On the fragmentation of self-gravitating discs

Farzana Meru

Submitted by Miss Farzana Karim Meru to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics, September 2010.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signed: .....
Miss Farzana Karim Meru

Date: .....

Abstract

I have carried out three-dimensional numerical simulations of self-gravitating discs to determine under what circumstances they fragment to form bound clumps that may grow into giant planets. Through radiation hydrodynamical simulations using a Smoothed Particle Hydrodynamics code, I find that the disc opacity plays a vital role in determining whether a disc fragments. Specifically, opacities that are smaller than interstellar Rosseland mean values promote fragmentation (even at small radii, $R < 25$AU) since low opacities allow a disc to cool quickly. This may occur if a disc has a low metallicity or if grain growth has occurred. Given that the standard core accretion model is less likely to form planets in a low metallicity environment, I predict that gravitational instability is the dominant planet formation mechanism in a low metallicity environment. In addition, I find that the presence of stellar irradiation generally acts to inhibit fragmentation (since the discs can only cool to the temperature defined by stellar irradiation). However, fragmentation may occur if the irradiation is sufficiently weak that it allows the disc to attain a low Toomre stability parameter.

With specific reference to the HR 8799 planetary system, I find that it is only possible for fragments to form in the radial range where the HR 8799 planets are located ($R \approx 24 - 68$ AU) if the disc is massive. In such a high mass regime, mass transport occurs in the disc causing the surface mass density to alter. Therefore, fragmentation is not only affected by the disc temperature and cooling, but also by any restructuring due to the gravitational torques. The high mass discs also pose a problem for the formation of this system because the protoplanets accrete from the disc and end up with masses greater than those inferred from observation and thus, the growth of planets would need to be inhibited. In addition, I find that further subsequent fragmentation at small radii also takes place.

By way of analytical arguments in combination with hydrodynamical simulations using a parameterised cooling method, I explore the fragmentation criteria which in the past, has placed emphasis on the cooling timescale in units of the orbital timescale, β. I find that at a given radius the surface mass density (i.e. disc mass and profile) and star mass also play a crucial role in determining whether a disc fragments or not as well as where in the disc fragments form. I find that for shallow surface mass density profiles ($p < 2$, where $\Sigma \propto R^{-p}$), fragments form in the outer regions of the disc. However for steep surface mass density profiles ($p \gtrsim 2$), fragments form in the inner regions of a disc. In addition, I find that the critical value of the cooling timescale in units of the orbital timescale, $\beta_{\rm crit}$, found in previous simulations is only applicable to certain disc surface mass density profiles and for particular disc radii and is not a general rule for all discs. I obtain an empirical fragmentation criteria between the cooling timescale in units of the orbital timescale, β, the surface mass density, the star mass and the radius. Finally, I carry out crucial resolution testing by performing the highest resolution disc simulations to date. My results cast some serious doubts on previous conclusions concerning fragmentation of self-gravitating discs.

Contents

List of Figures

List of Tables

Declaration

This thesis contains work that has been published with Professor Matthew R. Bate.

The results of Chapter 3 have been published in *EXOPLANETS AND DISKS: THEIR FORMATION AND DIVERSITY: Proceedings of the International Conference.* AIP Conference Proceedings, Volume 1158, pp. 139-140 (2009).

These results were also published in *Monthly Notices of the Royal Astronomical Society*, 2010, Volume 406, Issue 4, pp. 2279-2288.

The results of Chapter 5 have been accepted by *Monthly Notices of the Royal Astronomical Society* in August 2010 and are currently in press.

The results of Chapter 4 have been submitted to *Monthly Notices of the Royal Astronomical Society*.

Acknowledgements

I would like to thank my supervisor, Professor Matthew R. Bate, for his help, guidance and excellent scientific insight that has helped me to progress through my PhD. His scientific abilities are truly inspiring. I would also like to thank him for allowing me to make the most of a number of opportunities that I have been presented with during my PhD including numerous conferences, schools and research programmes.

I would also like to thank Dr. Kacper Kornet, Professor Jim Pringle, Professor Cathie Clarke and Dr. Daniel Price for very useful scientific discussions, as well as the rest of the University of Exeter Astrophysics group for inspiring discussions.

I would also like to thank my mentors, Professor Tim Naylor and Professor Arif Babul, for the immense support and priceless advice that they have given me with regards to my PhD and my future career.

In addition, I would like to thank my fiancé, Asif Wallani, my parents, Karim and Khadija Meru, and my sister, Zahra Meru, for their support throughout my PhD.

The calculations reported in this thesis were performed using the University of Exeter's SGI Altix ICE 8200 supercomputer and Intel Nehalem (i7) cluster. The disc images were produced using SPLASH (Price 2007). In addition, some of the simulations were run using the Grape cluster at the University of California, Santa Cruz, for which I would like to thank the ISIMA (2010) programme.

Farzana Meru
University of Exeter, U.K.
22nd September 2010

Chapter 1

Introduction

1.1 Planet formation background

Planet formation has been an area which has been studied in great detail, long before the discovery of the first extra-solar planets in the early 1990s. In 1992, three planets were discovered around the pulsar PSR 1257+12 (Wolszczan & Frail 1992) followed by a discovery around another pulsar the following year, PSR B1620-26 (Backer et al. 1993). The first extra-solar planet was discovered in 1995 around a main sequence star, 51 Peg (Mayor & Queloz 1995). Since these discoveries, this area of research has soared and has been widely explored in both observations as well as theory. From the observational point of view, new instruments have been designed that continue to push the detection limits further out to detect a vast diversity of planets as well as the planet formation environments. New techniques are being developed to put constraints on the various parameters of the planets (such as the mass, radii and atmospheric properties) and the parent stars (such as spectral type, age, metallicity and luminosity). In addition, these techniques also constrain the dynamics of the planets and the disc environments in which they form. On the other hand, theoretical advances have escalated as theories are being developed and compared to current observations with the help of increased computing capabilities that are widely used to test these theories.

Planets have been known to orbit the Sun in a circular motion and approximately in the same plane. It is this fact that drove astronomers to explore the concept of discs and the formation of planets in these discs that surround the central star. This propelled the immense amount of research into disc theory and is how observational astronomers began to look at discs.

Giant planets are particularly interesting for two main reasons: firstly, the majority of extra-solar planets discovered are giant and therefore we know more about their properties and can thus use this to match theory and observations with fewer assumptions. Secondly, giant planets in our solar system dominate its mass and angular momentum and so understanding the formation and evolution of these planets may well provide much more information as to how the terrestrial planets evolved under the influence of the giant planets.

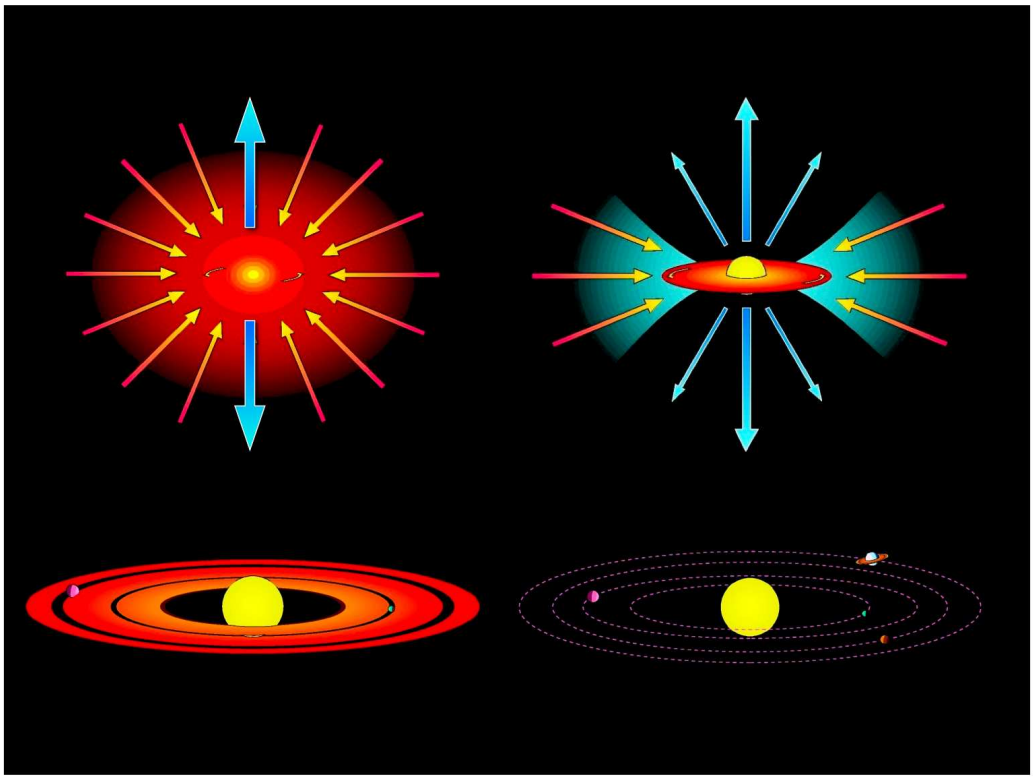

Figure 1.1: Schematic diagram showing the different stages of star formation. The early phase is the Class 0 phase (top left) where the molecular cloud is contracting and a stellar core forms. During the Class I phase (top right) a disc forms and is embedded in an envelope of gas. The penultimate stage involves an isolated star-disc system in which planets may form (Class II; bottom left) and in the Class III stage (bottom right) a disc is no longer present. Reproduced with permission from Mark McCaughrean.

1.2 Planet formation within the context of star formation

It is well known that planets form out of a disc of gas and dust surrounding the parent star. However, it is important to contextualise this and understand the evolutionary process that occurs well before the process of planet formation takes place as the earlier stages may influence the later planet formation stages. For example, the early evolution may affect the size and mass of the star and the disc out of which the planets form, which may ultimately affect the formation mechanism and properties of the planets that are formed.

The period of time over which giant planet formation around a star is believed to occur is only a small phase in the star's evolution. The early stages of star formation can be divided into 4 classes (Andre et al. 1993 for Class 0; Adams et al. 1987 for Class I - III, see Shu et al. 1987 for a review; see Figure 1.1). These stages were definitions that came from the mid-infrared spectra of observed Classical T Tauri systems at various stages in the evolutionary process. The Class 0 stage involves a molecular cloud of gas (mostly hydrogen and helium) and dust (metals) which is rotating and collapsing. This cloud is $\sim 10^4$AU in size and begins to collect at the centre to make up a stellar core which accretes gas onto it at a rate of $\sim 10^{-5} M_\odot$/yr and takes $\sim 10^4$ years. While

the gas is rotating, it forms a disc around the stellar core (though the simulations of Bate (1998) show that the outer disc forms before the stellar core) with accretion continuing to take place onto the star and disc at a rate of $\sim 10^{-6} M_{\odot}/yr$.

In the molecular cloud, the gas loses energy faster than momentum. The centripetal forces are more dominant around one axis. Therefore, as the gas rotates and the fluid elements collide, the motions around this dominant rotation axis are retained while the motions parallel to it are damped. The collisions of gas elements enable them to remove as much kinetic and internal energy as possible while keeping the same angular momentum. Given that a circular orbit is one where the elements can maintain the same amount of angular momentum with a minimum amount of energy (Pringle 1981), the particles move towards a circular orbit and form a disc around the central star.

Due to the gravitational attraction of the centrally accreting protostar, the envelope radius decreases to $\sim 10^3$AU during this stage. Since the accretion rate is lower, the timescale for this period increases to $\sim 10^5$ years. This is called the Class I stage and during this period, the disc mass is fairly high (O (10^{-1}) $M_{\odot}$) as the envelope is constantly accreting onto the disc.

The penultimate stage is the Class II stage where the envelope has completely disappeared, mostly because it has been accreted onto the central protostar and accretion disc but also because it may have been dispersed from the system by outflows. The central star is now called a Classical T Tauri star and continues to accrete gas onto it from the surrounding accretion disc. The disc radius can be as much as $\sim 10^2 - 10^3$ AU and this phase lasts $\sim 10^6$ years after which the accretion disc is completely depleted and a discless Class III weak-line T Tauri star remains.

The processes that take place to remove the disc that surrounds the star in the Class II phase to become the discless star in the Class III phase is a very active area of research (see Hollenbach et al. 2000 for a review on disc dispersal mechanisms). The disc matter may have accumulated together to form one or more planets. Hollenbach et al. (2000) argue that since the mass that goes into the planets is much smaller than the mass in the disc, planet formation only accounts for a small percentage of the disc dispersal. They argue that other more dominant mechanisms are involved and therefore observationally constraining the timescale for these disc dispersal mechanisms may indicate what other processes go on in the evolution of the disc in addition to planet formation. In addition, the timescale for these processes provide a limit as to how quickly planets form and therefore may constrain the planet formation and evolution mechanisms.

The matter that makes up the disc may have been accreted onto the central star: the viscosity in the disc causes it to spread such that most of the gas loses energy and angular momentum, while a small amount of gas moves out to large radii to conserve angular momentum in the system (Jeffreys 1924; see review by Pringle 1981). The timescale relevant for this is the viscous timescale which is small enough to match the observed data for $R < 10$ AU.

Alternatively, the gas and dust may have been driven out of the system by far ultraviolet- or X-ray-induced photoevaporation, which is most significant at $\sim 1 - 10$ AU and in the outer disc at $\gtrsim 30$ AU (Gorti et al. 2009). In addition, photophoresis, stellar or disc winds may drive matter out of the system (though Hollenbach et al. 2000 do not find winds to be a particularly effective mechanism). Finally, jets and tidal stripping of a disc due to close encounters may also play a part

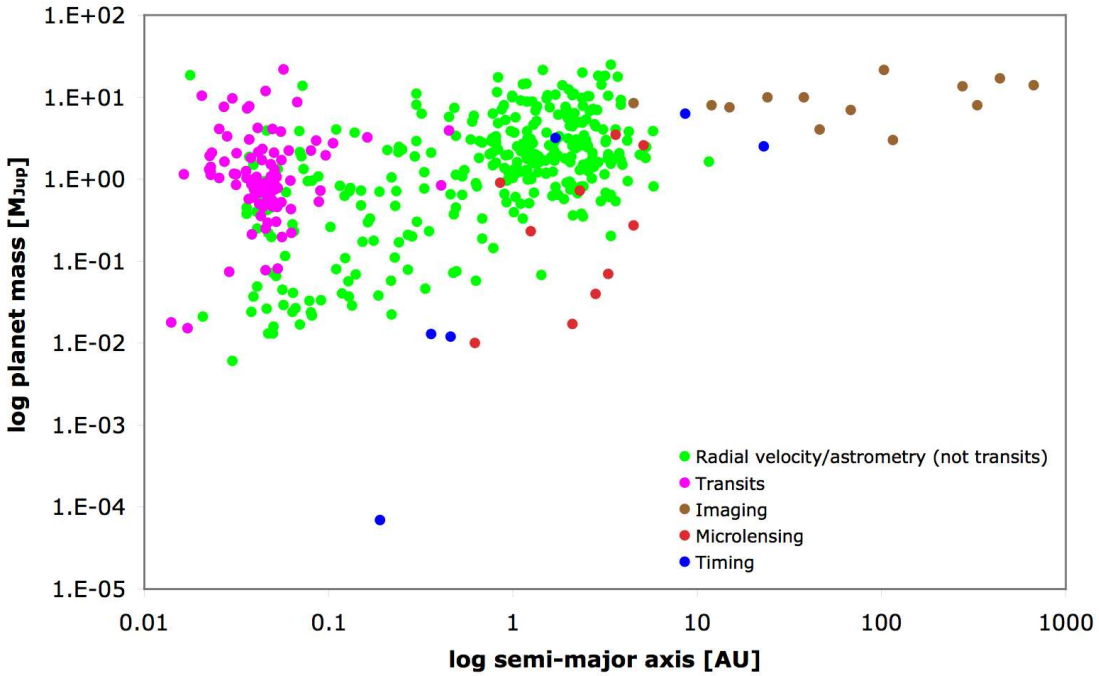

Figure 1.2: Plot showing the mass against semi-major axis of all the extra-solar planets discovered, separated out by their detection technique. The earlier planets discovered were more massive and close to the parent star while recently, detection limits have been pushed so that smaller and more distant planets are being detected. Data obtained from http://exoplanet.eu/

but the latter point may only be important for $R \gtrsim 100$ AU in dense clusters. Hollenbach et al. (2000) find that photoevaporation in conjunction with viscous disc spreading can cause an entire disc to be dispersed within $\sim 10^7$ years. Within the backdrop of other processes involved in the evolution of the disc from the Class II to Class III phases, planet formation and evolution is also expected to take place. The evolution process described above shows that since disc formation is part of the stellar evolution process, and since one disc dispersal mechanism is planet formation, a natural expectation is that planets are not rare, as I will illustrate in Section 1.3.1.

1.3 Observations

1.3.1 Extra solar planet observations

At the time of writing this thesis, 461 planets have been discovered. Figure 1.2 shows the mass against semi-major axis of all the planets that have been discovered. Due to the faint nature of the planets, it is very difficult to detect planets directly as the light from the parent star is so strong in comparison to the planet and so most planets have been discovered indirectly. A number of techniques have been developed to find extra-solar planets which in themselves differ so that they are able to probe different stars and reach different regions of the parameter space. These techniques are also complimentary since they confirm the observations using different techniques

and can also be combined to deduce the properties of the planets.

The first technique used to discover planets was the *pulsar timing* method. Pulsars emit radio waves and as their name suggests, they do this in a regular way. The presence of a planet causes the regularity of the radio waves that are observed to be disrupted since a planet causes the pulsar to move from its own orbit.

By far the main method by which extra-solar planets have been found has been through the *radial velocity* technique and is based on the idea that the motion of a planet has an effect on the motion of its parent star (as with the pulsar timing method). This technique is used when the plane consisting of the planet and the star has a component along the line of sight (with the signal being the strongest when the plane is parallel to the line of sight). The influence of the planet causes the velocity of the star along the line of sight to change which can be measured using Doppler spectroscopy. This measures the change in the spectral lines from the light emitted from the star due to the Doppler effect. This technique is frequently used to detect planets near to their parent star because the planets need to be observed for at least one orbit to confirm the planet detection. Also, the velocity perturbations are larger if the planet is closer to the parent star, making close-in planets easier to detect.

The second most successful way of discovering planets has been the *transiting planet* method. This technique only works when the planetary system is seen edge-on or as close to edge-on as possible and appeals to the brightness change as a result of the planet passing in front of or behind the central star. Since the planet is not as bright as the star, as it passes in front of it (the primary eclipse), it blocks out part of the light that the star emits. Consequently, the flux observed is reduced. Similarly, as the star passes around the back of the star (secondary eclipse), the total flux of the planet and the star is also reduced again and a smaller brightness dip is seen. This allows observers to put constraints on the star and planet properties using orbital dynamics. However, its disadvantage is that it is far more likely to detect short-period planets because this technique requires the system to be viewed as close to edge on as possible and as the orbit gets wider, the chances of being able to view such a scenario from Earth becomes increasingly smaller. Therefore, as with the radial velocity method, this technique results in a bias where more planets closer to the central star are discovered. The transit method is used to determine the radius of the planet, while the radial velocity method can determine the minimum planet mass, but not the absolute mass. However, the combination of the two techniques is powerful in that it allows both the planet mass and radius to be deduced. More recently, the transit timing technique has been developed further (termed transit timing variation) whereby additional planets may be detected due to a variation in the main transiting lightcurve of the first planet. This may also give clues as to the formation history. At the time of writing this thesis, no detections of additional planets have been made using this method.

Microlensing has been a way in which planets have also been detected. This involves two stars being almost exactly aligned, with the star closest to the observer (the lensing star) harbouring a planet. The measurement is unfortunately not reproducible but has been powerful in the past at detecting Earth-mass planets. It is a particularly useful technique to observe planets in between the Earth and the Galactic centre since there are a large number of background stars available.

Astrometry is an old technique based on the principle that the presence of a planet affects the motion of the central star (as with the radial velocity and pulsar timing techniques). The motion of a star due to the presence of the planet is measured directly which then allows the presence of a planet to be inferred. In contrast to other methods that are biased towards low period planets, this technique is more powerful for long period planets. However, long period planets also require long observation times. While discoveries of planets have been claimed using this technique, they have not been confirmed and so it has yet to prove its success.

The above methods use indirect techniques to infer the presence of a planet as well as its properties and those of the parent star. The final method, *direct imaging*, has found it difficult to detect planets in the past as the brightness of the star overwhelms the brightness of any planets that it may harbour. However, if the star's luminosity is sufficiently low and if the planet is sufficiently far away such that its brightness may be discerned from that of the parent star, the planet may be found. With recent improvements in telescope power in conjunction with the adaptive optics technique, planets have begun to be imaged directly. The first of these planets was Fomalhaut b (Kalas et al. 2008) which shows the presence of a planet at $\approx$ 119 AU. The second system, also discovered around the same time, was a multiple planet system around the A star, HR8799 (Marois et al. 2008). I discuss this system in more detail in Section 1.5. More recently, a low mass companion ($\approx 8\mathrm{M_J}$) has been detected around the young solar analogue 1RXS J160929.1-210524 at $\approx$ 330 AU from the parent star (Lafrenière et al. 2008, 2010) while the young, $\approx$ 10Myr old, star β Pictoris has also recently been found to host a giant planet with mass $\approx 9M_{\mathrm{J}}$ close to the parent star at $\approx 8-15$ AU (Lagrange et al. 2009, 2010). The former case is the lowest mass companion detected to date at such a large distance, while the latter is the closest planet to the central star to have been detected by the direct imaging technique. This therefore emphasises how the detection limits are continually being revised.

Since the discovery of the first extra-solar planets, planet observations have increased with time. Each method has developed significantly and each method has been able to detect multiple planet systems (49 multiple planet systems in total at the time of writing this thesis), the first of which was υ Andromedae found using the radial velocity technique (Butler et al. 1999). A vast number of stars have planets surrounding them: given the number of extra-solar planet detections is continually rising, it is almost expected that a star will harbour a planet. The question is no longer *does it harbour a planet?* but is instead *what are the properties of its planet and how did it form?*

1.3.2 Disc observations

Disc observation began when astronomers questioned how the planets in our solar system formed in a single plane around a central star. The observations of discs, which measure the excess emission above that due to the stellar photosphere (as a result of the dust in the discs which absorb and thermally re-emit), is used to provide us with numerous properties about the structure of discs and hence the environment in which planets may form. Such properties may include the disc mass, size, structure and metallicity. Figure 1.3 shows the regions in which different wavebands probe different areas of a disc: the optical waveband can only give information about

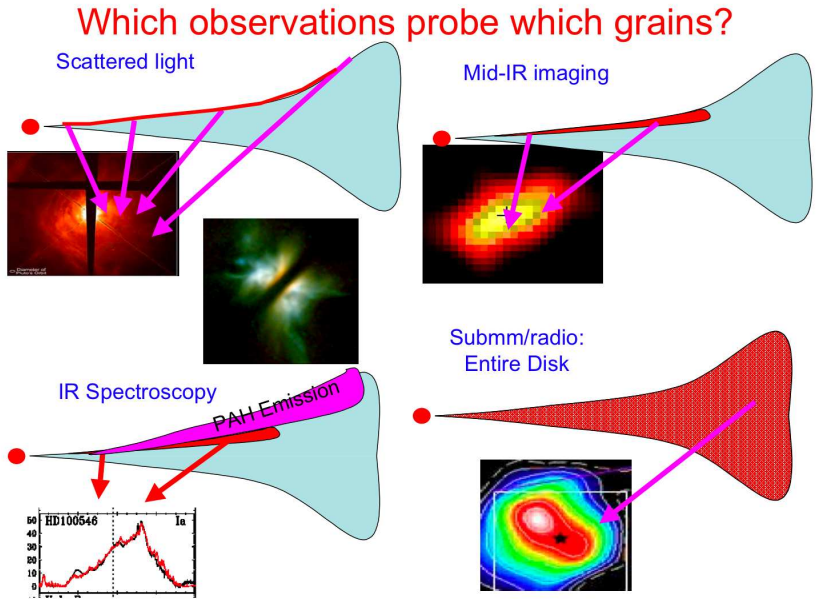

Figure 1.3: Schematic diagram showing the regions of a disc that different wavebands probe when observing discs. Scattered optical light only allows the surface to be seen. The various infrared wavebands probe slightly further. The longer wavelengths allow the inner regions of a disc to be observed. Image used with permission from Carsten Dominik, Kees Dullemond and Michiel Min.

the surface of the disc since we see the light that is scattered from the disc surface. The mid-infrared waveband probes the region a little below the surface, but still restricts us to the area close to the surface. Infrared spectroscopy can help us to probe further into the disc, vertically, while the far-infrared probes further out in the disc radially. The sub-millimetre waveband, on the other hand, can probe most of the disc since it is sensitive to cold dust and is currently the best tracer of mass. However, the next generation of instruments, for example the Atacama Large Millimeter Array (ALMA), will be much more powerful. This instrument will use the millimetre waveband to probe even further into the disc interior with much higher resolution and sensitivity than previous millimetre telescopes. This will provide more details on the structure of the discs, such as the presence of spiral structures and the structure of a disc which harbours a planet during its formation phase. Combining the different wavebands together provides a powerful technique for inferring the various properties of the disc.

Specifically for planet formation, determining the disc mass is critical since it is important that there is enough matter in the discs to make a planet (as well as potentially lose some of the disc mass to other disc dispersal processes as described in Section 1.2). The disc radius is also important since this indicates to us how large the planet formation environment can be. The metallicity is also important since this determines the composition of the planets that may form. We know from observations of our own solar system that giant planets can consist of materials other than hydrogen and helium (see review by Guillot & Gautier 2009) and as will be discussed in Chapter 3, metallicity may determine the method by which planets may form.

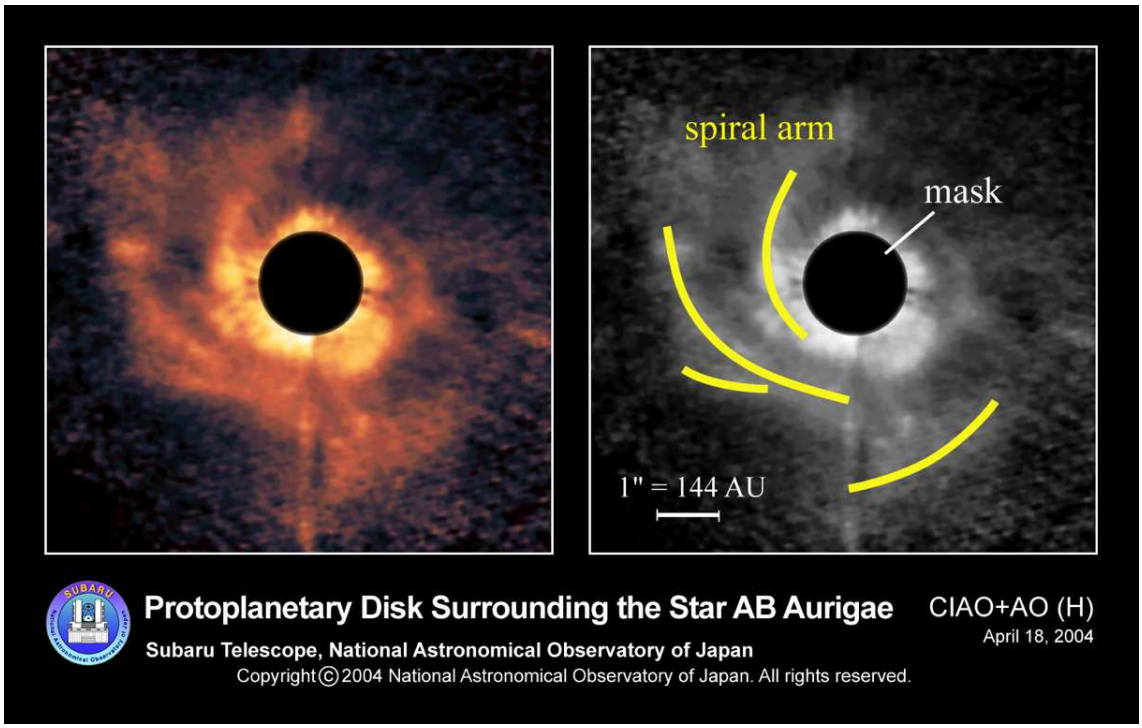

Figure 1.4: Near infrared image of the spiral structure in the disc surrounding the Herbig Ae star, AB Aurigae, observed by Fukagawa et al. (2004). The spiral structures extend to > 500 AU. Credit: National Astronomical Observatory of Japan (NAOJ)

In Section 1.4.2, I discuss the *gravitational instability* method. One characteristic of the discs that form planets by this method is that they have strong spiral structures. Figure 1.4 shows a double spiral structure observed in a disc extending to > 500 AU around the star AB Aurigae, a 2.4$M_{\odot}$, 4 Myr old pre-main sequence Herbig Ae star (Fukagawa et al. 2004). The image was obtained in the near infrared waveband, using the Subaru Telescope and was the first near infrared image of a spiral structure in a disc. Prior to this, similar spiral structure was observed (at $R >$ 100 AU) in the optical waveband in a disc surrounding the Herbig Be star, HD 100546 (Grady et al. 2001). However, such structures are typically difficult to see. With the next generation of instruments (e.g. ALMA) these features will be more easily identifiable.

For planet formation, a critical inference from disc observation is the disc dispersal timescale which indicates the maximum timescale on which planet formation can occur. The timescale within which the disc is present around the star is a key driver in determining *how* planet formation occurs and it is this parameter that has driven much of the research into planet formation theory in the recent past. The disc is present around the central star for a very short period of the overall lifetime of the star ($\approx$ 0.1% for a solar mass star) which provides only a narrow window of opportunity during which a planet can form. The timescale on which a disc lasts has been determined by observations: early work by Strom et al. (1993) showed that solar-type pre-main sequence stars contained discs that survive for up to 10 Myr and suggested that the disc survival times were much shorter (< 1 Myr) for more massive stars. More recently, Briceño et al. (2001) carried out a large scale survey of Orion OB1, consisting of pre-main sequence stars with masses

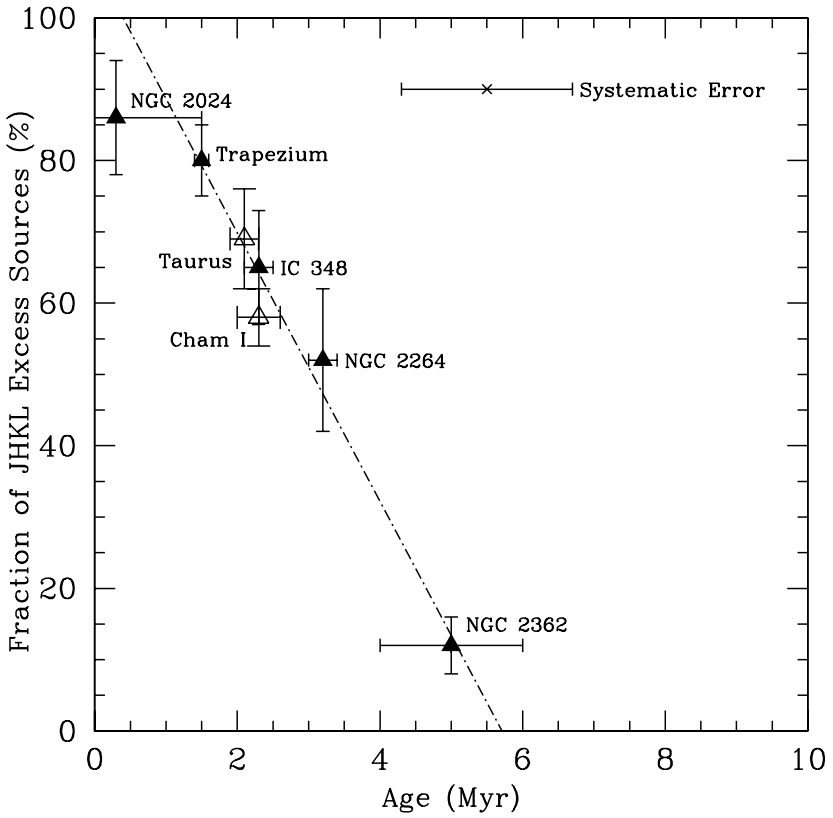

Figure 1.5: Fraction of sources with infrared excess (indicative of the presence of a disc) determined by Haisch et al. (2001b). The lack of excess around older objects suggests that discs are depleted within ≈ 6 Myr.

$M \approx 0.6 - 0.9 M_{\odot}$, and showed that the amount of near-infrared and Hα emission (characteristic of the presence of a disc and disc accretion, respectively) considerably decreased for stars that have ages between ≈ 3 – 10 Myr compared to those between ≈ 1 – 3 Myr. Haisch et al. (2001b) combine a number of infrared observational data of circumstellar discs around stars with ages ranging from 0.3 – 30 Myr and with star masses ranging from $0.13 - 1.3 M_{\odot}$. They show (Figure 1.5) that at young ages, the disc frequency is high ($\gtrsim$ 80%). However, by ≈ 3 Myr, only 50% of stars harbour a disc and by ≈ 6 Myr, all the discs have been dissipated. Haisch, Lada, & Lada (2001a) show that discs around more massive stars (earlier spectral types than G) dissipate more quickly (in $\lesssim 2 - 3$ Myr) than those around less massive stars (spectral types G and later) which is also consistent with the view that mass accretion rates onto the central star is faster for discs with high mass stars. Mamajek et al. (2004) carry out mid-infrared (N-band) observations of ≈ 30 Myr old stars in the Tucana-Horologium association and find that no excess emission is seen and suggest that the inner disc dissipates on timescales that coincide with the timescale on which it takes for accretion to finish. Comparing their results to other N-band surveys, they show a similar trend of decreasing infrared excess as Haisch et al. (2001b).

Carpenter et al. (2006) carry out observations of stars with masses varying between ≈ 0.1 and 20 $M_{\odot}$ in the 5 Myr old Upper Scorpius OB association and find that a considerable fraction

(19%) of lower mass (K and M) stars show infrared excesses similar to Class II sources in Taurus, while more massive (B, A, F and G) stars show little or no infrared excess. Hernandez et al (2007) carry out a study of the stars in σ Orionis, a 3 Myr old cluster with a range of stellar masses ($\lesssim 0.1$ to $\approx 7 M_{\odot}$) and confirm that disc evolution is more rapid for higher mass stars ($> 1 M_{\odot}$) since disc fractions are much lower around A-type and intermediate mass T Tauri stars. The results suggest that the discs around more massive stars will disperse faster than those around K and M stars, thus decreasing the *window of opportunity* for planet formation. In Chapter 3, I discuss the effects of metallicity on planet formation by gravitational instability. Yasui et al. (2009) carry out near-infrared observations of two young star-forming clusters (≈ 0.5 Myr) in the extreme outer galaxy which is known to be metal-poor (approximately 10% of solar metallicity) and suggest that disc dispersal in a low metallicity environment is very rapid (< 1 Myr). This clearly has implications for giant planet formation because if planets are found in a low metallicity environment, they will need to have formed very rapidly, which can then put constraints on the formation mechanism.

It is important to note that in deriving the lifetime of discs around stars, the dust components of these discs have been observed. To determine the total mass of a disc one has to assume a value for the dust to gas ratio. The canonical value assumed is ≈ 0.01. This is a value which is an estimate since the real value is unknown. The solar metallicity content is $Z = 0.02$, where Z is the mass fraction of metals. Given that the metallicity is made up of dust as well as volatile materials, then roughly half of this will make up the dust component i.e. 1%. If we assume that the composition of the Sun is representative of the material that the disc that once surrounded it was also composed of, we can infer that the dust to gas ratio in the disc is ≈ 0.01. Such an assumption is reasonable given that much of the material that makes up the Sun would have been accreted onto it from the accretion disc and also because the Sun and disc were born from the same molecular cloud. Critically, if the gas component in the discs remain longer than the dust components, this may impact the formation process by which planets form. To study the gas in discs, high resolution spectroscopy is required to measure the accretion rate onto the central star (primarily via a measurement of the Hα emission line), which is an indicator of the presence of gas in the disc. Mohanty et al. (2005) find that the disc accretion rate, $\dot{M}$, increases with star mass (where $\dot{M} \propto M_{\star}^2$). This is consistent with the view that the lifetime of discs around more massive stars is smaller than those around low mass stars. Lada et al. (2006) carry out near- and mid-infrared photometry of the IC 348 cluster and find a correlation between the equivalent width of the Hα emission line and the slope of the infrared excess spectral energy distribution and deduce that the gas and dust in discs evolve on similar timescales (with the dust more likely to persist for longer than the gas). Other authors (e.g. Pascucci et al. 2006; Sicilia-Aguilar et al. 2006) have also carried out studies that suggest that gas and dust evolution occurs at approximately the same rate. On the other hand, since terrestrial planets are thought to form on much longer timescales (≈ 50 Myr), the gas and dust evolution may occur on different timescales, though this may be observationally difficult to detect as the grain sizes in the discs become larger.

In addition to the disc dispersal mechanisms discussed above, a further evolution between the Class II and Class III phases involves the formation of debris discs, which are gasless discs made up of the remnants of planetesimals. These discs indicate that a planet forming process

(whether successfully or unsuccessfully) has occurred. While primordial discs consist of grains which grow over time, debris discs consist of grains that decrease in size over time since the collisions of planetesimals that have not been able to grow cause them to break up.

1.4 Planet formation theory

There are two ways in which giant planets have been hypothesised to form in situ: core accretion (Safronov 1969; Goldreich & Ward 1973; Pollack et al. 1996) and gravitational instability (Cameron 1978; Boss 1997; review by Durisen et al. 2007). The site of planet formation is common for both methods i.e. the gas and dust disc that forms around the central star.

1.4.1 Core accretion

The *core accretion* method was first introduced by Safronov (1969) and later studied in more detail by Goldreich & Ward (1973), Perri & Cameron (1974), Harris (1978) and Mizuno (1980) (see Pollack (1984) and Lissauer & Stevenson (2007) for reviews). It involves the sub-micron sized dust particles in the disc colliding, coalescing and growing larger by gravitationally attracting the dust material in its surrounding region to form a solid core of $\approx$ 100 km in size. This method is how terrestrial planets are hypothesised to form. As the core becomes larger, the gravitational force becomes more important than the gas drag (Pollack et al. 1996). The escape velocity of the gas surrounding the solid core (typically $M_{\rm core} \approx 10-15M_{\oplus}$) becomes so large that the gas cannot escape and instead, a phase of runaway gas accretion onto the core takes place, ultimately forming a gas giant planet. This happens when the gas and solid accretion rates onto the core become comparable. This method has typically been favoured for two reasons: firstly, observations of giant planets in our own solar system suggest that a solid core is likely (though it is currently uncertain whether Jupiter does in fact have such a core or not; Guillot 1999; Saumon & Guillot 2004). Secondly, the majority of extra solar planets that had initially been discovered were found very close to the central star. Gas giant planets are typically thought to form beyond the snow line so that there is enough mass available to make up a large core and also because their formation must take place beyond the dust sublimation radius. While the hot Jupiters that have been observed inside the snow line would not have formed there in situ by core accretion, this is still a favoured mechanism since they may have formed reasonably close and possibly then migrated inwards. However, the earlier planets were discovered by radial velocity and transiting methods which are typically biased towards close-in planets.

On the other hand, core accretion has historically presented two key problems. The first problem is a temporal issue: Pollack et al. (1996) showed that the timescale on which the gas giant planets in our solar system could have formed by core accretion may be too large (10^6 – 10^8 Myr) such that the gas in the disc is depleted before the gas giant planet is fully formed. Also, since the formation timescale increases with radius, planets at larger radii may struggle to form in time. This is known as the *timescale problem*. Secondly, simulations of this method model the growth of planets typically starting with kilometre-sized planetesimals, but while the growth of particles from small grains to metre-sized objects appears to be straight-forward, how to get from

metre-sized objects to kilometre-sized planetesimals is unknown. This is the so-called *metre-sized barrier*.

1.4.2 Gravitational instability

Gravitational instability (Kuiper 1951; Cameron 1978; Boss 1997; review by Durisen et al. 2007) typically occurs in early-stage discs that are massive enough such that the disc's self-gravity plays a part in driving its evolution. In contrast to the core accretion method, the key component that is important in forming the core is the gas, and hence gaseous cores form. This may well be the case for Jupiter (Saumon & Guillot 2004), though the capture of solid material to form a core after the fragmenting stage in a gravitationally unstable disc has also been proposed (Helled, Podolak, & Kovetz 2006)

Since gravitational instability is not thought to operate close to the central star, it has not been thought to be the dominant mechanism by which giant planets form as it was unable to describe the observations of close-in giant planets. Recent observations of planets at large distances (Kalas et al. 2008; Marois et al. 2008) encourage us to revisit this theory. Nero & Bjorkman (2009) have argued analytically that Fomalhaut b and at least the outer planet of the HR 8799 system could have formed by gravitational instability as the cooling timescales are likely to be small enough such that fragmentation is possible. Moreover, discs in their early stages are thought to be massive (Eisner & Carpenter 2006) suggesting that gravitational instability must play a role in the evolution of a disc in the late Class I and early Class II stages.

In this method, gravitationally unstable discs are subject to destabilising gravitational forces and stabilising pressure and rotational forces. For a differentially rotating disc, if there are perturbations in the density, velocity and gravitational potential which can be written in the form of an exponential function such that either the amplitude or the phase is slowly varying (known as the *WKB approximation* or the *tight-winding limit*, discussed below), then substituting into the mass continuity, momentum and Poisson equations yields a dispersion relation for a gaseous disc of the following form (see Binney & Tremaine (2008) for a detailed derivation):

$$(m\Omega - \omega)^2 = \kappa_{\rm ep}^2 - 2\pi G\Sigma|k| + k^2 c_{\rm s}^2 \tag{1.1}$$

where m is the azimuthal wavenumber (or the disturbance mode), Ω is the angular velocity, $\kappa_{\rm ep}$ is the epicyclic frequency, $c_{\rm s}$ is the sound speed in the disc, Σ is the surface mass density, G is the gravitational constant and k is the radial wavenumber, related to the wavelength by $k = 2\pi/\lambda$. For axisymmetric disturbances, $m = 0$ and the dispersion relation is given by:

$$\omega^2 = \kappa_{\rm ep}^2 - 2\pi G\Sigma|k| + k^2 c_{\rm s}^2. \tag{1.2}$$

If $\omega^2 > 0$, the disc is stable and if $\omega^2 < 0$, a perturbation exists whose amplitude grows exponentially such that the disc is unstable. We may now consider the effects of pressure and rotation separately. If the disc is not rotating such that $\kappa_{\rm ep} = 0$ but pressure effects are involved so that $c_{\rm s} > 0$, then the disc is stable if

$$|k| > \frac{2\pi G\Sigma}{c_s^2}, \tag{1.3}$$

i.e. if the wavelength is small. Therefore, pressure can only stabilise a disc on short wavelengths. Now consider the scenario where the disc is not pressure supported ($c_s = 0$) but is rotationally supported ($\kappa_{ep} > 0$). The disc is then stable if

$$|k| < \frac{\kappa_{ep}^2}{2\pi G\Sigma}, \tag{1.4}$$

i.e. if the wavelength is large. Consequently, rotation can only stabilise the disc on large wavelengths. This shows that neither pressure nor rotation can stabilise the disc against gravitational collapse by themselves. However, it may be possible for a combination of both to maintain stability in a disc. Since equation 1.2 is a quadratic in k, for stability at all wavelengths, we require that the minimum value of the right hand side of equation 1.2 is positive. By a trivial differentiation of equation 1.2 with respect to the wavenumber, $d\omega^2/dk$, we can see that

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

Substituting this into equation 1.2 and requiring that $\omega_{min}^2 > 0$ for stability, we can show that

$$Q > 1 \tag{1.6}$$

is required for stability, where

$$Q = \frac{c_s \kappa_{ep}}{\pi\Sigma G}, \tag{1.7}$$

and Q is known as the Toomre stability parameter (Toomre 1964). Toomre (1964) showed that for an infinitesimally thin disc to fragment, the stability parameter must be less than a critical value, $Q_{crit} \approx 1$. Since Σ, c_s and κ_{ep} are all functions of the disc radius, Q is also a function of radius. Therefore, $Q < 1$ is an indication of *local* axisymmetric instability near the radius R. Consequently, if a perturbation passes through a region where $Q < 1$, it will grow while it is in that region.

The WKB, or equivalently, tight-winding approximation makes the assumption that the radial wavelength is much smaller than the radius, i.e. $|kR| >> 1$ or $\lambda/R << 1$. In the disc scenario, the phase of the perturbation wave varies with time. The dispersion relation (equation 1.2) assumes that the WKB approximation is valid. The largest wavelength possible in a disc is $\lambda = \zeta H$, where H is the disc scaleheight and ζ is a constant with value $\zeta = O(1)$. We therefore, require that the disc aspect ratio $H/R << 1$, i.e. that the disc is thin, for this stability criterion to be valid. Protoplanetary discs are indeed thin with typical values of $H/R \approx 0.05$ so the Toomre stability criterion is valid for these discs.

In reality, since discs do have a finite thickness, the vertical component of gravity is essentially diluted causing the disc to stabilise. Goldreich & Lynden-Bell (1965) show analytically that a uniformly rotating isothermal disc considering the third dimension is $\approx$ 30% more stable

than the equivalent razor-thin disc. It is important to note that this analytical work was carried out considering a uniformly rotating disc rather than a differentially rotating disc but nevertheless, illustrates the effect of disc thickness. However, the thin disc approximation is still valid in these analyses. In addition, the equation of state also affects the critical value of the Toomre stability parameter, with lower values of the ratio of specific heats, γ, (i.e. a softer equation of state) requiring a higher value of the critical Toomre stability parameter, and hence making it easier to form fragments: Johnson & Gammie (2003) carry out local numerical simulations of isothermal discs and show that fragmentation can occur for $Q \lesssim 1.4$.

The above stability criterion can be explained physically (Toomre 1964; Binney & Tremaine 2008) by considering the stability of a circular region (located anywhere in a disc) of radius ΔR with mass, $M_{\rm p} = \pi\Sigma_o(\Delta R)^2$. The pressure acting on the region is given by p_o and the spin angular momentum, $S \approx \Omega(\Delta R)^2$. If the area of the patch is reduced to a fraction $(1 - f)$ of the original area, where $f << 1$, the pressure, rotational and gravitational forces on the patch will change. The pressure and gravitational potential of the region will change by $p_1 \approx f p_o \approx f c_{\rm s}^2 \Sigma_o$ and $fGM_{\rm p}/(\Delta R)^2$, respectively. Since the magnitude of the pressure and gravitational forces per unit mass are given by $F_p = |-\nabla p/\Sigma|$ and $F_g = |-\nabla\phi| = GM_{\rm p}/R^2$, then the additional pressure force acting on the region (in an outwards direction) is given by:

$$F_{\rm p,1} \approx \frac{f c_{\rm s}^2}{\Delta R}, \tag{1.8}$$

while the additional gravitational force acting on the region (in an inwards direction) is

$$F_{\rm g,1} \approx \frac{fGM_{\rm p}}{(\Delta R)^2} \approx fG\pi\Sigma_o. \tag{1.9}$$

If only these two forces existed (i.e. neglecting the rotational force), then the region is stable if $|F_{p,1}| > |F_{g,1}|$ i.e.

$$\Delta R \lesssim \frac{c_{\rm s}^2}{\pi G \Sigma_o} \equiv R_p. \tag{1.10}$$

This shows that stability can be achieved on small scales via pressure forces. Since the region will also consist of motions due to the rotation of the disc, the rotational forces cannot be neglected. The magnitude of the centrifugal force per unit mass (in the outwards direction), $|F_c| \approx \Omega^2 \Delta R \approx S^2/(\Delta R)^3$, changes since the spin angular momentum per unit mass of the region is conserved around its own centre. Therefore the additional centrifugal force is given by:

$$F_{\rm c,1} \approx \frac{fS^2}{(\Delta R)^3}. \tag{1.11}$$

The region is therefore also stable if $|F_{c,1}| > |F_{g,1}|$ i.e.

$$\Delta R \gtrsim \frac{\pi G \Sigma_o}{\Omega^2} \equiv R_c, \tag{1.12}$$

which shows that stability can be achieved on large scales via centrifugal forces. Figure 1.6 shows a schematic diagram of the regions over which the pressure and rotational forces stabilise the disc

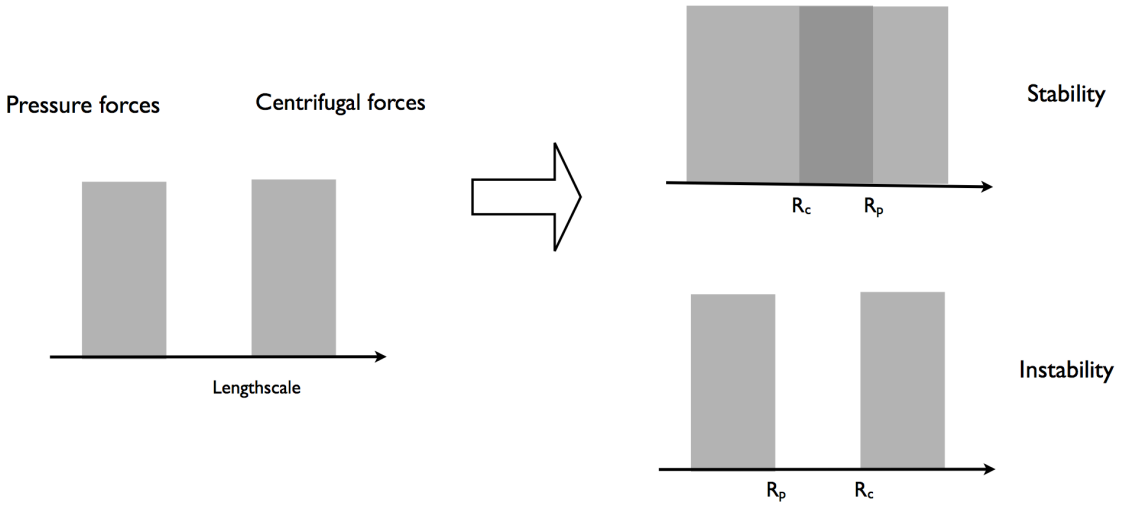

Figure 1.6: Schematic diagram showing the lengthscales over which pressure and rotational forces act to stabilise a disc. Pressure forces stabilise a disc on small scales while rotational forces stabilise a disc on large scales. If the largest scale over which pressure forces stabilise a disc, $R_{\rm p}$, is larger than the smallest scale over which rotational forces stabilise the disc, $R_{\rm c}$, then the disc is stable against collapse. On the other hand, if $R_{\rm c} > R_{\rm p}$, the disc may fragment into clumps.

against gravitational collapse. If $R_p < R_c$ then a region of wavelength space exists where the disc is not stabilised and gravitational collapse occurs. However, if on the other hand $R_p \gtrsim R_c$, then the disc is stabilised on all wavelengths. This is provided if

$$\frac{c_{\rm s}\Omega}{\pi\Sigma_o G} > 1, \tag{1.13}$$

which once again is the Toomre stability criterion derived above using the dispersion relation, though this time with more physical arguments. Note that the epicyclic frequency in equation 1.7 is replaced by the angular frequency here. Pringle (1977) showed that in an accretion disc around a star where the aspect ratio of the disc is small, the circular velocity is almost Keplerian right down to a narrow boundary layer on the surface of a star. Since in Keplerian discs, the epicyclic frequency is approximately equal to the angular frequency, this replacement is adequate.

Linear stability theory requires that the Toomre condition is satisfied for fragmentation to occur. Gravitational forces are key to whether the disc perturbation structure appears in the first place. However, it does not indicate what the intensity of the structures are, nor how long they will last. For non-linear evolution of gravitational instability, the amount of thermal energy is important and therefore the cooling in a disc must also be considered. Gammie (2001) showed that in addition to the stability criterion above, the disc must cool at a fast enough rate. Using shearing sheet simulations, he showed that if the cooling timescale can be parameterised as

$$\beta = t_{\rm cool}\Omega, \tag{1.14}$$

where

$$t_{\rm cool} = u\left(\frac{{\rm d}u_{\rm cool}}{{\rm dt}}\right)^{-1}, \tag{1.15}$$

u is the internal energy and ${\rm d}u_{\rm cool}/{\rm dt}$ is the total cooling rate, then for fragmentation we require $\beta \lesssim 3$, for a ratio of specific heats $\gamma = 2$ (in two dimensions). Rice, Lodato, & Armitage (2005) carried out three-dimensional simulations using a Smoothed Particle Hydrodynamics (SPH) code and showed that this cooling parameter is dependent on the equation of state. They showed that fragmentation can occur if $\beta < \beta_{\rm crit}$ where $\beta_{\rm crit} \sim 6 - 7$ for discs with $\gamma = 5/3$ and $\beta_{\rm crit} \sim 12 - 13$ for discs with $\gamma = 7/5$.

Viscosity may drive the evolution of a disc by causing the mass to spread and accrete onto the central star on the viscous timescale. Observations of Classical T Tauri stars (combining measurements of the accretion luminosity from the UV excess with the mass, radius and temperature of the star) suggest that the accretion rates from the discs onto the central stars are $\approx 10^{-9} - 10^{-7}{\rm M}_{\odot}$ /yr (Hartigan et al. 1995; Gullbring et al. 1998). Studies of emission lines from infalling gas onto the star suggest that the accretion rates may be as low as a few $\times 10^{-12}{\rm M}_{\odot}$ /yr, though these are the accretion rates for smaller stars (Muzerolle et al. 2003; Mohanty et al. 2005). The question that requires answering is *what drives this evolution?* Kinematic viscosity due to collisions between molecules of gas would cause the matter from the disc to accrete onto the central star very slowly and is too low to describe the observed accretion rates. Therefore, the source of the viscosity is not known. Given the unknown nature of the viscosity, a parametrisation, α, is introduced so that the stress tensor of the unknown viscosity is written in terms of the vertically integrated pressure (Pringle & Rees 1972; Shakura & Syunyaev 1973)

$$T_{R\phi} = \left|\frac{{\rm dln}\Omega}{{\rm dln}R}\right|\alpha\Sigma c_{\rm s}^2, \tag{1.16}$$

where $T_{R\phi}$ is the only non-vanishing component of the stress tensor in the accretion disc. Since this component of the stress tensor can also be written in the form

$$T_{R\phi} = \nu\Sigma R\left|\frac{{\rm d}\Omega}{{\rm d}R}\right|, \tag{1.17}$$

the kinematic viscosity, ν, can also be written in the form

$$\nu = \alpha c_{\rm s} H, \tag{1.18}$$

where $H = c_{\rm s}/\Omega$ is the isothermal scaleheight in the disc. The viscosity may come from turbulent motions which are likely to be subsonic. Since the largest size of the turbulent eddies are given by the disc scaleheight, the constant of proportionality, α, that describes the magnitude of the viscosity will be smaller than unity.

Assuming the mass accretion rates onto the central star quoted above are indicative of the mass accretion rate through a disc whose temperature is $\approx$ 10 K, an order of magnitude estimate can relate the mass accretion rate, $\dot{M}$, to the viscosity parameter via the Toomre equation (assuming $Q \approx 1$ for a gravitationally unstable disc) and

$$\dot{M} = 3\pi\nu\Sigma, \tag{1.19}$$

which suggests that the α parameter in real discs is $\approx 10^{-4} - 10^{-2}$. Alternatively, the α parameter in real discs can be approximated using the viscous timescale,

$$t_\nu = \frac{R^2}{\nu}, \tag{1.20}$$

and comparing this to the disc dispersal timescales. For a disc with aspect ratio, $H/R \approx 0.05$, and assuming the temperature of the disc is $\approx$ 10 K, considering disc lifetimes of $\approx 1 - 10$ Myr, the value of the α parameter at 50 AU is $\approx 3 \times 10^{-2} - 3 \times 10^{-3}$. A number of possibilities for the source of the viscosity have been suggested such as turbulent viscosity and magneto-rotational-instability. Gravitational instability may provide one such source of turbulent viscosity since this causes torques and hence gravitational stresses in a disc. Both of the above approximations assume that the disc disperses viscously, but the former method also assumes that the source of viscosity is gravitational torques. The only relevant non-zero component of the gravitational stress in a disc is given by (Lynden-Bell & Kalnajs 1972; Lodato & Rice 2004)

$$T^g_{R\phi} = \left|\frac{\mathrm{dln}\Omega}{\mathrm{dln}R}\right| \nu_{\mathrm{GI}}\Omega\Sigma = \Sigma\frac{\mathbf{g_r}\mathbf{g}_\phi}{4\pi G\rho}. \tag{1.21}$$

where $\mathbf{g_r}$ and $\mathbf{g}_\phi$ are the radial and azimuthal components of the gravitational acceleration due to the disc self-gravity, $\nu_{\mathrm{GI}} = \alpha_{\mathrm{GI}} c_s H$ is the viscosity associated with the turbulence caused by gravitational instability and α_{GI} is the gravitational stress. The resulting dissipation rate per unit area as a result of this is given by

$$Q^+ = T^g_{R\phi}|R\Omega'|. \tag{1.22}$$

In a disc where the cooling rate per unit area is given by

$$Q^- = \frac{U}{t_{\mathrm{cool}}} = \frac{\Sigma c_s^2}{\gamma(\gamma-1)t_{\mathrm{cool}}}, \tag{1.23}$$

Gammie (2001) and Rice et al. (2005) showed that in a steady state disc where the dominant form of heating is that due to gravitational instabilities such that $Q^+ = Q^-$, the gravitational stress in a disc can be linked to the cooling timescale by

$$\alpha_{\mathrm{GI}} = \frac{4}{9}\frac{1}{\gamma(\gamma-1)}\frac{1}{\beta}. \tag{1.24}$$

Therefore, the rapid cooling required for fragmentation can also be interpreted as a maximum gravitational stress that a disc can support without fragmenting, which Gammie (2001) and Rice et al. (2005) show to be $\alpha_{\mathrm{GI,max}} \approx 0.06$.

The concept of a fast cooling needed for fragmentation is very clear from previous work. However, the value of the critical cooling timescale, β_{crit} (and therefore, by equation 1.24, $\alpha_{\mathrm{GI,max}}$), does not appear to be too clear cut: Rice et al. (2003a) found that for a $0.1\mathrm{M}_\odot$ disc with surface mass density profile, $\Sigma \propto R^{-7/4}$, extending to a radius, $R_{\mathrm{out}} = 25$ AU around a $1\mathrm{M}_\odot$ star, the disc

fragments using $\beta = 3$ but not for $\beta = 5$, whereas for a disc with mass $M_{\rm disc} = 0.25{\rm M}_{\odot}$, the disc fragments for $\beta = 5$. On the other hand, Rice et al. (2005) suggest that the fragmentation boundary is independent of the disc to star mass ratio. Clarke, Harper-Clark, & Lodato (2007) showed that the critical value of β (below which fragmentation will occur if the stability criterion is met) may depend on the disc's thermal history: if the timescale on which the disc's cooling timescale is decreased is slower than the cooling timescale itself (i.e. a gradual decrease in β) then the critical value may decrease by up to a factor of 2. More recently, Cossins, Lodato, & Clarke (2010) showed that the critical value varies with the temperature dependence of the cooling law. In addition, they carry out a simulation of a self-gravitating disc with surface mass density profile, $\Sigma \propto R^{-3/2}$ (c.f. Rice et al. (2005) who used $\Sigma \propto R^{-1}$), with ratio of specific heats, $\gamma = 5/3$, and show that the critical value $\beta_{\rm crit} \approx 4$. Using equation 1.24, this is equivalent to $\alpha_{\rm GI,max} \approx 0.1$ which brings the result of $\alpha_{\rm GI,max} \approx 0.06$ described above into question. Yet a number of papers have been produced which base their work on the concept of a single critical value of β (or equivalently, a maximum gravitational stress value) (e.g. Clarke 2009; Rafikov 2009; Cossins et al. 2010; Kratter et al. 2010).

The above fragmentation criteria are based on the assumption that the dominant form of dissipation in the disc is due to internal heating processes. Previous simulations without external irradiation have considered isolated discs with simple cooling prescriptions (e.g. Lodato & Rice 2004) and with radiative transfer (e.g. Boss 2001; Cai et al. 2006; Mayer et al. 2007). Johnson & Gammie (2003) suggested that discs with external irradiation are likely to be effectively isothermal and can therefore be treated as such. Matzner & Levin (2005) analytically considered externally irradiated discs and concluded that stellar irradiation quenches fragmentation. Cai et al. (2008) carried out simulations with external irradiation and found that their discs are more resistant to fragmentation and proposed that these results may be extended to discs with stellar irradiation. Stamatellos & Whitworth (2008) also carried out simulations taking into account the effects of stellar irradiation and found this to be a stabilising factor. Rafikov (2009) analytically explored fragmentation in gravitationally unstable discs including the effects of stellar irradiation and suggested that fragmentation can only occur beyond $\approx$ 120AU. Dodson-Robinson et al. (2009) carried out a linear stability analysis on irradiated discs to show that gravitational instability takes place for systems with a large disc to star mass ratio. However, while fragmentation in gravitationally unstable discs may be less likely than previously thought, we still do not know in *what situations* discs may fragment when modelling them realistically with radiative transfer and by considering the effects of stellar irradiation. It is therefore important to deduce when fragmentation may occur when simulating discs with more detailed energetic conditions, and just how realistic or unrealistic fragmentation is in real discs.

Boss (2002) carried out simulations of gravitationally unstable discs and varied the opacity from 0.1× to 10× the Rosseland mean opacities and found that the fragmentation results were insensitive to the dust grain opacity. However, given that a reduced opacity is more likely to allow energy to stream out of a disc more easily causing it to cool and promote fragmentation, while in a high opacity disc the converse is true, it is interesting to consider what opacity values allow and do not allow fragmentation. Given that a disc's opacity gives somewhat an indication of how metal-

rich it is or how large or small the grain sizes are, we may then make preliminary conclusions on the disc conditions that are likely to promote fragmentation, which is one focus of this thesis.

1.4.3 Other secondary effects

For completeness, I also briefly describe the secondary mechanisms by which planets may end up in the location at which they are observed. These include scattering and migration. Scattering involves a highly dynamical process through which planets interact by close encounters causing smaller planets to be ejected from a system or even destroyed through collision. Rasio & Ford (1996) carried out numerical simulations of two planet systems and found that in $\approx 50\%$ of cases, direct collisions occurred. In the remaining simulations that did not involve collisions, one planet was typically ejected, leaving the other planet in an eccentric orbit at a smaller radius. If the two planets were of similar mass, the outer planet was typically ejected while if the planets were of unequal mass, the lower mass planet was ejected.

Independently, Weidenschilling & Marzari (1996) carried out simulations of three Jupiter-mass planets on circular orbits in a single system and found the most common result was that one planet was ejected on a hyperbolic trajectory. The ejected planet was not necessarily the outermost planet. In addition, some of their simulations showed that collisions and ejections of two planets did also occur. Following the scattering process, the remaining planets were on eccentric orbits and moved to both smaller and larger radii compared to their original radii.

Following a highly dynamical phase where multiple planets are present, ejections and collisions will still occur and the remaining more massive planets will stabilise by orbiting in mean motion resonances. Scattering may be a process by which higher order mean motion resonances are achieved (Raymond et al. 2008), or a process which may form dynamically stable planetary systems at large radii ($\approx 10^2 - 10^5$ AU; Veras et al. 2009).

Migration, on the other hand, involves the interaction between a planet (or multiple planets) and the disc and has primarily been used to explain the large number of hot Jupiters that have been observed. The earlier migration study was in fact carried out with reference to the interactions between the satellite, Mimas, and the Cassini Division in the rings around Saturn: Goldreich & Tremaine (1978) studied the properties of the gap and found that the inner edge of the gap coincided with the inner Lindblad radius.

The planet in the disc exerts a torque on the gas and the gas also exerts a torque onto the planet. This gives rise to Lindblad torques interior and exterior to the planet's orbit which have different magnitudes, thus causing a net torque. In addition, the motion of the gas in the horseshoe region of the planet's orbit also exerts a corotation torque onto the planet. Goldreich & Tremaine (1980) investigated the angular momentum and energy transfer between a disc and a satellite and considered the effects on the eccentricity due to the Lindblad and corotation torques. With specific application to Jupiter, they suggested that the angular momentum transfer would have been very rapid causing significant changes to Jupiter's orbit and to the disc structure in $\approx$ O(10^3) yrs. Migration regimes have been studied extensively, the two main areas being the Type I and Type II regimes, originally determined through an analytical model by Ward (1997).

As well as the interaction between the planet and the gaseous disc, the interaction between

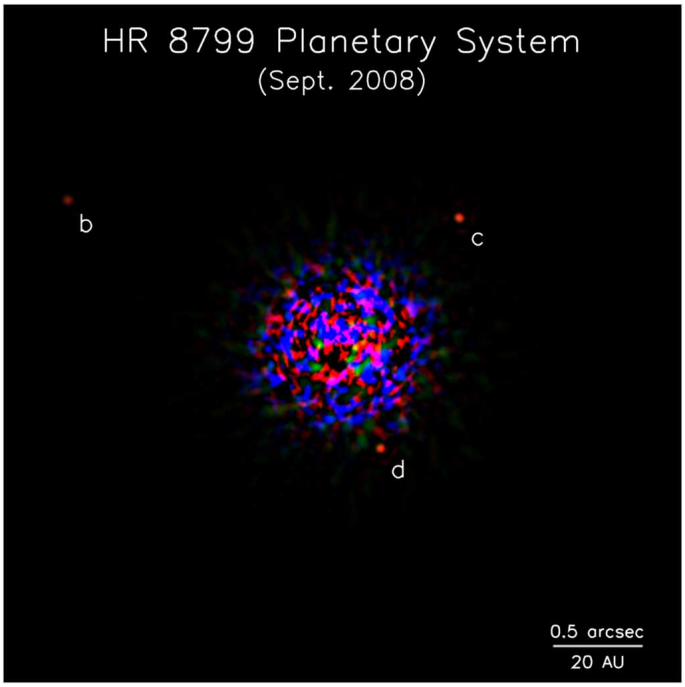

Figure 1.7: The HR 8799 planetary system observed in the near-infrared waveband (Marois et al. 2008). This planetary system was the first multiple planet system detected using the direct imaging technique.

the planet and the planetesimal disc has also been studied. Murray et al. (1998) showed that if the solid surface mass density is larger than a critical value (so that the mass interior to the planet is approximately the same mass as the planet itself), long-range inward migration can take place. In addition, Trilling et al. (1998) consider the mass loss from a migrating planet as it approaches the central star (due to the planet filling its Roche lobe). They suggest that planets may lose only a little or no mass as they approach the central star and therefore, migration may explain the large number of detections of hot Jupiters. They also suggest that a large number of planets may end up close to the central star. The details of the interactions between the planet and the disc are highly dependent on the mass of the planet (see review by Lubow & Ida 2010). Typically, migration acts to move the planet inwards to smaller radii (Ward 1997), though outward migration of a pair of planets in mean motion resonance has also been suggested (e.g. Crida et al. 2009).

1.5 Formation of the HR 8799 planetary system

The HR 8799 planetary system (Marois et al. 2008) is a recently discovered system and was the first multiple-planet system detected by the direct imaging technique (Figure 1.7). It consists of planets at projected separations of $\approx$ 24, 38 and 68 AU around a $\approx 1.5 M_{\odot}$ (Gray & Kaye 1999) A star with current luminosity of $L \approx 4.9 L_{\odot}$ (Gray & Kaye 1999) and is located $\approx$ 39 parsecs away (van Leeuwen 2007). Estimates of the star's age range between 30 and 160 Myr (Marois et al. 2008).

Currently, its formation mechanism is unknown. Dodson-Robinson et al. (2009) show that

even in the most favourable scenarios (where the mass of the nebula is high), the maximum radius at which core accretion may form a planet in this system is 35AU (assuming a disc lifetime of 5 Myr), and therefore this method may only explain the formation of the innermost planet. In addition, it is important to note that the core accretion scenario is less likely in systems with low metallicities (Kornet et al. 2005). The HR 8799 star is known to be a metal-poor, λ Bootis star with metallicity $[M/H] = -0.47$ (Gray & Kaye 1999) so it is reasonable to assume that its disc was similarly metal-poor. Therefore, core accretion is even less likely to have formed the HR 8799 planetary system.

Dodson-Robinson et al. (2009) also rule out the possibility of the planets being formed in the inner regions by core accretion and then moving outwards by planetesimal-driven migration. This is because for this to have a significant effect, the planet must interact with planetesimals whose combined mass is of the order of the planet mass and since their estimates of the combined mass was much smaller than the HR 8799 planet masses (even in the most favourable of scenarios), this possibility was ruled out. In addition, they also rule out the possibility of Type III migration based on previous work that suggests that outward Type III migration reverses and becomes inwards migration, as well as the requirement that the planet is near the inner edge of the disc. Crida et al. (2009) suggest that outward migration of a pair of planets in mean motion resonance may be possible, but currently, outward migration of three planets in a gas disc has not been considered.

Dodson-Robinson et al. (2009) also consider the possibility of the planets being formed close in by core accretion and then being scattered outwards. While they do not rule this method out, they suggest that obtaining a stable system would not be easy (also suggested by Veras et al. 2009) and that if this was the mechanism by which HR 8799 formed, then the system has been observed at a particular phase in its evolution where it appears to be dynamically stable. Further observations of other multiple-planet systems at large radii would be required to advocate or rule out this mechanism.

The other in-situ mechanism that may describe the formation of the HR 8799 planets is gravitational instability. Analytical estimates suggest that gravitational instability is unable to describe the formation of planets inside $R \sim 70 - 120$ AU (for a $1M_{\odot}$ star and using interstellar Rosseland mean opacities (Rafikov 2009; Clarke 2009)). However, for lower opacities, the radius outside of which fragments form moves inwards to smaller radii. This may occur if grain growth has occurred or if the metallicity is reduced, which as identified, is certainly likely to be the case for this system.

Nero & Bjorkman (2009) have argued analytically that at least the outer planet of the HR 8799 system could have formed by gravitational instability and this may occur for disc masses of between ≈ 0.1 and $0.6M_{\odot}$, but that HR 8799 c and d (the inner two planets) are too far in to form by gravitational instability. The HR 8799 system is therefore an interesting case because at first, no one formation mechanism appears to fully describe its formation.

In addition, the subsequent evolution of the planets may also give some clues as to the formation mechanism of this system. Fabrycky & Murray-Clay (2010) suggest that the timescale for dynamical stability of this system (assuming it is circular and viewed face-on) is two orders

of magnitude smaller than the estimated age of the star though they do suggest that the system may be stable if the inner two planets are locked in a mean motion resonance of 2:1 and either the outermost planet is small ($\lesssim 10 M_J$) and not in resonance with the middle planet, or if a double resonance exists such that the outer two planets are also in a 2:1 resonance. Furthermore, the evolution and growth of the planets has been discussed by Kratter et al. (2010) who suggest that the formation of the HR 8799 system may not have occurred by gravitational instability as the growth of the fragments would cause the planets to be more massive than the observations suggest. Such studies on the subsequent evolution of the planets following the formation stage is important to understand what formation mechanisms may be possible. In Chapter 4, I revisit this system by considering both the formation and evolution of the fragments to deduce whether the HR 8799 system could have formed by gravitational instability.

1.6 Core accretion, gravitational instability, or both?

The current literature suggests no single method can describe the formation of all the extra-solar planets discovered. In addition, the scenarios in which each mechanism is more likely to be dominant, are somewhat mutually exclusive. Firstly, core accretion typically acts closer to the central star, while gravitational instability is thought to form planets at larger radii. Secondly, core accretion is less likely in a low metallicity environment whereas the current literature does not suggest any preference for where gravitational instability might act (though the opacity arguments presented in Section 1.4.2 suggest that gravitational instability may be more likely to take place in low metallicity environments). Thirdly, since less massive discs typically deplete slower than those that are more massive, and since it is expected that less massive stars would only be able to support less massive discs, it may be expected that planets around late-type stars are formed by core accretion since the discs last longer (and are also less likely to be gravitationally unstable). Conversely, early-type stars are more likely to harbour more massive discs which may open up the door for gravitational instability to act in such an environment. On the other hand, the planetesimal accretion rate onto a core is proportional to the solid surface mass density (Safronov 1969). Therefore, while more massive discs have shorter lifetimes core accretion may still be a possible formation mechanism in such discs.

Boley (2009) proposed that core accretion may be a method by which planets may form at small radii ($\sim$ O(10)AU) while gravitational instability may be the mechanism by which planets may form at larger radii ($\gtrsim$ O(100)AU) though a hybrid scenario of forming gas giants in the same system by both core accretion and gravitational instability has yet to be modelled.

I have suggested here that the two in situ planet formation mechanisms may not be competing mechanisms, as has been proposed in the past, but in fact, may be complementary (and may even be further complemented by the prospect of other secondary effects such as migration or scattering).

1.7 Thesis focus

The focus of this thesis is on the historically less dominant planet formation mechanism, gravitational instability, with a focus on the fragmentation conditions in protoplanetary discs. The recent observations of extra-solar planets at large radii from the central star make this method more appealing and in light of recent observations of planets at large radii, a re-investigation of the gravitational instability method is warranted. In this thesis, I do this largely by way of numerical simulations, though some analytical work is also carried out.

Chapter 3 explores the effects of opacity on the fragmentation of a gravitational unstable disc. I carry out three-dimensional radiation hydrodynamical simulations of self-gravitating discs to determine whether discs that are modelled realistically using radiative transfer can be subject to fragmentation. I investigate the effects that the opacity has on the cooling and hence fragmentation conditions. In Chapter 4 I consider what disc conditions may have been required for a disc to fragment and form the HR 8799 system.

Concurrently, in Chapter 5, I also explore the fragmentation conditions in discs simulated without radiative transfer, but using a simpler parameterised cooling technique, as has been done so in the past (e.g. Rice et al. 2005). I consider in more detail, by way of analytical work as well as a wide parameter space of simulations, what disc and star conditions as well as cooling conditions are required for fragmentation. While past work has considered the fragmentation boundary, a wide parameter study has not been carried out. To determine the fragmentation conditions in more detail it is first important to understand the conditions in simplified disc models and use these as a basis for building in more complex physics. Consequently, I choose to carry out the simulations without including radiative transfer. Given that Chapter 4 and 5 were done in parallel, I use some findings from the results of Chapter 5 in Chapter 4.

Finally, previous work on the fragmentation criteria have not carried out significant resolution tests so it is unclear whether the results have indeed converged. I therefore carry out a resolution test in Chapter 6 to test the reliability of past results.

In Chapter 7 I make conclusions on the work presented in this thesis and discuss how the thesis could be extended further.

Chapter 2

Numerical Method

In this chapter, I describe the numerical and physical details concerning the methods employed to carry out the simulations presented in this thesis.

2.1 Numerical details

For hydrodynamical simulations, there are two fundamental types of method that are usually employed to carry out the simulations: grid-based hydrodynamical codes and particle based hydrodynamical codes. Grid-based codes generally use an Eulerian description which essentially considers the properties (and change in properties) of the fluid at a fixed spatial position on a grid. In each grid cell, the fluid equations are solved at the grid boundaries. Particle-based codes, on the other hand, use a Lagrangian method to solve the fluid equations at points, and generally do not have a specific structure. This method involves using dynamic fluid components where the fluid properties are determined in the comoving frame of a fluid element rather than considering the properties at a fixed spatial position. Both methods offer advantages and disadvantages. In the work carried out for this thesis, I have used a Smoothed Particle Hydrodynamics (SPH) method primarily because it offers an advantage over grid-based codes for problems with high density contrasts. It is able to provide accurate results for problems involving strong shocks and complex geometries (Steinmetz & Mueller 1993). Given the nature of this research involves gravitationally unstable discs with strong spiral density contrasts as well as shocked regions and more importantly, fragmenting regions, SPH is well suited to this problem as computational time is significantly reduced in low density regions allowing more computational time to be focused on high density regions. Furthermore, SPH deals with free boundaries much more easily than grid-based codes. In all the simulations carried out for this thesis, the outer radial boundaries of the discs are free boundaries, as are the vertical boundaries for the simulations presented in Chapters 5 and 6.

2.1.1 Smoothed Particle Hydrodynamics

SPH was originally developed by Gingold & Monaghan (1977) and Lucy (1977) to simulate astrophysical hydrodynamical scenarios. The simulations carried out in this thesis use a version that was originally developed by Benz (1990) and further developed by Bate, Bonnell, & Price (1995),

Whitehouse, Bate, & Monaghan (2005), Whitehouse & Bate (2006) and Price & Bate (2007). SPH is closely related to N-body codes in that both are particle codes. While both consider long-range gravity forces, SPH also considers pressure and viscosity. Particle based codes compute variables at a particular spatial point. However, each particle in an SPH code represents a volume of fluid with variables that are representative of the surrounding region (as opposed to the values at a particular point), such as density and pressure. SPH calculates the properties of the volume of fluid by considering the *smoothing length* of each particle, which I discuss in more detail below. P. Bodenheimer, G. P. Laughlin, M. Rózyczka, & H. W. Yorke (2007) state that a simple SPH method requires the following:

1. Initial conditions of the particles (co-ordinates, velocities and mass) such that the particle density can be calculated.
2. Calculation of the gravitational force
3. A physical law relating the pressure and density
4. A time advancing scheme to advance the co-ordinates and velocities of the particles
5. A scheme for determining the ratio of the change in pressure to the density, $\nabla P/\rho$
6. A method for determining the smoothing length.

In this chapter, I will cover these aspects in more detail.

The concept of *smoothing* essentially calculates an average of a particular quantity. If the variable A is being calculated for a particle, its approximate value at position r can be expressed as

$$A(r) = \int_V A(r')\delta(r - r')\mathrm{d}^3r'. \tag{2.1}$$

where r' is a location within a smoothing length, h, around position r. If the smoothing kernel, $W(r - r', h)$, which describes the mass distribution around a particle, is defined as

$$\lim_{h\to 0} W(r - r', h) = \delta(r - r'), \tag{2.2}$$

with the kernel normalised so that

$$\int_0^\infty W(r - r', h)\mathrm{d}^3r' = 1, \tag{2.3}$$

then

$$< A(r) > = \int_V A(r')W(r - r', h)\mathrm{d}^3r'. \tag{2.4}$$

If the kernel is even, equation 2.4 is approximately

$$< A(r) > \approx A(r) + O(h^2) \tag{2.5}$$

and therefore it is accurate to second order in space.

The SPH equations also require the use of W' and perhaps W'' in addition to W and so the kernel must be of a form where these are continuous. In the SPH code used for this thesis, the kernel is based on a cubic spline (Monaghan & Lattanzio 1985) and has the following form:

$$W(q,h) = \frac{\sigma}{h^d} \begin{cases} 1 - \frac{3}{2}q^2 + \frac{3}{4}q^3 & \text{for } 0 \leq q < 1, \\ \frac{1}{4}(2-q)^3 & \text{for } 1 \leq q < 2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

where d is the number of dimensions, σ is the normalisation constant equal to 2/3, $10/(7\pi)$ and $1/\pi$ in one, two and three dimensions respectively, and $q = r/h$. All the simulations presented here are carried out in three dimensions. The smoothing length is essentially the width of the smoothing kernel.

Expression 2.4 shows the approximate value of the quantity A expressed in the form of an integral. However, for computational purposes, this needs to be in discrete form. This is carried out via a density-weighted form:

$$< A_i >= \sum_{j=1}^{j=N} \frac{m_j}{\rho_j} A(r_j) W(r_{ij}, h) \tag{2.7}$$

where $r_{ij} = r_i - r_j$ and N is the number of particles within a smoothing length of particle i. Therefore, the first quantity that is calculated in the SPH code is the density of a particle from which other quantities (such as the pressure) are calculated. To calculate the average density, the variable ρ can be substituted for A into equation 2.7 to give:

$$< \rho_i >= \sum_{j=1}^{j=N} m_j W(r_{ij}, h). \tag{2.8}$$

It can also be shown (using Stokes' theorem and vector identities) that for the scalar quantity, A, and the vector quantity, $\mathbf{A}$, that

$$< \nabla A(r) > \; = \int_V A(r') \nabla W(r - r', h) \mathrm{d}^3 r' \tag{2.9}$$

and

$$< \nabla \cdot \mathbf{A}(\mathbf{r}) > \; = \int_V \mathbf{A}(\mathbf{r}') \cdot \nabla W(r - r', h) \mathrm{d}^3 r' \tag{2.10}$$

which in discrete form, become

$$< \nabla A(r) >_i = \rho_i \sum_j m_j \left(\frac{A_i}{\rho_i^2} + \frac{A_j}{\rho_j^2} \right) \nabla_i W_{ij} \tag{2.11}$$

and

$$< \nabla \cdot \mathbf{A}(r) >_i = -\frac{1}{\rho_i} \sum_j m_j (\mathbf{A}_i - \mathbf{A}_j) \cdot \nabla_i W_{ij} = \rho_i \sum_j m_j \left(\frac{\mathbf{A}_i}{\rho_i^2} + \frac{\mathbf{A}_j}{\rho_j^2} \right) \cdot \nabla_i W_{ij}, \tag{2.12}$$

respectively, where $\nabla_i W_{ij}$ is the gradient of the kernel. Other forms such as $\nabla^2 A$ and $\nabla \wedge \mathbf{A}$ can also be derived. Using equations 2.7- 2.8 and 2.11- 2.12, many of the physical quantities required to model the fluid can be calculated.

2.1.2 Smoothing length

The smoothing length is a very important variable to consider in SPH simulations. It defines the lengthscale of the region surrounding the particle being considered, over which the properties of the particle are calculated. Clarke & Carswell (2003) discuss the concept of a region over which local variables (i.e. most variables other than gravity such as pressure, temperature and density) are defined. They state that the size of the region must be

(i) much smaller than the lengthscale over which a variable changes its property,

(ii) large enough that it contains enough particles to represent a fluid rather than a small, discrete number of particles, and

(iii) for a collisionless system, the lengthscale must be larger than the mean free path of the system.

In SPH, these criteria determine the smoothing length. The smoothing length is the lengthscale over which there are enough particles to allow an averaging of a particle's variable to take place, or equivalently, it can be described as the resolution since for a fixed number of particles in a smoothing length, a smaller smoothing length indicates a larger overall number of particles in any one simulation, and hence a higher resolution in the system being modelled.

There are many implementations for the smoothing length, both time dependent and independent. The smoothing length used in these simulations is a time independent smoothing length that is spatially adaptive to the density but also maintains energy and entropy conservation (Springel & Hernquist 2002; Monaghan 2002; Price & Monaghan 2004, 2007). It is important to note, however, that though the smoothing length is not explicitly time-dependent, since it varies with the density which is time-dependent, the smoothing length is implicitly also time-dependent. The implementation used for the three-dimensional simulations presented in this thesis, described in detail in Price & Bate (2007), sets the smoothing length according to

$$h_i = \eta \left(\frac{m_i}{\rho_i} \right)^{\frac{1}{3}} \tag{2.13}$$

where m_i and ρ_i are the mass and density of particle i respectively, and the dimensionless parameter, $\eta = 1.2$, such that the average number of neighbours that each particle has is ≈ 60. The simulations carried out in Chapters 5 and 6 extend the work of Rice et al. (2005) and I compare my simulations to theirs. It is important to note that Rice et al. (2005) set their smoothing length using a different time-dependent method which approximately fixes the number of neighbours that

each particle has to ≈ 50 but their implementation does not conserve entropy. I carry out some initial benchmarking simulations in an attempt to reproduce their results and discuss the impact of this in Chapters 5 and 6.

The resolution is important to consider in numerical simulations involving gravitational collapse. If an adequate resolution is not used, simulations which should show fragmentation may not show this and incorrect conclusions may form. Bate & Burkert (1997) carried out SPH resolution tests and provided two key conclusions

(i) They showed that since the smoothing and gravitational softening lengths are spatially varying, it is not possible to have a fixed spatial resolution limit but it is possible to have a fixed mass resolution limit. They showed that the minimum resolvable mass, given by

$$M_{\rm res} \approx M_{\rm tot}\left(\frac{2N_{\rm neigh}}{N_{\rm tot}}\right), \tag{2.14}$$

must be less than the Jeans mass, given by

$$M_{\rm J} = \left(\frac{5R_{\rm g}T}{2G\mu}\right)^{\frac{3}{2}}\left(\frac{4}{3}\pi\rho\right)^{-\frac{1}{2}} \tag{2.15}$$

such that the density of the region to be resolved must satisfy

$$\rho < \left(\frac{3}{4\pi}\right)\left(\frac{5R_{\rm g}T}{2G\mu}\right)^{3}\left(\frac{N_{\rm tot}}{2N_{\rm neigh}}\frac{1}{M_{\rm tot}}\right)^{2}. \tag{2.16}$$

This is equivalent to saying that the smoothing length must be much smaller than the Jeans length. Since the smoothing length is spatially varying, this condition is satisfied provided the total number of particles is adequate.

(ii) For codes that allow different values of the gravitational softening length, ϵ, and the smoothing length, h, the relative values of these lengths are important. They showed that for a marginally stable Jeans mass clump, if the lengthscale over which gravity is softened, ϵ, is greater than the hydrodynamical smoothing length, h, the pressure forces, $F_{\rm p}$ may be higher than the gravitational forces, $F_{\rm g}$, and fragmentation may be inhibited. Conversely, if $\epsilon < h$, then $F_{\rm g} > F_{\rm p}$ and fragmentation may be artificially enhanced. To avoid such problems, it is required that $\epsilon \approx h$. Since the implementation used here involves the same width for both the smoothing and gravitational softening kernels, this condition is always met.

To illustrate the effects of a different gravitational softening length (as well as the use of different viscosity prescriptions), I carry out the "Wengen 4: Protoplanetary disk test" (available at http://www.astrosim.net/code/doku.php?id=home:codetest:hydrotest:wengen:wengen4). This is a code comparison test for a gravitationally unstable disc with set initial conditions. However, it is important to note that the creators of the Wengen test do in fact state that the results may be dependent on the softening used. Mayer & Gawryszczak (2008) carry out this test using a fixed number of neighbours and a fixed gravitational softening length. Using the Balsara (1995) switch implementation of the artificial viscosity (see Section 2.1.5) the authors find that the disc fragments. Using the SPH code with spatially adaptive smoothing lengths used for this thesis with a fixed viscosity, I carry out the same Wengen test and find that fragmentation does not

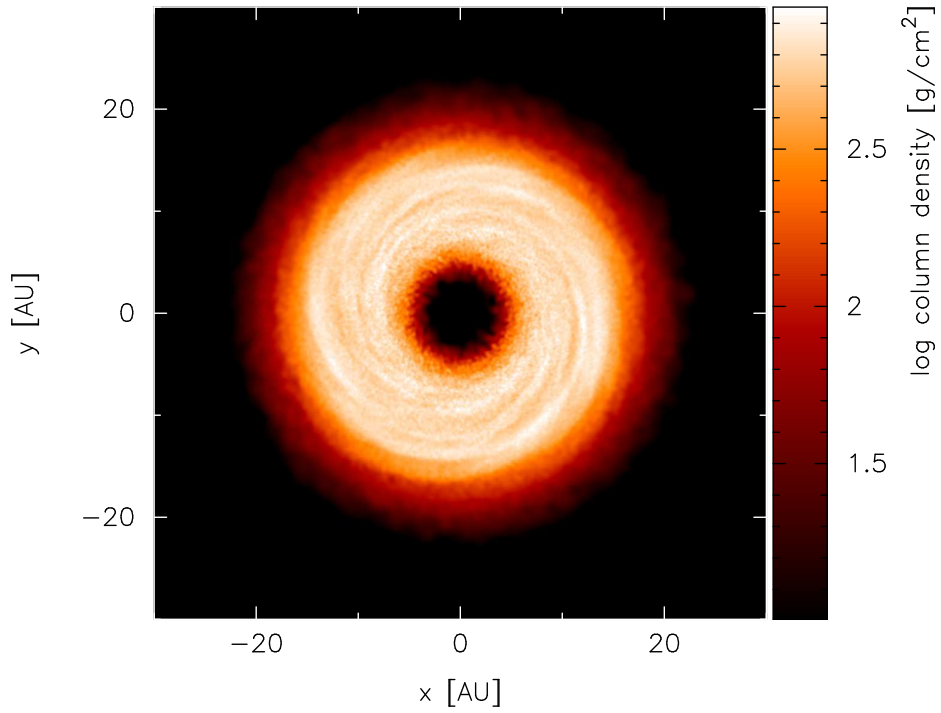

Figure 2.1: Final image of the Wengen test disc. Previous authors who have simulated this disc with a fixed gravitational softening and using a Balsara switch implementation for the artificial viscosity, find that a fragment forms. However, the same simulation carried out here with a spatially adaptive softening and a fixed viscosity shows no fragmentation. Thus, fragmentation results may be numerically dependent.

occur (Figure 2.1). This illustrates that the fragmentation results may be affected by numerical effects such as the softening lengths as well as the artificial viscosity prescription and therefore it is important to consider such numerical aspects carefully. Since the code used for the simulations in this thesis is spatially varying so that condition 2.16 is adhered to and since the gravitational softening and smoothing lengths are equivalent, provided the correct number of particles is used, the method employed here tackles the resolution issues well. The Wengen test shows that my calculations are more resistant to fragmentation than some other codes employed within the planet formation community and so any fragmentation that is seen in my calculations is more certain.

In a particular spatial range, if the number of particles is increased by a factor, f, the resolution, and hence the smoothing length, is increased by a factor $f^{\frac{1}{d}}$, where d is the number of dimensions. In Chapter 6, I carry out a resolution test on some earlier work carried out by Rice et al. (2005). Their simulations used 250,000 particles. In the resolution test, I decrease the linear resolution by a factor of 2 (i.e. using 31,250 particles) and also increase the linear resolution by a factor of 2 and 4 (using 2 million and 16 million particles, respectively).

2.1.3 Hydrodynamical equations

There are three main equations that need to be solved in hydrodynamical simulations. These are the mass continuity, momentum and thermal energy equations given in Lagrangian form by

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\nabla\cdot\mathbf{v} = 0, \tag{2.17}$$

$$\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\frac{\nabla p}{\rho} - \nabla\Phi, \tag{2.18}$$

and

$$\frac{\mathrm{D}u}{\mathrm{D}t} = -\frac{p}{\rho}\nabla\cdot\mathbf{v}, \tag{2.19}$$

where $\mathbf{v}$ is the velocity vector, p is the thermal pressure, Φ is the gravitational potential and u is the internal energy. The internal energy equation is derived from the first law of thermodynamics in the absence of external heating or cooling.

To put these fluid equations into a discrete SPH format, the code uses a Lagrangian for the self-gravitating gas (Price & Monaghan 2007),

$$L = \sum_{i=1}^{i=N} m_i\left(\frac{1}{2}v_i^2 - \Phi_i - u_i\right), \tag{2.20}$$

(where i denotes any one particle), the first law of thermodynamics and a method by which the total density can be estimated. Firstly, a particle's density is calculated using equation 2.8 using its current smoothing length. However at the same time, as seen in equation 2.13, the smoothing length is calculated using the particle's current density. This is therefore carried out iteratively using a Newton-Raphson iterative method and if necessary, combining this with a bisection scheme. Once the density and smoothing length have been obtained, the properties of the fluid particles are then evolved over time using the discrete forms of the fluid equations. Using the method described by Price & Monaghan (2007), the mass continuity equation in discrete form is

$$\frac{\mathrm{D}\rho_i}{\mathrm{D}t} = \frac{1}{\Omega_i}\sum_j m_j\mathbf{v}_{ij}\cdot\nabla_i W_{ij}(r_{ij}, h_i) \tag{2.21}$$

where

$$\Omega_i = \left(1 - \frac{\partial h_i}{\partial\rho_i}\sum_j m_j\frac{\partial}{\partial h_i}W_{ij}(r_{ij}, h_j)\right), \tag{2.22}$$

and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. The momentum equation (without the gravitational term) and energy equation are

$$\frac{\mathrm{D}\mathbf{v}_{i,\mathrm{hydro}}}{\mathrm{D}t} = -\sum_j m_j\left(\frac{P_i}{\Omega_i\rho_i^2}\nabla_i W_{ij}(r_{ij}, h_i) + \frac{P_j}{\Omega_j\rho_j^2}\nabla_i W_{ij}(r_{ij}, h_j)\right) \tag{2.23}$$

and

$$\frac{\mathrm{D}u_i}{\mathrm{D}t} = \frac{P_i}{\Omega_i\rho_i^2}\sum_j m_j\mathbf{v}_{ij}\cdot\nabla_i W_{ij}(r_{ij}, h_i), \tag{2.24}$$

respectively.

Since the Lagrangian has been used, these forms of the momentum, continuity and energy equations are conserved. In order to model shocks well, artificial viscosity is required to convert bulk motion into heat. This requires the momentum and energy equation to be modified, the description of which is in Section 2.1.5. Furthermore, an addition of the gravitational force term requires equation 2.23 to be modified further (see Section 2.1.4).

The calculations carried out for this thesis simulate the energetics in two different ways. In all of the simulations, the heating effects due to work done on the gas and due to viscous heating (through the inclusion of artificial viscosity to capture shocks) is included. In Chapters 5 and 6, the cooling in the disc is taken into account using the cooling parameter, β (equation 1.14), which cools the gas on a timescale given by equation 1.15 while in Chapters 3 and 4, radiative transfer is used. I discuss the effect that these two techniques have on the energy equation in Sections 2.2.1 and 2.2.2, respectively.

2.1.4 Gravitational force calculations

The gravitational force calculation carried out in the SPH code adopted here is in principle similar to a force calculation carried out in an N-body code. In discrete form, the gravitational force (per unit mass) between particle i and all other particles in the system is given by:

$$\frac{\mathrm{D}^2\mathbf{r}_i}{\mathrm{D}t^2} = -\sum_{j=1,j\neq i}^{N} \frac{Gm_\mathrm{j}\mathbf{r}_\mathrm{ij}}{|\mathbf{r}_\mathrm{ij}|^3} = -\nabla\Phi_i \tag{2.25}$$

where Φ_i is the gravitational potential for particle i. The force calculations may prove to be problematic if they are carried out between two very close particles as it will cause the acceleration, given by equation 2.25, to become very large. The problem is simply that the particle is not a point masses, but is a *volume* of fluid of finite size and therefore when two particles come close together, they should not act as point masses. However equation 2.25 alone would treat the particles as point masses. The gravitational forces are calculated using the *Binary Tree Method* which not only determines which particles are the nearest neighbours to particle i, but also calculates the gravitational force due to these particles individually. For more distant particles (further away than $2h$), the gravitational force contributions from individual particles is not necessary and instead the combined contribution is sufficient. This is carried out using the *node opening criterion*. The smallest node possible is a group of two particles, each of which is the nearest neighbour to the other i.e. they are mutual nearest neighbours. A node consisting of three particles is one where this pair of particles and another single particle are mutual nearest neighbours. The mutual nearest neighbour can be found for this triple to make a node consisting of four particles. This process can be repeated until a node of n particles is produced. In addition, a node of n particles can also be made using two smaller nodes rather than a node and a single particle. The node of n particles has size h_node which is equal to the sum of the maximum separation between the smaller nodes that make up the node of n particles and/or the individual particles themselves, and the greatest value of $2h_i$ of the particles that make up the node of n particles.

To calculate the gravitational force on a particle j, the separation, $r_{j,\mathrm{node}}$ between particle j and a node is considered. If the separation is less than $2h_\mathrm{node}$ i.e. if $h_\mathrm{node}/r_{j,\mathrm{node}} > 0.5$, then the

node will either be split into a single particle and a node of $n-1$ particles, or into two smaller nodes. The gravitational force between particle j and the single particle is calculated directly. For the node that now consists of $n-1$ particles, or for the two smaller nodes, the node opening criterion is repeated. If at any stage the node opening criterion is not satisfied, then the gravitational force between particle j and the node is calculated.

The gravitational potential is given by (Price & Monaghan 2007)

$$\Phi_i(r) = -G\sum_j m_j\phi(|r_{ij}|, h_i) \tag{2.26}$$

where ϕ is the gravitational softening kernel which avoids the problems outlined above and is related to the smoothing kernel by

$$W(|r_{ij}, h_i|) = -\frac{1}{4\pi r^2}\frac{\partial}{\partial r}\left[r^2\frac{\partial}{\partial r}\phi(|r_{ij}|, h_i)\right]. \tag{2.27}$$

This results in an additional gravity term added to equation 2.23:

$$\frac{\mathrm{D}\mathbf{v}_{i,\mathrm{gravity}}}{\mathrm{D}t} = -G\sum_j m_j\left[\frac{\phi'_{ij}(h_i)+\phi'_{ij}(h_j)}{2}\hat{\mathbf{e}}_{ij}\right] - \frac{G}{2}\sum_j m_j\left[\frac{\zeta_i}{\Omega_i}\nabla_i W_{ij}(h_i) + \frac{\zeta_j}{\Omega_j}\nabla_i W_{ij}(h_j)\right], \tag{2.28}$$

where $\phi'_{ij} = \partial\phi/\partial\mathbf{r}_{ij}$, $\hat{\mathbf{e}_{ij}} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|$ and

$$\zeta_i \equiv \frac{\partial h_i}{\partial\rho_i}\sum_j m_j\frac{\partial\phi_{ij}(h_i)}{\partial h_i}. \tag{2.29}$$

The first term in equation 2.28 is due to the gravitational force term that has been softened while the second is required to ensure energy conservation and takes into account the gradient in the smoothing length. In the past, the second term has not been taken into account thus resulting in only one of entropy or energy being conserved (usually energy is conserved). The contributions to the gravitational force on particle i from its neighbours (within a distance of $2h$) is considered from each individual particle. For the more distant particles, the gravitational force between individual particles is simply that of two point masses.

2.1.5 Artificial viscosity

In a fluid, as particles approach a shock front, bulk kinetic energy is converted to heat which causes a discontinuity in the density, pressure and energy at the shock front. However, such discontinuities cannot be treated exactly in numerical schemes and consequently, some smearing of the discontinuous density, pressure and energy profiles takes place at the shock front. The discontinuities are on the smallest lengthscales (the smoothing length) and since this is not resolved, particle oscillation may result which is a numerical artefact. To convert bulk motions into heat and dampen these oscillations, SPH requires the use of artificial viscosity which spreads the shock front so that it can be resolved. A number of forms of artificial viscosities have been proposed (Lucy 1977; Wood 1981; Monaghan & Gingold 1983; Loewenstein & Mathews 1986; Evrard

1988; Hernquist & Katz 1989). We use a common form of artificial viscosity by Monaghan & Gingold (1983) which is a time-independent fixed artificial viscosity. This adds an additional term to the momentum equation:

$$\frac{\mathrm{D}\mathbf{v}_{i,\nu}}{\mathrm{D}t} = -\sum_j m_j \Pi_{ij} \nabla_i W_{ij}, \tag{2.30}$$

where

$$\Pi_{ij} = \begin{cases} \frac{-\alpha_{\mathrm{SPH}}\ c_{\mathrm{s},ij}\ \mu_{ij} + \beta_{\mathrm{SPH}}\ \mu_{ij}^2}{\rho_{ij}} & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0 \\ 0 & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} > 0, \end{cases} \tag{2.31}$$

$$\mu_{ij} = \frac{h\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{\mathbf{r}_{ij}^2 + \eta^2}, \tag{2.32}$$

$c_{\mathrm{s},ij} = (c_i + c_j)/2$, $\rho_{ij} = (\rho_i + \rho_j)/2$, $\eta = 0.01h^2$ and ensures equation 2.32 does not diverge and α_{SPH} and β_{SPH} are the SPH viscosity parameters. This form includes artificial viscosity when particles are approaching each other (i.e. at a shock front) but not when the particles are moving away from each other. Viscosity comes in the form of bulk and shear viscosity. α_{SPH} provides a bulk term which dissipates kinetic energy as particles approach each other at the shock front to avoid particle oscillations once they have passed through the shock front. This term dominates when velocity differences between particles are small. The role of β_{SPH} is to stop particle interpenetration which may occur if the velocity differences between particles are highly supersonic and is thus important for high Mach numbers. The additional term in the energy equation as a result of this is

$$\frac{\mathrm{D}u_{i,\nu}}{\mathrm{D}t} = \frac{1}{2}\sum_j m_j \Pi_{ij} v_{ij} \cdot \nabla_i W_{ij}(r_{ij}, h_i). \tag{2.33}$$

In the discs modelled in this thesis, artificial viscosity is applied everywhere, not only at the shock front, and therefore a corollary of including it is that it adds shear viscosity and causes dissipation in the form of a Shakura & Syunyaev (1973) α-viscosity given by (equation 1.18)

$$\nu_{\mathrm{av}} = \alpha_\nu c_s H \tag{2.34}$$

where ν_{av} is the shear viscosity in the disc due to artificial viscosity, c_{s} is the sound speed, $H = \frac{c_{\mathrm{s}}}{\Omega}$ is the scaleheight in the disc set by the condition of vertical isothermality and α_ν is the turbulent viscosity parameter which describes the stress tensor in terms of the gas pressure given by

$$\alpha_\nu = 0.05\alpha_{SPH}\left\langle\frac{h}{H}\right\rangle. \tag{2.35}$$

Equation 2.35 is the form of the artificial viscosity in the continuum limit using the SPH kernel adopted in this code (Meglicki et al. 1993, M. R. Bate, private communication). The dissipation as a result of this is given by (using equations 1.17 and 2.34)

$$D(R) = T^{\nu}_{R\phi}|R\Omega'| = \left|\frac{\mathrm{dln}\Omega}{\mathrm{dln}R}\right|\alpha_\nu c_s^2 \Sigma |R\Omega'| \tag{2.36}$$

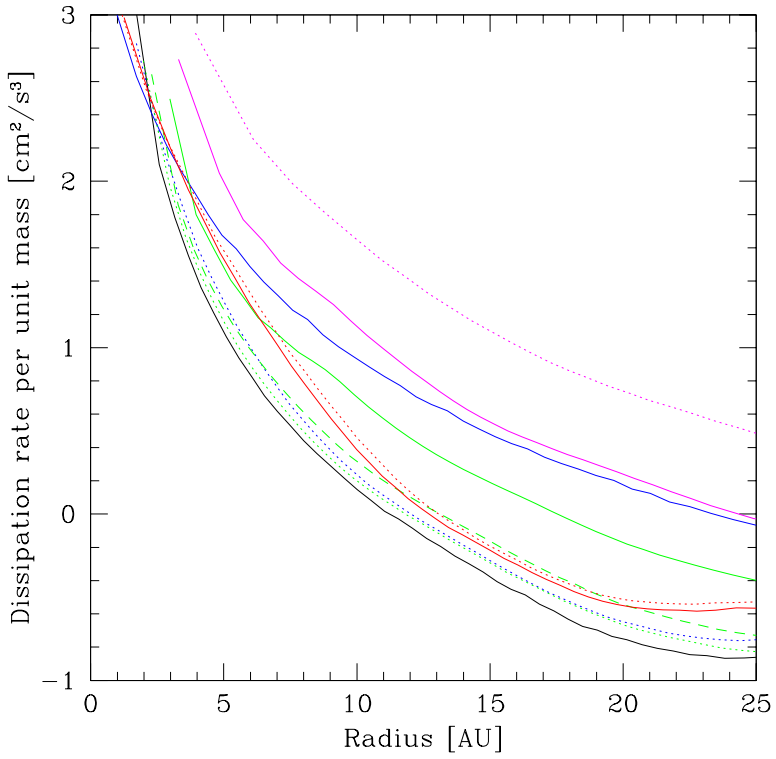

Figure 2.2: Plot showing the dissipation rate per unit mass against radius for a disc simulated with various different artificial viscosity parameters. The black solid line uses $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.1, 0.2)$ as done so by Lodato & Rice (2004) and Rice et al. (2005). Blue lines refer to values of $\alpha_{\rm SPH} < 0.1$ ($(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.01, 0.02)$ and $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.07, 0.14)$ are the blue solid and dotted lines, respectively). Green lines represent a viscosity higher than the (0.1, 0.2) by up to a factor of 10 ($(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.13, 0.26)$, $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.3, 0.6)$ and $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (1.0, 2.0)$ are the green dotted, dashed and solid lines, respectively). Magenta lines represent even higher viscosities $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (3.0, 6.0)$ and $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (10.0, 20.0)$ are the magenta solid and dotted lines, respectively). The red lines represent no change in $\alpha_{\rm SPH}$ but $\beta_{\rm SPH} = 0.01$ and 0.0 for the solid and dotted lines, respectively. A simulations with $(\alpha_{\rm SPH},\beta_{\rm SPH}) = (0.0, 0.0)$ was also carried which, as expected, produced no dissipation at all.

where $T^{\nu}_{R\phi}$ is the resulting stress tensor due to the artificial viscosity. The additional term in the energy equation as a result of this is

$$\frac{{\rm D}u_{i,\nu}}{{\rm D}t} = \frac{1}{2}\sum_j m_j \Pi_{ij} v_{ij} \cdot \nabla_i W_{ij}(r_{ij}, h_i) \tag{2.37}$$

If the artificial viscosity is too large, the hydrodynamical evolution of the disc may be driven by it, while if it is too small, it will lead to inaccurate modelling of shocks (Bate 1995). Numerical modelling of hydrodynamical astrophysical situations therefore requires careful thought as to what the artificial viscosity parameters should be so that shocks are modelled correctly without significant contributions to the energetics due to their inclusion.

The values of $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$ are typically ≈ 1 and 2, respectively, as these values reduce the root mean square velocity dispersion of the particles as they pass through the shock (Bate 1995).

A number of studies have used lower values to ensure that the angular momentum transport in discs modelled using SPH is due to the gravitational instability and not due to artificial viscosity (Lodato & Rice 2004, 2005). I have tested various values of the SPH artificial viscosity parameters. Figure 2.2 shows the viscous dissipation rate (in discs simulated with radiative transfer without self-gravity) resulting from different values of $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$. It can be seen that a change in $\beta_{\rm SPH}$ does not significantly affect the dissipation rate while a change in $\alpha_{\rm SPH}$ does: using very high or very low values of $\alpha_{\rm SPH}$ causes the viscous dissipation rate to increase. While the increase in dissipation rate is expected for high values of $\alpha_{\rm SPH}$, the increase with very low values is somewhat counterintuitive. Since the artificial viscosity also dampens random particle motions, using a low value of $\alpha_{\rm SPH}$ will mean that the particles will have too much kinetic energy (provided by the bulk motions) which will then be dissipated by the artificial viscosity (the resulting dissipation being more than the dissipation if the artificial viscosity value was slightly higher). This is purely a numerical effect. Therefore, another careful balance is required here. Figure 2.2 shows that the minimum viscous dissipation due to artificial viscosity occurs when $\alpha_{\rm SPH} \approx 0.1$. Since my thesis is concerned with the energetics in discs, reducing the dissipation rate is important and hence I choose this value. Bate (1995) shows that low values of $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$ are reasonable for low Mach number shocks ($\mathscr{M} \lesssim 2$). In these simulations, I expect the shocks to have low Mach numbers and so the value of $\alpha_{\rm SPH}$ chosen here provides a good compromise between requiring shocks to be modelled correctly as well as low dissipation rates. Since typically, $\beta_{\rm SPH} \sim 2\alpha_{\rm SPH}$, I choose $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$ to be 0.1 and 0.2, respectively. These are the same values as those implemented by Lodato & Rice (2004) and Rice et al. (2005) which is important since the work in this thesis involves some comparison with these authors.

A number of methods have been devised to decrease the amount of artificial viscosity in regions where they are not required. One such method is the Balsara (1995) switch which reduces the artificial viscosity in places where the ratio of the divergence to the curl of the velocity field is small i.e. in regions of strong vorticity. This uses a modified form of the μ_{ij} term in equation 2.32:

$$\mu_{ij} = \frac{h\mathbf{v}_{ij}\cdot\mathbf{r}_{ij}}{\mathbf{r}_{ij}^2+\eta^2}\left(\frac{f_i+f_j}{2}\right) \tag{2.38}$$

where

$$f_i = \frac{|\nabla\cdot\mathbf{v}|_i}{|\nabla\cdot\mathbf{v}|_i + |\nabla\wedge\mathbf{v}|_i + 0.0001c_{s,i}/h_i}. \tag{2.39}$$

This is a form of the viscosity used by Lodato & Rice (2004). Another method is a time dependent method by Morris & Monaghan (1996) which assigns a viscosity parameter, $\alpha_{\rm SPH}$, to each individual particle. This viscosity parameter decays for each particle over time, but increases as the particle approaches a shock front.

I have chosen not to include the Balsara (1995) switch since it reduces the viscosity in shearing flows but given that my simulations involve both shocks and shearing flows, the reduction may not handle the shocks well. In addition, a reason against using both the Balsara (1995) and Morris & Monaghan (1996) methods is that I want to ensure a controlled test which is not possible if the numerical viscosity is constantly changing. I have therefore chosen the fixed viscosity

method of Monaghan & Gingold (1983).

2.1.6 Timestepping

To integrate the second order ordinary differential force equation given by the sum of equations 2.23, 2.28 and 2.30, one can decompose it into two first order ordinary differential equations:

$$\frac{\mathrm{D}\mathbf{r}_\mathrm{i}}{\mathrm{D}t} = \mathbf{v}_i \tag{2.40}$$

and

$$\frac{\mathrm{D}^2\mathbf{r}_i}{\mathrm{D}t^2} = \frac{\mathrm{D}\mathbf{v}_i}{\mathrm{D}t} = \frac{\mathrm{D}\mathbf{v}_{i\mathrm{hydro}}}{\mathrm{D}t} + \frac{\mathrm{D}\mathbf{v}_{i,\mathrm{gravity}}}{\mathrm{D}t} + \frac{\mathrm{D}\mathbf{v}_{i,\nu}}{\mathrm{D}t} \tag{2.41}$$

A first order ordinary differential equation can be solved either explicitly or implicitly. An explicit method uses the values of a variable, A, and its differential with respect to time, at a particular timestep, n, to determine its value at the next timestep, $n+1$, as follows:

$$A_i^{n+1} = A_i^n + \mathrm{d}t\left(\frac{\mathrm{d}A_i}{\mathrm{d}t}\right)^n. \tag{2.42}$$

An implicit scheme solves the ordinary differential equation using the Backwards Euler method as follows:

$$A_i^{n+1} = A_i^n + \mathrm{d}t\left(\frac{\mathrm{d}A_i}{\mathrm{d}t}\right)^{n+1} \tag{2.43}$$

Explicit schemes are typically easier and less computationally demanding but they can suffer from numerical instability. Most of the equations in the SPH code used here are solved explicitly. At each timestep, the velocities and positions of the particles are updated using a form of the second order Runge-Kutta-Fehlberg (Fehlberg 1968) integration scheme developed by Benz (1984). This scheme carries out two force calculations per timestep. It is important to ensure that the code is not only accurate but also numerically stable i.e. such that the errors introduced due to the numerical scheme used to solve the differential equations explicitly do not grow exponentially. This may occur if the timestep, Δt, is limited to a value smaller than a certain critical value associated with the typical timescale of the problem being considered. In hydrodynamical simulations, one such typical timescale is the timescale on which sound propagates. The Courant-Friedrichs-Lewy (CFL) condition (Courant et al. 1928) is the most common timescale condition in hydrodynamical simulations and states that

$$\Delta t_i \leq C_{CFL}\frac{\Delta x_i}{c_\mathrm{s}} \tag{2.44}$$

where C_{CFL} is a constant less than unity, which for SPH becomes

$$\Delta t_i \leq C_{CFL}\frac{h_i}{c_\mathrm{s}}. \tag{2.45}$$

The constant C_{CFL} is typically taken to be $C_{CFL} = 0.3$. In addition, since viscosity is included, the

timestepping is also limited by the viscous timescale. In the SPH code employed here, these two timestepping conditions are combined to give (Hernquist & Katz 1989):

$$\Delta t_{cv,i} = \frac{C_o h_i}{h_i|\nabla \cdot \mathbf{v}_i| + c_i + 1.2(\alpha c_i + \beta \max_j |\mu_{ij}|)}. \tag{2.46}$$

where $C_o = 0.3$ and $\mu_{ij} = 0$ for diverging particles (i.e. $\mathbf{v}_{ij} \cdot \mathbf{r}_{ij} > 0$). The term $h_i|\nabla \cdot v_i|$ is an additional term used to improve the accuracy when $h_i|\nabla \cdot v_i| > c_i$ i.e. when supersonic motions are involved.

In addition, the SPH code also limits the timestep due to a force condition to ensure that smaller timesteps are used if particles suffer large accelerations (Monaghan 1992):

$$\Delta t_{f,i} \leq \left(\frac{h_i}{|a_i|}\right)^{\frac{1}{2}}. \tag{2.47}$$

Finally, in the non-radiative transfer calculations carried out in Chapters 5 and 6, since I apply an explicit cooling rate given by equations 1.14 and 1.15, it is important to ensure that the following timestep criteria is also met:

$$\Delta t_{c,i} \leq C_\beta \frac{\beta}{\Omega}, \tag{2.48}$$

where C_β is a constant less than unity, β is the cooling timescale in units of the orbital timescale and Ω is the angular velocity. In Chapter 5, I investigate the effects of varying the constant C_β in equation 2.48 on the critical cooling timescale $\beta_{\rm crit}$ by testing values of $C_\beta = 0.3$, 0.03 and 0.003.

Combining the timestep criteria from equations 2.46 to 2.48, the timestep for each individual particle is determined by

$$\Delta t_i = \min\left(\Delta t_{cv,i},\ \Delta t_{f,i},\ \Delta t_{c,i}\right). \tag{2.49}$$

and therefore, the SPH code does not advance all the particles on the same timestep. The radiation hydrodynamic equations are solved using an implicit scheme since the electromagnetic radiation propagates much faster than the speed of sound (and therefore an explicit scheme would require very small timesteps). Though implicit schemes are not subject to numerical instability it is still important to limit the timestep to ensure numerical accuracy. The timestepping that is used above is more than adequate to ensure the numerical accuracy of the results and has been tested by Whitehouse et al. (2005).

2.1.7 Particle types

The simulations presented in this thesis are of discs that are modelled using SPH particles to represent the gas in the discs. These are particles that have properties associated with them such as mass, spatial and velocity co-ordinates, density and temperature. However, they do not have a specified radius (hence the need for smoothing and gravitational softening lengths). The central stars are modelled using a sink particle (Bate et al. 1995) which only have the property of mass and spatial and velocity co-ordinates associated with them. Sink particles are able to accrete gas

particles subject to particular boundary conditions: particles are accreted onto the sink particle if they move within a radius $R_{\rm in}$ of the sink particle or if they move within $R_{\rm in}$ and $R_{\rm acc}$ (where $R_{\rm acc} > R_{\rm in}$) and are gravitationally bound to the sink particle.

In regions of high density, the calculations are carried out on very small timesteps. This means that for discs which have fragmented into high density clumps, the simulations will progress very slowly due to the high computational requirements. Therefore, it would not be possible to follow the simulations past the fragmentation stage. However, in very high density regions where the fragments are not likely to shear apart during the evolution of the disc, the fragments can be turned into sink particles, thus allowing the discs to evolve further on reasonable computational timescales. In Chapter 4, I follow the evolution of protoplanetary discs following the fragmentation stage by turning a clump of particles that are denser than a critical value (see Table 4.2) into sink particles. As with the sink particles used for the central stars, the newly formed sink particles also carry the same accreting conditions as described above.

2.1.8 Speeding up the SPH code

As described in Section 2.1.4, since gravity is a long-range force, the calculation of the gravitational force on any one particle needs to consider not just the gravitational effects of the particles within its locality, but also the gravitational effects of the particles that are further away. To calculate the force on particle i, the summation, given by equation 2.25 has to take place over all particles. To do this the computational time taken would scale as O(N^2), where N is the number of particles used in the simulation. The method adopted to reduce this computational expense is termed a *Binary Tree Method*, described in Section 2.1.4. The binary tree is determined at the start of the simulation. It is then updated at each timestep to take into account the changes due to particle movement and is rebuilt at every longest timestep. Since it does not have to be reproduced from the start at each and every timestep, it is a very efficient method. The tree method reduces the computational time from O(N^2) to O(Nlog$N$) as it identifies regions to be discarded (such as if a large sink particle exists in between particle i and j, the gravitational effect of particle j on i will be negligible). This reduction in computational expense thus allows much higher resolution simulations to be possible. For simulations that use radiative transfer, it is unclear what the computational time scaling is with the number of particles, though it is more computationally expensive since there are additional equations that need to be solved iteratively (see Section 2.2.2). Nevertheless, even in these simulations, the tree format helps to reduce the computational time.

In addition, since increasing the number of particles causes the smoothing length to decrease, the timesteps on which the particles can be advanced must also decrease since the particles cannot be advanced by more than a smoothing length in a timestep. Since the smoothing length scales as $N^{-1/d}$, where d is the number of dimensions, the timesteps taken decrease by this factor. Therefore, not only does the computational time increase due to an increase in number of particles, it also increases further due to the smaller timesteps. Therefore, the SPH code can be speeded up by decreasing the number of particles, but one must ensure that the resolution requirements are adequate enough as discussed in Section 2.1.2.

Another computationally demanding component is the regular need to evaluate the kernel

and its derivatives. To save time, this is evaluated for a range of values at the start of a simulation and stored in a table. When it is required, the value of the kernel is obtained by interpolating within the table, thus avoiding the need for heavy computational effort too frequently.

The SPH code also carries out the timestepping of the particles in set blocks such that the timesteps are integer powers of 2 times the smallest timestep (Bate 1995). As mentioned in Section 2.1.6, the SPH code does not advance all the particles on the same timestep. Particles in lower density regions are advanced on longer timesteps which saves computational time. In addition, particles are not advanced individually, but together with all the particles that are being moved on that same timestep.

The SPH code used here is fully parallelised using OpenMP and MPI. All the simulations use OpenMP which is a form of the code that allows it to be advanced in a parallel way on all the processes on any one computer node (cores that share the same global memory). Each computer node on the University of Exeter supercomputer consists of 8 compute cores. For the higher resolution simulations, I used MPI as well as OpenMP which offers the advantage that many computer nodes (and the processes that exist on those nodes) can be combined and used together to increase the parallel power. However, as the number of nodes used increases, the overheads associated with combining these nodes also increase, thus making a large number of nodes only adequate if the number of particles being used in any one simulation is high. For some simulations where 2 million particles were used, I carried out the simulations using MPI on a maximum of 2 nodes. For simulations involving 16 million particles, I carried out a number of timing tests which showed that generally, if 8 computer nodes were used (which is an adequate number of nodes given the computational demands by all the users of the University's supercomputer), the most optimum number of processes is 16 i.e. so that 4 OpenMP threads are used per MPI process.

In addition, I found that as the discs evolved, some particles may have moved far away from the bulk of the disc such that the gravitational effects are minimal. For simulations consisting of 16 million particles, this proved to slow the code down since the smoothing lengths become large and many neighbours may be found. However, since their contributions from such a large distance was not significant, in order to speed up the computing time, I chose to remove any particles that moved out to further than twice the original disc radius. This meant that not only did the code speed up due to not needing to incorporate these far-away particles into the calculations, but it also speeded up as the reduced number of particles meant that smaller array sizes and hence smaller executables could be used, reducing the memory requirements, thus saving overall computational time through better cache use.

2.2 Thermodynamics

The simulations presented in this thesis employ one of two types of thermodynamics. Chapters 3 and 4 employ radiative transfer, which considers the transfer of energy between individual particles. I also carry out simulations with a parameterised cooling method in Chapters 5 and 6, which simplifies the disc thermodynamics and can be used to understand some of the concepts behind fragmentation that would otherwise be more difficult to understand if complex thermodynamics

are employed.

2.2.1 Parameterised cooling

Parts of this thesis (Chapters 5 and 6) involve simulations which are very closely related to those of Rice et al. (2005). As mentioned in Section 2.1.3, the heating effects due to work done on the gas as well as the viscous heating is included in these simulations. The cooling in the disc is taken into account using the cooling parameter, β (equation 1.14), which cools the gas on a timescale given by equation 1.15. This requires an additional term in the energy equation (2.19):

$$\frac{\mathrm{D}u_{\mathrm{cool}}}{\mathrm{D}t} = -\frac{u}{t_{\mathrm{cool}}} = -\frac{u\Omega}{\beta}. \tag{2.50}$$

2.2.2 Radiative transfer

In Chapters 3 and 4 I carry out simulations which incorporate radiative transfer. The version of SPH used to carry out the simulations presented in these chapters includes time-dependent radiative transfer using the flux-limited diffusion approximation (Whitehouse et al. 2005; Whitehouse & Bate 2006) with two temperatures: that of the gas and that of the radiation field. The radiative transfer code has undergone significant testing by Whitehouse et al. (2005). As mentioned in Section 2.1.3, this requires additional terms in the momentum and internal energy equations:

$$\frac{\mathrm{D}\mathbf{v}_{RT}}{\mathrm{D}t} = \frac{\chi}{c_{\mathrm{s}}}\mathbf{F}, \tag{2.51}$$

and

$$\frac{\mathrm{D}u_{\mathrm{RT}}}{\mathrm{D}t} = -4\pi\kappa_P B + c_s\kappa_E E. \tag{2.52}$$

where χ is the total frequency-independent opacity (including absorption and scattering), $\mathbf{F}$ is the is the total frequency-integrated radiation flux, B is the Planck function, κ_E and κ_P are the energy mean and Planck mean absorption opacities. Equation 2.51 describes the increase in momentum of the gas due to the radiation flux. The first and second terms on the right hand side of equation 2.52 are due to the emission and absorption of energy by the gas, respectively, i.e. correspond to the energy transfer between the radiation field and the gas. In addition, there is a further radiation hydrodynamical equation which is the change in the energy density of the radiation field given by:

$$\rho\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{E}{\rho}\right) = -\nabla\cdot\mathbf{F} - \nabla v : \mathbf{P} + 4\pi\kappa_P\rho B - c\kappa_E\rho E, \tag{2.53}$$

where $\mathbf{P}$ and E are the total frequency-integrated radiation pressure tensor and radiation energy density. The change in the radiation momentum density and pressure are described by the first two terms whereas the final two terms correspond to the emission and absorption of the energy by the gas (i.e. these are directly correlated with the two terms on the right hand side of equation 2.52).

The method employed to simulate the radiative transfer is a flux-limited diffusion method which uses Fick's law of diffusion:

$$\mathbf{F} = -D\nabla E \tag{2.54}$$

where the diffusion constant, D, is given by

$$D = \frac{c_s \lambda_F}{\chi\rho} \tag{2.55}$$

where λ_F is the flux limiter given by

$$\lambda_F(R_F) = \frac{2 + R_F}{6 + 3R_F + R_F^2} \tag{2.56}$$

and R_F is a dimensionless quantity given by $R_F = |\nabla E|/(\chi\rho E)$. The flux limiter ensures that in the optically thin regions where $\chi\rho \to 0$, the radiation flux does not become unphysical (i.e. it does not move faster than the speed of light). The method employed here uses the flux limiter of Levermore & Pomraning (1981). It ensures that in the optically thick limit, $\lambda_F \to 1/3$, which assumes that the radiation field is isotropic and that this holds everywhere in the optically thick region (the Eddington approximation). In the optically thin limit $\lambda_F \to 1/R_F$ so that the radiation moves at the speed of light. In these two limits, flux limited diffusion works very well as it describes the transfer of energy by diffusion in the optically thick region and the streaming out of energy at the speed of light in the optically thin region. However, while it maintains the correct rate of energy transportation in the optically thin region, it does not maintain directionality since the energy transport process is diffusion rather than free-streaming. The boundary between the two regions is not as well described as it is in the extremes but it is important to note that to describe these regions well enough, the full time-dependent radiation transport equations need to be solved which is very computationally expensive. It is therefore necessary to make this trade-off between computational expense and physical accuracy.

2.2.3 Opacity

The discs simulated are assumed to be in local thermal equilibrium and the opacities are assumed to be grey Rosseland mean values. The opacities are based on the opacity tables of interstellar molecular dust grains produced by Pollack, McKay, & Christofferson (1985), and on Alexander (1975) for the higher temperature gaseous contributions. Strictly speaking, the Rosseland mean opacity should be used in optically thick regions since the Rosseland mean is an average of $1/\kappa_\nu$ over all frequencies, while the Planck mean opacity averages κ_ν directly over all frequencies and should be used in the optically thin regions. However, since Semenov et al. (2003) shows that both the Rosseland and Planck mean opacities give very similar values in the temperature range being considered for the simulations in this thesis ($\lesssim 400$ K), I use the Rosseland mean opacities for all regions of the disc.

Figure 2.3 shows the change in Rosseland mean opacity with temperature for a density $\rho = 1 \times 10^{-8}\mathrm{g/cm}^3$. Though the exact values at which the features of this graph exist may change with density, the overall shape of the opacity curve is the same for all density values. Table 2.1 summarises how the opacity scales with temperature in the regimes dominated by different types

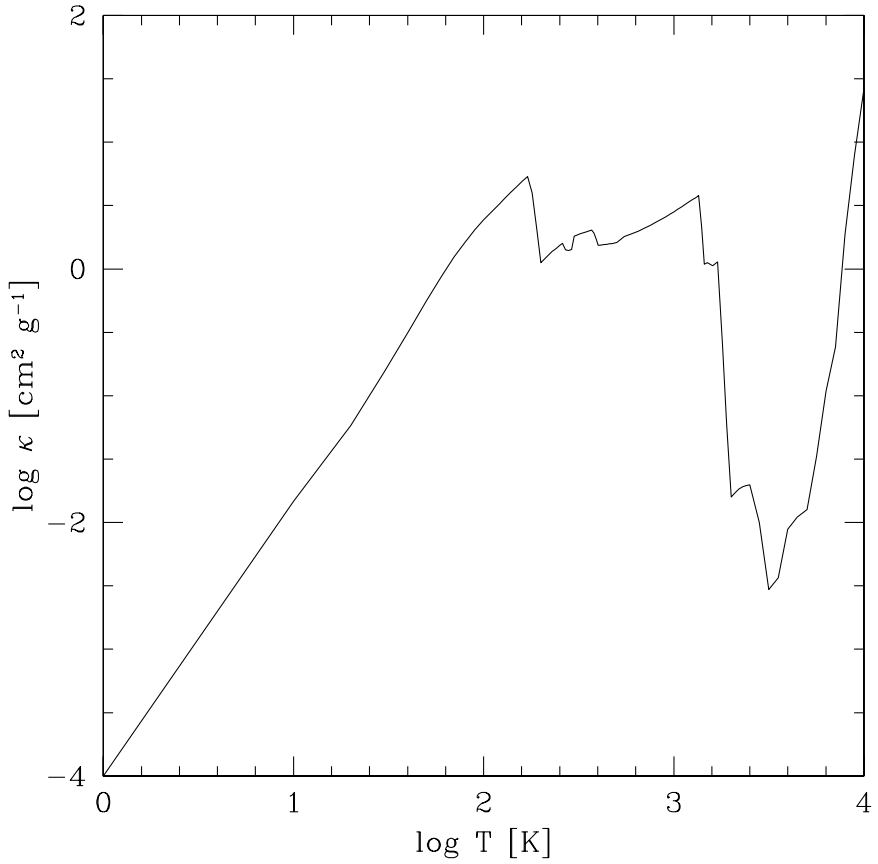

Figure 2.3: Plot showing the change in opacity with temperature for a density $\rho = 1 \times 10^{-8}\mathrm{g/cm}^3$. The Rosseland mean interstellar opacity tables used in these simulations include the dust opacities produced by Pollack, McKay, & Christofferson (1985), as well as the gas opacities produced by Alexander (1975).

Opacity regime	κ_0 [cm^2/g]	a	b	Minimum temperature [K]	Maximum temperature [K]
Ice	2×10^{-4}	0	2	0	1.6681×10^2
Ice sublimes	2×10^{16}	0	-7	1.6681×10^2	2.02677×10^2
Dust grains	1×10^{-1}	0	1/2	2.02677×10^2	$2.28677 \times 10^3 \rho^{2/49}$
Dust grains sublime	2×10^{81}	1	-24	$2.28677 \times 10^3 \rho^{2/49}$	$2.02976 \times 10^3 \rho^{1/81}$
Molecules	1×10^{-8}	2/3	3	$2.02976 \times 10^3 \rho^{1/81}$	$1 \times 10^4 \rho^{1/21}$
Hydrogen scattering	1×10^{-36}	1/3	10	$1 \times 10^4 \rho^{1/21}$	$3.11952 \times 10^4 \rho^{4/75}$
Bound-free & free-free scattering	1.5×10^{20}	1	-5/2	$3.11952 \times 10^4 \rho^{4/75}$	$1.79393 \times 10^8 \rho^{2/5}$
Electron scattering	3.48×10^{-1}	0	0	$1.79393 \times 10^8 \rho^{2/5}$	-

Table 2.1: Opacity scaling with density and temperature in each of the dominant regimes. The opacity is given by $\kappa = \kappa_0 \rho^a T^b$. The temperature range in which each regime applies is also given. The scaling with temperature is illustrated in Figure 2.3.

of particles (reproduced from Cossins et al. (2010); original data from Bell & Lin (1994)). The initial increase in opacity with temperature ($T \lesssim 100$ K) is a regime dominated by ice grains. The sharp decrease in opacity at ≈ 100 K is due to the sublimation of ice grains. Following this, the opacity becomes dust dominated and increases again with temperature until the temperature becomes so large ($\approx$ 1000 K) that the dust grains sublime at what is termed the *opacity gap*. Following the dust sublimation, the dominant contribution to the opacity is due to the gas and once again the opacity begins to increase. When simulating discs, the opacity plays an important role in whether the discs cool or not as it essentially determines how quickly energy can leave the disc and as explained in Chapter 1, the cooling in a disc is important for fragmentation. Therefore, sudden drops in the opacity play an important role, and in particular, the opacity gap is where major changes to fragmentation results may take place. Johnson & Gammie (2003) and Cossins et al. (2010) discuss the importance of the energetics when the temperature reaches that associated with the opacity gap. However, in the simulations presented here, the temperature of the discs are too low for this to be considered.

For the temperature range being considered in the discs simulated in this thesis ($\lesssim 400$ K), the key contribution to the opacity is from the dust. Semenov et al. (2003) compared the opacity tables produced by a number of authors and showed that for the temperature range being investigated here, the Rosseland mean opacity tables of Pollack et al. (1985), Bell & Lin (1994) and Semenov et al. (2003) agree very well.

The opacity values correspond to density and temperatures of up to $1\mathrm{g/cm}^3$ and 10000K, respectively. The exact values used in the simulations are obtained by interpolating between values in logarithmic space. A summary of how the opacity tables were used is available in Whitehouse & Bate (2006). For non-interstellar opacity simulations, I scale these values by the required factor (see Chapters 3 and 4 for simulation details). Reducing the opacity may be equivalent to decreasing the metallicity or increasing the grain sizes which would only have an effect on the dust opacity and not on the gas opacity. Therefore, the scaling is only carried out on the dust opacity values (the details of how the opacity is reduced can be found in Section 2.4 of Ayliffe & Bate (2009)). The models assume that the presence of dust only affects the value of the opacity in the disc and do not include the effects of dust on the disc dynamics. In reality, dust dynamics are important: since dust does not feel a pressure force, it moves with a Keplerian velocity, while gas moves at a sub-Keplerian velocity since it feels the pressure force. A reduction in opacity may also represent increased grain sizes. As grains grow to large sizes, they start to decouple from the gas. Rice et al. (2004) show that ≈ 100 cm sized particles in self-gravitating discs that are only loosely coupled to the gas may concentrate into the regions of pressure maxima i.e. in the spiral structures, thus potentially promoting the growth of planetesimals by core accretion. They discuss that for objects with sizes between 1-100cm, the maximum radial drift velocity may be sufficiently large (Weidenschilling 1977) so that the dust dynamics may be important. However, in the simulations presented in this thesis, the grain sizes in the discs may only be as large as millimetre or centimetre-sized and so the dynamical effects of the dust may not be too critical. Such smaller particles are likely to be coupled to the gas and therefore the presence of the particles is not likely to affect the dynamics. Therefore this assumption is valid.

2.2.4 Equation of state

I carry out investigations using a parameterised cooling method as well as using more detailed thermodynamics by way of radiative transfer. For both calculations, I use the ideal gas equation of state

$$P = \frac{R_{\rm g}}{\mu}\rho T, \tag{2.57}$$

where $R_{\rm g}$ is the gas constant, μ is the mean molecular weight and ρ and T are the density and temperature of the gas, respectively. For the parameterised cooling method, the internal energy is given by

$$u = \frac{R_{\rm g}T}{\mu(\gamma - 1)} = \frac{3}{2}\frac{R_{\rm g}T}{\mu}, \tag{2.58}$$

where $\gamma = 5/3$ and μ, the mean molecular weight, ≈ 2.38. For the simulations using radiative transfer, the equation of state assumes that the gas is composed of hydrogen (70%) and helium (28%) and includes the rotational and vibrational modes of molecular hydrogen, dissociation of molecular hydrogen and ionisation of both hydrogen and helium, as done so by Whitehouse & Bate (2006), though the equation of state has been corrected following Boley et al. (2007). It omits the contribution due to metals and assumes an ortho-/para-hydrogen ratio of 3:1. The discs are cold enough ($\lesssim$ 400 K) such that the rotational and vibrational degrees of freedom are not excited so that effectively, $\gamma = 5/3$ (which is required only to convert between the specific internal energy, u, and the sound speed, $c_{\rm s}$). The models assume that the presence of dust does not affect the equation of state. For simulations which change the opacity due to increased grain sizes, the equation of state is unaffected.

For those simulations which use interstellar opacity values, only 2% of the gas is composed of metals and so the effect due to the metal content on the equation of state is small. For those simulations that use a reduced opacity (and therefore a reduction in metallicity) the metal content would be even lower and would not have a significant effect on the equation of state and therefore this assumption is justified. However, for those simulations which increase the metallicity by way of increasing the opacity (by a factor of 10), the metallicity content will increase such that it has a significant effect on the relative amounts of non-metal to metal content in the disc. This increase in metallicity would affect the specific heat capacity, which in turn would affect the pressure forces and hence the disc stability. However, the details of this are dependent on the specific metal content and a detailed investigation into this is beyond the scope of this thesis.

2.2.5 Boundaries

Previous work involving simulations of self-gravitating discs have discussed the concept of boundaries at great length. It is thought that, particularly in radiative transfer simulations, the boundaries may play a key role in the fragmentation results. Cai et al. (2010) carry out a radiative transfer code comparative study by simulating the same disc that Boss (2007) simulated with the same initial conditions. Despite attempting to remove all possible differences between the two codes,

Cai et al. (2010) find that fragmentation inside tens of AU is not possible while Boss (2007) does find that fragmentation occurs. The suggestion for the lack of convergence between the two codes was attributed to different treatments of the optically thin boundaries.

Radial boundaries

The radial boundaries are modelled in exactly the same way in all the simulations carried out in this thesis. As mentioned in Section 2.1.7, particles are accreted onto the star if they move within a radius $R_{\rm acc}$ of the star or if they move within $R_{\rm acc} \leq R < R_{\rm in}$ and are gravitationally bound to the star. The detailed values of $R_{\rm acc}$ and $R_{\rm in}$ are described in each chapter. For simulations involving the formation of sink particles part way through the simulation, the boundaries between the disc and the new sink particles are treated in exactly the same way as the boundary between the disc and the central star. At the outer edge, the disc is free to expand.

Vertical boundaries and stellar irradiation

In all the simulations, the particles are free to move vertically away from the disc (though the gravitational effects of the star and disc will keep most of it close to the disc). In the simulations carried out without radiative transfer, there is no need for a vertical boundary to deal with the energetics as the heating and cooling in the discs is purely due to internal processes and does not rely on an external source or sink of energy into or out of the disc.

For the radiative transfer calculations, a flux-limited diffusion approximation is used to simulate the transfer of energy between particles. I use a two-layer approach to simulate the midplane and surface regions. To model the energy loss from the disc surface, a boundary layer of particles maintains a fixed temperature profile such that any energy that is passed to these boundary particles is effectively radiated away. Physically, I assume that the atmosphere temperature is not set by the disc itself but by some other process such as stellar or external irradiation. The particles forming the disc boundary are variable since some particles may move across the boundary from the optically thin to the optically thick region and vice versa.

In Chapter 3, the vertical location of the boundary between the optically thick part of the disc and the optically thin atmosphere is located at the maximum of: the height above the midplane where the optical depth, $\tau = 1$, or the height above the midplane where $z_{\rm b,H} = 1.75H$, whichever is greater, i.e.

$$z_{\rm b} = \max\left(1.75H,\ \sqrt{2}H\left|{\rm erf}^{-1}\left[1 - \frac{1}{\kappa\Sigma}\right]\right|\right), \tag{2.59}$$

where ${\rm erf}^{-1}$ is the inverse error function. Appendix A shows the detailed derivation of the $\tau = 1$ boundary height, $z_{{\rm b},\tau=1}$:

$$z_{{\rm b},\tau=1} = \sqrt{2}H\left|{\rm erf}^{-1}\left[1 - \frac{1}{\kappa\Sigma}\right]\right|. \tag{2.60}$$

The reason for imposing the restriction that the boundary height is the maximum out of the height at which $\tau = 1$ or $1.75H$ is because in the outer radial regions of the discs, the entire

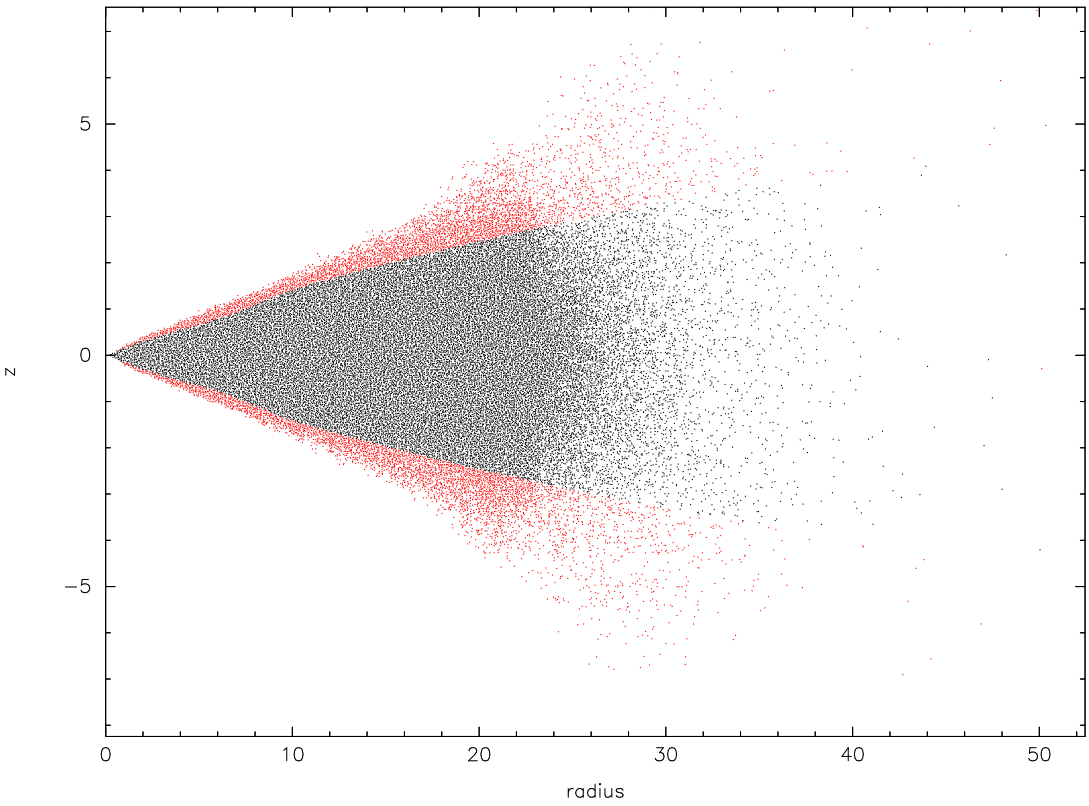

Figure 2.4: Plot showing the vertical height against radius of a simulated disc. The red particles are the boundary particles while the black particles are those that make up the bulk of the disc. The boundary particles have been determined by comparing their vertical height to equation 2.59.

vertical extent of the disc may be optically thin. Since I am investigating the evolution of the disc as a result of the radiative transfer processes in the disc, it is not reasonable to set the temperature of the entire vertical extent of the disc, even if it is optically thin. Nevertheless, in these regions I expect that the disc will be nearly vertically isothermal and therefore, the temperature that is set at the boundary would be the temperature that the disc adopts. The choice of $1.75H$ is made to ensure that a reasonable number of particles exist in the boundary region at all times during the simulation (between $\approx 7 - 15\%$) so that there are an adequate number of particles to simulate the effects of energy transfer from the disc to the external medium but ensuring that too many particles are not present so that the entire evolution of the disc is determined by the temperature profile of the boundary particles.

The calculation of the boundary height using equation 2.59 ensures that the boundary is at the vertical position where the optical depth, $\tau \lesssim 1$, and the choice ensures that the "bulk" of the disc (i.e. consisting of particles that lie closer to the disc midplane) is simulated using radiative transfer, rather than simplified energetic calculations. The remaining "boundary" particles have their temperature held at a constant temperature profile. Figure 2.4 gives an example of a disc simulated in this way and shows the vertical height against radius of all the particles used in this particular disc setup. The boundary particles are highlighted in red and the particles making up the bulk of the disc are highlighted in black.

I have carried out tests to see what the effects of the exact location of the vertical boundary

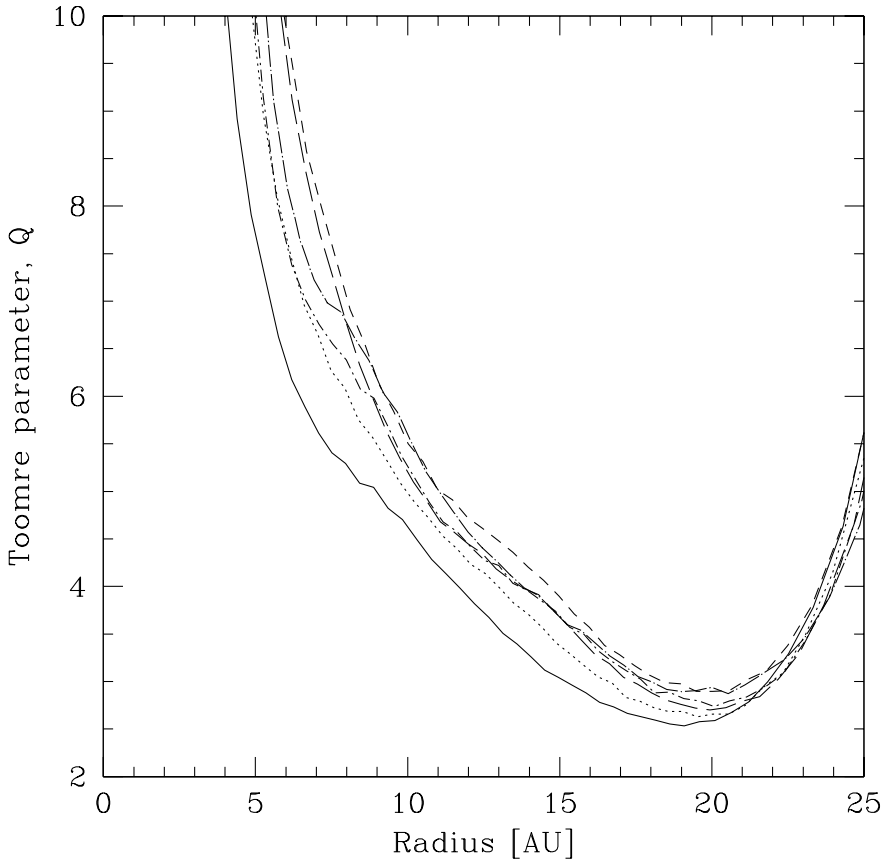

Figure 2.5: Toomre stability profiles of a $0.1M_{\odot}$, 25 AU disc (around a $1M_{\odot}$ star) evolved with the boundary height scaled to 0.5× (solid line), 0.75× (dotted line), 1.0× (short-dashed line), 1.1× (long-dashed line), 1.25× (dot-short dashed line and 1.5× (dot-long dashed line) the location of the boundary height calculated using equation 2.59. Moving the boundary height higher up into the optically thin region does not affect the results significantly, but moving it down into the optically thick region causes the disc temperature to become artificially low.

has on the disc energetics. I scaled the calculated boundary height by factors of 0.5, 0.75, 1.0, 1.1, 1.25 and 1.5. Figure 2.5 shows the resulting Toomre stability profiles of each of these simulations and shows that decreasing the boundary height into the optically thick region causes the Toomre stability profile to decrease as it causes the overall disc temperature to decrease. However, increasing the boundary height does not affect the stability profiles significantly. Therefore, I have found that provided the boundary is located in the optically thin surface region of the disc and provided that a layer of boundary particles exists over the entire disc surface (such that the energy can be radiated away from the surface), the disc energetics remain unchanged.

The boundary calculation in Chapter 3 calculates the boundary at the start of the simulation and does not re-evaluate the boundary as the simulation progresses. This is a reasonable assumption to make since the boundary is high up in the disc and does not play a part in the fragmentation of the disc since in these regions the density is low and fragmentation typically occurs in the midplane. Nevertheless, the boundary calculation in Chapter 4 uses an improved method so that the $\tau = 1$ boundary is periodically re-evaluated during the simulation. This is carried out after each largest particle timestep which typically takes place at least 40 times each outer rotation period. In Appendix A, I show that to calculate the boundary height, the opacity and surface mass density need to be evaluated. In cases where the boundary height is calculated after the start of the simulation, these values will change and so initial values cannot be used. Appendix B shows how the $\tau = 1$ boundary is calculated part way through a simulation.

Chapter 3

Opacity effects on fragmentation

3.1 Introduction

In this chapter, I model the evolution of massive self-gravitating discs using a global three-dimensional SPH code including radiative transfer and the effects of stellar irradiation. In particular, I explore the parameter space in terms of dust opacity, disc temperature and size in order to scope out if, and under what conditions, a self-gravitating disc may fragment. In Section 3.2, I describe the code used. In Section 3.3, I outline my simulations, including the disc setup and discuss the parameter space. I present my results and compare with previous studies in Sections 3.4 and 3.5, respectively.

3.2 Numerical setup

My simulations are carried out using an SPH code originally developed by Benz (1990) and further developed by Bate, Bonnell, & Price (1995), Whitehouse, Bate, & Monaghan (2005), Whitehouse & Bate (2006) and Price & Bate (2007). It is a Lagrangian hydrodynamics code, ideal for simulations that require a large range of densities to be followed, such as fragmentation scenarios. The version used to carry out the simulations presented here includes radiative transfer using the flux-limited diffusion approximation (Whitehouse et al. 2005; Whitehouse & Bate 2006) with two temperatures: that of the gas and that of the radiation field.

In order to model shocks, SPH requires artificial viscosity. I use a common form of artificial viscosity by Monaghan & Gingold (1983), which uses the parameters $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$. A corollary of including artificial viscosity is that it adds shear viscosity and causes dissipation. If this viscosity is too large, the evolution of the disc may be driven artificially, while if it is too small, it will lead to inaccurate modelling of shocks (Bate 1995). I have tested various values of the SPH parameters and find that a value of $\alpha_{\rm SPH} \sim 0.1$ provides a good compromise between these factors. Since typically, $\beta_{\rm SPH} \sim 2\alpha_{\rm SPH}$, I choose $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$ to be 0.1 and 0.2, respectively. These are the same values as those implemented by Lodato & Rice (2004). However, I have chosen to fix the artificial viscosity parameters whereas Lodato & Rice (2004) used the Balsara (1995) switch which reduces the artificial viscosity in places where the ratio of the divergence to the curl of the velocity field is small i.e. in regions of strong vorticity. The reasoning behind my choice is two-fold: (i) the Balsara switch reduces the viscosity in shearing flows but given that I have both

shocks and shearing flows, the reduction does not handle the shocks well; (ii) I want to ensure a controlled test which is not possible if the numerical viscosity is constantly changing.

3.2.1 Opacity & equation of state

The discs simulated are assumed to be in local thermal equilibrium and the opacities are assumed to be grey Rosseland mean values. The opacities are based on the opacity tables of interstellar molecular dust grains produced by Pollack, McKay, & Christofferson (1985), and on Alexander (1975) for the higher temperature gaseous contributions. A summary of how the opacity tables were used is available in Whitehouse & Bate (2006). For non-interstellar opacity simulations, I scale these values by the required factor (see Table 3.1 for simulation details). The details of how this was done can be found in Section 2.4 of Ayliffe & Bate (2009).

The equation of state used in these simulations assumes that the gas is composed of hydrogen (70%) and helium (28%) and includes dissociation of molecular hydrogen, ionisation of both hydrogen and helium, and the rotational and vibrational modes of molecular hydrogen, as done so by Whitehouse & Bate (2006), though the equation of state has been corrected following Boley et al. (2007). It omits the contribution due to metals. My discs are cold enough such that the rotational and vibrational degrees of freedom are not excited so that effectively, $\gamma = 5/3$. My models assume that the presence of dust only affects the value of the opacity in the disc. I do not include the effects of dust on the disc dynamics, nor do I consider its effects on the equation of state.

3.2.2 Stellar irradiation

As mentioned earlier, radiative transfer is simulated using the flux-limited diffusion approximation. I use a two-layer approach to simulate the midplane and surface regions. In order to model the energy loss from the disc surface, a boundary layer of particles maintains a fixed temperature profile such that any energy that is passed to these boundary particles is effectively radiated away. The vertical location of the boundary between the optically thick part of the disc and the optically thin atmosphere is located at the maximum of: the height above the midplane where the optical depth, $\tau = 1$, or the height above the midplane where $z_{\rm b} = 1.75H$, where H is the isothermal scale height given by $H = c_{\rm s}/\Omega$. Therefore, the boundary is at the vertical position where the optical depth, $\tau \lesssim 1$, and my choice ensures that the "bulk" of the disc (i.e. consisting of particles that lie closer to the disc midplane) is simulated using radiative transfer, rather than simplified energetic calculations. The remaining "boundary" particles have their temperature held at a constant temperature profile. Physically, I assume that the atmosphere temperature is not set by the disc itself but by some other process such as stellar or external irradiation. The particles forming the disc boundary are variable since some particles may move across the boundary from the optically thin to the optically thick region and vice versa. I have carried out tests to see what the effects of the exact location of the vertical boundary has on the disc energetics and have found that provided the boundary is located in the optically thin surface region of the disc and provided that a layer of boundary particles exists over the entire disc surface, the disc energetics remain unchanged.

Simulation name	Disc radius [AU]	Opacity scaling factor	$Q_{\rm min}$	Fragment?	Fragmentation time
Reference	25	1	2	n	-
Kappa10	25	10	2	n	-
Kappa0.1	25	0.1	2	n	-
Kappa0.01	25	0.01	2	n	-
Qmin1	25	1	1	n	-
Qmin0.75	25	1	0.75	n	-
Qmin0.75-Kappa0.1	25	0.1	0.75	n	-
Qmin0.75-Kappa0.01	25	0.01	0.75	y	9.7 ORPs
Qmin0.5	25	1	0.5	y	1.8 ORPs
L-Qmin1	300	1	1	n	-
L-Qmin1-Kappa10	300	10	1	n	-
L-Qmin1-Kappa0.1	300	0.1	1	y	1.5 ORPs
L-Qmin0.75	300	1	0.75	y	< 1 ORPs

Table 3.1: Summary of the simulations carried out. The opacity scalings refer to multiples of interstellar Rosseland mean opacity values as described in Section 3.2. $Q_{\rm min}$ refers to the minimum value of the Toomre parameter (at the outer edge of the disc) at the start of the simulation.

3.3 Simulations

Table 3.1 summarises the parameters and fragmentation results of the simulations presented here. Each simulation was run either beyond the point at which the disc attained a steady state (for > 6 outer rotation periods, ORPs), or until it fragmented (defined as regions which are at least three orders of magnitude denser than their surroundings). The magnitude and profile of the boundary temperatures are the same as those of the initial discs so that the discs start in thermal equilibrium with their vertical boundaries (atmospheres). I expect the discs to heat up initially due to work done and viscous heating. Following an initial transient phase, the bulk of the disc may or may not cool to re-establish thermal equilibrium with the boundary. I discuss a Reference case first before turning my attention to exploring the parameter space.

3.3.1 Reference case

The reference disc is set up in exactly the same way as a disc simulated by Lodato & Rice (2004): a 1 $M_\odot$ star with a 0.1 $M_\odot$ disc made of 250,000 SPH particles, spanning $0.25 \leq R \leq 25$AU. The initial surface mass density and temperature profiles of the disc are $\Sigma \propto R^{-1}$ and $T \propto R^{-\frac{1}{2}}$, respectively. The magnitudes of these are set such that the Toomre stability parameter (equation 1.7) at the outer edge of the disc, $Q_{\rm min} = 2$. This gives an aspect ratio, $H/R \sim 0.05$. I model the 1 $M_\odot$ star in the centre of the disc using a sink particle (Bate et al. 1995). At the inner disc boundary, particles are accreted onto the star if they move within a radius of 0.025 AU of the star or if they move into $0.025 \leq R < 0.25$AU and are gravitationally bound to the star. At the outer edge, the disc is free to expand.

The initial Reference disc is a Toomre stable disc. Given that the boundary temperature profile is the same as that of the initial disc and that $Q_{\rm min} = 2$, I do not expect this disc to fragment. However, I use the Reference disc as a fiducial case. In particular, I am concerned with the cooling rates that are present in the disc once it is in an equilibrium state. I emphasize my use of terminology here: when referring to the disc being in *thermal equilibrium with the boundary*,

I am describing the bulk of the disc being a similar temperature to the disc boundary (which is assumed to be determined by stellar irradiation), whereas an *equilibrium state* disc refers to the dissipative and cooling rates being balanced such that the Toomre stability profiles do not change with time.

3.3.2 Exploring the parameter space

Given that a motivation of this work is to determine if, and under what circumstances, fragmentation in realistically modelled self-gravitating discs may occur, I explore the parameter space in a number of ways. One parameter is the opacity: I re-run the Reference simulation with opacity values scaled to 10×, 0.1× and 0.01× the interstellar opacity values (simulations Kappa10, Kappa0.1 and Kappa0.01, respectively). This may be equivalent to a disc with differing metallicities or grain sizes. I make two assumptions here: i) the change in metallicity does not affect the equation of state (as described in Section 3.2.1) and ii) there are no spatial or temporal variations in the grain size distributions. As with the Reference case, these discs are simulated purely to analyse the energetics since I do not expect these discs to fragment.

I then choose to explore the initial and boundary temperature conditions by decreasing the magnitude of the disc temperature whilst maintaining the same surface mass density as the Reference case. I do this by changing the initial Toomre stability parameter profiles such that $Q_{\rm min} = 1, 0.75$ and 0.5 (simulations Qmin1, Qmin0.75 and Qmin0.5, respectively). This is equivalent to reducing the disc aspect ratios to $H/R \sim 2.2\times10^{-2}$, 1.7×10^{-2} and 1.1×10^{-2}, respectively. I reiterate that the boundary temperature is the same as the temperature of the initial disc and hence this setup not only changes the disc temperature profile, but it also changes the boundary temperature profile.

Furthermore, I consider a combination of the above factors by simulating discs with $Q_{\rm min} = 0.75$ and opacities that are 0.1× and 0.01× the interstellar opacity values.

The unfavourable conditions for fragmentation at small radii have been discussed at great length in the past (e.g. Rafikov 2005; Stamatellos & Whitworth 2008; Boley 2009; Rafikov 2009; Clarke 2009). I therefore expand my parameter space to include discs that are a factor of 12 larger with a radii range of $3 \leq R \leq 300$AU. These discs have the same mass as the 25AU discs and are set up so that $Q_{\rm min} = 1$. I simulate three different opacity values (1×, 10× and 0.1× the interstellar Rosseland mean opacities). In addition, I also simulate a large disc with $Q_{\rm min} = 0.75$ with interstellar opacity values.

In order to keep these disc masses and initial Toomre stability profiles the same as the smaller 25AU discs, I require both the surface mass density and absolute temperature to be reduced. These discs are therefore not only larger, but also colder than their equivalent (in terms of initial Toomre stability profiles) small discs.

3.4 Results

The simulations have been analysed in three main ways:

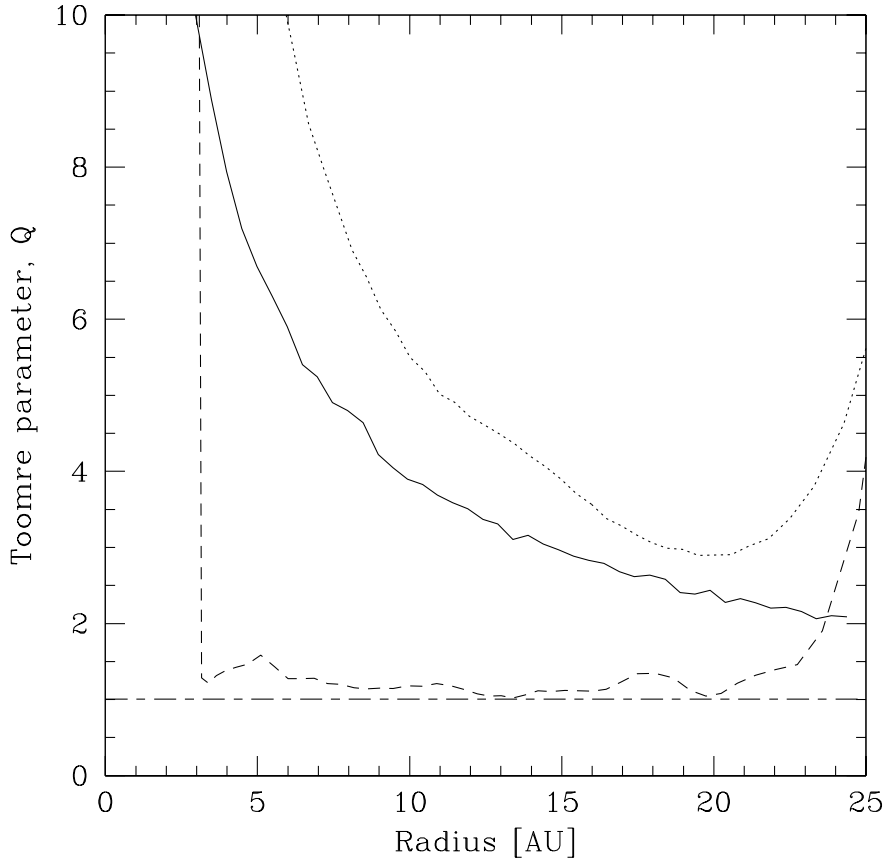

Figure 3.1: Azimuthally averaged values of the Toomre parameter at the start (solid line) and at a time t = 6.4 ORPs (dotted line) for the Reference simulation. The disc is unable to cool rapidly due to the internal heating and hence its end state is more stable than the initial disc. Also shown is the equivalent disc simulated by Lodato & Rice (2004, short dashed line) which cools using simplified cooling (with $\beta = 7.5$) rather than by radiative cooling, and also does not consider the effects of stellar irradiation. The critical value of $Q_{\rm crit} = 1$ is also marked.

(i) I compare the azimuthally averaged Toomre stability profiles of the initial and final (or in the case of fragmenting discs, shortly before fragmentation) discs which indicates whether the bulk of the discs were able to reach a state of thermal equilibrium with their boundary. The surface mass density does not change significantly throughout the simulations and hence changes in the Toomre stability parameter are due to changes in the disc temperature. This enables us to determine which discs are more likely to fragment. Note that I assume $\kappa_{\rm ep} = \Omega$ in equation 1.7.

(ii) I examine the timescale on which the discs cool (by considering the energy passed from the gas to the radiation within the disc as well as that which is assumed to be instantly radiated away from the disc surface by the boundary particles). In past simulations that have neglected the heating effects of stellar irradiation (e.g. Gammie 2001; Rice et al. 2005), the cooling, C, in a steady-state disc balances the heating due to gravitational stresses, $H_{\rm GI}$, and the heating due to artificial viscosity, H_ν, such that

$$C = H_{\rm GI} + H_\nu. \tag{3.1}$$

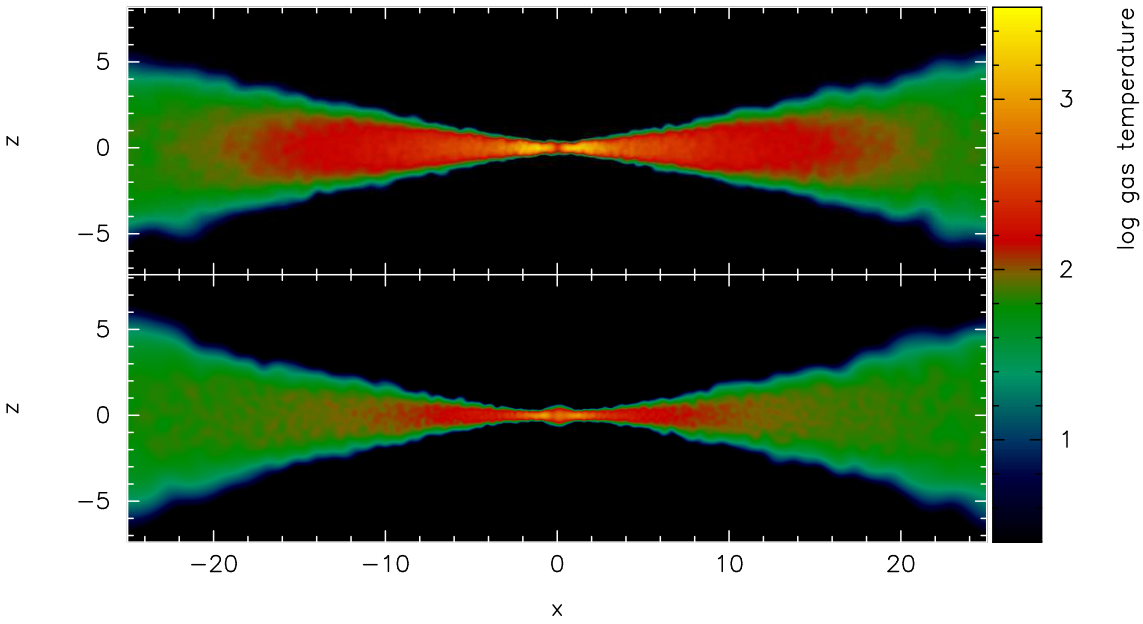

Figure 3.2: Logarithm of the gas temperature (in K) rendered in cross-sectional views of the Reference (top panel) and Kappa0.01 (bottom panel) discs at a time $t = 6.4$ ORPs. The surface of the Reference disc is clearly colder than the midplane whilst the midplane of the Kappa0.01 disc is closer to a state of thermal equilibrium with the boundary. Axis units are in AU.

If the artificial viscosity is low, $C \approx H_{\rm GI}$. In this case, the cooling timescale in units of the orbital timescale, β, can be related to the gravitational stress in the disc, $\alpha_{\rm GI}$ (Gammie 2001, Section 1.4.2):

$$\alpha_{GI} = \frac{4}{9}\frac{1}{\gamma(\gamma-1)}\frac{1}{\beta}. \tag{3.2}$$

Gammie (2001) and Rice et al. (2005) have shown that the maximum gravitational stress that a disc can support is $\alpha_{\rm GI} = 0.06$, beyond which fragmentation will occur. In discs that do not take into account heating due to external irradiation, this condition is equivalent to requiring the cooling timescale in terms of the orbital timescale, β, to be smaller than the critical values, described in Section 1.4.2, for fragmentation.

In my steady-state discs, not only does the cooling have to balance the heating due to the gravitational instabilities and the numerical viscosity, but it also has to balance the heating due to stellar irradiation, $H_{\rm SI}$, such that:

$$C = H_{\rm GI} + H_{\nu} + H_{\rm SI}. \tag{3.3}$$

In what follows, I calculate the parameter, ψ, which I define to be the timescale on which the disc cools in units of the orbital timescale. Without irradiation, $\psi = \beta$. When including heating due to stellar irradiation, the ψ parameter does not specifically tell us about the fraction of the cooling that balances the gravitational instabilities and therefore cannot be used directly to decide whether a disc should fragment or not. However, it *does* still give an indication as to what the cooling rate is in the discs which has been shown to be important when determining whether a disc is likely to

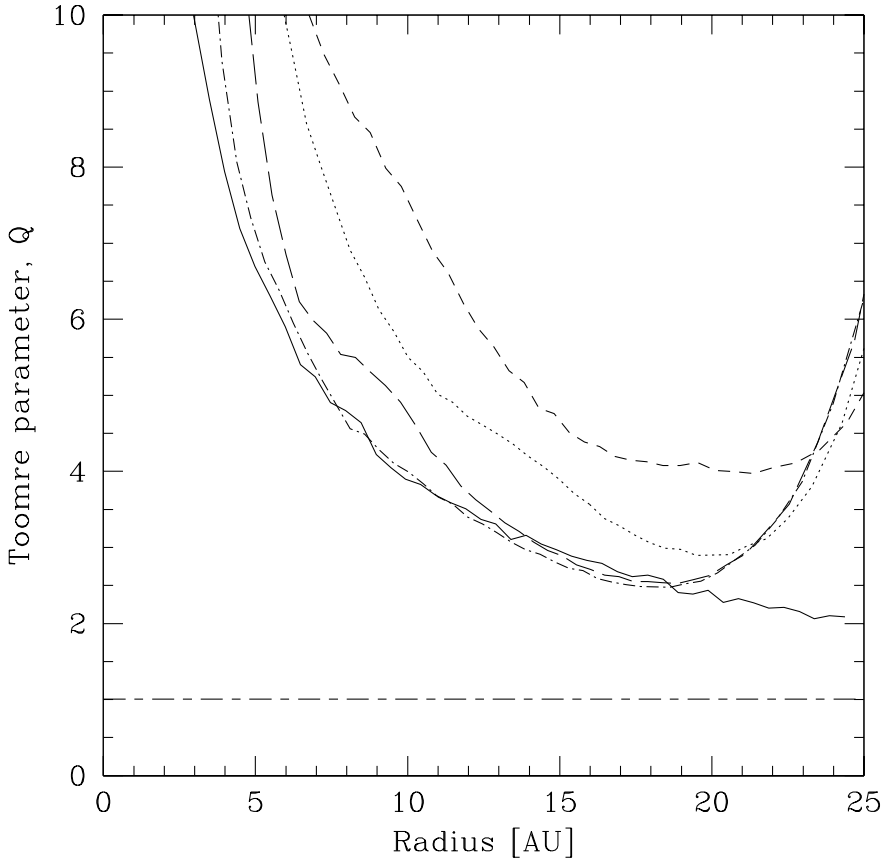

Figure 3.3: Azimuthally averaged values of the Toomre stability parameter at the start (solid line) and at a time t = 6.4 ORPs, for the Reference (dotted line), Kappa10 (short dashed line), Kappa0.1 (long dashed line) and Kappa0.01 (dot-dashed line) simulations. The critical value of $Q_{\rm crit}$ = 1 is also marked. Decreasing the opacity causes the disc to cool more efficiently.

fragment or not (Gammie 2001; Rice et al. 2005).

(iii) I examine images of the discs for signs of fragmentation which I define as clumps in the disc which are at least three orders of magnitude denser than their surroundings.

3.4.1 Reference case

Figure 3.1 shows the Toomre stability profile of the initial and final Reference disc. The disc is not able to cool rapidly enough in response to the internal heating and consequently the disc midplane becomes hotter than the boundary (top panel of Figure 3.2). Figure 3.1 also shows the final Toomre stability profile for the equivalent disc simulated by Lodato & Rice (2004) which used simplified cooling rather than radiative cooling. I have included this comparison to emphasize how significant the differences can be between discs simulated using simplified cooling parameters and discs modelled not only with more detailed radiative cooling but also incorporating the effects of stellar irradiation.

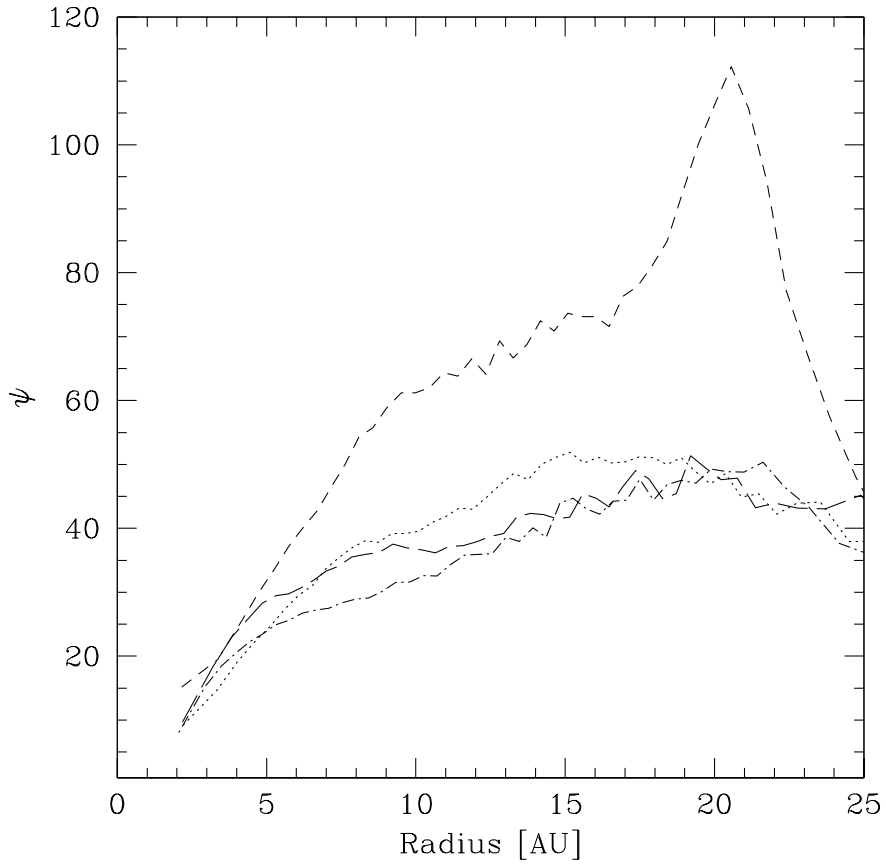

Figure 3.4: Cooling timescale, ψ, profiles for the Reference (dotted line), Kappa10 (short dashed line), Kappa0.1 (long dashed line) and Kappa0.01 (dot-dashed line) simulations at a time $t = 6.4$ ORPs. Decreasing the opacity causes the cooling rate in the disc to increase (and ψ to decrease), but only until the disc can reach thermal equilibrium with its boundary.

3.4.2 Opacity effects

Figure 3.3 shows the effect on the Toomre stability parameter of changing the opacity in the discs to 10×, 0.1× and 0.01× the interstellar Rosseland mean values (simulations Kappa10, Kappa0.1 and Kappa0.01, respectively). I see that decreasing the disc's opacity enhances its ability to reach a state of thermal equilibrium with the boundary since the radiation leaves the disc far more effectively and hence the disc cools faster. This is particularly evident in the cross-sectional plots showing the temperature structure in the vertical direction (Figure 3.2). Figure 3.4 shows the ψ profile of these simulations. This figure shows that a low opacity disc has a greater ability to radiate away the disc's energy. The decrease in the cooling timescale, ψ, does not continue at very low opacities because the stellar irradiation sets the boundary temperature and therefore, the minimum disc temperature.

Figures 3.3 and 3.4 particularly show that the low opacity discs are able to cool fast enough to reach a state of thermal equilibrium with their boundaries. Thus, if the conditions were right (i.e. the boundary temperature was lower so that the Toomre stability parameter was able to reach $Q \lesssim 1$), the low opacity disc may fragment. I therefore turn my attention to the disc absolute

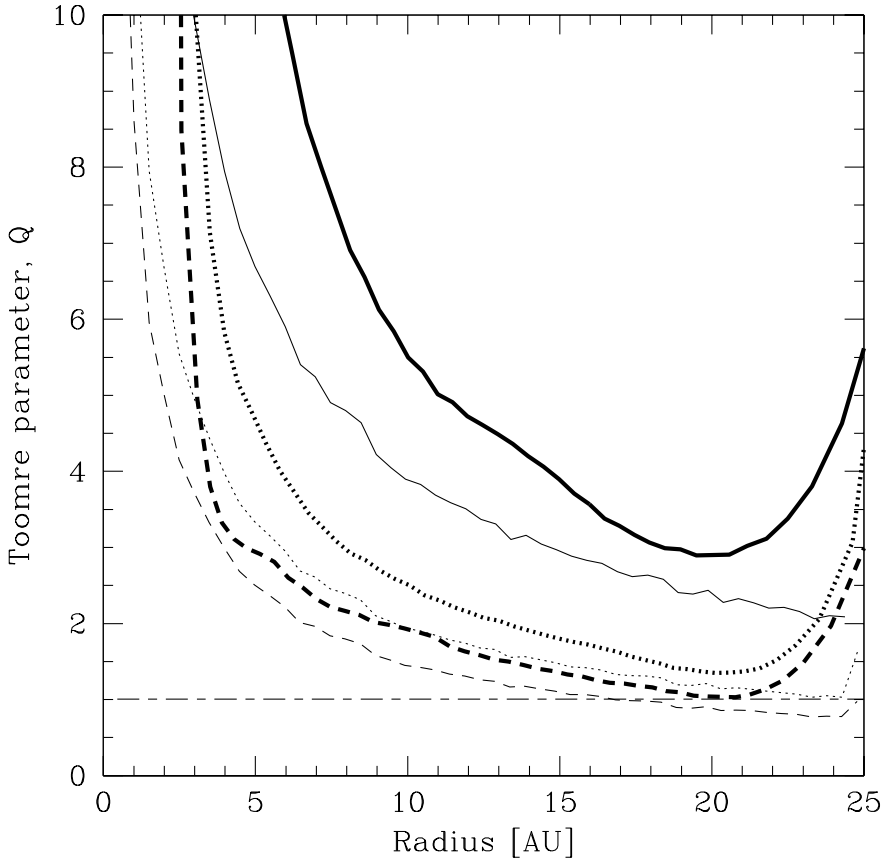

Figure 3.5: Initial (thin lines) and final (at t = 6.4 ORPs; heavy lines) Toomre stability profiles for the Reference (solid line), Qmin1 (dotted line) and Qmin0.75 (dashed line) simulations. The critical value of $Q_{\rm crit}$ = 1 is also marked. None of the three discs which have interstellar opacity values can cool rapidly enough to maintain thermal equilibrium with their boundaries.

temperature.

3.4.3 Colder discs

Figure 3.5 shows the initial and final Toomre stability profiles for the Reference, Qmin1 and Qmin0.75 simulations. The interesting aspect about the Qmin0.75 and Qmin1 cases are that though the initial and boundary conditions are either unstable or marginally stable, the discs still do not fragment because they are unable to cool rapidly and heat up so that they end up being at or above the marginal state (e.g. Figure 3.6). The cooling timescales, ψ, for the discs are as low as $\approx$ 20 (Figure 3.7) suggesting that for the discs to fragment, an efficient energy removing mechanism is needed such that the cooling timescale is lower than the Qmin0.75 (dashed line) curve in Figure 3.7.

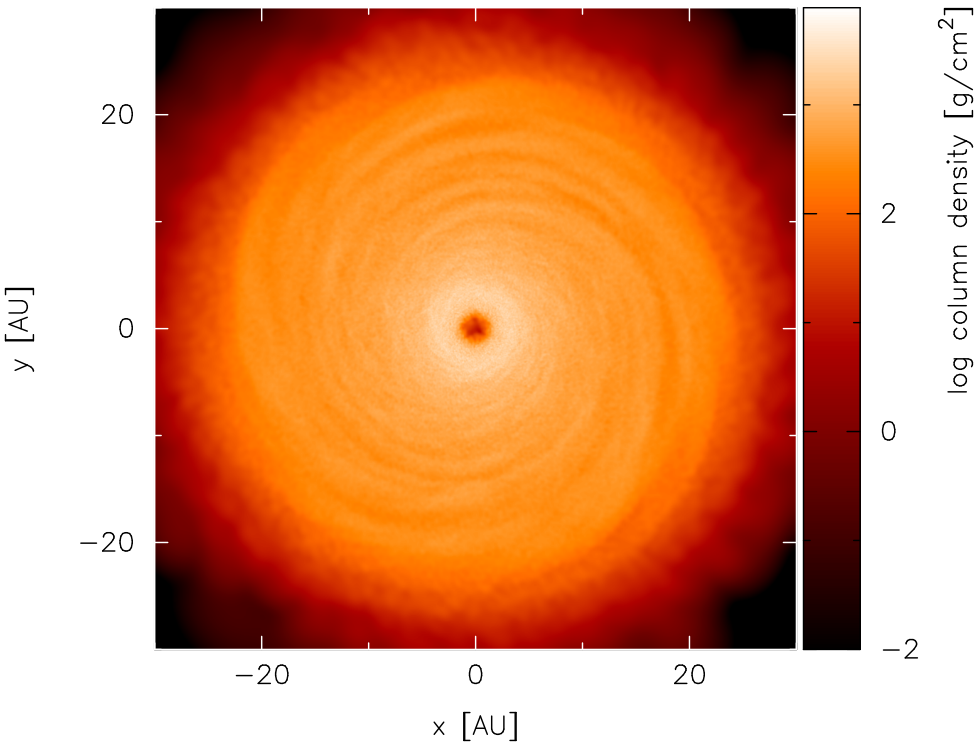

Figure 3.6: Surface density rendered image of the Qmin0.75 disc at a time t = 6.4 ORPs. At the end of the simulation, the disc has not fragmented despite initially being in a critical state because with interstellar opacities, it heats up.

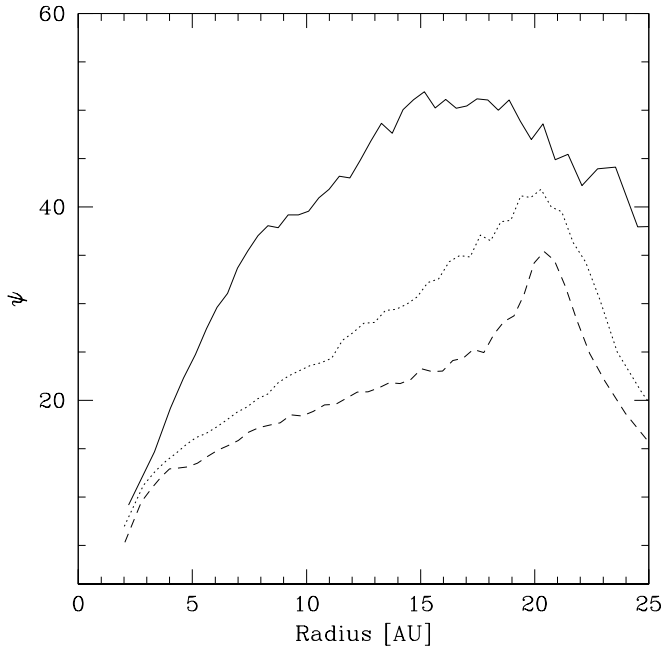

Figure 3.7: Cooling timescale, ψ, profiles for the Reference (solid line), Qmin1 (dotted line) and Qmin0.75 (dashed line) simulations at time t = 6.4 ORPs. With interstellar opacities, these discs are simply not able to cool rapidly enough to obtain a ψ value that is low enough for fragmentation.

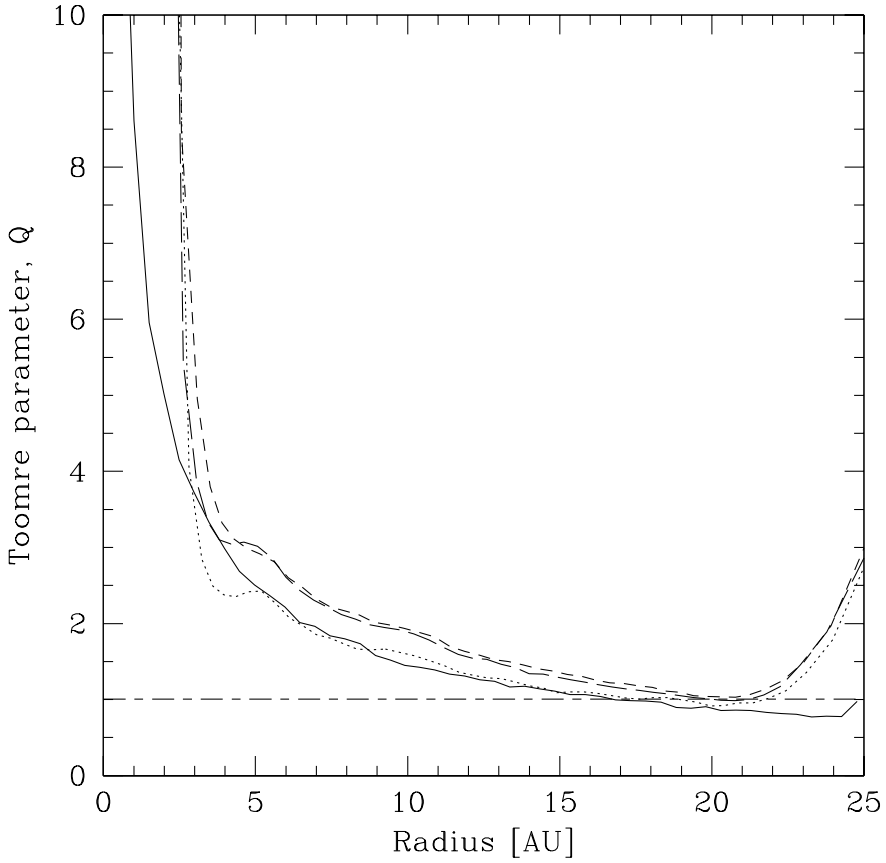

Figure 3.8: Toomre stability profiles at t = 6.4 ORPs for the Qmin0.75 (short dashed line) and Qmin0.75-Kappa0.1 (long dashed line) simulations and at t = 8 ORPs (just before it begins to fragment) for the Qmin0.75-Kappa0.01 (dotted line) simulation in comparison to the initial (solid line) Toomre stability profile for these discs. The critical value of $Q_{\rm crit}$ = 1 is also marked. The lower opacity disc cools rapidly enough to attain a state of thermal equilibrium with its boundary. It eventually fragments at t = 9.7 ORPs

3.4.4 Low temperature, low opacity discs

For completeness, I also simulate a disc with $Q_{\rm min}$ = 0.5 and find that though this disc also heats up, it cannot heat fast enough to become Toomre stable before it fragments (due to its initial conditions).

Until now, I have identified under what conditions discs are *more likely* to fragment (i.e. sub-interstellar opacities and cooler temperatures). Combining these conditions, I simulate two further discs with outer Toomre stability parameters of 0.75 with 0.1× and 0.01× interstellar opacity values (simulations Qmin0.75-Kappa0.1 and Qmin0.75-Kappa0.01, respectively). With such low opacity values, the Qmin0.75-Kappa0.01 disc would be equivalent to a low metallicity disc or a disc with grain sizes ranging between millimetre and centimetre sizes which is realistic given current observations (Calvet et al. 2002; Testi et al. 2003; Rodmann et al. 2006; Lommen et al. 2007).

Figure 3.8 shows that the disc with 0.1× interstellar opacity values is slightly cooler than

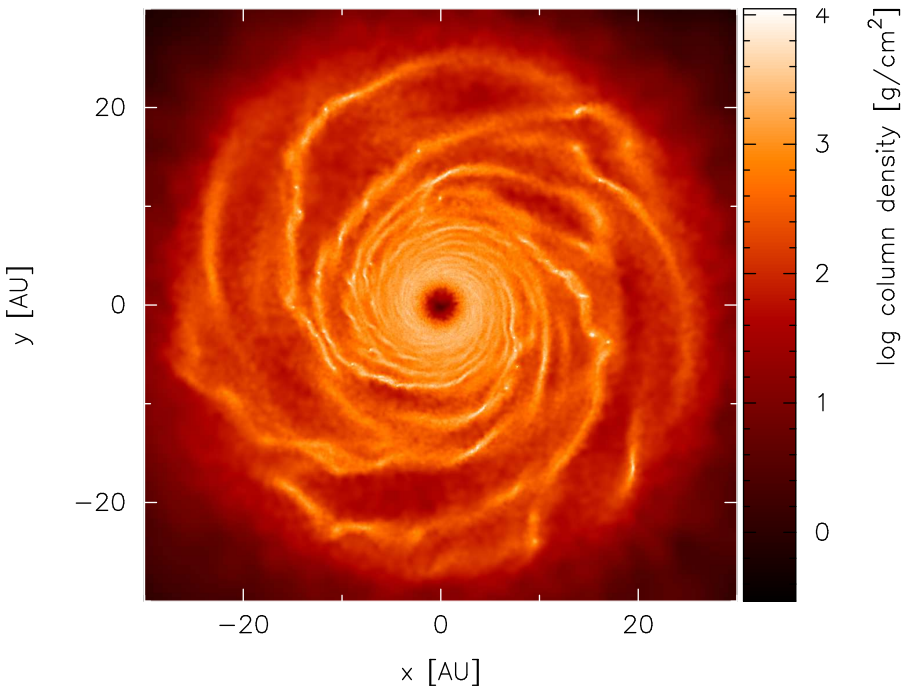

Figure 3.9: Surface density rendered image of the fragmented Qmin0.75-Kappa0.01 disc at a time of $t = 10.5$ ORPs. The disc not only requires reduced irradiation, but also low opacities are essential for it to cool rapidly enough to fragment.

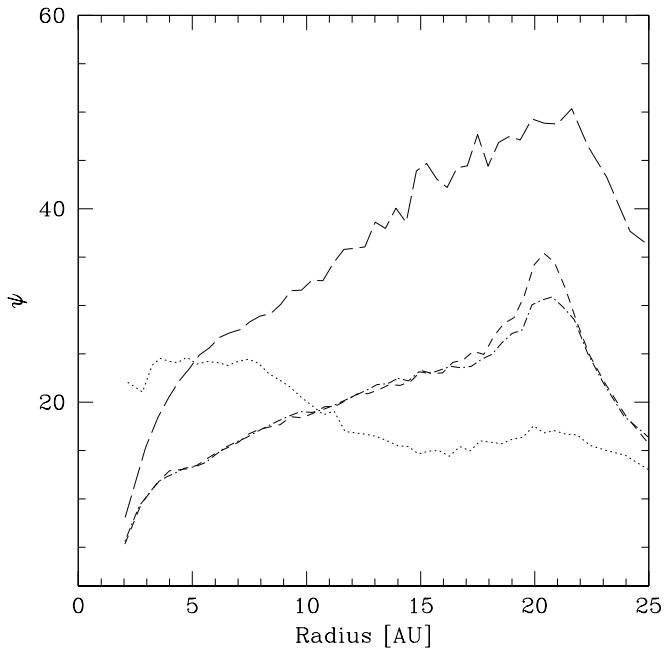

Figure 3.10: Cooling timescale, ψ, profiles for the Qmin0.75-Kappa0.01 simulation (dotted line) at $t = 8$ ORPs, just before it begins to fragment, in comparison to the Qmin0.75 (short dashed line), Kappa0.01 (long dashed line) and Qmin0.75-Kappa0.1 (dot-dashed line) simulations at $t = 6.4$ ORPs. The fragmenting disc has a low ψ value (≈ 15) in the outer part of the disc where it fragments.

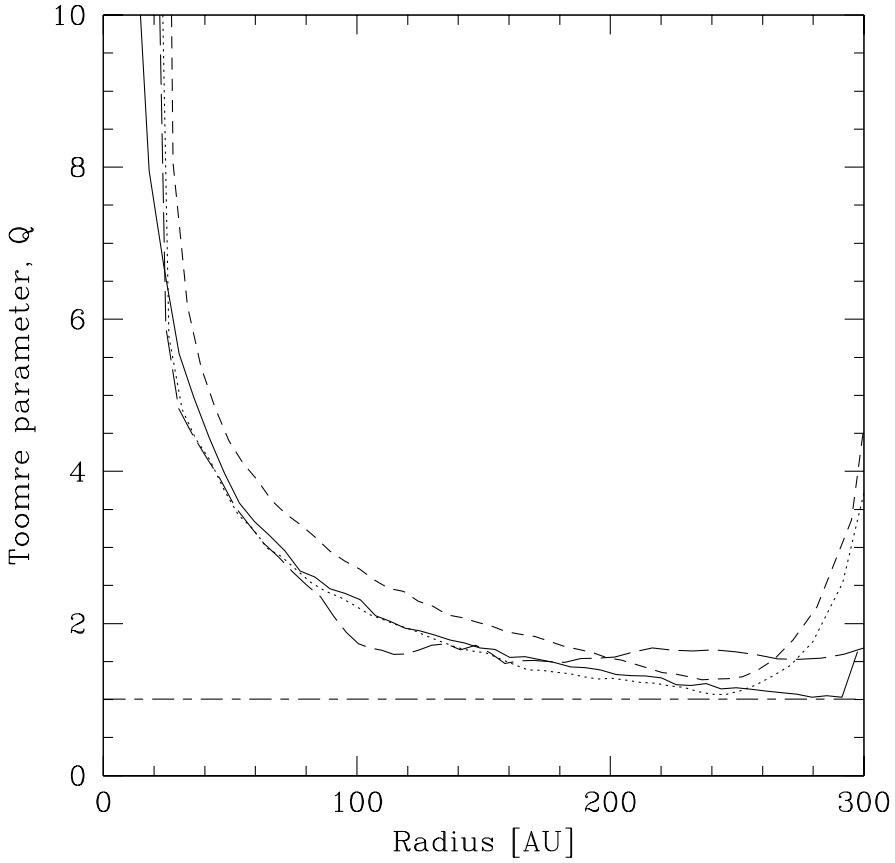

Figure 3.11: Toomre stability profiles for simulations L-Qmin1 (dotted line), L-Qmin1-Kappa10 (short dashed line) and L-Qmin1-Kappa0.1 (long dashed line) at a time of $t = 6.4$ ORPs as well as the boundary profile for these discs (solid line). The critical value of $Q_{\rm crit} = 1$ is also marked. Larger discs do not require as low opacities as smaller discs to attain a state of thermal equilibrium with their boundaries.

the case with interstellar opacities, but is still unable to cool rapidly enough to maintain a state of thermal equilibrium with its boundary. However, the disc with $0.01\times$ interstellar opacity values is able to since its Toomre stability profile at the end of the simulation is close to its initial value. I find that the disc in the latter simulation does indeed fragment (Figure 3.9), though it takes ~ 9.7 ORPs to do so. This is because:

(i) the disc goes through a transient phase where it heats up since the initial disc is not quite in hydrostatic equilibrium.

(ii) although the radiation is able to leave quickly, the disc is enveloped in a *thermal blanket* due to the effects of stellar irradiation, thus causing the disc to cool more slowly.

(iii) as the disc midplane cools, its cooling rate also decreases since the temperature gradient between the disc midplane and surface becomes smaller.

Figure 3.10 shows the cooling timescale, ψ, of simulation Qmin0.75-Kappa0.01 (1.7 ORPs prior to fragmentation) in comparison to simulations Kappa0.01 (which has $Q_{\rm min} = 2$), Qmin0.75 (which uses interstellar opacities) and Qmin0.75-Kappa0.1. With a reduced opacity, the Qmin0.75-Kappa0.01 disc cools on a timescale fast enough such that it fragments i.e. it cools fast enough

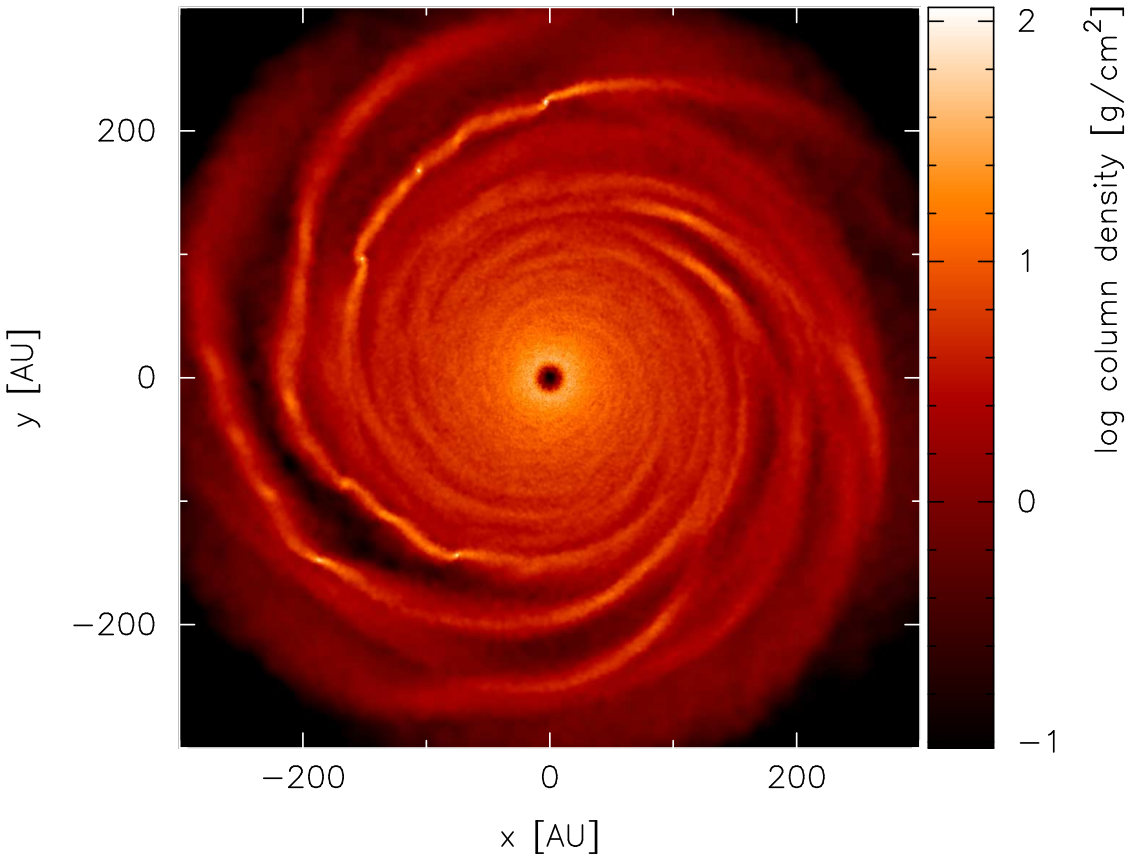

Figure 3.12: Surface density rendered image of the large, low opacity disc, L-Qmin1-Kappa0.1, at a time of t = 2.9 ORPs. A low opacity is also required for a large disc to fragment, though the opacity does not have to be as low as the 25AU discs.

so that the heating due to gravitational instabilities, numerical viscosity and stellar irradiation do not heat the disc significantly above the boundary temperature. This figure shows that the cooling timescale $\psi \approx 15$ in the outer parts of the disc, where fragmentation occurs. I emphasise the need to express caution when considering the absolute value of ψ in a non-steady state fragmenting disc, since it is dependent on the disc temperature which is constantly changing. I discuss this in detail in Section 3.5. A key interesting aspect about this simulation is that the fragments form even though the disc is small ($R < 25$ AU), a result contrary to what has been suggested in the past (see Section 3.5 for further discussion).

3.4.5 300 AU discs

I now turn my attention to larger discs that are usually considered more likely to promote fragmentation.

Simulations L-Qmin1-Kappa10, L-Qmin1 and L-Qmin1-Kappa0.1 show that though it is easier for larger discs to maintain a state of thermal equilibrium with their boundaries (Figure 3.11), higher opacity discs still struggle to do so, thus implying that low opacities are still important for discs to fragment.

However, despite simulating a large disc with a marginal Toomre stability profile ($Q_{\rm min} = 1$;

simulation L-Qmin1), fragmentation still does not occur for discs with interstellar opacities except for discs whose fragmentation is determined by the initial conditions (e.g. Simulation L-Qmin0.75 which fragments in $t < 1$ ORP). I find that though the opacities do not have to be as small as those in the 25AU discs (0.01× interstellar values), nevertheless, I still require discs with sub-interstellar opacities (0.1× interstellar values). Figure 3.12 shows the disc of simulation L-Qmin1-Kappa0.1 which fragmented after more than an outer rotation period of evolving. Though for fragmentation to occur, the opacity in this disc does not have to be decreased as much as for the smaller disc simulated in Qmin0.75-Kappa0.01, since it is colder, the grain sizes still correspond to sizes ranging between millimetre and centimetre which, as mentioned earlier, is entirely realistic.

3.5 Comparison with previous work

My simulations show that contrary to past studies, it is possible for discs to fragment at small radii (< 25AU) if the disc temperature and the opacity are low enough. The latter point may be the case if the disc is metal-poor or if grain growth has occurred to produce grains that are larger than interstellar sizes. The larger grain size is certainly a reasonable assumption since there is evidence for grains of up to centimetre sizes in discs (Calvet et al. 2002; Testi et al. 2003; Rodmann et al. 2006; Lommen et al. 2007), which could provide the low opacities necessary for my fragmenting discs. With respect to disc metallicity, it is well known that core accretion is not as efficient in metal-poor environments (Kornet et al. 2005). However, I find that gravitational instability is enhanced in such conditions. Boss (2002) carried out simulations of gravitationally unstable discs with varying opacities (from 0.1× to 10× the interstellar Rosseland mean values) in order to explore fragmentation in different metallicity discs and found that the fragmentation results were insensitive to the dust grain opacity. He also found that the disc midplanes could radiate energy to the disc surfaces very rapidly as the timescale for temperature equilibration by radiative diffusion between the disc midplane and the surface was smaller than the orbital timescale regardless of the opacity scaling. However, for my discs the converse is true and I find that the opacity, and hence metallicity, does play a part in the likelihood of fragmentation. My results suggest that gravitational instability may be the dominant giant planet formation mechanism in low metallicity discs.

The results of the small fragmenting discs are also in contrast to previous work which has suggested via simulations (e.g. Stamatellos & Whitworth 2008; Boley 2009) and analytical work (e.g. Rafikov 2005, 2009; Clarke 2009) that fragmentation at such small radii is not possible. However, this is due to the lower opacities in my discs. Rafikov (2009) analytically explored self-gravitating discs including the effects of stellar irradiation and suggested that fragmentation inside of ≈ 120AU is not possible. However, firstly he uses interstellar Rosseland mean opacities (which is not an unreasonable assumption to initially make) and secondly he assumes that the fragmentation boundary occurs when the "effective α parameter" (of the form of Shakura & Syunyaev 1973) due to the gravitational torques $\alpha_{\rm GI} \sim 1$. Other authors (e.g. Clarke 2009) assume the fragmentation boundary occurs when $\alpha_{\rm GI} \sim 0.06$ (Gammie 2001; Rice et al. 2005) which occurs at a radius $R \sim 70$ AU. This is still at a larger radii than my small fragmenting disc, though Clarke (2009)

also assumes interstellar Rosseland mean opacities. Equations A1-A5 of Clarke (2009) show that if the opacity is decreased by two orders of magnitude, as is the case for my small fragmenting disc, and using $\alpha_{\rm GI} \sim 0.06$ as the value at which fragmentation can occur, the radius outside of which fragmentation occurs is at $R \approx 24$ AU (a similarly small radius was also found by Kratter et al. 2010). This is much smaller than the canonical value of ~ 70 AU, although still not quite as small as $R \approx 15$ AU at which I find fragmentation. Given that the works of Clarke (2009) and Kratter et al. (2010) were analytical and while we have performed direct global three-dimensional radiation hydrodynamical simulations, this level of agreement is reasonable

I can also compare my results to the analytical work of Rafikov (2005). Based on a combination of the stability and cooling criteria of Toomre (1964) and Gammie (2001), respectively, Rafikov (2005) analytically derived a constraint on the disc temperature that is required for planet formation via gravitational instability. Using this constraint (equation 5 of Rafikov 2005) and taking into account the reduced opacity in my discs, the Rafikov fragmentation conditions for the surface mass density and temperature in the disc become

$$\Sigma \geq \frac{\Omega^{\frac{11}{10}}}{Q_{\rm crit}\pi G}\left(\frac{k}{\mu}\right)^{\frac{3}{5}}\left[\frac{1}{\zeta\sigma\kappa_o}\right]^{\frac{1}{10}}, \tag{3.4}$$

and

$$T \geq \left[\left(\frac{k}{\mu}\right)\frac{\Omega}{\zeta\sigma\kappa_o}\right]^{\frac{1}{5}}, \tag{3.5}$$

respectively, where k is the Boltzmann constant, μ is the mean particle mass, σ is the Stefan-Boltzmann constant, κ_o is the constant of proportionality for the Rosseland mean opacity expression for this temperature and low opacity regime given by $\kappa = \kappa_o T^2$ and has a value $\kappa_o = 2 \times 10^{-6}$g cm^{-2} K^{-2}, and $\zeta = 2\beta(\gamma - 1)$ and also absorbs any O(1) factors that have not been accurately considered in a proper calculation of the cooling time. Rafikov (2005) assumes that $\zeta \sim 1$. I find that my fragmenting disc is in agreement with these conditions to within a factor of ~ 2. As with the Clarke (2009) comparison, this level of agreement is reasonable given that my simulations are global, three-dimensional and use a realistic radiative transfer method. Consequently, my simulation results are consistent with previous analytical work, but they emphasise the importance of opacity in determining the radius outside of which fragmentation may occur.

Another difference between my simulations and those in the past that have used simplified cooling is that the heating in previous simulations has been dominated by internal heating processes i.e. heating due to gravitational instability and viscous processes, whereas my simulations involve additional external heating due to stellar irradiation. This additional *thermal blanket* causes the Toomre stability parameter to remain high, thus inhibiting fragmentation (consistent with the results of Cai et al. 2008 and Stamatellos & Whitworth 2008). However, in scenarios where the stellar irradiation is not so strong such that the Toomre stability parameter is small, fragmentation is possible, but only if the opacity is also decreased such that the disc can cool easily.

The cooling timescale, ψ, parameter cannot give any indication as to what the gravitational stress in a disc is since it incorporates the cooling in response to the heating due to stellar irradiation

as well as the gravitational stresses. However, it is still interesting to compare the values in my discs to the critical values of β for fragmentation obtained from simulations using more simplified energetics. Previous simulations using simplified cooling have suggested that for discs with $\gamma = 5/3$, fragmentation requires that the total cooling timescale in units of the orbital timescale, $\beta \lesssim 7$ (Rice et al. 2005). My non-fragmenting discs which have reached a steady state are consistent with this result since for these discs $\psi > 7$. For my fragmenting disc, I find that the cooling timescale in units of the orbital timescale can be as much as $\psi \approx 15$. However, I express caution when interpreting this result. It is important to note that the Gammie (2001) and Rice et al. (2005) conditions indicate the minimum cooling time (and hence a maximum gravitational stress) that a disc can support *without* fragmenting. In my fragmenting disc, the temperature continues to change as the disc cools, and hence the cooling rate is not constant since it is temperature dependent.

Chapter 4

Formation of the HR 8799 planetary system by gravitational instability

4.1 Introduction

In Chapter 3, I showed that the opacity in a disc can have a significant effect on whether it can fragment. In particular, I showed that a lower opacity (which may be due to a low metallicity environment or due to the presence of larger grains) allows a disc to cool more rapidly since energy can stream out of the disc easily. This allows a disc to fragment at smaller radii than that expected if higher opacities are used. I showed that while for interstellar opacities a $0.1\mathrm{M}_{\odot}$ disc is unable to fragment at small radii, the same disc can fragment at a radius of $R \approx 15$ AU for opacities that are two orders of magnitude smaller than interstellar Rosseland mean values. I also suggested that since HR 8799 is a metal poor, λ Bootis star with metallicity $[M/H] = -0.47$ (Gray & Kaye 1999), it is reasonable to assume that the disc that once surrounded the star, out of which the planets formed, was also metal-poor and therefore the formation of this planetary system may well have been by gravitational instability.

In this chapter, I explore the possibility of the formation of the HR 8799 system by gravitational instability in detail. I carry out numerical simulations to determine firstly *if* fragments can form in situ, by gravitational instability, at the radii at which the planets in this system are observed and secondly, what disc conditions are required to form these fragments. In addition, I also consider the subsequent growth and radial movement of the fragments in the disc. In Section 4.2, I outline the changes to the numerical method used in the simulations presented in this chapter compared to those presented in Chapter 3. In Section 4.3, I outline the simulations performed and present the results in Section 4.4. Finally, I discuss the results in Section 4.5.

4.2 Numerical details

The numerical method adopted in the simulations presented in this chapter is almost exactly the same as that used in Chapter 3. The vertical location of the boundary between the optically thick part of the disc and the optically thin atmosphere is determined slightly differently to Chapter 3:

previously, the boundary location was determined at the start of the simulation and then kept fixed throughout while the new method re-evaluates the boundary location at least 40 times per outer rotation period (ORP). The boundary height is the maximum out of: the height above the midplane where the optical depth, $\tau = 1$, or the height above the midplane where 10% of the particles at any radius are in the boundary. Therefore, the boundary is at the vertical position where the optical depth, $\tau \lesssim 1$, and the choice ensures that the "bulk" of the disc (i.e. consisting of particles that lie closer to the disc midplane) is simulated using radiative transfer, rather than simplified energetic calculations. The detailed calculations of the boundary height location are presented in Section 2.2.5 and Appendix B.

For some of the simulations presented in this chapter, I follow the mass and radial evolution of the fragments after they have formed. As discussed in Section 2.1.7, the computational expense increases in high density regions. For these simulations, I therefore follow the evolution by turning a clump of particles that are denser than a critical value (see Table 4.2) into sink particles. The newly formed sink particles also carry the same accreting conditions as those used for the central stars (see Section 2.1.7 for details). In these calculations, the central star and other sink particles formed are modelled using an accretion radius of 1 AU i.e. I assume that any gas within 1 AU of the sink particle can potentially be accreted onto it. Once formed, a planetary core typically accretes from a radius on the order of the Hill radius,

$$R_{\rm H} = a \sqrt[3]{\frac{m_{\rm p}}{3M_\star}}, \tag{4.1}$$

where $m_{\rm p}$ and a are the mass and semi-major axis of the protoplanet, respectively, and $M_\star$ is the mass of the central star. As will be shown in Section 4.4.2, the masses of the fragmenting clumps are initially $\approx$ O(1)$M_{\rm J}$. The first fragments that form in a disc (that I eventually simulate using sink particles) form between 20 and 70 AU. For a 1.5$M_\odot$ star and a 1$M_{\rm J}$ fragment forming between 20 and 70 AU, $R_{\rm H} \approx 1 - 4$ AU. Therefore using an accretion radius of 1 AU is reasonable but also conservative enough so that the final mass of the fragments are not overestimated. Note that I also carry out simulations with lower star masses such that $M_\star \approx 0.7 - 0.9 M_\odot$. However, the Hill radius is still similar to the above estimate since the star mass is cube-rooted in equation 4.1.

4.3 Simulations

4.3.1 Disc setup

The HR 8799 system has been observed to have a debris disc with an inner warm component between $\approx 6 - 15$ AU, a planetesimal disc between $\approx 90 - 300$ AU and an outer halo between $\approx 300 - 1000$ AU (Su et al. 2009). Since the planets are observed at projected separations of 24, 38 and 68 AU (Marois et al. 2008), I carry out simulations of discs whose radial extent is $1 < R <$ 100 AU using 250,000 SPH particles. Since the metallicity of HR 8799 is a factor of three smaller than that of the Sun, I simulate discs with opacities that are a maximum of 0.3 times the interstellar Rosseland mean values. I also carry out simulations where the opacities are 0.1 and 0.2 times the interstellar Rosseland mean values which may be the case if grain growth has also occurred in the

Simulation name	$Q_{\rm min}$	$M_\star$ [$M_\odot$]	$M_{\rm disc}$ [$M_\odot$]	$M_{\rm disc}/M_\star$	L [$L_\odot$]	p	Opacity scaling factor	Fragment radius [AU]	Sink?	Fragment time [yrs]
Mstar1.5-Q1.1-p1.75-0.3IS	1.1	1.5	1.2	0.8	4.3	1.75	0.3	-		-
Mstar1.5-Q1.1-p1.75-0.1IS	1.1	1.5	1.2	0.8	4.3	1.75	0.1	< 15 AU		390
Mstar1.5-Q1-p1.75-0.3IS	1	1.5	1.3	0.9	4.3	1.75	0.3	-		-
Mstar1.5-Q1-p1.75-0.1IS	1	1.5	1.3	0.9	4.3	1.75	0.1	< 15 AU		388
Mstar1.5-Q0.9-p1.75-0.3IS	0.9	1.5	1.4	0.9	4.3	1.75	0.3	< 20 AU		723
Mstar1.5-Q0.9-p1.75-0.2IS	0.9	1.5	1.4	0.9	4.3	1.75	0.2	< 15 AU		325
Mstar1.5-Q0.9-p1.75-0.1IS	0.9	1.5	1.4	0.9	4.3	1.75	0.1	< 15 AU		297
Mstar1.5-Q1.1-p1.5-0.1IS	1.1	1.5	0.8	0.5	4.3	1.5	0.1	-		-
Mstar1.5-Q1-p1.5-0.1IS	1	1.5	0.9	0.6	4.3	1.5	0.1	-		-
Mstar1.5-Q0.9-p1.5-0.3IS	0.9	1.5	0.9	0.6	4.3	1.5	0.3	-		-
Mstar1.5-Q0.9-p1.5-0.1IS	0.9	1.5	0.9	0.6	4.3	1.5	0.1	-		-
Mstar1.5-Q0.8-p1.5-0.3IS	0.8	1.5	1.1	0.7	4.3	1.5	0.3	8-30 AU		2274
Mstar1.5-Q0.7-p1.5-0.3IS	0.7	1.5	1.2	0.8	4.3	1.5	0.3	25-60 AU	Y	700
Mstar1.5-Q0.7-p1.5-0.1IS	0.7	1.5	1.2	0.8	4.3	1.5	0.1	55-65 AU	Y	700
Mstar1.5-Q0.9-p1.25-0.1IS	0.9	1.5	0.7	0.5	4.3	1.25	0.1	-		-
Mstar1.5-Q0.7-p1.25-0.1IS	0.7	1.5	0.9	0.6	4.3	1.25	0.1	-		-
Mstar0.8-Q1-p1.75-0.3IS	1	0.8	0.8	1.0	1.45	1.75	0.3	-		-
Mstar0.8-Q1-p1.75-0.1IS	1	0.8	0.8	1.0	1.45	1.75	0.1	< 15 AU		348
Mstar0.7-Q0.9-p1.75-0.3IS	0.9	0.7	0.9	1.3	1.45	1.75	0.3	-		-
Mstar0.7-Q0.9-p1.75-0.1IS	0.9	0.7	0.9	1.3	1.45	1.75	0.1	< 10 AU		243
Mstar0.9-Q0.9-p1.5-0.1IS	0.9	0.9	0.7	0.8	2	1.5	0.1	-		-
Mstar0.8-Q0.8-p1.5-0.1IS	0.8	0.8	0.7	0.9	1.7	1.5	0.1	-		-
Mstar0.8-Q0.7-p1.5-0.3IS	0.7	0.8	0.8	1.0	1.7	1.5	0.3	-		-
Mstar0.8-Q0.7-p1.5-0.1IS	0.7	0.8	0.8	1.0	1.7	1.5	0.1	50-65 AU	Y	1275
Mstar0.9-Q0.7-p1.25-0.1IS	0.7	0.9	0.7	0.8	2	1.25	0.1	-		-
Mstar0.9-Q0.6-p1.25-0.1IS	0.6	0.9	0.6	0.7	2	1.25	0.1	-		-

Table 4.1: Summary of the simulations carried out. The opacity scalings refer to multiples of interstellar Rosseland mean opacity values as described in Section 3.2.1. $Q_{\rm min}$ refers to the minimum value of the Toomre parameter (at the outer edge of the disc) at the start of the simulation. The luminosity has been determined using the stellar evolution models of Siess et al. (2000, Appendix C). p refers to the initial surface mass density, Σ, profile in the disc where $\Sigma \propto R^{-p}$. The penultimate column indicates if the simulation has been run with sink particles to follow the evolution of the fragments further (more details of these simulations can be found in Table 4.2).

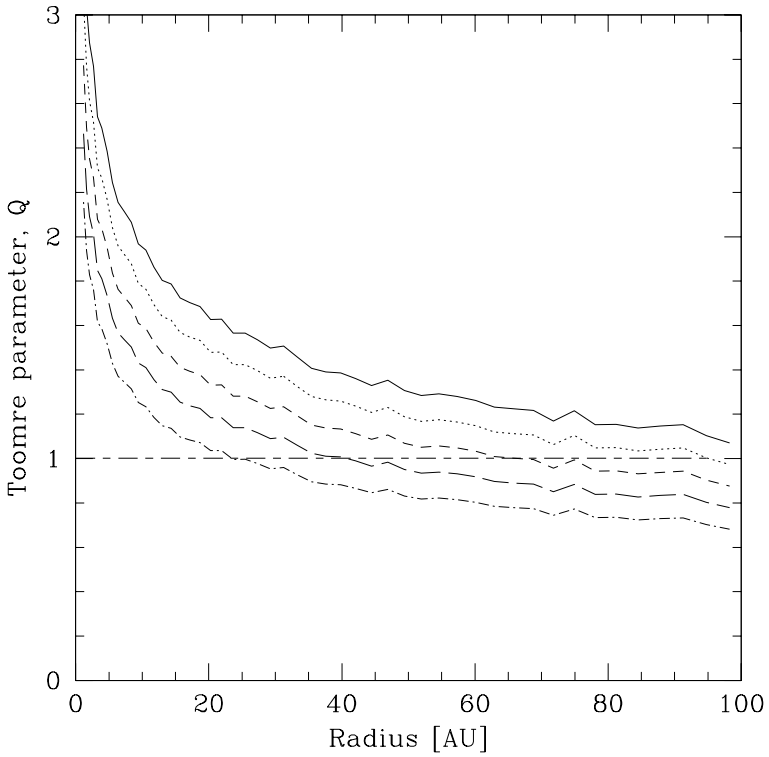

Figure 4.1: Initial Toomre stability profiles for the simulations carried out with a constant star mass, $M_\star = 1.5 M_\odot$, and a surface mass density profile, $\Sigma \propto R^{-3/2}$ set up so that the minimum value of the Toomre parameter at the outer edge, $Q_{\rm min} = 1.1$ (solid line), 1.0 (dotted line), 0.9 (short dashed line), 0.8 (long dashed line) and 0.7 (dot-short dashed line). The critical value of $Q_{\rm crit} \approx 1$ is also marked. Many of these discs are so massive that they are initially Toomre unstable. However, the internal heating processes cause the discs to heat up so that rapid fragmentation due to the unstable initial conditions does not occur.

disc. I assume the initial and boundary temperature profile to be $T \propto R^{-1/2}$, and set the absolute value of the temperature from the luminosity, L, of the central star (using $L = 4\pi R^2 \sigma T^4$, where σ is the Stefan-Boltzmann constant). I carry out two separate sets of simulations: the first simulations assume that the mass of the central star remained constant over time (i.e. the mass of the star when the planets formed was similar to its observed mass today), while the second approach assumes that the star mass will have grown over time. Table 4.1 summarises the simulations presented in this chapter and the key results in terms of fragmentation. If the HR 8799 system was to have formed by gravitational instability, the disc would have had to have been Toomre unstable. I have therefore chosen initial disc setups with low Toomre stability values to allow the discs to have some chance of fragmenting. However, this does not guarantee fragmentation since the cooling rate also needs to be high (Gammie 2001; Rice et al. 2005; Chapter 3).

4.3.2 Constant star mass

The simulations with a constant star mass assume that the mass of the star when the planets formed was similar to its current mass (and therefore aside from planet formation, I assume that the re-

mainder of the disc is dispersed and not accreted onto the central star). Using the stellar evolution models of Siess et al. (2000, Appendix C) for metallicity, $Z = 0.01$, I find that the star's luminosity at an age of 1 Myr would have been $L = 4.3\mathrm{L}_{\odot}$. Note that I choose not to use the outcome of the stellar evolution model at a time earlier than 1 Myr because the result from the model may be very dependent on initial model conditions. I initially set up discs with a uniform Toomre stability profile over the entire radial range of the disc (so that $\Sigma \propto R^{-7/4}$) with $Q = 1.1$ and 1.0. The luminosity and the Toomre stability value sets these disc masses to $M_{\mathrm{disc}} = 1.2$ and $1.3\mathrm{M}_{\odot}$, respectively, and I run these simulations with opacity values that are 0.3 and 0.1 times the interstellar Rosseland mean values (simulations Mstar1.5-Q1.1-p1.75-0.3IS, Mstar1.5-Q1.1-p1.75-0.1IS, Mstar1.5-Q1-p1.75-0.3IS and Mstar1.5-Q1-p1.75-0.1IS). I also set up another disc with the same surface mass density profile but with $Q = 0.9$ over the entire radial extent of the disc such that $M_{\mathrm{disc}} = 1.4\mathrm{M}_{\odot}$, and evolve this disc with opacity values of 0.1, 0.2 and 0.3 times the interstellar Rosseland mean values (simulations Mstar1.5-Q0.9-p1.75-0.1IS, Mstar1.5-Q0.9-p1.75-0.2IS and Mstar1.5-Q0.9-p1.75-0.3IS, respectively).

Concurrently to the simulations carried out for this chapter, I also carried out the simulations in Chapter 5. These show that the surface mass density profile plays a key role in where fragmentation occurs in a disc, with steeper surface mass density profiles promoting fragmentation in the inner regions. I therefore also simulate discs with two different shallower surface mass density profiles, $\Sigma \propto R^{-3/2}$ and $\Sigma \propto R^{-5/4}$, (keeping the initial and boundary absolute temperature and profile the same, using $L = 4.3\mathrm{L}_{\odot}$). This results in an initial Toomre stability profile that decreases with radius such that $Q \propto R^{-1/4}$ and $Q \propto R^{-1/2}$, respectively. For the intermediate surface mass density profile, $\Sigma \propto R^{-3/2}$, I set up two discs with minimum values of Q at the outer edge of the discs, $Q_{\mathrm{min}} = 1.1$ and 1.0 (with disc masses $M_{\mathrm{disc}} = 0.8$ and $0.9\mathrm{M}_{\odot}$, respectively) which I run using opacity values that are 0.1 times the interstellar Rosseland mean values. I also set up two discs with $Q_{\mathrm{min}} = 0.9$ and 0.7 such that the disc masses are 0.9 and $1.2\mathrm{M}_{\odot}$, respectively, and I run these for opacity values of 0.1 and 0.3 times the interstellar Rosseland mean values. In addition, I simulate a disc with $Q_{\mathrm{min}} = 0.8$ ($M_{\mathrm{disc}} = 1.1\mathrm{M}_{\odot}$) using a value of 0.3× the interstellar opacity values. Figure 4.1 shows the initial Toomre stability profiles for the discs with $\Sigma \propto R^{-3/2}$ and various values of Q_{min}. It can be seen that many of the discs start off in a state such that they are Toomre unstable in the outer regions of many of the discs but as seen in Chapter 3, the discs heat up due to the internal heating processes. It is therefore important to note that the fragmentation that is seen does not take place immediately as a result of the initial conditions.

Finally, I decrease the surface mass density profile further and set up discs with $\Sigma \propto R^{-5/4}$ such that the disc mass is more spread out. I set up two discs where $Q_{\mathrm{min}} = 0.9$ and 0.7, which have disc masses of 0.7 and $0.9\mathrm{M}_{\odot}$, respectively (simulations Mstar1.5-Q0.9-p1.25-0.1IS and Mstar1.5-Q0.7-p1.25-0.1IS, respectively) which are run using opacities that are 0.1 times the interstellar Rosseland mean values.

4.3.3 Evolving star mass

The second approach assumes that 90% of the disc material was eventually accreted onto the star while ≈ 10% was driven away from the system by jets and outflows (and a negligible amount went

into forming planets) and hence:

$$\begin{aligned} M_{\star,\text{today}} &= M_{\star,\text{1Myr}} + M_{\text{disc}} - M_{\text{jets}} \\ &= M_{\star,\text{1Myr}} + 0.9 M_{\text{disc}}. \end{aligned} \tag{4.2}$$

I set $M_{\star,\text{today}} = 1.5\text{M}_\odot$ according to observations of HR 8799 (Gray & Kaye 1999). The luminosity at an age of 1 Myr, L, determines the mass of the star at that age. For a given surface mass density profile and initial Toomre stability value at the outer edge of the disc, Q_{min}, the luminosity at an age of 1 Myr also sets the disc mass. I use the stellar evolution models of Siess et al. (2000, Appendix C) for metallicity, $Z = 0.01$, to determine the values of L, $M_{\star,\text{1Myr}}$ and M_{disc}, and do this by finding the value of the luminosity at an age of 1 Myr, from the stellar evolution models, that allows equation 4.2 to be satisfied.

As with the simulations with constant star mass, I carry out simulations firstly with surface mass density profiles, $\Sigma \propto R^{-7/4}$, such that the initial Toomre stability profile is uniform throughout the disc. These are set up so that $Q_{\text{min}} = 1.0$ and 0.9, i.e. with $M_{\text{disc}} = 0.8$ and $0.9\text{M}_\odot$ around stars with masses 0.8 and $0.7\text{M}_\odot$, respectively, so that the disc to star mass ratios are ≥ 1.0. The stars have luminosities, $L = 1.45\text{L}_\odot$. The discs are run for both 0.3 and 0.1 times the interstellar Rosseland mean opacity values.

I also simulate discs with surface mass density profiles, $\Sigma \propto R^{-3/2}$. The discs with $Q_{\text{min}} = 0.9$ and 0.8 ($M_{\text{disc}} = 0.7$ simulated around star masses of 0.9 and $0.8\text{M}_\odot$ with luminosities $L = 2$ and $1.7\text{L}_\odot$, respectively) are run with opacity values of 0.1 times the interstellar values. I also set up a further disc with $M_{\text{disc}} = 0.8\text{M}_\odot$ around a $0.8\text{M}_\odot$ (with $Q_{\text{min}} = 0.7$ and $L = 1.7\text{L}_\odot$) and run this for both values of 0.3 and 0.1 times the interstellar opacity values.

Finally, I simulate discs with even shallower surface mass density profiles with $\Sigma \propto R^{-5/4}$: I model a central star with $M_\star = 0.9\text{M}_\odot$ (so that $L = 2\text{L}_\odot$) surrounded by discs with $M_{\text{disc}} = 0.7$ and $0.6\text{M}_\odot$ (such that $Q_{\text{min}} = 0.7$ and 0.6, respectively). These discs are simulated with opacity values of 0.1 times the interstellar Rosseland mean values.

4.4 Results

4.4.1 Fragment locations and required disc masses

Surface mass density profile, $\Sigma \propto R^{-7/4}$

The initial simulations are carried out with a disc set up such that the initial Toomre stability parameter, Q, is constant over the entire radial extent of the disc. I find that for the discs with initial Q values of 1.1 and 1.0, the discs around $1.5\text{M}_\odot$ stars do not fragment for opacity values of 0.3 times the interstellar values and instead settle into a state where $Q > 1$ (Figure 4.2). Figure 4.3 shows the temperature rendered cross-sectional view of the disc in simulation Mstar1.5-Q1-p1.75-0.3IS which clearly shows that the temperature of the disc midplane is hotter than the boundary temperature. As observed in Chapter 3, for higher opacities, the energy is unable to stream out of the disc very fast and consequently, the disc is unable to cool rapidly enough. However, when both of these discs are simulated using a lower value of the opacity (0.1× the interstellar values),

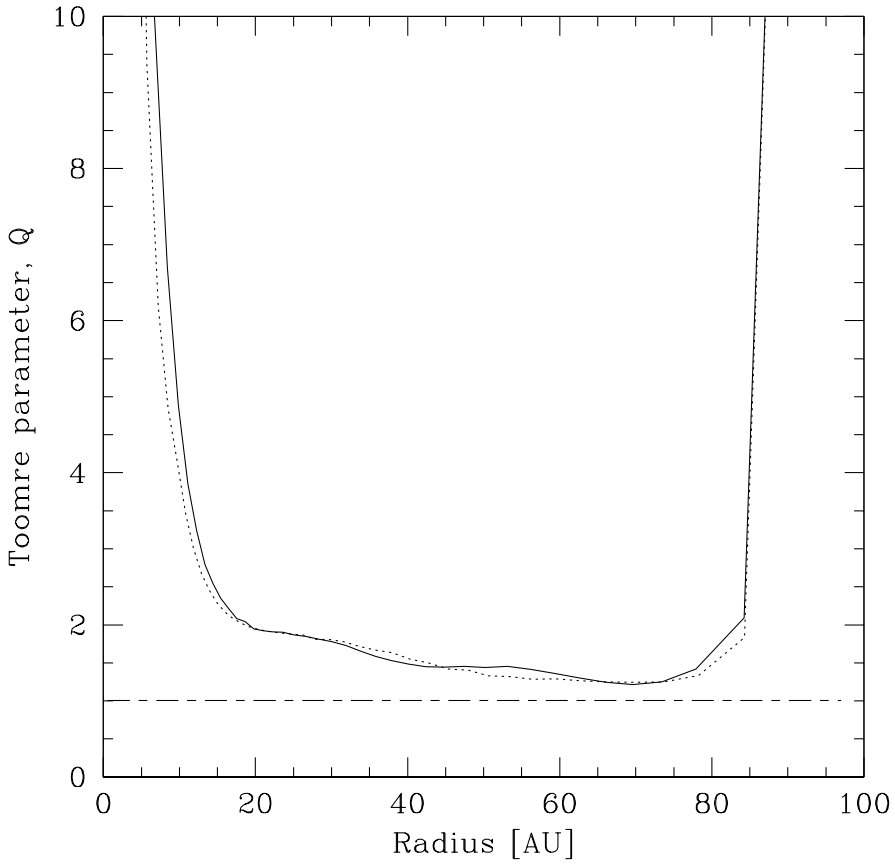

Figure 4.2: Final Toomre stability profiles of the discs with initially constant values of Q = 1.1 (solid line, $M_{\rm disc}$ = 1.2M$_\odot$, simulation Mstar1.5-Q1.1-p1.75-0.3IS) and 1.0 (dotted line, $M_{\rm disc}$ = 1.3M$_\odot$, simulation Mstar1.5-Q1-p1.75-0.3IS) for the simulations carried out with a constant star mass, $M_\star$ = 1.5M$_\odot$, and a surface mass density profile, $\Sigma \propto R^{-7/4}$ and with opacity values that are 0.3× the interstellar values. The internal heating is too high for the disc to maintain low Toomre stability values.

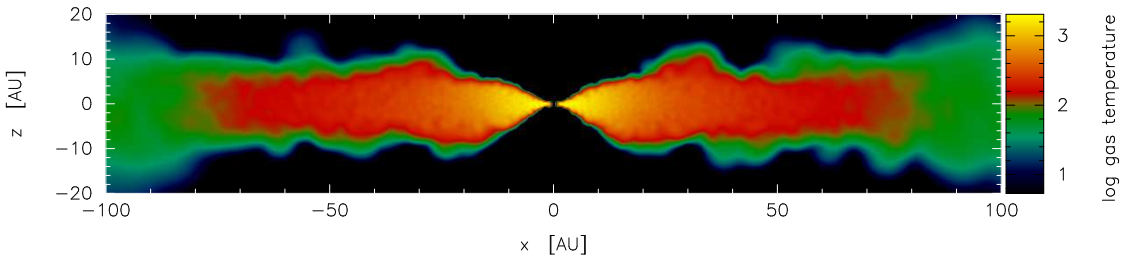

Figure 4.3: Temperature rendered cross-sectional plot of the final disc in simulation Mstar1.5-Q1-p1.75-0.3IS. The internal heating is too high for the disc to maintain a state of thermal equilibrium with the boundary temperature (set by irradiation from the star) as the surface over most of the disc is colder. The simulation uses opacity values that are 0.3× the interstellar values and simulates a 1.3M$_\odot$ disc with a surface mass density profile, $\Sigma \propto R^{-7/4}$, around a 1.5M$_\odot$ star. Temperature units are in Kelvin.

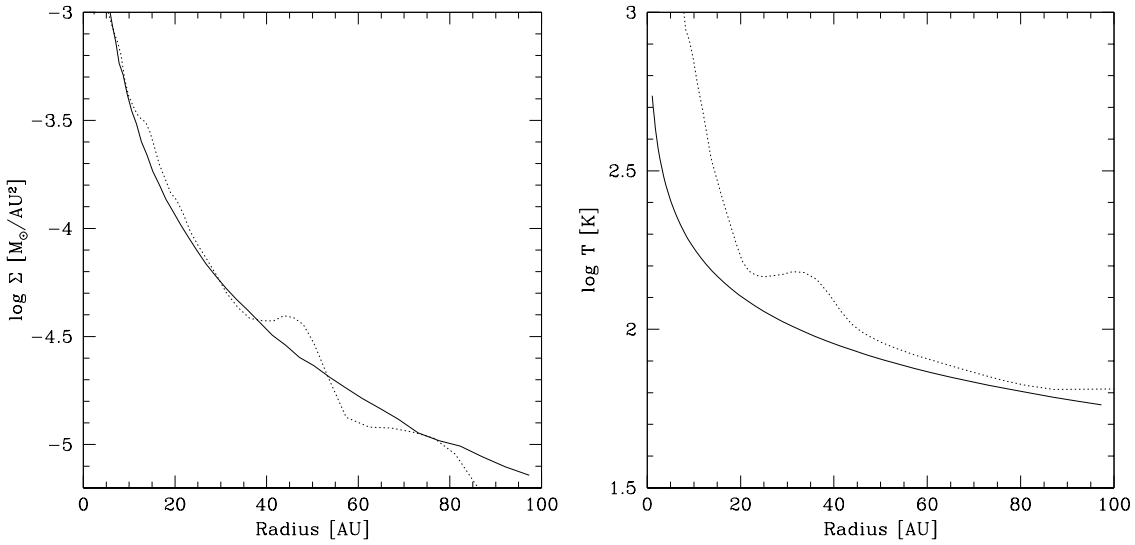

Figure 4.4: Surface mass density (left panel) and temperature (right panel) profiles at the start (solid line) and a short while before fragmentation (t = 373 yrs; dotted line) of simulation Mstar1.5-Q1.1-p1.75-0.1IS (a 1.2$M_\odot$ disc with $\Sigma \propto R^{-7/4}$ surrounding a star with $M_\star = 1.5M_\odot$). The solid line in the temperature graph also represents the boundary temperature. While the temperature in the disc increases, the restructuring of the disc (due to mass transport as a result of gravitational torques) allows it to attain a low Toomre stability parameter at $\lesssim$ 15 AU, thus allowing it to eventually fragment. The fragmenting disc is shown in Figure 4.5.

the discs are able to cool rapidly enough to fragment (within 400 yrs). This is also the case for the equivalent simulation with initial Q = 1.1 (simulation Mstar1.5-Q1.1-p1.75-0.1IS) but is somewhat unexpected for this disc since as discussed in Chapter 3, it is expected that the lowest temperature that the disc can adopt would be the boundary temperature and hence the lowest value of Q after it has evolved would be 1.1. It is therefore expected that this disc would not be able to fragment (though this value is also not too far from the critical value, $Q_{\rm crit} \approx 1$). However, I find that the high disc mass causes gravitational torques that cause mass transport in the disc and hence results in the disc restructuring itself (Figure 4.4) and so even though the temperature increases in the disc, the surface mass density increase as matter moves inwards, causes the value of Q to be reduced enough so that fragmentation occurs. In this particular case, the fragments form out of a single, dense spiral arm (Figure 4.5).

Simulations Mstar1.5-Q0.9-p1.75-0.3IS, Mstar1.5-Q0.9-p1.75-0.2IS and Mstar1.5-Q0.9-p1.75-0.1IS, employ the same disc setup as the above described discs but are slightly more massive (1.4$M_\odot$) and are run with values of 0.3, 0.2 and 0.1 times the interstellar opacity values. All three discs fragment but the disc with 0.1× the interstellar opacity values fragments quicker (within $\approx$ 300 yrs, compared to $\approx$ 325 yrs and $\approx$ 725 yrs for 0.2× and 0.3×, respectively). This is as expected since lower opacity values allow the energy to stream out of the disc much more rapidly and therefore, it is able to maintain a state of thermal equilibrium with its boundary much more quickly. Figure 4.6 illustrates the final state of these discs which shows that as the opacity is increased, the number of fragments appears to decrease (though it is important to note that the simulations have been run for different lengths of time).

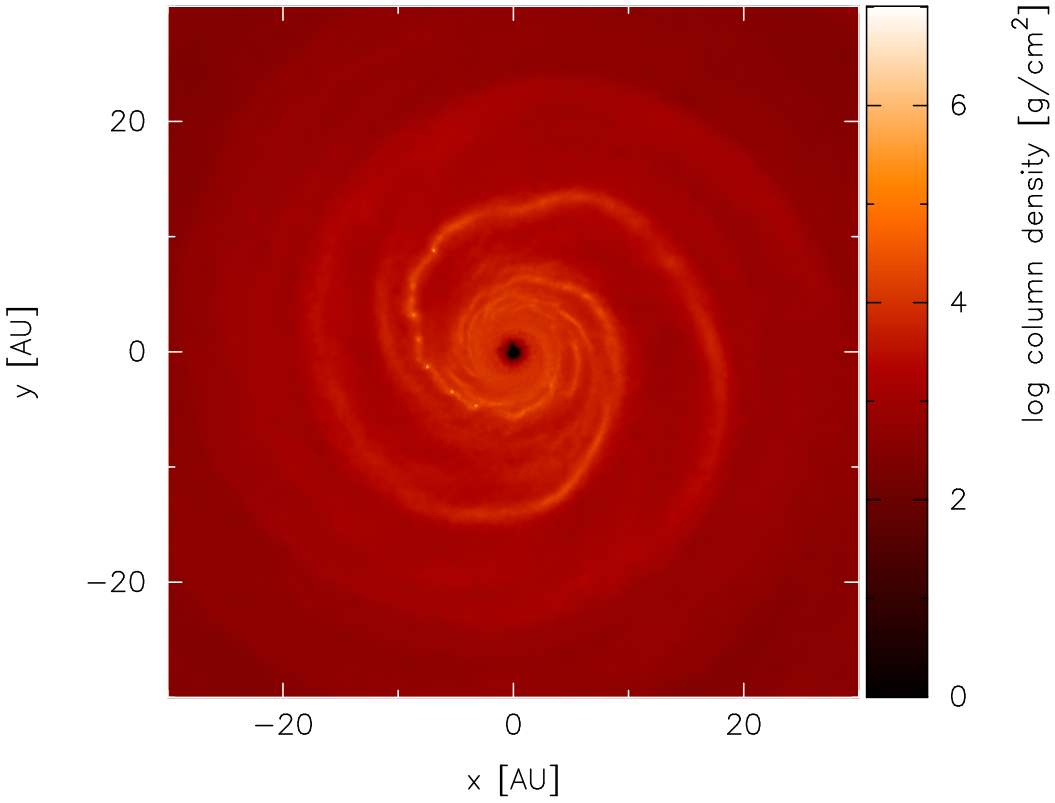

Figure 4.5: Surface mass density rendered image of an initially 1.2$M_\odot$ disc with surface mass density profile, $\Sigma \propto R^{-7/4}$ (such that $Q = 1.1$), surrounding a 1.5$M_\odot$ star (simulation Mstar1.5-Q1.1-p1.75-0.1IS), at time $t = 398$ yrs. The fragments form out of a single dense spiral arm. The surface mass density and temperature profiles for this disc a short while before fragmentation are shown in Figure 4.4.

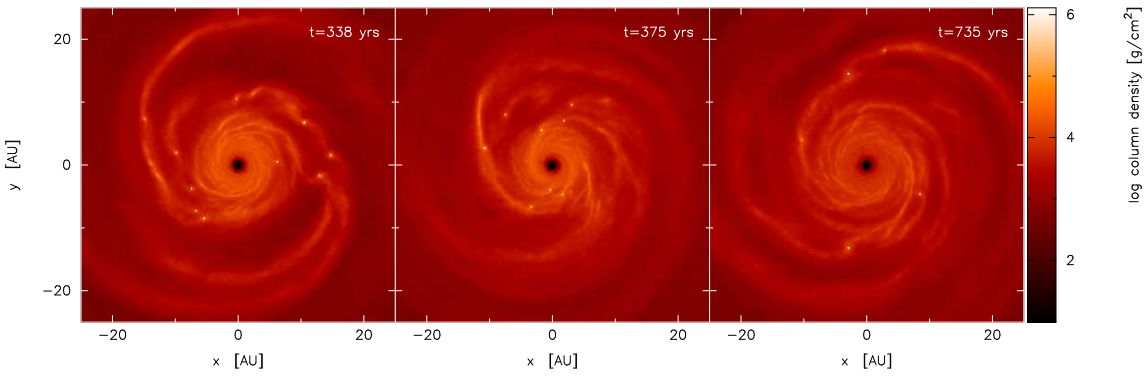

Figure 4.6: Surface mass density rendered images of the discs in simulations Mstar1.5-Q0.9-p1.75-0.1IS (left panel), Mstar1.5-Q0.9-p1.75-0.2IS (middle panel) and Mstar1.5-Q0.9-p1.75-0.3IS (right panel). The simulations are identical except that they have been run with opacity scalings of 0.1 (left panel), 0.2 (middle panel) and 0.3 (right panel). The simulation with the lower opacity fragments quicker and qualitatively appears to produces more fragments than those with a higher opacity, though the simulations have also been run for different lengths of time.

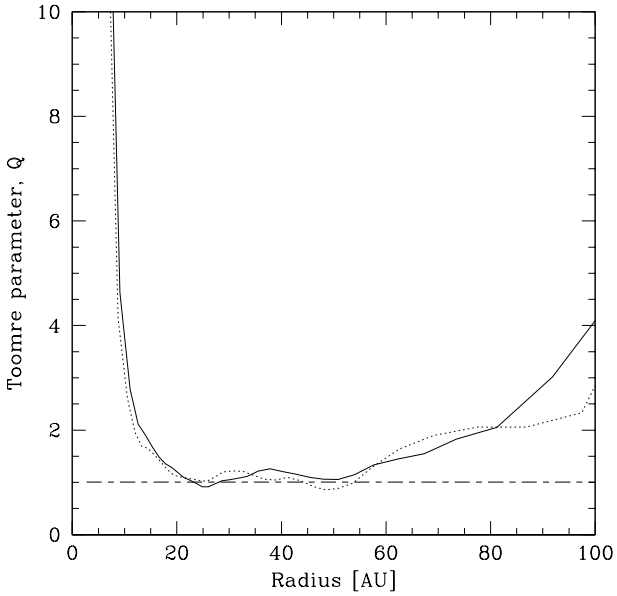

Figure 4.7: Final Toomre stability profiles of the discs with initially constant values of $Q = 1$ (solid line, $M_{\rm disc} = 0.8{\rm M}_{\odot}$, simulation Mstar0.8-Q1-p1.75-0.3IS) and 0.9 (dotted line, $M_{\rm disc} = 0.9{\rm M}_{\odot}$, simulation Mstar0.7-Q0.9-p1.75-0.3IS) for the simulations carried out with an evolved star mass, $M_{\star} = 0.8$ and $0.7{\rm M}_{\odot}$, respectively. The discs have a surface mass density profile, $\Sigma \propto R^{-7/4}$, and are modelled using opacity values that are 0.3× the interstellar values. The discs do not fragment but settle into a state where the Toomre stability parameter, $Q \approx 1$, over a large portion of the disc.

The simulations carried out assuming an evolved star mass show that using opacities with values 0.3× the interstellar values, the discs do not fragment at all. However, the Toomre stability profiles do begin to approach a state where $Q \approx 1$ (Figure 4.7). On the other hand, the simulations with opacity scalings of 0.1 do fragment. Figure 4.8 shows that while the temperature increases in both the discs with $Q_{\rm min} = 0.9$ (more so in the case where the opacity scaling is 0.3), the disc modelled using a scaling of 0.1 does not restructure itself as it fragments rapidly whereas the disc modelled using a scaling of 0.3 does not fragment and has time to restructure itself. In the fragmenting case (simulation Mstar0.7-Q1-p1.75-0.1IS), the fragments form out of a single spiral arm (Figure 4.9).

In all the simulations carried out with a surface mass density profile of $\Sigma \propto R^{-7/4}$, the fragments form at small radii, $R < 20$ AU, regardless of whether the discs were setup assuming a constant star mass or assuming the star mass evolved. This radius is too small for the observed HR 8799 planetary system.

Surface mass density profile, $\Sigma \propto R^{-3/2}$

In Chapter 5 (which was carried out concurrently to this chapter), I show that a shallower surface mass density profile promotes fragmentation further out in a disc. Given that the fragments that form in the discs with surface mass density profile, $\Sigma \propto R^{-7/4}$, form at smaller radii than required for the observed HR 8799 system, I carry out simulations of discs with a shallower profile, $\Sigma \propto R^{-3/2}$. I find that while for a steeper surface mass density profile, discs with $Q_{\rm min} = 1.1$, 1.0 or 0.9

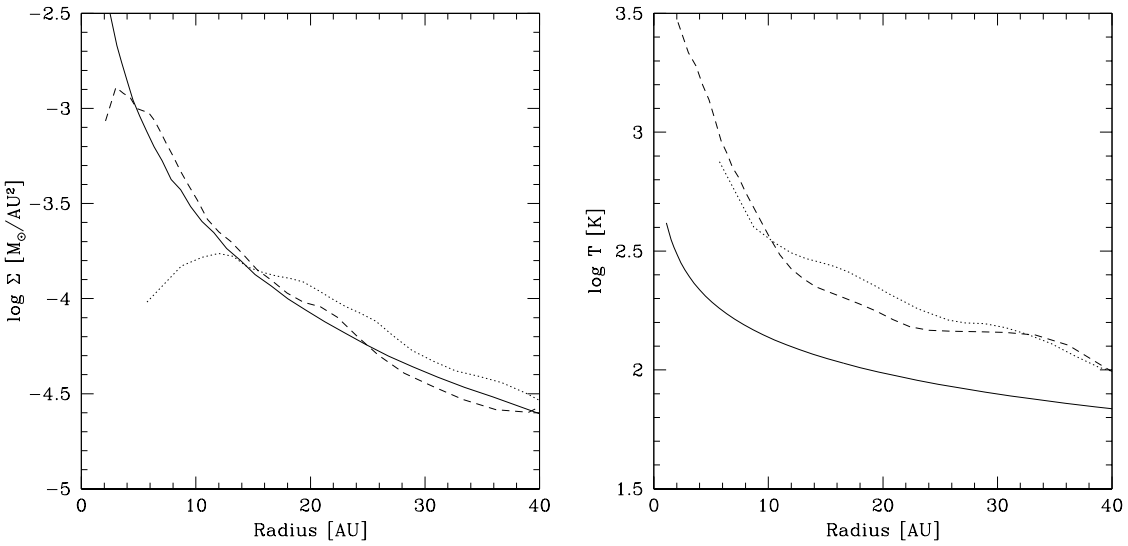

Figure 4.8: Surface mass density (left panel) and temperature (right panel) profiles at the start (solid line) and at a later time for a $0.9M_\odot$ disc around a $0.7M_\odot$ star (with $Q_{\rm min} = 0.9$ and $L = 1.45L_\odot$) modelled using opacity scalings of 0.3 (dotted line, simulation Mstar0.7-Q0.9-p1.75-0.3IS) and 0.1 (dashed line, simulation Mstar0.7-Q0.9-p1.75-0.1IS) for the inner 40 AU of the disc. The graphs for a scaling of 0.1 have been produced at a time shortly before fragmentation. The temperature in both discs increases significantly compared to the initial and boundary value (solid line). In the non-fragmenting case, the disc restructures itself significantly but does not fragment, while in the fragmenting case, the fragments form too rapidly before much restructuring can take place.

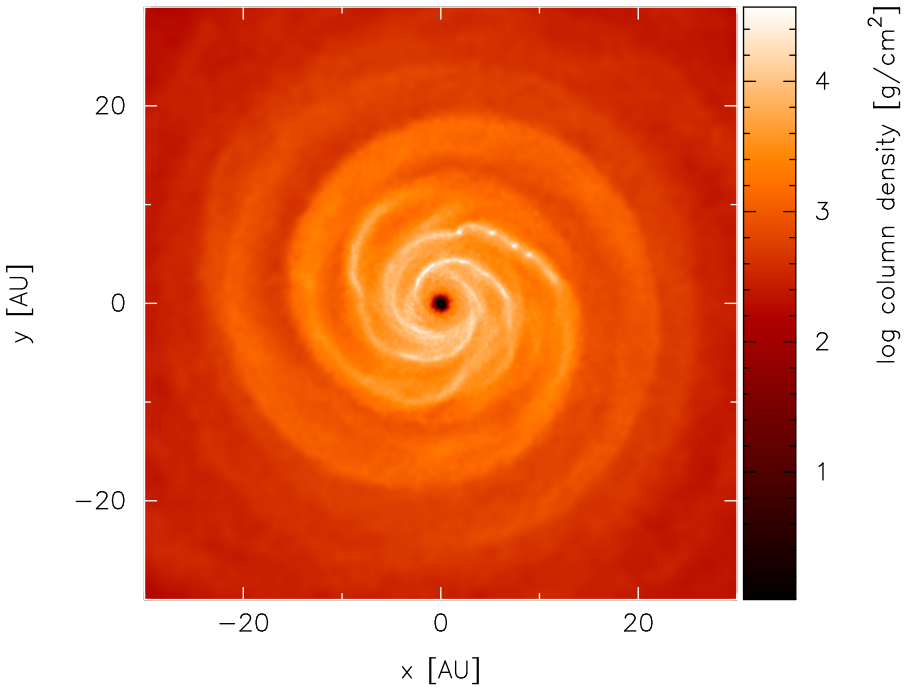

Figure 4.9: Surface mass density rendered image of an initially $0.9M_\odot$ disc with surface mass density profile, $\Sigma \propto R^{-7/4}$ (such that $Q = 1$), surrounding a $0.7M_\odot$ star (simulation Mstar0.7-Q1-p1.75-0.1IS), at time $t = 350$ yrs. The fragments form out of a single dense spiral arm.

are able to fragment using opacities that are 0.1× the interstellar values (assuming a constant star mass, $M_\star = 1.5\mathrm{M}_\odot$), for a shallower surface mass density profile with the same values of $Q_{\rm min}$, the discs are unable to fragment since the combination of the decreased disc mass as well as the initial setup of the disc mass being less concentrated in any one area (due to the shallower surface mass density profile) causes fragmentation to be suppressed.

Figure 4.10 (top left panel) shows the Toomre stability profile for simulation Mstar1.5-Q0.9-p1.5-0.1IS which shows that Q is marginally less than unity between $R = 55 - 70$AU. This is partly because the disc temperature has decreased below the boundary value at radii larger than $\approx$ 60 AU (Figure 4.10, bottom panel), but primarily because the disc has restructured itself. Figure 4.10 (top right panel) shows a plot of the surface mass density profile at the start and at a time, $t = 3172$ yrs. It can be seen that over time, the mass flows to smaller radii causing the surface mass density to increase in most of the disc. However, while it causes the Toomre stability parameter to decrease over most of the disc compared to the initial profile, this is not sufficient enough to cause it to fragment.

On the other hand, increasing the disc mass such that $Q_{\rm min} = 0.8$ (i.e. $M_{\rm disc} = 1.1\mathrm{M}_\odot$, simulation Mstar1.5-Q0.8-p1.5-0.3IS) allows fragmentation to occur (with opacities that are 0.3× the interstellar values). However, while the fragments form in regions further out than for a disc with a steeper surface mass density profile (simulation Mstar1.5-Q0.9-p1.75-0.3IS), the fragments still form too close to the central star ($R \approx 8-30$ AU). Upon increasing the disc mass to $M_{\rm disc} = 1.2\mathrm{M}_\odot$ such that $Q_{\rm min} = 0.7$ (simulation Mstar1.5-Q0.7-p1.5-0.3IS), I find that the disc does fragment, close to the region where the HR 8799 planets are observed at. Figure 4.11 shows an image of this disc once the first three fragments have formed. They initially form at $\approx$ 25, 49 and 60 AU in sequence starting from the outermost fragment first. This figure shows the evolved state of the disc once the fragments have been replaced by sink particles, represented by black stars (see Section 4.4.2).

For the simulations that assume an evolving star mass, while a disc with $Q_{\rm min} = 0.9$ fragments if $\Sigma \propto R^{-7/4}$ (with 0.1× the interstellar opacity values), the disc with the same value of $Q_{\rm min}$ does not fragment with a shallower surface mass density profile of $\Sigma \propto R^{-3/2}$ (simulation Mstar0.9-Q0.9-p1.5-0.1IS). For this simulation as well as simulations Mstar0.8-Q0.8-p1.5-0.1IS and Mstar0.8-Q0.7-p1.5-0.3IS, the Toomre stability values over most of the disc are close to $Q \approx 1$ (Figure 4.12). For the simulation where $Q_{\rm min} = 0.9$, the Toomre parameter for most of the disc is smaller than the initial value (Figure 4.13, top left panel). Figure 4.13 (bottom panel) shows that the temperature has increased from the initial value for most of the disc and therefore, the reduced Toomre stability values are due to the change in surface mass density due to the movement of the mass in the disc (Figure 4.13, top right panel).

Comparing simulations Mstar0.8-Q0.7-p1.5-0.3IS and Mstar0.8-Q0.7-p1.5-0.1IS, the discs used are identical with $Q_{\rm min} = 0.7$ ($M_{\rm disc} = 0.8\mathrm{M}_\odot$ around a $0.8\mathrm{M}_\odot$ star with $L = 1.7\mathrm{L}_\odot$). This disc is simulated using opacities that are 0.3 and 0.1 times the interstellar Rosseland mean values. I find that the former case does not fragment, but the latter case does. The three fragments that form do so out of a single dense spiral arm. They form close together at $\approx$ 50, 57 and 65 AU (Figure 4.14, top panel). It is interesting to note that for the evolving mass cases, the fragmenting

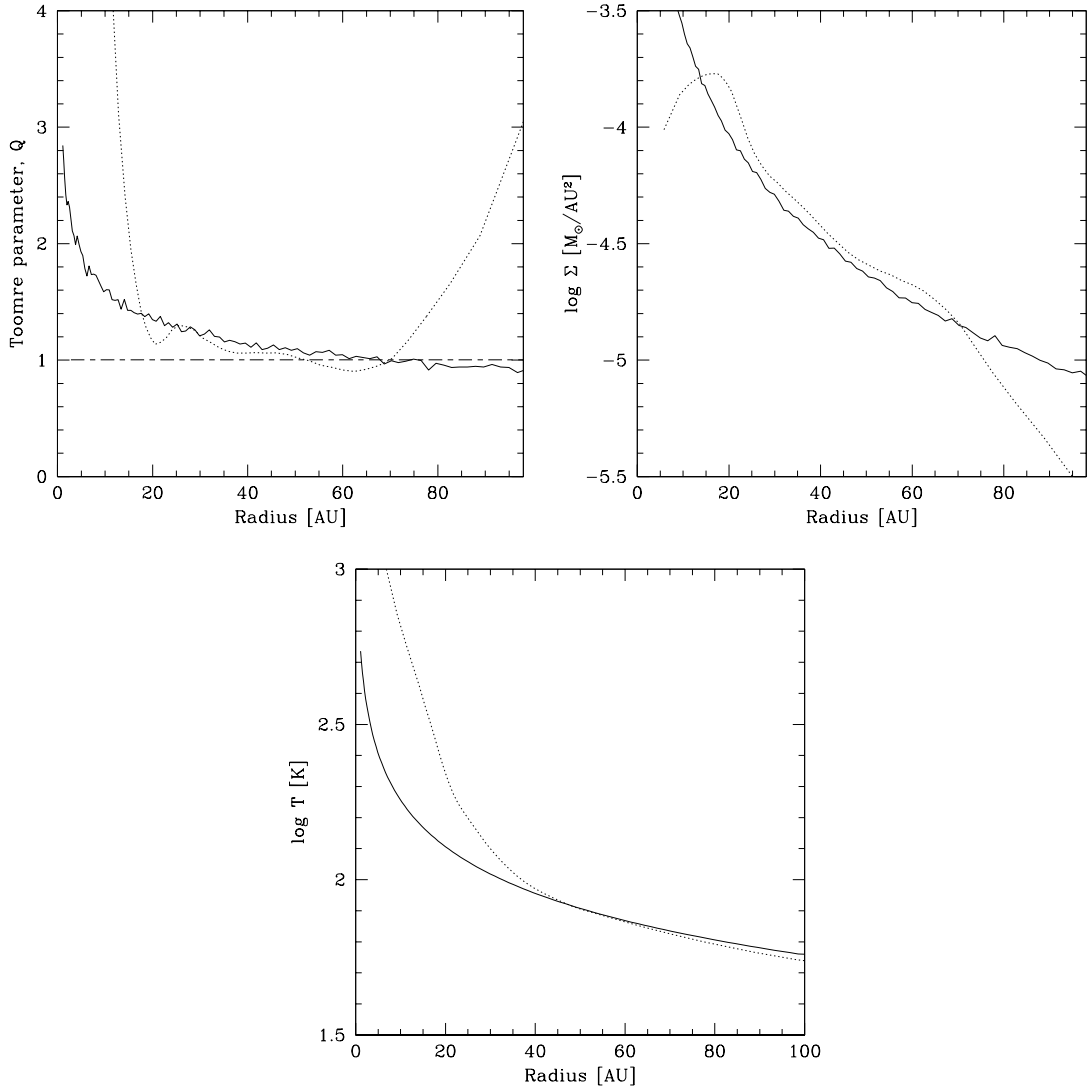

Figure 4.10: Toomre stability (top left panel), surface mass density (top right panel) and temperature (bottom panel) profiles of the initial (solid line) and final (dotted line) state of the disc in simulation Mstar1.5-Q0.9-p1.5-0.1IS which is a $0.9M_{\odot}$ disc surrounding a $1.5M_{\odot}$ star (with $Q_{\rm min} = 0.9$). This simulation was carried out using opacity values that are $0.1\times$ the interstellar Rosseland mean values. The initial temperature profile is also the same as the boundary profile set by the irradiation from the central star. Q is marginally less than the critical value between $R \approx 55 - 70$ AU. Though the final temperature (dotted line) is mostly higher than the boundary temperature (solid line), the final Toomre stability profile is below the initial value since the disc has restructured itself such that the surface mass density in the final state (dotted line) is higher than the initial surface mass density (solid line) for most of the disc. The high gravitational torques in the disc causes mass transport which results in the restructuring. The critical value of $Q_{\rm crit} \approx 1$ is also marked on the Toomre stability plot.

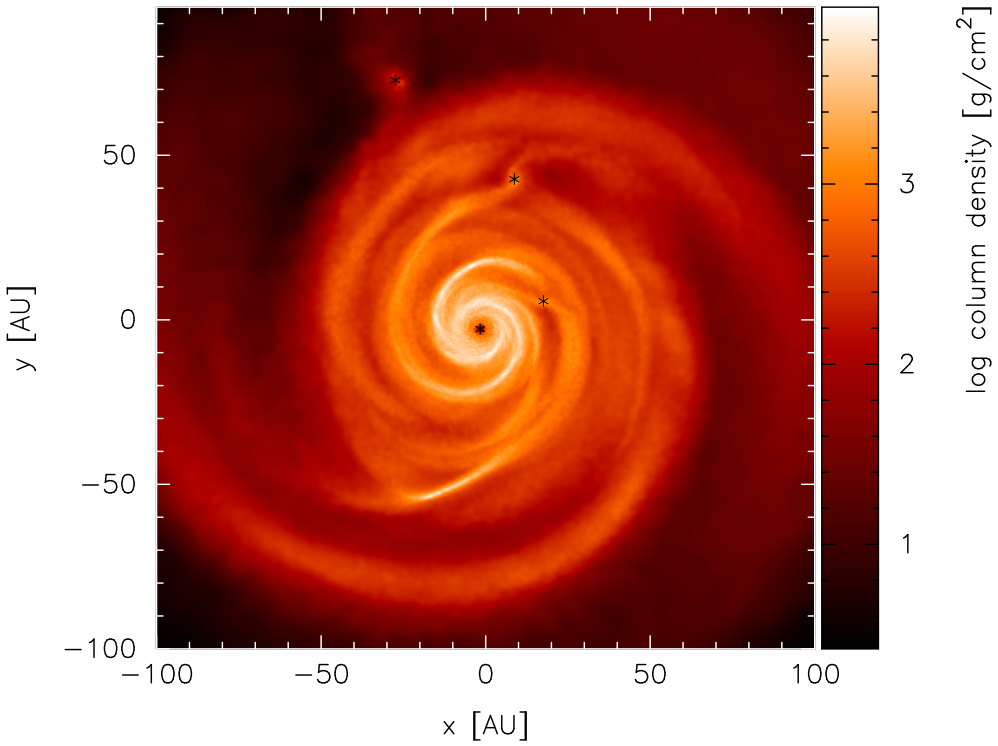

Figure 4.11: Surface mass density rendered image of the fragmenting disc in simulation Mstar1.5-Q0.7-p1.5-0.3IS at t = 1125 yrs. The central star and the fragments are represented by black stars. The fragments form at ≈ 25, 49 and 60 AU and evolve to ≈ 19, 48 and 77 AU, respectively, i.e. a little more spread out than the locations of the observed HR 8799 planets.

simulations have all involved large disc to star mass ratios ($\gtrsim 1$).

Surface mass density profile, $\Sigma \propto R^{-5/4}$

A small number of simulations were performed with a surface mass density profile, $\Sigma \propto R^{-5/4}$. I found that none of these simulations showed any signs of fragmentation (regardless of the assumption of the evolution of the mass of the central star). For a particular value of the minimum Toomre parameter at the outer edge of the disc, $Q_{\rm min}$, the total disc mass for a shallower surface mass density profile will be smaller than for a steeper surface mass density profile. In Chapter 5, I show that the disc mass does play a part in whether fragmentation occurs, with a higher disc mass promoting fragmentation. In addition, a steeper surface mass density profile allows more mass to be concentrated in the inner regions thus promoting fragmentation in those areas. I find that indeed, the discs modelled with a shallow surface mass density profile, $\Sigma \propto R^{-5/4}$, are unable to fragment, even though the minimum values of the Toomre parameter at the outer edge are as low as $Q_{\rm min} = 0.6 - 0.9$. This therefore means that if the disc that surrounded HR 8799 had a shallow surface mass density profile, then this disc would have been very Toomre unstable (with a mass larger than $M_{\rm disc} \approx 0.6 - 0.9 {\rm M}_{\odot}$).

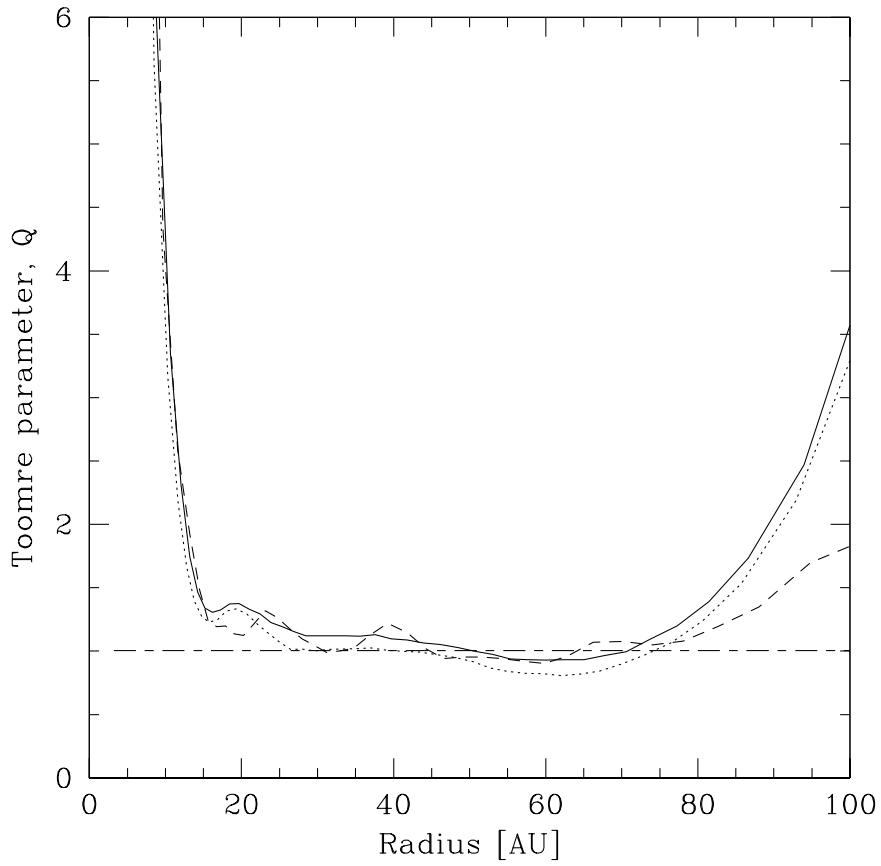

Figure 4.12: Toomre stability profiles of the final state of the discs with surface mass density profiles, $\Sigma \propto R^{-3/2}$, in simulations Mstar0.9-Q0.9-p1.5-0.1IS ($M_{\rm disc}$ = 0.7M$_\odot$, $M_\star$ = 0.9M$_\odot$, $Q_{\rm min}$ = 0.9, solid line), Mstar0.8-Q0.8-p1.5-0.1IS ($M_{\rm disc}$ = 0.7M$_\odot$, $M_\star$ = 0.7M$_\odot$, $Q_{\rm min}$ = 0.8, dotted line) and Mstar0.8-Q0.7-p1.5-0.3IS ($M_{\rm disc}$ = 0.8M$_\odot$, $M_\star$ = 0.8M$_\odot$, $Q_{\rm min}$ = 0.7, dashed line). The former two simulations are run with opacity scalings of 0.1 while the latter simulation is run with a scaling of 0.3. The discs have settled into a marginally stable state with $Q \approx 1$. The critical value of $Q_{\rm crit} \approx 1$ is also marked.

Simulation name	$Q_{\rm min}$	$M_\star$ [M$_\odot$]	$M_{\rm disc}$ [M$_\odot$]	$M_{\rm disc}/M_\star$	L [L$_\odot$]	p	Opacity scaling factor	Max. sink mass [M$_{\rm J}$]	Final no of frag.	Final sim. time [yrs]	$\rho_{\rm crit}$ [g/cm^2]
Mstar1.5-Q0.7-p1.5-0.3IS	0.7	1.5	1.2	0.8	4.3	1.5	0.3	48	35	1328	1×10^{-5}
Mstar1.5-Q0.7-p1.5-0.1IS	0.7	1.5	1.2	0.8	4.3	1.5	0.1	22	23	907	1.9×10^{-5}
Mstar0.8-Q0.7-p1.5-0.1IS	0.7	0.8	0.8	1.0	1.7	1.5	0.1	58	12	1663	1×10^{-5}

Table 4.2: Summary of the simulations carried out with sink particles. The opacity scalings refer to multiples of interstellar Rosseland mean opacity values as described in Section 3.2.1. $Q_{\rm min}$ refers to the minimum value of the Toomre parameter (at the outer edge of the disc) at the start of the simulation. The maximum sink mass is the largest mass sink particle that is present at the final simulation time.

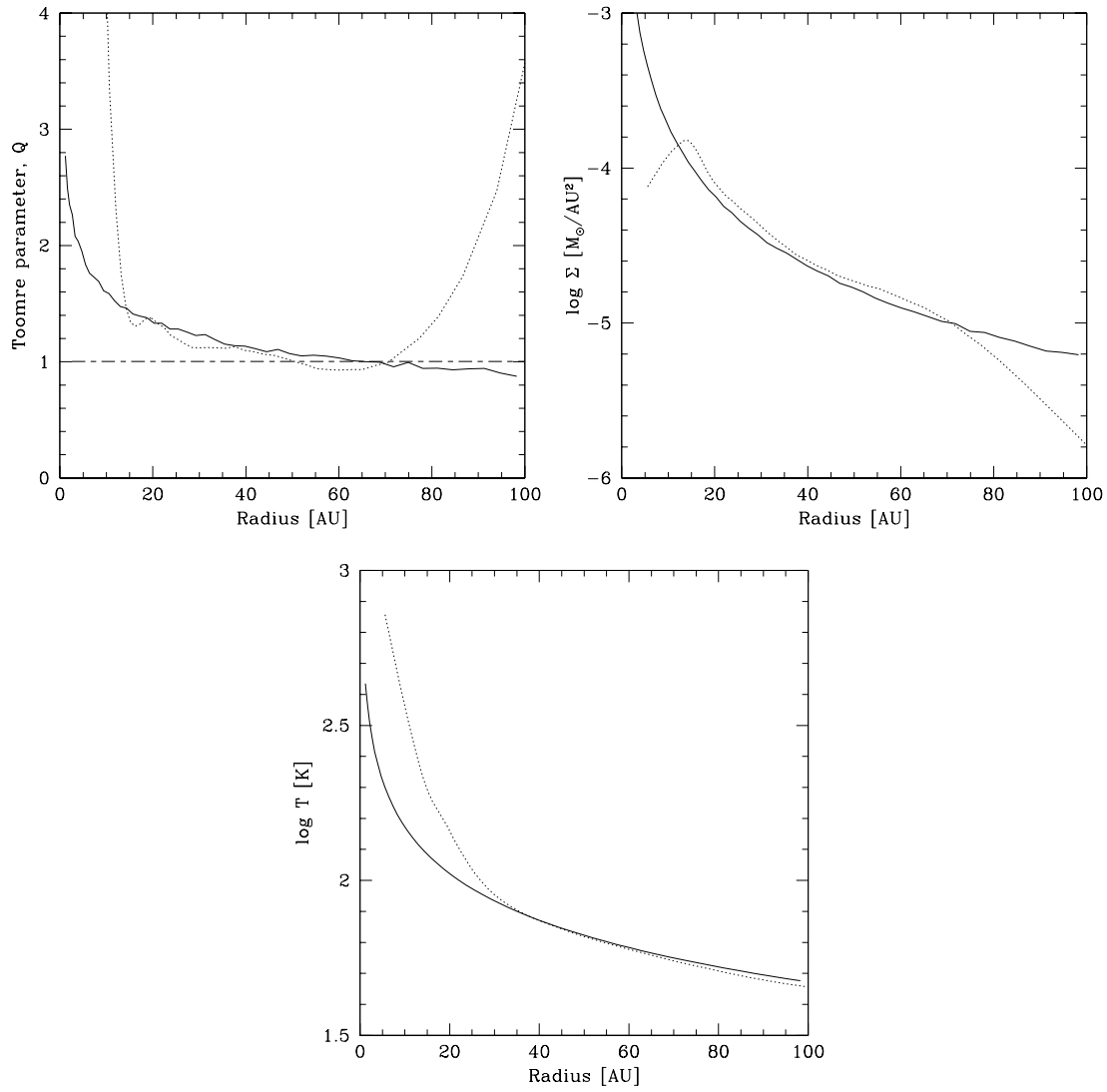

Figure 4.13: Toomre stability (top left panel), surface mass density (top right panel) and temperature (bottom panel) profiles of the initial (solid line) and final (dotted line) state of the discs in simulation Mstar0.9-Q0.9-p1.5-0.1IS. This simulation involves a $0.7M_{\odot}$ disc with surface mass density profile, $\Sigma \propto R^{-3/2}$ (such that $Q_{\rm min} = 0.9$), around a $0.9M_{\odot}$ star. The initial temperature profile is also the same as the boundary profile set by the irradiation from the central star. The temperature has increased over most of the disc and therefore it is the increase in surface mass density that causes the final Toomre stability profile to be smaller than the initial for the majority of the disc. The critical value of $Q_{\rm crit} \approx 1$ is also marked on the Toomre stability plot.

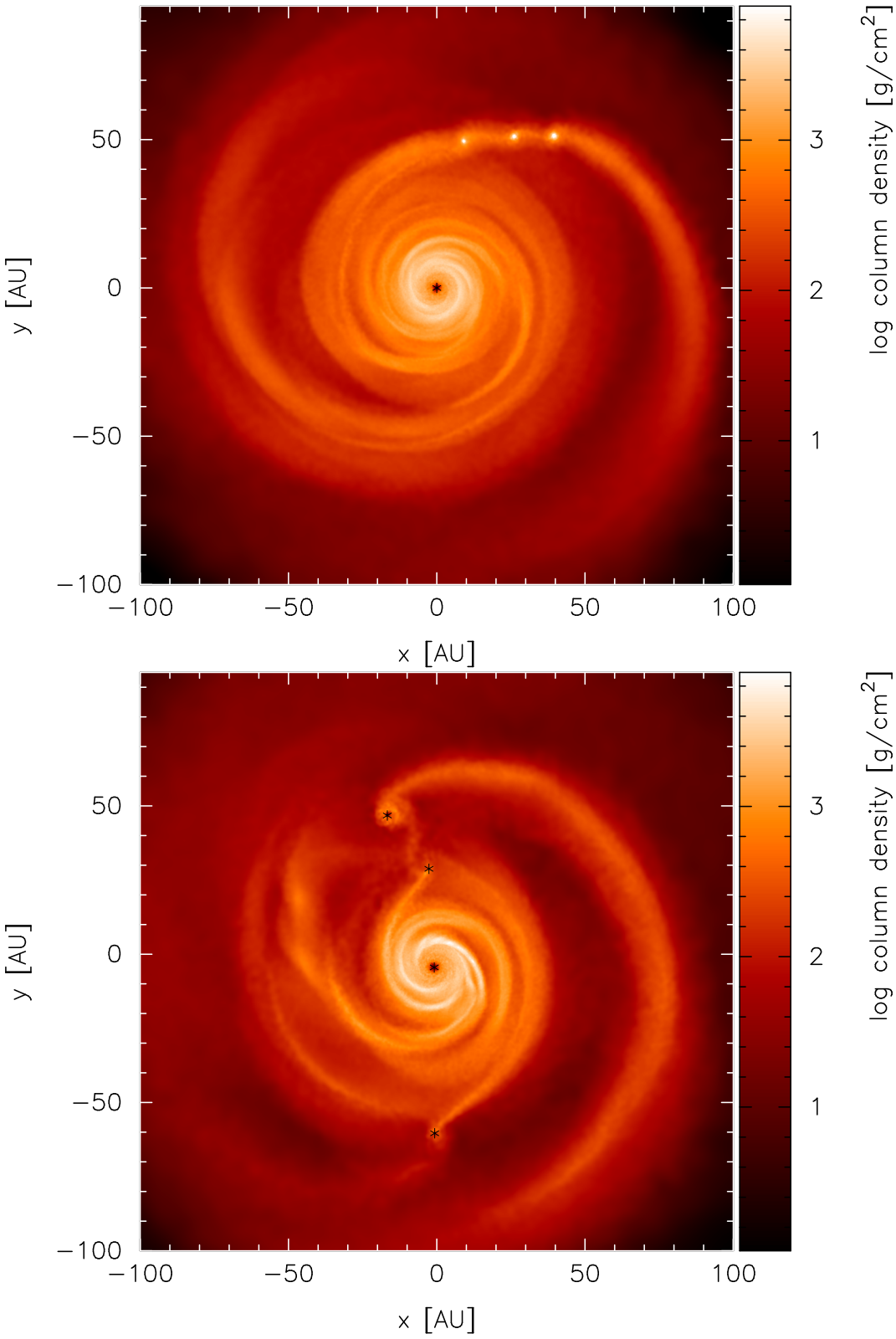

Figure 4.14: Surface mass density rendered images of the fragmenting disc in simulation Mstar0.8-Q0.7-p1.5-0.1IS at t = 1300 yrs (top panel) and t = 1550 yrs (bottom panel). The three fragments form close together at 50, 57 and 65 AU in a single dense spiral arm (top panel), but then evolve to radii of 29, 50 and 61 AU, i.e. within the radii range that the HR 8799 planets are observed (bottom panel). In the bottom panel, the fragments have been replaced by sink particles which, along with the central star, are represented by black stars.

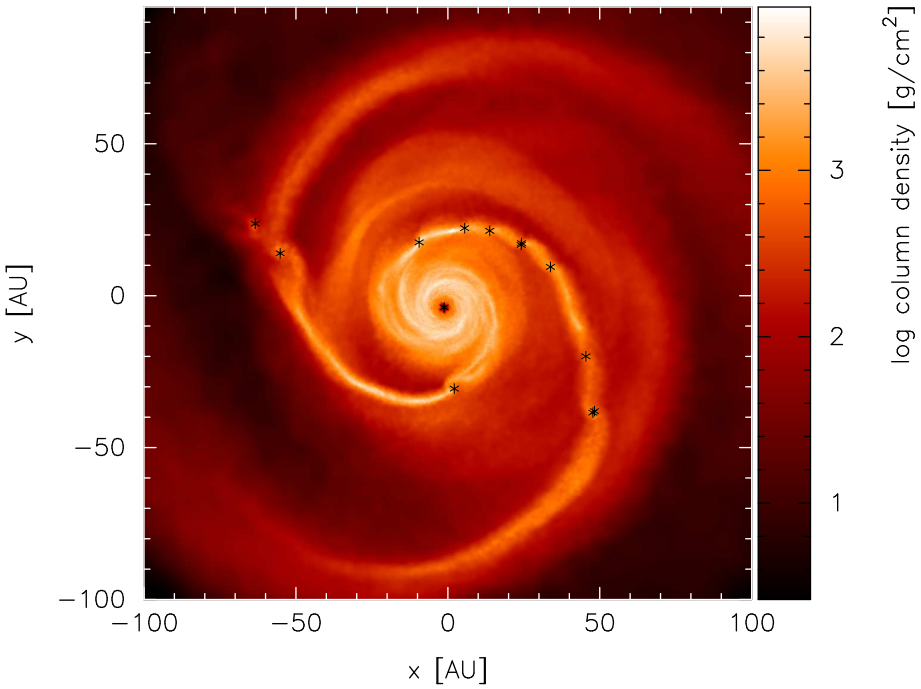

Figure 4.15: Surface mass density rendered image of the fragmenting disc in simulation Mstar1.5-Q0.7-p1.5-0.3IS (a 1.2$M_\odot$ disc surrounding a 1.5$M_\odot$ star with initial $Q_{\rm min}$ = 0.7) at t = 1178 yrs. The simulation was run using an opacity scaling of 0.3. After the first few fragments form (Figure 4.11), further fragmentation occurs in a single dense spiral arm. The fragments have been replaced by sink particles which, along with the central star, are represented by black stars.

4.4.2 Growth and radial evolution of fragments

Table 4.2 summarises the results of those simulations that have been run beyond the fragmenting stage using sink particles to follow the evolution of the fragments. In all the simulations carried out, I find that the fragments form in the outer regions followed by subsequent fragmentation occurring at a later time in the interior regions of the discs. In addition, I find that multiple fragments form out of the same dense spiral arm.

Figure 4.11 shows the fragmenting disc in simulation Mstar1.5-Qmin0.7-p1.5-0.3IS (with the fragments that have been turned into sink particles represented by black stars). The first three fragments form at ≈ 25, 49 and 60 AU in sequence starting from the outermost fragment first. I find that they evolve to ≈ 19, 48 and 77 AU, i.e. spanning a radial range greater than that of the planets currently observed in the HR 8799 planetary system. I note the presence of fairly strong spiral structures inside of ≈ 60 AU, particularly at small radii, suggesting that possible subsequent fragmentation may take place. Indeed, further fragmentation does take place. Figure 4.15 shows an image of further fragments forming out of one dense spiral arm, with the fragments that form having been turned into sink particles). When the first three sink particles form, they have masses of approx 1.2, 2 and 0.75M_J. However, by the time the fourth fragment forms (39 years after the first fragment), the fragments grow in mass to ≈ 21, 9.7 and 8.3M_J respectively. With the exception of the first fragment, these are close to the masses of observed HR 8799 planets. However, they all grow to masses larger than 20M_J. At the end of the simulation, the maximum fragment mass is

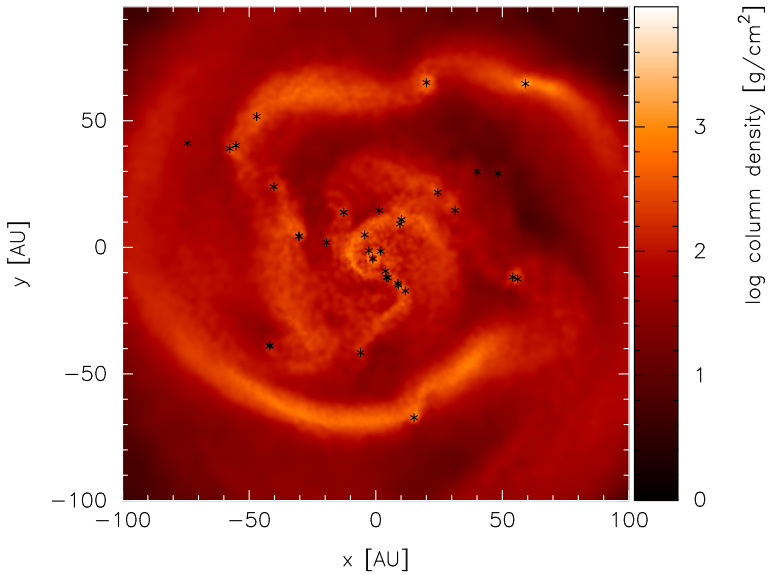

Figure 4.16: Surface mass density rendered image of the fragmenting disc in simulation Mstar1.5-Q0.7-p1.5-0.3IS (a $1.2M_\odot$ disc surrounding a $1.5M_\odot$ star with initial $Q_{\rm min} = 0.7$) at $t = 1328$ yrs. The simulation was run using an opacity scaling of 0.3. Following the initial formation of the fragments (Figure 4.11) and subsequent fragmentation (Figure 4.15), the formation of many more fragments causes the disc to be completely disrupted. The fragments have been replaced by sink particles which, along with the central star, are represented by black stars.

$\approx 48M_{\rm J}$ (at a time of 1328 years) and there is virtually no evidence of a disc having been present since it has been completely disrupted by the formation of the planets (Figure 4.16).

Figure 4.17 shows the fragmenting disc in simulation Mstar1.5-Qmin0.7-p1.5-0.1IS (with the fragments that have been turned into sink particles represented by black stars). The first four fragments form at similar radii out of two spiral arms at 57, 62, 62 and 55 AU at $t = 692$, 708, 710 and 721 yrs, respectively. These fragments have initial masses of 0.6, 1, 1.4 and $0.9M_{\rm J}$, respectively. By the time the fifth fragment forms at $\approx$ 44 AU (at $t = 777$ yrs), the first four fragments have evolved to $\approx$ 50, 63, 68 and 51 AU respectively, i.e. the fragments closer to the central star have moved to smaller radii whereas those at larger distances from the star have moved further away from the star. This effect is likely to be due to a combination of dynamical interactions between the planets as well as between the planets and the disc, in addition to gas drag. Their masses have also grown significantly to ≈ 8.4, 7.5, 5.4 and $7.9M_{\rm J}$, respectively. Following this, the spiral arms interior to this region become more dense and begin to fragment. By the end of the simulation (907 yrs), the maximum mass that the sink particles reach is $20M_{\rm J}$ and a total of 28 fragments are seen.

The simulations described above are identical to each other but were run using different opacities. In Section 4.4.1, I suggested that an increased opacity causes the number of fragments to be reduced. However, this does not appear to be the case when the simulations are run for longer using sink particles.

In simulation Mstar0.8-Q0.7-p1.5-0.1IS ($Q_{\rm min} = 0.7$ with a disc mass, $M_{\rm disc} = 0.8M_\odot$ around a $0.8M_\odot$ star), the first three fragments that form do so out of a single dense spiral arm. They form close together at $\approx$ 50, 57 and 65 AU (Figure 4.14, top panel). I find that they evolve

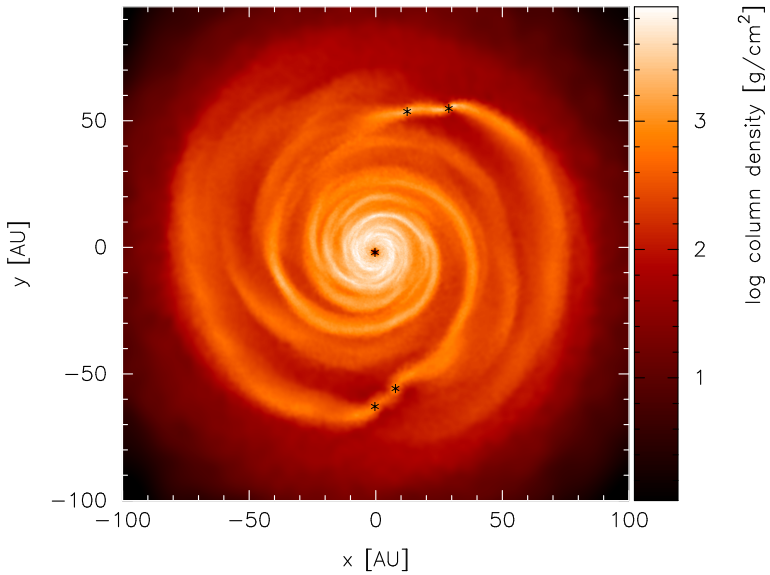

Figure 4.17: Surface mass density rendered image of the fragmenting disc in simulation Mstar1.5-Q0.7-p1.5-0.1IS (a 1.2$M_\odot$ disc surrounding a 1.5$M_\odot$ star with initial $Q_{\rm min} = 0.7$) at $t = 721$ yrs. The simulation was run using an opacity scaling of 0.1. The first fragments form at similar radii ($R \approx 55 - 62$ AU) out of two spiral arms. The fragments have been replaced by sink particles which, along with the central star, are represented by black stars.

to ≈ 29, 50 and 61 AU within 250 yrs (Figure 4.14, bottom panel). The masses are initially ≈ 3, 2 and 1.8M_J, respectively, and by the time the fourth fragment forms, the masses are ≈ 17, 37 and 36M_J, respectively. Further fragments form at smaller radii as the simulation progresses and by the end of the simulation ($t = 1663$ yrs), the maximum fragment mass is very large, ≈ 58M_J.

4.5 Discussion

The results presented here show that it is possible for fragments to form in the radial range of interest of the HR 8799 planetary system ($R \approx 24 - 68$ AU) but only if the disc masses are as high as ≈ 0.8 – 1.2$M_\odot$ (with disc to star mass ratios of between 0.8 and 1.1). In the case of the lower mass disc, for fragmentation to occur, a reduction in the opacity is required, which is feasible if grain growth has occurred such that a combination of a low metallicity and larger grain sizes cause the overall reduced opacity. For the higher mass discs, fragmentation does occur in the radial range required for the HR 8799 planetary system for both opacity values of 0.1× and 0.3× the interstellar Rosseland mean values, though the first fragments that form in the former case are typically concentrated between $R \approx 50$ and 68 AU. While I expect that more massive stars would harbour more massive discs, these disc masses are much larger than expected. However, quite massive discs have been observed early in the star formation process with masses ≈ 0.1$M_\odot$ (Eisner & Carpenter 2006; Andrews & Williams 2007) and even larger disc masses up to ≈ 0.3$M_\odot$ (Eisner & Carpenter 2006) (though the stellar masses for these discs are mostly in the range $M_\star \approx 0.1 - 1.0 M_\odot$), so such large disc masses may be plausible early on. In Chapter 5 I show that fragmentation is promoted if the disc to star mass ratio is effectively increased. The results presented here are consistent with this conclusion since high disc to star mass ratios are required, but it is also important to note

that this is dependent on the cooling rate (which, for example, may be determined by the opacity scaling; Chapter 3) and the surface mass density profile. It is also important to note that I have only considered the effects of the stellar luminosity in these simulations. However, given that the discs are massive, the accretion luminosity is likely to be important and may therefore affect the boundary temperature.

In addition, I find that a moderate surface mass density profile ($\Sigma \propto R^{-3/2}$ is needed for fragments to form in the required range. If the surface mass density profile is too steep, the fragments form too far in to describe the formation of the HR 8799 planets, whereas too shallow values require the discs to be so gravitationally unstable to begin with that it is questionable whether the discs would be able to reach such a state (without fragmenting) in the first place. However, even when simulating the discs that show fragmentation in the radial range of interest, using sink particles, subsequent fragmentation does occur at small radii, so it may be the case that fragmentation at small radii is unavoidable. This is somewhat counterintuitive given that previous work has suggested that gravitational instability is only able to form planets at large distances. However, previous simulations of gravitationally unstable discs typically do not consider the evolution beyond the fragmentation stage. In addition, this may be an artefact of the fact that massive discs are being used in these simulations whereas previous simulations of gravitationally unstable discs typically consider low mass discs (O(0.1)$M_\odot$). Lodato & Rice (2005), however, consider the evolution of massive discs (up to $M_{\rm disc} = 1M_\odot$ around a $1M_\odot$ star so that the disc to star mass ratio is 1) and find that the surface mass density changes during the evolution (whereas in low mass discs, this is not the case, e.g. Lodato & Rice (2004) and Chapter 3). This is evident here (also discussed later in Chapter 5) and as shown in many of the simulations presented in this chapter, can even cause a disc to restructure itself such that it gets to the marginally stable state with $Q \approx 1$ and may even fragment (simulation Mstar1.5-Q1.1-p1.75-0.1IS). Therefore, while in the previous chapter, fragmentation was heavily dependent on the disc temperature and whether the disc is able to cool enough to allow it to fragment, here, while the cooling is clearly important (since some discs modelled with opacities that are $0.1\times$ the interstellar values fragment while the equivalent discs modelled with $0.3\times$ do not), an additional factor to be considered is the restructuring of the discs due to the mass transport resulting from the gravitational torques caused by the high disc masses.

In addition, a number of simulations presented here show that multiple fragments form out of a single spiral arm that is dense enough. This may be characteristic of fragmentation in massive self-gravitating discs. Lodato & Rice (2005) show that massive discs ($M \geq 0.5M_\odot$) are dominated by low order spiral modes. The images of discs presented in this chapter qualitatively agree with this since the spirals are typically dominated by a smaller number of spiral arms than many tightly wound spirals (as is the case with the simulations presented in Chapter 3, see Figure 3.6). Since the densest part of the discs are indeed the spiral arms and it is in the dense regions that fragmentation is likely to occur, it is somewhat unsurprising that multiple fragments form out of the same spiral arm.

Aside from whether such massive discs exist in reality, the issue of growth of fragments into planets needs to be investigated further. Kratter et al. (2010) showed analytically that if fragments

formed, their subsequent evolution would cause them to grow to much larger masses than the observed masses of the HR 8799 planets (for a disc with values of the effective viscosity that are expected for a gravitationally unstable disc). With such high disc masses, it is certainly feasible that the fragments may accrete a significant amount of matter since the Hill radius at such distances is large. In my simulations, the fragment masses are initially approximately a few Jupiter masses but the fragments continue to increase in mass as they accrete from the massive disc. In fact, they grow very rapidly to sizes that are much larger than typical large mass planets and many exceed the deuterium-burning limit of $\approx 13\mathrm{M_J}$ (Spiegel et al. 2010) such that they would be classed as brown dwarfs. This is in agreement with the simulations of Stamatellos & Whitworth (2009) who modelled a system with a high disc to star mass ratio ($M_{\rm disc} = M_\star = 0.7\mathrm{M}_\odot$) and found that many more brown dwarfs formed than planetary mass objects and that if planetary mass objects do form, then they accrete mass to end up with a final mass above the deuterium-burning limit. In addition, the formation and growth of fragments occurs very rapidly (< 1700 yrs), and it is possible that the discs do not last for very long if many fragments form.

The radial movement of the fragments is a further important evolutionary aspect to consider. Given the dynamic activity of the disc once fragments form due to the movement of spiral structures and the resulting mass transport, additional torques may be exerted onto the fragments causing them to move to different radii, particularly when the fragments have just formed and are not too massive. There is evidence of this behaviour here. In addition, such mass transport may also cause the density in other regions to become enhanced and may result in further fragmentation. This is seen in the simulations carried out using sink particles since the spirals in the inner regions become dense, after the formation of fragments in the outer regions, and later fragment.

For the initial simulations that did not use sink particles, the number of fragments that form from the higher opacity simulations are fewer than the number in the low opacity simulations, and these also take more time to form. This is somewhat intuitive (and also expected from the results of Chapter 3) since a lower opacity allows energy to stream out of a disc much more rapidly, thus allowing a disc to cool enough to fragment. However, when sink particles are being considered, this does not necessarily appear to be the case. The sink particles may collectively accrete so much of the disc mass that the cooling due to a lower opacity may not be the dominant process that determines the evolution of the fragments: instead the evolution may be governed by how much disc mass there is left to form fragments and the dynamical interactions between the fragments, i.e. for subsequent fragmentation, the cooling in the disc may not be as critical as in the earlier stages, but this needs to be investigated further.

Additionally, one might expect the maximum mass of a fragment to be inversely proportional to the number of fragments (or equivalently, it is expected that the mass growth rate would decrease with increasing number of fragments), since a larger number of fragments means that more of the gas in the disc is subject to being accreted and hence any one fragment is unlikely to grow too large compared to the other fragments. This is not necessarily the case here (compare simulations Mstar1.5-Q0.7-p1.5-0.3IS and Mstar1.5-Q0.7-p1.5-0.1IS though this may also be an artefact of the different values of the critical density used, see Table 4.2). Also, my simulations show that more massive discs do not necessarily produce the most massive fragments (though they

do seem to produce more fragments). However, a more thorough investigation is required since the simulations have been run for different amounts of time following the fragmentation stage and using different values of the critical density.

Finally, it is important to note that the trends identified here are generally independent of the evolutionary assumptions of the central star mass.

Chapter 5

Fragmentation criteria extended: physical factors affecting the fragmentation boundary

5.1 Introduction

In Chapters 3 and 4, I carried out radiation hydrodynamical simulations to investigate the conditions under which fragmentation may occur in self-gravitating discs, considering different opacity values and taking into account the effects of stellar irradiation. In this chapter, I revisit previous results on the fragmentation criteria of Toomre (1964), Gammie (2001) and Rice et al. (2005). In Section 1.4.2, I review the two quantities that have historically been used to determine whether a disc is likely to fragment. If the stability parameter (Toomre 1964), $Q \lesssim 1$, and if the cooling timescale in units of the orbital timescale, $\beta < \beta_{\rm crit}$, then the disc will fragment. In Section 1.4.2, I also raise the question of whether a single value of $\beta_{\rm crit}$ (and therefore by equation 1.24, a single value of $\alpha_{\rm GI,max}$) indeed exists since the literature suggests that this may not necessarily be the case for all discs.

However, there is a more fundamental question that arises from previous numerical studies using a parameterised cooling method. In many simulations (e.g. Rice et al. 2003a, 2005; Clarke et al. 2007), the fragments all form in the outer parts of the discs. If the fragmentation criterion for a self-gravitating disc only depends on β, then if fragments form, one would expect them to form at all radii since all radii have the same value of β. This implies that the cooling timescale is not the only parameter on which fragmentation depends.

In this chapter, I investigate the criteria for fragmentation in greater detail. In Section 5.2 I analytically investigate how fragmentation may be expected to depend on various disc parameters. I test this analytical theory by carrying out global three-dimensional simulations, the numerical setup of which I describe in Section 5.3. In Section 5.4 I make initial comparisons between my simulations and previous simulations by Rice et al. (2005) as well as discussing the implications of the disc setup. In Section 5.5, I present my simulations, the results of which I describe in Section 5.6. I finally discuss my results in Section 5.7.

5.2 Analytical view

As discussed in Section 1.4.2, for a disc to fragment one criteria is that the Toomre stability parameter, $Q \lesssim 1$. Making the approximation that $\kappa_{\rm ep} \approx \Omega = \sqrt{GM_\star/R^3}$ and using $H = c_{\rm s}/\Omega$, where H is the isothermal scale height of the disc, the Toomre stability criterion becomes a condition on the aspect ratio, H/R:

$$\frac{H}{R} \lesssim \frac{\pi\Sigma R^2}{M_\star}. \tag{5.1}$$

Approximating the surface mass density as $\Sigma \approx M_{\rm disc}/(\pi R^2)$, equation 5.1 becomes (Gammie 2001)

$$\frac{H}{R} \lesssim \frac{M_{\rm disc}}{M_\star}. \tag{5.2}$$

The surface mass density of a disc can also be written in the form of a power-law, $\Sigma = \Sigma_o(R_o/R)^p$, where Σ_o is the surface mass density at radius R_o, and p is a constant for any one disc. Substituting this into equation 5.1, the condition for fragmentation becomes:

$$\frac{H}{R} \lesssim \frac{\pi\Sigma R^2}{M_\star} = \frac{\pi\Sigma_o R_o^p}{GM_\star} R^{(2-p)}. \tag{5.3}$$

Equations 5.2 and 5.3 show the following:

(i) Increasing the disc mass or decreasing the star mass is likely to promote fragmentation since a greater portion of the disc is likely to fulfill the above criteria.

(ii) The surface mass density profile may play a part in the fragmentation of a disc. If $p = 2$, the disc is scale-free and each radius is equivalent to any other radius: the right hand side (RHS) of equation 5.3 is constant with radius and the ratio of the cooling timescale to the orbital timescale, β, is also a constant. Therefore, if the disc settles into a quasi-steady state where the internal heating due to the gravitational instabilities balances the cooling, I expect Q also to be constant with radius (i.e. the left hand side, LHS, of equation 5.3, H/R, is also a constant). Consequently, if $p = 2$, either the entire disc should settle into a quasi-steady state or the entire disc should fragment. I note that in a fragmenting disc, because the heating, cooling and fragmentation timescales are all proportional to the dynamical timescale in the disc, the fragmentation should occur first (in absolute terms) at small radii.

For $p < 2$, the RHS of equation 5.3 increases with radius. Although, H/R is likely to increase with radius as well, since H/R will typically increase more slowly than the RHS of equation 5.3, this condition is more likely to be satisfied in the outer regions of a disc. Conversely, for $p > 2$, the RHS of equation 5.3 decreases with radius and hence the condition is more likely to be satisfied at small disc radii.

(iii) For a $p < 2$ disc with a low enough β such that it can fragment, the exact value of β may determine how much of the disc satisfies condition 5.3. If the cooling in a disc is fast such that β is small, the temperature and hence sound speed, $c_{\rm s}$ will decrease more rapidly than in a disc where β is higher. Consequently, since $H \propto c_{\rm s}$, the aspect ratio will be lower at any particular radius and hence the disc is more likely to satisfy condition 5.3 for smaller β. Since gravitational

instability typically develops on a dynamical timescale, $t_{\rm dyn} \propto 1/\Omega \propto R^{3/2}$, the instability will develop in the inner regions first and therefore fragmentation will first occur as close to the inner regions as possible where the fragmentation criteria are satisfied. Since more of the disc satisfies the above criteria for decreasing β, the radius at which the first fragment forms will also decrease. I therefore expect the fragmentation radius to move inwards with more efficient cooling.

(iv) Crucially, equation 5.3 shows that the radius is important and that for a shallow surface mass density profile ($p < 2$), there does not appear to be a limit for fragmentation (if the disc cools fast enough): provided the disc is large enough, condition 5.3 will be satisfied (since typically, an increase in H/R with radius will be much smaller than the increase in the RHS of equation 5.3).

5.3 Numerical setup

My simulations are carried out using an SPH code originally developed by Benz (1990) and further developed by Bate, Bonnell, & Price (1995) and Price & Bate (2007). It is a Lagrangian hydrodynamics code, ideal for simulations that require a large range of densities to be followed, such as fragmentation scenarios.

I include the heating effects in the disc due to work done on the gas and artificial viscosity to capture shocks. The cooling in the disc is taken into account using the cooling parameter, β (equation 1.14), which cools the gas on a timescale given by equation 1.15.

In order to model shocks, SPH requires artificial viscosity. I use a common form of artificial viscosity by Monaghan & Gingold (1983), which uses the parameters $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$. A corollary of including artificial viscosity is that it adds shear viscosity and causes dissipation. If this viscosity is too large, the evolution of the disc may be driven artificially, while if it is too small, it will lead to inaccurate modelling of shocks (Bate 1995). As discussed in Section 2.1.5, various values of the SPH parameters have been tested and I find that a value of $\alpha_{\rm SPH} \sim 0.1$ provides a good compromise between these factors. Since typically, $\beta_{\rm SPH} \approx 2\alpha_{\rm SPH}$, I choose $\alpha_{\rm SPH}$ and $\beta_{\rm SPH}$ to be 0.1 and 0.2, respectively, which are fixed throughout the simulations. Furthermore, my work will begin with a comparison with Rice et al. (2005) and so I use the same values used by them. I use an adiabatic equation of state with ratio of specific heats, $\gamma = 5/3$.

5.3.1 Numerical effects on fragmentation results

Rice et al. (2005) showed that for a disc with ratio of specific heats, $\gamma = 5/3$, the critical value of the cooling timescale in units of the orbital timescale required for fragmentation is $\beta_{\rm crit} \sim 6 - 7$. The SPH code used for the simulations presented in this thesis differs in the way the smoothing length of the particles is set from that of Rice et al. (2005): whilst their code sets the smoothing length by approximately fixing the number of neighbours that each particle has to ≈ 50, the current version uses a variable smoothing length which does not fix the number of neighbours but allows the smoothing length to be spatially adaptive whilst maintaining energy and entropy conservation (Springel & Hernquist 2002; Monaghan 2002), with my particular implementation described by Price & Bate (2007).

Simulation name	β	p	q	$Q_{\rm min}$	Initial Q profile	Fragments?
Benchmark1	7	1	0.5	2	Decreasing Q	n
Benchmark2	6.5	1	0.5	2	Decreasing Q	n
Benchmark3	6	1	0.5	2	Decreasing Q	n
Benchmark4	5.8	1	0.5	2	Decreasing Q	n
Benchmark5	5.7	1	0.5	2	Decreasing Q	n
Benchmark6	5.6	1	0.5	2	Decreasing Q	n
Benchmark7	5.5	1	0.5	2	Decreasing Q	y
Benchmark8	5	1	0.5	2	Decreasing Q	y
Benchmark9	5	1	-1	2	Flat Q	y
Benchmark10	6	1	-1	2	Flat Q	n
Benchmark11	5	1	-1	1	Flat Q	y

Table 5.1: Summary of the benchmarking simulations described in Section 5.4. p and q are the initial surface mass density and temperature profiles, $\Sigma \propto R^{-p}$ and $T \propto R^{-q}$, respectively. Simulations Benchmark1-8 have been set up in the same way as Rice et al. (2005) whereas simulations Benchmark9-11 have been set up with a uniform Toomre stability profile over the entire disc.

All explicit hydrodynamical simulations must limit the timestep based on the Courant condition. My SPH code also applies a force condition and a viscous timestep condition (see Monaghan 1992 for a review). In the simulations presented here, since I apply an explicit cooling rate, it is important to ensure that the following timestep criteria is also met (equation 2.48):

$$\Delta t \leq C_\beta \frac{\beta}{\Omega}, \tag{5.4}$$

where C_β is a constant less than unity. I investigate the effects of varying the constant C_β on the critical cooling timescale $\beta_{\rm crit}$ by testing values of $C_\beta = 0.3$, 0.03 and 0.003. I find that this does not have a significant effect on the fragmentation results and so I use $C_\beta = 0.3$ for the simulations presented here. However, the timestep criterion may become more important for small β or for particles at small radii. Therefore, for those simulations carried out with small values of β (≤ 3) or where fragmentation occurs at small radii ($\lesssim 5$ AU), the simulations have been repeated with $C_\beta = 0.03$ to confirm that this does not play a part in the results.

5.4 Benchmarking simulations

Table 5.1 summarises the parameters and fragmentation results of the simulations presented here. Each simulation was run either beyond the point at which the disc attained a steady state (for > 6 outer rotation periods, ORPs), or until it fragmented.

The simulations presented by Rice et al. (2005) also used an SPH code. However, since the way the smoothing length is set in my code differs to the code used by Rice et al. (2005), and since it is uncertain as to whether their timestepping considered the cooling timescale, I simulate the same disc that Rice et al. (2005) simulated in order to initially find the critical cooling timescale in units of the orbital timescale, $\beta_{\rm crit}$. This is done by setting up a 1 $M_\odot$ star with a 0.1 $M_\odot$ disc made of 250,000 SPH particles, spanning $0.25 \leq R \leq 25$AU. The initial surface mass density and temperature profiles of the disc are $\Sigma \propto R^{-1}$ and $T \propto R^{-\frac{1}{2}}$, respectively. The magnitudes of

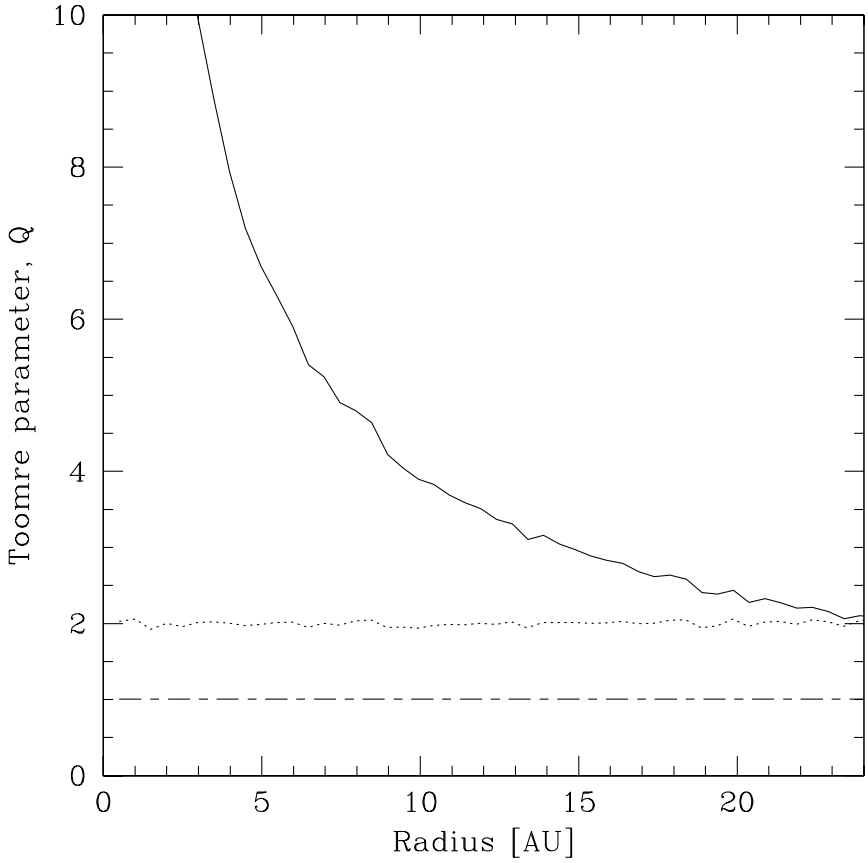

Figure 5.1: Azimuthally averaged values of the Toomre parameter for the initial discs with decreasing Toomre stability profile (simulations Benchmark1-8), set up in the same way as Rice et al. (2005, solid line), and with a flat Q profile with $Q = 2$ (simulations Benchmark9-11; dotted line). The critical value of $Q_{\rm crit} = 1$ is also marked.

these are set such that the Toomre stability parameter (equation 1.7) at the outer edge of the disc, $Q_{\rm min} = 2$. This gives an aspect ratio, $H/R \approx 0.05$. I model the 1 $M_\odot$ star in the centre of the disc using a sink particle (Bate et al. 1995). At the inner disc boundary, particles are accreted onto the star if they move within a radius of 0.025 AU of the star or if they move into $0.025 \leq R < 0.25$AU and are gravitationally bound to the star. At the outer edge, the disc is free to expand.

The simulation was run using a ratio of specific heats, $\gamma = 5/3$ and hence according to Rice et al. (2005), $\beta_{\rm crit} \approx 6 - 7$. I find that the critical value is ≈ 5.6 since this is the lowest value of β that the discs can have without fragmenting (compare simulations Benchmark1-8). According to equation 1.24, this is equivalent to a critical value of the gravitational stress, $\alpha_{\rm GI,max} \sim 0.07$ which is similar to the value of ~ 0.06 obtained by Rice et al. (2005). Given the differences between the codes, I consider this level of agreement acceptable. I therefore compare my remaining simulations to this value of $\beta_{\rm crit}$.

Figure 5.1 shows the initial Toomre stability profile of the Rice et al. (2005) disc (solid line) that is replicated here. As a simulation is started, the disc heats up due to the heating from gravitational instability, the resulting compression and viscous heating, and cools on the cooling

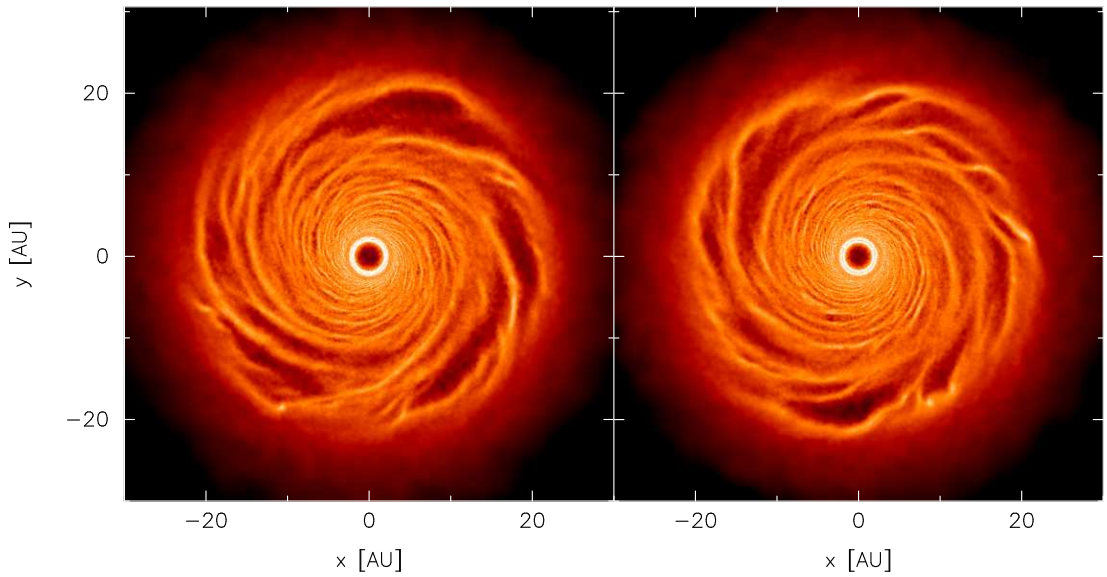

Figure 5.2: Surface mass density rendered image of the first fragments forming in the simulations using a decreasing Toomre stability profile (simulation Benchmark8, left image) and a disc set up with a flat Q profile with $Q = 2$ (simulation Benchmark9, right image). The discs were run with $\beta = 5$. In both cases the discs first fragment at $R_f \approx 20$ AU confirming that the initial temperature profile does not play a part in the evolution of the discs. The colour scale is a logarithmic scale ranging from $\log \Sigma = -7$ (dark) to -3 (light) $M_\odot/AU^2$.

timescale defined by the cooling parameter, β. Consequently, it is expected that the initial disc temperature profile would not play a part in the resulting evolution of the disc. I therefore test this by setting up a disc with the same surface mass density profile, $\Sigma \propto R^{-1}$, but with a temperature profile, $T \propto R$, so that its initial Toomre stability profile is *flat* (i.e. constant over the entire disc) with $Q = 2$ (Figure 5.1; dotted line). These discs were run for $\beta = 5$ and 6. Figure 5.2 shows the images of the evolved disc with decreasing Toomre stability profile and the flat Q disc, run with a cooling time, $\beta = 5$ (simulations Benchmark8 and Benchmark9, respectively). The first fragments form at ≈ 20 AU in both discs irrespective of the initial temperature profile. Figure 5.3 shows the final Toomre stability profiles of both discs run with $\beta = 6$ (simulations Benchmark3 and Benchmark10). Neither of these discs fragment and both discs evolve into a steady-state with very similar Toomre stability profiles. It can be seen that the change in initial temperature profile does not make a difference to the final results since with $\beta = 5$ both discs fragment at the same radius and in the non-fragmenting cases, the final Toomre stability profiles are very similar. Since the temperature in a disc evolves, it is reassuring that the initial temperature profile of the disc does not play a part in the outcome.

As mentioned in Section 1.4.2, current wisdom is that according to fragmentation theory, if the Toomre stability parameter is below unity and the timescale on which the disc cools is faster than a critical value, then the disc should fragment. Therefore, if a disc was set up so that its initial Toomre stability profile was flat with $Q = 1$, it would be expected that fragments would form everywhere in the disc soon after the simulation is started. Figure 5.4 shows the results of this simulation (Benchmark11). It can be seen that despite starting the simulation in a marginally stable state where any part of the disc may fragment soon after the simulation is started, the disc

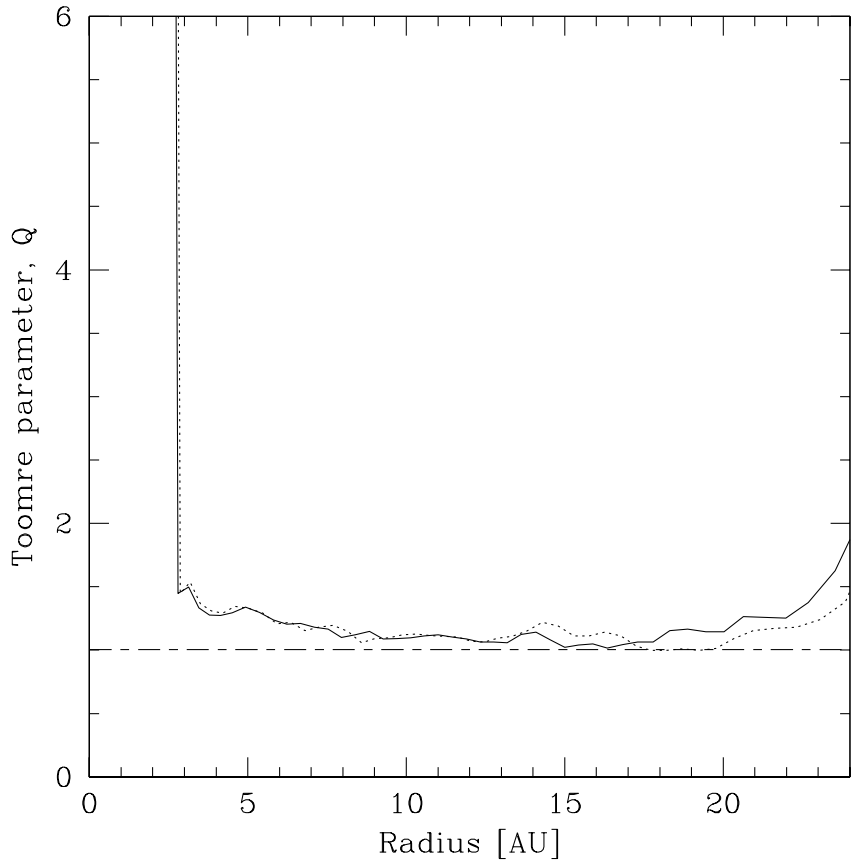

Figure 5.3: Azimuthally averaged values of the Toomre parameter for the discs with initially decreasing (solid line) and flat (dotted line) Toomre stability profiles (simulations Benchmark3 and 10, respectively). The discs were run with $\beta = 6$. Despite having different initial temperature profiles, both discs reach a steady-state with very similar Toomre stability profiles, confirming that the initial temperature does not play a part in the evolution of the discs. The critical value of $Q_{\rm crit} = 1$ is also marked.

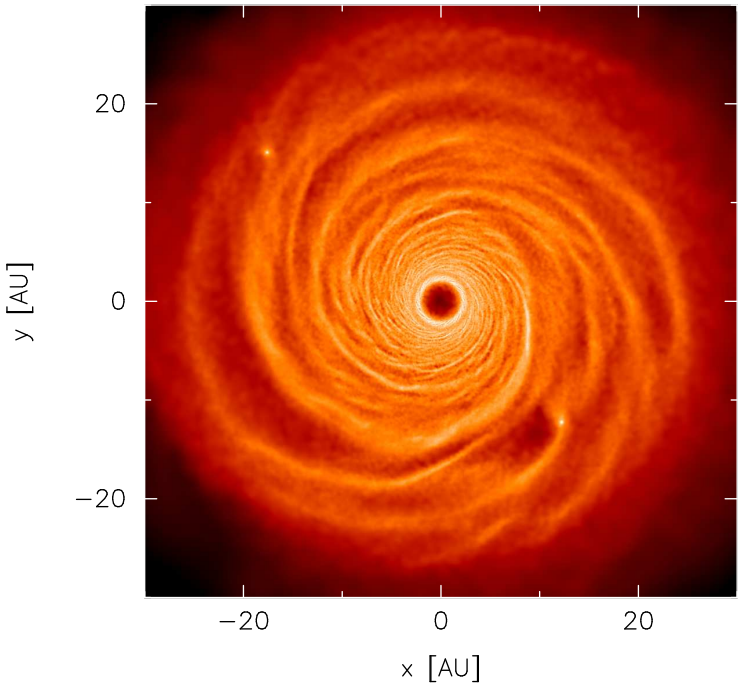

Figure 5.4: Surface density rendered image of the fragmenting disc in simulation Benchmark11 with an initial flat Q profile with $Q = 1$. The simulation was run with $\beta = 5$. Despite the initial disc being in a state of marginal stability such that, in theory, any part of the disc may fragment, the disc only fragments in the outer regions. The colour scale is a logarithmic scale ranging from $\log \Sigma = -8$ (dark) to -3 (light) $\mathrm{M}_{\odot}/\mathrm{AU}^2$.

only fragments in the outer regions. This illustrates that the disc fragmenting in the outer regions cannot be related to the initial value of the Toomre stability profile, Q, and more importantly, fragmentation cannot be a function of β alone.

5.5 Main simulations

In this section, I describe the initial conditions for all the individual numerical simulations I have performed to test my analytical predictions from Section 5.2. Table 5.2 provides a summary of the initial conditions as well as the radius at which the first fragment forms in the discs that do fragment.

I set up a series of *Reference* discs with $M_{\mathrm{disc}} = 0.1\mathrm{M}_{\odot}$ consisting of 250,000 SPH particles, spanning $0.25 \leq R \leq 25\mathrm{AU}$, surrounding a $1\mathrm{M}_{\odot}$ star, modelled using a sink particle. The inner and outer radial disc boundaries have been set up in the same way as the benchmarking discs in Section 5.4. All the discs in this section have been set up with a flat Q profile. Therefore, as the surface mass density profile is varied, the initial temperature profile, q, is also varied accordingly, where $T \propto R^{-q}$. The Reference discs have been set up so that $\Sigma \propto R^{-1}$ and $T \propto R$, normalised so that $Q = 2$. I highlight where I have differed from these initial conditions in the remaining simulations.

The analytical work presented in Section 5.2 suggests that for shallow surface mass density profiles, $p < 2$, fragments would form in the outer regions of the discs, while for discs with

Simulation name	β	p	Σ_o [$M_\odot/(AU)^2$]	M_{disc} [$M_\odot$]	$M_\star$ [$M_\odot$]	$M_{disc}/M_\star$	Q	Disc radius [AU]	R_f [AU]	$\frac{\Sigma R_f^2}{M_\star}$
Reference-beta6	6	1	6.4×10^{-4}	0.1	1	0.1	2	25	-	-
Reference-beta5.5	5.5	1	6.4×10^{-4}	0.1	1	0.1	2	25	22	1.4×10^{-2}
Reference-beta5	5	1	6.4×10^{-4}	0.1	1	0.1	2	25	20	1.3×10^{-2}
Reference-beta4	4	1	6.4×10^{-4}	0.1	1	0.1	2	25	20	1.3×10^{-2}
Reference-beta3	3	1	6.4×10^{-4}	0.1	1	0.1	2	25	8	5.1×10^{-3}
Reference-beta2	2	1	6.4×10^{-4}	0.1	1	0.1	2	25	3	1.9×10^{-3}
Reference-beta1	1	1	6.4×10^{-4}	0.1	1	0.1	2	25	2.5	1.6×10^{-3}
p1.5-beta4	4	1.5	1.8×10^{-3}	0.1	1	0.1	2	25	-	-
p1.5-beta3.5	3.5	1.5	1.8×10^{-3}	0.1	1	0.1	2	25	18	7.5×10^{-3}
p1.5-beta3	3	1.5	1.8×10^{-3}	0.1	1	0.1	2	25	1.7	2.3×10^{-3}
p2-beta3.5	3.5	2	3.5×10^{-3}	0.1	1	0.1	2	25	-	-
p2-beta3	3	2	3.5×10^{-3}	0.1	1	0.1	2	25	0.45	3.5×10^{-3}
p2-beta2	2	2	3.5×10^{-3}	0.1	1	0.1	2	25	0.3	3.5×10^{-3}
p2.5-beta4	4	2.5	4.4×10^{-3}	0.1	1	0.1	2	25	-	-
p2.5-beta3.5	3.5	2.5	4.4×10^{-3}	0.1	1	0.1	2	25	0.4	7.0×10^{-3}
p2.5-beta3.5-Q5	3.5	2.5	4.4×10^{-3}	0.1	1	0.1	5	25	0.3	8.1×10^{-3}
p2.5-beta3	3	2.5	4.4×10^{-3}	0.1	1	0.1	2	25	0.3	8.1×10^{-3}
p2.5-beta2	2	2.5	4.4×10^{-3}	0.1	1	0.1	2	25	0.35	7.5×10^{-3}
p1-Mstar2	5	1	6.4×10^{-4}	0.1	2	0.05	2	25	-	-
p1-Mstar0.5	5	1	6.4×10^{-4}	0.1	0.5	0.2	2	25	13	1.7×10^{-2}
p1-Mdisc0.2	5	1	1.3×10^{-3}	0.2	1	0.2	2	25	14	1.8×10^{-2}
p1-Mdisc0.05	5	1	3.2×10^{-4}	0.05	1	0.05	2	25	-	-
p1-beta1-Mdisc0.01	1	1	6.4×10^{-5}	0.01	1	0.01	2	25	6.5	4.2×10^{-4}
p1-beta2-Mdisc0.01	2	1	6.4×10^{-5}	0.01	1	0.01	2	25	15	9.6×10^{-4}
p1-beta2.5-Mdisc0.01	2.5	1	6.4×10^{-5}	0.01	1	0.01	2	25	17	1.1×10^{-3}
p1-beta3-Mdisc0.01	3	1	6.4×10^{-5}	0.01	1	0.01	2	25	-	-
p1-beta8-Mdisc0.3	8	1	1.9×10^{-3}	0.3	1	0.3	2	25	-	-
p1-beta10-Mdisc0.5	10	1	3.2×10^{-3}	0.5	1	0.5	2	25	-	-
p1-beta5-Mdisc1	5	1	6.4×10^{-3}	1	1	1	2	25	5.5	3.5×10^{-2}
p1-beta7-Mdisc1	7	1	6.4×10^{-3}	1	1	1	2	25	-	-
p1-beta10-Mdisc1	10	1	6.4×10^{-3}	1	1	1	2	25	-	-
p1-beta15-Mdisc1	15	1	6.4×10^{-3}	1	1	1	2	25	-	-
p1-beta6-extended	6	1	6.4×10^{-4}	0.2	1	0.2	2	50	24.5	1.6×10^{-2}
p1-beta7-extended	7	1	6.4×10^{-4}	0.2	1	0.2	2	50	29	1.9×10^{-2}
p1-beta8-extended	8	1	6.4×10^{-4}	0.2	1	0.2	2	50	30	1.9×10^{-2}
p1.5-beta4-extended	4	1.5	1.8×10^{-3}	0.15	1	0.15	2	50	33	1.0×10^{-2}
p1-beta6-Mdisc0.1-extended	6	1	3.2×10^{-4}	0.1	1	0.1	2	50	40	1.3×10^{-2}
p1-beta6-Mdisc0.2-Mstar2-extended	6	1	6.4×10^{-4}	0.2	2	0.1	2	50	34	1.1×10^{-2}

Table 5.2: Summary of the main simulations. p describes the initial surface mass density profile, $\Sigma \propto R^{-p}$, and Σ_o is the normalisation constant required to produce a disc with mass M_{disc}. The final column represents the RHS of equation 5.3 for the location at which the first fragment forms, R_f. The simulations were run with an initial flat Toomre stability profile, Q.

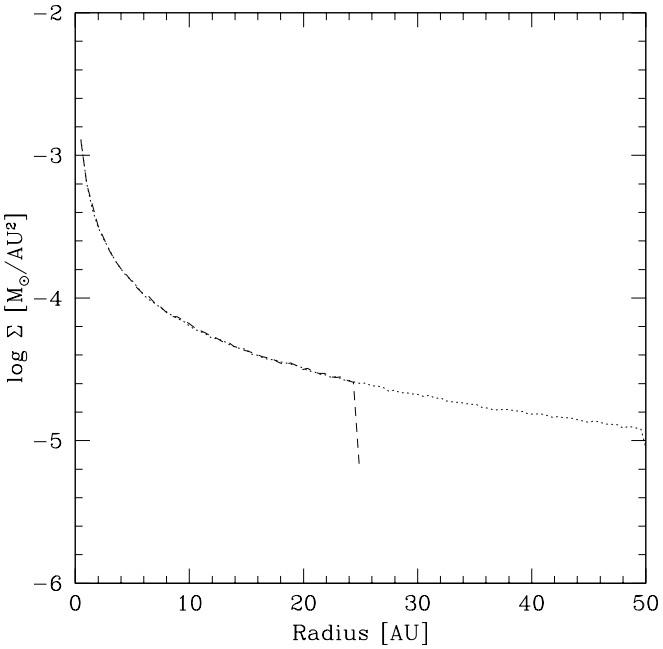

Figure 5.5: Initial surface mass density profiles of the discs used in simulations p1-beta6 (dashed line) and p1-beta6-extended (dotted line). The extended disc has the same surface mass density profile as the smaller disc.

steeper surface mass density profiles, $p \gtrsim 2$, the disc would fragment in the inner regions. I therefore test different values of the slope of the surface mass density profiles using $p = 1$ (simulations Reference-beta5.5 and Reference-beta6), 1.5 (simulations p1.5-beta3.5 and p1.5-beta4), 2.0 (simulations p2-beta2, p2-beta3 and p2-beta3.5) and 2.5 (simulations p2.5-beta2, p2.5-beta3.5 and p2.5-beta4). In addition, I also carry out a further simulation which is the same as simulation p2.5-beta3.5 but with an initial flat Q profile, $Q = 5$ (i.e. so that the initial temperature is 25/4 times hotter than the disc in simulation p2.5-beta3.5), to test the effects of an initially hotter disc on the location of fragmentation (simulation p2.5-beta3.5-Q5).

The analysis also suggests that for a disc with a fast enough cooling timescale such that it would fragment, the location at which the first fragment would form would move inwards to smaller radii as the cooling timescale is decreased. I therefore test the effect of decreasing β on the fragment location by simulating the Reference disc (i.e. with a surface mass density profile, $p = 1$) with different values of the cooling timescale, $\beta = 5.5$, 5, 4, 3, 2 and 1 (simulations Reference-beta5.5, Reference-beta5, Reference-beta4, Reference-beta3, Reference-beta2 and Reference-beta1, respectively).

I also argued that varying the disc or star mass would affect fragmentation. I test the effects of doubling and halving the star mass in simulations p1-Mstar2 and p1-Mstar0.5, respectively, and compare these to the Reference-beta5 simulation. I also carry out extensive tests of the effects of varying the disc mass firstly by doubling and halving the disc mass (simulations p1-Mdisc0.2 and p1-Mdisc0.05, respectively) and secondly by considering more extreme disc masses of $0.01 M_\odot$ (simulations p1-beta0.3-Mdisc0.01, p1-beta1-Mdisc0.01, p1-beta2-Mdisc0.01,

p1-beta2.5-Mdisc0.01 and p1-beta3-Mdisc0.01), $0.3M_\odot$ (simulation p1-beta8-Mdisc0.3), $0.5M_\odot$ (simulation p1-beta10-Mdisc0.5) and $1M_\odot$ (simulations p1-beta5-Mdisc1, p1-beta10-Mdisc1 and p1-beta15-Mdisc1) whilst maintaining a central star mass of $1M_\odot$.

The analytical work presented in Section 5.2 also showed that for a shallow surface mass density profile ($p < 2$), the radius of the disc may be the limiting aspect that causes a disc not to fragment. I therefore test a series of *extended discs* which have outer radii, $R_{\rm out} = 50$AU. Simulations p1-beta6-extended, p1-beta7-extended and p1-beta8-extended are set up so that Σ_o and p are the same as in the Reference discs (Figure 5.5). However, to extend the disc to $R_{\rm out} = 50$AU, the disc masses are increased to $M_{\rm disc} = 0.2M_\odot$. I run this simulation for $\beta = 6$, 7 and 8 (i.e. values that are larger than the critical values identified in Section 5.4). (Note that in order to keep the mass of the individual SPH particles the same as in the Reference simulations, I use 500,000 particles to make up this disc). In addition I also set up an extended disc with a surface mass density profile, $p = 1.5$, which has a disc mass of $1.5M_\odot$ (so that Σ_o is the same as in simulation p1.5-beta4). I run this simulation using $\beta = 4$ (simulation p1.5-beta4-extended). (As before, since I wish to keep the mass of the SPH particles the same as in simulation p1.5-beta4, I use 375,000 particles in this disc.)

Furthermore, I progress the analysis of extended discs by simulating two further discs (using 500,000 particles): the first is the same as that in p1-beta6-extended but using a total disc mass of $M_{\rm disc} = 0.1M_\odot$ (simulation p1-beta6-Mdisc0.1-extended) so that the total disc mass is the same as in p1-beta6 but Σ_o is smaller. The second is also the same as in p1-beta6-extended but the central star mass is also $M_\star = 2M_\odot$ so that the disc to star mass ratio is kept constant (simulation p1-beta6-Mdisc0.2-Mstar2-extended). Both of these discs are run with $\beta = 6$.

5.6 Results

For the analysis that follows, the key aspect about the fragments that will be considered will be the first fragment that forms. This is because subsequent evolution of the disc following an initial fragmentation stage may involve additional complexities that are beyond the scope of this chapter. Table 5.2 summarises the key fragmenting results. The radius at which the first fragments form, $R_{\rm f}$, has been determined by eye from the disc images. It is important to note that as seen in past simulations (e.g. Lodato & Rice 2004, Chapter 3), the surface mass density profile does not change significantly during the simulations, particularly for low mass discs ($M_{\rm disc} \lesssim 0.2M_\odot$). I highlight where the surface mass density profiles do change significantly and discuss the effects of this.

5.6.1 Fragmentation dependency on the surface mass density profile

Figures 5.6-5.9 show that as the surface mass density profile steepens, the location at which the first fragment forms moves to smaller radii in the disc. The analytical theory presented in Section 5.2 shows that for a shallow surface mass density fall off where $p < 2$, the fragments form in the outer regions of the disc (provided the cooling criterion is also satisfied). This is indeed the case for simulations with $p = 1$ and $p = 1.5$ (simulations Reference-beta5.5 and p1.5-beta3.5, respectively) as Figures 5.6 and 5.7 show that the fragments form at $R_{\rm f} \approx 20$ AU and ≈ 19 AU,

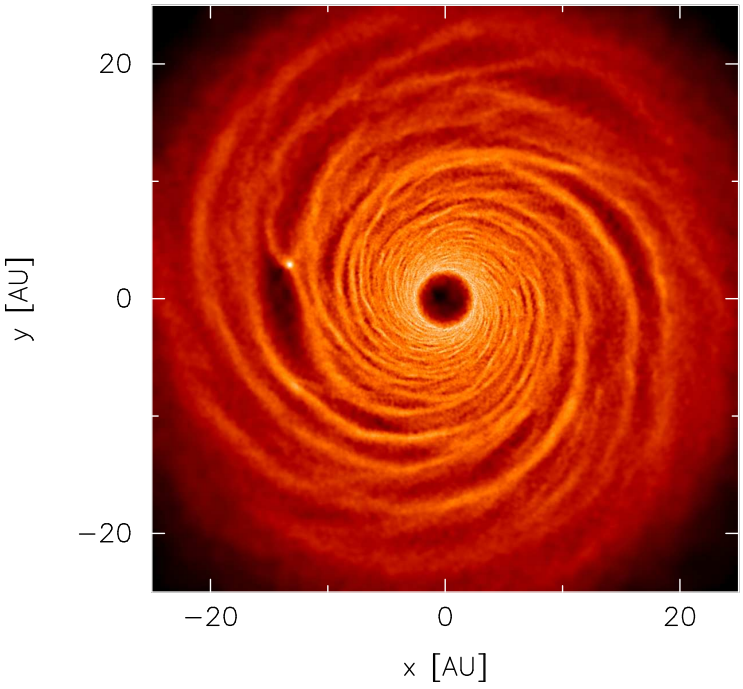

Figure 5.6: Surface mass density rendered image of the fragmenting disc with initial surface mass density profile $\Sigma \propto R^{-1}$. The simulation (Reference-beta5.5) uses $\beta = 5.5$. The fragment forms in the outer regions of the disc, confirming the analytical predictions in Section 5.2. The colour scale is a logarithmic scale ranging from $\log \Sigma = -6$ (dark) to -3 (light) $M_\odot/AU^2$.

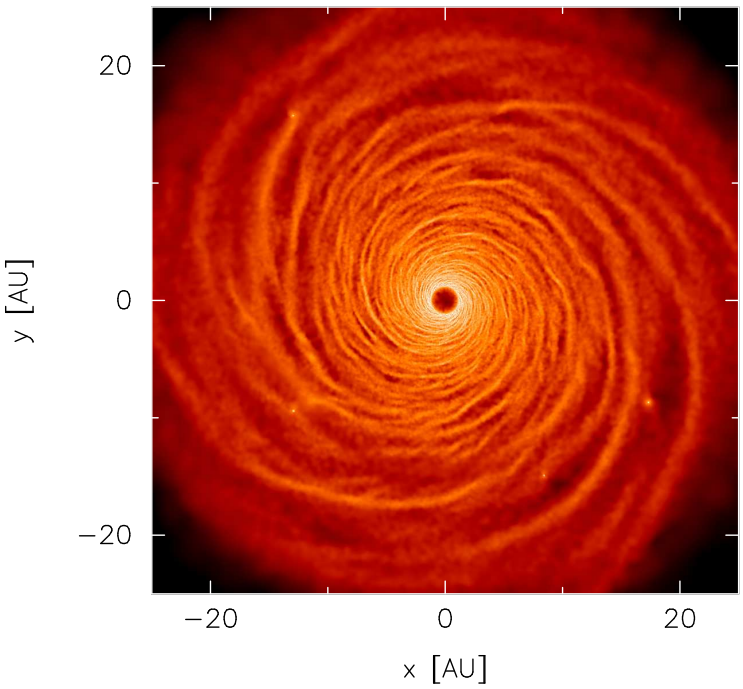

Figure 5.7: Surface mass density rendered image of the fragmenting disc with initial surface mass density profile $\Sigma \propto R^{-3/2}$. The simulation (p1.5-beta3.5) uses $\beta = 3.5$. The fragment forms in the outer regions of the disc, confirming the analytical predictions in Section 5.2. The colour scale is a logarithmic scale ranging from $\log \Sigma = -7$ (dark) to -2 (light) $M_\odot/AU^2$.

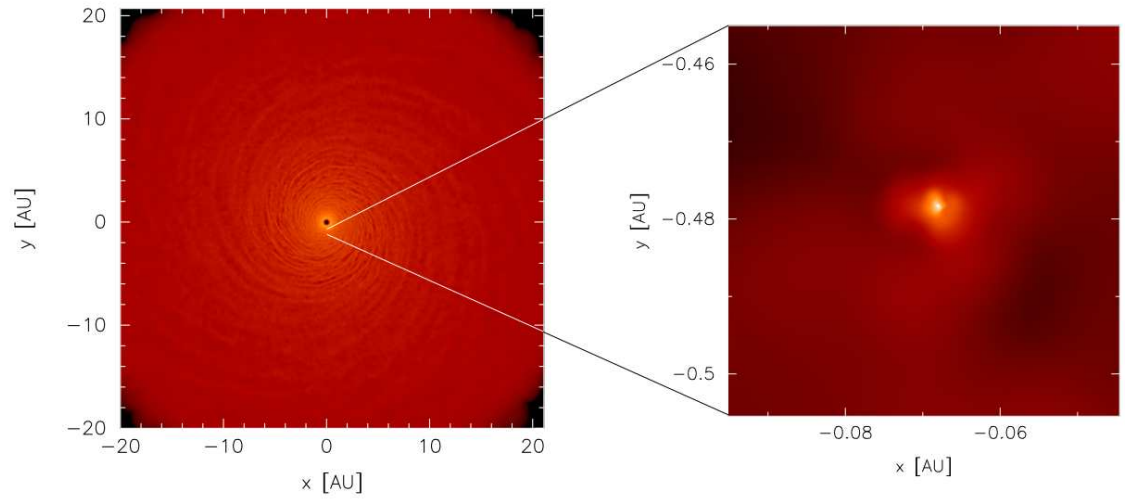

Figure 5.8: Surface mass density rendered image of the fragmenting disc with initial surface mass density profile $\Sigma \propto R^{-2}$ (simulation p2-beta3). The simulation used $\beta = 3$. The fragment forms in the inner regions of the disc as shown by the zoomed in image of the disc, confirming the analytical predictions in Section 5.2. The colour scale is a logarithmic scale ranging from $\log \Sigma = -11$ (dark) to 2 (light) $M_\odot/AU^2$ in the zoomed out image and from $\log \Sigma = -3.5$ (dark) to 1 (light) $M_\odot/AU^2$ in the zoomed in image.

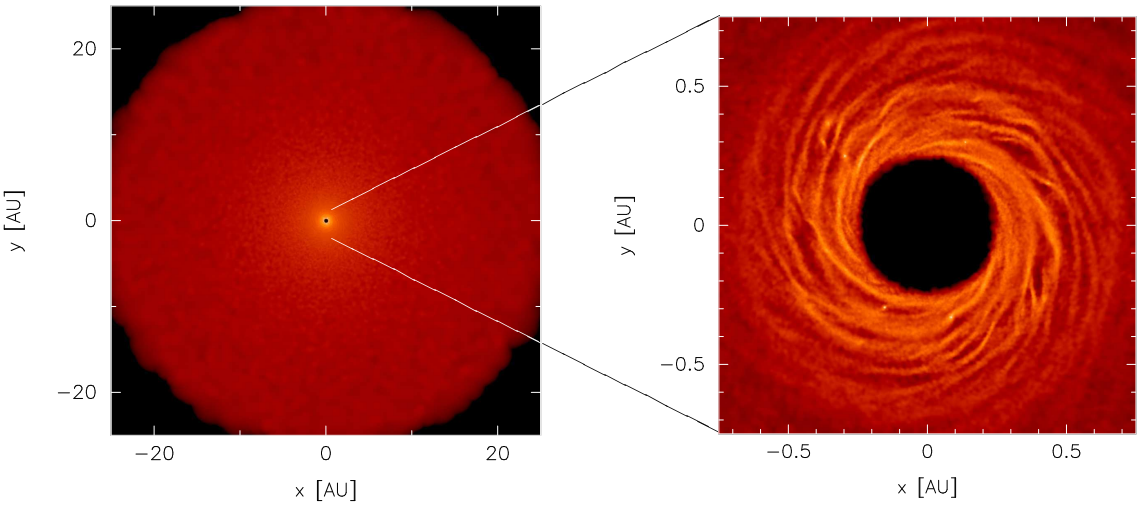

Figure 5.9: Surface mass density rendered image of the fragmenting disc with initial surface mass density profile $\Sigma \propto R^{-5/2}$ (simulation p2.5-beta3.5). The simulation used $\beta = 3.5$. The fragment forms in the inner regions of the disc, confirming the analytical predictions in Section 5.2. The colour scale is a logarithmic scale ranging from $\log \Sigma = -12$ (dark) to 3 (light) $M_\odot/AU^2$ in the zoomed out image and from $\log \Sigma = -4$ (dark) to 0.4 (light) $M_\odot/AU^2$ in the zoomed in image.

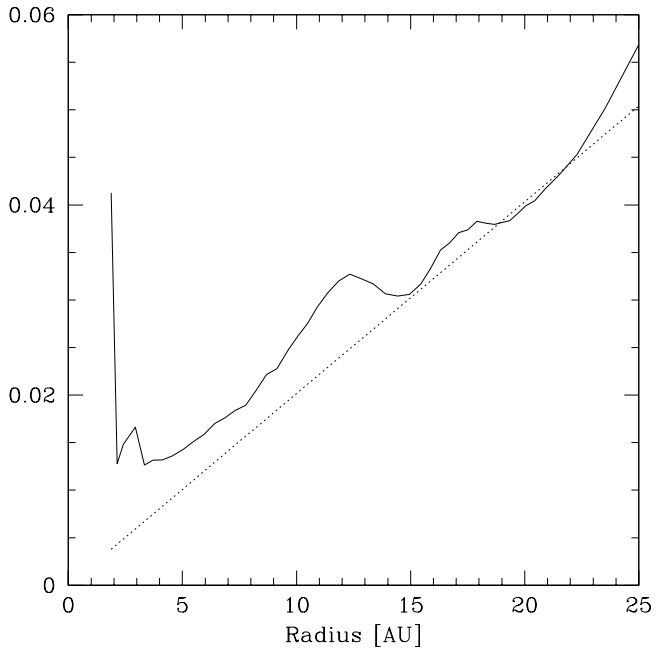

Figure 5.10: Plot of disc aspect ratio, H/R (solid line), against the RHS of equation 5.3 (dotted line) for simulation Reference-beta5.5. Condition 5.3 is satisfied at $\approx$ 20AU where the disc first fragments, confirming the analytical predictions in Section 5.2.

respectively. Figures 5.10 and 5.11 show the radial profile of the aspect ratios (calculated by azimuthally averaging the sound speed at each radii and dividing by ΩR at that radii) in the discs compared to the RHS of equation 5.3 at a time shortly before the discs fragment. It can be seen that condition 5.3 is satisfied at the region in which the first fragment forms shortly after. The oscillations in H/R are due to temperature fluctuations since although the cooling rate in the disc changes smoothly with radius, the heating of the disc occurs primarily in the spiral shocks. This therefore confirms the analytical predictions for shallow surface mass density profiles ($p < 2$) presented in Section 5.2. It is important to note that for a flat Q profile, the temperature profiles in the discs are an increasing function of radius ($T \propto R$) for $p = 1$ and a constant temperature profile for $p = 1.5$, yet the discs still fragment in the outer regions (again, re-emphasising that the initial temperature profile does not play a part in the disc evolution).

The analytical theory for $p \gtrsim 2$ suggested that if the disc was to fragment, it would do so in the inner regions of the disc. Figures 5.8 and 5.9 show that this is indeed the case. Figure 5.12 shows that the analytical condition is just satisfied for simulation p2-beta3 at the location at which the fragment forms. Figure 5.13 shows that for simulation p2.5-beta3.5, the analytical condition is also satisfied at the location at which fragmentation occurs soon after.

I therefore show that as the surface mass density profile is steepened so that more of the mass is concentrated in the inner regions of the disc, fragmentation moves towards smaller radii. It is important to note that the trend that fragmentation moves to smaller radii for steeper surface mass density profiles is valid even when considering a uniform value of β (compare simulations Reference-beta3, p1.5-beta3, p2-beta3 and p2.5-beta3 which are run using $\beta = 3$ and fragment at

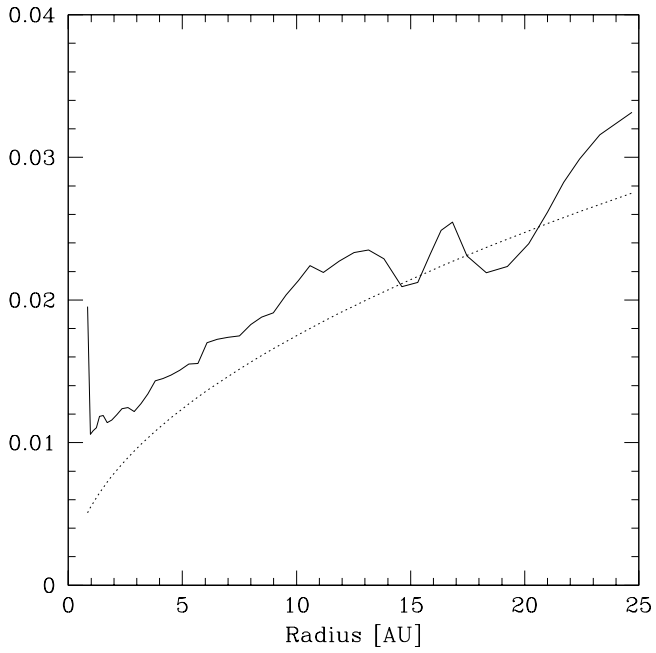

Figure 5.11: Plot of disc aspect ratio, H/R (solid line), against the RHS of equation 5.3 (dotted line) for simulation p1.5-beta3.5. Condition 5.3 is satisfied at $\approx$ 19AU where the disc first fragments, confirming the analytical predictions in Section 5.2.

$\approx$ 8, 1.7, 0.45 and 0.3 AU, respectively).

In addition, the results summarised in Table 5.2 show that a single value of $\beta_{\rm crit}$ is not applicable over all surface mass density profiles since the minimum value of β that a disc can have without fragmenting varies with the surface mass density profile.

Simulation p2.5-beta3.5-Q5 was the same as simulation p2.5-beta3.5 but had an initial disc that was hotter by a factor of 25/4. The results of this simulation show that the disc still fragmented in the inner regions.

5.6.2 Effect of the cooling timescale, β, on the fragment location

In Section 5.2 the analytical work presented suggested that for a fragmenting disc with a shallow surface mass density profile ($p < 2$), a decrease in the value of β would cause the location at which the first fragment forms to move inwards to smaller radii.

Figure 5.14 shows the radius at which the first fragment forms for different values of β (simulations Reference-beta5.5, Reference-beta5, Reference-beta4, Reference-beta3, Reference-beta2 and Reference-beta1). This shows a clear trend that the radius of fragmentation moves inwards for more efficient cooling.

5.6.3 The influence of star mass on fragmentation

In Section 5.2, I showed that decreasing the star mass is more likely to cause conditions 5.2 and 5.3 to be satisfied over a larger part of the disc and hence the disc is more likely to fragment. I test three

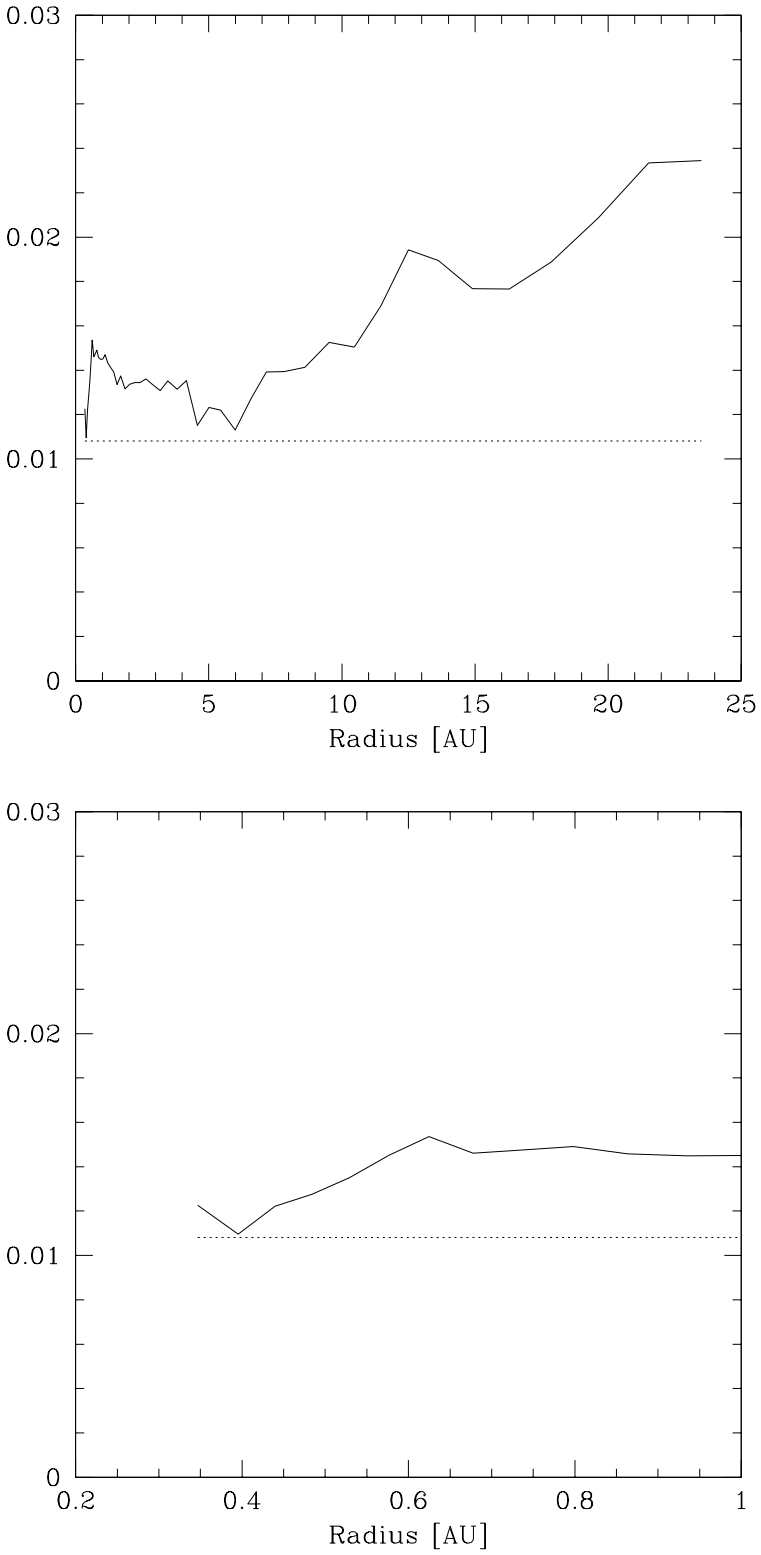

Figure 5.12: Plot of disc aspect ratio, H/R (solid line), against the RHS of equation 5.3 (dotted line) for simulation p2-beta3 for the radial range of the entire disc (upper panel) as well as zoomed into the inner regions (lower panel). Condition 5.3 is marginally satisfied at $\approx$ 0.4AU where the disc first fragments, confirming the analytical predictions in Section 5.2.

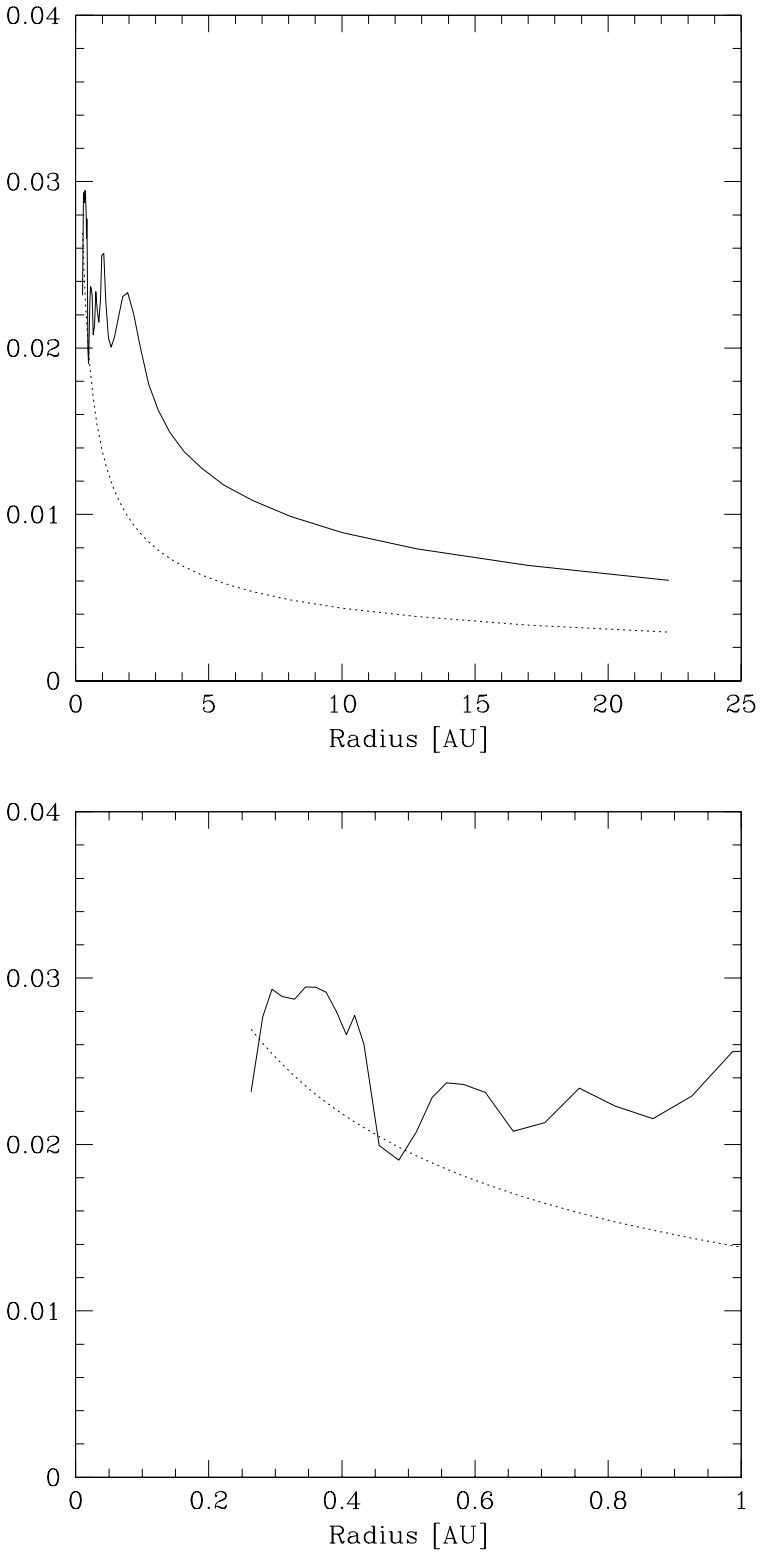

Figure 5.13: Plot of disc aspect ratio, H/R (solid line), against the RHS of equation 5.3 (dotted line) for simulation p2.5-beta3.5 for the radial range of the entire disc (upper panel) as well as zoomed into the inner regions (lower panel). Condition 5.3 is satisfied at ≈ 0.4AU where the disc first fragments, confirming the analytical predictions in Section 5.2.

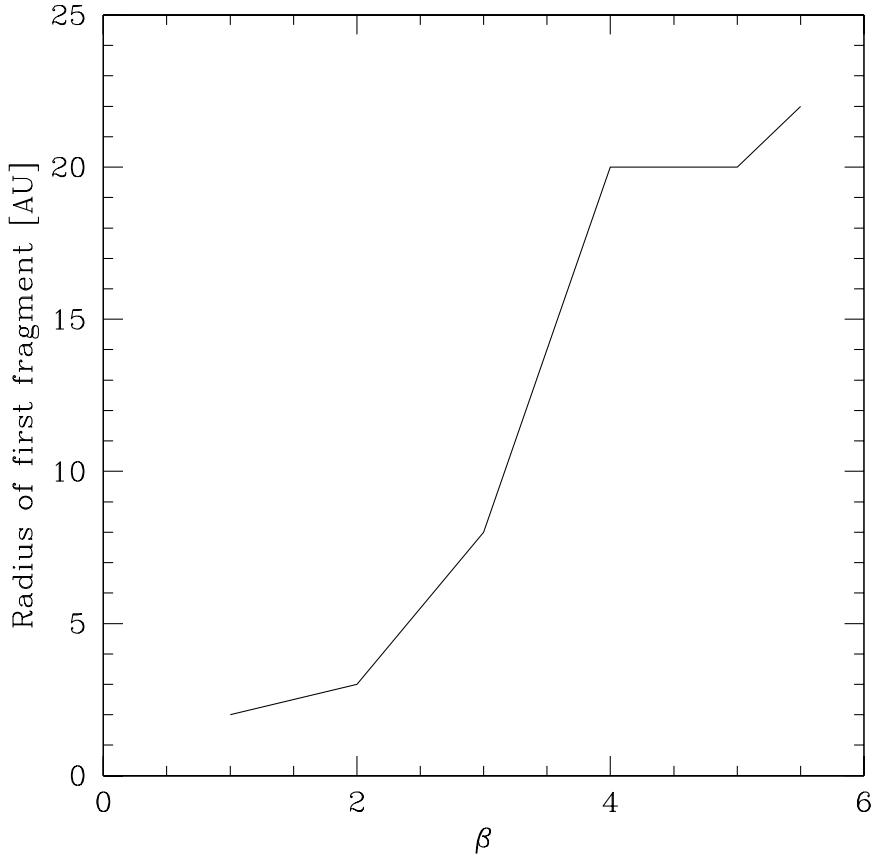

Figure 5.14: The radius at which the first fragment forms in the Reference simulations. The discs in these simulations are identical with a surface mass density profile, $p = 1$, but were run with different values of the cooling timescale in units of the orbital timescale, β. The radius at which the first fragment forms moves inwards with more efficient cooling.

identical discs with star masses of 0.5, 1 and 2 $M_\odot$ (simulations p1-Mstar0.5, Reference-beta5 and p1-Mstar2, respectively) which are run with the same cooling timescale, $\beta = 5$. It can be seen from Table 5.2 that the Reference-beta5 disc first fragments at $R_f \approx 20$ AU. However, when the star mass is halved, the first fragment forms at $R_f \approx 13$ AU. Since the RHS of condition 5.3 $\propto R^2/M_\star$, if the star mass is halved for the same value of β (and hence the same value of the LHS of condition 5.3), the radius at which the first fragment forms, R_f, decreases by a factor of $\sqrt{2}$. Conversely, doubling the star mass makes it harder for the condition to be satisfied and indeed the disc in simulation p1-Mstar2 does not fragment. Instead it settles into a state of marginal stability with $Q \approx 1$.

5.6.4 The influence of disc mass on fragmentation

The analytics presented in Section 5.2 showed that increasing the disc mass (and hence increasing Σ_o) allows conditions 5.2 and 5.3 to be satisfied over a larger part of the disc and hence the disc is more likely to fragment. I initially test this by comparing the results of simulations p1-Mdisc0.05, Reference-beta5 and p1-Mdisc0.2 which are identical discs except that the disc

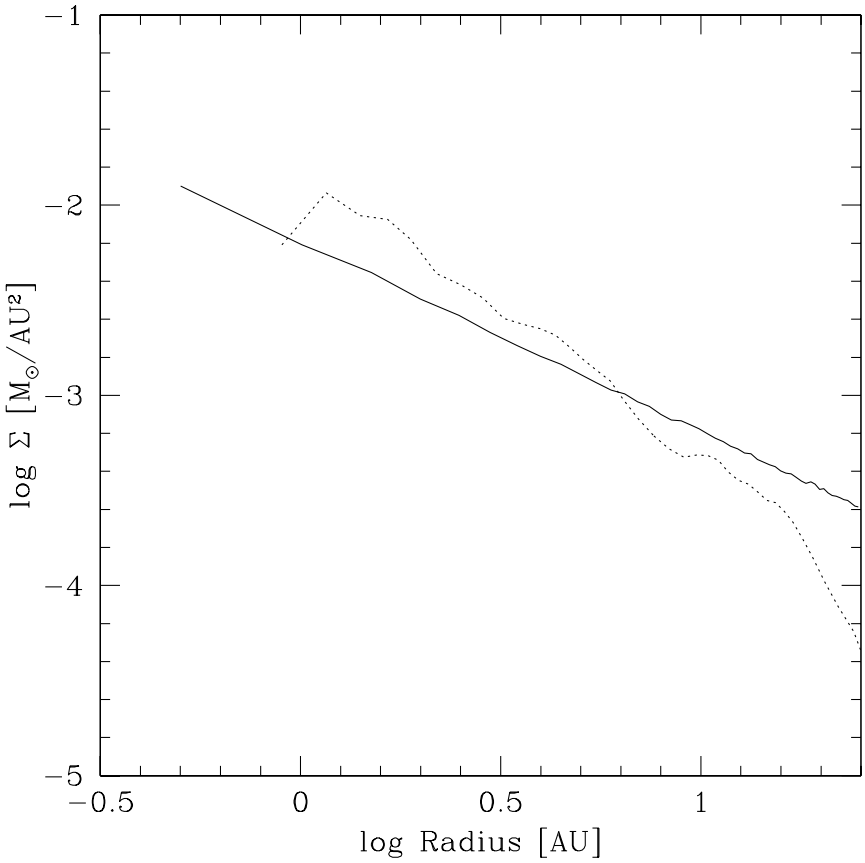

Figure 5.15: Surface mass density profiles for simulation p1-beta7-Mdisc1 at the start (solid line) and at a time more than 4 ORPs later (dotted line). Unlike the low mass simulations whose surface mass density profiles do not change throughout the simulations, the profile for this disc steepens causing a change in the effective values of Σ_o and p.

masses are $0.05M_\odot$, $0.1M_\odot$ and $0.2M_\odot$, respectively. Table 5.2 shows that doubling the disc mass from $0.1M_\odot$ to $0.2M_\odot$ does indeed cause the fragmentation conditions derived in Section 5.2 to be satisfied over a larger portion of the disc since the first fragments form at $R_f \approx 20$ AU and ≈ 14 AU, respectively. Since the RHS of condition 5.3 $\propto \Sigma R^2$, if the disc mass (and hence Σ) is doubled for the same value of β (and hence the same value of the LHS of condition 5.3), the radius at which the first fragment forms, R_f, decreases by a factor of $\sqrt{2}$. However, halving the disc mass makes it harder for the conditions to be satisfied and consequently, the disc does not fragment.

In addition, I also simulate a very low mass disc ($M_{\rm disc} = 0.01M_\odot$) and found that it fragments if $\beta = 1$, 2 and 2.5 but not for $\beta = 3$ (simulations p1-beta1-Mdisc0.01, p1-beta2-Mdisc0.01, p1-beta2.5-Mdisc0.01 and p1-beta3-Mdisc0.01, respectively). As β increases, the fragment location moves out in the disc, as found in Section 5.6.2. It is clear that, as with varying the star mass, the disc mass plays a crucial role in the fragmentation and the condition for fragmentation cannot simply be described using a single critical value of the cooling timescale.

In addition, I simulate higher mass discs, $M_{\rm disc} = 0.3$ and $0.5M_\odot$, which are run using $\beta = 8$ and 10 respectively (simulations p1-beta8-Mdisc0.3 and p1-beta10-Mdisc0.5) as well as discs

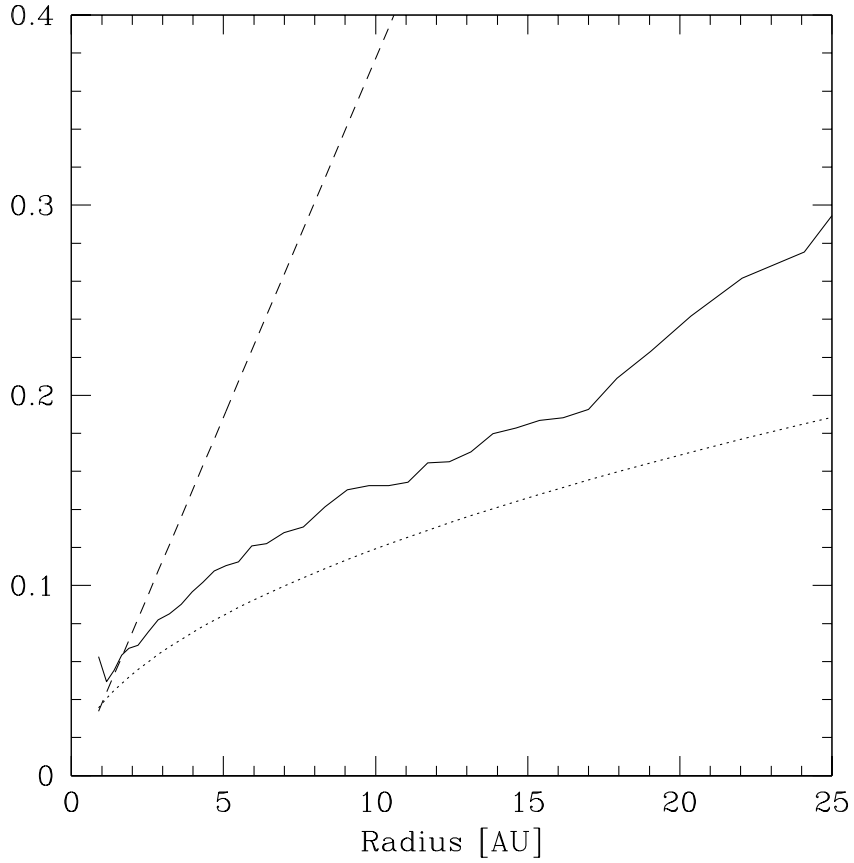

Figure 5.16: Plot of disc aspect ratio, H/R (solid line), for simulation p1-beta7-Mdisc1, against the RHS of equation 5.3 using the initial values of Σ_o and p (dashed line) and the new values of Σ_o and p (dotted line) determined after the disc has evolved for > 4ORPs by which time its surface mass density profile has changed. The condition is satisfied using the initial values of Σ_o and p but not using the new values and hence the disc does not fragment.

with $M_{\rm disc} = 1{\rm M}_\odot$ which I simulate using $\beta = 5$, 10 and 15 (simulations p1-beta5-Mdisc1, p1-beta10-Mdisc1 and p1-beta15-Mdisc1, respectively). I find that with the exception of simulation p1-beta5-Mdisc1, the discs do not fragment. Figure 5.15 shows the surface mass density profile of simulation p1-beta7-Mdisc1 at the start and more than 4 ORPs after the start of the simulation. It can be seen that unlike the lower mass discs, the profile steepens and the value of Σ_o increases. This is the case for all the non-fragmenting high mass disc simulations. Figure 5.16 shows the plot of the aspect ratio profile of this simulation against the RHS of condition 5.3 and shows that condition 5.3 is just satisfied in the inner regions. However, since during the simulation the disc mass redistributes itself, the surface mass density profile changes and consequently, using the newly obtained values of Σ_o and p, condition 5.3 is not satisfied. The only high mass disc that does fragment (simulation p1-beta5-Mdisc1), does so because the cooling time is so rapid that fragmentation occurs before the disc has had the chance to restructure itself.

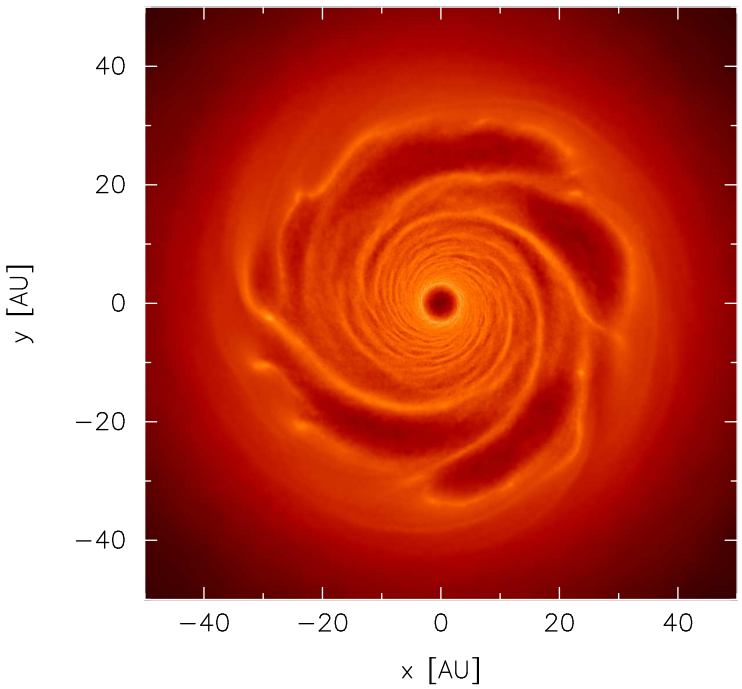

Figure 5.17: Surface mass density rendered image of the fragmenting disc in simulation p1-beta8-extended with initial surface mass density profile $\Sigma \propto R^{-1}$, but extending to 50 AU rather than 25 AU. This simulation was run with $\beta = 8$. According to Rice et al. (2005), this disc should not fragment since the cooling timescale, β, is larger than the critical value previously obtained with a radius of 25 AU. This simulation shows that the fragmentation criterion is more complex than a single critical cooling parameter. The colour scale is a logarithmic scale ranging from $\log \Sigma = -8$ (dark) to -2 (light) $M_\odot/AU^2$.

5.6.5 The role of the disc radius on fragmentation

In Section 5.2, I showed that for shallow surface mass density profiles ($p < 2$), fragmentation might occur for *any* value of the cooling timescale, if the disc is large enough.

Simulation Reference-beta6 (a disc with $R_{\rm out} = 25$ AU) did not fragment and though I did not run the same simulation with $\beta = 7$ or 8, I would expect that they would also not fragment. However, extended discs with the same values of Σ_o and p as Reference-beta6 do indeed fragment for $\beta = 6$ (simulation p1-beta6-extended), $\beta = 7$ (p1-beta7-extended) and $\beta = 8$ (p1-beta8-extended; Figure 5.17) with the first fragments forming at $R_{\rm f} \approx 25$, 29 and 30 AU, respectively. Similarly, I also simulate an extended disc with $p = 1.5$ and $\beta = 4$ (simulation p1.5-beta4-extended) and show that while the same disc truncated at $R_{\rm out} = 25$AU does not fragment, the extended disc does indeed fragment (at $R_{\rm f} \approx 33$AU).

In addition, given that in Section 5.6.4 I showed that the disc mass plays a part in whether fragmentation occurs or not, I simulate a $0.1 M_\odot$ disc which extends to $R_{\rm out} = 50$AU (simulation p1-beta6-Mdisc0.1-extended). The surface mass density profile is the same as in simulation p1-beta6-extended ($p = 1$) but Σ_o is decreased. The results show that the disc fragments at $R \approx 40$AU (c.f. ≈ 25 AU for simulation p1-beta6-extended). Therefore, while the disc mass affects where in the disc the first fragment forms, the conclusion that a small disc which does not fragment for a particular value of β may fragment at larger radii for the same value of β is still valid. Furthermore,

I can see that if the disc to star mass ratio is kept constant at $M_{\rm disc}/M_\star = 0.1$ for an extended disc, the disc fragments (at $R_{\rm f} \approx 34$ AU; simulation p1-beta6-Mdisc0.2-Mstar2-extended) while the small disc with the same disc to star mass ratio does not fragment, further corroborates the fact that the radius of the disc is important.

5.7 Discussion

It has previously been accepted that for a self-gravitating disc whose only source of heating is internally generated from the gravitational instability, the disc will fragment if the cooling timescale is short enough (Gammie 2001). However, I find that fragmentation at a given radius is not only dependent on the cooling timescale, β, but also on the disc surface density (i.e. disc mass and profile) and the star mass.

This is in contrast to Rice et al. (2005) who suggested that the fragmentation criterion is independent of the disc mass though in agreement with Rice et al. (2003a) who found that for a higher disc mass, fragmentation was easier: using $\beta = 5$, they found that a $0.25{\rm M}_\odot$ disc fragmented while a $0.1{\rm M}_\odot$ disc did not, but instead required a lower value of β. In particular, I highlight that in the past, it has been thought that a massive disc is required for fragmentation to occur. However, I show that it is indeed possible for low mass discs ($M_{\rm disc} \sim {\rm O}(0.01){\rm M}_\odot$) to fragment if the cooling in the discs is rapid enough. On the other hand, for high mass discs ($M_{\rm disc} \geq 0.3{\rm M}_\odot$ within $R_{\rm out} = 25$ AU), the discs do not fragment, unless the cooling time is fast, due to a steepening of the surface mass density profile and an increase in Σ_o making condition 5.3 harder to satisfy. Lodato & Rice (2005) also find that with larger disc masses, matter is redistributed causing the surface mass density profile to steepen.

In addition, I find that the critical value of β found for one particular surface mass density profile is not applicable to another disc mass distribution. My simulations show that for a steeper surface mass density profile, the cooling timescale required for a disc to fragment is smaller. Cossins et al. (2010) found that for a disc with $\Sigma \propto R^{-3/2}$, $\beta_{\rm crit} \sim 4-4.5$ whereas Rice et al. (2005) found that $\beta_{\rm crit} \sim 6-7$ for a disc with $\Sigma \propto R^{-1}$. In addition, Rice et al. (2003a) found that for a surface mass density profile of $\Sigma \propto R^{-7/4}$, the fragmentation boundary for a $0.1{\rm M}_\odot$ disc around a $1{\rm M}_\odot$ star is between $\beta = 3$ and 5 which Rice et al. (2005) note is inconsistent with their results. However, my results explain these previous inconsistencies present in the literature.

I also find that for $p < 2$, if a disc does not fragment for a particular cooling timescale in units of the orbital timescale, β, a larger disc with the same surface mass density profile may well fragment (compare simulations Reference-beta6 and p1-beta6-extended or p1.5-beta4 and p1.5-beta4-extended, and also p1-beta7-extended or p1-beta8-extended). Therefore, a critical cooling timescale can only be specified for a particular surface mass density at a particular disc radius and can therefore not be a general rule. The previous fragmentation criterion found by Rice et al. (2005) was for a disc to star mass ratio of 0.1 with $M_\star = 1{\rm M}_\odot$, a surface mass density profile of $\Sigma \propto R^{-1}$ and $R_{\rm out} = 25$ AU.

Finally, I also show that the mass of the star plays a part in whether a disc fragments or not. Therefore, as shown by the RHS of condition 5.3, whether a disc fragments or not is essentially a

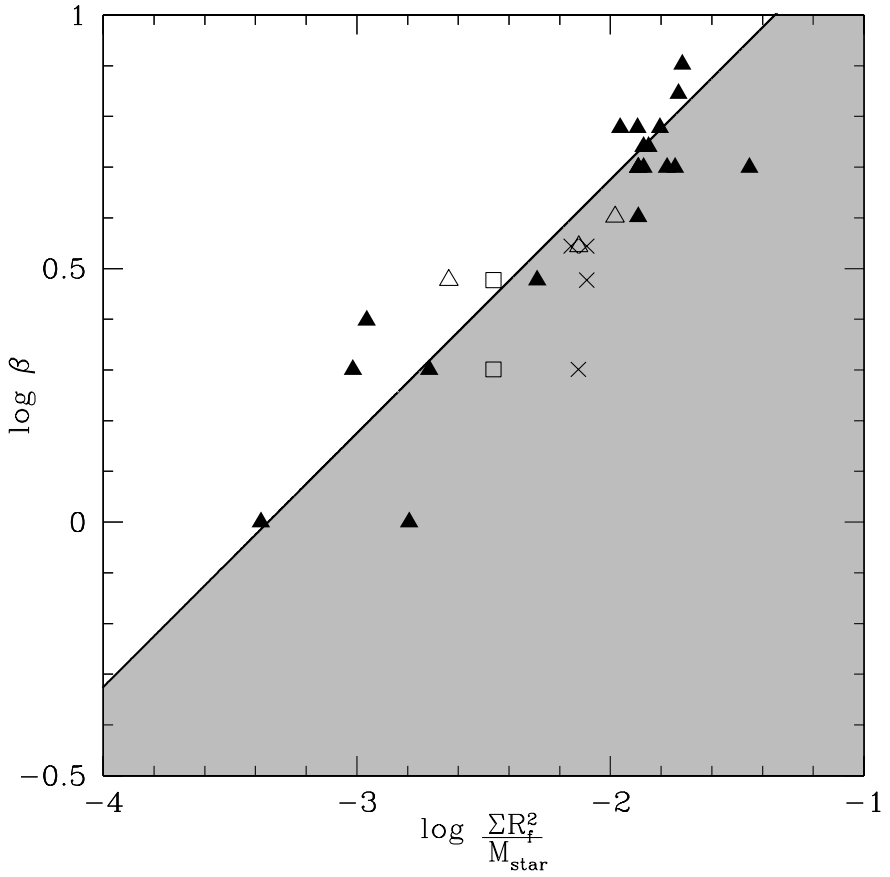

Figure 5.18: Logarithmic graph showing the trend between β and $\Sigma R_{\rm f}^2/M_\star$ determined by considering the location at which the first fragment forms in the discs, $R_{\rm f}$. The results include those simulations with a surface mass density profile, $p = 1$ (filled triangles), $p = 1.5$ (open triangles), $p = 2$ (open squares) and $p = 2.5$ (crosses). It is clear that a single critical value of β is not the case for all discs and that there is a relation between β, $M_{\rm disc}$, $M_\star$ and the surface mass density profile, p, that determines whether fragmentation occurs or not. The trendline has been determined by considering discs with shallow surface mass density profiles, $p < 2$ only as those discs with $p \gtrsim 2$ will always fragment in the innermost regions first. The grey shaded region is where I expect subsequent fragmentation may take place in discs with $p < 2$.

"trade-off" between the local surface mass density and the star mass.

5.7.1 The link between β, $M_{\rm disc}$, $M_\star$, the surface mass density profile, p, and fragmentation

Section 5.6 shows that the fragmentation criterion is clearly a complex problem which cannot simply depend on a single critical cooling timescale as has previously been thought to be the case. Equation 5.3 and the results presented here clearly show that there is a link between the cooling timescale (in terms of β), the disc mass (or more accurately, the local surface mass density) and the star mass.

Such a link can be explained physically. As a disc cools, gravitational instability develops resulting in density fluctuations above and below the unperturbed density, $\delta\rho/\rho$. The spiral

Simulation name	β	p	Σ_o [$M_\odot/(AU)^2$]	M_{disc} [$M_\odot$]	$M_\star$ [$M_\odot$]	$M_{disc}/M_\star$	Q_{min}	Disc radius [AU]	R_f [AU]	$\frac{\Sigma R_f^2}{M_\star}$
2m-beta5.5	5.5	1	6.4×10^{-4}	0.1	1	0.1	2	25	13	8.4×10^{-3}
2m-beta6	6	1	6.4×10^{-4}	0.1	1	0.1	2	25	15	9.6×10^{-3}
2m-beta5.5-Mdisc0.2	5.5	1	1.3×10^{-3}	0.2	1	0.2	2	25	7	9.0×10^{-3}
2m-beta6-Mdisc0.2	6	1	1.3×10^{-3}	0.2	1	0.2	2	25	7	9.0×10^{-3}
2m-beta6-Mstar2	6	1	6.4×10^{-4}	0.1	2	0.05	2	25	-	-

Table 5.3: Summary of the higher resolution simulations with 2×10^6 SPH particles. p describes the initial surface mass density profile, $\Sigma \propto R^{-p}$, and Σ_o is the normalisation constant required to produce a disc with mass M_{disc}. The final column represents the RHS of equation 5.3 for the location at which the first fragment forms, R_f. The simulations were run with an initially decreasing Toomre stability profile, Q, with $Q \propto R^{-3/4}$ and minimum value at the outer edge, Q_{min}.

Simulation name	β	p	Σ_o [$M_\odot/(AU)^2$]	M_{disc} [$M_\odot$]	$M_\star$ [$M_\odot$]	$M_{disc}/M_\star$	Q	Disc radius [AU]	R_f [AU]	$\frac{\Sigma R_f^2}{M_\star}$
F-beta8	8	1	6.4×10^{-4}	0.1	1	0.1	2	25	-	-
F-beta5	5	1	6.4×10^{-4}	0.1	1	0.1	2	25	-	-
F-beta5-Mdisc0.2	5	1	1.3×10^{-3}	0.2	1	0.2	2	25	9.7	1.2×10^{-2}
F-beta4	4	1	6.4×10^{-4}	0.1	1	0.1	2	25	17	1.1×10^{-2}
F-beta4-Mdisc0.05	4	1	3.2×10^{-4}	0.05	1	0.05	2	25	-	-
F-beta4-Mdisc0.2	4	1	1.3×10^{-3}	0.2	1	0.2	2	25	7.2	9.0×10^{-3}

Table 5.4: Summary of the preliminary simulations carried out using the 2-dimensional grid-based polar hydrodynamics code FARGO. p describes the initial surface mass density profile, $\Sigma \propto R^{-p}$, and Σ_o is the normalisation constant required to produce a disc with mass M_{disc}. The final column represents the RHS of equation 5.3 for the location at which the first fragment forms, R_f. (Note that since the outputs for these particular runs were very infrequent, the values of R_f are indicative and may not be accurate.) The simulations were run with an initially flat Toomre stability profile, Q.

structures involve shocks which produce heat that may balance the disc's cooling, thus reaching an equilibrium state. If the disc mass was irrelevant, for any particular star mass, and β, one would expect the fluctuations, $\delta\rho/\rho$, to be the same in all discs with the same surface density profile. However, comparing a low mass disc with a high mass disc with the same relative density fluctuations ($\delta\rho/\rho$), the density enhancement, $\delta\rho$, in the higher-mass disc will clearly be greater. At some disc mass, this enhancement will be self-gravitating (i.e. it will become a fragment), while in a lower-mass disc the density fluctuation will not form a fragment (unless the value of β is lowered). Similarly, if the disc is kept the same and the star's mass is increased, a given density enhancement may be sheared apart by the differential rotation so that a low value of β will be required for fragmentation.

Figure 5.18 shows a graph of β against $\Sigma R_f^2/M_\star$, where R_f is the radius at which the first fragment forms, and includes all the fragmenting simulations presented in this chapter (including the Benchmarking simulations). When interpreting these results, one should also note that an increase in $\Sigma R^2/M_\star$ does not necessarily imply an increased disc to star mass ratio: it is possible to have a low disc to star mass ratio but using a steeper surface mass density profile. I can see that there is a clear trend that as the RHS of equation 5.3 increases, so too does the value of β that will allow a fragment to form. The trendline presents the relation:

$$\beta = \eta \left(\frac{\Sigma R_{\rm f}^2}{M_\star} \right)^\delta, \tag{5.5}$$

where $\delta \approx 1/2$ and the constant of proportionality, $\eta \approx 47$, which I find using a least squares fit. I have also performed a few calculations with higher resolution (2×10^6 SPH particles). Table 5.3 shows a summary of these simulations. I find that as the disc mass is increased, the radius at which the first fragment forms moves to smaller radii (compare simulations 2m-beta5.5 and 2m-beta5.5-Mdisc0.2 or 2m-beta6 and 2m-beta6-Mdisc0.2) while increasing the star mass causes fragmentation to be suppressed (compare simulations 2m-beta6 and 2m-beta6-Mstar2). Note that in the high star mass case, the simulation is run for ≈ 2 ORPs longer than the equivalent low star mass case and yet no fragmentation occurs. In addition, I have performed preliminary calculations using the two-dimensional hydrodynamics code, FARGO. This is a polar grid-based code and I use a fixed grid of 512 and 1536 cells in the radial and azimuthal directions, respectively. Table 5.4 shows a summary of the simulations carried out. I find a similar trend in that increasing the disc mass promotes fragmentation at smaller radii (compare simulations F-beta4 and F-beta4-Mdisc0.2) or even causes a non-fragmenting disc to fragment (compare simulations F-beta5 and F-beta5-Mdisc0.2) while decreasing it causes fragmentation to be suppressed (compare simulations F-beta4 and F-beta4-Mdisc0.05). I find that this trend of β increasing with $\Sigma R^2/M_\star$ is maintained, although the exact values of η and δ may change slightly if I was able to perform all the calculations with high resolution using SPH or with FARGO.

It is also important to note that the trendline in Figure 5.18 has only been produced using the results of simulations with $p < 2$. This is because for $p \gtrsim 2$, the disc will always fragment in the innermost regions first. Consequently, if the results for $p \gtrsim 2$ were included, this would cause the trendline to be somewhat skewed.

The trendline can give some very useful information for those simulations with $p < 2$. The grey region is where I predict subsequent fragmentation is feasible. Traversing the plot in a vertical direction downwards from the trendline into the grey region, at any particular value of $\Sigma R^2/M_\star$ the disc will fragment at all values of β less than the limit given by the trendline (though the radius being considered will not necessarily be the first location at which fragmentation occurs). Similarly, traversing the plot in a horizontal direction from the trendline to the right side of the plot into the grey region, for a particular value of β, fragmentation is possible at a particular radius if the disc mass is increased or star mass is decreased. Similarly, *for a particular value of β and any one combination of the disc to star mass ratio*, fragmentation is also possible at larger radii than the location specified by the trendline i.e. to the right hand side of the trendline. Therefore, the trendline predicts the minimum possible radius at which fragments could in theory form in discs with shallow surface mass density profiles, $p < 2$.

For discs with $p \gtrsim 2$ which fragment in the inner regions, I expect that subsequent fragmentation may take place further out in the disc as far out as given by the trendline in Figure 5.18. In other words, for these discs, I expect there to be a maximum radius outside of which fragmentation will not occur (since the surface mass density fall-off is steep, the outer regions of these discs may struggle to have enough mass for gravitational instability to be significant). However, since I stop

the simulations soon after the first fragment forms due to the increased computational resources required to follow the simulations further, I am unable to test this prediction. Further work needs to be done in this area and is beyond the scope of this chapter. I also note that in real discs, for fragmentation to occur in the inner regions, the cooling time would have to be very small since the dynamical time at small radii would also be very small. Such short cooling times may not be possible in real discs.

It is also important to note that these simulations have been carried out using a ratio of specific heats, $\gamma = 5/3$. As shown by Rice et al. (2005), the ratio of specific heats plays a key role in the fragmentation boundary. I therefore also anticipate a further dependency on the equation of state. Furthermore, for high mass discs ($M_{\rm disc} \gtrsim 0.3{\rm M}_{\odot}$), I find that the initial surface mass density conditions (i.e. Σ_o and p) cannot be used to determine whether the disc will fragment or not. This is because as the disc evolves, the surface mass density profile steepens causing Σ_o and p to change. Consequently, though parts of the disc may start off in the grey shaded region of Figure 5.18 and hence may be expected to fragment, the disc may restructure itself on a timescale faster than the cooling timescale such that it moves out of this region and hence does not fragment.

Following the work of Gammie (2001) and Rice et al. (2005), a number of authors have used the critical cooling timescale, $\beta_{\rm crit}$ (or gravitational stress, $\alpha_{\rm GI,max}$) to predict fragmentation in realistic discs (e.g. Clarke 2009; Rafikov 2009; Cossins et al. 2010; Kratter et al. 2010). In light of the new results presented here, I would encourage previous conclusions based on these critical values to be revisited.

5.7.2 Implications of the new fragmentation criteria on the results of Clarke (2009)

Clarke (2009) produced an analytical model for the structure of a gravitationally unstable disc which is subject to realistic cooling, and determined a boundary between fragmentation and no fragmentation for different values of the steady state mass accretion rate through the disc (Figure 5.19). However, the fragmentation boundary was produced by assuming that $\alpha_{\rm GI,max} = 0.06$, which as shown in this chapter, is only applicable for certain star and disc conditions. I now briefly investigate the impact of the new results on the fragmentation boundary as deduced by Clarke (2009).

Figure 5.18 shows that the trend appears to be present at least between $\beta \approx 1 - 8$, or equivalently, for $\Sigma R^2/M_{\star} \approx 4.5 \times 10^{-4} - 2.9 \times 10^{-2}$ (calculated using equation 5.5). Appendix A of Clarke (2009) shows the steady state analytical solutions for a self-gravitating disc and I reproduce (without derivation) the key formulae here. The opacity regimes being considered in this section are the regimes where ice grains are dominant, ice grains sublime, dust grains dominate and dust grains sublime. The gravitational stress in each of these four regions are given by

$$\alpha_{\rm GI,ice\ grains} = 0.4\left(\frac{R}{100\ {\rm AU}}\right)^{\frac{9}{2}}, \tag{5.6}$$

$$\alpha_{\rm GI,ice\ sublimes} = 2.5 \times 10^{-3}\left(\frac{R}{100\ {\rm AU}}\right)^{\frac{9}{14}}\left(\frac{\dot{M}}{10^{-6}\ {\rm M}_{\odot}{\rm yr}^{-1}}\right)^{\frac{6}{7}}, \tag{5.7}$$

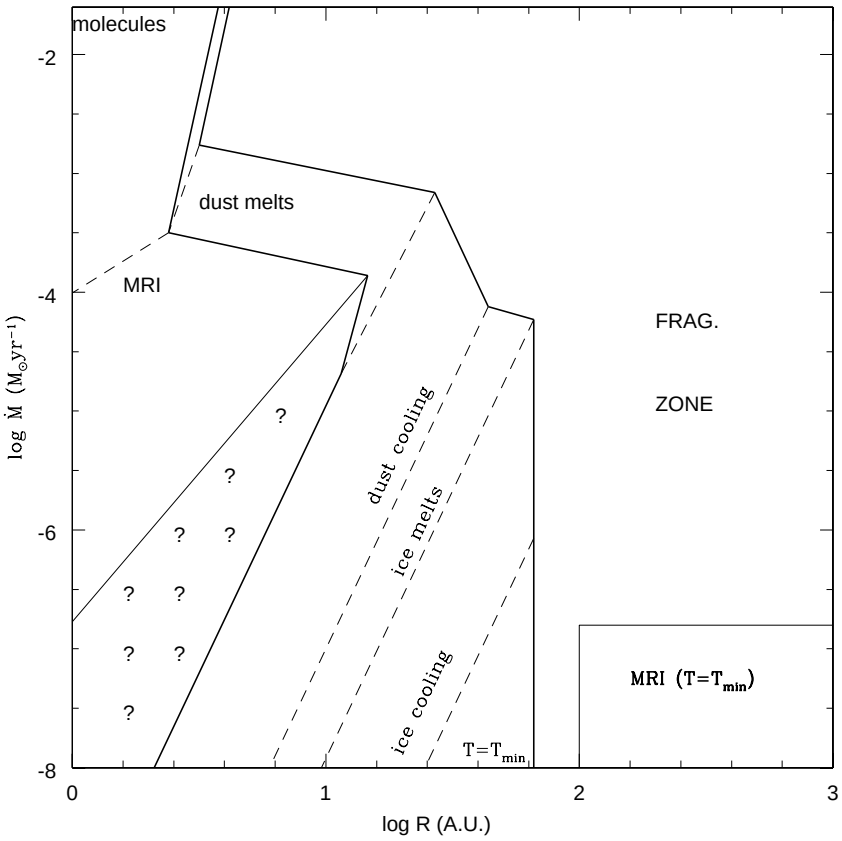

Figure 5.19: Logarithmic plot from Clarke (2009) showing the steady state mass accretion rate against radius for a disc around a $1M_{\odot}$ star. The boundary between fragmentation and no fragmentation is the solid line that lies immediately to the left of the region marked *frag. zone*.

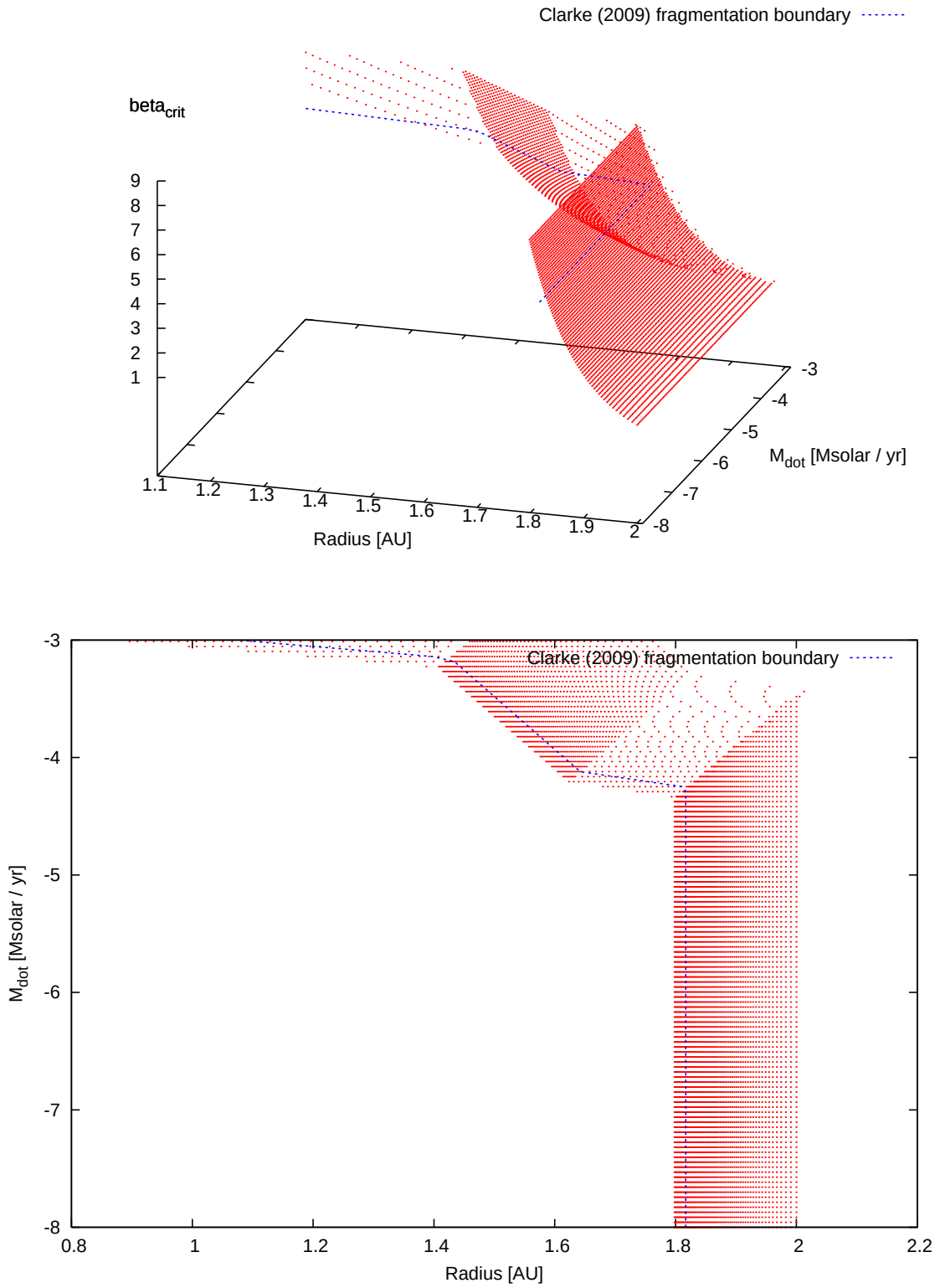

Figure 5.20: Variation of the critical radius at which a disc around a $1M_{\odot}$ star will fragment for mass accretion rates ranging between $\dot{M} = 10^{-8} - 10^{-3} M_{\odot} yr^{-1}$ and for critical values of the cooling timescale in units of the orbital timescale ranging between $\beta_{crit} = 1 - 8$. The upper panel shows the three-dimensional plot whereas the lower panel shows the projection of this onto the $\dot{M} - R$ plane. For low accretion rates, the critical radius can vary by as much as $\approx$ 37 AU . The fragmentation boundary as identified by Clarke (2009) for $\alpha_{GI,max} = 0.06$ is highlighted in blue.

$$\alpha_{\rm GI,dust\ grains} = 4.4 \times 10^{-2} \left(\frac{R}{100\ {\rm AU}} \right)^{\frac{9}{4}} \left(\frac{\dot{M}}{10^{-6}\ {\rm M_{\odot} yr^{-1}}} \right)^{\frac{1}{2}}, \tag{5.8}$$

and

$$\alpha_{\rm GI,dust\ sublimes} = 2.1 \times 10^{-4} \left(\frac{R}{100\ {\rm AU}} \right)^{0.41} \left(\frac{\dot{M}}{10^{-6}\ {\rm M_{\odot} yr^{-1}}} \right)^{0.945}. \tag{5.9}$$

The boundaries between *ice grains* and *ice grains subliming*, *ice grains subliming* and *dust grains*, and *dust grains* and *dust grains subliming* are located at

$$R_{\rm ice\ grains-sublime} = 27 \left(\frac{\dot{M}}{10^{-6}\ {\rm M_{\odot} yr^{-1}}} \right)^{\frac{2}{9}}\ {\rm AU}, \tag{5.10}$$

$$R_{\rm ice\ sublime-dust} = 17 \left(\frac{\dot{M}}{10^{-6}\ {\rm M_{\odot} yr^{-1}}} \right)^{\frac{2}{9}}\ {\rm AU}, \tag{5.11}$$

and

$$R_{\rm dust\ grains-sublime} = 5.5 \left(\frac{\dot{M}}{10^{-6}\ {\rm M_{\odot} yr^{-1}}} \right)^{0.24}\ {\rm AU}, \tag{5.12}$$

respectively. I recalculate the location of the fragmentation boundary using the above equations and vary the critical value of β (or equivalently, $\Sigma R^2/M_{\star}$), and hence $\alpha_{\rm GI,max}$. Figure 5.20 (upper panel) shows a surface that defines at what radius a disc around a $1{\rm M_{\odot}}$ star will fragment for values of the steady state mass accretion rate between $\dot{M} = 10^{-8} - 10^{-3}$ and for $\beta_{\rm crit} = 1 - 8$ (i.e. for $\Sigma R^2/M_{\star} \approx 4.5 \times 10^{-4} - 2.9 \times 10^{-2}$ or $\alpha_{\rm GI,max} = 0.05 - 0.4$). Figure 5.20 (lower panel) plots the projection of this graph onto the $\dot{M} - R$ plane and shows that the critical radius for fragmentation can vary by as much as ≈ 37 AU between the values of $\beta_{\rm crit}$ considered here. This may have significant consequences for conclusions on whether or not certain planets observed may have formed by the gravitational instability method. I discuss the implications of inaccurately determining the fragmentation boundary on the interpretation of observational results, in light of further results, in Chapter 6. In addition, it is important to note that the fragmentation boundaries presented in this section have used interstellar opacity values (as done so by Clarke 2009). The critical radii presented in this section will also vary with opacities other than interstellar values (as discussed in Chapter 3).

Following Clarke (2009), Cossins et al. (2010) show that the cooling timescale in units of the orbital timescale, β, is strongly dependent on the local temperature and since the different opacity regimes have different temperature dependencies, this may affect the local cooling time. This in turn affects the local gravitational stress, and hence may adjust where the fragmentation boundary is located. In light of these results by Cossins et al. (2010), the fragmentation boundaries described in this section would also need to be adjusted to reflect this.

Finally, Clarke (2009) discusses that the critical radius for fragmentation is also dependent on the central star mass such that $R_{\rm crit} \propto M_{\star}^{1/3}$. Therefore, not only will the star's mass affect the value of $R_{\rm crit}$ due to this relation discussed by Clarke (2009), but it will also affect $R_{\rm crit}$ through the part that it plays in determining the value of $\beta_{\rm crit}$ for any one system (through equation 5.5).

Chapter 6

Resolution effects on the fragmentation boundary

6.1 Introduction

In Chapter 5, I explored the physical factors that may affect the critical cooling timescale below which a disc will fragment into bound objects if the disc is also Toomre unstable. When carrying out numerical simulations, it is important to ensure that results converge with higher resolution. This can be computationally demanding if the resolution is increased significantly. For example, since the computational time required for the SPH code to carry out the non-radiative transfer simulations presented in this thesis scales as $N\ \log(N)$, where N is the number of particles used (see Section 2.1.8), an increase in the number of particles by a factor of 8 would increase the run time by more than a factor of 9. However, this is assuming the timesteps that are used remain the same. In reality, as shown in Section 2.1.8, the timesteps also decrease by a factor $N^{-1/d}$, where d is the number of dimensions, since the smoothing length is reduced resulting in a run time increase by more than a factor of 18. Nevertheless, convergence testing is very critical to ensure accuracy of results. With recent advances in computational power combined with fully MPI-parallelised codes, higher resolution testing is possible. In this chapter I explore the effects of resolution on the critical cooling timescale for fragmentation (in discs using simplified cooling rather than radiative transfer) and explore the effects this may have on the radius at which fragmentation may occur.

In Section 2.1.2 I showed that the linear resolution scales as $f^{\frac{1}{d}}$, where f is the factor by which the number of particles are increased by and d is the number of dimensions. While in the previous chapter I extended the work of Rice et al. (2005) considering the physical effects on fragmentation, in this chapter I extend their work to consider the effects of resolution. Rice et al. (2005) carried out a resolution test on one of their discs by decreasing the number of particles from 250,000 to 125,000. They found that the fragmentation results appeared to be unaffected by this. However, given that the simulations were carried out in three-dimensions, it is important to note that this resolution test was equivalent to decreasing the linear resolution (and hence the smoothing length) by only $\approx 21\%$, so it is unsurprising that significant differences were not seen. In addition, the earlier work by Gammie (2001) carried out a resolution test for simulations that

$\alpha_{\rm GI}$	$\beta_{\rm crit}$	$R_{\rm crit}$ ($M_\star = 1{\rm M}_\odot$)	$R_{\rm crit}$ ($M_\star = 1.5{\rm M}_\odot$)	$R_{\rm crit}$ ($M_\star = 2.1{\rm M}_\odot$)
0.018	22.2	50 AU	57 AU	64 AU
0.035	11.4	58 AU	66 AU	75 AU
0.07	5.6	68 AU	78 AU	87 AU
0.14	2.3	79 AU	90 AU	101 AU
0.28	1.4	92 AU	105 AU	115 AU

Table 6.1: Table showing how the critical radius of fragmentation according to Clarke (2009) may be affected for a disc surrounding a 1, 1.5 and 2.1$M_\odot$ star if the value of $\alpha_{\rm GI,max}$ is not correctly determined. The value of $\Sigma R^2/M_\star$ is kept constant.

did not fragment but he did not carry out a resolution test on the fragmentation boundary which, as discussed in Chapter 1, involves different thermodynamics. Most authors who simulate self-gravitating discs consider the resolution criterion set by Bate & Burkert (1997) (see Section 2.1.2; Mayer et al. 2002; Lodato & Rice 2004; Stamatellos & Whitworth 2008; Forgan et al. 2009). Some authors carry out a resolution test (in addition to ensuring the resolution criterion is satisfied) by increasing or decreasing the number of particles by a factor of 2 (e.g. Lodato et al. 2007; Clarke et al. 2007; Cossins et al. 2009). However large changes in resolution have not been tested.

6.2 The impact of inaccurate values of $\alpha_{\rm GI,max}$ on physical and observational conclusions

In Section 1.4.2 I showed that in a disc which is dominated by internal heating processes due to gravitational instability, the cooling timescale in units of the orbital timescale, β, can be related to the gravitational stress in the disc by

$$\alpha_{\rm GI} = \frac{4}{9}\frac{1}{\gamma(\gamma-1)}\frac{1}{\beta}. \tag{6.1}$$

The simulations presented in Chapter 5 (Benchmark1-8) showed that for a $0.1{\rm M}_\odot$ disc simulated using 250,000 particles with surface mass density profile, $\Sigma \propto R^{-1}$, spanning a radial range of $0.25 \leq R \leq 25$AU and surrounding a $1{\rm M}_\odot$ star, the critical value of the cooling timescale in units of the orbit timescale, $\beta_{\rm crit} \approx 5.6$ (for a disc modelled using a ratio of specific heats, $\gamma = 5/3$). This critical value is equivalent to $\alpha_{\rm GI,max} \approx 0.07$. Clarke (2009) produced an analytical model for the structure of a gravitationally unstable disc which is subject to realistic cooling. She showed that for optically thick discs that are sufficiently low in temperature that they are dominated by ice grains,

$$\alpha_{\rm GI} = 0.4\left(\frac{R}{100\ {\rm AU}}\right)^{\frac{9}{2}}, \tag{6.2}$$

for a disc with interstellar opacities and surrounding a $1{\rm M}_\odot$ star, where R is the radius being considered. This relationship shows that for a maximum value of the gravitational stress, a critical radius, $R_{\rm crit}$, can be found outside of which fragmentation can occur (for a disc with a shallow surface mass density profile). Since for the disc described above, $\alpha_{\rm GI,max} \approx 0.07$, then, $R_{\rm crit} \approx$

Property	parameters used
Star mass	$1M_{\odot}$
Disc mass	$0.1M_{\odot}$
Radial range	$0.25 \leq R \leq 25$AU
Initial surface mass density profile	$\Sigma \propto R^{-1}$
Initial temperature profile	$T \propto R^{-1/2}$
Ratio of specific heats	$\gamma = 5/3$

Table 6.2: Table showing the properties of the discs simulated in this chapter.

68 AU. The critical radius also scales as $M_{\star}^{1/3}$ (as described by Clarke 2009). One might then use this to compare to observational data and conclude whether a system may or may not have formed via gravitational instability. With reference to our own solar system, these values would suggest that the planets could not have formed via gravitational instability.

Table 6.1 shows the effect of varying $\alpha_{\rm GI}$ by a factor of 2 and 4 for central star masses of 1.0, 1.5 and 2.1 $M_{\odot}$ (assuming no change in the value of $\Sigma R^2/M_{\star}$ for simplicity of illustration). For HR 8799 ($M_{\star} \approx 1.5M_{\odot}$), using a value of $\alpha_{\rm GI,max} \approx 0.07$ which is close to the currently accepted value of $\alpha_{\rm GI,max}$, the fragmentation boundary might be expected to be at $\approx$ 78 AU. However, if the value of $\alpha_{\rm GI,max}$ is incorrectly determined, one might incorrectly conclude whether the outer planet, HR 8799 b, could indeed have formed by gravitational instability. Similarly, an incorrect value of $\alpha_{\rm GI,max}$ may affect the conclusions on the formation mechanism of the planet around the $\approx 2.1M_{\odot}$ A, Fomalhaut (exo http://exoplanet.eu). Fomalhaut harbours a planet at a distance of $\approx$ 119 AU. Using a large value of $\alpha_{\rm GI}$, one might conclude that the formation of it by gravitational instability is uncertain, while a lower value makes such a formation mechanism more likely.

6.3 Simulations

The disc and star properties used to carry out the simulations in this chapter are exactly the same as those used in simulations Benchmark1-8 in Chapter 5 and are summarised in Table 6.2. In Chapter 5, I carried out simulations using 250,000 particles, the same resolution used by Rice et al. (2005). In this chapter, I simulate the same disc at higher resolutions by using 2 million and 16 million particles (i.e. the smoothing length is decreased by a factor of 2 and 4, respectively). In addition, it is important to note that lower resolution testing can also provide interesting convergence results. Therefore, I carry out additional simulations using 31,250 particles (so that the smoothing length is a factor of 2 higher). I simulate the discs using various values of the cooling timescale in units of the orbital timescale, β, to determine the critical value, $\beta_{\rm crit}$, at different resolutions. Table 6.3 summarises the main simulation parameters and the key fragmenting results. The simulations were run either for at least 6 ORPs or until the discs fragmented. These are defined as regions which are at least two orders of magnitude denser than their surroundings. At the time of writing this thesis however, one of the 2 million particle simulations ($\beta = 11$) and two of the 16 million particle simulations ($\beta = 15$ and 20) are not complete and I highlight these in the results.

Simulation name	No of particles	β	Fragmented?
31k-beta2	31,250	2.0	Yes
31k-beta2.5	31,250	2.5	Yes
31k-beta3	31,250	3.0	Yes
31k-beta3.5	31,250	3.5	No
31k-beta4	31,250	4.0	No
250k-beta5	250,000	5.0	Yes
250k-beta5.5	250,000	5.5	Yes
250k-beta5.6	250,000	5.6	borderline
250k-beta6	250,000	6.0	No
250k-beta6.5	250,000	6.5	No
250k-beta7	250,000	7.0	No
250k-beta7.5	250,000	7.5	No
2m-beta5.5	2 million	5.5	Yes
2m-beta6	2 million	6	Yes
2m-beta6.5	2 million	6.5	Yes
2m-beta7	2 million	7	Yes
2m-beta8	2 million	8	Yes
2m-beta10	2 million	10	borderline
2m-beta10.5	2 million	10.5	borderline
2m-beta11	2 million	11	incomplete
2m-beta15	2 million	15	No
16m-beta10	16 million	10	Yes
16m-beta15	16 million	15	incomplete
16m-beta18	16 million	18	borderline
16m-beta20	16 million	20	incomplete

Table 6.3: Table showing the simulations carried out and the key fragmenting results. Note that the simulations with 250,000 particles were carried out in Chapter 5. The simulations labelled as *borderline* are those that show fragments forming which then shear apart in less than 1 ORP. The simulations that have not finished at the time of writing this thesis are labelled as *incomplete*.

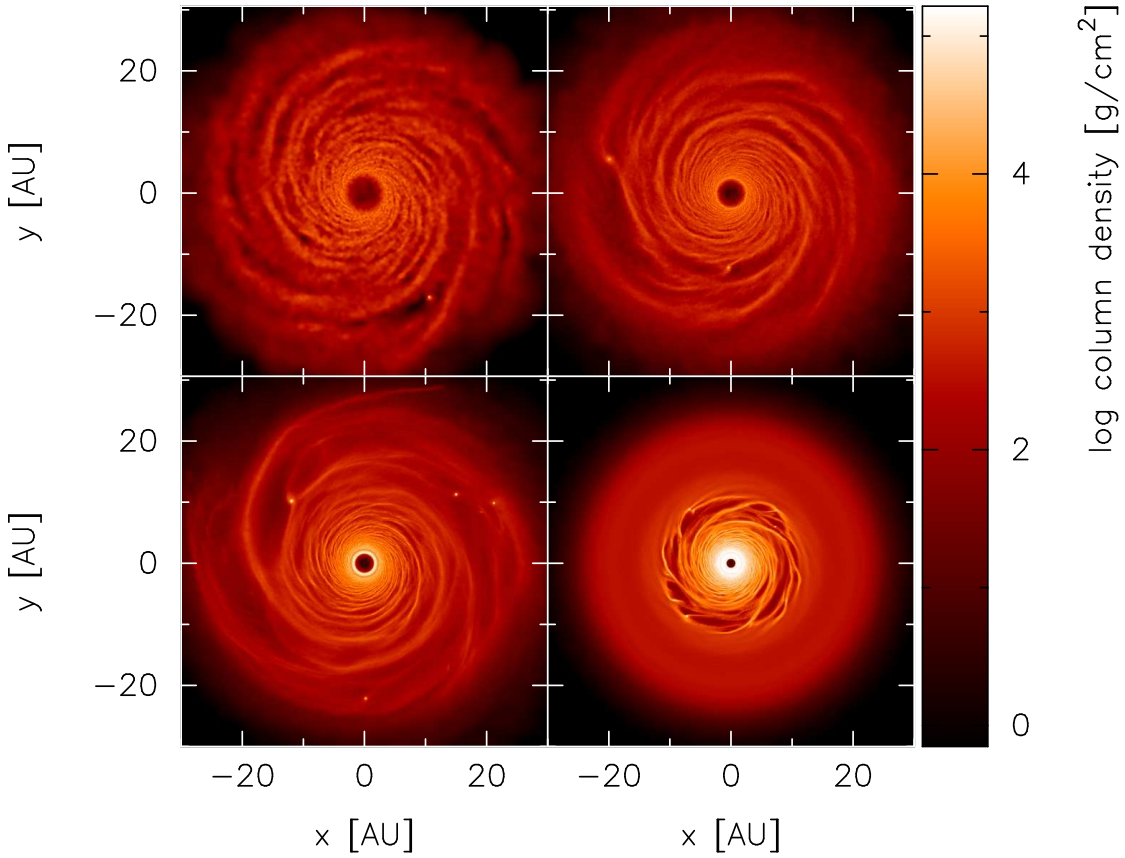

Figure 6.1: Surface mass density rendered images of the fragmenting discs with 31,250 (upper left), 250,000 (upper right), 2 million (bottom left) and 16 million (bottom right) particles using values of the cooling timescale in units of the orbital timescale of β = 3, 5.5, 8 and 10, respectively (simulations 31k-beta3, 250k-beta5.5, 2m-beta8 and 16m-beta10, respectively). The images are produced at time t = 5.3, 6.4, 5.3 and 2.5 ORPs, respectively.

6.4 Results

It is important to note that two of the 16 million particle simulations did not complete at the time of writing this thesis. As discussed in Section 2.1.8, increasing the number of particles significantly increases the time taken to run these simulations despite a number of methods having been adopted to speed up the simulations. The 16 million particle simulations run with β = 10 and 18 have taken $\approx$ 120 and 140 days, respectively. The former simulation was run using 8 compute nodes while the latter was run using 16 nodes. As will be shown, the fragmenting results of the incomplete simulations (those run with β = 15 and 20) can be inferred from the other high resolution simulations.

6.4.1 Fragmentation boundary

Table 6.3 highlights the key fragmenting results for the simulations carried out in this chapter. The simulations that show no fragments at all are classed as non-fragmenting discs. Those discs that fragment and where the fragments remain without being sheared apart (satisfying the criteria defined above) are classed as fragmented discs. There are four simulations that are classed as

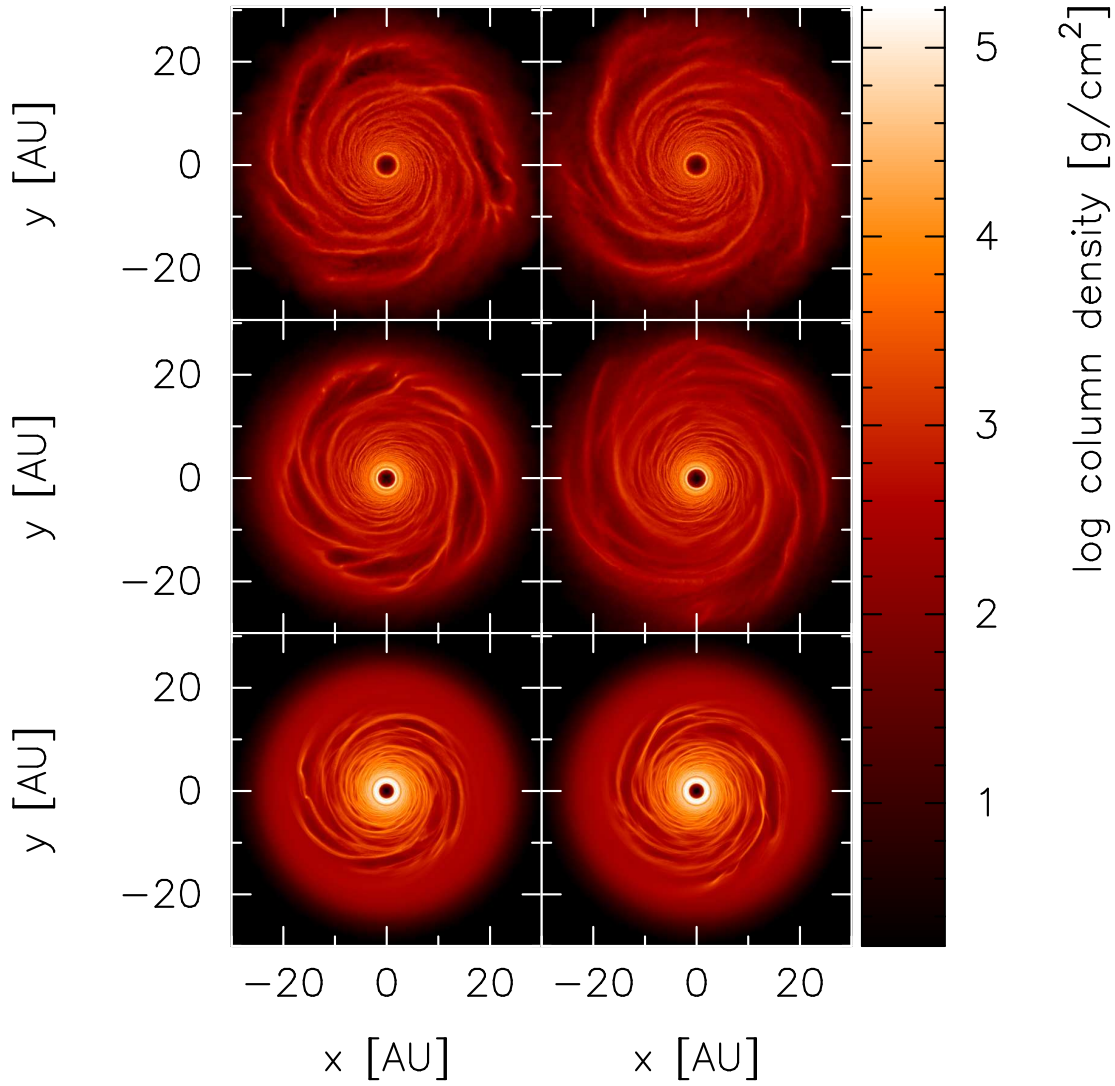

Figure 6.2: Surface mass density rendered images of the borderline cases (simulations 250k-beta5.6, 2m-beta10, 16m-beta18) for discs set up with 250,000 (β = 5.6), 2 million (β = 10) and 16 million (β = 18) particles (top, middle and bottom panels, respectively). The left panels show a hint of fragmentation at times, t = 3.8, 4.8 and 6.0 ORPs (top, middle and bottom panels, respectively). The right panels show the equivalent simulations a short time later at times, t = 4.2, 5.8 and 6.2 ORPs (top, middle and bottom panels, respectively). Within 1 ORP, the fragments have been sheared apart, classing these simulations as *borderline*.

borderline (simulations 250k-beta5.6, 2m-beta10, 2m-beta10.5 and 16m-beta18). These are discs which show signs of fragmentation but the fragments shear apart rapidly (within 1 ORP) and no further signs of fragmentation are seen.

Figure 6.1 shows images of the fragmenting discs in simulations 31k-beta3, 250k-beta5.5, 2m-beta8 and 16m-beta10 for β values of 3, 5.5, 8 and 10 simulated using 31,250, 250,000, 2 million and 16 million particles, respectively. Figure 6.2 shows images of the borderline discs in simulations 250k-beta5.6, 2m-beta10, 16m-beta18. The left panels show the discs appearing to form fragments. However, within 1 ORP, the fragments shear apart and do not appear to form again. The disc with 250,000 particles has been run for long enough that it has now regained a marginally stable state with $Q \approx 1$ (Figure 6.3). The higher resolution simulations have higher values of the Toomre stability parameter as they have been heated due to the gravitational instability but have not been run long enough for them to cool back down to a marginally stable state.

Figure 6.4 summarises the results in Table 6.3. The solid black line divides the fragmenting and non-fragmenting simulations and has been included by eye as a fit to where the boundary may be. Table 6.3 shows that for 250,000, 2 million and 16 million particles, the critical value of the cooling timescale in units of the orbital timescale, $\beta_{\rm crit} \approx 5.6$, 10 and 18, respectively, since these have been identified as *borderline* cases. For the discs simulated with 31,250 particles, I assume that the critical value is between the fragmenting case of $\beta = 3.0$ and the non-fragmenting case of $\beta = 3.5$ and thus take $\beta_{\rm crit} \approx 3.25$. It can clearly be seen that with the data that is available, the dividing line between the fragmenting and non fragmenting cases must be linear and therefore a convergence has not been reached. Figure 6.4 also shows how the trend may continue if a convergence does not take place at even higher resolution as well as how the trend may be if convergence is achieved.

6.4.2 Dissipation and cooling rates

The lack of convergence shows that at higher resolution, the discs are able to fragment much more easily since the critical value of the cooling timescale, $\beta_{\rm crit}$, in units of the orbital timescale is larger. A larger value of $\beta_{\rm crit}$ means that the cooling rate does not need to be as high for the cooling to overcome the heating due to gravitational instability, and hence for it to fragment. It is therefore useful to compare the total cooling and dissipation rates (per unit mass) in the discs (Figure 6.5) for simulations that are just on the non-fragmenting side of the boundary shown in Figure 6.4, i.e. the discs are strongly self-gravitating but do not fragment. To be able to analyse the discs more easily, I consider simulations that are not borderline cases. The total dissipation rates have been measured in the simulations and have been averaged over 1 ORP. The total cooling rate is calculated using

$$C = \frac{u}{t_{\rm cool}} = \frac{u\Omega}{\beta} = \frac{c_{\rm s}^2\Omega}{\gamma(\gamma-1)\beta}. \tag{6.3}$$

Figure 6.5 shows the dissipation and cooling rates for simulations 31k-beta3.5 and 250k-beta6 which are discs modelled with $\beta = 3.5$ and 6 using 31,250 and 250,000 particles, respectively. These discs have achieved a steady state since the total dissipation rate balances the total cooling

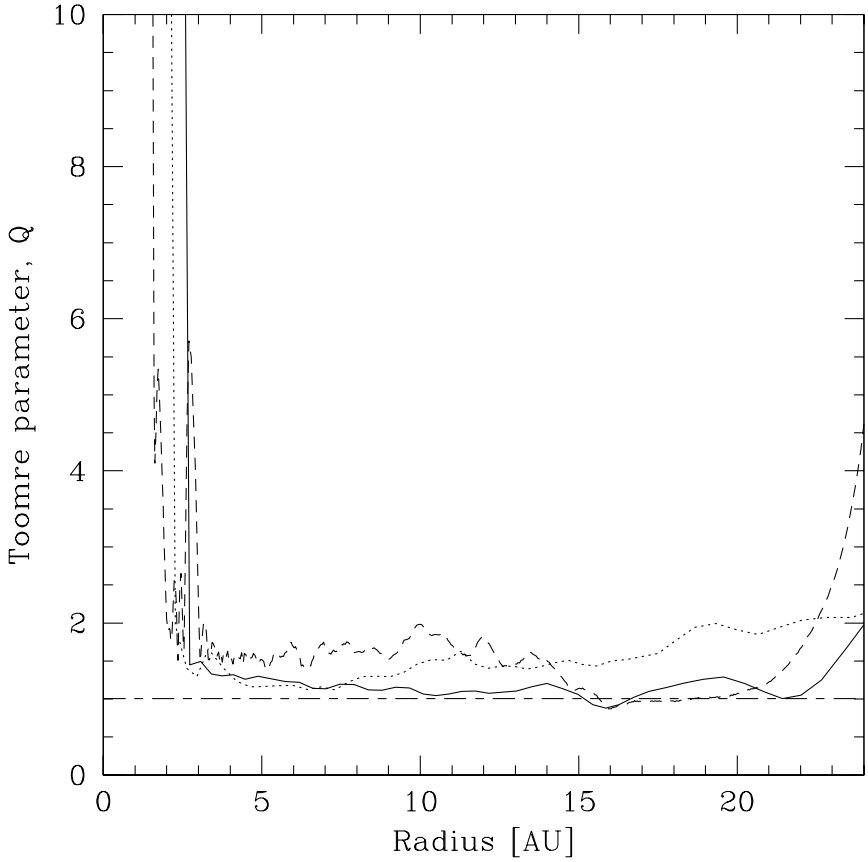

Figure 6.3: Toomre stability profiles for the borderline simulations for discs set up with 250,000 (solid line), 2 million (dotted line) and 16 million (dashed line) particles (simulations 250k-b5.6, 2m-b10 and 16m-b18, respectively) at time t = 6.4, 6.4 and 6.3 ORPs. These plots are produced at times Δt = 2.3, 0.7 and 0.1 ORPs after the fragments are identified to have sheared apart. The simulation with 250,000 particles has evolved for a longer time after the fragments have sheared apart and hence its Toomre stability profile is closer to the marginal state of $Q \approx 1$. The simulations with 2 million and 16 million particles have not evolved as far and so their Toomre stability profiles are higher due to the heating from the gravitational instability. The critical value of $Q_{\rm crit}$ = 1 is also marked.

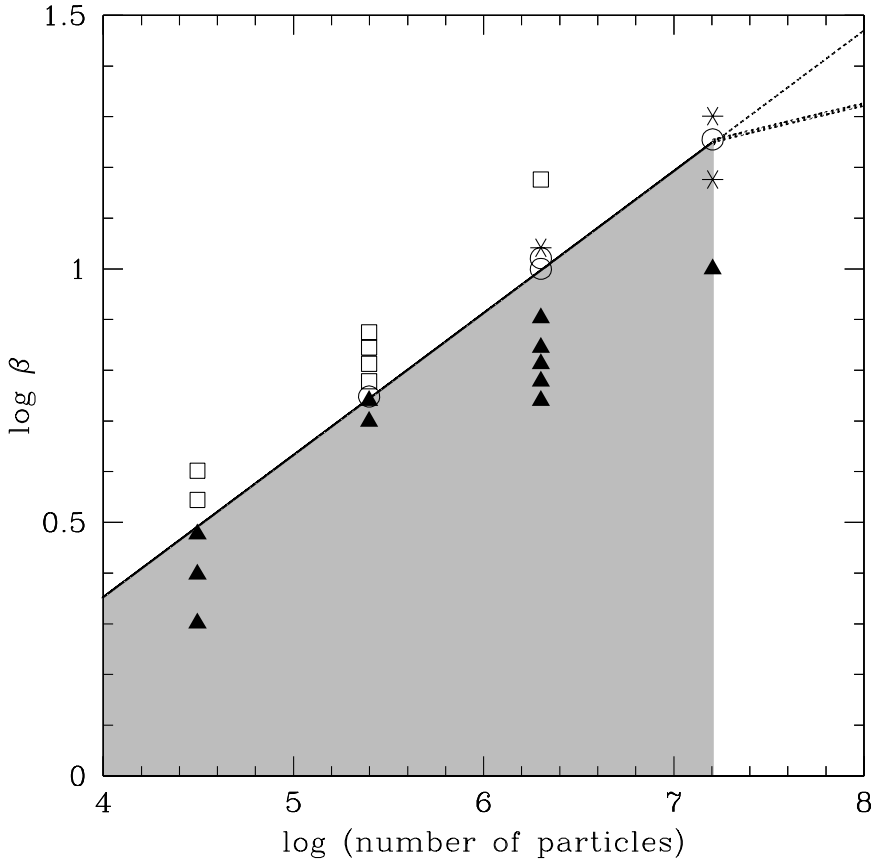

Figure 6.4: Graph of β against resolution of the non-fragmenting (open squares), fragmenting (solid triangles) and borderline (open circles) simulations. Also included are the simulations that have not finished (asterix). The borderline simulations are those that fragment but whose fragments are sheared apart and no further evidence of fragmentation is seen. The solid black line shows a dividing line between the fragmenting and non-fragmenting cases and the grey region is where fragmentation can take place. The graph shows no evidence of convergence of results with increased resolution. The thin dotted line shows how the trend will continue if convergence is not reached with higher resolution than 16 million particles whereas the heavy dotted line shows how the trend might continue with higher resolution if convergence begins to take place.

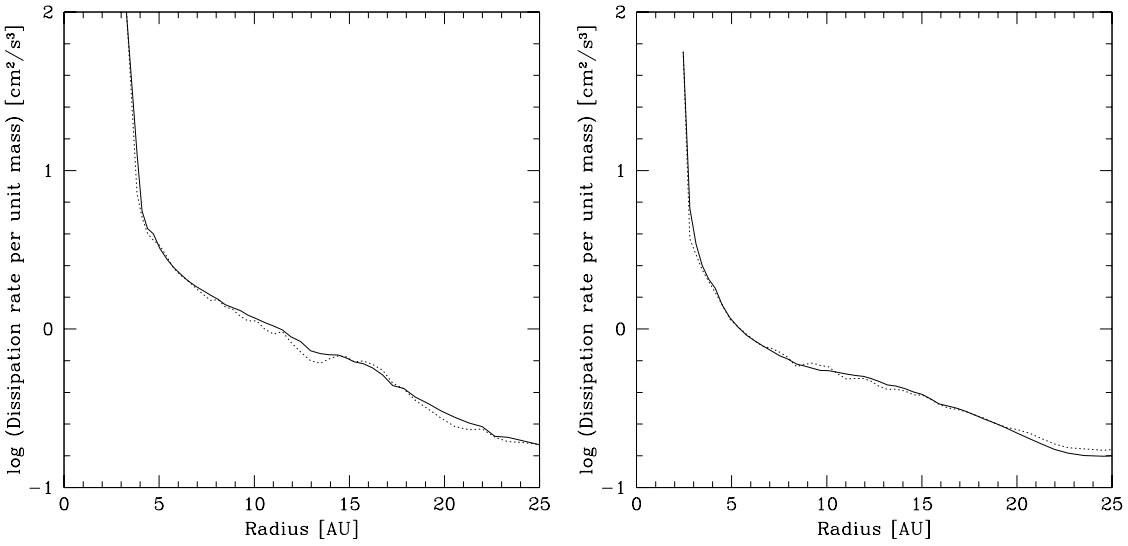

Figure 6.5: Graphs of total dissipation rate per unit mass (solid line) against radius for discs with 31,250 (left panel) and 250,000 (right panel) particles using $\beta = 3.5$ and 6, respectively (simulations 31k-beta3.5 and 250k-beta6, respectively). Overlaid is the cooling rate per unit mass (dotted line). The discs are in a steady state as the dissipation and cooling rates balance each other throughout the disc. The data for the simulations with 2 million and 16 million particles is unavailable at the time of writing this thesis.

rate. At the time of writing this thesis, the data for the simulations with 2 million and 16 million particles is unavailable.

In the simulations carried out, there are several contributions to the dissipation in the discs. The primary contribution is expected to be from the gravitational stress. The relevant component of the gravitational stress tensor is given by (Lynden-Bell & Kalnajs 1972; Lodato & Rice 2004, equation 1.21)

$$T^{\mathrm{g}}_{R\phi} = \Sigma \frac{\mathbf{g_r} \mathbf{g}_\phi}{4\pi G \rho}, \tag{6.4}$$

where $\mathbf{g_r}$ and $\mathbf{g}_\phi$ are the radial and azimuthal components of the gravitational acceleration due to the disc self-gravity. Using equation 1.22, the dissipation rate (per unit mass) due to the gravitational stress (for a Keplerian disc) is given by

$$D_{\mathrm{grav}} = \left| \frac{\mathrm{d}\ln\Omega}{\mathrm{d}\ln R} \right| \Omega \frac{\mathbf{g_r} \mathbf{g}_\phi}{4\pi G \rho} \approx \frac{3}{2} \Omega \frac{\mathbf{g_r} \mathbf{g}_\phi}{4\pi G \rho}. \tag{6.5}$$

The flow of the fluid will also cause a Reynolds stress given by

$$T^{\mathrm{Re}}_{R\phi} = \Sigma \delta \mathbf{v}_R \delta \mathbf{v}_\phi = \Sigma (\mathbf{v_R} - \overline{\mathbf{v_R}})(\mathbf{v}_\phi - \overline{\mathbf{v}_\phi}) \tag{6.6}$$

where $\mathbf{v_R}$ and $\mathbf{v}_\phi$ are the components of the fluid velocity in the radial and azimuthal directions, respectively and the overbar represents the average of these quantities in the surrounding region of fluid. The dissipation rate per unit mass associated with this flow is given by:

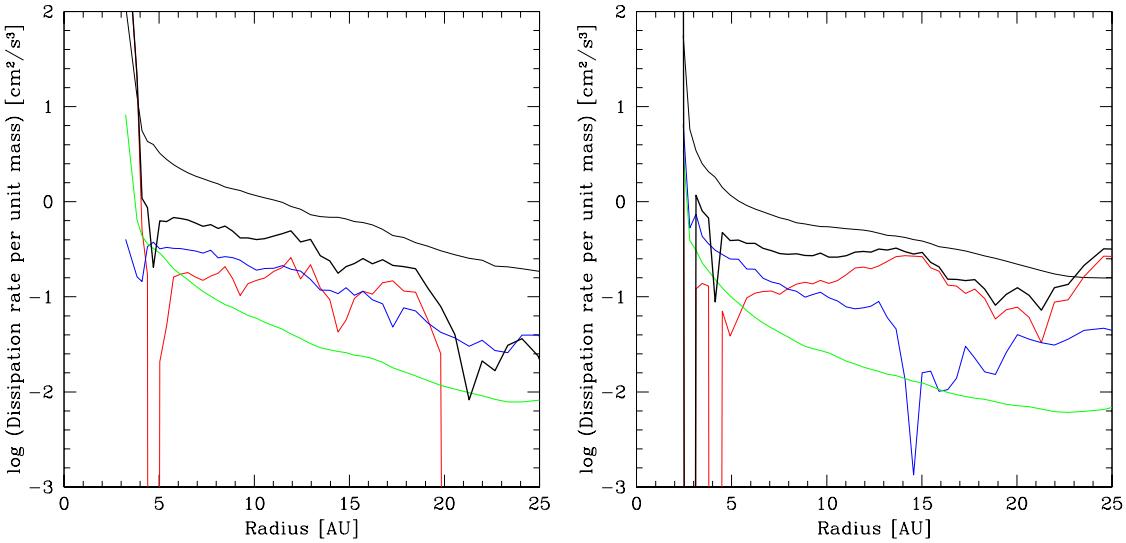

Figure 6.6: Graphs of dissipation rate per unit mass against radius averaged over 1 ORP for a disc with 31,250 (left panel) and 250,000 (right panel) particles (simulations 31k-beta3.5 and 250k-beta6, respectively). The thin black line shows the total dissipation rate per unit mass as measured and recorded during the simulations. Also plotted are the dissipation rates per unit mass due to artificial viscosity (green line), Reynolds stress (blue line) and gravitational stress (red line). The heavy black line shows the sum of these three stresses and would be expected to lie on top of the thin black line. It is clear that there is an additional source of heating present in the simulations that is not expected and is likely to be caused by *additional numerical dissipation*. The data for the simulations with 2 million and and 16 million particles is unavailable at the time of writing this thesis.

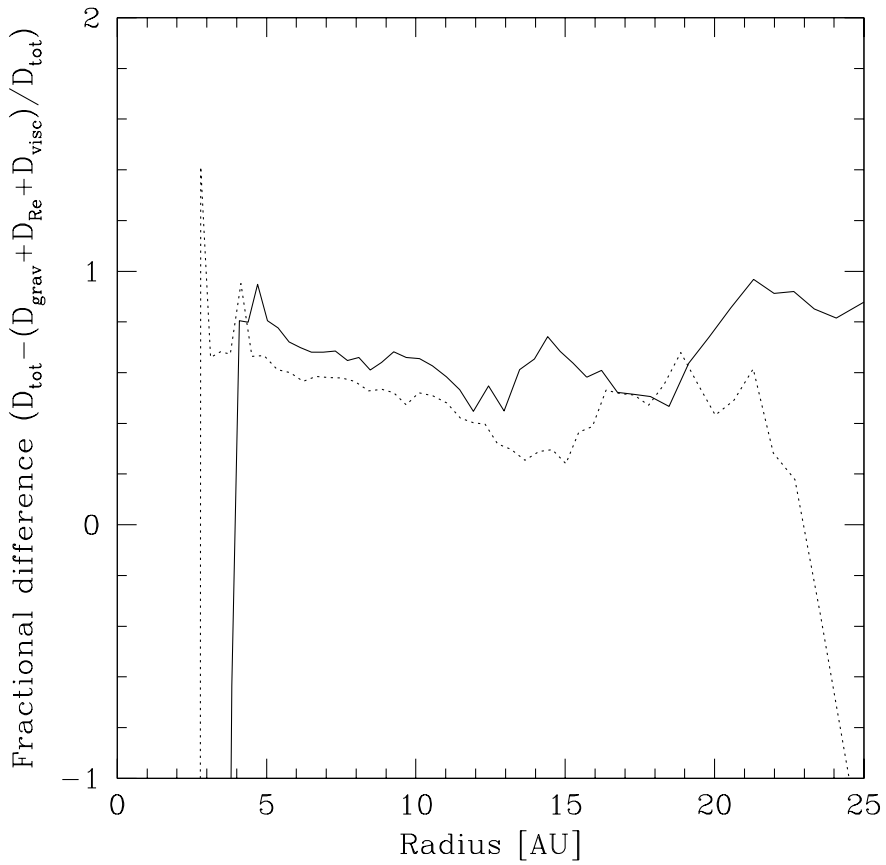

Figure 6.7: Graph of fractional difference between the dissipation rate calculated in the simulations and the total dissipation rate due to the gravitational stress, Reynolds stress and artificial viscosity for a disc with 31,250 (solid line) and 250,000 (dotted line) particles using β = 3.5 and 6, respectively (simulations 31k-beta3.5 and 250k-beta6 respectively). The fractional difference decreases as the resolution increases. The data for the simulations with 2 million and 16 million particles is unavailable at the time of writing this thesis.

$$D_{\rm Re} = \left|\frac{{\rm dln}\Omega}{{\rm dln}R}\right|\Omega\delta v_R\delta v_\phi. \tag{6.7}$$

In addition, since there is also artificial viscosity present, the analytical estimate of the dissipation associated with the shear component (i.e. no contributions from the bulk viscosity or random motions damped by artificial viscosity) is given by (equations 2.34 and 2.36):

$$D_\nu = 0.05\left|\frac{{\rm dln}\Omega}{{\rm dln}R}\right|^2\alpha_{\rm SPH}\frac{h}{H}c_s^2\Omega. \tag{6.8}$$

Figure 6.6 shows the dissipation rate due to the gravitational stress, the Reynolds stress and the artificial viscosity for the simulations 31k-beta3.5 and 250k-beta6. This figure also shows the total dissipation rate calculated directly from the code as well as the sum of the above three dissipation rates averaged over 1 ORP. It can clearly be seen that the total dissipation rate is generally higher than the sum of the dissipation rates due to gravitational stress, Reynolds stress and artificial viscosity. Therefore it can be seen that there are other form(s) of dissipation that are unexpected and is likely due to numerical dissipation. I will refer to the unknown numerical dissipation as *additional numerical dissipation*. Figure 6.7 shows the fractional difference between the total dissipation rate measured in the simulations and the sum of the three calculated components of the dissipation, averaged over 1 ORP. It can be seen that as the resolution increases, the fractional difference between the total dissipation rate measured in the simulations and the sum of the dissipation due to gravitational stress, Reynolds stress and artificial viscosity decreases. This suggests that as the resolution increases further, convergence may be reached but testing this will require a significant amount of computing time which is beyond the scope of this thesis and may not be possible for a number of years.

6.5 Discussion

In Section 2.1.2, I describe the Wengen test which shows that variations in a numerical technique may have significant effects on fragmentation results (using a spatially adaptive softening and a fixed viscosity, I do not find fragmentation whereas Mayer & Gawryszczak (2008) found that fragmentation took place when using a fixed gravitational softening and a Balsara switch implementation for the artificial viscosity; Figure 2.1). Indeed, in Chapter 5 I show that a variation in the way the smoothing length is set and a variation in how the timestepping is carried out can play a part in the outcome (for a disc made up of 250,000 particles, I find that the critical value of the cooling timescale in units of the orbital timescale, $\beta_{\rm crit} \approx 5.6$, whereas Rice et al. (2005) found $\beta_{\rm crit} \approx 6 - 7$. This shows that numerical effects do play a part in the outcome of simulations and that these need to be thoroughly tested and convergence reached to ensure accurate conclusions.

Resolution is one such aspect that needs to be considered. Previous work has typically considered the resolution criterion of Bate & Burkert (1997) (see Section 2.1.2). Some resolution testing of SPH simulations of self-gravitating discs has taken place but this has typically involved varying the particle number by a factor of 2. Lodato et al. (2007) and Cossins et al. (2009) found

that a factor of 2 variation in the particle number did not affect their results. However, these were for non-fragmenting discs. Similarly, Gammie (2001) carried out resolution testing of gravitationally unstable regions, but this was also for regions that do not fragment. Clarke et al. (2007) carried out a simulation of a fragmenting self-gravitating disc and increase the number of particles from 250,000 to 500,000 and suggest that fragmentation may be affected by resolution. However, on the other hand, Rice et al. (2005) suggest that their fragmentation results are unaffected by resolution (though they decrease their particle number from 250,000 to 125,000). However, large changes in resolution on the fragmentation have not been tested.

Lodato & Rice (2004, 2005) show that 250,000 particles is adequate enough to simulate a self-gravitating disc (with $\alpha_{\rm SPH}, \beta_{\rm SPH} = 0.1, 0.2$) to ensure that the artificial dissipation is smaller than the dissipation due to self-gravity by more than an order of magnitude. However, the total dissipation compared to the sum of the various contributions to the dissipation has not been considered.

The variations in fragmenting results with different resolutions show that convergence has clearly not taken place. The dissipation rate graphs suggest another form of dissipation is present in the disc and that it also varies with resolution. The question that is yet to be answered is *what is the source of the additional numerical dissipation since the SPH code used here conserves energy?* A possible reason for the additional numerical dissipation may be extra particle oscillations (originally provided by the bulk motions in the disc) that are not fully damped by the artificial viscosity but are later dissipated as heat. This needs to be investigated further. In addition I have only used a fixed viscosity method in these simulations. While it would be worrying if the viscosity prescription used had such a profound effect on the results, this needs to be explored.

Once the cause has been identified, a further question to answer is whether the effects are primarily important in simulating disc configurations or whether it plays a significant part in other geometries. For example, if the additional dissipation was mainly being caused in situations with shocked scenarios, the problems identified here are unlikely to cause inaccurate results of other scenarios modelled using SPH, where shocks are not so common.

The simulations carried out with 16 million particles have taken between 4-7 months to run on the University of Exeter's supercomputer, mostly using 64 compute cores on 8 computer nodes (though one simulation was run using 128 compute cores on 16 computer nodes). Depending on the nature of what is causing convergence not to occur, even higher resolution simulations will be required for accurate modelling. However, these would take much longer. In addition, these discs are modelled with a simple parameterisation for the cooling. Such high resolution simulations carried out with radiative transfer would take much longer and thus is not likely to be feasible in the near future. However, this does not mean that it is not possible to achieve scientific results. It may simply be that only results *relative* to another scenario can be carried out (though this is largely dependent on the nature of the additional numerical dissipation and whether it is fixed for a particular resolution).

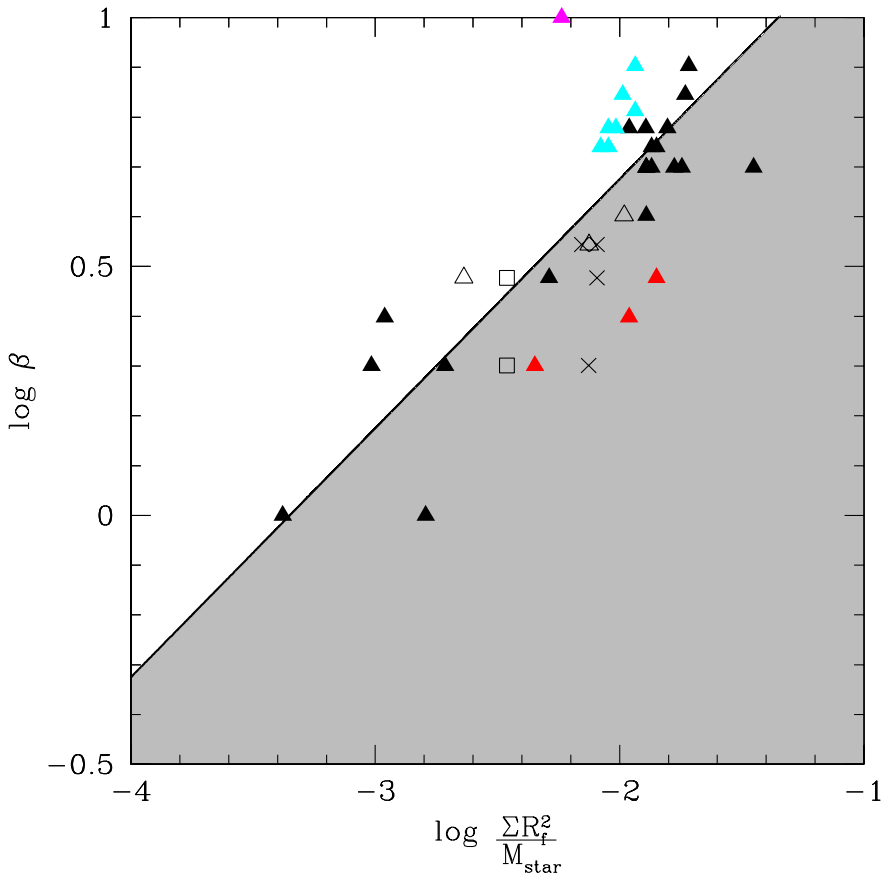

Figure 6.8: Logarithmic graph showing the trend between β and $\Sigma R_{\rm f}^2/M_\star$ determined by considering the location at which the first fragment forms in the discs, $R_{\rm f}$. This graph is identical to that in Figure 5.18 but overlaid with the fragmenting results from this chapter using 31,250 (red triangles) and 16 million (magenta triangle) particles, as well as the fragmenting 2 million (cyan triangles) particle simulation results presented in this chapter and in Table 5.3 in Chapter 5. This graph suggests that the trend identified in Chapter 5 is evident. The exact value of the constant of proportionality certainly seems like it would be different for different resolutions. It is not clear from the few simulations run here whether the slope will also change with resolution.

No of particles	$\beta_{\rm crit}$	$\alpha_{\rm GI,max}$	$R_{\rm crit}$ ($M_\star = 1{\rm M}_\odot$)	$R_{\rm crit}$ ($M_\star = 1.5{\rm M}_\odot$)	$R_{\rm crit}$ ($M_\star = 2.1{\rm M}_\odot$)
31,250	3.25	0.12	77 AU	88 AU	99 AU
250,000	5.6	0.07	68 AU	78 AU	87 AU
2 million	10.0	0.04	60 AU	69 AU	77 AU
16 million	18.0	0.02	53 AU	60 AU	67 AU

Table 6.4: Table showing how the critical radius of fragmentation according to Clarke (2009) may be affected for a disc surrounding a 1, 1.5 and 2M$_\odot$ star for the different values of $\beta_{\rm crit}$ identified for discs with 31,250, 250,000, 2 million and 16 million particles. The value of $\Sigma R^2/M_\star$ is kept constant.

6.5.1 Implications on results of previous chapters

Figure 6.8 shows a logarithmic graph of β against $\Sigma R_{\rm f}^2/M_\star$, where $R_{\rm f}$ is the radius at which the first fragment forms. This figure is the same as Figure 5.18 but has the results from this chapter overlaid. The red, cyan and magenta triangles are the results for the simulations with 31,250, 2 million (including the results from Chapter 5) and 16 million particles, respectively. Though a full parameter study has not been carried out at different resolutions, the existence of a trend identified in Chapter 5 still appears to be present. However, as discussed in Chapter 5, a change of resolution is likely to affect the values of η and δ, deduced to be ≈ 47 and 0.5 for a disc with 250,000 particles. Figure 6.8 shows that the constant of proportionality (i.e. η) is likely to be different for different resolutions. A more fuller study is required to determine whether the slope (i.e. δ) would also change with resolution.

In Chapter 4, I showed that to form the HR 8799 planets by gravitational instability, a large disc mass is required. Such disc masses are unlikely to be the case in reality and current observational efforts have not identified discs larger than $\approx 0.1 - 0.4{\rm M}_\odot$ (Andrews & Williams 2007; Eisner & Carpenter 2006). I therefore suggested that the HR8799 system may not have formed by this mechanism. However, a higher resolution study may produce quite significant differences. Since it appears that fragmentation is somewhat *easier* at higher resolution i.e. that the cooling that is needed to balance the heating is not as high as for low resolution simulations, the lower disc masses that did not fragment in Chapter 4 may well do so at higher resolution (due to the results found in this chapter as well as because the discs will be better resolved and hence the fragments will be more easily discernible as discussed by Bate & Burkert 1997). However, given that convergence has not been reached, it is uncertain as to whether an accurate conclusion can be made even if high resolution simulations are carried out. On one hand, if at low resolution it is possible for a reasonable mass disc (i.e. in line with current observations) to fragment, one may conclude that gravitational instability may be a formation mechanism since even higher resolutions would also be expected to show this. However, if the cause of a lack of convergence is an inherent property of the SPH code such that an increased resolution is not likely to allow convergence to be reached, then absolute conclusions will be very difficult to obtain.

6.5.2 Implications of a lack of convergence on observational interpretation

Table 6.4 shows the equivalent value of $\alpha_{\rm GI,max}$ that is associated with the value of $\beta_{\rm crit}$ identified for each resolution considered in this chapter, as well as the critical radius outside of which fragmentation may take place according to Clarke (2009) for central star masses of $1M_{\odot}$, $1.5M_{\odot}$ and $2.1M_{\odot}$. It can be seen that the critical radius of fragmentation identified for a disc with 16 million particles is smaller than the value identified for a disc with 250,000 by $> 20\%$. This can have profound consequences on the interpretation of observational results. Table 6.4 shows that while the value of $\alpha_{\rm GI,max}$ used from simulations using 250,000 particles may cause one to conclude that the outer planet around HR 8799 may possibly have formed via gravitational instability (since its projected separation is ≈ 68 AU), the results using 16 million particles suggest that there is a larger possibility that this planet could have formed by gravitational instability. For a less marginal case of Fomalhaut b which has a separation of ≈ 119 AU, the results do not appear to be so critical when determining whether it is a potential candidate for a planet formed by gravitational instability.

Chapter 7

Conclusions

I have carried out three-dimensional numerical simulations of self-gravitating discs to determine under what circumstances they fragment to form bound clumps that may grow into giant planets. The nature of the simulations are hydrodynamical and are carried out both with and without radiative transfer. I have initially carried out radiation hydrodynamical simulations to investigate the evolution of massive self-gravitating discs ($M_{\rm disc}/M_{\star} = 0.1$). I consider discs with opacities ranging from $0.01\times$ to $10\times$ the interstellar Rosseland mean values. I also consider the effects of changing the initial and boundary temperatures of the discs as well as simulating different disc sizes (with outer disc radii, $R_{\rm out} = 25$ and 300 AU).

I find that the disc opacity is very important in determining whether a disc is likely to fragment. In particular, I find that fragmentation is promoted in discs with opacity values lower than interstellar Rosseland mean values since this allows radiation to leave the disc quickly. On the other hand, I find that the presence of a *thermal blanket* as a result of the stellar irradiation inhibits fragmentation since the discs are only able to cool to the boundary temperature. However, I also show that under certain circumstances, fragmentation may occur. My results demonstrate that for fragmentation, weak irradiation is required such that the boundary temperature and hence Toomre stability parameter is low, in addition to low enough opacities (even in large, cool discs) since this allows more efficient cooling so that the disc's temperature does not increase significantly (due to internal dissipation) above the boundary temperature.

Low opacities may exist in low metallicity discs or discs with larger grain sizes. This is a particularly important and timely result given the recent discoveries of wide orbit planets (Kalas et al. 2008; Marois et al. 2008) and the future emphasis for surveys of planets on such wide orbits. I show that it is possible for fragmentation to occur in gravitationally unstable discs even at radii where the innermost planet of the HR 8799 system is located ($R \gtrsim 24$ AU). Furthermore, HR 8799 is known to be a metal-poor, λ Bootis star with metallicity $[M/H] = -0.47$ (Gray & Kaye 1999) so it is reasonable to assume that its disc was similarly metal-poor. I have shown that such a scenario favours fragmentation and therefore, my results in Chapter 3 indicate that all three planets of the HR 8799 system may well have formed via gravitational instability. Though a hybrid core accretion and gravitational instability scenario for planet formation may also be a possibility for this system, my calculations in Chapter 3 show that such a hybrid scenario may not be necessary.

I have also carried out three-dimensional radiation hydrodynamical simulations to investi-

gate specifically whether the HR 8799 planetary system could have formed in situ by gravitational instability. I have shown that the formation of fragments at a specified location in a disc involves a convoluted interplay between the disc mass, star mass (and hence luminosity) and the surface mass density profile. I find that the disc masses required for fragmentation in the radial range of interest ($R \approx 24 - 68$ AU) are $\approx 0.8 - 1.2 M_{\odot}$. While in lower mass gravitationally unstable discs the surface mass density does not change significantly such that the disc temperature and cooling are the key aspects to consider with respect to the evolution of a disc (Chapter 3), in high mass discs, I find that the surface mass density profile does change which also affects the stability of a disc. The disc adjusts both the surface mass density and the temperature to try and reach a state where $Q \approx 1$. These adjustments may help to promote fragmentation.

If such massive discs, required for fragmentation to occur in the radial range of interest, are found early in the star formation process, the evolution of the fragments needs to be considered in more detail to see (i) whether the fragments grow to the required sizes, (ii) whether they evolve to different radii and (iii) whether subsequent fragments form in the disc. My simulations show that while fragments do form and evolve within the radial range of interest, further subsequent fragmentation occurs at smaller radii out of the massive disc. In addition, the fragments grow to sizes much larger than those expected for the HR8799 planetary system. Therefore, while it is possible (given a high disc mass) for fragments to form at the required radii, the growth of planets would need to be inhibited for this system to have formed by gravitational instability.

Concurrently to carrying out radiation hydrodynamical simulations, I also investigate the fragmentation criteria in self-gravitating discs modelled using a parameterised cooling technique rather than radiative transfer. Considering simplified methods such as this can provide insight that may be useful to understand more complex scenarios. I present an analytical approach to examine the fragmentation of self-gravitating protoplanetary discs, and confirm the results using global three-dimensional numerical simulations. My key result is that fragmentation does not simply depend on the disc cooling timescale, β, but also on the ratio of the surface mass density at radius, R, to the stellar mass, i.e. $\Sigma R^2 / M_{\star}$. I find that fragmentation occurs when

$$\beta < \eta \left(\frac{\Sigma R_{\rm f}^2}{M_{\star}} \right)^{\delta}, \tag{7.1}$$

where $\delta \approx 1/2$ and $\eta \approx 47$. For a power-law surface mass density, $\Sigma \propto R^{-p}$, this relation predicts the innermost radii at which subsequent fragments can form in a disc with shallow surface mass density profiles ($p < 2$) as well as the radius *outside of which fragments cannot form* in discs with steep surface mass density profiles, ($p \gtrsim 2$). Generally, I find that an increase in the steepness of the disc surface mass density profile promotes fragmentation at smaller radii. In addition, the results show that the fragmentation boundary as discussed by Clarke (2009) could vary by as much as ≈ 37 AU for low mass accretion rates. However, while the existence of a trend may remain, the exact values of η and δ are likely to be affected by the resolution. In addition, I show that the critical cooling timescale in units of the orbital timescale, $\beta_{\rm crit}$, determined from previous simulations is not a converged quantity. The simulations carried out using 16 million particles are currently the highest resolution simulations of gravitationally unstable discs currently known to have been

performed and despite such high resolution, a converged quantity currently does not exist and it remains questionable as to whether such a value would even exist. Depending on the nature of what is causing the lack of convergence, it may be possible to cautiously determine trends, upper and lower limits, relative differences between different disc scenarios, modelled at a lower resolution. However, given that I have shown that such an uncertainty exists in the fragmentation results of massive discs, earlier conclusions that are based on this theory may not necessarily hold as the values previously determined are now redundant.

7.1 Future studies

The results from Chapter 3 show crucially that a low metallicity environment is more likely to form planets via gravitational instability. Given that observed discs have inferred surface mass density profiles that are shallow ($p < 2$), it is expected that fragments are likely to have formed in the outer regions of discs. We have entered a generation where detection of planets further away from the parent star is now beginning to be possible and observational limits are constantly being pushed. The next generation of instruments and techniques are likely to populate more of the unpopulated region of the extrasolar planet mass-radius plot that I showed in Chapter 1. As this begins to occur, the theory that low metallicity and gravitational instability go hand-in-hand can then be observationally tested.

Though not discussed in this thesis, I have also carried out extensive work on shadowed discs. Such a scenario may occur if a warp or a puffed up inner region is present in a disc. Shadowed discs have been considered in the past to explain the spectral energy distributions of many discs around Herbig Ae stars. However, the concept of shadowing in the early stages of disc evolution when the disc is massive enough to be self-gravitating has not been considered. The results of Chapter 3 show that stellar irradiation acts to inhibit fragmentation. If shadowing was present, this would remove the effects of stellar irradiation from a part of, or even the whole of, the disc surface. This may potentially cause a self-gravitating disc to fragment (which may otherwise not do so in the presence of stellar irradiation). This problem is in fact numerically challenging because the boundary conditions require careful attention to ensure the conditions imposed do not play a part in the outcome of the simulations. My work to date involving this topic suggests that the optically thin discs may not be as affected by the boundary conditions as the optically thick discs. However, a careful study is required.

The simulations presented here are those of isolated discs in the early Class II stage of star formation. However, there is no reason to believe that if fragmentation was to occur then it can only take place during this stage. In fact, the earlier Class I stage is likely to involve more massive discs whose high mass is sustained and even growing due to the mass continually settling onto the disc during the embedded stage. This may then cause the discs to be more gravitationally unstable and more susceptible to fragmentation. As shown in this thesis, high mass discs ($\gtrsim 0.3\mathrm{M}_{\odot}$) clearly behave very differently to lower mass discs ($< 0.3\mathrm{M}_{\odot}$) given that significant restructuring occurs in high mass discs. The evolution of such higher mass discs are greatly understudied, even moreso in the embedded phase of disc evolution, and deserve more attention.

The results presented in Chapter 4 show that the disc masses need to be high for fragmentation to occur in the radial range of interest of the HR 8799 system. However, this can be problematic since more fragments form at a later stage in the disc's evolution and in addition, the fragment masses become too large. The latter point is almost certainly a consequence of the high disc mass and the former point may also be affected by it. The simulations in Chapter 4 used disc masses such that the discs were initially Toomre unstable over a large radial range and yet some of these simulations did not show fragments. This was partly because the discs were not able to cool rapidly enough. Therefore. while they may have satisfied the Toomre criterion, they may not have satisfied the cooling criterion. However, a key assumption in these simulations is that the opacity is not more than ten times smaller than the interstellar Rosseland mean value. If the opacity was reduced further (i.e. due to grain growth), the lower mass discs may then fragment. It would be interesting to take an approach similar to that used in Chapter 3 and determine what values of the opacity are required to allow a smaller mass disc (with a mass more in line with current observations) to fragment (which can then be used to determine what grain sizes might be expected which can also be compared to current observations).

The simulations carried out in Chapter 4 have considered massive discs. However, it is uncertain whether the subsequent fragmentation in the inner regions of the discs is purely an artefact of high mass discs or whether this is also characteristic of subsequent fragmentation in low mass discs. Should this be the case, this would bring into question the concept of whether gravitational instability is indeed only able to act in the outer regions of a disc, or whether the trigger needs to be in the outer regions so that further fragmentation at smaller radii can then take place. Given that many simulations of self-gravitating discs are stopped shortly after the fragmentation stage, the conclusion that fragmentation only occurs in the outer regions of a disc may be somewhat biased. While the subsequent fragmentation and evolution of fragments has been considered (Mayer et al. 2002; Rice et al. 2003b), it still remains a largely unexplored area and requires further investigation. In particular, understanding how earlier fragmentation affects the later fragmentation needs to be considered. In addition to exploring the subsequent fragmentation and mass growth rate of a fragment in a disc, it is also important to understand the torques on the planet and the effects on the radial migration under the influence of strong self-gravity. This is work that I have already started to carry out.

In Chapter 1, I suggest that no single planet formation mechanism can describe all the observations and therefore both methods may act but in different regimes. However, to date, no simulations have been carried out which attempt to simulate both mechanisms in the same disc, i.e. what I termed *a hybrid scenario*. This would provide useful insight into whether the two mechanisms are indeed able to operate in the same environment or whether they can only take place in mutually exclusive environments. The obvious suggestion that has been proposed in the past is that core accretion acts closer to the parent star while gravitational instability acts further away. This is an example of where the two mechanisms may operate in the same environment. However, core accretion is more likely to take place in a high metallicity environment while I have shown that gravitational instability is likely to be the dominant mechanism in the low metallicity environment. This suggests that the two mechanisms would not be able to operate in the same

disc (unless there was, for example, a radial metallicity distribution, which may occur if there was a radial dust distribution). It is also important to note that the radiative transfer simulations that have been performed here do not consider spatially varying opacities whereas in reality this might be the case. The effect of this on the fragmentation would be interesting to consider.

In Chapter 5, I showed that in a disc with a steep surface mass density profile ($p \gtrsim 2$), the fragments form in the inner regions of the disc first. Due to high computational power, these simulations were not run further beyond the stage when the first fragments formed. However, I predicted that for a $p = 2$ disc, the fragments should be able to form everywhere and in a disc where $p > 2$, subsequent fragments can form further out in a disc until the radius that determines the trendline for a particular surface mass density, star mass and cooling timescale. This prediction needs to be tested. In addition, I showed that I have begun to confirm the trend presented in Chapter 5 with a grid-based hydrodynamics code, FARGO. I intend on continuing this study with this code to confirm the trends being seen. Furthermore, in Chapter 5 I showed preliminary results on how the radius outside of which fragmentation could take place is affected by the different values of the critical cooling timescale (or equivalently, different values of the ratio $\Sigma R^2/M_\star$) when considering more realistic cooling. A more thorough investigation into this would be interesting. However, given that a particular value of the critical timescale has not been established for any particular system given the resolution issues outlined in Chapter 6, it will not be possible to determine a single radius outside of which a particular system will allow fragmentation to occur, but instead a range of values may then need to be considered.

The resolution issues identified in Chapter 6 are very worrying since many previous results and conclusions have been based on earlier results of low resolution simulations (both published and also the results presented in this thesis). It is vital to understand first and foremost, why the simulations in Chapter 6 do not converge, as this potentially affects and could somewhat nullify previous results of gravitationally unstable discs.

Appendix A

Boundary height calculation

The vertical boundary height, $z_{b,\tau=1}$ between the optically thick and optically thin region of the disc is located at the point where the optical depth, $\tau = 1$. Using the expression for the optical depth

$$\tau = -\int_{\infty}^{z_b} \kappa\rho \mathrm{d}z, \tag{A.1}$$

and using

$$\rho = \rho_o e^{-\frac{z^2}{2H^2}}, \tag{A.2}$$

where ρ_o is the density of the midplane of the disc, equation A.1 becomes

$$\tau = -\int_{\infty}^{z_b} \kappa\rho_o e^{-\frac{z^2}{2H^2}} \mathrm{d}z. \tag{A.3}$$

Using the standard result

$$\int_{t_b}^{\infty} e^{-t^2} \mathrm{d}t = \frac{\sqrt{\pi}}{2}(1 - \mathrm{erf}(t_s)), \tag{A.4}$$

equation A.3 becomes

$$\tau = \kappa\rho_o H \sqrt{\frac{\pi}{2}}\left(1 - \mathrm{erf}\frac{z_b}{\sqrt{2}H}\right) = 1, \tag{A.5}$$

which upon rearrangement, gives

$$\frac{z_b}{H} = \sqrt{2}\mathrm{erf}^{-1}\left[1 - \sqrt{\frac{2}{\pi}}\frac{1}{H\kappa\rho_o}\right]. \tag{A.6}$$

Since

$$\int_{-\infty}^{\infty} \rho \mathrm{d}z = \rho_o \int_{-\infty}^{\infty} e^{-\frac{z^2}{2H^2}} \mathrm{d}z, \tag{A.7}$$

and using the standard result

$$\int_{-\infty}^{\infty} e^{-t^2} \mathrm{d}t = \frac{\sqrt{\pi}}{2}, \tag{A.8}$$

upon rearranging, the midplane density can be expressed as

$$\rho_o = \frac{\Sigma}{H}\sqrt{\frac{2}{\pi}}, \tag{A.9}$$

which when substituted into equation A.6, gives

$$\frac{z_\mathrm{b}}{H} = \sqrt{2}\mathrm{erf}^{-1}(x) \tag{A.10}$$

where

$$x = 1 - \frac{1}{\kappa\Sigma}. \tag{A.11}$$

The opacity was calculated using the boundary temperature profile and assuming a low density in the optically thin upper layers of the disc, $\rho = 1 \times 10^{-15}\mathrm{g/cm}^2$, and the surface mass density is the initial surface mass density of the disc. The inverse error function is an odd function and is negative for negative values of x. Therefore, I use the absolute values for the boundary height:

$$\frac{z_\mathrm{b}}{H} = \sqrt{2}|\mathrm{erf}^{-1}(x)|. \tag{A.12}$$

Appendix B

Improved boundary height calculation

The vertical boundary height, $z_{\mathrm{b},\tau=1}$ between the optically thick and optically thin region of the disc is located at the point where the optical depth, $\tau = 1$. While in Appendix A, the calculation of the boundary height is only carried out at the start of a simulation, the calculation described here gives the flexibility of being able to re-evaluate the boundary height part way through a simulation. Using the expression for the optical depth

$$\tau = -\int_{\infty}^{z_{\mathrm{b}}} \kappa\rho \mathrm{d}z, \tag{B.1}$$

and calculating the opacity using the boundary temperature profile and assuming a low density, $\rho = 1 \times 10^{-15}\mathrm{g/cm}^2$, the equation becomes

$$\tau = \kappa \int_{z_{\mathrm{b}}}^{\infty} \rho \mathrm{d}z = \kappa\Sigma_{\mathrm{b}}, \tag{B.2}$$

where κ is taken out of the integral since I assume that in the optically thin region of the boundary, the disc is vertically isothermal and so is a constant at each radius. Σ_{b} is the surface mass density calculated using only the boundary particles. Since

$$\Sigma = \frac{N m_{\mathrm{p}}}{\text{area of annulus}} \tag{B.3}$$

where N is the number of particles in the boundary and m_{p} is the mass of one particle (note that all the gas particles have the same mass). If all the particles in the annulus are ordered from the furthest away from the midplane to the closest, the particle that represents the boundary between the optically thick and optically thin region is

$$N_{\mathrm{b}} = \frac{\text{area of annulus}}{\kappa m_{\mathrm{p}}}. \tag{B.4}$$

The boundary height is therefore the height at which the $N_{\mathrm{b}}^{\mathrm{th}}$ particle is located. To deal with the case where the $\tau = 1$ boundary does not exist (i.e. further out in the disc where the entire vertical extent of the disc is optically thin, I define the boundary height as being the maximum out of: the height at which the $N_{\mathrm{b}}^{\mathrm{th}}$ particle is located or the height at which 10% of the particles in the annulus being considered are in the boundary. The reasoning behind this is the same as discussed

in Section 2.2.5: that an adequate number of particles is required in the boundary region to ensure the loss of energy from the disc by radiation is modelled correctly, but also ensuring that too many particles do not make up the boundary such that the evolution of the disc is determined by the temperature profile of the boundary particles.

Appendix C

Seiss stellar evolution models

Luminosity [$L_\odot$]	Mass [$M_\odot$]
1.029×10^{-1}	0.100
3.776×10^{-1}	0.200
7.229×10^{-1}	0.300
7.835×10^{-1}	0.400
9.536×10^{-1}	0.500
1.187	0.600
1.447	0.700
1.726	0.800
2.037	0.900
2.392	1.000
2.704	1.100
3.112	1.200
3.511	1.300
3.849	1.400
4.300	1.500
4.774	1.600
5.216	1.700
5.750	1.800
6.289	1.900
6.899	2.000
8.255	2.200
1.090×10^{1}	2.500
1.349×10^{1}	2.700
2.120×10^{1}	3.000
8.126×10^{1}	3.500
3.488×10^{2}	4.000
5.733×10^{2}	5.000
1.035×10^{3}	6.000
1.737×10^{3}	7.000

Table C.1: Luminosity and corresponding star mass at an age of 1 Myr, determined using the stellar evolution models of Siess et al. (2000), using a metallicity of $Z = 0.01$

Bibliography

http://exoplanet.eu

Adams, F. C., Lada, C. J., & Shu, F. H. 1987, ApJ, 312, 788

Alexander, D. R. 1975, ApJS, 29, 363

Andre, P., Ward-Thompson, D., & Barsony, M. 1993, ApJ, 406, 122

Andrews, S. M. & Williams, J. P. 2007, ApJ, 659, 705

Ayliffe, B. A. & Bate, M. R. 2009, MNRAS, 393, 49

Backer, D. C., Foster, R. S., & Sallmen, S. 1993, Nature, 365, 817

Balsara, D. S. 1995, Journal of Computational Physics, 121, 357

Bate, M. 1995, PhD thesis, Univ. Cambridge

Bate, M. R. 1998, ApJL, 508, L95

Bate, M. R., Bonnell, I. A., & Price, N. M. 1995, MNRAS, 277, 362

Bate, M. R. & Burkert, A. 1997, MNRAS, 288, 1060

Bell, K. R. & Lin, D. N. C. 1994, ApJ, 427, 987

Benz, W. 1984, A&A, 139, 378

Benz, W. 1990, in Numerical Modelling of Nonlinear Stellar Pulsations Problems and Prospects,, ed. J. R. Buchler (Kluwer, Dordrecht), 269

Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Boley, A. C. 2009, ApJL, 695, L53

Boley, A. C., Hartquist, T. W., Durisen, R. H., & Michael, S. 2007, ApJL, 656, L89

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

Boss, A. P. 2001, ApJ, 563, 367

Boss, A. P. 2002, ApJL, 567, L149

Boss, A. P. 2007, ApJL, 661, L73

Briceño, C., Vivas, A. K., Calvet, N., et al. 2001, Science, 291, 93

Butler, R. P., Marcy, G. W., Fischer, D. A., et al. 1999, ApJ, 526, 916

Cai, K., Durisen, R. H., Boley, A. C., Pickett, M. K., & Mejía, A. C. 2008, ApJ, 673, 1138

Cai, K., Durisen, R. H., Michael, S., et al. 2006, ApJL, 636, L149

Cai, K., Pickett, M. K., Durisen, R. H., & Milne, A. M. 2010, ApJL, 716, L176

Calvet, N., D'Alessio, P., Hartmann, L., et al. 2002, ApJ, 568, 1008

Cameron, A. G. W. 1978, Moon and Planets, 18, 5

Carpenter, J. M., Mamajek, E. E., Hillenbrand, L. A., & Meyer, M. R. 2006, ApJL, 651, L49

Clarke, C. J. 2009, MNRAS, 396, 1066

Clarke, C. J. & Carswell, R. F. 2003, Principles of Astrophysical Fluid Dynamics

Clarke, C. J., Harper-Clark, E., & Lodato, G. 2007, MNRAS, 381, 1543

Cossins, P., Lodato, G., & Clarke, C. 2010, MNRAS, 401, 2587

Cossins, P., Lodato, G., & Clarke, C. J. 2009, MNRAS, 393, 1157

Courant, R., Friedrichs, K., & Lewy, H. 1928, Mathematische Annalen, 100, 32

Crida, A., Masset, F., & Morbidelli, A. 2009, ApJL, 705, L148

Dodson-Robinson, S. E., Veras, D., Ford, E. B., & Beichman, C. A. 2009, ApJ, 707, 79

Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, in Protostars and Planets V,, ed. B. Reipurth, D. Jewitt, & K. Keil (Univ. Arizona Press, Tucson), 607–622

Eisner, J. A. & Carpenter, J. M. 2006, ApJ, 641, 1162

Evrard, A. E. 1988, MNRAS, 235, 911

Fabrycky, D. C. & Murray-Clay, R. A. 2010, ApJ, 710, 1408

Fehlberg, E. 1968, NASA technical report

Forgan, D., Rice, K., Stamatellos, D., & Whitworth, A. 2009, MNRAS, 394, 882

Fukagawa, M., Hayashi, M., Tamura, M., et al. 2004, ApJL, 605, L53

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

Gingold, R. A. & Monaghan, J. J. 1977, MNRAS, 181, 375

Goldreich, P. & Lynden-Bell, D. 1965, MNRAS, 130, 97

Goldreich, P. & Tremaine, S. 1980, ApJ, 241, 425

Goldreich, P. & Tremaine, S. D. 1978, Icarus, 34, 240

Goldreich, P. & Ward, W. R. 1973, ApJ, 183, 1051

Gorti, U., Dullemond, C. P., & Hollenbach, D. 2009, ApJ, 705, 1237

Grady, C. A., Polomski, E. F., Henning, T., et al. 2001, AJ, 122, 3396

Gray, R. O. & Kaye, A. B. 1999, AJ, 118, 2993

Guillot, T. 1999, Planet. Space Sci., 47, 1183

Guillot, T. & Gautier, D. 2009, ArXiv e-prints

Gullbring, E., Hartmann, L., Briceno, C., & Calvet, N. 1998, ApJ, 492, 323

Haisch, Jr., K. E., Lada, E. A., & Lada, C. J. 2001a, AJ, 121, 2065

Haisch, Jr., K. E., Lada, E. A., & Lada, C. J. 2001b, ApJL, 553, L153

Harris, A. W. 1978, in Lunar and Planetary Inst. Technical Report, Vol. 9, Lunar and Planetary Institute Science Conference Abstracts, 459–461

Hartigan, P., Edwards, S., & Ghandour, L. 1995, ApJ, 452, 736

Helled, R., Podolak, M., & Kovetz, A. 2006, Icarus, 185, 64

Hernquist, L. & Katz, N. 1989, ApJS, 70, 419

Hollenbach, D. J., Yorke, H. W., & Johnstone, D. 2000, Protostars and Planets IV, 401

Jeffreys, H. 1924, The Earth: Its Origin, History, and Physical Constitution (Cambridge University Press)

Johnson, B. M. & Gammie, C. F. 2003, ApJ, 597, 131

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

Kornet, K., Bodenheimer, P., Różyczka, M., & Stepinski, T. F. 2005, A&A, 430, 1133

Kratter, K. M., Murray-Clay, R. A., & Youdin, A. N. 2010, ApJ, 710, 1375

Kuiper, G. P. 1951, in 50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics, ed. J. A. Hynek, 357–+

Lada, C. J., Muench, A. A., Luhman, K. L., et al. 2006, AJ, 131, 1574

Lafrenière, D., Jayawardhana, R., & van Kerkwijk, M. H. 2008, ApJL, 689, L153

Lafrenière, D., Jayawardhana, R., & van Kerkwijk, M. H. 2010, ApJ, 719, 497

Lagrange, A., Bonnefoy, M., Chauvin, G., et al. 2010, Science, 329, 57

Lagrange, A., Gratadour, D., Chauvin, G., et al. 2009, A&A, 493, L21

Levermore, C. D. & Pomraning, G. C. 1981, ApJ, 248, 321

Lissauer, J. J. & Stevenson, D. J. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil, 591–606

Lodato, G., Meru, F., Clarke, C. J., & Rice, W. K. M. 2007, MNRAS, 374, 590

Lodato, G. & Rice, W. K. M. 2004, MNRAS, 351, 630

Lodato, G. & Rice, W. K. M. 2005, MNRAS, 358, 1489

Loewenstein, M. & Mathews, W. G. 1986, Journal of Computational Physics, 62, 414

Lommen, D., Wright, C. M., Maddison, S. T., et al. 2007, A&A, 462, 211

Lubow, S. H. & Ida, S. 2010, ArXiv e-prints

Lucy, L. B. 1977, AJ, 82, 1013

Lynden-Bell, D. & Kalnajs, A. J. 1972, MNRAS, 157, 1

Mamajek, E. E., Meyer, M. R., Hinz, P. M., et al. 2004, ApJ, 612, 496

Marois, C., Macintosh, B., Barman, T., et al. 2008, Science, 322, 1348

Matzner, C. D. & Levin, Y. 2005, ApJ, 628, 817

Mayer, L. & Gawryszczak, A. J. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 398, Astronomical Society of the Pacific Conference Series, ed. D. Fischer, F. A. Rasio, S. E. Thorsett, & A. Wolszczan, 243–+

Mayer, L., Lufkin, G., Quinn, T., & Wadsley, J. 2007, ApJL, 661, L77

Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. 2002, Science, 298, 1756

Mayor, M. & Queloz, D. 1995, Nature, 378, 355

Meglicki, Z., Wickramasinghe, D., & Bicknell, G. V. 1993, MNRAS, 264, 691

Mizuno, H. 1980, Progress of Theoretical Physics, 64, 544

Mohanty, S., Jayawardhana, R., & Basri, G. 2005, ApJ, 626, 498

Monaghan, J. J. 1992, ARA&A, 30, 543

Monaghan, J. J. 2002, MNRAS, 335, 843

Monaghan, J. J. & Gingold, R. A. 1983, Journal of Computational Physics, 52, 374

Monaghan, J. J. & Lattanzio, J. C. 1985, A&A, 149, 135

Morris, J. P. & Monaghan, J. J. 1996, J. Comp. Phys.

Murray, N., Hansen, B., Holman, M., & Tremaine, S. 1998, Science, 279, 69

Muzerolle, J., Hillenbrand, L., Calvet, N., Briceño, C., & Hartmann, L. 2003, ApJ, 592, 266

Nero, D. & Bjorkman, J. E. 2009, ApJL, 702, L163

P. Bodenheimer, G. P. Laughlin, M. Rózyczka, & H. W. Yorke , ed. 2007, Numerical Methods in Astrophysics: An Introduction

Pascucci, I., Gorti, U., Hollenbach, D., et al. 2006, ApJ, 651, 1177

Perri, F. & Cameron, A. G. W. 1974, Icarus, 22, 416

Pollack, J. B. 1984, ARA&A, 22, 389

Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icarus, 124, 62

Pollack, J. B., McKay, C. P., & Christofferson, B. M. 1985, Icarus, 64, 471

Price, D. J. 2007, Publications of the Astronomical Society of Australia, 24, 159

Price, D. J. & Bate, M. R. 2007, MNRAS, 377, 77

Price, D. J. & Monaghan, J. J. 2004, MNRAS, 348, 139

Price, D. J. & Monaghan, J. J. 2007, MNRAS, 374, 1347

Pringle, J. E. 1977, MNRAS, 178, 195

Pringle, J. E. 1981, ARA&A, 19, 137

Pringle, J. E. & Rees, M. J. 1972, A&A, 21, 1

Rafikov, R. R. 2005, ApJL, 621, L69

Rafikov, R. R. 2009, ApJ, 704, 281

Rasio, F. A. & Ford, E. B. 1996, Science, 274, 954

Raymond, S. N., Barnes, R., Armitage, P. J., & Gorelick, N. 2008, ApJL, 687, L107

Rice, W. K. M., Armitage, P. J., Bate, M. R., & Bonnell, I. A. 2003a, MNRAS, 339, 1025

Rice, W. K. M., Armitage, P. J., Bonnell, I. A., et al. 2003b, MNRAS, 346, L36

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. 2004, MNRAS, 355, 543

Rodmann, J., Henning, T., Chandler, C. J., Mundy, L. G., & Wilner, D. J. 2006, A&A, 446, 211

Safronov, V. S. 1969, Evoliutsiia doplanetnogo oblaka. (Nakua, Moscow)

Saumon, D. & Guillot, T. 2004, ApJ, 609, 1170

Semenov, D., Henning, T., Helling, C., Ilgner, M., & Sedlmayr, E. 2003, A&A, 410, 611

Shakura, N. I. & Syunyaev, R. A. 1973, A&A, 24, 337

Shu, F. H., Adams, F. C., & Lizano, S. 1987, ARA&A, 25, 23

Sicilia-Aguilar, A., Hartmann, L. W., Fürész, G., et al. 2006, AJ, 132, 2135

Siess, L., Dufour, E., & Forestini, M. 2000, A&A, 358, 593

Spiegel, D. S., Burrows, A., & Milsom, J. A. 2010, ArXiv e-prints

Springel, V. & Hernquist, L. 2002, MNRAS, 333, 649

Stamatellos, D. & Whitworth, A. P. 2008, A&A, 480, 879

Stamatellos, D. & Whitworth, A. P. 2009, MNRAS, 392, 413

Steinmetz, M. & Mueller, E. 1993, A&A, 268, 391

Strom, S. E., Edwards, S., & Skrutskie, M. F. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine, 837–866

Su, K. Y. L., Rieke, G. H., Stapelfeldt, K. R., et al. 2009, ApJ, 705, 314

Testi, L., Natta, A., Shepherd, D. S., & Wilner, D. J. 2003, A&A, 403, 323

Toomre, A. 1964, ApJ, 139, 1217

Trilling, D. E., Benz, W., Guillot, T., et al. 1998, ApJ, 500, 428

van Leeuwen, F. 2007, A&A, 474, 653

Veras, D., Crepp, J. R., & Ford, E. B. 2009, ApJ, 696, 1600

Ward, W. R. 1997, Icarus, 126, 261

Weidenschilling, S. J. 1977, MNRAS, 180, 57

Weidenschilling, S. J. & Marzari, F. 1996, Nature, 384, 619

Whitehouse, S. C. & Bate, M. R. 2006, MNRAS, 367, 32

Whitehouse, S. C., Bate, M. R., & Monaghan, J. J. 2005, MNRAS, 364, 1367

Wolszczan, A. & Frail, D. A. 1992, Nature, 355, 145

Wood, D. 1981, MNRAS, 194, 201

Yasui, C., Kobayashi, N., Tokunaga, A. T., Saito, M., & Tokoku, C. 2009, ApJ, 705, 54