SHEDDING LIGHT ON THE FORMATION OF STARS AND PLANETS

SHEDDING LIGHT ON THE FORMATION OF STARS AND PLANETS: NUMERICAL SIMULATIONS WITH RADIATIVE TRANSFER

By

PATRICK D. ROGERS, B.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2012) McMaster University
(Physics & Astronomy) Hamilton, Ontario

TITLE: Shedding Light on the Formation of Stars and Planets: Numerical Simulations With Radiative Transfer

AUTHOR: Patrick D. Rogers, B.Sc.(Saint Mary's University)

SUPERVISOR: Dr. James Wadsley

NUMBER OF PAGES: xii, 186

Abstract

We use numerical simulations to examine the fragmentation of protostellar discs via gravitational instability (GI), a proposed formation mechanism for gas-giant planets and brown dwarfs. To accurately model heating and cooling, we have implemented radiative transfer (RT) in the TreeSPH code Gasoline, using the flux-limited diffusion approximation coupled to photosphere boundary cooling. We present 3D radiation hydrodynamics simulations of discs that are gravitationally unstable in the inner 40 AU; these discs do not fragment because the cooling times are too long. In prior work, one of these discs was found to fragment; however, we demonstrate that this resulted from an over-estimate of the photosphere cooling rate. Fragmentation via GI does not appear to be a viable formation mechanism in the inner 40 AU.

We also present simulations of GI in the outer regions of discs, near 100 AU, where we find GI to be a viable formation mechanism. We give a detailed framework that explains the link between cooling and fragmentation: spiral arms grow on a scale determined by the linear gravitational instability, have a characteristic width determined by the balance of heating and cooling, and fragment if this width is less than twice their Hill radius. This framework is consistent with the fragmentation and initial fragment masses observed in our simulations. We apply the framework to discs modelled with the commonly-used β-prescription cooling and calculate the critical cooling rate for the first time, with results that are consistent with previous estimates measured from numerical experiments.

RT is fundamentally important in the star formation process. Non-ionizing radiation heats the gas and prevents small-scale fragmentation. Ionizing radiation from massive stars is an important feedback mechanism and may disrupt giant molecular clouds. We present methods and tests for our implementation of ionizing radiation, using the Optically-Thin Variable Eddington Tensor method.

Acknowledgements

This thesis is the culmination of work that has come together over the past six years. I owe a great thanks to the many people who have made it possible to carry out this work, and who have ensured that I had a good time doing it. It is impossible to thank everyone individually here, but I hope that I am able to convey my sense of gratitude.

I begin by thanking my supervisor, James. I largely based my decision to attend graduate school here at McMaster on my choice of supervisor, and I have always been happy with this choice. I have appreciated James' availability, his willingness to chat, and his insight into issues of both physics and numerics. I have admired his interest in astrophysics on every scale, from cosmological to planetary, and am certainly a more well-rounded astrophysicist as a result of our discussions. I must also add that I enjoy his accurate passing on the soccer pitch.

I also thank Ralph and Christine who, in addition to being on my supervisory committee, have taught me a number of courses. I appreciate their insight, interest, and useful comments regarding my research. In addition, I thank our collaborators, Tom Quinn and Lucio Mayer, who have been incredibly kind and helpful.

I have had a very positive time here at McMaster. In my experiences, the faculty, staff, and students of the department, and the astronomy group in particular, have come together to create an atmosphere of helpfulness, curiosity, collaboration, and sociability. It has been easy to make friends, which means that I can't possibly list all of them here. I thank my office-mates Rory and Rachel; the other members of our research group, particularly Sijing, Nic, Greg, and Elizabeth; as well as the other physicists and astronomers that I have studied with, particularly Charles-Phillippe, Nathan, Rob, Josh, Mike, Phil, Clare, Brendan, Blair, Annie, Andreas, Laura, Kate, and Allan.

I thank my family, who have instilled in me the curiosity and work-ethic that have brought this thesis to fruition. Above all, I thank my parents. Although they are not here to see the completion of this thesis, they are always in my thoughts.

Finally, I thank my wife Claire. She has been a tremendous part of my life over

the course of this work. I thank her for her patience, support, and understanding, particularly during those times when I had my nose stuck in a book or a computer for long periods: writing paper drafts, studying for comprehensives, and writing this thesis. She has lent me her ear many times, and is by now, without a doubt, an expert on the details of gravitational instability, numerical simulations, and debugging the Gasoline code. I am lucky to have her in my life.

Preface

Chapters 3 and 4 of this thesis are papers written by myself, Patrick D. Rogers, and previously published in *Monthly Notices of the Royal Astronomical Society* (*MNRAS*). Chapter 3 appeared as Rogers, P.D. & Wadsley, J.W. 2011, *MNRAS*, 414, 913-929 (doi:10.1111/j.1365-2966.2011.18523.x), while Chapter 4 appeared as Rogers, P.D. & Wadsley, J.W. 2012, *MNRAS*, 423, 1896-1908 (doi:10.1111/j.1365-2966.2012.21014.x). I wrote the papers and performed the research within, under the supervision of the co-author, James Wadsley. The previously published material has been reformatted in the style of the thesis. I grant an irrevocable, non-exclusive license to McMaster University and the National Library of Canada to reproduce the material as part of this thesis.

Contents

List of Figures

Chapter 1

Introduction

Radiative transfer (RT), the transport of electromagnetic radiation (including for example, visible light), is one of the fundamental physical processes in astronomy. For many of us, one of the first, and most basic, ways by which we experience astronomy is deeply rooted in this process: when we look up at the sky on a clear night, we see photons emitted from the surfaces of stars and transported to our retinae across the vast distances that separate these stars from the Earth.

RT is the primary means by which we learn about the universe. Telescopes collect light from astronomical sources, and the analysis of this light, the location of its source, its intensity, and its dependence on wavelength, informs us about the structure and physical state of the regions observed, including their densities, temperatures, chemistry, and motions.

Theoretical models are constructed to address the questions that arise from astronomical observations. These include fundamental questions of formation, such as how galaxies, stars, and planets are created. A detailed comparison between the predictions of theoretical models and observations allows us to determine which models best describe the cosmos. Because of the complexity of astrophysical systems, numerical simulations have become an indispensable tool in our construction of theoretical models.

In this thesis, we consider three-dimensional, dynamic (time-evolving) simulations of systems whose evolution depends on radiative transfer. The absorption of radiation

leads to heating, while its emission leads to cooling. Thermal pressure stabilizes gas against gravitational collapse; consequently, radiative transfer is a key process in determining the growth of structure, such as the formation of stars and planets.

It is an exciting era in the study of planet formation. The first planet around a solar type star was discovered only in 1995. In the fewer than two decades that have passed, over 3000 confirmed and candidate extrasolar planets have been detected. These observations have shown a great variety in the architecture of planetary systems, and offer valuable data against which we can compare theoretical models of planet formation.

The two prominent theories of gas-giant (ie. Jupiter-like) formation are the fragmentation via gravitational instability (GI) model, and the core accretion model. In both models, gas-giants (in which we include both planets and brown dwarfs, substellar companions with masses greater than 13 $M_{\rm Jupiter}$) form in the protostellar discs of gas and dust that surround young stars. Fragmentation via GI is a top-down process, in which a region of a massive protostellar disc is unstable, and collapses to form a gas-giant relatively quickly, in thousands of years. In constrast, the core-accretion model is a bottom-up process, in which a protoplanet grows through the conglomeration of the solids in the protostellar disc. When the protoplanet reaches sufficient mass, runaway accretion of the gaseous component occurs. This process is slow, taking millions of years. In Chapter 2, we review our current understanding of protostellar discs and gas-giant formation, from both observational and theoretical perspectives.

In this thesis, we present original research on the formation of gas-giants via GI. The conditions for a disc to be gravitationally unstable have previously been determined analytically. However, study of the transition from instability to fragmentation requires the use of numerical simulations. Previous work has shown that gravitationally unstable discs fragment only if they cool quickly. Realistic heating and cooling, and thus RT, must therefore be included in numerical simulations studying fragmentation via GI. In Chapter 3 we present our implementation of RT in the gravito-hydrodynamics code Gasoline, using the flux-limited diffusion approximation coupled to radiative cooling from a photosphere boundary condition. It is important

to carefully test simulation codes in order to be confident in their output. We present results for a number of standard test problems.

In Chapter 3, we also present 3D radiation hydrodynamics simulations of GI in the inner regions of protostellar discs, inside of 40 AU. Prior numerical simulations including RT have disagreed on the viablity of fragmentation via GI in this region. We show that fragmentation occurs in one of these simulations because of an overestimate of the radiative cooling rate in the disc. We find that fragmentation via GI is not a viable gas-giant formation mechanism inside of 40 AU.

In Chapter 4, we present a suite of 3D radiation hydrodynamics simulations of GI in the outer regions of protostellar discs, near 100 AU. These simulations show that fragmentation via GI is a viable gas-giant formation mechanism in this region. In addition, we use our suite of simulations to develop an improved understanding of the link between cooling and fragmentation. We present a framework in which spiral arms develop on a scale determined by the linear gravitational instability, have a characteristic width determined by the balance of heating and cooling, and fragment if this width is less than twice their Hill radius. This model is consistent with the fragmentation and initial fragment masses observed in our simulations. We use the model to calculate the critical cooling rate required for fragmentation, for the first time. The results are consistent with previous estimates based on numerical simulations.

RT is also fundamentally important in star formation. Heating from non-ionizing radiation affects the initial mass function of stars by preventing small-scale fragmentation and the over-production of brown dwarfs. Feedback from ionizing radiation produced by massive stars may disrupt the parent giant molecular cloud, and limit the overall star formation efficiency. In Chapter 5, we give a brief overview of some recent research that demonstrates the importance of RT in star formation. Non-ionizing radiation is well-modelled by our implementation of RT presented in Chapter 3. In Chapter 6, we present methods and tests for our implementation of ionizing radiation, using the OTVET method, in Gasoline.

In Chapter 7, we summarize the main results of this thesis. Our implementation of RT in Gasoline has expanded our numerical toolkit, making it possible to more

accurately simulate systems in which RT plays an important role. We conclude with a brief overview of some future research projects that make use of this expanded numerical toolkit to address interesting questions in galaxy, star, and planet formation.

Chapter 2

Protostellar discs and the formation of gas-giants

2.1 Introduction

The two most prominent theories of planet formation are the core accretion model and the fragmentation via gravitational instability (GI) model. The core accretion model explains the formation of all planets, including terrestrial, ice-giant, and gas-giant, within the same framework. In this framework, km-size planetesimals are built up through the agglomeration of dust grains within a disc of gas and dust that surrounds a central star. Two-body collisions, enhanced by gravitational attraction, then lead to the growth of some planetesimals to 1000-km protoplanetary sizes.

Further accretion of planetesimals, and possibly gas, leads to the continued growth of protoplanets. If a protoplanet grows large enough, to a crossover mass of roughly 10-15 $M_{\oplus}$, while there remains significant gas in the disc, runaway gas accretion occurs, and a gas-giant is formed. Protoplanets that are not able to reach the crossover mass before the dissipation of the gas disc result in the terrestrial or ice-giant planets.

Fragmentation via GI explains the formation of the gas-giant planets, though recent work has also suggested evolution subsequent to fragmentation may also explain the formation of solid planets. In this model, a massive disc of gas and dust surrounding a central star can be prone to gravitational instability, in which local regions of

the disc have self-gravity that is comparable to the stabilizing effects of the coriolis and pressure forces. Unstable discs form spiral overdensities that may then fragment to form gas-giant planets, depending on the thermodynamics of the disc. In contrast to the core accretion model, formation via GI happens very quickly, taking only a few orbital times. This can lead to the formation of gas-giants in hundreds, or thousands of years, rather than millions of years necessary in the core accretion model.

The direct formation of a gas-giant through fragmentation of a gas and dust disc via GI is the most prominent manner in which planet formation may be influenced by instabilities, but it is not the only one. A number of instabilities may play important roles in the formation of planetesimals. These processes include the gravitational fragmentation of a thin dust layer (Goldreich & Ward 1973; Youdin & Shu 2002), and the concentration of dust grains in the spiral arms of gravitationally unstable discs (Rice et al. 2004), in vortices (Klahr & Bodenheimer 2003; Mamatsashvili & Rice 2009), and in turbulent overdensities (Johansen et al. 2006) enhanced by the streaming instability (Youdin & Goodman 2005).

In both the core accretion and gravitational fragmentation models, planets are formed within the protostellar (or alternatively, protoplanetary) discs that surround young stars, including brown dwarf primaries. The properties of these discs— their sizes, temperatures, mass distributions, and lifetimes— are fundamental inputs, and constraints, to models of the planet formation process. We therefore begin, in Section 2.2, by reviewing the current research, both observational and theoretical, on the properties of protostellar discs, as well as their formation, evolution, and the relevant physical processes involved.

In Sections 2.3 and 2.4, we give overviews of the formation of gas-giants through core accretion and through GI. From this point on, we specifically use the term gas-giant companion, rather than gas-giant planet, as we are interested in the formation of gas-giants both below and above the mass limit for deuterium fusion (13 M_{Jup}), which separates planets from brown dwarfs. We are interested in the formation of gas-giants in protostellar discs (as opposed to formation in the prestellar core or molecular cloud), and not in a specific mass range.

Recent years have shown a dramatic increase in the number of observed gas-

giant companions. These observations are invaluable, as they provide much-needed constraints on theories of planet formation. In Section 2.5, we overview some of the observations of gas-giants and discuss some of the successes and limitations in explaining these observations through the current models of gas-giant formation.

2.2 Protostellar discs

Protostellar discs are intimately linked to both the formation of stars, as well as the formation of planets. Observations of the velocity gradients across prestellar cores suggest that cores have rotational motion (Goodman et al. 1993); this motion may reflect internal turbulence (Burkert & Bodenheimer 2000). During the gravitational collapse of a rotating prestellar core, the initial angular momentum of the core prevents a radial collapse of material forming the central star directly. Instead, a disc is built up over time from the inside out, as material from farther away in the core, with ever-higher initial angular momentum, collapses to the disc-plane at ever-larger radii. Extended protostellar discs have been directly imaged in Orion (O'Dell et al. 1993) using the Hubble Space Telescope (HST), and have been inferred through observations of mid-infrared excess around young stars (Strom et al. 1989).

Because of the angular momentum initially present in the prestellar core, much of the star's final mass must be accreted through the disc. As evidence of this, theoretical models of stellar accretion along magnetic field lines that thread a protostellar disc agree well with observed line profiles (Muzerolle et al. 2001). The evolution of protostellar discs, and the transport of material through them (see review by Armitage 2011) thus control the formation of the star itself.

In addition to controlling accretion onto the central star, protostellar discs are also the birthplaces of planetary systems. In this context, discs are often referred to as "protoplanetary". The terms "protostellar disc" and "protoplanetary disc" refer to the same structure, but at different times. The disc is protostellar during the early phases, when the disc mass may be comparable to the nascent protostar, and it is protoplanetary in the later stages, when the disc mass may only be a few percent of that of the star. These early phases (class 0 and I) last roughly 0.7 Myr (see Sections

2.2.1 and 2.2.4). In the context of this thesis, in which the focus will predominately be on massive discs, we will exclusively use the term protostellar, even in reference to low mass discs.

The fact that the planets in the solar system orbit around the Sun in the same sense, and that they orbit with small inclinations to a unique plane are good indications that the solar system formed out of a protostellar disc that initially surrounded the sun. The concept of a minimum-mass solar nebular (MMSN: Weidenschilling 1977; Hayashi 1981), computed by taking the planetary compositions at each radius and extrapolating to solar abundances, is a useful baseline for considering protostellar discs. For reference, the MMSN has a mass of 0.01-0.07 $M_{\odot}$ and a surface density profile $\Sigma \propto R^{-3/2}$, where R is the radius from the sun.

2.2.1 Classification

The significant amount of study of protostellar discs that has taken place has resulted in a corresponding body of terminology. In order to clarify this terminology, Evans et al. (2009a) have compiled the nomenclature into a useful "diskionary." In order to frame some of the discussion in the rest of this section, we begin by describing the classification scheme for protostellar discs from both an observational and theoretical point of view.

During the early-stages of the collapse of a prestellar core, the still-forming protostar and protostellar disc are obscured by the large column of material in the envelope that has yet to collapse. The large column of material, combined with the large opacity from micron-sized dust grains, leads to very large visual extinctions. Consequently, investigations of this embedded phase require observations at longer wavelengths (infrared, sub-millimetre, and millimetre), at which the opacity from dust grains is smaller.

Lada & Wilking (1984), using near-IR observations of the nearby star forming region ρ-Ophiucus, found that young stellar objects (YSOs) could be placed within three categories based on the slope of the spectral energy distribution (SED) at wavelengths in the near-IR. A classification scheme based on this result was put forward

by Lada (1987): class I sources show an SED that is broader than blackbody and increasing longward of 2 microns; class II sources show an SED that is broader than blackbody and constant or decreasing longward of 2 microns; while class III sources show an SED that is well-modelled by a reddened blackbody, with perhaps a small excess in the mid-IR.

This classification scheme has been explained in terms of an evolutionary sequence. Class I YSOs represent the early phase of collapse, in which the emission is dominated by the material in the envelope; class II YSOs have emission that is dominated by the star and the protostellar disc; and class III YSOs have emission that is characteristic of the star, with no strong evidence for a protostellar disc, although a debris disc may exist. This correspondence between observational classification and evolutionary state agrees with models of the collapse of rotating cores (Adams et al. 1987). A further extension of the classification scheme was introduced by Andre et al. (1993): class 0 YSOs represent the earliest phases of core collapse, during which the envelope emission (and mass) exceeds that of the star. Figure 2.1 depicts the classification scheme for YSOs.

Discs in the class 0-I phases are deeply embedded, with significant emission from the envelope. As a consequence, disentangling emission arising from the envelope, which may also include material in an outflow, from emission arising from the disc, is non-trivial. Separating the disc from the envelope requires molecular-line observations that can clearly show the rotation of the disc material, or detailed radiative transfer (RT) modelling. Due to these observational difficulties, we begin by discussing the later phases of protostellar discs, which are less obscured by the envelope, and therefore more amenable to study. We focus on observations of class II discs, in which the disc and star dominate the emission, in Section 2.2.2. These observations give us important information on the masses, and radial structure and size, and vertical structure of these discs. We focus on the lifetimes of these discs, as well as their evolution and dispersal, in Section 2.2.3. Finally, we address the earliest phases of protostellar discs, with recent observations of class 0-I discs in §2.2.4, and a discussion of the recent progress in modelling the formation of class 0-I discs in Section 2.2.5.

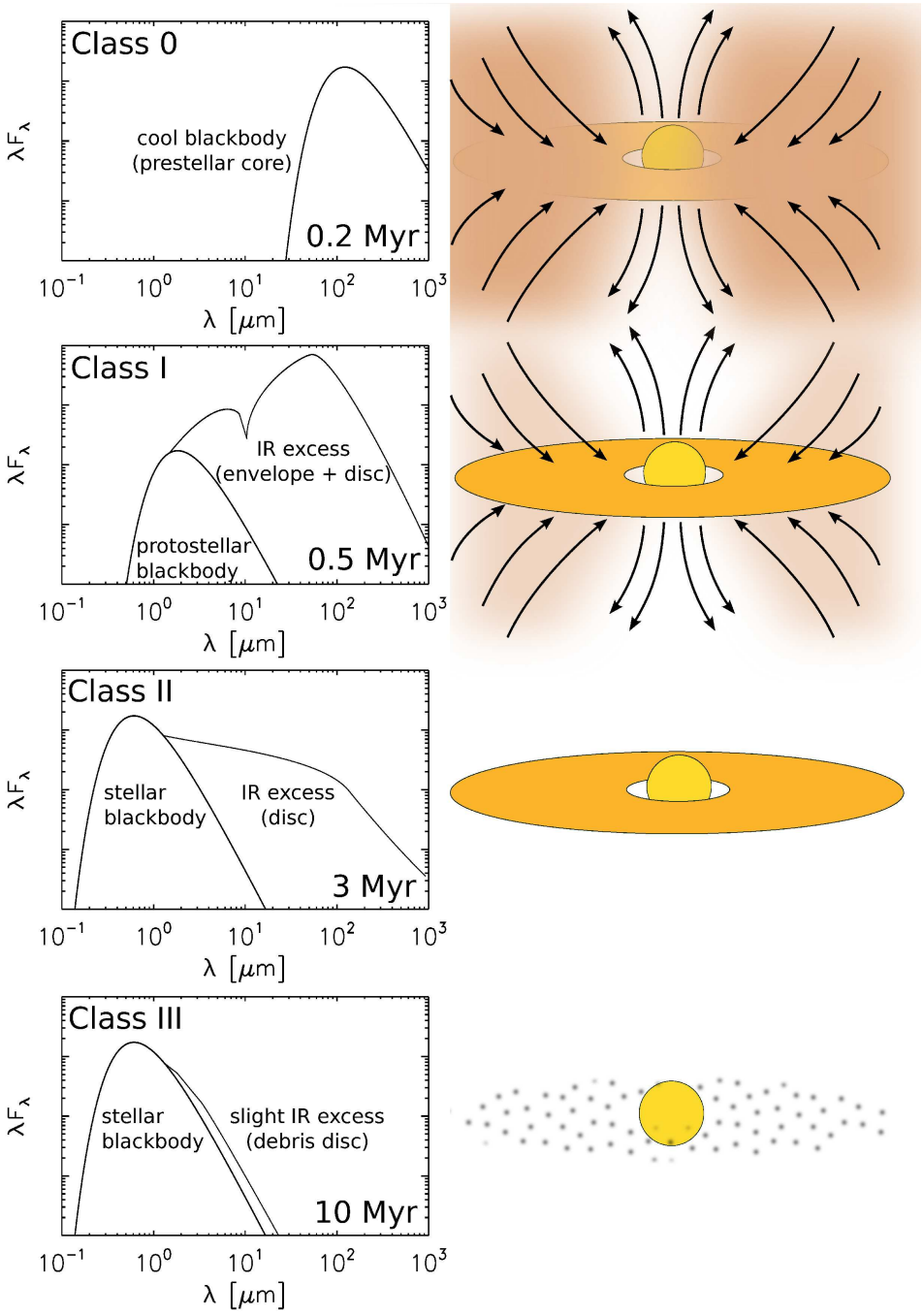

Figure 2.1: A depiction of the classification scheme for YSOs (and protostellar discs, Lada 1987; Andre et al. 1993) and the typical lifetime of each phase. The figure is based on one from Dauphas & Chaussidon (2011).

2.2.2 General characteristics of class II discs

The masses of class II discs can be determined from observations of dust emission in YSOs at mm and sub-mm wavelengths. At these wavelengths, much of the disc emission (roughly that originating from greater than 10 AU in the MMSN; Williams & Cieza 2011) is expected to be optically-thin. There is a relatively simple relation between the observed flux due to optically-thin emission and the total disc-mass (Beckwith et al. 1990), which has been used in a number of works (e.g. Andrews & Williams 2005, 2007b) to study the statistics of disc-masses for YSOs in nearby star forming regions. The result of these studies is that the median disc-to-star mass in the class II phase is relatively small, about 0.9 per cent (Williams & Cieza 2011). Significant uncertainties in the opacities used in the determination of these masses, largely due to the possible presence of dust grains larger than mm-size (Williams & Cieza 2011) make these mass estimates somewhat uncertain. Furthermore, mass-estimates based on observed accretion rates (Hartmann et al. 1998) tend to be an order of magnitude larger than optically-thin estimates (Andrews & Williams 2007b). This, along with the high occurence of observed exoplanets (Cassan et al. 2012), and the difficulty of forming planets in such low-mass discs, suggests that optically-thin observations may underestimate the total disc-mass, or that planet formation has already made significant progress by the class II phase (Greaves & Rice 2010).

Resolving the scale of protostellar discs in nearby star-forming regions at sub-mm wavelengths requires the use of interferometry. A combination of resolved interferometric continuum observations and SEDs have been used to determine the radial structure (in addition to mass) of a number of protostellar discs. Initial studies (e.g. Kitamura et al. 2002; Andrews & Williams 2007a) modelled the discs using power-laws in the surface density, $\Sigma \propto R^{-p}$, and the temperature, $T \propto R^{-q}$. These studies have generally found p in the range 0-1 (Williams & Cieza 2011) and q in the range 0.5-0.7. More recent work (Isella et al. 2009; Andrews et al. 2009) has instead modelled the surface density profiles with the self-similar solution for a viscous disc (Lynden-Bell & Pringle 1974; Hartmann et al. 1998),

$$\Sigma \propto (R/R_c)^{-\gamma} \exp\left[-(R/R_c)^{(2-\gamma)}\right], \tag{2.1}$$

where R_c is a scaling radius, interior to which lies roughly two-thirds of the disc mass. Andrews et al. (2009) found γ to be in the range (0.4, 1), while Isella et al. (2009) found a larger range of (-0.8, 0.8), where the values of $\gamma < 0$ represent inner regions of the disc that have increasing surface density as a function of radius. These are examples of "transitional discs."

The above interferometric studies also give constraints on the sizes of protostellar discs. Andrews et al. (2009) found R_c in the range 14-198 AU, while Isella et al. (2009) found R_c in the range 30-230 AU. This range of protostellar disc sizes is in agreement with HST observations of discs in Orion (Vicente & Alves 2005). From these observations, it is clear that protostellar discs can be very extended.

The vertical structure of protostellar discs is determined by the hydrostatic balance between the vertical component of gravity (from the central star and the self-gravity of the disc) and the vertical pressure gradient (from the density and temperature profiles, and possibly turbulence). The vertical temperature structure of the disc is set by a balance of heating and cooling, which are dominated by dust, due to its large opacity (line cooling is only a minor coolant, Dullemond et al. 2007). Dust is heated by the absorption of irradiation from the central star, absorption of reprocessed IR emission from nearby dust, and through viscous dissipation; it cools via IR emission.

Kenyon & Hartmann (1987) first determined that protostellar discs are flared, based on observations of the far-infrared excess in YSOs. A flaring geometry allows more of the star's light to be reprocessed by dust grains at large radii in the disc, where the temperature is low, thus contributing to the emission in the far-infrared. The flaring geometry of discs was later confirmed directly through HST images of protostellar discs, such as the "butterfly star" in Taurus (Wolf et al. 2003).

One expects an optically-thin layer of superheated dust on the surface of protostellar discs. This layer of dust is directly exposed to the radiation from the central star, and reprocesses this stellar light into the IR, heating the interior of the disc. Using a flared disc composed of two layers, a superheated surface layer, and a cooler interior layer, Chiang & Goldreich (1997) derived a solution for the vertical structure of a non-accreting protostellar disc assuming hydrostatic balance. Their model successfully matched observed SEDs, including the far-IR excess. More complex,

"one + one-dimensional" computational models employ vertical radiative transport in an azimuthally symmetric disc, in order to determine the vertical structure self-consistently and have also been used to model observed SEDs. D'Alessio et al. (1998) employed a 1+1D model including diffusive vertical radiative transport, as well as viscous dissipation to study the vertical structure of discs. From their work, three zones in the disc emerge: an outer zone dominated by irradiation, an intermediate zone where the surface temperature is determined by irradiation, but the midplane temperatue is determined by viscous dissipation, and an inner zone dominated by viscous dissipation.

We also note that fully two-dimensional (azimuthally symmetric) RT models have been developed for disc studies (e.g. Dullemond & Dominik 2004). These generally iterate between solving radiative transfer and pressure balance. Their advantage is that they can take into account radial diffusion of radiation, as well as self-shadowing effects. A number of models also include processes such as line-cooling, chemical networks, and/or distinct dust and gas temperatures in order to study the upper layers of protostellar discs (see the review by Dullemond et al. 2007).

2.2.3 Lifetimes and evolution of discs

The lifetimes of protostellar discs provide a fundamental constraint on any model of gas-giant formation: the gas-giant must be able to form within the lifetime of the disc.

Observations of a number of nearby clusters show that discs, the inner regions detected through their near-IR excess, have lifetimes of typically 3 Myr (Haisch et al. 2001). Furthermore, sub-mm observations (Andrews & Williams 2005, 2007b) have demonstrated that the outer-regions must be dispersed at roughly the same time as the inner regions, since the fraction of discs detected at sub-mm wavelengths (characteristic of the outer disc) is nearly identical to the fraction of discs detected in the near-IR. In some instances, disc lifetimes can be quite short. The Spitzer c2d survey ("From Cores to Discs" Cieza et al. 2007) examined 230 weak-lined T-Tauri stars (WTTSs, corresponding to Class III YSOs), stars that show no signs

of accretion (and thus have no evidence for an inner disc), in an effort to detect if any discs existed around these stars. The authors found that 50% of the youngest WTTSs had no evidence of having discs, indicating that many discs are dissipated on timescales of 1 Myr. In addition, the authors determined a transition timescale between optically thick discs and an absence of disc of 0.4 Myr. The fact that the disc lifetime is observed to be much longer than the time to clear the entire disc once the inner disc is gone is described as the "two-time-scale" problem (see further references in Williams & Cieza 2011).

The observations described above focused on the dust content of protostellar discs, since it is the easiest component to observe. Observations of accretion rates in young stars indicate that the timescale for accretion to end is roughly 2.3 Myr, with 60% of stars still accreting at an age of 2 Myr, and 2% accreting at an age of 10 Myr (Fedele et al. 2010). Although accretion only describes the innermost regions of the disc, this timescale for the lifetime of the gas disc is consistent with the inferred lifetimes based on observations of dust.

The limited lifetimes (a few Myr), and short dispersal time (a fraction of a Myr) show that protostellar discs evolve. In addition, observations of emission lines from young stars are consistent with accretion onto the star from the disc (Muzerolle et al. 2001), indicating mass transport. Here, we describe some of the mechanisms responsible for evolution of the disc.

Mass transport through the disc requires the transport of angular momentum. Shakura & Sunyaev (1973) argued that turbulence, whatever its source, results in a viscosity that scales as

$$\nu = \alpha c_{\rm s} H_{\rm NSG}, \tag{2.2}$$

where $H_{\rm NSG} = c_{\rm s}/\Omega$ is the non-self-gravitating scale height ($c_{\rm s}$ is the sound speed and Ω is the rotation rate). This parameterization is known as an "α-viscosity." The time evolution of the surface density of a viscous accretion disc (see Frank et al. 2002; Armitage 2011) is

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu \Sigma r^{1/2}\right)\right]; \tag{2.3}$$

in addition to mass transport, there is also a viscous heating per surface area of

$9/8\nu\Sigma\Omega^2$. In general, this viscosity drives accretion of material onto the star, while simultaneously spreading the disc to large radii. Changes to the disc structure take place on roughly the viscous timescale, $t_{\rm visc} = r^2/\nu$. If steady-state is assumed, one can balance the viscous heating (and irradiation) from the central star and the disc emission to compute the steady-state temperature structure of the disc. The numerical models of D'Alessio et al. (1998), which use an α-viscosity model, are one example of such a model.

It is not necessary to know the physical mechanism for viscosity in order to apply the α-viscosity model. Values for α can be determined by comparing α-viscosity models with observed accretion rates, yielding values of $\alpha \sim 10^{-2}$ (Hartmann et al. 1998). However, it is worthwhile to understand the source of viscosity in order to understand the applicability of an α-viscosity, as well as confirm the observed values. The magneto-rotational instability (MRI, Balbus & Hawley 1991) is one likely source of viscosity in protostellar discs. In the ideal-MHD limit, one expects the MHD turbulence that results from the MRI to be well-characterized as an α-viscosity, because flux-freezing makes the process inherently local in nature (Balbus & Papaloizou 1999). The importance of non-ideal effects, such as weak coupling between the magnetic field lines and the gas complicate the study of the MRI, but can also lead to interesting features, such as "dead-zones" (Gammie 1996). These are regions of low ionization fraction that are stable to the MRI (see the review by Armitage 2011 for a detailed discussion). For massive discs, in which self-gravity is important, GI is expected to significantly contribute to the angular momentum transport within the disc. GI is discussed in Section 2.4.

The solid component of the protostellar disc, initially composed of micron-sized dust grains that are well coupled to the gas component, evolves differently than the gas-component. The vertical component of gravity results in the settling of dust grains towards the midplane, with the larger dust grains expected to settle to a smaller scale height. In addition, the grain size-distribution evolves, with collisions and coagulation producing larger grains.

Theoretical models for these processes, coupled to 1+1D RT models, have found that the timescale for dust settling is short, roughly 10^5 years. Turbulent mixing,

however, prevents complete settling of the dust, with a steady-state achievable in less than 1 Myr (Dullemond & Dominik 2004). Similar calculations, have shown that the timescale to remove all small grains (up to 100 μm) occurs too fast (10^4 years) to be consistent with observed SEDs, and fragmentation of grains is necessary (Dullemond & Dominik 2005). More complex models, including fragmentation of grains, radial drift and the viscous evolution of the disc have been developed (Birnstiel et al. 2010). These demonstrate the well-known "metre-size barrier": grains larger than a few cm undergo destructive collisions, as well as rapid radial drift inward to the star.

The gas-component of a protostellar disc is eventually dissipated through photo-evaporation due to radiation from the central star, or nearby stars. The radiation responsible for photoevaporation can be divided into three types: 1) FUV, with 6 eV $< h\nu <$ 13.6 eV, 2) EUV, with 13.6 eV $< h\nu <$ 0.1 keV and 3) X-ray, with 0.1 keV $< h\nu$, where h is Plank's constant, and ν is the frequency of the light. The basic mechanisms are well-described in the review by Dullemond et al. (2007).

To summarize, EUV photons heat only the outer layers of the disc (due to the large opacity to ionizing radiation), but heat the gas to 10^4 K independent of the radius. FUV (and X-ray) photons by contrast, are able to penetrate more deeply into the disc, with a heating rate that depends on the radiation flux, the gas density, and the chemistry. Thus, the calculation of photoevaporation via FUV and X-ray is sensitive to the ability to calculate the disc's vertical structure accurately. Whether an evaporative flow is created depends on the gas's sound speed in comparison to the local escape speed.

Hollenbach et al. (1994) first considered EUV photoevaporation from a central star and derived mass-loss rates, which are based on the Strömgren condition, that ionizations balance recombinations. In essence, a thin-layer of ionized gas is created on the surface of the disc, similar to an HII region. The large sound-speed of this layer leads to an evaporative flow outside a critical radius, outside of which gravity is weak enough that the gas becomes unbound.

Alexander et al. (2006a,b) used the mass-loss rates of Hollenbach et al. (1994), with some improvements based on hydrodynamics simulations, combined with a viscous disc-model to consider the photoevaporation of the disc. In their "UV-switch" model,

mass is transported inwards through the disc via viscous accretion, while mass is lost at each radius via an EUV wind. At early times, viscous accretion dominates; however, after a few Myr, viscous accretion no longer balances the wind loss-rate in the interior of the disc and an inner hole develops. Direct EUV irradiation on the inner edge of the disc enhances mass loss and the entire disc is photoevaporated after roughly 10^5 years. This model is able to overcome the "two-time-scale problem" described above and explain both the observed lifetimes of protostellar discs, as well as the observed rapid disc dispersal timescale. Gorti et al. (2009) and Gorti & Hollenbach (2009), stating that EUV fluxes should be lower than those used by Alexander et al. (2006a,b), examined the role that FUV and X-rays (in addition to EUV) plays in driving photoevaportative flows. They found that FUV and X-ray were effective at driving mass loss at earlier times, and at larger radii than the EUV flux.

2.2.4 Observations of class 0-I discs

Now that we have considered the later phases of protostellar discs, as well as many of the important physical processes in determining disc structure and evolution, we turn to the earliest phases of the disc. As described previously, class 0-I YSOs have discs that are deeply embedded in envelope material, making the study of class 0-I discs significantly more difficult than for class II discs. However, even without resolving the disc emission, large surveys of YSOs have provided information regarding the timescales for disc formation, and some details of disc evolution.

Discs do not spend long in the class 0-I phases. Evans et al. (2009b) surveyed five nearby star forming regions (within 300 pc) using the Spitzer IR telescope as part of the c2d survey. The authors classified each YSO according to the slope of its SED. Based on the numbers of YSOs in each class, and a typical age for the class II phase of 2 Myr, they determined that class 0 and class I phases have lifetimes of only 0.16 and 0.54 Myr respectively.

Kenyon et al. (1990) found that a sample of YSOs in Taurus-Auriga had luminosities that were significantly less than was expected for the estimated rates of accretion. In fact, class I sources were not found to be more luminous than class II

sources. One proposed solution to this "luminosity problem" is that accretion rates are not constant, but are instead bursty. In this picture, the majority of YSOs are in a long-lived low-accretion state during which their accretion (and luminosity) is low. These low-accretion periods are punctuated by short-lived events of high-accretion (and luminosity) during which the star gains most of its mass. Bursty accretion may also explain the phenomenon of FU Orionis outbursts, in which the brightness of a star is observed to increase by up to 6 magnitude on short timescales (less than 100 years, see review by Hartmann & Kenyon 1996). In agreement with this picture of bursty accretion, Evans et al. (2009b) found that the luminosity of their sources could be reconciled with the expected accretion rates if roughly 50 % of the stellar mass was accreted in only 7 % of the class I lifetime.

To resolve the disc-scale (a few 100 AU) in nearby star forming regions requires an angular resolution of roughly 1", while the ability to penetrate the significant envelope requires long wavelengths. Sub-mm interferometry is thus a valuable tool in the study of these deeply embedded discs. In addition, long baselines can be used to filter out emission from the envelope, which is expected to occur on larger scales and thus dominate on short baselines. Interferometric observations are generally combined with RT codes and SED modelling, in order to disentangle the emission from the star, disc, and envelope.

Eisner et al. (2005) combined continuum 1.3 mm interferometry from the Owen Valley Radio Observatory (OVRO) and SED observations with a radiative transfer code. The authors examined five class I objects in Taurus and found disc masses in the range of 0.1-1 M$\odot$. In addition, the envelope infall rates determined were an order of magnitude larger than observed accretion rates onto the stars, pointing to bursty accretion. Jørgensen et al. (2007, 2009) used the Submillimetre Array (SMA) to survey 10 class 0 and 10 class I discs. With combined SMA and SED observations, and an RT code, they determined that the mean disc mass was 0.05 M$\odot$. Perhaps surprisingly, they found no trend in disc mass between class 0 and class I objects, despite a clear reduction in the amount of material in the envelope (from 1 to 0.1 M$\odot$), which suggests efficient transport of material through the disc. Enoch et al. (2011) combined continuum observations from the Combined Array for Research in

Millimeter-wave Astronomy (CARMA) with SEDs and RT modelling, determined a typical disc mass of 0.2 M$\odot$. The authors also found that 6 of nine class 0 YSOs have detectable discs, pointing to the rapid formation of discs (within the 0.2 Myr of the class 0 phase). These observations indicate that the masses of discs in the class 0-I phases are significantly greater than those in the class II phase, and that bursty accretion likely occurs.

The above work uses a combination of observations in addition to RT modelling to disentangle emission arising from the disc from that of the envelope. Some recent work (Brinch et al. 2007; Jørgensen et al. 2007, 2009) has used molecular line emission in CHO^+ J=3-2 in order to detect the disc directly by looking for Keplerian rotation. In the most highly embedded sources, however, the emission is optically thick. The Atacama Large Millimeter/sub-millimeter Array (ALMA) telescope will be able to detect less dense isotopologues, making the direct detection of disc emission easier in class 0-I YSOs.

2.2.5 Models of disc formation

Due to the difficulties in studying the very first stages of disc formation observationally, including the short timescale for disc formation (less than 0.2 Myr, Evans et al. 2009b) and the degree of obscuration from the infalling envelope, much of the progress in understanding the formation of discs has come from numerical simulations. Computational models generally begin with a gravitationally unstable prestellar core, and follow its collapse and the subsequent formation and evolution of the protostar and its protostellar disc. Because of the range in scales (over six orders of magnitude separate the scales of prestellar core and stellar radius), and the necessary integration times ($10^5 - 10^6$ years), this is a difficult process to model in 3D while including all of the relevant physics. Only recently, and for relatively few cases, has this been done. We begin with a historical overview of the progress made, and discuss some of recent results, and their correspondence with observation.

Larson (1969) performed some of the first simulations of prestellar collapse. He used 1D radiation hydrodynamics to follow the collapse of a 1 $M_\odot$ protostellar core,

and found three phases for the collapse. The material remained nearly isothermal until it reached a density of $\rho = 10^{-13}$ g cm^{-3}, at which point it became optically thick. This material at high density then formed a first core, which was near hydrostatic equilibrium, and about 4 AU in size. As matter continued to accrete on the core, the internal temperature increased until molecular hydrogen dissociated at the center, prompting the collapse to a stellar core of roughly 5 M_{Jup}. The remaining material fell onto the stellar core during a main accretion phase. Although these early calculations are important in understanding the formation of the protostar, the initial conditions lacked rotation, and made the formation of a disc impossible.

If one assumes that angular momentum is conserved as material from a rotating prestellar core collapses onto a protostellar disc, then one can map the initial angular momentum of the material to the radius at which it impacts the disc. Lin & Pringle (1990) used this principle to calculate accretion rates for the evolution of their 1D disc formation model, which included radiative cooling and angular momentum transport via parameterizations for both the turbulent viscosity and self-gravity. They concluded that protostellar discs should be several hundred AU in extent, and have mass comparable to the central star. Yorke et al. (1993) reached similar conclusions using 2D radiation hydrodynamic simulations of a collapsing 1 $M_{\odot}$ prestellar core: they formed discs 200-500 AU in extent, with masses similar to the protostar.

In recent years, with increased computing power, it has become possible to perform 3D numerical simulations of prestellar core collapse, with an ever increasing amount of realistic physics (RT, magnetic fields) included. 3D hydrodynamic simulations of prestellar core collapse using barotropic equations of state to mimic the thermodynamic effects of radiative transfer have examined the collapse of both uniformly rotating (Machida et al. 2010) and turbulent (Hayfield et al. 2011) prestellar cores. These works demonstrate that massive discs (in the range of $0.1 - 1.0$ $M_{\odot}$) on the scale of hundreds of AU form within 10^4 years. At early times, the discs were observed to be even more massive than the protostar. Consequently, these massive discs were prone to gravitational instability, demonstrated strong spiral modes, and in some cases fragmented. 3D radiation-hydrodynamics (RHD) simulations performed by Bate (2011) lead to similar conclusions: massive (0.22 $M_{\odot}$), extended (100 AU)

discs, susceptible to fragmentation are formed quickly (these calculations only follow to shortly after the formation of the stellar core, $\sim 6 \times 10^4$ years). We save a detailed discuss of GI and the conditions for fragmentation for Section 2.4; however, we do wish to emphasize here that discs that are susceptible to GI and fragmentation naturally arise in a number of simulations of protostellar disc formation.

The role of magnetic fields in disc formation is currently somewhat uncertain. Magnetic braking has been found to be extremely effective at removing angular momentum during the collapse of a prestellar core. In fact, in the ideal MHD case, it has been found to suppress the formation of any rotationally supported disc (Mellon & Li 2008; Seifried et al. 2011). This conflict between simulation and observation has been termed the "magnetic braking catastrophe." Non-ideal MHD effects, such as ambipolar diffusion (Duffin & Pudritz 2009) and ohmic dissipation (Machida & Matsumoto 2011) allow only for the formation of an inner-disc of less than 10 AU. This is much smaller than the hundred-AU scale common to observed discs.

One recently proposed resolution of this catastrophe finds fault in the idealized initial conditions, having uniform rotation, that are often used in prestellar core collapse simulations. Seifried et al. (2012) have performed magnetohydrodynamic simulations of the collapse (and fragmentation) of a turbulent 100 $M_\odot$ prestellar core. They found that individual protostars were able to form discs on the scale of 100 AU, with typical masses of 0.1 $M_\odot$, within roughly 10^4 years. The use of turbulent initial conditions, rather than those with uniform rotation, resulted in a lack of coherent rotation on scales greater than a few hundred AU. In this case, strong toroidal fields, which act to efficiently extract angular momentum, were not built up. This explanation is consistent with the turbulent radiation magnetohydrodynamic simulations of (Commerçon et al. 2011), in which large discs are formed for realistic magnetic field strengths.

Because of the computational expense of 3D simulations, only a relatively small number of simulations have been performed, and often only for a limited integration time. Simplified simulations of varying complexity, which are computationally cheaper, are useful in exploring some of the available parameter space. These have included one-zone models (Kratter et al. 2008), 1D radial model (Hueso & Guillot

2005; Rice et al. 2010), and 2D models.

Vorobyov & Basu (2010b) and Vorobyov (2010, 2011) used thin-disc (2D) hydrodynamic simulations, including irradiation and radiative cooling, to study the collapse of prestellar cores based on the axisymmetric, magnetically supercritical core of Basu (1997). Over a range of rotation rates for the prestellar core, the authors found median disc masses 0.09 and 0.14 $M_{\odot}$ during the class 0 and class I phase respectively, with a nearly constant (with significant scatter) disc-to-star mass ratio of 0.72 and 0.64 respectively. The median disc radii are 190 and 280 AU for each phase respectively. Much of the inner region of the disc is initially optically thick, which may skew observational estimates of disc mass, which assume optically thin emission. The authors also note that most of their discs are gravitationally unstable and fragment. However, most fragments do not survive, but instead migrate inwards into the central star and are accreted. This bursty accretion may offer a possible solution to the luminosity problem, as well as the FU Orionis phenomenon.

2.2.6 Summary

Our discussion of protostellar discs has been detailed because these discs are fundamental to the formation of gas-giants. Before considering the mechanisms for gas-giant formation, we summarize a few of the key points of this section. Observations of protostellar discs show that they are a common around young stars. Observations and models of formation both suggest discs are massive and extended in their class 0-I phases, which roughly lasts a combined 0.7 Myr. At these early times, discs are massive enough to be susceptible to GI and fragmentation, which may explain the bursty accretion that may lead to the luminosity problem and FU Orionis outburts. Discs evolve through mass transport, the rapid growth of dust grains to larger bodies, and are eventually dispersed by photoevaporation. Lifetimes of discs are roughly 3 Myr, although a significant number are dispersed on shorter, 1 Myr timescales.

2.3 Gas-giant formation via core accretion

In the core accretion model, gas-giant planets form in a disc through a bottom-up process, in which there is growth from dust grains to protoplanets that can then capture gaseous envelopes. Models indicate that this process generally takes a few Myr, which is comparable to the lifetime of the disc. Here, we give an overview of the formation of a gas-giant in the core accretion model (Lissauer & Stevenson 2007; Mordasini et al. 2010).

This scenario begins with the buildup from micron-size dust grains to km-size planetesimals. The process of dust growth via agglomeration is outlined by Dominik et al. (2007). One uncertainty in this process is how metre-size planetesimals grow to km-size. At the expected relative velocities, m-size objects are expected to destroy each other upon collision, rather than to stick together. In addition, planetesimals large enough to decouple from the gas experience a "headwind" as they orbit, as the gas travels with a sub-Keplerian velocity because of radial pressure gradients. The timescale for inward radial drift, and destruction, is shortest for m-size objects (less than 100 years, Mordasini et al. 2010). One possible solution to the "metre-size barrier" is the rapid concentration of m-size bodies that results from turbulence coupled to the streaming instability (Johansen et al. 2007).

Once km-sized planetesimals have formed, there is continued growth through two-body collisions. This growth is initially a runaway process, in which the larger planetesimals double their mass faster than smaller planetesimals (Wetherill & Stewart 1989). This results in a small number of protoplanetary embryos that are much larger than the remaining disc of planetesimals. At this stage, the embryos continue to accrete planetesimals in what is known as "oligarchic growth." Each embryo, however, only has a limited reservoir of plantesimals from which it can accrete, and there is a maximum mass, or "isolation mass" (Lissauer 1993) that includes the material within a few Hill radii, where the Hill radius is

$$R_{\rm Hill} = \left[\frac{M_{\rm embryo}}{3M_\star}\right]^{1/3} a, \tag{2.4}$$

where a is the semi-major axis, $M_{\rm embryo}$ is the mass of the embro, and $M_\star$ is the mass of the central star.

The above steps, along with subsequent collisions between embryos, describe the formation of solid planets. To produce a gas-giant, we must also consider the accretion of gas onto the embryo. A number of authors (Pollack et al. 1996; Alibert et al. 2005; Lissauer et al. 2009) have performed calculations of the growth of a single protoplanet embryo accreting both planetesimals and gas in a protostellar disc. These calculations model the embryo in 1D, assuming hydrostatic equilibrium. We describe their general results.

The embryo's accretion of planetesimals in the runaway phase, and growth to the isolation mass, is rapid (taking 10^5 years); however, the initial phase of gas accretion is relatively slow, as it is governed by the thermodynamics of the embryo's envelope. In this phase, the envelope extends to the Hill radius and must radiate away the energy that is input through gravitational contraction and plantesimal accretion in order to contract. This contraction allows more gas to become bound to the embryo and accrete onto the envelope. In addition, the embryo's Hill radius increases as its mass increases, allowing for accretion of gas and planetesimals over a larger zone of the disc.

When the core accretes enough gas so that its envelope and core are of roughly equal mass, then the self-gravity of the envelope becomes important, and the envelope contracts with the addition of more gas. At this crossover mass, roughly 10-15 $M_\oplus$, there is runaway gas accretion. Further accretion is no longer dependent upon the thermodynamics of the disc, but rather the availability of gas. The embryo rapidly grows in mass until accretion is limited, by the formation of a gap (Bryden et al. 1999), or the dissipation of the gas disc. Lissauer et al. (2009) showed that a Jupiter-mass gas-giant can be formed at 5 AU, within the roughly 3 Myr lifetime of the disc, given a surface density that is roughly three times that of the MMSN.

A further consideration, absent in the work of Lissauer et al. (2009), is the radial migration of the protoplanet during its formation in the gas disc. Radial migration of a protoplanet can be caused by two types of disc-planet interactions within a gas disc. In Type I migration (Goldreich & Tremaine 1980), an embedded protoplanet migrates

inwards (in a Keplerian disc) due to an imbalance of torques between material inside the orbit of the protoplanet and material outside of the orbit. However, estimates of the type I migration rate (e.g. Tanaka et al. 2002) are very short (roughly 1 Myr at 5 AU for an earth-mass core), and therefore pose a problem for gas-giant formation via core accretion (the linear rate must be reduced by an order of magnitude to agree with exoplanet observations, Ida & Lin 2008). The solution to the Type I migration problem will include non-isothermal effects (Paardekooper & Mellema 2006), as well as dead-zones, opacity transitions, and other migration traps (Hasegawa & Pudritz 2011 and references therein). Detailed models of gas-giant formation such as those of Lissauer et al. (2009) have not yet been carried out for realistic Type I migration.

Migration is not limited to the protoplanet embryo. In Type II migration (Lin & Papaloizou 1986; Bryden et al. 1999), the protoplanet is massive enough to open a gap in the disc, and migrates inwards on the viscous time of the disc, or more slowly for more massive planets (Papaloizou et al. 2007). As will be discussed below, the issue of migration is not limited to formation via core accretion. It is an important process in planet formation theory that requires further research.

2.4 Gas-giant formation via gravitational instability

2.4.1 Conditions for fragmentation

In the fragmentation via gravitational instability model (Cameron 1978; Boss 1997), gas-giants form through the rapid collapse of local regions in massive, unstable protostellar discs. Models of GI show that the process is rapid, taking hundreds to thousands of years, in contrast to the millions of years that are required in the core accretion model. The exact conditions in which GI is expected to form gas-giants is a topic of current research, and is the focus of the work presented in Chapters 3 and 4. Here, we outline recent research studying the role of GI in protostellar discs, and the criteria under which fragmentation, and the formation of gas-giants is expected.

Whether a thin disc is gravitationally unstable is determined by the Toomre Q

parameter (Toomre 1964; Binney & Tremaine 2008),

$$Q = \frac{c_{\rm s}\kappa}{\pi G\Sigma}, \tag{2.5}$$

in which G is the gravitational constant, Σ is the local surface density, and

$$\kappa = \sqrt{R\frac{d\Omega^2}{dR} + 4\Omega^2}. \tag{2.6}$$

is the epicyclic frequency. In the above, Ω is the local rotation rate; for a Keplerian disc, $\kappa = \Omega$. The Toomre Q is derived from a linear stability analysis for a local patch of a thin-disc. This analysis leads to the dispersion relation for an axisymmetric disturbance:

$$\omega^2 = \kappa^2 - 2\pi G\Sigma|k| + k^2 c_{\rm s}^2, \tag{2.7}$$

where k is the wavenumber of the perturbation, and ω is the frequency (or growth-rate, if $\omega^2 < 0$) of the perturbation. In this patch, the destabilizing effects of self gravity are countered by the stabilizing effects of shear (for long-wavelength perturbations) and pressure (for short-wavelength perturbations). When $Q < 1$, the disc is unstable to perturbations with wavelengths that satisfy

$$\lambda_{\rm Toomre}\left(1 - \sqrt{1-Q^2}\right) \leq \lambda_{\rm unstable} \leq \lambda_{\rm Toomre}\left(1 + \sqrt{1-Q^2}\right), \tag{2.8}$$

where

$$\lambda_{\rm Toomre} = 2\pi^2 G\Sigma/\kappa^2 \tag{2.9}$$

is the most unstable wavelength in the linear analysis. Massive, cold protostellar discs are expected to have regions with $Q < 1$, and will therefore be unstable. The finite-thickness of realistic discs has a stabilizing effect (Goldreich & Lynden-Bell 1965), and reduces the critical value to $Q \approx 0.75$ (Kim et al. 2002).

The Toomre analysis is a local linear analysis: it considers the stability of one patch in the disc. In addition, it is strictly only applicable for tightly-wound disturbances. Numerical simulations of GI have been very useful in studying the global stability of discs to GI, including the non-linear regime. These simulations have shown

that discs with regions of $Q \lesssim 1.5$ are unstable to the growth of non-axisymmetric perturbations; the disc develops multi-armed, trailing spiral overdensities, usually with multiple modes present (Durisen et al. 2007).

A variety of treatments for the thermodynamics of the gas have been used in numerical studies of GI. Locally isothermal simulations (Laughlin & Bodenheimer 1994; Boss 1997; Mayer et al. 2004), in which the temperature of the gas is a function of its initial (usually radial) position, assume that cooling is rapid; any excess thermal energy is radiated away instantaneously. At the other extreme, adiabatic simulations (Mayer et al. 2004) assume that cooling is inefficient; the gas is unable to radiate away the excess thermal energy from compressional or shock heating.

Other simulations have included cooling parameterized using the local cooling time, which is defined (for Lagrangian methods such as smoothed particle hydrodynamics, SPH) through the evolution of the specific internal gas energy, u:

$$\frac{Du}{Dt} = -\frac{u}{t_{\rm cool}}, \tag{2.10}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$ is the co-moving derivative (there is an analogous definition for Eulerian, grid-based methods). These prescriptions enforce a constant cooling rate per dynamical time,

$$t_{\rm cool}\Omega = \beta \tag{2.11}$$

(β-prescription, Gammie 2001; Rice et al. 2003; Meru & Bate 2011; Cossins et al. 2009), or a constant cooling rate, $t_{\rm cool} = \text{constant}$ (Mejía et al. 2005).

Parameterized cooling treatments such as these are useful in studying the role of thermodynamics in disc simulations, since the cooling rate is specified and is constant. However, they are not meant to represent realistic cooling, since discs do not have fixed valued of $t_{\rm cool}$ or β (as can be observed in our simulations, see Figures 3.11 and 3.14). The global nature of β-prescription cooling means that simulations that employ it are limited in what they can tell us about realistic discs. For example, since there is no spatial variation of the cooling rate, we can not learn about where in the disc fragmentation is expected; in addition, the character of the fragmentation is likely not realistic, since the whole disc is susceptible to fragmentation.

Realistic cooling results from radiative (and possibly convective) transport of energy within the disc, which depends on the optical depth, temperature, and density of disc material. A number of studies have used various local radiative cooling approximations based on vertical diffusive transport with a smooth transition to the free-streaming limit (Johnson & Gammie 2003; Stamatellos & Whitworth 2008; Boley 2009). The accurate determination of the thermodynamics of the disc requires the inclusion of radiative transfer in simulations, and a number of authors have examined GI using radiation hydrodynamics (Boss 2001; Mayer et al. 2007; Cai et al. 2006; Boley et al. 2007; Meru & Bate 2010a; Stamatellos et al. 2011; and Chapters 3 and 4).

There are two possible outcomes of the non-linear growth of GI in a disc: the amplitude of the overdensities can saturate, or the disc can fragment. Whether fragmentation occurs depends on the thermodynamics of the disc: unstable discs that cool quickly can fragment, while those that do not cool fast enough are stable to fragmentation, and maintain $Q \sim 1$. Before examining this cooling criterion for fragmentation in detail, we first note the influence of GI in discs that are stable to fragmentation. As discussed in Section 2.2.3, observed accretion onto YSOs (and the growth of the star itself) requires the transport of material, and angular momentum, through the disc. One mechanism for such transport is the sum of gravitational torques exerted by the spiral structure in gravitationally unstable discs (Larson 1984).

Laughlin & Bodenheimer (1994) examined the transport of mass and angular momentum in a gravitationally unstable ($Q_{\rm min} = 1.3$) disc, which did not fragment, using 3D SPH with a locally-isothermal equation of state. The authors found significant transport resulted from the gravitational torques. In addition, they found that the evolution of the surface density of the disc agreed well with an α-viscosity model, with $\alpha = 0.03$.

It is not immediately clear that GI should be well-characterized by an α-viscosity, since this model assumes that angular momentum transport and energy dissipation act locally, and gravity is an inherently non-local force. However, in the case that the GI results in the disc being near corotation (the pattern speed of the spiral structure matches the local rotation rate, $\Omega_{\rm P} = \Omega$), Balbus & Papaloizou (1999) demonstrated

that an α-viscosity treatment is applicable to gravitational torques.

A number of numerical simulations have since been carried out to test in what circumstances this condition is met. The locality of GI-induced transport shows a dependence on the thermodynamic treatment. Lodato & Rice (2004) and Lodato & Rice (2005), using β-prescription cooling, found that there was no significant global transport for both low (less than 0.1) and high (up to 1.0) disc-to-star mass ratios. In contrast, Mejía et al. (2005), using constant $t_{\rm cool}$ cooling, found that there was global transport, and that an α-viscosity underestimated the observed mass-transport rate by an order of magnitude. Boley et al. (2006), using radiative hydrodynamics, found that there was a limited range, near corotation, in which mass transport was well-characterized by an α-viscosity, in agreement with Balbus & Papaloizou (1999). Lodato & Rice (2004) argued that discs with aspect ratios (H/R, where H is the scale height) of $H/R < 0.1$ should be well characterized by an α-viscosity. For $Q \approx 1$ discs, $H/R \propto M_{\rm disc}/M_\star$. One would consequently expect more massive discs (larger disc-to-star mass ratio) to have more global transport and to be less well characterized by an α-viscosity. Cossins et al. (2009), using β-prescription cooling, and Forgan & Rice (2011), using radiative transfer, computed a non-local transport fraction, and found that it increased with the disc-to-star mass ratio. Forgan & Rice (2011) found that only light discs ($M_{\rm disc}/M_\star < 0.5$) with small aspect-ratios ($H/R < 0.1$) were well-characterized by an α-viscosity. From these results, it appears reasonable to model GI-driven transport using α-viscosity in the case of low-mass discs. High-mass discs can be treated using α-viscosity only in regions near corotation.

Gravitationally unstable discs often exhibit multiple spiral modes. Laughlin et al. (1997, 1998) examined the global stability of a polytropic disc to non-axisymmetric modes using analytic calculations and 3D hydrodynamics simulations. From both approaches, the authors found that the disc was initially unstable to the growth of a two-armed spiral mode. However, in the non-linear phase of growth, they found that there was significant coupling between spiral modes. This coupling naturally led to the growth of other spiral modes, $m = 0$ and 4, from the initial $m = 2$ mode. Non-linear coupling results in the growth of multiple modes.

Gravitational torques are important in driving the evolution of unstable discs

through mass-transport. However, our primary interest in this thesis is the formation of gas-giants, so we now turn to the question of when protostellar discs fragment. Numerical simulations have shown that discs with regions of $Q \lesssim 1.5$ develop spiral structure; in addition, simulations have shown that whether such gravitationally unstable discs fragment depends on the thermodynamics of the disc.

Simulations employing locally-isothermal equations of state (Boss 1997; Mayer et al. 2002, 2004) have shown that massive discs, with $Q < 1.4$, fragment into massive gravitationally bound clumps, potential precursors to gas-giant planets, with masses in the range of Saturn to super-Jupiter. These clumps have been found to be long-lived, even when a switch to an adiabatic equation of state is employed once a critical density is reached. If cooling is efficient, as it is in the locally-isothermal approximation, then GI is an effective mechanism for forming gas-giants quickly, within a few hundred years inside of 20 AU. However, if cooling is inefficient, then fragmentation is unlikely to occur. Simulations employing an adiabatic equation of state, in which there is shock heating but no cooling, show that such discs do not fragment, even when they are initially very unstable, with regions of $Q \approx 0.8$ (Pickett et al. 2000; Mayer et al. 2004).

Gammie (2001) investigated the thermodynamic conditions required for fragmentation using 2D shearing-box simulations of a patch of a gravitationally unstable disc. Rather than attempting to include realistic radiative cooling, he simplified the problem by including cooling via the β-prescription, equation (2.11). He then carried out a series of numerical experiments in which he varied the cooling rate, β, and established a cooling criterion for fragmentation: for local cooling faster than a critical cooling rate, $\beta = \beta_{\rm crit} < 3$, fragmentation was observed. In the cases where the cooling rate was too long for fragmentation, the disc settled into a self-regulated state, termed "gravitoturblence", in which heating from GI balanced the prescribed cooling and maintained $Q \approx 1$. We note that the specific value of the critical cooling rate determined by Gammie (2001) is applicable only to the case of the 2D power-law equation of state used in that work.

The idea of a critical cooling rate for fragmentation was confirmed using global 3D SPH simulations (Rice et al. 2005), as well as global 3D grid-based simulations

(Mejía et al. 2005). The physical explanation put forth for the existence of a critical cooling time has relied on the use of an α-viscosity model to describe GI in the disc (Gammie 2001; Rice et al. 2003). In this model, there is an equilibrium state in which the viscous heating is balanced by the prescribed cooling, For a given cooling rate, the necessary viscosity is determined by

$$\alpha = \frac{1}{9/4\gamma_{\rm 2D}\left(\gamma_{\rm 2D}-1\right)\beta}. \tag{2.12}$$

where $\gamma_{\rm 2D}$ is the two-dimensional adiabatic index, which can be mapped to the 3-D adiabatic index, γ. If there is a maximum viscous stress, $\alpha_{\rm crit}$, that GI can develop (Rice et al. 2005), then there is a corresponding maximum heating rate. If that heating rate is not sufficient to balance the prescribed cooling, then no equilibrium is attainable, and the disc fragments.

Rice et al. (2005) found that the critical cooling rate is a function of the adiabatic index of the gas, but that this is consistent with a maximum stress of $\alpha_{\rm crit} \approx 0.06$. However, recent work has suggested that a single value of $\beta_{\rm crit}$ may not exist, and that it may depend on the properties of the disc (Meru & Bate 2011), or the thermal history of the gas (Clarke et al. 2007). This picture is also complicated by the fact that 3D simulations used to determine $\beta_{\rm crit}$ appear not to have converged numerically (Meru & Bate 2010b). Finally, recent 2D shearing box simulations (Paardekooper 2012) suggest that fragmentation is possible for longer cooling times, but that it is a stochastic process. Fragmentation is possible for $\beta > \beta_{\rm crit}$, but may just be less likely, requiring long-integration times for the necessary local conditions to be satisfied.

In Chapter 4, we discuss the issues regarding the use of a critical cooling time in more detail. In addition, we present a more detailed physical framework in which to understand the role that cooling plays in fragmentation; in fact, this framework allows for the direct calculation of the critical cooling time. In the current discussion, we note that cooling plays a fundamental role in determining whether gravitationally unstable discs fragment: discs must cool quickly for fragmentation to occur. The critical cooling time is useful in examining the physical conditions necessary for realistic protostellar discs to fragment. One should, however, keep in mind that the cooling

rate required for fragmentation is likely not as simple as a single value for $\beta_{\rm crit}$, and that there is uncertainty regarding the particular numerical value of the cooling rate required for fragmentation.

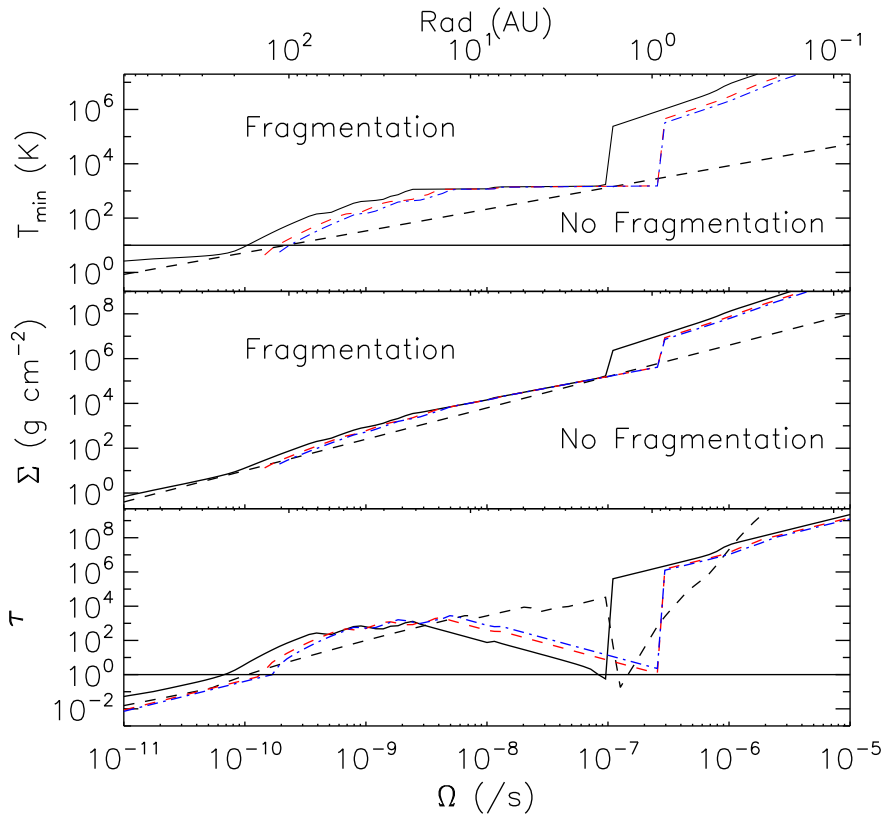

Figure 2.2: The disc properties required for fragmentation: the minimum temperature and surface density required to satisfy the Toomre Q and cooling criteria, and the associated optical depth. The requirements are given for the radiative cooling time of Rafikov (2005) (black line), the optically thick perturbative radiative cooling time of Nero & Bjorkman (2009) (blue, dot-dashed line), and the optically thick perturbative radiative cooling time of Kratter et al. (2010) (red, dashed line). The lower-limit, when the midplane temperature is equal to the photosphere temperature is given by the black, dashed line. The horizontal line in the temperature (opacity) plot denotes $T = 10$ K ($\tau = 1$). The conversion from rotation rate to radius is made for a central star of 1 $M_\odot$.

There are then two criteria that must be satisfied for disc fragmentation to occur: the disc must be gravitationally unstable ($Q < Q_{\rm min}$, the Toomre criterion) and cool quickly ($t_{\rm cool}\Omega < \beta_{\rm crit}$, the cooling criterion). These two criteria have been combined

to analytically determine the necessary physical conditions in a disc that are required for fragmentation (Rafikov 2005, 2007; Nero & Bjorkman 2009; Kratter et al. 2010).

Discs cool primarily through radiative cooling, for which the cooling time is (Rafikov 2005)

$$t_{\rm cool} \approx \frac{\Sigma c_{\rm s}^2}{\gamma - 1} \frac{\tau + 1/\tau}{2\sigma T^4}, \tag{2.13}$$

where the optical depth is $\tau \approx \kappa\Sigma/2$, the opacity is κ, and the midplane temperature is T (see Rafikov 2007 for a consideration of convection). A short cooling time requires a high temperature; however, if fragmentation is to occur, this high temperature must be combined with a correspondingly large surface density in order to keep Q near unity. Rafikov (2005), using the above radiative cooling, combined the requirements of the Q and the cooling criteria to determine the minimum temperature and surface density, as a function of rotation rate, that a disc must possess in order to fragment.

Figure 2.2 shows the results of the Rafikov (2005) analysis for opacities from D'Alessio et al. (1997). A central star of 1 $M_\odot$ is used to convert between rotation rate and radius. In addition to the cooling time from equation (2.13), we also plot the analysis for the optically thick perturbative cooling time of Nero & Bjorkman (2009),

$$t_{\rm cool} \approx \frac{\Sigma c_{\rm s}^2}{\gamma - 1} \frac{3\tau}{64\sigma T^4}, \tag{2.14}$$

and the optically thick perturbative cooling time, for an irradiated disc, of Kratter et al. (2010),

$$t_{\rm cool} \approx \frac{\gamma\Sigma c_{\rm s}^2}{\gamma - 1} \frac{3\tau}{32\sigma T^4}. \tag{2.15}$$

As pointed out by Rafikov (2005), the minimum temperatures and surface densities required for fragmentation are unlikely to be achieved in the inner regions of realistic protostellar discs, when observations of discs are considered. Inside of roughly 20 AU, the necessary temperatures are even above the sublimation temperature of dust (1200 K), and the necessary surface density is greater than 10^3 g cm^{-2}. At larger radii, the requirements for fragmentation are less stringent, and are more likely satisfied in realistic discs. At 100 AU, the necessary temperature is roughly 20 K, and the necessary surface density is roughly 50 g cm^{-2}.

The conditions necessary for fragmentation have also been considered through the use of 3D radiation hydrodynamics (RHD) simulations. Gravitationally unstable discs are generally massive enough to be optically thick, when Rosseland mean opacities from dust grains are considered. Consequently, the diffusion approximation is an accurate and, given the constraints of computing power, computationally viable method of incorporating radiative transport into 3D dynamic simulations of protostellar discs. In addition to diffusive transport, simulations must also incorporate radiative boundary conditions to model the cooling of the disc from its photosphere (and heating from irradiation). Because the cooling rate ultimately determines whether fragmentation takes place, an accurate boundary condition is fundamentally important in determining fragmentation.

Initial 3D RHD simulations of GI in the inner regions (within 40 AU) of protostellar discs gave conflicting results regarding the viability of fragmentation. Some results (Boss 2001; Mayer et al. 2007) found fragmentation in their simulations, while others (Mejía et al. 2005; Boley et al. 2007) found that the cooling times in this region were too long to satisfy the cooling criterion and allow fragmentation, in agreement with the above analytic treatment. Meru & Bate (2010a) found fragmentation in the inner disc only under extreme conditions, such as an initial $Q = 0.5$, or reduction in the opacity by a factor of 100.

Efforts have been made to determine the cause of the conflicting results from simulation, and it appears that the cause is the differing photosphere boundary treatments used. Cai et al. (2010) carried out a simulation in an attempt to replicate the results of Boss (2001), who found fragmentation. Cai et al. (2010) used the same initial conditions, and attempted to replicate the input physics; however, they did not find fragmentation, as Boss (2001) had. Since the only noticeable difference between the simulations was the choice of boundary conditions, Cai et al. (2010) concluded that this was the cause of the disagreement.

In Chapter 3 we present results from two 3D RHD simulations that indicate that fragmentation is unlikely in the inner regions of protostellar discs because the cooling times are too long to satisfy the cooling criterion. One simulation is of a disc similar to that used by Mejía et al. (2005) and Boley et al. (2007); we agree with their result that

fragmentation does not occur. In addition, we simulate one of the discs that Mayer et al. (2007) found to fragment; in this case, we disagree with the author's result: we find that fragmentation does not occur. Importantly, we demonstrate that their choice of photosphere boundary condition was flawed, as it overpredicts the amount of cooling in the disc. In fact, it overcools by the necessary amount to satisfy the cooling criterion and permit fragmentation. We demonstrate that our photosphere boundary condition yields robust cooling rates. Given the large mass of the disc (0.34 $M_\odot$ within 30 AU), this is a strong indication that the conditions in the inner disc are not suitable for fragmentation.

However, the analytic work presented in Figure 2.2 (Rafikov 2005, 2007; Nero & Bjorkman 2009; Kratter et al. 2010) does indicate that fragmentation via GI is viable in the outer regions of protostellar discs, near 100 AU. At these large radii, the cooling criterion is satisfied at fairly low temperatures and surface densities. At these large radii, heating through irradiation from the central star dominates over viscous accretion, and maintains high enough temperatures to satisfy the cooling criterion (Kratter et al. 2010). If a disc is massive enough to have $Q \approx 1$, then fragmentation is likely. This expectation has been confirmed by a number of RHD simulations (Boley 2009; Vorobyov & Basu 2010b; Stamatellos et al. 2011; Boss 2011; as well as the simulations presented in Chapter 4.).

2.4.2 The products of fragmentation at large radii

From both an analytical and a numerical simulation standpoint, the conditions for fragmentation can be satisfied in the outer regions of realistic protostellar discs (outside of roughly 100 AU). However, the end-result of the fragmentation process is less clear, and is a topic of current research. Current questions include the survival of the fragments, their masses, and their final orbital radii.

A number of mechanisms can disrupt a newly formed gas-giant, including tidal disruption from the central star, collisions with other fragments, and collisions with spiral arms. These disruption mechanisms were considered by Kratter & Murray-Clay (2011). If a fragment is to survive tidal disruption, it must collapse to a radius equal

to or smaller than a pressure-modified Hill radius (the standard Hill radius, equation (2.4), only considers shear). The contraction time of a clump depends on its cooling rate: faster cooling results in a quicker collapse and a higher likelihood for survival. If the gas is described as being $\gamma = 7/5$, however, even an adiabatic collapse leads to contraction within pressure-modified Hill radius. The survival of fragments during collisions with other fragments, or with spiral arms also depends on the ability of fragments to cool quickly, contract, and become more gravitationally bound.

In addition to efficient cooling, the internal temperature of the fragment also plays a key role in the rate of contraction. Once the temperature of the fragment becomes larger than the dissociation temperature of molecular hydrogen, then thermal energy is diverted into breaking molecular bonds, rather than providing pressure support, and a rapid collapse from AU-scales to a few R_{Jup} ensues. This is analagous to the collapse from first core to stellar core during star formation. For initial fragment masses, the timescale can be $10^4 - 10^5$ years (Boley et al. 2010). However, when grain growth and sedimentation in the fragment are considered, the timescale could be as short as 10^3 years (Helled & Bodenheimer 2011). However, no simulation including all relevant effects, including accretion of gas and planetesimals, rotation, radiative cooling, and grain evolution, has yet been performed, so the timescale for fragment contraction remains uncertain.

From simulations, the initial masses of fragments are expected to be in the gas-giant planet regime (a few M_{Jup}, Boley et al. 2010; and Chapter 4). Estimates of the fragment mass that use the Toomre length, the most unstable wavelength from the Toomre analysis, equation (2.9), grossly overestimate the observed masses of initial fragments in simulations. However, the subsequent growth of fragments may push the mass well into the brown dwarf regime, since the expected isolation, and gap-forming masses are much greater than the brown dwarf limit (Kratter et al. 2010). Mechanisms for limiting the mass-growth of fragments, such as the gap overlap of multiple fragments, or the dispersal of the disc require further study.

As is the case with planets formed via the core accretion scenario, gas-giants formed through GI are unlikely to remain at a single radius due to interaction with the gas disc. RHD simulations of gas-giant formation (Boley et al. 2010; Vorobyov &

Basu 2010b; and Chapter 4) show that fragments can migrate inwards from 100 AU on timescales of 10^4 years. More detailed studies of migration, which insert masses into gravitationally unstable discs and follow the evolution show a trend of similarly rapid inward migration, with some stochastic outward migration (Michael et al. 2011; Baruteau et al. 2011). The overall rate is consistent with Type I migration. The final outcome of this migration may be the destruction of the fragment via infall into the central star (Vorobyov & Basu 2010b), or through tidal disruption (Boley et al. 2010; Nayakshin 2010). Since the Hill radius, equation (2.4), decreases linearly with semi-major axis, an extended gas-giant migrating inward faster than it is contracting will be tidally disrupted by the central star. Fragments may also survive: the simulation of Michael et al. (2011) showed a 1 M_{Jup} gas-giant's migration was halted at 17 AU, at the inner Lindblad resonance of the dominant spiral mode. In systems with strong photoevapoation, the masses of surviving fragments may be substantially reduced by mass-loss from irradiation (Boss et al. 2002; Nayakshin 2010). This may change the mass of fragments from the regime of brown dwarfs to that of gas-giant planets. As with migration in the case of planets formed via core accretion, the migration of gas-giants formed through GI requires more study.

2.5 Observations of gas-giants

2.5.1 Gas-giant companions in the inner regions

Only recently, with the growing number of observations of planets outside of our own solar system, have we had significant statistical constraints for theories of planet formation. The first planet around another solar type star, 51 Peg b, was discovered less than twenty years ago (Mayor & Queloz 1995). The Jupiter-mass planet was detected using the radial velocity technique, in which the doppler-shift of a stellar spectra, due to the gravitational pull of an orbiting planet, is observed. Since 1995, a number of radial velocity surveys, such as CORALIE and HARPS, have begun, resulting in the detection of over 700 planets.

The other prevailing method of detecting extrasolar planets is the transit method,

in which the dimming of a star's brightness, caused by the transit of an orbiting planet in front of the host star, is observed. This method has been used to observe over 230 confirmed extrasolar planets. However, this number will grow quickly, as the Kepler space telescope has detected over 2300 extrasolar planet-candidates in only its first 16 months (Batalha et al. 2012). These candidates must be confirmed, but the false-positive fraction is expected to be low (Lissauer et al. 2012).

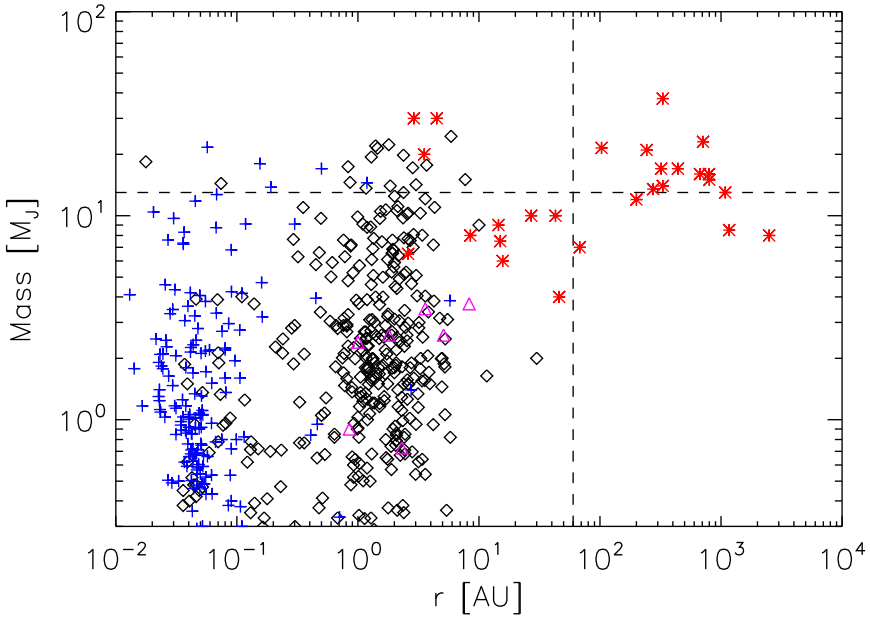

Figure 2.3: The current observational mass-radius relation for the 568 confirmed gas-giant companions. Companions discovered via the radial velocity technique are given by the black triangles, those discovered via the microlensing technique are given by the magenta triangles, those discovered via the transit technique are given by the blue crosses, and those discovered via direct imaging are given by the red asterisks. The vertical dashed line is a reference for r=60 AU, a rough estimate for the innermost region where GI can be expected to form a gas-giant around a sun-like star, while the horizontal dashed line is a reference for the deuterium burning limit of 13 M_{Jup}, above which a companion is considered to be a brown dwarf rather than a planet. The brown dwarf desert can be observed by the lack of objects in the upper left of the figure. The data is taken from www.exoplanet.hanno-rein.de/.

We focus on the observational constraints placed on the formation of gas-giant

companions (with $M > 0.3\ \mathrm{M_{Jup}}$, the mass of Saturn). Of the 763 confirmed extrasolar planets, 568 are gas-giants under this definition. The observed mass-radius relation is shown, for all detection techniques, in Figure 2.3. As can be seen, both the radial velocity method and the transit method are currently suited to detecting planets only on relatively short orbits, since the maximum orbit of a detectable planet is effectively limited by the duration of the survey.

There are a number of observational results for gas-giants that must be explained by a successful theory of the planet formation process, such as the orbital distribution of gas-giants, or mass-period relation. Hot Jupiters are gas-giants that orbit very close to their host star ($\lesssim 0.1$ AU), and include the first extrasolar planet, 51 Peg b. Observations show that such planets are rare: roughly 1% or fewer stars are host to hot Jupiters (Howard et al. 2010, 2011; Mayor et al. 2011). There also appears to be a maximum-mass for gas-giants that is a function of radius, such there is a lack of massive gas-giants (greater than 2 $\mathrm{M_{Jup}}$) on short (less than 100d, roughly 0.4 AU) orbits (Udry & Santos 2007; Mayor et al. 2011), while larger gas-giants exist at larger radii. In addition, the frequency of gas-giants is also an increasing function of radius (Mayor et al. 2011).

The overall occurrence of gas-giants with orbits less than 10 years was found to be roughly 10% from the published exoplanet data (Mayor et al. 2011). More specifically, the occurrence of gas-giants has been shown to be a decreasing function of metallicity (the mass-metallicity relation, Fischer & Valenti 2005; Mayor et al. 2011). Observations also show that the frequency of gas-giants is an increasing function of stellar-mass (Marcy et al. 2008).

There have been numerous studies pointing to the existence of a "brown dwarf desert" for companions with orbits less than 10 years (4.6 AU) of solar-type stars (Halbwachs et al. 2000, 2003; Udry et al. 2007; Sahlmann et al. 2011); that is, there appears to be a lack of companions that have masses in the range of 13-75 $\mathrm{M_{Jup}}$, as compared to planetary-mass or stellar-mass companions. This has been suggested as evidence that stellar and planetary mass companions have different formation mechanisms.

2.5.2 Accordance with models

Given the strong constraints from the Toomre and cooling criteria, Figure 2.2, and the growing consensus from 3D RHD simulations, in situ formation of gas-giants in the inner regions (within roughly 40 AU) via GI is unlikely. In this regards, core accretion is currently the most probable mechanism for the formation of the observed gas-giant population. However, since fragmentation via GI is expected at large radii (roughly 100 AU), and rapid inwards migration is observed in simulations, GI may play a role in the population of planets and gas-giants in the inner regions. The processing of solids in the inside of gas-giants formed via GI at large radii, followed by inwards migration and tidal disruption has been suggested as a formation scenario for rocky cores in the inner regions (Boley et al. 2010; Nayakshin 2010). These may subsequently re-accrete gas from the disc and once again form gas-giants. The plausibility of this scenario, and the impact on the inner region by fragmentation of the outer region, requires further investigation.

The formation of gas-giants via core accretion is able to explain a number of the observed features in the current exoplanet observations. Population synthesis codes, which employ simplified models of core accretion and migration to create a large number of planetary systems over a range of the possible parameter space (formation location, disc lifetimes, disc mass, metallicity, etc), have been useful in comparing the core accretion mechanism to the observed mass-period relation, Figure 2.3. These models (Ida & Lin 2004a,b, 2005, 2008; Mordasini et al. 2009, 2012) have been successful in reproducing a number of the observed trends. The decrease of gas-giant mass with decreasing semi-major axis has been explained based on the mass-dependence of Type II migration. The mass-metallicity relation has been explained based on the increased surface density of solids, and the resulting shorter formation time of cores that are able undergo runaway gas accretion.

An additional piece of evidence in support of core accretion is the lack of observed planets in the mass range 15-30 $M_{\oplus}$ (Mayor et al. 2011). Such a gap was predicted by population synthesis (Mordasini et al. 2009), based on the ability of such planets to rapidly accrete gas and become more massive.

There remain, however, some key issues with population synthesis models, such as the inclusion of Type I migration, which must be reduced from the analytic estimates by orders of magnitude in order to reproduce observations (Ida & Lin 2008). Improved treatments of Type I migration, based on recent simulations with more detailed physics are only beginning to be included (Mordasini et al. 2011).

2.5.3 Gas-giant companions at large radii

Radial velocity and transit surveys are only capable of finding gas-giants relatively close to the host star. Nominally, the duration of a given survey determines the maximum radius at which a planet can be detected, using the transit or radial velocity methods. However, the statistical analysis of radial velocity data in multi-planet systems can lead to the detection of planets at larger radii. Currently, the maximum radius at which a planet has already been detected using this method is 11.6 AU (Uma 47: Gregory & Fischer 2010).

The direct detection of the light from a gas-giant companion (direct imaging) is difficult due to the high contrast between the host star and the much cooler companion. Direct imaging favours the detection of gas-giant companions at large radii from the host star. The first detection of such a companion came in 2004, with the detection of a 4 M_{Jup} companion to the 25 M_{Jup} brown dwarf 2MASSWJ 1207334-393254 (Chauvin et al. 2004). In addition to favouring the detection of gas-giant companions on wide orbits, direct detection also favours the detection of relatively young (Myr) gas-giants, because they are still contracting and are therefore brighter than their older (Gyr) counterparts. Using photometry, and often including spectroscopy, the masses of gas-giants detected through direct imaging are determined by comparing the luminosity of the companion and its age, determined from the host star, to evolution models such as Chabrier et al. (2000). Since there are significant uncertainties regarding the initial formation of the gas-giants, the masses are somewhat uncertain (Marois et al. 2008).

There have been a number of subsequent detections of gas-giant companions via direct imaging, with 27 shown in Figure 2.3. These detections include companions at

large radii, such as HR8799 b (7 M_{Jup} at 68 AU, Marois et al. 2008), GQ Lup b (21.5 M_{Jup} at 103 AU, Neuhäuser et al. 2005), as well as companions at extremely large radii, such as HIP 75830 (23 M_{Jup} at 710 AU, Lafrenière et al. 2011), and HN Peg b (16 M_{Jup} at 795 AU, Luhman et al. 2007). The direct imaging of the four gas-giants of (9, 10, 10, and 7) M_{Jup} in the HR8799 system (Marois et al. 2008, 2010) is particularly interesting, since the planets cover a large range in radii (14, 27, 43, and 68) AU; this includes the inner region, which may favour formation via core accretion, as well as the outer region, which may favour formation via GI. Detection of spectroscopic signatures of accretion, such as Paβ, can point to the presence of a circumplanetary disc, such as in observations of GSC 06214-00210 b (Bowler et al. 2011). Such a disc may point to the lack of previous dynamical interactions with other bodies, thereby ruling out an origin via scattering from the inner disc.

Direct imaging surveys are beginning to provide data on the statistics of substellar companions on large orbits. Janson et al. (2011, 2012) surveyed 15 stars in the spectral class B2-A0, and 85 stars in the spectral class FGKM. No companions were observed and the authors used a model of formation via GI, along with their observational sensitivity, to place upper-limits on the number of gas-giants that can be formed through GI. They found that the limit is less than 30% for B2-A0 host stars, and less than 8% for FGKM stars. We caution that the uncertainty in the final mass of companions formed through GI makes the authors' estimates of a detection threshold uncertain. In addition, they did not take into account radial migration of the planets. Therefore, the upper-limits on formation via GI are expected to be too low. However, such observational surveys do show that gas-giant companions more massive than roughly 10 M_{Jup} are expected to be rare beyond about 50 AU. Lafrenière et al. (2007) placed an upper limit of 2% on the number of stars with at least one brown dwarf in the range 50-250 AU, with no assumptions regarding the mass or semi-major axis distribution. More sensitive surveys, that are able to probe down to the limits of the gas-giant mass regime, will be able to determine how common gas-giant companions at large radii are, and place constraints on their formation.

Gravitational microlensing is an additional technique that can detect gas-giants that are on large orbits. Surveys such as the Microlensing Observations in Astro-

physics (MOA) and the Optical Gravitational Lensing Experiment (OGLE) observe millions of stars in order to detect microlensing events, in which the brightness of a background star briefly increases due to gravitational lensing from a foreground body. For a large number of events, the timing of the lensing events can be used to determine the mass function of the gravitational lens population. Using this technique, Sumi et al. (2011) found that there is a large population of Jupiter-mass objects at radii larger than 10 AU from any host-star; in fact, there are roughly 2 objects per main-sequence star in the range 0.08-1 $M_{\odot}$.

This work suggests that there is a very large population of gas-giant planets that are either free-floating, or are on large (greater than 10 AU) orbits around their host star. More sensitive direct imaging surveys will determine the fraction of these planets that are companions on large orbits, but those already completed (Lafrenière et al. 2007; Janson et al. 2011, 2012) indicate that most of the population is free-floating.

The turnover in the initial mass function (Chabrier 2003) means that free-floating planets are unlikely to be produced in molecular clouds in the same way as stars. The most common explanation for free-floating planets has been that they are the result of planet-planet scattering in multiple planet systems. However, Veras & Raymond (2012), using n-body simulations to examine ejection frequency as well as observational data on the frequency and number of gas-giants around stars, found that scattering is unable to explain the large number of free-floating planets indicated by the microlensing observations. The authors note that scattering may explain the free-floating population if the frequency of Sumi et al. (2011) is over-estimated. Given that the estimate is based on only ten events, this appears plausible. If, however, the estimate is accurate, and there are nearly two free-floating Jupiter-mass per main-sequence star, then additional mechanisms will be required. These could include ejections from multiple-star systems, perturbations from the stellar cluster environment, and ejections promoted by post-main-sequence mass loss (see Veras & Raymond 2012 and references therein).

2.5.4 Accordance with models

Core accretion cannot be responsible for the in situ formation of the observed population of gas-giants beyond roughly 40 AU (Ida & Lin 2004a; Dodson-Robinson et al. 2009; Rafikov 2011), since the timescales to build cores massive enough for runaway gas accretion exceed the observed lifetimes of protostellar discs. In addition, planet-planet scattering is an unlikely explanation for multiple systems such as HR 8799 (Dodson-Robinson et al. 2009). More sensitive direct imaging surveys will help place constraints on the fraction of gas-giants at large radii that may have been scattered from the inner disc through the detection of any trend of frequency with the age of the system (Dodson-Robinson et al. 2009). Other mechanisms for transferring gas-giants formed through core accretion in the inner regions to the outer regions, such as the resonant scattering of Crida et al. (2009), require more attention.

In contrast, analytic considerations and numerical simulations agree that in situ formation of gas-giants is likely at large radii (roughly 100 AU and larger). Consequently, formation via GI appears the more natural explanation for gas-giants on large orbits. However, more theoretical work must be done in order to explain the observed masses, and the lack of migration to the inner disc. Vorobyov & Basu (2010a), for example, found from their simulations that most gas-giants formed in the outer region migrate inwards and accrete onto the star, but that there is occasionally a surviving population of gas-giants at large radii. It may be that keeping gas-giants on large orbits requires special conditions and may be rare. Better direct imaging observations, with detection limits to the bottom of the gas-giant mass regime will provide better constraints on formation via GI.

2.6 Conclusions

This chapter has been a review, from both a theoretical and observational standpoint, of the current understanding of gas-giant formation in protostellar discs. The physical conditions of the inner regions of discs are favourable to the in situ formation via core accretion, but not via GI, whereas the reverse is true in the outer regions. This has

led to the idea that there are two modes of gas-giant formation (Boley 2009) at work in protostellar discs: core accretion in the inner regions and GI in the outer regions. More research is necessary to consider the possibility, and consequences, of migration from one region to another.

In the next two chapters, we present our own research that has added to the understanding of the conditions necessary for fragmentation via GI in protostellar discs. This research has used carefully-tested 3D radiation hydrodynamics, with a robust photosphere boundary condition, to demonstrate that fragmentation does not occur in the inner regions of protostellar discs because cooling is not fast enough, whereas fragmentation can take place in the outer regions. In addition, we have developed a more detailed physical framework to explain the link between cooling and fragmentation. This physical model agrees well with the outcomes of our 3D RHD simulations, and can also be used to explain the cooling criterion and calculate the critical cooling time. This research has helped inform the conclusions put forward in this review.

Bibliography

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

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

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

Alibert Y., Mordasini C., Benz W., Winisdoerffer C., 2005, A&A, 434, 343

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

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

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

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

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

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

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

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

Baruteau C., Meru F., Paardekooper S.-J., 2011, MNRAS, 416, 1971

Basu S., 1997, ApJ, 485, 240

Batalha N. M. et al., 2012, ArXiv e-prints: 1202.5852

Bate M. R., 2011, MNRAS, 417, 2036

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

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

Birnstiel T., Dullemond C. P., Brauer F., 2010, A&A, 513, A79+

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

Boley A. C., Durisen R. H., Nordlund Å., Lord J., 2007, apj, 665, 1254

Boley A. C., Hayfield T., Mayer L., Durisen R. H., 2010, Icarus, 207, 509

Boley A. C., Mejía A. C., Durisen R. H., Cai K., Pickett M. K., D'Alessio P., 2006, ApJ, 651, 517

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

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

Boss A. P., 2011, ApJ, 731, 74

Boss A. P., Wetherill G. W., Haghighipour N., 2002, Icarus, 156, 291

Bowler B. P., Liu M. C., Kraus A. L., Mann A. W., Ireland M. J., 2011, ApJ, 743, 148

Brinch C., Crapsi A., Jørgensen J. K., Hogerheijde M. R., Hill T., 2007, A&A, 475, 915

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

Burkert A., Bodenheimer P., 2000, ApJ, 543, 822

Cai K., Durisen R. H., Michael S., Boley A. C., Mejía A. C., Pickett M. K., D'Alessio P., 2006, ApjL, 636, L149

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

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

Cassan A. et al., 2012, Nature, 481, 167

Chabrier G., 2003, PASP, 115, 763

Chabrier G., Baraffe I., Allard F., Hauschildt P., 2000, ApJ, 542, 464

Chauvin G., Lagrange A.-M., Dumas C., Zuckerman B., Mouillet D., Song I., Beuzit J.-L., Lowrance P., 2004, A&A, 425, L29

Chiang E. I., Goldreich P., 1997, ApJ, 490, 368

Cieza L. et al., 2007, ApJ, 667, 308

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

Commerçon B., Hennebelle P., Henning T., 2011, ApjL, 742, L9

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

Crida A., Masset F., Morbidelli A., 2009, ApjL, 705, L148

D'Alessio P., Calvet N., Hartmann L., 1997, ApJ, 474, 397

D'Alessio P., Canto J., Calvet N., Lizano S., 1998, ApJ, 500, 411

Dauphas N., Chaussidon M., 2011, Annual Review of Earth and Planetary Sciences, 39, 351

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

Dominik C., Blum J., Cuzzi J. N., Wurm G., 2007, Protostars and Planets V, 783

Duffin D. F., Pudritz R. E., 2009, ApjL, 706, L46

Dullemond C. P., Dominik C., 2004, A&A, 417, 159

Dullemond C. P., Dominik C., 2005, A&A, 434, 971

Dullemond C. P., Hollenbach D., Kamp I., D'Alessio P., 2007, Protostars and Planets V, 555

Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, Protostars and Planets V, 607

Eisner J. A., Hillenbrand L. A., Carpenter J. M., Wolf S., 2005, ApJ, 635, 396

Enoch M. L. et al., 2011, ApJS, 195, 21

Evans N. et al., 2009a, ArXiv e-prints: 0901.1691

Evans, II N. J. et al., 2009b, ApJS, 181, 321

Fedele D., van den Ancker M. E., Henning T., Jayawardhana R., Oliveira J. M., 2010, A&A, 510, A72

Fischer D. A., Valenti J., 2005, ApJ, 622, 1102

Forgan D., Rice K., 2011, MNRAS, 417, 1928

Frank J., King A., Raine D. J., 2002, Accretion Power in Astrophysics: Third Edition, Frank, J., King, A., & Raine, D. J., ed.

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

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

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

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

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

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

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

Gorti U., Hollenbach D., 2009, ApJ, 690, 1539

Greaves J. S., Rice W. K. M., 2010, MNRAS, 407, 1981

Gregory P. C., Fischer D. A., 2010, MNRAS, 403, 731

Haisch, Jr. K. E., Lada E. A., Lada C. J., 2001, ApjL, 553, L153

Halbwachs J. L., Arenou F., Mayor M., Udry S., Queloz D., 2000, A&A, 355, 581

Halbwachs J. L., Mayor M., Udry S., Arenou F., 2003, A&A, 397, 159

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

Hartmann L., Kenyon S. J., 1996, ARA&A, 34, 207

Hasegawa Y., Pudritz R. E., 2011, MNRAS, 417, 1236

Hayashi C., 1981, Progress of Theoretical Physics Supplement, 70, 35

Hayfield T., Mayer L., Wadsley J., Boley A. C., 2011, MNRAS, 417, 1839

Helled R., Bodenheimer P., 2011, Icarus, 211, 939

Hollenbach D., Johnstone D., Lizano S., Shu F., 1994, ApJ, 428, 654

Howard A. W. et al., 2011, ArXiv e-prints: 1103.2541

Howard A. W. et al., 2010, Science, 330, 653

Hueso R., Guillot T., 2005, A&A, 442, 703

Ida S., Lin D. N. C., 2004a, ApJ, 604, 388

Ida S., Lin D. N. C., 2004b, ApJ, 616, 567

Ida S., Lin D. N. C., 2005, ApJ, 626, 1045

Ida S., Lin D. N. C., 2008, ApJ, 673, 487

Isella A., Carpenter J. M., Sargent A. I., 2009, ApJ, 701, 260

Janson M., Bonavita M., Klahr H., Lafrenière D., 2012, ApJ, 745, 4

Janson M., Bonavita M., Klahr H., Lafrenière D., Jayawardhana R., Zinnecker H., 2011, ApJ, 736, 89

Johansen A., Klahr H., Henning T., 2006, ApJ, 636, 1121

Johansen A., Oishi J. S., Mac Low M.-M., Klahr H., Henning T., Youdin A., 2007, Nature, 448, 1022

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

Jørgensen J. K. et al., 2007, ApJ, 659, 479

Jørgensen J. K., van Dishoeck E. F., Visser R., Bourke T. L., Wilner D. J., Lommen D., Hogerheijde M. R., Myers P. C., 2009, A&A, 507, 861

Kenyon S. J., Hartmann L., 1987, ApJ, 323, 714

Kenyon S. J., Hartmann L. W., Strom K. M., Strom S. E., 1990, AJ, 99, 869

Kim W.-T., Ostriker E. C., Stone J. M., 2002, ApJ, 581, 1080

Kitamura Y., Momose M., Yokogawa S., Kawabe R., Tamura M., Ida S., 2002, ApJ, 581, 357

Klahr H. H., Bodenheimer P., 2003, ApJ, 582, 869

Kratter K. M., Matzner C. D., Krumholz M. R., 2008, ApJ, 681, 375

Kratter K. M., Murray-Clay R. A., 2011, ApJ, 740, 1

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

Lada C. J., 1987, in IAU Symposium, Vol. 115, Star Forming Regions, M. Peimbert & J. Jugaku, ed., pp. 1–17

Lada C. J., Wilking B. A., 1984, ApJ, 287, 610

Lafrenière D. et al., 2007, ApJ, 670, 1367

Lafrenière D., Jayawardhana R., Janson M., Helling C., Witte S., Hauschildt P., 2011, ApJ, 730, 42

Larson R. B., 1969, MNRAS, 145, 271

Larson R. B., 1984, MNRAS, 206, 197

Laughlin G., Bodenheimer P., 1994, ApJ, 436, 335

Laughlin G., Korchagin V., Adams F. C., 1997, ApJ, 477, 410

Laughlin G., Korchagin V., Adams F. C., 1998, ApJ, 504, 945

Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846

Lin D. N. C., Pringle J. E., 1990, ApJ, 358, 515

Lissauer J. J., 1993, ARA&A, 31, 129

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

Lissauer J. J. et al., 2012, ArXiv e-prints: 1201.5424

Lissauer J. J., Stevenson D. J., 2007, Protostars and Planets V, 591

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

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

Luhman K. L. et al., 2007, ApJ, 654, 570

Lynden-Bell D., Pringle J. E., 1974, MNRAS, 168, 603

Machida M. N., Inutsuka S.-i., Matsumoto T., 2010, ApJ, 724, 1006

Machida M. N., Matsumoto T., 2011, MNRAS, 413, 2767

Mamatsashvili G. R., Rice W. K. M., 2009, MNRAS, 394, 2153

Marcy G. W. et al., 2008, Physica Scripta Volume T, 130, 014001

Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348

Marois C., Zuckerman B., Konopacky Q. M., Macintosh B., Barman T., 2010, Nature, 468, 1080

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

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

Mayer L., Quinn T., Wadsley J., Stadel J., 2004, ApJ, 609, 1045

Mayor M. et al., 2011, ArXiv e-prints: 1109.2497

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

Mejía A. C., Durisen R. H., Pickett M. K., Cai K., 2005, ApJ, 619, 1098

Mellon R. R., Li Z.-Y., 2008, ApJ, 681, 1356

Meru F., Bate M. R., 2010a, MNRAS, 406, 2279

Meru F., Bate M. R., 2010b, MNRAS, L174+

Meru F., Bate M. R., 2011, MNRAS, 410, 559

Michael S., Durisen R. H., Boley A. C., 2011, ApjL, 737, L42

Mordasini C., Alibert Y., Benz W., Klahr H., Henning T., 2012, ArXiv e-prints: 1201.1036

Mordasini C., Alibert Y., Benz W., Naef D., 2009, A&A, 501, 1161

Mordasini C., Dittkrist K.-M., Alibert Y., Klahr H., Benz W., Henning T., 2011, in IAU Symposium, Vol. 276, IAU Symposium, A. Sozzetti, M. G. Lattanzi, & A. P. Boss, ed., pp. 72–75

Mordasini C., Klahr H., Alibert Y., Benz W., Dittkrist K.-M., 2010, ArXiv e-prints: 1012.5281

Muzerolle J., Calvet N., Hartmann L., 2001, ApJ, 550, 944

Nayakshin S., 2010, MNRAS, 408, L36

Nero D., Bjorkman J. E., 2009, ApjL, 702, L163

Neuhäuser R., Guenther E. W., Wuchterl G., Mugrauer M., Bedalov A., Hauschildt P. H., 2005, A&A, 435, L13

O'Dell C. R., Wen Z., Hu X., 1993, ApJ, 410, 696

Paardekooper S.-J., 2012, MNRAS, 421, 3286

Paardekooper S.-J., Mellema G., 2006, A&A, 459, L17

Papaloizou J. C. B., Nelson R. P., Kley W., Masset F. S., Artymowicz P., 2007, Protostars and Planets V, 655

Pickett B. K., Cassen P., Durisen R. H., Link R., 2000, ApJ, 529, 1034

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62

Rafikov R. R., 2005, ApjL, 621, L69

Rafikov R. R., 2007, ApJ, 662, 642

Rafikov R. R., 2011, ApJ, 727, 86

Rice W. K. M., Armitage P. J., Bate M. R., Bonnell I. A., 2003, MNRAS, 339, 1025

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

Rice W. K. M., Mayo J. H., Armitage P. J., 2010, MNRAS, 402, 1740

Sahlmann J. et al., 2011, A&A, 525, A95

Seifried D., Banerjee R., Klessen R. S., Duffin D., Pudritz R. E., 2011, MNRAS, 417, 1054

Seifried D., Banerjee R., Pudritz R. E., Klessen R. S., 2012, MNRAS, L442

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Stamatellos D., Maury A., Whitworth A., André P., 2011, MNRAS, 413, 1787

Stamatellos D., Whitworth A. P., 2008, A&A, 480, 879

Strom K. M., Strom S. E., Edwards S., Cabrit S., Skrutskie M. F., 1989, AJ, 97, 1451

Sumi T. et al., 2011, Nature, 473, 349

Tanaka H., Takeuchi T., Ward W. R., 2002, ApJ, 565, 1257

Toomre A., 1964, ApJ, 139, 1217

Udry S., Fischer D., Queloz D., 2007, Protostars and Planets V, 685

Udry S., Santos N. C., 2007, ARA&A, 45, 397

Veras D., Raymond S. N., 2012, MNRAS, 421, L117

Vicente S. M., Alves J., 2005, A&A, 441, 195

Vorobyov E. I., 2010, ApJ, 723, 1294

Vorobyov E. I., 2011, ApJ, 729, 146

Vorobyov E. I., Basu S., 2010a, ApjL, 714, L133

Vorobyov E. I., Basu S., 2010b, ApJ, 719, 1896

Weidenschilling S. J., 1977, Ap&SS, 51, 153

Wetherill G. W., Stewart G. R., 1989, Icarus, 77, 330

Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67

Wolf S., Padgett D. L., Stapelfeldt K. R., 2003, ApJ, 588, 373

Yorke H. W., Bodenheimer P., Laughlin G., 1993, ApJ, 411, 274

Youdin A. N., Goodman J., 2005, ApJ, 620, 459

Youdin A. N., Shu F. H., 2002, ApJ, 580, 494

Chapter 3

The importance of photosphere cooling in simulations of gravitational instability in the inner regions of protostellar discs

3.1 Introduction

The stability of self-gravitating discs can be characterized by the Toomre Q parameter (Toomre 1964):

$$Q = \frac{c_s \kappa_e}{\pi G \Sigma}, \tag{3.1}$$

where c_s is the sound-speed, κ_e is the epicyclic frequency (approximately the rotation rate, Ω for Keplerian discs), G is the gravitational constant, and Σ is the surface density. Low values of $Q \approx 1$, as would be found in a massive, cold disc, lead to gravitational instability (GI). Numerical work (Boss (1997); Mayer et al. (2004)) has shown that protostellar discs can be susceptible to GI and as a result can fragment, producing bound clumps that are potential precursors to Jovian, or larger extrasolar, planets.

However, these early works relied on simplified equations of state (either isothermal or isothermal with a switch to adiabatic during instability). Further numerical studies have demonstrated that fragmentation of these discs is dependent on the cool-

ing timescale of the gas. In general, unstable discs fragment if the cooling times are short:

$$t_{cool}\Omega < \eta, \tag{3.2}$$

where $\eta = 3$ for $\gamma = 2$ gas (Gammie 2001), $\eta = 7$ for $\gamma = 5/3$, and $\eta = 13$ for $\gamma = 7/5$ (Rice, Lodato, & Armitage 2005). (See Meru & Bate (2010b) for recent work concerning the critical values of cooling times.)

In order to accurately model energy transport and cooling, it is necessary to include radiative transfer in the detailed study of the stability of protostellar discs. This has been incorporated in a number of studies investigating the likelihood of fragmentation in the inner tens of AU of discs. However, there remains open debate as to the viability of planet formation via GI in this regime, since these studies have shown different outcomes (Boss (2007) and Mayer et al. (2007) observe fragmentation, while Cai et al. (2006) and Boley et al. (2007a) do not). Meru & Bate (2010a) observe fragmentation in the inner regions of protostellar discs; however, their discs begin in a very unstable state (with Toomre $Q < 1$ initially).

For studies of giant planet formation via GI, accurate photosphere boundary treatments are of fundamental importance since the photosphere controls the rate at which the disc is able to cool, which ultimately determines whether or not the disc is able to fragment if gravitationally unstable. Indeed, Cai et al. (2010) recently carried out a simulation in an attempt to compare with the results of Boss (2007) (which show fragmentation). The authors attempted to remove any differences between this simulation and the work of Boss (2007), using the same initial conditions and attempting to replicate the input physics. However, in contrast to the results of Boss (2007), the authors did not observe fragmentation, and concluded that "because everything else is the same, the disagreement must be due to differences in the treatment of the optically thin regions and the photospheric BC's."

We demonstrate that previous photosphere boundary conditions using edge-particles have over-estimated cooling-rates of discs by a large margin and can suffer systematic effects due to dynamics. We present a new photosphere treatment using edge-particles that is able to accurately determine photosphere areas, even for dynamic systems. The

method is shown to be able to satisfy the relaxation test proposed by Boley et al. (2007a); this is the first time that this important test case has been performed with SPH.

We use this new photosphere treatment, coupled to a new implementation of RT in the flux-limited diffusion (FLD) approximation in the TreeSPH code Gasoline (Wadsley, Stadel, & Quinn 2004), to study the viability of fragmentation via GI as a planet formation mechanism in the inner tens of AU of protostellar discs. We find that discs that are unstable in this regime do not fragment because they cannot cool fast enough.

Of particular importance, we rerun a simulation of a disc of mean molecular weight $\mu = 2.7$ that was observed to fragment in previous work (Mayer et al. 2007). In contrast to that result, we find that this disc does not fragment, despite allowing the disc-mass to increase to nearly three times the value needed for fragmentation as determined by Mayer et al. (2007). With our more accurate treatment of photosphere cooling, we observe that the cooling time is far too long to allow fragmentation. This result emphasizes the importance of accurate photosphere cooling treatments and the difficulty of forming giant planets through fragmentation in the inner regions of discs.

The structure of the paper is as follows: in §3.2 we outline our new implementation of FLD in Gasoline; in §3.3, we present test-cases of this implementation; in §3.4, we discuss flaws in previous photosphere boundary conditions and describe a new, more accurate photosphere boundary condition that satisfies the Boley et al. (2007a) relaxation test; in §3.5, we show the results of simulations of unstable protostellar discs; and in §3.6, we discuss the implications of this work.

3.2 Methods for FLD

3.2.1 Fundamental equations of radiation hydrodynamics

The equations of radiation hydrodynamics (RHD) to order (v/c) accuracy in a frame comoving with the fluid, assuming local thermodynamic equilibrium, and integrated over frequency are given by (see Mihalas & Mihalas (1984) and Turner & Stone

(2001)):

$$\frac{D\rho}{Dt} = -\rho\nabla\cdot\mathbf{v}, \tag{3.3}$$

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p + \frac{1}{c}\chi_F\rho\mathbf{F}, \tag{3.4}$$

$$\rho\frac{D}{Dt}\left(\frac{E}{\rho}\right) = -\nabla\cdot\mathbf{F} - \nabla\mathbf{v}:\mathbf{P} + 4\pi\kappa_P\rho B - c\kappa_E\rho E, \tag{3.5}$$

$$\rho\frac{D}{Dt}\left(\frac{e}{\rho}\right) = -p\nabla\cdot\mathbf{v} - 4\pi\kappa_P\rho B + c\kappa_E\rho E, \quad \text{and} \tag{3.6}$$

$$\frac{\rho}{c^2}\frac{D}{Dt}\left(\frac{\mathbf{F}}{\rho}\right) = -\nabla\cdot\mathbf{P} - \frac{1}{c}\chi_F\rho\mathbf{F}. \tag{3.7}$$

The above equations describe mass, momentum, internal gas energy, internal radiation energy, and radiative flux in terms of the comoving derivative, $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla$. In the above set of equations, ρ is the gas density, $\mathbf{v}$ is the fluid velocity, p is the gas pressure, E is the radiation energy density, e is the gas internal energy density, $\mathbf{F}$ is the radiative flux, and $\mathbf{P}$ is the radiation pressure tensor. Furthermore, c is the speed of light, κ_E is the energy mean absorption, κ_P is the Planck mean absorption, and χ_F is the flux mean total opacity (absorption plus scattering). We hereafter take $\kappa_E = \kappa_P = \chi_F = \kappa_R$, the Rosseland mean opacity. The integrated Planck function is $B = (\sigma_{SB}/\pi)T_g^4$, and one can relate a radiation temperature to the radiation energy density through the relation $E = 4\sigma_{SB}T_r^4/c$, where σ_{SB} is the Stefan-Boltzmann constant.

3.2.2 Flux-limited diffusion

The flux-limited diffusion approximation replaces the equation for the radiative flux with a simplified version:

$$\mathbf{F} = \frac{c\lambda}{\kappa_R\rho}\nabla E. \tag{3.8}$$

This form of the radiative flux contains λ, a *flux-limiter*, whose functional form tends to the correct limit for the flux in optically thick regions ($\mathbf{F} \rightarrow (c/3\kappa_R\rho)\nabla E$, resulting in a diffusion term in the radiation energy equation) and enforces causality in optically

thin regions, where the denominator in the above tends to zero ($\mathbf{F} \rightarrow cE$).

The flux-limiter smoothly transitions between the two limits and also specifies the angular dependence of the radiation field (Turner & Stone 2001). In optically thick regions, flux-limited diffusion is equivalent to using the Eddington approximation ($P_{ii} = 1/3E$) and assuming steady-state in the radiation energy equation.

In the FLD framework, the radiation pressure tensor is given by

$$\mathbf{P} = \mathbf{f}E, \tag{3.9}$$

where $\mathbf{f}$ is the Eddington tensor:

$$\mathbf{f} = 1/2(1-f)\mathbf{I} + 1/2(3f-1)\hat{n}\hat{n}, \tag{3.10}$$

with $\hat{n} = \nabla E/|\nabla E|$. The Eddington factor is

$$f = \lambda + \lambda^2 R^2 \tag{3.11}$$

where $R = |\nabla E|/(\kappa_R \rho E)$.

There are a number of flux-limiters in the literature; we choose to use that of Levermore & Pomraning (1981) in order to consistently compare our test-results to those of Whitehouse et al. (2005):

$$\lambda = \frac{2+R}{6+3R+R^2}. \tag{3.12}$$

3.2.3 Smoothed Particle Hydrodynamics

Smoothed Particle Hydrodynamics (SPH; Lucy (1977), Gingold & Monaghan (1977)) is a Lagrangian Method of solving the fluid equations that is widely used in astrophysics (see reviews by Monaghan (1992) and Price (2005)).

The method's foundation is the interpolation of fluid variables using sample-points of fixed mass (particles) that move with the fluid velocity. Thus, each particle (i) calculates a given fluid variable (A_i) using an interpolation (with interpolating, or smoothing, function W) of the values of that fluid variable (A_j) from nearby particles

(j):

$$A_i = \sum_j m_j \frac{A_j}{\rho_j} W_{ij}\left(\mathbf{r}_{ij}, h\right), \tag{3.13}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r_j}$ and h is the smoothing length. First derivatives in the SPH formalism are rather straightforward, since most of the quantities in the above equation are constants (they are particle quantities); the only exception is the smoothing function, W. Thus, the gradient of a fluid variable is:

$$\left(\nabla A\right)_i = \sum_j m_j \frac{A_j}{\rho_j} \nabla W_{ij}\left(\mathbf{r}_{ij}, h\right) \tag{3.14}$$

Second derivatives in SPH are noisy due to the disorder usually present in the particle distribution. Cleary & Monaghan (1999) showed that an integral approximation of the second derivative in the conduction equation— of the same form as the second derivative in the radiation energy equation in FLD— of the form

$$\left(\nabla\left(k\nabla A\right)\right)_i = \sum_j \frac{m_j}{\rho_i \rho_j} \frac{4k_i k_j}{k_i + k_j} \left(A_i - A_j\right) \frac{\nabla W_{ij}\left(\mathbf{r}_{ij}, h\right)}{\mathbf{r}_{ij}} \tag{3.15}$$

ensures the continuity of flux.

3.2.4 The implicit FLD energy equations in the SPH formalism

One can rewrite the FLD equations for internal radiation and internal gas energy in the SPH formalism given the above definitions. In addition, one must decide on a manner in which to time-integrate the equations.

The most straightforward manner in which to do this is using an explicit time integration scheme, one in which a fluid variable is updated using a rate-of-change calculated at the current timestep. The simplest such scheme is forwards-Euler:

$$A_i^{n+1} = A_i^n + \Delta t \left[\frac{DA_i}{Dt}\right]^n; \tag{3.16}$$

here, the superscript denotes the timestep and Δt is the timestep.

Unfortunately, explicit time-integration has the drawback that in order for the solution to be numerically stable, there are constraints on the size of the timestep that can be used. The constraining timesteps for explicitly integrating our equations of FLD (see Whitehouse & Bate 2004 for the criteria, but the most notable is the diffusion timestep: $\Delta t = \frac{\kappa\rho}{c\lambda}h^2$) can be much smaller than the timestep for hydrodynamics alone. Consequently, we use an implicit time-integration scheme, in which the rate-of-change in the fluid variable is calculated at the future timestep*. The simplest implicit scheme is backwards-Euler:

$$A_i^{n+1} = A_i^n + \Delta t \left[\frac{DA_i}{Dt}\right]^{n+1}, \tag{3.17}$$

which is of first-order accuracy.

It is a straightforward exercise to use the information in this section to write the FLD equations (3.5) & (3.6) for the radiative and the gas energies in the SPH formalism using a backwards-Euler time-integration. For each particle,the resulting equations for the specific radiation energy ($\xi = E/\rho$) and the specific gas energy ($u = e/\rho$) are

$$\begin{aligned} f_{\xi,i} &= \xi_i^{n+1} - \xi_i^n \\ &\quad - \Delta t \left(\sum_j \frac{m_j}{\rho_i \rho_j} \frac{\nabla W_{ij}}{\mathbf{r_{ij}}} \left[\frac{4k_i k_j}{k_i + k_j}\right] c\left(\rho_i \xi_i - \rho_j \xi_j\right) - R_i \xi_i - ac\kappa_i \left[\frac{\rho_i \xi_i}{a} - T_g\left(u_i\right)^4\right] \right)^{n+1} \\ &\quad = 0, \end{aligned} \tag{3.18}$$

Additional comment for thesis: in the protostellar disc simulations presented in this thesis, the diffusion timestep is a factor of 10^2 to 10^5 times shorter than the hydrodynamic (Courant) timestep. As such, an explicit time integration of the equations of FLD is expected to make any simulation very computationally expensive. We therefore use an implicit method so that we can time-integrate on the much longer hydrodynamic timestep, significantly reducing the computational expense.

and

$$f_{u,i} = u_i^{n+1} - u_i^n - \Delta t \left(\frac{1}{2} \sum_j \left(F_p(u_i) + F_p(u_j) + \Pi_{ij} \right) m_j \mathbf{v}_{ij} \cdot \nabla W_{ij} + ac\kappa_i \left[\frac{\rho_i \xi_i}{a} - T_g \left(u_i \right)^4 \right] \right)^{n+1} = 0. \tag{3.19}$$

Note that the quantity R_i is just the inner product $\nabla \mathbf{v} : \mathbf{f}$ (from $\nabla \mathbf{v} : \mathbf{P}$ in the radiation energy equation, (3.6)), Π_{ij} is the standard SPH viscosity used for capturing shocks (with the Balsara (1995) switch), and there are two functions of the specific gas energy: the gas temperature, $T_g(u)$, and the pressure function, $F_p(u) = P/\rho^2$. Our method does not rely on a specific relation between these functions and the specific gas energy (such as the linear relation $T_g = u/c_v$, where c_v is the heat capacity). Boley et al. (2007b) made the case that such a linear relation for the gas temperature may have an unrealistic effect on simulations of protostellar disk stability, in which one should accurately take into account the rotational and vibrational modes of molecular hydrogen. In this section, we use the general functions $T_g(u)$ and $F_p(u)$; however, it should be noted that we have specifically implemented an adiabatic equation of state (used in the simulation of §3.5.1), as well as the equation of state of Boley et al. (2007a) (used in the simulation of §3.5.2).

3.2.5 Solving the implicit FLD energy equations

For each timestep in a simulation of a system of n particles, we have a system of $2n$ non-linear equations to solve, equations (3.18) & (3.19), with the unknowns being the specific radiation and specific gas energies of each particles at the future timestep, ξ_i^{n+1} and u_i^{n+1}. The other quantities in the implicit energy equations that are expected to be at the future timestep are related to the positions and velocities of the particles. These are calculated in the standard manner of the kick-drift-kick integration scheme used in Gasoline using predictor for the velocity.

The non-linear method: equations and iterative solution

Our method of solving the system of the $2n$ implicit energy equations is similar to that of Whitehouse et al. (2005): a non-linear solution using a basic iterative method.

As per Whitehouse et al. (2005), for each particle we can rewrite the implicit specific gas energy equation, (3.19), in terms of the specific radiation energy and substitute the result into the implicit specific radiation energy equations, (3.18). This results in a single non-linear equation that can be solved for the specific gas energy (if one assumes, as in Whitehouse et al. (2005) that $T = u/c_v$, the resulting equation is a quartic).

This equation is of the form:

$$g_i(u_i^{n+1}) = AT(u_i^{n+1})^4 + BF_p(u_i^{n+1}) + Cu_i^{n+1} + D = 0, \tag{3.20}$$

with the following definitions (keeping close to Whitehouse et al. (2005)):

$$A = \Delta t \Gamma \left(\chi - 1\right) \tag{3.21}$$

$$B = -\Delta t \left(\chi - 1 - \beta\right) P_{d,i} \tag{3.22}$$

$$C = \left(\chi - 1 - \beta\right) \tag{3.23}$$

$$D = \beta \left(\xi_i^n + \Delta t D_{n,i}\right) - \left(\chi - 1 - \beta\right) \left(u_i^n + \Delta t P_{n,i}\right) \tag{3.24}$$

$$\Gamma = ac\kappa_i \tag{3.25}$$

$$\beta = \Delta t c \kappa_i \rho_i \tag{3.26}$$

$$\chi = \Delta t \left(D_{d,i} - R_i\right) \tag{3.27}$$

$$D_{d,i} = \sum_j \frac{m_j}{\rho_j} \frac{\nabla W_{ij}}{\mathbf{r_{ij}}} c \left[\frac{4\kappa_i \kappa_j}{\kappa_i + \kappa_j}\right] \tag{3.28}$$

$$D_{n,i} = -\sum_j \frac{m_j}{\rho_i} \frac{\nabla W_{ij}}{\mathbf{r_{ij}}} c \left[\frac{4\kappa_i \kappa_j}{\kappa_i + \kappa_j}\right] \xi_j^{n+1} \tag{3.29}$$

$$P_{d,i} = \sum_j \frac{1}{2} m_j \mathbf{v}_{ij} \cdot \nabla W_{ij} \tag{3.30}$$

$$P_{n,i} = \sum_j \frac{1}{2} m_j \mathbf{v}_{ij} \cdot \nabla W_{ij} \left(F_p(u_j^{n+1}) + \Pi_{ij}\right) \tag{3.31}$$

$$\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j \tag{3.32}$$

For each iteration (l) of our solution method we first evaluate the coefficients for the above combined-energy equation, (3.20), (for each particle) using the values of the energies from the last $(l-1)$ iteration. Thus, for each particle, we have an equation of the form

$$g_{i,l}(u_{i,l}^{n+1}) = A_{l-1}T(u_{i,l}^{n+1})^4 + B_{l-1}F_p(u_{i,l}^{n+1}) + C_{l-1}u_{i,l}^{n+1} + D_{l-1} = 0, \tag{3.33}$$

which is solved using a Newton-Raphson root-finder. We then compute each particle's specific radiation energy using a rearrangement of the implicit specific radiation energy equation, (3.18):

$$\xi_{i,l}^{n+1} = \frac{\xi_i^n + \Delta t\left(\left[\frac{D\xi_i}{Dt}\right]^n + D_{n,i,l-1} + \Gamma_{l-1}T(u_i^{n+1})^4\right)}{C_{l-1}}. \tag{3.34}$$

Since we update each particles' energies only after we have calculated new energies for all particles, we are using Jacobi's method (this maintains the symmetry in the coefficients of neighbour-particles).

Convergence is achieved if the change in each specific energy over the iteration is small compared to the corresponding change over the entire timestep, that is, if

$$\left[\frac{|u_l^{n+1} - u_{l-1}^{n+1}|}{|u_l^{n+1} - u^n|}\right]_{RMS} < \epsilon \quad \text{and} \quad \left[\frac{|\xi_l^{n+1} - \xi_{l-1}^{n+1}|}{|\xi_l^{n+1} - \xi^n|}\right]_{RMS} < \epsilon, \tag{3.35}$$

for which we typically use $\epsilon = 10^{-4}$. If convergence is not achieved, we simply continue the iterative procedure.

A caveat concerning convergence

The Jacobi iterative method is guaranteed to converge for linear systems, but there is no such guarantee for non-linear systems. Indeed, in simulations of protostellar discs using implicit non-linear FLD (such as that in §3.5), we have observed non-convergent behaviour for a subset of the particles located in the outer region of the

disc. Individual non-convergent particles have been found to oscillate between two values for an energy, making convergence impossible.

A simple example of an iterative method showing oscillations between multiple solutions is the logistic map,

$$x_{n+1} = a x_n \left(1 - x_n\right), \tag{3.36}$$

for which there is a single solution for small values of a, but multiple solutions for a larger than a critical value. We make an analogy between the non-linear Jacobi method and the logistic map and link the growth factor, a to the relaxation parameter, ω. Each iteration of the Jacobi method gives an estimate of the specific radiation energy ξ_{n+1} (for example). We can choose to adjust this estimate by choosing instead a value:

$$\xi_i^{'n+1} = \xi_i^n + \omega \left(\xi_i^{n+1} - \xi_i^n\right). \tag{3.37}$$

If $\omega < 1$, then this is known as under-relaxation (Press et al. 2007); if $\omega > 1$, over-relaxation; and by default, $\omega = 1$.

If non-convergence is detected in the non-linear method for a given time-step, we decrease the relaxation parameter ω in order to converge to one solution, analogous to decreasing the growth factor a in the logistic map. Presently, our method for decreasing the relaxation parameter is purely an empirical one. When a simulation shows non-convergence, we examine the convergence as a function of iteration number. We then restart the simulation with a relaxation parameter of

$$\omega = \omega_0 + (1 - \omega_0)\, e^{-(n - n_0)/r}, \text{ for } n > n_0 \tag{3.38}$$

where ω_0 is a lower-bound to the relaxation parameter (typically 0.3), n_0 is the iteration at which the relaxation parameter deviates from unity (the iteration at which we observe that convergence is stalled), and r is a iteration-scale over which the under-relaxation occurs (typically n_0). Since non-convergence has been rare in our simulations, and the above method has been robust in preventing repeated non-convergence, we have not developed a more sophisticated method.

3.3 Testing the FLD

As pointed out by Boley et al. (2007a), the debate regarding GI's viability as a planet-formation mechanism necessitates a thorough testing of RT schemes. Although our method for FLD is not new, we present results for a number of standard tests here in order reassure the reader of the accurate treatment of RT in Gasoline. Consequently, one can be confident in the results of the simulations presented in §3.5.

Most astrophysical problems of interest are inherently 3-D. In addition, known issues in 1-D SPH, such as wall-heating, are known to be different in 3-D. In shock problems, it is possible for particles to slip past other particles in 3-D, which is not the case in 1-D. We therefore perform all of the following, essentially 1-D, tests in 3-D using periodic boundary conditions*. We compare to results from a 1-D FLD SPH code (Whitehouse & Bate (2004), Whitehouse et al. (2005)), a 2-D FLD grid code (Turner & Stone (2001)), and an adaptive 1-D RT grid code (Sincell et al. (1999)).

3.3.1 Heating and cooling terms

This test, done by Turner & Stone (2001), Whitehouse & Bate (2004), and Whitehouse et al. (2005), tests (in isolation) the equilibration terms in the gas and radiation energy equations, (3.18) and (3.19). Two tests are performed, one in which $T_g > T_r$ initially (the cooling case) and one in which $T_g < T_r$ initially (the heating case). In both instances, we follow the time-evolution of the gas energy as it moves towards $T_g = T_r$. For a stationary gas of constant density in which the radiation energy dominates the gas energy (and hence can be assumed to be constant), the time-evolution of the gas energy is given by

$$\frac{de}{dt} = c\kappa\rho E - ac\kappa \left(\frac{e}{\rho c_v} \right)^4 . \tag{3.39}$$

The system used is the same as that of Whitehouse & Bate (2004): it is composed of uniform density, $\rho = 10^{-7}$ g cm^{-3}, gas with an opacity of $\kappa = 0.4$ cm^2/g, adiabatic

**Additional comment for thesis*: by performing these tests in 3-D, rather than in 1-D, we avoid the numerical differences between 1-D and 3-D mentioned above. However, to test the code's accuracy in treating the geometric effects of multiple dimensions, it would be ideal to have true 3-D tests.

index of $\gamma = 5/3$, and radiation energy density of $E = 10^{12}$ erg cm^{-3}. The gas energy density is initially set to either $e = \rho u = 10^{2}$ erg cm^{-3} (the heating case) or $e = 10^{10}$ erg cm^{-3} (the cooling case).

Figure 3.1 shows the results of the implicit non-linear method using a time-step that is the greater of 10^{-11} s or 5% of the elapsed time, much larger than the initial cooling time. The agreement between the analytic result (red, dashed line) and the results (cooling: cross, heating: diamond) is excellent.

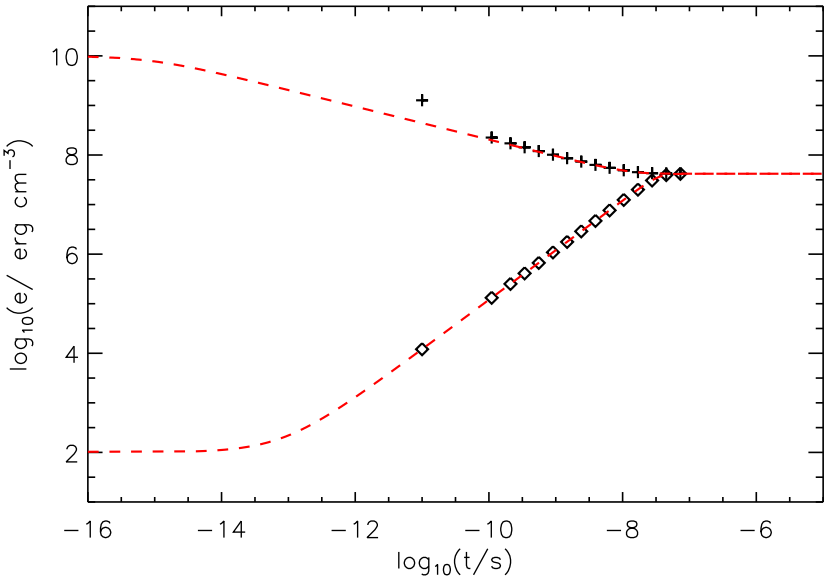

Figure 3.1: Testing the equilibration of the gas temperature with the radiation temperature for the cooling case ($T_g > T_r$; crosses) and the heating case ($T_g < T_r$; diamonds) using the implicit non-linear method with a time-step that is the greater of 10^{-11} s or 5% of the elapsed time. The analytic evolution — equation (3.39)— is given by the red, dashed-line. Only every 10^{th} time-step is plotted.

3.3.2 Propagating radiation fronts in optically-thin media

This test, also done by Turner & Stone (2001), Whitehouse & Bate (2004), and Whitehouse et al. (2005), tests the effectiveness of the flux-limiter to limit the flux to

cE in the free-streaming limit. The system is initially at a uniform density $\rho = 0.025$ g cm^{-3}, and the radiation energy density is $E = 10^{-1}$ erg cm^{-3} in the region [0.0, 0.1] cm and $E = 10^{-2}$ erg cm^{-3} in the region [0.1, 1.0] cm, with the gas energy in temperature-equilibrium. In this way, a radiation front is created within the gas. Using boundary conditions of $E = 10^{-1}$ erg cm^{-3} at $x = 0.0$ and $E = 10^{-2}$ erg cm^{-3} at $x = 1.0$ maintains this front for $t > 0$.

Figure 3.2 shows the results of this test at $t = 10^{-11}$ s for the implicit non-linear method using a variety of time-steps. The initial configuration is given by the solid line, while the results are given by the symbols. As one can see, the radiation front is quite diffusive, as seen in Turner & Stone (2001), Whitehouse & Bate (2004), and Whitehouse et al. (2005); however, its position is consistent with it travelling at the speed of light— this position is marked by the vertical dashed line.

Using a single timestep of 10^{-11} s, much larger than the Courant time of $\Delta t \sim h/2c \sim 10^{-13}$ s, results in a very diffuse front, as observed by Turner & Stone (2001) (more diffusive than Whitehouse et al. (2005), potentially due to the lower dimensionality, 1-D, and lower resolution of that work). This illustrates that the method remains stable for time-steps much larger than the physical time-scale, but can lose accuracy.

3.3.3 Optically thick (adiabatic) and optically thin (isothermal) shocks

This test, done by Whitehouse & Bate (2004) and Whitehouse et al. (2005), demonstrates the ability of the algorithm to capture the thermodynamic limits of radiative transport, as well as the transition regime. If a gas is unable to efficiently radiate away the thermal energy that it receives when it shocks, then it will be well characterized as being adiabatic; in contrast, if the gas *is* able to completely radiate this energy away, then it will be well characterized as being isothermal. We can examine the transition between these regimes simply by varying the opacity of the gas.

The system is set up in the same fashion as Whitehouse & Bate (2004), except that a slightly larger domain is used: $[-2 \times 10^{15}, 2 \times 10^{15}]$ cm. The gas is initially

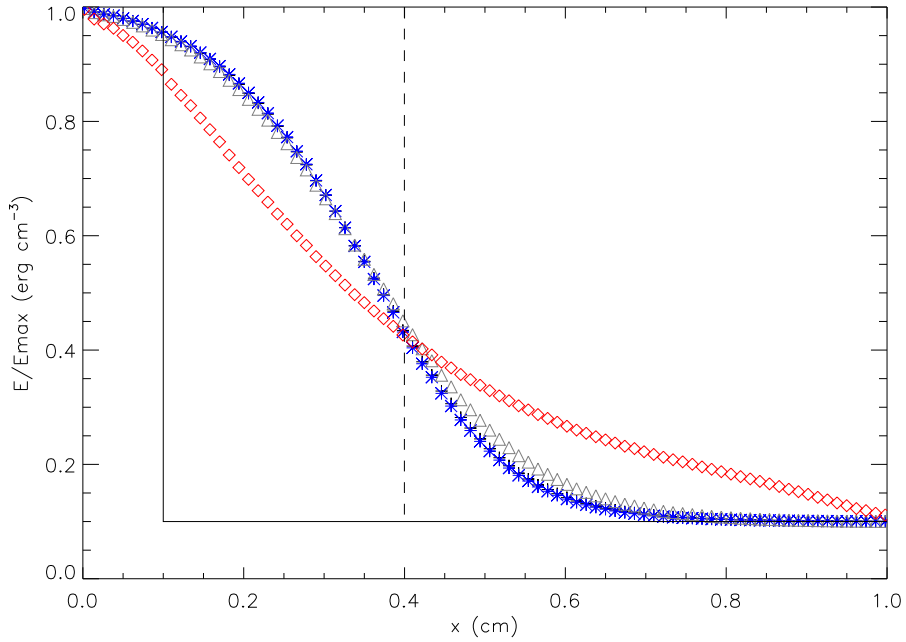

Figure 3.2: Testing the flux-limiter's ability to constrain the flux to cE . A radiation front (initial configuration given by the solid line) moves to the right in an optically-thin medium, with an expected position after $t = 10^{-11}$ s denoted by the dashed line. The results of the implicit non-linear method (particle values averaged over bins) for various time-steps are given by the symbols ($\Delta t = 10^{-11}$ s: diamonds; $\Delta t = 10^{-12}$ s: triangles; $\Delta t = 10^{-13}$ s: asterisks; $\Delta t = 10^{-14}$ s: crosses).

of uniform density $\rho = 10^{-10}$ g cm^{-3}, with $T_r = T_g = 1500$ K, an adiabatic index of $\gamma = 5/3$, and an opacity of one of $\kappa = \{4.0 \times 10^{1}, 4.0 \times 10^{-1}, 4.0 \times 10^{-3}, 4.0 \times 10^{-5}\}$ cm^2/g. The region $x < 0$ has a velocity equal to the sound speed, $v_x = c_s = 3.2 \times 10^5$ cm/s, in the positive x direction, while the region $x > 0$ has an equal speed, but in the negative x direction. This velocity configuration results in two shocks being created at the origin, one moving in the positive x direction and one moving in the negative x direction. The boundaries of the problem are implemented using reflective ghost particles that maintain the initial energies of the gas; these boundaries move towards the origin at the sound speed.

Figure 3.3 shows the results of this test using the implicit non-linear method. The left column is of density, while the right is of temperature. The top row has an

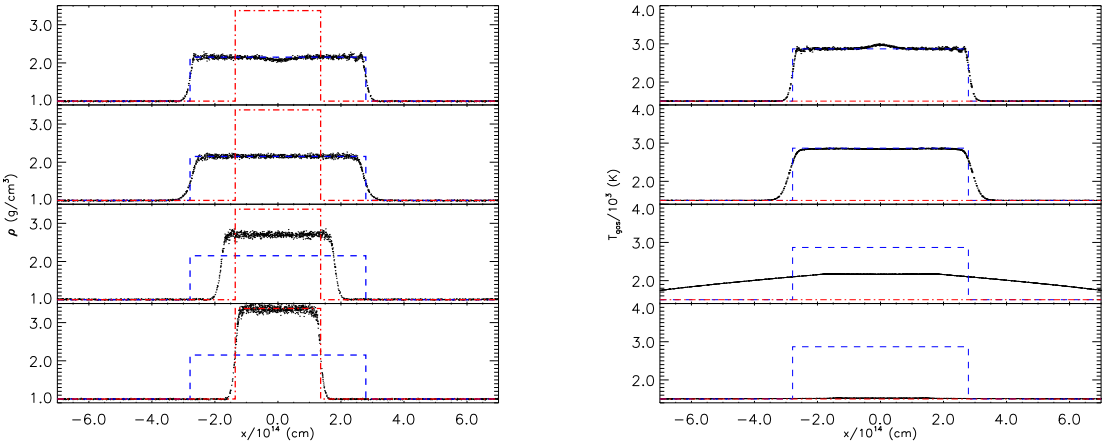

Figure 3.3: Testing the adiabatic and isothermal nature of radiative shocks, as well as the transition region. The left column is of density, while the right column is of temperature. The opacity changes with row, with values of (from top to bottom) $\kappa = \{4.0 \times 10^{1},\ 4.0 \times 10^{-1},\ 4.0 \times 10^{-3},\ 4.0 \times 10^{-5}\}$ cm^2/g. The results are given in black, while the analytic results are given by (for an isothermal shock) the red, dot-dashed line and (for an adiabatic shock) the blue, dashed line. Only a subset of the particles is plotted.

opacity of 4.0×10^1 cm^2/g, and moving downwards, the rows are arranged in order of decreasing opacity. The results are given in black, while the analytic results are given by (for an isothermal shock) the red, dot-dashed line and (for an adiabatic shock) the blue, dashed line. As is evident, the shock in the top row (with the largest opacity) closely follows the result for an adiabatic shock, while that in the bottom row (with the smallest opacity) closely follows the result for an isothermal shock. As expected, the shock with $\kappa = 4.0 \times 10^{-3}$ cm^2/g demonstrates the transition region between these limits. These results are consistent with those of Whitehouse & Bate (2004).

3.3.4 Subcritical and supercritical shocks

A subcritical shock is a radiative-hydrodynamic shock in which photons from the post-shock gas are unable to appreciably heat the pre-shock gas, resulting in a large temperature difference between pre- and post- shock gas. In contrast, in a supercritical shock photons from the post-shock gas *are* able to efficiently heat the pre-shock gas so that there is little temperature difference between the pre- and post-

shock gas.

Analytic approximations have been developed for these cases (Zel'dovich & Raizer (1967); with a concise overview given in Sincell et al. (1999)), and have been tested with a 1-D adaptive-grid radiative-hydrodynamics code (Sincell et al. 1999), as well as with FLD codes (Turner & Stone (2001); Whitehouse & Bate (2004); and Whitehouse et al. (2005)).

The problem is set up according to Sincell et al. (1999) with a uniform density $\rho = 7.78 \times 10^{10}$ g cm^{-3}, mean molecular weight $\mu = 0.5$, adiabatic index $\gamma = 5/3$, and a temperature gradient so that $T = 10 + 75x/(7 \times 10^{10})$ K (x is in cm). The radiation energy is initially in equilibrium with the gas energy. At $t = 0$ a piston (created with reflective ghost particles) moves supersonically into the gas at the origin with a velocity of $v = 6 \times 10^5$ cm/s for the sub-critical shock and $v = 1.6 \times 10^6$ cm/s for the super-critical shock.

Figure 3.4 shows the results using the implicit non-linear method for the sub-critical shock at a time $t \sim 10^{-11}$ s. The gas temperature is given by the black points, the radiation temperature by the blue points, and the analytic approximations by the red, dashed line. The left top panel shows the temperature profile, the right top panel shows the flux profile, and the bottom panel shows the Eddington factor profile. The analytic results are scaled to the post-shock temperature, T_1, and the peak flux, $F_0 = \sigma_{SB} T_1^4$, as in Sincell et al. (1999). The results are in agreement with the analytic approximations; however, the decline in gas-temperature in the post-shock region is too steep, and the flux is too large downstream of the shock. This was also observed by Turner & Stone (2001), who attributed this to an incorrect angular distribution of the radiation field. The Eddington factor shows a large spike at the shock front, which is not as visible in Turner & Stone (2001) or Whitehouse et al. (2005). This is caused by the large gradient in the radiation energy density at this location. As can be seen from the equation for the Eddington factor (3.11), for large values of ∇E, the Eddington factor approaches unity.

Figure 3.5 shows the results using the implicit non-linear method for the super-critical shocks at time $t \sim 10^{-11}$ s (see Figure 3.4 for details). As in Sincell et al. (1999), the analytic approximation uses $T_c = 1700$ K and $\tau_c = -5$. Without sufficient

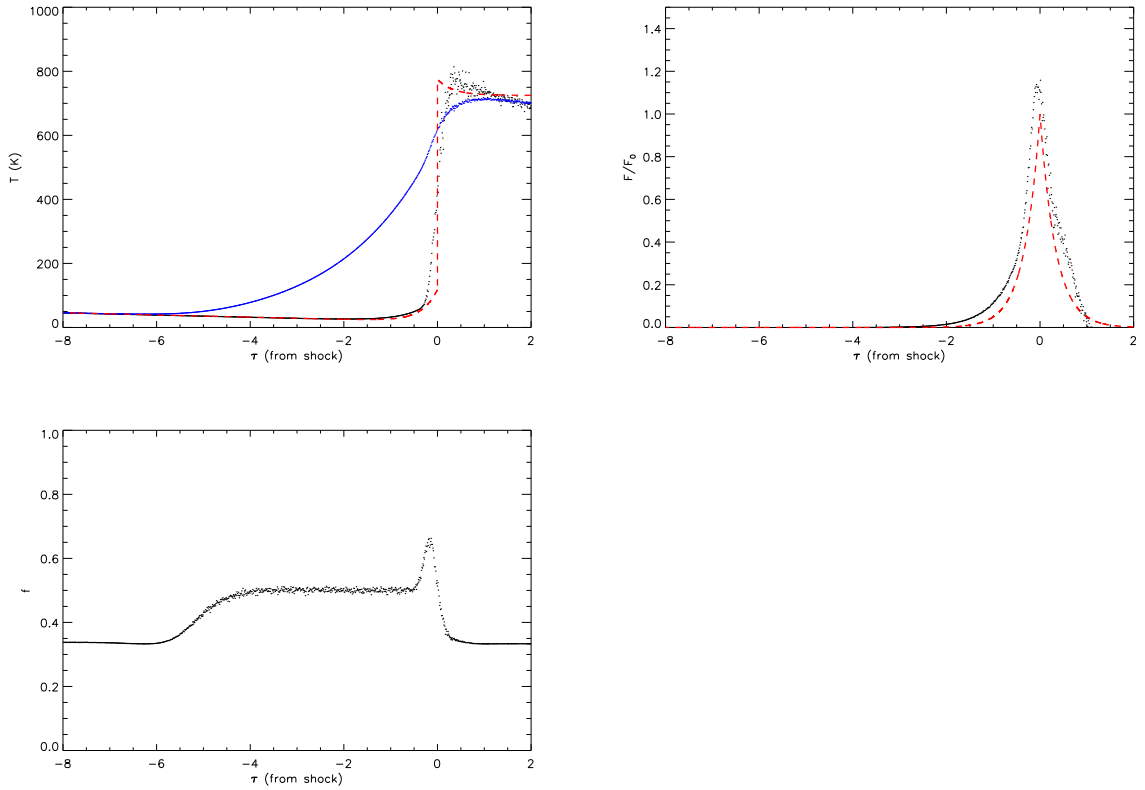

Figure 3.4: Simulating a subcritical shock using the implicit non-linear method. The gas temperature is given by the black points, the radiation temperature by the blue points, and the analytic approximations (Zel'dovich & Raizer 1967) by the red, dashed line. The top left panel shows the temperature profile, the top right panel shows the flux profile, and the bottom panel shows the Eddington factor profile. Only a subset of the particles is plotted.

resolution, we only weakly resolve the optically-thin temperature spike at the shock: we find a temperature jump of $\sim$ 130 K (compared to the expected 1600 K), and a thickness of $\tau \sim 0.4$ (compared to the expected $\tau = 0.15$). There is a large spike in the flux profile at the shock, which is likely caused by penetration of SPH particles (which cannot take place in 1-D tests). This feature can be removed by increasing the artificial viscosity parameters (we have observed this for values of $\alpha = 4$ and $\beta = 8$); however, the structure of the shock will be less well resolved if this is done. Consequently, one's choice of artificial viscosity parameters should be based on balancing

the errors that will occur in the flux profile and the shock structure.

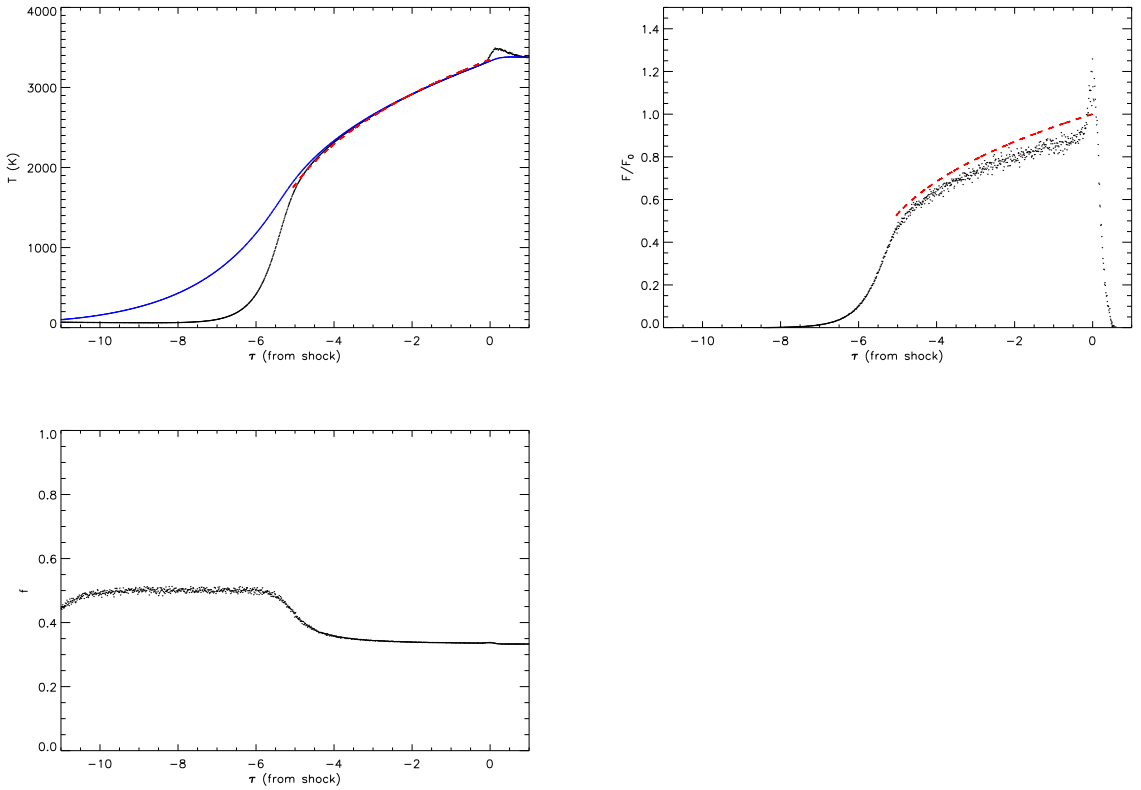

Figure 3.5: Simulating a supercritical shock using the implicit non-linear method. See Figure 3.4 for details.

3.4 Boundary cooling

3.4.1 Boundary cooling algorithm

The FLD algorithm described allows for the radiative transport of energy in all regimes of optical depth; however, in the SPH implementation of FLD, any radiative transport can only take place where there are particles present: no boundary conditions have as yet been specified.

In simulations of GI in protostellar discs, it is of fundamental importance to correctly consider the emission of energy from the disc's photosphere. Ultimately, it

is this emission that allows the disc to cool, and protostellar disc stability has been shown to be sensitive to the cooling rate (Gammie 2001).

There are a number of boundary conditions/photosphere treatments that couple to RT described in the literature. Fixed temperature boundaries are used by Boss (2001, 2007) (for gas below a critical density), as well as Meru & Bate (2010a) (for gas that is optically thin or above a critical scale height). Cai et al. (2006) allow the optically thin exterior of a disc to cool as much as is allowed by the gas's emissivity and couple this region to the optically thick region using an Eddington-like boundary condition. Forgan et al. (2009) couple FLD to an escape probability cooling approximation throughout the disc. Boley et al. (2007a) use vertical-ray transport to model RT (and cooling) in the z-direction, while using FLD in the other coordinates only in optically thick regions.

Our new photosphere boundary condition uses the basic idea of the boundary condition of Mayer et al. (2007): allowing the SPH particles on the "edge" of the disc to radiate away energy as if they were the edge of a plane-parallel atmosphere. However, there are a number of important flaws in the photosphere treatment of Mayer et al. (2007). As we demonstrate in §3.5.1, these flaws can lead to unphysical fragmentation in studies of GI in the inner regions of protostellar discs. We first review the boundary condition of Mayer et al. (2007) here and discuss its important flaws. We then overview our new boundary condition and demonstrate that it overcomes these flaws and is able to handle a published test-problem, the relaxation test of Boley et al. (2007a).

In the Mayer et al. (2007) boundary condition, particles are declared to be on the "edge" of the disc (edge-particles) based on three directions in polar coordinates: upwards ($+z$), downwards ($-z$) and out (r). Each particle searches within a distance $2h$, where h is the smoothing length, to determines whether any neighbour-particles reside within an "edge-detection angle" (EDA) of one of these three directions. In Mayer et al. (2007), this angle is treated as a free parameter and is varied between simulations. If no neighbour particles are found to lie within an EDA of one the three aforementioned directions, then this particle is marked as being an edge-particle. Each edge-particle has a surface area of $A_i = 4\pi h_i^2$ (the area of the photosphere for

which the edge-particle is responsible for), and loses specific internal gas energy at a rate of

$$\dot{u}_i = \frac{f A_i \sigma_{SB} T^4}{m_i}, \tag{3.40}$$

where $f = 1/2$ if the particle is an edge-particle in one of the three aforementioned directions, $f = 3/4$ if it an edge-particle in two of the three directions and $f = 1$ if it is an edge-particle in all three directions.

This method has a number of important flaws. The most important of these relate to the surface area estimate, which controls the rate at which a system will radiatively cool. We have tested the method on the relaxation test proposed by Boley et al. (2007a), which the method does not satisfy. This test will be discussed more thoroughly below; however, for the purposes of discussing the flawed surface area estimate of the Mayer et al. (2007) method, it suffices to say that the test consists of a column of gas, periodic in x and y, whose density structure adjusts in z in order to come to an (analytically derived) equilibrium in which internal heating is balanced by the radiative cooling of the column from two surfaces: one above and one below the midplane. This test has the useful feature that the total radiating area of the system is known: it is twice the x-extent times the y-extent.

We have found that the Mayer et al. (2007) boundary condition cannot satisfy the relaxation test because it generally over-estimates the total surface area of the system by a considerable amount. For this boundary condition, the total surface area is a decreasing function of the EDA used. For an EDA of 40° (consistent with a comparison of the simulation in §3.5.1 to that of Mayer et al. (2007)), we have found that the surface area is generally overestimated by a factor of approximately 6. Consequently, the system cools at a rate which is far higher than is expected. Furthermore, the total surface area of the system as calculated by the method has been found to vary by over 30% during dynamic evolution of the system. These variations of the effective emitting area from the expected value can have important effects on the cooling of gas, and consequently the outcome of simulations studying GI.

The smoothing length of a SPH particle is a measure of the inter-particle separa-

tion; however, using this quantity to estimate the local surface area of the photosphere is problematic for a number of reasons. For a particle on the edge of a system, the smoothing length will generally be over-estimated, since there will only be neighbours in the half of the search sphere interior to the edge. For a system of uniform density, this over-estimate will be a factor of $2^{1/3}$; however, this is decreased if the system has a strong density gradient towards the surface (as in a protostellar disc). Beyond this complication, the smoothing length is a quantity that evolves dynamically as the density distribution of the system changes. As discussed above with regards to the Boley et al. (2007a) test, this dynamical dependence has been shown to result in large changes of the total emitting area, even in a system where the total area is a well-defined constant.

We present our new photosphere boundary condition, which is based on the edge-particle idea of Mayer et al. (2007). Our method, however, is able to calculate surface areas that are consistent with expected values to within a few percent, even in dynamical systems. Moreover, our method is geometry independent and has no free parameters to adjust.

In order to determine which particles reside on the edge, we first determine individual geometry-independent outwards directions using $\nabla\rho_i$. Each particle then searches through its list of neighbour-particles to determine if any of these reside within an angle θ of this direction— it should be noted that we do not treat θ as a free-parameter, but use a fixed value of $\theta \approx 49°$. If no neighbour-particles are found within this angle, then the particle is deemed to be an edge-particle (this process is demonstrated in Figure 3.6 (left)). In order to remove any dependence that this procedure may have on particular neighbour-lists, we iterate, removing any non-edge particles from the search tree each iteration. Removing non-edge particles results in larger search-radii for the remaining edge-particles, which in turn results in a better measure of which particles are on the edge (this also aids finding edge-particle neighbours, as discussed in the next paragraph). We stop iterating once the number of edge-particles has converged.

Once the edge-particles are determined, we compute the surface area of the photosphere associated with each edge-particle. For each edge-particle, we find a fixed

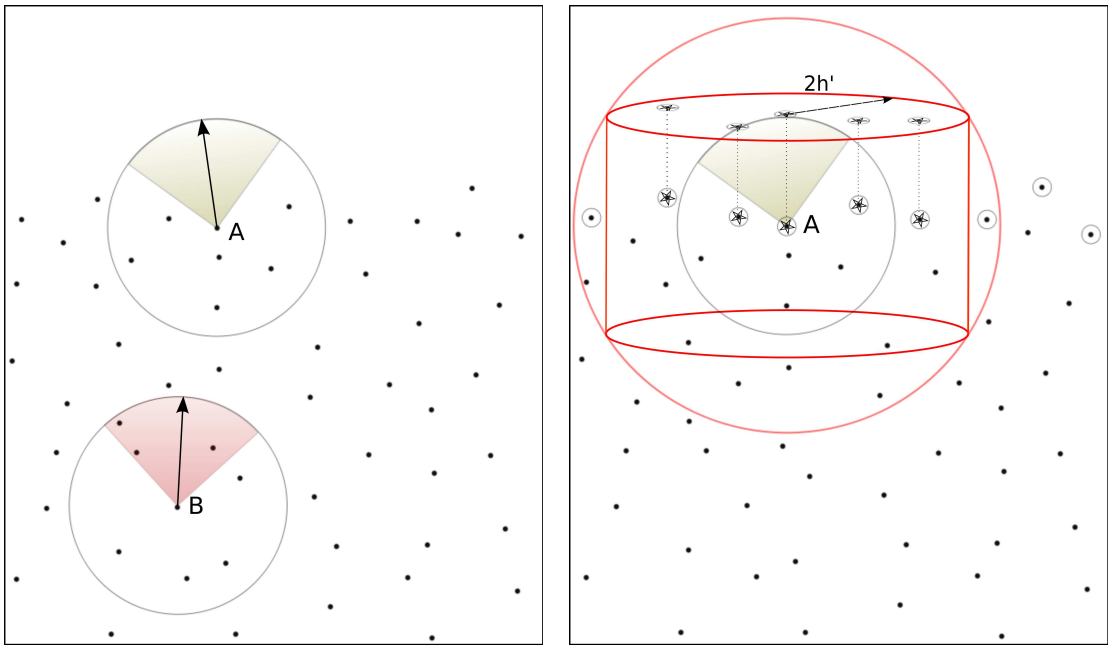

Figure 3.6: Determining the edge-particles and their emitting areas. In the left panel, particle A examines within its search-radius (circle) and finds that no particles reside within θ (the green region) of the direction $\nabla\rho_A$ (the arrow); therefore, particle A is an edge-particle. This condition is not satistfied for particle B. In the right panel, the edge-particle A searches within its search-radius (red circle) for other edge-particles (circles with points). Only edge-particles that are within the search-cylinder (circles with stars within the red cylinder) are projected onto the 2-D surface for the 2-D SPH column density estimate (using smoothing length $h' = \sqrt{2/3}h$).

number of edge-particle neighbours using a spherical search. We then project these edge-particles onto a 2-D surface, whose normal is a local average of $\nabla\rho$, and use a 2-D SPH estimate of the local column-density Σ_i to compute the local emitting area:

$$A_i = \frac{m_i}{\Sigma_i}. \tag{3.41}$$

In reality, since the surface area estimate relies on a 2-D SPH sum (for the *column* density), we do not actually include all of the edge-particle neighbours that are found in the search radius, but rather ignore those edge-particle neighbours that lie outside of a search-cylinder bounded by the search-sphere. This eliminates biased weighting in the inner region of the 2-D sum that can result from a spherical search in an envi-

ronment with a strong density gradient. We choose the search-cylinder of maximum volume that is bounded by the search-sphere. Figure 3.6 (right) demonstrates this process.

By using this 2-D SPH estimate in the calculation of the local photosphere area, rather than the smoothing length, we are able to calculate consistent surface areas even in dynamical situations. While we found that the Mayer et al. (2007) estimate varied by 30% in the relaxation test, the 2-D SPH estimate gave a total surface area that was within 2% of the expected value. Figure 3.7 demonstrates that this algorithm accurately computes the surface area of a disc. The surface area of edge-particles (binned) for an axisymmetric protostellar disc initial-condition is given by the diamonds and is compared to the surface area of a cylinder (since the disc is very flat, with an opening angle of $\approx 3°$, this is a very good approximation). While the total surface area estimate of the Mayer et al. (2007) boundary treatment is a function of the EDA, we note that the use of the 2-D SPH estimate results in a total surface area that is not sensitive to our choice of $\theta \approx 49°$; the total area of the disc in Figure 3.7 was found to vary by only 0.5 % for choices of θ in the range 40-60°.

Finally, in order to include the boundary cooling from each edge-particle, we simply include an additional term of

$$-\Delta t \frac{A_i}{m_i} \sigma \left(\frac{T}{0.81} \right)^4 = -\Delta t \frac{A_i}{m_i} \left(0.576 c \rho_i \xi_i \right) \tag{3.42}$$

in the particle's implicit specific radiation energy equation (3.18) and solve the system of equations as described previously— the factor of (1/0.81) reflects the fact that the edge-particles emit from $\tau = 0$ and must be adjusted to have the correct effective temperature, analogous to the solution for a plane-parallel atmosphere (Chandrasekhar 1960).

3.4.2 Relaxation test

We present the first SPH results for the relaxation test proposed by Boley et al. (2007a), which asseses the combination of FLD and radiative boundary cooling. This test is a fundamental one for codes that simulate GI in protostellar discs, since it

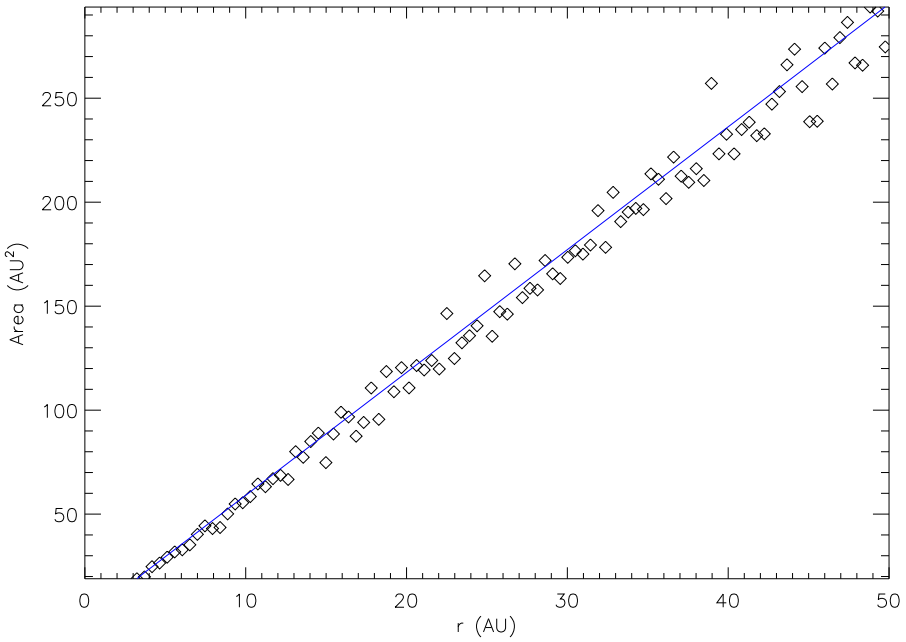

Figure 3.7: Testing the surface area estimates of edge-particles for a protostellar disc initial condition. The sum of the surface area of edge-particles within radial bins is given by the diamonds, while the surface area of a cylinder is given by the blue, solid line.

asseses the accuracy of radiative boundary cooling. This test consists of a 1D configuration of gas (our test is 3D with periodic boundaries in x and y) of constant opacity under a constant vertical (z) gravitational acceleration towards the midplane. The gas is initially in an arbitrary configuration and heated at a constant rate per mass

$$\Gamma = \frac{\sigma T_e^4 \kappa}{\tau_m}, \tag{3.43}$$

where T_e is the effective temperature and τ_m is the midplane optical depth. The gas is then allowed to relax to an equilibrium in which the bulk heating is balanced by cooling from the photosphere. The resulting temperature profile as a function of the optical depth is

$$T^4(\tau) = \frac{3}{4} T_e^4 \left[\tau \left(1 - \frac{\tau}{2\tau_m} \right) + \frac{1}{\sqrt{3}} \right]. \tag{3.44}$$

Figure 3.8 shows the results of the implicit non-linear method for $T_e = 100$ K,

and optical depths of $\tau_m = 100$ (left panel) and $\tau_m = 5$ (right panel). The gas temperature is given by the black points and the analytic result by the red, dashed line. The agreement is quite good, although there is a deviation of a few degrees for each case. This is the result of an underestimate of the emitting area by the edge-particles of a few percent, which causes the temperature to increase in order to emit the proper flux.

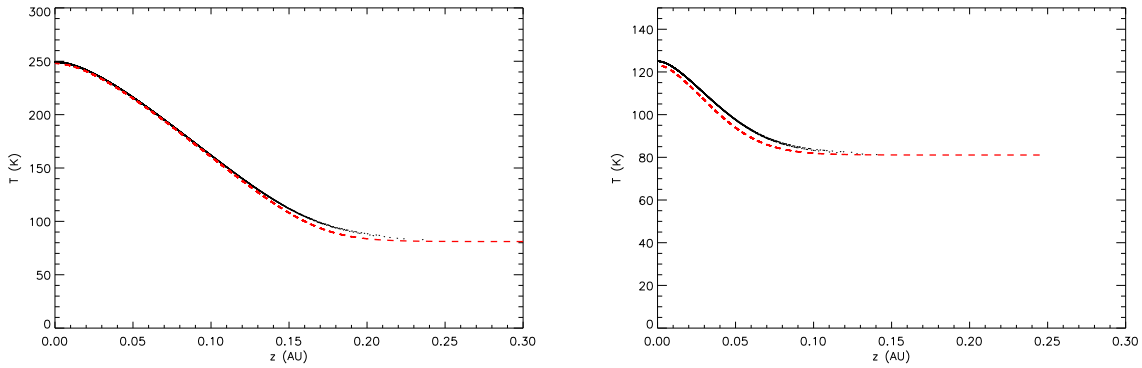

Figure 3.8: Testing the coupling of FLD and the boundary cooling for the implicit non-linear method using the relaxation test of Boley et al. (2007a). The bulk-heating of the gas is balanced by radiative cooling from the surface for $T_e = 100$ K and $\tau_m = 100$ (left) and $\tau_m = 5$ (right). The gas temperature is given by the black points, while the analytic solution, equation (3.44), is given by the red, dashed line. Only a subset of the particles is plotted.

3.5 Simulations of GI in the inner regions of discs

We use the combined FLD and photosphere boundary condition described above to investigate the viability of GI as a mechanism for giant planet formation in the inner tens of AU in protostellar discs. We have performed two sets of simulations for protostellar discs that evolve towards instability in a slow manner, and are consequently robust tests of the viability of fragmentation.

Mayer et al. (2007) carried out simulations with radiative transfer of a protostellar disc brought to instability through mass-gain and found that fragmentation was pos-

sible, though not certain, for disc masses in excess of 0.12 $M_\odot$ provided that the disc was of sufficient metallicity ($\mu \geq 2.4$). Our first case (Simulation A) is a rerun of a simulation of a particular disc with $\mu = 2.7$ and EDA = 40° that Mayer et al. (2007) found to fragment. We use parameters and initial conditions provided by the authors and find that, in contrast to their results, fragmentation does not occur in this disc. The fundamental difference between the simulation presented here, and that of Mayer et al. (2007) is an improved treatment of the photosphere boundary condition. This demonstrates the fundamental importance of using accurate photosphere treatments, such as that presented in §3.4, in simulations investigating GI.

Our second case (Simulation B) is of a 0.1 $M_\odot$ disc surrounding a 0.5 $M_\odot$ star that slowly transitions to instability through the growth of density perturbations. This disc is similar to one used by the Indiana University group (Mejía et al. 2005), but with a different density distribution in the outer regions. It is our goal to see whether fragmentation is possible in the inner regions of discs, and as discussed below, this change in initial condition is made to make our simulation even more likely to show fragmentation than those of the Indiana University group. However, as in the case of Simulation A, we find that although the disc becomes graviationally unstable, an accurate treatment of the photosphere results in cooling that is too slow to allow fragmentation.

Together, these simulations demonstrate that GI may not be a viable mechanism of planet formation in the inner tens of AU of protostellar discs. Though the simulated discs are gravitationally unstable, when an accurate photosphere treatment is taken into account, they are simply not able to cool fast enough for fragmentation to occur.

3.5.1 Simulation A: Mayer et al. (2007) Disc

Mayer et al. (2007) performed simulations with radiative transfer of a protostellar disc approximately 30 AU in radial extent surrounding a 1 $M_\odot$ star. As described by the authors, the SPH particles making up the system increased in mass over time. In this way, the disc transitioned from a state of stability to a state of instability in a smooth fashion. The authors then carried out a number of simulations beginning

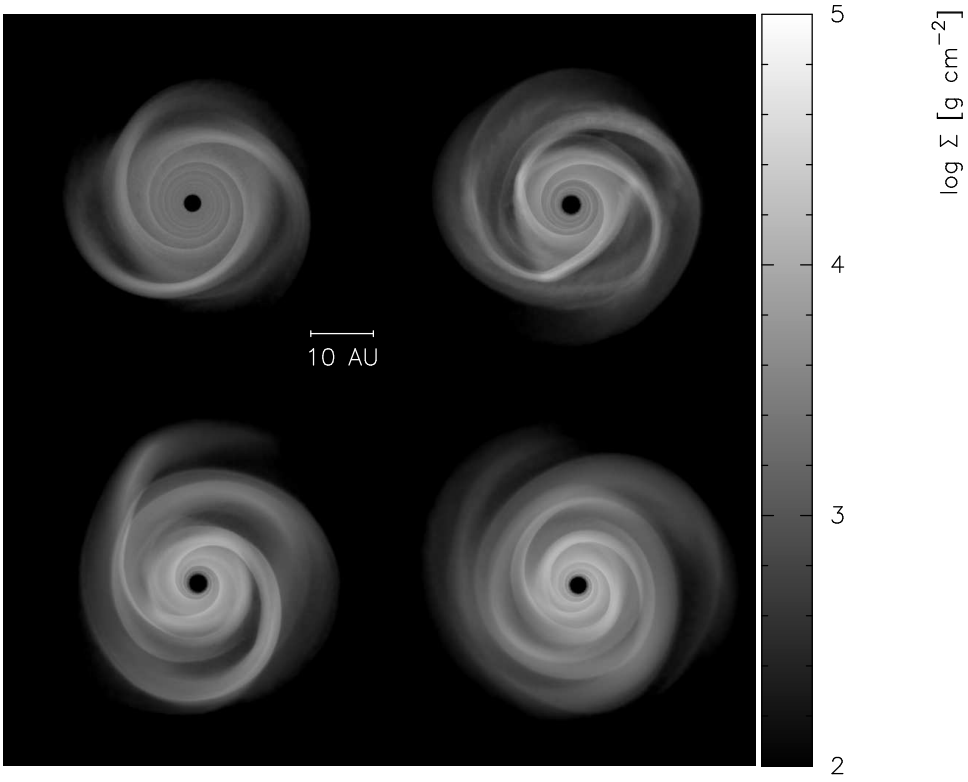

Figure 3.9: Simulation A: rerunning the Mayer et al. (2007) $\mu = 2.7$ simulation. Logarithmic surface density maps for four times during the simulation: $t = 1112$ (top left), 1217 (top right), 1322 (bottom left), and 1427 (bottom right) years). All surface density maps of discs in this paper were created using SPLASH (Price 2007).

from the disc as of about 10^3 years, and varied the opacity and mean molecular weight searching for conditions that allowed fragmentation. They observed fragmentation in a particular simulation with a mean molecular weight of $\mu = 2.7$ and EDA $= 40°$ (see Figure 1, bottom-right in Mayer et al. (2007)).

We rerun this simulation using the initial conditions (consisting of 10^6 SPH particles) as provided by the authors, beginning at a time of about 10^3 years. We use the same equation of state, an ideal gas with $\gamma = 7/5$, mean molecular weight, $\mu = 2.7$, and dust opacities (D'Alessio, Calvet, & Hartmann (1997)). Our implementation of FLD is different, but should result in the same physical outcome. The fundamental difference between the simulation presented here, and that of Mayer et al. (2007) is the improved treatment of the photosphere boundary condition. Radiation from

the central star is not considered. The Jeans-mass and disc scale height are resolved everywhere.

Logarithmic density maps for the disc are given for four times (t = 1112, 1217, 1322, and 1427 years) in Figure 3.9, and values for the midplane number density, midplane temperature, Toomre Q, and dimensionless cooling time ($t_{cool}\Omega$) for these times are given in Figures 3.10 and 3.11.

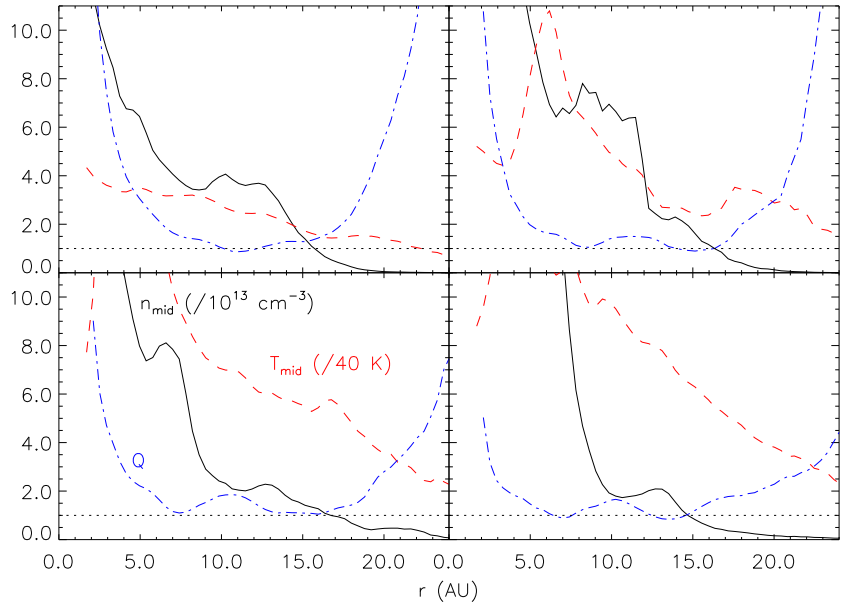

Figure 3.10: Simulation A: azimuthally averaged physical quantities of the disc, for the four times shown in Figure 3.9. The midplane number density (in units of 10^{13} cm^{-3}) is given by the solid, black line, the midplane temperature (in units of 40 K) is given by the red, dashed line, and the Toomre Q is given by the blue, dot-dashed line. The horizontal dotted line is a reference for Q = 1.

What is immediately apparent from Figure 3.9 is that, in contrast to Mayer et al. (2007), the disc does not undergo fragmentation (compare to Figure 1 (bottom right) of Mayer et al. (2007)). The disc exhibits strong spiral density enhancements, and regions of $Q \approx 1$, however, this does not lead to fragmentation.

We can understand why the disc does not fragment despite having regions of $Q \approx 1$ by considering the cooling time. For gas of $\gamma = 7/5$, numerical simulations

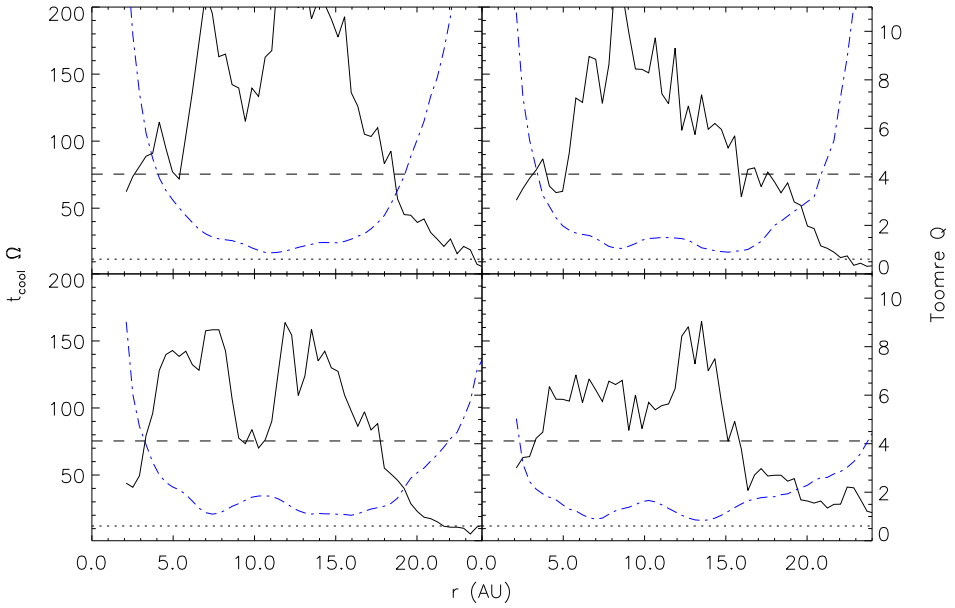

Figure 3.11: Simulation A: azimuthally averaged dimensionless cooling time (solid, black line) and Toomre Q (blue, dot-dashed line) of the disc, for the four times shown in Figure 3.9. The horizontal dotted line represents $t_{cool}\Omega = 12$, the critical value for fragmentation for $\gamma = 7/5$ (Rice et al. 2005), while the horizontal line represents a cooling time six times larger, an estimate of the critical value for the Mayer et al. (2007) boundary condition.

have shown that the critical value for the dimensionless cooling time is $\eta = 12$ (Rice, Lodato, & Armitage 2005). As shown in Figure 3.11, regions of low Toomre Q have dimensionless cooling times that are 6-20 times greater than this critical value. Therefore, although the disc is unstable, it is simply not able to cool fast enough for the instability to develop into fragmentation.

The inability of the disc to fragment can also be observed in the regulation of the Toomre Q to values near unity. The mass of the disc is an increasing function of time (due to the mass growth of the SPH particles) and acts to decrease the Toomre Q and drive the disc to instability. However, it is clear from Figure 3.11 that as the mass of the disc increases, the temperature of the disc increases in a corresponding fashion (due to shock heating) so that the Toomre Q does not significantly change,

with a minimum value that remains constant near $Q \approx 1$.

It is quite understandable that Mayer et al. (2007) observed fragmentation in their simulation, given the flaws observed in the photosphere boundary condition, as discussed in §3.4. At a time of 1217 years, our simulation shows short-lived density enhancements created in the spiral arms near 10 AU (one is visible to the right of disc-centre in Figure 3.9 (top right)). These enhancements occur at a similar time and radius as the fragments observed in Mayer et al. (2007). Given the over-estimate in photosphere surface area measured in their boundary condition, as well as the likely deviations of this surface area due to dynamical effects, it is quite conceivable that the local cooling time was less than the critical value needed for these density enhancements to become gravitationally bound fragments. We estimate a modified critical value of $\eta = 72$ that would take into account the measured over-estimate for the Mayer et al. (2007) photosphere and plot this value in Figure 3.11.

The results of this comparison reinforce the fundamental importance of the cooling rate, and accurate photosphere boundary conditions in simulations of GI in protostellar discs. The fact that this massive, centrally concentrated disc (at the end of our simulation, the disc contains 0.34 $M_\odot$ within 30 AU, significantly more than the 0.15 $M_\odot$ requirement of Mayer et al. (2007) for fragmentation) does not fragment, despite being unstable, points to the difficulty of forming giant planets in the inner regions of discs via fragmentation.

3.5.2 Simulation B: Mejía et al. (2005) Disc

We present a second simulation, in which we investigate the viability of fragmentation due to GI in a 0.11 $M_\odot$ protostellar disc surrounding a 0.5 $M_\odot$ star. The initial disc-profile is designed to be similar to that used in studies of GI by the Indiana University group (see Figure 1 of Mejía et al. (2005); Cai et al. (2006), Boley et al. (2006), Boley et al. (2007a)). The key difference is that our initial disc contains a significant amount of mass (0.04 $M_\odot$) beyond the region of low Q. We have made this change in order to avoid the "burst" phase seen in these studies, as we wish to consider a very "quiet" transition from the initial model to GI via the slow growth of perturbations caused

by the Poisson noise in the initial density distribution. As can be observed in Figure 3 of Boley et al. (2006), the burst phase radially redistributes a significant amount of mass, with the final radial extent of the disc nearly double the initial extent of 40 AU. This radial-redistribution moves material away from the radius at which the disc is initially most unstable, acting to stabilize the disc by increasing the Toomre Q. Thus, our disc, in which this significant radial-redistribution of mass does not take place, is a more stringent test of the likelihood of fragmentation.

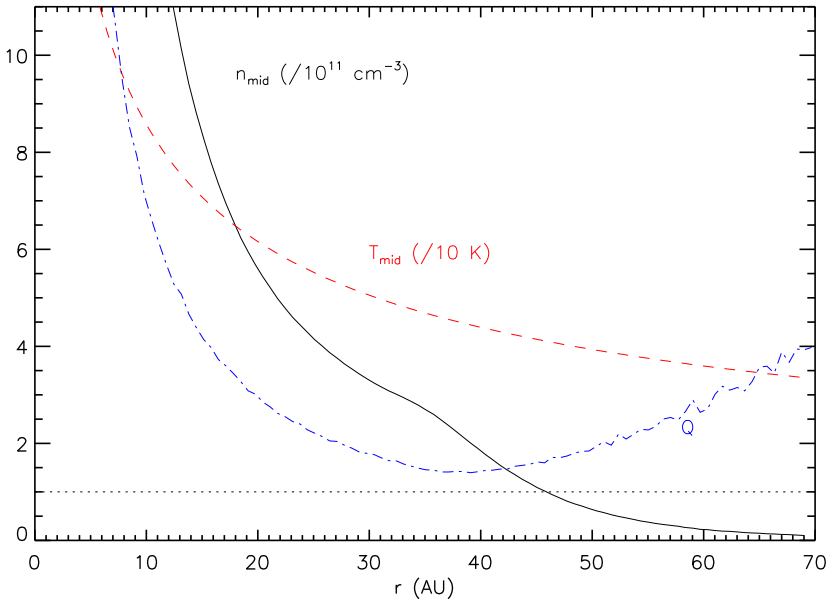

Figure 3.12: Simulation B: azimuthally averaged physical quantities of the initial disc model. The midplane number density (in units of 10^{11} cm^{-3}) is given by the solid, black line, the midplane temperature (in units of 10 K) is given by the red, dashed line, and the Toomre Q is given by the blue, dot-dashed line. The horizontal dotted line is a reference for Q = 1.

The initial, axisymmetric, model of 2.5×10^5 SPH particles was created in a manner similar to that of Shen et al. (2010)– see Figure 3.12 for the disc properties. The surface density profile is of the form $\Sigma \propto r^{-1/2}$ in the region between 3 and 34 AU, and $\Sigma \propto r^{-1/2}\left((r/r_m)^{-12} - 3(r/r_m)^{-8} + 3(r/r_m)^{-4}\right)$ in the region between 34 and 70 AU, with $r_m = 34$ AU. This profile results in a mass of 0.07 M$_\odot$ within 40 AU (the

same mass within the extent of the Mejía et al. (2005) disc), and an additional 0.04 $M_\odot$ beyond 40 AU. This additional 0.04 $M_\odot$ prevents a significant radial redistribution of matter during the simulation*.

The initial temperature profile is a power law: $T \propto r^{-1/2}$. Using the specified radial surface density and temperature profiles, the vertical structure satisfying hydrostatic equilibrium was calculated in an iterative manner, with both the gravity of the star and the self-gravity of the disc taken into account. The initial disc has a Q profile– using midplane values and $\kappa_e = \Omega$– that is a minimum of about 1.5 at a radius of 39.5 AU. We define one outer rotation period (ORP) as the Keplerian period at this radius, approximately 250 years.

The central star is modelled as a 0.5 M$\odot$ star particle with a gravitational softening of 3 AU; we do not include radiation emitted from the star. We use a mean-molecular weight of 2.3, realistic Rosseland-mean opacities (D'Alessio, Calvet, & Hartmann (1997); with a maximum grain size of 1 mm), as well as a consistent treatment of the internal energy of molecular hydrogen that takes into account translational, rotational, and vibrational modes, assuming an equilibrium ortho/para mixture (Boley et al. 2007b). The Jeans-mass and the disc scale height are resolved everywhere.

The disc is initially axisymmetric, but photospheric cooling gradually decreases Q to about 1 by 2 ORPs and the disc develops noticeable spiral density enhancements. By design, the growth of non-axisymmetric structure in the disc is a slow process. We observe that after about 9 ORPs, there is little variation in the azimuthally averaged midplane density, temperature, and Q in the inner 70 AU, demonstrating that the disc has reached a steady-state in which heating from GI is balanced by cooling from the photosphere.

We can once again understand why the disc does not fragment despite having regions of $Q \approx 1$ by considering the cooling time. This simulation's treatment for the internal energy of molecular hydrogens gives a varying adiabatic exponent in the

**Additional comment for thesis*: the power-law form of the surface density, $\Sigma \propto r^{-1/2}$, matches the profile used by Mejía et al. (2005). The power-law index of 0.5 is consistent with interferometric observations and SED fitting (see Chapter 2.2.2 and Williams & Cieza 2011). The profile outside of 34 AU was chosen to give a smooth falloff in mass outside of the region of minimum Q, in order to avoid the "burst" phase observed by Mejía et al. (2005).

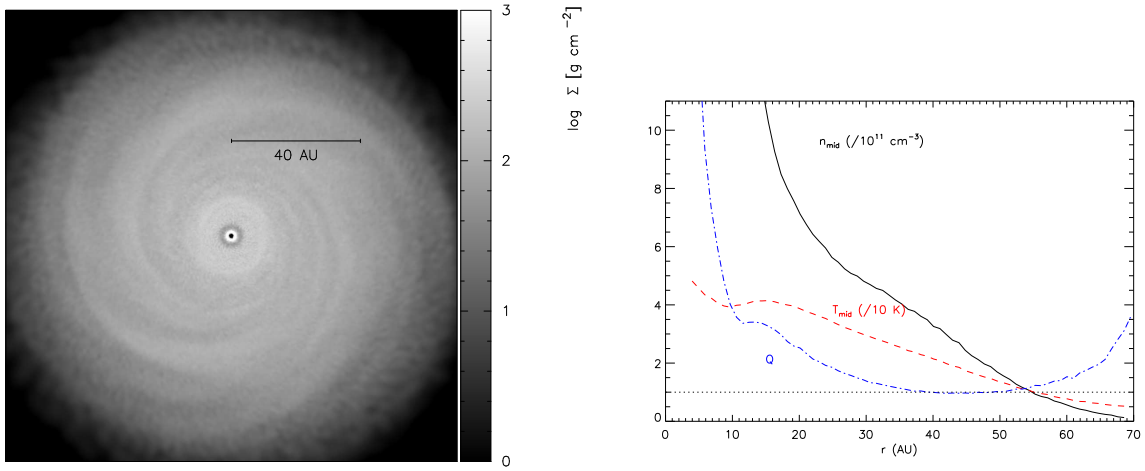

Figure 3.13: Simulation B: the disc after 10 ORPs. The left panel is of the disc's surface density, while the right panel gives azimuthally averaged properties of the disc (see Figure 3.12 for details).

range 1.36-1.8, which prevents a direct comparison of the cooling time to specific values of η in the literature. However, the cooling time of the disc in the region of $Q \approx 1$ (35-55 AU) is much larger than even large values of η (for $\gamma = 7/5$, $\eta = 13$) as can be seen in Figure 3.14.

3.6 Conclusions

We have presented a new implementation of RT in the FLD approximation for the TreeSPH code Gasoline, and have shown that it performs well on a number of standard tests.

In addition, we have demonstrated that previous photosphere boundary conditions using the smoothing lengths of edge-particles overestimate photosphere cooling and are susceptible to large variations in cooling rates due to dynamical effects. We have presented a new photosphere boundary condition using 2-D SPH sums over edge-particles that is able to calculate photosphere areas accurately even in dynamical systems. This boundary condition, coupled to FLD, is able to satisfy the relaxation test proposed by Boley et al. (2007a), an important test for codes that simulate protostellar discs; this is the first time this test has been performed with SPH.

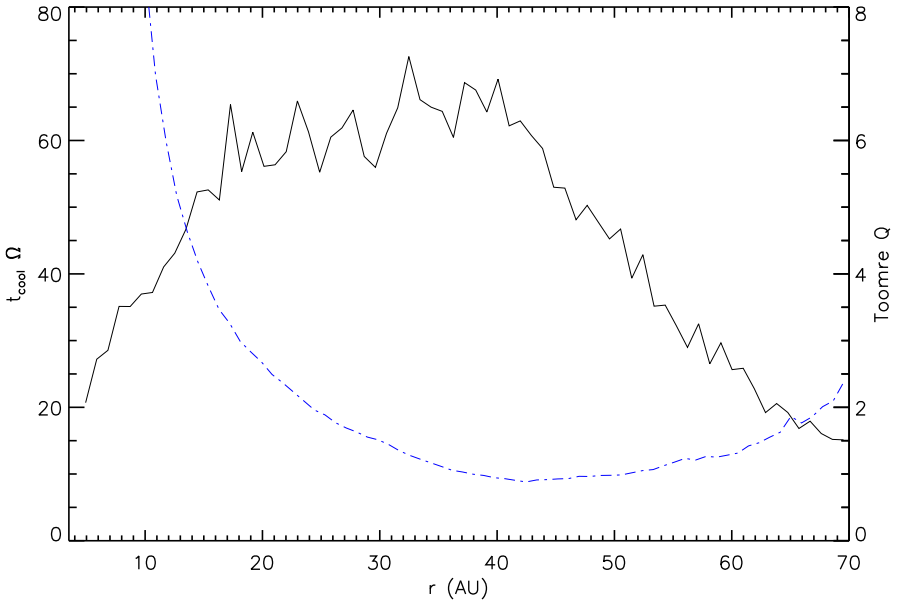

Figure 3.14: Simulation B: azimuthally averaged dimensionless cooling time (solid, black line) and Toomre Q (blue, dot-dashed line) of the disc after 10 ORPs. The cooling time in the region of $Q \approx 1$ (35-55 AU) is much larger than what one would expect would allow fragmentation.

In order to test the viability of fragmentation via GI as a mechanism for giant planet formation, we have performed two simulations, using the coupled FLD and photosphere boundary condition, of protostellar discs that are gravitationally unstable in the inner tens of AU. We have found that although the discs develop spiral density enhancements and have Toomre Q ≈ 1, they do not fragment because they are not able to cool fast enough.

One of our simulations, Simulation A, is a rerun of a previously published simulation in which the authors observed fragmentation (Mayer et al. 2007). Our simulation does not show fragmentation, and we attribute this difference to our new, more accurate, treatment of photosphere cooling. This result emphasizes the fundamental importance of using accurate photosphere treatments in simulations of GI in protostellar discs.

These numerical results, and others (Boley et al. (2007a), Forgan et al. (2009), Cai

et al. (2010)), as well as analytic studies (Rafikov 2005, 2007), suggest that GI may be an unlikely mechanism for planet formation in the inner tens of AU of protostellar discs. However, disc fragmentation is a complex process, with important roles played by GI, photosphere cooling, disc irradiation (Cai et al. 2008), mass accretion (Rafikov 2009), and dust evolution (Podolak et al. 2011). Further studies exploring a larger range of parameters and including all of the relevant physics are needed before the viability of GI as a mechanism for planet formation in the inner regions can be fully addressed. Nevertheless, analytic work (Nero & Bjorkman 2009) and numerical work (Boley 2009) have demonstrated the viability of GI as a mechanism of giant planet formation in the outer regions of protostellar discs. This is particularly interesting given the recent direct imaging observations (Kalas et al. (2008), Marois et al. (2008)) that have shown extrasolar planets to exist at such large radii. In future work, we will further investigate the detailed viability of GI as a planet formation mechanism.

3.7 Acknowledgements

We would like to thank the anonymous referee for comments that have greatly improved the quality of this manuscript. We would also like to thank Lucio Mayer and Thomas Quinn for useful discussions. This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC). Computations were carried out using the facilities of the Shared Hierarchical Research Computing Network (SHARCNET), which is located in Ontario, Canada.

Bibliography

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

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

Boley A. C., Durisen R. H., Nordlund Å., Lord J., 2007a, apj, 665, 1254

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

Boley A. C., Mejía A. C., Durisen R. H., Cai K., Pickett M. K., D'Alessio P., 2006, ApJ, 651, 517

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

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

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

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., Boley A. C., Mejía A. C., Pickett M. K., D'Alessio P., 2006, ApjL, 636, L149

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

Chandrasekhar S., 1960, Radiative transfer, Chandrasekhar, S., ed.

Cleary P. W., Monaghan J. J., 1999, Journal of Computational Physics, 148, 227

D'Alessio P., Calvet N., Hartmann L., 1997, ApJ, 474, 397

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

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

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

Kalas P. et al., 2008, Science, 322, 1345

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

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

Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348

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

Mayer L., Quinn T., Wadsley J., Stadel J., 2004, ApJ, 609, 1045

Mejía A. C., Durisen R. H., Pickett M. K., Cai K., 2005, ApJ, 619, 1098

Meru F., Bate M. R., 2010a, MNRAS, 406, 2279

Meru F., Bate M. R., 2010b, MNRAS, L174+

Mihalas D., Mihalas B. W., 1984, Foundations of radiation hydrodynamics, Mihalas, D. & Mihalas, B. W., ed.

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

Nero D., Bjorkman J. E., 2009, ApjL, 702, L163

Podolak M., Mayer L., Quinn T., 2011, ApJ, 734, 56

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, Numerical Recipes: The Art of Scientific Computing (Third Edition)

Price D., 2005, ArXiv Astrophysics e-prints: 0507472

Price D. J., 2007, PASA, 24, 159

Rafikov R. R., 2005, ApjL, 621, L69

Rafikov R. R., 2007, ApJ, 662, 642

Rafikov R. R., 2009, ApJ, 704, 281

Rice W. K. M., Lodato G., Armitage P. J., 2005, MNRAS, 364, L56

Shen S., Wadsley J., Hayfield T., Ellens N., 2010, MNRAS, 401, 727

Sincell M. W., Gehmeyr M., Mihalas D., 1999, Shock Waves, 9, 391

Toomre A., 1964, ApJ, 139, 1217

Turner N. J., Stone J. M., 2001, ApJS, 135, 95

Wadsley J. W., Stadel J., Quinn T., 2004, New A., 9, 137

Whitehouse S. C., Bate M. R., 2004, MNRAS, 353, 1078

Whitehouse S. C., Bate M. R., Monaghan J. J., 2005, MNRAS, 364, 1367

Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67

Zel'dovich Y. B., Raizer Y. P., 1967, Physics of shock waves and high-temperature hydrodynamic phenomena, Zel'dovich, Y. B. & Raizer, Y. P., ed.

Chapter 4

The fragmentation of protostellar discs: the Hill criterion for spiral arms

4.1 Introduction

The fragmentation of protostellar discs through gravitational instability (GI) is a possible mechanism for the formation of gas-giant planets and brown dwarfs. For a disc to be prone to fragmentation, there are generally thought to be two criteria that must be satisfied.

The first criterion is that a disc must be gravitationally unstable. This can be characterized by the Toomre Q parameter (Toomre 1964), which is the result of a linear stability analysis for a differentially rotating thin disc:

$$Q = \frac{c_s \kappa_e}{\pi G \Sigma}. \tag{4.1}$$

In the above, c_s is the sound speed of the gas, κ_e is the epicyclic frequency ($\kappa_e = \Omega$, the rotation rate, for Keplerian rotation), G is the gravitational constant, and Σ is the surface density. For low values of $Q \sim 1$, such as would be found in a massive, cold disc, a disc will be gravitationally unstable.

The second criterion for fragmentation is that a disc, in addition to being gravitationally unstable, must also cool quickly. Gammie (2001), using 2-D shearing-box

simulations, examined the stability of a local patch of a protostellar disc with a simplified cooling prescription:

$$t_{\rm cool}\Omega = \beta, \tag{4.2}$$

where $t_{\rm cool}$ is the cooling time and β is a constant. The cooling time here is defined through the evolution of the specific internal gas energy, u:

$$\frac{Du}{Dt} = -\frac{u}{t_{\rm cool}}, \tag{4.3}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$ is the co-moving derivative.

By performing simulations with different values for β, the author found the cooling criterion for fragmentation to be

$$t_{\rm cool}\Omega \leq \beta_{\rm crit} = 3; \tag{4.4}$$

this particular value of $\beta_{\rm crit}$ is valid only for the specific 2-D power-law equation of state used by the author. Subsequent work, using global 3-D SPH simulations (Rice, Lodato, & Armitage 2005), as well as global 3-D grid-based simulations (Mejía et al. 2005), have also found fragmentation for fixed cooling times faster than a critical rate.

Gammie (2001), as well as Rice et al. (2005), outlined a physical argument for the existence of a critical cooling time. If GI can be well-characterized by an α-viscosity model (Shakura & Sunyaev (1973); see Lodato & Rice (2004, 2005) for the applicability of this), then a steady-state can exist if the viscous heating by GI is balanced by the prescribed cooling. If this balance is achieved, then the required viscosity of the disc is determined by

$$\alpha = \frac{1}{9/4\gamma_{\rm 2\text{-}D}\left(\gamma_{\rm 2\text{-}D} - 1\right)\beta}. \tag{4.5}$$

Here, $\gamma_{\rm 2\text{-}D}$ is the two-dimensional adiabatic index. As discussed by Gammie (2001), this can be related to the 3-D adiabatic index, γ, for the non-self-gravitating case, where $\gamma_{\rm 2\text{-}D} = (3\gamma - 1)(\gamma + 1)$, as well as for the strongly self-gravitating case, where $\gamma_{\rm 2\text{-}D} = 3 - 2/\gamma$. If GI has a maximum α that it can attain, then it also has a

maximum heating rate (Rice et al. 2005). If the prescribed β cools the disc faster than this maximum heating rate, then no balance between heating and cooling can be achieved, and the disc fragments.

The β-prescription of cooling, however, is a simplified model; more generally, β will evolve with α in a given disc. In a realistic disc, heating and cooling are linked to the physical state of the disc. In this work, we consider fragmentation in this more realistic case, whereas the β-prescription of cooling does not allow Gammie (2001) to have done so.

Furthermore, the applicability of a single value of $\beta_{\rm crit}$ (or $\alpha_{\rm max}$) is in some doubt. Rice et al. (2005) found that the value of $\beta_{\rm crit}$ depends on the equation of state of the gas (the value of γ), but that this dependence is consistent with a unique value of $\alpha_{\rm max} \sim 0.06$. However, other simulation results in the literature have found a non-unique value for the critical cooling time, even for a fixed equation of state. Clarke et al. (2007), using a time-dependent $\beta(t)$, found that the thermal history of the gas is important in determining $\beta_{\rm crit}$, and can result in its reduction by a factor of two. Meru & Bate (2011) found $\beta_{\rm crit}$ to be a function of the distance from the central star, the local surface density, and the stellar mass.

Complicating the applicability of a unique $\beta_{\rm crit}$ are questions of numerical convergence. Meru & Bate (2010) recently found that their 3-D SPH simulations had not converged numerically; they found larger values of $\beta_{\rm crit}$ with increasing resolution. A similar increase in the critical cooling time with enhanced resolution was found by Paardekooper et al. (2011), using 2-D grid-based simulations. However, these authors were able to achieve numerical convergence if they slowly decreased the cooling rate with time, which allowed the entire disc to be in a state of gravito-turbulence while gradually approaching the fragmentation boundary. Lodato & Clarke (2011) have proposed that the non-convergence seen by Meru & Bate (2010) may be the result of heating via artificial viscosity, or smoothing of density enhancements in simulations with insufficient resolution, while Rice et al. (2011) have proposed that the non-convergence is the result of using a particle's specific internal energy in the cooling rate, rather than using a smoothed version of this quantity.

It is also worth noting that when considering the stability of protostellar discs

using the cooling criterion (e.g. Rafikov 2005), it is important to consider a disc as it would be in the non-linear state of GI, rather than as it would be in the axisymmetric state. The radiative cooling time is a function of temperature, opacity, and surface density, all of which are different in spiral arms, where fragmentation takes place, as compared to an axisymmetric disc. For example, Johnson & Gammie (2003) found that cooling times in the non-linear phase of GI could differ substantially from the cooling time of the initial axisymmetric initial condition.

In this work, we present a set of 3-D radiation hydrodynamic simulations of a massive, optically-thick, protostellar disc, unstable near 100 AU, around an A star. Rather than using a β-prescription for the cooling, these simulations include realistic heating and cooling of the disc, including cooling from the disc photosphere and irradiation from the central star. We do, however, vary the cooling rate in this set of simulations by scaling the dust opacity table by a constant factor. By reducing the opacity (which reduces the cooling time for an optically thick disc) over this set, we observe a transition from discs that are stable against fragmentation to discs that do fragment; this is consistent with the cooling criterion work of Gammie (2001).

We have used results from Cossins, Lodato, & Clarke (2009), and from this set of simulations to develop a simple, yet detailed, physical model for the fragmentation of a gravitationally unstable protostellar disc. In this model, spiral arms develop in an unstable disc on a characteristic scale related to the disc scale height (Cossins et al. 2009). The heating rate per unit mass of the disc from GI is proportional to the square of the amplitude of the surface density variations in the disc (Cossins et al. 2009); as spiral arms become more condensed, the heating rate per unit mass is increased. The cooling rate per unit mass of the disc from photospheric cooling is inversely proportional to the square of the surface density; as spiral arms become more condensed, the cooling rate per unit mass in the arms decreases. There is therefore a natural scale for the thickness of a spiral arm in a gravitationally unstable disc. This scale is set by a balance between heating from spiral waves and radiative cooling. It is worth noting that for faster cooling rates (shorter cooling times), this thickness will be decreased.

A second scale of interest in this model is the Hill radius, which, for an object of

mass M, is

$$H_{\rm Hill} = \left[\frac{GM}{3\Omega^2}\right]^{1/3}. \tag{4.6}$$

The importance of the Hill radius can be understood within the context of planet formation in a disc of planetesimals around a star. If a protoplanet embryo has formed in this disc, then it of interest to determine the radius over which it may further accrete planetesimals. The Hill radius sets this embryo's sphere of influence: material within the Hill radius is bound to the embryo and will be accreted. In essence, material within the Hill radius of an object is dominated by that object's gravity, while material outside of the Hill radius is dominated by the central star's gravity, which is equivalent to the role of the local shear in the Toomre criterion. For the purpose of this discussion, we define the Hill thickness as twice the Hill radius.

In this framework, we can extend the cooling criterion of Gammie (2001) with the Hill criterion for spiral arms. *In a gravitationally unstable disc, the natural thickness of the spiral arms is set by a balance between heating and cooling. Fragmentation occurs in this disc if there is a section of arm whose natural thickness is smaller than that section's Hill thickness.* Essentially, if a section of a spiral arm lies within its own Hill thickness, then shear will be unable to prevent the collapse of the arm, and fragmentation can take place.

In a gaseous disc, pressure can prevent fragmentation from taking place. By considering the Hill radius, we have not addressed the role that pressure plays in determining fragmentation and how it may modify the critical thickness of spiral arms necessary for fragmentation to take place. The correct determination of this scale requires the solution to a stability analysis of a spiral arm in a differentially rotating system. Since we do not have such a solution, we have chosen to consider the Hill thickness. The analysis of our simulations; however, does indicate that the Hill thickness is the correct scale to examine. The Hill criterion for fragmentation is consequently an empirical criterion.

Our model is consistent with the cooling criterion: as the cooling time decreases, spiral arms become thinner and more over-dense, becoming more likely to reside within their own Hill thickness, and consequently more likely to fragment. With

the Hill criterion, however, we have developed a more detailed, and more complete, physical picture of fragmentation. This picture can be applied to the general case of a disc with physical heating and cooling, or the more specific case of a disc with β-prescription cooling. In fact, it offers a means to calculate what the critical cooling time is for a given region of a given disc.

There have been previous attempts to characterize the fragmentation of spiral arms using a physical model. Durisen, Hartquist, & Pickett (2008) examined the stability of a spiral arm, considered as an isothermal sheet, and found that fragmentation was expected for low values of Q; in addition, the authors argued that strong compression in a spiral shock acts to suppress fragmentation. This implies that fragmentation should occur near corotation.

Boley et al. (2010) proposed that fragmentation takes place when the Toomre length (the most unstable radial wavelength) in a spiral arm lies inside the region around corotation found by Durisen et al. (2008). The physical models of Durisen et al. (2008) and Boley et al. (2010) are strictly only applicable to isothermal discs; however, Boley et al. (2010) did find agreement between the fragment masses predicted in the model, and those found in a simulation with radiative cooling. In contrast, our model of fragmentation presented in this work is derived for discs with cooling.

The structure of the paper is as follows. In §4.2, we overview our numerical methods as well as our set of simulations of gravitationally unstable, irradiated protostellar discs. In section §4.3, we give a detailed picture of our model of protostellar disc fragmentation and the Hill criterion. In addition, we demonstrate the model's consistency with the simulations of §4.2. In §4.4, we show that the Hill criterion is quantitatively consistent with the cooling criterion and discuss the predictive qualities of the model. Finally, in §4.5 we give our conclusions.

4.2 Numerical simulations of gravitationally unstable irradiated discs

4.2.1 Numerical methods

Our simulations were performed with the TreeSPH code Gasoline (Wadsley, Stadel, & Quinn 2004), with the addition of radiative transfer in the flux-limited diffusion approximation [FLD; Rogers & Wadsley (2011)]. As described by the authors, FLD is able to model the transfer of energy only in regions in which SPH particles reside. Because of limited resolution, any SPH representation of a protostellar disc naturally has two edges, representing the upper and lower atmospheres. Radiative cooling from the disc atmospheres is modelled by means of a photosphere boundary condition: the SPH particles on the "edge" of the disc (the edge-particles) are found, robust surface areas (the area of the photosphere for which an edge-particle is responsible) are calculated using a 2-D SPH estimate, and a plane-parallel cooling term is added to the radiative energy equation for the edge-particles. The radiative hydrodynamics has been tested on a number of standard problems, including the relaxation test of Boley et al. (2007a), which is particularly suited to protostellar disc simulations.

The conditions in the outer regions of discs (roughly 100 AU and beyond) are expected to be favourable to gravitational fragmentation, since the cooling criterion is likely satisfied there (Rafikov 2007). As pointed out by Kratter & Murray-Clay (2011), the heating of the outer regions is expected to be dominated by the irradiation of the disc's surface by the central star, rather than by viscous heating. Since our simulations focus on fragmentation at these large radii, it is fundamentally important to account for this heating via irradiation.

The photosphere boundary condition of Rogers & Wadsley (2011) offers a straightforward means by which this can be done. In addition to the cooling term in the specific radiation energy equation for each edge-particle, we have added a heating term of

$$\left[\frac{D\xi_i}{Dt}\right]_{\rm irrad} = \frac{A_i}{m_i}\sigma\left(T_{\rm irrad}\right)^4, \tag{4.7}$$

where ξ is the specific radiation energy, A_i is the surface area of the edge-particle, m_i

is the particle mass, σ is the Stefan-Boltzmann constant, and $T_{\rm irrad}$ is the temperature of the stellar irradiation.

Kratter, Murray-Clay, & Youdin (2010b) used the passive flared disc model of Chiang & Goldreich (1997), along with a stellar evolution model, to determine the equilibrium temperature distribution for a disc surrounding a 1.35 $M_\odot$ A star, which they found to be:

$$T = 40 \text{ K} \left(\frac{R}{70 \text{ AU}} \right)^{-3/7}. \tag{4.8}$$

Since we are not able to treat the super-heated dust layer of optical depth $\tau < 1$ in our simulations, it is appropriate to use this equilibrium temperature distribution as the $T_{\rm irrad}$ in the irradiation heating term, equation (4.7). In addition, we implement a floor of $T_{\rm irrad} = 20$ K to take into account the background radiation field.

There have been previous multidimensional simulations of protostellar discs that have made efforts to account for the effects of stellar and/or envelope irradiation. These include 3-D simulations (Cai et al. 2008; Stamatellos & Whitworth 2008; Boley 2009), axisymmetric simulations (Zhu et al. 2009), and thin-disc simulations (Vorobyov & Basu 2010).

4.2.2 Initial conditions and input parameters

The initial, axisymmetric model of 5×10^5 SPH particles was created in a manner similar to that of Shen et al. (2010)– see Figure 4.1 for the disc properties. The surface density profile has the form $\Sigma \propto r^{-1}$ in the region of 20-70 AU (there is a smooth increase of Σ from 10-20 AU). There is a smooth functional form of $\Sigma \propto r^{-1} \exp\left\{-4 \log(R) \left[0.5 \log(R) - \log(R_m)\right] / \log(R_o/R_m)\right\}$, with $R_m = 70$ AU and $R_o = 160$ AU, from 70-160 AU, after which there is a steep drop off of $\Sigma \propto r^{-15}$. There is roughly 0.61 $M_\odot$ within 200 AU.

This particular surface density distribution is motivated by the initial temperature profile, which is given by the equilibrium temperature in equation (4.8). The combination of temperature and surface density results in a broad region of the disc having an initial Q of roughly unity.

Using the initial surface density and temperature profiles, the vertical structure

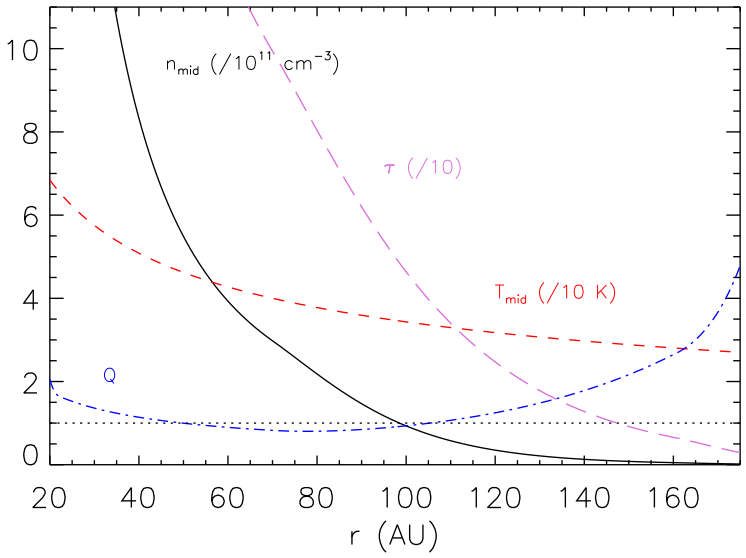

Figure 4.1: The physical quantities of the initial disc profile. The midplane number density (in units of 10^{11} cm^{-3}) is given by the solid, black line; the midplane temperature (in units of 10 K) is given by the red, dashed line; the optical depth (in units of 10) is given by the purple, long-dashed line; and the Q is given by the blue, dot-dashed line. The horizontal, dotted line is a reference for $Q = 1$.

satisfying hydrostatic equilibrium was calculated iteratively, taking into account both the gravity from the central star and the self-gravity of the disc. We have taken care to design this initial condition to ensure that it is "quiet," even though it is gravitationally unstable; that is, to ensure that there is a smooth evolution to spiral structure, without any strong radial redistribution of mass on short timescales (roughly the orbital period) due to transients from the initial setup. In this way, the azimuthally-averaged properties of the initial condition are reflective of the evolved disc.

The central star is modelled as a 1.35 $M_\odot$ sink particle with a radius of 10 AU. We use a mean molecular weight of 2.3 and realistic Rosseland-mean opacities (D'Alessio, Calvet, & Hartmann 1997). Although the code is capable of using a consistent treatment of the internal energy of molecular hydrogen that takes into account transla-

tional, rotational, and vibrational modes (Boley et al. 2007b), to simplify the analysis of our simulations, we use a fixed adiabatic index of $\gamma = 7/5$. In the outer regions of protostellar discs, of particular interest here, temperatures are typically below 100 K. Therefore, a choice of $\gamma = 5/3$ would be more physical; however, for the purposes of this study, which aims to understand the process of fragmentation in detail, the distinction is not important. The scale height is resolved by at least three smoothing lengths outside of 25 AU, and the Jeans length is resolved until shortly after fragmentation takes place. We define one ORP, one outer rotation period, to be the Keplerian period at our fiducial radius of 100 AU, roughly 863 years.

4.2.3 Simulations

We present a set of five simulations, each of which uses the initial conditions described above, the only difference being the opacity used. Simulation (A, B, C, D, E) has an opacity table that is scaled by a constant value of (1/10, 1/3, 1, 3, 10). Thus, Simulation C has an estimated physical opacity for solar metallicity, while Simulations A and B have reduced opacities, and Simulations D and E have increased opacities. Physical changes in opacity could be the result of grain growth (Birnstiel, Dullemond, & Brauer 2010), grain evolution via the passage of spiral arms (Podolak, Mayer, & Quinn 2011), or formation in an environment with a non-solar metallicity. Our goal, however, is not to reproduce different physical environments, but rather to explore the necessary conditions for gravitational fragmentation. In this context, a simple scaling of the opacity table is both sufficient and desirable.

All five simulations evolved in a similar fashion over the first 2.5 ORPs. High mode-number spiral structure developed slowly from SPH Poisson noise in each of the discs over this time until settling down to a transitioning state of two or three spiral arms. The transition from axisymmetric initial condition to spiral structure is observed to be smooth, with no strong transients.

The final states of the simulations are shown in Figure 4.2. Simulations C, D, and E have been evolved for roughly 8.5 ORPs without fragmentation having taken place (although strong spiral over-densities may persist), while Simulation A has

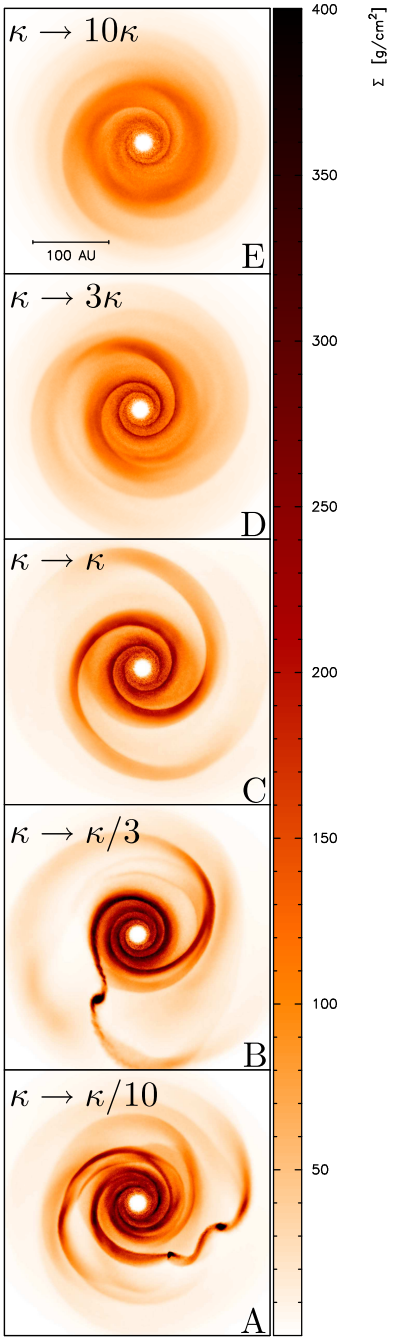

Figure 4.2: The final states of the simulated discs: surface density plots of Simulations E, D, C, B, and A are shown from top to bottom. As the opacity scaling is reduced from the physical value, fragmentation occurs. The discs are shown at respective times of 8.5, 8.5, 8.5, 3.5, and 2.9 ORPs.

fragmented with two objects forming, and Simulation B has fragmented with one object forming. This set of simulations, therefore, demonstrates a transition from non-fragmentation to fragmentation, as a function of the opacity scaling.

For a patch of an optically thick disc, the cooling time is approximately

$$t_{\rm cool} = \frac{1}{4}\frac{1}{\gamma - 1}\frac{c_s^2 \kappa}{\sigma T^4}\Sigma^2, \tag{4.9}$$

where κ is the opacity and T is the midplane temperature (Rafikov 2007). As can be observed in Figure 4.1, our initial condition is optically thick, with $\tau > 1$, out to at least 150 AU, even when the opacity is scaled by 1/10, as it is in Simulation A. This extends over the entire gravitationally unstable region (where $Q \lesssim 1.7$, Durisen et al. 2007). Hence, the cooling time is directly proportional to the opacity, and our set of simulations offers a means of exploring the fragmentation boundary as a function of cooling time in a manner similar to the simulations of Gammie (2001). The difference is that our simulations use realistic radiative cooling (even though the opacities may be scaled), rather than β-prescription cooling.

The cooling criterion suggests that reducing the cooling time (by reducing the opacity) will eventually lead to fragmentation. Thus, that Simulations A and B fragment is consistent with this picture. However, it is possible to use these simulations to better understand why exactly fragmentation takes place.

Figure 4.3 shows the five simulations at roughly the same time, shortly before fragmentation took place in Simulations A and B. The difference in the structure of the five discs at this time offers evidence of a detailed description for why Simulations A and B fragment, while Simulations C, D, and E do not. As can be observed, as the cooling time decreases (as the opacity decreases), spiral arms in a disc become thinner and more over-dense and this makes fragmentation more likely to occur.

4.3 Fragmentation model and the Hill criterion

We present a model of spiral arm fragmentation in gravitationally unstable discs, based on the observation from the simulations of the previous section that reduced

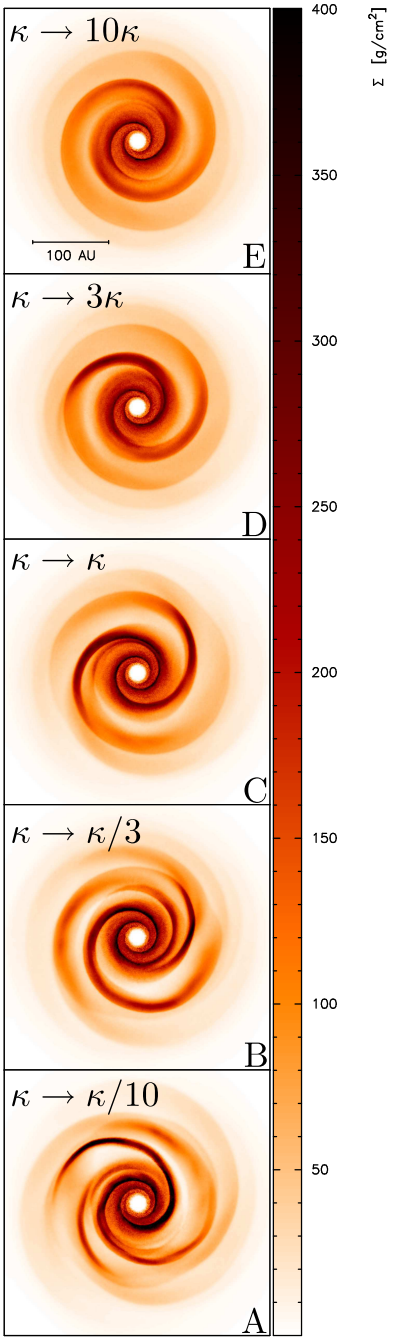

Figure 4.3: The simulated discs before fragmentation: surface density plots of Simulations E, D, C, B, and A are shown from top to bottom. As the opacity scaling is reduced, the spiral arms become thinner and more over-dense. The discs are shown at a time of 2.5 ORPs.

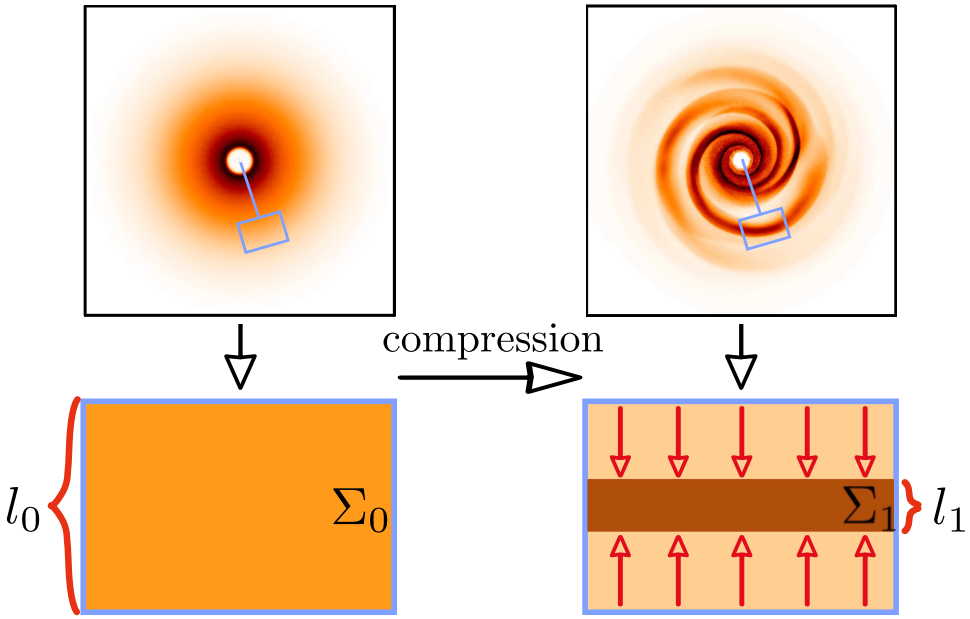

Figure 4.4: The creation of a spiral arm: a local patch of radial extent l_0 (left) in an axisymmetric disc collapses radially to form a section of a tightly-wound spiral arm (right) of radial extent l_1.

cooling times lead to thinner arms, which are more likely to fragment. This model can be broken into two components: the first is a model for the (roughly) steady-state spiral structure in an unstable disc; while the second is a criterion for the fragmentation of these spirals. Many of the details of our model are empirical in nature: we have used results from the simulations of Cossins et al. (2009), as well as our own set of simulations in determining some of the important parameters.

4.3.1 Spiral structure

We begin with a model for the spiral structure that results in a gravitationally unstable disc. We consider a patch of an initially axisymmetric disc that will develop spiral structure, such as our initial condition for the simulations of the preceding section. This patch is located at some distance, R, away from the central star and is of a radial extent l_0, with a characteristic surface density Σ_0. GI acts on the scale of l_0 to

collapse mass radially[1], resulting in the formation of a spiral arm of thickness l_1 and characteristic surface density Σ_1. This process is demonstrated in Figure 4.4.

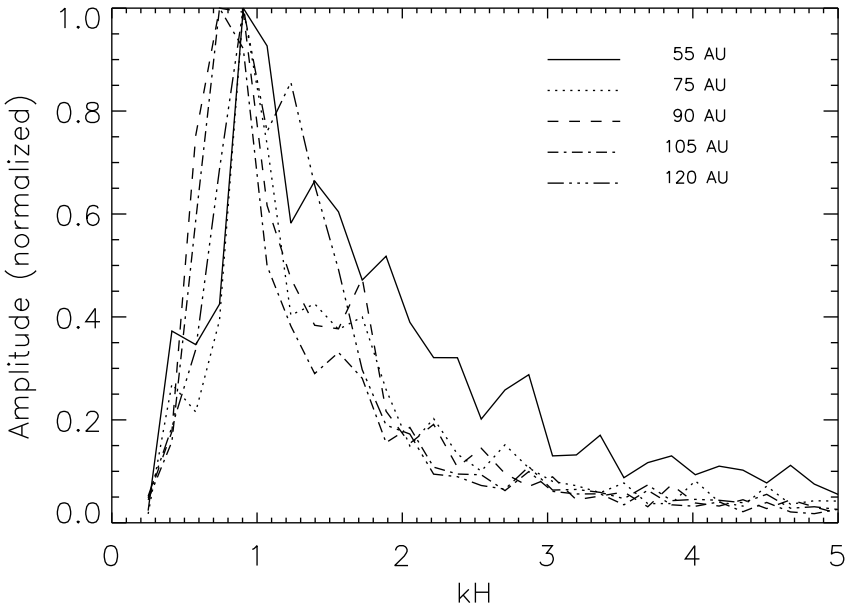

Figure 4.5: Radial Fourier analysis of Simulation B at 2.5 ORPs: the amplitude of the radial mode is given as a function of the wavenumber times the local scale height for a number of radii. Each amplitude curve is normalized by the peak amplitude for that radius. The Fourier analysis is carried out using the method of Cossins et al. (2009).

What is the appropriate scale for l_0? Cossins et al. (2009) performed a number of simulations of $Q \sim 1$ discs and found that, from a radial Fourier transform of these discs, the dominant radial wavenumber was typically

$$k \approx \frac{1}{H} = \frac{\pi G \Sigma}{c_s^2}. \tag{4.10}$$

Therefore, we expect the scale of spiral arm formation to be $l_0 = 2\pi H$. We have tested that this is consistent with our own simulations. Figure 4.5 shows the results of a radial Fourier transform of Simulation B at a time of 2.5 ORPs.

[1]The simulations of §4.2 show tightly-wound spiral structure with a typical winding angle of $i \sim 10°$; to first-order, a purely radial collapse is a fair approximation.

How many spiral arms are likely to form in our disc? Numerical studies (Lodato & Rice 2004, 2005) have shown that as the disc-to-star mass-ratio increases, $Q \sim 1$ discs show fewer spiral arms. Our simulations are of quite massive discs, with a disc-to-star mass-ratio of $M_d/M_\star = 0.45$. This high disc mass is necessary for the disc to have $Q \sim 1$ near 100 AU for our realistic irradiation temperature, and results in a typical arm number of $m = 2$ or 3.

The number of arms in a disc is likely the result of swing amplification, with significant amplification of a mode, m, requiring the swing amplification parameter,

$$X_m = \frac{\Omega^2 R}{2\pi G \Sigma m}, \tag{4.11}$$

to satisfy $1 < X_m < 3$ (Binney & Tremaine 2008). From the above, we can see that X_m roughly scales with the disc-to-star mass-ratio:

$$X_m \propto M_\star / M_d. \tag{4.12}$$

Thus, for low disc-to-star mass-ratios, only high-order modes will satisfy $1 < X_m < 3$, while for high disc-to-star mass-ratios, only low-order modes will.

What is the steady-state thickness, l_1, of the newly formed spiral arm? We posit that this scale is the result of a balance between heating of the disc through the spiral waves, and radiative cooling. Assuming that the spiral density wave deposits a fixed fraction, ϵ, of its energy into the disc per dynamical time, Cossins et al. (2009) showed that the heating rate per unit mass from spiral arms can written as

$$Q^+ = \epsilon \frac{c_s^2 \mathcal{M} \tilde{\mathcal{M}} \Omega}{2} \left(\frac{\delta\Sigma}{\Sigma} \right)^2, \tag{4.13}$$

where

$$\mathcal{M} = \frac{m\Omega_\mathrm{p}}{k c_s} \text{ and } \tilde{\mathcal{M}} = \frac{m\left(\Omega_\mathrm{p} - \Omega\right)}{k c_s} \tag{4.14}$$

are the radial phase Mach number and the Doppler-shifted radial phase Mach number, Ω_p is the pattern speed, and $(\delta\Sigma/\Sigma)$ is the spiral over-density, the fractional increase in surface density in an arm compared to the average surface density at that radius.

As outlined by the authors, the pattern speed, and hence the Mach numbers, can be calculated from the dispersion relation for a finite-thickness disc:

$$m^2 \left(\Omega_{\rm p} - \Omega\right)^2 = c_s^2 k^2 + \Omega^2 - \frac{2\pi G \Sigma |k|}{1 + |k| H}, \tag{4.15}$$

if the radial and azimuthal wavenumbers are known. From their simulations, the authors found a relatively constant value of $\epsilon \approx 0.2$ (see their Figure 15). For $Q \sim 1$ discs, which have dominant radial wavenumbers of $k \approx 1/H$, solving the above dispersion relation for the ratio of the pattern speed to the rotation rate yields $\Omega_{\rm p}/\Omega = 1 + 1/m$. We therefore expect the disc to be near corotation throughout the gravitationally unstable region of $Q \sim 1$, consistent with the results of Cossins et al. (2009).

In an irradiated disc, there is additional heating from the stellar irradiation, so that

$$Q^+ = \epsilon \frac{c_s^2 \mathcal{M} \tilde{\mathcal{M}} \Omega}{2} \left(\frac{\delta\Sigma}{\Sigma}\right)^2 + \frac{2\sigma T_{\rm irrad}^4}{\Sigma}. \tag{4.16}$$

In our simple analytic model, the spiral over-density can be calculated by assuming that some fraction, $1/f$, of the total mass per length within l_0, with a characteristic surface density Σ_0, is compressed into the spiral arm, of thickness l_1, and characteristic surface density Σ_1 :

$$\left(\frac{\delta\Sigma}{\Sigma}\right) = \frac{\Sigma_1 - \Sigma_0}{\Sigma_0} = \left(\frac{l_0}{f l_1} - 1\right). \tag{4.17}$$

In the above, and in the analytic calculations of §4.4, we have made use of a simplified top-hat surface density profile for a spiral arm, since we are unable to predict the true surface density profiles that will result from GI. The surface density Σ_1 can be thought of as the integrated average of the true surface density within the arm's thickness, l_1: $\Sigma_1 = \int_{l_1} \Sigma(R) dR / l_1$.

The heating from equation (4.16) is balanced by radiative cooling, for which the cooling rate per unit mass, using equation (4.9), is

$$Q^- = u/t_{\rm cool} = \frac{4}{\gamma} \frac{\sigma T^4}{\kappa \Sigma^2}, \tag{4.18}$$

where we have used $u = c_s^2/[\gamma(\gamma - 1)]$. Setting the above cooling rate equal to the heating rate of equation (4.16), and using the other information in this section, as well as the initial axisymmetric properties of our disc, leaves us with an equation with only one unknown: the thickness of the spiral arm, l_1, in our patch of interest (assuming that we know the proper midplane temperature for radiative cooling, T, in the spiral arm. This will be elaborated upon in §4.4).

4.3.2 Determining fragmentation

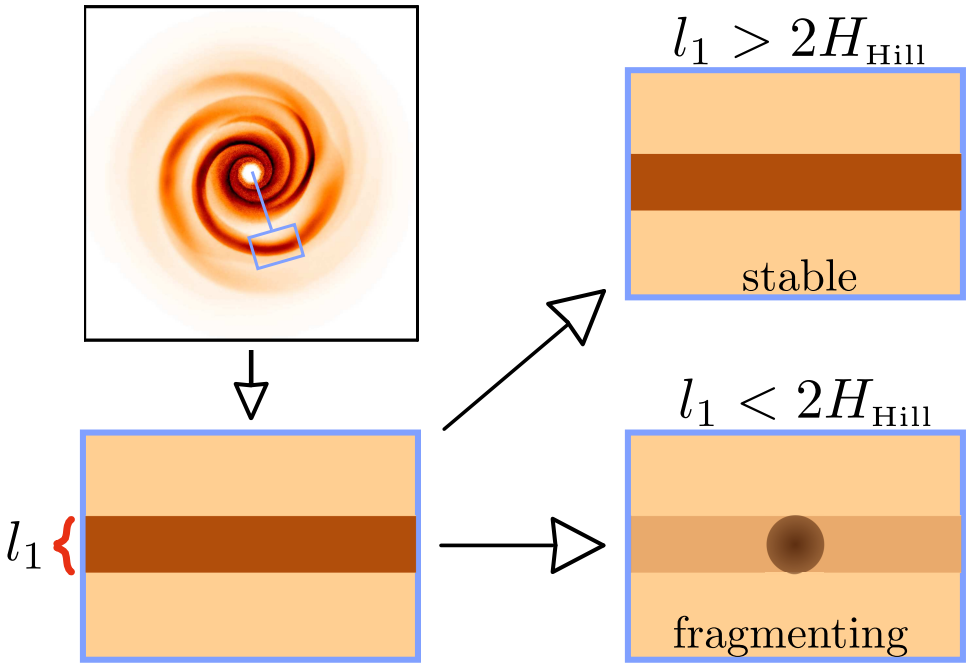

Figure 4.6: The Hill criterion for spiral arm fragmentation: if a section of spiral arm lies within its own Hill thickness, then that section of arm is free to collapse and fragmentation takes place. If a section of spiral arm lies outside of its own Hill thickness, then shear stabilizes the arm and fragmentation does not take place.

Once we know the steady-state thickness of the spiral arm, l_1, we can determine the mass of the section of arm within that thickness and use equation (4.6) to calculate the Hill radius for this section of the arm. If the section of arm has a thickness satisfying $l_1/(2H_{\rm Hill}) < 1$, then the section lies within its own Hill thickness. In the absence of pressure forces, this means that the section is bound, as the tidal force from the

central star (manifest as rotational shear) is less than the self-gravity of the section. Once the section is bound, fragmentation occurs. Conversely, if $l_1/(2H_{\rm Hill}) > 1$, then the section of arm is not bound and fragmentation does not occur.

In our simple analytic model, with a top-hat surface density profile as in equation (4.17), the section of arm has a mass $M = \Sigma_1 l_1^2$ and equation (4.6) gives a Hill radius of:

$$H_{\rm Hill} = \left[\frac{G\Sigma_1 l_1^2}{3\Omega^2}\right]^{1/3}. \tag{4.19}$$

The Hill thickness tells one about the ability of shear to prevent the fragmentation of the arm and is therefore expected to be an important scale. However, in comparing the radial thickness of the arm to the Hill thickness we are ignoring the role of pressure, despite strong radial pressure gradients present across the arm. It is therefore reasonable to expect that the critical thickness for arm fragmentation may be modified from the Hill thickness. There are, however, no strong pressure gradients along the arm (the azimuthal direction), thus fragmentation occurring in this direction should be determined by the Hill criterion.

Determining the correct scale for fragmentation requires a detailed calculation of the stability of a spiral arm accounting for differential rotation. In the absence of such a calculation, we posit that the correct scale to consider is indeed the Hill thickness. As described below, the results from our simulations are consistent with this. The Hill criterion for fragmentation, demonstrated in Figure 4.6, is thus empirically based.

4.3.3 Consistency with simulations

An analysis of the spiral arms formed in the simulations of §4.2 shows that their thicknesses, and stability, are consistent with the Hill criterion for fragmentation. In Figure 4.7, we show two examples of this analysis to illustrate this consistency. Our analysis focuses on the surface density of a radial slice of the disc (with a typical angular width of 5°). Over this slice, a spiral arm is evident as a large over-density. We find that arms are often asymmetric; consequently, we determine a thickness for an arm by fitting each side of the arm (with respect to the radius of highest Σ, $R_{\rm peak}$),

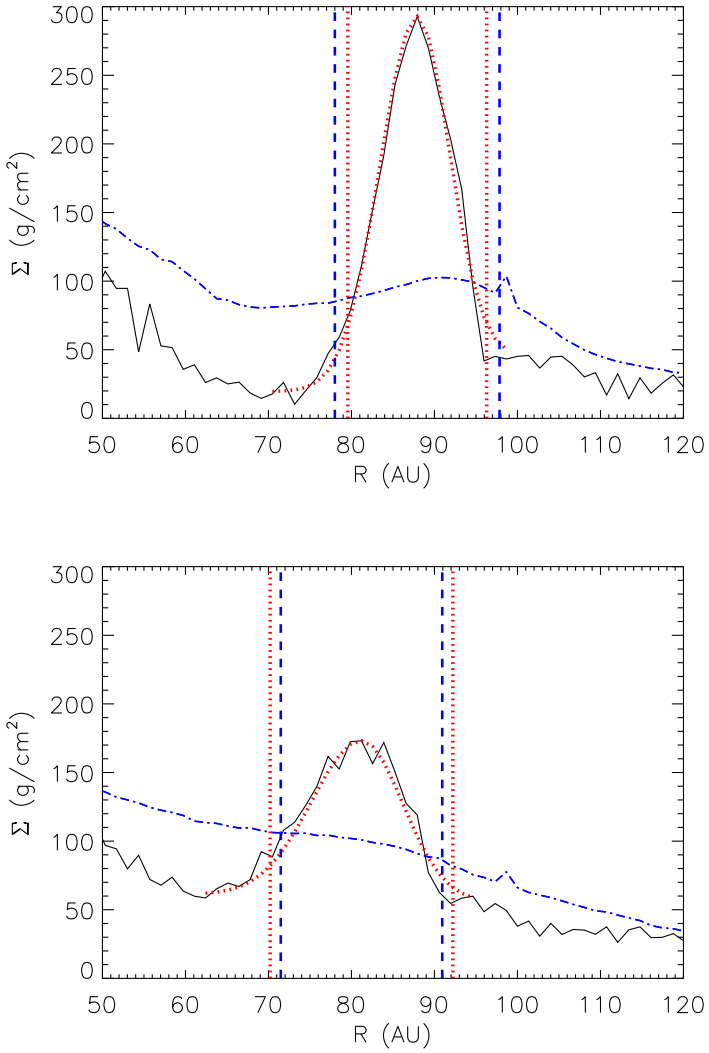

Figure 4.7: Comparison of a spiral arm's thickness to its Hill thickness. A radial slice of Simulation B (Simulation D) is shown in the top (bottom) panel. The surface density of the radial slice is given by the black line, while the azimuthally averaged surface density is given by the blue, dot-dashed line. As described in the text, the spiral arm (the large over-density) is fit (the red, dashed line) and the arm thickness is found (the vertical, red dashed lines). The mass within this arm thickness is computed to determine the Hill thickness (the vertical, blue dashed lines). Consistent with the Hill criterion, Simulation B (Simulation D) has an arm which falls inside of (outside of) its Hill thickness and fragmentation is (is not) observed shortly thereafter.

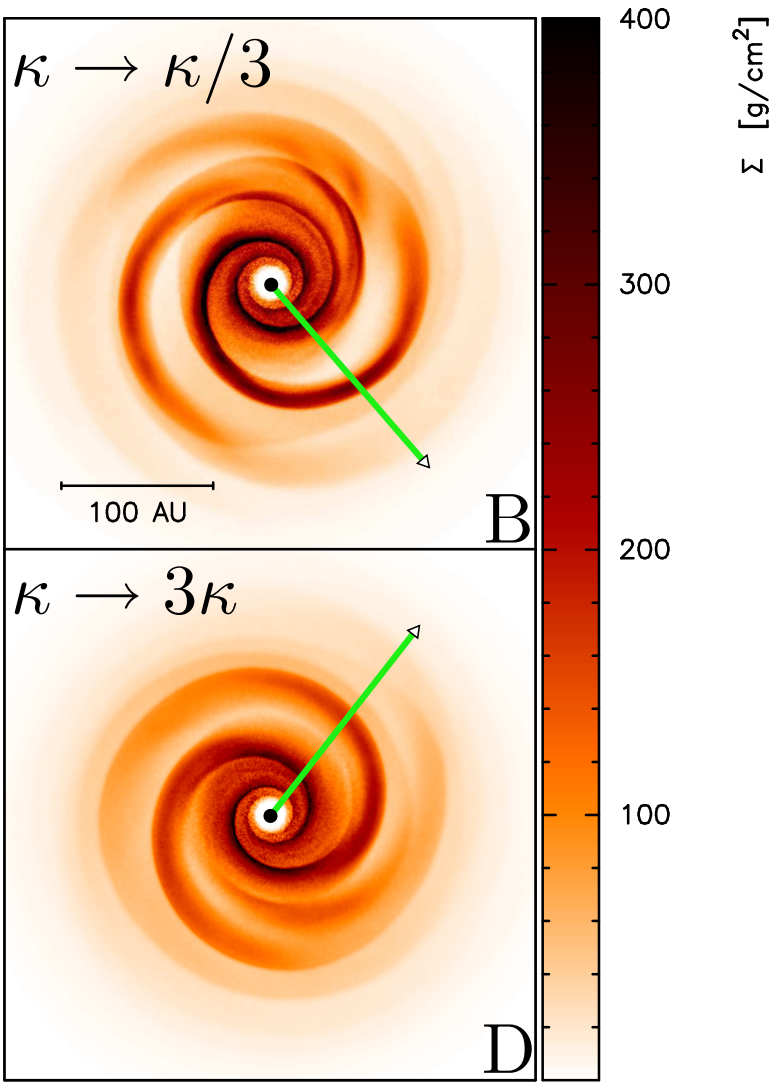

Figure 4.8: Location of the radial cross-sections in Figure 4.7.

with a Gaussian of the form

$$\Sigma_{\rm arm} = \Sigma_{\rm base} + \Sigma_{\rm peak} e^{\frac{\left(R-R_{\rm peak}\right)^2}{2b^2}}, \tag{4.20}$$

where $\Sigma_{\rm base}$ is the value of the surface density adjacent to the arm and $\Sigma_{\rm peak}$ is the maximum of the arm's surface density. The thickness of the arm is taken to be

$l_1 = a(b_{\rm left} + b_{\rm right})$, with $a = 2$, and the mass of the section of arm is determined using a numerical evaluation of

$$M_{\rm arm} = \int_{R_{\rm peak}-2b_{\rm left}}^{R_{\rm peak}+2b_{\rm right}} \Theta(R) R \Sigma(R) dR, \tag{4.21}$$

where $\Theta(R) = l_1/R$ is the angular extent of the section of arm. In contrast to our analytic model presented in the preceding section, we do not use a simplified top-hat model here, but integrate the fit to the measured arm profile over the thickness of the arm, since we know what the arm profile is in this case. This allows us to verify the Hill criterion's ability to determine fragmentation. With this verification, we can then be confident in applying the Hill criterion to our analytic model, as we do in §4.4.

Determining the thickness of a spiral arm is not a trivial task. Our use of Gaussian fitting functions is an attempt to be objective and reproducible and is motivated by the reasonably accurate fits that we achieve. Our choice of $a = 2$ as the coefficient in determining l_1 is motivated by having the majority (95%) of the over-density of the arm (the mass in the arm that is at $\Sigma > \Sigma_{\rm base}$) contained within the arm. To contain all of the mass in a Gaussian fit requires integrating out to an infinite distance, which is clearly problematic in defining the finite thickness of a spiral arm. We have found that using values of $a > 2$ in fitting the arms often results in including material that is clearly outside of the arm; that is, the calculated arm thickness is obviously too large when the radial surface density profiles are examined by eye.

Figure 4.7 (top) shows this analysis, at roughly 2.5 ORPs, for an arm in Simulation B, shortly before fragmentation. Consistent with the Hill criterion, the arm thickness is less than the Hill thickness and therefore fragmentation is expected to occur in this arm; indeed, this arm fragmented a short time after the time-step used for this analysis. In contrast, Figure 4.7 (bottom) shows this analysis, at roughly 2.3 ORPs, for an arm in Simulation D that never fragmented. Consistent with the Hill criterion, the arm thickness is greater than the Hill thickness and therefore fragmentation is not expected. The positions of the radial slices for this analysis are shown in Figure 4.8.

We use these two examples to illustrate the consistency of the fragmentation criterion with the simulations performed. More generally, we have found that lower opacity discs have arms that are consistently smaller with respect to their Hill thickness in comparison with higher opacity discs. Despite a large range in the opacity scalings used in the simulations (a factor of 100), we find that the arm thickness in the discs is generally similar to the Hill thickness. This can be observed for the two arms examined in Figure 4.7. This small range for the ratio of the arm thickness to the Hill thickness observed in the simulations may seem surprising; however, it can be understood through our analytic calculations for radiatively cooled discs, presented in §4.4.2. The ratio of the arm thickness to the Hill thickness is weakly dependent on the opacity.

In each of the cases where fragmentation takes place (this occurs in the reduced opacity discs), the arms that fragment are observed to fall within their own Hill thickness shortly before fragmentation. Indeed, all arms that smoothly evolve to lie within their own Hill thickness fragment.

4.3.4 Detailed evolution of spiral arms

The simplicity of the model described above is desirable, as it gives a straightforward, physical picture of the formation of spiral arms in a gravitationally unstable disc, as well as the physical criterion that determines whether or not those spiral arms fragment. However, unstable discs do exhibit a great deal of complexity; here, we discuss this complexity and comment on its implications for our model.

We have described an individual spiral arm's stability as being the result of whether or not it is contained within its Hill radius. In addition, we have described the steady-state thickness of a single spiral arm as being the result of a balance between heating and cooling. However, the simulations show that the spiral structure in the disc evolves with time: the number of arms in the disc is not constant, nor is the over-density of each arm.

An analysis of spiral arms in our simulations shows that individual spiral arms typically exist for a number (roughly three) of sound-crossing times; that is, they last

approximately a dynamical time (orbital period). The sound-crossing times of arms are shorter than the radiative cooling time (two to thirty times shorter, depending on the opacity scaling), suggesting that the arms can be treated as evolving in a quasi-static manner. This is important in applying the Hill criterion to determine stability.

Material in the arm is shock-heated, and radiatively cools. From the observed density contrasts between arms and inter-arm regions, we estimate the expected temperature increase in an adiabatic shock, and calculate the radiative cooling time to reach the observed temperature contrasts. We find that this cooling time is within the lifetime of the arms (it is equal to or up to five times shorter than the dynamical time, depending on the opacity scaling). This suggests that the arms have time to come to a balance between heating and cooling. In addition, the ability of the arm to readjust hydrodynamically on the short sound-crossing time also suggests that heating and cooling can come into balance, since the arm can quickly adjust to an imbalance.

Evolution of the disc as a whole, such as heating during a phase of pronounced spiral activity such as a "burst" (Mejía et al. 2005), does not prevent the application of our model. As long as the evolution of the disc takes place on a timescale that is longer than the sound-crossing time of a spiral arm, the balance between heating and cooling in a spiral arm is a reasonable approximation. Since a burst phase heats the disc on roughly the orbital time, the model is applicable. A balance between heating and cooling will be invalid, however, in cases where there are strong hydrodynamic interactions between spiral arms (or between a spiral and a fragment, see below), since this will produce heating on a timescale of roughly the sound-crossing time.

The Hill criterion describes the fragmentation of a spiral arm, but it does not necessarily determine whether or not this fragmentation leads to the formation of a long-lived object; this also depends on the cooling of the fragment and the complex environment of the disc in which the initial fragmentation takes place. Simulation C, for example, demonstrates an instance of "failed fragmentation," as shown in Figure 4.9. One of the spiral arms appears to have fragmented; however, the fragment is only short-lived: it shortly thereafter collided with the next spiral arm, without surviving.

The Hill criterion describes the formation of a fragment based on the inability of

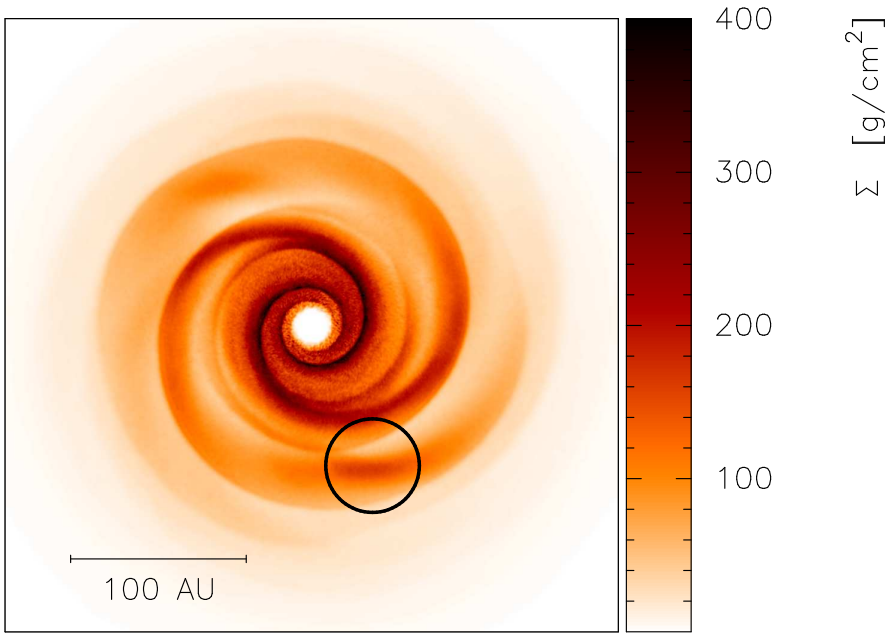

Figure 4.9: An instance of "failed fragmentation" in Simulation C. In this example, a fragment, visible within the black circle, is observed; however, it is not long-lived. Shortly after this time-step, the fragment collided with the spiral arm and did not survive.

shear to stabilize a section of a spiral arm. Further collapse occurs on the radiative cooling timescale of the fragment. If this timescale is long, then the fragment may still be quite diffuse, and easily disrupted by collisions with subsequent spiral arms. Indeed, Simulation B, which generally has shorter cooling times than Simulation C because of its opacity scaling, shows a fragment which formed, but then subsequently collided twice with spiral arms; in contrast to Simulation C, this object survived, as observed in Figure 4.2. Fragmentation, therefore, can be well characterized by our model; however, whether or not fragments survive also depends on the complex non-linear interactions between collapsing fragments and the spiral structure in the disc.

As described, the fragment of Simulation C was disrupted through a collision with the subsequent spiral arm. This resulted in a strong compression of the spiral arm; in fact, the compression was strong enough that the arm was observed to lie within its own Hill thickness. Nevertheless, the arm did not fragment. This does not conflict

with the Hill criterion because in this instance, the arm was not in a near steady-state. Since the timescale for the collision was much shorter than the sound-crossing time of the arm, the arm could not adjust. As a result, the increased over-density of the arm lead to an increased heating rate, see equation (4.13), and a reduced cooling rate, see equation (4.18). As a consequence of the imbalance between heating and cooling, pressure forces caused the arm thickness to expand on roughly the sound-crossing time, with the result that no fragmentation took place.

4.4 Consistency with the critical cooling criterion and predictive ability

The physical model developed in the preceding section can be used to examine whether or not fragmentation is likely to take place in a disc. In this section, we demonstrate that the predictions of this model are consistent with previous numerical results of discs evolved using the β-prescription of cooling, as well as the results of our suite of simulations discussed in §4.2. Specifically, we analyze the initial condition of our simulated disc described in §4.2 and adopt general values for the parameters. We use $l_0 = 2\pi H$, $\epsilon = 0.2$ (Cossins et al. 2009), $m = 2$ and $f = 1.5$ (characteristic of measurements from our set of simulations). We calculate the value of $\mathcal{M}\tilde{\mathcal{M}}$, equation (4.14), as outlined in Cossins et al. (2009), by solving the dispersion equation, (4.15); we use equation (4.10) for k, which results in $\mathcal{M}\tilde{\mathcal{M}} = Q^2\,(m+1)$. The values for these parameters will be correct to within $\mathcal{O}(1)$, but will likely have variation depending on the physical properties of a disc.

Caution is warranted when considering the stability of a disc based on its initial condition. Johnson & Gammie (2003) examined the stability of discs using 2-D shearing-box simulations that included radiative cooling. They found that the cooling times in the non-linear phase of the disc could differ by orders of magnitude from the cooling times in the initial condition. As a consequence, assessing the stability of the disc in its non-linear phase, via the cooling criterion, based on the initial cooling times can lead to the wrong conclusion.

In the following calculations, we use the initial condition of our disc simulations as a starting point in analytic calculations of stability. However, our model takes into account the increased surface density in the spiral arms that is characteristic of the non-linear phase. In addition, for irradiated discs, we take into account the increase in temperature and opacity in the spiral arms, based on results from our set of simulations.

4.4.1 Calculating the critical cooling time

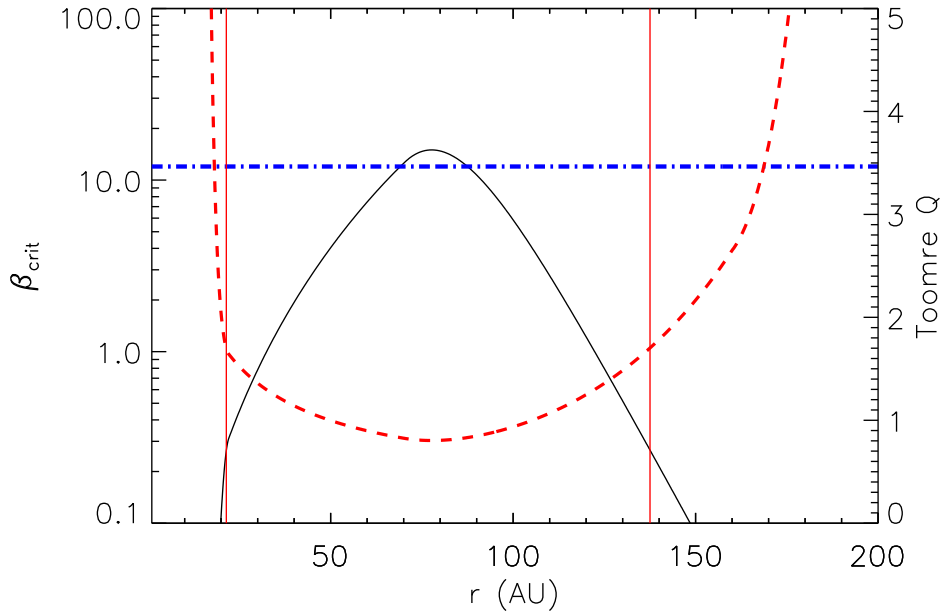

Figure 4.10: The critical cooling time (black line), as calculated for our disc initial condition. The horizontal blue, dot-dashed line represents the value as previously determined by numerical experiments, while the red, dashed line shows the Q profile. The vertical, red lines show the extent of the unstable region, with $Q < 1.7$. As can be seen, the value for $\beta_{\rm crit}$ in this region is consistent with the value from numerical experiment.

The critical cooling time, $\beta_{\rm crit}$, for a $Q \sim 1$ disc with $\gamma = 7/5$ has been found to be $\beta_{\rm crit} = 12$ from numerical experiments [Rice et al. (2005); with the caveat that numerical convergence has not been clearly demonstrated]. If we adopt a heating

rate without irradiation (consistent with the aforementioned simulations), given by equation (4.13) and balance this heating with a β-prescription cooling rate, given by

$$Q^- = \frac{c_s^2 \Omega}{\gamma(\gamma-1)\beta}, \tag{4.22}$$

then we can solve for the cooling rate that results in a certain arm thickness:

$$\beta = \frac{2}{\epsilon \mathcal{M}\tilde{\mathcal{M}}\gamma\left(\gamma-1\right)}\left[1 - \frac{l_0}{f 2^{3/2}}\left(\frac{2H_{\rm Hill}}{l_1}\right)^{3/2}\left(\frac{3f\Omega^2}{G\Sigma_0 l_0}\right)^{1/2}\right]^{-2}. \tag{4.23}$$

The critical cooling time can then be computed from the above equation by setting the arm thickness to exactly match the Hill thickness, $l_1/(2H_{\rm Hill}) = 1$. Since both the heating, Q^+, and cooling, Q^-, are proportional to the square of the sound-speed in this case, the increase in sound-speed that occurs in a spiral shock does not affect the critical cooling time.

The critical cooling time computed for our disc initial condition is shown in Figure 4.10. As can be seen, we do not find a unique value for the critical cooling rate, but rather a value that depends strongly on radius (due to the variation of disc properties with radius). Importantly, we observe that our model of disc fragmentation predicts critical cooling times in the unstable region of the disc (where $Q \sim 1$) that are consistent with the results of numerical experiments.

Equation (4.23) implies that fragmentation should occur for small values of $\mathcal{M}\tilde{\mathcal{M}}$, since these result in longer critical cooling times. The product $\mathcal{M}\tilde{\mathcal{M}}$ is smallest near corotation (in fact, it is formally zero there: see equation 4.14); therefore, fragmentation should occur near corotation, as suggested by Durisen et al. (2008). However, as discussed in §4.3, regions with $Q \sim 1$ are expected to be near corotation. Hence, fragmentation near corotation is consistent with fragmentation in regions with $Q \sim 1$.

The consistency of our model with numerical experiments is a useful check on our model. In addition, it is noteworthy that Figure 4.10 demonstrates the first calculation of the critical cooling time from a physical model of fragmentation. Previous estimates of $\beta_{\rm crit}$ have come only from numerical experiments.

4.4.2 Predictive ability of the model for irradiated discs

It is useful to check our model against the results of previous work using cooling in the form of a β-prescription. However, it is of particular interest to apply the model to the more realistic case of an irradiated disc with radiative cooling. Without considering GI, an irradiated disc has a natural equilibrium state in which the heating of a particular patch of disc from stellar irradiation is balanced by the radiative cooling of that patch. Here, we consider deviations from this equilibrium state due to GI.

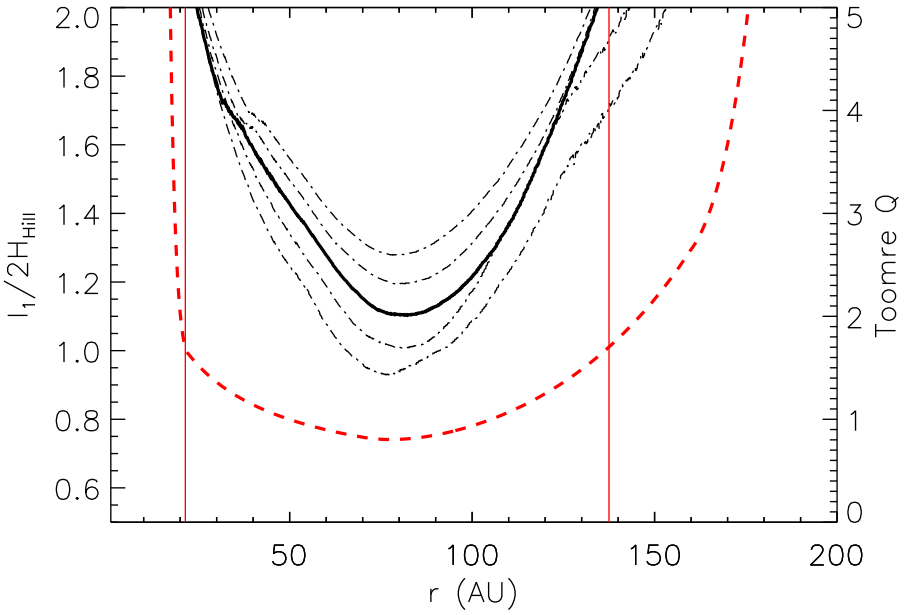

Figure 4.11: The spiral arm thickness (black curves) of an irradiated disc in units of the Hill thickness, as calculated for our disc initial condition. The arm thickness is calculated for the range of opacities used in the simulations. From the curve of greatest $l_1/(2H_{\rm Hill})$ to the curve of smallest $l_1/(2H_{\rm Hill})$ the opacity scalings are 10, 3, 1, 1/3, and 1/10 the physical opacity (the solid black curve). The Q profile is given by the red, dashed curve, while the vertical, red lines show the extent of the unstable region, with $Q < 1.7$.

Specifically, there is an additional heating of the disc from the spiral arms, given by equation (4.13), which will result in an increase in the midplane temperature of δT. We consider spiral arms in which this excess heating is balanced by a perturbative

radiative cooling. Following Kratter et al. (2010b), for excess heating transported to the photosphere via radiative diffusion, the perturbative radiative flux is

$$F = \frac{16\sigma}{3\kappa\Sigma}\left(T^4 - T_{\rm irrad}^4\right), \tag{4.24}$$

which results in a cooling rate per unit mass of

$$Q^- = \frac{F}{\Sigma} = \frac{16\sigma}{3\kappa\Sigma^2}\left(T^4 - T_{\rm irrad}^4\right), \tag{4.25}$$

where the midplane temperature is $T = T_{\rm irrad} + \delta T$.

We calculate the natural arm thickness for our disc initial condition from §4.2 by balancing the perturbative heating and cooling in the arm. This results in the following quartic:

$$\{C\}\, l_1^4 - \{B\}\, l_1^2 + \left\{\frac{2Bl_0}{f}\right\} l_1 - \left\{B\left(\frac{l_0}{f}\right)^2\right\} = 0, \tag{4.26}$$

where

$$B = \frac{\epsilon c_s^2 \mathcal{M}\tilde{\mathcal{M}}\Omega}{2} \tag{4.27}$$

and

$$C = \frac{16\sigma\left(T^4 - T_{\rm irrad}^4\right)}{3\kappa}\left(\frac{f}{\Sigma_0 l_0}\right)^2. \tag{4.28}$$

Once l_1 is calculated, we can calculate the ratio

$$\left(\frac{l_1}{2H_{\rm Hill}}\right) = \frac{l_1^{2/3}}{2}\left[\frac{3f\Omega^2}{G\Sigma_0 l_0}\right]^{1/3} \tag{4.29}$$

to determine the stability of the arm.

Computing the arm thickness requires knowledge of the midplane temperature of the spiral arm during the non-linear phase. Since we are unable to self-consistently predict this temperature, this unfortunately limits the predictive ability of the model's application to radiatively cooled discs. However, we can use the results of our simula-

tions in §4.2 to test the consistency of the model with our simulation results. We have measured the maximum midplane temperature for each opacity case as a function of radius, averaged over the one ORP before fragmentation took place in Simulations A and B. We use these measurements of T in determining the arm thickness from equation (4.26). In addition to the temperature perturbation, we take into account the resulting changes in sound speed and opacity.

Johnson & Gammie (2003) found that cooling times in the non-linear phase can differ by orders of magnitude from the cooling times in the initial disc, largely because of sharp variations in the opacity with temperature. However, these sharp variations occur in discs that have high temperatures; in particular, temperatures near the opacity gap, $T \sim 1200$ K. In contrast, the outer regions of irradiated discs typically have much lower temperatures of $T < 100$ K. For these temperatures, there is no sharp variation in $\kappa(T)$ and κ is not a function of the density. For the typical increase of $\delta T \sim T_{\rm irrad}/2$ observed in the region of fragmentation in our simulations, the resulting change in opacity is only about a factor of 1.25. In the outer regions of irradiated discs, opacity changes during the non-linear phase are not a large effect; however, we have taken them into account in our calculations.

The results of the arm thickness calculations, over the range of opacity scalings used in our simulations, are shown in Figure 4.11. We expect discs to fragment for $l_1/(2H_{\rm Hill}) \leq 1$, and the trend in Figure 4.11 is consistent with this picture: as found in our simulations, the increased-opacity discs are less likely to fragment (have larger $l_1/(2H_{\rm Hill})$) than the reduced-opacity discs. However, even though the increased-opacity discs are not expected to fragment, the arm thickness is expected to be within a factor of two of the Hill thickness in the region of $Q \sim 1$. The small range in $l_1/(2H_{\rm Hill})$ over the different opacity scalings, as seen in Figure 4.11, may seem counter-intuitive given that the opacity variation is a factor of 100. However, we can understand this weak dependence on opacity as the result of the non-linear nature of the self-regulation of heating and cooling in a spiral arm.

As described in §4.3.1, the region of the initial disc that goes into the formation of the arm, l_0, is the result of the initial GI and does not depend on the opacity of the disc. The thickness of the spiral arm is then the result of the balance of heating and

cooling in the arm. For radiative cooling in a spiral arm, the cooling rate per unit mass is $Q^- \propto 1/(\kappa\Sigma_1^2)$, while the heating rate per unit mass is $Q^+ \propto \Sigma_1^2$. A lower opacity results in a faster cooling rate, which requires a faster heating rate for balance to be achieved. This faster heating rate requires a higher surface density in the arm, but this acts to decrease the cooling rate. One can see then, that the balance between heating and cooling, which determines $l_1/(2H_{\rm Hill})$, has a non-linear dependence on the opacity. In fact, if we parameterize this dependence as a power-law, $l_1/(2H_{\rm Hill}) \propto \kappa^\nu$, in the region of instability, we find a weak opacity dependence, with $\nu \sim 0.1$.

As discussed in §4.3.3, this result is consistent with our simulations. Figure 4.7 (bottom) shows the analysis of an arm from Simulation D. This arm is stable, in that it does not lie within its Hill thickness; however, its thickness only exceeds the Hill thickness by a relatively small factor. All of the arms analyzed in the simulations have thicknesses that are similar to their Hill thicknesses.

From our calculation, only the lowest opacity case, corresponding to Simulation A, has $l_1/(2H_{\rm Hill}) < 1$, and would be expected to fragment; in fact, both Simulation A, and the second-lowest opacity case, Simulation B, have shown fragmentation. This discrepancy is likely simply the result of the choices of parameters used in the calculation. We have chosen values for a number of parameters in our model (l_0, f, ϵ, and $\mathcal{M}\tilde{\mathcal{M}}$) that are expected to be correct to within $\mathcal{O}(1)$; however, the exact values will likely have some variation. A change in these parameters shifts the curves of $l_1/(2H_{\rm Hill})$ vertically, but has little effect on the small range of $l_1/(2H_{\rm Hill})$ over opacity scalings. With an improved understanding of the growth of spiral structure, and the heating of spiral arms, the model's predictive abilities will be improved.

Furthermore, our analytic model considers the stability of a disc based on the stability of a statistically-averaged arm, as our choices of parameters, (l_0, f, ϵ, and $\mathcal{M}\tilde{\mathcal{M}}$), are characteristic of the average. For example, the value of ϵ, was determined by Cossins et al. (2009) by considering the overall self-regulation of their $Q \sim 1$ discs. It is unclear how much the value of ϵ in a given arm is expected to vary from this statistical average. For a disc to fragment, however, only a single arm is required to be unstable, rather than the statistically-averaged arm. It may then be important to understand the variance of the model parameters away from their characteristic

values, in order to more accurately predict fragmentation.

As with the application of the model to the case of β-prescription cooling in §4.4.1, we find, from our numerical solution of equation (4.26), that fragmentation is more likely for smaller values of $\mathcal{M}\tilde{\mathcal{M}}$ in the case of radiative cooling. This implies that fragmentation should take place near corotation.

4.4.3 Predicted masses of fragments in irradiated discs

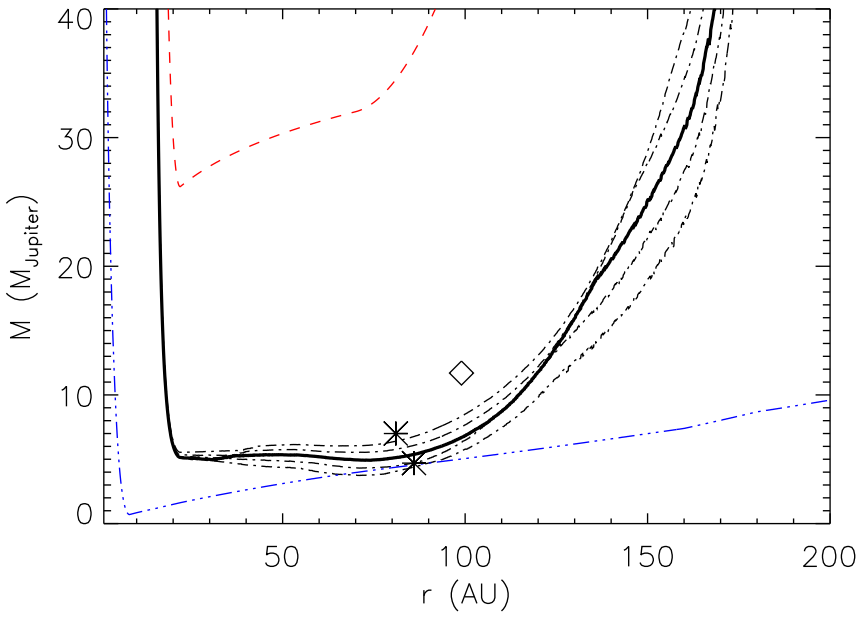

Figure 4.12: The predicted initial fragment mass (black curves), in Jupiter masses, of an irradiated disc, as calculated for our disc initial condition. The fragment mass is calculated for the range of opacities used in the simulations. From the curve of greatest mass to the curve of smallest mass, the opacity scalings are 10, 3, 1, 1/3, and 1/10 the physical opacity (the solid black curve). The Toomre mass is given by the red, dashed curve, while the predicted mass of Boley et al. (2010) is given by the blue, triple-dot-dashed curve. The asterisks represents the initial fragments from Simulation A, while the diamond represents the initial fragment form Simulation B.

The initial mass of fragments can be a useful constraint on the physical model of fragmentation (Boley et al. 2010; Forgan & Rice 2011). From the natural arm

thickness, l_1, calculated using equation (4.26), the fragment mass can be estimated as $M = \Sigma_1 l_1^2$. The model predictions are lower-bounds for the final mass of fragments, since significant accretion of disc-material can occur after fragmentation has taken place.

Figure 4.12 demonstrates the initial fragment mass expected for the range of opacity scalings used in our simulations. We compare our model's fragment mass with the Toomre mass, $M_{\rm T} = \pi\Sigma(l_0/2)^2$, as well as with the fragmentation model of Boley et al. (2010). In addition, we compare the model predictions with the initial fragments observed in Simulations A and B, for which masses have been determined using the SKID group-finder (Stadel 2001). SKID determines groups of SPH particles based on the gradient of density, and then performs an iterative unbinding procedure, which we have modified to include thermal energy.

As can be observed in Figure 4.12, our model predicts fragments with initial masses in the gas-giant regime, significantly less than the Toomre mass. The masses of the fragments from Simulations A and B are consistent with the predicted masses to within a factor of two*. It is interesting to note that the fragmentation model of Boley et al. (2010), strictly valid only for isothermal simulations, predicts similar mass fragments in the region where fragmentation is observed.

4.5 Discussion and conclusions

4.5.1 Implications for planet formation

Direct-imaging observations have shown the existence of gas-giant planets at large distances from their host A star, including HR 8799b, 7 $M_{\rm Jup}$ at a distance of 68 AU (Marois et al. 2008), and Fomalhautb, 3 $M_{\rm Jup}$ at a distance of 119 AU (Kalas et al. 2008). It is difficult to explain the existence of gas-giants at such distances from their host star in the core-accretion scenario, since the surface densities are typically too low to form the necessary rocky cores within the lifetime of the gas disc (Dodson-

**Additional comment for thesis*: we would like to emphasize that these are the *initial* masses of the fragments. The final masses are substantially larger due to subsequent accretion of disc material, and are in the brown dwarf regime.

Robinson et al. 2009; Rafikov 2011). However, more investigation is warranted in order to determine if such planets can be explained in the core-accretion scenario.

In comparison, fragmentation via GI has been shown to be a viable formation mechanism at large distances from the host star from both theoretical arguments (Rafikov 2007; Nero & Bjorkman 2009; Kratter et al. 2010b), as well as from numerical simulations with radiative transfer (Boley 2009; Vorobyov & Basu 2010; Stamatellos et al. 2011; Boss 2011).

The particular set of 3D radiation hydrodynamic simulations presented here were designed to investigate fragmentation at large radii ($\sim$ 100 AU) around A stars. At these distances, heating from stellar irradiation is expected to be the dominant heating source; we have included irradiation using the $T_{\rm irrad}$ expected for a 1.35 $M_\odot$ A star (Kratter et al. 2010b).

The results of our simulations show that GI can produce gravitationally bound objects at large distances from the star, given opacities on the low side of the expected range. Such opacities could be the result of grain-growth (Birnstiel et al. 2010), grain evolution via the passage of spiral arms (Podolak et al. 2011), or formation in an environment with a non-solar metallicity (HR 8799 is roughly 1/3 solar metallicity, Marois et al. 2008). Our simulations do not take these physical mechanisms into account, but rather use a simple scaling of the opacity table.

Although our simulations do not have the resolution to follow the evolution of bound objects as their central densities run away, it is interesting to consider the objects at the end-state of our simulations, shown in Figure 4.2. At the end of Simulation A (with an opacity scaled by 1/10), there are two brown dwarfs of masses 21 $M_{\rm Jup}$ and 15 $M_{\rm Jup}$, at respective distances of 62 AU and 95 AU; while at the end of Simulation B (with an opacity scaled by 1/3), there is one brown dwarf of mass 40 $M_{\rm Jup}$ at a distance of 95 AU. Neither the masses nor the distances of these objects represent their final state: all of the objects are accreting mass and migrating inwards at the end of the simulation. The long-term evolution, and migration, of objects formed at large distances through GI is an active area of research. In addition to providing a mechanism to form the gas-giants observed at large distances from their star, formation of objects through GI, coupled with inward migration and tidal

disruption (Boley & Durisen 2010; Nayakshin 2010) represents an interesting channel for the formation of rocky cores at smaller distances.

We conclude that GI in discs can produce brown dwarfs at large distances from A stars. We have, of course, only shown fragmentation for a single surface density and temperature profile. It is of interest to investigate a greater region of the parameter space with numerical simulations in order to explore the possibility of low-mass companions such as those observed by Marois et al. (2008) and Kalas et al. (2008)*.

4.5.2 Physical model of fragmentation

We have presented a new framework to explain the link between cooling and fragmentation in protostellar discs. This framework consists of two components. The first is a simple model for the formation of spiral arms, in which the thickness of a spiral arm is set by a balance between heating (through gravitational instability and irradiation) and radiative cooling. The second is a criterion for fragmentation: spiral arms that have a natural thickness smaller than their Hill thickness fragment, resulting in objects that may survive to become gas-giant planets or brown dwarfs.

This model of fragmentation is based on results from Cossins et al. (2009) as well as our suite of 3-D radiation hydrodynamics simulations of gravitational instability in an irradiated, optically-thick protostellar disc surrounding an A star. By reducing the opacity scaling, and consequently the cooling time, over the set of simulations, we have produced a suite that demonstrates the transition from non-fragmentation to fragmentation. From an analysis of these simulations, we have found that the critical scale for determining fragmentation is roughly the Hill thickness: those spiral arms that are found to fragment lie within twice their Hill radius, while those spiral arms that do not fragment extend beyond their Hill thickness. In the future, it would be of interest to have a robust calculation of the critical scale for fragmentation from a

**Additional comment for thesis*: the initial masses of the fragments produced in the simulations presented here are in the gas-giant planet regime. However, in the outer disc, the isolation and gap opening masses are in the brown dwarf regime, so fragments grow from planets to brown dwarfs through accretion of disc material. Fragments may remain in the gas-giant planet mass regime, and so explain the low-mass companions observed, through gap overlap of multiple fragments or dispersal of the disc; however, the rate of occurrence of such an outcome requires further study (see Section 2.4.2 as well as Kratter et al. 2010b).

stability analysis of a spiral arm in a differentially rotating system.

In comparison to the critical cooling time picture, our model of fragmentation is a more detailed, and more general, physical picture of fragmentation that is applicable to discs with realistic heating and cooling, in addition to discs with β-prescription cooling. Indeed, by coupling the Hill criterion to our simple model of spiral arm formation, heating, and cooling using a β-prescription, we have been able, for the first time, to calculate $\beta_{\rm crit}$. We find that there is not a single value for $\beta_{\rm crit}$, but that it depends on the local properties of the disc; in addition, our calculation is consistent with the value determined by numerical experiment.

We have also demonstrated how this model can be used to predict fragmentation, and fragment masses, in irradiated discs with radiative cooling. Applying the model to the initial condition of our simulated disc, for the various opacity scalings used, yields predictions that are consistent with the results of our simulations. An improvement in the predictive abilities of the model depends on a better understanding of several parameters that describe the formation and heating of the spiral arms.

Fragmentation, as determined by the Hill criterion, is weakly dependent on opacity, in agreement with Boss (2002) and Cai et al. (2006). Despite the factor of 100 in opacity scaling across the suite of simulations, the thickness of spiral arms is observed to be near the Hill thickness, as predicted by our analytic calculations.

In addition, the survival of a fragment can depend on the outcome of the initial fragment's collision with the remaining spiral arms. In one of our simulations, a fragment was destroyed through this process. The destruction of fragments through interactions with weak shocks has also recently been shown to be important in studies of GI employing the β-cooling prescription (Paardekooper 2012). The outcome of these interactions may be particularly sensitive to numerical methods.

This model has been developed in the context of protostellar discs; however, it may also be of use in the context of star formation in a disc near the Galactic centre (Levin & Beloborodov 2003), as well as star-cluster formation in optically-thick starburst galaxies, such as Arp 220.

In this work, we have considered isolated discs; that is, the effects of accretion from the surrounding envelope were ignored. However, accretion is expected to play an im-

portant role in gravitationally unstable discs [Boley (2009); Kratter et al. (2010a)], since it is accretion that will push the mass of the disc towards being sufficient for instability to set in, keep it unstable despite mass-transport, and contribute to heating. In future work, we intend to investigate the effects of accretion on the stability of protostellar discs, in the context of our model of fragmentation.

4.6 Acknowledgements

We would like to thank Ralph Pudritz for useful discussions. In addition, we would like to thank the referee, Richard Durisen, for thorough comments that have greatly improved this work. Surface density plots were created using SPLASH (Price 2007). This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC), and computations were carried out using the facilities of the Shared Hierarchical Research Computing Network (SHARCNET), which is located in Ontario, Canada.

Bibliography

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

Birnstiel T., Dullemond C. P., Brauer F., 2010, A&A, 513, A79+

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

Boley A. C., Durisen R. H., 2010, ApJ, 724, 618

Boley A. C., Durisen R. H., Nordlund Å., Lord J., 2007a, apj, 665, 1254

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

Boley A. C., Hayfield T., Mayer L., Durisen R. H., 2010, Icarus, 207, 509

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

Boss A. P., 2011, ApJ, 731, 74

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., Boley A. C., Mejía A. C., Pickett M. K., D'Alessio P., 2006, ApjL, 636, L149

Chiang E. I., Goldreich P., 1997, ApJ, 490, 368

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

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

D'Alessio P., Calvet N., Hartmann L., 1997, ApJ, 474, 397

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

Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, Protostars and Planets V, 607

Durisen R. H., Hartquist T. W., Pickett M. K., 2008, Ap&SS, 317, 3

Forgan D., Rice K., 2011, MNRAS, 417, 1928

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

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

Kalas P. et al., 2008, Science, 322, 1345

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

Kratter K. M., Murray-Clay R. A., 2011, ApJ, 740, 1

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

Levin Y., Beloborodov A. M., 2003, ApjL, 590, L33

Lodato G., Clarke C. J., 2011, MNRAS, 413, 2735

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

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

Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348

Mejía A. C., Durisen R. H., Pickett M. K., Cai K., 2005, ApJ, 619, 1098

Meru F., Bate M. R., 2010, MNRAS, 406, 2279

Meru F., Bate M. R., 2011, MNRAS, 410, 559

Nayakshin S., 2010, MNRAS, 408, L36

Nero D., Bjorkman J. E., 2009, ApjL, 702, L163

Paardekooper S.-J., 2012, MNRAS, 421, 3286

Paardekooper S.-J., Baruteau C., Meru F., 2011, MNRAS, 416, L65

Podolak M., Mayer L., Quinn T., 2011, ApJ, 734, 56

Price D. J., 2007, PASA, 24, 159

Rafikov R. R., 2005, ApjL, 621, L69

Rafikov R. R., 2007, ApJ, 662, 642

Rafikov R. R., 2011, ApJ, 727, 86

Rice W. K. M., Forgan D. H., Armitage P. J., 2011, MNRAS, 2111

Rice W. K. M., Lodato G., Armitage P. J., 2005, MNRAS, 364, L56

Rogers P. D., Wadsley J., 2011, MNRAS, 414, 913

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shen S., Wadsley J., Hayfield T., Ellens N., 2010, MNRAS, 401, 727

Stadel J. G., 2001, PhD thesis, UNIVERSITY OF WASHINGTON

Stamatellos D., Maury A., Whitworth A., André P., 2011, MNRAS, 413, 1787

Stamatellos D., Whitworth A. P., 2008, A&A, 480, 879

Toomre A., 1964, ApJ, 139, 1217

Vorobyov E. I., Basu S., 2010, ApJ, 719, 1896

Wadsley J. W., Stadel J., Quinn T., 2004, New A., 9, 137

Zhu Z., Hartmann L., Gammie C., McKinney J. C., 2009, ApJ, 701, 620

Chapter 5

The role of radiative transfer in star formation

5.1 Introduction

The previous three chapters of this thesis have focused on gas-giant planet and brown dwarf formation through the fragmentation of massive protostellar discs. As described, radiative transfer (RT) plays a critical role in this process, as it determines the cooling rate of the disc, as well as the heating rate in the outer disc by stellar irradiation. The heating and cooling rates in a gravitationally unstable disc determine whether or not fragmentation occurs.

In this chapter, we consider the important role that RT plays in the star formation process. Stars are predominately formed in giant molecular clouds (GMCs), and the transition from atomic gas to molecular gas depends on having a sufficient column of gas and dust for effective self-shielding against the galactic FUV field. In the initial phases of turbulent star formation, non-ionizing (IR) radiation released from gravitational contraction in high-density regions leads to the local heating of these regions, suppressing small-scale fragmentation. Once massive stars form, they produce large amounts of ionizing (UV) radiation, which creates HII regions that drive turbulence on the scale of the parent molecular cloud, and possibly disrupt the cloud as a whole, shutting off star formation. For massive stars, radiation pressure significantly affects

accretion. In massive molecular clouds, radiation pressure may also play a role in the cloud's disruption.

Our goal is to give context to the numerical methods presented in this thesis, and demonstrate their importance to simulations addressing fundamental questions concerning star formation. Our methods will be applied to such simulations in the near future. This chapter is not meant to be a comprehensive review of the field, but only a sample of the recent research. We first briefly overview the current understanding of star formation in Section 5.2. Our implementation of radiative transfer in the flux-limited diffusion (FLD) approximation, presented in Section 3.2, is the most common method of including non-ionizing radiation in simulations of star formation, as described in Section 5.3. The role of ionizing feedback from massive stars is presented in Section 5.4; this section gives context to Chapter 6, in which we present methods and tests for our implementation of ionizing RT in Gasoline.

5.2 A primer on star formation

The star formation process spans an incredible range in physical scale, from that of galaxies (kpc), to giant molecular clouds (10's of pc), to individual prestellar cores (0.1 pc), and ultimately, down to stellar radii (10^{-8} pc). Star formation also involves a number of physical effects, all of which are likely important: hydrodynamics (including turbulence), self-gravity, magnetic fields (including non-ideal effects), and radiative transfer. As a consequence, star formation is an active, and rich, field of research. Here, we give only a brief introduction to the subject, to give context to our discussion of the role of radiative transfer. For a more detailed consideration, see the reviews by McKee & Ostriker (2007), Zinnecker & Yorke (2007), and Klessen et al. (2009).

Most stars form in clusters (Lada & Lada 2003) created in GMCs, which are generally defined as having masses greater than 10^4 $M_\odot$. Most of the molecular mass in the Milky Way resides in GMCs with $M > 10^5$ $M_\odot$ (McKee & Ostriker 2007). The

timescale for gravitational collapse is the free-fall time:

$$t_{\rm ff} = \left(\frac{3\pi}{32G\rho}\right)^{1/2}, \tag{5.1}$$

which is about 4 Myr for a typical GMC number density of 100 cm^{-3}. As noted by Zuckerman & Evans (1974), GMCs in free-fall collapse, without substantial feedback, are not consistent with observations, as the resultant star formation rate would greatly exceed the 1 M$_\odot$ yr^{-1} of the Milky Way. Observations of GMCs in the LMC (Blitz et al. 2007) and M33 (Engargiola et al. 2003) indicate typical lifetimes of 10-30 Myr, with a short starless phase of 3-7 Myr.

These observations indicate that GMCs are either supported, and not in free-fall, or that they are dispersed relatively soon after the first stars are formed (or some combination of these factors). Magnetic fields have been proposed as one mechanism of support (Mouschovias & Spitzer 1976); in this picture, athough the cloud as a whole is magnetically-supported, prestellar cores increase in mass through ambipolar diffusion. When the mass of a core has grown large enough, gravitational collapse ensues. However, this basic picture of magnetically-controlled star formation is inconsistent with observed mass-to-magnetic-flux ratios (Crutcher et al. 2009).

Turbulence has been proposed as an alternative method of support for GMCs. The seminal work of Larson (1981) demonstrated that the observed scale-dependence of line-widths is indicative of turbulence, which is supersonic on scales larger than about 0.1 pc. Simulations of MHD turbulence (ie. Ostriker et al. 2001) show that it produces a range of density structures, specifically a log-normal probability-distribution function, for both the volume density and the column density.

One major goal of any theory of star formation is to explain the initial mass function (IMF) of stars, the number of stars formed per bin of stellar mass. The observed IMF shows a power-law at the high-mass end (Salpeter 1955), as well as a turnover at lower masses (~ 0.2 M$_\odot$; Chabrier 2003). In the turbulent fragmentation picture of star formation, overdensities are formed by multiple supersonic shocks. Some of these are transient, while others becoming self-gravitating prestellar cores that collapse to form stars. The core mass function (CMF) expected from turbulent

fragmentation has a similar form to the observed IMF: a high-mass power-law with turnover at a fraction of a solar mass (Padoan & Nordlund 2002). Observations also indicate that the CMF has a similar form to the IMF (ie. Reid & Wilson 2006; Alves et al. 2007), but with a higher turnover mass. A one-to-one transition from core to star, with an efficiency of 30% has been proposed (Alves et al. 2007), although this factor may simply be the result of projection effects (Ward et al. 2012). In addition, a one-to-one correspondence between core and star neglects the observed high rate of stellar multiplicity (Duquennoy & Mayor 1991), which would result from the small-scale fragmentation of cores.

The model of turbulent fragmentation predicts that only a small percentage of the total mass of a GMC is at high enough densities to be capable of star formation. However, turbulence decays on a sound-crossing time (roughly the free-fall time; Ostriker et al. 2001). It cannot, by itself, prevent the free-fall collapse of the GMC, and the resulting unrealistically high star formation efficiencies, unless there is a driving mechanism. In addition, when stars form, there is resultant feedback on the GMC in the form of protostellar outflows, radiation pressure, and HII regions created from the ionizing radiation of massive stars.

Turbulence in GMCs is consistent with driving at the largest scale (McKee & Ostriker 2007). Turbulence can be driven externally by, for example, cloud-cloud collisions (Tasker & Tan 2009); or internally, by stellar feedback. Protostellar outflows are unable to drive turbulence on the scale of GMCs (Matzner 2007); however, HII regions can operate on the largest scale (Matzner 2002). In addition to driving turbulence, analytic models suggest that the expansion of HII regions can act to disrupt the GMC as whole, unbinding it (Matzner 2002; Krumholz et al. 2006). Recent work has also suggested that momentum feedback of radiation pressure on dust grains plays an important role in disrupting GMCs, particularly the most massive (Krumholz & Matzner 2009; Murray et al. 2010).

5.3 The effects of non-ionizing feedback

The characteristic mass for gravitational fragmentation in a uniform medium is the Jeans mass:

$$M_{\rm Jeans} = \left(\frac{\pi c_{\rm s}^2}{G\rho}\right)^{1/2} \propto \left(\frac{T}{\rho}\right)^{1/2}; \tag{5.2}$$

only regions with $M > M_{\rm Jeans}$ collapse when perturbed. In addition, the stability of rotating systems to gravitational fragmentation is determined by the Toomre Q:

$$Q = \frac{c_{\rm s}\Omega}{\pi G\Sigma} \propto \sqrt{T}, \tag{5.3}$$

where we have assumed that the rotation is Keplerian. From the above, we can see that both the typical mass-scale for gravitational fragmentation, as well as the stability of discs to fragmentation are sensitive to the thermodynamics of the gas.

During the star formation process, gravitational contraction and shocks lead to the heating of gas and dust to 10's to 100's K, in comparison to the ambient temperature of 10-20 K. This excess thermal energy is radiated away, predominately in the IR, where dust grains are the dominant sources of opacity. This radiation heats the surrounding gas, raising the Jeans mass and preventing fragmentation of low-mass objects such as brown dwarfs. In addition, this radiative heating acts to stabilize the circumstellar disc of a protostar, decreasing the likelihood of fragmentation into low-mass objects. The luminosity of protostars from accretion and fusion contribute important radiative feedback that limits the subsequent fragmentation of the nearby material, including the protostar's disc and nearby filaments.

Only recently have 3D radiation hydrodynamics simulations of star formation in molecular clouds and protostellar cores been carried out, demonstrating the importance of the suppression of fragmentation from IR radiative heating. Most of this work has used FLD (see Section 3.2) to model the transport of IR radiation interacting with dust grains (work by the Berkeley group, led by Mark Krumholz, Richard Klein, and Chris McKee; Matthew Bate; Benoit Commerçon; while work by Thomas Peters, discussed in section 5.4, includes IR heating from protostars via ray-tracing, coupled to molecular-line cooling). FLD is expected to be a good approximation in optically

thick regions, including prestellar cores, or high-mass star forming clumps (Krumholz et al. 2012). In addition, these works assume that the gas and dust temperatures are well coupled, which is strictly only true in the high-density regions, requiring $n \gtrsim 10^8$ for 100 K gas, or $n \gtrsim 10^5$ for 10 K gas (Offner et al. 2009). The simulations of Matthew Bate do not include feedback on scales smaller than the sink radius, and are thus lower-limits on the ability of IR radiative heating to reduce fragmentation.

Simulations with non-ionizing RT of low-mass clouds [50 $M_\odot$ by Bate (2009) and 185 $M_\odot$ by Offner et al. (2009)], high-mass cores [100 $M_\odot$ by Krumholz et al. (2007)], as well as larger cluster-forming clumps [500 $M_\odot$ by Bate (2012) and 10^3 $M_\odot$ by Krumholz et al. (2012)] all show the same general effect of the radiative heating. Small scale fragmentation in filaments and discs near protostars is suppressed, as expected from the discussion regarding the Jeans mass and Toomre Q. The effect is dramatic: the RHD simulation of Bate (2012) produced only 10% of the brown dwarfs as a barotropic simulation of the same initial condition. Although fewer objects are formed, roughly the same mass is turned into stars, so that the typical stellar mass is increased. Simulations of the fragmentation of a 120 $M_\odot$ prestellar core by Commerçon et al. (2011) have demonstrated the manner in which radiative feedback and magnetic fields couple to suppress fragmentation. Strong magnetic fields can lead to effective magnetic braking, which increases the rate of accretion onto a protostar. Consequently, there is greater radiative feedback, and fragmentation is more highly suppressed than in a system with weak magnetic fields.

The simulations of cluster-forming clumps (Bate 2012; Krumholz et al. 2012) show that the inclusion of non-ionizing RT yields IMFs that are consistent with the observed. Interestingly, Krumholz et al. (2012) found that as star formation progressed in their simulation, the turnover mass shifted to higher masses, as continued accretion drove increased heating, leading to an increasing Jeans mass. The authors found that the addition of turbulence on a scale larger than the clump, and to a lesser extent, the inclusion of protostellar outflows, were necessary to match the turnover mass of their IMF to the observed.

In addition to affecting subsequent fragmentation, radiation from a massive protostar can strongly affect further accretion onto the star. For stars more massive than

about 20 $M_{\odot}$, the radiation pressure from photons on the dusty infalling envelope is enough to halt spherically-symmetric accretion, limiting the maximum stellar mass. Recent 3D RHD simulations (Krumholz et al. 2009) have demonstrated that accretion is not, in fact, halted by the radiation pressure because of multi-dimensional effects. Accretion can continue through the disc that is formed during the collapse of the prestellar core, and through filaments created from Rayleigh-Taylor instabilities. This latter effect may be enhanced by the use of frequency-integrated FLD, and may require better treatments to properly follow the accretion (Kuiper et al. 2012). Effective accretion via a disc has also been observed in the 2D and 3D RHD simulations, using ray-tracing coupled to FLD, of massive star formation carried out by Kuiper et al. (2010, 2011).

5.4 The effects of ionizing feedback

The simulations described above show that non-ionizing radiative heating primarily affects local regions of star formation, on the scale of 0.01-0.3 pc. In addition to the IR photons responsible for such radiative heating, massive stars produce large numbers of ionizing, UV, photons. These photons produce H II regions of ionized hydrogen, with characteristic temperatures of roughly 10^4 K, and sound speeds of roughly 10 km/s. Analytic models of HII regions in star-forming GMCs (Matzner 2002; Krumholz et al. 2006) indicate that HII regions dominate the energetics of the GMC, are able to drive turbulence on the scale of the GMC, thereby keeping the star formation efficiency low, and are ultimately able to disrupt GMCs on timescales of 10-30 Myr, consistent with observed lifetimes. In considering star-cluster formation on the scale of a GMC, it is fundamentally important to consider the effects of ionizing radiation. We note that for the largest GMCs, particularly those in starburst galaxies, analytic models indicate that radiation pressure from both non-ionizing and ionizing photons may play a dominant role in disruption (Krumholz et al. 2009; Murray et al. 2010). Observations of the giant HII region 30 Doradus indicate that both radiation pressure and HII gas pressure are important in driving the expansion (Lopez et al. 2011).

The expansion of a massive star's HII region into a uniform medium occurs in two stages (Spitzer 1978). In the first stage, UV photons create an ionization front, inside of which the gas is ionized and outside of which it is neutral. This front expands supersonically, in comparison to both the ionized gas and the neutral gas; as a consequence, the density remains unchanged during this phase. When the radius of the ionization front approaches the Strömgren radius, inside of which recombinations balance ionizations, the increased gas pressure of the 10^4 K gas becomes important, and there is a second stage, of pressure-driven expansion. In this stage, the HII region expands supersonically into the surrounding gas, sweeping up a thin shell of material.

The initial expansion of HII regions is more complicated than the classical picture presented above. Observations indicate that HII regions exist in an embedded, ultra-compact, phase for roughly 4×10^5 years (Churchwell 2002). This is much longer than would be expected for the classical picture of HII expansion. Keto (2002) found that the addition of the star's gravity and continued accretion to the analytic model significantly modifies the initial expansion of the HII region. Initially, the star is unable to sustain an HII region. As its mass grows and it produces more ionizing photons, an HII region develops; however, the escape velocity from the star exceeds the sound-speed of the ionized gas, and the HII region remains "gravitationally trapped." After sufficient mass-growth, the star is finally able to drive a pressure-expanding HII region. The inclusion of accretion along a disc results in the production of bipolar outflows (Keto 2007).

Only recently have 3D simulations of star-cluster formation that take into account ionizing feedback and self-gravity been performed. Peters et al. (2010) performed 3D simulations of the collapse of a 1000 $M_\odot$ molecular cloud 0.5 pc in size, including self-gravity, hydrodynamics, an approximate treatment for non-ionizing radiation, and ionizing radiation using a ray-tracing scheme. As expected, the growth of HII regions from the resulting massive protostars, over the simulated time of 0.75 Myr, differs from the classical picture. Continued accretion onto the massive stars leads to asymmetric, and unsteady expansion of the HII regions. In fact, HII regions were observed to contract, rather than expand, during periods of increased accretion onto a star. From the simulation results, the authors were able to successfully reproduce

the observed morphological classes of ultra-compact HII regions. This study did not, however, include turbulence, nor did it address the effects of ionizing feedback at the scale of the GMC.

Dale et al. (2007) have included an approximate treatment for feedback from ionizing radiation in their SPH code. The method is a Strömgren volume method, in which SPH particles determine the flux of ionizing radiation received from nearby sources by a radial integral of the density distribution along the line of sight, taking into account the ionization of neutral gas, and recombinations in the ionized gas. Based on the calculated flux, gas may be declared to be ionized, and heated to 10^4 K. The heating from non-ionizing radiation is treated only approximately using a barotropic equation of state. HII region expansion also differs from the classical picture on the largest scales because of the turbulent, self-gravitating nature of GMCs. Dale et al. (2012) used the Strömgren volume method described to examine the effects of ionizing feedback for a number GMCs with the same ratio of turbulent to gravitational energy, but different mean densities. The clouds, with masses of $10^4 - 10^6$ $M_\odot$ were initially of uniform density, with a turbulent velocity field. Star-cluster formation, and the resulting ionizing feedback, occurred relatively quickly. The authors found that the ionizing feedback did have large dynamical effects, with 1-10% of the MC mass ionized, and 1-60% becoming unbound within 3 Myr; the effects of the feedback decreased for the increasingly massive clouds. However, the overall star formation efficiency over this period was not largely affected by ionization.

In this discussion, we have focused on the ability of HII regions to disrupt GMCs and limit the star formation efficiency. However, HII region expansion has also been suggested as a means of triggering star formation through compression of the surrounding gas (Elmegreen & Lada 1977). Dale et al. (2012) suggest that the negative effects of ionizing radiation may be offset by triggered star formation in their runs, although more study is required.

The work of Dale et al. (2012) suggests that the effects of ionizing feedback on GMC evolution are more complicated than in the analytic models; in particular, limiting the star formation efficiency is more difficult than expected. Other effects are likely also important, including stellar winds and protostellar jets (Nakamura &

Li 2007), which are effective forms of feedback on smaller scales; as well as magnetic fields, which may increase the energy input of HII regions (Gendelev & Krumholz 2012). For the most massive clusters, whose escape velocities exceed the sound speed of ionized gas, radiation pressure may play an important role in GMC disruption. Supernovae will inject considerable energy into the GMC roughly 4 Myr after star formation.

Harper-Clark (2011) performed grid-based simulations of GMCs of 2×10^5 and 8×10^5 $M_\odot$ and included ionizing radiation (using a ray-tracing method), radiation pressure, supernovae, and magnetic fields. In contrast to the results of Dale et al. (2012), the authors found that feedback was effective at limiting the star formation efficiency (to 5-21% of the GMC mass) and disrupting the GMC. A number of simulations were carried out, varying the feedback mechanisms that were included. The results show that the gas-pressure from HII regions was the primary cause of the limited star formation efficiency. More simulations are thus required in order to determine the effectiveness of feedback to limit star formation and disrupt GMCs. More of the GMC parameter space must be explored for longer times, using higher resolution, and less approximative methods.

5.5 Conclusions

Radiative transfer is fundamentally important in our understanding of star formation, and a necessary ingredient in numerical simulations. On the small scales, heating from non-ionizing IR radiation increases the local Jeans mass, and stabilizes protostellar discs, limiting the production of brown dwarfs and influencing the low-mass end of the IMF. Massive stars ionize gas, and drive HII regions which unbind GMCs and are a source of turbulence. Radiation pressure affects the accretion of high-mass stars, and may also unbind GMCs, particularly the most massive. Our implementation of FLD in Section 3.2 is suitable for modelling non-ionizing radiation. In the following chapter, we present methods and tests for ionizing radiation.

Bibliography

Alves J., Lombardi M., Lada C. J., 2007, A&A, 462, L17

Bate M. R., 2009, MNRAS, 392, 1363

Bate M. R., 2012, MNRAS, 419, 3115

Blitz L., Fukui Y., Kawamura A., Leroy A., Mizuno N., Rosolowsky E., 2007, Protostars and Planets V, 81

Chabrier G., 2003, PASP, 115, 763

Churchwell E., 2002, ARA&A, 40, 27

Commerçon B., Hennebelle P., Henning T., 2011, ApjL, 742, L9

Crutcher R. M., Hakobian N., Troland T. H., 2009, ApJ, 692, 844

Dale J. E., Ercolano B., Bonnell I. A., 2012, ArXiv e-prints: 1205.0360

Dale J. E., Ercolano B., Clarke C. J., 2007, MNRAS, 382, 1759

Duquennoy A., Mayor M., 1991, A&A, 248, 485

Elmegreen B. G., Lada C. J., 1977, ApJ, 214, 725

Engargiola G., Plambeck R. L., Rosolowsky E., Blitz L., 2003, ApJS, 149, 343

Gendelev L., Krumholz M. R., 2012, ApJ, 745, 158

Harper-Clark E., 2011, PhD thesis, UNIVERSITY OF TORONTO (CANADA

Keto E., 2002, ApJ, 580, 980

Keto E., 2007, ApJ, 666, 976

Klessen R. S., Krumholz M. R., Heitsch F., 2009, ArXiv e-prints: 0906.4452

Krumholz M. R., Klein R. I., McKee C. F., 2007, ApJ, 656, 959

Krumholz M. R., Klein R. I., McKee C. F., 2012, ArXiv e-prints: 1203.2620

Krumholz M. R., Klein R. I., McKee C. F., Offner S. S. R., Cunningham A. J., 2009, Science, 323, 754

Krumholz M. R., Matzner C. D., 2009, ApJ, 703, 1352

Krumholz M. R., Matzner C. D., McKee C. F., 2006, ApJ, 653, 361

Kuiper R., Klahr H., Beuther H., Henning T., 2010, ApJ, 722, 1556

Kuiper R., Klahr H., Beuther H., Henning T., 2011, ApJ, 732, 20

Kuiper R., Klahr H., Beuther H., Henning T., 2012, A&A, 537, A122

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57

Larson R. B., 1981, MNRAS, 194, 809

Lopez L. A., Krumholz M. R., Bolatto A. D., Prochaska J. X., Ramirez-Ruiz E., 2011, ApJ, 731, 91

Matzner C. D., 2002, ApJ, 566, 302

Matzner C. D., 2007, ApJ, 659, 1394

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

Mouschovias T. C., Spitzer, Jr. L., 1976, ApJ, 210, 326

Murray N., Quataert E., Thompson T. A., 2010, ApJ, 709, 191

Nakamura F., Li Z., 2007, ApJ, 662, 395

Offner S. S. R., Klein R. I., McKee C. F., Krumholz M. R., 2009, ApJ, 703, 131

Ostriker E. C., Stone J. M., Gammie C. F., 2001, ApJ, 546, 980

Padoan P., Nordlund Å., 2002, ApJ, 576, 870

Peters T., Banerjee R., Klessen R. S., Mac Low M.-M., Galván-Madrid R., Keto E. R., 2010, ApJ, 711, 1017

Reid M. A., Wilson C. D., 2006, ApJ, 644, 990

Salpeter E. E., 1955, ApJ, 121, 161

Spitzer L., 1978, Physical processes in the interstellar medium

Tasker E. J., Tan J. C., 2009, ApJ, 700, 358

Ward R. L., Wadsley J., Sills A., Petitclerc N., 2012, ArXiv e-prints: 1207.5535

Zinnecker H., Yorke H. W., 2007, ARA&A, 45, 481

Zuckerman B., Evans, II N. J., 1974, ApjL, 192, L149

Chapter 6

Ionizing radiation hydrodynamics with OTVET: methods and tests

6.1 Introduction

In this chapter, we present methods and tests of our implementation of ionizing radiative transfer (RT) in the smoothed particle hydrodynamics (SPH) code Gasoline (Wadsley et al. 2004). Our implementation closely follows the work of Petkova & Springel (2009), which was the first to include RT in SPH using the Optically Thin Variable Eddington Tensor (OTVET) formulation of Gnedin & Abel (2001).

In star formation, non-ionizing radiation is continually reprocessed by dust grains, creating a diffuse radiation field that is well-treated by methods such as flux-limited diffusion. In contrast, ionizing photons are emitted by massive stars, and are directly absorbed by neutral hydrogen. Most recombinations lead to excited states of hydrogen, and produce non-ionizing IR photons, but a significant number of recombinations (40%, Osterbrock & Ferland 2006) do lead directly to the ground state, producing ionizing photons. However, if the mean-free path of the ionizing recombination photons is small compared to the size of the HII region, then it is reasonable not to model the resulting diffuse UV field, but to assume that these photons are absorbed locally (the on-the-spot approximation, Osterbrock & Ferland 2006). The

approximation works well for simple geometries, but results in unrealistically sharp shadows when inhomogoneities are present (Ercolano & Gritschneder 2011). With this commonly employed approximation, it is the direct emission from nearby sources (and absorptions along the line of sight) that determines the ionizing radiation field.

Consequently, most treatments of ionizing radiation use a ray-tracing scheme that solves the equation of radiative transfer along the line-of-sight from star to gas (Mellema et al. 2006; Rijkhorst et al. 2006; Wise & Abel 2011). Ray-tracing has previously been implemented in SPH codes (Pawlik & Schaye 2008; Gritschneder et al. 2009). However, given the complexity of adding ray-tracing to Gasoline, and the previous work done for our implementation of FLD, we have chosen to use the OTVET method in Gasoline. OTVET was designed for modelling ionizing RT, in which the radiation field is dominated by discrete sources. It represents a more direct, and faster, path to including ionizing RT in Gasoline. As a benefit, OTVET was designed for efficient modelling of RT in cosmological simulations with large numbers of sources. Our implementation will consequently be very useful for the cosmological simulations performed in our research group.

6.2 The Variable Eddington Tensor formulation of radiative transfer

The OTVET formulation was originally developed in order to include the effects of radiative transfer on galaxy formation in cosmological simulations. We therefore begin by considering the equation of radiative transfer including cosmological terms. In the context of this thesis, however, our aim is to use the OTVET formulation to consider the ionizing radiation from massive stars in star formation simulations at the scale of molecular clouds. Consequently, we quickly drop the cosmological terms, though they may easily be included in our implementation in the future. Descriptions of the OTVET method can also be found in Gnedin & Abel (2001) and Petkova & Springel (2009).

6.2.1 The equation of radiative transfer and its moments

The equation of radiative transfer in a comoving frame for an expanding universe is (Gnedin & Ostriker 1997):

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \frac{\mathbf{n}}{a}\frac{\partial I_\nu}{\partial \mathbf{x}} - \frac{H}{c}\left(\nu\frac{\partial I_\nu}{\partial \nu} - 3I_\nu\right) = -\kappa_\nu I_\nu + j_\nu, \tag{6.1}$$

where ν is the frequency, I_ν is the frequency-dependent intensity, κ_ν is the frequency-dependent absorption coefficient, j_ν is the frequency-dependent emission coefficient, a is the cosmological scale-factor, $H = \dot{a}/a$ is the Hubble constant, $\mathbf{n}$ is a direction vector, and c is the speed of light.

A common approach in solving the equation of RT is to consider the equation's moments, as well as the corresponding moments of the intensity, I_ν. The first three moments of I_ν are the mean-intensity,

$$J_\nu = \oint I_\nu d\Omega, \tag{6.2}$$

the flux,

$$F^i_\nu = \oint n^i I_\nu d\Omega, \tag{6.3}$$

and the radiation pressure tensor,

$$P^{ij}_\nu = \oint n^i n^j I_\nu d\Omega, \tag{6.4}$$

where Ω is the unit of solid-angle and the indices i and j run through the three elements of a vector in Cartesian space. The Eddington tensor, h^{ij}, defines the relationship between the radiation pressure tensor and the mean intensity:

$$P^{ij}_\nu = h^{ij}_\nu J_\nu. \tag{6.5}$$

We now drop the third and fourth terms on the left-hand side of equation (6.1), which represent the effects of the expansion of the universe, and set the scale factor

to $a = 1$. We write the first two moments of the RT equation:

$$\frac{1}{c}\frac{\partial J_\nu}{\partial t} + \frac{1}{a}\frac{\partial F^i_\nu}{\partial x^i} = -\kappa_\nu J_\nu + j_\nu, \tag{6.6}$$

and

$$\frac{1}{c}\frac{\partial F^j_\nu}{\partial t} + \frac{1}{a}\frac{\partial h^{ij} J_\nu}{\partial x^i} = -\kappa_\nu F^j_\nu. \tag{6.7}$$

We ignore the term of c^{-1} in the second-moment RT equation; the flux is then

$$F^j_\nu = -\frac{1}{\kappa_\nu}\frac{\partial h^{ij} J_\nu}{\partial x^i}. \tag{6.8}$$

Since the radiation energy density is $E_\nu = 4\pi J_\nu / c$, this form of the flux is similar to the flux-limited diffusion approximation, equation (3.8), used in Chapter 3. The difference in this case is that the direction of propagation of the flux in equation (6.8) is determined by the Eddington tensor, rather than solely by the gradient of the radiation energy density. The form of equation (6.8) can be described as anisotropic diffusion.

If we substitute this form of the flux into the first-moment RT equation, (6.6), then we can solve for the rate-of-change of the mean intensity:

$$\frac{\partial J_\nu}{\partial t} = c\frac{\partial}{\partial x_j}\left(\frac{1}{\kappa_\nu}\frac{\partial h^{ij} J_\nu}{\partial x^i}\right) - c\kappa_\nu J_\nu + c j_\nu, \tag{6.9}$$

This equation can be used to determine the time evolution of the radiation field, but only if we know the form of the Eddington tensor, h^{ij}.

6.2.2 The Optically Thin Variable Eddington Tensor formulation

Gnedin & Abel (2001) proposed closing the approximate first-moment RT equation, (6.9), assuming an optically-thin form for the Eddington tensor. For an optically-thin

medium, the radiation pressure tensor can be computed from

$$P^{ij}(\mathbf{x}) \propto \int d^3x' \rho_{\text{source}}(\mathbf{x}') \frac{(\mathbf{x}-\mathbf{x}')_i (\mathbf{x}-\mathbf{x}')_j}{(\mathbf{x}-\mathbf{x}')^4}, \tag{6.10}$$

and the Eddington tensor as

$$h^{ij}(\mathbf{x}) = P^{ij}(\mathbf{x}) / \text{Tr}(P(\mathbf{x})). \tag{6.11}$$

In the above, ρ_{source} is the density of photons produced at a location from, for example, a massive star.

6.3 Solution of the first-moment RT equation in SPH

6.3.1 The first-moment RT equation in SPH

The first-moment RT equation, (6.9), specifies the time-evolution of the frequency-dependent mean intensity, J_ν. A complete multi-frequency approach would require solving this equation for a number of frequency ranges, $d\nu$, that cover the spectrum. Since RT is computationally expensive, one generally focuses on solving equation (6.9) only for a small number of physically important frequency ranges. In this work, we consider a narrow frequency range of radiation corresponding to the ionization energy of atomic hydrogen, $h\nu_0 = 13.6$ eV, such that the mean-intensity has the form

$$J_\nu = J_0 f(\nu), \tag{6.12}$$

where $f(\nu)$ is a function describing the typical spectrum near ν_0. It is straightforward to generalize this approach to multiple frequency ranges, such as those corresponding to the ionization energies of helium, and the dissociation energy of molecular hydrogen.

To solve the first-moment RT equation in SPH requires an appropriate choice of variables. For the narrow frequency range that we are considering, the number

density of ionizing photons is

$$n_\gamma = \frac{1}{c}\int_\nu \frac{4\pi J_\nu}{h\nu} d\nu = \frac{1}{c}\frac{4\pi J_0}{h\nu_0}. \tag{6.13}$$

The total number density of hydrogen atoms, including neutral, $n_{\rm HI}$, and ionized, $n_{\rm HII}$, components is

$$n_{\rm H} = n_{\rm HI} + n_{\rm HII} = \frac{X_{\rm H}\rho}{m_{\rm H}}, \tag{6.14}$$

where $X_{\rm H}$ is the mass fraction of hydrogen, ρ is the density, and $m_{\rm H}$ is the mass of a Hydogren atom. SPH uses Lagrangian particles of fixed mass, but evolving volume. Therefore, rather than consider the number density of photons, it is more useful to consider a scaled photon fraction that is independent of volume, $\tilde{n}_\gamma = n_\gamma/n_{\rm H}$. With this choice of variables, the first-moment equation of RT, (6.9), upon multiplication by $4\pi/ch\nu$ and integration over ν, becomes:

$$\frac{\partial \tilde{n}_\gamma}{\partial t} = c\frac{\partial}{\partial x_j}\left(\frac{1}{\bar{\kappa}}\frac{\partial \tilde{n}_\gamma h^{ij}}{\partial x^i}\right) - c\bar{\kappa}\tilde{n}_\gamma + \tilde{s}_\gamma. \tag{6.15}$$

In the above, $\tilde{s}_\gamma$ is the production rate of ionizing photons (scaled by $n_{\rm H}$ to effectively be per unit mass) from, for example, nearby massive stars. If there is a fixed rate of production of ionizing photons, $\dot{N}_\gamma$, then, for an SPH particle of mass m, the source term is

$$\tilde{s}_\gamma = \dot{N}_\gamma \frac{m_{\rm H}}{m}. \tag{6.16}$$

The frequency-averaged absorption coefficient, $\bar{\kappa}$, is defined through the loss of photons; for our narrow frequency range,

$$\bar{\kappa}\tilde{n}_\gamma = \frac{1}{n_{\rm H}}\int_\nu \frac{4\pi}{ch\nu}\kappa_\nu J_\nu d\nu = \tilde{\sigma}\tilde{n}_\gamma \tilde{n}_{\rm HI} n_{\rm H}, \tag{6.17}$$

where $\tilde{\sigma}$ is the frequency-averaged absorption cross-section of neutral hydrogen. For frequencies with $\nu > \nu_0$, the cross section of neutral hydrogen is

$$\sigma_\nu = \sigma_0 \left(\frac{\nu_0}{\nu}\right)^4 \frac{\exp\left(4 - 4\arctan\left[\epsilon\right]/\epsilon\right)}{1 - \exp\left(-2\pi/\epsilon\right)}, \tag{6.18}$$

where $\sigma_0 = 6.30 \times 10^{-18}$ cm^2 is the cross-section at the resonance frequency, ν_0, and $\epsilon = \sqrt{\nu/\nu_0 - 1}$ (Osterbrock & Ferland 2006). From the above, one can see that absorption will be strongly peaked near ν_0. The frequency-averaged absorption coefficient, and the frequency-averaged absorption cross-section also depend on the spectrum of photons in the frequency range of interest.

The right hand side of the first-moment RT equation, (6.15), has three terms with straightforward physical interpretations: the first term is an anisotropic diffusion term that governs the conservative transport of photons, the second term governs the absorption of photons, and the third term governs the production of photons.

As with FLD, we introduce a flux-limiter, λ, to ensure that radiation does not propagate faster than the speed of light in optically thin regions, where $\bar{\kappa}$ becomes small. With this introduction, the first-moment equation is

$$\frac{\partial \tilde{n}_\gamma}{\partial t} = c\frac{\partial}{\partial x_j}\left(\frac{\lambda}{\bar{\kappa}}\frac{\partial \tilde{n}_\gamma h^{ij}}{\partial x^i}\right) - c\bar{\kappa}\tilde{n}_\gamma + \tilde{s}_\gamma. \tag{6.19}$$

To compare directly with the test results of Petkova & Springel (2009), we use a similar flux-limiter

$$\lambda\left(R\right) = \frac{1 + 0.1R}{1 + 0.1R + 0.1R^2}, \tag{6.20}$$

with

$$R = \frac{|\nabla \tilde{n}_\gamma|}{\bar{\kappa}\tilde{n}_\gamma}. \tag{6.21}$$

However, we note that the above flux-limiter is slightly different from that presented by Petkova & Springel (2009):

$$\lambda_{\text{P2009}}\left(R\right) = \frac{1 + 0.1R}{1 + 0.1R + 0.01R^2}. \tag{6.22}$$

This is to ensure that the correct limits for the optically thick case, $\lambda \to 1$ as $R \to 0$, and the optically thin case, $\lambda \to 1/R$ as $R \to \infty$, are satisfied. Since the test results in Petkova & Springel (2009) compare well with our tests in Section 6.5, we assume that there is an error in the paper, and not in their implementation.

To solve the above equation within the SPH framework, we make a final change

of variables to consider

$$N_i = m_i \tilde{n}_{\gamma,i}, \tag{6.23}$$

where the subscript i denotes the $i^{\rm th}$ SPH particle, in order to account for particles of differing mass. N is the number of photons per SPH particle, to within a constant factor. Particles of differing mass are not commonly used in SPH star-formation simulations; however, they are common in cosmological simulations. We therefore make this change of variables to maintain the generality of the implementation.

As discussed in Chapter 3, first derivatives in SPH are straightforward, as they require only the derivative of the kernel function W; however, second-derivatives taken in a similar fashion are noisy due to particle disorder. In our implementation of FLD, we used the approximation of Cleary & Monaghan (1999) for an isotropic second-derivative. Petkova & Springel (2009) developed a similar approximation for an anisotropic second-derivative. For a SPH particle, i, with neighbours denoted by j, the anisotropic derivative of the tensor $\mathbf{Q}$ is:

$$\frac{\partial}{\partial x_\alpha}\left(\frac{1}{s}\frac{\partial Q_{\alpha\beta}}{\partial x_\beta}\right)\bigg|_{\mathbf{x}_i} = 2\sum_j \frac{1}{s_{ij}} \frac{\mathbf{x}_{ij}^{\rm T}\left[\tilde{\mathbf{Q}}_j - \tilde{\mathbf{Q}}_i\right]\nabla_i W_{ij}}{|\mathbf{x}_{ij}|^2}\frac{m_j}{\rho_{ij}}. \tag{6.24}$$

In the above, s and ρ are symmetrized between neighbour-particles, in the form $1/s_{ij} = 1/2(1/s_i + 1/s_j)$, and $\mathbf{x}_{ij} = \mathbf{x}_j - \mathbf{x}_i$. For the fully anisotropic derivative,

$$\tilde{\mathbf{Q}} = \frac{5}{2}\mathbf{Q} - \frac{1}{2}{\rm Tr}\left(\mathbf{Q}\right)\mathbf{I}. \tag{6.25}$$

However, Petkova & Springel (2009) note that this may lead to numerical instability, as there is the possiblity of anti-diffusion of photons from a particle with fewer photons to a particle with more photons. Because of this, they also consider an anisotropy-limited second-derivative, in which equation (6.24) is evaluated with $\tilde{\mathbf{Q}} = \mathbf{Q}$. In our testing, the full-anisotropy derivative has led to numerical instability; therefore, we use only the anisotropy-limited derivative in the rest of this chapter.

If we symmetrize the exchange of photons between neighbour-particles, then we

can rewrite the first-moment equation of RT in the SPH formalism as

$$\frac{\partial N_i}{\partial t} = \sum_j w_{ij} \left(N_j - N_i\right) - c\bar{\kappa}_i N_i + m_i \tilde{s}_{\gamma,i}, \tag{6.26}$$

where

$$w_{ij} = \frac{2c}{\bar{\kappa}_{ij}} \frac{m_{ij}}{\rho_{ij}} \frac{\mathbf{x}_{ij}^{\mathrm{T}} \tilde{\mathbf{h}}_{ij} \nabla_i W_{ij}}{|\mathbf{x}_{ij}|^2}. \tag{6.27}$$

As mentioned, the subscript ij denotes a symmetric combination, except in the case of $\mathbf{x}$, in which it represents the distance vector between the particles.

6.3.2 Time-integration of the RT equation

The time integration of the first-moment equation of RT, (6.26), suffers from the same difficulties dealt with in the implementation of FLD in Chapter 3: the time-steps required for numerical stability in the explicit time-integration of RT are generally much shorter than the time-steps of interest in hydrodynamics. We take the same approach as with FLD, and solve the first-moment equation of RT using a backwards-Euler time-integration, which has the general form

$$y_i^{n+1} = y_i^n + \Delta t \left[\frac{Dy_i}{Dt}\right]^{n+1}, \tag{6.28}$$

for an integrated variable, y, time-step, Δt, and where the superscript n denotes the beginning of the current time-step. This integration allows for longer time-steps while maintaining numerical stability, but requires an iterative solution.

The backwards-Euler integration of the first-moment RT equation, (6.26), results in a linear-system of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{6.29}$$

in which the matrix elements are

$$A_{ij} = \delta_{ij} \left(1 + \sum_k \Delta t w_{ik} + \Delta t c\bar{\kappa}\right)^{n+1} - w_{ij}^{n+1}, \tag{6.30}$$

where δ_{ij} is the Kronecker delta, the known vector elements are

$$b_i = N_i^n + \Delta t \tilde{s}_{\gamma,i}^{n+1} m_i, \tag{6.31}$$

and the unknown solution vector is

$$x_i = N_i^{n+1}. \tag{6.32}$$

The elements of this sparse linear system of equations are computed within the SPH framework of Gasoline, and the solution vector is computed using the Conjugate-Gradient method within the Portable, Extensible Toolkit for Scientific Computation (PETSc; Balay et al. 1997, 2011a,b).

6.4 Coupling RT to hydrodynamics, heating and cooling, and the chemical network

6.4.1 Operator splitting

We have thus far described the propagation of ionizing photons. However, RT must also be coupled with the other relevant processes, such as hydrodynamics, ionizations and recombinations (the chemical network), and non-hydrodynamic heating and cooling processes. Similar to Petkova & Springel (2009), we adopt an operator-splitting approach, in which the time-integration of these processes is split into a number of steps: 1) acceleration from gravity and the OTVET Eddington tensor are computed; 2) acceleration and heating from hydrodynamics (artificial viscosity and PdV work) are computed; 3) the source term of photons for each particle, $\tilde{s}_{\gamma,i}$, is computed; 4) The photon number density is calculated from the first-moment RT equation, (6.19); and 5) the ionization fractions as well as heating and cooling from non-hydrodynamic processes are computed.

For the test-problems in Section 6.5, we use direct summation to compute the OTVET Eddington tensor via equations (6.10) and (6.11). Direct summation is

reasonable for small numbers of ionizing sources (there is only one in the tests), but scales as $N^2_{\rm sources}$, and can be expensive for large numbers of sources. Since the optically-thin Eddington Tensor has the same $1/r^2$ dependence as gravity, it can be efficiently calculated using a tree-method similar to that used for the gravitational acceleration calculation in Gasoline. OTVET thus scales as $N_{\rm sources} \log N_{\rm sources}$, and is well-suited to simulations with large numbers of sources. In general, for nearby tree-nodes containing sources, direct summation is employed, while for tree-nodes that are farther away, the total luminosity of the tree-node, rather than the individual sources, is used to calculate the Eddington tensor. Specifically, an angle-opening criterion similar to that used for the gravity calculation can be used to determine whether a tree-node must be opened. Calculation of the OTVET Eddington tensor in the gravitational tree-walk is currently being implemented within our research group (R. Woods).

We have also employed a simplified approach for the source terms of ionizing photons for the testing in Section 6.5. We currently specify one source of photons based on the position of the gas particles. In the future, we will calculate source terms based on the positions of the gas particles relative to nearby star/sink particles.

6.4.2 Chemical network

In this work, we consider only hydrogen-ionizing radiation, and therefore also consider a simplified chemical composition consisting only of atomic and ionized hydrogen. For this simplified chemical network, there is one governing equation:

$$\frac{\partial \tilde{n}_{\rm HII}}{\partial t} = c\bar{\sigma} n_{\rm H} \tilde{n}_{\rm HI} \tilde{n}_{\gamma} - \alpha n_{\rm H} \tilde{n}_{\rm e} \tilde{n}_{\rm HII}, \tag{6.33}$$

in which α is the recombination coefficient. We can use the relations $n_{\rm HI} = n_{\rm H} - n_{\rm HII}$ and $n_{\rm e} = n_{\rm HII}$ to close the above equation. The first term on the right hand side is the photoionization rate, while the second term is the recombination rate. For comparison with Petkova & Springel (2009), we have ignored collisional ionization in the above, although in general it is included.

The extension of this network to include other relevant components such as helium

or molecular hydrogen is straightforward, and has previously been incorporated within the cooling framework of Gasoline. In prior work, the photon number density has been based on uniform backgrounds (Shen et al. 2010) or more recently, approximative radiative transfer treatments (Christensen et al. 2012). The difference in this work, is that $\tilde{n}_\gamma$ is computed from the first-moment equation of RT.

6.4.3 Non-hydrodynamic heating and cooling

The Gasoline cooling framework already includes non-hydrodynamic heating and cooling processes such as photoionization heating, radiative recombination, collisional ionization, line-cooling, and bremsstrahlung. The photoionization heating is modified in this work, as it depends on the number density of ionizing photons. The heating rate per unit volume is

$$\Gamma = c\bar{\epsilon}\bar{\kappa}n_{\rm H}\tilde{n}_\gamma, \tag{6.34}$$

in which $\bar{\epsilon}$ is the frequency-averaged energy in excess of 13.6 eV of ionizing photons. We calculate the photoionization heating using the $\tilde{n}_\gamma$ computed from the first-moment equation of RT.

6.4.4 Time integration

The small time-steps required for numerical stability mean that explicit time-integration of the chemical network and internal energy is not feasible within hydrodynamic simulations. Gasoline uses a stiff equation solver to concurrently integrate both the chemical network and the internal energy, assuming that the hydrodynamic heating and the density are constant over the hydrodynamic time-step. Within this cooling framework, the addition of other chemical species, or heating and cooling processes is straightforward.

6.5 Test cases

We present code results for two tests taken from the Cosmological Radiative Transfer Codes Comparison Project (Iliev et al. 2006). Each test consists of a uniform-density

system of pure hydrogen (atomic and ionized), with a single source of ionizing photons. The tests do not include hydrodynamic evolution. The first test examines the expansion of an ionization (HII) front in isothermal gas, and has the benefit of an analytic solution. The second test examines the expansion of an ionization front in non-isothermal gas, for which heating and cooling are considered. This case has no analytic solution, but results can be compared to the those of other RT codes. Since we use methods very close to Petkova & Springel (2009), we use the parameters for the tests as stated by these authors, and compare directly with their results.

6.5.1 The isothermal ionization front

In this test, we consider an initially neutral, uniform density gas with $n_{\rm H} = 10^{-3}$ cm^{-3}, which is ionized by a single source of 13.6 eV photons, with $\dot{N}_\gamma = 5 \times 10^{48}$ photons/s, positioned at the centre of the simulation. The ionization cross-section is thus $\bar{\sigma} = 6.30 \times 10^{-18}$ cm^{-2}, while the recombination rate is taken to be $\alpha = 2.59 \times 10^{-13}$ cm^{-3}s^{-1}, characteristic of the temperature of 10^4 K. The evolution of the system is isothermal, and does not include hydrodynamics.

The Strömgren radius describes the region surrounding the source of photons in which ionizations are balanced by recombinations. If one assumes that the ionization front is sharp, that the region over which the gas transitions from being completely neutral to being completely ionized is infinitesimal, then this radius is

$$r_{\rm S} = \left(\frac{3\dot{N}_\gamma}{4\pi\alpha n_{\rm H}^2} \right)^{1/3} . \tag{6.35}$$

The time-evolution of such an ionization front, of radius $r_{\rm I}$, has the solution

$$r_{\rm I}(t) = r_{\rm S} \left[1 - \exp\left(t/t_{\rm recomb} \right) \right]^{1/3} , \tag{6.36}$$

in which $t_{\rm recomb} = 1/n_{\rm H}\alpha$ is the recombination time (Osterbrock & Ferland 2006). For the parameters of this test, $t_{\rm recomb} = 125$ Myr, and $r_{\rm S}$=5.38 kpc.

The solution for a non-sharp ionization front must be computed numerically from

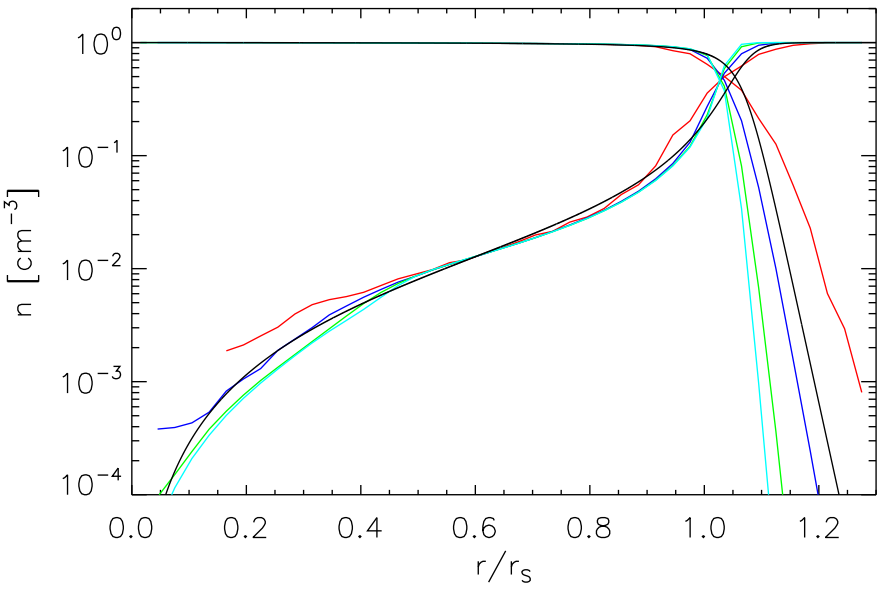

Figure 6.1: The isothermal ionization front test at $t = 4t_{\rm recomb} = 500$ Myr. The spherically-averaged radial ionization profiles, $\tilde{n}_{\rm HI}(r)$ and $\tilde{n}_{\rm HII}(r)$, from Gasoline are given by the (red, blue, cyan, and green) curves for particle numbers of (16^3, 32^3, 64^3, and 128^3) and a timestep of 0.5 Myr. The analytic solution is given by the black line.

analytic considerations (Osterbrock & Ferland 2006). In this case, $r_{\rm S}$ is defined as the radius at which $\tilde{n}_{\rm HI} = \tilde{n}_{\rm HII}$. For a single ionizing source, the radial photon density profile is diluted by geometry and attenuated by absorption:

$$n_\gamma(r) = \frac{\dot{N}_\gamma}{4\pi r^2 c} \exp\left[-\tau(r)\right], \tag{6.37}$$

where the optical depth is

$$\tau(r) = \int_0^r \bar{\sigma} n_{\rm H} \tilde{n}_{\rm HI}(r')\, dr'. \tag{6.38}$$

For a given radial ionization profile, $\tilde{n}_{\rm HI}(r)$, equation (6.37) can be integrated radially outwards from $r = 0$ to determine the radial photon density profile. To determine $r_{\rm S}(t)$, we couple the radial integration of (6.37) to a backwards-Euler time-

integration of the chemical network equation, (6.33), at each radius. We begin at $t = 0$ with an initially neutral medium. For each timestep, we use the previous timestep's $n_\gamma(r)$ to update the current step's $\tilde{n}_{\rm HI}(r)$ using equation (6.33). We then use this to update $n_\gamma(r)$ through the radial integration of equation (6.37). From this process, we compute $r_{\rm S}(t)$.

The equilibrium radial ionization profile can be computed from the above procedure for $t \gg t_{\rm recomb}$. It can also be calculated more directly, from the radial integration of (6.37) out from $r = 0$, taking $\tilde{n}_{\rm HI}(r = 0) = 0$ (the interior of the Strömgren sphere is almost completely ionized) and assuming that ionizations balance recombinations at every radius (using equation (6.33) with the left-hand-size set to zero).

Figure 6.1 shows the spherically-averaged radial ionization profile computed by Gasoline at $t = 4t_{\rm recomb} = 500$ Myr, for particle numbers of (16^3, 32^3, 64^3, and 128^3). One can see that the ionization front becomes sharper, more closely following the analytic solution, for higher resolutions. The code results agree well with the analytic profiles, although they do slightly underestimate $r_{\rm S}$. These results agree well with those of Petkova & Springel (2009).

Figure 6.2 shows the time evolution of the ionization front, $r_{\rm S}(t)$, for a particle number of 64^3 and time-steps of (0.05, 0.5, 5.0, and 50.0) Myr. The backwards-Euler time-integration of the photon transport means that the method is numerically stable for long time-steps. In fact, all of the time-steps used are much longer than the largest explicitly stable time-step of 10^{-3} Myr (Petkova & Springel 2009). The practical criterion for choosing a time-step is not numerical stability, but accuracy, which must be investigated through testing. From the results, shorter time-steps result in more accurate solutions, as expected, with the two smallest time-steps of 0.05 and 0.5 Myr giving results that follow the solution quite accurately. For the 5.0 Myr time-step, the expansion of the ionization front is slower than expected, but after $4t_{\rm recomb}$, $r_{\rm S}$ is within a few percent of the expected result. We find that a time-step of 50 Myr results in a rather inaccurate treatment of the ionization front. After $4t_{\rm recomb}$, $r_{\rm S}$ is only about 40% of the expected value. We find that our two longest time-steps of 5.0 and 50.0 Myr give less accurate solutions than the same time-steps of Petkova & Springel (2009); the reason for this discrepancy is unclear.

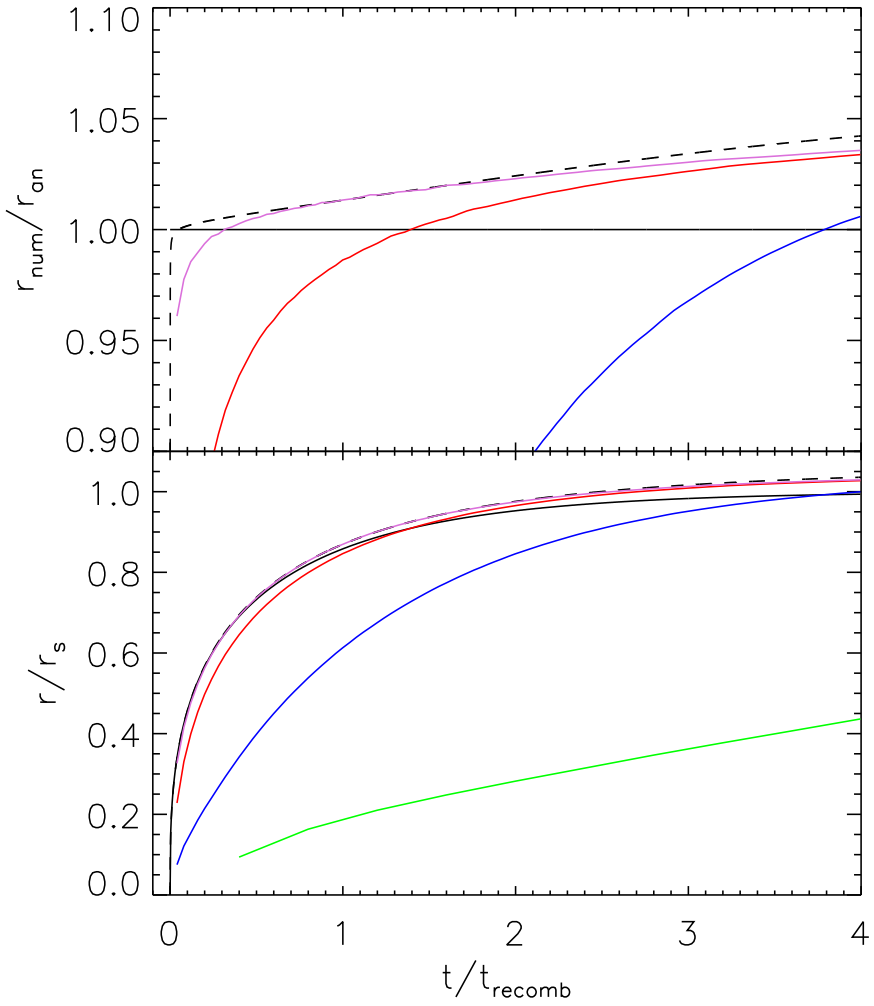

Figure 6.2: The time evolution of the isothermal ionization front. The lower panel shows $r_{\rm S}\,(t)$ computed from Gasoline for a particle number of 64^3. The results for time-steps of (0.05, 0.5, 5.0, and 50.0) Myr are given by the (orchid, red, blue, and green) curves. The analytic solution for a sharp-ionization front is given by the black curve, while the solution for a non-sharp-ionization front is given by the dashed curve. The upper panel shows the ratio of the numerically calculated $r_{\rm S}\,(t)$ to the analytically calculated $r_{\rm S}\,(t)$ for a sharp-ionization front, using the same colour and line conventions as in the lower panel.

6.5.2 The non-isothermal ionization front

We use the same initial setup of the previous test to consider the evolution of a non-isothermal ionization front. The gas temperature is set to initially be 100 K ev-

erywhere, but photoionization heats the gas, which also cools via recombinations, collisional ionizations, collisionally excited line-cooling, and bremsstrahlung. The photoionization heating rate is given by equation (6.34), for which we adopt a frequency-averaged excess energy of $\bar{\epsilon} = 29.65$ eV, and a frequency-averaged cross-section of $\bar{\sigma} = 1.63 \times 10^{-18}$ cm^2. The cooling rates are identical to Petkova & Springel (2009), and are those of Cen (1992). For completeness, cooling rates in erg cm^{-3} s^{-1} for recombination, $\Lambda_{\rm recomb}$, collisional ionization, $\Lambda_{\rm c\text{-}i}$, line-cooling, $\Lambda_{\rm line}$, and bremsstrahlung, $\Lambda_{\rm brem}$, are:

$$\Lambda_{\rm recomb} = 8.7 \times 10^{-27} \sqrt{T} \left(\frac{T}{10^3 \text{ K}}\right)^{-0.2} / \left[1 + \left(\frac{T}{10^6 \text{ K}}\right)^{0.7}\right] n_{\rm e} n_{\rm HII} \tag{6.39}$$

$$\Lambda_{\rm c\text{-}i} = 1.27 \times 10^{-21} \sqrt{T} \left(1 + \sqrt{\frac{T}{10^5 \text{ K}}}\right) \exp\left(\frac{-157809.1}{T}\right) n_{\rm e} n_{\rm HI} \tag{6.40}$$

$$\Lambda_{\rm line} = 7.5 \times 10^{-19} \sqrt{T} \left(1 + \sqrt{\frac{T}{10^5 \text{ K}}}\right)^{-1} \exp\left(\frac{-118348}{T}\right) n_{\rm e} n_{\rm HI} \tag{6.41}$$

$$\Lambda_{\rm brem} = 1.85 \times 10^{-27} \sqrt{T} n_{\rm e} n_{\rm HII}. \tag{6.42}$$

Figure 6.3 shows the evolution of the spherically-averaged radial temperature profile and ionization fraction for times of (20, 100, and 500) Myr and for particles numbers of (16^3, 32^3, and 64^3). Inside of $r_{\rm S}(t)$, the temperature of the gas is in the range $10^4 - 3 \times 10^4$ K. This temperature is the result of the balance between photoionization heating, which alone would heat the gas to $3.5{\times}10^5$ K, and the various cooling processes included. This temperature is an overestimate compared to observed HII regions because of the lack of metal line-cooling here. The gas outside of the ionized region remains near 100 K. The results for the different spatial resolutions have not completely converged, as the higher resolution runs show slightly smaller $r_{\rm S}(t)$ and slightly higher temperatures. However, the differences are not large, and appear to be decreasing with increased resolution. The results are similar to those found by Petkova & Springel (2009).

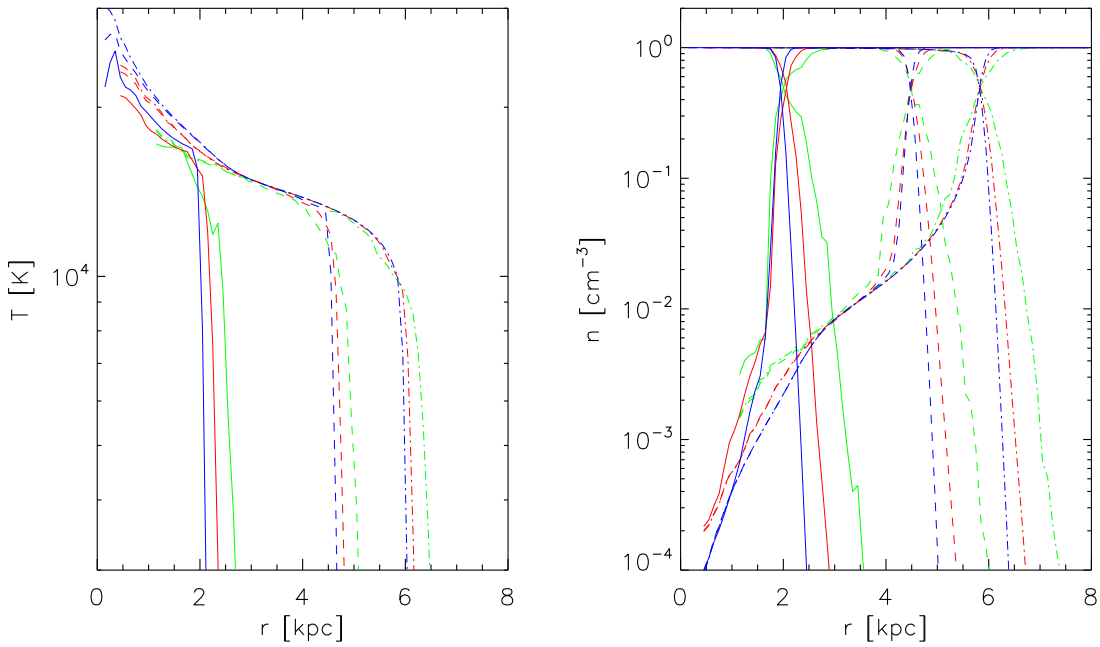

Figure 6.3: The non-isothermal ionization front. The left panel shows the radial temperature profile computed by Gasoline. The results for particle numbers of (16^3, 32^3, and 64^3) are given by the (green, red, and blue) curves. The results at times of (10, 100, and 500) Myr are given by the (solid, dashed, and dot-dashed) curved. The right panel shows the radial ionization profiles, $\tilde{n}_{\rm HI}(r)$ and $\tilde{n}_{\rm HII}(r)$, with the same colour and line conventions as the left panel.

6.6 Conclusions

We have presented methods and tests of our implementation of ionizing radiative transfer in the Gasoline TreeSPH code. The transport of photons is computed by solving the first moment equation of RT, which has been closed using the optically-thin variable Eddington tensor formulation of Gnedin & Abel (2001). Our implementation follows that of Petkova & Springel (2009), who first incorporated this method into SPH. Photon transport has been coupled to the gravity, hydrodynamics, chemical network (ionization fraction), and heating and cooling calculations present in Gasoline. The implementation has been successfully tested on two standard test-problems from Iliev et al. (2006): the isothermal ionization front, and the non-isothermal ion-

ization front.

Bibliography

Balay S. et al., 2011a, PETSc users manual. Tech. Rep. ANL-95/11 - Revision 3.2, Argonne National Laboratory

Balay S. et al., 2011b, PETSc Web page. Http://www.mcs.anl.gov/petsc

Balay S., Gropp W. D., McInnes L. C., Smith B. F., 1997, in Modern Software Tools in Scientific Computing, Arge E., Bruaset A. M., Langtangen H. P., eds., Birkhäuser Press, pp. 163–202

Cen R., 1992, ApJS, 78, 341

Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, ArXiv e-prints: 1205.5567

Cleary P. W., Monaghan J. J., 1999, Journal of Computational Physics, 148, 227

Ercolano B., Gritschneder M., 2011, MNRAS, 413, 401

Gnedin N. Y., Abel T., 2001, New A., 6, 437

Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581

Gritschneder M., Naab T., Burkert A., Walch S., Heitsch F., Wetzstein M., 2009, MNRAS, 393, 21

Iliev I. T. et al., 2006, MNRAS, 371, 1057

Mellema G., Iliev I. T., Alvarez M. A., Shapiro P. R., 2006, New A., 11, 374

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Pawlik A. H., Schaye J., 2008, MNRAS, 389, 651

Petkova M., Springel V., 2009, MNRAS, 396, 1383

Rijkhorst E.-J., Plewa T., Dubey A., Mellema G., 2006, A&A, 452, 907

Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581

Wadsley J. W., Stadel J., Quinn T., 2004, New A., 9, 137

Wise J. H., Abel T., 2011, MNRAS, 414, 3458

Chapter 7

Conclusions

7.1 Summary

It is an exciting era in the study of planet formation. It has been fewer than two decades since the first planet was discovered around another star. A variety of observational approaches have been used to detect an ever-increasing number of extrasolar planets, roughly 3000 confirmed and candidate planets at the time of this writing. These observations have demonstrated the surprising variety of planetary systems, showing that the architecture of our solar system is by no means the rule. In particular, gas-giant companions (planets and brown dwarfs), have been discovered over a wide range of orbital radii from their host stars, from hot Jupiters, with orbits smaller than that of Mercury, to gas-giant companions with orbits larger than 100 AU in radius.

Observations are critical to the verification of theoretical models of gas-giant formation. In addition, observations of the birthplaces of gas-giants, protostellar discs composed of gas and dust, provide constraints on these models. Observations indicate that protostellar discs are common around young stars, that they have a massive phase during their evolution, and that they have typical lifetimes of a few million years.

In this thesis we have presented original research on the formation of gas-giants (Jupiter-like planets and brown dwarfs) through the fragmentation of protostellar

discs via gravitational instability (GI), one of the two prominent models of gas-giant formation. Fragmentation via GI is a top-down process, in which a region of a massive protostellar disc, composed of gas and dust, collapses to form a gas-giant quickly, in thousands of years. In contrast, the core accretion model is a bottom-up process, in which protoplanets grow through the conglomeration of solids in the protostellar disc, growing to masses capable of runaway accretion of the gaseous component of the disc; in this model gas-giants are formed slowly, in millions of years.

Fragmentation via GI requires that a protostellar disc is gravitationally unstable. The condition for instability has been previously determined analytically; however, to determine whether instability leads to fragmentation, and the formation of gas-giants, requires the use of numerical simulations. A number of such simulations have previously been performed, and have indicated that gravitationally unstable discs fragment only if they cool relatively quickly.

To study the viability of fragmentation via GI as a gas-giant formation mechanism requires simulations which accurately model the physical heating and cooling in these discs. Therefore, we have implemented radiative transfer (RT) in the gravito-hydrodynamics code Gasoline, using the commonly-used flux-limited diffusion (FLD) approximation. To model net radiative cooling of the disc, we have coupled FLD to a radiative photosphere boundary condition. The implementation of new physics in a simulation code requires careful testing, and validation of the method and the implementation. We have carefully tested our implementation of RT, including the photosphere boundary, on a number of standard test problems and find that it is accurate. The implementation and testing of RT were presented in Chapter 3.

Our radiation hydrodynamics simulations of protostellar discs that are gravitationally unstable inside of 40 AU show that GI is not a viable mechanism of gas-giant formation in this region. Consistent with analytic models, discs are not able to cool fast enough in this region to transition from instability to fragmentation. Prior numerical simulations of GI, however, disagreed on the likelihood of fragmentation in this region. For one of these prior simulations, we have demonstrated that fragmentation resulted from an over-estimate of the radiative cooling rate in the disc. These simulations were presented in Chapter 3.

We have also performed simulations of protostellar discs that are gravitationally unstable at large radii, near 100 AU. Consistent with analytic models, we have found that fragmentation via GI is a viable gas-giant formation mechanism at these distances. Fragments are found to initially have masses in the gas-giant planet regime; however, they quickly gain mass through accretion of disc material and become brown dwarfs. In addition, we have used our suite of simulations to investigate the link between cooling and fragmentation. From an analysis of our simulations, we have found that spiral arms only fragment if they are smaller than twice their Hill radius. We have developed a model of spiral arm formation and fragmentation in which spiral arms develop due to the linear gravitational instability, have a characteristic width determined by the balance of heating and cooling, and fragment if this width is less than their Hill thickness. This model is consistent with the fragmentation and initial fragment masses in our simulations. In addition, the model has been used in the first calculation of the critical cooling rate required for fragmentation; the results are consistent with previous estimates from numerical experiments. These simulations, and the fragmentation model were presented in Chapter 4.

Radiative transfer is also a fundamentally important physical process in star formation. The numerical methods for RT presented in Chapter 3 can be used to model the non-ionizing radiation that is important in preventing small-scale fragmentation and the over-production of brown dwarfs in numerical simulations of star formation. Ionizing radiation is an important feedback mechanism from massive stars, and can possibly disrupt star forming giant molecular clouds. We have presented numerical methods and test for our implementation of ionizing RT in Chapter 6.

7.2 Future prospects

Numerical simulations are indispensable tools in the study of complex, and non-linear, astrophysical systems such as those discussed in this thesis: protostellar discs, prestellar cores, and giant molecular clouds. Much of this thesis was devoted to the incorporation of radiative transfer into the Gasoline code for 3D dynamical simulations. We have thereby expanded our numerical toolkit, making it possible to more

accurately simulate systems in which RT plays an important role. In this thesis, we have performed simulations of protostellar discs, and gained an improved understanding of the fragmentation process, and its viability as a gas-giant formation mechanism. However, the inclusion of RT in Gasoline has opened up a number of avenues of research that we have not had time to explore in detail. We conclude by briefly outlining some that we find particularly interesting.

7.2.1 Protostellar discs and gas-giant formation

The simulations of GI presented in this thesis began with isolated (non-accreting), gravitationally unstable, axisymmetric initial conditions, with radial surface density and temperature profiles consistent with observations. We did not address how the discs evolved to this gravitationally unstable state, nor did we consider the role of accretion in maintaining the instability against mass transport. Radiation hydrodynamics simulations of the formation of protostellar discs from the collapse of prestellar cores will be important in determining the properties of discs in the earliest phases of evolution. It is during these phases that discs are massive and prone to fragmentation. Such simulations will self-consistently address how common the gravitational fragmentation of protostellar discs is.

More study is required to determine what becomes of the objects formed from gravitational fragmentation at large radii. Simulations, including those presented here, have shown that inward migration is common (with the possible end-result of accretion by the central star); however, observations have detected gas-giants at large radii. In what circumstances do fragments remain at large radii without significant migration, and how common is this outcome? The 2D simulations of Vorobyov & Basu (2010a,b) suggest that the majority of fragments migrate inwards and are accreted by the central star, while only a small number may survive at large radii. The gas-giants observed at large radii may represent a rare outcome of fragmentation. The study of migration is complicated by the formation of multiple fragments, their interactions with each other, their interactions with and accretion from the disc, and subsequent accretion of the remaining envelope mass onto the disc. It is likely that a large suite

of simulations will be necessary to determine the statistical likelihood of different migration outcomes. In addition to affecting the positions of fragments (and their very survival), migration also affects their masses. An understanding of migration is necessary for a comparison of models to observations.

The survival of gas-giants formed through GI depends on their ability to quickly contract. Simulations of the evolution of fragments therefore require realistic cooling, and a realistic equation of state. Of particular importance is the treatment of the dissociation of molecular hydrogen, which allows for the rapid collapse of the fragments to a more bound state. This equation of state has beens implemented in Gasoline to study the evolution of fragments in radiation hydrodynamics simulations (Galvagni et al. 2011). If rapid contraction does not occur, what are the consequences of gas-giant destruction through tidal disruption, or interactions with spiral arms? Are short-lived gas-giants able to effectively process the solid component of the disc and promote planetesimal or protoplanet growth?

It is important to determine the long-term evolution of gas-giants formed through GI because this allows a comparison of the final objects, their masses, positions, and eccentricities, to the ever-increasing number of extra-solar planet observations. Population synthesis models have been a useful tool to statistically compare the products of the core accretion model to observations. A similar statistical comparison of the products of GI to observations is desired.

7.2.2 Star formation

With the implementation of both non-ionizing (IR) and ionizing (UV) RT in gasoline, simulations of star formation are an obvious avenue of future research. Previous simulations of star-cluster formation including non-ionizing RT have focused on small star forming clumps ($M \leq 10^3$ $M_\odot$). Our research group is currently carrying out isothermal simulations of a giant molecular cloud of a more realistic mass of 5×10^4 $M_\odot$ (J. Wadsley). These simulations, or portions thereof, will be recomputed with non-ionizing RT included to study its effects on the initial mass function.

Before performing simulations with ionizing RT, we must test the coupling to hy-

drodynamics on standard tests, such as those of Iliev et al. (2009). After this testing has been done, simulations examining the effectiveness of ionizing feedback in star cluster simulations can be carried out. To date, no simulations have included both ionizing and non-ionizing RT, although both play an important role. Currently, non-ionizing and ionizing RT are independently implemented in Gasoline; however, coupling the two forms of RT together using operator splitting should be straightforward. Accurately following the coupling of different frequencies, through the production of IR photons from the recombination of hydrogen ionized by UV photons for example, is likely the biggest challenge. Testing will be required to determine how well the different methods couple.

Our implementation of ionizing radiation can easily be extended to other wavelengths of interest, such as the Lyman-Werner (LW) band, which is important in the dissociation of molecular hydrogen. The formation and destruction of molecular hydrogen have been implemented in Gasoline (Christensen et al. 2012). Modelling LW RT would be useful in galaxy-scale simulations of the self-consistent formation of giant molecular clouds, which would lead to a better understanding of the clouds' formation, and improved initial conditions for star-cluster simulations.

7.2.3 Galaxy formation

Radiative transfer plays an important role in cosmological simulations of galaxy formation. Including RT in simulations allows for the calculation of self-consistent radiation fields produced by the stellar population. A self-consistent UV field can be used to examine reionization, and to determine accurate local cooling rates. Currently, a uniform UV background, which is a function of redshift, is used in Gasoline. Stars form from molecular gas, yet star-formation "recipes" used in numerical simulations often calculate the star formation rate as a function of the total gas density. A self-consistent LW field can be used to model the formation and destruction of molecular gas, making it possible to calculate the star formation rate as a function of the density of molecular gas (Christensen et al. 2012). Stellar feedback regulates the star formation in galaxy-formation simulations, but resolution-limits require that it

be included as a sub-grid model. High resolution simulations of feedback from star clusters in a local patch of a galactic disc will be carried out by our research group (B. Keller), with the aim of creating more realistic feedback models. Feedback in the form of radiation pressure, HII regions, and supernovae will be included.

7.2.4 Numerical improvements and approximations

The inclusion of RT in 3D dynamical simulations can be computationally expensive, even for approximative methods. The protostellar disc simulations, with non-ionizing RT, presented in this thesis are roughly ten times slower than similar isothermal simulations (this factor of 10 slowdown for our implicit time integration can be compared to the factor of 10^2 to 10^5 slowdown that would be necessary for an explicit time integration). Initial testing indicates that the expense of ionizing radiation is similar to that of the gravity calculation. The timing of production simulations, rather than tests, will give a better idea of the computational cost. Further improvements in the computational implementation of FLD, such as in the parallelization of the code (Matthew Bate, private communication), or a symmetric linearization of the FLD equations and solution using PETSc, may lead to substantial improvements in computational efficiency of the code. In addition, approximative treatments of RT may prove useful, and allow for larger simulations to be carried out for longer durations. The implementation of RT presented in this thesis will be valuable in testing the accuracy of such approximations.

Bibliography

Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, ArXiv e-prints: 1205.5567

Galvagni M., Hayfield T., Mayer L., Saha P., 2011, in IAU Symposium, Vol. 280, IAU Symposium, p. 167P

Iliev I. T. et al., 2009, MNRAS, 400, 1283

Vorobyov E. I., Basu S., 2010a, ApjL, 714, L133

Vorobyov E. I., Basu S., 2010b, ApJ, 719, 1896