The role of gravitational instability in early protostellar evolution and massive planet formation: a numerical perspective

Author(s): Hayfield, Tristen John

Publication Date: 2011

Permanent Link: https://doi.org/10.3929/ethz-a-6618026

Rights / License: In Copyright - Non-Commercial Use Permitted
The Role of Gravitational Instability in Early Protostellar Evolution and Massive Planet Formation: A Numerical Perspective
DISS. ETH NO. 19540

The Role of Gravitational Instability in Early Protostellar Evolution and Massive Planet Formation: A Numerical Perspective

A dissertation submitted to

ETH ZÜRICH

for the degree of

Doctor of Sciences

presented by

Tristen John Hayfield
B.Sc. Hons., Physics, McMaster University
Born July 1, 1982
Citizen of Canada

Accepted on the recommendation of

Prof. Dr. Simon Lilly, examiner
Prof. Dr. Lucio Mayer, co-examiner
Prof. Dr. Ralf Klessen, co-examiner

2011
Contents

1 Introduction ................................................................. 15
  1.1 Molecular Clouds .................................................. 15
    1.1.1 Virial Theorem ............................................. 16
    1.1.2 The Energy Balance in Clouds ......................... 17
    1.1.3 Turbulence .................................................. 18
    1.1.4 Magnetic Fields ............................................ 19
    1.1.5 Fragmentation .............................................. 20
    1.1.6 Cores ......................................................... 21
  1.2 Forming Protostellar Systems ................................... 22
  1.3 Gravitational Instability ......................................... 23
  1.4 Disc Fragmentation ................................................ 24
  1.5 Simulations .......................................................... 25
  1.6 Smoothed Particle Hydrodynamics ............................... 26
    1.6.1 The kernel interpolant .................................... 27
    1.6.2 Gradients and error analysis ............................ 29
    1.6.3 hydrodynamics .............................................. 31
    1.6.4 Artificial Viscosity ....................................... 33
  1.7 Aims of this thesis ................................................. 34
## Contents

2 Properties of Prestellar Discs .... 37
   2.1 Introduction .................. 38
   2.2 The Simulations ............... 41
      2.2.1 Initial Conditions ........... 41
      2.2.2 Thermodynamics ............. 42
      2.2.3 Resolution ................ 44
      2.2.4 Disc Identification ........ 44
   2.3 Results ....................... 45
      2.3.1 Surface densities and temperatures ..... 48
      2.3.2 Disc Evolution .............. 52
      2.3.3 Stability and Fragmentation .... 66
   2.4 Discussion .................... 71
      2.4.1 Fragmentation .............. 71
   2.5 Conclusions ................... 76

3 Clumps in the Outer Disk ......... 79
   3.1 Introduction .................. 80
   3.2 Models ....................... 81
   3.3 Clump Mass .................... 86
   3.4 Initial Angular Momentum ...... 92
   3.5 Contraction Timescale ......... 94
   3.6 Consequences of Clump Disruption .... 99
      3.6.1 Dust Processing and Growth .... 100
      3.6.2 Core Formation ............. 101
      3.6.3 FU Ori ..................... 102
   3.7 Conclusions ................... 103

4 Resolving mixing in SPH .......... 115
   4.1 Introduction .................. 116
4.2 Smoothed Particle Hydrodynamics . . . . . . 118
4.3 Optimised SPH . . . . . . . . . . . . . . . . 122
  4.3.1 A general derivation of SPH . . . . . 122
  4.3.2 Energy versus entropy forms of SPH . 125
  4.3.3 Errors: choosing the free functions . . 126
  4.3.4 Stability: the choice of kernel function 136
  4.3.5 The local mixing instability & RT dens-
      ities . . . . . . . . . . . . . . . . . . . . . 151
4.4 Implementation . . . . . . . . . . . . . . . . 155
4.5 The Kelvin-Helmholtz instability . . . . . . 156
  4.5.1 Numerical set-up . . . . . . . . . . . 157
  4.5.2 Results . . . . . . . . . . . . . . . . 160
4.6 The Sod shock tube . . . . . . . . . . . . . . 165
4.7 The blob test . . . . . . . . . . . . . . . . . 166
4.8 Conclusions . . . . . . . . . . . . . . . . . . 169
4.9 The effect of neighbour number in TSPH . . 170
4.10 The blob test setup . . . . . . . . . . . . . . 171
Summary

This thesis explores the role of gravitational instabilities in the early stages of star formation, touching on a pair of fundamental questions. The first: What is it that favours the production of multiple systems? In chapter 2 of this thesis we make steps toward answering this question. We find that in the cloud environment, nascent binaries are more stable in the face of rapid accretion than isolated systems. A second question addressed is, is there a plausible way of producing giant planets quickly? In chapter 3 we make substantial progress on this question by showing that under realistic conditions, a young, massive protostellar disc can fragment in the outer regions. Although the second question is quite different from the first, the two questions are linked by the prestellar environment in which we examine them, and by the role played by accretion and by gravitational instabilities in their resolution. The third question examined in this thesis looks at numerical methods and aims to serve the scientific community by improving the fidelity of our results, and opening new avenues by technical means: How can we improve the quality of our sim-
ulations? In chapter 4 we solve some fundamental problems with smoothed particle hydrodynamics, a simulation method widely used in the community. In detail:

**Chapter 2:** We present high-resolution 3D smoothed particle hydrodynamics simulations of the formation and evolution of protostellar discs in a turbulent molecular cloud. Using a piecewise polytropic equation of state, we follow the evolution of an isolated and a binary protostellar system in the same environment. In both cases the discs are sufficiently massive to develop gravitational instabilities. The isolated system accretes gas with steadily increasing specific angular momentum until an \( m = 2 \) mode develops into material arms that fragment, forming clumps with initial masses of \( 5.5 \, M_{\text{jup}} \) and \( 7.4 \, M_{\text{jup}} \), and growing to masses of \( 39 \, M_{\text{jup}} \) and \( 14 \, M_{\text{jup}} \), respectively. The clumps have accretion rates of \( 10^{-5} \, M_{\odot} \, \text{yr}^{-1} \). While the discs in the binary system are strongly self-gravitating, we find that they are stable against fragmentation due to disc truncation and mass profile steeping by tides, accretion of high specific angular momentum gas by the secondary, and angular momentum being redirected into the binary’s orbit.

**Chapter 3:** We explore the initial conditions for fragments in the extended regions (\( r \gtrsim 50 \, \text{AU} \)) of gravitationally unstable disks. We combine analytic estimates for the fragmentation of spiral arms with 3D SPH simulations to show that initial fragment masses are in the gas giant regime. These initial fragments will have substantial angular momentum, and should form disks with radii of a few AU. We show that clumps will survive for multiple orbits before they undergo a second, rapid collapse due to H2 dissociation and that it is possible
to destroy bound clumps by transporting them into the inner disk. The consequences of disrupted clumps for planet formation, dust processing, and disk evolution are discussed. We argue that it is possible to produce Earth-mass cores in the outer disk during the earliest phases of disk evolution.

**Chapter 4:** Standard formulations of smoothed particle hydrodynamics (SPH) are unable to resolve mixing at fluid boundaries. We use an error and stability analysis of the generalised SPH equations of motion to prove that this is due to two distinct problems. The first is a leading order error in the momentum equation. This should decrease with increasing neighbour number, but does not because numerical instabilities cause the kernel to be irregularly sampled. We identify two important instabilities: the clumping instability and the banding instability, and we show that both are cured by a suitable choice of kernel. The second problem is the local mixing instability (LMI). This occurs as particles attempt to mix on the kernel scale, but are unable to due to entropy conservation. The result is a pressure discontinuity at boundaries that pushes fluids of different entropy apart. We cure the LMI by using a weighted density estimate that ensures that pressures are single valued throughout the flow. This also gives a better volume estimate for the particles, reducing errors in the continuity and momentum equations. We demonstrate mixing in our new Optimised Smoothed Particle Hydrodynamics (OSPH) scheme using a Kelvin Helmholtz instability (KHI) test with density contrast 1:2, and the ‘blob test’ – a 1:10 density ratio gas sphere in a wind tunnel – finding excellent agreement between OSPH and Eulerian codes.
Zusammenfassung

gemeinde als Hilfe dienen. Wie können wir die Qualität unserer Simulationen erhöhen? In Kapitel 4 lösen wir einige grundlegende Probleme von «Smoothed Particle Hydrodynamics», eine weitverbreitete Methode. Im Detail:

**Kapitel 2**: Wir stellen hochauflösende 3D «smoothed particle hydrodynamics»-Simulationen der Entstehung und Entwicklung protostellarischer Scheiben in einer Molekülwolke dar. Mittels einer teilweise polytropischen Zustandsgleichung folgen wir der Entwicklung eines protostellarischen Sterns und eines protostellarischen Doppelsterns in der gleichen Umgebung. In diesen beiden Fällen sind die Scheiben genügend schwer, um Schwerkraftsinstabilitäten zu erzeugen. Das Einzelsystem häuft Gas an, das einen stetig ansteigenden Drehimpuls hat, bis sich ein $m = 2$ Modus zu materiellen Armen entwickelt. Die Arme fragmentieren, erzeugen zwei Gasklumpen mit Massen von $5.5 \, M_{\text{jup}}$ und $7.4 \, M_{\text{jup}}$, und wachsen bis zu Massen von $39 \, M_{\text{jup}}$ beziehungsweise $14 \, M_{\text{jup}}$ heran. Die Gasklumpen haben im Durchschnitt Anhäufungsgeschwindigkeiten von $10^{-5} \, M_{\odot} \, \text{yr}^{-1}$. Obwohl sich die Scheiben im Doppelsystem durch ihre eigene Schwerkraft stark beeinflussen, finden wir, aufgrund der Verkürzung der Scheibe, Massprofillansteigung der Gezeiten, der Anhäufung von Gas welches einen hochspezifischen Drehimpuls besitzt, die der sekundäre Stern verursacht hat und zuletzt die Umleitung von Drehimpuls in den Doppelsternorbit, stabil gegen Fragmentierung sind.

**Kapitel 3**: Wir erforschen die Anfangsbedingungen von Gasklumpen in den erweiterten Regionen schwerkraftsinstabilier Scheiben. Wir kombinieren analytische Schätzungen für

mischen, aber dazu nicht fähig sind aufgrund der Konservierung der Entropie. Das Ergebnis ist eine Druckdiskontinuität an der Scherschicht, die die Flüssigkeiten mit verschiedenen Entropien auseinander drückt. Wir beseitigen die LMI mittels einer gewichteten Dichteschätzung, die gewährleistet, dass der Druck einwertig in der ganzen Strömung ist. Diese führt auch zu einer besseren Volumenschätzung für die Teilchen, und die Fehler in der Kontinuitätsgleichung und Momentumgleichung nehmen ab. Wir erläutern das Mischen in unserem neuen Optimised Smoothed Particle Hydrodynamics (OSPH) Schema mit einem Kelvin-Helmholtz-Instabilität (KHI) Test, mit Dichtekontrast 1:2, und mit dem »Blob Test«, ein 1:10 Dichtekontrast-Gasphäre in einem Windtunnel, und finden eine ausgezeichnete Übereinstimmung zwischen OSPH und Eulerischen Koden.
Chapter 1

Introduction

1.1 Molecular Clouds

All star formation occurs within molecular clouds. Molecular clouds are so named because they are dark and opaque and are comprised primarily of cold, molecular hydrogen. They span a large range of sizes and masses. The largest clouds, called giant molecular cloud complexes, span some 100 pc and have masses in the range of \( \sim 10^6 \, M_\odot \), and the smallest clouds are \( \sim 1 \, \text{pc} \) and have masses of \( \sim 10 \, M_\odot \). The temperatures of molecular clouds are low, and are well-constrained by observation. Observations indicate a minimum temperature of 8 K, with typical temperatures of 10–13 K (Kirk et al., 2007). Molecular clouds are supersonically turbulent. Remarkably, there is an observed trend in the magnitude of turbulence in molecular clouds, and their size, such that the line-of-sight
velocity dispersion $\sigma(l)$ varies as the length scale $l^{1/2}$ (Larson, 1981). Larson also found two other scaling relations for molecular clouds, namely that of self-gravitating equilibrium $2\sigma l^2/GM \sim 1$, and that the density $\rho$ scales inversely with cloud size as $l^{1.1}$. Molecular clouds are thought to have lifetimes between 1–10 Myr, inferred from the ages of stars found within them. Stars form from dense regions in the clouds, named cores, whose self-gravity outweighs the physical processes providing support, namely thermal pressure, turbulent pressure, and magnetic fields.

1.1.1 Virial Theorem

One of the basic questions we can ask about molecular clouds are whether or not they are in hydrostatic equilibrium. We can approach this question using the virial theorem. Euler’s equation in the comoving frame with gravitational and magnetic source terms reads (Stahler & Palla, 2005)

$$\frac{Dv}{Dt} = -\frac{\nabla p}{\rho} - \nabla \Phi_g + \frac{1}{\rho c_s} j \times B$$

(1.1)

where $\frac{D}{Dt}$ is the convective derivative, $\rho$ is the density, $p$ is the pressure, $\Phi_g$ is the gravitational potential, $j$ the current density, and $B$ the magnetic field. The equation 1.1 is a relation between local fluid properties. To understand how the global properties are related, we take the scalar product with the position $r$ and integrate over the cloud volume to get

$$\frac{1}{2} \frac{\partial^2 I}{\partial t^2} = 2T + 2U + W + M$$

(1.2)
where $I$ is the scalar moment of inertia, and $T$, $U$, $W$ and $M$ are the kinetic, thermal, gravitational, and magnetic energies respectively. Surface terms have been ignored, which restricts the discussion to strongly self-gravitating molecular clouds. For clouds in approximate equilibrium, the left hand side of 1.2 is approximately zero.

### 1.1.2 The Energy Balance in Clouds

Using simple (i.e. dimensionally motivated) approximations, we can obtain a basic picture of the energy budget in clouds. The thermal energy of a uniform cloud is $M_{cl} \epsilon$ where $\epsilon$ is the specific internal energy, while to a constant of order unity the gravitational binding energy is $G M_{cl}^2 / R$, thus their ratio is, in terms of typical values (Stahler & Palla, 2005)

$$\frac{U}{|W|} \simeq 3 \times 10^{-3} \left( \frac{M_{cl}}{10^5 M_\odot} \right)^{-1} \left( \frac{R}{25 \text{ pc}} \right) \left( \frac{T}{15 \text{ K}} \right). \quad (1.3)$$

It’s clear clouds are not supported by thermal pressure. Repeating the same calculation, assuming a uniform magnetic field in the cloud we get a magnetic field energy of $R^3 |B|^2 / 6\pi$ and a final ratio of (Stahler & Palla, 2005)

$$\frac{M}{|W|} \simeq 0.3 \left( \frac{B}{20 \ \mu \text{G}} \right)^2 \left( \frac{R}{25 \text{ pc}} \right)^4 \left( \frac{M_{cl}}{10^5 M_\odot} \right)^{-2}. \quad (1.4)$$

At least on the largest scales in molecular clouds, observations indicate that the magnetic energy budget is relatively close to
virial. What type of support is being provided by the magnetic field? The field cannot be uniform throughout the cloud, or else clouds would be observed to be flattened along field lines. It is more likely then that hydromagnetic waves are providing support. Finally, the primary source of kinetic energy in a cloud is due to random internal motions, i.e. the velocity dispersion $\sigma$. This gives us $\frac{1}{2}M_{cl}\sigma^2$ as the kinetic energy in the cloud, and for typical values we get a ratio of (Stahler & Palla, 2005)

$$\frac{T}{|W|} \approx 0.5 \left( \frac{\sigma}{4 \text{ km/s}} \right)^2 \left( \frac{M_{cl}}{10^5 \text{ M}_\odot} \right)^{-1} \left( \frac{R}{25 \text{ pc}} \right). \quad (1.5)$$

From this we see that there is a great deal of energy in internal motions in molecular clouds, and ratio is close to virial on these scales. It is this large amount of energy in random motions which underlies the notion of turbulent support or “pressure”.

### 1.1.3 Turbulence

Molecular clouds are observed to be supersonically turbulent. That is, observations of tracers such as CO show linewidths ($\sim 2 \text{ km/s}$ for a typical cloud) much broader than thermal ($\sim 1.9 \times 10^4 \text{ cm/s}$). Furthermore, the Larson scaling relations imply that the underlying motions are self-similar. On the smallest scale, the viscous scale, the turbulence is dissipated and needs replenishment. The observed scaling relations, and the need for a cascade of energy from larger scales to smaller
scales leads one to conclude that either clouds are short-lived, or their turbulence is somehow driven on the largest scales. It is not likely that the turbulence is internally driven however, since massive stars are more likely to blow the cloud apart (Blitz & Shu, 1980), and the outflows from low-mass stars cannot reach the largest cloud scales (Brunt et al., 2009). Furthermore, clouds in the solar neighbourhood are not thought to be long-lived. This came to light when attempts to find T-tauri stars older than 1–3 Myr failed (Herbig et al., 1986). Although it’s clear that some initial seeding of the turbulence is required, the timescale of subsequent driving mechanisms is constrained by the relatively short lifetimes of molecular clouds.

1.1.4 Magnetic Fields

Magnetic fields have the potential to counterbalance gravity in molecular clouds and retard their collapse. Spitzer (1968) derives a critical mass $M_B$ using the virial theorem for uniformly magnetised clouds

$$M_B = \frac{1}{\pi} \left( \frac{5}{9G} \right)^{1/2} \Phi_B$$

where $\Phi_B$ is the flux through the cloud. Clouds with masses greater than $M_B$ can collapse under self-gravity. Such clouds are called “magnetically supercritical”, conversely clouds with masses less than the critical mass are “magnetically subcritical”. Molecular clouds are generally considered to be mag-
netically supercritical (McKee, 1989). This is because a magnetically subcritical cloud would require an external confining pressure, comparable to its internal magnetic pressure, to stop it from expanding, whereas a subcritical cloud can be held together by self gravity. Observations seem to confirm the general picture of magnetically supercritical clouds (Bourke et al., 2001; Crutcher, 1999). There is however, indirect evidence that molecular clouds could in fact be supported by Alfvénic turbulence. Heyer et al. (2009) repeated observations done by Solomon et al. (1987), and found that not only were cloud masses in the sample generally 2–5 times smaller than originally thought, they observed that $\sigma/R^{1/2}$ scales with surface density as $\Sigma^{1/2}$. This is in contradiction to Larson’s third observed scaling relation, and is consistent with simple models of magnetically supported clouds (Mouschovias, 1987). As the observational picture evolves over the decades, we can still say that true role of magnetic fields in clouds can only be clarified with yet more observations.

1.1.5 Fragmentation

Stars are produced by gravitational fragmentation of gas in molecular clouds. One of the simplest ways of conceptualising the gravitational fragmentation of gas is an analysis due to Jeans (1902). Consider an infinite, isothermal, uniform gas with sound speed $c_s$ and density $\rho_0$. Although not strictly rigorous, a naïve perturbation analysis of the system shows that
perturbations larger than

\[ \lambda_j = \left( \frac{\pi c_s^2}{G \rho_0} \right)^{1/2} \]  \hspace{1cm} (1.7)

where \( G \) is the gravitational constant, are unstable, and will grow exponentially in amplitude. From the jeans length one can derive via dimensional means a jeans mass

\[ M_j \simeq \lambda_j^3 \rho_0 \]  \hspace{1cm} (1.8)

which yields an approximate minimum fragment mass. Since the gas in a molecular cloud is isothermal up to densities of \( 10^{-13} \text{ g cc}^{-1} \), one can use this to derive an absolute minimum fragment mass for molecular clouds, see Low & Lynden-Bell (1976) and also section 2.2.3.

### 1.1.6 Cores

Cores are dense objects observed within molecular clouds with subsonic internal motions. Let’s assume that the energy balance in a thermally supported core is approximately given by (Hartmann, 2009)

\[ \frac{GM_C}{R_C} \sim c_s^2 = \frac{\kappa T}{\mu m_H} \]  \hspace{1cm} (1.9)

where \( M_C, R_C \) are the core mass and radius, \( T \) is the temperature, \( \mu \) the mean molecular weight, and \( m_H \) the hydrogen mass. A typical cloud temperature of \( 10^4 K \), and mean molecular weight of 2.3 give a sound speed of \( 1.9 \times 10^4 \text{ cm/s} \), for a
solar mass core, the length scale $R_C$ should be $\sim 0.1$, which seems to be borne out by observations (e.g. Caselli et al., 2002; Goodman et al., 1993).

Initially, cores were considered long-lived objects undergoing a slow, quasistatic contraction hindered by magnetic fields. Ions are forced to travel along field lines by the Lorenz force, and ion-neutral collisions couple the neutral component indirectly to the field. A sufficiently low ion-fraction in a (UV-shielded) deeply embedded core will allow neutrals to diffuse past the ions in a process known as ambipolar diffusion (Mestel & Spitzer, 1956). Eventually the core would become unstable to collapse on a timescale of $10^7$ years.

Although debate on the issue of the role of magnetic fields in cores continues due to in part to the difficulty of observing magnetic fields in cores, cores are observed to have quite irregular structures, implying that they are dynamically evolving objects (Myers et al., 1991) and combined with the short lifetimes of molecular clouds, it seems that the magnetically-dominated, quasistatic picture of core evolution is unlikely.

### 1.2 Forming Protostellar Systems

A gravitationally bound molecular cloud core will collapse on a free-fall time of $\sim (G\rho)^{-1/2}$. The collapse will proceed either until the density is high enough for the gas to become optically thick and effectively assume a stiffer, adiabatic equation of state, or until the angular momentum radius

$$r_j = j^2/(GM)$$

(1.10)
is reached, with $M$ being the mass of the collapsing object enclosed within the radius $r_j$. After free-fall, we expect the mass accretion rate onto the system to approach

$$\dot{M} \sim \frac{c_s^3}{G}$$  (1.11)

asymptotically over the course of several free-fall times. Early on in the collapse, the accretion rate can easily be several times this value. The collapse will initially produce a central, pressure supported bulge, and since the collapse is an inside-out process, as it proceeds, streams of gas of increasing angular momentum will collide with each other (or the disc material) around the central object, producing an outward growing disc.

## 1.3 Gravitational Instability

The disc is built up during inside-out collapse, and growing amounts of mass start arriving on the disc having higher angular momentum. Gas will invariably start arriving on the disc faster than viscous processes can transport it inward, and the disc mass will grow until it becomes comparable that of the central object. This will lead to production of gravitational instabilities. The Toomre $Q$ parameter (Toomre, 1964)

$$Q = \frac{c_s \kappa}{\pi G \Sigma}$$  (1.12)

where $\kappa$ is the epicyclic frequency and $\Sigma$ is the surface density, is a measure of stability. If $Q > 1$, then the disc is stable to nonaxisymmetric perturbations, otherwise it is unstable. If
he disc is unstable, perturbations will grow on the orbital time to produce spiral density waves. In the early phases of disc formation, gravitational instabilities (GIs) will initially form spiral density waves in the disc, which then self-regulate by transporting mass inward and angular momentum outward. In practice, in numerical simulations, at $Q \sim 2$ one sees the production of spiral waves, while the threshold for runaway instability is $Q \sim 1.5$ (Papaloizou & Savonije, 1991).

### 1.4 Disc Fragmentation

If GIs enter the nonlinear regime they will either saturate or fragment to produce clumps. Whether or not fragmentation occurs depends on the heating and cooling processes in the region of nonlinearity. The spiral arms will first typically steepen into shocks, providing a gravitationally-driven source of heating. Adiabatic compression further heats the gas, along with mass transport, turbulent source terms and shock heating from other sources. If the material can cool on the order of an orbital time, fragmentation can occur (Gammie, 2001). Complicating the matter, the opacities, and hence the cooling rate will be sensitive to localised effects within the spiral arm. When fragmentation does occur (in simulations), clumps of gas having masses of order $\sim M_J$ are produced (e.g. Boss 1997b; Boss 2002, 2008; Mayer et al. 2004, 2007). There is thus a considerable interest in the fate of the clumps and whether or not this is a viable mechanism for producing giant planets or substellar companions. We revisit this topic in chapters 2 and 3.
1.5 Simulations

Observations of molecular clouds, stars, and even planets are now plentiful, and provide useful constraints on theories of star and planet formation. At the same time, simple, idealised analyses of these systems provide a framework for interpreting observations and driving observational campaigns. Lying between these two extremes are numerical simulations. Numerical systems can be modelled self-consistently, and they allow one to explore problems not only in greater detail and complexity, but also well into nonlinear regimes. Furthermore, numerical models provide us with data that are not readily available from observations, such as full phase-space and thermodynamic information. Results from numerical simulations can then both stimulate observations, and be used to inform the assumptions of more complicated analytical models.

The most basic concept behind dynamical modelling in astrophysics is to discretise a problem domain, in both space and time, of a system governed by a set of differential equations, and starting with some initial conditions, integrate the system forward in time accordingly. In this thesis we will mainly be considering the equations of hydrodynamics, coupled with self gravity and thermodynamic source terms (cooling).

Perhaps the most intuitive way to go about this is to divide the domain into a fixed cartesian grid of cubes, and to solve the equations on the grid, updating the contents of each cube using fluxes at subsequent time steps. Such a setup is known as an Eulerian scheme. Eulerian schemes are relatively easy to analyse, and represent the most mature, robust schemes devel-
CHAPTER 1. INTRODUCTION

oped to date (e.g. Fryxell et al., 2000; Teyssier, 2002; Miniati 
& Colella, 2007). Problems involving self-gravity often un-
dergo changes in the relevant physical scales as the simulation 
progresses. To address this issue, codes which adaptively dis-
cretise space according to criteria have been developed, known 
as adaptive mesh refinement (AMR) codes. While these codes 
are no doubt successful, they suffer some drawbacks: They are 
difficult to implement, Eulerian schemes for hydrodynamics 
can sometimes suffer due to their lack of Galilean invariance, 
they do not couple in the most straightforward fashion to tree-
codes for solving gravity, and implementing flexible boundary 
conditions is difficult.

1.6 Smoothed Particle Hydrodynamics

Alternatively, one may choose to view the discretisation as a 
collection of particles representing the fluid flow. Typically, 
the problem domain is discretised into particles of equal mass 
using the density distribution. The equations of motion for 
each particle are then solved in the comoving frame, and the 
particles are advected with the flow. This is known as a La-
grangian method. The prototypical Langrangian scheme is 
smoothed particle hydrodynamics (SPH) (Gingold & Mon-
aghan 1977; Lucy 1977). SPH has a number of advantages 
over grid methods. Abandoning the grid and employing par-
ticles which move with the flow gives the method a natural 
level of adaptivity. Regions of high density are naturally sam-
pled with many resolution elements. The discrete equations of
motion of the particles are simple, and relatively easy to integrate. In contrast to grid codes, SPH codes couple naturally to N-body gravity solvers. Also in contrast to Eulerian codes, free boundaries are quite naturally treated.

1.6.1 The kernel interpolant

The basic idea behind SPH is to use a kernel to interpolate basic fields (such as density) originating from a discrete, unordered point set; the particles. Formally, starting in the continuous limit, we use the integral

\[ A(x) = \int f(x') W(x - x') dx' \]  

(1.13)

where \( f(x) \) is a field to be interpolated, and \( W(x) \) is a kernel, to create the integral interpolant \( A(x) \). The kernel \( W \) is a symmetric function obeying the condition

\[ \int W(x) dx = 1. \]  

(1.14)

If we choose the kernel \( W(x) = \delta(x) \), then we can recover the original function, as \( A(x) = f(x) \). The kernel is then typically parametrised by \( h \), called the smoothing length, which is a characteristic length of the kernel. Having

\[ \lim_{h \to 0} W(x; h) = \delta(x) \]  

(1.15)

is necessary for convergence.
In SPH, the density is a fundamental quantity. How do we estimate it? The integral

$$\tilde{\rho}(x) = \int \rho(x') W(x - x') \, dx'$$

(1.16)

is estimated not using the density $\rho$ but instead the mass $m$, an integral of $\rho$ which, by assumption, is conserved in the volume surrounding a particle. The density estimate is then instead

$$\tilde{\rho}(x) = \int W(x - x') \, dm'.$$

(1.17)

In the discrete limit, $dm'$ is $m_j \delta(x_j - x') \, dx'$, where $m_j$ and $x_j$ are the mass and position, respectively, of the $j$th particle in the neighbourhood of $x$. In modern day simulations, finite-support kernels are used, and the integral becomes

$$\tilde{\rho}_i = \sum_{j=1}^{N} m_j W_{ij}$$

(1.18)

where the notation $B_i$ denotes $B(x_i)$, $B_{ij}$ denotes $B(x_i - x_j)$, and $N$ is the number of particles in the support of $W$. In practice, we fix $N$ and vary the size of the kernel such that the $N$th neighbour is where the kernel’s support ends. This fixes the quality of the density estimate, while allowing for adaptivity in the scales resolved. The density estimate enters all subsequent interpolations via $dx = (dm)/\rho$. Using this identity we get the following for a general SPH interpolant in the discrete limit

$$A_i = \sum_{j=1}^{N} \frac{m_j}{\rho_j} f_j W_{ij}$$

(1.19)
1.6. SMOOTHED PARTICLE HYDRODYNAMICS

1.6.2 Gradients and error analysis

A straightforward application of the derivative operator to 1.13 yields

$$\nabla A(x) = \int f(x') \nabla W(x - x') dx'$$ \hspace{1cm} (1.20)

or in the discrete limit

$$\nabla_i A_i = \sum_{j=1}^{N} \frac{m_j}{\rho_j} f_j \nabla_i W_{ij}$$ \hspace{1cm} (1.21)

where $\nabla_i$ reminds us that the derivatives are with respect to the coordinates of the $i$th particle.

The basic tool for assessing the quality of the interpolations is the Taylor expansion. A Taylor expansion of $f$ about $x$ in equation 1.13 gives

$$A(x) = \int \left( f(x) + (x - x') f'(x) + ... \right) W(x - x') dx'$$ \hspace{1cm} (1.22)

due to the aforementioned symmetry of $W$, the second term ($f'$) in parenthesis should be zero upon integration, making the interpolation errors second order, in the continuous limit. In the discrete limit 1.22 is

$$A_i = \sum_{j=1}^{N} \frac{m_j}{\rho_j} \left( f_i + x_{ij} f'_i + ... \right) W_{ij}$$ \hspace{1cm} (1.23)

the first term tends to be well-behaved, except around discontinuities in the particle distribution. The antisymmetry of the
second term acts to suppress it, but only perfect symmetry in
the particle distribution can ensure that this term be zero. This
is not achieved in practice, and for \( N = 32 \), an interpolation of
a uniform function, over a uniform, but noisy particle distribu-
tion in 3D exhibits fluctuations of more than 10\%. Repeating
the Taylor expansion for equation 1.20, we get

\[
\nabla A(x) = \int (f(x) + (x - x') f'(x) + ...) \nabla W(x - x') dx'.
\]

(1.24)

In the continuous limit, the \( f \) term disappears due to the an-
tisymmetry of the kernel’s derivative, and the errors are first
order. In the discrete limit, we have

\[
\nabla_i A_i = \sum_{j=1}^{N} \frac{m_j}{\rho_j} (f_j + x_{ij} f_i' + ...) \nabla_i W_{ij}.
\]

(1.25)

Unfortunately, this method leaves us with an error term of or-
der \( h^{-1} \) in the discrete limit, imperfectly suppressed by the
antisymmetry of the gradient kernel. The error then depends
on the absolute normalisation of \( f_i \), which can dwarf the true
gradient. In practice, because we cannot expect the particles to
be ordered, this operator gives noisy gradients. This estimate
becomes even worse when there is a discontinuity in the parti-
cle distribution (this is discussed extensively in chapter 4). We
can ameliorate the situation by directly subtracting the error
term, which gives us the following operator

\[
\nabla_i A_i = \sum_{j=1}^{N} \frac{m_j}{\rho_j} (f_j - f_i) \nabla_i W_{ij}.
\]

(1.26)
This is a much cleaner estimate of the gradient, however it is symmetric under the interchange of the indices \( i \) and \( j \), which makes it unsuitable for some purposes as we shall see later. This is still an order zero estimate of the gradient, however, and higher order estimates can be constructed using the taylor expansion; we work through the case of the momentum equation explicitly in 4.3.3.

### 1.6.3 Hydrodynamics

The equations of hydrodynamics in the comoving frame read

\[
\frac{D\rho}{Dt} = -\rho \nabla v \tag{1.27}
\]

\[
\frac{Dv}{Dt} = -\nabla p \rho \tag{1.28}
\]

where \( \frac{D}{Dt} \) is the convective derivative \( \frac{\partial}{\partial t} + v \cdot \nabla \), 1.27 is the continuity equation, and 1.28 is Euler’s equation. Assuming adiabatic flow, we also have

\[
\frac{du}{dt} = -\frac{P}{\rho} \nabla \cdot v \tag{1.29}
\]

by combining the first law of thermodynamics with equation 1.27. To close the system of equations, we also have an equation of state; in the work that follows we deal with ideal gases, so

\[
\frac{p}{\rho} = (\gamma - 1)\epsilon \tag{1.30}
\]
where $\epsilon$ is the specific internal energy, and $\gamma$ is the ratio of specific heats, is employed.

The Euler equations can be derived from the Lagrangian for hydrodynamics (e.g. Bennett 2006):

$$L = \int \left( \frac{1}{2} \rho v^2 - \rho u \right) dV$$

(1.31)

and in many modern derivations of the equations of motion for SPH, equation 1.31 is discretised, rather than equations 1.27–1.29.

Replacing the volume element $dV$ with the volume per SPH particle $m/\rho$, we obtain

$$L = \sum_{j} m_{j} \left( \frac{1}{2} v_{j}^2 - u_{j} \right)$$

(1.32)

and the standard SPH equations of motion then follow from the Euler-Lagrange equations:

$$\frac{d\rho_{i}}{dt} = \sum_{j} m_{j} v_{ij} \cdot \nabla W_{ij}$$

(1.33)

$$\frac{dv_{i}}{dt} = - \sum_{j} m_{j} \left[ \frac{P_{i}}{\rho_{i}^2} + \frac{P_{j}}{\rho_{j}^2} \right] \nabla W_{ij}$$

(1.34)

$$\frac{du_{i}}{dt} = \frac{P_{i}}{\rho_{i}^2} \sum_{j} m_{j} v_{ij} \cdot \nabla W_{ij}.$$  

(1.35)

Note that equation 1.33 is automatically satisfied by time derivative of the SPH density estimate (equation 1.33). For
1.6. SMOOTHED PARTICLE HYDRODYNAMICS

this reason, equation 1.33 is often referred to as the \textit{integral form} of the continuity equation. In chapter 4 we explore alternative formulations of equations 1.33–1.35 based on free functions which disappear in the continuous limit, but which are present in the discrete limit.

1.6.4 Artificial Viscosity

The equations of hydrodynamics quite naturally admit solutions with discontinuities. Problems with discontinuous initial conditions have solutions which are combinations of shock waves, rarefactions and contact discontinuities. Even a continuous initial condition can quite trivially evolve into a discontinuity. A simple example would be that of a wave train of non-infinitesimal amplitude. The density peaks in the wave correspond to peaks in the sound speed, and thus travel slightly faster than the troughs. This leads to a self-steepening wave which approaches a form having a vertical jump from peak to trough, and threatens to become multiple-valued. This is physically impossible, and instead a shock wave results. At the shock front, energy is dissipated and the variables describing the system suffer discontinuities. One remarkable fact about shocks is that it can be shown that as long as the dissipation has the right general form, the jumps in the variables remain independent of its magnitude. This motivates the introduction of dissipation known as \textit{artificial viscosity} into SPH. The aim is to prevent the formation of discontinuities, which pose numerical difficulties, by introducing an artificial dissipation scale which is \textit{above grid}. In SPH it has the form (Monaghan,
\[ \frac{\partial \mathbf{v}}{\partial t_{\text{visc}}} = \sum_{j=1}^{N} -\frac{m_j}{\bar{\rho}_{ij}} (\alpha \bar{c}_{ij} \mu_{ij} + \beta \mu_{ij}^2) \nabla_i W_{ij} \quad (1.36) \]

where

\[ \mu_{ij} = \frac{h \mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{r_{ij}^2 + 0.01 h^2}, \quad (1.37) \]

the notation \( \bar{x}_{ij} = 0.5 \ast (x_i + x_j) \), \( \alpha \) and \( \beta \) are parameters controlling the strength of the viscosity typically being set to 1 and 2, respectively, \( c \) is the sound speed, and \( h \) is the smoothing length. The effect of this artificial viscosity is to make \( h \) the dissipation scale, and thus the shock is spread out over several smoothing lengths. Typically, this is fine, however one should be aware that this transition zone can have unanticipated interactions with other physics, e.g. cooling. These basic equations, along with a time integration scheme, form the basis of SPH.

### 1.7 Aims of this thesis

The main drive of this thesis is the examination of low-mass star, and massive planet formation from a numerical perspective, concluding with a reevaluation of the numerical methods themselves. Several themes link these topics. Self-gravity plays not only a role in driving early prestellar evolution, but also, quite possibly, one in the production of massive planets. Working at the technological limits of current simulations...
brings many questions of numerics to mind, including resolution requirements for basic physical processes. There is then finally the question of whether and to what extent the numerical methods employed allow one to capture the physics of interest.

The outline of this thesis is as follows. In chapter 2 we examine the early stage of collapse of molecular cloud cores, and the build-up material into a prestellar object and disc, within the cloud environment. In chapter 3 we simulate a massive, extended protostellar disc subject to rapid accretion and realistic cooling, characterising and making predictions regarding fragmentation. In chapter 4 we undergo a rigorous examination of smoothed particle hydrodynamics and solve several key problems with the method.

In chapter 2, we perform high resolution numerical simulations of turbulent molecular clouds - with sufficient resolution to faithfully capture the collapse of individual cores into prestellar systems. We then explore the stability of rapidly accreting, self-gravitating prestellar systems produced in these simulations.

In chapter 3, we simulate a single, massive, extended self-gravitating protoplanetary disc, putting in realistic assumptions about the thermodynamics and cooling. We show that fragmentation can indeed occur and give quantitative predictions of the properties of these giant planet progenitors, using a novel model based on the detailed dynamics of fragmentation. We use analytical models combined simulation results to discuss the further evolution of the clumps.

In chapter 4, we try to understand the problem of mixing in
SPH. By analysing the discrete equations of motion and drawing from results in the literature, we uncover the fundamental issues preventing mixing. We identify three separate numerical instabilities, namely the clumping instability, the banding instability, and the local mixing instability, which hinder the convergence of the method. We demonstrate that fixing these instabilities allows one to model mixing by simulating a 1:2 kelvin helmholtz instability and demonstrating, for the first time with an SPH-based method, the destruction of a 1:10 density ratio cloud in a wind (the “blob” test).
Chapter 2

The Properties of Prestellar Discs in Isolated and Binary Prestellar Systems

In this chapter we explore with numerical simulations the stability of isolated and binary protostellar systems embedded in the cloud environment. In the early, rapidly accreting stage of protostellar formation examined, gravity plays a dominant role in triggering instabilities and moving mass through the disc. We find that several properties of binary systems stabilise them in the face of high accretion rates, and puts their discs on a different evolutionary path to that of an isolated system.

\[1\] This chapter is based on the paper Hayfield, Mayer, Wadsley & Boley (2010)
in the same environment. In contrast, the isolated system is susceptible to growing gravitational instabilities and lead to the production of massive substellar companions.

2.1 introduction

The overall paradigm under which star formation occurs remains under debate. One hypothesis is that stars are thought to form as a collection of low mass fragments in collapsing clumps of gas, and then undergo competitive accretion as the fragments accrete gas from their common reservoir (Bonnell et al. 2001; Bonnell & Bate 2006). The main competing hypothesis is gravitational collapse, where massive star-forming clumps collapse and form multiple cores. Each star forms from the gas that is available in its own core, with limited accretion from other material in the parent clump (Krumholz et al., 2005). Regardless of the star formation paradigm, in the case of solar-type stars, the end result is most likely membership in a binary or multiple system (Duquennoy & Mayor 1991; Eggenberger et al. 2004). Scenarios for the formation of such systems abound. Amongst the simplest, most idealised scenarios are the fission of a bar-unstable core (Durisen et al. 1986; Burkert et al. 1997) and the fragmentation of centrally condensed, rotating, magnetised cores (Boss, 1997a). More complicated hypotheses appeal to the chaos of the cloud environment, such as core-core collisions (Turner et al., 1995), protostellar encounters (Shen et al., 2010), dynamical capture in unstable multiple systems (Bate et al., 2002), and accretion-

Star formation scenarios involving the rapid collapse of a protostellar core following the loss of support against gravity require the formation of a massive accretion disc around the central object (e.g. Vorobyov & Basu 2007 and Walch et al. 2009). Such discs undergo a short-lived $\sim 0.1 \text{ Myr}$ stage where the disc is massive relative to its host ($0.1 < M_d/M_* < 1$) and where gravitational instabilities operate to transport mass through the disc onto the central protostar (Vorobyov & Basu, 2007). Massive, accreting, and extended protostellar discs have been shown to be susceptible to fragmentation, which could be responsible for a range of phenomena such as FU Orionis events and early dust processing if the clumps are disrupted (Boley et al., 2010), and the formation of substellar companions otherwise (Vorobyov & Basu 2007 and Boley 2009).

Given the problem that both the standard core accretion planet formation timescale and protostellar disc lifetimes are typically a few Myr (Haisch et al., 2001), the idea of creating giant planets in a few orbital times via gravitational instabilities was revived (Boss, 1997b) and has since been the subject of sustained interest (Boss 2002, 2008; Mayer et al. 2004, 2007; Pickett et al. 2000; Pickett & Durisen 2007; Boley et al. 2006; Boley 2009). Analytical works constraining inner disc fragmentation (Rafikov, 2005, 2007), and the short cooling times required to form long-lived clumps within $\sim 10 \text{ AU}$ in simulations of protoplanetary discs, along with observations of massive planets on wide orbits (e.g. Marois et al. 2008),
have lead to a shift in focus to outer disc (> 40 AU) fragmentation (e.g. Boley 2009). Whether the extended disc models used in these studies are similar to discs formed from collapsed molecular cloud cores remains to be seen.

In light of the likely connection between outer disc fragmentation and early protostellar systems, we examine in this paper the formation of discs in detail using 3D SPH simulations of core collapse in turbulent molecular clouds. In particular we compare the early evolution in mass, surface density, specific angular momentum, and disc stability between two different systems under near-identical conditions: an isolated and a binary system. Previous studies of gravitational instabilities in binary systems have yielded mixed results, with some finding that the perturbing companion hinders fragmentation through disc truncation and tidal heating (Nelson 2000; Mayer et al. 2005) and alternatively promotes fragmentation, also through tidal perturbations (Boss, 2006). However, the evolved systems considered in these studies may not be as susceptible to fragmentation as their protobinary counterparts would have been due to the enhanced importance of gravitational instabilities during protostellar disc formation.

The paper is laid out as follows: in Section 2 we discuss the simulations, initial conditions and the clump identification procedure. In Section 3 we present the results and analysis. Further discussion is provided in Section 4, and our conclusions are laid out in Section 5.
2.2. THE SIMULATIONS

2.2 The Simulations

All calculations were run with Gasoline (Wadsley et al., 2004), a parallel implementation of TreeSPH. We use a fixed number $N$ of smoothing neighbors, with the main runs using $N = 32$. Artificial viscosity is the standard prescription (Gingold & Monaghan, 1983) with $\alpha = 1$, $\beta = 2$, controlled with a Balsara switch (Balsara, 1989).

2.2.1 Initial Conditions

We start with a spherical and uniform cloud, with a radius of 0.188 pc, a mass of 50 $M_\odot$, and a temperature of 10 K, as in Bate et al. 2003. It was seeded with supersonic turbulent velocities and is marginally self-bound. To simulate interstellar turbulence, the velocity field of the cloud was generated on a grid as a divergence-free Gaussian random field with an imposed power spectrum $P(k) \propto k^{-4}$. The resulting velocity dispersion $\sigma(l)$ varies as $l^{1/2}$, and is consistent with the Larson scaling relations (Larson, 1981). The velocities were then interpolated from the grid to the particles. Finally, the condition that the cloud be marginally self-bound gives a normalization for the global velocity dispersion of $1.17 \text{ km s}^{-1}$.

The Isolated and Binary Systems

The fiducial cloud free fall time is $t_{\text{ff}} = \sqrt{3\pi/32G\rho_{cl}} = 1.92 \times 10^5 \text{ yr}$, with $\rho_{cl} = 1.2 \times 10^{-19} \text{ g cc}^{-1}$ being the initial cloud
density. The evolution of the cloud was followed up to \( t = 1.2 \, t_{\text{ff}} \).

At \( t = 0.88 \, t_{\text{ff}} \) a core, whose collapse we follow in detail, has already formed a small central protostar. The protostar accretes rapidly via a connecting filament until the filament itself produces a fragment. The fragment survives first pericenter with the protostar and becomes a companion.

We produced an isolated system for comparison under almost identical conditions by tracing the particles comprising the companion back in time to \( t = 0.88 \, t_{\text{ff}} \) and stirring them by randomly exchanging their velocities. This procedure conserves the energy and linear momentum but not the angular momentum of the stirred particles. There were approximately 5300 particles, totaling 0.053 \( M_\odot \), that had their velocities perturbed.

### 2.2.2 Thermodynamics

We model the thermodynamics of collapse with a piecewise polytropic equation of state (Tohline 1982 and Bate 1998):

\[
P = \begin{cases} 
\kappa_0 \rho & \rho < 10^{-13}, \\
\kappa_1 \rho^{7/5} & 10^{-13} \leq \rho,
\end{cases} \tag{2.1}
\]

with \( \kappa_0 \) chosen so that the sound speed \( c_s = 1.84 \times 10^4 \, \text{cm s}^{-1} \) and the subsequent \( \kappa_1 \) chosen to ensure pressure continuity. Theoretical estimates of temperatures in molecular clouds find that their temperatures should range between 5–10 K over densities ranging from \( 10^{-19}–10^{-13} \, \text{g cc}^{-1} \) (Larson 1985; Low &
2.2. THE SIMULATIONS

Lynden-Bell 1976; Masunaga & Inutsuka 2000). Observations indicate a somewhat higher minimum temperature of 8 K, with typical temperatures of 10–13 K (Kirk et al., 2007). The equation of state thus captures the approximately isothermal behaviour of the cloud in the intermediate density regime. At high densities the gas transitions to being adiabatic with exponent $\gamma = 7/5$. This equation of state is a simplification of the internal heating and cooling processes in a molecular cloud and, in the high density regime, will minimize the potential for fragmentation. Furthermore, through our choice of softening ($\epsilon_g = 2.3$ AU) combined with the polytropic equation of state, we allow the first hydrostatic core to be marginally resolved, while still inhibiting its subsequent dissociating collapse (Larson, 1969), which can be computationally demanding to resolve. Three-dimensionsal collapse simulations with more sophisticated thermodynamics, and including radiative transfer, but of isolated cores with no turbulent cloud environment, have been performed by (Whitehouse & Bate 2006; Bate 2010; Tomida et al. 2010; Commerçon et al. 2010). While proper treatments of both thermodynamics and radiative transfer are important, such simulations are computationally expensive and so far can only be carried out at lower resolution, and with shorter integration times than required in our calculations. We shall explore the problem with more detailed thermodynamics and radiative transfer in a future publication.
Chapter 2. Properties of Prestellar Discs

2.2.3 Resolution

The simulations employ 5 M particles. The particle masses $m_p$ are thus $1.0 \times 10^{-5} \, M_\odot$ in the isolated and binary accretion simulations. The minimum jeans mass $M_j$, defined as $M_j = 2.92 \, c_s^2 / (G^{3/2} \rho_0^{1/2})$, is that obtained at the transition density to the adiabatic equation of state, and is $M_j^* = 1.7 \times 10^{-3} \, M_\odot = 1.8 \, M_{\text{jup}}$. Previous work has shown that molecular clouds should have a minimum jeans mass (Low & Lynden-Bell, 1976), with the gas becoming optically thick to its own cooling radiation at $\sim 10^{-13} \, g \, cc^{-1}$. In our 5 M particle runs $M_j^*/m_p = 167$ and the jeans mass remains adequately resolved at all times. A fixed gravitational softening $\epsilon_g = 2.3 \, \text{AU}$ is used in all simulations.

2.2.4 Disc Identification

In order to compare and separate the systems in the isolated and binary simulations, a working definition of what comprises a protostellar system is required. We define a protostellar system within this context as being a self-bound gas structure with a peak density greater than $6.7 \times 10^{-12} \, g \, cc^{-1}$. The density threshold was chosen to be well above the critical density in order to identify only bound structures that are in the adiabatic regime. To determine whether any cold, isothermal gas is bound to a given system, we require that it is bound below a threshold binding energy $E_b = 2.6 \times 10^{34} \, \text{erg}$. The binding energy criterion is imposed in order to exclude particles that are only marginally or not uniquely bound to the
system, e.g. lying in the protostellar envelope or flowing between the primary and secondary. The value was chosen empirically to be as small as possible while minimising noise in the results.

To put this into practice we use the SKID (Stadel, 2001) group finder. SKID works by pushing tracer particles along density gradients to find local maxima, and linking them with friends-of-friends. Once group assignments are found we then remove unbound particles to create self-bound groups, computing:

$$E = \frac{m_p}{2} |v - v_{cm}|^2 + U + E_{th}$$ (2.2)

as the total energy, with $v$ and $v_{cm}$ the velocity and center of mass velocity, $U$ the gravitational potential, and $E_{th}$ the thermal energy. The center of mass frame is updated throughout the unbinding procedure.

\section{Results}

At $t = 0.88t_{ff}$ a nascent protostellar system has begun to form from the collapse of its enveloping core. In Fig. 2.1, the surface density and velocity field, projected through a 1000 AU$^3$ box cut around the prestellar object, are shown. The box contains 0.29 M$_\odot$ of gas. The central prestellar object formed via fragmentation of a filament 1000 AU in extent. We see in the integrated transverse velocity field that accretion primarily occurs along the filament, as there is only a small torus of opening angles around the system from which low density material
Figure 2.1: The column density and velocity field in a 1000 AU$^3$ volume, containing 0.29 M$_\odot$ of gas, around the central prestellar object. The time is $t = 0.88t_{\text{ff}}$, at the onset of core collapse.
Figure 2.2: The protostellar object formed at the onset of its parental core’s collapse at $t = 0.88t_{\text{ff}}$. The object is 1.7 kyr old at this stage with a disc of 25 AU in radius, and a mass of 0.055 $M_\odot$. 
is observed to make a direct approach. Most gas collides with the filament and is subsequently funneled onto the system.

In Fig. 2.2, the surface density of the prestellar system, in projection along its axis of rotation, is shown. The system is \(0.055 \, M_\odot\) and the disc is 25 AU. The filaments feeding the disc at this stage have densities ranging from \(7 \times 10^{-15} \, g \, cc^{-1}\) up to the critical adiabatic density \(10^{-13} \, g \, cc^{-1}\). The system accretes rapidly until \(t = 0.893t_{\text{ff}}\) whereupon a fragment of mass \(8 \times 10^{-3} \, M_\odot\) forms in the low density filamentary material, close to the Jeans mass of \(6.4 \times 10^{-3} \, M_\odot\). The fragment survives first pericenter and becomes a companion. The isolated system, in contrast, continued accreting material rapidly, building up the disc around the protostellar system.

### 2.3.1 Surface densities and temperatures

**Surface Density Maps**

In this section, we compare the divergent evolution of the isolated and binary systems. To accommodate the change in dynamical times at disc scales we set the units to kyr and count time relative to the formation of the system at \(0.88t_{\text{ff}}\). In Fig. 2.3 we see the surface density maps of the isolated and the binary systems in the left and right columns respectively, shown at times 4.4 kyr, 5.8 kyr, and 7.1 kyr after formation of the primary (or equivalently, the isolated system). Differences appear early between the two systems. In the topmost row, we see that the primary in the binary system is similar to the isolated system in morphology and surface density except for
Figure 2.3: The surface densities of the isolated and binary systems, respectively in the left and right columns, and shown at times 4.4 kyr, 5.8 kyr, and 7.1 kyr from the top row to bottom, after formation.
Figure 2.4: Mass-weighted temperature maps of the isolated and binary systems, respectively in the left and right columns, and shown at times 4.4 kyr, 5.8 kyr, and 7.1 kyr from the top row to bottom, after formation.
a tidal disturbance of the disc caused by an initial close passage of the secondary. In the central row the differences are more pronounced. At 5.8 kyr we see that the isolated system has continued to grow in mass and the disc now has a radius of 50 AU. Between 4.4 kyr and 5.8 kyr approximately 1 orbital time has passed in the binary system and the secondary is at its apogee. The primary has shrunk in extent as the secondary has accreted some of the gas in its outer regions. The secondary is also rapidly accreting high specific angular momentum gas from the two filaments feeding the system. In the bottom row at 7.1 kyr we see that the isolated system has continued to grow in mass and extent but has become gravitationally unstable and an $m = 2$ spiral arm has developed into a material arm. At 7.1 kyr the binary system has completed another orbit, both systems have grown in mass, and the same pattern of tidal disturbance of the primary is observed.

**Temperature Maps**

In Fig. 2.4, mass-weighted temperature maps of the binary and isolated systems, at times 4.4 kyr, 5.8 kyr and 7.1 kyr, are plotted, as in Fig. 2.3. As we are using a piecewise polytropic equation of state (Equation 2.1), some of the gas bound at the outer edge of the system(s) and some interarm gas is still in the isothermal phase. In the isolated system, increasing amounts of mass end up at large radii and at low densities, due to its higher specific angular momentum $j$ (see 2.3.2 for elaboration on these points). In the binary system however, pictured in the right column of Fig. 2.4, almost all of the mass identified as
bound to each system lies at high densities (and is thus adiabatic) and within their respective tidal radii.

2.3.2 Disc Evolution

Disc Mass

In Fig. 2.5, the accretion histories of the isolated system, primary, and secondary are plotted. The mass at a given time is determined by the procedure outlined in 2.2.4, and is thus the combined protostellar and disc mass. The isolated system accretes gas steadily and at a high rate throughout the simulation, with typical accretion rates of $3 - 5 \times 10^{-5} \, M_\odot \, yr^{-1}$. After the initial collapse, the central protostar, defined as the mass contained within $r < 5 \, AU \simeq 2 \epsilon$, has an accretion rate of $5 - 6 \times 10^{-6} \, M_\odot \, yr^{-1}$. Taking the asymptotic rate from 1D protostellar collapse theory, $\dot{M} = m_\odot c_s^3/G$ (Stahler & Palla, 2005), we find that $\dot{M} = 2 \times 10^{-6} \, M_\odot \, yr^{-1}$, and so $m_\odot$, a constant of order unity, is approximately 2.5 at the protostar. This is not a high accretion rate, as 1D collapse simulations asymptote to $m_\odot = 2$ (Stahler & Palla, 2005). The accretion rate onto the entire system however, is high, with $m_\odot \sim 10$, and is perhaps due to embedding within a filament. Both accretion rates are consistent with other studies, such as those found in Walch et al. 2009.

As expected, the net mass accreted by the binary system is close to that of the isolated one. The difference between the two is because there is some mass present in the binary that can not be uniquely assigned either to the primary or secondary.
2.3. RESULTS

Figure 2.5: Total mass accreted (disc + protostar) versus time for the isolated, primary, and secondary systems, and the net mass in the binary system.
Figure 2.6: The ratio of disc mass to accreted stellar mass versus time, for the isolated, primary systems.
2.3. **RESULTS**

The binary system is plotted starting at 3.8 kyr, instead of its formation time at 3330 yr, because during the first pericenter the identification procedure has difficulties separating the two objects uniquely. In the binary system the secondary, due to its orbit, preferentially accretes high specific angular momentum gas from the filaments and accretes mass at a higher rate than the primary, and at later times the mass ratio tends toward unity. This behaviour has been noted before in the literature (e.g. Bate & Bonnell 1997).

In Fig. 2.6, the ratio of stellar mass (the mass within 5 AU, chosen to correspond with the limits of our softening) to disc mass for the isolated and primary systems are plotted. The isolated system continues gaining disc mass faster than it can be transported to the star, while the primary evolves steadily towards stability. As systems normally become moderately self-gravitating at a mass ratio of \( \sim 0.1 \), we see that these newly formed protostellar systems are in a regime dominated by gravitational instability.

**Cumulative Mass Fraction and Mass Profiles**

The discs in both the isolated and binary systems are all massive relative to the central protostar and exhibit strong spiral arms as seen in Figs. 2.3. We expect that tidal torques will, in addition, play a role in triggering mass transport within the discs in the binary as noted for example in (Mayer et al., 2005). In Figs. 2.7 & 2.8 are plotted the surface densities averaged in annuli of the isolated disc and of the primary of the binary system. We see from Fig. 2.7 that in the isolated system the trend
Figure 2.7: Evolution of the surface density profile of the isolated system. $\sigma$ tends to increase at all radii with time.
Figure 2.8: Evolution of the surface density profile of the primary in the binary system. $\sigma$ decreases steadily at outer radii and increases at inner radii due to tidal interactions.
Figure 2.9: The mass profile evolution of the isolated system. As time progresses, increasing amounts of mass relative to the total lies at large radii. We illustrate this by plotting the evolution of the half-mass radius in black.
Figure 2.10: The mass profile evolution of primary. In contrast to Fig. 2.9 we see the relative fraction of mass at large radii decreasing with time. We illustrate this by plotting the evolution of the half-mass radius in black.
as time progresses is toward increasing amounts of mass at all radii, while from Fig. 2.9 it is evident that the relative amount of mass at large radii is increasing, illustrated by the outward movement of the half-mass radius. The surface density profile of the primary in the binary system (see Fig. 2.8) shows increasing surface density within 20 AU and clearly decreasing surface density outside of 30 AU throughout the simulation. This is reflected in the cumulative mass profile in fig. 2.10 as a decreasing overall fraction of mass found at larger radii, and the clear inward movement of the half-mass radius (while the total mass slowly increases). The source of this difference is likely increased mass transport from gravitationally amplified tidal perturbations. To illustrate the role of tidal effects we model the primary and secondary as point masses and compute the Jacobi radius $r_j$ at 5.8 kyr given the nominal values $M = 0.125 \, M_\odot$, $m = 0.05 \, M_\odot$ and $R_0 = 75 \, \text{AU}$ (see Figs. 2.5 & 2.3) for the primary mass, secondary mass and separation, respectively. Given these parameters we compute $r_j = 0.41 \, R_0 = 31 \, \text{AU}$, thus we expect tidal effects to, at the very least, limit the disc of the primary to 44 AU. Looking at the cumulative mass profile, we see that at our disc-finding procedure is roughly consistent with such a simple model, with very little bound mass found outside of 30 AU.

**Temperature Profiles**

In Figs. 2.11 & 2.12 are plotted the temperature profiles in the isolated and binary systems, respectively. At times 4.4 kyr and 5.8 kyr, the temperature declines steadily from 250 K to 10 K
Figure 2.11: Azimuthally mass-averaged temperatures in the isolated system. At later times the high $j$ gas has formed an extended disc which includes cold interarm gas around the primary, leading to a break in the temperature profile.
Figure 2.12: Azimuthally mass-averaged temperatures in the primary of the binary system. The disc remains limited by tidal forces, and almost all bound gas is adiabatic.
at 40 AU and ∼ 60 AU respectively, whereupon it remains at the temperature floor for larger radii, however there is little mass at these radii (see Fig. 2.9). At time 7.1 kyr, the disc becomes more extended, and the temperature profile flattens at large radii. The disk contains a mixture of isothermal and adiabatic gas, with the spiral arms containing most of the adiabatic component. In the binary system (Fig. 2.12), the temperature profiles of the primary lack the low temperature, low density component seen in the isolated system due to truncation of the disc at the tidal radius, and do not evolve much after 5.8 kyr.

**Specific Angular Momentum**

In Fig. 2.13 are plotted the combined disc and protostellar specific angular momenta for the isolated system, primary and secondary. Each measure is computed in the centre-of-mass frame of the given systems and is thus a measure of the spin (i.e. disc) component of the specific angular momentum. The isolated system rapidly increases in specific angular momentum, and after 7.1 kyr, becomes unstable to fragmentation. In the binary system the primary has a small spike in its specific angular momentum $j$ at 4.2 kyr due to weakly bound material appearing to be bound to the primary while the secondary reaches pericenter. This is also seen in the mass plot at the same time. The primary exhibits very little evolution in $j$ thereafter, while from Fig. 2.5 we see that it continues to accrete mass in the interim; it accretes at constant $j$. In contrast the secondary increases steadily in $j$ along with $M$.

Part of the reason that the binary increases more slowly in
Figure 2.13: Evolution of the specific angular momentum $j$ of the isolated and binary systems. $j$ grows rapidly for the isolated system and leads to a more rapid buildup of an extended disc, whereas tidal effects and lower $j$ leads to smaller discs in the binary system.
specific angular momentum than the isolated system, is that it is able to store momentum in the binary orbit. To measure this effect, we use the relationship

$$L = \sum_i R_i \times P_i = R \times P + \sum_i r_i \times p_i$$  \hspace{1cm} (2.3)$$

where $L$ is the angular momentum, $R_i$ are the positions of each particle, $P_i$ are the particle momenta, $R$ is the barycentre, $P$ the velocity of the barycentre, $r_i$ and $p_i$ positions and momenta relative to the barycentric frame, and decompose the binary into spin and centre of mass angular momenta. We further get the specific angular momenta using the appropriate mass-weightings along with Equation 2.3. At $t = 7.1 \text{ kyr}$, we find that in the barycentric frame, the isolated system has a specific angular momentum of $7.0 \times 10^{19} \text{ cm}^2 \text{s}^{-1}$. In the binary system, the binary orbit itself has $6.1 \times 10^{19} \text{ cm}^2 \text{s}^{-1}$, the primary has $3.0 \times 10^{19} \text{ cm}^2 \text{s}^{-1}$ and the secondary $1.3 \times 10^{19} \text{ cm}^2 \text{s}^{-1}$ (also visible from Fig. 2.13). The specific angular momentum of the binary orbit dominates that found in the other components.

We quantify alignment of the specific angular momentum of the disc with that of the environment by computing the angle $\theta = \cos^{-1}(\hat{j}_d \cdot \hat{j}_e)$, with $\hat{j}_d$ being the unit vector of the specific angular momentum of the disc and $\hat{j}_e$ being that of the material in the environment in a box of dimensions $2000^3 \text{ AU}^3$ surrounding the system. At the time of formation of the system (Figs. 2.1 & 2.2) we find $\theta = 36^\circ$. In the isolated system we observe angles of $13^\circ$, $7.7^\circ$, $12^\circ$ at times $4.4 \text{ kyr}$, $5.8 \text{ kyr}$ and $7.1 \text{ kyr}$, respectively. In the binary system the $\hat{j}_d$ of the pri-
mary is $18^\circ, 7.3^\circ, 5.3^\circ$ at times 4.4 kyr, 5.8 kyr and 7.1 kyr. The vectors tend to be oriented perpendicular to the major axes of the filaments feeding the disc. Such relatively small misalignments are consistent with other works suggesting filaments tumbling about one of their shorter axes as the origin of protostellar angular momentum (Walch et al., 2009) and with observations of protostellar systems on these length scales, which find that most gas is extended in the direction perpendicular to outflows (Tobin et al., 2010).

### 2.3.3 Stability and Fragmentation

Throughout the simulation(s), the discs of both the isolated and binary systems are massive and gravitationally unstable. To compare stability we compute the Toomre $Q = c_s \kappa / \pi G \sigma$ parameter locally using mass-weighted projections, through a $200 \times 200 \times 200 \text{AU}^3$ box, followed by azimuthal averages, of the disc. The results are shown in Figs. 2.14 & 2.15. Although the $Q$ curve for the isolated system drops below unity in some regions at 4.4 kyr it does not fragment right away. As the envelope is accreted, the disc evolves to a state where $Q$ lies between one and two out to 90 AU. The disk remains in the unstable regime, but $Q$ varies substantially as the system accretes. In contrast, the primary of the binary system has a minimum $Q$ of 1.3 at later times and, except initially, rises above two at radii greater than 25 AU. The binary system is thus significantly more stable than the isolated one, and this is borne out by the fact that it does not fragment at later times (up to 20 kyr).
Figure 2.14: Azimuthally averaged Toomre Q for the isolated system at times 4.4 kyr, 5.8 kyr and 7.1 kyr.
Figure 2.15: Azimuthally averaged Toomre Q for the binary system at times 4.4 kyr, 5.8 kyr and 7.1 kyr. The binary system remains stable for as long as the simulation was run (20 kyr).
Figure 2.16: Toomre $Q$ parameter of the isolated system averaged in an annulus from 60–70 AU in times preceding clump formation. A clump later forms at $9.2 \text{ kyr}$ at 65 AU.
Figure 2.17: A surface density map of the isolated system at $t = 11 \text{k yr}$, approximately one orbital time after the second fragment formed.
At 9.2 kyr we find that the isolated system becomes unstable to fragmentation, forming two clumps, one at $r = 65$ AU, followed by another at $r = 100$ AU in the disc. The first clump (C1) formed in one of the material arms of the disc from adiabatic gas, with an initial mass of $5.5 \, M_{\text{jup}}$ and specific angular momentum of $1.8 \times 10^{18} \text{ cm}^2 \text{s}^{-1}$. The second clump (C2) formed at 10 kyr, had a mass of $7.4 \, M_{\text{jup}}$ and specific angular momentum of $3.2 \times 10^{18} \text{ cm}^2 \text{s}^{-1}$. We seek to understand formation of the first clump from adiabatic gas by looking at the behavior of $Q$ in the run-up to fragmentation. In Fig. 2.16 is plotted the $Q$ parameter averaged in an annulus ranging from 60–70 AU. $Q$ starts off initially very high but drops as continuing accretion builds up the disc in this region. Both clumps are depicted in Fig. 2.17 at 11 kyr. At this time, clump C1 has made approximately 1.5 orbits and has a mass of $39 \, M_{\text{jup}}$, and clump C2 one half-orbit, has grown to $14 \, M_{\text{jup}}$. Both clumps accrete at a rate of roughly $10^{-5} \, M_{\odot} \text{ yr}^{-1}$.

2.4 Discussion

2.4.1 Fragmentation

Accretion and specific angular momentum

In the simulations we performed the isolated system becomes prone to fragmentation via the development of massive material arms, even though the gas is adiabatic. Instabilities in the isolated system are driven by steady accretion of gas with increasing specific angular momentum. The disc, sufficiently
massive to generate spiral arms, accumulates a reservoir of gas lying at large radii (i.e. with high specific angular momentum). This reservoir is then swept up by the spiral arms, which when sufficiently massive, become prone to fragmentation; such instabilities have been documented in similar contexts in the literature (e.g. Bonnell 1994; Whitworth et al. 1995; Hennebelle et al. 2004). At the same time, Offner et al. (2008) observe that in simulations of freely decaying turbulence the resulting protostellar systems are more prone to fragmentation, and attribute it to high rates of accretion. Seeing as the isolated system in the present study has an accretion rate 2–3 times that of the nominal 1D asymptotic rate, perhaps the tendency toward fragmentation observed by Offner et al. (2008) is due to a combination of higher accretion rates and accretion of gas with high specific momentum. Furthermore, it could be that the continued driving of turbulence is reducing the specific angular momentum of accreted gas, or acting as a viscosity and depleting the reservoirs of gas which would otherwise be swept up in dense material arms.

The binary system remains stable throughout the simulation. Although the discs are sufficiently massive to be self-gravitating, and the system is accreting, a number of effects work against outer disc fragmentation. The mutual tidal limiting of the discs, as discussed in section 2.3.2, is a major barrier to fragmentation. Though the binary separation evolves (increases) throughout the simulation, by the end of our run the separation is $\sim 75$ AU, and given the tendency of the binary mass ratio to approach unity (e.g. Bate & Bonnell 1997), there is very little chance for outer disc fragmentation to occur.
2.4. DISCUSSION

Longer simulations looking at the long term evolution of the binary separation should be undertaken to see if they separate sufficiently to allow extended discs to grow. The tendency for the secondary to accrete high specific angular momentum gas as seen in this paper and in other studies of binaries (e.g. Bate & Bonnell 1997), acts strongly to deplete the reservoir of gas in the outer part of the system and shuts out the possibility for the growth of the same Toomre instability which lead to fragmentation of the isolated system. Also, the enhanced inward transport of mass due to tidal torques shown in 2.3.2 acts to stabilise the disc against fragmentation, an effect that has also been observed by (e.g. Mayer et al. 2005).

Accretion and gas temperatures

In the isolated system two clumps form by the end of the simulation. They form in the outer reaches of the disc where temperatures were low, and where some interarm gas is still isothermal at 10 K, although clump C1 forms in a material arm from adiabatic gas with a temperature of $\sim 30$ K. Recent simulations of star formation including a flux-limited diffusion treatment for radiative transfer and some modelling of protostellar accretion emphasise the importance of radiative feedback on the envelope. To estimate the effect we compute the
accretion luminosity of the protostar \((\text{Stahler & Palla, 2005})\): 

\[
L_{\text{acc}} = \frac{GM_* \dot{M}}{R_*} = 61 L_\odot \left( \frac{\dot{M}}{10^{-5} \text{M}_\odot \text{yr}^{-1}} \right) \left( \frac{M_*}{1 \text{M}_\odot} \right) \left( \frac{R_*}{5 \text{R}_\odot} \right)^{-1} \tag{2.4}
\]

with \(M_* = 0.11 \text{M}_\odot\) being the protostellar mass, which we compute by taking all mass within \(r < 5 \text{AU} \simeq 2\epsilon\); \(\dot{M} = 5 \times 10^{-6} \text{M}_\odot \text{yr}^{-1}\) is the (time-averaged) accretion rate onto the protostar, and \(R_*\) the protostellar radius that, since it is sub-grid, we assume a fiducial value of \(5 \text{R}_\odot\). From this we get a typical accretion luminosity of \(L_{\text{acc}} = 3.4 L_\odot\). Clump C1 forms at \(r_c = 65 \text{AU}\). To try to get an upper limit on the temperature at that radius, we assume that the disc presents an absorbing surface of height \(h = 10 \text{AU}\) spanning \(r_c \pm 5 \text{AU}\), and that all of the accretion luminosity crossing this surface deposits all of its energy in this region, yielding a heating rate of \(L_{\text{acc}}h/2r_c = 10^{33} \text{erg s}^{-1}\). We assume D’Alessio et al. (2001) opacities and a 1 \(\mu\text{m}\) grain size, and a background irradiation temperature of 30 K, and using the existing density field of the gas within \(r_c \pm 5 \text{AU}\) at the time of fragmentation, we calculate the temperature at which the cooling rate balances the heating rate of the gas to be 50 K. We wish to emphasise that the assumptions made here are extremely conservative with the aim of bracketing the upper end of conceivable outer disc temperatures. Given the weak dependence of \(Q\) on \(T\), it does not appear that the accretion luminosity will be able to affect the disc stability at these large radii.
Clump masses and accretion rates

The isolated system forms two clumps having masses 5.5 $M_{\text{jup}}$ and 7.4 $M_{\text{jup}}$. What kind of clump masses should one expect in the outer disc? Because the disc is in a disordered state at the time of clump formation we do not compare the clump masses to those computed from axisymmetric models, but instead we compute the local Jeans mass of the clump-forming gas, which for clump C1 was 4.7 $M_{\text{jup}}$, and that of clump C2 was 4.5 $M_{\text{jup}}$. The particle mass in the simulations was $10^{-5} M_\odot$, hence in the clumps the Jeans mass is resolved with $\sim 450$ particles, or $\sim 14$ smoothing kernels. The fragmentation phenomena are thus well-resolved according to the criteria set forth in the literature (Bate & Burkert 1997; Nelson 2006). In addition, the initial angular momentum radii of the clumps are 0.31 AU and 0.73 AU, respectively, and should they undergo a second collapse in $\sim 10^4$ years due to $\text{H}_2$ dissociation, any accretion onto the final body will take place through an accretion disc (Boley et al., 2010). We observe final clump masses of 39 $M_{\text{jup}}$ and 14 $M_{\text{jup}}$ for C1 and C2, respectively, however whether or not dissociative collapse will occur before these masses are achieved is highly sensitive to assumptions about the dust. We emphasise here that the connection between our final obtained clumps and the substellar companions (the planets or brown dwarfs) produced from them is uncertain. A dissociating collapse of the clumps and the establishment of an accretion disc is a certainty, however other dynamical instabilities may form, such as a bar instability, and can play a role in redistributing mass and angular mo-
mentation. Because the simulation lacks the necessary physics and resolution to follow the internal evolution of the clumps, we refrain from speculating on the final object masses. Nevertheless, we can estimate accretion luminosities for our final clumps using our clump parameters. Both clumps C1 and C2 had accretion rates of $\sim 10^{-5} \, M_\odot \, yr^{-1}$. Computing the accretion luminosity using Equation 2.4, inputting the measured accretion rate, clump masses of $5.5-39 \, M_\odot$, and assuming a fiducial final size of $1 \, R_{\text{jup}} = 7 \times 10^9 \, \text{cm}$, gives accretion luminosities ranging between 16–110 $L_\odot$. Such relatively high luminosities are achieved at this accretion rate because the radius of Jupiter is $\sim 50$ times smaller than our fiducial one for the central protostar ($5 \, R_\odot$). With such extremely high luminosities we can see that mass accretion simply cannot continue unabated without feedback playing some role. Our final clump masses and luminosities thus fall on the upper end of the range of expected values.

2.5 Conclusions

The primary motivation of this paper was to compare the evolution of an isolated and a binary system during the early stages of prestellar development within the same environment. Of particular interest was the potential for fragmentation of such early, accreting and self-gravitating discs. The following are our findings:

- The initial collapse occurs within a filament that tends
to collimate infalling gas onto the central prestellar objects, whose specific angular momenta tend to remain aligned to that of the environment (and perpendicular to the embedding filament) on scales of 1000 AU.

- Tidal interactions act to strip cold gas from members of the binary, concentrate the mass profiles, with an increasing fraction of mass in the disc interior relative to the exterior, and limit the growth of the specific angular momentum of members of the binary. The end result is an increased robustness to fragmentation in the face of high accretion rates found in collapsing cores.

- The binary is able to maintain stability while accreting gas of increasing specific angular momentum by storing the angular momentum in its orbit. The orbital component dominates the spin components.

- The secondary preferentially accretes high specific angular momentum gas, at a higher rate, driving the mass ratio toward unity.

- The isolated system studied accretes mass at increasingly higher rates with continuously increasing specific angular momentum, leaving large amounts of cold gas at large radii and leading to susceptibility of the system to fragmentation.

- Both fragments accrete at $\sim 10^{-5} \, M_\odot \, yr^{-1}$, with calculated accretion luminosities between 16–110 $L_\odot$. 
CHAPTER 2. PROPERTIES OF PRESTELLAR DISCS
Chapter 3

Clumps in the Outer Disk by Disk Instability: Why They are Initially Gas Giants and the Legacy of Disruption

1

In this chapter we move to smaller scales and simulate a single protostellar system with higher resolution, and with more detailed thermodynamics. Like in the previous chapter, the system is undergoing accretion which drives gravitational

\footnote{This chapter is based on the paper Boley, Hayfield, Mayer \\ & Durisen (2010)}
instabilities. We use a cooling scheme which is based on the local optical depth and realistic dust opacities. The result is fragmentation in the outer parts of the disc, producing clumps with a range of masses and evolutionary paths. Using the lagrangian nature of SPH, we trace the clump forming material to an earlier time in the simulation and use it to create quantitative predictions for clumps formed by this mechanism. One clump survives for several orbits before it is tidally disrupted, leading us to speculate on the potential of such a process for dust processing in the disc.

3.1 introduction

Protoplanetary disks are likely to be massive during their initial phases of evolution. Collapse calculations demonstrate that gravitationally unstable disks do form (e.g. Vorobyov & Basu, 2009) and that their evolution includes phases of strong gravitational instability (e.g. Vorobyov & Basu, 2006). In the inner disk \( r \lesssim 50 \) AU, cooling times relative to local dynamical times are too long for the instability to result in fragmentation, and the disk reaches a self-regulating state (e.g. Boley et al., 2006; Durisen et al., 2007). There may be some exceptions due to changes in disk chemistry (Mayer et al., 2007), but these situations require further study. In contrast, fragmentation in the outer, extended disk \( r \gtrsim 50 \) AU becomes quite possible if the Toomre (1964) \( Q \) can be driven toward unity by, e.g. mass loading (Boley, 2009). Whether these fragments typically produce brown dwarfs (Stamatellos et al., 2007; Sta-
matellos & Whitworth, 2009) or gas giant planets is a topic of debate. However, as we argue here, clumps that become destroyed can be just as important to planet formation and disk evolution as clumps that remain bound.

In this paper, we study disk fragmentation conditions and make estimates for initial fragment masses; we describe the initial angular momenta of fragments; and we discuss the consequences of clump disruption for planet formation. We present simulation data and describe toy models in section 2. In section 3, we derive proper estimates for fragment masses, and find consistency between our estimates and simulation data. We estimate the expected initial angular momenta of clumps in section 4, which are also in rough agreement with simulation data. We use a polytropic model in section 5 to estimate the time during which a clump could be tidally disrupted, including accretion effects. These sections culminate to show that clump disruption is a real possibility during disk evolution. For section 6, we speculate on the consequences of disrupted clumps for dust processing, core formation, and the FU Orionis phenomenon. Our conclusions are given in section 7.

3.2 Models

We present two parameterized models. Model A is used to show what we expect for fragmentation around A stars. We assume a temperature profile $T = 350 (1 AU/r)^{1/2} + 10$ K, a mean molecular weight $\mu = 2.3$, and the epicyclic frequency $\kappa \approx \Omega \approx \Omega_{\text{Keplerian}}$ for central star $M = 1.5 M_\odot$. The sur-
face density $\Sigma$ is given by the Toomre (1964) parameter $Q = c_s \kappa / (\pi G \Sigma)$, where $c_s$ is the local sound speed. The entire disk does not need to be gravitationally unstable ($Q \lesssim 1.7$, Durisen et al., 2007), and the total disk mass for the model could vary considerably depending on the size of the low-$Q$ region. If Model A’s disk has a $Q = 1.5$ between $r \sim 100$ and 150 AU, the mass in this region would be about $0.15 \ M_\odot$. Our second model, Model M, represents a disk orbiting an M star, embedded in an envelope with a 30 K irradiation temperature. For $Q = 1.5$, the mass contained between $r \sim 100$ and 150 AU is $0.04 \ M_\odot$. In addition to these parameterized models, we present a simulation with initial conditions (ICs) based on the simulation “SIMA” from Boley (2009) just before the disk fragments. We refer to this model for the rest of the paper as SPHSIM to avoid confusion with SIMA and the analytic models described above. In SPHSIM, a $0.3M_\odot$ star is surrounded by an $r \sim 400$ AU disk that is accreting mass from an envisaged envelope at $\sim 10^{-5}M_\odot \ yr^{-1}$. This accretion rate is consistent with what one expects in a protostellar collapse with a background temperature of 30 K. At the time our simulation begins, the disk has a mass of $\sim 0.19M_\odot$. The inner radius is at $r \sim 18$ AU and an outer radius of $r \sim 510$ AU. We do not include additional mass loading, so SPHSIM represents a last-burst scenario, i.e. the final phase of fragmentation that this disk is likely to experience. The data were interpolated from the CHYMERA (Boley, 2007) cylindrical grid to an SPH realization via a density-weighted Monte Carlo sampling.

SPHSIM is run using GASOLINE (Wadsley et al., 2004), which is a multipurpose code designed to model structure at
3.2. MODELS

various scales, e.g. planetary disks as well as cosmological structure formation. One million particles are followed, with a particle mass $\sim 2 \times 10^{-4} M_J$. The spline force softening is set to 2 AU for the star and 0.5 AU for all other particles. The radiative cooling algorithm is the same as that described in Boley (2009), but adapted to SPH. The cooling is calculated from $\nabla \cdot F = - (36\pi)^{1/3} s^{-1} \sigma (T^4 - T_{\text{irr}}^4) (\Delta \tau + 1/\Delta \tau)^{-1}$, where $s = (m/\rho)^{1/3}$ and $\Delta \tau = s\kappa \rho$ for the local opacity $\kappa$, particle mass $m$, and density $\rho$. The factor $36\pi$ comes from defining the radius of a resolution element as $r = (3m/(4\pi \rho))^{1/3}$ and setting the ratio between the radiative flux and the divergence of the flux to be the area over the volume of the resolution element. D’Alessio et al. (2001) opacities are used, with a 1 $\mu$m maximum grain size. The irradiation temperature $T_{\text{irr}} = 30$ K everywhere. This cooling approximation is good for the outer disk regions, where midplane optical depths are $\lesssim 1$. The disadvantage to this approximation is that it neglects the effects of radiation from a collapsing clump on its surrounding medium.

In addition to the radiative cooling algorithm described above, GASOLINE has been augmented with the Read et al. (2010) OSPH modifications in order to address the SPH limitations outlined, most recently, by Agertz et al. (2007). We find that OSPH is superior to other suggested modifications (e.g. Price, 2008) because (1) OSPH works for a self-gravitating fluid, (2) recovers the correct timescale for the growth of the Kelvin-Helmholtz instability without introducing new free parameters, and (3) allows for convergence testing of the hydrodynamics by increasing the number of neighbors without
being compromised by the tensile instability. We also note that GASOLINE uses a fixed gravitational softening, while the SPH smoothing length $h$ is variable, defined by 32 neighbors. Using a variable softening length can induce fluctuations in the potential energy of particles, which inevitably leads to errors in energy conservation. On the other hand, Nelson (2006) showed that fixed softening could lead either to either enhanced or suppressed clumping. Although there are ways of improving energy conservation that would permit the use of variable softening lengths (Price & Monaghan, 2007), this has yet to be implemented in GASOLINE and represents a future code development project. Instead, we have chosen to ensure that energy is conserved and set the mass resolution such that the gravitational softening length is larger than the typical SPH smoothing length in dense structure, e.g. spiral arms (see discussion in Mayer et al., 2004). The fragments that form in SPHSIM (below) have a median softening to smoothing length ratio of about 4, so SPHSIM should err on the side of slowing the collapse of fragments. Moreover, GASOLINE showed highly satisfactory results in the Wengen Test 4 comparison project, where a highly unstable disk was followed in detail by several codes. In particular, Gawryszczak & Mayer (2008) reported strikingly good agreement between GASOLINE and the FLASH AMR code when following fragmentation in a self-gravitating disk.

Three clumps form in SPHSIM (C1, C2, and C3), and their initial properties are listed in Table 1. We computed the num-
number of particles per Jeans mass according to

\[ \frac{M_{\text{Jeans}}}{M_{\text{particle}}} = 2.92 \frac{c_s^3}{(M_{\text{particle}} G^3/2 \rho_{\text{peak}}^{1/2})}, \]

for sound speed \( c_s \) and peak density \( \rho_{\text{peak}} \) in the fragmenting spiral arm. As \( \frac{M_{\text{Jeans}}}{M_{\text{particle}}} \sim 2000 \), fragmentation is numerically well-resolved, satisfying the Bate & Burkert (1997) criterion (see also Nelson, 2006). Particles are flagged as members using SKID (Stadel, 2001), which groups particles according to density gradients and then completes an iterative unbind for each particle in the group. When comparing the results to SIMA, it should be noted that Boley (2009) reported a clump mass of 20 \( M_J \) for the end of the simulation. This estimate included mass growth for about 1.75 orbits, so it does not represent the clump’s initial mass, which is between 4 and 5 \( M_J \). Boley expressed reservation in the Letter about accepting the final masses from his simulations because, e.g. the radiative effects of the clump on its surroundings were not modeled, and the resolution was too low to follow the evolution of the clump itself. In addition, SIMA only forms one clump, while SPHSIM forms three. Although integrated or time-averaged quantities between two realizations of a simulation should give comparable answers, detailed structure, especially clumps, are extremely sensitive to initial conditions (see Wengen 4 Comparison Project\(^2\)). The difference in the number of fragments between SPHSIM and SIMA are not considered by us to be failures of either model.

\(^2\)www.astrosim.net
CHAPTER 3. CLUMPS IN THE OUTER DISK

3.3 Clump Mass

In this section, we calculate initial clump masses for unstable disks. Why do we care about initial clump masses? Although clumps are likely to accrete, the evolution of the system will depend on its initial state. For example, a non-rotating, one Jupiter-mass clump will contract for about a few $\times 10^5$ yr, until molecular hydrogen dissociates, causing rapid collapse (e.g. Tohline, 2002). We refer to this contraction timescale as the clump’s primary contraction time ($\tau_1$). For five Jupiter masses, $\tau_1$ is reduced to a few $\times 10^4$ yr, and for an initial mass of ten Jupiter masses, the clump will collapse in $< 10^4$ yr (Helled & Bodenheimer, 2010). A factor of ten in the initial mass can change the clump’s initial evolution timescale by a factor of about 100. The spatial scale of fragmentation, which is related to the initial mass, also determines the rotational angular momentum (next section). Once the core of the clump collapses, this angular momentum barrier should lead to the formation of a circumplanetary/brown dwarf disk. This disk is expected to control the long-term accretion history of the clump, as material entering the Hill sphere of the collapsed planet will have angular momentum from disk shear. In order to avoid confusion between fragmentation, i.e. the formation of the clump, and the rapid collapse that follows dissociation of H$_2$, we refer to the latter as dissociative collapse.

From a practical standpoint, it is possible to reasonably constrain initial masses using global simulations, while the subsequent evolution can only be modeled poorly at this time. Simulating clump growth requires resolving convection, the
3.3. CLUMP MASS

photosphere, chemical changes, the clump’s effect on the surrounding medium, the core/disk transition region, and the gas flow into the Hill sphere. Even at our resolution of about 5000 particles per Jupiter mass, this is a daunting task, and best addressed by high-resolution simulations of individual clumps. For all of these reasons, we argue that constraining initial clump masses is fundamental to understanding the fragmentation process in disks and gas giant planet formation.

**What do we expect for initial clump masses?** In order to calculate the fragment mass, we need to know the local surface density and size scale for fragmentation. A back-of-the-envelope estimate is the Toomre mass

\[ M_T = \pi \left( \frac{\lambda_T}{2} \right)^2 \Sigma \]  

(3.1)

(e.g. Nelson, 2006), where the Toomre wavelength,

\[ \lambda_T = \frac{2c_s^2}{(G \Sigma)}, \]  

(3.2)

is roughly the most unstable radial wavelength for local sound speed \( c_s \) and smooth surface density \( \Sigma \). The surface density can be calculated from the Toomre parameter using \( \Sigma = \frac{c_s \Omega}{\pi G \langle Q \rangle} \), and this value can be used to find the Toomre mass for a given \( c_s \) and \( \langle Q \rangle \), where we have taken \( \kappa \approx \Omega \). The brackets are used to denote that \( \langle Q \rangle \) is a smooth, axisymmetric quantity. We can rewrite the Toomre wavelength

\[ \lambda_T = 2\pi \langle Q \rangle \frac{f_e f_g H}{}, \]  

(3.3)

where the local scale height

\[ H = \frac{c_s}{(f_e f_g \Omega)}, \]  

(3.4)
and \( f_e \) and \( f_g \) are shape factors of order unity that depend on the equation of state and on self-gravity effects, respectively. Setting \( \langle Q \rangle = 1 \) and calculating a Toomre mass based on the most unstable radial wavelength for the unperturbed axisymmetric disk (\( \lambda_T \)) includes material over a radial extent that is \( 2\pi H \). Gravitational instabilities can produce strong spiral waves and local density perturbations, making such use of the axisymmetric measure \( \lambda_T \) over large scales in the non-linear regime, where fragmentation occurs, extremely dubious. The Toomre mass is strictly an estimate for the disk mass that becomes incorporated into one wavelength of the resulting spiral waves. Fragmentation is best described in the context of spiral arms.

Instead of using measurements that correspond to an axisymmetric disk, we use length scales and surface density perturbations that are appropriate for spiral arms. The radial extent of fragmentation can be estimated using the results of Durisen et al. (2008, hereafter DHP2008), who used the virial theorem to show that a disk, under isothermal conditions, is most susceptible to fragmentation within a region \( \delta r \) from the corotation of a spiral wave. They demonstrated this behavior using isothermal hydrodynamics simulations. Other studies have confirmed that fragments tend to form near the corotation of spiral waves, even when radiative physics is included (Boley & Durisen, 2008; Boley, 2009). DHP2008 found that an isothermal spiral shock, with a corotation at \( r \), is stable against fragmentation for
3.3. **CLUMP MASS**

\[
\left( \frac{\delta r}{r} \right)^2 > \frac{4\pi^2 f_{\text{DHP}}^2 f_g}{81(\sin i)^2 \langle Q \rangle m^2}, \tag{3.5}
\]

where \( m \) is the number of spiral arms and \( i \) is the pitch angle of the spiral, which is typically \( i \approx 10^\circ \) in gravitationally unstable disks (see Boley & Durisen, 2008; Cossins et al., 2009). Using the DHP2008 definition for \( f_{\text{DHP}} \), we find that

\[
f_{\text{DHP}} = \frac{H M^2 m}{(\pi r)},
\]

where \( M \) is the Mach number for the isothermal shock. Strictly, we are using \( M \) to indicate the density enhancement in the spiral arm over the smooth distribution, and this should be kept in mind when we quantify our results. If the shock truly is isothermal, then the Toomre wavelength within the overdensity will be

\[
\lambda'_T = 2c_s^2/(G \Sigma M^2) = \lambda_T / M^2. \tag{3.6}
\]

We expect fragmentation to occur when \( \lambda'_T \) becomes equal to \( 2\delta r \). The ratio of these quantities is

\[
\left( \frac{2 \delta r \mathcal{M}^2}{\lambda_T} \right)^2 = \frac{132 f_g^3}{81 \pi^2 \langle Q \rangle^3}. \tag{3.7}
\]

Fragmentation should occur at corotation when

\[
f_g \approx 1.8 \langle Q \rangle. \tag{3.8}
\]

Now that we have an estimate for the location and radial extent of fragmentation, we need an expression for the surface density perturbation, relative to the axisymmetric density,
that will permit fragmentation. We also need to know the azimuthal width of the shock, which can be estimated by assuming it is similar to the full height ($2H$) of the disk. For estimating $H$ within a spiral arm, we refer to Boley & Durisen (2006), who showed that the scale height in the post-shock region of a self-gravitating, isothermal spiral shock should be reduced by the factor

$$F(M) = \left(\frac{q + M^2}{q + 1}\right)^{1/2}$$

(3.9)

due to the increased gas density, where $q \equiv$ external gravity or self-gravity for the axisymmetric disk. The self-gravity of a disk at its scale height is well approximated by $2\pi G \Sigma$, while the star’s gravity by $\Omega c_s$. Combining these terms yields

$$q \approx \langle Q \rangle / 2.$$  (3.10)

However, the self-gravity of the unperturbed disk should be included in our definition of $f_g$ as well. We approximate this effect by setting

$$f_g \approx F(M/q^{1/2}) = \left(\frac{\langle Q \rangle + 4M^2/\langle Q \rangle}{\langle Q \rangle + 2}\right)^{1/2}$$

(3.11)

for $\langle Q \rangle < 2$ and $f_g = F(M)$ otherwise. Multiple simulations have shown that an initial $\langle Q \rangle \lesssim 1.4$ is required for an isothermal disk to fragment (e.g. Tomley et al., 1994; Nelson et al., 1998; Johnson & Gammie, 2003; Mayer et al., 2004; Durisen et al., 2007). Using $\langle Q \rangle = 1.4$ in equation (8) gives
3.3. CLUMP MASS

$f_g \approx 2.5$, which corresponds to $M \approx 2.7$ in equation (11). As $\langle Q \rangle$ is lowered, the density perturbation that is required to induce fragmentation is also lowered.

Now that we have estimates for the radial extent of the fragmenting region ($\lambda_T'$), for the width of the spiral ($2H$, using $f_g$ and equation 4), and for the surface density perturbation over the axisymmetric disk ($\Sigma M^2$), we can find the initial clump mass within the context of a fragmenting spiral arm:

$$M_f = 2\lambda_T \frac{\Sigma c_s}{\Omega f_g}.$$  \hspace{1cm} (3.12)

Figure 1 shows the initial mass if it were calculated assuming $M_T = \pi \Sigma (\lambda_T/2)^2$, assuming the Model A temperature profile, and by using our estimate for $M_f$; we assumed $\langle Q \rangle = 1.4$. The curve for Model M shows what we expect from the model parameters, and the actual clump masses are shown by symbols. C1 and C2 match our $M_f$ estimate well, and the initial masses are in the gas giant regime. C3 also has a mass in the gas giant range, but our $M_f$ calculation is an overestimate. This may be due to differences between prompt and delayed fragmentation (see DHP2008 for a detailed discussion). C1 and C2 form near corotation (prompt fragmentation), while C3 appears to form during the collision between the wake of C1 and an outer arm (delayed fragmentation). In contrast to our $M_f$ estimates, $M_T$ indicates that initial masses should be in the range of brown dwarfs for $r \gtrsim 80$ AU, even for $\langle Q \rangle = 1$. This has led to recent claims (Kratter et al., 2010) that, even initially, gas giant-mass clumps should be atypical, which is inconsistent with our analysis and our simulation data. Figure
CHAPTER 3. CLUMPS IN THE OUTER DISK

2 shows a close-up snapshot of C1 before and just after fragmentation. Fragmentation is clearly confined to the spiral arm. The pre-fragment material has an $H \approx 1.2$ AU and a $\delta r \approx 9$ AU. For comparison, we estimate that $H$ should be $\approx 2.2$ AU and $\delta r \approx 8.5$ AU using our analysis above.

3.4 Initial Angular Momentum

The radial extent of fragmentation suggests that clumps should have substantial angular momentum from the shear in the disk. The specific angular momentum of a newly-formed clump can be approximated by

$$J_{\text{init}} \approx \frac{1}{3}(\Omega(r + \delta r)(r + \delta r)^2 - \Omega(r)r^2),$$

which for $\delta r \ll r$, $J_{\text{init}} \approx 1/6(GM_{\text{star}}/r)^{1/2}\delta r$, where $\delta r = \lambda'/T/2$. This estimate assumes that the difference between the orbital angular momentum of a clump’s outermost material and of the centroid of fragmentation ($r$) goes entirely into rotation. The factor of 1/3 comes from assuming we have a rigid rod of length $2\delta r$ for the moment of inertia, based on the shape of the collapsing region in the left panel of Figure 2. Figure 3 shows the initial angular momenta for hypothetical fragments that we expect would form in Models A and M, as a function of radius, according to equation (13). Clumps that form at $r \sim 100$ AU should have $J \sim \text{few} \times 10^{18}\text{cm}^2\text{s}^{-1}$, which is roughly consistent with the simulation data. This also strongly suggests that a circumplanetary disk can form with a radius of
3.4. **INITIAL ANGULAR MOMENTUM**

approximately an AU (Fig. 4) after subsequent clump contraction. The expected angular momentum radius, i.e. where we expect rotation to limit further contraction, can be estimated by

\[ r_J \approx \frac{\delta r^2 M}{36r M_f}. \]  

(3.14)

Using our estimate for the fragmentation mass \( M_f \) from section 3, this becomes

\[ r_J \approx \frac{\pi^2 \langle Q \rangle^2 v_K f_g r}{144 c_s M^4}, \]  

(3.15)

where \( v_K \) is the Keplerian orbital speed. The sizes of the clumps in the SPHSIM (Table 1) are consistent with this estimate (Fig. 4), even though we overestimate \( J \) by a factor of about two. This suggests that thermal pressure is still an important component in a clump’s initial size. Indeed, the initial \( T/|W| \) for the clumps is roughly 0.2, where \( T \) is the total rotational energy and \( W \) is its potential energy. These clumps are initially more like rapidly rotating spheroids than true central object+disk systems. During the contraction phase, the clump may become susceptible to dynamical instabilities, for example, the bar instability (e.g. Durisen et al., 1986). Such a dynamic event would rapidly rearrange the angular momentum distribution of the object, leading to rapid outward transfer of angular momentum and overall expansion of the clump. Convection could also play a role in redistributing angular momentum, but its overall effect is uncertain. Unless some mechanism can transfer angular momentum inward from the outer
mass shells to the central regions, our estimate should be valid, and shows that the high-$J$ material in the clump will be unable to collapse to the size scales of Jupiter. Once dissociative collapse is reached, a disk should form. To illustrate this point further, we compare Jupiter’s rotational angular momentum to a wide-orbit clump’s initial $J$. It is unclear whether Jupiter’s angular momentum is consistent with the $J$s for planets on wide orbits, but it provides a reference value. For this estimate, we assume that Jupiter is a rigidly rotating sphere with a radius $\sim 7 \times 10^9$ cm and a rotational period of 10 hr. This gives $J_{\text{Jupiter}} \sim 3 \times 10^{15}$ cm$^2$ s$^{-1}$, which is about three orders of magnitude smaller than a clump’s initial $J$.

We have ignored the possibility that a large fraction of the angular momentum given to the clump from shear goes into altering the orbital angular momentum of the clump. We expect for this to affect our result by a factor of order unity. Likewise, the orientation of the arm just before fragmentation and the difference between the assumed moment of inertia (the rigid rod) and the actual one will also cause errors of order unity. Nevertheless, the SPHSIM data show that our estimate for $J$ is valid to an order of magnitude and provides a reasonable upper limit.

### 3.5 Contraction Timescale

So far, we have defined the mass scales and initial $J$ that we expect for fragmentation. Whether these clumps can become bound objects depends, in part, on their primary contraction
3.5. **CONTRACTION TIMESCALE**

timescale. Once central temperatures reach $T_c \sim 2000$ K, $H_2$ begins to dissociate, and a rapid collapse to sizes of a few Jupiter radii will follow because energy goes into dissociation instead of thermal support. After the $H_2$ collapse, the clump will be stable against tidal disruption owing to its small size. In contrast, a clump that is transported into the inner disk before it undergoes dissociative collapse could become tidally destroyed. This is illustrated by Figure 5, which shows a snapshot of a clump being tidally disrupted in SPH SIM. Two clumps interact, and C2 from Table 1 is put on an eccentric orbit. As it approaches periastron, the clump’s volume becomes much greater than its Hill volume, and it is destroyed. The disruption of C2 takes place about 1500 yr after its formation. Is this disruption physically motivated? Should the clump have already undergone dissociative collapse? In this section, we determine whether the primary contraction timescale ($\tau_1$) is long enough to make clump disruption a real possibility in protoplanetary disks.

Helled (2009, private communication) graciously shared her data with us (from Helled & Bodenheimer, 2010), which show that clumps with masses of a few $M_J$ roughly follow an $n = 2.3$ polytrope. In order to calculate $\tau_1$ for a variety of conditions, we have developed a poor-man’s gas giant evolution code using polytropes. By assuming initial radii and masses as calculated above, an initial polytrope solution can be determined. Using the polytrope profile and D’Alessio et al. (2001) Rosseland mean gray opacities, a photosphere can be calculated, which gives us the luminosity $L = 4\pi R_{\text{eff}}^2 \sigma T_{\text{eff}}^4$. The time step between iterations is $\Delta t = 0.01 U/L$, where
$U$ is the internal energy of the polytrope. For a given step, the total energy of the polytrope is updated according to $E = U + W - \Delta t L$. The potential energy of the system is then calculated as $W = 3E\frac{\gamma - 1}{3\gamma - 4} - \frac{3GM\Delta M}{R(5-n)}$, where $\gamma = 7/5$ is the thermodynamic adiabatic index, which is not necessarily the same as the structural adiabatic index $\gamma_P = 1 + 1/n$. The term with $\Delta M$ accounts for mass that is accreted over $\Delta t$. Once the new potential energy is determined, a new radius can be calculated using $R = -\frac{3GM^2}{((5-n)W)}$, and the new internal energy can be calculated from the virial, i.e. $U = -W/(3(\gamma - 1))$. With the new mass and radius, the density and temperature profiles can be updated. This continues until $T_c > 2000$ K. For all calculations here, $n = 2.3$ and the maximum grain size assumed in the opacities is $1\mu m$. Unfortunately, we ignore clump rotation for these calculations, even though, as discussed above, we expect it to influence subsequent clump evolution. To reiterate, this omission is expected to have two principal effects: (1) Rotation is extra support, and the contraction should be altered, especially when dynamical instabilities set in. (2) A nonspherical contraction will likely alter the amount of material in the clump that is exposed to high temperatures, but our calculations present an order of magnitude estimate.

We now consider the structure and contraction of hypothetical clumps using the method described above. Polytrope Clump 1 (PC1) is followed without accretion, and its mass is set to $3M_J$ with an initial polytrope radius $\sim 5$ AU. The photosphere is initially at a radius $\sim 3$ AU, with a temperature of 22 K. The background irradiation could affect the contraction timescale, but we do not address that detail here. The cen-
3.5. **CONTRACTION TIMESCALE**

The central temperature and density evolution for PC1 is shown in the $\rho-T$ plane in Figure 6. Symbols indicate $10^4$ yr intervals. The contraction time for PC1 is $8 \times 10^4$ yr, which is very similar to the results of Helled & Bodenheimer. This timescale is about 100 orbits at 100 AU for a 1.5 $M_\odot$ star, giving a large window of opportunity for, e.g., clump-wave and clump-clump interactions to transport clumps inward or outward. However, we also need to address whether accretion will decrease the contraction timescale such that disruption becomes unlikely.

We assume that the clump is accreting at its maximum rate, where $\dot{M}$ onto the clump is limited solely by the rate that mass can be delivered to the planet’s Hill sphere. For simplicity, consider a cylinder centered on the fragment, with Hill radius $R_H$. The flow of material into this cylinder should be dominated by the shear flow around the fragment, which yields

$$
\dot{M} = 2 \int_{0}^{\pi/2} \frac{\Omega(r_0)}{2} \Sigma(r_0) R_H^2 \cos i \sin i \, di, \quad (3.16)
$$

where $r_0$ is the orbital radius and $i$ is the angle from the perpendicular to the radial direction. We have assumed that $\Sigma$, the surface density at $r_0$, is constant over perturbations of $R_H$ from $r_0$. Evaluating the integral gives

$$
\dot{M} \approx 2 \times 10^{-7} M_\odot \text{ yr}^{-1} \left( \frac{r_0}{100 \text{ AU}} \right)^{1/2} \left( \frac{M}{M_J} \right)^{2/3} \times \left( \frac{M_{\text{star}}}{M_\odot} \right)^{-1/6} \frac{\Sigma}{10 \text{ g cm}^{-2}}, \quad (3.17)
$$
for clump mass $M$. We have chosen to normalize the function to $\Sigma = 10 \text{ g cm}^{-2}$ because this surface density corresponds to an unstable disk with $M_\text{star} = M_\odot$ and $T = 10 \text{ K}$ at $r = 100 \text{ AU}$. As the clump accretes, the size of its Hill sphere grows, allowing it to capture more mass. As it becomes more massive, the evolution time toward rapid dissociative collapse decreases.

Figure 7 shows the ratio of the clump mass just before $T_c = 2000 \text{ K}$ to the initial mass, as a function of the initial fragment mass. Although we expect initial clump masses to typically be a few $M_J$, we show higher masses for completeness. Each fragment is assumed to grow at the rate given by equation (17), and has an initial polytrope radius of 5 AU. For calculating $\dot{M}$, we assume a primary mass of 1 $M_\odot$, $\Sigma = 10 \text{ g cm}^{-2}$, and an $r_0 = 100 \text{ AU}$. The accretion rate is updated every time step, allowing $\dot{M}$ to grow as the clump grows. For comparison with PC1, a maximally accreting 1 $M_J$ clump (PC2) is shown on the $\rho$-$T$ plane along with PC1 (Fig. 6). This clump grows to about 14 $M_J$ before it reaches dissociative collapse. Recall that this mass actually represents a rapidly rotating clump, which should form a proto-gas giant/brown dwarf + disk system after dissociative collapse. The timescale to reach collapse is also shown in Figure 7, which indicates that fragments with initial masses less than about 6 $M_J$ remain susceptible to tidal disruption for $\sim 10$ orbits at 100 AU around a 1.5 $M_\odot$ star. As the initial fragment mass increases, accretion becomes less important to the precollapse evolution, but the time available for disruption continues to decrease.

We do not claim that equation (17) represents the rate that a
3.6. CONSEQUENCES OF CLUMP DISRUPTION

clump must accrete material; instead, we use it only as a rough upper limit for the accretion rate. As discussed above, rotation could affect the contraction timescale and lead to dynamical instabilities in the rapidly rotating clump. Additional caveats include the following: (1) The material entering the clump’s Hill sphere may have non-negligible thermal pressure. (2) The energy gained due to the accretion of gas may not be efficiently radiated away. (3) Some material entering the Hill sphere may be unable to shed enough angular momentum to become part of the clump. (4) Convection could rearrange the mass and angular momentum distribution of the clump, leading to additional instabilities. It is unclear how these assumptions will affect our accretion estimate in detail, but we expect that these processes will decrease the average mass accretion rate of the clump, which will increase the time a clump remains susceptible to disruption. We conclude that the disruption of clumps during the earliest phases of disk evolution is quite possible.

3.6 Consequences of Clump Disruption for Planet Formation

The work in the previous sections demonstrates that clumps can remain in their primary contraction phase long enough after their formation to become tidally disrupted. The arguments are based on analytic work as well as simulation data. In this section, we discuss three consequences that clump disruption, as a general mechanism, can have on disk evolution. Each
subsection represents a proper study on its own, so our aim here is to present a foundation for subsequent work. The first topic, dust processing, is the most strongly connected to work presented in this manuscript, while the last topic is much more speculative.

3.6.1 Dust Processing and Growth

Figure 8 shows the dust mass in PC1 that is above temperatures of 1000 K, assuming no mixing or settling and a dust-to-gas ratio of $1/100$. Mixing and an enhancing the dust-to-gas ratio by enrichment will allow for more solids to be processed. Helled & Bodenheimer (2010) show that enrichment by planetesimals is unlikely for clumps that form on wide orbits, but this does not rule out enrichment at birth. As discussed in section 3, fragmentation is most likely to form at the corotation of spiral waves. These spirals will trap solids due to the pressure difference between the wave and the surrounding disk (Haghighipour & Boss, 2002; Rice et al. 2004), ensuring that the dust is enriched where the disk is most likely to fragment. According to Figure 8, for roughly $5 \times 10^4$ yr, $> 1M_\oplus$ of dust will experience temperatures over 1000 K for our assumptions. The figure also shows that a clump would need to survive for $\sim 10^4$ yr before significant high-temperature, high-pressure dust processing is expected to occur. Because C2 only survives for about 1500 yr, this particular clump is not expected to process dust thermally, unless it occurs in the disruption process itself. However, there could have been substantial grain growth during this time, as meter-sized objects
can grow in $\sim 1000$ yr for the conditions in the clump (e.g., Weidenschilling, 2000). Once the clump is disrupted, this material will be distributed into the disk. Because a clump on an eccentric orbit can experience multiple phases of mass loss due to changes in its Hill sphere, complete disruption is not necessarily required to liberate processed dust.

For clumps that do experience high temperatures, their partial or entire disruption may be a mechanism for processing Calcium-Aluminium-Inclusions (CAIs) or CAI-like particles. CAIs are composed of refractory minerals that are stable at temperatures above 1400 K (e.g., Scott 2007). The advantage of this formation mechanism is that it can occur as soon as the disk begins to form; CAIs are among the oldest objects in the Solar System. It is also a fallacy to assert that all of this material will be lost due to accretion onto the star. GIs are known to transport material over large radii inward and outward, with significant mixing and stirring (e.g., Boss, 2004; Boley & Durisen, 2006). The retention of such material should be addressed, which we leave to future work, but we do not expect for all of this processed material to be lost.

### 3.6.2 Core Formation

Clump disruption also may be capable of producing Earth-mass cores early in a disk’s evolution. These cores could provide a substantial head start for the core-accretion mechanism, and are not subject to the meter-barrier problem. To illustrate this possibility, we refer to Helled & Schubert (2008), who found that cores $\sim 1M_\oplus$ can form in contracting proto-gas gi-
ants. In their models, core formation was halted when the core temperature reached 1300 K, which begins silicate sublimation for the small grain sizes that they considered. For clump masses $> 5M_J$, they found that core formation was inhibited due to the fast evolution and high temperatures in the core. We note that larger grains and highly refractory materials can survive beyond 1300 K (see, e.g. Fig. 2 in Scott, 2007), increasing the time (mass range) available for core formation.

Any core that forms from a failed gas giant will not necessarily be lost due to type 1 migration, although it is not expected to stay at the disruption radius. Laughlin et al. (2004) showed that the magnetorotational instability can lead to density fluctuations that will put cores on a random walk through the disk. GIs should cause a similar behavior, but this needs to be explored in future work. The cores that form in failed gas giants could become the first planetoids formed in planetary systems, and become seeds for the growth of gas and ice giants, or left as dwarf planets in an outer disk. Earth-size objects in the Kuiper belt would be consistent with clump disruption.

3.6.3 FU Ori

A massive clump will remain susceptible to tidal disruption for $10^4$ to $10^5$ yr. During this time, transport of a clump into the innermost regions of the disk ($r < 1$ AU) may become possible through clump scattering and/or disk torques. This topic warrants further study, but we only emphasize this as a real possibility during the early phases of disk evolution while heavy
infall is present on the disk (Vorobyov & Basu, 2006; Boley, 2009). For a $\sim 0.3 M_\odot$ star, the Hill radius of a $3 M_J$ object is $\sim 0.3$ AU at a disk radius of 2 AU. For our simple contraction model, we find that the radius of PC1 is comparable to this size just as it hits $T_c \sim 2000$ K. Clump disruption inside a few AU can suddenly supply the inner disk with mass, which may lead to a thermal instability (Bell & Lin, 1994), and consequently, an FU Orionis outburst. Note that we do not require the clump to make it all the way to the star. For a very eccentric scattering event, such that the clump arrives at perihelion at nearly the escape speed, a Jupiter-mass fragment around a $0.5 M_\odot$ primary will cause a maximum change in the primary’s radial velocity of about 60 m/s. Unfortunately, the radial velocity drifts of, e.g. FU Ori have only been constrained to 300 m/s (Petrov & Herbig 2008). Larger clump/primary ratios and smaller perihelion passages may allow for an observable radial drift in other systems.

3.7 Conclusions

We have demonstrated that initial clump masses are expected to be in the gas giant regime. Our analytic mass estimates are consistent with hydrodynamics simulations with radiative cooling. These clumps will have significant angular momentum, suggesting that disks $\sim$ AU in radius should form after dissociative collapse; the initial clump mass represents what will eventually become a core-disk system. Even when mass is assumed to be accreting at its maximum rate, at least with
our analytic estimates, 1 \( M_J \) clumps are only expected to grow to masses \( \sim 10 \, M_J \) before \( \text{H}_2 \) dissociation causes rapid collapse, consistent with known masses for the planets in HR 8799 (Marois et al., 2008). Further growth of the system may be regulated by the circumplanetary/brown dwarf disk due to the angular momentum of the new material entering the Hill sphere. Clump \( \tau_1 \) contraction timescales, even for several \( M_J \), will be \( 10^4 \) to \( 10^5 \) yr, giving sufficient time for clumps to be transported into the inner disk and to be tidally destroyed. Clumps will have very different environments from the typical conditions in the outer disk, and they represent factories for processing dust and building large solid bodies. Clump disruption therefore represents a mechanism for processing dust, modifying grain growth, and building large, possibly Earth-mass, objects during the first stages of disk formation and evolution.
Table 3.1: Initial clump parameters from the simulation SPH-SIM. The value $R_s$ is the radius that a sphere would have for the same volume as the clump. The formation radius is indicated by $a$. The masses and $J$s for C1 and C2 are consistent with our estimates in sections 3 and 4. C3 seems to form as a result of an interaction between an existing clump’s spiral wake and a wave, and is consistent with delayed fragmentation. The last column indicates the formation time after the start of SPH-SIM.
Figure 3.1: Mass estimates calculated by assuming the Toomre mass \( M_T = \pi \Sigma (\lambda_T / 2)^2 \) and by calculating \( M_f \) as described in the text. The curve associated with Model M represents the fragmentation mass that one would expect in the simulation. The curve labeled “T(K) Model A” gives the temperature profile for Model A. The simulation data (crosses) are in good agreement with the \( M_f \) estimates.
3.7. CONCLUSIONS

Figure 3.2: The particle distribution for C1 just before and after fragmentation. Green represents particles that are included in the initial mass given in Table 1.
Figure 3.3: Specific angular momenta for Models A and M. Model M shows what we expect to see in SPHSIM. The simulation data are indicated by crosses.
Figure 3.4: Angular momentum radius $r_J$ for or Models A and M. Model M shows what we expect to see in SPHSIM. The simulation data for $R_s$ are indicated by crosses. Because the $J$s for the clumps are lower than expected by a factor of order unity, the correspondence between $R_s$ and $r_J$ may be due to thermal support.
Figure 3.5: Surface density snapshot from the simulation SPHSIM. The innermost clump (now quite elongated) is being tidally destroyed, and does not survive for more than $1/4$ more of an orbit. C1, near the top of the snapshot, is becoming bar unstable.
Figure 3.6: The $\rho - T$ plane for the contracting polytrope models. Symbols indicate roughly $10^4$ yr intervals. Once the central temperature reaches 2000 K, rapid collapse should occur. PC1 is a 3 $M_J$ clump contracting in isolation. PC2 starts at 1 $M_J$ and accretes mass as fast as gas can flow into the clump’s Hill sphere.
Figure 3.7: The ratio of $M_{2000}$ to the initial fragment mass $M_i$, where $M_{2000}$ is the mass of the clump when the central temperature reaches 2000 K, leading to rapid collapse. The contraction time in kyr is shown with the dashed line, and is on the same scale as the mass ratio curve. Even with accretion, clumps with initial masses below $\sim 6 M_J$ remain susceptible to disruption for $10^4$ yr. For masses greater than roughly 8 $M_J$, the clump collapses before the mass can be doubled. By $M_i = 20 M_J$, accretion is marginalized.
Figure 3.8: Dust mass in regions with $T > 1000$ K. We have assumed no settling or mixing and a dust-to-gas ratio of 1/100.
Chapter 4

Resolving mixing in smoothed particle hydrodynamics

Seeing as results from smoothed particle hydrodynamics simulations underpin the results the preceding chapters, it is natural to take a look at the method itself. At the time of the study a discussion had been reignited in the literature by (Agertz et al., 2007) about some of the deficiencies in the method when it came to mixing. The simple physics of mixing at a shear layer or the destruction of a cloud in a supersonic wind (the “blob” test) are quite universal topics in astrophysics. The cloud in a supersonic wind could quite easily

1This chapter is based on the paper Read, Hayfield & Agertz (2010)
be, for example, a gas clump sent flying through a protostellar disc after a three-body interaction. The question of whether and to what extent turbulence acts on the clump is then relevant. In this work we identify the problems preventing mixing in SPH and propose solutions to them, showing that the solutions work on several test problems, including being the first ones to demonstrate a solution which also works for the “blob” test.

4.1 introduction

Smoothed Particle Hydrodynamics (SPH) was first introduced as a tool for studying stellar structure (Gingold & Monaghan 1977; Lucy 1977), but has since found wide application in all areas of theoretical astrophysics (Monaghan, 1992), in engineering (Libersky et al., 1993), and beyond (e.g. Hieber & Koumoutsakos 2008).

Although there are many varieties of SPH, the central idea is to represent a fluid by discrete particles that move with the flow (Monaghan 1992; Price 2005). Typically these particles represent the fluid exactly, though in some variants the fluid is advected on top of the particles (Dilts 1999; Maron & Howes 2003). The key advantages over Eulerian schemes\(^2\) are its Lagrangian nature that makes it Galilean invariant, and its particle nature that makes it easy to couple to the fast multipole method for gravity that scales as \(O(N)\) (Dehnen 2000; Green-

\(^2\)This does not apply to Lagrangian moving mesh schemes that are Galilean invariant (Springel, 2010).
4.1. *INTRODUCTION*

gard & Rokhlin 1987). However, SPH has problems correctly integrating fluid instabilities and mixing at boundaries (Morris 1996b; Dilts 1999; Ritchie & Thomas 2001; Marri & White 2003; Agertz et al. 2007). Several different reasons have been suggested for this in the literature so far. Morris (1996b) and Dilts (1999) argue that the problem owes to errors in the SPH gradients that do not show good convergence for irregular particle distributions. Price (2008) argue that the problem owes to the fact that entropies are discontinuous at boundaries, while the densities are smooth. This gives spurious pressure blips at boundaries that drive fluids of different entropy apart. They find that adding thermal conductivity at boundaries to smooth the entropies gives improved mixing in SPH. Wadsley et al. (2008) make a similar argument, phrasing the problem in terms of an inability for SPH particles to mix and generate entropy on the kernel scale. They find that adding a heat diffusion term to model subgrid turbulence gives improved mixing in SPH. Finally, Ritchie & Thomas (2001) suggest that the problem lies in the SPH density estimate. They introduce a new temperature weighted density estimate that is designed to give smoother pressures at flow boundaries, thus combating the spurious boundary pressure blip.

In this paper, we perform an error and stability analysis of SPH in its most general form to understand why mixing fails. In doing this, we show that all of the above authors correctly identified one of two distinct problems with mixing in SPH. The first is an $O(h^{-1})$ error in the momentum equation identified by Morris (1996b) and Dilts (1999). The second relates to entropy conservation on the kernel scale, as addressed directly
CHAPTER 4. RESOLVING MIXING IN SPH

by Price (2008) and Wadsley et al. (2008), and indirectly by Ritchie & Thomas (2001). Having identified the problem, we present a new method – Optimised Smoothed Particle Hydrodynamics (OSPH) – that, given sufficient resolution, correctly resolves multiphase fluid flow.

This paper is organised as follows. In §4.2 and §4.3, we briefly review standard SPH schemes and introduce our new OSPH scheme. We show that there are two distinct problems with mixing in SPH: the \( E_0 \) error in the momentum equation, and the ‘local mixing instability’ (LMI), and we show how both can be cured. In §4.4, we present our implementation of OSPH in the GASOLINE code (Wadsley et al., 2004). In §4.5, we use a Kelvin Helmholtz instability (KHI) test with density contrast 1:2 and 1:8 to demonstrate mixing in OSPH. We show the effect of turning on each of the OSPH improvements one at a time, arriving at a solution that is in excellent agreement with the Eulerian code RAMSES (Teyssier, 2002). In §4.6, we use the standard Sod shock tube test to demonstrate that OSPH can successfully model shocks. In §4.7, we revisit the ‘blob test’ introduced in Agertz et al. (2007), finding excellent agreement between OSPH and the Eulerian code FLASH (Fryxell et al., 2000). Finally, in §4.8 we present our conclusions.

4.2 Smoothed Particle Hydrodynamics

In SPH, the fluid is represented by discrete particles that move with the flow. The density of each particle is estimated by a weighted sum over its neighbours:
4.2. SMOOTHED PARTICLE HYDRODYNAMICS

\[ \rho_i = \sum_j^N m_j W(|r_{ij}|, h_i) \quad (4.1) \]

where \( h_i \) and \( m_j \) are the smoothing length and mass of particle \( i \) and \( j \), respectively; we define \( r_{ij} = r_i - r_j \) and similarly for other vectors; and \( W \) is a symmetric kernel that obeys the normalisation condition:

\[ \int_V W(|r - r'|, h) d^3r' = 1 \quad (4.2) \]

and the property:

\[ \lim_{h \to 0} W(|r - r'|, h) = \delta(|r - r'|) \quad (4.3) \]

In the limit \( N \to \infty, h \to 0 \) (and using \( m_j/\rho_j \to d^3r' \)) equation 4.1 recovers the continuum flow density.

The equations of motion for SPH are then derived by discretising the Euler equations – the continuity, momentum and energy equations:

\[ \frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v} \quad (4.4) \]
\[ \frac{d\mathbf{v}}{dt} = -\frac{\nabla P}{\rho} \quad (4.5) \]
\[ \frac{du}{dt} = -\frac{P}{\rho} \nabla \cdot \mathbf{v} \quad (4.6) \]

where \( \rho, \mathbf{v} \) and \( u \) are the density, velocity and internal energy per unit mass of the flow, respectively.
CHAPTER 4. RESOLVING MIXING IN SPH

The Euler equations can be derived from the Lagrangian for hydrodynamics (e.g. Bennett 2006):

\[ L = \int \left( \frac{1}{2} \rho v^2 - \rho u \right) dV \]  
(4.7)

and in many modern derivations of the equations of motion for SPH, equation 4.7 is discretised, rather than equations 4.4-4.6.

Replacing the volume element \( dV \) with the volume per SPH particle \( m/\rho \), we obtain (Price 2005):

\[ L = \sum_j m_j \left( \frac{1}{2} v_j^2 - u_j \right) \]  
(4.8)

and the standard SPH equations of motion then follow from the Euler-Lagrange equations:

\[ \frac{d\rho_i}{dt} = \sum_j^N m_j v_{ij} \cdot \nabla_i W_{ij} \]  
(4.9)

\[ \frac{dv_i}{dt} = - \sum_j^N m_j \left[ \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right] \nabla_i W_{ij} \]  
(4.10)

\[ \frac{du_i}{dt} = \frac{P_i}{\rho_i^2} \sum_j^N m_j v_{ij} \cdot \nabla_i W_{ij} \]  
(4.11)

where \( W_{ij} = W(|\mathbf{r}_{ij}|, h_i) \).

Note that equation 4.9 is automatically satisfied by time derivative of the SPH density estimate (equation 4.1). For this
reason, equation 4.1 is often referred to as the *integral form* of the continuity equation.

The above system of equations are closed by the equation of state:

\[ P_i = u_i (\gamma - 1) \rho_i \] (4.12)

This standard approach to deriving the SPH equations of motion has the advantage that the resulting equations are *coherent*\(^3\) by construction – that is they are consistent with a Lagrangian. This gives very good conservation properties for the flow. It is also straightforward to calculate the necessary correction terms that arise if the smoothing lengths are a function of space and time \( h = h(\mathbf{r}, t) \) (see e.g. Nelson & Papaloizou 1994; Price 2005). We do not include these correction terms in this paper.

However, this standard derivation leads to a scheme that cannot correctly model fluid mixing processes (see §4), which motivates us to move to a more general derivation. Discretising each of the Euler equations separately leads to a free function for each equation: \( \eta, \phi \) and \( \zeta \), and a different smoothing kernel for each. This is the approach we take next in §4.3. In §4.3.3 and §4.3.4, we will then use an error and stability analysis of these more general equations of motion to constrain the new functions \( \eta, \phi \) and \( \zeta \) and our new kernels. By choosing these new free functions and kernels such that they minimise the integration error, we will arrive at a new scheme that can,

\(^3\)Also called *consistent* (Oger et al., 2007).
with sufficient resolution, correctly resolve multiphase fluid flow.

4.3 Optimised Smoothed Particle Hydrodynamics

In the previous section, we presented a standard derivation of the SPH equations of motion. However, this standard derivation leads to a scheme that cannot correctly model fluid mixing processes (see §4). In this section, we move to a more general derivation of the SPH equations of motion. We show that, in general, we have a free function for each of the Euler equations: \( \eta \), \( \phi \) and \( \zeta \), as well as a different smoothing kernel for each. There is also a freedom in the energy equation in the choice of integration variable (energy or entropy; §4.3.2). In §4.3.3 and §4.3.4, we will then use an error and stability analysis of these more general equations of motion to constrain the new functions \( \eta \), \( \phi \) and \( \zeta \) and our new kernels. By choosing these new free functions and kernels such that they minimise the integration error, we will arrive at a new scheme that can, with sufficient resolution, correctly resolve multiphase fluid flow.

4.3.1 A general derivation of SPH

In general, we have some freedom in how we discretise the Euler equations (equations 4.4-4.6) to obtain the equations of
4.3. **OPTIMISED SPH**

motion for SPH (see e.g. Monaghan, 1992; Price, 2005; Ross-wog, 2009). The gradients in the Euler equations can be expanded to include a new free function for each equation: $\eta$, $\phi$ and $\zeta$:

\[
\frac{d\rho}{dt} = \eta \left[ \mathbf{v} \cdot \nabla \left( \frac{\rho}{\eta} \right) - \nabla \cdot \left( \frac{\rho\mathbf{v}}{\eta} \right) \right] \quad (4.13)
\]

\[
\frac{d\mathbf{v}}{dt} = -\left[ \frac{P\phi}{\rho^2} \nabla \left( \frac{\rho}{\phi} \right) + \frac{1}{\phi} \nabla \left( \frac{P\phi}{\rho} \right) \right] \quad (4.14)
\]

\[
\frac{d\mathbf{u}}{dt} = \frac{P}{\rho^2} \zeta \left[ \mathbf{v} \cdot \nabla \left( \frac{\rho}{\zeta} \right) - \nabla \cdot \left( \frac{\rho\mathbf{v}}{\zeta} \right) \right] \quad (4.15)
\]

In the continuum form, above, $\eta$, $\phi$ and $\zeta$ cancel. But in the discrete SPH form, they remain giving a useful additional freedom (Price, 2005):

\[
\frac{d\rho_i}{dt} = \sum_j^N m_j \frac{\eta_i}{\eta_j} \mathbf{v}_{ij} \cdot \mathbf{H}_{ij} \mathbf{r}_{ij} \quad (4.16)
\]

\[
\frac{d\mathbf{v}_i}{dt} = -\sum_j^N m_j \left[ \frac{P_i \phi_i}{\rho_i^2} \frac{\phi_i}{\phi_j} + \frac{P_j \phi_j}{\rho_j^2} \frac{\phi_j}{\phi_i} \right] \mathbf{K}_{ij} \mathbf{r}_{ij} \quad (4.17)
\]

\[
\frac{d\mathbf{u}_i}{dt} = \frac{P_i}{\rho_i^2} \sum_j^N m_j \frac{\zeta_i}{\zeta_j} \mathbf{v}_{ij} \cdot \mathbf{L}_{ij} \mathbf{r}_{ij} \quad (4.18)
\]

where $\mathbf{H}_{ij} = \left[ H(|\mathbf{r}_{ij}|, h_i) + H(|\mathbf{r}_{ij}|, h_j) \right] / 2$, $\mathbf{K}_{ij}$ and $\mathbf{L}_{ij}$ are symmetrised smoothing kernels – one for each Euler equation. Standard SPH (SPH from here on) is a special case of
the above with \( \eta = \phi = \zeta = 1 \) and \( H_{ij} r_{ij} = K_{ij} r_{ij} = L_{ij} r_{ij} = \nabla_i \bar{W}_{ij} \).

Equation 4.16 casts the continuity equation in differential form. This is problematic since, in this case, the particles no longer represent the fluid exactly. Instead they represent a moving mesh on which the Euler equations are solved. This leads to the danger that high density regions will contain few particles leading to large errors (Maron & Howes, 2003). For this reason, we use instead a generalised integral form for the continuity equation:

\[
\rho_i = \sum_j N \eta_i \eta_j \bar{W}_{ij} \tag{4.19}
\]

which, taking the time derivative, gives:

\[
\frac{d\rho_i}{dt} = \sum_j N \eta_i \eta_j v_{ij} \cdot \nabla_i \bar{W}_{ij} + \epsilon \tag{4.20}
\]

where:

\[
\epsilon = \sum_j N \eta_j \left( \frac{\dot{\eta}_i}{\eta_i} - \frac{\dot{\eta}_j}{\eta_j} \right) \eta_i \eta_j \bar{W}_{ij} \tag{4.21}
\]

and \( \dot{\eta} = \frac{dn}{dt} \).

This reduces to the continuity equation (equation 4.16) under the kernel constraint: \( H_{ij} r_{ij} = \nabla_i \bar{W}_{ij} \), and for \( \epsilon = 0 \). The latter can be satisfied by construction if \( \eta_i = \eta_j \) (as is the case for SPH), or if \( \dot{\eta} = 0 \). However, in the continuum limit \( (N \to \infty, h \to 0) \), \( \epsilon \to 0 \) and so \( \epsilon \) will vanish with increasing
resolution. For this reason, equation 4.19 gives a valid approximation to the continuity equation for any choice of \( \eta \), with \( \epsilon \) simply contributing an additional error term.

### 4.3.2 Energy versus entropy forms of SPH

A final freedom in the equations motion for SPH comes from the energy equation. Equation 4.18 is the standard energy form of SPH, but there is also an entropy form (Goodman & Hernquist 1991; Springel & Hernquist 2002). Instead of the internal energy, \( u \), we evolve a function \( A(s) \) – the *entropy function* – that is a monotonic function of the entropy \( s \) defined by the equation of state:

\[
P_i = A_i(s) \rho_i^\gamma
\]  

(4.22)

Away from shocks and in the absence of thermal sources or sinks, \( A_i \) is a constant of motion. Thus, taking the time derivative of equation 4.22 and substituting for equation 4.12, we recover:

\[
\frac{du_i}{dt} = \frac{P_i}{\rho_i^2} \frac{d\rho_i}{dt}
\]  

(4.23)

by construction. Schemes that obey equation 4.23 are called *thermodynamically consistent*.

In practice, we find – for the tests presented in this paper – that the energy and entropy forms of SPH give near-identical results, provided that equation 4.23 is satisfied (for adiabatic flow). We use the thermodynamically consistent energy form
throughout this paper. This gives us the constraints: $\zeta = \eta$ and $L_{ij} r_{ij} = H_{ij} r_{ij} = \nabla_i W_{ij}$, which we apply from here on. We also use $K_{ij} r_{ij} = \nabla_i W_{ij}$, as in standard SPH. This is not a formal requirement, but ensures that coherence is recovered in the limit of constant density.

### 4.3.3 Errors: choosing the free functions

In this section, we perform an error analysis of the generalised equations for SPH (equations 4.17, 4.18 and 4.19) derived in §4.3.1. We will then choose our free functions $\eta$, $\phi$ and $\zeta$ so that these errors are minimised.

#### Error analysis

We assume that both the pressure and velocity of the flow are smooth. In this case, we can Taylor expand to give:

$$P_j \simeq P_i + h x_{ij} \cdot \nabla_i P_i + O(h^2)$$  \hspace{1cm} (4.24)

and

$$v_j \simeq v_i + h (x_{ij} \cdot \nabla_i) v_i + O(h^2)$$  \hspace{1cm} (4.25)

where $x_{ij} = r_{ij} / h$, and we have assumed a constant smoothing length $h$.

Substituting equations 4.24 and 4.25 into the continuity and momentum equations gives:

$$\frac{d\rho_i}{dt} \simeq -\rho_i (R_i \nabla_i) \cdot v_i + \epsilon + O(h)$$  \hspace{1cm} (4.26)
and

$$\frac{dv_i}{dt} \simeq -\frac{P_i}{\rho_i h} E_{0,i} - \frac{(V_i \nabla_i) P_i}{\rho_i} + O(h) \quad (4.27)$$

where $E_{0,i}$ is a dimensionless error vector given by:

$$E_{0,i} = \sum_j^{N} \frac{m_j}{\rho_j} \left[ g_{ij} + g_{ij}^{-1} \right] \nabla_i^x W_{ij} \quad (4.28)$$

and $R_i$ and $V_i$ are dimensionless error matrices given by:

$$R_i = \sum_j^{N} \frac{m_j}{\rho_j} f_{ij} S_{ij} \quad ; \quad V_i = \sum_j^{N} \frac{m_j}{\rho_j} g_{ij}^{-1} S_{ij} \quad (4.29)$$

with:

$$S_{ij} = \frac{1}{x} \frac{\partial W_{ij}}{\partial x} \left( \begin{array}{ccc} x_{ij}^2 & x_{ij} & x_{ij} z_{ij} \\ y_{ij} x_{ij} & y_{ij}^2 & y_{ij} z_{ij} \\ z_{ij} x_{ij} & z_{ij} y_{ij} & z_{ij}^2 \end{array} \right) \quad (4.30)$$

where $\nabla_i^x = h \nabla_i$; $x_{ij} = (x_{ij}, y_{ij}, z_{ij})$; $x = |x_{ij}|$; $f_{ij} = \frac{\rho_j}{\rho_i} \frac{\eta_i}{\eta_j}$; and $g_{ij} = \frac{\rho_j}{\rho_i} \frac{\phi_i}{\phi_j}$.

The accuracy of the continuity equation (4.26) is given by the extent to which $\epsilon = 0$ (see equation 4.21) and $R_i = I$, the identity matrix. The accuracy of the momentum equation (4.27) is given by the extent to which $E_{0,i} = 0$ and $V_i = I$. (The energy equation behaves similarly to the continuity equation with $\epsilon = 0$.)
Minimising errors: the continuity equation

Let us consider how accurately equation 4.26 approximates its Euler equation equivalent (equation 4.4). First, consider standard SPH where $\eta = 1$ and $\epsilon = 0$ by construction. Typically in the literature, the error is calculated only in the continuum limit ($N \to \infty; h \to 0$; see e.g. Price (2005)). In this case, the sums become integrals, and (using $m_j/\rho_j \to d^3x'$) we obtain terms like:

$$\lim_{N \to \infty} R_{33}(x) = \int_V d^3x' f(x, x') \frac{(z - z')^2}{|x - x'|} \frac{\partial W}{\partial x}$$

(4.31)

and

$$\lim_{N \to \infty} R_{12}(x) = \int_V d^3x' f(x, x') \frac{(x - x')(y - y')}{|x - x'|} \frac{\partial W}{\partial x}$$

(4.32)

where the notation $33$ refers to element $[3, 3]$ in the matrix $R$.

If we assume smooth densities, then we can Taylor expand $f$ also to obtain:

$$f = \frac{\rho(r')}{\rho(r)} \simeq 1 + h \frac{(x - x')}{\rho} \cdot \nabla \rho + O(h^2)$$

(4.33)

and we see that, by symmetry of $W$, $R = I$ to $O(h^2)$. In fact, Taylor expanding to an order higher than above, it is straightforward to show that the whole continuity equation is accurate to $O(h^2)$ in the limit $N \to \infty$ (see e.g. Price 2005). A similar argument applies to the other SPH equations of motion and
leads to the conclusion that SPH is accurate to $O(h^2)$. However – and this is a key point – this formal calculation is only valid for smoothly distributed particles in the limit $N \to \infty$. In practical situations, where we have a finite number of particles within the kernel and these are not perfectly smoothly distributed, the leading order errors in the continuity equation appear at $O(0)$ and are contained within the matrix $R$. We will quote orders of error from here on in this finite particle limit.

We can think of each term of $R$ as a finite sum approximation to a dimensionless integral that should be either 0 (for the off diagonal terms), or 1 (for the diagonal terms). For smooth particle distributions, this approximation is a good one since $f_{ij} \approx 1$, while $m_j/\rho_j$ gives a good estimate of the volume of each particle within the kernel. However, if the particles are distributed irregularly on the kernel scale – for example at a sharp density step – then $f_{ij}$ can grow arbitrarily large, while $m_j/\rho_j$ becomes a poor volume estimate. We will demonstrate this in §4.5.

We can improve matters by choosing $\eta = \rho$, which fixes $f = 1$ always. However, the integral form of the continuity equation then becomes:

$$\rho_i = \sum_j^N \frac{\rho_i}{\rho_j} m_j W_{ij}$$

which must be solved iteratively and is not guaranteed to converge. Worse still, $\epsilon$ is now no longer zero and contributes an additional error.

Ritchie & Thomas (2001) present an interesting solution
to this dilemma. If the pressures are approximately constant across the kernel \((P_i \simeq P_j)\) then, for the energy form of SPH (see equation 4.12 and §4.3.2), \(\frac{\rho_i}{\rho_j} \simeq \frac{u_j}{u_i}\) and equation 4.34 is well approximated by the integral continuity equation:

\[
\rho_i = \sum_j N \frac{u_j}{u_i} m_j W_{ij} \quad (4.35)
\]

This can be solved without the need for iteration.

The above suggests that we use \(\eta = 1/u\). Thermodynamic consistency then requires that we set \(\zeta = \eta = 1/u\) (see §4.3.2).

There may be some advantage, however, to using the entropy form of SPH. In this case, the equation of state is given by equation 4.22. For approximately constant pressure across the kernel, we now have that \(\frac{\rho_i}{\rho_j} = \left(\frac{A_j}{A_i}\right)^\frac{1}{\gamma}\), and the integral continuity equation becomes:

\[
\rho_i = \sum_j N \left(\frac{A_j}{A_i}\right)^\frac{1}{\gamma} m_j W_{ij} \quad (4.36)
\]

This has the advantage that, in the absence of shocks or thermal sources/sinks, \(\dot{A} = 0\) and so the error term \(\epsilon = 0\) by construction (see equation 4.21). In practice, however, we find no appreciable difference between the energy and entropy forms of SPH for the tests presented in this paper. This suggests that \(\epsilon\) is not a significant source of error.

Equations 4.35 and 4.36 retain the desirable integral form for the density, while giving significantly improved error properties. They also have a second important advantage that we
discuss in \S 4.3.5. We refer to equation 4.35 as the ‘RT’ density estimator for the energy form of SPH; and equation 4.36 as the RT density estimator in the entropy form.

**Minimising errors: the momentum equation**

The momentum equation (equation 4.27) is more problematic than the continuity equation. Its accuracy is governed not only by the extent to which \( V_i = I \), but primarily by the leading \( E_0,i \) term that should vanish.

First, consider the situation in standard SPH where \( \phi = 1 \) and \( g_{ij} = \rho_j/\rho_i \). As for the continuity equation, in the continuum limit (\( N \to \infty; h \to 0 \)), \( E_0 = 0 + O(h^2) \) since \( \nabla^2_i W_{ij} \) is antisymmetric. However, this analysis is only relevant if the particles are smoothly distributed on the kernel scale. For irregularly distributed particles, \( g_{ij} \) can grow arbitrarily large, while \( m_j/\rho_j \) is not guaranteed to be a good volume estimate. In such situations, \( E_0 \) contributes a significant error. Worse still, moving to higher resolution is not guaranteed to help. In order for the SPH integration to converge as \( h \to 0 \), we require that \( E_{0,i} \) shrinks faster than \( P_i \rho_i^{-1} \). This requires some care in making sure that \( h \) does not shrink too fast as the number of particles is increased.

A density step is an extreme example of an irregular particle distribution, and this suggests that the \( E_0 \) error is at least in part responsible for SPH’s failure to correctly model mixing processes between different fluid phases. We demonstrate this in \S 4.5.

There are three key problems with ensuring that \( P_i \rho_i^{-1} E_{0,i} \)
Figure 4.1: Stability plots for the Cubic Spline (CS) kernel (equation 4.45) in OSPH. The plots show contours of the frequency $\omega^2/k^2/c_s^2$ of plane waves impacting a regular lattice of particles, as a function of the wavenumber $k$ and the smoothing length $h$, in units of the inter-particle spacing $dx = 1$. From left to right the plots show $(k_x, k_y, k_z) = k(1, 0, 0), k(1, 1, 0)$ and $k(1, 1, 1)$. The three rows show the longitudinal wave and the two transverse waves for each of these orientations. Also marked by the blue dashed lines are the $h$ that corresponds to 32 neighbours (bottom line) and 128 neighbours (top line). OSPH is unstable if $\omega^2 < 0$ (grey regions).
4.3. **OPTIMISED SPH**

Core Triangle (CT) kernel:

Figure 4.2: As Figure 4.1, but for the Core Triangle (CT) kernel.
High Order Core Triangle HOCT4 kernel:

Figure 4.3: As Figure 4.1, but for the High Order Core Triangle (HOCT) kernel with $n_k = 4$. 
will shrink with increasing resolution. The first is the function $g_{ij}$. In SPH, this is the ratio $\rho_j / \rho_i$ which is large when there are large density gradients. We can significantly improve on this if we choose our free function $\phi = \rho$. In this case, we have $g_{ij} = g_{ij}^{-1} = 1$ by construction, and $g_{ij}$ no longer contributes to the $E_0$ error even for large density gradients across the kernel. The second problem relates to kernel scale smoothness. If particles clump or band on the kernel scale, then we will have poor kernel sampling and $E_0$ will not approach its integral limit even at very high resolution. Ensuring that this does not happen means ensuring that our OSPH scheme is stable to perturbations. We discuss this next in §4.3.4. The third and final problem is the volume estimate of each particle $m_j / \rho_j$. This will be poor if the particles are irregularly distributed on the kernel scale (for example at a density step) leading to a large $E_0$ error. We discuss this further in §4.3.5.

The choices $\zeta = \eta = 1/u; \phi = \rho$ and the kernel constraints $\overline{H}_{ij}r_{ij} = \overline{L}_{ij}r_{ij} = \overline{K}_{ij}r_{ij} = \nabla_i \overline{W}_{ij}$ are the first important ingredients in our OSPH scheme. These choices mean that we are no longer coherent, but this only introduces tolerable $O(h^2)$ errors in the energy conservation (Hernquist &

---

4It is interesting to note that other work in the literature has also found that $\phi = \rho$ is the preferred choice if density gradients are large (Oger et al. 2007; Price 2005; Ritchie & Thomas 2001; Dilts 1999). But we could not find a detailed proof similar to that presented here. Interestingly, Marri & White (2003) find empirically that $\phi = \rho^{3/2}$ gives the best performance for multiphase flow in their tests. Our analytic results here suggest that this is not the optimal choice, though perhaps the inclusion of cooling and/or other physics makes a difference.
4.3.4 Stability: the choice of kernel function

In §4.3.3, we used an error analysis of the generalised SPH equations of motion to show that the dominant source of error in SPH is in the momentum equation – the $E_0$ error. We showed that choosing the free functions $\zeta = \eta = 1/u; \phi = \rho$ and the kernel constraints $H_{ij} r_{ij} = L_{ij} r_{ij} = K_{ij} r_{ij} = \nabla_i W_{ij}$ should minimise both this error and errors in the continuity equation, and we called these choices OSPH.

In OSPH, provided the particles are regularly distributed on the kernel scale, we can make $E_0$ arbitrarily small simply by increasing the neighbour number. However, if the particles are irregularly distributed, $E_0$ can shrink very slowly with increasing resolution. In this section, we show that for large neighbour number the cubic spline kernel typically used in SPH calculations is unstable to both particle clumping (§4.3.4) and particle banding (4.3.4), and we derive a new class of kernels that are stable to both even for large neighbour number. In §4.5, we show that these new kernels give significantly im-

---

5Note that it is possible to construct pseudo-coherent versions of OSPH using $\zeta = \eta = \phi = 1/u$ for the energy form, or $\zeta = \eta = \phi = 1/A^{1/5}$ for the entropy form. Introducing ‘grad $h$’ terms as in Nelson & Papaloizou (1994), such schemes can then be made to conserve energy exactly in the limit of constant timesteps. However, they are only truly coherent up to the approximation that the $\epsilon$ error in the continuity equation is small (see equation 4.21). Nonetheless, it would be interesting to explore such schemes in future work.
4.3. **OPTIMISED SPH**

proved performance at no additional computational cost.

**The clumping instability**

The clumping instability\(^6\) can be derived from a linear 3D stability analysis of the OSPH equations of motion. Following Morris 1996a and Morris 1996b, we imagine a lattice of equal masses \(m\) of equal separation \((\Delta x_0, \Delta y_0, \Delta z_0)\) with initial density \(\rho_0\) and pressure \(P_0\). We perturb these with a linear wave of the form:

\[
\begin{align*}
    x_i &= x_{0,i} + a \exp[i(k \cdot x_{0,i} - \omega t)] \\
    \rho_i &= \rho_0 + D \exp[i(k \cdot x_{0,i} - \omega t)] \\
    P_i &= P_0 + c_s^2 D \exp[i(k \cdot x_{0,i} - \omega t)]
\end{align*}
\]  

(4.37)  

(4.38)  

(4.39)

and similar for particle \(j\), where \(c_s^2 = \frac{\partial P}{\partial \rho} = \gamma P_0/\rho_0\) is the sound speed assuming an adiabatic equation of state (\(\gamma = 1\) gives an isothermal equation of state), \(a = (X, Y, Z)\) is the amplitude of the perturbation and \(k = (k_x, k_y, k_z)\) is the wave vector.

To simplify the analysis, we assume that we have a lattice symmetry such that for every displacement vector \(x_{0,i,j} = x_{0,i} - x_{0,j}\) to a neighbour, there is also one at \(-x_{0,i,j}\). Then, plugging equations 4.37, 4.38 and 4.39 in to 4.17, discarding terms higher than first order, and connecting \(D\) to \(X, Y, Z\)

---

\(^6\)Also called the tensile instability.
through the continuity equation\(^7\)

\[
D = m \sum_j (1 - e^{i \phi_{ji}}) \left[ \frac{\partial \overline{W}_{ij}}{\partial x_i} X + \frac{\partial \overline{W}_{ij}}{\partial y_i} Y + \frac{\partial \overline{W}_{ij}}{\partial z_i} Z \right],
\]

we obtain the 3D OSPH dispersion relation\(^8\):

\[
\omega^2 a = \left[ \frac{2m P_0}{\rho_0^2} \sum_j H(\overline{W}_{0,ij})(1 - \cos k \cdot x_{0,ij}) + \right.
\]

\[
(\gamma - 2) \frac{m^2 P_0}{\rho_0^3} (q_i \wedge q_i) \cdot a
\]

(4.41)

where \(q_i \wedge q_i\) is the outer product of \(q_i\), \(H(\overline{W})\) is the Hessian\(^9\) of \(\overline{W}\):

\[
H_{aa} = \frac{\partial^2 \overline{W}}{\partial x_a^2} = \frac{d^2 \overline{W}(r)}{dr^2} \frac{x_a^2}{r^2} + \frac{d \overline{W}(r)}{dr} \frac{1}{r} \left( 1 - \frac{x_a^2}{r^2} \right)
\]

(4.42)

\[
H_{ab} = \frac{\partial \overline{W}}{\partial x_a \partial x_b} = \frac{d^2 \overline{W}(r)}{dr^2} \frac{x_a x_b}{r^2} - \frac{d \overline{W}(r)}{dr} \frac{x_a x_b}{r^3}
\]

(4.43)

\(^7\) We use here the full OSPH continuity equation in the entropy form (equation 4.16 with \(\eta = 1/A^{1/\gamma}\)). However, for plane waves on a constant density lattice \(A_j/A_i = 1\) and so this is identical to the SPH continuity equation with \(\eta = 1\).

\(^8\) This is actually identical to the SPH dispersion relation derived under the same assumptions in 3D (Morris, 1996b).

\(^9\) Recall that the outer product of two vectors is a matrix, while the Hessian is a square matrix of second order partial derivatives.
and \( q_i \) is given by:

\[
q_i = \sum_j \sin k \cdot x_{0,ij} \nabla_i W_{0,ij} \tag{4.44}
\]

Our scheme is stable if \( \omega^2 \geq 0 \). It is also desirable for the numerical sound speed to equal the true sound speed: \( \omega^2/k^2 = c_s^2 \).

In SPH it is typical to use the cubic spline (CS) kernel given by:

\[
W = \frac{8}{\pi h^3} \begin{cases} 
1 - 6x^2 + 6x^3 & 0 < x \leq \frac{1}{2} \\
2(1 - x)^3 & \frac{1}{2} < x \leq 1 \\
0 & \text{otherwise}
\end{cases} \tag{4.45}
\]

where \( x = r/h \) is the distance from the centre of the kernel in units of the smoothing length.

In Figure 4.1, we show contours of \( \omega^2/k^2/c_s^2 \) as a function of wavenumber \( k \) and the smoothing length \( h \) in units of the inter-particle spacing \( dx = 1 \), for the CS kernel. We assume an adiabatic equation of state with \( \gamma = 5/3 \). From left to right the plots show \( (k_x, k_y, k_z) = k(1, 0, 0), k(1, 1, 0) \) and \( k(1, 1, 1) \). The three rows show the longitudinal wave and the two transverse waves for these orientations.

From Figure 4.1, it is clear that the CS kernel in 3D is unstable to longitudinal waves for \( h \gtrsim 2 \), and very unstable to transverse waves. The unstable longitudinal waves drive the clumping, or tensile instability that causes particles to clump on the kernel scale (Schuessler & Schmitt 1981; Thomas & Couchman 1992; Herant 1994; Morris 1996a; Monaghan 2000).
CHAPTER 4. RESOLVING MIXING IN SPH

Figure 4.4
4.3. OPTIMISED SPH

Figure 4.4: A Kelvin-Helmholtz instability (density ratio $R_\rho = 2$) at $\tau_{KH} = 1$ modelled with SPH, TSPH and OSPH using CS, CT and HOCT4 kernels (see equations 4.45, 4.46 and 4.47). From left to right the plots show, in a slice of width $dx = 1$ about the z-axis: density contours; a zoom-in on the particle distribution around one of the rolls; the magnitude of the $|E_0|$ error (see equation 4.28) as a function of $y$; and the pressure in a slice of width $dx = 1$ about the x-axis, as a function of $y$. The circles on the density contour plots mark the size of the smoothing kernel, $h$. 
The clumping instability is a problem because it means that increasing the neighbour number will not give improved sampling of the kernel, and the $E_0$ error will remain large. However, the situation is dramatically improved if we add a constant central core to the kernel gradient $\frac{\partial W}{\partial r}$. This gives a constant force term at the centre of the Kernel that physically prevents clumping. We choose a kernel that is maximally similar to the CS kernel, while obeying $\frac{\partial W}{\partial r} = \text{const. \forall r < \alpha}$, where $\alpha$ is the core size. This leads us to the Core Triangle (CT) kernel:
4.3. **OPTIMISED SPH**

**Figure 4.6**

- **TSPH-CT-128**
- **TSPH-HOCT4-442**
- **OSPH-HOCT4-442**
Figure 4.6: Long term evolution of the KH instability in TSPH and OSPH versus the Eulerian code RAMSES. From left to right, the panels show density contours in a slice of width $dx = 1$ about the z-axis at times $\tau_{KH} = 1, 2$ and 3.
Figure 4.7: A Sod shock tube test in SPH (top) and OSPH (bottom). From left to right the panels show the variation in density, velocity, pressure and temperature across the shock, respectively. The blue lines give the analytic solution. This test was performed in 3D.
SPH-CS-32; 31,686 particles in the blob

TSPH-HOCT4-442; 126,744 particles in the blob

OSPH-HOCT4-442; 126,744 particles in the blob

FLASH; 128 × 128 × 384, 8785 cells in the blob, no refinement

Figure 4.8: The blob test in SPH, TSPH, OSPH and the Eulerian code FLASH. From left to right the plots show density contours at times $\tau_{KH} = 0, 1, 2$ and 3. The contour bar gives logarithmic density in cgs units.
Figure 4.9: The magnitude of the $|E_0|$ error in a slice of width $dx = 1$ about the z-axis, as a function of $y$ (top); and the pressure in a slice of width $dx$ about the x-axis and $dx$ about the z-axis, as a function of $y$ (bottom) for the blob test in SPH-32 (left), TSPH-HOCT4-442 (middle) and OSPH-HOCT4-442 (right) at time $\tau_{KH} = 1$. 
Chapter 4. Resolving Mixing in SPH

\[
W = \frac{N}{h^3} \begin{cases} 
-12\alpha + 18\alpha^2 & 0 < x \leq \alpha \\
1 - 6x^2 + 6x^3 & \alpha < x \leq \frac{1}{2} \\
2(1 - x)^3 & \frac{1}{2} < x \leq 1 \\
0 & \text{otherwise}
\end{cases}
\] (4.46)

where \( \beta = 1 + 6\alpha^2 - 12\alpha^3 \), \( N = 8/\left[\pi \left(6.4\alpha^5 - 16\alpha^6 + 1\right)\right] \), and the core size is fixed at \( \alpha = 1/3 \) by the requirement that \( \frac{\partial^2 W}{\partial r^2} \) be continuous.

Figure 4.2 shows stability plots for the CT kernel. The CT kernel has greatly improved stability for the longitudinal waves (top row) compared to the CS kernel and should give significantly improved performance for large neighbour number. We demonstrate this in §4.5.

Note that for all of the kernels we use in this paper, we consistently apply the kernel for the density estimate and its gradient for the energy and momentum equations. For small neighbour number, the central triangle in the CT kernel will degrade the quality of the density estimate. However, in this paper we typically use large neighbour numbers (> 100). In this case, very few particles sample the inner regions of the kernel and the bias introduced in the density is negligible. (The quality of the density estimate in OSPH can be seen in the Sod shock tube test in §4.6.) We found in tests that retaining the CS kernel just for the density estimate gives near-identical results.
4.3. **OPTIMISED SPH**

**The banding instability**

The clumping instability is a result of unstable longitudinal waves. A related instability – the banding instability – is a result of unstable transverse waves. For both the CT and CS kernels, there are broad bands of instability to transverse waves (see Figures 4.1 and 4.2). If the neighbour number is carefully chosen to lie in a stable region, banding will not occur. However, banding can still be excited at boundaries if $h$ changes there, moving into an unstable region.

For both the CS and CT kernels, it is difficult to find a suitable neighbour number for which the kernel is stable to all transverse and longitudinal modes. This suggests hunting for an even more stable kernel. A full search is beyond the scope of this paper. Here, we present a simple class of kernels that improve stability by moving to higher order (Morris, 1996b). Following Price (2005), we generalise our CT kernel to order $n_k$ to obtain the following class of kernels that we call the High Order Core-Triangle (HOCT) kernels:

\[
W = \frac{N}{h^3} \begin{cases} 
Px + Q & 0 < x \leq \alpha \\
(1 - x)^{n_k} + A(\gamma - x)^{n_k} + B(\beta - x)^{n_k} & \alpha < x \leq \beta \\
(1 - x)^{n_k} + A(\gamma - x)^{n_k} & \beta < x \leq \gamma \\
(1 - x)^{n_k} & \gamma < x \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]  

(4.47)
where:

\[ A = \frac{1 - \beta^2}{\gamma^{n_k-3}(\gamma^2 - \beta^2)} \quad (4.48) \]
\[ B = -\frac{1 + A \gamma^{n_k-1}}{\beta^{n_k-1}} \quad (4.49) \]

\[ P = -n_k(1 - \alpha)^{n_k-1} - n_k A(\gamma - \alpha)^{n_k-1} - n_k B(\beta - \alpha)^{n_k-1} \quad (4.50) \]
\[ Q = (1 - \alpha)^{n_k} + A(\gamma - \alpha)^{n_k} + B(\beta - \alpha)^{n_k} - P \alpha \quad (4.51) \]

and \( \alpha \) and \( N \) are calculated numerically for a given choice of \( n_k \). Continuity requires that \( \alpha \) solves the equation:

\[ 0 = (1 - \alpha)^{n_k-2} + A(\gamma - \alpha)^{n_k-2} + B(\beta - \alpha)^{n_k-2} \quad (4.52) \]

where \( \beta \) and \( \gamma \) are free parameters. In this paper we choose \( \beta = 0.5, \gamma = 0.75 \). Other choices, and indeed other high-order kernels, may give better results than those presented here. We tabulate values for \( A, B, P, Q, \alpha \) and \( N \) as a function of \( n_k \) in Table 4.1. Notice that the core size \( \alpha \) decreases with \( n_k \).

Stability plots for the HOCT4 kernel (with \( n_k = 4 \)) are given in Figure 4.3. Notice the improvement over the CT kernel, particularly for the transverse waves. There are two bands where the kernel is fully stable to both longitudinal and transverse waves on a lattice: 96 neighbours and 442 neighbours, corresponding to \( h = 2.86 \) and \( h = 4.75 \), respectively. We use the latter choice since this also gives very low \( E_0 \). The CT kernel also has a stability band for \( h \sim 4.75 \), but this is narrower than for the HOCT4 kernel, while the HOCT4 kernel with this many neighbours gives better spatial resolution.
(It is important to realise that the smoothing length for different kernels takes on a different meaning in terms of spatial resolution. We suggest a resolution criteria based on the numerical sound speed versus the true sound speed for longitudinal waves: \( \omega^2/k^2/c_s^2 \). Spatial scales are well resolved if \( \omega^2/k^2/c_s^2 \approx 1 \). By this definition, our choice of 442 neighbours \((h = 4.75)\) for the HOCT4 kernel gives a very similar spatial resolution to 128 neighbours \((h = 3.14)\) for the CT kernel.)

Note that our stability analysis only applies for particles arranged on a lattice. Hexagonal close-packed particles, randomly arranged particles, and indeed boundaries may have different preferred stability regions. A full analysis is beyond the scope of this present work.

The banding instability is not as problematic as the clumping instability for the tests we present in §4.5 and §4.7. Unlike the clumping instability, it does not seem to (directly) play a major role in preventing mixing from occurring in SPH (see §4.5.2).

### 4.3.5 The local mixing instability & RT densities

Our error analysis in §4.3.3 missed one very important error term. This is because the Taylor expansion assumed that both the pressures and velocities in the flow are smooth. Unfortunately, in SPH at sharp boundary this is not the case. The reason for this is easiest to understand using the entropy form of
SPH, as follows (similar arguments apply also for the energy form).

Imagine a density step of ratio $R_\rho = \rho_1/\rho_2$ initially in pressure equilibrium, such that the entropy function (equation 4.22) is given by $A_1/A_2 = 1/R_\rho^\gamma$. Now imagine that we perturb the boundary very slightly by pushing a low density particle towards it. The particle’s entropy is conserved, but its density increases very rapidly proportional to $R_\rho$. This leads to an increase in pressure: $P_1 = P + \kappa_1 R$, where $\kappa_1$ is some constant that depends on the perturbation size and the kernel. On the other side of the boundary, if we push a high density particle towards the low density region, however, its density will rapidly decrease giving a decrease in pressure: $P_2 = P - \kappa_2 R_\rho$. This drives us towards a pressure discontinuity at the boundary which drives an associated error in the momentum equation. It can be thought of as a fundamental result of particles trying to mix on the kernel scale, but being

Table 4.1: Parameters for a selection of High Order Core-Triangle (HOCT) kernels. See equation 4.47 for details and definitions.

<table>
<thead>
<tr>
<th>$n_k$</th>
<th>$A$</th>
<th>$B$</th>
<th>$P$</th>
<th>$Q$</th>
<th>$\alpha$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.4</td>
<td>-9.4</td>
<td>-1.81</td>
<td>1.028</td>
<td>0.317</td>
<td>3.71</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>-18.8</td>
<td>-2.15</td>
<td>0.98</td>
<td>0.214</td>
<td>6.52</td>
</tr>
<tr>
<td>5</td>
<td>4.27</td>
<td>-37.6</td>
<td>-2.56</td>
<td>0.962</td>
<td>0.161</td>
<td>10.4</td>
</tr>
<tr>
<td>8</td>
<td>10.1</td>
<td>-300.8</td>
<td>-3.86</td>
<td>0.942</td>
<td>0.0927</td>
<td>30.75</td>
</tr>
</tbody>
</table>
unable to as a result of entropy conservation. We call this the local mixing instability (LMI).

Although not phrased in terms of the LMI, the LMI is a recognised problem in the literature and there are essentially two classes of solution. We can generate entropy at the boundary to give smooth entropies and therefore smooth pressures, as in (Wadsley et al., 2008) and Price (2008); or we can try to obtain sharper densities that are consistent with the discrete entropies. This is the approach adopted by Ritchie & Thomas (2001), and the approach we take in this paper. The key advantage of sharpening the densities is that we do not need to specify a subgrid mixing model.

The sharper densities we require are exactly what we get from the density estimate given in equation 4.36, and originally proposed by Ritchie & Thomas (2001) – the ‘RT’ density estimate. Consider the perturbation discussed above, but now using the RT density estimate\(^{10}\). A low density particle which has half of its kernel in the high density phase (an extreme example), will have a density:

\[
\rho_{\text{low}} = \sum_{j}^{{N}_l} m_j \overline{W}_{ij} + \sum_{j}^{{N}_r} \left( \frac{A_j}{A_i} \right)^{\frac{1}{\eta}} m_j \overline{W}_{ij} \tag{4.53}
\]

where \(N_l\) is the number of particles in the low density region, \(N_r\) is the number in the high density region, and we have used

\(^{10}\)We use here the entropy form given in equation 4.36 since we use the entropy form of SPH in this analysis. If instead, we use the energy form of SPH then we should use instead equation 4.35.
the fact that the ratio \( (A_j/A_i)^{\frac{1}{\gamma}} = 1 \) for the low density region.

If the simulation is adiabatic and started in pressure equilibrium, then for the high density region \( (A_j/A_i)^{\frac{1}{\gamma}} = 1/R_\rho \), and since the high density particles sample the kernel \( R_\rho \) times more often than the low density particles, we recover:

\[
\rho_{\text{low}} = \sum_j^N m_j A_j \gamma W_{ij}
\]

which is identical to a particle in the low density region. A similar derivation applies for a high density particle at the boundary. Thus, the RT density estimate ensures that densities remain sharp. It is straightforward to show that it also ensures the pressures are single valued throughout the flow. Substituting the RT density estimator (equation 4.36) into equation 4.22, we obtain:

\[
P_i = \left[ \sum_j^N m_j A_j^{\frac{1}{\gamma}} W_{ij} \right]^\gamma
\]

Notice that the entropy function \( A_j \) now appears inside the sum, whereas in standard SPH it would appear as \( A_i \) outside of the sum. This difference ensures that the \( P_i \) will be everywhere single valued throughout the flow – even at boundaries. A similar derivation can be made for the energy form of SPH, in which case we should use the density estimate given in equation 4.35.

The RT density estimate is robust to particle mixing on the kernel scale and should lead to a dramatically reduced LMI.
4.4. IMPLEMENTATION

We demonstrate this in §4.5. Furthermore, the RT density estimate ensures that our error analysis in §4.3.3 is valid by construction since it ensures smooth pressures (recall that we assumed that both the pressures and the velocities were smooth, but not the densities). And, since the RT density estimate leads to sharper densities, it gives improved volume estimates for the particles. This suggests that we can expect the RT density estimate to reduce $E_0$ at boundaries. We demonstrate this also in §4.5.

Note that the RT density estimate is chosen to ensure single valued pressures throughout the flow. However, when extracting results from a simulation, it is the positions of the particles themselves that describe the state of the fluid. This suggests using the density estimate in equation 4.1 for calculating the observable flow density, rather than the RT density estimate. This is the approach we adopt in this paper, though the difference is negligible.

4.4 Implementation

We implemented OSPH in GASOLINE (Wadsley et al., 2004), a parallel implementation of TreeSPH that uses a fixed number $N$ of smoothing neighbours\footnote{Allowing for varying neighbour number is needlessly dissipative (Nelson & Papaloizou 1994; Attwood et al. 2007).}, and a standard prescription for the artificial viscosity as in Gingold & Monaghan (1983) with $\alpha = 1$, $\beta = 2$, controlled with a Balsara switch (Balsara,
1989). We used variable timesteps controlled by the Courant time with a Courant factor of 0.4.

The improved stability and error properties of OSPH motivate a full re-examination of the standard SPH artificial viscosity. This is beyond the scope of this present work. However, we note that the improved stability in OSPH means that particles better follow characteristics of the flow, while the gradients in the Balsara switch will be less noisy. Both of these effects should act to decrease the viscosity in regions of steady flow. (Note that all numerical schemes carry numerical viscosity, whether it is manifested through limited resolution or artificial shock-capturing viscosity. Indeed, these viscous terms are vital for successfully modelling shocks.) In §4.5, §4.6 and §4.7, we show that our OSPH results agree very well with analytic expectations, and with the results from Eulerian codes. This suggests that the viscosity prescription in OSPH is not a significant source of error. Certainly it is not responsible for SPH’s inability to model mixing processes.

### 4.5 The Kelvin-Helmholtz instability

In this section, we use a 1:2 and 1:8 density ratio shearing fluid simulation to test mixing in OSPH. We use the naming convention XSPH-K-N, where X denotes the variety of SPH, K the choice of kernel, and N the neighbour number (see Table 4.2).
4.5. **THE KELVIN-HELMHOLTZ INSTABILITY**

<table>
<thead>
<tr>
<th>Flavour of SPH</th>
<th>(\eta)</th>
<th>(\phi)</th>
<th>(\zeta)</th>
<th>Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPH</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>CS</td>
</tr>
<tr>
<td>TSPH</td>
<td>1</td>
<td>(\rho)</td>
<td>1</td>
<td>CS, CT, HOCT4</td>
</tr>
<tr>
<td>OSPH</td>
<td>(1/u)</td>
<td>(\rho)</td>
<td>(1/u)</td>
<td>HOCT4</td>
</tr>
</tbody>
</table>

Table 4.2: The different flavours of SPH we explore in this work. The free functions \(\eta\), \(\phi\) and \(\zeta\) are defined in equations 4.16, 4.17 and 4.18. The Cubic Spline (CS) kernel is given by equation 4.45, the Core-Triangle (CT) kernel is given by equation 4.46, and the fourth order High-Oder Core Triangle (HOCT4) kernel is given by equation 4.47 with \(n_k = 4\).

### 4.5.1 Numerical set-up

A Kelvin-Helmholtz instability (KHI) occurs when two shearing fluids are subjected to an infinitesimal perturbation at the boundary layer. The result of the perturbation is a linearly growing phase in which the layers start to interpenetrate each other, progressively developing into a vortex in the non-linear phase that mixes the two fluid layers. The growth-rate of the instability is in general a complicated function of the shear velocity, fluid densities, compressibility, interface thickness, gravity, viscosity, surface tension, magnetic field strength etc. In this test, we are only interested in the behaviour of inviscid, incompressible (i.e. with bulk motions very much less than the sound speed) perfect fluids neglecting gravity. In this case, the
linear growth rate of the KHI is (Chandrasekhar 1961):

\[ w = k \left( \frac{\rho_1 \rho_2}{\rho_1 + \rho_2} \right)^{1/2} v, \]  

(4.56)

where \( k = \frac{2\pi}{\lambda} \) is the wavenumber of the instability, \( \rho_1 \) and \( \rho_2 \) are the densities of the respective layers and \( v = v_1 - v_2 \) is the relative shear velocity. The characteristic growth time for the KHI is then:

\[ \tau_{\text{KH}} \equiv \frac{2\pi}{w} = \frac{(\rho_1 + \rho_2)\lambda}{(\rho_1 \rho_2)^{1/2} v}. \]  

(4.57)

This is a particularly challenging test for SPH/OSPH because the velocity due to particle noise can approach the sound speed which can wash out the physical velocity perturbation relevant for this test.

We set up the problem in three dimensions using a periodic thin slab defined by \( x \in \{-0.5, 0.5\}, y \in \{-0.5, 0.5\} \) and \( z \in \{-1/64, 1/64\} \). The domain satisfied:

\[ \rho, T, v_x = \begin{cases} \rho_1, T_1, v_1 & |y| < 0.25 \\ \rho_2, T_2, v_2 & |y| > 0.25 \end{cases} \]  

(4.58)

The density and temperature ratio were

\[ R_\rho = \frac{\rho_1}{\rho_2} = \frac{T_2}{T_1} = \frac{c_2^2}{c_1^2}, \]  

(4.59)

ensuring that the whole system was pressure equilibrium. The two layers were given constant and opposing shearing velocities, with the low density layer moving at a Mach number
\( M_2 = -v_2/c_2 \approx 0.11 \) and the dense layer moving at \( M_1 = M_2 \sqrt{R_\rho} \). The density ratios considered in this work are small which assures a subsonic regime where the growth of instabilities can be treated using equation 4.57 (Vietri et al., 1997).

To trigger instabilities, velocity perturbations were placed on the two boundaries of the form:

\[
v_y = \delta v_y \left[ \sin \left( \frac{2\pi (x + \lambda/2)}{\lambda} \right) \exp\left( -\left(10(y - 0.25)\right)^2 \right) \right.
\]

\[
- \left. \sin \left( \frac{2\pi x}{\lambda} \right) \exp\left( -\left(10(y + 0.25)\right)^2 \right) \right] \tag{4.60}
\]

where the perturbation velocity \( \delta v_y/v = 1/8 \) and \( \lambda = 0.5 \) is the wavelength of the mode.

Equal mass particles were placed in lattice configurations to satisfy the setup described above. To satisfy pressure equilibrium everywhere, in TSPH the temperatures were adjusted at boundaries to be coherent with the smoothed density step measured by equation 4.1. This was not done for the OSPH simulations since these sharpen the densities using the discrete initial temperatures.

The low density region \( \rho_2 \) was set up using 256 particles in the \( x \)-direction and the appropriate number of particles in the other dimensions to satisfy a fixed inter-particle distance. The high density region \( \rho_1 \) was created in the same way with 320 particles in the \( x \)-direction. We adopted a periodic simulation domain.

The RAMSES simulation used the same numerical set-up as described above, but in 2D rather than in a thin slab. We performed the \( R_\rho = 2 \) simulation using the LLF Riemann solver (Toro, 1999) on a \( 256 \times 256 \) fixed Cartesian grid. The
CHAPTER 4. RESOLVING MIXING IN SPH

LLF solver is rather diffusive and is used in order to suppress the growth of undesirable small scale KHI s arising from grid irregularities.

We note that all numerical schemes carry numerical viscosity, whether it is manifested through limited resolution or artificial shock-capturing viscosity. A detailed study of this effect on the KHI and the relation to physical viscosity is beyond the scope of this paper.

4.5.2 Results

Figure 4.4 shows our results for the KHI test (density ratio $R_\rho = 2$) at $\tau_{KH} = 1$ modelled with SPH, TSPH and OSPH, using three different kernels: CS, CT and HOCT4, and different neighbour numbers as marked on each plot (see also Table 4.2). From left to right, the panels show, in a slice of width $dx = 1$ about the z-axis: density contours of the simulation box, a zoom in on the particle distribution around one of the rolls, the magnitude of the $|E_0|$ error (equation 4.28) as a function of $y$, and the pressure as a function of $y$ in a slice of width $dx = 1$ about the x-axis.

The clumping instability

Using the standard CS kernel, SPH-CS-128 (top row, Figure 4.4) and TSPH-CS-128 (second row) give poor results that improve very slowly with increasing neighbour number. This can be seen both in the lack of strong evolution on the boundary,
4.5. THE KELVIN-HELMHOLTZ INSTABILITY

and in the large $|E_0|$ error, even for 128 neighbours. TSPH-CS-128 gives slightly better results than SPH-CS-128, showing the first beginnings of a KHI roll, but both are in poor agreement with the RAMSES results (bottom row).

The reason for the poor performance in both SPH-CS-128 and TSPH-CS-128 is the clumping instability (§4.3.4). Particles gather together on the kernel scale, giving poor kernel sampling, and poor associated error. This can be seen in the particle distribution for SPH-CS-128 and TSPH-CS-128 (second row, Figure 4.4) which show visible holes and overdensities in the particle distribution. Using instead the CT kernel introduced in §4.3.4, the results improve dramatically (third row, Figure 4.4). Now the errors reduce for increasing neighbour number (see Appendix 4.9). With 128 neighbours, we successfully resolve a KH roll up to $\tau_{KH} = 1$ with the correct growth time.

It has been noted previously in the literature that putting a small core inside a cubic spline kernel suppresses the clumping instability (Thomas & Couchman 1992; Herant 1994), but its importance for modelling multiphase flow was not realised. One alternative fix includes adding a negative pressure term (Monaghan, 2000), which in tests we find works also. However, we prefer changing the kernel to introducing new forces since we may then still estimate our errors through $|E_0|$.

The banding instability

In addition to the clumping instability, there is also an instability to transverse waves – the banding instability (§4.3.4). For
the KHI tests we present here, the banding instability occurs only on the boundary and appears to be relatively benign. This is shown in Figure 4.5, that shows a zoom in on the boundary at $\tau_{KH} = 1$ for TSPH-CT-128, TSPH-HOCT4-442 and OSPH-HOCT4-442. The TSPH-CT-128 simulation has a kernel and neighbour number combination that are unstable to transverse waves (see Figure 4.2), and banding is clearly visible on the boundary. However, TSPH-HOCT4-442 should be stable to transverse waves, yet the banding persists. Only in our full scheme, OSPH-HOCT4-442, is the banding is gone.

To understand the above results, we ran an additional test that we omit for brevity – TSPH-HOCT4-96. This simulation showed little boundary evolution because the low neighbour number and associated large $|E_0|$ significantly damped the KHI. However, interestingly, there was no banding observed on the boundary (recall that the HOCT4 kernel with 96 neighbours should be stable to both transverse and longitudinal wave perturbations).

Taken together, our results suggest that the observed banding at the boundary is a result of a transverse wave instability driven by the local mixing instability (LMI; §4.3.5). Where there is little evolution at the boundary and the kernel is chosen to be stable to transverse waves, the banding disappears, as was the case for our extra TSPH-HOCT4-96 simulation. Where there is strong evolution at the boundary, as was the case for TSPH-HOCT4-442, the LMI drives banding irrespective of the choice of kernel. Only in our full scheme, OSPH-HOCT4-442, where the LMI is cured and the kernel is stable to transverse waves, is the banding cured.
4.5. THE KELVIN-HELMHOLTZ INSTABILITY

The $E_0$ error

Away from boundaries, the $E_0$ error in TSPH decreases with the neighbour number, as expected for smooth flow (see Appendix 4.9). However, on the boundary the $|E_0|$ error grows by 2-3 orders of magnitude. Increasing the neighbour number does result in better long-term evolution, but the results improve very slowly. This is shown in Figure 4.6. Notice that TSPH-HOCT4-442 resolves two wraps of the KH roll at $\tau_{KH} = 2$, whereas TSPH-CT-128 only manages one. However, even in TSPH-HOCT4-442, the long-term evolution eventually degrades. By $\tau_{KH} = 3$, the results are ‘gloopy’, rather similar to simulations that explicitly model fluid surface tension (see e.g. Herrmann 2005).

The poor $E_0$ on the boundary is the result of a poor volume estimate for each particle $m_j/\rho_j$ (see §4.3.3). However $E_0$ is not solely responsible for the gloopy behaviour. There is a second problem – similar to a numerical surface tension term – that needs to be solved extra to minimising $E_0$. This is the local mixing instability (LMI) error (§4.3.5).

The local mixing instability error

The right panels of Figure 4.4 show the pressure as a function of $y$ in a slice of width $dx = 1$ about the $z$-axis and width $dx = 1$ about the $x$-axis. In SPH and TSPH there is a clear pressure discontinuity on the boundary. This is caused by the local mixing instability (LMI) discussed in §4.3.5.

Notice that the pressure blip is larger in TSPH than in SPH,
yet the KHI roll progresses further in TSPH than in SPH. This apparent paradox is the result of the improved performance in TSPH. As the KHI roll progresses in TSPH, particles are pushed closer to the boundary making the LMI worse and increasing the pressure blip. In SPH, there is a larger gap at the boundary due to the larger surface tension error. This leads to less evolution and a smaller associated pressure blip. We will see a similar effect occurring in the blob test in §4.7.

As discussed in §4.3.5, the LMI should be cured by the RT density estimate (equation 4.35). This is shown in Figure 4.4, third row which shows the results for our full OSPH scheme. With the RT densities, the pressure at the boundary has a much smaller blip, while $E_0$ is reduced by over an order of magnitude. This latter effect occurs since the RT densities also give an improved volume estimate for each particle (see §4.3.3 and §4.3.5). The long term evolution is now in excellent agreement with the RAMSES results (compare Figure 4.4 third and bottom rows).

Although OSPH gives significantly improved results compared to SPH, the scheme is numerically expensive. Simulations with larger density gradients require very high resolution. This is shown in Figure 4.6, second from bottom row. This shows the long term evolution of a KHI test with density ratio $R_\rho = 8$ in OSPH. The solution should be similar to the $R_\rho = 2$ simulation, but it is not. The ‘gloopy’ behaviour indicative of large surface tension errors has returned. Further increasing the neighbour numbers would reduce this problem, but at increased numerical cost. We will address this issue in future work (Hayfield & Read in prep.).
4.6 The Sod shock tube

Before we embark on the blob test in §4.7, it is worth checking that our new OSPH scheme can still correctly resolve shocks. To test this, we use a standard Sod shock tube test (Sod, 1978).

The Sod shock tube consists of a 1D tube on the interval $[-0.5, 0.5]$ with a discontinuous change in properties at $x = 0$ designed to generate a shock. The left state is described by $\rho_l = 1.0, P_l = 1.0, v_l = 0$, and the right state by $\rho_r = 0.125, P_r = 0.1, v_r = 0$, where $\rho, P$ and $v$ are the density, pressure and velocity along the $x$ axis. We use an adiabatic equation of state with $\gamma = 1.4$. The subsequent evolution of the problem has a self-similar analytic solution that has a number distinct features which quite generally test a code’s conservation properties, artificial viscosity, ability to handle nonlinear waves, and shock resolution.

Figure 4.7 shows the results for the Sod shock tube test at time $t = 0.2$ in SPH (top) OSPH (bottom). Since we are primarily concerned with the 3D performance of the code, the test was performed in 3D on the union of a $24 \times 24 \times 300$ lattice on the left, with a $12 \times 12 \times 150$ lattice on the right, giving a 1D resolution of 450 points. We use 442 neighbours for this test in both SPH and OSPH to ensure that any difference is not simply due to improved kernel sampling in OSPH.

For SPH, the only strong disagreement with the analytic solution is in the pressures that have a blip at $x = 0.2$, and the temperatures that overshoot at $x = 0.2$. The former feature is due to the LMI (see §4.3.5 and §4.5.2). The latter feature is seen in all SPH Sod shock tube tests and results from the well-
known ‘wall heating’ effect (Noh, 1987). This is an error due to the artificial viscosity prescription and is beyond the scope of this present work.

For OSPH, the results are even better than for SPH. The pressure blip is now gone, while the temperature overshoot at $x = 0.2$ is reduced. Only the velocities appear to be worse, with some remaining dispersion at $x = 0.2$. This owes to the jump in density at this point, and the associated jump in $|E_0|$. This gives a force error at the discontinuity which introduces some dispersion into the velocities. In SPH this cannot occur since the LMI causes a pressure blip at the boundary that prevents mixing. We will discuss this issue further in a forthcoming paper (Hayfield & Read in prep.).

### 4.7 The blob test

The KHI test presented in §4.5 is a worst-case scenario for OSPH, since it has a pure adiabatic sharp boundary. For many practical situations, boundaries will be less sharp, while physical entropy generation due to shocks and/or cooling will suppress the LMI. We give a practical example of this in this section using the blob test described in Agertz et al. (2007). A spherical cloud of gas of radius $R_{cl}$ is placed in a wind tunnel with periodic boundary conditions. The ambient medium is ten times hotter and ten times less dense than the cloud so that it is in pressure equilibrium with the latter. We refer to the initial density contrast between the cloud and the medium as $R_{\rho,ini}$. The wind velocity ($v_{\text{wind}} = c_s M$) has an associated
4.7. THE BLOB TEST

Mach number $M = 2.7$. This leads to the formation of a bow shock after which the post-shock subsonic flow interacts with the cloud and turns supersonic as it flows past it.

The blob test is useful for investigating how different hydrodynamics codes model astrophysical processes important for multiphase systems, such as shocks, ram-pressure stripping and fragmentation through KH and Rayleigh-Taylor (RT) instabilities. As $\tau_{\text{KH}} < \tau_{\text{RT}}$, an approximate timescale of the cloud destruction is that of the full growth of the largest KH mode, i.e. the wavelength of the cloud’s radius. This can be obtained by considering the post-shock flow on the cloud and its time-dependence as the shock weakens and the cloud is accelerated. A full analysis of this test is presented in Agertz et al. (2007) and gives $\tau_{\text{KH}} \approx 1.6\tau_{\text{cr}}$ where $\tau_{\text{cr}} = 2R_{\text{cl}}R_{\rho,\text{ini}}^{1/2}/v$ is the crushing time and the velocity $v$ refers to the streaming velocity in the reference frame of the cloud. After this time, the cloud is expected to show a more complicated non-linear behaviour leading to disruption. The original blob test was initialised in a glass-like configuration obtained using simulated annealing using a standard SPH code. Since we now use OSPH rather than SPH, we must set up new ICs for the blob. Our new IC set up is described in detail in Appendix 4.10. Unlike the previous blob test, where perturbations were seeded by random noise in the particle distribution, here we deliberately seed an inward growing mode on the front surface of the blob. This makes comparison between OSPH and FLASH simpler, since then the morphology of the blob is less dependent on small scale numerical noise.
The results are presented in Figure 4.8, where we compare SPH, TSPH and OSPH with increasing resolution with similar results from the Eulerian code \textsc{flash} (Fryxell et al., 2000). The SPH results (top panels) are similar to those presented in Agertz et al. (2007). The blob is squashed by the shock, but does not break up. There are no visible surface KHI or Rayleigh-Taylor instabilities. TSPH (second row) gives significantly improved results. The central depression is now resolved and the blob is mostly destroyed by $\tau_{\text{KH}} = 3$. However, the density remains clumpy as compared to the \textsc{flash} simulation (bottom row). Our full OSPH scheme (third row) gives excellent agreement with the \textsc{flash} results. There are clear surface KHI and RT instabilities and the blob breaks up fully by $\tau_{\text{KH}} = 3$. The precise details of the break up in \textsc{flash} and OSPH are different. However, these differences are smaller than those we observed between \textsc{flash} simulations of varying resolution. They are caused by the non-linear break up of the blob that is affected by resolution-scale perturbations.

Figure 4.9 shows $E_0$ and the pressure blips for the blob test in SPH (left), TSPH (middle) and OSPH (right) at $\tau_{\text{KH}} = 1$. In SPH and TSPH, the two fluid phases (marked by the black and grey solid circles) remain well separated at all times. In both cases the pressure distribution shows discontinuities. By contrast, in OSPH the fluids are already mixed at $\tau_{\text{KH}} = 1$, while the pressures are smooth and single-valued throughout the flow.

There is a more modest improvement in $E_0$ between SPH, TSPH and OSPH than that seen in the KHI tests presented in §4.5. OSPH gives a $E_0$ smaller by a factor $\sim 5$ compared to the
4.8. CONCLUSIONS

SPH and TSPH simulations, whereas in the KHI tests, there was an improvement of over an order of magnitude. There are two reasons for this. Firstly, since the SPH simulation shows little evolution at the boundary, the initial $E_0$ is relatively well conserved. By contrast, TSPH shows significant boundary evolution due to improved mixing. This can actually worsen $E_0$ at the boundary since the particles are still unable to properly interpenetrate as a result of the LMI. Figure 4.8 clearly shows, however, that TSPH gives improved mixing. Secondly, entropy generation at the shock softens the density step around the blob, leading to an improved $E_0$ even for the SPH simulation.

4.8 Conclusions

Standard formulations of smoothed particle hydrodynamics (SPH) cannot resolve fluid mixing and instabilities at flow boundaries. We have used an error and stability analysis of the generalised SPH equations of motion to show that mixing fails for two distinct reasons. The first is a leading order error in the momentum equation. This should decrease with increasing neighbour number, but does not because numerical instabilities cause the kernel to be irregularly sampled. We identified two important instabilities: the clumping instability and the banding instability, and we showed that both are cured by a suitable choice of kernel. The second problem is the local mixing instability (LMI). This occurs as particles attempt to mix on the kernel scale, but are unable to due to entropy
conservation. The result is a pressure discontinuity at boundaries that pushes fluids of different entropy apart. We cured the LMI by using a weighted density estimate proposed by Ritchie & Thomas (2001). We showed that this both reduces errors in the continuity equation and allows individual particles to mix at constant pressure.

We demonstrated mixing in our new Optimised Smoothed Particle Hydrodynamics (OSPH) scheme using a Kelvin Helmholtz instability (KHI) test with density contrast 1:2, and the ‘blob test’ – a 1:10 density ratio gas sphere in a wind tunnel – finding excellent agreement between OSPH and Eulerian codes.

OSPH is a multiphase Lagrangian method that conserves momentum, mass and entropy, and demonstrates that it is possible to model multiphase fluid flow using SPH. OSPH remains a low-order method, requiring large neighbour number to keep the $E_0$ error small. We will address this problem in a forthcoming paper, where we use the lessons learnt in this present work to move to higher order particle methods (Hayfield & Read in prep.).

### 4.9 The effect of increasing neighbour number in TSPH

In this appendix we show the effect of increasing neighbour number for TSPH-CT (i.e. without the clumping instability). The results for the same KH instability test shown in Figure...
4.10. **THE BLOB TEST SETUP**

4.4 are shown in Figure 4.10 for 32 and 64 neighbours. Notice that with increasing particle number, the error vector $E_0$ is reduced, the pressure blip at the boundary is reduced, and the results improve.

### 4.10 The blob test setup

The hydrodynamical properties of the blob test are described in §4.7. We use a periodic simulation box of size, in units of the cloud radius $R_{cl}$, $\{L_x, L_y, L_z\} = \{10, 10, 30\}$ and we centre the cloud at $\{x, y, z\} = \{5, 5, 5\}$. The equal mass SPH particles constituting the ambient medium and the cloud are arranged in lattice configurations to achieve the relevant density.
contrast $R_{\rho, ini} = 10$. The particle temperatures ($T \sim P/\rho$) are then assigned to achieve pressure equilibrium where the local density measurement of equation 4.1 is used for consistency. The wind velocity ($v_{\text{wind}} = c_s M$) has an associated Mach number $M = 2.7$, where the sound speed is $c_s = \sqrt{\gamma P/\rho}$ using an adiabatic index $\gamma = 5/3$.

In the original blob test described in Agertz et al. (2007), SPH particle noise was used to trigger instabilities. This procedure is not applicable when using a noise free lattice configuration. Hence, we use spherical harmonics to apply large scale perturbations to the surface layer of the cloud. The full perturbation, in spherical coordinates centred on the cloud, can be expressed as

$$v_{\text{pert}}(r, \theta, \phi) = \delta v R(r) \text{Re}[Y(\theta, \phi)^m)]/C,$$

where $R(r) = \exp \left(2(r - r_{\text{cl}})/r_{\text{cl}} \right)$ (the radial component), is defined for $r \leq r_{\text{cl}}$ and the spherical harmonic is, adopting $l = 5, m = 3$:

$$Y_5^3 = \frac{-1}{32} \sqrt{\frac{385}{\pi}} e^{3i\phi} \sin^3 \theta (9 \cos^2 \theta - 1). \quad (4.61)$$

The constant $C$ simply normalises the real part of the harmonic to reach a maximum value of 1. We chose a subsonic perturbation $\delta v = -0.06 v_{\text{wind}}$. 


Acknowledgements

I am very grateful to my supervisor Lucio Mayer for his support and encouragement, and for allowing me the freedom to explore my own interests and ideas, for letting me to make mistakes and learn from them, and encouraging me to take up collaborations with colleagues during my time here.

I am honored to have collaborated with several brilliant post-docs in my time here, Aaron Boley, Justin Read, and Michela Mapelli. They have set a standard which I hope can live up to in the future.

I am grateful for the many enjoyable moments I have shared with my ETH and University colleagues over the years: Oscar Agertz, Christian Knobel, Simone Callegari, Yinjie Peng, Andrea Banzatti, Pawel Kampczyk, Richard Wenzel, Marina Galvagni, and Tobias Motz to name but a few.

I thankful that Timo Schmidt’s German is better than mine.

I am eternally indebted to Tina Wentz, Aaron Manalaysay, and Valérie Keller for being great friends and for being of great
assistance me.

My greatest thanks goes out my mother Janice and my stepfather Marc Guilbeault for always believing in me.
Bibliography


Bennett A., 2006, Lagrangian Fluid Dynamics


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


Chandrasekhar S., 1961, Hydrodynamic and hydromagnetic stability


Greengard L., Rokhlin V., 1987, Journal of Computational Physics, 73, 325


Herrmann M., 2005, Journal of Computational Physics, 203, 539


Jeans J. H., 1902, Royal Society of London Philosophical Transactions Series A, 199, 1


BIBLIOGRAPHY


Noh W., 1987, Journal of Computational Physics, 72, 78


Price D., 2005, ArXiv Astrophysics e-prints


Spitzer L., 1968, Diffuse matter in space


Stahler S. W., Palla F., 2005, The Formation of Stars


Tohline J. E., 1982, Fundamentals of Cosmic Physics, 8, 1


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


