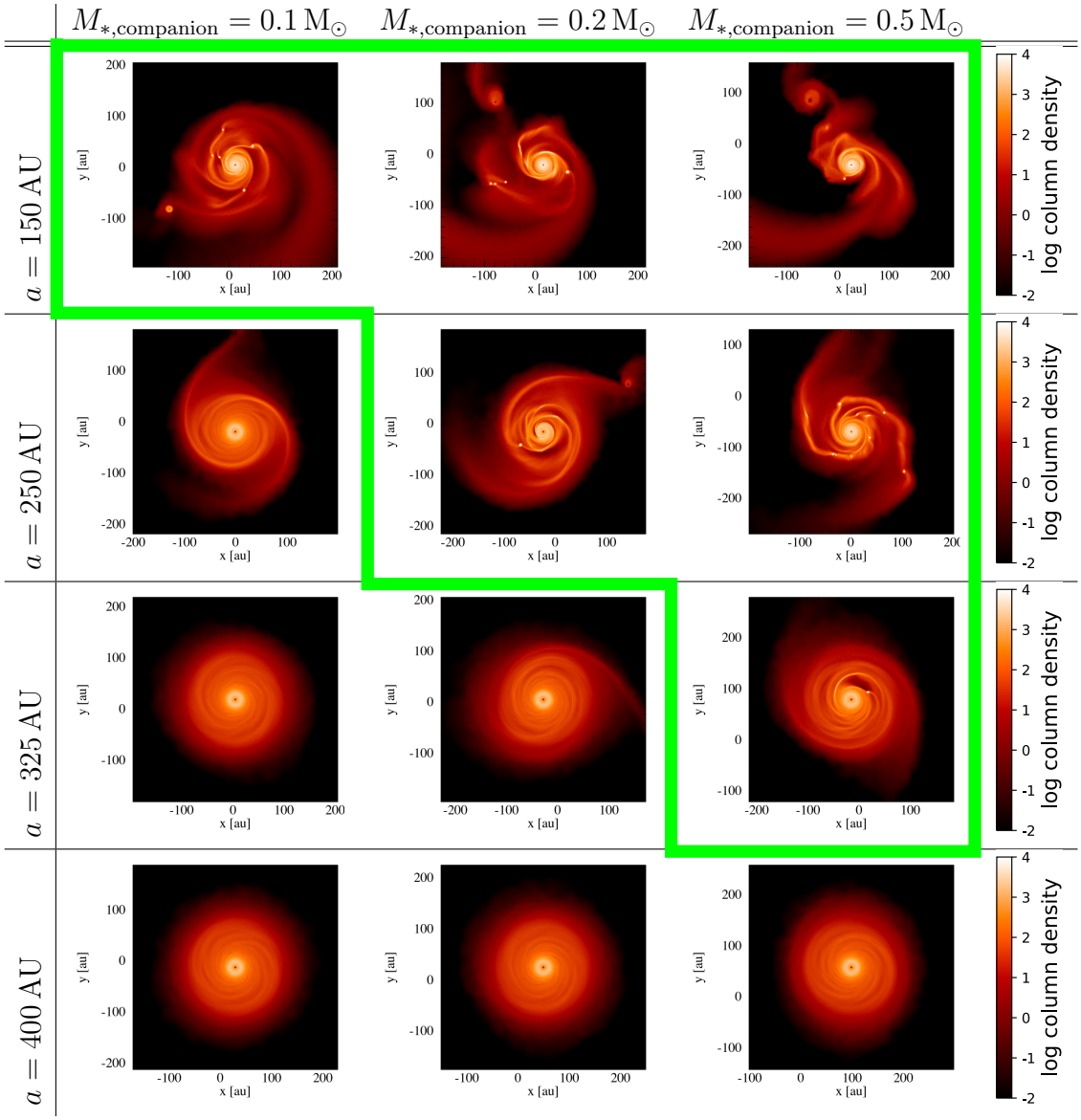

Figure 4.7 *Final states of the discs where we vary the mass and the semi-major axis of the companion star. Disc setup parameters are summarised in Table 4.6 and outlined in detail in Section 4.2.1.5. Green boxes are included to highlight disc configurations which resulted in fragmentation, when their reference run analog did not.*

We find that the disc-star interaction for intermediate separation binaries can be beneficial for fragmentation, with the exact range of ideal semi-major axes being a function of the orbital eccentricity, inclination and companion star mass. A plot summarising the companion's orbital parameters which are found to trigger fragmentation is included in Figure 4.8, where we also highlight the minimum radius at which fragments formed in each disc.

In general, the companion will drive an $m = 2$ spiral through the disc, and fragmentation occurs at the inner region of one, or both, of the spirals as a result of the enhanced surface density pushing the disc over the fragmentation threshold. Heating of the disc induced by the companion is balanced by efficient cooling (see Figure 4.3) and the instability is able to grow until fragmentation occurs.

We find that this is true for intermediate separation binaries with $150\,\mathrm{AU} \leq a \leq 250\,\mathrm{AU}$ ($116\,\mathrm{AU} \leq r_{\mathrm{peri,actual}} \leq 186\,\mathrm{AU}$), when considering circular binary orbits in the plane of the disc.

For wide orbit binaries the spiral induced by the companion becomes progressively weaker with increasing binary separation. When $a \gtrsim 400\,\mathrm{AU}$ and $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$, the disc's final surface density profile and $Q-$profile are almost identical to the counterpart disc from the reference run with no companion.

Very short separation binary encounters, where the companion passes through the outer edge of the disc, become prohibitive to fragmentation. As the companion star passes through the disc, material is ejected and the remaining surface density profile is modified to be much steeper in the inner disc, and truncated at a distance slightly smaller than the distance of periastron passage. Hence a much more compact and lower mass disc remains, and no fragmentation can occur.

When including an eccentricity in the binary orbit, we find a similar range in $r_{\mathrm{peri,actual}}$ capable of triggering fragmentation. From the suite considering non-circular orbits with moderate eccentricities ($e = 0.25, 0.5$) we find that semi-major axes $150\,\mathrm{AU} \leq a \leq 400\,\mathrm{AU}$ ($92\,\mathrm{AU} \leq r_{\mathrm{peri,actual}} \leq 163\,\mathrm{AU}$) can induce fragmentation. When considering highly eccentric orbits, with $e = 0.75$, none of our simulations fragment. This is generally because the high eccentricity causes the companion to pass through the disc at periastron passage.

When including an orbital inclination for the companion, we find its influence to become progressively lesser as we move its orbit away from the plane of the disc. When $i = 60°$ we find the sweet spot in binary semi-major axis to be between

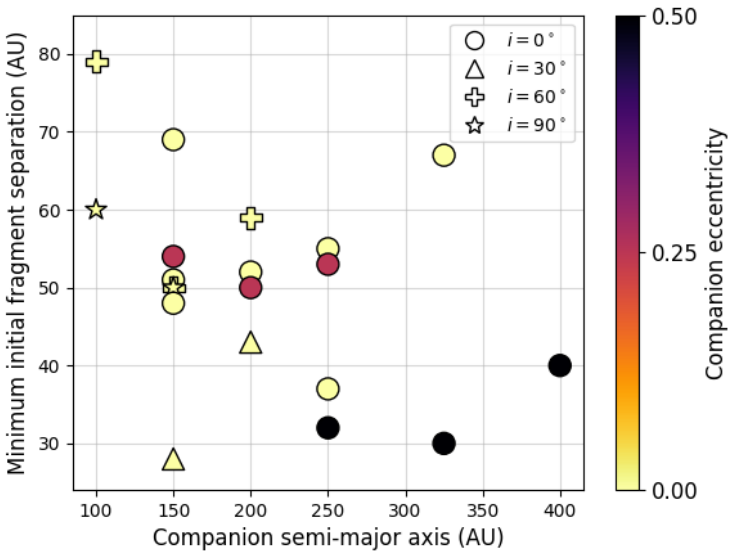

Figure 4.8 *Summary of model parameters found to trigger fragmentation throughout this work, including the minimum radius at which fragments formed in each disc. A total of 20 systems presented here resulted in fragmentation. Companion orbital eccentricities are distinguished by different colors. Companion orbital inclinations are distinguished by different plot markers.*

$100\,\mathrm{AU} \leq a \leq 200\,\mathrm{AU}$ ($74\,\mathrm{AU} \leq r_{\mathrm{peri,actual}} \leq 152\,\mathrm{AU}$), which is reduced to being between $100\,\mathrm{AU} \leq a \leq 150\,\mathrm{AU}$ ($75\,\mathrm{AU} \leq r_{\mathrm{peri,actual}} \leq 116\,\mathrm{AU}$) when considering companions with $i = 90°$. High inclination binary companions ($i = 60°, 90°$) which pass through the disc outer edge are less destructive than when the binary orbit is in the plane of the disc, hence fragmentation can occur for slightly shorter separations when $i = 60°$ and $i = 90°$ compared to when $i = 0°$.

The sweet spot found in binary separation is broadened as we increase the companion star's mass from $0.2\,\mathrm{M}_{\odot}$ to $0.5\,\mathrm{M}_{\odot}$, as the higher mass companion drives a stronger spiral mode through the disc. We find companions with semi-major axes as wide as $a = 325\,\mathrm{AU}$ ($r_{\mathrm{peri,actual}} = 256\,\mathrm{AU}$) can trigger fragmentation when $M_{*,\mathrm{companion}} = 0.5\,\mathrm{M}_{\odot}$. Equally, when considering less massive companions, the sweet spot in binary separation is narrowed. Only one of our simulations, with $a = 150\,\mathrm{AU}$ ($r_{\mathrm{peri,actual}} = 110\,\mathrm{AU}$), results in fragmentation when $M_{*,\mathrm{companion}} = 0.1\,\mathrm{M}_{\odot}$. In the disc configurations which fragment for more than one value of $M_{*,\mathrm{companion}}$ (when $a = 150\,\mathrm{AU}$ and $250\,\mathrm{AU}$), we find that the discs fragment faster with increasing companion mass.

4.4.2 Comparison to previous theoretical work

Previous work considering the possibility of fragmentation induced by the presence of a binary star companion consists of three key papers in Nelson (2000), Mayer et al. (2005) and Boss (2006), with their results discussed in the review paper Mayer et al. (2007a). Nelson (2000) and Mayer et al. (2005) found that the presence of a companion suppresses any instability due to significant tidal heating in spiral shock waves, thus stabilising the disc. Boss (2006), however, concluded that binary companions may promote fragmentation, finding that spiral waves generated from the tidal interaction between the disc and the companion would typically go on to form dense, self-gravitating clumps. In Mayer et al. (2007a) the authors largely attribute the differences in their results to the use of an artificial viscosity in Nelson (2000) and Mayer et al. (2005), which isn't included in Boss (2006), and would contribute significantly toward heating of the disc in the presence of a shock wave, given a sufficiently large artificial viscosity was included.

Various authors have also investigated the role of stellar flyby events in promoting or suppressing fragmentation in discs which would be marginally stable in isolation. The early work of Boffin et al. (1998), Watkins et al. (1998a) and Watkins et al. (1998b) found that, when considering isothermal discs, previously non-fragmenting discs would fragment during star-disc and disc-disc interactions. However later models which included more realistic cooling found that heating of the disc during the stellar encounter was sufficient to stabilise it against fragmentation (Lodato et al., 2007; Forgan & Rice, 2009).

Until now, most work has considered either simple cooling prescriptions, where the cooling time is proportional to the local orbital time ($\beta-$cooling, Equation 1.27), isothermal discs, or have not included an artificial viscosity which will capture heating from shocks. Including a more realistic algorithm to approximate radiation transport in our models allows us to model disc cooling more accurately, hence we can capture whether the disc is able to radiate away the additional energy generated through tidal heating during the binary encounter.

The aforementioned works generally also considered much more compact discs than we have modelled here (in the case of Nelson, 2000; Mayer et al., 2005; Boss, 2006; Lodato et al., 2007), or discs with $R_{\rm out} = 1000\,{\rm AU}$ (in the case of Boffin et al., 1998; Watkins et al., 1998a,b). However, owing to the simpler methods used to model disc cooling, their models are scale-free and can be scaled to different

physical units for comparison with the results here. Hence when comparing to the works of Boffin et al. (1998); Watkins et al. (1998a,b); Nelson (2000); Mayer et al. (2005); Boss (2006); Lodato et al. (2007) we can use their ratios $r_{\rm peri}/r_{\rm out,disc}$ for direct comparison to our results. This is not the case for the results from Forgan & Rice (2009), who use the Forgan et al. (2009) hybrid radiative transfer method, and considered $R_{\rm out} = 40\,{\rm AU}$ discs. For the coplanar binary encounters considered here, we find that companions with $0.92 \leq r_{\rm peri}/r_{\rm out,disc} \leq 1.86$ trigger fragmentation. Only Lodato et al. (2007) considered binary encounters within this range, finding that no fragmentation occurred in their simulations.

Of the previous works which include a similar cooling approximation as we use here, we find that our results broadly indicate the same thing. Forgan & Rice (2009) find that small separation, disc-penetrating encounters heat the disc material, whilst angular momentum transport and mass stripping result in a more stable disc configuration after the encounter. Large separation encounters have very little effect, becoming less significant as the periastron distance increases. However intermediate separation encounters may modify the surface density profile of the disc, without causing significant heating, such that the disc is more unstable over a larger range of radii after the encounter. None of the discs in Forgan & Rice (2009) fragment, but the authors suggest that there could be some region of parameter space in periastron distance which may act to promote fragmentation.

Meru (2015), who used the flux-limited diffusion approximation (e.g. Mayer et al., 2007b), also found that further fragmentation may be triggered in a disc which has fragmented already. The fragment which has initially formed causes material to be channelled inwards, increasing the density of the inner spirals, causing fragmentation.

Here, we have presented a suite of simulations which model realistic cooling using the Forgan & Rice (2009) radiative transfer approach. We find that efficient cooling is able to prevent the disc temperature from increasing significantly during the binary's periastron passage, and fragmentation can occur in the spiral regions of enhanced surface density which are driven by the companion.

4.4.3 Comparison to observations

Binaries are often neglected from observational and theoretical exoplanetary science, as they complicate the modelling of planet formation, as well as the detection and characterisation of planetary systems. Most of the work (theoretical and observational) conducted so far on planets in binaries has focused on close-in binaries (separations of tens of AU), generally agreeing that tight binaries ($<$ 50–100 AU) hinder planet formation (Bergfors et al., 2013; Kraus et al., 2012, 2016; Kaib et al., 2013). However, the first planets discovered in binary systems showed distinct orbital and physical properties from the rest of the planetary population, hinting at the possibility that binary companions could dramatically reorient the orbital configuration of planetary systems (Zucker & Mazeh, 2002). Observations of binary star systems suggest that stellar multiplicity at wider separations may play a key role in the formation of high-mass gas giant planets and brown dwarfs. Various surveys have found an excess of outer companions to stars with massive hot Jupiters or short-period stellar and substellar companions when compared to field stars, suggesting that binary star systems on separations of a few hundred AU may be favourable sites for the formation of these inner companions.

Beginning with their survey of solar-type spectroscopic binaries (SB), Tokovinin et al. (2006) found an excess of wide tertiary stellar companions for SBs with periods from 1–30 days, rising to a frequency of 96% for SBs with periods $<$ 3 days. In their series of "Friends of hot Jupiters" papers, Ngo et al. (2016) searched for stellar companions to 77 systems hosting hot Jupiters. They found that 47 ± 7% of stars hosting hot Jupiters have a binary companion with separations between 50–20,000 AU (a value 3 times higher than found for field stars), although Moe & Kratter (2021) concluded that this excess was not significant after accounting for remaining statistical biases. Nonetheless, Ngo et al. (2016) still observed a significant deficit of tight binary companions, with separations 50–100 AU, compared to wider systems, consistent with the idea that shorter-period binaries may be detrimental to planet formation (Wang et al., 2014; Kraus et al., 2016).

Using direct imaging data, Fontanive et al. (2019) searched for wide-orbit binary companions to 38 stars known to host very massive hot Jupiters or brown dwarfs (7–60 M_{Jup}) on short periods ($<$1 AU), finding a binary fraction close to 80% for these systems on separations of 20–10,000 AU, twice as high as for field stars, with a significance confirmed in Moe & Kratter (2021). Again, they observed a lack of binaries with separations of tens of AU, and instead found

an excess of intermediate separation binaries, with a peak in binary separation at ∼250 AU. The binary frequency for massive giant exoplanets and brown dwarfs ($M > 7\,M_{Jup}$) was found to be higher than for lower mass hot Jupiters (0.2–$4\,M_{Jup}$), suggesting that the stellar companion's influence may facilitate the formation of high-mass giant planets and brown dwarfs. The systems probed in Fontanive et al. (2019) also have a lower mean metallicity, consistent with that of the field (Moe et al., 2019), compared to hosts to genuine hot Jupiters like those studied in Ngo et al. (2016). Given the strong correlation seen between metallicity and the ability to form gas giant planets via core accretion (Mordasini et al., 2012; Jenkins et al., 2017), the high-mass inner substellar companions targeted in Fontanive et al. (2019) are therefore likely to have formed via GI rather than core accretion as for the lower-mass hot Jupiters.

Recently, Fontanive & Bardalez Gagliuffi (2021) reached similar conclusions, finding that giant planets have a substantially larger raw stellar multiplicity fraction than sub-Jovian planets, and that this trend further increases up to a ∼30% raw binary fraction for massive planet and brown dwarfs ($M > 7\,M_{Jup}$) on very short orbital separations (< 0.5 AU), with the most massive and shortest-period substellar companions almost exclusively observed in multiple-star environments. These systems thus appear to follow the architectures of stellar spectroscopic binaries, systematically observed as part of hierarchical triple systems (Tokovinin et al., 2006). Notably, Fontanive & Bardalez Gagliuffi (2021) showed that these extreme inner companions, with few analogues in (seemingly) single-star systems, were predominantly found to be in binaries with separations of a few hundred AU, and mostly on separations < 250 AU (despite a strong incompleteness at these separations) for substellar companions with masses above $7\,M_{Jup}$, consistent with results from Fontanive et al. (2019). In comparison, they found a peak around 600 AU (subject to the same incompleteness biases) for binaries hosting lower-mass planets or warm and cool gas giants on wider orbital separations, and these systems showed similar planet properties to the population of exoplanets orbiting single stars.

These results suggest that very wide binaries have no meaningful impact on the architectures of planetary systems, and confirm the idea that very tight binary systems have a negative impact on planet formation. In particular, it appears that binaries with tens to a few hundred AU separations prevent planet formation for sub-Jovian and giant planets with masses up to a few M_{Jup}, while wider binaries can harbour such planets but without affecting their orbital properties. This

indicates that the exoplanet population issued from core accretion only exists in binary configurations that are not disruptive to planet formation and do not influence the resulting planet properties. On the other hand, the higher-mass population of giant planets and brown dwarfs on short-period orbits, likely formed via GI, are predominantly seen in intermediate separation binaries of few hundred AU separations (Fontanive & Bardalez Gagliuffi, 2021), which must thus play a role in their existence.

Here, we show with simulations of self-gravitating discs that such intermediate separation binary companions may assist in the formation of giant planets through means of the gravitational instability. We find that when introducing a stellar companion at a few hundred AU into a disc configuration which would previously not fragment, fragmentation may be induced as the companion drives strong spirals which push the disc over the limit for instability. We find this to be true for binaries with semi-major axes between ∼100–400 AU for the explored parameter space, with some dependency on binary orbital eccentricity, inclination and companion mass. This is consistent with the binary projected separations observed by Fontanive et al. (2019) and Fontanive & Bardalez Gagliuffi (2021), a peak in the observed distribution at around 250 AU. We also note that in our simulations the orbital properties of these intermediate separation binaries remain mostly unchanged after a full orbital period, as the companions do not pass directly through the disc hence the drag that they experience from the disc material is minimal.

Shorter-period binaries and highly-eccentric systems inhibit fragmentation as the disc-star interaction becomes destructive. As the companion passes through periastron it will pass through the disc, leaving a compact, lower mass disc remaining. Hence we would expect a lower frequency of GI-born planets within tight binary systems on separations of tens of AU, which is also consistent with the shortfall of such systems in observations (Wang et al., 2014; Kraus et al., 2016; Ngo et al., 2016; Fontanive et al., 2019).

4.4.4 Outlook and implications for short-period, massive planets

Our work provides a viable formation pathway for the high-mass giant planets and brown dwarfs observed around components of multiple star systems (Fontanive

et al., 2019; Fontanive & Bardalez Gagliuffi, 2021). However, these substellar companions are actually observed on very short orbital periods (<1 AU), much tighter than the typical formation locations from disc fragmentation.

In some cases, these objects could have been scattered by the binary stellar companion onto highly eccentric orbits and then been tidally circularised onto their current orbits (Rice et al., 2015). However, as discussed earlier, this is only possible for a subset of the systems presented in Fontanive et al. (2019).

Another possibility is that these objects may have naturally migrated to their current locations. Baruteau et al. (2011) showed that fragments forming in young, massive discs will undergo rapid, type I migration before having chance to open a gap, and may be able to reach the inner disc within a few orbital periods. However, it is uncertain as to what fraction of fragments will survive this migration, and what their eventual masses will be after tidal downsizing (Nayakshin, 2010a; Boley et al., 2010).

We also find indications that fragmentation triggered by the binary companion may be occurring closer in than is usually found for discs in isolation. In Figure 4.3, considering the disc with an $a = 250\,\mathrm{AU}$ companion, the $Q-$profile reaches a minimum of $Q = 0.79$ at $R = 63\,\mathrm{AU}$, hence the first fragment initially forms at $R = 56\,\mathrm{AU}$. In Figure 4.8, we plot the minimum separation at which fragments form in each of our discs, including all systems with $M_{\mathrm{disc}} = 0.2\,\mathrm{M}_{\odot}$ which resulted in fragmentation. Of the 20 discs included in the plot, we find 9 systems produce fragments within $R = 50\,\mathrm{AU}$, and 2 form fragments within $R = 30\,\mathrm{AU}$.

It may then be that a combination of scattering, fragments forming close in, and rapid inward migration can produce the giant planets and brown dwarfs observed on very short orbital periods. Whilst Baruteau et al. (2011) considered the subsequent migration of single fragments forming in a self-gravitating discs, it is not known how a binary companion or the formation of multiple fragments may affect this. We leave this question as subject of future work.

4.5 Conclusions

Observations of systems with close-in massive planetary and brown dwarf companions suggest that almost all host a binary stellar companion on a wider orbit (Fontanive et al., 2019). Also, the properties of the close-in objects are consistent

with them having formed via fragmentation in a gravitationally unstable disc (Fontanive & Bardalez Gagliuffi, 2021). However, disc fragmentation is only likely to operate in the outer parts of such discs, requiring that these objects somehow move from where they formed onto the close-in orbits they now occupy.

In some cases, the close-in object could have been scattered by the binary stellar companion and then undergone tidal circularisation onto its current close-in orbit (Rice et al., 2015; Fontanive et al., 2019). However, in many cases the tidal circularisation timescale is far too long for this to be a viable pathway for these systems. That such systems still typically host binary stellar companions suggests that these stellar companions still play a role in their formation.

To investigate this, we have conducted a series of 3D SPH simulations of self-gravitating discs with a binary stellar companion, exploring the companion's orbital parameter space for configurations which may trigger fragmentation in a marginally gravitationally unstable disc. We find a "sweet spot" in which intermediate separation binaries can induce fragmentation, with the exact set of ideal orbital parameters being a function of the companion's semi-major axis, eccentricity, inclination and mass.

Radiation transport is modelled using the Forgan et al. (2009) hybrid approach. For the discs modelled here, with outer radii $R_{\rm out} = 100\,{\rm AU}$, we find that efficient cooling during intermediate separation ($100\,{\rm AU} \lesssim a \lesssim 400\,{\rm AU}$) binary encounters allows disc fragmentation to occur in a spiral region of enhanced surface density driven by the companion star. Short separation disc-penetrating ($a \lesssim 100\,{\rm AU}$) encounters are generally destructive, as mass stripping and disc heating entirely wipe out any instability. This is also true of highly eccentric binary orbits, which result in the companion passing through the disc. However, highly inclined ($i \gtrsim 60°$) disc-penetrating encounters can be less destructive, allowing shorter separation encounters to trigger fragmentation than when the binary orbit is in the plane of the disc. Wide orbit binary encounters ($a \gtrsim 500\,{\rm AU}$) have little effect on the disc properties, with the companion's influence becoming progressively lesser with increasing binary separation.

The range of binary separations found to promote fragmentation is consistent with the projected separations of the systems which display an excess of close-in giant planets and brown dwarfs (Wang et al., 2014; Kraus et al., 2016; Ngo et al., 2016; Fontanive et al., 2019; Fontanive & Bardalez Gagliuffi, 2021). As our results show that intermediate separation binary systems could be favourable

sites for the formation of massive substellar objects, we suggest that triggered fragmentation may contribute to the excess of massive planets and brown dwarfs observed around these systems. The question now remains how these fragments, initially formed on wide orbits, might have migrated to the very short separations ($< 1\,\mathrm{AU}$) where they are now currently observed, and will be the subject of future work.

	a	$M_{\rm disc}$	$M_{*,\rm primary}$	$M_{*,\rm companion}$	e	i	$r_{\rm peri,calc}$	$r_{\rm peri,actual}$	Fragmented?	$t_{\rm frag}$
Reference runs:	-	$0.1\,M_\odot$	$1\,M_\odot$	-	-	-	-	-	✗	-
	-	$0.2\,M_\odot$	$1\,M_\odot$	-	-	-	-	-	✗	-
	-	$0.3\,M_\odot$	$1\,M_\odot$	-	-	-	-	-	✓	700 yrs
	-	$0.4\,M_\odot$	$1\,M_\odot$	-	-	-	-	-	✓	500 yrs
Varying binary separation:	100 AU	$0.1\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	100 AU	87 AU	✗	-
	100 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	100 AU	78 AU	✗	-
	100 AU	$0.3\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	100 AU	70 AU	✓	600 yrs
	100 AU	$0.4\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	100 AU	64 AU	✓	500 yrs
	250 AU	$0.1\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	250 AU	213 AU	✗	-
	250 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	250 AU	186 AU	✓✓	3050 yrs
	250 AU	$0.3\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	250 AU	NR	✓	720 yrs
	250 AU	$0.4\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	250 AU	NR	✓	540 yrs
	500 AU	$0.1\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	500 AU	430 AU	✗	-
	500 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	500 AU	373 AU	✗	-
	500 AU	$0.3\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	500 AU	NR	✓	680 yrs
	500 AU	$0.4\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	500 AU	NR	✓	445 yrs
	1000 AU	$0.1\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	1000 AU	862 AU	✗	-
	1000 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	1000 AU	758 AU	✗	-
	1000 AU	$0.3\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	1000 AU	NR	✓	650 yrs
	1000 AU	$0.4\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	1000 AU	NR	✓	500 yrs
Additional separation runs:	150 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	150 AU	116 AU	✓✓	750 yrs

	200 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	200 AU	150 AU	✓✓	1500 yrs
	325 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	325 AU	240 AU	✗	-
	400 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	0°	400 AU	299 AU	✗	-
Varying eccentricity:	150 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	113 AU	92 AU	✓✓	900 yrs
	150 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	75 AU	63 AU	✗	-
	150 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	38 AU	39 AU	✗	-
	200 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	150 AU	119 AU	✓✓	1300 yrs
	200 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	100 AU	81 AU	✗	-
	200 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	50 AU	46 AU	✗	-
	250 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	188 AU	147 AU	✓✓	2000 yrs
	250 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	125 AU	102 AU	✓✓	1770 yrs
	250 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	63 AU	54 AU	✗	-
	325 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	244 AU	200 AU	✗	-
	325 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	163 AU	134 AU	✓✓	2750 yrs
	325 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	81 AU	69 AU	✗	-
	400 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	300 AU	236 AU	✗	-
	400 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	200 AU	163 AU	✓✓	3800 yrs
	400 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	100 AU	84 AU	✗	-
	500 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.25	0°	375 AU	295 AU	✗	-
	500 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.5	0°	250 AU	204 AU	✗	-
	500 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0.75	0°	125 AU	104 AU	✗	-
Varying inclination:	100 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	30°	100 AU	77 AU	✗	-
	100 AU	$0.2\,M_\odot$	$1\,M_\odot$	$0.2\,M_\odot$	0	60°	100 AU	74 AU	✓✓	835 yrs

	100 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	90°	100 AU	75 AU	✓✓	835 yrs
	150 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	30°	150 AU	114 AU	✓✓	720 yrs
	150 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	60°	150 AU	115 AU	✓✓	1045 yrs
	150 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	90°	150 AU	116 AU	✓✓	1340 yrs
	200 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	30°	200 AU	151 AU	✓✓	1600 yrs
	200 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	60°	200 AU	152 AU	✓✓	2840 yrs
	200 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	90°	200 AU	153 AU	✗	-
	250 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	30°	250 AU	187 AU	✗	-
	250 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	60°	250 AU	189 AU	✗	-
	250 AU	0.2 $M_\odot$	1 $M_\odot$	0.2 $M_\odot$	0	90°	250 AU	190 AU	✗	-
Varying companion mass:	150 AU	0.2 $M_\odot$	1 $M_\odot$	0.1 $M_\odot$	0	0°	150 AU	110 AU	✓✓	900 yrs
	150 AU	0.2 $M_\odot$	1 $M_\odot$	0.5 $M_\odot$	0	0°	150 AU	112 AU	✓✓	550 yrs
	250 AU	0.2 $M_\odot$	1 $M_\odot$	0.1 $M_\odot$	0	0°	250 AU	181 AU	✗	-
	250 AU	0.2 $M_\odot$	1 $M_\odot$	0.5 $M_\odot$	0	0°	250 AU	197 AU	✓✓	1550 yrs
	325 AU	0.2 $M_\odot$	1 $M_\odot$	0.1 $M_\odot$	0	0°	325 AU	237 AU	✗	-
	325 AU	0.2 $M_\odot$	1 $M_\odot$	0.5 $M_\odot$	0	0°	325 AU	256 AU	✓✓	4000 yrs
	400 AU	0.2 $M_\odot$	1 $M_\odot$	0.1 $M_\odot$	0	0°	400 AU	292 AU	✗	-
	400 AU	0.2 $M_\odot$	1 $M_\odot$	0.5 $M_\odot$	0	0°	400 AU	315 AU	✗	-

Table 4.7 *Summary of all discs simulated here, and whether they did or did not fragment. Black ticks: disc fragmented and so did its reference run analog. Double green ticks: disc fragmented when the reference run analog did not. Black crosses: did not fragment.* $r_{\rm peri,actual}$ *=NR ("not reached") denotes systems where the simulation ended before reaching* $r_{\rm peri,actual}$*, as the disc had already fragmented.*

Chapter 5

AB Aurigae: possible evidence of planet formation through the gravitational instability

This chapter contains work from Cadman et al. (2021). Whilst Cadman et al. (2021) was led by myself, the project was undertaken in collaboration with various listed co-authors, hence some parts of the text may contain contributions from these co-authors.

5.1 Motivation

Owing to recent advances in high resolution infrared (IR) imaging we are now capable of observing the planet formation process taking place. Observations of these discs have revealed substructures indicative of the presence of planetary companions, such as rings (ALMA Partnership et al., 2015b; Andrews et al., 2016; Avenhaus et al., 2018; Bertrang et al., 2018; Dipierro et al., 2018; Huang et al., 2018a), gaps (Andrews et al., 2011; Perez et al., 2015; Ginski et al., 2016; van Boekel et al., 2017) and spirals (Garufi et al., 2013; Grady et al., 2013; Benisty et al., 2015; Pérez et al., 2016; Tang et al., 2017; Huang et al., 2018b; Dong et al., 2018), and recently it has even become possible to directly image giant protoplanets which are still forming (Keppler et al., 2018; Müller et al., 2018; Haffert et al., 2019; Boccaletti et al., 2020). Study of these systems may reveal crucial insights into the underlying physics governing the planet formation

process.

AB Aurigae is a $2.4 \pm 0.2\,\mathrm{M}_{\odot}$, Herbig Ae/Be star (DeWarf et al., 2003), at a distance $d \approx 162.9 \pm 1.5\,\mathrm{pc}$ (Gaia Collaboration et al., 2018). Various authors find an age for the star-disc system to be somewhere between $1 - 4\,\mathrm{Myr}$ (van den Ancker et al., 1997; DeWarf et al., 2003; Piétu et al., 2005). Measurements of the disc surrounding AB Aurigae find an extended, $R_{\mathrm{out}} = 400 - 500\,\mathrm{AU}$, low-mass disc, where $M_{\mathrm{d}} = 0.01\,\mathrm{M}_{\odot}$, with up to a factor ≈ 10 uncertainty on the mass estimate (DeWarf et al., 2003; Andrews & Williams, 2005; Corder et al., 2005; Semenov et al., 2005). The stellar accretion rate of $\dot{M} = 1.3 \times 10^{-7}\,\mathrm{M}_{\odot}\mathrm{yr}^{-1}$ (Salyk et al., 2013) is unusually high for a $1 - 4\,\mathrm{Myr}$ old system, as the depleted disc mass at this late stage limits the available amount of accretable material.

The AB Aurigae disc has been studied extensively owing to its complex substructure, with authors reporting multiple rings (Piétu et al., 2005; Hashimoto et al., 2011; Tang et al., 2012, 2017), bright inner spirals (Piétu et al., 2005), extended CO spirals (Tang et al., 2012), and the possible presence of multiple, planetary-mass companions (Piétu et al., 2005; Tang et al., 2012, 2017). Recent high resolution, scattered light observations of AB Aurigae performed by Boccaletti et al. (2020) using SPHERE provide some of the most spectacular images of a protoplanetary disc to date, revealing detailed spiral features, and placing new constraints on the properties of any potential companions (see Figure 5.1). A kink in the inner spiral at $R \approx 30\,\mathrm{AU}$ is found to be consistent with the presence of a protoplanet with a mass of $4{-}13\,\mathrm{M}_{\mathrm{Jup}}$ (hereafter referred to as planet P1), which is also consistent with conclusions from previous authors (Piétu et al., 2005; Tang et al., 2012, 2017). The authors also report a point-source located at the outer edge of the inner disc, which is characterised by a gas and dust cavity at $R \approx 140\,\mathrm{AU}$, for which they tentatively derive a planetary mass of $3\,\mathrm{M}_{\mathrm{Jup}}$ (hereafter referred to as planet P2). Throughout this chapter we aim to explore the likely formation history of planet P1.

In the core accretion model of giant planet formation (Mizuno, 1980; Pollack et al., 1996), growth proceeds through the steady collisional accumulation of planetesimals onto a rocky core, which may eventually become massive enough for the onset of accretion of a gaseous envelope. Currently this model provides the most popular explanation for the formation of giant planets. However, it has been shown that formation timescales, which may be anywhere up to 10 Myr, may exceed typical disc lifetimes (Haisch, Lada & Lada, 2001). This is especially problematic for giant planets on wide orbits (such as those observed in the

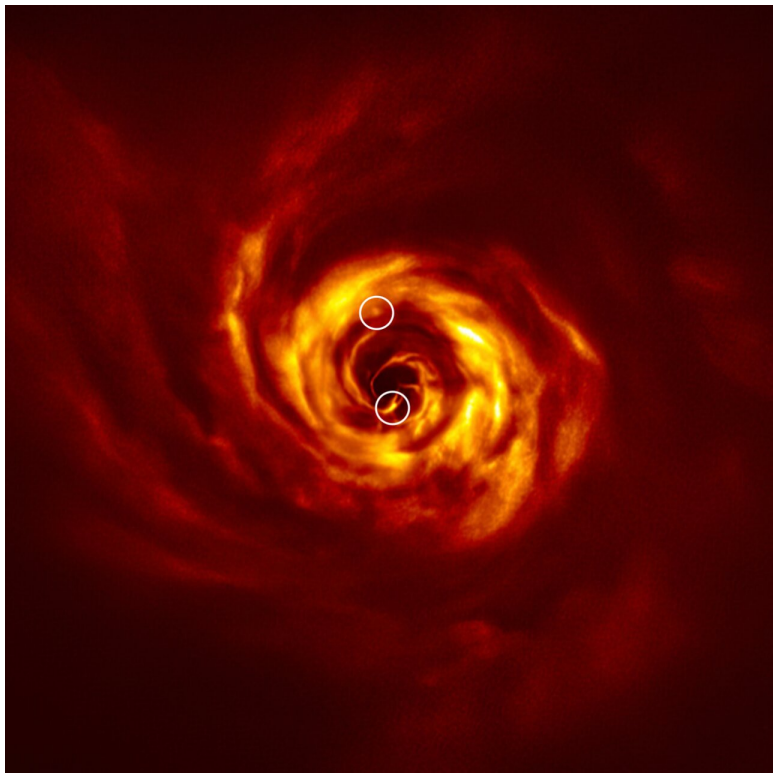

Figure 5.1 *Image of the disc surrounding AB Aurigae obtained using VLT's SPHERE instrument. The locations of the possible protoplanets which have been identified by Boccaletti et al. (2020) are highlighted. Image credit: ESO/Boccaletti et al. (2020).*

AB Aurigae disc) where the planetesimal surface densities will be low, hence formation timescales will be long.

In the GI model of planet formation (Boss, 1997), unstable regions of the disc may directly collapse to rapidly form giant gaseous protoplanets or brown dwarfs. Fragmentation may occur if the disc is sufficiently massive and able to cool quickly, hence the conditions for fragmentation will likely only be satisfied in the outer disc where the gas is optically thin (Clarke, 2009; Rice & Armitage, 2009). Analytical calculations of the Jeans mass in self-gravitating discs (e.g. Section 3.6, Forgan & Rice, 2011, 2013a; Cadman et al., 2020a) suggest that fragmentation may initially form objects with masses between a few and a few 10s of Jupiter masses. Hence wide-orbit, giant planets and brown dwarfs (such as those observed in the AB Aurigae disc) are expected to be the most likely outcome of disc fragmentation. Subsequent dynamical evolution, migration, tidal stripping and growth may then follow, during which the fragment may contract to form a compact planetary/brown dwarf mass object, or be entirely torn apart and destroyed (Nayakshin, 2010a,b, 2011; Forgan & Rice, 2013b; Nayakshin & Fletcher, 2015; Forgan et al., 2018; Humphries et al., 2019).

We find ourselves in a unique position with the protoplanets observed in the AB Aurigae disc, as most of the exoplanets discovered to date have already undergone significant migration and dynamical evolution since their formation making it more challenging to probe their formation histories. The young age of AB Aurigae places strict time constraints on the possible formation histories of the observed planets, thus it provides a good testbed for planet formation theories. In this chapter we focus on the formation history of planet P1 (core accretion vs. GI), and whether the AB Aurigae system could be evidence of planet formation through GI. This chapter is organised as follows. In Section 5.2 we calculate the likely timescale for planet P1 to form through core accretion, and in Section 5.3 we evaluate the possibility that the planet may have formed directly through GI during AB Aurigae's early evolution. We determine the critical disc-to-star mass ratio for fragmentation in Section 5.3.1, and use viscous evolution models in Section 5.3.2 to predict whether the disc may have ever been massive enough to fragment at some point in the recent past. We place new constraints on the current mass of the disc in Section 5.3.2.2, and in Section 5.3.3 we calculate the Jeans mass in a gravitationally unstable, AB Aurigae-like disc. We discuss our results and draw conclusions in Sections 5.4 and 5.5 respectively.

5.2 Formation through core accretion

5.2.1 Core accretion timescale

5.2.1.1 Methods

To model the formation timescale of a gas giant planet through core accretion, we use a similar approach to that outlined in Ida & Lin (2004). We begin by assuming that either a $M_{\rm core,init} = 0.01\,{\rm M}_{\oplus}$ or a $M_{\rm core,init} = 0.1\,{\rm M}_{\oplus}$ core, with density $\rho_{\rm core} = 3.2\,{\rm gcm}^{-2}$, has formed at a semi-major axis, a, which we vary between 5 AU and 50 AU. For simplicity, we consider planet growth in-situ and neglect any migration through the disc, an assumption which is discussed in Section 5.4.

Core growth proceeds at a rate (Safronov, 1969),

$$\dot{M}_{\rm core} = \pi R_c^2 \Sigma_{\rm p} \Omega f_g, \tag{5.1}$$

where R_c is the radius of the core, $\Sigma_{\rm p}$ is the local planetesimal surface density, Ω is the angular frequency and f_g is the gravitational enhancement factor, calculated using the equations from Greenzweig & Lissauer (1992). The local planetesimal surface density, $\Sigma_{\rm p}$, is defined as the surface density of dust within a radial annulus defined by the protoplanet's Hill radius, $R_{\rm H}$, where,

$$R_H = a\left(\frac{M_{\rm p}}{3M_*}\right)^{1/3}, \tag{5.2}$$

where M_* is the mass of the host star and $M_{\rm p}$ is the total planet mass, equal to the sum of the core and envelope masses.

Whilst the core mass is still low, growth initially proceeds through planetesimal accretion, and we update $M_{\rm p}$ using Equation 5.1 at each timestep. A planet may begin to retain a gaseous envelope if the core exceeds the critical mass for the onset of gas accretion, $M_{\rm d,crit}$, where (Ikoma et al., 2000),

$$M_{\rm d,crit} = 10\left(\frac{\dot{M}_{\rm core}}{10^{-6}{\rm M}_{\oplus}{\rm yr}^{-1}}\right)^{0.25}\left(\frac{\kappa}{1{\rm gcm}^{-2}}\right)^{0.25}{\rm M}_{\oplus}, \tag{5.3}$$

where κ is the planetesimal opacity, for which we use $\kappa = 1\,\mathrm{gcm}^{-2}$.

We use a simple approach to calculate the accretion rate of a gaseous envelope onto the core, $\dot{M}_{\mathrm{gas}}$, based on the Kelvin-Helmholtz cooling timescale, τ_{KH}, of the protoplanet, where,

$$\tau_{\mathrm{KH}} = 10^9 \left(\frac{M_{\mathrm{p}}}{\mathrm{M}_{\oplus}} \right)^{-3} \text{years}, \tag{5.4}$$

and,

$$\dot{M}_{\mathrm{gas}} = \frac{M_{\mathrm{p}}}{\tau_{\mathrm{KH}}}. \tag{5.5}$$

This approximation is only valid provided that there is sufficient disc gas present for the planet to accrete, and envelope accretion will cease if the planet is able to deplete all the gas available within its feeding zone. This can be defined in terms of an upper mass limit for in-situ formation, known as the gas isolation mass, $M_{\mathrm{g,iso}}$, where,

$$M_{\mathrm{g,iso}} = 50 \left(\frac{\Sigma_{\mathrm{g}}}{2400 \mathrm{gcm}^{-2}} \right)^{1.5} \left(\frac{a}{1\mathrm{AU}} \right)^{3} \left(\frac{M_*}{\mathrm{M}_{\odot}} \right)^{-1/2} \mathrm{M}_{\oplus}, \tag{5.6}$$

where Σ_{g} is the local gas density. We prevent further growth once $M_{\mathrm{p}} \geq M_{\mathrm{g,iso}}$.

We set up the gas component of the disc with a total mass $M_{\mathrm{gas}} = 0.6\,\mathrm{M}_{\odot}$, hence a disc-to-star mass ratio, $q = 0.25$, and with $\Sigma_{\mathrm{g}} \propto R^{-1}$. The surface density profile of the gas disc is evolved using the one dimensional model outlined in Rice & Armitage (2009), where we assume a radially constant, fixed value for the Shakura-Sunyaev viscous-α of $\alpha = 10^{-3}$ (Shakura & Sunyaev, 1973). The planetesimal component of the disc is set up as,

$$\Sigma_{\mathrm{p}} = f_{\mathrm{dust}} \eta_{\mathrm{ice}} \left(\frac{R}{R_0} \right)^{-1}, \tag{5.7}$$

where f_{dust} is a scale factor such that we set Σ_{p} at 5 AU to be $2\,\mathrm{gcm}^{-2}$, $3\,\mathrm{gcm}^{-2}$, $5\,\mathrm{gcm}^{-2}$ and $10\,\mathrm{gcm}^{-2}$, and η_{ice} is a constant where,

$$\eta_{\rm ice} = \begin{cases} 4.2, & \text{if } a \geq a_{\rm ice} \\ 1, & \text{if } a < a_{\rm ice}, \end{cases} \tag{5.8}$$

and $a_{\rm ice}$ is the ice line located at,

$$a_{\rm ice} = 2.7 \left(\frac{M_*}{\rm M_\odot} \right)^2 \text{AU}. \tag{5.9}$$

In each case, we allow the planets to evolve in the disc for a maximum of 10 Myr.

5.2.1.2 Results

Figure 5.2 shows the resultant planet growth tracks using this formalism, considering the setups with $M_{\rm core,init} = 0.01\,\rm M_\oplus$. Planet formation begins with a phase of core growth, which may either be slow or rapid depending on the local planetesimal surface density. This phase tends to plateau once the local planetesimal surface density is depleted, at which point the planet mass remains approximately constant. The critical core mass for the onset of gas accretion is proportional to the planetesimal accretion rate onto the core, and as the heating from accretion ceases the contraction of a gas envelope may ensue. Wide-orbit, giant planet formation is generally favoured near to, and just beyond the ice line due to the enhancement in the local planetesimal surface density (Equation 5.8). We calculate $a_{\rm ice} \approx 15.6\,\text{AU}$ for a star of mass $2.4\,\rm M_\odot$. If the local planetesimal surface density is particularly high, for example near to the ice line in Figure 5.2c, the core mass may pass straight through the critical mass without plateauing. If the local planetesimal surface density is low, for example at a large semi-major axis in Figure 5.2d, the core may never experience significant growth.

In Table 5.1 we show the results of 32 runs, where we measure the time for the core to have accreted a significant envelope, of mass equal to the core mass ($M_{\rm p} > 2M_{\rm core}$), and to reach a total planetary mass of $M_{\rm P1} = 4\,\rm M_{Jup}$, equal to the lower limit of the estimated mass for planet P1. We find it challenging to produce a planet of at least $4\,\rm M_{Jup}$ in an AB Aurigae-like disc in $\lesssim 1 - 4\,$Myr. In the majority of setups considered here the planet will either reach its isolation mass before reaching the mass of planet P1, as seen in Figure 5.2a, or will not grow rapidly enough to reach $M_{\rm P1}$ within the duration modelled here. To rapidly form

a planet this massive generally requires a significant core has initially formed, in a disc with an extremely high planetesimal surface density, with a planet on a shorter orbit than where planet P1 is currently located.

The planetesimal surface densities considered here, with $\Sigma_{\mathrm{P,5AU}} = 2\,\mathrm{gcm}^{-2}$, $3\,\mathrm{gcm}^{-2}$, $5\,\mathrm{gcm}^{-2}$ and $10\,\mathrm{gcm}^{-2}$ correspond to total planetesimal masses across the disc of $0.012\,\mathrm{M}_{\odot}$, $0.024\,\mathrm{M}_{\odot}$, $0.048\,\mathrm{M}_{\odot}$ and $0.072\,\mathrm{M}_{\odot}$. These are equivalent to initial dust-to-gas ratios of 0.01, 0.02, 0.04 and 0.06 if we assume that the dust and gas in the disc decoupled when the gas mass was $1.2\,\mathrm{M}_{\odot}$ (therefore when $q = 0.5$), and that all of this dust then went on to form planetesimals. However, if any of the dust was depleted by some other mechanism or did not go on to form planetesimals, which would likely be the case, then the planetesimal surface densities used here would demand much higher initial dust-to-gas ratio prior to decoupling. Therefore given the generously high planetesimal surface densities used here, we would consider these core accretion timescales as optimistic lower limits to what we might expect in a realistic disc.

Mechanisms which may be capable of speeding up these core accretion timescales and have not been included in the models here, such as pebble accretion, are discussed in Section 5.4.2.

5.3 Formation through the Gravitational Instability

5.3.1 Critical mass limit for fragmentation

5.3.1.1 Methods

We use the PHANTOM (Price et al., 2018) SPH code to determine the critical mass limit for fragmentation in a disc around a $2.4\,\mathrm{M}_{\odot}$ star, analogous to AB Aurigae. We represent the disc with $N = 1 \times 10^{6}$ SPH particles, distributed between $R_{\mathrm{in}} = 2.5\,\mathrm{AU}$ and $R_{\mathrm{out}} = 400\,\mathrm{AU}$ with a surface density profile $\Sigma \propto R^{-1}$ and sound speed profile $c_{\mathrm{s}} \propto R^{-0.25}$. We modify PHANTOM such that we model radiative cooling using the hybrid radiative transfer method (Forgan et al., 2009) outlined in Section 2.4.1.4. We assume that disc irradiation leads to a constant background temperature, which we represent as, $T_{\mathrm{irr}} = 10\,\mathrm{K}$. Artificial

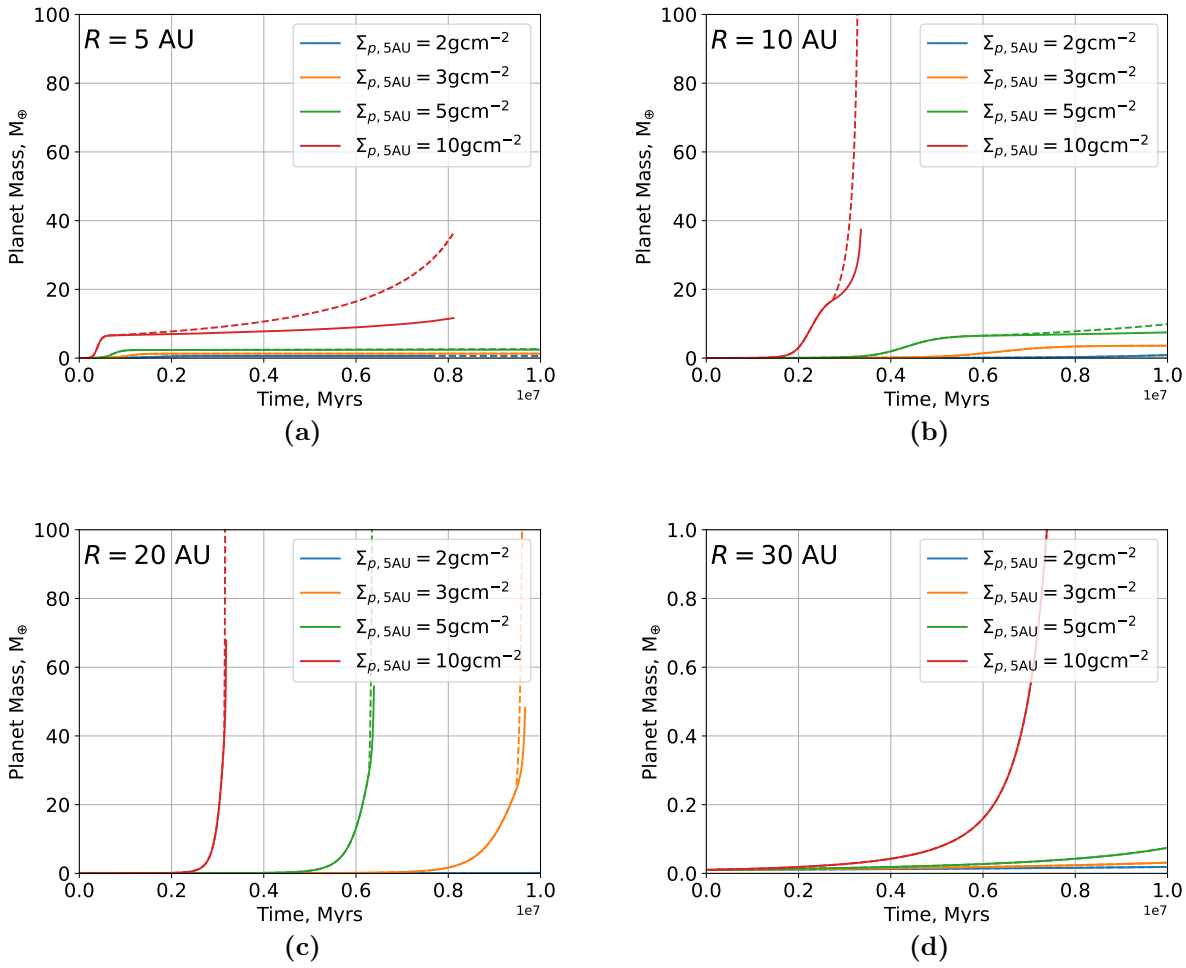

Figure 5.2 *Evolution of planet core masses (solid line) and core + envelope masses (dashed lines) for in-situ core accretion planet formation at radii $R = 5\,AU$ (5.2a), $10\,AU$ (5.2b), $20\,AU$ (5.2c) and $30\,AU$ (5.2d) from the stellar host. In each case the models begin with an initial core mass $M_{\rm core,init} = 0.01\,M_{\oplus}$ at $t = 0$. We vary the planetesimal surface densities in the disc such that $\Sigma_{\rm P,5AU} = 2\,gcm^{-2}$, $3\,gcm^{-2}$, $5\,gcm^{-2}$ and $10\,gcm^{-2}$, which correspond to total planetesimal masses across the disc of $0.012\,M_{\odot}$, $0.024\,M_{\odot}$, $0.048\,M_{\odot}$ and $0.072\,M_{\odot}$.*

$M_{core,init}$ $M_\oplus$ (1)	R AU (2)	$\Sigma_{p,5AU}$ gcm^{-2} (3)	t_{env} Myr (4)	t_{P1} Myr (5)
0.01	5	2	-	-
0.01	5	3	-	-
0.01	5	5	-	-
0.01	5	10	6.43	-
0.01	10	2	-	-
0.01	10	3	-	-
0.01	10	5	-	-
0.01	10	10	3.14	-
0.01	20	2	-	-
0.01	20	3	9.56	-
0.01	20	5	6.33	-
0.01	20	10	3.16	3.18
0.01	30	2	-	-
0.01	30	3	-	-
0.01	30	5	-	-
0.01	30	10	8.77	8.82
0.1	10	2	-	-
0.1	10	3	-	-
0.1	10	5	-	-
0.1	10	10	1.88	-
0.1	20	2	6.81	-
0.1	20	3	3.95	4.05
0.1	20	5	2.60	2.66
0.1	20	10	1.30	1.32
0.1	30	2	-	-
0.1	30	3	9.92	-
0.1	30	5	6.61	6.72
0.1	30	10	3.33	3.38
0.1	50	2	-	-
0.1	50	3	-	-
0.1	50	5	-	-
0.1	50	10	-	-

Table 5.1 *Results of the core accretion models. (1) Initial core mass. (2) Semi-major axis of core. (3) Planetesimal surface density at 5 AU. (4) Time before the planet reaches runaway growth, where the envelope mass exceeds the core mass. (5) Time before the planet mass reaches 4 M_{Jup}*

disc viscosity is modelled using the standard $\alpha - \beta$ viscosity prescription, where we use $\alpha_{\rm SPH} = 0.1$ and $\beta_{\rm SPH} = 0.2$.

Each disc is allowed to evolve for a maximum of $t = 15,550\,\rm yrs$, equal to 3 orbital periods at $R_{\rm out} = 400\,\rm AU$, or until fragments form and the computational timestep becomes prohibitively long for the simulations to continue. We calculate the thermalisation timescale, $t_{\rm therm,i}$, from Forgan et al. (2009) for each of our disc final states, which represents the time for the disc material to reach thermal equilibrium. We find that in the discs that don't fragment within $15,550\,\rm yrs$, $\max(t_{\rm ttherm,i}) \ll 1\,\rm kyr$. It is therefore reasonable to assume that if these discs have not fragmented by this point, they will likely not do so in the future.

5.3.1.2 Results

Final states of these SPH simulations are shown in Figure 5.3, where we vary the initial disc mass between $0.2\,\rm M_\odot - 0.35\,\rm M_\odot$, which correspond to disc-to-star mass ratios $q = 0.08 - 0.15$.

The final states of these disc models show that we expect a disc similar to AB Aurigae to fragment and form multiple clumps if $M_{\rm d} \geq M_{\rm d,crit} = 0.3\,\rm M_\odot$ ($q \geq 0.125$), and to display non-axisymmetric substructure if $M_{\rm d} \geq 0.25\,\rm M_\odot$ ($q \geq 0.1$). For $M_{\rm d} \leq 0.2\,\rm M_\odot$ ($q \leq 0.08$), it is unlikely that the gravitational instability will lead to the growth of significant spirals and, in the absence of a perturber, it should be almost entirely axisymmetric. Therefore given the current low mass state of the AB Aurigae disc we predict that it should be gravitationally stable, as expected.

When also considering a set of discs with outer radii $R_{\rm out} = 300\,\rm AU$ and $R_{\rm out} = 500\,\rm AU$ we find that this critical disc-to-star mass ratio has some dependence on disc size, with more extended discs being more stable. When $R_{\rm out} = 500\,\rm AU$ we find the critical disc mass for fragmentation to be $M_{\rm d,crit} = 0.35\,\rm M_\odot$ ($q_{\rm crit} = 0.15$), and when $R_{\rm out} = 300\,\rm AU$ we find $M_{\rm d,crit} = 0.3\,\rm M_\odot$ ($q_{\rm crit} = 0.125$).

5.3.1.3 Subsequent migration of the clumps

Fragmentation will only occur if the disc is able to radiate energy away at a rate faster than the clump will collapse, hence primarily operates at large radii from the central star where the disc opacity is low thus it can cool efficiently. In

the disc with $q = 0.125$, the fragment forms at $a \approx 200\,\mathrm{AU}$, much further out than the current semi-major axis of planet P1. 2D hydrodynamical simulations indicate that once fragments form in a gravitationally unstable disc they will rapidly migrate to the inner regions within a few orbital periods (Baruteau et al., 2011). Computation times become prohibitively long for us to model the long-term migration of clumps using SPH, as to resolve the high densities at the clump centres requires long integration times. Instead we can use the analytic calculations from Nayakshin (2010a) to get an idea for the approximate migration timescales of a protoplanet formed in an AB Aurigae-like disc. For type I migration, the time to move between radii a_{out} to a_{in} will be,

$$\Delta t_{\mathrm{mig,I}} = \int_{a_{\mathrm{out}}}^{a_{\mathrm{in}}} \frac{t_{\mathrm{mig,I(a)}}}{a} da, \tag{5.10}$$

where,

$$t_{\mathrm{mig,I}}(a) = \left(\frac{M_{\mathrm{p}}}{M_*}\Omega\right)^{-1} \frac{H}{a}, \tag{5.11}$$

and for type II migration,

$$\Delta t_{\mathrm{mig,II}} = \int_{a_{\mathrm{out}}}^{a_{\mathrm{in}}} \frac{t_{\mathrm{mig,II(a)}}}{a} da, \tag{5.12}$$

where,

$$t_{\mathrm{mig,II}}(a) = \frac{1}{\alpha\Omega}\left(\frac{H}{a}\right)^{-2}, \tag{5.13}$$

where H is the disc scale height at $R = a$.

Whether a planet is in the type I or type II regime can be established in terms of a transition mass, M_t, which roughly corresponds to the mass at which protoplanets become capable of gap-opening. For $M \leq M_t$ (lower-mass, faster migrating protoplanets) the planet will be in the type I regime, and for $M \geq M_t$ (higher-mass, slower migrating protoplanets) the planet will be in the type II regime, where,

$$M_t = 2M_*\left(\frac{H}{R}\right)^3. \tag{5.14}$$

We can calculate the time for planet P1 to migrate from $a_{\rm out} = 200\,{\rm AU}$ to $a_{\rm in} = 30\,{\rm AU}$, by substituting $M_* = 2.4\,{\rm M}_\odot$, $M_{\rm p} = 4\,{\rm M}_{\rm Jup}$, $\alpha = 0.06$ for a saturated disc, and calculating the azimuthally averaged disc scale height, taken from the SPH disc where $M_{\rm d} = 0.3\,{\rm M}_\odot$. Integrating Equations 5.10 and 5.12 we calculate $\Delta t_{\rm mig,I} = 6.9\,{\rm kyr}$ and $\Delta t_{\rm mig,II} = 1.0\,{\rm Myr}$. Note that the value of α used here should be considered an upper limit as α will decrease as the planet migrates. Thus the calculated $t_{\rm mig,II}$ would be a lower limit.

From Equation 5.14 we calculate the transition mass for gap opening to be $M_t = 2.4\,{\rm M}_{\rm Jup}$, which would place planet P1 comfortably in the type II regime. Baruteau et al. (2011) however suggest that GI protoplanets will migrate inwards much faster than the gap opening timescale, and that their migration may be better explained in the type I regime. $\Delta t_{\rm mig,I}$ and $\Delta t_{\rm mig,II}$ are likely more representative of lower and upper limits on the migration timescale of planet P1, and the subsequent migration of a GI protoplanet will be best explained by a combination of both regimes. In either case, these simple calculations demonstrate that, to a first approximation, it should be entirely possible for a fragment formed on a wide orbit to migrate inward to the current location of planet P1 within the current lifetime of the AB Aurigae disc.

5.3.2 Viscous evolution models of AB Aurigae

Despite AB Aurigae's disc mass being too low to be gravitationally unstable currently, it will likely have been much more massive in the past prior to mass depletion through stellar accretion and photoevaporative winds, as massive discs will rapidly evolve away from an initially high mass state (Hall et al., 2019). Viscous evolution models use analytic prescriptions to calculate the evolution history of a protoplanetary disc's surface density profile. We use them here to calculate the likely mass evolution history of AB Aurigae, and predict whether the disc mass may have previously exceeded the critical mass for fragmentation found in Section 5.3.1.

5.3.2.1 Methods

Full details of the model used here to calculate the evolution of a disc whose primary source of viscosity is provided by self-gravity can be found in Rice & Armitage (2009). We also outline the basic equations here.

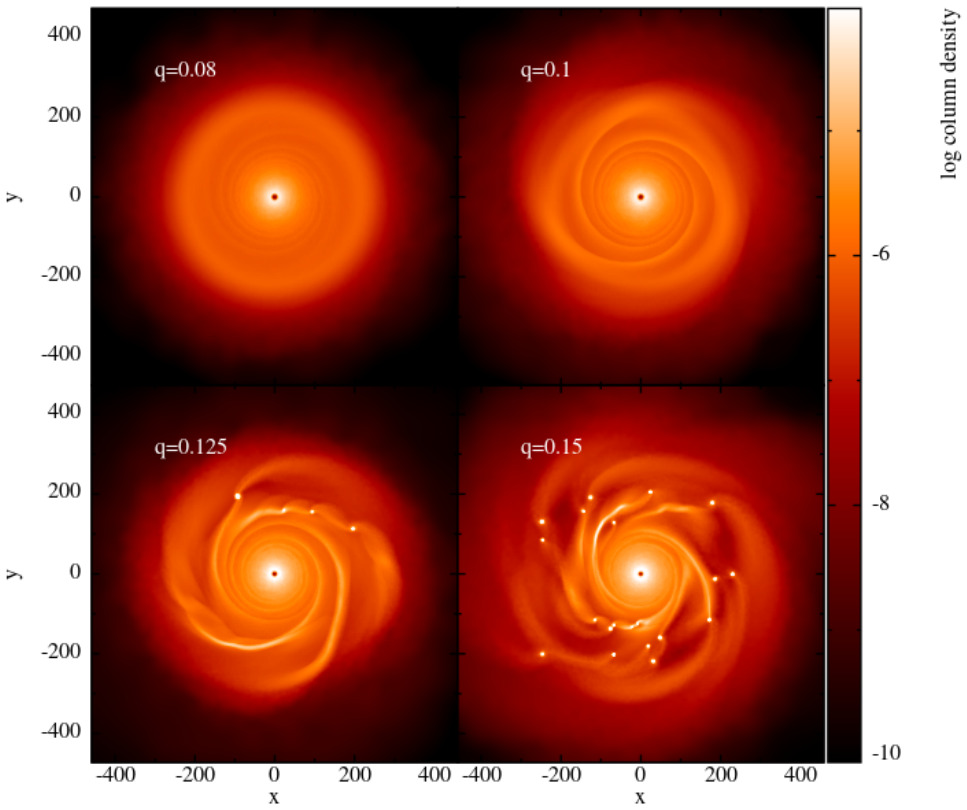

Figure 5.3 *SPH models of an AB Aurigae-like disc. Each disc is set up with* $M_* = 2.4\,M_\odot$, $R_{\rm out} = 400\,AU$, $N = 1 \times 10^6$ *and* $\Sigma \propto R^{-1}$, $c_{\rm s} \propto R^{-0.25}$. *We vary the disc-to-star mass ratios within the range* $q = 0.08 - 0.15$ *(*$M_{\rm d} = 0.2 - 0.35\,M_\odot$*). We find the critical disc-to-star mass ratio for fragmentation in an AB Aurigae-like disc to be* $q_{\rm crit} = 0.125$ *(*$M_{\rm d,crit} = 0.3\,M_\odot$*).*

Viscous evolution of the surface density, $\Sigma(r,t)$, can be modelled using the one-dimensional prescription from Lynden-Bell & Pringle (1974); Pringle (1981), which is also derived in Section 2.3.1,

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right)\right] - \dot{\Sigma}_{\rm wind}, \tag{5.15}$$

where we now also include $\dot{\Sigma}_{\rm wind}$ which represents the photoevaporative mass loss due to irradiation from the central star. Here we implement the x-ray photoionization model described in detail in Owen et al. (2011) and assume a moderate x-ray luminosity of $1 \times 10^{30}\,{\rm erg\,s^{-1}}$, noting that the rate of photoevaporative mass loss scales linearly with x-ray luminosity. Disc viscosity, $\nu(r,t)$, is modelled using the Shakura-Sunyaev viscous$-\alpha$ prescription (Equation 2.86; Shakura & Sunyaev, 1973). The sound speed, $c_{\rm s}$, is calculated by solving Equation 1.21, where we force the disc to be in a marginally unstable state with $Q = 1.5$.

The volume density can then be calculated as $\rho = \Sigma/2H$, and the temperature, T, optical depth, τ, and ratio of specific heats, γ, can be determined by interpolation of the equation of state table from Stamatellos et al. (2007) using the Rosseland mean opacities from Bell & Lin (1994).

To calculate the viscous-α term from Equation 2.86, we must first determine the disc cooling time, which requires that we calculate the radiative cooling term (Hubeny, 1990),

$$\Lambda = \frac{16\sigma}{3}(T^4 - T_{\rm irr}^4)\frac{\tau}{1+\tau^2}, \tag{5.16}$$

and determine the local cooling time as $t_{\rm cool} = U/\Lambda$, where the energy per unit surface area is,

$$U = \frac{c_{\rm s}^2\Sigma}{\gamma(\gamma-1)}. \tag{5.17}$$

In a disc where the primary source of viscosity comes from self-gravity, the effective viscous-α term can be calculated using Equation 1.28. We set a lower limit, $\alpha_{\rm min}$, below which we assume that GI is no longer the dominant source of viscosity but instead, in a sufficiently ionized disc, MRI may dominate, for

example. If $\alpha < \alpha_{\rm min}$ we set $\alpha = \alpha_{\rm min}$ and recalculate the disc properties, now no longer requiring the disc to be gravitationally unstable with $Q = 1.5$.

Equations 1.21 and 1.28 can then be solved to calculate $c_{\rm s}$ and α, from which we can use Equation 2.86 to calculate ν. Equation 5.15 can then be integrated to determine the time evolution of the disc's surface density profile, hence its mass evolution.

5.3.2.2 A note on the current mass of the AB Aurigae disc

Protoplanetary disc masses are notoriously challenging to measure. They often rely on empirical conversions between a disc's flux density and its mass, which require uncertain assumptions about the disc optical depth, metallicity, dust-to-gas ratio and grain size distribution. Combined with uncertainties in the flux measurement and distance toward the system, mass estimates may be uncertain by up to an order of magnitude, and are usually considered to represent lower bounds. Estimates of the disc mass surrounding AB Aurigae find a low mass disc, with $M_{\rm d} = 0.01\,{\rm M}_{\odot}$ and uncertainty up to a factor ≈ 10 (Andrews & Williams, 2005; Corder et al., 2005; Piétu et al., 2005; Semenov et al., 2005).

The accretion rate onto the star may also provide us a with rough estimate of the disc mass, as it is indicative of the mass reservoir available to the star from the disc. A protoplanetary disc is expected to settle into a steady-state with a constant mass accretion rate given by Equation 3.1. This can be rearranged to show that a disc with sound speed profile, $c_{\rm s} = c_{\rm s,0} R^{-0.25}$, and surface density profile, $\Sigma = \Sigma_0 R^{-1}$, may have a radially constant viscous-α given by,

$$\alpha = \frac{1}{3\pi} \frac{\dot{M}\sqrt{GM_*}}{c_{\rm s,0}^2 \Sigma_0}. \tag{5.18}$$

We can substitute in for $c_{\rm s,0}$ by assuming a flattened disc with $H/R = 0.1$ at $R = 100\,{\rm AU}$, and substituting $H = c_{\rm s}/\Omega$, where H is the local disc scale height. Similarly, we can substitute Σ_0 for the disc outer radius, $R_{\rm out} = 400\,{\rm AU}$ and disc mass to obtain an equation in terms of α and $M_{\rm d}$,

$$\alpha = \frac{200}{3} \frac{\dot{M} R_{\rm out}}{\sqrt{{\rm G}M_*}} \frac{\sqrt{100{\rm AU}}}{M_{\rm d}}. \tag{5.19}$$

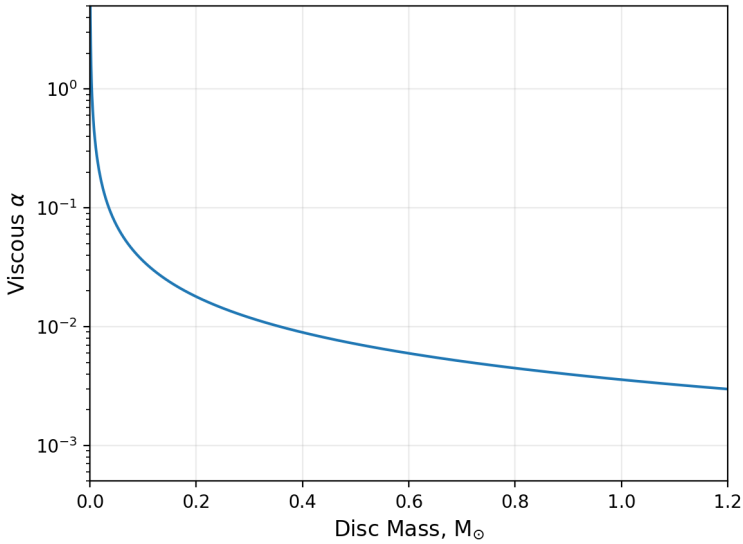

Figure 5.4 *Viscous−α vs. disc mass for a steady-state disc with $\dot{M} = 1.3 \times 10^{-7}\,M_\odot yr^{-1}$, equal to the mass accretion rate measured in AB Aurigae. We calculate α as a function of disc mass using Equation 5.19, which assumes that the disc has a radially constant viscous−α.*

We plot this equation in Figure 5.4 for a star of mass $2.4\,\mathrm{M}_\odot$ and mass accretion rate $\dot{M} = 1.3 \times 10^{-7}\,\mathrm{M}_\odot\mathrm{yr}^{-1}$ (Salyk et al., 2013). In Table 5.2 we show calculated disc masses corresponding to α values 0.1, 0.01 and 0.001.

From Figure 5.4 we see that for a very-low mass disc ($M_\mathrm{d} \leq 0.1\,\mathrm{M}_\odot$) to have an accretion rate $\dot{M} = 1.3 \times 10^{-7}\,\mathrm{M}_\odot\mathrm{yr}^{-1}$ would require a disc viscosity much higher than we would usually expect from a quasi-stable disc, with $\alpha \geq 0.1$. If instead the disc is still massive, with $M_\mathrm{d} \geq 0.1\,\mathrm{M}_\odot$, the viscous-$\alpha$ required to explain the high accretion rate decreases significantly. In a quasi-stable disc we might typically expect $\alpha \approx 10^{-2} - 10^{-4}$ (Hartmann et al., 1998; Rafikov, 2017).

We do not attempt to propose an exact disc mass for AB Aurigae here, but instead wish to highlight that in order to explain the system's high accretion rate may require that the disc is more massive than has previously been suggested, and that it is likely at least as massive as the upper bound of the current disc mass estimates.

5.3.2.3 Results

With this in mind we use these viscous evolution models to predict how long ago the AB Aurigae disc may have been massive enough to exceed the critical mass limit for fragmentation, where $M_{\rm d} = M_{\rm d,crit} = 0.3\,{\rm M}_{\odot}$, assuming the system to have a current disc mass approximately equal to the upper bound on the mass estimate, $M_{\rm d} = 0.1\,{\rm M}_{\odot}$.

In order to do this, we set up discs with initial masses $M_{\rm d} = M_{\rm d,crit} = 0.3\,{\rm M}_{\odot}$ and evolve them forward in time until their mass has been depleted to $M_{\rm d} = 0.1\,{\rm M}_{\odot}$, assuming $\alpha_{\rm min}$ values in the range $0.01 - 0.05$. Discs are set up with initial parameters similar to what we might expect in a young AB Aurigae disc, with $M_{*} = 2.4\,{\rm M}_{\odot}$, $R_{\rm out,init} = 400\,{\rm AU}$, surface density profile $\Sigma \propto R^{-1}$ and temperature profile $T \propto R^{-0.75}$. We assume again that irradiation leads to a constant background temperature $T_{\rm irr} = 10\,{\rm K}$.

The results of these models are shown in Figure 5.5. To illustrate how long in the recent past the AB Aurigae disc may have been massive enough to exceed the fragmentation threshold, we have plotted the disc mass evolution in reverse order. Hence $t = 0$ represents the disc in its current state, with $M_{\rm d} = 0.1\,{\rm M}_{\odot}$, and the x-axis measures Myrs in the past. For example, in the case of the $\alpha_{\rm min} = 0.05$ model, we predict that the AB Aurigae disc may have been more massive than $M_{\rm d,crit} = 0.3\,{\rm M}_{\odot}$ approximately 1.3 Myrs ago.

Discs with higher viscous$-\alpha$ values will evolve at a faster rate, hence the time between the disc mass being in its current state, with $M_{\rm d} = 0.1\,{\rm M}_{\odot}$, and exceeding the critical mass limit, $M_{\rm d,crit} = 0.3\,{\rm M}_{\odot}$, will be shorter.

These models again reiterate how it is challenging to reconcile the current low estimated disc mass with the high measured accretion rate, leading us to conclude that AB Aurigae is either currently more massive than observations suggest, or was almost certainly so in its recent past. In the most highly accreting case,

α	$M_{\rm d}$ (${\rm M}_{\odot}$)
0.1	0.04
0.01	0.36
0.001	> 1.2

Table 5.2 *Disc masses corresponding to* $\alpha = 0.1, 0.01$ *and* 0.001 *in Figure 5.4.*

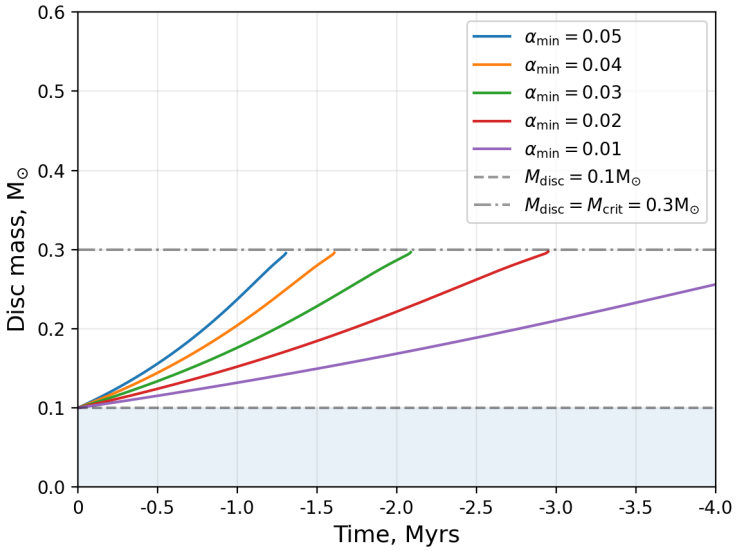

Figure 5.5 *Mass evolution of a disc similar to AB Aurigae, calculated using the viscous evolution models outlined in Section 5.3.2.1. The plot begins with a disc mass equal to the current mass of AB Aurigae, with* $M_{\rm d} = 0.1\,M_{\odot}$ *at* $t = 0$*, and illustrates how long in the recent past the disc mass may have exceeded the critical mass limit for fragmentation,* $M_{\rm d,crit} = 0.3\,M_{\odot}$*. Hence the x-axis measures Myrs in the past. We vary the value of* $\alpha_{\rm min}$*, which represents a background viscous-*α *value generated by some process other than disc self-gravity.*

with $\alpha_{\rm min} = 0.05$, we find the accretion rate when $M_{\rm d} = 0.1\,{\rm M}_{\odot}$ to be $\dot{M} = 4.80 \times 10^{-8}\,{\rm M}_{\odot}{\rm yr}^{-1}$, and in the lowest accreting case with $\alpha_{\rm min} = 0.01$ we find $\dot{M} = 9.0 \times 10^{-8}\,{\rm M}_{\odot}{\rm yr}^{-1}$, both of which are significantly lower than the currently measured value of $\dot{M} = 1.3 \times 10^{-7}\,{\rm M}_{\odot}{\rm yr}^{-1}$ (Salyk et al., 2013).

Crucially though, the plots in Figure 5.5 demonstrate how we can trace the AB Aurigae disc back to a previously higher mass state, and how the disc mass may have exceeded the fragmentation threshold in the recent past. When assuming a moderate background $\alpha_{\rm min}$ we find that the disc mass may have exceeded $M_{\rm d,crit}$ within the past $\approx 1.25 - 4\,{\rm Myr}$ (i.e. within the current estimated lifetime of the disc). Thus it is plausible that a young AB Aurigae disc may have fragmented to form one or multiple giant gaseous protoplanets during its early evolution.

5.3.3 Jeans mass in an AB Aurigae-like disc

The local Jeans mass in a self-gravitating disc can be used, in the case where a region of the disc fragments, to estimate the masses of the bound clumps that will form. In Chapter 3 and Cadman et al. (2020a) we derived a revised equation for the Jeans mass in an irradiated self-gravitating disc, which is stated in Equation 3.11. We can use this equation, and use the same approach as in Section 5.3.2.1 to calculate how the Jeans mass varies as a function of $\dot{M}$ and $R_{\rm out}$. We assume the disc to be marginally unstable, with $Q = 1.5$, and use Equation 1.21 to obtain $c_{\rm s}$, and solve equation 1.28 to obtain α for use in Equation 3.1, allowing us to calculate the Jeans mass for a range of disc outer radii and accretion rates.

In Figure 5.6 we plot Equation 3.11, for a disc around a $2.4\,{\rm M}_{\odot}$ star, with $\dot{M}$ between $1 \times 10^{-9}\,{\rm M}_{\odot}{\rm yr}^{-1}$ and $1 \times 10^{-4}\,{\rm M}_{\odot}{\rm yr}^{-1}$, and $R_{\rm out}$ between 50 AU and 500 AU. We assume that disc irradiation leads to a constant background temperature, and consider two cases where $T_{\rm irr} = 10\,{\rm K}$ and $T_{\rm irr} = 50\,{\rm K}$. Higher disc temperatures reduce the effective viscous$-\alpha$, whilst also providing greater pressure support against direct collapse, thus stabilising the system against GI. Hence, for a given $\dot{M}$ and $R_{\rm out}$ the Jeans mass increases as a function of irradiation.

A gravitationally unstable disc may fragment if a collapsing clump is able to cool and radiate energy away at a rate faster than the local dynamical time. This condition can be expressed in terms of a critical value of the dimensionless cooling parameter, $\beta_c = t_{\rm cool}\Omega$, which in turn can be expressed in terms of a critical viscous$-\alpha$. We typically expect this value to be somewhere between $\alpha_{\rm crit} \approx 0.06 - 0.1$ (Gammie, 2001; Rice et al., 2005; Baehr et al., 2017), thus we include contours of $\alpha = 0.01$ and $\alpha = 0.1$ in Figure 5.6 to indicate regions of the parameter space that may fragment.

These plots reiterate that at an earlier stage of AB Aurigae's evolution, when the mass accretion rate was likely higher than it currently is, it is entirely plausible that the disc may have been gravitationally unstable and may have fragmented, as these higher accretion rate states lie in an unstable region of parameter space.

For a given disc radius the minimum Jeans mass doesn't vary much whether we assume fragmentation can only occur for $\alpha \geq 0.01$ or $\alpha \geq 0.1$. Assuming that $\alpha_{\rm crit} = 0.1$ we find the minimum Jeans masses at $R = 200\,{\rm AU}$, 300 AU and 400 AU to be $1.6\,{\rm M}_{\rm Jup}$, $2.5\,{\rm M}_{\rm Jup}$ and $3.4\,{\rm M}_{\rm Jup}$ respectively when $T_{\rm irr} = 10\,{\rm K}$, and

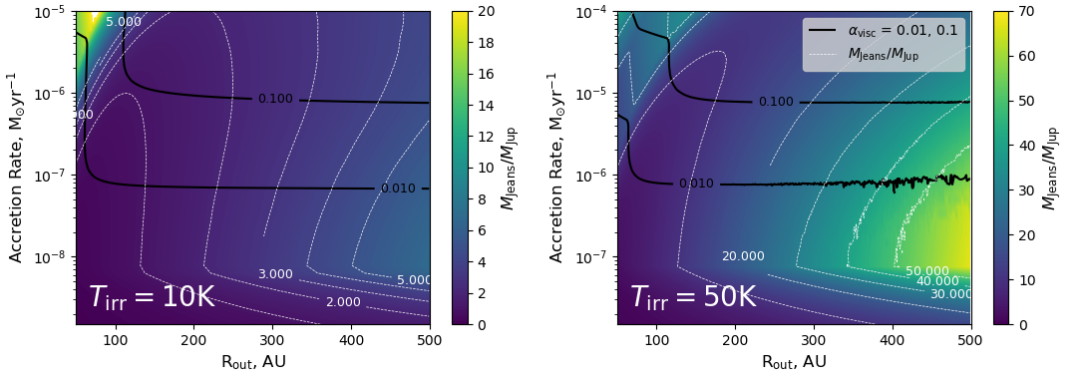

Figure 5.6 *Jeans mass in a self-gravitating disc surrounding a 2.4 $M_\odot$ star. We consider two values for the disc irradiation temperature,$T_{\rm irr} = 10\,K$ (left) and $T_{\rm irr} = 50\,K$ (right). We expect a disc to be unstable against fragmentation for $\alpha_{\rm crit} \approx 0.06 - 0.1$, thus we plot contours of $\alpha = 0.01$ and $\alpha = 0.1$ to indicate regions of parameter space which would likely be unstable against fragmentation.*

to be $10.3\,\mathrm{M_{Jup}}$, $12.4\,\mathrm{M_{Jup}}$ and $13.3\,\mathrm{M_{Jup}}$ respectively when $T_{\rm irr} = 50\,\mathrm{K}$, roughly coinciding with what we observe from the mass of planet P1.

5.4 Discussion

5.4.1 Implications for formation through core accretion

Significant fine tuning of the model parameters is required in Section 5.2 to form planet P1 through core accretion within the strict time constraint of the system's measured age. To form a planet of $4\,\mathrm{M_{Jup}}$ within $1 - 4\,\mathrm{Myr}$ generally requires a planetesimal surface density much higher than would usually be expected, with a total planetesimal mass across the disc $\geq 0.072\,\mathrm{M_\odot}$ when $\mathrm{M_{core,init}} = 0.01\,\mathrm{M_\oplus}$, and $\geq 0.024\,\mathrm{M_\odot}$ when $\mathrm{M_{core,init}} = 0.1\,\mathrm{M_\oplus}$. When $\Sigma_{\rm p,5AU} = 2\,\mathrm{gcm^{-2}}$, hence with a total planetesimal mass across the disc of $0.012\,\mathrm{M_\odot}$, we generally see very slow planet growth.

It is important to note however that we have only considered a simple formalism for our modelling of core accretion here, and that processes not included in our models, such as planet migration, pebble accretion and disc instabilities, may be capable of accelerating initial growth. We discuss the effect of these next.

5.4.2 Limitations of the core accretion models

Migration allows the planet to sample a wider region of the disc, therefore preventing the local planetesimal surface density becoming depleted as rapidly as when it grows in-situ. When we include core migration in Section 5.2 the planets generally grow at a faster rate. However we chose to only consider in-situ formation here, as including migration causes all the cores to migrate to the inner disc ($a \lesssim 3\,\mathrm{AU}$) away from the location where we currently find planet P1, and toward the regions of higher planetesimal surface density where they accrete at a faster rate. Some other mechanism, such as planet-planet scattering, would then be required to explain planet P1's subsequent migration back out to $a \approx 30\,\mathrm{AU}$. When modelling in-situ formation at the current semi-major axis of planet P1, we see only slow growth when $M_{\mathrm{core,init}} = 0.1\,\mathrm{M}_{\oplus}$, and almost no growth when $M_{\mathrm{core,init}} = 0.01\,\mathrm{M}_{\oplus}$.

Instabilities in discs may be capable of generating large over-densities of solids, hence they have been suggested as possible mechanisms for accelerated planetesimal growth and, in extreme cases, fragmentation of the disc solids under their self-gravity. The spiral arms of young, GI discs have been shown to cause strong dust-trapping (Rice et al., 2004), whilst the gravitational collapse of filaments generated in the streaming instability (Youdin & Goodman, 2005; Youdin & Johansen, 2007) may form planetesimals of radii $100 - 1000\,\mathrm{km}$ (Johansen et al., 2007, 2011, 2012), thus providing a possible mechanism for the initial formation of rocky cores. Whilst refraining from modelling the detailed physics of dust trapping through disc instabilities, we can crudely represent local grain enhancements by simply increasing the total dust-to-gas ratio in the disc, which by default will increase the planetesimal surface density local to the accreting core. We account this by increasing the total planetesimal surface density by up to a factor of 6 in Section 5.2.

Mechanisms for accelerated growth and rapid core formation become necessary as core accretion faces challenges when establishing how the first planetesimals are able to grow beyond metre sizes. The initial stages of growth are believed to be slow, as dust grains may encounter growth barriers beyond metre sizes (Brauer et al., 2008; Mordasini et al., 2010). It has been shown, as consequence of intrinsic gas-dust drag in the disc, that grains of a critical size will radially migrate and be accreted onto the star within a fraction of the disc lifetime (the radial drift barrier, Weidenschilling, 1977). Further, solids of millimetre to centimetre sizes, with

Stokes number close to 1, are expected to have high relative azimuthal velocities, hence grain-grain collisions may become destructive, resulting in shattering (the fragmentation barrier, Birnstiel et al., 2012), or neutral and result in recoiling (the bouncing barrier, Zsom et al., 2010), both of which prevent a positive outcome of coagulation. In our model we assume that a core of mass $0.01\,\mathrm{M}_\oplus$ or $0.1\,\mathrm{M}_\oplus$, with $R_{\mathrm{core,init}} = 1.6 \times 10^3\,\mathrm{km}$ and $R_{\mathrm{core,init}} = 3.5 \times 10^3\,\mathrm{km}$ respectively, has already formed at $t = 0$, therefore avoiding the detailed physics of this initial phase of core growth. Note that these initial core sizes are consistent with, but slightly larger than, the planetesimals expected to form through direct collapse of the dust disc during the streaming instability (Johansen et al., 2007, 2011, 2012).

Possibly most importantly, we note that we do not include a prescription for pebble accretion in our model (for a review see Johansen & Lambrechts, 2017). Accretion of millimetre to centimetre sized pebbles onto planetesimal cores may have the potential to generate significantly faster growth rates than the planetesimal-planetesimal accretion we consider here. Pebbles are thought to be abundant in protoplanetary discs, since it is a natural outcome from the fragmentation and bouncing barriers. Pebbles of millimetre-centimetre sizes are coupled to the gas in the disc. The gas component orbits at sub-Keplerian velocities due to the outward gas pressure. The solids, which are orbiting at Keplerian velocities, will experience a drag force which, in a smooth, laminar disc, will cause them to radially drift inward. This migration of pebbles can lead to them being transported to within the path of the growing planetesimal core, constantly replenishing the pebbles within the planetesimal's feeding zone and preventing it from reaching its isolation mass as quickly as they do in Section 5.2. If the planetesimal is gravitationally massive and capable of perturbing the velocities of nearby solids, the pebbles may enter into complex trajectories, orbiting and eventually settling down into its gravitational potential well. If the planetesimal's gravitational cross section exceeds its geometric cross section, pebble accretion may become the dominant growth mechanism. In their review paper Johansen & Lambrechts (2017) show that pebble accretion may be capable of resolving many of the timescale problems associated with core accretion, whilst being able to explain the formation of all planet types.

5.4.3 Implications for formation through GI

In Section 5.3.1 we used SPH simulations to determine the critical mass limit for fragmentation in a disc surrounding a $2.4\,\mathrm{M}_{\odot}$ star, finding that for a $R_{\mathrm{out}} = 400\,\mathrm{AU}$ disc, $M_{\mathrm{d,crit}} = 0.3\,\mathrm{M}_{\odot}$ ($q_{\mathrm{crit}} = 0.125$). Whilst we have mostly focused our discussion on the case of single fragment formation from our SPH simulations, it is also likely that multiple clumps may form in a disc with a mass slightly higher than $M_{\mathrm{d,crit}}$ (see Figure 5.3). The initial formation of multiple protoplanets may then also provide an explanation for the wider-orbit planet P2 which has also been inferred, located at a distance $a \approx 140\,\mathrm{AU}$ from the parent star (Boccaletti et al., 2020). We have refrained from analysing the formation history of planet P2, due to its slightly more tentative detection, choosing instead to focus on planet P1. However it would seem that the formation of a $3\,\mathrm{M}_{\mathrm{Jup}}$ planet at $a \approx 140\,\mathrm{AU}$ may be even more challenging to explain in the core accretion paradigm than is the case for planet P1, as the gas and dust surface densities in the disc will drop off as $\Sigma \propto R^{-1}$, hence will be exceedingly low at such a large radius. As we see only minimal core growth at $R = 30\,\mathrm{AU}$ in Figure 5.2d, it is likely that growth at $R = 140\,\mathrm{AU}$ would be near-negligible. It may then be the case that in fact planets P1 and P2 represent two survivors from several fragments which could have initially formed.

Despite the AB Aurigae disc being far too low mass to be gravitationally unstable currently, models of the system's viscous evolution in Section 5.3.2 suggest that it may have been much more massive when it was younger, potentially exceeding the critical mass limit for fragmentation. It seems reasonable to expect that the disc might have previously fragmented in an extended system such as AB Aurigae, as previous studies suggest that fragmentation is inevitable in GI discs at radii, $R \gtrsim 50 - 100\,\mathrm{AU}$ (Rafikov, 2005; Whitworth & Stamatellos, 2006; Clarke, 2009; Forgan & Rice, 2011). Further, in Chapter 3 and Cadman et al. (2020a); Haworth et al. (2020) we used hydrodynamic simulations to demonstrate that, whilst lower mass stars may support gravitationally stable massive discs, susceptibility to fragmentation increases as a function of stellar mass, and that discs around higher mass stars ($M_{*} \geq 2\,\mathrm{M}_{\odot}$) may fragment for relatively low disc-to-star mass ratios. AB Aurigae being an extended disc around a higher mass star therefore seems to be an ideal candidate system to search for surviving products of GI.

If the disc had been able to fragment whilst it was young, it is not necessarily true that the clumps will have survived the 1−4 Myr lifetime of the AB Aurigae system.

We find that fragments may initially form on wide-orbits with $R \gtrsim 200\,\mathrm{AU}$, and use analytic calculations to predict initial clump masses $1.6 - 13.3\,\mathrm{M_{Jup}}$. However subsequent evolution is inevitable, and the fragments will rapidly migrate through the disc (Baruteau et al., 2011).

In the tidal downsizing hypothesis of planet formation (Nayakshin, 2010a,b, 2011) GI embryos will cool and contract as they migrate. Dust sedimentation may lead to the formation of a solid core, potentially of mass comparable to that of a terrestrial planet (Boss, 1998). If the embryo's outer layers contract slowly whilst migration occurs rapidly then tidal stripping from the parent star may occur once the embryo reaches the inner disc, as its physical radius may exceed its Hill sphere (Nayakshin, 2010a). It is possible that many of the initially formed fragments may be entirely destroyed during this tidal downsizing process (Nayakshin & Fletcher, 2015; Humphries et al., 2019). Population synthesis calculations find this may be the true of $\approx 50\%$ of GI protoplanets, with the remaining objects eventually residing at $a \gtrsim 20\,\mathrm{AU}$ (Forgan & Rice, 2013b), although when including fragment-fragment scattering this survival fraction may be significantly less (Forgan et al., 2018). The initial formation of multiple clumps would then be necessary if any are to survive beyond this early phase of evolution. Accretion of material onto the protoplanets will also occur as they migrate through the disc. Kratter et al. (2010) showed that most GI fragments will grow well beyond the mass limit for Deuterium burning, and that any GI-born planets likely represent the low mass tail of the eventual GI fragment mass distribution. The Jeans mass estimates that we present in Section 5.3.3 therefore represent those shortly after collapse only, as dynamical evolution will significantly influence the embryo's eventual mass.

We also tentatively suggest that the previous disc mass estimates ($M_\mathrm{d} \approx 0.01\,\mathrm{M_\odot}$) (DeWarf et al., 2003; Andrews & Williams, 2005; Corder et al., 2005; Semenov et al., 2005) appear too low to be consistent with the high stellar accretion rate (Salyk et al., 2013), which is indicative of the presence of a large mass reservoir. 1D calculations which assume the disc to be in a quasi-steady state with a radially constant viscous$-\alpha$ suggest a lower limit for the current disc mass as $M_\mathrm{d} \gtrsim 0.1\,\mathrm{M_\odot}$ (see Fig. 5.4). This rough lower limit is in fact consistent with the upper bound of the uncertainty on the current disc mass estimates. However even when assuming this slightly higher disc mass, we still find the calculated accretion rates from our viscous evolution models in Section 5.3.2.3 to be significantly lower than the accretion rate measured from the system. On the unusually high stellar accretion rate, Tang et al. (2012) suggest a possible explanation is the presence of an inner

disc, characterised by a gas/dust cavity observed at $R \approx 100\,\mathrm{AU}$, which is being replenished through accretion from the remnant envelope above and below the disc midplane. This would suggest that the measured accretion rate does not represent that of a settled, $R_{\mathrm{out}} = 400\,\mathrm{AU}$ disc as we have assumed here, and would allow for the existence of a low mass disc whilst being consistent with a high accretion rate. We only attempt to further highlight this discrepancy between the measured disc mass and accretion rate, and note that the current mass of the disc does not significantly affect the overall conclusions from this chapter in regards to the formation history of planet P1.

Since the writing of this chapter, Currie et al. (2022) published a direct imaging result of a protoplanet in the AB Aurigae disc at $\sim 93\,\mathrm{AU}$, also suggesting that this may be indicative of planet formation through GI. This recent result adds further strength to our hypothesis that the AB Aurigae disc may have been potentially massive enough to undergo fragmentation in its recent past, possibly forming multiple clumps which may still be present and evolving in the disc today.

5.5 Conclusions

In this chapter we have analysed the possible formation history of the $4-13\,\mathrm{M_{Jup}}$ planet observed at $a \approx 30\,\mathrm{AU}$ within the protoplanetary disc surrounding AB Aurigae (Piétu et al., 2005; Tang et al., 2012, 2017; Boccaletti et al., 2020). The young age of the star-disc system places strict constraints on the core accretion formation timescale, which we find challenging to explain within its $1-4\,\mathrm{Myr}$ measured age. The planet's high mass and wide-orbit are indicative of a planet which may have instead formed through gravitational instability in the natal AB Aurigae disc.

The key results are as follows.

1. Typical in-situ core accretion formation timescales for planet P1 exceed the system's measured age. Fine tuning of the model parameters is required in order to form a planet of $4\,\mathrm{M_{Jup}}$ within $1-4\,\mathrm{Myr}$, including significant enhancement of the planetesimal surface density in the disc, and, in most cases, that a large planetesimal core with $M_{\mathrm{core,init}} = 0.1\,\mathrm{M_\oplus}$ has already formed near to the snow line at $t = 0$. At the current semi-major axis of planet P1 ($a = 30\,\mathrm{AU}$) we find extremely slow in-situ growth due to the

low disc surface densities at wide orbits. We do not include a prescription for pebble accretion in our models here, but note that it may be capable of speeding up planet growth.

2. A disc surrounding a $2.4\,\mathrm{M}_\odot$ star, analogous to young AB Aurigae, would have fragmented if its initial mass exceeded $M_{\mathrm{d,crit}} = 0.3\,\mathrm{M}_\odot$ ($q_{\mathrm{crit}} = 0.125$). If the disc mass is slightly higher than $M_{\mathrm{d,crit}}$ several fragments may form. Formation of multiple fragments would allow margin for some fragment destruction, which is likely inevitable during their subsequent dynamical evolution of GI protoplanets.

3. Viscous evolution models of the AB Aurigae disc suggest that it may have been massive enough to exceed $M_{\mathrm{d,crit}}$ during its early lifetime whilst the disc was still young and massive. We find that a $0.1\,\mathrm{M}_\odot$ disc may have been more massive than $M_{\mathrm{d,crit}} = 0.3\,\mathrm{M}_\odot$ approximately $1.25 - 4\,\mathrm{Myr}$ ago.

4. Fragments will initially form on wide orbits and then rapidly migrate inwards. Typical migration timescales of a GI protoplanet which formed at $R \approx 200\,\mathrm{AU}$ within a young AB Aurigae disc are found to be shorter than the current age of the system. We use analytic calculations to determine type I and type II migration timescales, finding that for migration from $R_{\mathrm{out}} = 200\,\mathrm{AU}$ to $R_{\mathrm{in}} = 30\,\mathrm{AU}$, $\Delta t_{\mathrm{mig,I}} = 6.9\,\mathrm{kyr}$ and $\Delta t_{\mathrm{mig,II}} = 1.0\,\mathrm{Myr}$ when considering disc conditions taken from our hydrodynamic simulations.

5. Calculations of the Jeans mass in a moderately irradiated proto-AB Aurigae disc represent what the initial fragment masses might have been immediately after formation. We find that $M_{\mathrm{J}} = 1.6 - 13.3\,\mathrm{M}_{\mathrm{Jup}}$, which is consistent with the masses of the planets P1 and P2 in the AB Aurigae disc.

6. Although we focus our discussion on the formation history of planet P1, we highlight that planet P2 found at $a \approx 140\,\mathrm{AU}$ with an estimated mass $M_{\mathrm{P2}} = 3\,\mathrm{M}_{\mathrm{Jup}}$ may be even more challenging to reconcile with formation through core accretion.

We therefore propose that planets P1 and P2 which have been discovered through scattered light observations of the AB Aurigae disc (Boccaletti et al., 2020) may stand as evidence of planet formation through GI.

Chapter 6

The observational impact of dust trapping in self-gravitating discs

This chapter contains work from Cadman et al. (2020b). Whilst Cadman et al. (2020b) was led by myself, the project was undertaken in collaboration with various listed co-authors, hence some parts of the text may contain contributions from these co-authors.

6.1 Motivation

Discs around very young stars are typically heavily embedded and optically thick at optical wavelengths (Dunham et al., 2014). They will, however, emit thermal infrared (IR) radiation and may be resolved by high-resolution, sub-mm observations with ALMA. Thanks to recent observational advances, spiral substructure, characteristic of massive self-gravitating protoplanetary discs, is now within our observing capabilities. For example, the Disk Substructures at High Angular Resolution Project (DSHARP) ALMA survey recently performed an in depth analysis of 20 nearby protoplanetary discs (see Figure 1.4), 3 of which exhibit possible spiral substructure resembling GI (Pérez et al., 2016; Andrews et al., 2018b; Huang et al., 2018b).

Non-axisymmetric disc features are, however, not unique to GI, and may be explained through alternative mechanisms such as planet-disc interactions (Lin & Papaloizou, 1986; Tanaka et al., 2002). It may be possible to distinguish between planet and GI induced spiral structure through scattered light vs. sub-

mm observations, as dust trapping in spiral regions is likely to be more effective in gravitationally unstable discs (Rice et al., 2004; Dong et al., 2015a; Juhász et al., 2015).

As was shown in Section 2.2.3, the negative gas pressure gradient in discs results in gas-dust drag and inward radial migration of the faster orbiting dust grains. Micron-sized grains will typically be strongly coupled to the gas, hence will orbit with the same sub-Keplerian velocities as the gas, hence they will closely trace the gas distribution. Objects larger than metre sizes will be decoupled from the gas and will orbit with approximately Keplerian velocities. Intermediate, ∼mm-sized grains will however experience a large drag and undergo significant radial drift.

In smooth, laminar discs radial drift results in migration toward the disc centre where gas pressure is maximum. However, the propagation of GI induced spiral density perturbations will generate a non-axisymmetric pressure gradient where the peaks of the spirals behave as local pressure maxima. Dust grains will migrate and concentrate at the peaks of the spirals resulting in enhanced emission and potential accelerated planetesimal growth (Rice et al., 2004, 2006). Dipierro et al. (2014, 2015) have previously shown that GI induced spiral structure should be detectable with ALMA at moderate distances ($d \sim 140\,\mathrm{pc}$), and that dust trapping in the spirals will produce detectable signatures in their observed spectral index maps.

In this chapter we build on previous work by Hall et al. (2016) who developed a semi-analytic formalism for determining the structure of self-gravitating protoplanetary discs, performed 3D Monte Carlo radiative transfer on these models and produced synthetic disc images using the ALMA simulator. We add to these models by including a prescription for dust grain enhancement in the spiral density waves. These models allow us to produce a suite of discs at little computational expense when compared to approaches such as SPH. Therefore we are able to efficiently explore a wide range of disc parameter space and produce observational predictions for telescopes such as ALMA.

In Sections 6.2 and 6.3 we present our disc models, and describe the radiative transfer approach as well as how we used the ALMA simulator in our analysis. In Section 6.4 we use SPH to model the extent to which we might expect grains to be enhanced in self-gravitating discs, allowing us to inform our semi-analytic prescription. In Section 6.5 we discuss grain growth and the grain fragmentation

threshold. In Section 6.6 we discuss our disc parameter setup and in Section 6.7 we use our models to generate synthetic observations of self-gravitating discs at distances comparable to the Taurus star-forming region. In Section 6.8 we apply our models to three discs from the DSHARP survey, analysing whether or not their observed substructure may be the result of self-gravity. In Section 6.9 we discuss and draw conclusions.

6.2 Disc models - Setup

We setup our discs using the 1D models introduced by Clarke (2009) (see also Rice & Armitage, 2009; Forgan & Rice, 2013c) and further developed by Hall et al. (2016) to include 3D structure such as the spiral density waves characteristic of self-gravitating discs. These models are described in detail in Hall et al. (2016) and summarised in Section 6.2.1. We refer the reader to Hall et al. (2016) for a comparison of this formailsm's ability to accurately reproduce self-gravitating spiral shape and amplitudes from SPH simulations. Dust grain enhancement is imposed semi-analytically, in line with what we might expect from spiral density structure in self-gravitating discs, and is described in Section 6.2.2.

6.2.1 Self-gravitating disc models

To set up our disc models we begin by using the same formalism as the 1D models outlined in Section 3.2. We assume that irradiation leads to a constant background temperature of $T_{\rm irr} = 10\,\mathrm{K}$ here.

These are then developed into 3D models in the same way as is described in detail in Hall et al. (2016). We assume that the azimuthal location of the spirals, $\theta_{\rm spiral}$, will be logarithmic in shape as,

$$\theta_{\rm spiral} = \frac{1}{b}\log\left(\frac{r}{a}\right), \tag{6.1}$$

where a and b are constants defining the shape of the spirals. Here we use $a = 13.5$ and $b = 0.38$, in line with that used in Hall et al. (2016).

At each azimuthal location in the disc, $\theta_{\rm x,y}$, we calculate a fractional over-density, $\delta\Sigma/\Sigma$, characterised by a spiral amplification factor, S, such that (Cossins et al.,

2009),

$$\frac{\langle \delta\Sigma \rangle}{\langle \Sigma \rangle} = S\alpha^{1/2}, \tag{6.2}$$

where here we define $S = 2$.

This fractional over-density is imposed sinusoidally at each azimuthal location in the disc, $\theta_{\mathrm{x,y}}$ such that,

$$\delta\Sigma(\phi) = \langle \delta\Sigma \rangle \cos(m\phi). \tag{6.3}$$

Here, m is the azimuthal wavenumber (i.e. the number of spiral arms) and ϕ is the phase difference between the location of the spiral arms and each azimuthal position in the disc,

$$\phi = \theta_{\mathrm{spiral}} - \theta_{\mathrm{x,y}}. \tag{6.4}$$

We expect that the azimuthal wavenumber will be roughly related to the disc-to-star mass ratio, q, as (Cossins et al., 2009; Dong et al., 2015b),

$$m \approx 1/q. \tag{6.5}$$

We use this in Equation 6.3 to impose an azimuthal wavenumber in a disc of mass-ratio, q, assuming a symmetrical response (with $m = 2, 4, 8...$) and rounding m to the nearest appropriate value.

Finally, the disc's vertical density profile is calculated using Equation 2.60, where the self-gravitating scale height is calculated using Equation 2.61.

6.2.2 Grain concentration

In the presence of spiral density waves, dust grains will radially migrate and concentrate at their density maxima (Rice et al., 2004). The extent of this radial migration will be strongly dependent on grain size, a. Small grains of $\sim \mu$m scale will be strongly coupled to the gas in the disc, will experience very little radial

drift and will closely trace the gas distribution. The largest particles of $\sim$ m scale will be decoupled and will be unaffected by the disc gas pressure, therefore orbiting with approximately Keplerian velocities.

For intermediate-sized dust grains of $\sim$ mm $-$ cm scale, the impact of the gas drag will be significant. Radial drift velocities will be large and, hence, grain concentration at the spiral pressure maxima will be high. The gas-dust coupling is characterised by the Stokes number,

$$\mathrm{St} = \frac{a\rho_\mathrm{s}\Omega}{\rho c_\mathrm{s}}, \tag{6.6}$$

where ρ_s is the internal density of the dust grains and ρ is the local gas density.

The solution of the momentum equation suggests that the radial drift velocity has a $1/(\mathrm{St}+\mathrm{St}^{-1})$ relation (Weidenschilling, 1977). We therefore propose a grain enhancement factor of the form,

$$\eta_i = 1 + \frac{2d}{\mathrm{St}_i + \mathrm{St}_i^{-1}} - \frac{\mathrm{St}_i}{200}, \tag{6.7}$$

where d is a constant, to be determined later, that represents the peak dust concentration factor in spirals. Here, η_i is defined as the local grain enhancement factor relative to the mean dust-to-gas ratio in the disc for the ith grain size. The local dust surface density for the ith grain size, $\Sigma_{\mathrm{d},i}$, will then be enhanced as,

$$\Sigma_{\mathrm{d},i} = \langle\epsilon_i\rangle(\Sigma_0 + \eta_i\delta\Sigma), \tag{6.8}$$

where $\langle\epsilon_i\rangle$ is the average dust-to-gas ratio for each grain size in the disc. Here we use the canonical value of $\langle\epsilon\rangle = 0.01$ to represent the total dust-to-gas ratio over all grain sizes.

Particles with $\mathrm{St}_i \ll 1$ will be strongly coupled to the gas, experience minimal radial drift and will therefore have $\eta_i \approx 1$. The dust surface density will exactly trace the gas distribution in this case, with $\Sigma_{\mathrm{d},i} = \langle\epsilon_i\rangle(\Sigma_0 + \delta\Sigma)$. Large solids with $\mathrm{St}_i \gg 1$ will be entirely decoupled from the gas and will have constant surface density across the disc, with $\eta_i \approx 0$ and $\Sigma_{\mathrm{d},i} = \langle\epsilon_i\rangle\Sigma_0$. Note that we set a lower limit of $\eta_i = 0$ here. Intermediate sized grains with $\mathrm{St}_i \approx 1$ will generate peak enhancement factors of $\eta_i \approx 1 + d$, and therefore dust surface densities,

$\Sigma_{\mathrm{d},i} = \langle\epsilon_i\rangle(\Sigma_0 + (1+d)\delta\Sigma)$.

From Equation 6.3, regions coincident with the spiral peaks, where $m\phi = 0°$, will experience maximum enhancement by a factor $\Sigma_0 + \eta_i\langle\delta\Sigma\rangle$, as $\delta\Sigma = \langle\delta\Sigma\rangle$ in these regions. Dust surface density in inter-arm regions, where $m\phi = 180°$, will equally be depleted by a factor $\Sigma_0 - \eta_i\langle\delta\Sigma\rangle$, as $\delta\Sigma = -\langle\delta\Sigma\rangle$ here.

To avoid $\Sigma_{\mathrm{d},i}$ becoming negative in inter-arm regions, we employ a correction factor,

$$\Sigma_{\mathrm{d},i,\mathrm{corr}} = \begin{cases} \eta_i\langle\delta\Sigma\rangle - \Sigma_0, & \text{if } \Sigma_0 + \eta_i\delta\Sigma < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6.9}$$

Thus our resultant dust surface density becomes,

$$\Sigma_{\mathrm{d},i} = \frac{\langle\epsilon_i\rangle(\Sigma_0 + \eta_i\delta\Sigma + \Sigma_{\mathrm{d},i,\mathrm{corr}})\Sigma_0}{\Sigma_0 + \Sigma_{\mathrm{d},i,\mathrm{corr}}}. \tag{6.10}$$

This ensures $\Sigma_{\mathrm{d}} > 0$ by increasing our dust distribution by a factor $\Sigma_{\mathrm{d,i,corr}}$ in cases where $\Sigma_0 + \eta_i\delta\Sigma < 0$. The denominator is a normalisation which ensures our mean dust surface density remains unchanged by $\Sigma_{\mathrm{d,i,corr}}$, thus ensuring mass conservation.

6.2.3 Monte Carlo Radiative Transfer

The dust temperatures are calculated using the TORUS radiation transfer code (Harries et al., 2019). Radiative equilibrium is calculated using the Monte Carlo technique originally described in Lucy (1999) and outlined in Section 2.4.2. Our discs are illuminated by a central star, whose radiation field is here represented by 10^9 photon packets. These photon packets are emitted from the star isotropically and proceed to undergo a random walk through the grid, experiencing both absorption and scattering, until they escape the computational domain and the dust temperatures can be calculated assuming radiative equilibrium. Another cycle of 10^9 photon packets are then emitted, now with these updated temperatures, until the dust temperatures are found to converge and continuum images can be produced.

Our disc is constructed within a mesh of grid cells, where we use a mass resolution

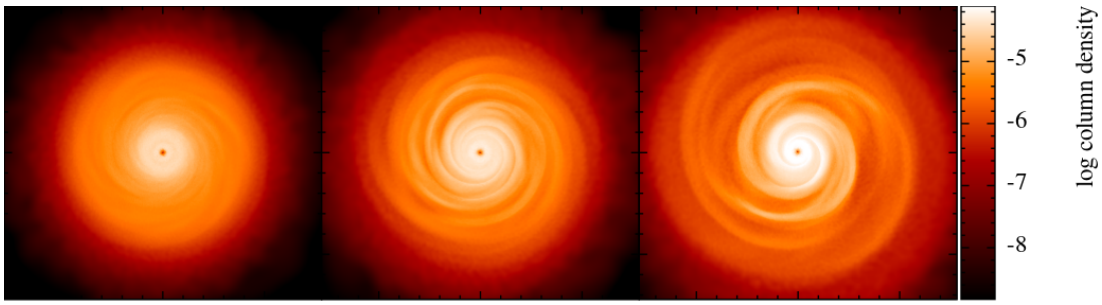

Figure 6.1 *Surface density structure of self-gravitating SPH discs with* $R_{\rm out} = 100\,AU$ *after evolving for 5 outer orbital periods (*$t = 31420\,yrs$*). Discs are constructed with 500,000 SPH gas particles and have mass ratios* $q = 0.2, 0.3, 0.4$ *from left to right.*

criteria of $1 \times 10^{-4}\,\mathrm{M}_{\odot}$ when subdividing parent cells into 2^D child cells, where $D = 3$ for our 3-dimensional domain.

6.3 ALMA simulations

The output continuum images from TORUS are then used as inputs to the ALMA simulator in the Common Astronomy Software Application (CASA) package (version 5.1) (McMullin et al., 2007) to produce synthetic ALMA images from our disc models. We use ALMA cycle 7 array configurations to produce these images, exploring various array sizes and resolutions in order to find optimal configurations for each observing frequency.

We apply unsharp image masking (Malin, 1977) to generate residual images from our synthetic observations by subtracting a smoothed radial profile of the image flux from itself. This technique highlights any non-axisymmetric features in our images, specifically spiral arms, by reducing the image flux range without reducing its dynamical range. We subtract a 2D Gaussian profile of FWHM closely matched to the beam size of our simulated images (we use 0.05"x0.05" here), and scaled with the peak image flux.

6.4 SPH models - Determining peak grain enhancement

Our semi-analytic prescription of dust trapping in Equation 6.7 requires that we determine the expected peak dust concentration factor, d, in disc spiral arms. To do this we employ the 3D SPH code PHANTOM (Price et al., 2018) to numerically model the behaviour of dust particles in self-gravitating discs.

We set up three sets of discs with mass ratios $q = 0.2$, 0.3 and 0.4 around a central star of mass $M_* = 1\,\mathrm{M}_\odot$. Each disc has initial inner and outer radii $R_{\mathrm{in}} = 1\,\mathrm{AU}$ and $R_{\mathrm{out}} = 100\,\mathrm{AU}$, and are set up with initial surface density profiles $\Sigma \propto R^{-1.5}$ and initial sound speed profiles $c_s \propto R^{-0.5}$. We use artificial viscosity terms $\alpha_{\mathrm{SPH}} = 0.1$ and $\beta_{\mathrm{SPH}} = 0.2$. Cooling is modelled using the polytropic cooling approximation introduced in Stamatellos et al. (2007) and outlined in Section 2.4.1.4.

We use 500,000 SPH particles to represent the gas disc and initially evolve the discs for 5 outer orbital periods without adding any dust particles. We then add 125,000 dust SPH particles and allow the discs to evolve for a further orbital period. The final states of the gas-only discs are shown in Figure 6.1. For each set of discs we run 20 separate simulations for 20 different grain sizes distributed log-normally between $0.1\,\mu$m and 200 cm. To minimise computational expense, we neglect the self-gravity of these dust particles and treat them as test particles only.

Dust-gas mixtures are modelled using two evolution models; the two-fluid method where the dust and gas are represented by two distinct particle populations coupled by a drag term (Laibe & Price, 2012a,b), and the one-fluid method where the mixture is represented by gas particles only and the grain fraction is evolved along with the gas density for each particle (Price & Laibe, 2015). The one-fluid method is implemented for smaller particle sizes at which the terminal velocity approximation is valid (i.e. when the stopping time is shorter than the computational timestep, see Youdin & Goodman, 2005), thus it is not appropriate for modelling larger grains. We find an appropriate grain size boundary at which to switch between these two methods at $a \approx 2\,\mathrm{mm}$. We therefore model all discs with $a \leq 2\,\mathrm{mm}$ using the one-fluid method, and discs with $a > 2\,\mathrm{mm}$ using the two-fluid method.

After evolving the dusty discs for a further orbital period, peak dust-to-gas ratios are determined by taking a radial slice of the disc, of azimuthal width 5°, and fitting a Gaussian distribution to the dust mass fraction at the spiral location. A demonstration of this is shown in Figure 6.2; we fit curves to a radial slice of the $q = 0.4$ disc, where the spiral's peak is located at $\approx 60 - 70$ AU. In Figure 6.3 we fit log-normal curves to the best-fit dust-to-gas ratio peaks from the $q = 0.3$ and $q = 0.4$ discs. We exclude the $q = 0.2$ disc from the remainder of this analysis as only weak spiral structure develops, therefore we observe only moderate grain enhancement.

Grain enhancement generally increases with increasing disc mass, primarily due to stronger spiral structure as we increase the disc-to-star mass ratio. This results in larger density gradients, greater radial drift velocities, and stronger concentration of grains. It is possible that grain concentration may continue to increase with increasing disc mass above $q = 0.4$. However, for mass ratios $q \gtrsim 0.5$ discs become susceptible to fragmentation for the stellar mass considered here. This will act to disrupt any spiral arm structure thus limiting grain concentration. We therefore only model disc masses up to $q = 0.4$.

Grains become most concentrated for sizes $a \approx 200 - 500$ mm, with peak dust-to-gas ratios $\epsilon \approx 0.06$ and $\epsilon \approx 0.07$ in the $q = 0.3$ and $q = 0.4$ discs respectively, giving values of $d \approx 5$ and $d \approx 6$ respectively for Equation 6.7. For the discs generated in Section 6.6, with disc masses $q \lesssim 0.3$, we assume a maximum value of $d = 5$ in our models.

6.5 Grain growth and the fragmentation threshold

Appropriate maximum grain sizes for use in the grain size distributions from the equations in Section 6.2 can be obtained using models of grain growth in protoplanetary discs. Grain growth proceeds through steady coagulation and accumulation during grain-grain collisions (Testi et al., 2014). The tendency for grains to stick together and grow during these collisions will depend on their collisional velocities. Particles with St < 1 (i.e. smaller grains) have smaller relative azimuthal velocities, hence when they collide they will likely coalesce in a so-called *hit-and-stick* process (Chokshi et al., 1993; Dominik & Tielens, 1997).

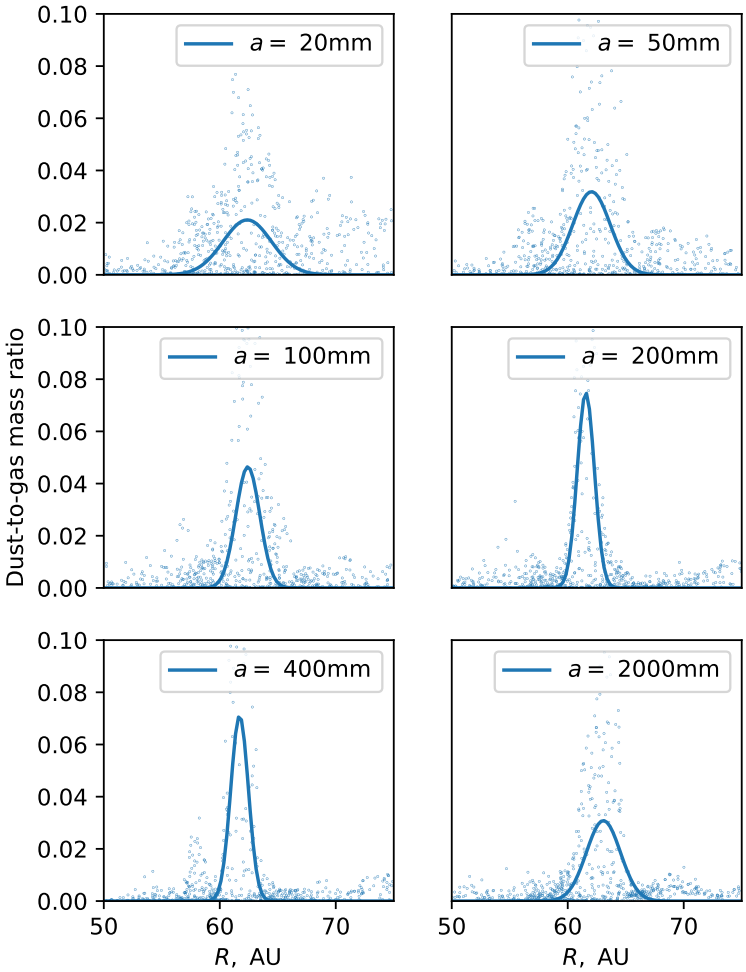

Figure 6.2 *Gaussian fits to the dust-to-gas mass ratios taken from a radial slice of the $q = 0.4$ disc, setup as described in Section 6.4. We plot how the dust-to-gas ratios vary for grain sizes $a = 20\,mm$, $50\,mm$, $100\,mm$, $200\,mm$, $400\,mm$ and $2000\,mm$. Grains sizes $a \approx 200 - 500\,mm$ become highly concentrated reaching peak dust-to-gas ratios $\epsilon \approx 0.07$ here.*

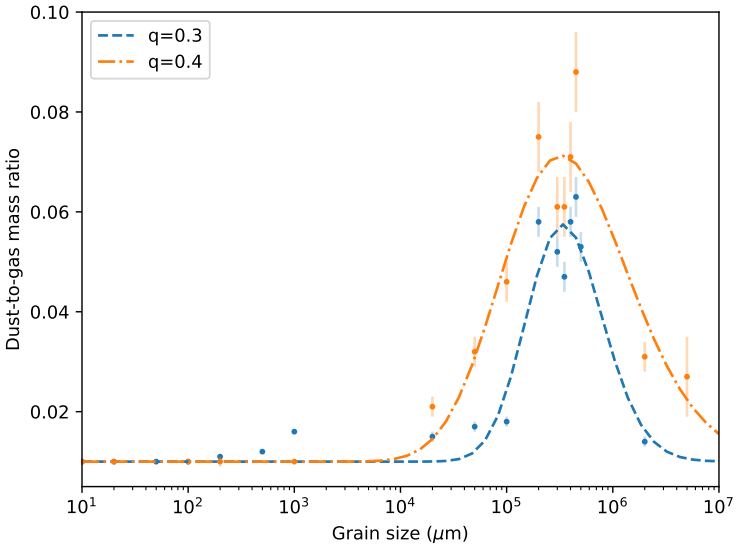

Figure 6.3 *Best-fit dust-to-gas ratios in SPH discs with mass ratios* $q = 0.3, 0.4$ *and* $R_{\rm out} = 100\,AU$*. Each disc consists of 500,000 gas particles, 125,000 dust particles and has been allowed to evolve for 6 outer orbital periods (*$t = 37700 yrs$*). We show the points with* 1σ *error bars obtained from their best-fit values. Log-normal curves are fitted to the data.*

Larger particles will have higher relative azimuthal velocities, reaching a constant maximum value for $\mathrm{St} \geq 1$. Kothe et al. (2013) find a power-law mass dependence for the affinity of solids, $v_{\rm th} \propto m^{-3/4}$, with less massive solids having a greater threshold velocity for sticking. As particles grow, their impact velocities will increase accordingly and collisions will result in particles either bouncing off each other, compacting their densities in the process (Güttler et al., 2010; Zsom et al., 2010), or shattering into several smaller fragments. These two growth barriers, known as the *bouncing barrier* and the *fragmentation threshold* respectively, may consequently limit the maximum size to which grains are able to grow through collisional accumulation, therefore limiting our value of $a_{\rm max}$.

The particle size at which the bouncing barrier is reached will depend on a number of factors such as particle porosity, density and material, and is therefore non-trivial to calculate analytically. Instead, we reason that the wealth of smaller, micron-sized solids dominating the dust-mass budget in discs (see Williams & Cieza, 2011) requires regular replenishment through a cycle of growth and fragmentation, as otherwise these smaller grain sizes would quickly be depleted as they grow (Dullemond & Dominik, 2008; Brauer et al., 2008; Birnstiel et al., 2010). This indicates that particles are able to grow to at least as large as the

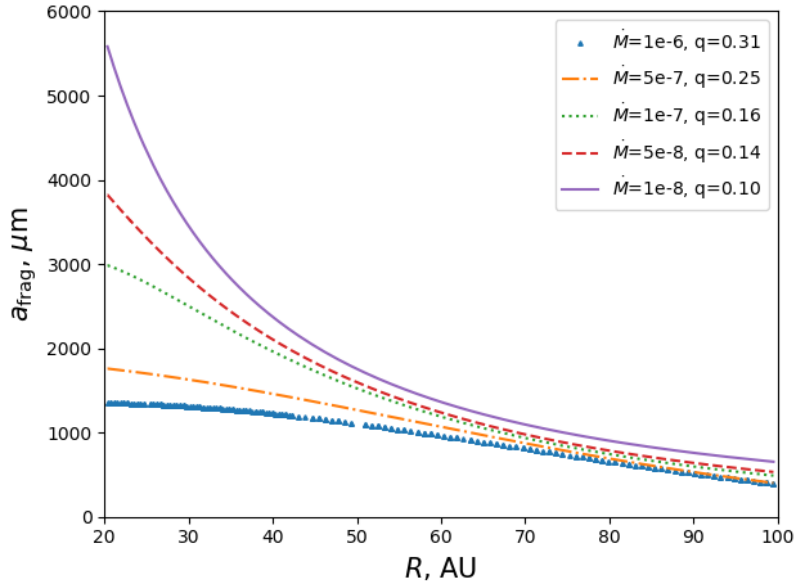

Figure 6.4 *Radial distribution of the fragmentation threshold from Equation 6.11 for mass accretion rates* $\dot{M} = 1 \times 10^{-6}\, M_{\odot}\, yr^{-1}$, $5 \times 10^{-7}\, M_{\odot}\, yr^{-1}$, $1 \times 10^{-7}\, M_{\odot}\, yr^{-1}$, $5 \times 10^{-8}\, M_{\odot}\, yr^{-1}$ *and* $1 \times 10^{-8}\, M_{\odot}\, yr^{-1}$, *and where* $v_{\rm frag} = 10\, ms^{-1}$.

fragmentation threshold, and we therefore use this to define $a_{\rm max}$ in our models.

Dipierro et al. (2015) show that the fragmentation threshold leads to a maximum grain size of,

$$a_{\rm max} = \frac{4\langle \Sigma_{\rm g} \rangle}{3\pi\alpha\rho_{\rm s}} \frac{v_{\rm frag}^2}{\langle c_{\rm s} \rangle^2}, \tag{6.11}$$

where $v_{\rm frag}$ is the fragmentation threshold velocity representing the maximum relative velocity that particles can withstand before collisions result in shattering. We use the azimuthally averaged gas surface density, $\langle \Sigma_{\rm g} \rangle$, and sound speed, $\langle c_{\rm s} \rangle$, as spiral features are short lived and grain growth timescales typically exceed these. We can estimate the viscous$-\alpha$ here by assuming that in a quasi-steady, self-gravitating disc dominated by turbulent motion, the viscous stress will saturate at a maximum value $\alpha = 0.06$ (Rice et al., 2005), therefore defining the limiting maximum grain size.

We use this to set our value of $a_{\rm max}$ in our disc models assuming two cases of $v_{\rm frag} = 10\,{\rm ms}^{-1}$ and $v_{\rm frag} = 30\,{\rm ms}^{-1}$. The mid-plane distributions of $a_{\rm frag}$ are plotted in Figures 6.4 and 6.5 for discs of outer radius, $R_{\rm out} = 100\,{\rm AU}$, and mass accretion rates ranging from $\dot{M} = 1 \times 10^{-8}\,{\rm M}_{\odot}\,{\rm yr}^{-1}$ to $\dot{M} = 1 \times 10^{-6}\,{\rm M}_{\odot}\,{\rm yr}^{-1}$.

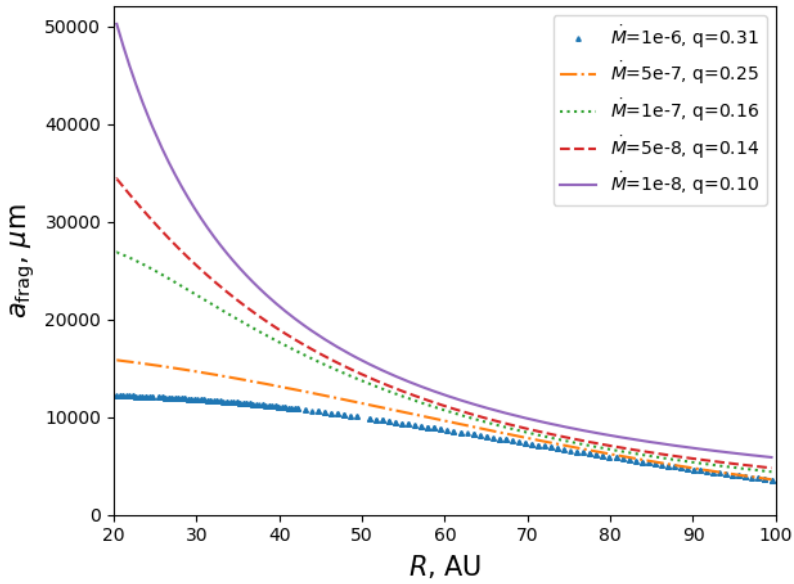

Figure 6.5 *Radial distribution of the fragmentation threshold from Equation 6.11 for mass accretion rates* $\dot{M} = 1 \times 10^{-6}\,M_{\odot}\,yr^{-1}$, $5 \times 10^{-7}\,M_{\odot}\,yr^{-1}$, $1 \times 10^{-7}\,M_{\odot}\,yr^{-1}$, $5 \times 10^{-8}\,M_{\odot}\,yr^{-1}$ *and* $1 \times 10^{-8}\,M_{\odot}\,yr^{-1}$, *and where* $v_{\rm frag} = 30\,ms^{-1}$.

The fragmentation threshold decreases with increasing $\dot{M}$ (i.e. with increasing disc mass), and becomes smallest in the outer disc where $a_{\rm frag}$ is comparable for all disc masses. In the most massive discs when $v_{\rm frag} = 10\,{\rm ms}^{-1}$, grains can only grow to $\sim$mm sizes before collisions become destructive, with this maximum grain size in the inner disc decreasing by a factor of ~ 5 as we increase the disc mass from $q = 0.1$ to $q = 0.31$. For the higher threshold of $v_{\rm frag} = 30\,{\rm ms}^{-1}$ the value of $a_{\rm frag}$ increases by a factor $v_{\rm frag}^2$ for all disc masses (a factor 9), and grains can grow to $a_{\rm max} \sim$cm sizes here.

6.6 Disc models - Setup parameters

With the additional information from Sections 6.4 and 6.5, it is now possible to use our models to predict for which disc parameters we expect self-gravitating disc substructure will be observable with ALMA. We setup discs as described in Section 6.2 exploring a range of parameter space in disc masses, grain sizes and observing frequencies.

Our central star is modelled with $M_* = 1\,{\rm M}_{\odot}$, $R_* = 2.325\,{\rm R}_{\odot}$ and $T_{\rm eff} = 4350\,{\rm K}$. We assume a canonical dust-to-gas ratio of 0.01, and represent our grains as

$\dot{M}$ ($M_\odot yr^{-1}$) (1)	M_{disc}/M_* (2)	m (3)
1×10^{-6}	0.31	4
5×10^{-7}	0.25	4
2.81×10^{-7}	0.22	4
1.58×10^{-7}	0.19	4
1×10^{-7}	0.16	8
5×10^{-8}	0.14	8
2.81×10^{-8}	0.12	8
1.58×10^{-8}	0.11	8
1×10^{-8}	0.10	8

Table 6.1 *(1) Mass accretion rates used for the discs setup in Section 6.6 and analysed in Section 6.7. (2) Calculated disc-to-star mass ratios. (3) Number of input spiral modes for each disc.*

Draine & Lee (1984) silicates with size distribution,

$$n(a) \propto a^{-q}, \tag{6.12}$$

distributed between minimum and maximum grain sizes a_{min} and a_{max}, and assume $q = q_{ism} = 3.5$ (Mathis et al., 1977). We set $a_{min} = 0.1\,\mu$m and vary the value of a_{max} to represent different stages of grain growth, using values $a_{max} = 10\,\mu$m (minimal grain growth), 1 mm, 10 cm, 100 cm, $a_{frag,10ms^{-1}}$ and $a_{frag,30ms^{-1}}$ (the grain fragmentation thresholds as described in Section 6.5). We use 50 dust grain sizes distributed logarithmically between $0.1\,\mu$m and $2 \times 10^6\,\mu$m, and set the grain fraction for any grain size greater than a_{max} in each case to be zero.

We generate discs with 9 different mass accretion rates (Equation 3.1), where an increase in $\dot{M}$ roughly corresponds to an increase in disc mass. We use values of $\dot{M} = [1 \times 10^{-6}, 5 \times 10^{-7}, 2.81 \times 10^{-7}, 1.58 \times 10^{-7}, 1 \times 10^{-7}, 5 \times 10^{-8}, 2.81 \times 10^{-8}, 1.58 \times 10^{-8}, 1 \times 10^{-8}]\,M_\odot yr^{-1}$, which correspond to disc-to-star mass ratios, $q \approx 0.31, 0.25, 0.22, 0.19, 0.16, 0.14, 0.12, 0.11$ and 0.10 respectively. Using the relation between mass ratio and the number of spiral modes in Equation 6.5, and assuming a symmetrical response where we have an even number of modes, each of these discs are set up with $m = 4$ and $m = 8$ for the more massive and less massive discs respectively. A summary of these disc setups is laid out in Table 6.1.

Continuum images of these discs are generated for observing frequencies 115 GHz

$f_{\rm obs}$ (1)	$t_{\rm obs}$ (2)	Antenna Config (3)	PWV Level (4)
115 GHz	1800 s	alma.cycle7.8	5.186 mm
230 GHz	1800 s	alma.cycle7.8	1.796 mm
690 GHz	1800 s	alma.cycle7.6	0.472 mm

Table 6.2 *Input parameters used here for generating synthetic images with* CASA. *(1) ALMA observing frequency. (2) Simulated observing time. (3) ALMA antenna configuration used. (4) Precipitable Water Vapour (PWV) level.*

($\lambda = 2.6$ mm), 230 GHz ($\lambda = 1.3$ mm) and 690 GHz ($\lambda = 0.4$ mm), corresponding to ALMA observing bands 3, 6 and 9 respectively. We consider discs at a distance of 140 pc, comparable to those in the Taurus star forming region. Example TORUS output images produced in this way are shown in Figure 6.6 for discs with $a_{\rm max} =$ 1 mm and accretion rates from Table 6.1.

We then use these continuum disc images as inputs to the CASA tasks SIMOBSERVE and SIMANALYZE and generate synthetic ALMA observations. Observing times, antenna configurations and PWV values used as inputs to CASA are laid out in Table 6.2. Unsharp image masking is applied to these synthetic observations in order to highlight any non-axisymmetric disc features present, as described in Section 6.3. We demonstrate the process of generating synthetic ALMA observations and then unsharp masked residual images from TORUS continuum profiles in Figure 6.7.

6.7 Disc models - Results

Our focus here is to analyse the parameter space in which self-gravitating disc structure may be observable with ALMA. We present our results in this section considering the effects of varying disc mass, grain size distribution and observing frequency on our ability to distinguish spiral structure in our disc model. Galleries of unsharp masked synthetic disc images where we explore this parameter space can be found in Section 6.10.

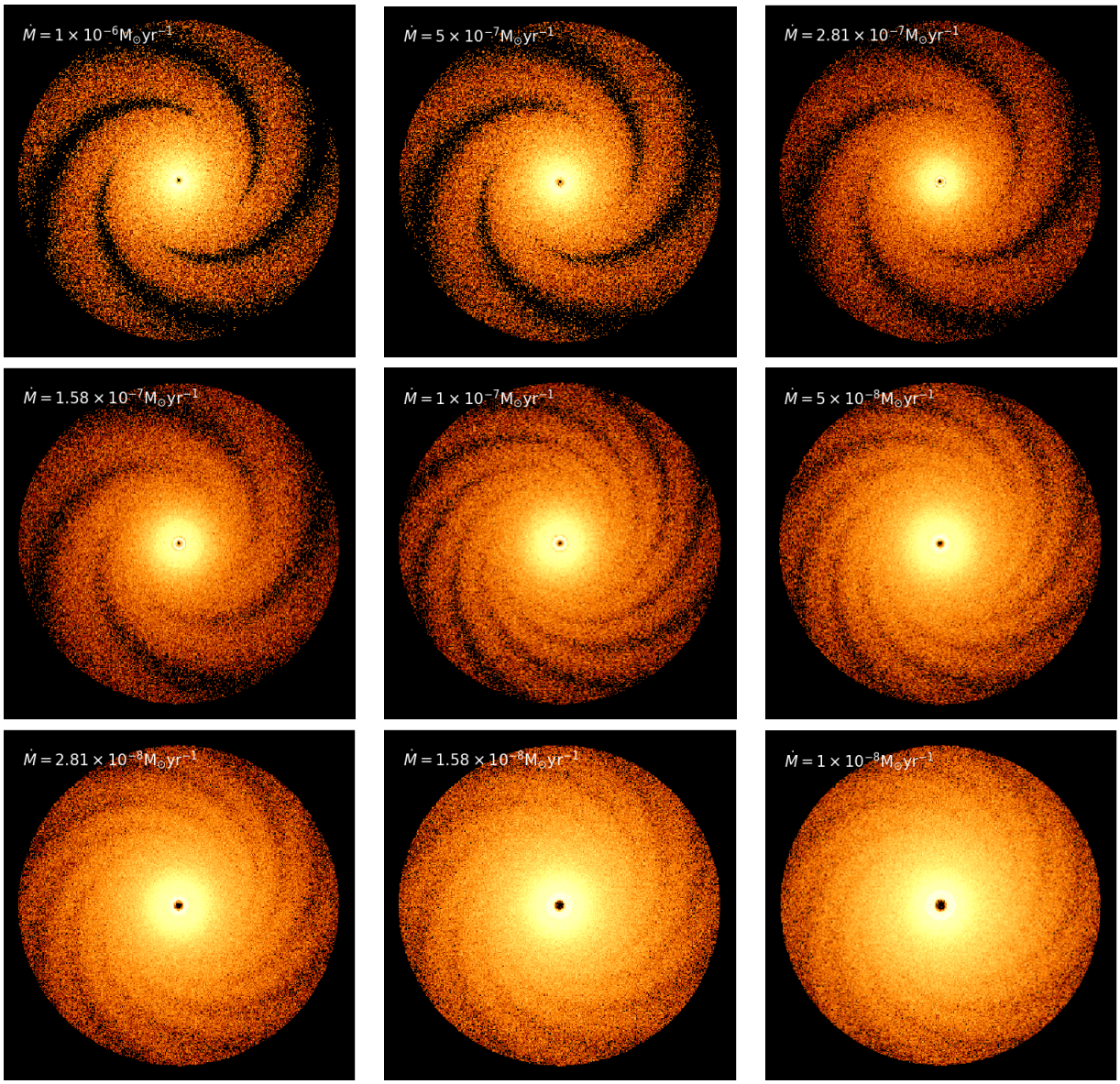

Figure 6.6 TORUS *disc continuum images at 230GHz ($\lambda = 1.3\,mm$). Discs are set up with $R_{\rm out} = 100\,AU$, grain size distributions $n(a) \propto a^{-3.5}$ with $a_{\rm min} = 0.1\,\mu m$ and $a_{\rm max} = 1\,mm$, and mass accretion rates (from left to right) Top: $\dot{M} = 1 \times 10^{-6}\,M_\odot\,yr^{-1}$, $5 \times 10^{-7}\,M_\odot\,yr^{-1}$, $2.81 \times 10^{-7}\,M_\odot\,yr^{-1}$. Middle: $1.58 \times 10^{-7}\,M_\odot\,yr^{-1}$, $1 \times 10^{-7}\,M_\odot\,yr^{-1}$, $5 \times 10^{-8}\,M_\odot\,yr^{-1}$. Bottom: $2.81 \times 10^{-8}\,M_\odot\,yr^{-1}$, $1.58 \times 10^{-8}\,M_\odot\,yr^{-1}$, $1 \times 10^{-8}\,M_\odot\,yr^{-1}$.*

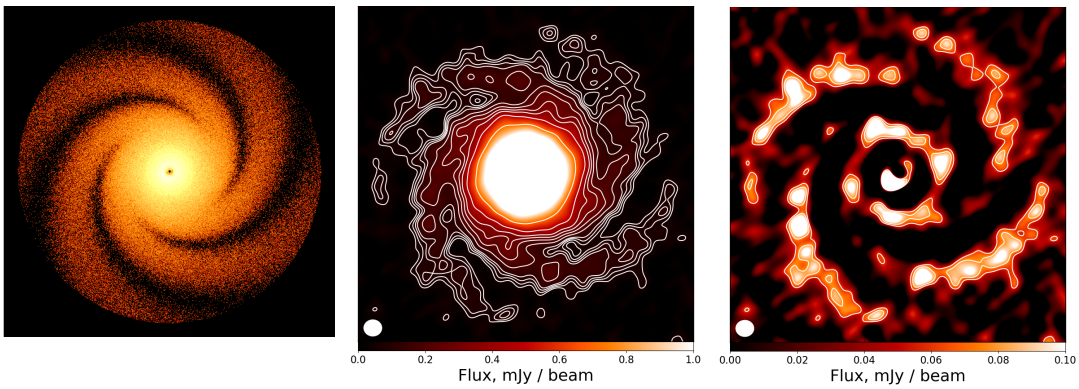

Figure 6.7 *Demonstration of the process generating unsharp masked disc images from* TORUS *radiation transfer continuum profiles. Left: Output continuum disc image from* TORUS*. Middle: Synthetic ALMA observation using* CASA*. Right: Unsharp masked residual image. Discs have properties* $\dot{M} = 5 \times 10^{-7}\, M_{\odot} yr^{-1}$*,* $R_{\rm out} = 100\, AU$*,* $a_{\rm max} = 1\, mm$ *and are observed at* $f_{\rm obs} = 115\, GHz$ *(*$\lambda = 2.6\, mm$*) with observation exposure time, array configuration and PWV level laid out in Table 6.2.*

6.7.1 Analysing the impact of grain enhancement

We begin this section by first demonstrating the impact of grain enhacement on observability. We showed in Section 6.4 that dust trapping of ∼cm sized grains significantly enhances dust-to-gas ratios in spiral arm regions, therefore equally acting to remove dust from interarm regions. Spiral structure consequently becomes sharper and more distinct, producing higher flux ratios between arm and interarm regions due to enhanced and depleted emission at these locations respectively.

We illustrate our grain enhancement prescription in Figure 6.8 by plotting how dust-to-gas ratio varies across our disc model for grains with sizes of $a = 10\,\mu$m, 1 mm and 10 cm, in a disc with $\dot{M} = 1 \times 10^{-6}\,{\rm M}_{\odot}{\rm yr}^{-1}$, $R_{\rm out} = 100\,{\rm AU}$ and grain size distribution $n(a) \propto a^{-3.5}$ with $a_{\rm min} = 0.1\mu$m and $a_{\rm max} = 100\,$cm. Grains of $a = 10\mu$m with ${\rm St} \ll 1$ exactly trace the gas distribution and display an entirely uniform dust-to-gas ratio across the disc. The Stokes number, and therefore also grain concentration factor, η, scales with grain size up to ${\rm St} = 1$. As we consider larger grain sizes up to $a = 10\,$cm, grains become enhanced in the spiral arms and clear non-axisymmetric dust-to-gas ratios start to emerge.

It is useful here to quantify observability of spiral structure in terms of the ratio of the RMS fluxes in the disc arm and interarm regions (i.e. $F_{\rm RMS,arm}/F_{\rm RMS,iarm}$). Arm and interarm regions in our resultant disc images can be located using

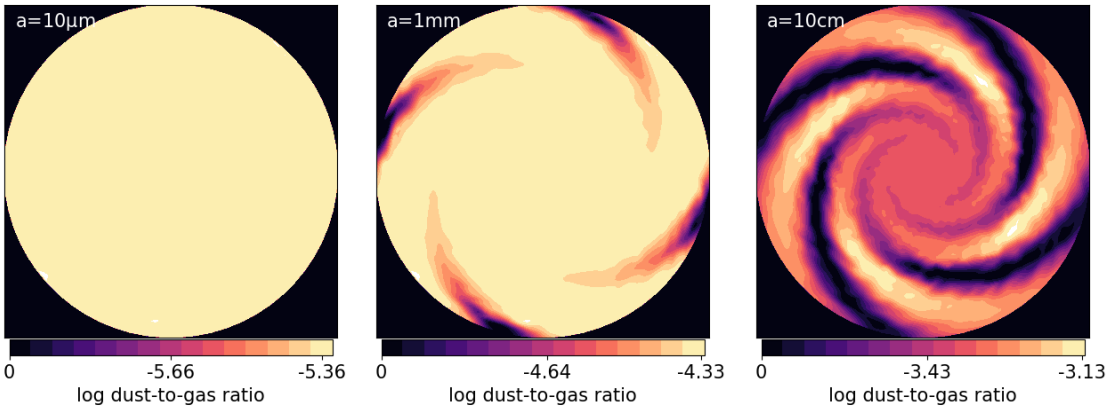

Figure 6.8 *Plotted are the dust-to-gas ratios for individual grain species of different sizes in a disc with* $\dot{M} = 1 \times 10^{-6}\, M_{\odot} yr^{-1}$, $R_{\rm out} = 100\, AU$ *and grain size distribution* $n(a) \propto a^{-3.5}$ *with* $a_{\rm min} = 0.1 \mu m$ *and* $a_{\rm max} = 100\, cm$. *We plot the distributions for grains of sizes* $a = 10 \mu m$ *(Left),* $a = 1\, mm$ *(Middle) and* $a = 10\, cm$ *(Right). We demonstrate the impact of our grain enhancement prescription outlined in Section 6.2.2 as* $\sim$*cm sized grains become highly concentrated in the disc spiral arms. Note that the colourbars are scaled to the maximum dust-to-gas mass ratio in each respective grain size bin,* $\epsilon_{i,\rm max}$.

Equation 6.1, and we calculate the RMS fluxes between radii $70 - 100$ AU where we find spiral structure to be most prominent. In Figure 6.9 we plot how these flux ratios vary with mass accretion rate, and show comparison plots for models that do not include dust grain enhancement in Figure 6.10. Flux ratios are calculated using the synthetic ALMA observations prior to unsharp masking. Example like-for-like unsharp masked disc images are also included for reference in Figure 6.11.

For the same disc parameters we calculate considerably higher flux ratios when including dust trapping in our model, most notably when the dust mass budget is dominated by millimetre/centimetre grains (i.e. when $a_{\rm max} =$ mm$-$cm sizes). Previously blurred arm and interarm regions become distinct as millimetre emission is concentrated in the spiral peaks. The key implication here is that with grain enhancement generating stronger spiral structure for the same mass discs, we should expect to detect self-gravitating disc structure for lower disc masses than previously predicted, if sufficient grain growth has occurred. In discs with no grain growth, or in which grains have grown well beyond centimetre sizes, the lack of dust mass in millimetre/centimetre aggregates is detrimental to the observability of disc substructure.

Given the short potential lifetime of a disc's self-gravitating phase its important to

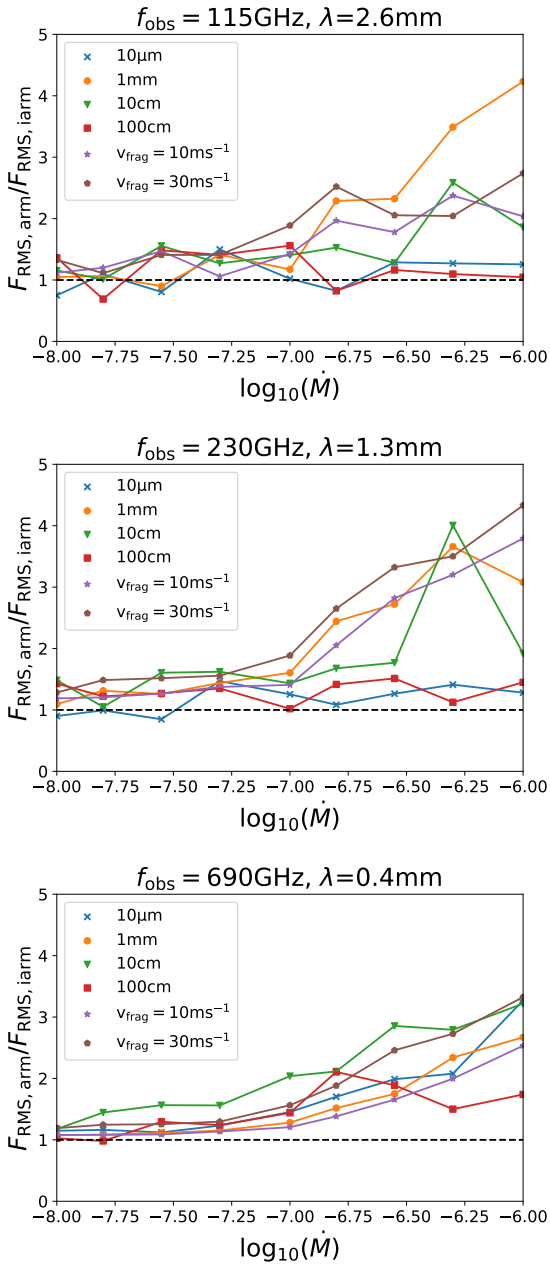

Figure 6.9 *Ratios of RMS fluxes in spiral arm regions to RMS fluxes in interarm regions ($F_{\mathrm{RMS,arm}}/F_{\mathrm{RMS,iarm}}$) plotted against mass accretion rate, $\log_{10}(\dot{M})$, for the discs modelled in Section 6.6 and presented in Section 6.10. These plots are generated using the synthetic ALMA observations prior to performing unsharp image masking.*

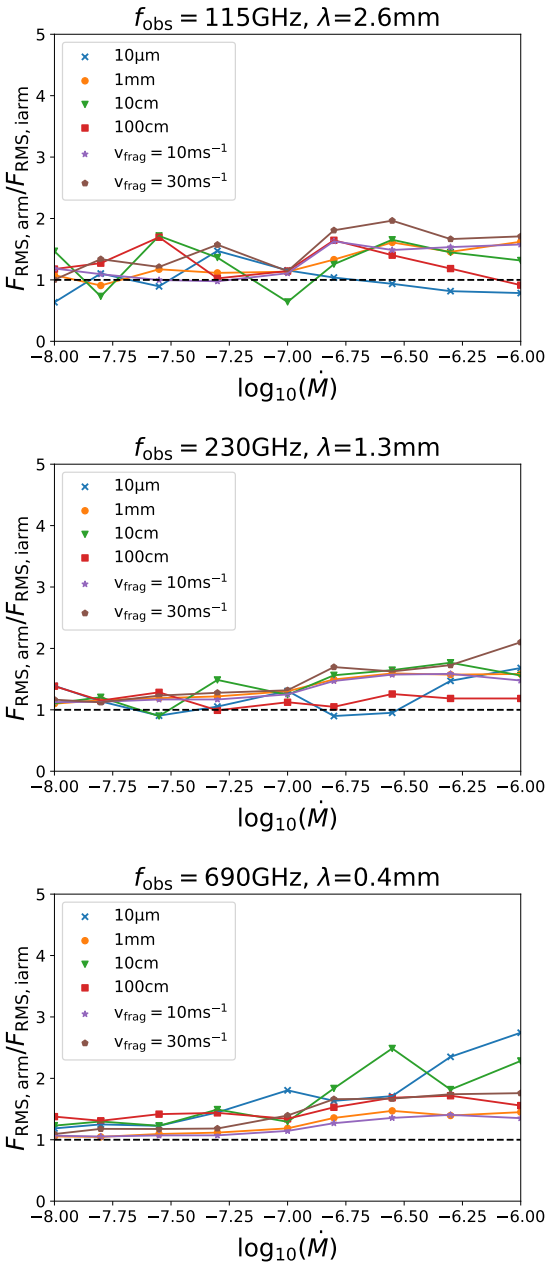

Figure 6.10 *Ratios of RMS fluxes in spiral arm regions to RMS fluxes in interarm regions ($F_{\mathrm{RMS,arm}}/F_{\mathrm{RMS,iarm}}$) plotted against mass accretion rate, $\log_{10}(\dot{M})$, for the discs modelled in Section 6.6. Here we do not account for grain enhancement in spiral arm regions, therefore reducing the prominence of spiral structure in discs compared to their counterparts in Figure 6.9. These plots are generated using the synthetic ALMA observations prior to performing unsharp image masking.*

note how fast grains can actually grow, and thus what likely maximum grain size would be present in young, self-gravitating discs. Models of grain growth predict that millimetre and centimetre-sized grains form rapidly on timescales $\lesssim 10^5$ yrs (Dullemond & Dominik, 2005; Laibe et al., 2008). It therefore seems reasonable to expect grains to have grown to at least as large as the fragmentation threshold before the end of a disc's self-gravitating phase, and that enhanced emission in spiral regions from these larger grains may be significant.

Multi-wavelength observations of discs allow us to probe grain growth and dust trapping through calculation of the disc opacity spectral index, β (Dipierro et al., 2015). In the Rayleigh-Jeans limit of an optically thin disc the dust opacity at sub-mm wavelengths will approximately scale as $\kappa \propto \nu^{\beta}$, where for interstellar dust grains $\beta_{\rm ism} \approx 1.7$. Observations of discs show $\beta_{\rm disc} < \beta_{\rm ism}$ (e.g. Testi et al., 2003; Ricci et al., 2010) which can be naturally accounted for by the presence of larger grains in the disc and therefore grain growth (Draine, 2006). In Figure 6.12 we calculate the β-parameter from our synthetic ALMA observations, considering fluxes $\nu_1 = 460\,{\rm GHz}$ and $\nu_2 = 100\,{\rm GHz}$, and discs with $\dot{M} = 5 \times 10^{-7}\,{\rm M_{\odot}yr^{-1}}$ and $a_{\rm max} = 1\,{\rm mm}$ and $10\,{\rm cm}$. The pixelwise β can be calculated as,

$$\beta = \frac{\ln F_1 - \ln F_2}{\ln \nu_1 - \ln \nu_2} - 2, \tag{6.13}$$

where F_1 and F_2 are the pixelwise fluxes at frequencies ν_1 and ν_2 respectively. Spiral regions display the lowest β values due to dust trapping of larger grains, whilst depletion of these same grains in inter-spiral regions produces comparatively higher β values. Inner disc regions are optically thick and consequently also display low β values. We calculate mean β-values 1.197 and 0.525 for $a_{\rm max} = 1\,{\rm mm}$ and $a_{\rm max} = 10\,{\rm cm}$ respectively, where the higher β value is consequence of less grain growth in the $a_{\rm max} = 1\,{\rm mm}$ disc. Note that both of these discs display $\beta_{\rm disc} < \beta_{\rm ism}$.

Through calculation of the $\beta-$parameter in our disc model we therefore demonstrate how it is possible to retrieve information about the underlying grain distribution in discs, and how our model may be used to probe grain properties in discs which have been observed at multiple wavelengths.

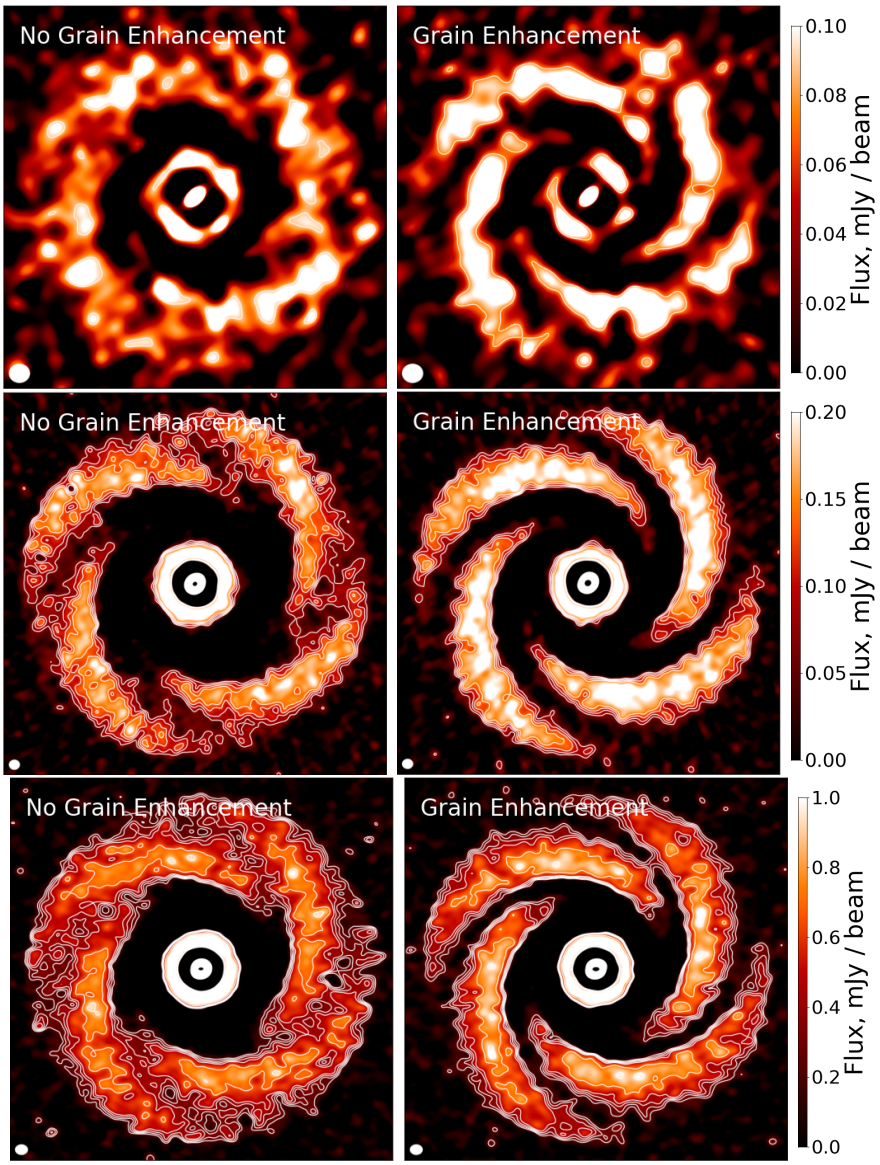

Figure 6.11 *Unsharp masked disc images for* $\dot{M} = 1 \times 10^{-6}\,M_{\odot}\,yr^{-1}$ *and* $a_{\rm max} = v_{\rm frag,30ms^{-1}}$*, observed at frequencies Top:* $115\,GHz$ *(*$\lambda = 2.6\,mm$*), Middle:* $230\,GHz$ *(*$\lambda = 1.3\,mm$*) and Bottom:* $690\,GHz$ *(*$\lambda = 0.4\,mm$*). We compare like for like disc models with our prescription for grain enhancement included (right column) and not included (left column) in the disc models.*

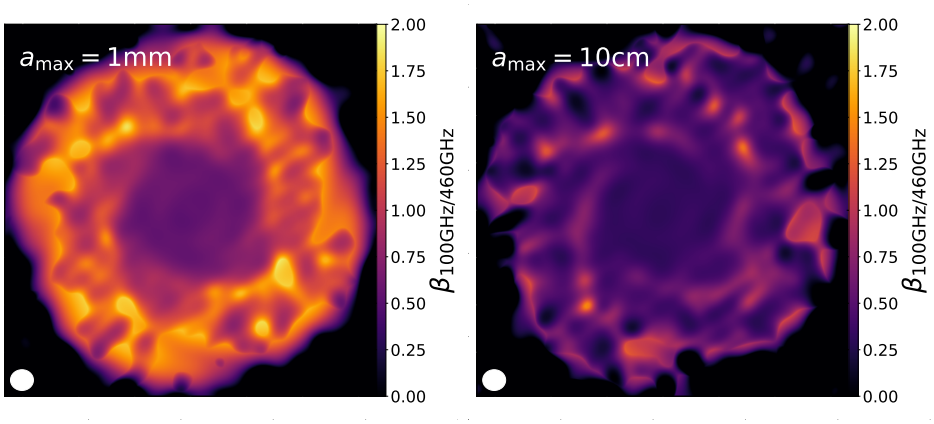

Figure 6.12 *Pixelwise opacity spectral index, β, derived from synthetic observations of discs with $\dot{M} = 5 \times 10^{-7}\, M_{\odot} yr^{-1}$ and grain size distributions $n(a) \propto a^{-3.5}$ where $a_{\rm max} = 1\, mm$ (Left) and $a_{\rm max} = 10\, cm$ (Right).*

6.7.2 Observing self-gravitating discs in Taurus

We now wish to make observational predictions of self-gravitating discs, considering those at a distance $d \sim 140\,$pc comparable to the Taurus star-forming region. We setup a suite of discs as described in Section 6.6 and refer the reader to the unsharp masked disc images presented in Section 6.10 for this discussion.

Spiral amplitude in our models increases as $\delta\Sigma/\Sigma \propto \alpha^{-1/2}$ (Equation 6.2), hence is an increasing function of accretion rate (see equation 3.1). This is illustrated in Figure 6.13 for discs with $a_{\rm max} = 1\,$mm observed at $f_{\rm obs} = 115\,$GHz. Low $\dot{M}$ discs generally exhibit no observable substructure for any grain size distribution, whilst the most massive discs tend to be capable of generating detectable spirals at all frequencies considered here. This does however depend on how much grain growth has occurred, as we require that the dust mass budget is dominated by millimetre/centimetre grains ($a_{\rm max} =$ mm–cm sizes) if we are to resolve any spirals.

Dust emissivity peaks for $\lambda \approx 2\pi a$ (Armitage, 2010), therefore emission from millimetre grains will peak at $\approx$mm wavelengths. The corresponding wavelengths to the observing frequencies considered here are 2.6 mm, 1.3 mm and 0.4 mm for frequencies of 115GHz, 230GHz and 690GHz respectively. When the dust mass budget is dominated by micron grains or metre-sized objects (i.e. $a_{\rm max} = 10\,\mu$m or $a_{\rm max} = 100\,$cm) disc substructure becomes invisible at the ALMA bands considered here as the arm-interarm contrast is low. We illustrate this in Figure 6.14 which shows how emission from spiral regions varies with grain size

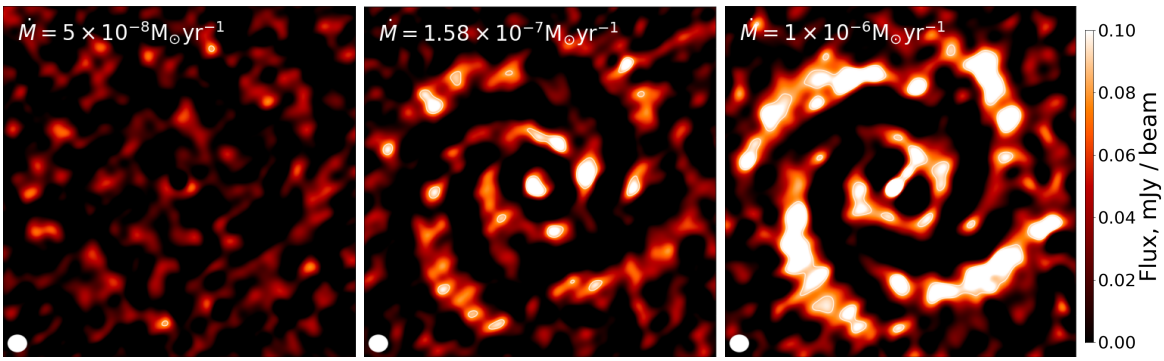

Figure 6.13 *Unsharp masked disc images observed at* $f_{\rm obs} = 115\,GHz$ *(*$\lambda = 2.6\,mm$*) in* CASA*. Each disc model has maximum grain size* $a_{\rm max} = 1\,mm$*,* $R_{\rm out} = 100\,AU$ *and* $\dot{M} = 5 \times 10^{-8}\,M_\odot yr^{-1}$ *(Left),* $\dot{M} = 1.58 \times 10^{-7}\,M_\odot yr^{-1}$ *(Middle),* $\dot{M} = 1 \times 10^{-6}\,M_\odot yr^{-1}$ *(Right). Observation exposure time, array configuration and PWV level used for these observations are laid out in Table 6.2.*

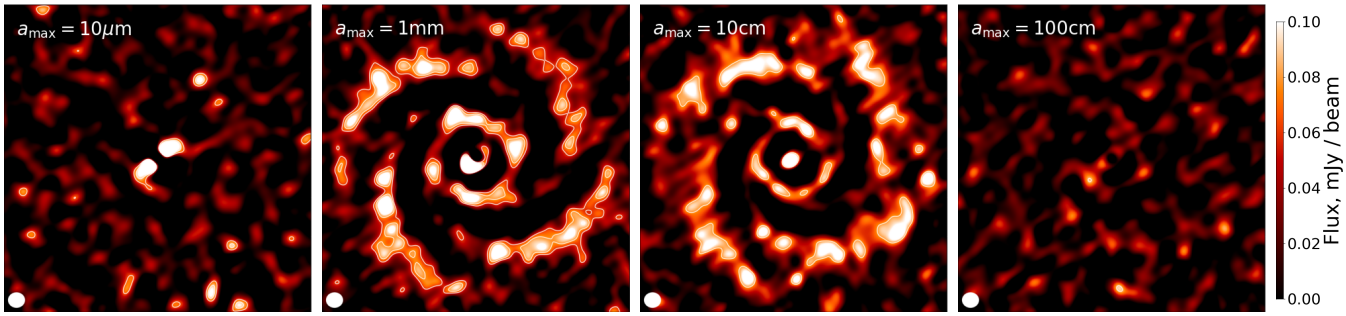

Figure 6.14 *Unsharp masked disc images observed at* $f_{\rm obs} = 115\,GHz$ *(*$\lambda = 2.6\,mm$*) in* CASA*. Each disc has* $\dot{M} = 5 \times 10^{-7}\,M_\odot yr^{-1}$*,* $R_{\rm out} = 100\,AU$ *and we vary* $a_{\rm max}$ *in the grain size distributions as* $10\,\mu m$ *(Left),* $1\,mm$ *(Left middle),* $10\,cm$ *(Right middle) and* $100\,cm$ *(Right). Observation exposure time, array configuration and PWV level used for these observations are laid out in Table 6.2.*

distribution in discs with $\dot{M} = 5 \times 10^{-7}\,\mathrm{M_\odot yr^{-1}}$ observed at $f_{\rm obs} = 115\,\mathrm{GHz}$. Substructure only becomes recognisable in discs with unfavourable grain size distributions when we observe at shorter wavelengths ($f_{\rm obs} = 690\,\mathrm{GHz}$, $\lambda = 0.4\,\mathrm{mm}$), but only in the most highly accreting cases.

Without including dust trapping in their model, Hall et al. (2016) previously found a narrow region of parameter space within which self-gravitating discs would display spirals observable with ALMA. They predicted a 100 AU disc must be accreting in the range $1 \times 10^{-7}\,\mathrm{M_\odot yr^{-1}} \lesssim \dot{M} \lesssim 1 \times 10^{-6}\,\mathrm{M_\odot yr^{-1}}$, where the maximum accretion rate here is set by the limit at which discs become susceptible to fragmentation. We suggest that in fact spiral emission may be distinct for lower accretion rates than previously predicted, if sufficient grain growth has occurred.

The discs in Figures 6.20 and 6.21 observed at 230 GHz and 690 GHz respectively continue to display detectable spiral structure down to the lowest $\dot{M}$ considered here, as long as the dust mass budget is dominated by millimetre/centimetre grains. Note however that we are observing these discs face-on and therefore in favourable conditions for resolving spiral features. Inclining and rotating these discs may well obscure them. However, we would still expect to be able to detect spirals to lower $\dot{M}$ than previously suggested.

It is also intriguing that we calculate the fragmentation threshold to fall almost exactly coincident with the ideal $a_{\rm max}$ values for detecting spirals (see Figures 6.4 and 6.5). We should therefore not be surprised if we find that in fact the grain size distributions of self-gravitating discs fall within this ideal region of parameter space.

6.8 Analysing discs from the DSHARP sample

We now turn our model to analysing real observational data of potentially self-gravitating discs. The recent DSHARP survey studied 20 nearby protoplanetary discs using ALMA, with 3 of these discs exhibiting spiral substructure (Andrews et al., 2018b; Huang et al., 2018b). The ALMA continuum images from this survey of the Elias 2-27, WaOph 6 and IM Lup discs are shown in Figure 6.15. We use our models to investigate if the observed spiral substructure in these 3 systems can be explained through the gravitational instability.

Although well within the capability of our models, a complete examination of the potential parameter space of these discs is beyond the scope of the work presented here. Instead, we simply model these 3 systems using the disc parameters derived in Andrews et al. (2018b) and Huang et al. (2018b), and make predictions as to whether we should expect these systems to produce self-gravitating spiral substructure observable with ALMA. The disc parameters used are laid out in Table 6.3. We setup these discs with dust size distribution $n(a) \propto a^{-3.5}$, with $a_{\rm min} = 0.1\mu$m and set $a_{\rm max}$ as the fragmentation threshold where $v_{\rm frag} = 10\,{\rm ms}^{-1}$ (Equation 6.11), and use the canonical dust-to-gas ratio of 0.01.

Residual images in Huang et al. (2018b) are produced by deprojecting the discs and subtracting their median axisymmetric radial profiles. We do the same here by binning each disc into 1 AU-wide radial bins and subtracting the median

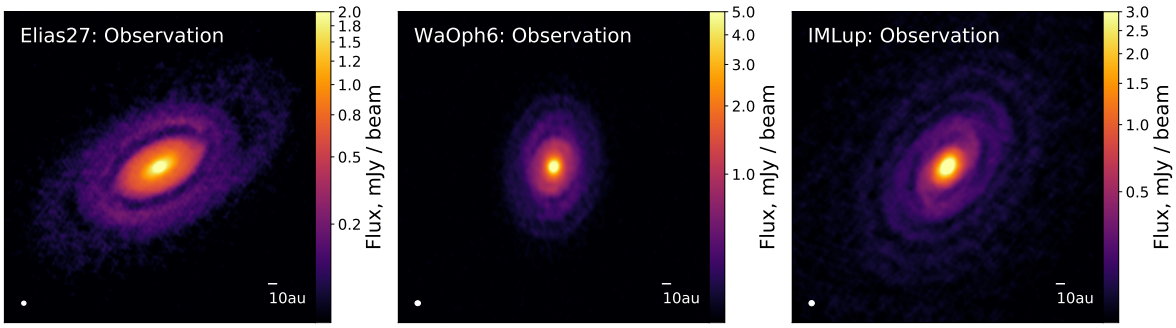

Figure 6.15 *ALMA 240GHz (1.3mm) continuum images of Elias 2-27 (Left), WaOph 6 (Middle) and IM Lup (Right) (Andrews et al., 2018b). ALMA antenna configurations, observing times and PWV levels for these observations are laid out in Table 6.4.*

azimuthal fluxes. We re-derive the residual images for each of the original DSHARP observations in this way, as well as for our disc models. For each disc observation and model, we show deprojected continuum and residual images (with PA $= 0°$ and $i = 0°$), presenting our results in Figures 6.16, 6.17 and 6.18. In each case we provide reference colorbars for direct comparison between the fluxes of the disc models and observations, and each disc model and counterpart observation is plotted between the same flux range for ease of comparison.

Logarithmic spiral structure is imposed in each disc model using values of a and b (Equation 6.1) derived by Huang et al. (2018b). The values of a and b used here are laid out in Table 6.3.

We produce synthetic observations of each disc using CASA with observing setups consistent with those outlined in Andrews et al. (2018b). We observe each disc for $t_{\rm obs} = 3600\,$s using array configuration C40-8. For each observation we use PWV values at the upper bound of the quoted range in Andrews et al. (2018b), setting values 1.35 mm, 1.30 mm and 1.05 mm for Elias 2-27, IM Lup and WaOph 6 respectively. Input parameters for CASA used for each disc are laid out in Table 6.4.

6.8.1 Elias 2-27

Elias 2-27 is a 0.8 Myr M0 star located in the ρ Oph star forming region at a distance $d = 116^{+19}_{-10}\,$pc (Gaia Collaboration et al., 2018; Andrews et al., 2018b). The residual profile of the Elias 2-27 continuum image (Figure 6.16) shows two symmetric spiral arms extending from $R_{\rm in} \sim 50\,$AU to $R_{\rm out} \sim 230\,$AU, with

Disc	$\log M_*$ ($M_\odot$)	R_* (AU)	$\log T_{\rm eff}$ (K)	$R_{\rm spirals}$ (AU)	$\log \dot{M}$ ($M_\odot\,{\rm yr}^{-1}$)	d (pc)	i (°)	PA (°)	a (AU)	b
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Elias 2-27	$-0.31^{+0.15}_{-0.11}$	2.3	3.59 ± 0.03	50-230	-7.2 ± 0.5	116^{+19}_{-10}	56.2	118.8	110.9	-0.282
WaOph 6	$-0.17^{+0.17}_{-0.09}$	3.2	3.62 ± 0.03	25-75	-6.6 ± 0.5	123±2	47.3	174.2	45.9	0.238
IM Lup	$-0.05^{+0.09}_{-0.13}$	2.5	3.63 ± 0.03	25-110	-7.9 ± 0.4	158±3	47.5	144.5	43	-0.181

Table 6.3 *Disc model parameters used in our modelling of the DSHARP discs in Section 6.8. Columns are as follows. (1) Disc being modelled. (2) Log stellar mass. (3) Stellar radius. (4) Log effective temperature of the star. (5) Spiral inner and outer radii considered here. (6) Log mass accretion rate. (7) Distance to the system. (8) Disc inclination. (9) Disc position angle. (10) Best-fit logarithmic spiral a (Equation 6.1). (11) Best-fit logarithmic spiral b (Equation 6.1).*

Disc	$f_{\rm obs}$ (1)	$t_{\rm obs}$ (2)	Antenna Config (3)	PWV Level (4)
Elias 2-27	240 GHz	3600 s	C40-8	1.35 mm
WaOph 6	240 GHz	3600 s	C40-8	1.30 mm
IM Lup	240 GHz	3600 s	C40-8	1.05 mm

Table 6.4 *Input parameters used here for generating synthetic images with* CASA *for the modelled DSHARP discs. (1) ALMA observing frequency. (2) Simulated observing time. (3) ALMA antenna configuration used. (4) Precipitable Water Vapour (PWV) level.*

PA = 118.8° and $i = 56.2°$ (Huang et al., 2018b).

The spiral structure of Elias 2-27 is probably the most well-studied of the three discs here. The system has previously been modelled using both both grid-based and SPH simulations, with authors such as Meru et al. (2017), Tomida et al. (2017) and Hall et al. (2018) all finding GI to be a plausible explanation for the observed morphology. Estimates of the Toomre parameter in the disc however suggest that Elias 2-27 should be gravitationally stable at all radii (Pérez et al., 2016), but this comes with the caveat that estimates of Q are subject to high levels of uncertainty. Further research where the constraints on the disc mass and temperature are improved may lead to different conclusions.

We set up our disc model with $\log M_*(\mathrm{M}_\odot) = -0.31$, $R_* = 2.3\,\mathrm{AU}$, $\log T_{\mathrm{eff}}(\mathrm{K}) = 3.59$ and $\log \dot{M}(\mathrm{M}_\odot \mathrm{yr}^{-1}) = -7.2$ (Andrews et al., 2018b; Huang et al., 2018b). Logarithmic spiral structure is imposed with $a = 76.0\,\mathrm{AU}$ and $b = -0.29$ extending from $R = 50 - 230\,\mathrm{AU}$, where we use a mask to remove the inner 50 AU from our observations to avoid the spirals being washed out by the brighter central region.

Our model calculates Elias 2-27 to have a disc mass $M_{\mathrm{disc}} = 0.13\,\mathrm{M}_\odot$ inside $R_{\mathrm{out}} = 230\,\mathrm{AU}$, and therefore $q = 0.27$. Figure 6.16 shows the resultant synthetic observations generated from our models, exhibiting clear self-gravitating spiral structure in both the deprojected continuum and residual images.

6.8.2 WaOph 6

WaOph 6 is a 0.3 Myr K6 star located in the ρ Oph star forming region at a distance $d = 123 \pm 2\,\mathrm{pc}$ (Gaia Collaboration et al., 2018; Andrews et al., 2018b). After subtracting the axisymmetric radial profile, two compact spiral arms are revealed which extend from $R_{\mathrm{in}} \sim 25\,\mathrm{AU}$ to $R_{\mathrm{out}} \sim 75\,\mathrm{AU}$, with PA = 174.2° and $i = 47.3°$ (Huang et al., 2018b).

In their analysis of the morphology of gravitationally unstable discs, Dong et al. (2015b) suggest that for a disc to be gravitationally unstable it must be compact ($R \leq 100\,\mathrm{AU}$) and highly accreting at a rate $\dot{M} \geq 10^{-6}\,\mathrm{M}_\odot \mathrm{yr}^{-1}$. WaOph 6 has the highest accretion rate and the most compact spiral structure of the 3 discs in question here, both of which are close to matching these suggested criteria.

We setup our disc model with $\log M_*(\mathrm{M}_\odot) = -0.17$, $R_* = 3.2\,\mathrm{AU}$, $\log T_{\mathrm{eff}}(\mathrm{K}) =$

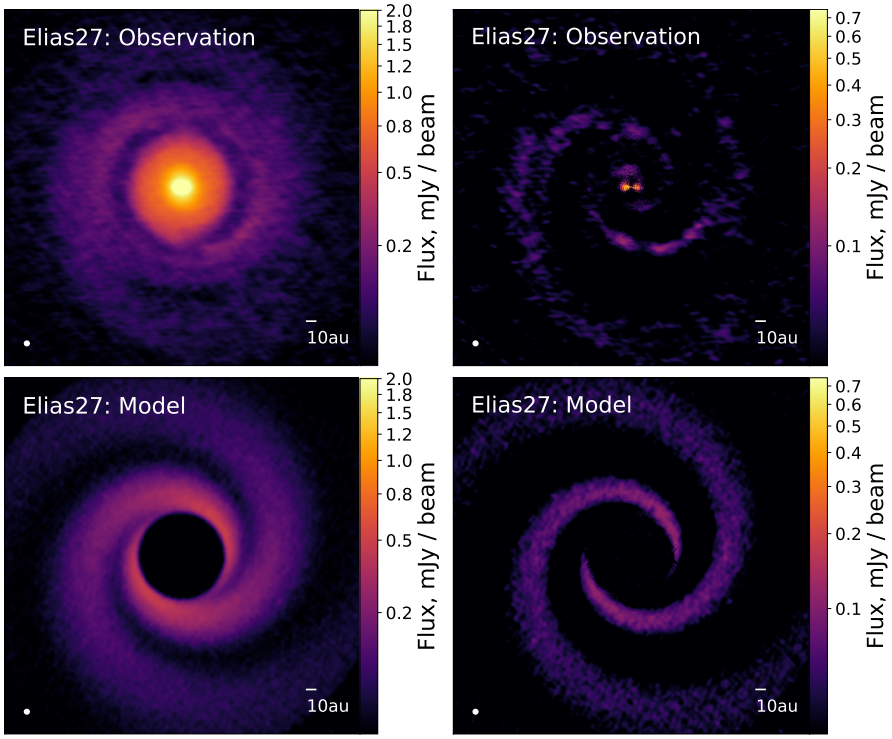

Figure 6.16 *Elias 2-27 discs images. Top: Deprojected ALMA continuum image (left) and residual profile (right). Bottom: Deprojected disc model continuum image (left) and residual profile (right). Input properties for the disc models and observation parameters are laid out in Tables 6.3 and 6.4.*

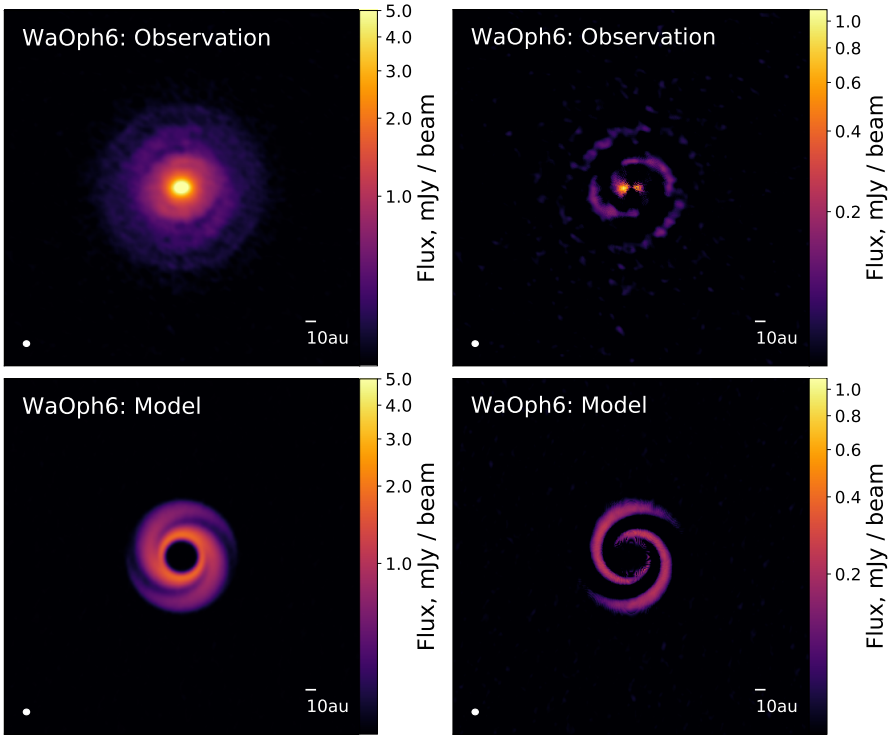

Figure 6.17 *WaOph 6 discs images. Top: Deprojected ALMA continuum image (left) and residual profile (right). Bottom: Deprojected disc model continuum image (left) and residual profile (right). Input properties for the disc models and observation parameters are laid out in Tables 6.3 and 6.4.*

3.62 and $\log\dot{M}(\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = -6.6$ (Andrews et al., 2018b; Huang et al., 2018b). Logarithmic spirals are imposed with $a = 34.0\,\mathrm{AU}$ and $b = 0.24$ extending from $R = 25 - 75\,\mathrm{AU}$, where again we mask the inner 25 AU of the disc images.

We calculate WaOph 6 to have a disc mass $M_{\mathrm{disc}} = 0.16\,\mathrm{M}_{\odot}$ and therefore $q = 0.24$. Our models reproduce distinct observable, self-gravitating spiral structure in both the deprojected continuum and residual images shown in Figure 6.17.

6.8.3 IM Lup

IM Lup is a 0.5Myr K5 star in the Lupus II cloud at a distance $d = 158 \pm 3\,\mathrm{pc}$ (Gaia Collaboration et al., 2018; Andrews et al., 2018b). Residual profiles of the IM Lup continuum images reveal two spirals extending from $R_{\mathrm{in}} \sim 25\,\mathrm{AU}$ to $R_{\mathrm{out}} = 110\,\mathrm{AU}$, with PA= 144.5° and $i = 47.5°$ (Huang et al., 2018b).

Previous detection of any spiral structure in the IM Lup system has been minimal, with observed substructures being classified as two concentric rings at

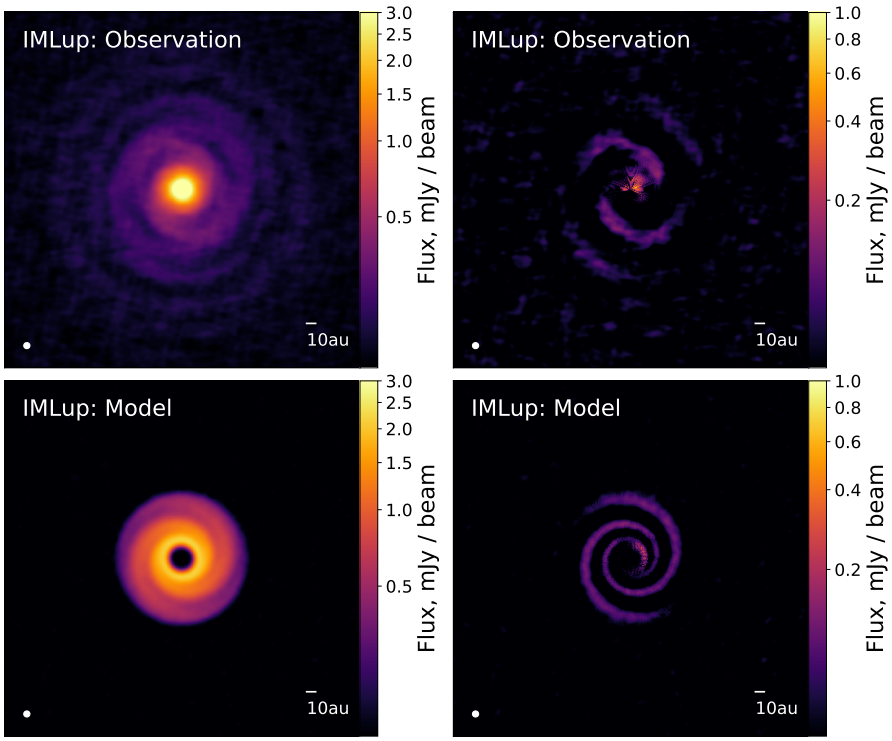

Figure 6.18 *IM Lup discs images. Top: Deprojected ALMA continuum image (left) and residual profile (right). Bottom: Deprojected disc model continuum image (left) and residual profile (right). Input properties for the disc models and observation parameters are laid out in Tables 6.3 and 6.4.*

$R \approx 95\,\mathrm{AU}$ and $R \approx 320\,\mathrm{AU}$, and only tenuous reports of the possibility of tightly wound spirals (Avenhaus et al., 2018). Cleeves et al. (2016) report a massive, gravitationally stable disc with a minimum Toomre parameter $Q_{\mathrm{min}} = 3.7$ at $R = 70\,\mathrm{AU}$ and an extended CO disc to $R = 970\,\mathrm{AU}$, making IM Lup one of the largest protoplanetary discs detected to date.

We model the disc here out to $R_{\mathrm{out}} = 110\,\mathrm{AU}$, consistent with the radial extent of the observed spiral structure reported in Huang et al. (2018b). Our disc model is setup with $\log M_*(\mathrm{M}_\odot) = -0.05$, $R_* = 2.5\,\mathrm{AU}$, $\log T_{\mathrm{eff}}(\mathrm{K})$ and $\log \dot{M}(\mathrm{M}_\odot \mathrm{yr}^{-1}) = -7.9$ (Andrews et al., 2018b; Huang et al., 2018b). We impose logarithmic spiral structure with $a = 59.0\,\mathrm{AU}$ and $b = 0.18$ extending from $R = 25 - 110\,\mathrm{AU}$.

We calculate IM Lup to have a disc mass $M_{\mathrm{disc}} = 0.098\,\mathrm{M}_\odot$ and $q = 0.11$ within $R = 110\,\mathrm{AU}$, and therefore the lowest disc-to-star mass ratio of the three discs modelled here. The deprojected disc images in Figure 6.18 show tightly wound spiral structure in the continuum and residual images, with geometry and spiral fluxes closely matching those observed in the inner disc of the IM Lup system.

6.8.4 Conclusions on DSHARP sample

We present the results of our semi-analytic analysis of the observed spiral structure in the Elias 2-27, WaOph 6 and IM Lup systems. We note again that the purpose of this simple functional formalism is not to exactly reproduce, but to approximate, the likely spiral morphologies and fluxes of the 3 systems in question, and to investigate whether systems of their quoted disc and stellar properties should be capable of generating detectable non-axisymmetric substructure when observed with ALMA. We do this by imposing logarithmic spiral structure characteristic of GI, with self-consistently calculated spiral amplitudes and realistic grain distributions. All 3 of the models presented here produce detectable spirals of comparable structure and fluxes to their observed DSHARP counterparts, indicating that GI may be the dominant mechanism responsible for the observed substructure in these discs.

For Elias 2-27, WaOph 6 and IM Lup we derive disc masses $M_{\rm disc} = 0.13\,{\rm M}_{\odot}, 0.16\,{\rm M}_{\odot}$ and $0.098\,{\rm M}_{\odot}$ and disc-to-star mass ratios $q = 0.27, 0.24$ and 0.11 within their respective outer radii. Common assumption is that GI requires $q \gtrsim 0.5$, therefore rendering these discs too low mass to generate prominent self-gravitating structure. However it may be possible for discs to display self-gravitating spirals for much lower mass ratios than previously thought, with the critical mass ratio having a strong dependence on the host star mass and disc opacity (Veronesi et al., 2019; Cadman et al., 2020a; Haworth et al., 2020). We therefore should caution against discarding GI as a plausible mechanism based off this criterion alone.

It is important to note here that whilst we shouldn't be surprised that our models accurately reproduce the spiral form of the systems considered here, as the geometry is imposed in Equation 6.1, we should be more concerned with how accurately our models are able the reproduce the spiral flux amplitudes of the observed systems, as these are determined self-consistently from the disc mass accretion rate and the viscous-α. The self-consistently calculated spiral amplitudes in our models all generate comparable fluxes to their counterpart observations, indicating that self-gravity may be a plausible explanation to these 3 systems.

In our model we assume that some grain growth has occurred up to the fragmentation threshold. We note again here that models of grain growth generally suggest that centimeter aggregates form rapidly on timescales $\sim 10^5$ yrs

(Dullemond & Dominik, 2005; Laibe et al., 2008), therefore given the ages of these 3 systems our assumption seems reasonable. If, however, we modelled these systems assuming no grain growth, it is likely that we would not find any signatures of GI. Therefore, if these discs are indeed gravitationally unstable, our models suggest that significant grain growth must have also occurred. Future multi-wavelength observations of these systems, and derivation of the discs' β-parameter (Equation 6.13), will help to establish if this is the case.

An alternative explanation for the observed spiral structure in the DSHARP discs may be the presence of a stellar or planetary-mass companion. Planet-disc gravitational interactions can generate disc perturbations, and massive companions may be capable of triggering two-armed symmetric spiral responses similar to those observed in DSHARP (e.g. Dong et al., 2015a, 2016; Bae & Zhu, 2018a,b; Kurtovic et al., 2018). However in order to drive the spiral modes observed, for example in the Elias 2-27 system, would require a wide-orbit companion of potentially tens of Jupiter masses, thus rendering any companion likely detectable at sub-mm/IR wavelengths (Meru et al., 2017). To our knowledge no companion has as yet been detected in any of the 3 discs observed here. More commonly associated features of planet-disc interactions are the presence of annular substructures such as rings and planet-driven gaps. Elias 2-27, WaOph 6 and IM Lup all display these features, as do a total of 18 discs in the DSHARP sample (Huang et al., 2018a). The DSHARP collaboration report no companion detections in any of these 18 discs despite many of the observed features being suggestive of massive companions which ought to be observable at such high angular resolution. It may then be the case that either massive planets are fainter than previously thought (Dong et al., 2018), or that the observed rings are driven by lower mass, fainter planets which remain invisible to the DSHARP survey. If the latter, then these lower mass companions may not be capable of driving the observed spiral structure in Elias 2-27, WaOph 6 and IM Lup alone, but a combination of both GI and planet-disc interactions may be a plausible scenario (e.g. Pérez et al., 2016).

More detailed analysis of these systems, investigating the effect of varying accretion rate, disc irradiation, the dominant spiral mode and grain size distribution will be the subject of future work.

6.9 Summary and conclusion

We present our updated self-consistent, semi-analytic model of self-gravitating discs that also includes a prescription for dust trapping. We make use of the efficient nature of the model by generating a suite of disc models at little computational expense, and examine the parameter space within which we predict self-gravitating discs will generate spiral structure that can be resolved when imaged with ALMA. Monte-Carlo radiative transfer is employed here to produce synthetic observations of these model discs, allowing us to make realistic predictions about the strength of the perturbations and the grain size distribution required to generate observable spiral structure.

Realistic dust trapping is modelled using a semi-analytic prescription in which particles with $\mathrm{St} = 1$ may reach grain concentration factor $\eta \approx 6$ at the density peaks of the spiral perturbations, where η represents the local dust enhancement relative to the mean dust-to-gas ratio in the disc, assumed to be 0.01 in all the models considered here. We find that particles of millimetre and centimetre sizes concentrate most strongly in spiral arms resulting in significantly enhanced millimetre emission in these regions. When the dust mass budget is dominated by these millimetre and centimetre sized grains we find self-gravitating structure to be observable in much lower mass discs than previously predicted. Through calculation of the grain fragmentation threshold in the discs modelled here we find that grains may only grow to as large as a few centimetres before grain-grain collisions become destructive. Therefore it may be the case that grain size distributions in self-gravitating discs satisfy this dust mass budget criterion.

Our synthetic unsharp masked images of discs in the Taurus star forming region ($d \sim 140\,\mathrm{pc}$) exhibit distinguishable spiral structure for disc masses as low as $q = 0.1$ given that sufficient grain growth has occurred. These images are generated using realistic ALMA observing setups with reasonable observing times and PWV levels. We do however note that we only consider face-on discs during this evaluation and that inclining and rotating them may well obscure any substructure, likely most adversely in low mass discs with the weakest spirals.

Through multi-wavelength observations and derivation of the $\beta-$parameter we show how it is possible to retrieve information about grain growth and the dust-to-gas ratio distribution from our model discs. Through comparison of our predicted $\beta-$values to those calculated from future multi-wavelength observations of self-

gravitating discs, it may be possible to utilise our disc model to examine grain distributions in the observed discs.

Applying our disc model to systems from the DSHARP sample, we find the quoted disc parameters for Elias 2-27, WaOph 6 and IM Lup suggest that they are all capable of driving observable, self-gravitating spiral structure providing that grains have grown to as large as the fragmentation threshold. We calculate disc-to-star mass ratios $q = 0.27, 0.24$ and 0.11, within their published outer radii, respectively for the 3 systems. A more detailed analysis exploring the potential parameter space of the DSHARP sample will be left to future work.

6.10 Gallery of discs

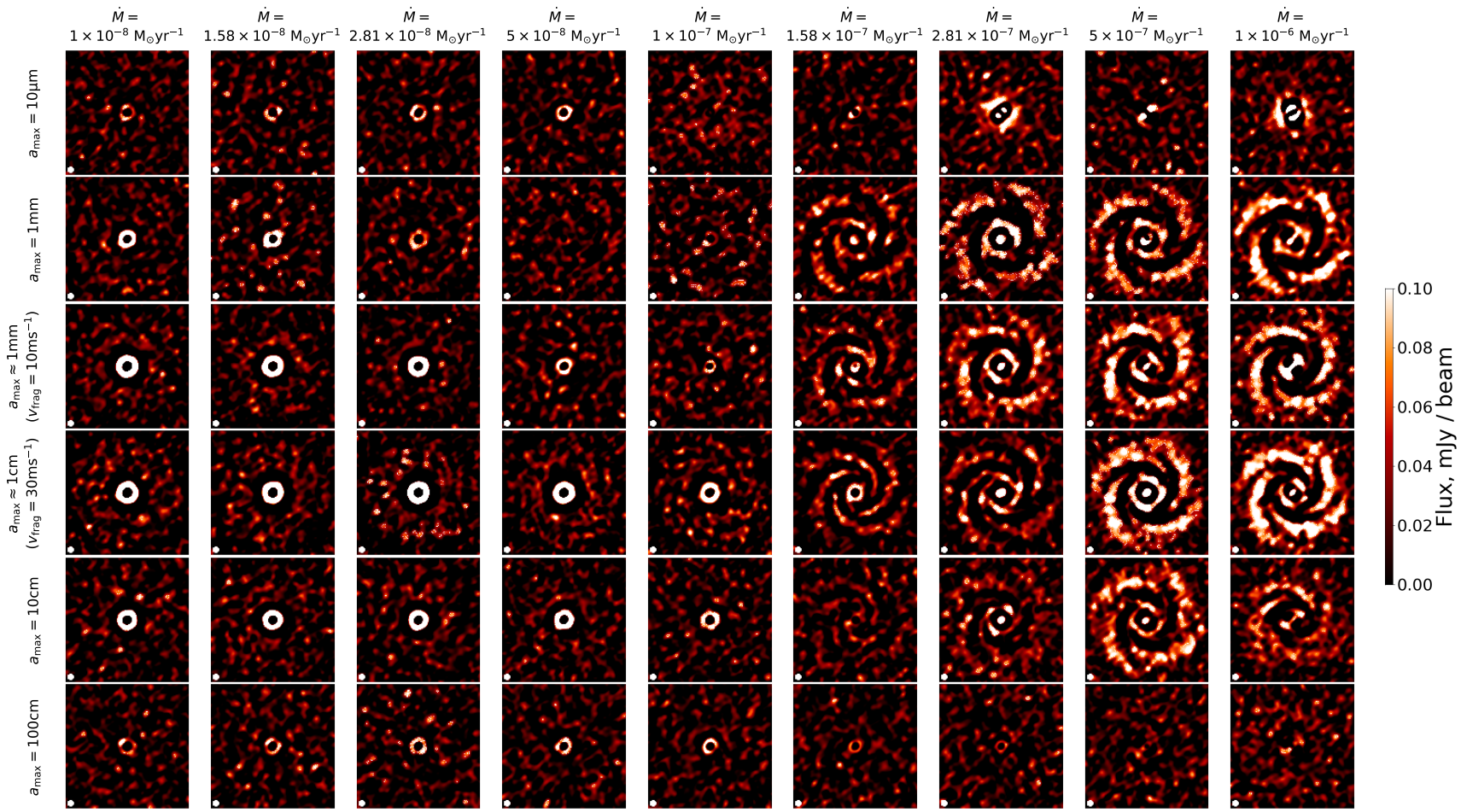

Figure 6.19 *Gallery of discs observed at* $f_{\rm obs} = 115 GHz$ ($\lambda = 2.6\,mm$). *Disc setups are described in Section 6.6.* CASA *observing inputs are laid out in Table 6.2.*

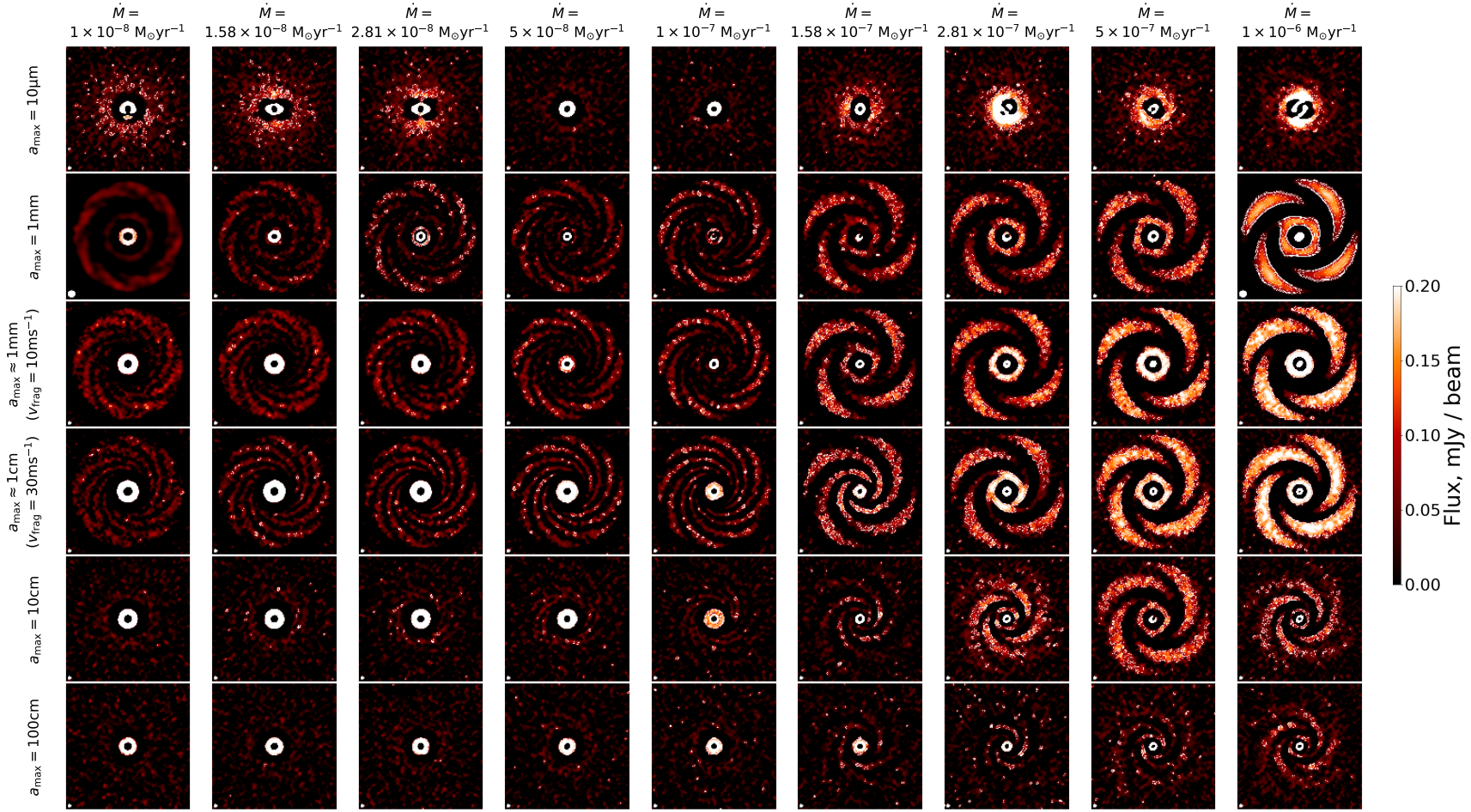

Figure 6.20 *Gallery of discs observed at* $f_{\rm obs} = 230 GHz$ *(*$\lambda = 1.3\,mm$*). Disc setups are described in Section 6.6.* CASA *observing inputs are laid out in Table 6.2.*

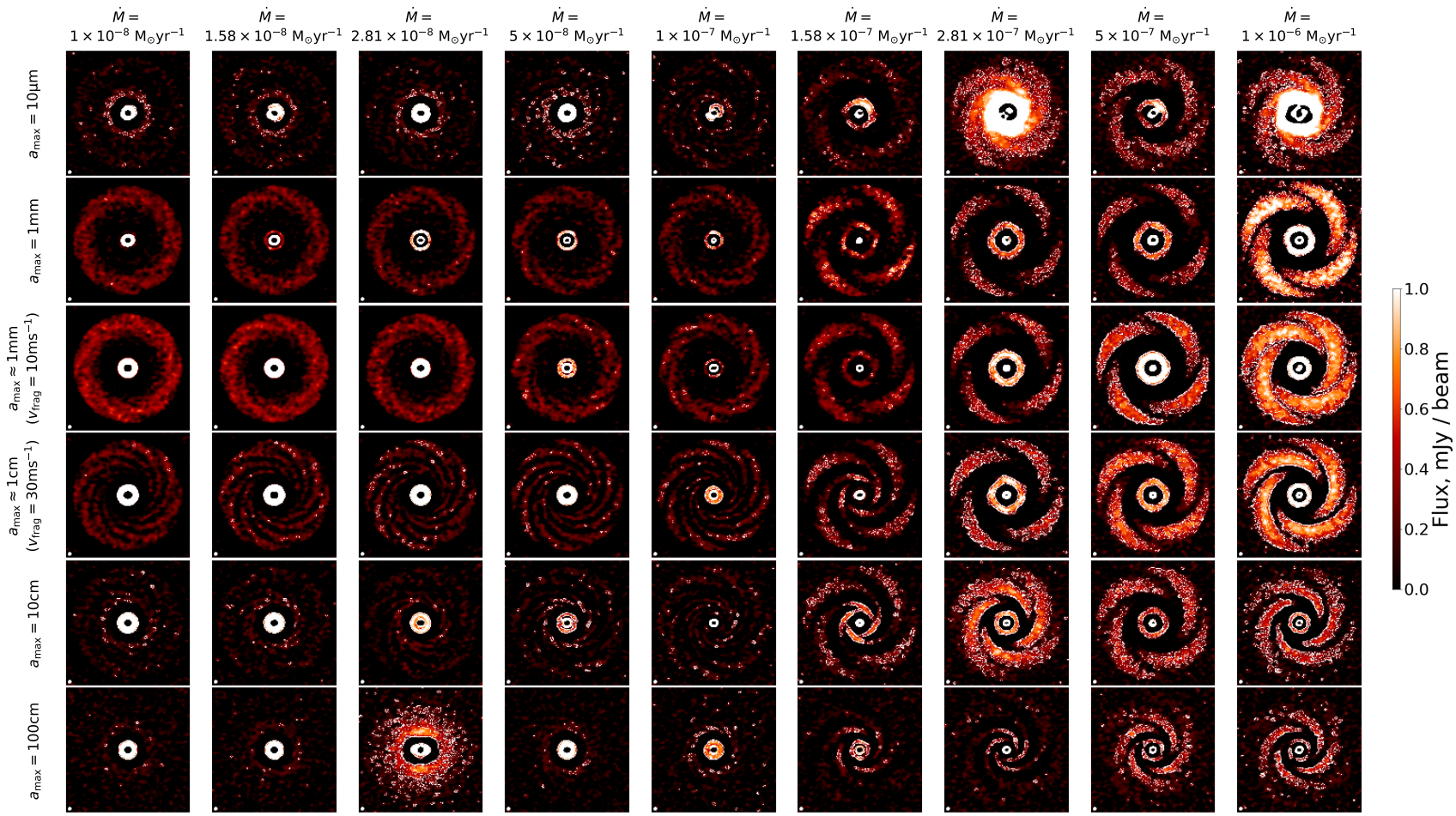

Figure 6.21 *Gallery of discs observed at* $f_{\rm obs} = 690\,GHz$ *(*$\lambda = 0.4\,mm$*). Disc setups are described in Section 6.6.* CASA *observing inputs are laid out in Table 6.2.*

Chapter 7

Outlook

7.1 Thesis summary

The gravitational instability model of planet formation provides a promising pathway for the formation of wide-orbit, giant planets whose formation histories remain puzzling to today's models of core accretion. At the time of writing, the jury is still out as to whether planets can form through GI, although promising strides are being made toward settling this debate. Today's surveys are capable of imaging discs at the earliest stages of their formation, revealing systems with substructures resembling GI (Pérez et al., 2016; Andrews et al., 2018a; Huang et al., 2018b), as well as discs which may be in the process of fragmenting (Tobin et al., 2016; Boccaletti et al., 2020). Results such as these allow us to test our theories, and help us to understand how commonplace GI systems and GI-born planets may be in nature.

The research undertaken throughout this thesis aims to contribute to our understanding of GI's role in protoplanetary disc evolution. I began by modelling how a disc's susceptibility to fragmentation may be influenced by the mass of its parent star. Using SPH simulations to explore the disc-star parameter space I found that discs become more susceptible to fragmentation as the star's mass is increased. This is because the critical disc-to-star mass ratio for fragmentation decreases with increasing stellar mass. This indicates that wide-orbit, giant planets and brown dwarfs should be more commonly found around higher mass stars, consistent with results from recent observational surveys (Vigan et al., 2012; Nielsen et al., 2013; Lannier et al., 2016; Nielsen et al., 2019). Additionally,

these results predict that low mass stars can support high mass discs which don't fragment, allowing for large mass reservoirs to be available for core accretion planet formation in these systems, hence higher planet formation efficiencies.

In Chapter 4 I used SPH to explore the parameter space of self-gravitating discs which are part of binary star systems. Recent observations suggest that close-in, massive planets and brown dwarfs are almost always accompanied by an outer stellar companion (Fontanive et al., 2019), suggesting that the companion may play an important role in the formation of these systems. Such planets/BDs are distinct from "hot Jupiters" ($M \approx 0.2 - 4\,\mathrm{M_{Jup}}$) as they are more massive ($M > 7\,\mathrm{M_{Jup}}$), and their stellar hosts exhibit a significantly lower mean metallicity than hosts of hot Jupiters; a possible telltale sign that these planets formed through GI. Using SPH I identified a sweet spot in the binary companion's parameter space where the companion can trigger fragmentation in a disc that did not fragment in isolation. Intermediate separation binaries may drive a spiral arm through the disc, and combined with efficient cooling the spiral may fragment. The range of identified favourable binary separations is consistent with the projected separations of the systems which are found to have an excess of close-in, giant planets, suggesting that this triggered fragmentation may have played a role in the formation of these systems. The question remains how these GI-born fragments, initially formed on wide orbits, might then migrate to the short orbits on which they are currently observed if GI is to be a viable formation mechanism for these systems.

I then use SPH combined with 1D models to analyse the young disc surrounding AB Aurigae, focusing on the formation histories of the protoplanets which have recently been identified in the disc through high resolution imaging. AB Aurigae has an estimated age of $1 - 4\,\mathrm{Myr}$ and shows strong evidence that at least 2 wide-orbit, massive protoplanets are in the process of forming within the disc. Focusing on the formation history of the $4 - 13\,\mathrm{M_{Jup}}$ protoplanet observed at $R \approx 30\,\mathrm{AU}$ from the star, I show that core accretion models struggle to form such a planet within the current lifetime of the disc. Instead I show that, given the current properties of the system, the disc was likely massive enough to undergo fragmentation in the not-so-distant past, and the expected properties of a GI fragment are consistent with the properties of the observed protoplanet. These results indicate that the AB Aurigae may be the first system to show evidence of ongoing GI planet formation. Since writing this thesis observations of AB Aurigae have detected further ongoing clump formation in the disc, revealing

another super-Jupiter mass object at $R \approx 93\,\mathrm{AU}$ which is indicative of planet formation through GI (Currie et al., 2022).

Finally, I use semi-analytic models combined with synthetic disc imaging to make observational predictions of self-gravitating discs, focusing on the impact of dust-trapping in the disc's spiral arms. I use SPH to show that dust-to-gas ratios may be enhanced by a factor $d \approx 5-6$ at the peaks of GI spirals, leading to enhanced millimeter emission in these regions thus improving our prospects of observing GI discs. The results indicate that systems with disc-to-star mass ratios as low as $q = 0.1$ may be capable of driving spirals which are detectable with ALMA, as long as sufficient grain growth up to $\sim$mm-cm sizes has already occurred in the disc. These disc masses are lower than have been previously predicted by models which don't consider the effect of dust trapping. We also use these models to show that the amplitudes of the observed spirals in the discs around Elias 2-27, WaOph 6 and IM Lup are all consistent with spirals driven by GI. Future high resolution, multi-wavelength observations and improved mass constraints for these discs will help to reveal whether this is in fact the case.

7.2 Ongoing work

The research in chapters 3 and 4 focuses on the disc conditions required for fragmentation during the very early phases of the disc's evolution. However, it is currently not possible to compare these predictions to systems in nature, as observations of very young, gravitationally unstable discs are rare as the GI phase is short lived hence challenging to observe. Understanding the long term evolution of these systems and the fragments which form through GI will help us to make predictions as to what the resultant planetary systems containing GI-born planets should look like, and identify and distinguishable features from non-GI planets.

Modelling the long term evolution of systems is computationally expensive using full hydrodynamic treatments such as SPH. Instead, 1D population synthesis codes such as those in Forgan & Rice (2013b); Forgan et al. (2018) can be used to generate population statistics of planets which formed through GI, which can be compared to synthetic populations of planets which formed through core accretion (e.g. Mordasini, 2018), and used to determine any telltale signatures of GI planetary systems.

Of particular interest to the work in Chapter 3, we expect the eventual planetary systems which form around higher mass stars to look different to those which form around lower mass stars as we predict that a higher frequency of GI planets will form. Models such as those in Forgan & Rice (2013b); Forgan et al. (2018) allow us to generate synthetic population statistics which can be compared to observed population statistics to test this theory. Currently work is being done to build on the models in Forgan & Rice (2013b); Forgan et al. (2018) before progressing with this, where a stage of disc formation through cloud collapse is being incorporated, and the disc is now being evolved with the planets simultaneously.

Equally, highly parallelisable grid-based hydrodynamic codes such as FARGO (Benítez-Llambay & Masset, 2016), which can be run on GPUs, can be used to follow the long term evolution of GI systems. Of interest to the work in Chapter 4, following how a companion influences the migration rates of GI fragments may help us to understand the frequency with which wide-orbit, giant protoplanets evolve to become close-in, giant planets/brown dwarfs like those found in Fontanive et al. (2019). Whilst we demonstrated that the companion's influence can trigger fragmentation, suggesting that GI planets and brown dwarfs may be more common in binary systems, we still need to understand how they might have migrated to the short orbits on which the observed population of planets are currently found. It has already been shown that scattering of fragments onto eccentric orbits followed by tidal circularisation can account for around half of the systems observed in Fontanive et al. (2019), however other mechanisms which may accelerate migration must be explored if triggered fragmentation is to explain the full excess of close-in, giant planets which are found in these binary systems.

7.3 Conclusion

With the advent of the *James Webb Space Telescope* (JWST)-era of observational astronomy upon us, high resolution images of potentially self-gravitating discs will hopefully only become more frequent. Telescopes such as ALMA and the VLT have already provided us with some mind-blowing images, such as the DSHARP sample shown in Figure 1.4 and AB Aurigae shown in Figure 5.1. Equally, large scale surveys of planetary systems are steadily building a more coherent picture of population statistics. These observational insights inspire and inform theorists, providing them with a testbed to examine their ideas, as has often been the case

in the research throughout this thesis.

As things stand, the general consensus is that planet formation probably primarily occurs through core accretion. If planet formation through GI is possible, it is likely that it only occurs at large distances from the host star, forming only the most massive planets. There is now growing observational evidence that gravitationally unstable discs and GI-born planets do exist in nature. Wide-orbit, giant planets are now often being discovered through direct imaging surveys, some of which have characteristics indicative of formation through GI (e.g. Vigan et al., 2017; Nielsen et al., 2019; Vigan et al., 2021). Equally, direct imaging surveys of close-in, giant planets which find an excess of outer binary companions (Fontanive et al., 2019) indicate that triggered fragmentation may play an important role in their formation histories. With the recent observations of L1448 IRS3B (Tobin et al., 2016) and AB Aurigae (Boccaletti et al., 2020; Currie et al., 2022) we may now have the first direct evidence of discs undergoing fragmentation. There is also now mounting evidence that the spirals observed in systems such as Elias 2-27 can be successfully explained through GI (Hall et al., 2018; Paneque-Carreño et al., 2021).

We therefore find ourselves in an exciting time for exoplanet discovery, protoplanetary disc imaging and theories of planet formation. GI is emerging as a very promising complimentary theory to core accretion where the growing wealth of data combined with theoretical developments will only serve to improve our understanding of the earliest stages of planet formation. Whilst many unanswered questions of course still remain, I hope that this body of work acts to at least answer some of them, and that it may inspire future research in protoplanetary disc theory.

Bibliography

ALMA Partnership et al., 2015a, ApJ, 808, L3

ALMA Partnership et al., 2015b, ApJ, 808, L3

Andre P., Montmerle T., 1994, ApJ, 420, 837

Andre P., Ward-Thompson D., Barsony M., 1993, ApJ, 406, 122

Andrews S. M., 2020, ARA&A, 58, 483

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

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

Andrews S. M., Rosenfeld K. A., Wilner D. J., Bremer M., 2011, ApJ, 742, L5

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

Andrews S. M., Terrell M., Tripathi A., Ansdell M., Williams J. P., Wilner D. J., 2018a, ApJ, 865, 157

Andrews S. M., et al., 2018b, ApJ, 869, L41

Armitage P. J., 2010, Astrophysics of Planet Formation. Cambridge University Press, doi:10.1017/CBO9780511802225

Armitage P. J., 2020, Astrophysics of Planet Formation, 2 edn. Cambridge University Press, doi:10.1017/9781108344227

Aumann H. H., et al., 1984, ApJ, 278, L23

Avenhaus H., et al., 2018, ApJ, 863, 44

Bae J., Zhu Z., 2018a, ApJ, 859, 118

Bae J., Zhu Z., 2018b, ApJ, 859, 119

Baehr H., Klahr H., Kratter K. M., 2017, ApJ, 848, 40

Baruteau C., Meru F., Paardekooper S.-J., 2011, MNRAS, 416, 1971

Bell K. R., Lin D. N. C., 1994, ApJ, 427, 987

Benisty M., et al., 2015, A&A, 578, L6

Benz W., 1990, in Buchler J. R., ed., Numerical Modelling of Nonlinear Stellar Pulsations Problems and Prospects. p. 269

Benítez-Llambay P., Masset F. S., 2016, The Astrophysical Journal Supplement Series, 223, 11

Bergfors C., et al., 2013, MNRAS, 428, 182

Bertrang G. H. M., Avenhaus H., Casassus S., Montesinos M., Kirchschlager F., Perez S., Cieza L., Wolf S., 2018, MNRAS, 474, 5105

Birnstiel T., Dullemond C. P., Brauer F., 2010, A&A, 513, A79

Birnstiel T., Klahr H., Ercolano B., 2012, A&A, 539, A148

Boccaletti A., et al., 2020, A&A, 637, L5

Bodenheimer P., Pollack J. B., 1986, Icarus, 67, 391

Bodenheimer P., Yorke H. W., Rozyczka M., Tohline J. E., 1990, ApJ, 355, 651

Boffin H. M. J., Watkins S. J., Bhattal A. S., Francis N., Whitworth A. P., 1998, MNRAS, 300, 1189

Boley A. C., Hayfield T., Mayer L., Durisen R. H., 2010, Icarus, 207, 509

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

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

Boss A. P., 1998, ApJ, 503, 923

Boss A. P., 2006, ApJ, 641, 1148

Bowler B. P., et al., 2010, ApJ, 709, 396

Brandner W., 2006, Planet Formation: Theory, Observations, and Experiments. Cambridge Astrobiology, Cambridge University Press, doi:10.1017/CBO9780511536571

Brauer F., Dullemond C. P., Henning T., 2008, A&A, 480, 859

Cadman J., Rice K., Hall C., Haworth T. J., Biller B., 2020a, MNRAS, 492, 5041

Cadman J., Hall C., Rice K., Harries T. J., Klaassen P. D., 2020b, MNRAS, 498, 4256

Cadman J., Rice K., Hall C., 2021, MNRAS, 504, 2877

Cadman J., Hall C., Fontanive C., Rice K., 2022, MNRAS, 511, 457

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102

Chokshi A., Tielens A. G. G. M., Hollenbach D., 1993, ApJ, 407, 806

Chorin A., Marsden J., 1993, A Mathematical Introduction to Fluid Mechanics. Springer Publishing, doi:https://doi.org/10.1007/978-1-4612-0883-9

Clarke C. J., 2009, MNRAS, 396, 1066

Cleary P. W., Monaghan J. J., 1999, Journal of Computational Physics, 148, 227

Cleeves L. I., Öberg K. I., Wilner D. J., Huang J., Loomis R. A., Andrews S. M., Czekala I., 2016, ApJ, 832, 110

Corder S., Eisner J., Sargent A., 2005, ApJ, 622, L133

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

Currie T., et al., 2022, Nature Astronomy,

DeWarf L. E., Sepinsky J. F., Guinan E. F., Ribas I., Nadalin I., 2003, ApJ, 590, 357

Dipierro G., Lodato G., Testi L., de Gregorio Monsalvo I., 2014, MNRAS, 444, 1919

Dipierro G., Pinilla P., Lodato G., Testi L., 2015, MNRAS, 451, 974

Dipierro G., et al., 2018, MNRAS, 475, 5296

Dominik C., Tielens A. G. G. M., 1997, ApJ, 480, 647

Dong R., Zhu Z., Rafikov R. R., Stone J. M., 2015a, ApJ, 809, L5

Dong R., Hall C., Rice K., Chiang E., 2015b, ApJ, 812, L32

Dong R., Zhu Z., Fung J., Rafikov R., Chiang E., Wagner K., 2016, ApJ, 816, L12

Dong R., et al., 2018, ApJ, 860, 124

Dotter A., 2016, ApJS, 222, 8

Draine B. T., 2006, ApJ, 636, 1114

Draine B. T., Lee H. M., 1984, ApJ, 285, 89

Dullemond C. P., Dominik C., 2005, A&A, 434, 971

Dullemond C. P., Dominik C., 2008, A&A, 487, 205

Dunham M., et al., 2014, in Beuther H., Klessen R., Dullemond C., Henning T., eds, , Protostars and Planets VI. University of Arizona Press, pp 195–218

Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 607 (arXiv:astro-ph/0603179)

Dutrey A., et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 317 (arXiv:1402.3503), doi:10.2458/azu˙uapress˙9780816531240-ch014

Eisner J. A., et al., 2018, ApJ, 860, 77

Fischer D. A., Valenti J., 2005, ApJ, 622, 1102

Flaherty K. M., Hughes A. M., Rosenfeld K. A., Andrews S. M., Chiang E., Simon J. B., Kerzner S., Wilner D. J., 2015, ApJ, 813, 99

Flaherty K. M., et al., 2017, ApJ, 843, 150

Flaherty K. M., Hughes A. M., Teague R., Simon J. B., Andrews S. M., Wilner D. J., 2018, ApJ, 856, 117

Flaherty K., et al., 2020, ApJ, 895, 109

Fontanive C., Bardalez Gagliuffi D., 2021, Frontiers in Astronomy and Space Sciences, 8, 16

Fontanive C., Rice K., Bonavita M., Lopez E., Mužić K., Biller B., 2019, MNRAS, 485, 4967

Forgan D., Rice K., 2009, MNRAS, 400, 2022

Forgan D., Rice K., 2011, MNRAS, 417, 1928

Forgan D., Rice K., 2013a, MNRAS, 430, 2082

Forgan D., Rice K., 2013b, MNRAS, 432, 3168

Forgan D., Rice K., 2013c, MNRAS, 433, 1796

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

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

Forgan D. H., Hall C., Meru F., Rice W. K. M., 2018, MNRAS, 474, 5036

Gaia Collaboration et al., 2018, A&A, 616, A1

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

Garufi A., et al., 2013, A&A, 560, A105

Gillon M., et al., 2017, Nature, 542, 456

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Ginski C., et al., 2016, A&A, 595, A112

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

Grady C. A., et al., 2013, ApJ, 762, 48

Greenzweig Y., Lissauer J. J., 1992, Icarus, 100, 440

Güttler C., Blum J., Zsom A., Ormel C. W., Dullemond C. P., 2010, A&A, 513, A56

Haffert S. Y., Bohn A. J., de Boer J., Snellen I. A. G., Brinchmann J., Girard J. H., Keller C. U., Bacon R., 2019, Nature Astronomy, 3, 749

Haisch Karl E. J., Lada E. A., Lada C. J., 2001, ApJ, 553, L153

Hall C., Forgan D., Rice K., Harries T. J., Klaassen P. D., Biller B., 2016, MNRAS, 458, 306

Hall C., Forgan D., Rice K., 2017, MNRAS, 470, 2517

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

Hall C., Dong R., Rice K., Harries T. J., Najita J., Alexander R., Brittain S., 2019, ApJ, 871, 228

Harries T. J., Haworth T. J., Acreman D., Ali A., Douglas T., 2019, Astronomy and Computing, 27, 63

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

Hashimoto J., et al., 2011, ApJ, 729, L17

Haworth T. J., Facchini S., Clarke C. J., Mohanty S., 2018, MNRAS, 475, 5460

Haworth T. J., Cadman J., Meru F., Hall C., Albertini E., Forgan D., Rice K., Owen J. E., 2020, MNRAS, 494, 4130

Hendler N., Pascucci I., Pinilla P., Tazzari M., Carpenter J., Malhotra R., Testi L., 2020, ApJ, 895, 126

Hernquist L., Katz N., 1989, ApJS, 70, 419

Howell S. B., et al., 2014, PASP, 126, 398

Huang J., et al., 2018a, ApJ, 869, L42

Huang J., et al., 2018b, ApJ, 869, L43

Hubeny I., 1990, ApJ, 351, 632

Humphries J., Vazan A., Bonavita M., Helled R., Nayakshin S., 2019, MNRAS, 488, 4873

Humphries J., Hall C., Haworth T. J., Nayakshin S., 2021, MNRAS, 502, 953

Ida S., Lin D. N. C., 2004, ApJ, 604, 388

Ikoma M., Nakazawa K., Emori H., 2000, ApJ, 537, 1013

Janson M., Bonavita M., Klahr H., Lafrenière D., Jayawardhana R., Zinnecker H., 2011, ApJ, 736, 89

Jeans J. H., 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jenkins J. S., et al., 2017, MNRAS, 466, 443

Johansen A., Lambrechts M., 2017, Annual Review of Earth and Planetary Sciences, 45, 359

Johansen A., Oishi J. S., Mac Low M.-M., Klahr H., Henning T., Youdin A., 2007, Nature, 448, 1022

Johansen A., Klahr H., Henning T., 2011, A&A, 529, A62

Johansen A., Youdin A. N., Lithwick Y., 2012, A&A, 537, A125

Johnson J. A., et al., 2007, ApJ, 665, 785

Juhász A., Benisty M., Pohl A., Dullemond C. P., Dominik C., Paardekooper S. J., 2015, MNRAS, 451, 1147

Kaib N. A., Raymond S. N., Duncan M., 2013, Nature, 493, 381

Keppler M., et al., 2018, A&A, 617, A44

Kothe S., Blum J., Weidling R., Güttler C., 2013, Icarus, 225, 75

Kratter K., Lodato G., 2016, ARA&A, 54, 271

Kratter K. M., Matzner C. D., 2006, MNRAS, 373, 1563

Kratter K. M., Murray-Clay R. A., Youdin A. N., 2010, ApJ, 710, 1375

Kraus A. L., Ireland M. J., Hillenbrand L. A., Martinache F., 2012, ApJ, 745, 19

Kraus A. L., Ireland M. J., Huber D., Mann A. W., Dupuy T. J., 2016, AJ, 152, 8

Kuiper G. P., 1951, Proceedings of the National Academy of Science, 37, 1

Kurtovic N. T., et al., 2018, ApJ, 869, L44

Lada C. J., 1987, in Peimbert M., Jugaku J., eds, Vol. 115, Star Forming Regions. p. 1

Laibe G., Price D. J., 2012a, MNRAS, 420, 2345

Laibe G., Price D. J., 2012b, MNRAS, 420, 2365

Laibe G., Gonzalez J. F., Fouchet L., Maddison S. T., 2008, A&A, 487, 265

Lannier J., et al., 2016, A&A, 596, A83

Larson R. B., 1969, MNRAS, 145, 271

Larson R. B., 1981, MNRAS, 194, 809

Lesur G., et al., 2022, arXiv e-prints, p. arXiv:2203.09821

Levin Y., 2003, arXiv e-prints, pp astro–ph/0307084

Levin Y., 2007, MNRAS, 374, 515

Lin D. N. C., Papaloizou J., 1986, ApJ, 307, 395

Lin D. N. C., Pringle J. E., 1990, ApJ, 358, 515

Lin C. C., Shu F. H., 1964, ApJ, 140, 646

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

Lloyd J. P., 2011, ApJ, 739, L49

Lodato G., Rice W. K. M., 2004, in Bertin G., Farina D., Pozzoli R., eds, American Institute of Physics Conference Series Vol. 703, Plasmas in the Laboratory and in the Universe: New Insights and New Challenges. pp 266–271 (arXiv:astro-ph/0309570), doi:10.1063/1.1718465

Lodato G., Meru F., Clarke C. J., Rice W. K. M., 2007, MNRAS, 374, 590

Lucy L. B., 1977, AJ, 82, 1013

Lucy L. B., 1999, A&A, 344, 282

Lynden-Bell D., Pringle J. E., 1974, MNRAS, 168, 603

Machida M. N., Inutsuka S.-i., Matsumoto T., 2010, ApJ, 724, 1006

Malin D. F., 1977, AAS Photo Bulletin, 16, 10

Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348

Marois C., Zuckerman B., Konopacky Q. M., Macintosh B., Barman T., 2010, Nature, 468, 1080

Mathis J. S., Rumpl W., Nordsieck K. H., 1977, ApJ, 217, 425

Matzner C., Levin Y., 2005, ApJ, 628, 817

Mayer L., Wadsley J., Quinn T., Stadel J., 2005, MNRAS, 363, 641

Mayer L., Boss A., Nelson A. F., 2007a, arXiv e-prints, p. arXiv:0705.3182

Mayer L., Lufkin G., Quinn T., Wadsley J., 2007b, ApJ, 661, L77

McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI. p. 127

Meru F., 2015, MNRAS, 454, 2529

Meru F., Bate M. R., 2010, MNRAS, 406, 2279

Meru F., Bate M. R., 2011, MNRAS, 411, L1

Meru F., Juhász A., Ilee J. D., Clarke C. J., Rosotti G. P., Booth R. A., 2017, ApJ, 839, L24

Mizuno H., 1980, Progress of Theoretical Physics, 64, 544

Moe M., Kratter K. M., 2021, MNRAS, 507, 3593

Moe M., Kratter K. M., Badenes C., 2019, ApJ, 875, 61

Monaghan J. J., 1992, ARA&A, 30, 543

Morales J. C., et al., 2019, Science, 365, 1441

Mordasini C., 2018, in Deeg H. J., Belmonte J. A., eds, , Handbook of Exoplanets. p. 143, doi:10.1007/978-3-319-55333-7_143

Mordasini C., Klahr H., Alibert Y., Benz W., Dittkrist K.-M., 2010, arXiv e-prints, p. arXiv:1012.5281

Mordasini C., Alibert Y., Benz W., Klahr H., Henning T., 2012, A&A, 541, A97

Müller A., et al., 2018, A&A, 617, L2

Nakagawa Y., Nakazawa K., Hayashi C., 1981, Icarus, 45, 517

Nayakshin S., 2010a, MNRAS, 408, L36

Nayakshin S., 2010b, MNRAS, 408, 2381

Nayakshin S., 2011, MNRAS, 413, 1462

Nayakshin S., Fletcher M., 2015, MNRAS, 452, 1654

Nelson A. F., 2000, ApJ, 537, L65

Nero D., Bjorkman J. E., 2009, ApJ, 702, L163

Ngo H., et al., 2016, ApJ, 827, 8

Nielsen E. L., et al., 2013, ApJ, 776, 4

Nielsen E. L., et al., 2019, AJ, 158, 13

Owen J. E., Ercolano B., Clarke C. J., 2011, MNRAS, 412, 13

Paardekooper S.-J., 2012, MNRAS, 421, 3286

Paczynski B., 1978, Acta Astron., 28, 91

Paneque-Carreño T., et al., 2021, ApJ, 914, 88

Papaloizou J. C. B., Terquem C., 2005, Reports on Progress in Physics, 69, 119

Perez S., et al., 2015, ApJ, 798, 85

Pérez L. M., et al., 2016, Science, 353, 1519

Piétu V., Guilloteau S., Dutrey A., 2005, A&A, 443, 945

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62

Price D. J., Laibe G., 2015, MNRAS, 451, 813

Price D. J., et al., 2018, Publ. Astron. Soc. Australia, 35, e031

Pringle J. E., 1981, ARA&A, 19, 137

Rafikov R. R., 2005, ApJ, 621, L69

Rafikov R. R., 2017, ApJ, 837, 163

Ricci L., Testi L., Natta A., Neri R., Cabrit S., Herczeg G. J., 2010, A&A, 512, A15

Rice W. K. M., Armitage P. J., 2003, ApJ, 598, L55

Rice W. K. M., Armitage P. J., 2009, MNRAS, 396, 2228

Rice W. K. M., Armitage P. J., Bate M. R., Bonnell I. A., 2003, MNRAS, 339, 1025

Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2004, MNRAS, 355, 543

Rice W. K. M., Lodato G., Armitage P. J., 2005, MNRAS, 364, L56

Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2006, MNRAS, 372, L9

Rice W. K. M., Mayo J. H., Armitage P. J., 2010, MNRAS, 402, 1740

Rice W. K. M., Armitage P. J., Mamatsashvili G. R., Lodato G., Clarke C. J., 2011, MNRAS, 418, 1356

Rice K., Lopez E., Forgan D., Biller B., 2015, MNRAS, 454, 1940

Ricker G. R., et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Rybicki G. B., Lightman A. P., 1986, Radiative Processes in Astrophysics

Safronov V., 1969, in Evolution of the protoplanetary cloud and formation of the earth and the planets.

Salyk C., Herczeg G. J., Brown J. M., Blake G. A., Pontoppidan K. M., van Dishoeck E. F., 2013, ApJ, 769, 21

Santos N. C., Israelian G., Mayor M., 2004, A&A, 415, 1153

Schlaufman K. C., 2018, ApJ, 853, 37

Semenov D., Pavlyuchenkov Y., Schreyer K., Henning T., Dullemond C., Bacmann A., 2005, ApJ, 621, 853

Shakura N. I., Sunyaev R. A., 1973, A&A, 500, 33

Smith B. A., Terrile R. J., 1984, Science, 226, 1421

Spitzer Lyman J., 1942, ApJ, 95, 329

Stamatellos D., Whitworth A. P., 2009, MNRAS, 392, 413

Stamatellos D., Whitworth A. P., Bisbas T., Goodwin S., 2007, A&A, 475, 37

Tanaka H., Takeuchi T., Ward W. R., 2002, ApJ, 565, 1257

Tang Y. W., Guilloteau S., Piétu V., Dutrey A., Ohashi N., Ho P. T. P., 2012, A&A, 547, A84

Tang Y.-W., et al., 2017, ApJ, 840, 32

Teague R., et al., 2016, A&A, 592, A49

Testi L., Natta A., Shepherd D. S., Wilner D. J., 2003, A&A, 403, 323

Testi L., et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 339 (arXiv:1402.1354), doi:10.2458/azu˙uapress˙9780816531240-ch015

Tobin J. J., et al., 2016, Nature, 538, 483

Tokovinin A., Thomas S., Sterzik M., Udry S., 2006, A&A, 450, 681

Tomida K., Machida M. N., Hosokawa T., Sakurai Y., Lin C. H., 2017, ApJ, 835, L11

Toomre A., 1964, ApJ, 139, 1217

Veronesi B., Lodato G., Dipierro G., Ragusa E., Hall C., Price D. J., 2019, MNRAS, 489, 3758

Vigan A., et al., 2012, A&A, 544, A9

Vigan A., et al., 2017, A&A, 603, A3

Vigan A., et al., 2021, A&A, 651, A72

Vorobyov E. I., Basu S., 2010, ApJ, 714, L133

Wang J., Xie J.-W., Barclay T., Fischer D. A., 2014, ApJ, 783, 4

Watkins S. J., Bhattal A. S., Boffin H. M. J., Francis N., Whitworth A. P., 1998a, MNRAS, 300, 1205

Watkins S. J., Bhattal A. S., Boffin H. M. J., Francis N., Whitworth A. P., 1998b, MNRAS, 300, 1214

Weidenschilling S. J., 1977, MNRAS, 180, 57

Weidenschilling S. J., 1980, Icarus, 44, 172

Whitworth A. P., Stamatellos D., 2006, A&A, 458, 817

Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67

Wolszczan A., Frail D. A., 1992, Nature, 355, 145

Youdin A. N., Goodman J., 2005, ApJ, 620, 459

Youdin A., Johansen A., 2007, ApJ, 662, 613

Young M. D., Clarke C. J., 2015, MNRAS, 451, 3987

Zsom A., Ormel C. W., Güttler C., Blum J., Dullemond C. P., 2010, A&A, 513, A57

Zucker S., Mazeh T., 2002, ApJ, 568, L113

van Boekel R., et al., 2017, ApJ, 837, 132

van den Ancker M. E., The P. S., Tjin A Djie H. R. E., Catala C., de Winter D., Blondel P. F. C., Waters L. B. F. M., 1997, A&A, 324, L33