



# SPH and its Application to Stellar Disruption

Lee Cullen

Submitted in total fulfilment of the requirements of the degree of  
Doctor of Philosophy

2010

Theoretical Astrophysics Group  
Department of Physics and Astronomy  
The University of Leicester

---

# ABSTRACT

## SPH and its Application to Stellar Disruption

by

Lee Cullen

In this thesis we study Smoothed Particle Hydrodynamics (SPH) which is a numerical method for simulating fluid flow used widely in astrophysics. In SPH artificial viscosity is necessary for the correct treatment of shocks, but often generates unwanted dissipation away from shocks, particularly in poorly resolved flows. In this study we address this problem by refining the method proposed by Morris & Monaghan (1997). The new scheme uses the rate of change of the velocity divergence,  $D_t(\nabla \cdot \mathbf{v})$ , to indicate a shock and focuses on eliminating viscosity away from shocks. The new method works as least as well as any previous technique in the strong-shock regime, but becomes virtually inviscid away from shocks. In particular sound waves or oscillations of self-gravitating gas spheres are hardly damped over many periods.

We also look at stability issues for SPH, in particular the well known clumping instability. We perform numerical tests of the stability analysis performed by Morris (1996) and find that there are bands of unstable regions as suggested by Read et al. (2010). We also demonstrate that a cored kernel can greatly reduce the clumping instability.

Finally we apply the SPH method to extend the stellar disruption work of Lodato et al. (2009) to orbits with a range of pericentre distances. We find that the light curve produced by this disruption is closer to the predicted  $L \propto t^{-5/3}$  (Rees, 1988) for encounters that are closer to the black hole than the tidal disruption radius. For encounters that are further from the black hole than the tidal disruption radius, the light curve deviates from this predicted power law. We also look at how elliptical orbits can effect the stability of the star. We find that in elliptical orbits a star can be disturbed further from the black hole than in the parabolic case. We then consider the fate of the S2 star and conclude that when it becomes a red giant and expands, S2 will be tidally disrupted.

---

# ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Walter Dehnen for his advice and helpful comments. Thanks also to Dr Graham Wynn and Dr Justin Read for many discussions on SPH. I would also like to thank all the PhD students in the theoretical astrophysics group for their help and support. Thanks to all the members of the physics 5-a-side football team for the much needed breaks from physics! Most importantly I would like to thank Sophie for all her support.

I also acknowledge STFC for the studentship I have recieved during the course of my studies at Leicester University.

---

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 A Conservative SPH Scheme</b>	<b>3</b>
2.1 Introduction . . . . .	3
2.2 Basic SPH Principles . . . . .	7
2.3 The Density Estimate . . . . .	8
2.4 SPH Smoothing Lengths . . . . .	10
2.4.1 Interpretation of SPH Smoothing Lengths . . . . .	10
2.4.2 Adapting SPH Smoothing Lengths . . . . .	10
2.4.3 Estimating the Velocity Divergence . . . . .	13
2.4.4 The Resolution Element of SPH . . . . .	14
2.5 SPH Fluid Equations . . . . .	17
2.5.1 Continuity Equation . . . . .	18
2.5.2 The Momentum Equation . . . . .	19
2.5.3 Internal Energy . . . . .	21
2.5.4 Equation of States . . . . .	22
2.6 Errors in the Density Estimate . . . . .	23
2.6.1 The Bias and Variance for Poisson Noise . . . . .	23
2.6.2 Numerical Results - Fixed $\mathbf{h}$ . . . . .	26
2.6.3 Numerical Results -Varying $\mathbf{h}$ . . . . .	29
2.7 Gradient Estimators . . . . .	31
2.7.1 An Alternative Gradient Operator . . . . .	32
2.8 Gravity . . . . .	34
2.9 Time Integration . . . . .	36
2.10 Code Details . . . . .	39
2.10.1 Hydrodynamics Calculation . . . . .	40



---

2.10.2	Periodic Boundary Conditions . . . . .	41
2.10.3	External Potentials . . . . .	41
2.10.4	SPHfalcON Specific Algorithms . . . . .	41
2.11	Summary . . . . .	42
<b>3</b>	<b>Artificial Dissipation in SPH</b>	<b>44</b>
3.1	Introduction . . . . .	44
3.2	Artificial Viscosity . . . . .	46
3.2.1	“Standard” SPH Viscosity . . . . .	47
3.2.2	Balsara Switch . . . . .	50
3.2.3	Morris & Monaghan Viscosity Switch . . . . .	50
3.3	Towards an Improved Viscosity Trigger . . . . .	54
3.3.1	Preliminary Switch 1 . . . . .	56
3.3.2	Preliminary Switch 2 . . . . .	58
3.3.3	Preliminary Switch 3 . . . . .	63
3.4	The New Viscosity Method . . . . .	64
3.4.1	Defining The Source Term . . . . .	65
3.4.2	Adjusting the Viscosity . . . . .	67
3.4.3	Estimating $D_t(\nabla \cdot v)$ . . . . .	68
3.4.4	Limiting The Viscosity . . . . .	70
3.4.5	Viscosity Algorithms . . . . .	74
3.4.6	Comparing The Viscosity Source Terms . . . . .	75
3.4.7	A Direct Comparison of Viscosity . . . . .	76
3.5	Is $\alpha_{\min} > 0$ required? . . . . .	77
3.6	Viscosity Suppression Tests . . . . .	79
3.6.1	Sound-wave steepening . . . . .	79
3.6.2	1D Converging Flow Test . . . . .	82
3.6.3	2D Keplerian Ring . . . . .	83
3.6.4	Softened Disc . . . . .	91
3.6.5	2D Shear Flows . . . . .	92
3.6.6	A shearing shock . . . . .	95
3.6.7	An Oscillating Polytropic Sphere . . . . .	97
3.6.8	Oscillating Polytrope Orbiting a Point Mass . . . . .	98
3.7	Shock capturing tests . . . . .	100
3.7.1	Sod shock tube test . . . . .	100
3.7.2	1D Ram $\mathcal{M} = 50$ Shock . . . . .	101

---

---

3.7.3	Particle Penetration . . . . .	102
3.7.4	Evrard Test . . . . .	105
3.8	Summary of New Viscosity Method . . . . .	107
3.9	Missing Discontinuities - Artificial Mass Diffusion . . . . .	109
3.10	Artificial Thermal Conductivity . . . . .	111
3.10.1	Monaghan (1997) Conductivity . . . . .	111
3.10.2	Rosswog & Price (2007) Conductivity . . . . .	112
3.10.3	Price (2008) Conductivity . . . . .	112
3.10.4	Problems With Existing Methods . . . . .	113
3.10.5	Where Should Conductivity be Applied? . . . . .	115
3.10.6	Towards an Improved Conductivity Scheme . . . . .	116
3.10.7	Tests With Fixed Strength . . . . .	118
3.10.8	Strength of Conductivity Required . . . . .	121
3.11	Conclusions . . . . .	126
<b>4</b>	<b>SPH Stability</b>	<b>130</b>
4.1	Introduction . . . . .	130
4.2	Initial Conditions in SPH . . . . .	131
4.2.1	Sobol Initial Conditions for the 1D Sod Shock Tube . . . . .	134
4.3	Stability of Initial Condition . . . . .	135
4.3.1	Undamped Initial Conditions . . . . .	135
4.3.2	Damped Initial Conditions . . . . .	138
4.3.3	Setting up Initial Condition . . . . .	139
4.4	Small Amplitude Perturbations and Stability . . . . .	140
4.4.1	Stability Analysis . . . . .	142
4.4.2	Stability Analysis With Viscosity . . . . .	145
4.4.3	Numerical Results . . . . .	148
4.4.4	3D Gaussian Position Perturbations . . . . .	160
4.4.5	Effect of Kernel on Particle Clumping . . . . .	165
4.5	2D Sound-wave Steepening . . . . .	165
4.6	2D Sod Shock Tube . . . . .	166
4.7	Conclusion . . . . .	167
<b>5</b>	<b>Stellar Disruption</b>	<b>169</b>
5.1	Introduction . . . . .	169
5.2	Theory . . . . .	170

---

5.2.1	Tidal Disruption . . . . .	170
5.2.2	The Light Curve . . . . .	172
5.3	Initial Conditions . . . . .	174
5.4	Varying Pericentric Distance . . . . .	176
5.4.1	Mass Loss from the Star . . . . .	176
5.4.2	Specific Energy Distribution . . . . .	179
5.4.3	The Light Curve . . . . .	180
5.5	Varying Polytropic Index of the Star . . . . .	181
5.5.1	The Specific Energy Distribution . . . . .	182
5.5.2	Light Curves . . . . .	185
5.6	Spinning Star . . . . .	188
5.7	Mass Ratio . . . . .	190
5.8	Hyperbolic Orbits . . . . .	194
5.9	Elliptical Orbits . . . . .	199
5.9.1	The Fate of S2 . . . . .	202
<b>6</b>	<b>Conclusions</b>	<b>204</b>
6.1	Summary . . . . .	204
6.2	Future Work . . . . .	206
	<b>Appendices</b>	<b>209</b>
<b>A</b>	<b>Comparison of Viscosity Magnitudes</b>	<b>210</b>
A.1	Homologous Flow . . . . .	210
A.2	Sound wave . . . . .	210
A.3	Strong Shock . . . . .	211
<b>B</b>	<b>Viscosity in the Continuum Limit</b>	<b>214</b>
<b>C</b>	<b>Polytrope Oscillations</b>	<b>217</b>
C.1	Stable Polytropic Setup . . . . .	217
C.2	Applying a Perturbation . . . . .	219
C.2.1	Mass Conservation . . . . .	219
C.2.2	Energy Equation . . . . .	220
C.2.3	Momentum Conservation . . . . .	220
C.2.4	Linear Adiabatic Wave Equation (LAWE) . . . . .	220
C.3	Solutions to LAWE . . . . .	222

C.3.1	Boundary Conditions . . . . .	222
C.3.2	The Solution . . . . .	223
<b>References</b>		<b>228</b>

---

# 1

## Introduction

Many astrophysical problems involve complex systems where the underlying physical laws are well understood, but are too complex to solve analytically. An example of this is the dynamics of the solar system, the positions of the planets with respect to the Sun are determined by the gravitational interaction of each planet with the Sun and each other (as well as minor bodies such as comets). The physical laws of gravity are well understood as are the laws of dynamics. However, finding an analytical solution for the positions of each planet at a future time given an initial starting point is impossible; instead a numerical method is required. The same is true for fluid dynamics, the governing equations are well understood but describing the flow of a fluid is impossible for all but the most simple cases.

Fluid dynamics plays an important role in astrophysics, from the dynamics of galactic mergers, to accretion discs and star formation, to name a few. Therefore in order to try to understand these processes one must turn to computational methods which model the fluid flow in these systems. Computational fluid dynamics (CFD) is not only important in astrophysics, it has a wide variety of applications including engineering, coastal defence and the entertainment industry (e.g. computer games). Furthermore, CFD techniques have recently been applied to the mechanics of solid bodies (Das & Cleary, 2009).

The method of choice for many astrophysics simulations is Smoothed Particle Hydrodynamics (SPH) which is a particle based method invented by Lucy (1977) and Gingold & Monaghan (1977). In chapter 2 we introduce the fundamentals of

the SPH method and show how it can be coupled to N-body schemes. This allows SPH to be easily applied to situations involving self-gravitating fluid dynamics, of which there are many in astrophysics. We demonstrate how SPH can be derived self-consistently from a conservative approach in which mass, energy, momentum, angular momentum and entropy are all conserved by construction. In this conservative approach only a density estimation, system Lagrangian and one extra constraint linking density and smoothing length are required. We also consider sources of error in the density estimation for various particle distributions.

As the SPH scheme described in chapter 2 conserves entropy by construction an artificial method is required to dissipate entropy at shocks. Many implementations of SPH, and therefore many simulations using SPH, currently use a dissipation scheme which is not optimal. The consequence of this being that weak shocks are over-smoothed and strong shocks are not smoothed enough. Due to the complex and non-linear nature of fluid dynamics, the incorrect capturing of shocks can often drastically alter the results of the simulation. Whilst some progress has been made in this area (Morris & Monaghan, 1997), there are still problems with over damping of fluid features, which we aim to resolve by the new dissipation scheme described in chapter 3.

As with any numerical scheme there are areas where the method performs well and there are areas of weakness. Some of these weaknesses are addressed in chapters 2 and 3. In chapter 4 we look at numerical stability of the SPH scheme. In particular we focus on the well known clumping instability (Morris, 1996) in which particles can cluster together. Much theoretical analysis has been done to try to understand clumping (Morris, 1996; Cha, 2002; Read et al., 2010), in chapter 4 we perform numerical tests to verify this stability analysis.

In chapter 5 we apply the SPH method to the problem of the tidal disruption of stars by super-massive black holes. We extend the work of Evans & Kochanek (1989) and Lodato et al. (2009) to include orbits of various pericentre distances. We consider not only parabolic orbits but also elliptical and hyperbolic orbits. This is an ideal problem to study using SPH as the simulation domain consists of mostly empty space with the areas of interest located where the gas is.

Finally chapter 6 concludes this study.

---

# 2

## A Conservative SPH Scheme

### 2.1 Introduction

As there are many uses for CFD there exists many different schemes to model the fluid flow which differ slightly depending on the type of flow one is interested in. However, in order to produce a numerical representation of a fluid all methods must solve the Euler equations which underpin fluid dynamics, these are

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (2.1a)$$

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v}^T + P) = 0, \quad (2.1b)$$

$$\frac{\partial E}{\partial t} + \nabla \cdot (\mathbf{v}(E + p)) = 0. \quad (2.1c)$$

Traditional methods of simulating fluid dynamics involve attempting to solve the Euler equations by representing the fluid with a grid containing many cells. The Euler equations are then solved using a finite volume method, usually based on that of Godunov (1959), although many refinements to the scheme have since been suggested such as higher order schemes and adaptive grid methods (e.g. AMR codes such as `Ramses` Teyssier, 2002). The conservative variables of the fluid<sup>1</sup> are considered to be piecewise constant over the mesh cell at a given time. At each cell

---

<sup>1</sup>Those variables that appear explicitly in the conservative Euler equations, that is  $\rho$ ,  $\rho \mathbf{v}$  and  $E$ .

interface there exists a conservation law (Euler's equation) and piecewise constant data containing a single discontinuity. This type of problem is well known as the Riemann problem. Therefore the solution at the cell interface may be obtained using a Riemann solver (see for example Toro, 1999). The properties of the fluid in each cell are then evolved according to Euler's equations by computing the fluxes entering and leaving the grid cell. As it is possible to solve the Riemann problem numerically for 1D initial conditions with two piecewise constant states, Riemann solvers provide invaluable solutions to shock tube problems, which can be compared to numerical schemes. For general initial value problems, the solution can be thought of as a superposition of the solutions of local Riemann problems.

An alternative to grid-based methods are particle-based methods, in which the fluid is replaced by a set of particles which are free to move. This freedom allows one to model the Lagrangian representation of the fluid where one follows individual fluid parcels as they move through space. The Euler equations, written in Lagrangian form are

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (2.2a)$$

$$\frac{d\mathbf{v}}{dt} + \frac{1}{\rho} \nabla P = 0, \quad (2.2b)$$

$$\frac{du}{dt} + \frac{P}{\rho} \nabla \cdot \mathbf{v} = 0. \quad (2.2c)$$

The most common method of doing this is smoothed particle hydrodynamics (SPH) (Lucy, 1977; Gingold & Monaghan, 1977), which is the subject of this chapter. The particles are a representation of the distribution of the underlying fluid (similar to a Monte Carlo representation, although randomness should be avoided). Early SPH implementations found particle estimates for these fluid equations, which lead to a plethora of methods depending on how gradient operations were performed. However, more recent implementations simply define an estimate for the density at a particle's position and a particle Lagrangian which is representative of the underlying system. This naturally results in the solution to the Euler equations.

Both grid based and particle based methods have their advantages and disadvantages, therefore the two methods may perform better in different situations. The main advantage of SPH over grid-based methods is that particles are free to move throughout the whole simulation without the need for boundaries. However, in a grid-based methods one has to know the simulation domain in advance. Further-



more grid based methods require surrounding cells so that gas can be advected into the empty space. Not only can this surrounding medium interact with the region of interest, extra computational effort is required as one must consider interactions between the empty grid cells. This advantage of SPH is particularly important in astrophysics, where there are usually no physical boundaries. Also one often models systems (i.e. stars or galaxies) which are well separated from each other. A grid code would therefore require one to model lots of empty space whereas in SPH the particles would be located in the places of interest. A grid code may also run into problems if the region of interest expands to engulf the surrounding empty grid cells because there are no truly open boundaries for the gas to move into. The other main advantage of SPH is that there are automatically more particles in higher density regions, thus increasing the resolution of the scheme. This is again of concern for astrophysics where one often simulates large dynamic ranges. In standard grid codes, the grid has a fixed resolution at all points which cannot change, although adaptive grid methods are going some way to allow adaptive resolution (e.g. adaptive mesh refinement, AMR). SPH is also Galilean invariant, whereas AMR methods are not, which can cause problems with the choice of coordinate system (e.g. Tasker et al., 2008) and can lead to numerical diffusion. Also AMR methods are not isotropic or translational invariant as shown by Tasker et al. (2008) who find that the translation of a polytrope around a periodic box causes a significant change in the polytrope.

The main disadvantage of SPH over grid based methods is that SPH is often more expensive computationally. The Riemann problem only has to be solved between neighbouring grid cells, whereas SPH interactions take place between neighbouring particles, of which there are many. Furthermore, as the grid cells are fixed the neighbouring grid cells only need to be found once. However, as SPH particles are free to move, a given particle's neighbours are constantly changing, and therefore an efficient algorithm is required to find neighbouring particles. Another disadvantage of SPH is that in order to get good resolution in low density regions, the total number of particles needs to be large as most of the particles will be in the high density regions. Grid codes (with fixed grid size) have the same density of grid cells in low density regions allowing good resolution. This can be an issue if the region of primary interest is not the high density region, or low density regions are important in the development of some physical process. An example of where this is a problem is the simulation of the Kelvin-Helmholtz instability, good resolution is required in the low density region to fully resolve the instability. Even in 2D with only a 1:2

density ratio, this can dramatically increase the total number of particles required. Another issue with SPH is the generation of initial conditions. As particles are free to roam, any configuration of particles that satisfies the target density profile can be used. However, randomly placing particles to achieve this often leads to extremely noisy results. Finally there has been concern over the ability of SPH to handle mixing (Agertz et al., 2007), leading to problems in SPH capturing Kelvin-Helmholtz and Rayleigh-Taylor instabilities. However, recent developments by Price (2008) and Read et al. (2010) have made these types of problems solvable in SPH.

An issue of concern is that SPH and grid based schemes sometimes disagree significantly for even the most basic calculations. Perhaps the most famous example of this is the Santa Barbara cluster comparisons of Frenk et al. (1999) where the entropy of the core of the cluster is different for the two types of simulation. Both the SPH methods and grid methods have converged and therefore both schemes use a sufficiently high resolution, they have however converged on different answers. Unfortunately due to the complexity of the simulation it is not clear from simple analytical arguments which method is correct (if any).

More recently hybrid methods have been suggested which aim to remove the problems associated with both SPH and grid based methods. One approach (Inutsuka, 2002) combines SPH with a Riemann solver by replacing the pressure in the SPH momentum equation (see equation 2.39) with the pressure found by solving the Riemann problem between the SPH particle and its neighbour. More recently Springel (2009) attempted to find a hybrid between the mesh-less regime of SPH and the grid methods based on Riemann solvers. The author invented a scheme which contains both moving interpolation points and grid cells. This is done by allowing particles to move freely like in SPH. However, fluid properties are computed in a grid like manner by firstly constructing Voronoi cells and computing the flux into and out of each Voronoi cell using a Riemann solver. The main drawback of this method is the building of the Voronoi tessellation, which is very computationally expensive. Furthermore, whilst the method is promising, as a new scheme its usefulness in astrophysical simulations has not been fully explored. Also any potential problems with the scheme are unknown whereas with SPH and grid based techniques which have been around longer, the regimes where the schemes are likely to fail are at least known, if not fully understood.

There are many implementations of SPH in literature, all following some basic SPH principles, of which there are many useful reviews (Monaghan, 1992, 2005;

Price, 2004). Here we give an overview of the basic implementation of a conservative SPH scheme, that is a method that conserves energy, momentum, angular momentum and entropy by construction. Whilst there are many possible implementations of SPH where the equations vary slightly, using a conservative approach gives a unique set of SPH equations.

As we are interested in astrophysical applications of SPH we may ignore some of problems associated with SPH such as free surface flows and incompressible flows as these rarely occur in astrophysics. This removes some of the difficulties associated with SPH found in engineering applications, but as we shall see the SPH scheme still contains some difficulties that need to be overcome.

## 2.2 Basic SPH Principles

A fluid can be approximated by a discrete set of particles where the particles are interpolation points. At each interpolation point the properties of the fluid are calculated. The integral interpolant of a function  $A(\mathbf{x})$  is given by the convolution

$$\langle A(\mathbf{x}) \rangle = \int A(\mathbf{x}') W(\mathbf{x} - \mathbf{x}', h) d\mathbf{x}' \quad (2.3)$$

where  $W$  is an interpolating kernel with the following two properties

$$\begin{aligned} \int W(\mathbf{x} - \mathbf{x}', h) d^3\mathbf{x}' &= 1 \\ \lim_{h \rightarrow 0} W(\mathbf{x} - \mathbf{x}', h) &= \delta(\mathbf{x} - \mathbf{x}') \end{aligned} \quad (2.4)$$

and  $h$  is the smoothing length. The smoothing length defines the radius at which we truncate the interpolation. Interpolation points  $\mathbf{x}'$  which are further away from  $\mathbf{x}$  than  $h$  are considered negligible.

No assumptions about the kernel have been made although most (if not all) SPH applications use an isotropic kernel. The reason for this is that as space is isotropic, the kernel should be symmetric about the particle in question so that all neighbouring particles of equi-distance are weighted equally, giving no preference for direction. An isotropic kernel also means that the derivative of the kernel is an odd function which is a desirable property as the sign of the gradient will be direction

dependent. Therefore the kernel usually has the extra property that

$$W(\mathbf{x} - \mathbf{x}') = W(\mathbf{x}' - \mathbf{x}). \quad (2.5)$$

We often use the shorthand notation for the kernel

$$W(|\mathbf{x}_i - \mathbf{x}_j|, h) \equiv W_{ij}(h). \quad (2.6)$$

We may define the dimensionless kernel  $w$  via

$$W(|\mathbf{x}_i - \mathbf{x}_j|, h) \equiv \frac{1}{h^\nu} w\left(\frac{|\mathbf{x}_{ij}|}{h}\right) \quad (2.7)$$

where  $\nu$  is the number of spatial dimensions. For gradients of the kernel we define the dimensionless kernel gradient  $\tilde{w}$  via

$$\nabla_i W(|\mathbf{x}_i - \mathbf{x}_j|, h) \equiv \frac{\mathbf{x}_i - \mathbf{x}_j}{h^{\nu+2}} \tilde{w}\left(\frac{|\mathbf{x}_i - \mathbf{x}_j|}{h}\right). \quad (2.8)$$

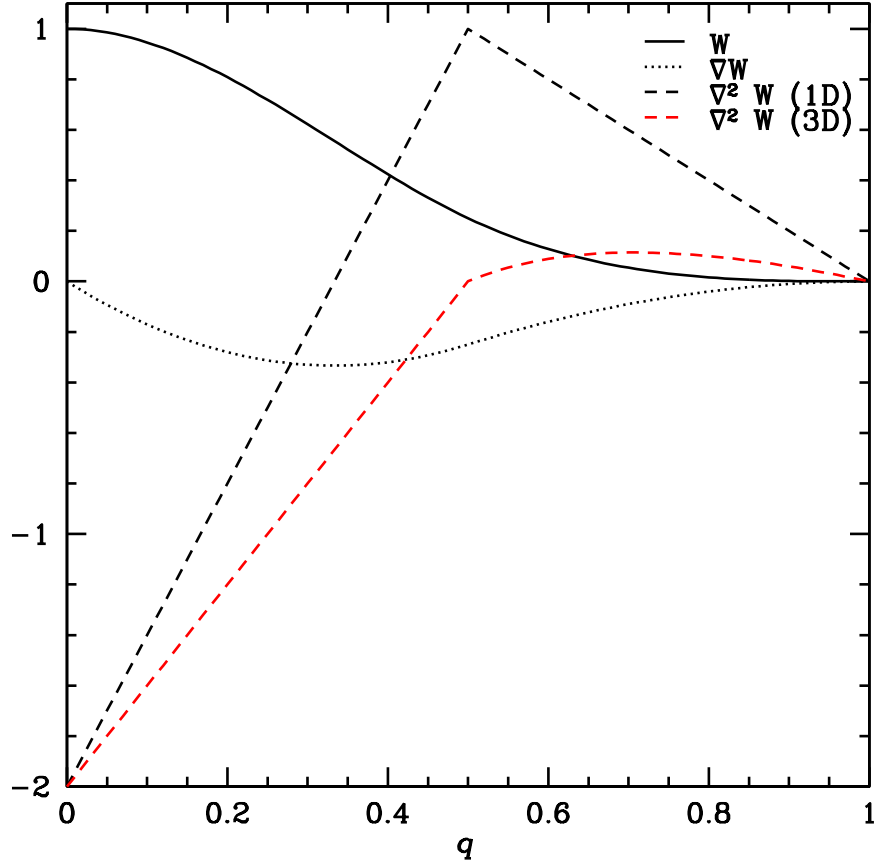
The kernel function should be weighted such that particles close to the core contribute towards the estimate more than particles further from the centre. Therefore a kernel with similar functional form to a Gaussian is suitable. We require the kernel and derivatives to be zero at the edge of the kernel and the first two derivatives to be smooth functions with no discontinuities. Therefore we use the standard cubic spline kernel used in most implementations of SPH.

$$w(q) = W_0 \begin{cases} (1 - 6q^2 + 6q^3) & 0 < q < 0.5 \\ 2(1 - q)^3 & 0.5 < q < 1 \end{cases} \quad (2.9)$$

Here  $q = |\mathbf{x}_i - \mathbf{x}_j|/h$  and  $W_0$  is a constant arising from the normalisation condition (equation 2.4).  $W_0 = 4/3, 40/7\pi, 8/\pi$  in one, two and three dimensions respectively. Note that the cubic spline is sometimes defined with  $q = 2|\mathbf{x}_i - \mathbf{x}_j|/h$  and the limits are twice the limits we use here.

## 2.3 The Density Estimate

For numerical simulations we need to discretise the fluid. Dividing the fluid into a set of  $N$  discrete points, the interpolant can be approximated as a sum over all the



**Figure 2.1:** The cubic spline kernel (equation 2.9), and its 1st and 2nd derivatives (Without normalisation). Note that the second derivative of the kernel depends on the number of dimensions used due to the appearance of  $\nabla \cdot \mathbf{x}$ .

points. This is done by writing an integral containing a density over a volume, as a summation over the mass elements. Therefore we obtain an *estimate* of the density as

$$\hat{\rho}_i = \sum_j m_j W_{ij}(h). \quad (2.10)$$

Here the quantity  $\hat{\rho}_i$  is the SPH estimate of the fluid density at position  $\mathbf{x}_i$ . It should be emphasised that this is an *estimate* of the underlying fluid field, it does not give the fluid field itself. This point is often overlooked in SPH literature.

## 2.4 SPH Smoothing Lengths

### 2.4.1 Interpretation of SPH Smoothing Lengths

There are two interpretations of the SPH smoothing length. The first is the scatter interpretation which assumes that each particle has a mass which is smeared out in space according to the kernel and the smoothing length. Properties of the fluid are found by summing contributions from neighbouring particles. The second is the gather interpretation. Particles are point markers in the fluid, the properties at any point are found by weighting the contribution of each neighbour according to the kernel (Hernquist & Katz, 1989).

When the smoothing length is constant for all particles, the scatter and gather methods are equivalent. In the case of variable smoothing lengths, the two interpretations are not the same. Suppose we are calculating properties of particle  $i$  by summing over a set of  $j$  particles. In the scatter method, the smoothing length used by the kernel is the  $j$ 'th particle's as we are summing contributions from smoothed out particles at a point in space which happens to be at the same point as particle  $i$ . In the gather approach, the smoothing length of particle  $i$  is used because the contributions from neighbouring particles are weighted according to where they are in the kernel of particle  $i$ . Most implementations of SPH use the gather approach which is more consistent with the particle based approach of SPH. We also adopt the gather approach, that is  $W_{ij}(h) = W_{ij}(h_i)$

### 2.4.2 Adapting SPH Smoothing Lengths

The full power of SPH comes from the ability to adapt to the local conditions of the simulation, which is derived from the Lagrangian nature of the scheme. Regions of high density automatically have more particles and lower density regions have fewer particles. However, if we were to use a fixed smoothing length, the resolution of the simulation would be only slightly increased in the high density region. This is because the fluid properties would still be smoothed over the same volume, the smoothing would simply be averaged over more particles, which may reduce the noise.

When the local conditions of the particles are changed, the smoothing lengths must be adjusted in some way. The smoothing lengths must be given two types of adjustment. The first is an adjustment at a fixed time to allow the smoothing

length to be constrained accurately. The second adjustment is a prediction of the smoothing length at the next time-step based on the information at the current time-step. This allows for fast convergence of the first adjustment.

Early methods to adjust the smoothing lengths (Hernquist & Katz, 1989) used a prescription based on the ratio of the number of neighbours a particle has, to the desired number of neighbours. This essentially ensures that the particles have a fixed number of neighbours to within some tolerance (usually to within two particles). However, this method means that the smoothing lengths of particles are discontinuous and will suddenly jump as particles move into or out of each others smoothing sphere. A situation could also arise in low density regions where many neighbouring particles are at the same distance and therefore the actual number of neighbours is much higher than the desired number. The disadvantage of this method is demonstrated by Attwood et al. (2007) where the affect of the tolerance parameter on the oscillations of a polytropic sphere are investigated. They find that increasing the tolerance parameter increases the numerical dissipation in the SPH scheme, damping the amplitude of the polytrope oscillations more quickly.

More recently several authors (Springel & Hernquist, 2002; Price & Monaghan, 2004) suggest constraining the smoothing lengths by ensuring that the total SPH mass inside a particle smoothing sphere is constant for all particles (to within some tolerance parameter). This is done by coupling the smoothing length to the density as

$$\mu_i = \hat{\rho}_i h_i^\nu \quad (2.11)$$

and requiring that the parameter  $\mu_i$  be equal to some global constant  $M$  (where  $\nu$  is the number of dimensions used). In order to achieve this, the parameter  $\mu_i$  is calculated for every particle using

$$\mu_i = \sum_j m_j w_{ij} \quad (2.12)$$

which is identical to equation (2.11) by virtue of the density estimator (2.10). We may use a Newton-Raphson scheme to adjust the smoothing lengths such that

$$M = \mu + (\tilde{h} - h) \left. \frac{\partial \mu}{\partial h} \right|_t + \mathcal{O}((\tilde{h} - h)^2) \quad (2.13)$$

thus giving the adjustment as

$$\tilde{h}_i = h_i + \frac{M - \mu_i}{\partial \mu_i / \partial h|_t}. \quad (2.14)$$

However, we found that a more accurate method for the iteration scheme can be obtained if we use a Newton-Raphson iteration in  $\log \mu - \log h$  space:

$$\log(M) = \log(\mu) + \left( \log(\tilde{h}) - \log(h) \right) \frac{\partial \log(\mu)}{\partial \log(h)} \Big|_t + \mathcal{O}((\tilde{h} - h)^2). \quad (2.15)$$

thus giving the adjustment as

$$\tilde{h}_i = h_i \left( \frac{M}{\mu_i} \right)^{f_i/\nu} \quad (2.16)$$

where the quantity  $f_i$  is found by taking the partial derivative of  $\mu$  with respect to  $h$  at a fixed time;

$$f_i = \nu \frac{\mu_i}{h_i} \frac{\partial \mu_i}{\partial h_i} \Big|_t^{-1}. \quad (2.17)$$

The rate of change of  $\mu$  with respect to the smoothing length can be found by differentiating equation (2.12)

$$\frac{\partial \mu_i}{\partial h_i} \Big|_t = -\frac{1}{h_i^3} \sum_j m_j x_{ij}^2 \tilde{w}(q). \quad (2.18)$$

We found this method to converge with typically two digits improvement per iteration. However, we found that if  $\mu$  is significantly too small the scheme converged much slower, or in some cases not at all. Therefore we refine the model to explicitly exclude the  $h_i$  independent self-contributions, i.e.

$$\tilde{h}_i = h_i \left( \frac{M - m_i w(0)}{\mu_i - m_i w(0)} \right)^{\tilde{f}_i/\nu} \quad (2.19)$$

with

$$\tilde{f}_i = \nu \frac{\mu_i - m_i w(0)}{h_i} \frac{\partial \mu_i}{\partial h_i} \Big|_t^{-1}. \quad (2.20)$$

This model is used when  $\mu_i < M$ . In order to get a faster convergence we also use



an initial prediction of the smoothing lengths using the total time derivative

$$\frac{d\mu_i}{dt} = \left. \frac{\partial \mu_i}{\partial t} \right|_{h_i} + \left. \frac{\partial \mu_i}{\partial h_i} \right|_t \dot{h}_i. \quad (2.21)$$

The requirement that  $d\mu_i/dt = 0$  leads to an expression for the rate of change of the smoothing length

$$\dot{h}_i = - \left. \frac{\partial \mu_i}{\partial t} \right|_{h_i} \bigg/ \left. \frac{\partial \mu_i}{\partial h_i} \right|_t. \quad (2.22)$$

We obtain the partial derivative of  $\mu$  with respect to time as

$$\left. \frac{\partial \mu_i}{\partial t} \right|_{h_i} = \frac{1}{h_i^2} \sum_j m_j \mathbf{x}_{ij} \cdot \mathbf{v}_{ij} \tilde{w}(q). \quad (2.23)$$

Thus the smoothing length may be predicted as

$$h_i(t + \delta t) = h_i(t) e^{\delta t \frac{\dot{h}_i(t)}{h_i(t)}} \quad (2.24)$$

which avoids negative values for  $h_i$ .

### 2.4.3 Estimating the Velocity Divergence

In a conservative SPH scheme, the velocity divergence is the only SPH spatial gradient that needs to be computed<sup>2</sup>. This can readily be computed by applying the constraint  $d(h^\nu \hat{\rho})/dt = 0$ ,

$$\nu \hat{\rho} h^{\nu-1} \dot{h} + h^\nu \frac{d\hat{\rho}}{dt} = 0. \quad (2.25)$$

The Euler equation for the rate of change of density (see equation 2.2a) may be used to obtain

$$\widehat{\nabla \cdot \mathbf{v}} = \frac{\nu \dot{h}}{h}. \quad (2.26)$$

---

<sup>2</sup>However, one may need to compute other gradients for extensions to the conservative SPH scheme.

From the definition of  $\dot{h}$  given by equation (2.22) we may write

$$\widehat{\nabla \cdot \mathbf{v}} = -\frac{\nu}{h} \left. \frac{\partial \mu}{\partial h} \right|_t^{-1} \left. \frac{\partial \mu}{\partial t} \right|_{h_i}. \quad (2.27)$$

Using equations (2.21), (2.17) and (2.11) we may write

$$\widehat{\nabla \cdot \mathbf{v}}_i = -f_i \frac{1}{\hat{\rho}_i h_i^{\nu+2}} \sum_j m_j \mathbf{x}_{ij} \cdot \mathbf{v}_{ij} \tilde{w}_{ij}(h_i) \quad (2.28)$$

$$= -\frac{f_i}{\hat{\rho}_i} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla W_{ij}(h_i). \quad (2.29)$$

It should also be noted at this point that the velocity in equation (2.28) is the SPH particle velocity. This is because the velocity arises from the temporal derivative of the particle positions.

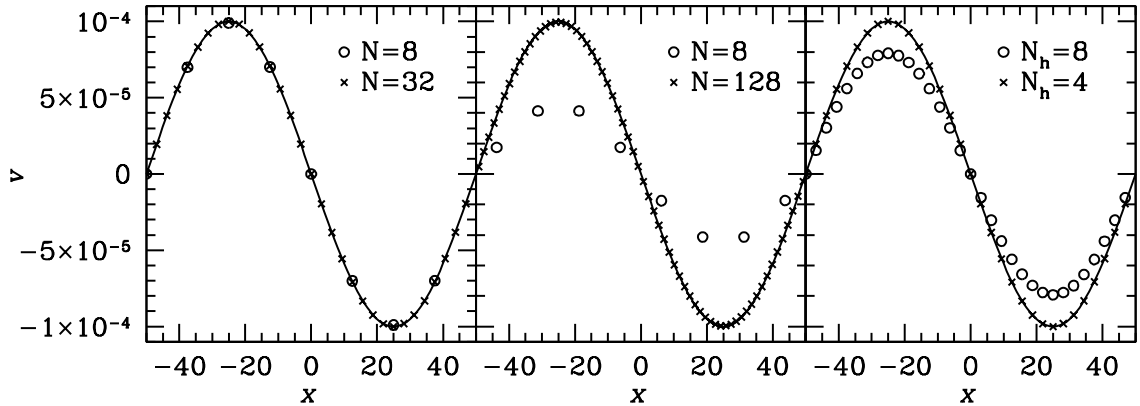
#### 2.4.4 The Resolution Element of SPH

The idea of a resolution element in SPH is not as straight forward as that of a grid code where each cell can be thought of as a discrete carrier of information. Agertz et al. (2007) argues that in SPH, particles are not an independent resolution element as the particle is effectively a weighted average over the neighbouring particles. Therefore the resolution element is set not only by the total number of particles but also by the number of neighbouring particles. This is also true to some degree for grid codes, the grid elements are not independent of each other as the fluxes entering and leaving the grid cell depend on the state of the neighbouring cells ( $2^\nu$  for standard grid codes). However, in modern grid codes many interpolation tricks are employed which increase the resolution above that of the standard 1st order Godunov scheme.

Clearly in SPH the number of neighbours used will impact on the resolution of a simulation. The more neighbours used, the larger the size of the region smoothed over, decreasing the sharpness of the results (especially at shocks) and the accuracy of the results. There are however other factors that can increase or decrease the resolution of an SPH simulation, perhaps the most important being artificial viscosity<sup>3</sup>. Applying a large amount of viscosity can provide extra smoothing of a shock

---

<sup>3</sup>Artificial viscosity is a dissipation mechanism required to resolve shocks. This will be discussed in full in section 3



**Figure 2.2:** A 1D sound wave with amplitude  $v_a = 10^{-4}$  after 5 dynamical times. **Left:** SPH varying the number of particles. **Middle:** Ramses varying the number of grid cells. **Right:** SPH with 32 particles, varying the number of neighbours.

front for example.

Having too low a resolution in a simulation can not only affect the sharpness of the results, it can also affect the accuracy of the results. Fig. 2.2 shows the results of the propagation of a sound wave with a velocity amplitude of  $v_a = 10^{-4}$  after 5 sound crossing times. The SPH runs are with no artificial viscosity in order to allow the wave to remain undamped. We observe that SPH is able to resolve the wave with just 8 particles per wavelength (the number of neighbours used here was 4). This result is particularly encouraging for astrophysics as we often have regions which are at the resolution limit, therefore it is encouraging that sound waves in these regions will not be dissipated away. However, with the grid code Ramses (Teyssier, 2002) we find that when only 8 grid cells are used the wave is severely damped. By decreasing the resolution with Ramses we have increased the numerical dissipation. This is because the grid code has a dissipation mechanism built into it via the Riemann solver, which can not be turned off. However, increasing the resolution quickly reduces the dissipation. With 32 particles we find that the amplitude of the sound wave is only damped by 2%.

We see that in SPH the resolution is affected much more by changing the number of neighbouring particles rather than changing the total number of particles. SPH is able to resolve the wave well with only 8 particles when 4 neighbours are used (the smoothing length is a quarter of the wavelength). However, with 32 particles and 8 neighbours (the smoothing length is an eighth of the wavelength) the wave appears damped as shown in the right panel of Fig. 2.2. In fact the wave has not

been damped, the propagation of the wave is not at the sound speed and therefore we are seeing the wave at a point after it has reached its maximum amplitude (note that this is not the case with *Ramses*, the wave really is damped when using too few grid cells). With 8 neighbours we find that the propagation speed of the wave is at 104% of the sound speed.

The result is clearly not due to the smoothing length being too large in comparison to the wave length as the results with 8 SPH particles have shown (the smoothing length in this case is larger). Therefore it is not always possible to increase SPH resolution or accuracy by simply using more particles, the number of neighbours used is also an important factor<sup>4</sup>.

### Effect of Artificial Viscosity

Artificial viscosity (which will be discussed in full in Chapter 3) can also play an important role in the resolution of an SPH scheme. As shown by Meglicki et al. (1993) the viscous force supplied by artificial viscosity is in the form of both a bulk and shear viscosity with coefficients  $\eta = \frac{1}{2}\alpha\kappa h c\rho$  and  $\xi = \frac{5}{3}\eta$ , where the factor  $\kappa$  is of order unity but depends on the functional form of the kernel. The parameter  $\alpha$  controls the strength of viscosity. This implies that artificial viscosity is resolution dependent: the higher the resolution the smaller artificial viscosity is. Thus a straightforward way to reduce unwanted dissipation is to increase the resolution by using more particles with a fixed number of neighbours. This is a desirable property as it ensures that when we have a shock the length scale over which the shock is smoothed is of the order of the smoothing length, which is of course resolution dependent.

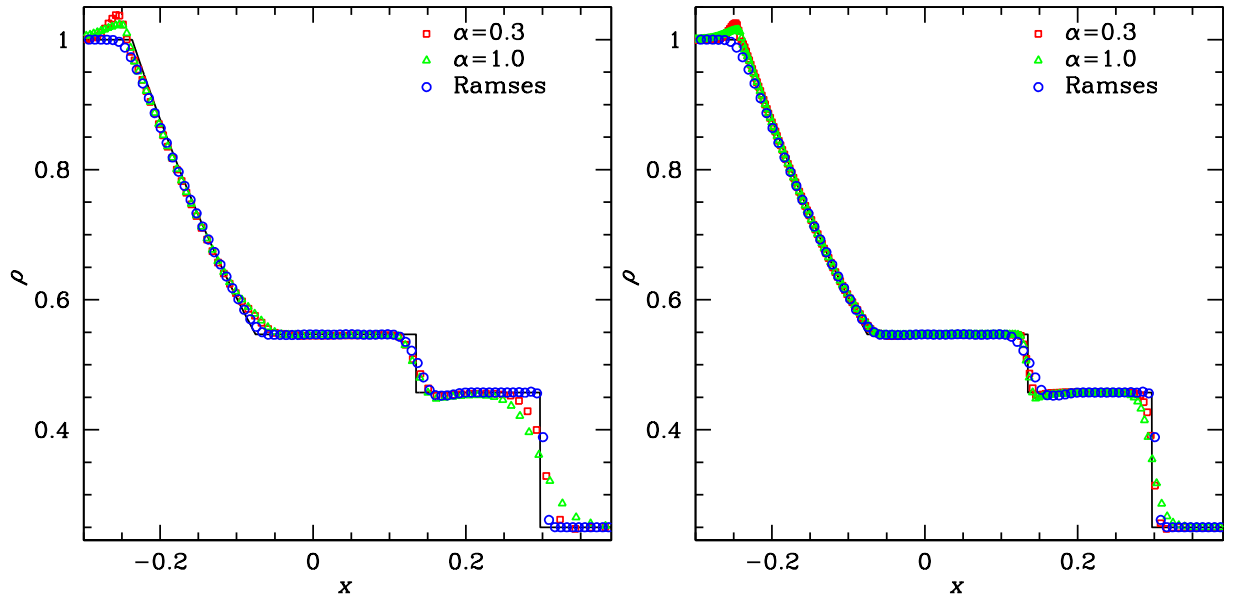
Meglicki et al. (1993) also demonstrated that the parameter  $\alpha$  can determine the size of the region the shock is smoothed over. We need enough viscosity to resolve the shock but not so much that the shock is overly smoothed. In order to demonstrate this we look at the effects of varying the strength of viscosity on the Sod (1978) shock tube test. This is where two regions initially at rest, separated at  $x = 0$ , have different densities and pressures. The region  $x < 0$  has  $\rho = 1$ ,  $P = 1$  and the region  $x > 0$  has  $\rho = 0.25$  and  $P = 0.1795$ .

We see from Fig. 2.3 that the contact wave<sup>5</sup> is unaffected by changing the vis-

---

<sup>4</sup>In 1D the optimal number of neighbours is 4 as we are using cubic spline interpolation which requires 4 constraints per interpolation point.

<sup>5</sup>This is a density discontinuity with equal pressures and velocities.



**Figure 2.3:** Comparing the results of the Sod shock tube test with varying strength viscosity to Ramses. For Ramses a total of 256 grid cells were used. The SPH runs in the *left* plot are with a total of 255 particles, in the *right* plot the SPH runs matched the number of particles in the low density region to that of Ramses, giving a total of 640 particles.

cosity, however the shock wave is smoothed out more when using a higher viscosity. We note that the smoothing in the contact wave is approximately the same for both SPH and the grid code when using the same number of particles as grid cells. However, when the number of SPH particles in the low density region matches the number of grid cells of Ramses (right hand plot), we find that the contact wave remains much sharper in SPH. We also observe that SPH still smooths out the shock more than the grid code. This is unavoidable as both codes smooth shocks over the neighbouring particles, in SPH this is 4 particles. In a grid code only the nearest grid cells are neighbours and so the smoothing is considerably less.

## 2.5 SPH Fluid Equations

The Euler equations written in Lagrangian coordinates (equation 2.2), fully describe (along with an equation of state linking pressure, internal energy and density) the properties of a parcel of fluid at all points in time and space. It is possible to derive SPH forms of these equations by applying a SPH spatial gradient opera-

tor to the fluid quantities. However there are many possible forms for the SPH gradient operator, leading to a set of SPH equations that are not unique. Furthermore, depending on the gradient operator used, the conservation and error properties may vary vastly. Instead we adopt the conservative approach proposed by Springel & Hernquist (2002).

### 2.5.1 Continuity Equation

The most common method to estimate the density is to use equation (2.10).

$$\hat{\rho}_i = \sum_j m_j W_{ij}(h_i).$$

This method relies on calculating the density based on the coordinates of surrounding particles, as well as the smoothing length of the particle we are concerned with. This is the most fundamental SPH equation as the density only depends on the interpolation kernel, the positions and masses of the particles. No assumptions have been made about the differentiability of the density field, if the density is discontinuous then the kernel will smooth the discontinuity away giving a steady transition of density over the SPH smoothing length.

Alternatively one may integrate the density using the continuity equation given by equation (2.2a). The continuity equation for varying smoothing lengths may be obtained by using the estimate for the velocity divergence (see equation 2.28):

$$\begin{aligned} \widehat{\frac{d\rho_i}{dt}} &= -\hat{\rho}_i \widehat{\nabla \cdot \mathbf{v}} \\ &= f_i \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W \left( \frac{|\mathbf{x}_i - \mathbf{x}_j|}{h_i} \right). \end{aligned} \quad (2.30)$$

Most implementations of SPH use the density summation equation, although there are some advantages of using the continuity equation. The main difference when using the continuity equation is that the density of the particles does not change unless the particles have a velocity difference with respect to each other. This allows for true discontinuities in the density estimate, which may be useful in some applications, such as modelling of water waves (Monaghan, 1992). The main disadvantage of using the continuity equation is that mass is not exactly conserved. For example,

the rate of change of the density on particle  $i$  due to particle  $j$  is

$$\widehat{\frac{d\rho_{i \rightarrow j}}{dt}} = m_j (\mathbf{v}_i - \mathbf{v}_j) \frac{|x_{ij}|}{h_i^5} \tilde{w}(q) \quad (2.31)$$

whereas the change on particle  $j$  due to particle  $i$  is

$$\widehat{\frac{d\rho_{j \rightarrow i}}{dt}} = m_i (\mathbf{v}_j - \mathbf{v}_i) \frac{|x_{ji}|}{h_j^5} \tilde{w}(q). \quad (2.32)$$

Assuming equal mass particles and a fixed smoothing length we see that the two expressions are the same (even without these assumptions they have the same sign), violating mass conservation.

Furthermore there is a more fundamental difference between the two forms of the continuity equation, as pointed out by Price (2008). The density summation equation represents an integral formulation as we have not assumed that the density is differentiable, whereas taking the time derivative assumes that the density is differentiable. Price (2008) shows that the summation form of the continuity equation contains an extra surface term compared to the differential form of the continuity equation. This surface integral term, given by  $\int [\rho' \mathbf{v}' W] \cdot d\mathbf{S}$ , vanishes in most cases but not at flow discontinuities. Therefore at flow discontinuities information is lost when using the continuity equation. As a result most codes use the density summation (2.10) equation. However, the continuity equation may be used by applying a smoothing to the density field, this is done in the form of an artificial dissipation (this is discussed in more detail in chapter 3).

### 2.5.2 The Momentum Equation

Instead of manipulating the gradient operator, which leaves many possible forms of the momentum equation (see section 2.7), one can instead use the Lagrangian of the system to derive an expression for the momentum equation. This ensures that the momentum, angular momentum, energy and entropy of the system are conserved. Here we give a brief overview of this method as first proposed by Springel & Hernquist (2002). The equation of motion for the Lagrangian for a set of  $N$  constraints is given by

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{q}}_i} - \frac{\partial L}{\partial \mathbf{q}_i} = \sum_k^N \lambda_k \frac{\partial \phi_k}{\partial \mathbf{q}_i} \quad (2.33)$$

with the SPH Lagrangian given by

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \sum_k^N m_k \dot{\mathbf{x}}_k^2 - \frac{1}{\gamma - 1} \sum_k^N m_k K_k \hat{\rho}_k^{\gamma-1}. \quad (2.34)$$

The set of functions  $\phi = \phi(\mathbf{x}, h)$  are obtained by applying the  $N$  constraints for the adjustments of the smoothing length, given by

$$\phi_k = \rho_k h_k^\nu - \mu = 0. \quad (2.35)$$

One may calculate the Lagrangian multipliers  $\lambda$ , by considering the equation of motion for the set independent variables  $h$ . This gives the Lagrangian multipliers as

$$\lambda_i = \frac{m_i K_i \hat{\rho}_i^{\gamma-2}}{h_i^\nu} (1 - f_i) \quad (2.36)$$

where  $f_i$  is given by equation (2.17). Now that the Lagrangian multipliers are known, we may solve the equation of motion for the set of independent variables  $\mathbf{x}_i$  giving

$$m_i \frac{d\widehat{\mathbf{v}}_i}{dt} = - \sum_k^N m_k \frac{f_k \hat{P}_k}{\hat{\rho}_k^2} \left. \frac{\partial \hat{\rho}_k}{\partial \mathbf{x}_i} \right|_{h_k}. \quad (2.37)$$

At this point no SPH information has been used except for the constraint describing how the smoothing lengths are updated and the SPH form of the Lagrangian. We have not defined how to calculate any other SPH quantities such as gradients or even the SPH density. Therefore we can say that equation (2.37) is the general equation of motion for the SPH Lagrangian (2.34) when the smoothing lengths are constrained using (2.11). At this point we need to provide an expression for the SPH estimate of the derivative of the density. Using the SPH estimate for the density given by equation (2.10) we may obtain

$$\frac{\partial \hat{\rho}_k}{\partial \mathbf{x}_i} = \sum_j m_j \frac{\partial}{\partial \mathbf{x}_i} W \left( \frac{|x_k - x_j|}{h_k} \right) = \sum_j m_j (\delta_{ik} - \delta_{ij}) \nabla W_{kj}(h_k). \quad (2.38)$$



Using the identity  $\nabla_i W_{ij} = -\nabla_j W_{ij}$  and noting that  $k$  is just a dummy index we may obtain the SPH estimate for the acceleration as

$$\widehat{\frac{d\mathbf{v}_i}{dt}} = - \sum_j m_j \left[ \frac{f_i \hat{P}_i}{\hat{\rho}_i^2} \nabla_i W_{ij}(h_i) + \frac{f_j \hat{P}_j}{\hat{\rho}_j^2} \nabla_i W_{ij}(h_j) \right]. \quad (2.39)$$

This formulation has the advantage that linear and angular momentum, energy and entropy are all conserved. This pair-wise conservation is useful for accretion disc problems, where the transport of angular momentum can lead to the transfer of matter through the disc. Therefore, this pair-wise conservation of angular momentum prevents spurious transport in the disc and one is able to model possible transport mechanisms, rather than numerical artifacts. However, it has the disadvantage that in the case of constant pressure the acceleration is non zero if the densities are different. This situation typically would occur at contact discontinuities such as Kelvin-Helmholtz simulations.

### 2.5.3 Internal Energy

The internal energy equation for the case of variable smoothing lengths is found from equation (2.2c) and the estimate for the velocity divergence (2.28)

$$\widehat{\frac{du_i}{dt}} = \frac{f_i \hat{P}_i}{\hat{\rho}_i^2} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}(h_i). \quad (2.40)$$

For an ideal gas, an alternative of integrating the internal energy equation is to integrate the entropy equation. This is the preferred method of some SPH practitioners and is the method currently used in **Gadget-2** (Springel, 2005). When integrating the entropy equation the internal energy and pressure can be computed from the density and entropy (see equation 2.42). As entropy is a conserved quantity it only changes at shocks, where it is generated through artificial viscosity (see chapter 3). Hernquist (1993) demonstrated for a non-conservative SPH scheme, internal energy was well conserved but there were large errors in the entropy when using the energy formulation. However, when the entropy formulation was used, the entropy was well conserved but there were large errors in the internal energy. The reason for this is that the correction terms due to the variable smoothing lengths were not accounted for. When using the correction terms both energy and entropy were well conserved (see Springel & Hernquist, 2002). Tests have shown that there is no difference in

the results obtained when using the entropy or internal energy formalism.

### 2.5.4 Equation of States

In SPH internal energy and density are the only fluid variables that are evolved in their own right, therefore an equation of state to calculate the pressure from the internal energy and density is needed (pressure is needed for the momentum equation). The pressure is estimated using the estimated density and internal energy in an equation of state, that is  $\hat{P}_i = P(\hat{u}_i, \hat{\rho}_i)$ . The three equation of states we will use during this work are described below.

An isothermal equation of state, in which the pressure is calculated using the equation

$$P = c_s^2 \rho. \quad (2.41)$$

As the gas is isothermal the internal energy is not integrated, in fact we do not even store the internal energy. Typically the sound speed is constant for all particles. An isentropic equation of state (sometimes called polytropic equation of state), in which the pressure is evaluated using

$$P = K \rho^\gamma \quad (2.42)$$

where  $\gamma$  is the polytropic index of the gas. The polytropic constant,  $K$ , is given as a parameter and remains constant in the simulation. In the cases where the entropy is to remain constant, artificial viscosity acts to damp away kinetic energy, expelling it from the system. This of course means that energy is not conserved. Finally an adiabatic gas equation of state, in which the pressure is calculated using

$$P = (\gamma - 1)u\rho. \quad (2.43)$$

This is equivalent to (2.42) in the absence of viscosity. However, this time the internal energy is integrated. This of course means that the entropy function  $K$  is not constant but is given by

$$K = (\gamma - 1)u\rho^{1-\gamma}. \quad (2.44)$$

It can be shown that in the absence of viscosity,  $K$  is constant. Taking the time

derivative of (2.44) gives

$$\frac{dK}{dt} = -(\gamma - 1)^2 u \rho^{-\gamma} \frac{d\rho}{dt} + (\gamma - 1) \rho^{1-\gamma} \frac{du}{dt}. \quad (2.45)$$

From the first law of thermodynamics we know that

$$\frac{du}{dt} = \frac{P}{\rho^2} \frac{d\rho}{dt}. \quad (2.46)$$

Using (2.46) together with (2.43) we find that the rate of change of  $K$  is zero. Therefore in the case of no artificial viscosity,  $K$  should remain constant for each particle when using the ideal gas equation of state. When artificial viscosity is applied, the entropy is increased as energy has been transferred between particles. It is of course possible that each particle may have a different value for  $K$ .

## 2.6 Errors in the Density Estimate

The density estimation given by equation (2.10) is a smoothed version of the true underlying fluid density. We can quantify how useful this density estimation is by performing an error analysis. Whilst many authors (e.g. Price, 2004) have performed an error analysis on the density, the error given is the usually the bias (i.e. the average error) while the noise in the estimate is usually ignored. Furthermore the error analysis is usually performed assuming that the particles are distributed on grids, which is often not the case in SPH, although this assumption is often not explicitly stated. Particles may be distributed on glasses, sheared grids which have become anisotropic, or in some cases particles may be placed randomly. For randomly placed particles we can look to analytic expressions for the bias and variance of the density estimate, this will be discussed in section 2.6.1. However, in other particle distributions such as grids and glasses no analytic expression is easily obtainable. Therefore in section 2.6.2 we perform numerical tests to determine the bias and variance in some simple yet frequently occurring test cases.

### 2.6.1 The Bias and Variance for Poisson Noise

To achieve a given density field there are an infinite number of ways particles can be placed. For example one may create  $p$  realisations of the density field (with  $N$  particles per realisation) by randomly placing  $N$  particles for each distribution.

Therefore if we estimate the density at the point  $x$  for an ensemble of particle distributions, we expect a distribution in the estimated density. This distribution may be characterised by the bias and variance which give an indication of the error we may expect in the density estimate for a typical simulation. The bias of an estimator is the difference between the estimators expected value and the true value of the parameter being estimated, that is

$$\text{Bias}[\hat{\rho}(x)] = \langle \hat{\rho}(x) \rangle - \rho(x). \quad (2.47)$$

The variance of a distribution is the expected value of the square of the deviation of that variable from its expected value

$$\text{var}[\hat{\rho}(x)] = \langle \hat{\rho}(x)^2 \rangle - \langle \hat{\rho}(x) \rangle^2. \quad (2.48)$$

For a random distribution of points the density estimate  $\hat{\rho}(x)$  is a random variable due to its dependence on the random positions of the other particles. The expectation value, denoted  $\langle \hat{\rho} \rangle$ , is the mean of the random variable  $\hat{\rho}$ , and can be found as an ensemble average of the density estimate over all possible particle distributions. This is essentially the same as taking the limit of  $N \rightarrow \infty$  (see equation 2.3). Therefore the expectation value is

$$\langle \hat{\rho}(x) \rangle = \frac{1}{h^\nu} \int \rho(x_j) w\left(\frac{|x - x_j|}{h}\right) dx_j. \quad (2.49)$$

It is important to note that in this kernel estimation approach, the point at which the density is estimated is **not** at a given particles position, i.e.  $x \neq x_j$ . The kernel estimation is trying to estimate the density at a point in space (which is not one of the particle positions) due to distribution of the surrounding particles. Therefore the self-contribution should not be taken in computing the density as it introduces a bias.

In SPH this is not the case, the self-contribution is taken and therefore the expectation value contains an additional term, causing a bias when using a random distribution of particles (see Fig. 2.4). The reason SPH takes a self-contribution stems from the original motivation behind SPH. In SPH one uses cubic spline interpolation such that properties at a point in space are a weighted contribution from all points within the smoothing sphere (i.e. the gather approach, see section 2.4.1). As in SPH we always interpolate at a particle position, we not only include surrounding

particles, the particle its self must provide a self-contribution to the density. For randomly placed particles this self-contribution causes a bias in the density estimate. This is therefore a strong argument against using randomly placed particles for SPH simulations as we shall see in the following section.

Following Silverman (1986) we may Taylor<sup>6</sup> expand  $\rho(x_j)$  around  $x$  to find the expectation as

$$\langle \hat{\rho}(x) \rangle = \frac{1}{h^\nu} \int \left[ \sum_{\mathbf{k}} \frac{(\mathbf{x}_j - \mathbf{x})^{\mathbf{k}}}{\mathbf{k}!} \nabla^{\mathbf{k}} \rho(x) \right] w \left( \frac{|\mathbf{x} - \mathbf{x}_j|}{h} \right) d\mathbf{x}_j. \quad (2.50)$$

We may make a substitution such that  $\mathbf{x}_j = h\mathbf{y} + \mathbf{x}$  and thus  $d\mathbf{x}_j = h^\nu d\mathbf{y}$  and expand to second order to obtain the bias as

$$\text{Bias}[\hat{\rho}(x)] = \frac{h^2}{2} \Delta \rho(x) \int \mathbf{y}^2 w(|\mathbf{y}|) d\mathbf{y} + \text{Higher order terms}. \quad (2.51)$$

Therefore in uniform density situations, the bias should be zero (to order  $\mathcal{O}(h^2)$ ). The variance is also given by Silverman (1986) as

$$\begin{aligned} \text{var}[\hat{\rho}(x)] &= \frac{1}{nh^{2\nu}} \int \rho(x_j) w \left( \frac{x_j - x}{h} \right)^2 dx_j \\ &\quad - \frac{1}{n} \left[ \frac{1}{h^\nu} \int \rho(x_j) w \left( \frac{x_j - x}{h} \right) dx_j \right]^2. \end{aligned} \quad (2.52)$$

Using the expression for the bias and the expectation we may express this as

$$\text{var}[\hat{\rho}(x)] = \frac{1}{n} \frac{1}{h^{2\nu}} \int \rho(x_j) w \left( \frac{x_j - x}{h} \right)^2 dx_j - \frac{1}{n} (\rho(x) + \text{Bias}[\hat{\rho}(x)])^2. \quad (2.53)$$

As we know the bias is  $\mathcal{O}(h^2)$ , assuming  $n$  is large and  $h$  is small we may express  $\frac{1}{n} (\rho(x) + \text{Bias}[\hat{\rho}(x)])^2 \approx \mathcal{O}(n^{-1})$ . Thus the variance can be found by assuming a Taylor expansion to first order.

$$\text{var}[\hat{\rho}(x)] \approx \frac{1}{nh^\nu} \rho(x) \int w(|\mathbf{y}|)^2 d\mathbf{y} + \mathcal{O}(n^{-1}). \quad (2.54)$$

Therefore we have a trade off between the bias and variance. Increasing  $h$  will de-

---

<sup>6</sup>We use multi-index notation where  $\mathbf{k} \equiv (k_x, k_y, k_z)$ ,  $|\mathbf{k}| = k_x + k_y + k_z$  and  $\mathbf{k}! = k_x!k_y!k_z!$ . For example, the  $|\mathbf{k}| = 1$  term has three contributions  $(1, 0, 0)$ ,  $(0, 1, 0)$  and  $(0, 0, 1)$ . Therefore if  $\mathbf{r} = (x, y, z)$  the product  $\mathbf{r}^{\mathbf{k}} \nabla^{\mathbf{k}}$  for  $\mathbf{k} = 1$  unpacks to  $x\partial/\partial x + y\partial/\partial y + z\partial/\partial z$ .

crease the random variations and therefore the variance but increases the systematic error (the bias) in the estimate. Decreasing  $h$  will reduce the bias but increase the variance. (Silverman, 1986). Furthermore the variance will depend on the number of distributions we are averaging over. Using more distributions (increasing  $n$ ) reduces the variance. The bias is unaffected by using more distributions (provided we use large enough  $n$ ).

## 2.6.2 Numerical Results - Fixed $h$

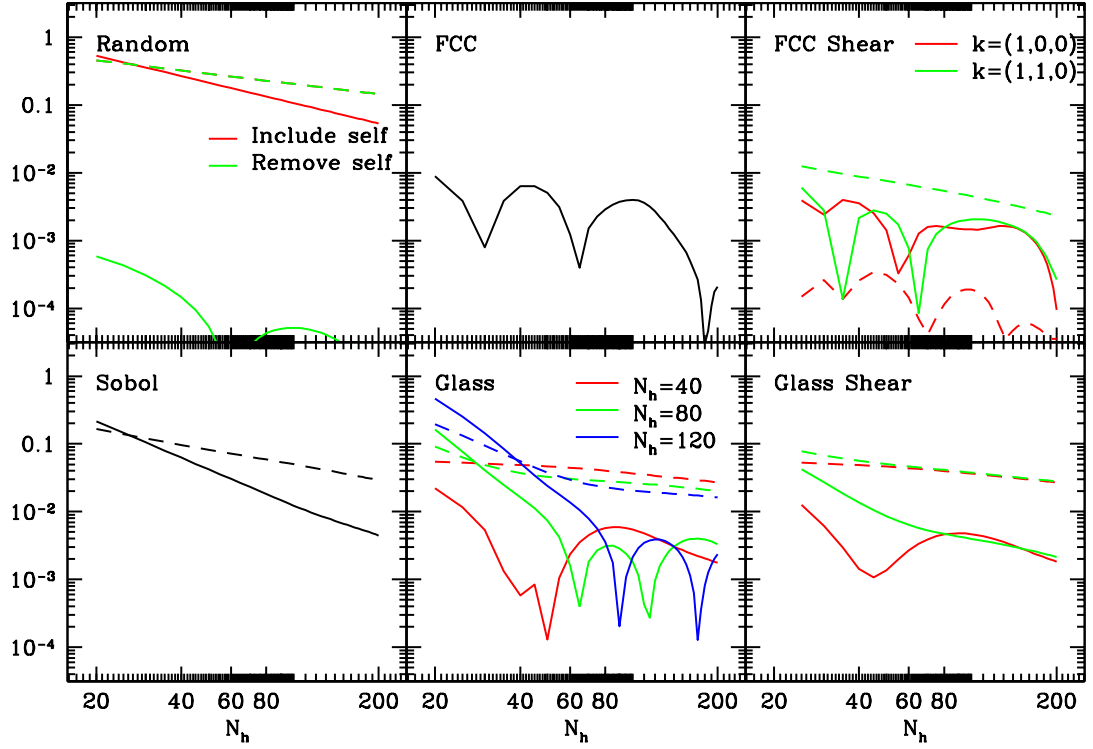
### Randomly Placed Particles

We firstly consider the bias and standard deviation for a uniform density field constructed using randomly positioned particles. As all of the particles are independent and have the same expected density, rather than taking an ensemble average over all possible distributions, we may take the average over all particles. This allows us to obtain estimates of the bias and standard deviation which can be compared to the analysis performed in the previous section. For random variables this approach should give a good approximation to the analysis. For other particle distributions the error properties may be vastly different from the analysis performed above.

We use a fixed smoothing length in order to be consistent between the numerical results and the analysis performed above. The smoothing length used depends on the number of neighbours and the total number of particles as

$$h \propto \left( \frac{N_h}{N} \right)^{1/3}. \quad (2.55)$$

Fig. 2.4 shows the bias and standard deviation of the density estimate for a uniform density field. We find that when the self contribution of the density is removed the bias is zero to within machine accuracy. However, including the self contribution to the density introduces a bias with the functional form  $\text{Bias} \propto h^{-3}$ . As stated earlier the expectation of the density estimate (equation 2.49) assumes that the density is estimated at the point  $x$  which is *not* one of the particle locations. When the self contribution is included, as is required in SPH, a bias is introduced as information about the particles position is already used. This demonstrates the importance of avoiding particle noise in SPH simulations. We see that the standard deviation can be readily reduced by increasing the number of neighbours (and therefore smoothing length), as expected from the analysis of the previous section. However, increasing



**Figure 2.4:** Bias (*solid lines*) and standard deviation (*dashed lines*) for the SPH density estimate for various uniform density particle distributions. The smoothing length is fixed for a given neighbour number according to equation (2.55).

neighbour number increases the computational cost and may introduce other problems such as particle clumping (see section 4.3 and Read et al., 2010).

Fig. 2.4 also shows the bias and standard deviation for a Sobol sequence, these are number generators that more evenly fill the domain (Sobol, 1967, see also section 4.2). In this case the self contribution is included as we are interested in how SPH deals with this type of particle distribution. We find that the variance is lower by a factor  $\sim 10$  than the random distribution for  $N_h = 40$ . We also find that when excluding the self contribution the bias is larger, indicating that the self contribution is required to reduce the bias.

### Ordered Particle Distributions

In a FCC (face-centred-cubic) grid the standard deviation is zero as all possible distributions are identical as particles may only be placed on lattice positions. Therefore taking different distributions equates to either simply shifting all lattice posi-

tions by a constant factor or swapping particles with each other. The bias, shown in Fig. 2.4, is small (but non zero) compared to the random distribution and Sobol sequence as the particles are not randomly distributed, they are correlated. Therefore the analysis performed in the previous section is not strictly correct as we assumed the density was a random variable. We find that the bias passes through zero<sup>7</sup> and changes sign at around  $N_h \approx 35$  and  $N_h \approx 65$ . Therefore using  $N_h \approx 35$  or  $N_h \approx 65$  is well justified as the bias in the density estimate is smallest around these points.

Fig. 2.4 shows the bias and standard deviation in the density estimate for a glass. The glass was initially created by applying random perturbations with a Gaussian distribution to a FCC lattice. The particles were then evolved with artificial viscosity until their kinetic energies became suitably small ( $\sim 10^{-5}$ ). We tested three glass configurations all using  $\sim 10^6$  particles. Each glass was created using a different number of neighbours, in this case 40, 80 and 120<sup>8</sup>. The smoothing length was then reset in order to perform the bias analysis.

We see that the bias for the  $N_h = 40$  glass drops sharply around  $N_h = 40$ . However, the bias for the  $N_h = 80$  glass dips before and after  $N_h = 80$ . This is also true when using  $N_h = 120$  neighbours to make the glass. The minimum in the bias is again small but non-zero and is comparable to that of the FCC lattice. The standard deviation on the other hand is much larger than in the FCC lattice as all particles are not equivalent, we have effectively introduced noise into the particle distribution.

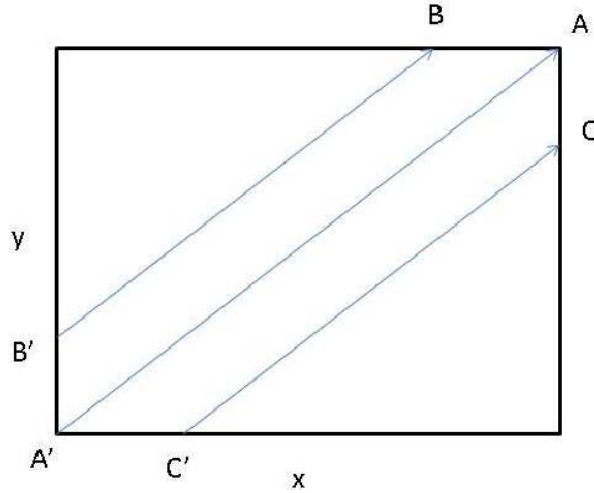
## Shear Flows

Shear flows can be tricky to model in SPH as particles distributions that are initially isotropic quickly become anisotropic due to the shearing motions of the particles. To test the accuracy of the density estimate we set up a uniform density box in 3D, placing particles on an FCC lattice. We then applying a shearing in two directions,  $\mathbf{k} = (1, 0, 0)$  and  $\mathbf{k} = (1, 1, 0)$ . The particles positions were evolved without hydrodynamic forces until  $t = 0.3$ . Evolving the particle positions for  $\mathbf{k} = (1, 1, 0)$  is somewhat tricky with periodic boundary conditions. All particles move along a constant line in time and particles move outside the periodic box once they reach the end of their line. When this happens the particles are placed back at the start of their line (see Fig. 2.5)

<sup>7</sup>In Fig. 2.4 and Fig. 2.6 we plot absolute values of the bias

<sup>8</sup>The smoothing length is allowed to vary according to 2.11



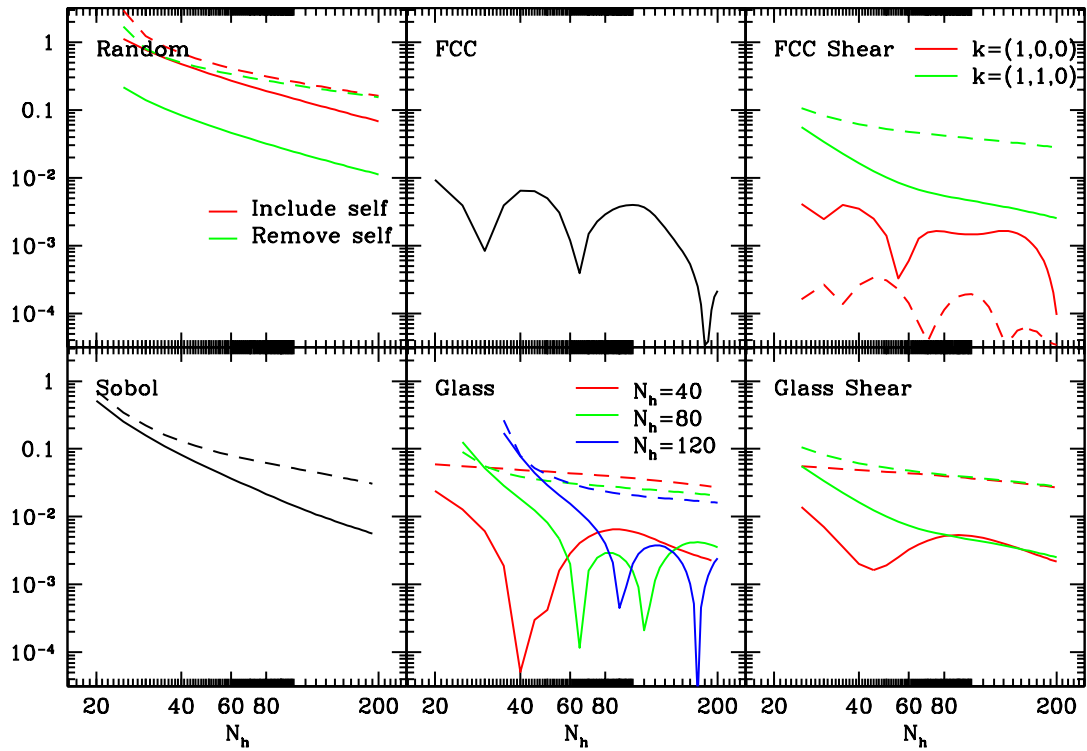


**Figure 2.5:** Shear flow in the  $k = (1, 1, 0)$  direction, as particles leave the periodic box at  $A$ ,  $B$  and  $C$  they re-enter it at  $A'$ ,  $B'$  and  $C'$ .

We applied the shear flow to both an FCC lattice and a glass (made with  $N_h = 40$ ) as shown in Fig. 2.4. We find that the bias remains of the same order as the distributions without the shear flow imposed. However, the standard deviation becomes significantly higher, especially for flows in the  $k = (1, 1, 0)$  direction. The implication of this is that whilst on average the properties of the flow are the same, the estimates have become noisy. This can be a severe problem for SPH as many techniques rely on having smoothed, noise-free data (for example shock detection schemes, see section 3.4.3). Both the bias and variance are lower when shearing an FCC lattice rather than a glass.

### 2.6.3 Numerical Results -Varying $h$

The analysis and numerical tests performed above assumed a fixed SPH smoothing length, however in practice this is not done: the smoothing length is varied to allow the resolution to adapt to the local conditions. The density is strongly linked to the smoothing length via the kernel. Therefore we perform the numerical tests allowing a varying smoothing length. The results for the FCC lattice, glass and shear flow, as shown in Fig. 2.6, are similar to that of using a fixed smoothing length. However, we find that for the random distribution a significant bias is introduced when using randomly placed particles and removing the self contribution to the density. When the self contribution is allowed the random distribution shows a similar bias to a fixed smoothing length, although it falls faster for lower neighbour number. This



**Figure 2.6:** Bias squared (*solid lines*) and variance (*dashed lines*) for the SPH density estimate for various uniform density particle distributions. The smoothing length is allowed to vary for a given neighbour number according to equation (2.11).

is also true for the Sobol sequence, the bias falls faster than the fixed  $h$  for smaller neighbour number.

These results show that the standard deviation (noise) error can easily be reduced by selecting ordered particle distributions which are free from particle noise. We find that randomness in particle positions should be avoided wherever possible as it introduces a large bias and standard deviation. Although the Sobol sequence attempts to limit the clustering of particles by distributing them more evenly, we still find an unacceptably high bias and standard deviation. Whenever possible a FCC lattice should be used to construct initial conditions as the noise in the density estimate is greatly reduced, as is the bias. Failing this one should relax the initial conditions into a glass which can significantly reduce the bias errors, especially when using  $N_h = 40$ . The reduction of both bias and noise errors can be important for other aspects of the SPH scheme, in particular in the setting of artificial viscosity. This will be discussed in more detail in chapter 3.

## 2.7 Gradient Estimators

A conservative SPH scheme does not require an estimate for the SPH gradient of a quantity, only  $\nabla \cdot \mathbf{v}$  is required which can be computed directly from the adaption of smoothing lengths (see equation 2.28). However, one may need to compute gradients for extensions to the SPH scheme, for example in viscosity or MHD (magneto hydrodynamics). Also if one wanted to move away from a conservative SPH scheme a gradient operator would be required for deriving the momentum equation. The expected value for the gradient of a quantity can be simply expressed as

$$\langle \nabla A(\mathbf{x}) \rangle = \int A(\mathbf{x}') \nabla W(\mathbf{x} - \mathbf{x}', h). \quad (2.56)$$

This discretises as

$$\widehat{\nabla_i A_i} = \sum_j m_j \frac{\hat{A}_j}{\hat{\rho}_j} \nabla W \left( \frac{|\mathbf{x}_i - \mathbf{x}_j|}{h} \right). \quad (2.57)$$

However, it is easy to show that a significant improvement can be made on this estimation. Taking a Taylor expansion of  $A(x_j)$  around  $x_i$  we obtain

$$\begin{aligned} \langle \widehat{\nabla_i A_i} \rangle &= A_i \sum_j \frac{m_j}{\rho_j} \nabla W \left( \frac{|\mathbf{x} - \mathbf{x}_j|}{h} \right) \\ &\quad + \nabla_i A_i \cdot \sum_j \frac{m_j}{\hat{\rho}_j} (\mathbf{x}_j - \mathbf{x}_i) \nabla W_{ij} \\ &\quad + \mathcal{O}(h^2). \end{aligned} \quad (2.58)$$

The summation part of the first term in this expression is simply the SPH estimate for the gradient of unity. In conditions where there is particle noise this will be non-zero. Therefore we may simply correct away this error by using a gradient estimator of the form

$$\widehat{\nabla_i A} = \sum_j m_j \frac{\hat{A}_j - \hat{A}_i}{\hat{\rho}_j} \nabla W \left( \frac{|\mathbf{x}_i - \mathbf{x}_j|}{h} \right). \quad (2.59)$$

This trick is often referred to as the “second golden rule of SPH” (Monaghan, 1992) and is readily obtained by taking the density inside the gradient operator

$$\nabla A = \rho^{-1} [\nabla (A\rho) - A \nabla (\rho)]. \quad (2.60)$$

The summation part of the second term in equation (2.58) gives a matrix,  $\sum_j \frac{m_j}{\rho_j} (\mathbf{x}_j - \mathbf{x}_i) \nabla W_{ij} = \underline{\mathbf{M}}_i$ . The extent to which the estimate of the gradient is equal to the underlying fluid gradient is determined by how well  $\underline{\mathbf{M}}_i = \underline{\mathbf{I}}$ . In principle this error can easily be corrected for by dividing the SPH estimate for the gradient by this matrix (Price, 2004; Read et al., 2010).

The error in the SPH estimate of a gradient is usually quoted as  $\mathcal{O}(h^2)$ , which is the third term in the expansion given by equation (2.58). However, as we have just demonstrated this is only valid for flows where particle noise is not an issue. Therefore the estimate for the gradient in typical SPH simulations, which may contain particle noise or anisotropic particle distributions, is usually lower order than this.

### 2.7.1 An Alternative Gradient Operator

We may also rewrite the density inside the gradient operator in the following way,

$$\nabla A = \frac{A}{\rho} \frac{\partial \rho}{\partial \mathbf{x}} + \rho \frac{\partial}{\partial \mathbf{x}} \left( \frac{A}{\rho} \right) \quad (2.61)$$

which can be discretised as

$$\widehat{\nabla_i A_i} = \sum_j m_j \left[ \frac{\hat{A}_i}{\hat{\rho}_i} + \frac{\hat{\rho}_i \hat{A}_j}{\hat{\rho}_j^2} \right] \nabla W_{ij}. \quad (2.62)$$

We can determine the error in this estimate by performing a Taylor expansion of  $A(\mathbf{x}_j)$  around  $\mathbf{x}_i$  (Price, 2004)

$$\langle \widehat{\nabla_i A_i} \rangle = A_i \sum_j m_j \left( \frac{1}{\hat{\rho}_i} + \frac{\hat{\rho}_i}{\hat{\rho}_j^2} \right) \nabla W_{ij} \quad (2.63)$$

$$+ \nabla_i A_i \cdot \sum_j \frac{m_j \hat{\rho}_i}{\hat{\rho}_j^2} (\mathbf{x}_j - \mathbf{x}_i) \nabla W_{ij} + \mathcal{O}((x_j - x_i)^2). \quad (2.64)$$

If we were to remove the first error term the expression for the SPH gradient would be

$$\widehat{\nabla A} = \hat{\rho}_i \sum_j m_j \frac{\hat{A}_j - \hat{A}_i}{\hat{\rho}_j^2} \nabla W_{ij}. \quad (2.65)$$

This is strikingly similar to what we had before. Therefore when computing the gradient of an SPH quantity, if we treat the first error term we always obtain an

expression similar to equation (2.59). Therefore treating the first error term will always ensure that the gradient estimate is zero if the fluid quantity is constant.

The expression (2.62) is a common choice for the gradient operator used for the acceleration equation when not deriving it from a Lagrangian

$$a_i = -\frac{\nabla P}{\rho} = -\sum_j m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \nabla W_{ij}. \quad (2.66)$$

The advantage of this method is that it is symmetric between particle pairs unlike the other expressions we have derived. This is particularly useful for the acceleration equation as it allows linear and angular momentum conservation. The force on particle  $i$  due to particle  $j$  is

$$F_{i \rightarrow j} = -m_i m_j \left[ \frac{A_i}{\rho_i^2} + \frac{A_j}{\rho_j^2} \right] \frac{\mathbf{x}_{ij}}{h^{\nu+2}} \tilde{w}(q). \quad (2.67)$$

Similarly the force on particle  $j$  due to particle  $i$  is

$$F_{j \rightarrow i} = -m_i m_j \left[ \frac{A_i}{\rho_i^2} + \frac{A_j}{\rho_j^2} \right] \frac{\mathbf{x}_{ji}}{h^{\nu+2}} \tilde{w}(q) \quad (2.68)$$

thus only differing by the sign of the force ( $\mathbf{x}_{ij} = -\mathbf{x}_{ji}$ ), therefore automatically obeying Newton's 3rd law as the sum of all forces evaluates to zero.

However, this method has the disadvantage that the gradient may be non zero even when  $A$  is constant. This can be problematic in situations where contact discontinuities are important, such as the simulation of Kelvin-Helmholtz instabilities (KHI). This is where two fluids with different densities and equal pressures shear each other, leading to the growth of the KHI. This growth can be severely suppressed as particles at the boundaries tend to be spuriously repelled from each other, which inhibits mixing. If we were to correct for the first error term the derivative would be exactly zero when the function is constant. However, as soon as we correct for the first error term we break the symmetry between particles and therefore lose the momentum conserving property.

However, there have been suggestions in the literature (e.g. Abel, 2010) to use the SPH gradient (2.59) to obtain a momentum equation which gives

$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \frac{P_j - P_i}{\rho_i \rho_j} \nabla W_{ij}(h_i). \quad (2.69)$$

This equation has the advantage that in the case of constant pressure the acceleration is exactly zero. However, it can easily be shown that this equation violates both momentum and angular momentum conservation (Monaghan, 1992). The conventional wisdom of most SPH practitioners is that it is better to use the second golden rule of SPH to obtain the symmetric form of the momentum equation given by equation (2.66).

There are also many other methods one could use to calculate the gradient of an SPH quantity, for example Hernquist & Katz (1989) use  $\sqrt{P_i P_j}/(\rho_i \rho_j)$  as yet another symmetric combination for the gradient operator in order to compute the acceleration. This form of the momentum equation is symmetric for fixed smoothing length, but for variable smoothing length the equation is not symmetric as the kernel only takes particle  $i$ 's smoothing length. However one can use the average of the smoothing lengths inside the kernel or use the average of the two kernels with each particles smoothing length to symmetrise the contribution from the kernel. The down side of doing this is the gradient of the kernel with respect to the smoothing length has not been taken into account.

In all of these derivations of the momentum equation some form of dissipative term is required to smooth out discontinuities and generate entropy at shocks. This is usually applied in the form of artificial viscosity, described in full in section 3.

## 2.8 Gravity

From star and planet formation to the collisions and dynamics of galactic clusters, gravity plays an extremely important role in astrophysical processes. Therefore in order for an SPH code to be useful to astrophysicists, the fluid must interact gravitationally, not only with itself but also with external bodies. As simulations involving millions of SPH particles are not uncommon, we require an algorithm to compute the gravitational forces that does not scale too steeply with the number of bodies. Interestingly SPH does not suffer from this problem as it scales<sup>9</sup> as  $\mathcal{O}(NN_h)$ . A naive method to compute gravitational forces between interpolation points would be to simply perform a double loop over all the particles, computing the distance between two particles and then computing the force. As the simulation size increases this algorithm will become extremely expensive as the number of computations required scales as  $\mathcal{O}(N^2)$ .

---

<sup>9</sup>Although this scaling ignores the fact that smaller time-steps are required at higher resolutions.

Many gravity codes use a method which allows the force from a set of distant particles to be computed faster than a naive  $\mathcal{O}(N^2)$  method. Firstly one builds an oct-tree comprised of nodes which are either bodies (particles) or cells where a cell represents a group of particles. The force on each particle is calculated by walking the tree from the root node and progressing downwards into each daughter node. If the particle and cell in question are well separated (which is determined by an acceptance criterion) then instead of opening the cell and computing the force exerted by each daughter node, the potential from the cell is approximated via a multipole expansion. If the cell is not well separated the force is evaluated by summing contributions from the daughter nodes of the cell, which may either be a particle-particle force or a cell-particle force. The tree walk has a complexity  $\mathcal{O}(\log(N))$  per body, giving a total complexity of  $\mathcal{O}(N \log(N))$  (for an overview see Barnes & Hut, 1986).

An alternative method for computing long range forces is the Fast Multipole Method (FMM) where particles are sorted into a hierarchy of nested grids. The multipole moments of each grid cell are computed and the forces between grid cells are computed by a multipole expansion. In this case the cells not only generate a force, they also receive a force, i.e. they are both sinks and sources. This formally reduces the complexity to  $\mathcal{O}(N)$ . However, the FMM method is rarely used in astrophysics as the method is non-adaptive and to achieve the same accuracy of a tree code it becomes expensive (for more details see Greengard & Rokhlin, 1997).

Whilst these methods can improve the scaling properties of a code it should be noted that the overall computational cost per force calculation is not the same with each method. The direct summation method is a simple and fast method to calculate the force whereas building a tree or FMM structure is relatively expensive and therefore the cost for a *single* force calculation is larger. In other words, whilst the direct method may scale as  $\alpha N^2$  and the tree code as  $\beta N \log(N)$ , the direct method will be faster for small  $N$  as  $\beta \gg \alpha$ . Both Pringle (1995) and Li et al. (2009)<sup>10</sup> find that for less than around 1000 particles the direct summation method is faster than the tree code. However, the exact point at which this occurs depends on the implementation details and will vary from code to code.

For the number of particles used in a typical astrophysical simulation, typically at least 100,000 (but usually more like millions), the tree code provides a much needed speed up of the gravitational force calculation. Whilst a tree code can reduce the

---

<sup>10</sup>Although they considered coulomb forces rather than gravity.

computational effort required the force calculated will only be an approximation of the true force. However for many astrophysical problems an approximation of the gravitational force is preferable to small number of particles or long running times. In any case the particles are usually not real gravitational points (such as stars in a galaxy) they are a Monte Carlo representation of the mass distribution function.

One drawback of the tree code is that the force  $F_{A \rightarrow B} \neq F_{B \rightarrow A}$ . The reason for this is that  $F_{A \rightarrow B}$  is computed by taking a multipole expansion of the cell containing particle  $B$ , whilst  $F_{B \rightarrow A}$  is computed by taking a multipole expansion of the cell containing particle  $A$ . Therefore the momentum of the system is not conserved.

A vast improvement on the tree code is to use a mutual gravitational force solver as described by (Dehnen, 2000, 2002) which is essentially an adaptive FMM scheme based on a hierarchical tree of cubic cells. The forces between two well separated cells are calculated in a symmetric way in which both cells are sources and sinks.

The algorithm works by firstly building a oct-tree where each cell has up to eight daughter nodes. Each daughter node can be either a single body or another cell. After the tree build we compute each cell's mass, centre of mass, and quadrapole moments. Starting at the root-root interaction we decide what type of interaction two nodes should take. Firstly a body self interaction is ignored as a single particle cannot gravitationally attract itself. However, a cell may interact with itself gravitationally as the cell is a collection of particles. An interaction between well separated nodes (determined by a multipole acceptance criteria) is solved by computing the potential, force and the potentials higher derivatives at the centre of the sink cell from the multipole expansion of the source cell. If the cells are not well separated the node with the largest maximum distance from the centre of mass is split into, at most, eight new cells and mutual interactions between the daughter cells are computed. After this interaction phase the Taylor coefficients are accumulated and the expansion is evaluated at every body's position. For more details see Dehnen (2000, 2002).

## 2.9 Time Integration

In any numerical scheme time integration is necessary to move the state of the system forward in time. Therefore the choice of time integrator is of upmost importance and should not be taken lightly. A good time integrator should have good conservation characteristics, not only in the short term (e.g. a single step), but also long term



conservation. A computational requirement is that the equations being modelled may be stably integrated without having to resort to a prohibitively short time step, this is particularly important in SPH as the force computation is the most expensive part of the algorithm. Therefore we aim to get from  $t_{start}$  to  $t_{end}$  using the fewest number of time steps possible. The Euler method is a well known example of an unstable integrator with slow convergence of error with step size.

A more useful and commonly used integrator is the Leapfrog method. Unlike other integrators (such as Runge Kutta algorithms) the leapfrog is time reversible due to its time symmetry. The time reversibility of the leapfrog algorithm ensures that the Hamiltonian is conserved, leading to the conservation of energy. A comparison of the RK4 and leapfrog methods can be seen in Springel (2005) and Quinn et al. (1997), where the authors compare the accuracy of the energy conservation for elliptical orbits. Although the RK4 algorithm has a smaller error for a particular time step the total error grows with each time step. The leapfrog algorithm however conserves the energy to the same accuracy for an infinite amount of orbits. The reason for this lies in the time reversibility of the leapfrog scheme. Suppose the error in the integration of an orbit is  $\delta E$ . Then the time reversed orbit must also have an error of  $\delta E$  as the time reversed orbit is equivalent to an orbit forward in time where  $v \rightarrow -v$ . Therefore in order to be time reversible (zero error after the forward and backward orbit) the total error over the orbit must be zero. However, the error over segments of the orbit do not have to be zero as the energy calculated is an approximation to the true energy of the system.

Due to its long term stability, the leapfrog algorithm is a common choice for astrophysical problems. We use the second order symplectic Kick-Drift-Kick Leapfrog integrator. In N-body dynamics the acceleration is dependant only on the positions of the particle and therefore can be computed after the drift step as this gives the particle positions at the end of the step. However, in SPH the rate of change of SPH quantities may depend on itself or other quantities which are being kicked, for example the rate of change of the internal energy  $du/dt$ , depends on the velocity. In order to compute these derivatives after the drift step we must know the values at the end of the step. Therefore a prediction of the SPH quantities are needed. We

adapt the usual leapfrog algorithm for SPH.

$$\mathbf{v}_p = \mathbf{v}_0 + \tau \mathbf{a}_0 \quad (2.70a)$$

$$\mathbf{v}_{1/2} = \mathbf{v}_0 + \frac{\tau}{2} \mathbf{a}_0 \quad (2.70b)$$

$$S_p = S + \tau \dot{S}_0 \quad (2.70c)$$

$$S_{1/2} = S_0 + \frac{\tau}{2} \dot{S}_0 \quad (2.70d)$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \tau \mathbf{v}_{1/2} \quad (2.70e)$$

$$\mathbf{a}_1 = \dot{v}(\mathbf{x}, v_p, S_p) \quad (2.70f)$$

$$\dot{S}_1 = \dot{S}(\mathbf{x}_1, \mathbf{v}_p, S_p) \quad (2.70g)$$

$$\mathbf{v}_1 = \mathbf{v}_{1/2} + \frac{\tau}{2} \mathbf{a}_1 \quad (2.70h)$$

$$S_1 = S_{1/2} + \frac{\tau}{2} \dot{S}_1. \quad (2.70i)$$

Here the subscript  $_p$  denotes predicted quantities and  $S$  is any SPH quantity. In the acceleration calculation (2.70f) the density used is either the SPH density calculated by the summation equation or the predicted density if the continuity equation is integrated. We should note here that individual variables are not required to store all of these quantities. For example, a single variable for  $\mathbf{x}$  per particle is all that is required as we may overwrite  $\mathbf{x}$  at the drift step. That is in C/C++ code we may use  $\mathbf{x} \stackrel{\pm}{=} \tau \mathbf{v}_{1/2}$ .

We also allow for a variable time step in the KDK scheme. This is introduced via the notion of time-step levels, each particle is allowed to sit on any time-step level and move between levels (there are some restrictions placed on this however). The longest time-step allowed corresponds to the synchronisation time at which all particles are moved together; this is taken as a parameter at initialisation. Other time-step levels may be shorter by factors of 2. In a scheme such as this any particle may move on to a shorter time step at any time. However, one must be careful allowing particles to move onto a longer time-step. A particle may only move onto a longer time step if the shorter step is at the start of the longer step. Failure to do this causes particles to become out of sync (Bate, 1995). Furthermore one must note that the use of variable time-steps breaks the momentum and energy conservation properties of SPH, however the violation is usually small.

The placement of particles on time-step levels may be determined by several parameters. The traditional method for gas dynamics is to use the Courant condition

(2.71a). A modified version of this has been suggested by Bate (1995) which includes contributions from the characteristic velocity and viscosity terms (2.71b). We are able to create many other time-step parameters by taking a ratio of the quantity to the rate of change of that quantity (equations 2.71c to 2.71e). We may also derive an expression based on the smoothing length and acceleration (2.71f).

$$\tau = \frac{f_{cs}h}{C_s} \quad (2.71a)$$

$$\tau = \frac{f_{cm}h}{C_s + h|\nabla \cdot v| + 1.2(\alpha C_s + \beta h|\nabla \cdot v|)} \quad (2.71b)$$

$$\tau = \frac{f_{ca}C_s}{a} \quad (2.71c)$$

$$\tau = \frac{f_{hd}h}{\dot{h}} \quad (2.71d)$$

$$\tau = \frac{f_{du}U}{\dot{U}} \quad (2.71e)$$

$$\tau = f_{ha}\sqrt{\frac{h}{a}}. \quad (2.71f)$$

There are many other methods used for N-body codes involving quantities such as the gravitational potential. However, these often do not have the correct units (of time) or are not Galilean invariant; therefore they are undesirable. Note that time step parameters involving the acceleration are not invariant to the transformation to a reference frame undergoing a constant acceleration.

If more than one of the parameters are used, the particles are placed on the shortest time-step level given by the parameters to ensure accurate integration. The factors in front of the time-step levels, the minimum time-step and number of time-step levels are set at run time.

## 2.10 Code Details

Thus far we have outlined the SPH method for computation of the density, momentum and internal energy of a fluid. We have discussed refinements to SPH such as variable smoothing lengths and have explored the resolution element of SPH. We have discussed the method used to compute the forces due to self gravity and how to integrate the system.

Finally we will now discuss how each of these methods are integrated into a single code, in particular how the SPH and gravity computations take place. We do

not discuss other aspects of the code such as memory management etc.

The work contained in the remainder of this thesis involves applying the SPH method to a set of problems. To do this two separate codes were used. The main code for astrophysical simulations used was **SPHfalcON** which is a fully 3D SPH code with self-gravity, external potentials and variable time-steps<sup>11</sup>. In order to test various SPH algorithms and properties, we found it was useful to have a code that could also work in 1D and 2D (1DSPH and 2DSPH). These codes are essentially simpler versions of **SPHfalcON** (with reduced functionality) but still use exactly the same SPH algorithms applied to 1D and 2D respectively.

### 2.10.1 Hydrodynamics Calculation

This section outlines the hydrodynamical algorithms common to the three codes mentioned above. In our code we use the standard SPH representations of the fluid equations given (2.10), (2.39) and (2.40). We use the notion of sweeps, which are essentially a double loop. The first (outer) loop is over the SPH particles, whilst the inner loop is over the particles neighbour list.

The SPH computation starts by predicting  $h$ ,  $f$ ,  $\rho$  and  $\alpha$  for each particle (both active and inactive)<sup>12</sup> based on the current information. The smoothing lengths are predicted using equation (2.24). The density is predicted based on the prediction of the smoothing lengths, using

$$\rho_i(t + \delta t) = \rho_i(t) \left( \frac{h_i(t + \delta t)}{h_i(t)} \right)^{-3}. \quad (2.72)$$

The  $f$  factors due to the taking gradient of the kernel with respect to the smoothing length are also predicted for inactive particles. This is done using the relationship of equation (2.11) and assuming that  $\mu_i$  is constant. This leads to

$$f_i(t + \delta t) = f_i(t) \frac{\ln h_i(t + \delta t)}{\ln h_i(t)}. \quad (2.73)$$

We firstly build a neighbour list by constructing a tree. We then adjust the smoothing lengths according to the method outlined in section 2.4.2

We then perform an SPH sweep, computing  $\mu_i$ ,  $\partial\mu_i/\partial t$  and  $\partial\mu_i/\partial h_i$  using equa-

---

<sup>11</sup>SPHfalcON is the SPH extension of the N-body code **GyrfalcON** Dehnen (2002)

<sup>12</sup>An active particle is one where the force computation is taking place. An inactive particle is one where the force computation is not taking place

tions (2.11), (2.21) and (2.18). In between sweeps one and two we perform some ‘book keeping’, setting the density,  $P/\rho^2$  and the sound speed of each particle using the desired equation of state. We also set  $\alpha_i$  for active particles. In the second SPH sweep we compute  $dv/dt$  and  $du/dt$  using the SPH equations and the neighbour list. We then destroy the neighbour list.

### 2.10.2 Periodic Boundary Conditions

We allow for periodic boundary conditions in each coordinate direction which can be controlled individually, that is the boundary size can vary in each coordinate direction. We may even turn periodic boundaries off in one direction whilst keeping them in the others. If the distance between two particles in the coordinate direction of interest is larger than the half period of the boundary then the shortest path is to go around the periodic boundary. Therefore we adjust the distance by the full period in that direction.

### 2.10.3 External Potentials

An external potential can be used to apply a force to all particles. Typical external potentials used are Keplerian potentials and plummer potentials. This allows particles to act as though they were in a fixed gravitational field without having to compute the self gravity between particles.

### 2.10.4 SPHfalcON Specific Algorithms

#### Type of particles

In SPHfalcON we allow for two different types of particle, pure N-body particles and SPH particles. N-body particles only take part in the gravitational force interaction whereas SPH particles take part in both gravity and hydrodynamics. One may achieve pure hydrodynamics by using only SPH particles and setting the gravitational constant to zero. In the SPHfalcON code we use the notion of active and inactive particles. Active particles are particles that require a force computation whereas for inactive particles do not require a force computation as they are on a longer time step which is not currently being evaluated.

## Gravity Calculation

To compute the gravitational force we use the method described by Dehnen (2000, 2002) as outlined in section 2.8.

## SPHfalcON Calculations

The force computation takes place in the following order

1. If we are using periodic boundary conditions we first check that the particles positions are contained within the periodic box.
2. If we are using varying softening lengths we set the softening length of the SPH and N-body particles.
3. Build the tree
4. Set the gravitational forces
  - (a) Compute self gravity if turned on
  - (b) Compute external gravity if turned on
5. Computes SPH forces
  - (a) Predict  $h$  for active particles and  $h, f, \rho$  and  $\alpha$  for inactive particles.
  - (b) Sweep 0 - Adjust the smoothing lengths for active particles.
  - (c) Sweep 1 - Compute  $\mu_i$  and derivatives for active particles. We also compute the density and pre-factors (such as  $\hat{P}/\hat{\rho}^2$ ).
  - (d) Sweep 2 - Compute SPH forces and rate of change of internal energy for active particles.

## 2.11 Summary

In this chapter we have detailed a method of simulating fluid flows using smoothed particle hydrodynamics (SPH) (Lucy, 1977; Gingold & Monaghan, 1977). SPH is a mesh-free Lagrangian approach as opposed to methods based on grid cells, such as Godunov's scheme. This mesh-free property of SPH makes the method easy to apply to a wide range of astrophysical problems.

The basic principle behind this method is to divide the fluid up into discrete interpolation points known as particles which are free to move. The particles have a spatial distance known as the smoothing length over which their properties are smoothed by the kernel. The density of each SPH particle may be computed by summing weighted contributions of each neighbouring particle according to their distance from the particle of interest. As the particles are free to move this means that regions of higher density automatically have a higher number of particles, and therefore the resolution of the scheme adapts to the local density of the fluid.

Although the size of the smoothing length can be fixed in both space and time, this does not take advantage of the full power of SPH. By allowing each particle's smoothing length to vary in time, the resolution of a simulation can be made to automatically adapt itself depending on local conditions. The smoothing length is automatically adjusted such that the mass enclosed in a smoothing sphere is constant. SPH forms of Euler's equations may be derived self-consistently from the Lagrangian of the system, allowing one to compute the fluid forces acting on each particle.

In astrophysics, one often wants to supplement a hydrodynamical calculation with gravity. SPH allows one to model gravity by the use of either an external potential or N-body type point masses. These point masses interact with the SPH particles via gravity only and therefore are often used to represent dark matter particles. Alternatively, one may allow SPH particles to interact with each other gravitationally, thus allowing the gas to be self-gravitating, as is the case in many astrophysical situations. Incorporating other astrophysical processes such as radiative transfer (Whitehouse & Bate, 2004; Bastien et al., 2004; Stamatellos & Whitworth, 2005) and magnetic fields (Dolag et al., 1999; Price, 2004) has had some success and is still an active area of research. SPH has also been applied to many other areas, such as deformation and fracture of solids (Jutzi et al., 2008).

SPH is far from complete, with many problems still unsolved. For example, there are still stability issues with SPH such as the well known particle clumping problem (Imaeda & Inutsuka, 2002). Furthermore mixing at discontinuities can also be problematic in SPH (Agertz et al., 2007), leading to problems in simulations of effects such as Rayleigh-Taylor and Kelvin-Helmholtz instabilities. However more recent research (Price, 2008; Read et al., 2010) suggest that these problems can be overcome. Finally we detailed the specific implementation of SPH used in section 2.10.

---

# 3

## Artificial Dissipation in SPH

### 3.1 Introduction

The formulation of SPH from a Lagrangian as described in §2.5 (see also Springel & Hernquist, 2002) automatically conserves mass, energy, entropy, linear and angular momentum. This conservation of entropy means that SPH is a *dissipation-less* scheme. The dissipation-less properties of SPH can be observed by simulating a small amplitude 1D sound wave, the amplitude and period of the wave are preserved (see §2.4.4).

However, in a real fluid entropy is not conserved, in fact entropy generation will occur readily at shock fronts where collisions between particles randomise their velocities thus generating heat and entropy. This basic collisional mechanism is inherent to all fluids (except for dust and collisionless plasma, which therefore may not be considered fluids) and prevents the flow from becoming multi-valued. This process occurs on the scale of the mean free path of the fluid, which is many orders of magnitude below the resolution of any numerical scheme. In this sense, SPH is a collisionless method at the molecular level. As we are not able to resolve this process, an artificial prescription is required to mimic this entropy generation.

In the formulation of SPH, we have assumed that all SPH quantities are smooth functions with no discontinuities. This smoothness is also assumed when we estimate derivatives: we assume that we are differentiating a smooth function without any



discontinuities. The smoothing of all quantities is a fundamental requirement for SPH, as the particles are just a smoothed representation of the underlying fluid.

Even in nature, fluid properties at shocks are not discontinuous, they change steadily over the mean free path of the fluid. The fluid simply appears discontinuous as we are not easily able to observe scales as small as the fluid mean free path. In a SPH scheme (or any other numerical method) we have no way of resolving length scales of the order of the mean free path of the fluid, the number of particles (or grid cells) required would be of the same order of the number of particles in the fluid, thus becoming prohibitively expensive. Grid codes attempt<sup>1</sup> to simulate the fluid by modelling fluid discontinuities as sharp jumps as seen at the macro level in nature. However, SPH is not designed to resolve fluid discontinuities on the scales smaller than the resolution. By smoothing the SPH quantities we are allowing for the fact that the fluid does change continuously across shocks, however we do not have the resolution to get the size of this region correct. The method preserves the position of the shock and the post and pre-shock values, however nothing can be determined about the region over which the shock occurs. Therefore, a “fundamental” difference between SPH and grid codes is that at the resolution limit SPH smooths fluid properties whereas grid codes introduce discontinuities.

In order to smooth out fluid properties at discontinuities in SPH, dissipation is required. As we are not able to resolve scales as small as the mean free path of a fluid, we are forced to take a sub-resolution model of this smoothing process. The aim of these artificial methods are to only smooth the fluid properties where the flow is approaching a discontinuity. As the method is artificial we should only introduce it to *mimic* and *not model* a real physical process, therefore we should aim to reduce artificial dissipation in a flow that is not discontinuous or approaching a discontinuity. For example, in sound waves some parts of the wave are converging but the velocity is not changing discontinuously. Therefore dissipation is not required to further smooth the velocity profile. However, as the sound wave steepens the velocity will start to approach a discontinuity, at which point dissipation is required.

In this chapter we will discuss the artificial dissipation process used in SPH schemes. We will firstly look at artificial viscosity, which is used to smooth discontinuities in velocity, and propose an improved algorithm to get the strength of this smoothing correct. Secondly we will look at other dissipative processes (such

---

<sup>1</sup>The subject of many grid code papers focus on how to sharpen discontinuities by making the scheme higher order, for example Sod (1978); van Leer (1979); Colella (1990); Toro (1999); Kermani & Stockie (2003)

as thermal conductivity Price, 2008) which are used to smooth away discontinuities in other fluid quantities.

## 3.2 Artificial Viscosity

Artificial viscosity was primarily introduced into the SPH scheme in order to resolve shocks and prevent particle penetration.<sup>2</sup> Particle penetration in SPH can occur as the repulsive pressure force is reduced to zero as particles become arbitrary close, in other words the force is softened in a similar manner to N-body methods. This is done automatically by the cubic spline kernel as the first derivative approaches zero inside a critical radius as the particles approach each other (see Fig. 2.1). As SPH is entropy conserving, a mechanism is needed to generate entropy at shock fronts. This is achieved by dissipating velocity differences between neighbouring SPH particles and converting the excess kinetic energy into thermal energy. Unfortunately, since viscosity is a dissipative process, the corresponding SPH equations cannot be derived from a variational principle, and we are back to ad-hoc methods for deriving them.

It should be noted that in this chapter we will not discuss *physical* viscosity. Physical viscosity is employed when the fluid modelled exhibits resistance to deformation by shear stresses and can be thought of as the “thickness” of a fluid. In a Newtonian fluid, the shear stress between layers is proportional to the velocity gradient in the direction perpendicular to the layers. For example, if one wanted to model honey with SPH, a physical viscosity would be required to take into account the shear viscosity between layers of the honey. In practice many SPH users abuse the SPH artificial viscosity in order to mimic a physical viscosity, usually by applying a non time varying fixed viscosity (see for example Murray (1998)). The aim of artificial viscosity is not to model this physical viscosity, rather is it present only to allow entropy generation at shocks and to introduce a smoothing at shock fronts in order to avoid discontinuities. The aim is to provide a sub-resolution scheme that mimics the effect of particle collisions on the scale of the mean free path. Therefore we should seek a scheme that only applies artificial viscosity at shock fronts. However, if one wanted to model physical viscosity in this way, an extra term can be added to the proposed artificial viscosity scheme to achieve this.

In the following sections we will outline the current methods used for SPH arti-

---

<sup>2</sup>Particle penetration occurs when particles of opposing fluid streams pass directly through each other.

ficial viscosity and discuss the problems with each scheme.

### 3.2.1 “Standard” SPH Viscosity

The traditional form of artificial viscosity (Monaghan, 1992) adds the following terms to the momentum and energy equations to convert the kinetic energy of a particle into heat energy:

$$\left(\frac{d\mathbf{v}_i}{dt}\right)_{\text{AV}} = - \sum_j m_j \Pi_{ij} \nabla_i \bar{W}_{ij}, \quad (3.1a)$$

$$\left(\frac{du_i}{dt}\right)_{\text{AV}} = \frac{1}{2} \sum_j m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \nabla_i \bar{W}_{ij}, \quad (3.1b)$$

where we use the averaged kernel  $\nabla_i \bar{W}_{ij}$  in order to conserve energy and momentum between particle pairs,

$$\nabla_i \bar{W}_{ij} = \frac{1}{2} (\nabla_i W(|\mathbf{x}_{ij}|, h_i) + \nabla_i W(|\mathbf{x}_{ij}|, h_j)). \quad (3.2)$$

The artificial viscosity term is given by

$$\Pi_{ij} = \begin{cases} \frac{-\alpha \bar{c}_{ij} \mu_{ij} + \beta \mu_{ij}^2}{\bar{\rho}_{ij}} & \text{for } \mathbf{v}_{ij} \cdot \mathbf{x}_{ij} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

and

$$\mu_{ij} = \frac{\bar{h}_{ij} \mathbf{v}_{ij} \cdot \mathbf{x}_{ij}}{\mathbf{x}_{ij}^2 + 0.01 \bar{h}_{ij}^2} \quad (3.4)$$

where we have used the notation  $a_{ij} = a_i - a_j$  and  $\bar{a}_{ij} = (a_i + a_j)/2$ . Since  $\Pi_{ij} = 0$  for receding particle pairs, artificial viscosity does not affect expanding flows. The functional form (3.3) of SPH artificial viscosity may seem rather ad-hoc, but amongst several methods tested this turned out to be the most useful (Gingold & Monaghan, 1983). Moreover, it is equivalent to the form of dissipation used implicitly in Riemann solvers (Monaghan, 1997), which are used in virtually all modern grid codes. Furthermore, Meglicki et al. (1993) showed that in the continuum limit (assuming both approaching and receding pairs), the linear term of the viscosity given by

equation (3.3) may be written as

$$F_i = \frac{\alpha h \kappa}{2\rho} [\partial_j (c\rho \partial_j v_i + c\rho \partial_i v_j + c\rho \delta_{ij} \partial_k v_k)], \quad (3.5)$$

where  $\kappa$  is of order unity but depends on the form of the kernel. The viscosity term in the Navier Stokes equation may be written as

$$F_i = \frac{1}{\rho} \partial_i \sigma_{ij} = \frac{1}{\rho} \partial_i \left[ \eta \left( \partial_i v_j + \partial_j v_i - \frac{2}{3} \delta_{ij} \partial_k v_k \right) + \xi \delta_{ij} \partial_k v_k \right]. \quad (3.6)$$

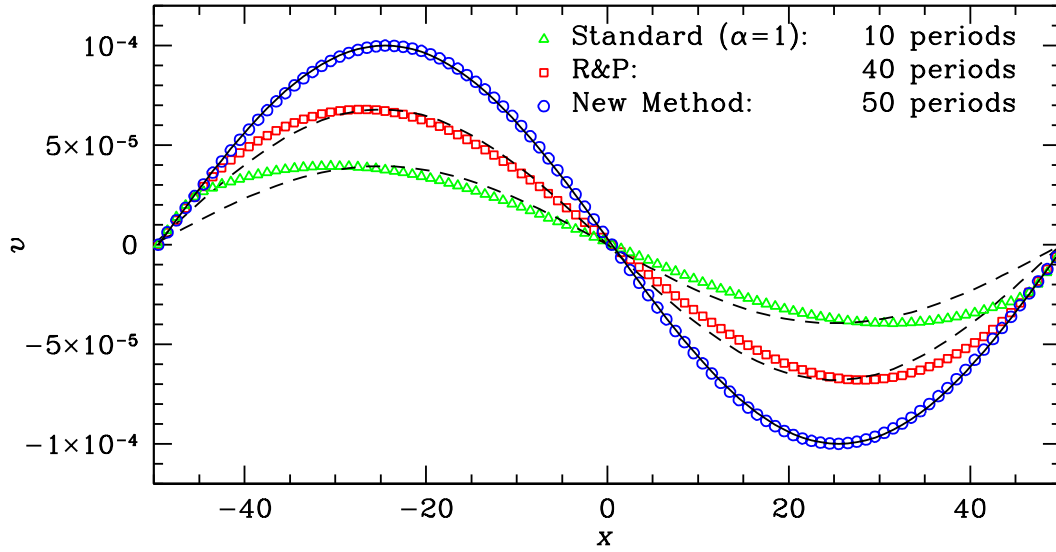
The first term gives the shear viscosity and the second term gives the bulk viscosity. By comparing the SPH form and the Navier Stokes equation we see that for SPH

$$\eta = \frac{1}{2} \alpha \kappa h c \rho, \quad (3.7a)$$

$$\xi = \frac{5}{3} \eta. \quad (3.7b)$$

Therefore the linear term in SPH viscosity gives both a shear and bulk viscosity and the form of artificial viscosity given above is the SPH version of the linear viscous stress tensor in the Navier Stokes equation. The quadratic term only becomes large when there are high Mach number shocks. Note that the artificial viscosity is resolution dependent as it depends on the smoothing length: the higher the resolution the smaller the effect. This is an advantageous property as a higher resolution will allow shocks to be resolved more sharply. This is exactly the desired behaviour as the aim of artificial viscosity is to smooth away “fluid discontinuities” (i.e. changes in the flow variables over the mean free path) that can not be resolved. Thus, a straightforward though expensive way to reduce unwanted dissipation in standard SPH simulations is to increase the resolution. In fact, one motivation for reducing artificial viscosity is to avoid this purely numerical necessity for high resolution and to choose the resolution on a more physical basis. Furthermore it is often impractical to do this as increasing resolution not only increases the computational cost per time-step, it also increases the number of time steps required (due to the Courant conditions, see equation 2.71a) and the amount of memory required to run and store the simulation.

There are several problems associated with this “always on” type of artificial viscosity described above. Firstly the parameter  $\alpha$  is fixed, and therefore viscosity is always switched on for approaching pairs, i.e. whenever there is converging flow.



**Figure 3.1:** A 1D sound wave with velocity amplitude  $v = 10^{-4}$  and sound speed  $c_s = 1$ . The “standard” SPH viscosity has quickly damped the amplitude to less than half its initial value after only 10 periods. The M&M method also damps the sound wave, but much slower. The new method, described in §3.4, preserves the amplitude of the sound wave by keeping the artificial viscosity low.

However, it is possible to have a converging flow without a shock, an example of which is a sound wave; some parts of the wave travel faster and therefore converge (see Fig. 3.1). This unwanted dissipation will almost certainly affect a simulation, potentially severely so. In some cases, such as the sound wave example, it may be possible to select the magnitude of the viscosity to minimise the adverse effects whilst still allowing the reproduction of shocks. In general however, the effect of artificial viscosity will be unknown prior to the simulation and, possibly, even afterwards. For example if one wanted to observe the effect of a perturbing massive body on a pulsating star, it may be very difficult to distinguish the effect of artificial viscosity from that of the perturbing body. It is known that artificial viscosity will act to damp any small scale oscillations and therefore it is impossible to reliably say anything about how the perturbing body effected the pulsating star. Also the user is required to select the strength of the artificial viscosity scheme before the simulation proceeds. Inexperienced users of SPH may select inappropriate values leading to spurious results (this of course could also happen with experienced users of SPH).

In most SPH applications to date, the above treatment with  $\alpha = 1$  is used. In particular, the widely used code **GADGET 2** (Springel, 2005) uses a fixed viscosity  $\alpha$  chosen at the start of the simulation (though Dolag et al., 2005, have implemented

into GADGET 2 the improved method described in §3.2.3 below). Clearly, in complex simulations, where strong and weak shocks are present as well as converging flows, any fixed choice for  $\alpha$  is unsatisfactory, leading either to poor treatment of strong shocks (e.g. particle penetration and noise), over-damping of converging flows, or both.

### 3.2.2 Balsara Switch

The purpose of artificial viscosity is to allow for entropy generation across shocks and to stop particle penetration. However, as Meglicki et al. (1993) demonstrated, the viscosity described above contains both a bulk viscosity and an unwanted shear term. This shear viscosity may seriously compromise simulations of shear flows, such as in a differentially rotating gas disc where it may cause spurious angular momentum transport. In order to try to reduce this effect, Balsara (1995) introduced a switch to reduce viscosity in the presence of shear flows. This was done by multiplying  $\Pi_{ij}$  with a reduction factor  $\bar{B}_{ij} = (B_i + B_j)/2$  with

$$B_i = \frac{|\nabla \cdot \mathbf{v}_i|}{|\nabla \cdot \mathbf{v}_i| + |\nabla \times \mathbf{v}_i|}. \quad (3.8)$$

However, whilst this works well for pure shear or pure bulk flows, a problem occurs when there is a mixture of shear and bulk flow. If the shear flow is high then the viscosity is reduced, even though there may still be a strong bulk flow. Furthermore, the viscosity is only reduced, it is not completely eliminated, therefore the unwanted effects caused by shear viscosity would still occur but on a longer timescale.

### 3.2.3 Morris & Monaghan Viscosity Switch

In order to provide a more general viscosity method that could be used in a wide variety of applications, *without* any “tweaking”<sup>3</sup>, Morris & Monaghan (1997) (M&M) suggested an adaptation to the standard SPH viscosity scheme. Each particle is afforded its own individual parameter  $\alpha_i$  which is adaptive. This is implemented in equation (3.3) by replacing  $\alpha$  with  $\bar{\alpha}_{ij} = (\alpha_i + \alpha_j)/2$  and setting  $\beta \propto \bar{\alpha}_{ij}$ . Each particle’s viscosity parameter is evolved individually according to the differential

---

<sup>3</sup>i.e. without having to change the parameters of the scheme for various situations

equation

$$\frac{d\alpha_i}{dt} = -\frac{\alpha_i - \alpha_{\min}}{\tau_i} + S_i. \quad (3.9)$$

The parameter  $\alpha_{\min}$  constitutes a lower limit for the artificial viscosity such that  $\alpha_i = \alpha_{\min}$  for non-convergent flows (i.e. vanishing source term). Morris & Monaghan suggest  $\alpha_{\min} = 0.1$  thus reducing the artificial viscosity away from shocks by an order of magnitude compared to standard SPH. For a convergent flow, on the other hand,  $\alpha_i$  grows above that value, guaranteeing the proper treatment of shocks. In the post-shock region, the flow is no longer convergent and  $\alpha_i$  decays back to  $\alpha_{\min}$  on the time scale  $\tau_i$  given by

$$\tau_i = \frac{\ell h_i}{c_i} \quad (3.10)$$

where  $c$  is the sound speed and  $h$  is the smoothing length. The parameter  $\ell$  controls the number of smoothing lengths that the viscosity decays over after the shock has passed. M&M suggest a value of  $\ell = 2.5 \rightarrow 5$  (note that they use a kernel of size  $2h$  whereas we use  $h$ ). They proposed to adapt the strength of artificial viscosity according to the local convergence, giving the source term as

$$S_i = \max\{-(\nabla \cdot v)_i, 0\}. \quad (3.11)$$

This source term allows the growth of the artificial viscosity whenever there is a converging flow (i.e. negative velocity divergence). For a constant source term the viscosity parameter will eventually reach an asymptotic value when the source term and decay term exactly cancel. Solving equation (3.9) for  $\dot{\alpha} = 0$  gives the asymptotic value as

$$\alpha_s = \alpha_{\min} + S_i \tau_i. \quad (3.12)$$

It is possible to solve equation (3.9) for  $\alpha$  directly by replacing the source term with the asymptotic value and integrating to obtain

$$\alpha_i \leftarrow \alpha_s + (\alpha_i - \alpha_s)e^{-\delta t/\tau_i}. \quad (3.13)$$

Although it is theoretically possible for the viscosity parameter to become arbitrarily large, in practice the viscosity parameter will never reach this asymptotic

value as the source term is usually sharply peaked. This means that large source terms are integrated for a short time, preventing arbitrarily large values of viscosity. Morris & Monaghan (1997) determined the maximum value of their viscosity switch at an infinite shock front as

$$\alpha_{\max} = \alpha_{\min} + \ln \frac{\gamma + 1}{\gamma - 1} \quad (3.14)$$

which is  $\approx 1.9$  for  $\gamma = 1.4$ .

More recently, Rosswog et al. (2000) proposed firstly to combine this method with the Balsara switch and secondly to alter the adaption equation (3.9) to<sup>4</sup>

$$\frac{d\alpha_i}{dt} = \frac{\alpha_{\min} - \alpha_i}{\tau_i} + S_i (\alpha_{\max} - \alpha_i) \quad (3.15)$$

with  $\alpha_{\max} = 1.5$ , while Price (2004) advocated  $\alpha_{\max} = 2$ . It is still possible to solve the differential equation using (3.13), however the asymptotic value changes to

$$\alpha_s = \frac{\alpha_{\min} + \alpha_{\max} S_i \tau_i}{1 + S_i \tau_i}. \quad (3.16)$$

The effect of this alteration is firstly to prevent  $\alpha_i$  exceeding  $\alpha_{\max}$  and secondly to increase  $\dot{\alpha}_i$  for small  $\alpha_i$ . However, this has an adverse effect when there are strong shocks, preventing the viscosity reaching the required value. As a simple example to see where this occurs, consider a shock where  $S\tau \gg 1$ . The Morris & Monaghan method would give a large  $\alpha_s$  whilst still giving  $\alpha \lesssim 2$ . However, the Rosswog et al. method gives  $\alpha_s \approx \alpha_{\max}$ . Thus, when this is integrated over the relatively short time that the source term is applied, the magnitude of the viscosity parameter will be considerably less. This can be observed from Fig. 3.29 which shows the two schemes for a  $\mathcal{M} = 50$  ram shock.

For the purposes of this study we refer to the M&M method as the differential equation given by (3.9), with asymptotic value (3.12) and parameters  $\alpha_{\min} = 0.1$ ,  $\beta = 2 \bar{\alpha}_{ij}$  and  $\ell = 5$ . We refer to the R&P (Rosswog/Price) method as the differential equation given by (3.15) with asymptotic value (3.16) and parameters  $\alpha_{\min} = 0.1$ ,  $\alpha_{\max} = 2$ ,  $\beta = 2 \bar{\alpha}_{ij}$  and  $\ell = 5$ . Despite the problems with strong shocks as pointed out above, the R&P method is the current state-of-the-art for SPH and is implemented in PHANTOM (by D. Price) and VINE (Wetzstein et al., 2008).

---

<sup>4</sup> This is equivalent to keeping (3.9) but multiplying the source term (3.11) by  $(\alpha_{\max} - \alpha)$ , which is what Rosswog et al. actually did.



### Problems with the M&M Scheme

Tests with the M&M switch have shown that the viscosity is increased by an appropriate amount to resolve shocks, giving accurate post and pre-shock solutions, smoothing the shock over a few particles. There is however room for improvement in this method. For modelling inviscid or low-viscosity fluids, as in most astrophysical applications, any finite lower viscosity limit is undesirable as it results in unwanted dissipation, for example the dampening of sound waves, as demonstrated in Fig. 3.10, or stellar pulsations (see §3.6.7). A minimum viscosity should not be necessary for the whole flow as viscosity is only required at shock fronts and in the post shock regions in order to suppress ringing. Morris & Monaghan (1997) suggested a value of  $\alpha_{\min} = 0.1$ , claiming this is required in order to ‘maintain order amongst the particles away from shocks’, though they did not demonstrate nor argue in any detail that this value was necessary or sufficient for this purpose. Later, Rosswog et al. used  $\alpha_{\min} = 0.05$ , though without exposing their method to a decent number of test problems. More recently, Price (2004) used  $\alpha_{\min} = 0.1$  and demonstrated the validity of this scheme with numerous tests. However, as far as we are aware, nobody has yet come up with a general method (i.e. valid under all possible conditions) with  $\alpha_{\min} = 0$ . Furthermore, as we demonstrate in sections 3.5 and 4.4.4, the damping of noise only requires a very small viscosity which is provided by the viscosity schemes to be considered.

Another noticeable problem with M&M’s method is the delay between the peak in the viscosity parameter  $\alpha_i$  and the shock front. This is demonstrated in Fig. 3.7 where the particles’ viscosities are still increasing when the shock arrives. One reason for this lag is that integrating the differential equation (3.9) increases  $\alpha$  until the asymptotic value (3.12) is reached or the source term is reduced. As a consequence, it takes several time steps for  $\alpha$  to reach the value required for resolving the shock, at which point the shock may already have passed through (this becomes worse when  $\alpha$  has to grow from  $\alpha_{\min} = 0$ ). A second reason for this lag is the source term (3.11) peaking at the exact shock position for a symmetrically smoothed shock. In practice, the situation is slightly better as it peaks one particle distance in front of the shock (Morris & Monaghan, 1997, see also Fig. 3.7). However, immediately behind the shock the (SPH model of the) flow is still converging and hence  $\alpha_i$  is still increasing towards the asymptotic value.

A further problem with the method of Morris & Monaghan is that it does not distinguish between convergent flows and shocks. Thus in converging flows, such

as the pulsating star, viscosity will be increased at the contraction phase of the pulsation, reducing the amplitude of the pulsations. Even if we set the floor value of the viscosity ( $\alpha_{\min}$ ) to zero, we would find that whenever there is converging flow the viscosity parameter would be increased inducing unwanted dissipation.

### 3.3 Towards an Improved Viscosity Trigger

In this section we present a set of preliminary source terms that were considered as possible candidates for an improved viscosity scheme. These toy switches enabled us to get a feel of the required properties for a successful artificial viscosity scheme. They helped rule out unsuccessful viscosity triggers and helped identify problems that may be encountered with the new scheme. These initial tests allowed us to investigate why particular methods were successful in some tests whilst failing in others. This steered us towards the final scheme which will be discussed in full in §3.4.

To attempt to improve the artificial viscosity algorithm we need to distinguish between converging flows and shocks. There are many potential methods of doing this and therefore we tested a number of alternatives to the Morris & Monaghan (1997) switch. We make two general alterations to the M&M method which are required for an improved viscosity scheme.

#### Minimum Viscosity

Firstly we require that SPH is inviscid away from shocks, and therefore we set  $\alpha_{\min} = 0$ . This enforces no artificial dissipation away from shocks allowing flows such as sound waves to remain undamped. This is a requirement of all the schemes we will discuss below. Enforcing this change can sometimes cause problems with ringing in the post shock region, caused by the viscosity decaying too quickly after the shock. Normally this is suppressed by employing a minimum viscosity which allows any post-shock oscillations to be damped away, helping the particles remain ordered. Therefore in order to compensate for having no minimum viscosity, we increase the decay time after the shock by setting  $\ell = 10$ . We demonstrate in sections 3.5 and 4.4.4 that a minimum viscosity is not required to suppress numerical noise, as suggested by Morris & Monaghan (1997).

### Adjusting the Particle's Viscosity

As discussed in the beginning of this section, increasing  $\alpha_i$  by integrating a differential equation (the solution to which is equation 3.13) is often too slow, in particular if  $\alpha$  has to grow from  $\alpha_{\min} = 0$ , as we intend to. Therefore, we instead set  $\alpha_i$  directly to an appropriate local<sup>5</sup> value  $\alpha_{\text{loc},i}$  whenever this exceeds the current value for  $\alpha_i$ . If the local value is lower than the current value of the viscosity parameter, which is the case in post shock regions,  $\alpha_i$  is reduced slowly via integration. This is done to avoid post shock ringing which can occur when reducing the viscosity too quickly in the post shock region. Allowing the viscosity to jump in this way ensures that the viscosity parameter has reached the value required for the local conditions, rather than the local conditions some time ago. To do this we redefine the differential equation for  $\alpha$  as

$$\frac{d\alpha}{dt} = \frac{\alpha_{\text{loc}} - \alpha}{\tau}. \quad (3.17)$$

The viscosity is therefore set using the rule

$$\alpha_i \leftarrow \begin{cases} \alpha_{\text{loc}} & \text{for } \alpha_{\text{loc}} > \alpha_i \\ \alpha_{\text{loc}} + (\alpha_i - \alpha_{\text{loc}})e^{-\delta t/\tau} & \text{for } \alpha_{\text{loc}} < \alpha_i. \end{cases} \quad (3.18)$$

The M&M scheme may be recovered by setting  $\alpha_{\text{loc}} \rightarrow S\tau$ . However, using an asymptotic value in this way (equation 3.18) allows for arbitrarily large viscosities as we jump directly to this value. This is avoided in a scheme that integrates the viscosity as the source term is large for only a short period of time. By allowing the viscosity to jump to the required value in this manner we must take extra care to only detect shock fronts. If the new switch incorrectly identifies a region of the flow as a shock (e.g. mistakes a rarefaction wave as a shock) the viscosity will be increased rapidly causing damping where it is not required. However if this is done with the M&M method the viscosity may not be increased by much as the source term is integrated over many time-steps. Therefore the effect on the simulation may not be as severe with the M&M method, although in practice we find the M&M switch does in certain situations introduce significant damping where it is not required.

The way of defining  $\alpha_{\text{loc}}$  therefore determines how the viscosity behaves. We test

---

<sup>5</sup>We define the asymptotic value as the value obtained by integrating equation 3.9 (or a similar equation) with constant source term. We define the local value as that used when setting  $\alpha = \alpha_{\text{loc}}$ .

a number of preliminary methods in order to understand what factors are important in constructing a viscosity scheme. In the following sections we focus our efforts on reducing viscosity away from shocks whilst still retaining the ability to capture weak shocks where only a small amount of artificial viscosity is required.

### 3.3.1 Preliminary Switch 1

We can define a shock as occurring whenever the velocity difference across a resolution element is compressive and exceeds the sound speed. We can determine when this happens by defining a quantity

$$\epsilon = -\frac{h \nabla \cdot \mathbf{v}}{c}. \quad (3.19)$$

In order to reduce viscosity in flows away from shocks, a simple solution is to multiply the normal M&M source term by a pre-factor which is small whenever  $\epsilon \ll 1$  and around unity when  $\epsilon \geq 1$

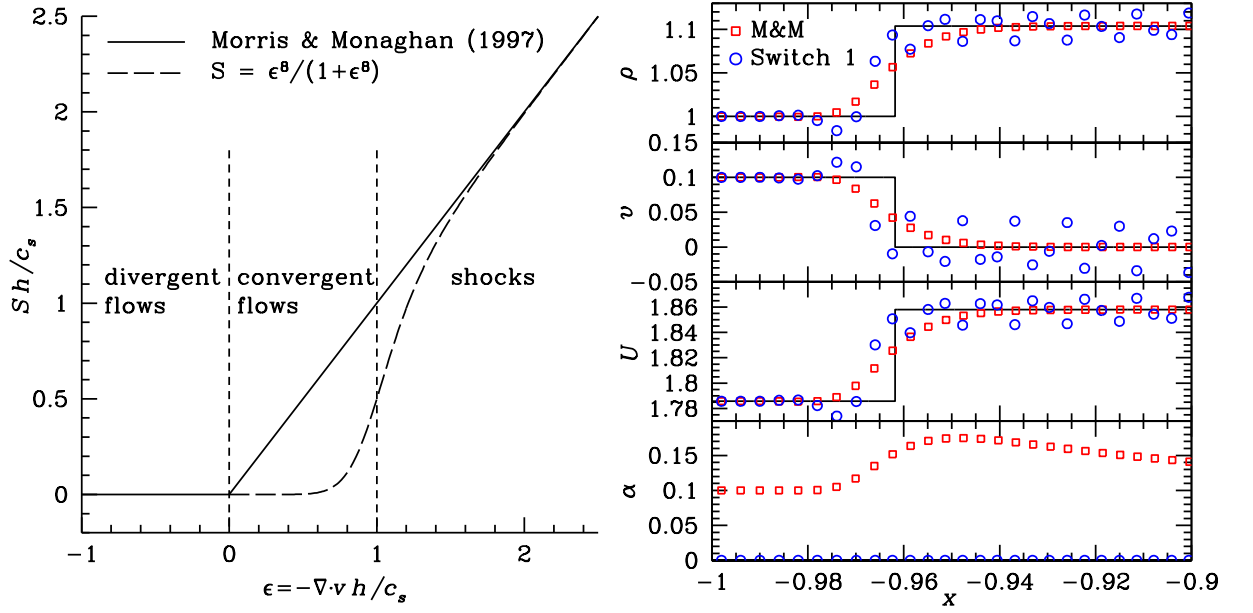
$$S = \frac{a\epsilon^b}{1 + \epsilon^b} [\max\{0, -\nabla \cdot \mathbf{v}\}]. \quad (3.20)$$

For the pre-factor we would typically<sup>6</sup> use  $a = 1$  and  $b = 8$ . The scheme can be seen in the left panel of Fig. 3.2. The source term remains low in mildly converging flows and allows a smoother transition from divergent to convergent flow. When a shock is detected the source term increases rapidly, allowing the shock to be captured. We define the local value of viscosity (i.e. that value of viscosity to be jumped to) as  $\alpha_{\text{loc}} = S\tau$ , which is the definition of the M&M asymptotic value. This is inadvisable in general as discussed earlier because it allows an arbitrarily large value for  $\alpha$ . However, as we only test this trigger in low viscosity regimes to expose its weaknesses, using  $\alpha_{\text{loc}} = S\tau$  is sufficient.

Whilst this method may work well for high Mach number shocks, the method can not detect discontinuities that are subsonic. These are shocks (discontinuities in pressure and velocity) where the difference in velocity across the shock is less than the sound speed and typically occur due to the steepening of sound waves (note that this occurs in the well known Sod (1978) shock tube test). A simple example to demonstrate this is the 1D subsonic ram test. The initial conditions

---

<sup>6</sup>This set of parameters allow the source term to increase sharply as  $\epsilon \rightarrow 1$ . However any value of  $b$  would reduce the viscosity in flows with  $\epsilon < 1$



**Figure 3.2:** *Left:* Morris & Monaghan source term and switch 1 source term (equation 3.20). The modified scheme will reduce viscosity whenever parameter  $\epsilon$  is small. **Right:** A comparison of the Morris & Monaghan switch and switch 1 (equation 3.20) for a subsonic ram shock. The modified source term keeps viscosity too low in this case and causes noise in the post-shock region.

consist of a constant pressure and density gas with the sound speed set to unity. A discontinuity in the velocity is applied such that  $v = 0.1$  for  $x < 0$  and  $v = -0.1$  for  $x > 0$ . As SPH is a smoothed scheme we provide a smoothing to the initial velocity discontinuity as suggested by Monaghan (1997)

$$v = \frac{v_L + v_R e^{x/d}}{1 + e^{x/d}} \quad (3.21)$$

where  $d$  is the largest initial particle separation. The initial velocity discontinuity results in two shocks which propagate in each direction symmetrically about the origin.

The right panel of Fig. 3.2 compares the results for this scheme and the M&M method. Whilst we found that the new scheme performed well in strong shocks and purely converging flows, the method was unsuccessful in the subsonic shock regime as demonstrated by this test. The scheme keeps the viscosity low as the divergence in the velocity is small compared to the sound speed, leading to noise in the post-shock region.

We may decrease the parameter  $b$  in order to allow a larger pre-factor for cases when  $\epsilon < 1$ . In this subsonic ram test we find that  $\epsilon \approx 0.15$  and therefore this pre-factor will always reduce the source term to around a tenth of that of the M&M switch (even with  $b = 1$ ). However, we find that the reduction in the local value compared to the asymptotic value is offset by allowing  $\alpha$  to jump directly. In practice the asymptotic value is never reached by integrating equation (3.9), whereas the local value is the actual value to be used. This allows the viscosity to quickly reach the required strength when using  $b = 1$ , allowing for good shock resolution. However, the method would still lead to a dampening of converging flows such as sound waves as the viscosity is increased before the flow becomes discontinuous. Also doing this for a supersonic shock ( $v = \pm 2$ ) causes too much viscosity to be applied as the asymptotic value jumped to is extremely high. This leads to noise in the region between the shocks. This occurs because there is no upper limit on the local value  $\alpha_{\text{loc}}$ , it can become arbitrarily large. A solution to this would be to use a local value defined in a similar way to the asymptotic value for the R&P method; that is equation (3.16). This would limit  $\alpha_{\text{loc}} \rightarrow \alpha_{\text{max}}$  in the limit of  $S \rightarrow \infty$ .

This method is therefore unsuccessful when there are subsonic discontinuities as they can not be differentiated from converging flows by this method.

### 3.3.2 Preliminary Switch 2

Switch 1 attempted to detect regions of the flow where the velocity difference between two particles exceeds the sound speed. However, this is not sufficient as it results in no dissipation being applied at subsonic shocks. As shocks can be viewed as unresolved discontinuities in the pressure and velocity, we attempt to detect unresolved pressure gradients. This is done by determining regions where the pressure gradient is large compared to the local pressure. If this is the case we have detected a region where smoothing needs to be applied. A possible pre-factor to determine this is

$$\epsilon_i = \frac{h_i |\nabla P_i|}{P_i}. \quad (3.22)$$

Therefore  $\epsilon$  will detect situations where the SPH estimate of the local pressure change (over a smoothing length) is large compared to the particle pressure. It is preferable to use pressure gradients to velocity gradients as pre-factors involving velocity are difficult to construct whilst still ensuring Galilean invariance. We also

found that the equivalent method for detecting velocity gradients (replacing  $\nabla P$  and  $P$  with  $\nabla \cdot v$  and  $|v|$ ) did not perform as well, the viscosity parameter tended to be slightly smaller leading to post-shock ringing.

The estimation of the pressure gradient can be done in two ways. Firstly the SPH gradient estimator can be used,

$$\nabla P = \sum_j m_j \frac{P_i - P_j}{\rho_j} \nabla W_{ij}(h_i) \quad (3.23)$$

or secondly the acceleration can be used to calculate the gradient in the pressure

$$|\nabla P| = \rho |a|. \quad (3.24)$$

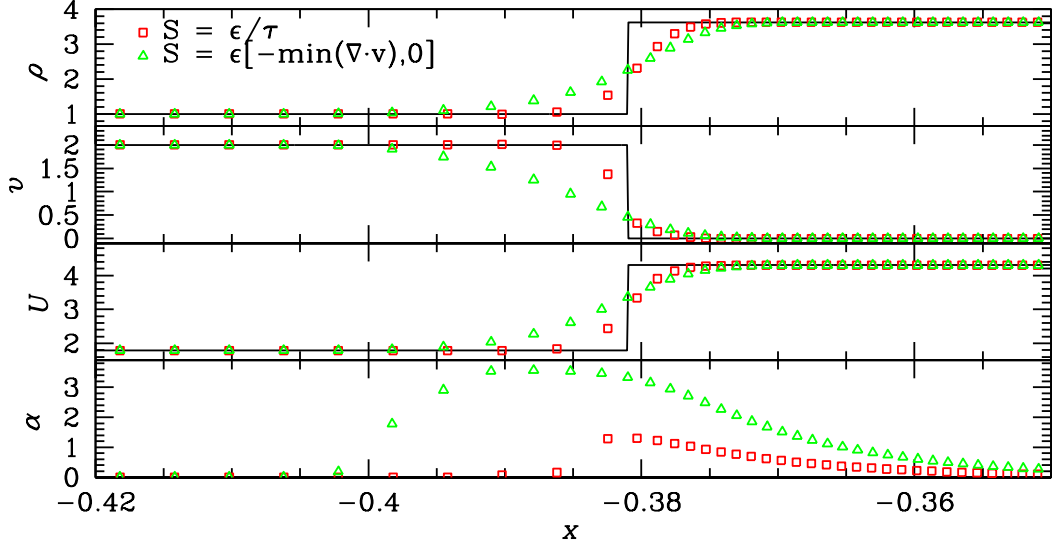
We prefer the use of equation (3.24) as we find the SPH pressure gradient is more susceptible to numerical noise. By using the acceleration we also remove the need for the storage and calculation of an extra quantity. Furthermore the particle velocities and positions are evolved according to the pressure gradient used in the momentum equation rather than using the SPH gradient estimator. Therefore in order to remain consistent we should use the pressure gradient from the acceleration. This gives the pre-factor as

$$\epsilon_i = \frac{h_i |a_{\text{SPH}}|}{P_i / \rho_i}. \quad (3.25)$$

As the particle velocities come directly from the SPH acceleration, using the acceleration allows us to detect where the particle velocities are about to become large compared to the local sound speed. As well as being able to detect strong shocks (i.e. where the velocity jump is large compared to the sound speed) the switch should also detect subsonic shocks as we do not use a functional form to restrict  $\epsilon$ , as was the case for switch 1. Furthermore, the value of  $\epsilon$  will map linearly to the shock strength, causing more damping in strong shocks as is required. However, this may cause problems for the strongest shocks as the local value will become far too high. The source term can then be constructed as

$$S = \epsilon [-\min\{\nabla \cdot v, 0\}]. \quad (3.26)$$

However, we find that this method tends to increase the viscosity to a much higher value than required, leading to an overly damped shock. The reason for this is not



**Figure 3.3:** *Effect of the source term on the ram shock with  $v = \pm 2$ . The source term (3.26) increases the viscosity too much causing an overly damped shock.*

only that  $\epsilon$  reaches a large value; it is also due to the large value of the velocity divergence which enhances the viscosity. We found a better method is to divide by the local dynamical time  $\tau$  such that,

$$S = \epsilon \frac{1}{\tau}, \quad (3.27)$$

essentially defining  $\alpha_{\text{loc}} = \epsilon$  (again setting  $\alpha_{\text{loc}} = S\tau$ ). This is very similar to the M&M source term which in effect takes the asymptotic value as the ratio of two time scales,  $\nabla \cdot \mathbf{v}$  and  $\tau = \ell h/c$ . In this method the local value ( $\alpha_{\text{loc}}$ ) is the ratio of the square of two velocity scales,  $h|\mathbf{a}_{\text{SPH}}|$  and  $P/\rho = c^2/\gamma$ .

The difference between these two source terms can be seen in Fig. 3.3 where we show the results of a ram test with  $v = \pm 2$ . The source term (3.26) increases the viscosity parameter to a higher value damping the shock over a much larger region than the source term given by equation (3.27). As we are jumping to a local value, rather than integrating towards an asymptotic value, we require the local value to be smaller than the M&M asymptotic value. In strong shocks, the source term given by (3.26) will cause the local value to become larger than the M&M asymptotic value, over-dampening the shock. Therefore the source term given by equation (3.27) is recommended. However, when using this source term we still need to use the velocity divergence to determine whether the flow is converging or diverging as

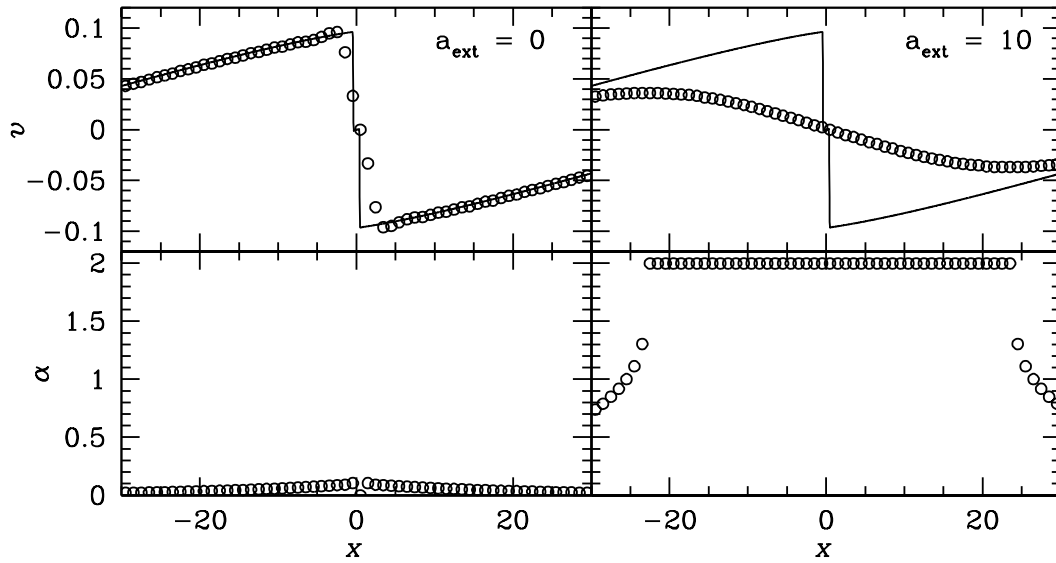


taking the pressure gradient does not distinguish between the two cases. We only apply viscosity when there is a converging flow, otherwise we set  $\epsilon = 0$  regardless of the size of the pressure gradient. It is possible for  $\epsilon$  to reach a relatively large value (such as  $\epsilon \gg 2$ ) in situations where there is a large jump in pressure (e.g. a  $1 : 10^{-4}$  jump). In some strong shock situations it is entirely possible for this value to be much larger, causing an extremely large viscosity parameter. Such a large value for viscosity is unnecessary, we find that a value of  $\alpha = 2$  is more than sufficient for all cases but the strongest shocks where the  $\beta$  term is required in any case. Therefore a modification is required for strong shocks in order to prevent excessive damping.

We find (from tests not presented here) that this method works well for pure hydrodynamical problems, shocks are well resolved over a few particles and the viscosity parameter is kept low away from shock regions. For problems involving other forces, such as gravity, the switch fails as the viscosity is increased to a much higher level than required. The method relies on the detection of sharp pressure gradients which produce a force accelerating particles towards each other, this is done by taking the absolute value of the SPH acceleration to obtain the pressure gradient. However, when self-gravity is involved it is possible to have gravity supported pressure gradients that are in *exact* equilibrium, producing no net force. This occurs in stars for example, the self gravity of the star is balanced by the internal pressure gradients within the star producing a system in (almost) exact equilibrium and therefore no need for artificial viscosity. However, the switch would detect a pressure gradient which could be large compared to the pressure, especially at the edge of the star, causing the increase in viscosity.

A simple modification to the viscosity switch would be to take the total acceleration to determine an effective pressure gradient instead of the SPH acceleration. By doing this we are effectively considering gravity (and any other external force) as an extra pressure gradient acting on the particles. Whilst this would work well for an isolated star, there are other situations where this would not work. The most general example where this would fail can easily be seen by considering the effect of imparting a constant external acceleration on a system of particles. An astrophysical application where this may occur is the motion of a star in a background potential, which may be caused by other SPH modelled stars, point masses or an analytical potential.

As a simple example we consider the steepening of a 1D sound wave with and without an external acceleration. When an external acceleration is applied a con-



**Figure 3.4:** The steepening of a 1D sound wave with the switch 2 using 100 SPH particles. An constant external acceleration is applied to the right hand plot causing an increase in the viscosity, thus damping the wave. Note that the velocities are corrected for the acceleration to allow ease of comparison. The results are compared to that of a with *Ramses* (Teyssier, 2002) using 4096 grid cells.

stant value of  $\mathbf{a}_{\text{ext}} = 10$  is added to all particle accelerations. Clearly the total force experienced by the particles is no longer the same as the SPH force. However, gas properties such as density and internal energy remain the same as the only physical quantities changed are position and velocity, both of which are changed by the same amount for all bodies. Clearly then the hydrodynamical results remain unchanged. In Fig. 3.4 the two SPH runs are compared to a high resolution run with *Ramses* (Teyssier, 2002) with no external acceleration using 4096 grid cells (compared to 100 SPH particles). When there is no external acceleration the method does a good job in keeping the viscosity low until the wave starts to steepen. When the wave steepens a strong pressure discontinuity is detected causing an increase in the viscosity parameter, allowing the shock to be resolved over around 5 SPH particles. When an external acceleration is applied the wave is severely damped due to the increase of viscosity. As the total acceleration is taken to compute the effective pressure gradient, the switch sees this external acceleration as a large pressure gradient and so increases the viscosity parameter.

Therefore the acceleration is clearly not a sufficient pre-factor in all cases. This is mainly because the acceleration is not a local quantity, unlike the velocity divergence, which would be unaffected in the examples described above.

### 3.3.3 Preliminary Switch 3

In order for the viscosity capturing method to be successful when there are external forces, the switch must be based on a local quantity which is both Galilean and gauge invariant. As the previous method of using the acceleration was successful, especially in at low Mach number flows, we consider a source term that uses the divergence of the acceleration.

The acceleration divergence is a local quantity which satisfies both Galilean and gauge invariance: applying a constant external force simply adds a constant value onto the particles acceleration. Therefore taking a source term based on acceleration divergence would produce the same results regardless of any *constant* external acceleration. Furthermore, suppose a point mass acts gravitationally on a polytrope. If the point mass is sufficiently far away then  $\nabla \cdot \mathbf{a}$  is unaffected as all particles in the polytrope receive roughly the same acceleration from the point mass. However, if the point mass is moved closer then the polytrope will be tidally stretched. At this point the acceleration divergence will be affected by the point mass and viscosity is introduced.

The divergence of the acceleration allows the detection of regions that have spatially varying accelerations. If the acceleration is spatially varying then clearly the relative particle velocities are changing, potentially resulting in a steepening of velocity gradients, eventually forming a shock. Typically,  $\nabla \cdot \mathbf{a} < 0$  and  $\nabla \cdot \mathbf{a} > 0$  immediately in front of and behind the shock respectively with  $\nabla \cdot \mathbf{a}$  averaging to zero across the shock due to Gauss' theorem<sup>7</sup>. This is advantageous as we require the particle's viscosity to be increased in advance so they are ready for the approaching shock. This is achieved by having a method that jumps directly to a local value, rather than integrating a differential equation, which may be too slow. However, in order for this method to be successful we must have a scheme which triggers at worst when the shock arrives. In this case the method triggers in a small region (of the order of the smoothing length) before the shock, guaranteeing the viscosity will "see" the shock coming.

Regions of the flow that are compressing immediately before the shock can be detected in a similar way to converging flows with the M&M source term. In analogy with with equation (3.11) we define

$$A = \max\{-\nabla \cdot \mathbf{a}, 0\} \tag{3.28}$$

---

<sup>7</sup>Gauss' theorem states that  $\int (\nabla \cdot \mathbf{F}) dV = \int \mathbf{F} \cdot \mathbf{n} dS$

There are a number of possible methods to construct a source term of units  $[\text{time}^{-1}]$  based on an acceleration divergence. To determine the ideal source term it is advisable to consider a number of simple yet common test cases (see table 3.1). After considering these simple test cases and experimenting with various methods we settled on

$$S_a = \frac{1}{\ell} \begin{cases} (h/c)fA & \text{for } fA < c^2/h^2, \\ \sqrt{fA} & \text{otherwise.} \end{cases} \quad (3.29)$$

where  $f$  is a factor of the order unity (we used  $f = 1.5$ ). The viscosity was set using equation (3.41) with  $\eta = 0.5$ . We found this method to be very successful in both weak and strong shocks and the method was invariant to external accelerations.

However we found that unwanted dissipation was introduced in situations where a net acceleration is balanced by rotation leading to a dynamic equilibrium, such as galactic discs. In idealised situations (which we consider here), the particles are accelerated but are in dynamic equilibrium so there is no change in density, no shocks and no particle penetration. Therefore there is no reason to apply an artificial dissipation. We can determine the functional form of  $\nabla \cdot \mathbf{a}$  in a general rotating disc<sup>8</sup>. The acceleration in a rotating disc is  $a_r = -r\omega^2$  and  $a_\theta = 0$ , therefore the acceleration divergence is

$$\nabla \cdot \mathbf{a} = \frac{1}{r} \frac{\partial}{\partial r}(-r^2\omega^2) + \frac{1}{r} \frac{\partial}{\partial \theta}(0) = -2\omega \left( \omega + r \frac{d\omega}{dr} \right) \quad (3.30)$$

which may be positive or negative and can vary through out the disc. In comparison the velocity divergence is exactly zero in this case as in a rotating disc  $v_r = 0$  and  $v_\theta = r\omega(r)$

$$\nabla \cdot \mathbf{v} = \frac{1}{r} \frac{\partial}{\partial r}(0) + \frac{1}{r} \frac{\partial}{\partial \theta}(r\omega) = 0. \quad (3.31)$$

### 3.4 The New Viscosity Method

Based on the results of the preliminary tests we may construct a new scheme which adjusts the SPH particles viscosity parameter such that viscosity is only applied at velocity discontinuities. As discussed previously we seek a scheme that allows an

---

<sup>8</sup>The divergence operator in cylindrical polar coordinates is  $\nabla \cdot \mathbf{F} = \frac{1}{r} \frac{\partial}{\partial r}(rF_r) + \frac{1}{r} \frac{\partial}{\partial \theta} F_\theta$

absence of dissipation away from shocks and can increase the viscosity parameter in the region before a shock. In this section we describe the shock indicator and how to adjust the particles viscosity based on this shock indicator. We then consider situations where our estimator may not be accurate and how we can overcome this problem. In sections 3.5, 3.6 and 3.7 we expose the new scheme to a variety of test problems.

### 3.4.1 Defining The Source Term

In light of the results of the previous section it is clear that we require a source term that behaves like  $\nabla \cdot \mathbf{a}$  in shocks and converging flows but behaves like  $\nabla \cdot \mathbf{v}$  in rotation. A source based on  $\nabla \cdot \mathbf{v}$  is physically sensible as a velocity divergence indicates a change in density, and therefore a possibility that viscosity is required. However a higher derivative is also useful as it allows the shock to be detected earlier and is better able to distinguish pre and post shock regions from converging flows and sound waves. We can in principle swap the spatial and temporal derivatives taken on the velocity so that  $\nabla \cdot \mathbf{a} \rightarrow D_t(\nabla \cdot \mathbf{v})$  where we denote the time derivative operator  $D_t = d/dt$ . We can show how  $D_t(\nabla \cdot \mathbf{v})$ ,  $\nabla \cdot \mathbf{v}$  and  $\nabla \cdot \mathbf{a}$  are related by noting that<sup>9</sup>

$$\frac{D}{Dt} \nabla \mathbf{v} = v^{\alpha, \beta t} + v^\gamma v^{\alpha, \beta \gamma} = \dot{\underline{\mathbf{V}}}^{\alpha \beta}. \quad (3.32)$$

As the acceleration can be written as

$$a^\alpha = \frac{D}{Dt} v^\alpha = v^{\alpha, t} + v^\gamma v^{\alpha, \gamma}, \quad (3.33)$$

taking the spatial derivative of this gives

$$\frac{\partial a^\alpha}{\partial x^\beta} = v^{\alpha, \beta t} + v^{\gamma, \beta} v^{\alpha, \gamma} + v^\gamma v^{\alpha, \gamma \beta} = \underline{\mathbf{A}}^{\alpha \beta}. \quad (3.34)$$

We may use equation 3.32 to rewrite the expression for  $\partial a^\alpha / \partial x^\beta$  as

$$\frac{\partial a^\alpha}{\partial x^\beta} = \frac{D}{Dt} \nabla \mathbf{v} + v^{\gamma, \beta} v^{\alpha, \gamma}. \quad (3.35)$$

---

<sup>9</sup>where we have used the convective derivative  $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \nabla$

As  $v^{\alpha,\beta} = \nabla \mathbf{v} = \underline{\mathbf{V}}^{\alpha\beta}$  by definition we may write

$$\frac{D}{Dt} \nabla \mathbf{v} = \nabla \mathbf{a} - (\nabla \mathbf{v})^2. \quad (3.36)$$

By taking the trace of this matrix equation we notice that the difference between  $D_t(\nabla \cdot \mathbf{v})$  and  $\nabla \cdot \mathbf{a}$  is the trace of  $(\nabla \mathbf{v})^2$ . The advantage of swapping the order of the derivatives in this way is that in cases where the density is not changing (and so the velocity divergence is zero) there is no viscosity trigger. This is not the case if using  $\nabla \cdot \mathbf{a}$ , the density may be constant but  $\nabla \cdot \mathbf{a}$  is not necessarily zero, as we demonstrated via equation (3.30). Furthermore, from differentiating the continuity equation we obtain

$$D_t(\nabla \cdot \mathbf{v}) = -\frac{d^2 \ln \rho}{dt^2}. \quad (3.37)$$

$D_t(\nabla \cdot \mathbf{v}) < 0$  indicates a non-linear density increase and a steepening of the flow, as is typical for any pre-shock region. Conversely, in the post-shock region  $D_t(\nabla \cdot \mathbf{v}) > 0$ . This suggests that we should take only negative values into account for any artificial viscosity switch. In analogy with equation (3.11) we may construct the source term as

$$S = \frac{h}{v_{\text{sig}}} \max \{-D_t(\nabla \cdot \mathbf{v}), 0\} \quad (3.38)$$

where

$$v_{\text{sig}} = \max_j \{\bar{c}_{ij} - \min \{0, \mathbf{v}_{ij} \cdot \hat{\mathbf{x}}_{ij}\}\}. \quad (3.39)$$

The signal velocity<sup>10</sup> was used rather than the sound speed as we found better results in low sound speed regimes where there was noise in the velocity divergence. Furthermore we find that using the signal velocity removes the need for the more complicated functional form of the source term given by equation (3.29) for the strong shock regime. This is because both the signal velocity and  $D_t(\nabla \cdot \mathbf{v})$  become large in strong shocks as particles rapidly approach each other. This helps reduce the source term slightly in strong shock regimes where  $D_t(\nabla \cdot \mathbf{v})$  is very high (see also table 4.2).

---

<sup>10</sup>In order to remain consistent we also redefine  $\tau = \ell h / v_{\text{sig}}$ .

### 3.4.2 Adjusting the Viscosity

As there is no upper limit on the size of a source term, a local value based on  $\alpha_{\text{loc}} = S\tau$  would allow arbitrarily large values of  $\alpha$  to be obtained. This was observed with tests using the preliminary switches. Instead we use a method motivated by the method of Rosswog et al. (2000) by defining a more general differential equation for  $\alpha$ ,

$$\frac{d\alpha}{dt} = -\frac{\alpha}{\tau} + S(\alpha_{\text{max}}^\eta - \alpha_i^\eta)^{1/\eta}. \quad (3.40)$$

We determine a local value from this by assuming the equation is integrated with a constant source term until the asymptotic value is reached. In this case the asymptotic value is given by

$$\alpha_s = \alpha_{\text{max}} \frac{S\tau}{[1 + (S\tau)^\eta]^{1/\eta}}. \quad (3.41)$$

Of course in the new scheme we just jump directly to this local value (where we use  $\eta = 1$ )

$$\alpha_{\text{loc}} = \alpha_{\text{max}} \frac{\xi}{1 + \xi}. \quad (3.42)$$

This method of limiting  $\alpha$  to  $\alpha_{\text{max}}$  was required, as our experiments showed, to reflect the difference between an asymptotic value and the value actually used. The R&P method used  $\xi = S\tau$ , however we find that to account for the difference between integrating and jumping, it was appropriate remove the constant  $\ell$  from  $\tau$ ; that is set  $\xi = S\tau/\ell = Sh/v_{\text{sig}}$ . This means that for small values of  $S$ ,  $\alpha_{\text{loc}}$  is smaller than the asymptotic value by a factor of  $\ell$ . This reduction allows for the fact that the asymptotic value is never reached in practice (but only in the limit of late times and constant source term), whilst  $\alpha_{\text{loc}}$  is the actual value to be used. For large  $S$  the rate at which  $\alpha_{\text{loc}}$  approaches the limit  $\alpha_{\text{loc}} \rightarrow \alpha_{\text{max}}$  depends on the value of  $\eta$ ; we found  $\eta = 1$  allowed sufficient growth of  $\alpha$  with large source terms. After passing through a shock,  $\alpha_{\text{loc}}$  will quickly return to zero whenever  $\alpha_i > \alpha_{\text{loc}}$ . In this decay we use  $\ell = 10$  in equation (3.10) for  $\tau$ . This implies that the viscosity decays twice as slowly as in previous methods and avoids some occasional minor post-shock ringing not present in methods with  $\alpha_{\text{min}} > 0$ . However, the traditional  $\ell = 5$  also gives satisfactory results for most problems.

### 3.4.3 Estimating $D_t(\nabla \cdot \mathbf{v})$

There are four slightly different ways to estimate  $D_t(\nabla \cdot \mathbf{v})$ . The first is to simply use the rate of change of  $\nabla \cdot \mathbf{v}$  found by computing the change in  $\nabla \cdot \mathbf{v}$  over a time step,

$$D_t(\nabla \cdot \mathbf{v}) = \frac{\delta \nabla \cdot \mathbf{v}}{\delta t}, \quad (3.43)$$

where  $\nabla \cdot \mathbf{v}$  is estimated from equation (2.28). Whilst this is a crude estimate we find that in most situations it is accurate enough to provide a useful viscosity source term. The reason for this is that the viscosity parameter  $\alpha$  does not usually need to be too accurate, an error of  $\pm 10\%$  from the optimal value will in most cases suffice. However, in some SPH simulations, where formally both the velocity divergence and  $D_t(\nabla \cdot \mathbf{v})$  should equal zero we find that the SPH estimate of both  $\nabla \cdot \mathbf{v}$  and  $D_t(\nabla \cdot \mathbf{v})$  is non-zero. We find this particularly troublesome in shearing flows such as 2D Keplerian discs (see section 3.6.3). To investigate the cause of this we firstly consider the SPH estimate for  $\nabla \cdot \mathbf{v}$  which can be written as follows (using equations 2.18, 2.21, 2.22 and that  $D_t(h^\nu \rho) = 0$ )

$$\widehat{\nabla \cdot \mathbf{v}} = -\frac{\dot{\rho}}{\rho} = \nu \frac{\dot{h}}{h} = \nu \frac{\sum_j m_j \mathbf{v}_{ij} \cdot \mathbf{x}_{ij} \tilde{w}_{ij}}{\sum_j m_j x_{ij}^2 \tilde{w}_{ij}}. \quad (3.44)$$

Although this estimate for the velocity divergence may be satisfactory for the SPH density estimation method, it is not necessarily an accurate estimate of the velocity divergence of the underlying fluid. To illustrate this we may consider the matrix

$$\underline{\mathbf{D}}_i^{\alpha\beta} = \sum_j \mathbf{v}_{ij}^\alpha \mathbf{x}_{ij}^\beta e_{ij}, \quad (3.45)$$

where  $e_{ij}$  is a weighting factor and the trace of  $\underline{\mathbf{D}}$  corresponds to the numerator in equation (3.44). Taking a Taylor expansion of  $\mathbf{v}_j$  around  $\mathbf{x}_i$  gives

$$\mathbf{v}_j^\alpha \approx \mathbf{v}_i^\alpha + (\mathbf{x}_j - \mathbf{x}_i) \underline{\mathbf{V}}_i^{\alpha\beta} \quad (3.46)$$

where  $\underline{\mathbf{V}}_i^{\alpha\beta} = \nabla \mathbf{v}$ . Substituting this into equation (3.45) we obtain

$$\underline{\mathbf{D}}_i^{\alpha\beta} = \sum_j \mathbf{x}_{ij}^\alpha \underline{\mathbf{V}}_i^{\alpha\beta} \mathbf{x}_{ij}^\beta e_{ij} = \underline{\mathbf{V}}_i^{\alpha\beta} \sum_j \mathbf{x}_{ij}^\alpha \mathbf{x}_{ij}^\beta e_{ij} = \underline{\mathbf{V}}_i^{\alpha\beta} \underline{\mathbf{T}}_i^{\alpha\beta} \quad (3.47)$$



where

$$\underline{\mathbf{T}}_i^{\alpha\beta} = \sum_j \mathbf{x}_{ij}^\alpha \mathbf{x}_{ij}^\beta e_{ij}. \quad (3.48)$$

The trace of the matrix  $\underline{\mathbf{V}}_i^{\alpha\beta}$  gives the divergence of the velocity. In order to try to reduce the noise cause by the anisotropy in the particle distribution we may use

$$\nabla \cdot \mathbf{v}_i = \text{tr}(\underline{\mathbf{T}}_i^{-1} \underline{\mathbf{D}}_i) \quad (3.49)$$

as an improved estimate for  $\nabla \cdot \mathbf{v}$ . Comparing (3.45) and (3.48) to (3.44), we see that the simple estimator corresponds to  $\widehat{\nabla \cdot \mathbf{v}} = \nu \text{tr}(\underline{\mathbf{D}})/\text{tr}(\underline{\mathbf{T}})$  with the weights<sup>11</sup>  $e_{ij} = m_j \tilde{w}(r_{ij})$ . If we split  $\underline{\mathbf{V}}$  into its isotropic part (divergence), the symmetric traceless part  $\underline{\mathbf{S}}$  (shear), and the antisymmetric part  $\underline{\mathbf{R}}$  (vorticity),

$$\underline{\mathbf{V}} = \nu^{-1} \nabla \cdot \mathbf{v} \underline{\mathbf{I}} + \underline{\mathbf{S}} + \underline{\mathbf{R}} \quad (3.50)$$

and insert it into equation (3.44), we find for the simple estimator

$$\widehat{\nabla \cdot \mathbf{v}} = \nabla \cdot \mathbf{v} + \nu \text{tr}(\underline{\mathbf{S}} \tilde{\underline{\mathbf{T}}})/\text{tr}(\underline{\mathbf{T}}) \quad (3.51)$$

where  $\tilde{\underline{\mathbf{T}}}$  denotes the anisotropic (traceless) part<sup>12</sup> of  $\underline{\mathbf{T}}$ . Owing to the symmetry of  $\underline{\mathbf{T}}$  and that the matrix  $\underline{\mathbf{R}}$  contains zeros on the diagonals, the vorticity is harmless, producing no contribution. Thus, the simple estimator (3.44) contains an  $\mathcal{O}(h^0)$  error term, which originates from anisotropy of  $\underline{\mathbf{T}}$  in conjunction with velocity shear. For perfectly symmetric particle distributions  $\tilde{\underline{\mathbf{T}}} = 0$ , but in general  $\tilde{\underline{\mathbf{T}}} \neq 0$  such that in the presence of strong shear even a small residual  $\tilde{\underline{\mathbf{T}}}$  results in a failure of the simple estimator (3.44).

This typically happens in differentially rotating discs, where the velocity field is divergent-free but contains shear and even in the absence of noise  $\tilde{\underline{\mathbf{T}}} \neq 0$  owing to the shearing particle distribution. It should emphasised that this problem is *not* caused by particle noise, the particles may be evenly spaced and the problem will still occur.

Alternatively one may try to estimate  $D_t(\nabla \cdot \mathbf{v})$  from information from a single

<sup>11</sup>For the weighting factors we may use  $e_{ij} = m_j w_{ij}$ ,  $m_j w_{ij}/\rho_j$ ,  $m_j \tilde{w}_{ij}$  or  $m_j \tilde{w}_{ij}/\rho_j$ . Whilst we found that there was little difference between any of the methods,  $m_j \tilde{w}_{ij}$  was slightly more accurate.

<sup>12</sup>One can show that as  $\underline{\mathbf{S}}$  and  $\underline{\mathbf{T}}$  are symmetric  $\text{tr}(\underline{\mathbf{S}} \underline{\mathbf{T}}) = \text{tr}(\underline{\mathbf{S}} \tilde{\underline{\mathbf{T}}})$

time step. We therefore may estimate

$$D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\dot{\underline{\mathbf{V}}}) = \text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2). \quad (3.52)$$

The matrix  $\underline{\mathbf{A}}$  (see equation 3.34) may be obtained in a similar way to the matrix  $\underline{\mathbf{V}}^{\alpha\beta}$ , by firstly obtaining

$$\underline{\mathbf{J}}_i^{\alpha\beta} = \sum_j \mathbf{a}_{ij}^\alpha \mathbf{x}_{ij}^\beta e_{ij}, \quad (3.53)$$

and then inverting the matrix equation

$$\underline{\mathbf{J}}_i^{\alpha\beta} = \underline{\mathbf{A}}_i^{\alpha,\beta} \underline{\mathbf{T}}_i^{\alpha\beta}. \quad (3.54)$$

Finally, one can estimate  $D_t(\nabla \cdot \mathbf{v})$  using the acceleration divergence rather than the trace of the matrix  $\underline{\mathbf{A}}$ ,

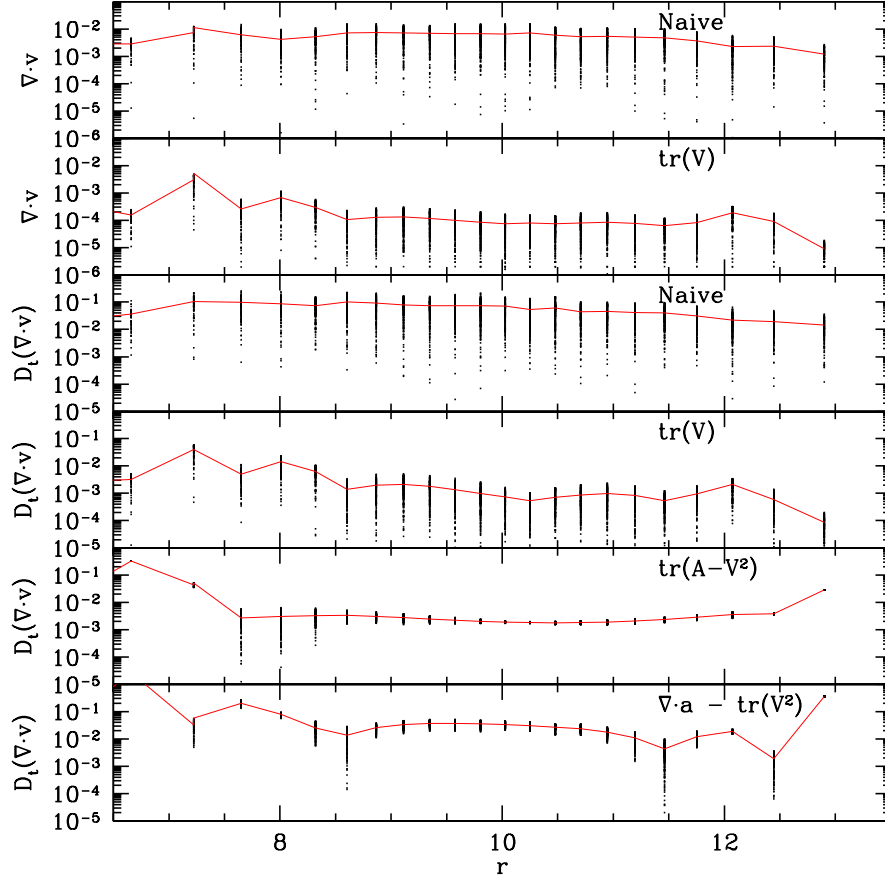
$$D_t(\nabla \cdot \mathbf{v}) = \nabla \cdot \mathbf{a} - \text{tr}(\underline{\mathbf{V}}^2). \quad (3.55)$$

However, this may be more susceptible to particle anisotropy as the acceleration divergence has not been corrected.

We can compare the various estimators for the 2D Keplerian disc test. We use a sound speed of  $c_s = 0.01$ . In order to remove the destabilising effects caused by artificial viscosity we run the simulation with  $\alpha = 0$  for all particles at all times (see section 3.6.3 for more details). Fig. 3.5 shows the accuracy of the various estimates. We see that the estimate using the naive SPH  $\nabla \cdot \mathbf{v}$  is the most susceptible to the anisotropy in the particle distribution. The azimuthally averaged values for  $D_t(\nabla \cdot \mathbf{v})$  were around an order of magnitude higher when using the naive estimate of  $\nabla \cdot \mathbf{v}$  rather than the improved estimate for  $\nabla \cdot \mathbf{v}$  and  $\text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2)$ . The method given by equation (3.55) is (as expected) less accurate because  $\nabla \cdot \mathbf{a}$  has not been treated for the errors due to the particle anisotropies. We find that the best method is that given by equation (3.52) as this not only substantially reduces the azimuthally averaged values for  $D_t(\nabla \cdot \mathbf{v})$  is also reduces the spread in  $D_t(\nabla \cdot \mathbf{v})$  at each radii.

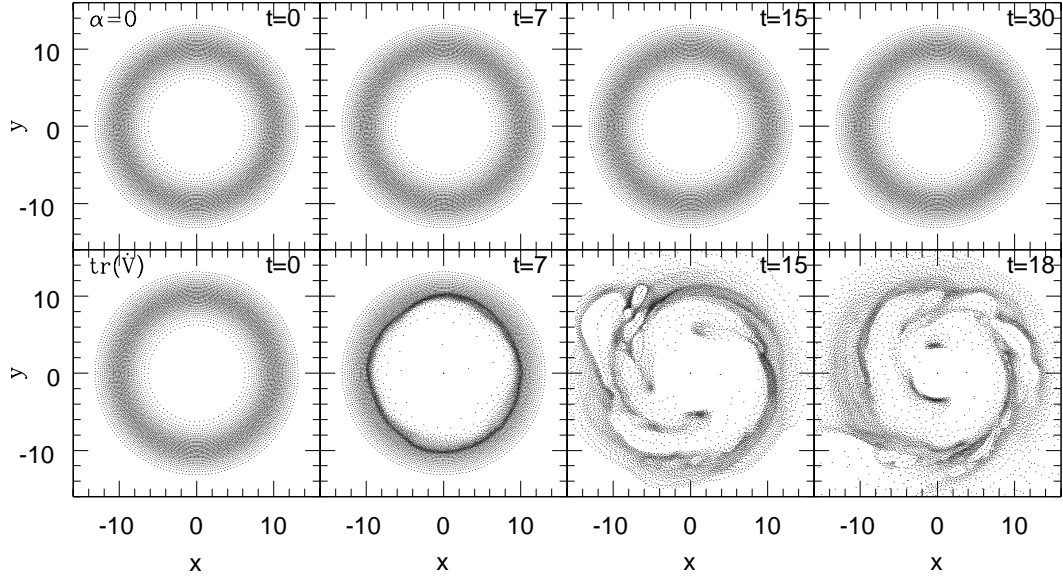
### 3.4.4 Limiting The Viscosity

We have managed to substantially reduce the error in the estimate of both  $\nabla \cdot \mathbf{v}$  and  $D_t(\nabla \cdot \mathbf{v})$  using an improved estimate for  $\nabla \cdot \mathbf{v}$ . However, we still see that in



**Figure 3.5:** Accuracy of the various estimates for  $\nabla \cdot \mathbf{v}$  and  $D_t(\nabla \cdot \mathbf{v})$  for the Keplerian ring test with  $c_s = 0.01$ . The black points indicate values for each particle. The red lines show the azimuthally averaged RMS values.

some cases, errors in the estimate of  $D_t(\nabla \cdot \mathbf{v})$  can still cause a substantial amount of artificial viscosity. Fig. 3.6 illustrates this by looking at the effect of viscosity on a Keplerian disc. With no artificial viscosity the disc remains stable. However, when the artificial viscosity scheme based on  $\text{tr}(\mathbf{V})$  is used the disc quickly becomes disrupted. This is because the estimator for  $D_t(\nabla \cdot \mathbf{v})$  still contains noise due to anisotropies in the particle distribution. The correction outlined in the last section is only a 1<sup>st</sup> order correction, clearly in some cases higher order terms can contribute, causing an error in the estimate. As the sound speed is low the numerical noise still appears quite large in the context of the simulation. This leads to a relatively large viscosity of around  $\alpha = 0.5$ . To reduce this noise we can use a limiter whereby we reduce the source term if there is strong shear and if we do not detect shocks. To



**Figure 3.6:** *Keplerian disc test with period of  $P = 2\pi$  at the density peak. **Top** shows the evolution of the disc with no viscosity, **bottom** shows the disc using the source term given by (3.52)*

do this we apply a limiter  $\kappa$  to the source term,

$$S = \kappa \frac{h}{v_{\text{sig}}} \max\{-D_t(\nabla \cdot \mathbf{v}), 0\}. \quad (3.56)$$

A good indicator of a shock is the quantity

$$R_i = \frac{\sum_j \text{sign}(\nabla \cdot \mathbf{v})_j w_{ij}}{\sum_j w_{ij}} \quad (3.57)$$

where  $R = -1$  in a shock as the flow around a shock is converging. In flows where noise dominates, we expect that  $\nabla \cdot \mathbf{v}$  will approximately average to zero and therefore  $R \rightarrow 0$ . Whilst this tells us if a shock is occurring it does not tell us anything about the strength of the shock.

The limiter can be constructed by careful consideration of how the viscosity should behave in situations of shear flows and shocks. If we have only shocks and no shear then the limiter should be  $\kappa = 1$  as we do not wish to reduce viscosity. We also require artificial viscosity whenever there is noise in the velocity field as this prevents disorder in the particle distribution. If there is noise in the velocity field then both  $\nabla \cdot \mathbf{v}$  and the shear component of  $\nabla \mathbf{v}$  will be of the same order. If there is a large amount of shear and no shock, such as the Keplerian disc test, we need

to reduce artificial viscosity as the anisotropy in the particle distribution combined with the shear causes noise in the estimate of  $D_t(\nabla \cdot \mathbf{v})$ . Finally as we shall see in section 3.6.6, if we have a shock and large shear we need to reduce the viscosity slightly. This is because anisotropy in the particle distribution combined with the shear causes an overestimate of  $D_t(\nabla \cdot \mathbf{v})$ , which increases the viscosity too much.

A simple method of doing this has already been proposed by Balsara (1995) where the viscosity is limited if the divergence in the velocity field is much smaller than the curl in the velocity field. This was done to eliminate this exact problem, namely that in shear flows, anisotropies in the particle positions cause a non-zero velocity divergence. In this case we would not use the curl we would use the shear component of  $\nabla \mathbf{v}$ .

However, we find that the Balsara limiter is generally not strong enough to limit the viscosity in shear flows. A simple modification to the Balsara limiter would be to square each term such that  $\kappa = |\nabla \cdot \mathbf{v}|^2 / (|\nabla \cdot \mathbf{v}|^2 + |\nabla \times \mathbf{v}|^2)$ . However, with this we find that in situations where there are both shocks and shear flows (see section 3.6.6), the viscosity can be over limited. Therefore we can include the shock indicator 3.57 to boost the contribution from  $\nabla \cdot \mathbf{v}$  in shocks,

$$\kappa = \frac{[a(1-R)^n |\nabla \cdot \mathbf{v}|]^k}{[a(1-R)^n |\nabla \cdot \mathbf{v}|]^k + |\mathbf{V}_s|^k}. \quad (3.58)$$

This leads to a only a slight reduction in viscosity in the presence of strong shocks and shear flows. Furthermore in situations where  $\nabla \cdot \mathbf{v}$  is small due to noise and the shear is large, the limiter is small leading to a reduction in the viscosity. We found though careful experimentation that using  $a = 2$ ,  $n = 4$  and  $k = 2$  gave good results in a variety of situations.

To compute the shear in any matrix  $\underline{\mathbf{M}}$  we firstly obtain the traceless part of the matrix as

$$\hat{\underline{\mathbf{M}}} = \underline{\mathbf{M}} - \frac{\text{tr}(\underline{\mathbf{M}})}{\nu}. \quad (3.59)$$

The traceless part may then be split into a symmetric component, which represents the shear. This may be obtained as

$$\underline{\mathbf{M}}_s = \frac{1}{2} (\hat{\underline{\mathbf{M}}} + \hat{\underline{\mathbf{M}}}^T). \quad (3.60)$$

We may then take the absolute values of these matrices as an indicator of the noise.

The absolute value of the matrix may be obtained by summing the square of each component, thus

$$|\underline{\mathbf{M}}| = \sqrt{\sum_{\alpha\beta} M^{\alpha\beta} M^{\alpha\beta}}. \quad (3.61)$$

### 3.4.5 Viscosity Algorithms

We compute the shear limited  $\alpha_{\text{loc}}$  according to the following pseudo-code

---

**Algorithm 1** Compute  $\alpha_{\text{loc}}$ 


---

$\xi = \left(\frac{h}{v_{\text{sig}}}\right)^2 \max(-D_t(\nabla \cdot \mathbf{v}), 0)$	// Compute factor required for $\alpha_{\text{loc}}$
$R = \sum_j \text{sign}(\nabla \cdot \mathbf{v})_j w_{ij} / \sum_j w_{ij}$	// Compute shock indicator
$X = 2(1 - R)^4  \nabla \cdot \mathbf{v} $	
$\kappa = X^2 / (X^2 +  \underline{\mathbf{V}}_s ^2)$	// Compute limiter factor
$\xi = \xi \times \kappa$	// Reduce factor for $\alpha_{\text{loc}}$ .
$\alpha_{\text{loc}} = \alpha_{\text{max}} \xi / (1 + \xi)$	// Compute $\alpha_{\text{loc}}$ using (3.42 )

---

This method works by substantially reducing the viscosity if the shear is significantly larger than  $\nabla \cdot \mathbf{v}$  and no shock detected. If a shock is detected, the viscosity is only slightly reduced in the presence of a strong shear flow. In the case of noise in the velocity field,  $\nabla \cdot \mathbf{v}$  and the shear are of the same order and therefore the viscosity is able to damp away this noise. Finally, if there is no shear, the limiter has no affect on the viscosity source term. The exact form of this limiter may be improved upon but it does give a good reduction of viscosity in shear flows (see section 3.6.3). The viscosity is adjusted according to the algorithm below.

---

**Algorithm 2** Adjust the Viscosity

---

$\tau = \ell h / v_{\text{sig}}$	// Compute $\tau$
<b>if</b> $\alpha < \alpha_{\text{loc}}$ <b>then</b>	
$\alpha = \alpha_{\text{loc}}$	// Jump to $\alpha_{\text{loc}}$
<b>else</b>	
$\alpha = \alpha_{\text{loc}} + (\alpha - \alpha_{\text{loc}}) \exp(-\delta t / \tau)$	// Decay down towards $\alpha_{\text{loc}}$
<b>end if</b>	

---

Situation	Example	AV req.	$\nabla \cdot \mathbf{v}$	$\nabla \cdot \mathbf{a}$	$D_t(\nabla \cdot \mathbf{v})$	$S_{D_t(\nabla \cdot \mathbf{v})}$
Static Equilib.	Polytrope	No	0	0	0	0
Dynamic Equilib.	Rotating Disc	No	0	$-2\omega(\omega + r\omega')$	0	0
Weakly Converging flow	Homologous flow	No	$-3b$	0	0	0
	sound wave	No	$kv_a$	$k^2cv_a$	$k^2cv_a$	$c/v_{\text{sig}}hk^2v_a$
Velocity Discont.	Strong Shock	Yes	$\sim (\delta v/h)$	$\sim f(\delta v/h)^2$	$\sim (\delta v/h)^2$	$\sim (\delta v/h)$

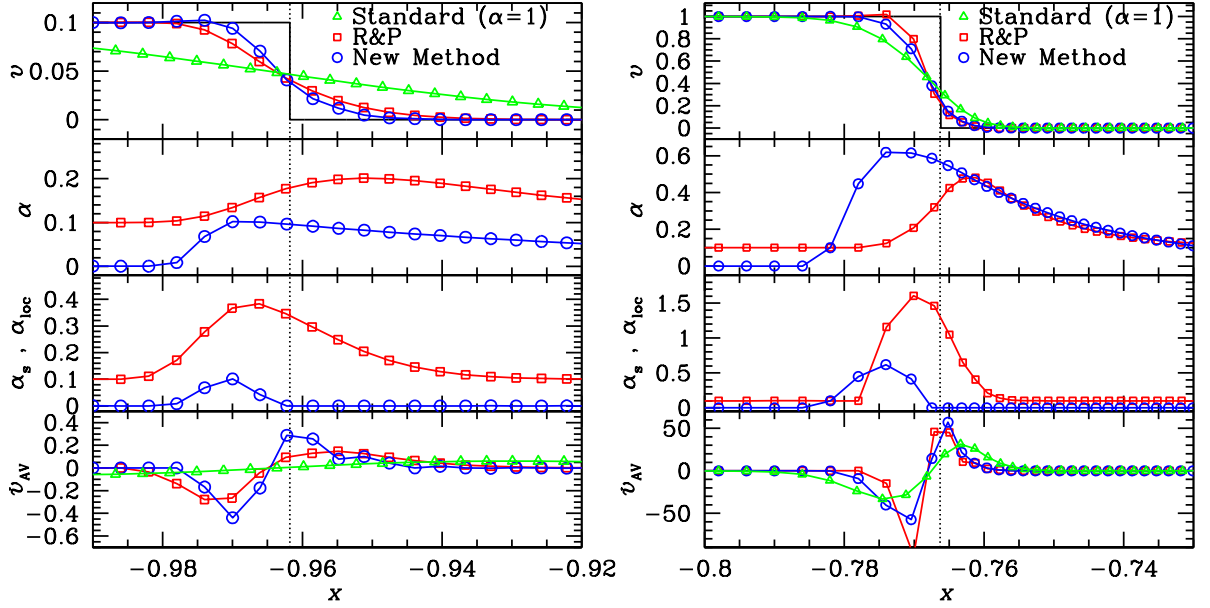
**Table 3.1:** Typical amplitudes of  $\nabla \cdot \mathbf{v}$ ,  $\nabla \cdot \mathbf{a}$  and  $D_t(\nabla \cdot \mathbf{v})$  for some simple situations.

### 3.4.6 Comparing The Viscosity Source Terms

We can determine how the new source term will behave compared to previous methods in a variety of simple yet common situations. Following Morris & Monaghan (1997) we firstly consider a simple homologous flow  $\mathbf{v} = -b\mathbf{x}$  with  $b > 0$ , which approximates certain astrophysical problems involving collapse and does not require artificial dissipation (Morris & Monaghan, 1997). In this case we find  $S_v = 3b$  but  $D_t(\nabla \cdot \mathbf{v}) = 0$ . Therefore in this case the viscosity is triggered in the M&M method but not in the new scheme and so the new scheme is clearly superior.

Secondly we consider a sound wave with amplitude  $v_a \ll c$  and wave number  $k \ll h^{-1}$  as an example of a well resolved, weakly convergent flow. In this case the M&M source term has an amplitude of  $kv_a$ , whereas the new scheme has an amplitude of  $c/v_{\text{sig}}hk^2v_a \approx hk^2v_a$ . Since  $v_{\text{sig}} \simeq c \gg v_a$  we have  $\alpha_{\text{loc}} \simeq \alpha_{\text{max}}h^2k^2(v_a/c)$ , while for the M&M method the asymptotic value  $\alpha_s \simeq \alpha_{\text{min}} + hk(v_a/c)\ell$ . In the limit  $kh \rightarrow 0$  of a well-resolved wave,  $\alpha_{\text{loc}} \rightarrow 0$  faster than  $\alpha_s \rightarrow \alpha_{\text{min}}$ , such that even with  $\alpha_{\text{min}} = 0$  the M&M method would be more viscous than our new scheme.

Finally, consider the situation of a strong shock with velocity discontinuity such that  $\delta v \gg c$ . The M&M method has been extensively tested in this regime and performs well, so the new scheme should have approximately the same source term in order to perform well in strong shocks. Assuming that the shock is smoothed over one kernel width, we find a maximum amplitude of  $\approx \delta v/h$  for  $\nabla \cdot \mathbf{v}$  while for  $D_t(\nabla \cdot \mathbf{v})$  the maximum amplitude is  $\approx (\delta v/h)^2$  (the exact values depend on the shock conditions and the functional form of the smoothing kernel, see appendix A for more details). However as we take a division by the signal velocity  $v_{\text{sig}} \approx \delta v$ , the new source term is comparable to the M&M method. Table 3.1 summarises the source terms for the various methods for some typical test cases.



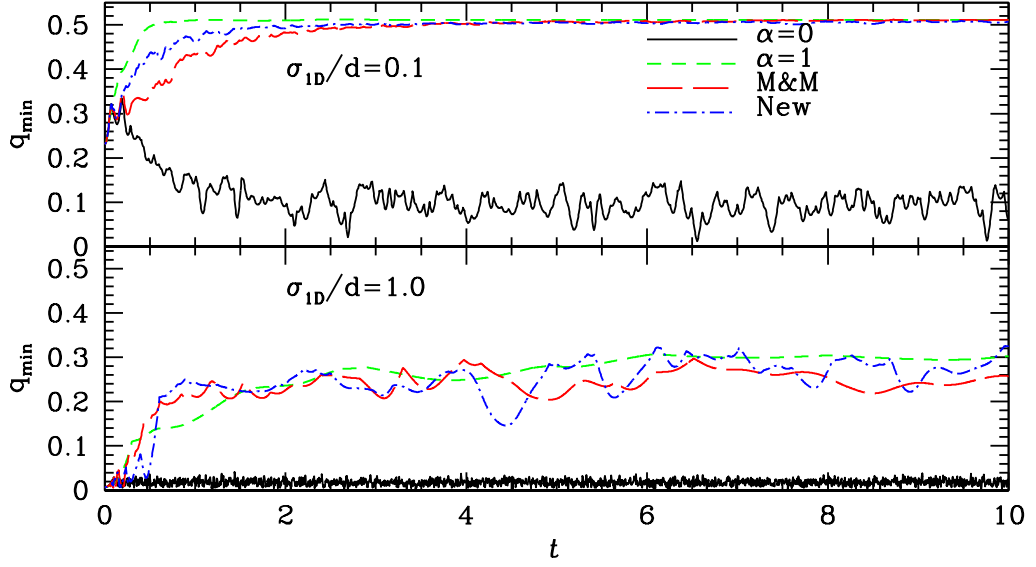
**Figure 3.7:** Comparing standard SPH, the M&M and our new method for a 1D ram shock with  $\Delta v = 0.1$  (left) and  $\Delta v = 1$  (right), using an adiabatic equation of state with  $\gamma = 1.4$ . We compare the velocity, viscosity parameter, its asymptotic value and the viscous deceleration. The shock propagates from right to left.

### 3.4.7 A Direct Comparison of Viscosity

In Fig. 3.7 we compare this new scheme with the R&P method as proposed by Price (2004) and with standard SPH for a weak (left) and strong (right) 1D ram shock (an initial velocity discontinuity of  $v = \pm\Delta v$  at  $x = 0$ , resulting in two shocks of  $\Delta v$  propagating in either direction from the origin). In both regimes the peak in  $\alpha_s$  and  $\alpha_{loc}$  respectively is one particle farther in front of the shock with our new method than the R&P method, which reflects the superiority of  $D_t(\nabla \cdot \mathbf{v})$  over  $\nabla \cdot \mathbf{v}$  to detect an approaching shock. This, combined with setting the viscosity parameter directly to the required value, results in the peak in  $\alpha$  to occur two particle separations *before* the shock for our new method, while for the R&P method it peaks a similar length *behind* the shock.

With our new method, the viscous deceleration (bottom panels in Fig. 3.7) sets in about three particle separations before both the weak and the strong shock, resulting in good shock capturing properties in either case. The R&P method, on the other hand, decelerates the flow much earlier for a weak shock than for a strong shock, exactly opposite to what one would want. This results in significant over-damping





**Figure 3.8:** Time evolution of  $q_{\min}$ , defined in equation (3.62), for SPH simulations started from noisy initial conditions (see text). All SPH schemes with artificial viscosity suppress the noise equally well.

of weak shocks (which also pertains to density and internal energy – not shown in Fig. 3.7), while our method smooths both shocks over four particle separations (top panels in Fig. 3.7), corresponding to the optimal SPH resolution. This difference enables the new method to better differentiate between converging flows and weak shocks, allowing converging flows to remain undamped.

As the new scheme detects the shock arriving earlier it removes the need for a minimum viscosity, which is one of the main drawbacks of the M&M/R&P method. The absence of this minimum in the M&M method leads to post-shock ringing as the peak in the source term occurs too close to the shock and a finite time is required to integrate the source term.

Note that standard SPH is hopeless: it over-smooths the strong shock and is completely incapable of dealing with the weak shock.

### 3.5 Is $\alpha_{\min} > 0$ required?

A major novelty of our method is the complete absence of artificial viscosity away from shocks. Hence, if  $\alpha_{\min} > 0$  was indeed required to suppress noise in the particle distribution, as previously argued, then our method should fail in this regard. Noise in SPH can emerge from shocks, carelessly generated initial conditions, or amplifi-

cation of rounding errors. In order to suppress shock-generated noise (‘post-shock ringing’), our method uses twice the decay time  $\tau$  as the M&M method by setting  $\ell = 10$ .

Here, we assess the stability against noise amplification by considering the time evolution of noisy initial conditions. We generate these by adding random displacements to particle positions representing noise-free hydrostatic equilibrium (the vertices of a face-centred-cubic grid, i.e. densest-sphere packing). We consider two cases with the displacements in each dimension drawn from a normal distribution with RMS amplitude equal to the nearest-neighbour distance or a tenth of it, respectively.

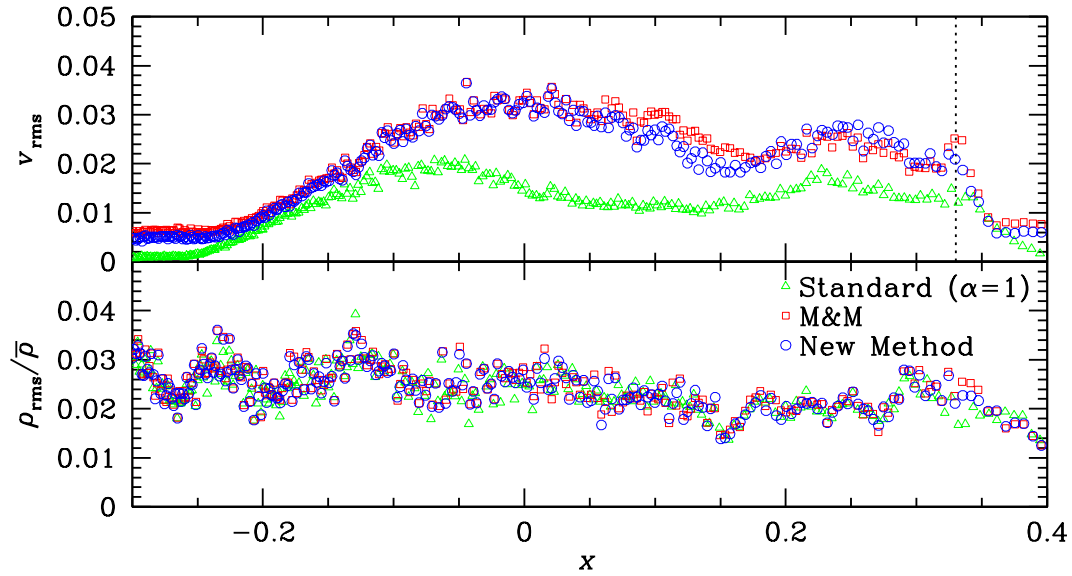
The time evolution of those noisy initial conditions generally has one of three possible outcomes, which can be distinguished by monitoring

$$q_{\min} \equiv \min_j \{(|\mathbf{x}_{ij}|/h_i)\}. \quad (3.62)$$

Either the particles settle back close to the original grid ( $q_{\min}$  approaches its grid value  $q_{\text{grid}}$ ), form a glass ( $q_{\min}$  approaches a finite value  $< q_{\text{grid}}$ ), or form dense clumps (‘clumping instability’,  $q_{\min} \sim 0$ ). Fig. 3.8 plots the evolution of  $q_{\min}$  for  $N_h = 40$  SPH neighbours when  $q_{\text{grid}} \approx 0.529$ . Clumping only occurs when no viscosity is used at all, while for any other scheme the particles settle back onto the grid or form a glass. Remarkably, even the time evolution of  $q_{\min}$  is very similar between the schemes tested (only the M&M method is a bit slow in suppressing low-amplitude noise). This clearly demonstrates that our novel viscosity scheme is as good as any other for maintaining particle order.

We can also consider the noise in post shock regions with the M&M scheme and the new method. This gives an indication of the noise level in typical SPH simulations. Post-shock noise occurs because the shock-induced compression disrupts the original particle order. The conventional wisdom among many SPH users is that a small amount of viscosity is required to reduce this post-shock ringing, hence the use of  $\alpha_{\min} = 0.1$ . However, we argue that if the shock is captured correctly, the decay in the viscosity parameter after the shock should be sufficient in preventing post-shock ringing.

In Fig. 3.9, we plot the amplitudes of the velocity and density noise in 3D simulations of the standard Sod (1978) shock tube test (see also §3.7.1). The three methods have similar levels of density noise, but standard SPH is less noisy in the



**Figure 3.9:** *The RMS amplitudes of density and velocity fluctuations for 3D simulations of the Sod (1978) shock tube test. Initial conditions were prepared using a glass. The shock propagates to the right and is indicated by the dotted line; the velocity jump at the shock is 0.63.*

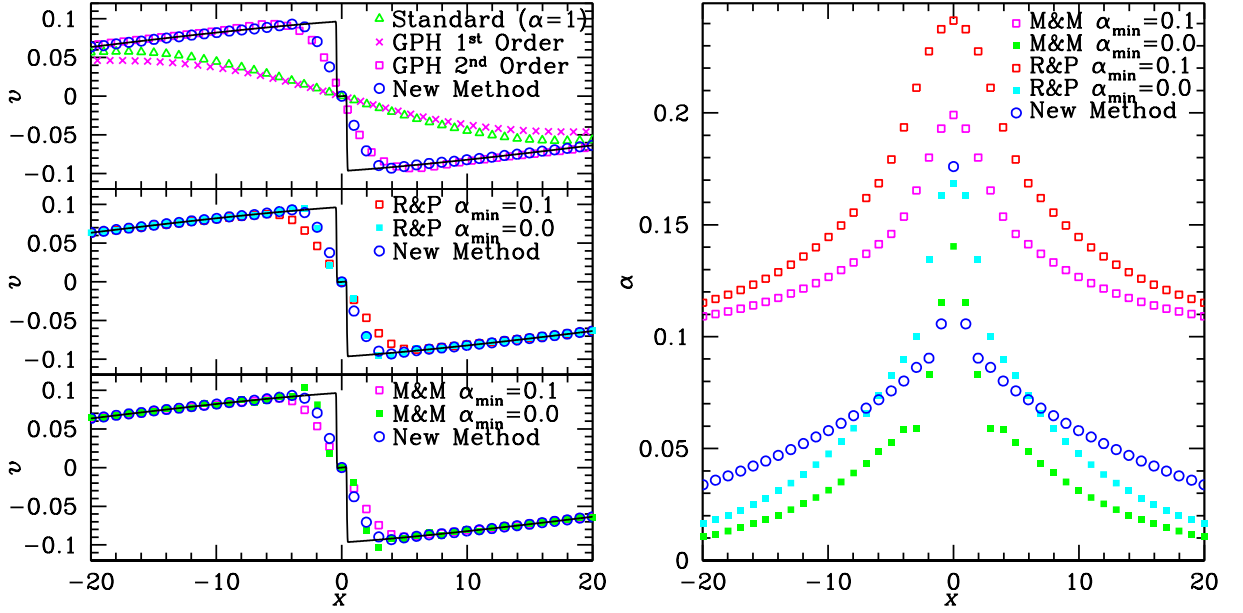
velocities, which is not surprising given its stronger viscosity. However, between the two viscosity suppressing schemes there is little difference, even though  $\alpha_{\min} = 0$  for our method. Similar results were obtained for other shock tests and we conclude that our method is no worse than M&M’s for maintaining particle order, demonstrating that  $\alpha_{\min} = 0$  does not significantly increase the noise error in post shock regions.

## 3.6 Viscosity Suppression Tests

We now present some tests to validate our new method in low-Mach-number regimes, where standard or R&P/M&M SPH gives too much unwanted dissipation. Therefore, our goal in this section is to demonstrate that our new method is better than or at least as good as the R&P/M&M switch.

### 3.6.1 Sound-wave steepening

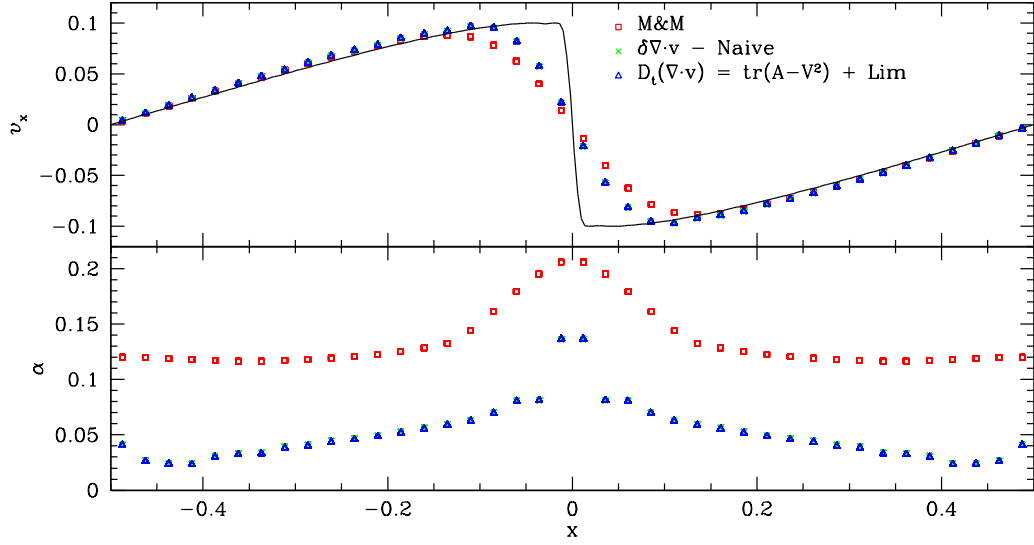
The steepening of sound waves is a simple example demonstrating the importance of distinguishing between converging flows and shocks. As the wave propagates, density and pressure oscillate adiabatically, resulting in variations of the sound speed. As a



**Figure 3.10:** Steepening of a 1D sound wave: velocity (left) and viscosity parameter (right) vs. position for standard SPH, M&M’s switch, R&P’s switch, our new method, and Godunov particle hydrodynamics of first and second order (GPH, Cha & Whitworth, 2003), each using 100 particles per wavelength. The solid curve in the left panel is the solution obtained with a high-resolution grid code. Our new method reproduces the discontinuity well over just five particles.

consequence, the density peak of the wave travels faster than the trough, eventually trying to overtake. This cannot happen as the fluid will become multivalued at that point, instead a shock forms transferring entropy. The time taken for the wave to steepen depends on the sound speed of the wave and the velocity amplitude of the wave; the larger the velocity amplitude compared to the sound speed the faster the wave steepens. This is a sensitive test for the viscosity switch as we require the suppression of viscosity at early times in order to allow the development of the shock with a good resolution. When the shock has formed the viscosity must be increased quickly to allow the transfer of entropy across the shock.

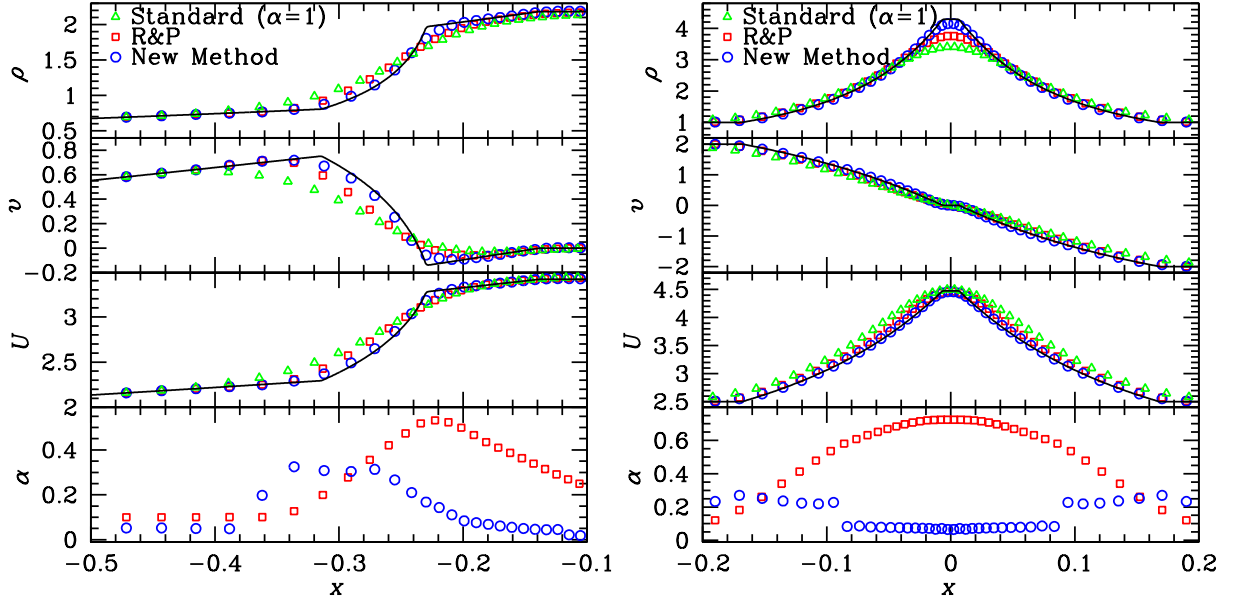
In our test, a 1D sound wave with a velocity amplitude 10% of the sound speed is used (adiabatic equation of state with  $\gamma = 1.4$ ). Fig. 3.10 compares the velocity field at the moment of wave steepening for various SPH schemes, each using 100 particles, with a high-resolution grid simulation (Ramses using 4096 cells, Teyssier, 2002). The new method resolves the shock better than both the M&M and R&P schemes (with  $\alpha_{\min} = 0.1$ ) as the viscosity is kept low until the shock forms, reducing



**Figure 3.11:** *Steepening of a sound wave in 2D. **Top** shows velocity and **bottom** shows viscosity parameter vs. position. We compare the various methods for estimating  $D_t(\nabla \cdot \mathbf{v})$  and demonstrate that the limiter is does no harm in this case. Solid line shows the Ramses result.*

unwanted dissipation. Furthermore, M&M with  $\alpha_{\min} = 0$  is unable to fully smooth the velocity across the shock, leading to oscillations in the velocity either side of the shock. This is because the viscosity is not increased quickly enough at the instance of the shock forming. The R&P method with  $\alpha_{\min} = 0$  also performs well as the viscosity is slightly higher in low viscosity regimes, allowing good smoothing of the shock over a few particles.

In Fig. 3.10, we also show results from a GPH (Godunov-type particle hydrodynamics, Cha & Whitworth, 2003) simulation. This scheme only differs from SPH by using the pressure  $P^*$ , found by solving the Riemann problem between the particle and its neighbour, in the momentum and internal energy equations. This avoids the need for explicit artificial viscosity as the Riemann solver can describe all solutions of the 1D Euler equations with two constant initial states. This substitution does not affect the energy or momentum conservation (Cha, 2002), and indeed we find that both are well conserved. There is a significant difference between the first-order and second-order GPH scheme. The former is comparable to standard SPH and also to an Eulerian Godunov grid code using the same Riemann solver without (artificial-viscosity suppressing) interpolation (not shown). The second-order GPH scheme, on the other hand, resolves the discontinuity almost as well as the new scheme.



**Figure 3.12:** A 1D converging flow test with initially constant density and pressure and velocities given by equation (3.63) using an adiabatic equation of state with  $\gamma = 1.4$ . **Left:** run for  $v_a = 1$  at  $t = 0.3$ ; **Right:** run for  $v_a = 2$  at  $t = 0.1$ . The solid lines are the result of a high-resolution Eulerian grid-code simulation using *Ramses*.

We may also perform the sound wave steepening in higher dimensions. The results in 2D are shown in Fig. 3.11 where we compare the M&M scheme to the new method. We find that the all methods for estimating  $D_t(\nabla \cdot \mathbf{v})$  are equivalent in this case and that the limiter has no affect on the viscosity. This is because in this case the flow is converging and so the ratio  $R = -1$  and there is no shear flow, thus viscosity is not limited.

### 3.6.2 1D Converging Flow Test

This test is similar to sound-wave steepening in that it requires good treatment of convergent flows and weak shocks. The initial conditions are uniform pressure and

density and a continuously varying flow velocity

$$v = \begin{cases} 4(1+x)v_a & -1.00 < x < -0.75, \\ v_a & -0.75 < x < -0.25, \\ -4xv_a & -0.25 < x < 0.25, \\ -v_a & 0.25 < x < 0.75, \\ 4(1-x)v_a & 0.75 < x < 1.00. \end{cases} \quad (3.63)$$

As there is no analytical solution, we compare the results to a high-resolution grid-code simulation (*Ramses* simulation with 4096 grid cells compared to 100 SPH particles). We run tests for  $v_a = 1$  and  $v_a = 2$  as shown in the left and right panels of Fig. 3.12.

While the R&P switch certainly improves upon standard SPH, it still over-smoothes the velocity profile as the viscosity is increased before a shock has formed. This is particularly evident in the velocity profile of the  $v_a = 2$  case (right panel) near  $x = 0$ . The new switch keeps the viscosity low, in the  $v_a = 2$  case, an order of magnitude lower than with the R&P method. In fact, the agreement between our method and the high-resolution grid code is as good as one can possibly expect at the given resolution, in particular the velocity plateau and density amplitude around  $x = 0$  in the  $v_a = 2$  case are correctly modelled.

### 3.6.3 2D Keplerian Ring

This test deals with the evolution of a non self gravitating gaseous ring, orbiting around a central point mass. Initially the ring is in equilibrium as pressure forces, the attraction by the point mass, and inertial (centrifugal) forces cancel. As the Keplerian orbits are differentially rotating the flow is shearing and any viscosity causes the ring to spread (Lynden-Bell & Pringle, 1974). This type of problem was one of the original motivations for the Balsara switch (see §3.2.2), as the SPH form of the artificial viscosity (equation 3.3) gives not only a bulk viscosity but also a shear viscosity (Meglicki et al., 1993). The Balsara switch was introduced to better model such differentially rotating configurations by reducing viscosity in the presence of shear flows.

Maddison et al. (1996) performed pressure-free simulations of rings with a Gaussian radial density profile. Using standard SPH artificial viscosity, they found a

spreading of the ring consistent with the predictions from Lynden-Bell & Pringle. However they also found an instability to develop from the inner edge of the ring, which quickly breaks up the disc. They blame this on a viscous instability (Lyubarskij et al., 1994), which causes eccentric orbits at the inner edge of the ring to become more eccentric due to the viscous deceleration peaking at apo-centre.

Imaeda & Inutsuka (2002) find a similar effect in SPH simulations of gaseous (not pressure-free) Keplerian discs, and blame it on alleged inadequacy of the SPH scheme itself. Triggered by these claims, Price (2004) and Monaghan (2006) performed SPH simulations of the same and similar problems and found no such numerical instabilities. However, while pressure forces likely act to stabilise the viscous instability, gaseous rings are potentially unstable to non-axisymmetric perturbations (Papaloizou & Pringle, 1984, 1985; Goldreich & Narayan, 1985). These grow on a hydrodynamical time scale and were effectively diminished by Monaghan's (2006) choice of a very low sound speed compared to the orbital velocity.

In our test, we use  $GM = 1000$  for the central point mass and use a gas ring with Gaussian surface density given by

$$\rho = \rho_0 e^{-(r-a)^2/b}, \quad (3.64)$$

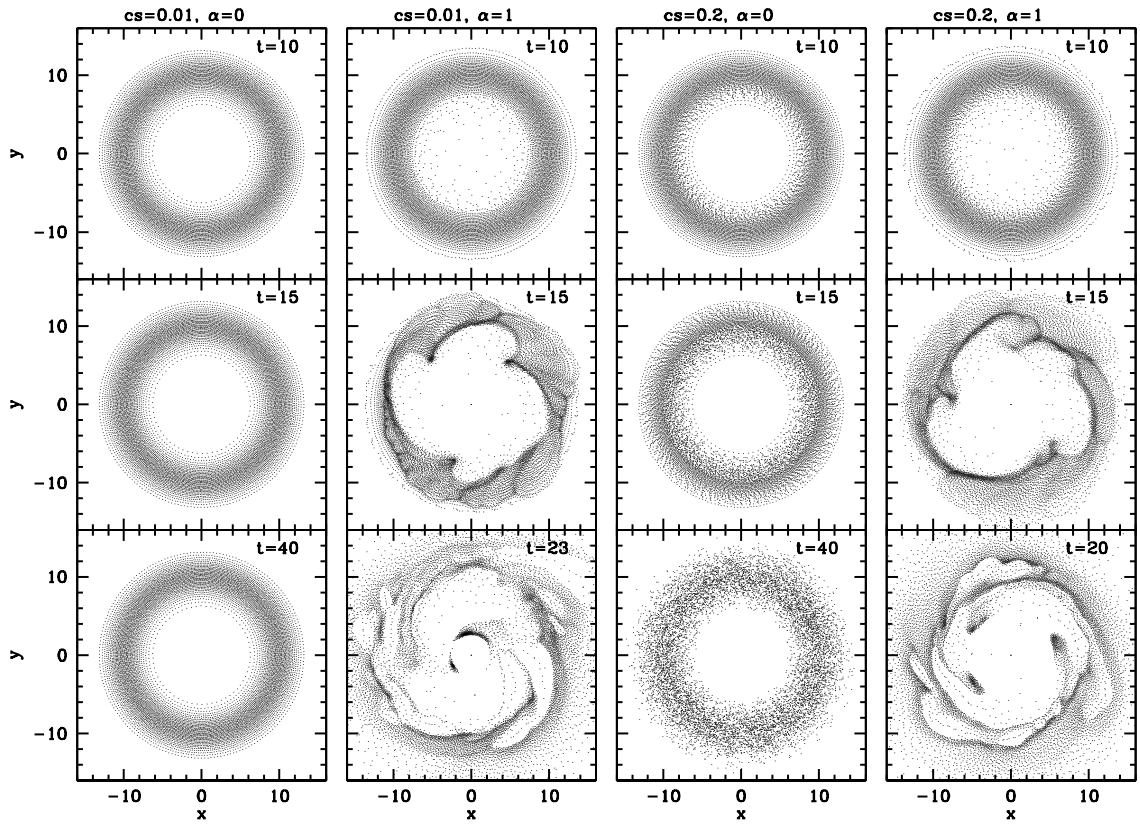
centred on  $a = 10$  with width  $b = 5$  represented by  $N = 9745$  particles initially placed according to the method of Cartwright et al. (2009). The rotation velocities were set by firstly computing the total acceleration due to the point mass and hydrodynamic forces for each particle. In order for the disc to be stable, the total acceleration must be inwards so that it can be balanced by rotation, which was the case for the parameters used in this study. The particles were given a rotation such that  $a = v^2/r$ , this ensured that the disc was initially in exact dynamic equilibrium.

As we used a relatively low sound speed in these tests, the hydrodynamic force was initially significantly smaller than the gravitational force. Thus the velocity profile is relatively unaffected by changing the sound speed. This implies an orbital period of  $\approx 2\pi$  and velocity of  $v_{\text{orb}} \approx 10$  at the centre of the ring. We also applied a very small softening to the point mass force such that

$$\mathbf{F} = \frac{GM}{(r^2 + e^2)^{3/2}} \mathbf{r}, \quad (3.65)$$

where  $e = 0.1$ . This was used so that particles that strayed into the inner regions of the disc did not receive a large kick. This allowed the simulation to proceed for



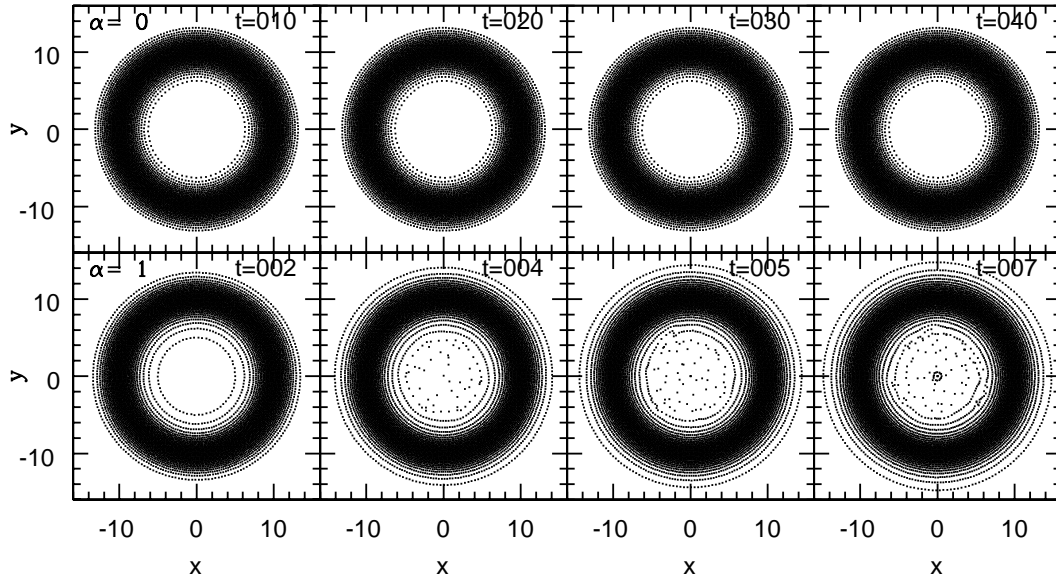


**Figure 3.13:** *The effect of sound speed on the stability of the Keplerian disc with and without artificial viscosity is demonstrated.*

longer, allowing the break up of the disc to be monitored.

We firstly compare the effect of sound speed, viscosity and hydrodynamics on the stability of the rings. Fig. 3.13 shows the disc at various times with and without artificial dissipation using two sound speeds of  $c_s = 0.01$  and  $c_s = 0.2$ . When using a low sound speed ( $c_s = 0.01$ ) without artificial dissipation, the particles in the disc remain ordered and the density remains on its initial profile after around five rotations of the densest part of the disc. However, increasing the sound speed changes the stability properties of the disc, we find that at  $t = 15$  ( $\approx 2.4$  orbits of the densest part of the disc) the particle distribution has become noisy and by  $t = 30$  ( $\approx 4.8$  rotations) this noise has grown. This is also observed in the density profile (not shown), as time progress the density profile moves further from the initial distribution.

When an artificial viscosity of  $\alpha = 1$  is used, particles at the inner edge of the

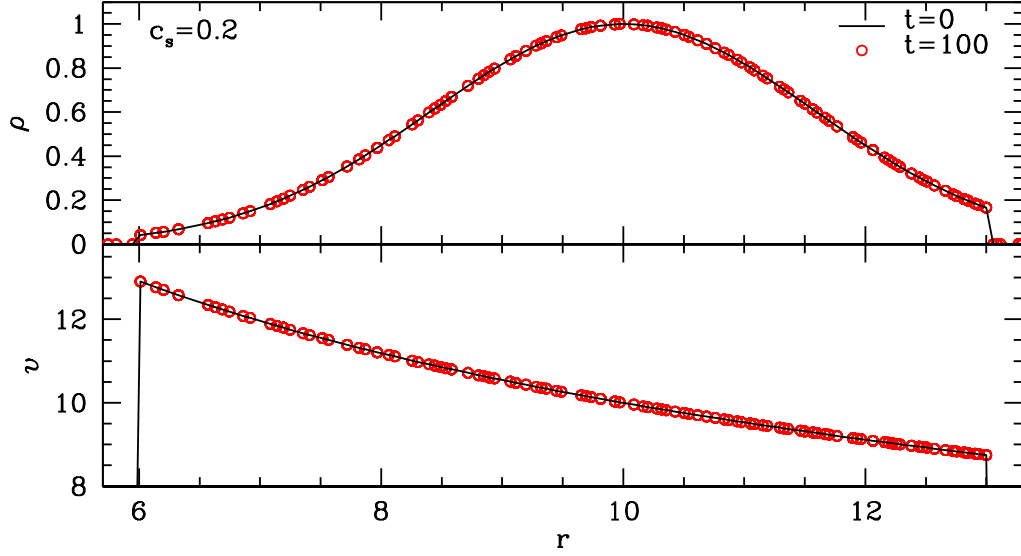


**Figure 3.14:** *The Keplerian ring test with no hydrodynamic forces and  $c_s = 0.2$ . The effect of artificial viscosity on the stability of the disc is demonstrated.*

disc move inwards. This leads to a few particles receiving large accelerations as they approach the point mass. Particles just outside the inner edge of the disc are accelerated outwards, causing parts of the disc to increase in density. This destroys the dynamic equilibrium of the disc which quickly leads to the disc breaking up. We observe that in this case with  $\alpha = 1$ , both the  $c_s = 0.01$  and  $c_s = 0.2$  discs are disrupted on the same time scale.

We can also investigate the effect of viscosity on the stability of the ring by turning off the hydrodynamic forces so that the only forces acting are gravity and artificial viscosity. This allows us to observe the effect of only the artificial viscosity, without the hydrodynamic instability. Fig. 3.14 shows the results of the Keplerian disc with  $c_s = 0.2$  when hydrodynamic forces are turned off. When no artificial viscosity is applied the disc remains unperturbed as expected. This is because by turning off hydrodynamic forces and artificial dissipation we are essentially following a set of test particles which do not interact with each other. However, when viscosity is turned on particles from the inner edge of the disc move inwards towards the point mass. Particles that are at the outer edge of the disc move outwards, therefore the disc spreads as predicted by Lynden-Bell & Pringle.

We may also look at how the Keplerian ring behaves in other numerical schemes, in this case we use **Ramses** (Teyssier, 2002). In grid based methods artificial viscosity is not used, with dissipation being provided implicitly by the Riemann solver.

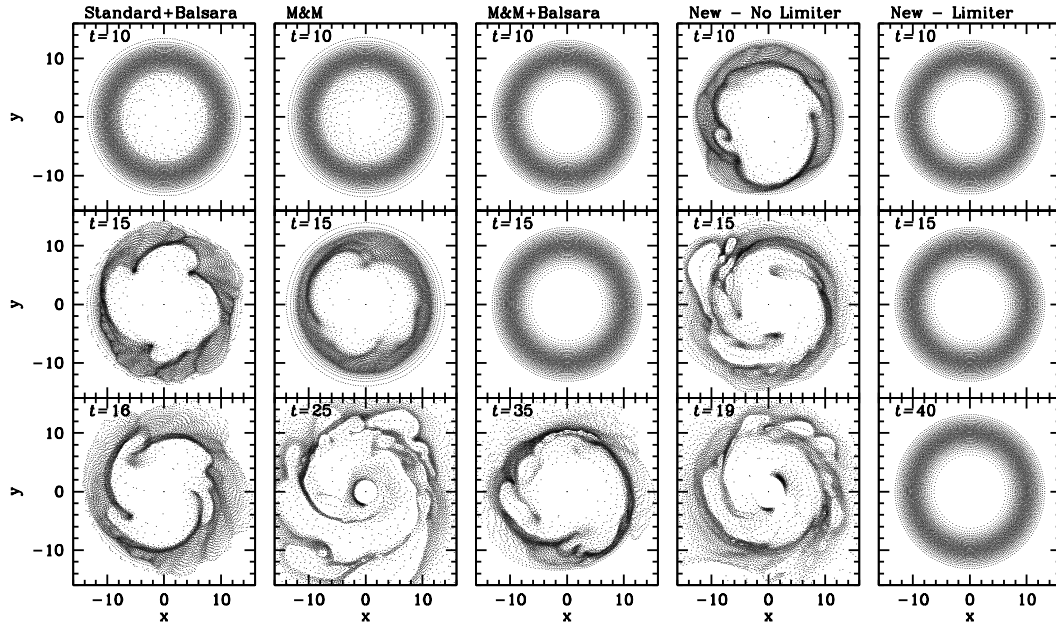


**Figure 3.15:** *Density and velocity profile of the Keplerian ring with  $c_s = 0.2$  at  $t = 0$  and  $t = 100$  with Ramses.*

Fig. 3.15 shows that the  $c_s = 0.2$  ring remains on the initial density and velocity profile after more than fifteen orbits of the density peak. By comparison the standard SPH scheme (with no viscosity) causes the disc to become clumpy in less than three rotations.

There appears to be two separate instabilities in SPH, both of which are not present in the Ramses simulation. There is a viscous instability caused by the application of artificial viscosity. It is clear that this type of instability would not be present in the Ramses simulation, as there is no artificial dissipation in the grid based scheme. The rate at which the viscous instability occurs depends on the type of viscosity method used (i.e. how large  $\alpha$  is) and to a lesser extent the sound speed. The dynamic instability (first noticed by Imaeda & Inutsuka, 2002) which occurs in the SPH scheme is again not present in the Ramses simulation. The rate at which the dynamic instability occurs depends on the sound speed, with it occurring quicker for larger sound speeds.

We now investigate the effect of different artificial viscosity schemes on the simulation of the disc. As we are primarily interested in the effects of artificial viscosity, we choose a sound speed of  $c_s = 0.01$  such that the hydrodynamic instability is suppressed. Fig. 3.16 shows the particle distribution in the disc at various times for the M&M method. We see that the viscosity quickly causes the onset of the instability when using the M&M switch. When this is augmented with the Balsara



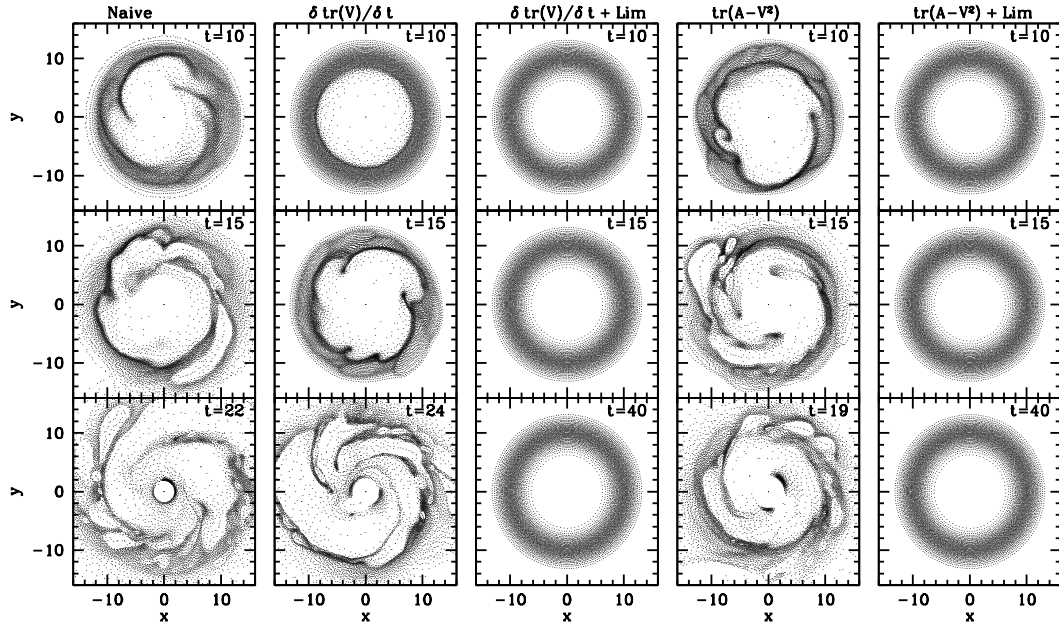
**Figure 3.16:** A comparison of the new scheme to a Balsara viscosity and the M&M method for the Keplerian disc test.

limiter we find that the onset of the viscous instability is delayed, however, the disc is still quickly disrupted. This is because the Balsara limiter can only reduce the effects of artificial viscosity; it cannot completely eliminate viscosity. Therefore as the viscosity is lowered the time for the viscous instability to take effect is longer.

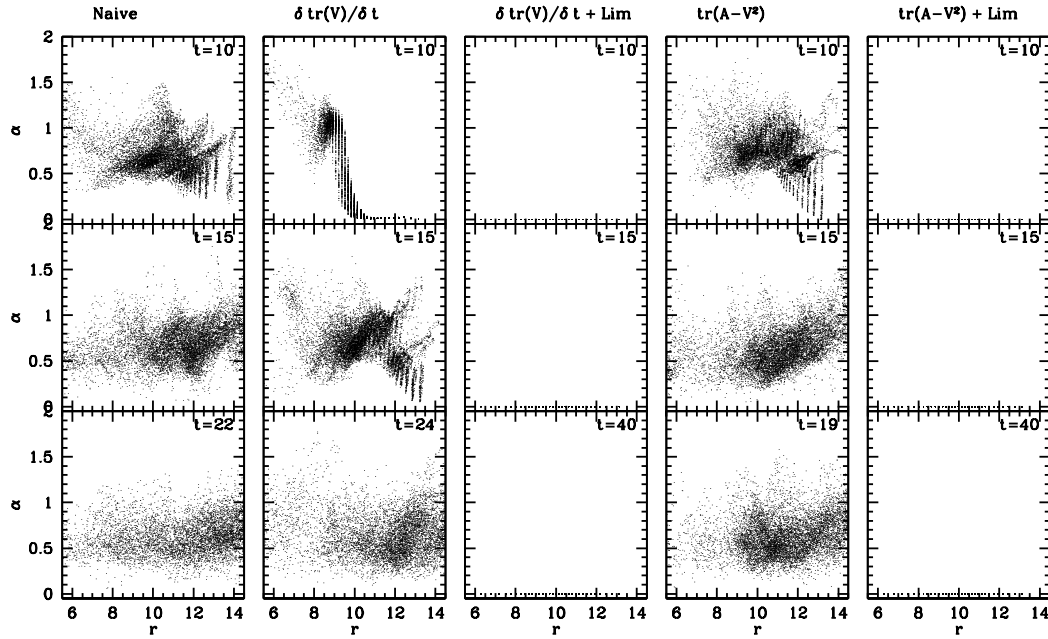
In the new scheme (Fig. 3.16) without the shear limiter, viscosity is introduced quickly as was the case for the M&M method. The induced viscosity in the new scheme (and the M&M scheme with  $\alpha_{\min} = 0$ ) is due to anisotropies in the particle distribution which then, via the shear velocity field, causes the estimate of velocity divergence to be non zero. The limiter, outlined in section 3.4.4, is able to significantly reduce the viscosity (see Fig. 3.18) allowing the disc to remain as stable as when no artificial dissipation was used.

Fig. 3.17 shows the particle profiles for the new scheme with different estimators for  $D_t(\nabla \cdot \mathbf{v})$ . Using the naive estimate of  $\nabla \cdot \mathbf{v}$  (equation 3.44) to compute  $D_t(\nabla \cdot \mathbf{v})$  (equation 3.43) causes particle disorder to be quickly introduced. Using the corrected estimator for  $\nabla \cdot \mathbf{v}$  (equation 3.49) or computing  $D_t(\nabla \cdot \mathbf{v})$  directly (equation 3.52) delays the onset slightly but still does not avoid the problem. This is because although the problem of anisotropies in the particle distribution is significantly reduced (see Fig. 3.19), any slight noise in the estimate for  $D_t(\nabla \cdot \mathbf{v})$  will

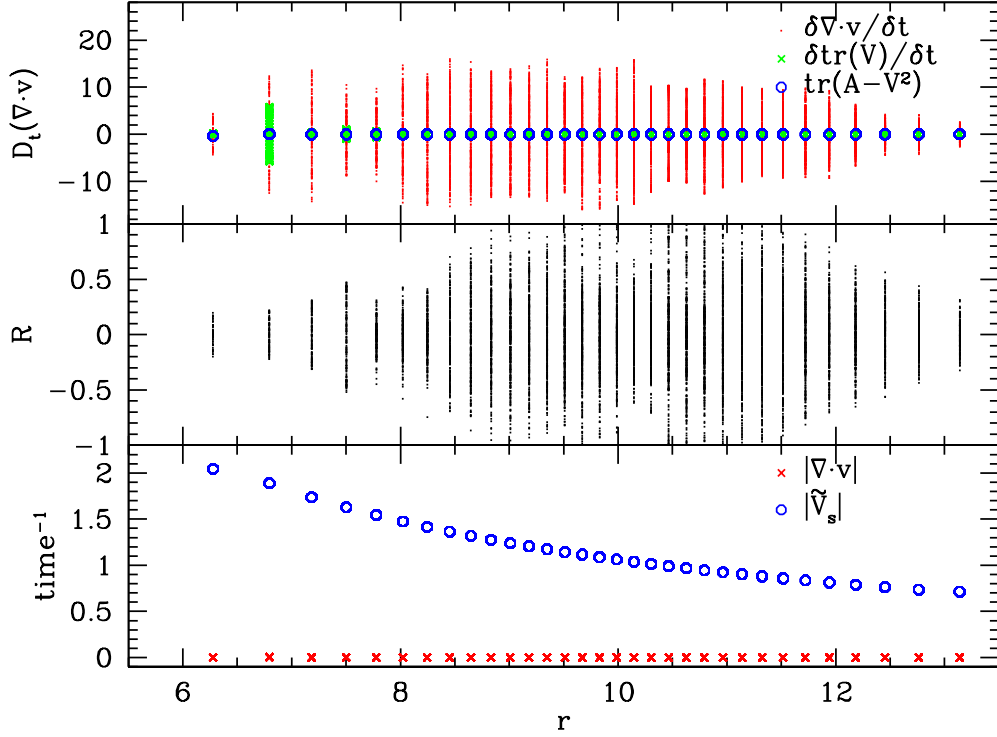




**Figure 3.17:** Particle distributions plotted at various times for the Keplerian ring test using the new viscosity scheme with  $c_s = 0.01$ .



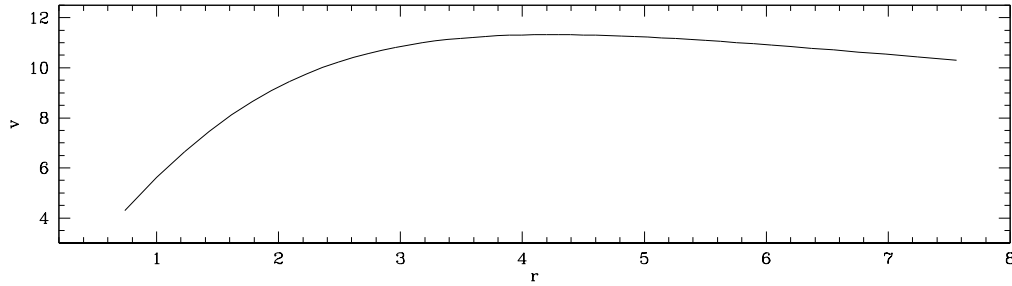
**Figure 3.18:** The viscosity parameter plotted at various times for the Keplerian ring test using the new viscosity scheme with  $c_s = 0.01$ .



**Figure 3.19:** The Keplerian ring test at  $t = 1$  ( $1/6^{\text{th}}$  of a rotation). **Top**) shows the three estimations for  $D_t(\nabla \cdot \mathbf{v})$ . In red and green we show  $D_t(\nabla \cdot \mathbf{v}) = \delta \text{tr}(\mathbf{V})/\delta t$  with the red using the naive estimate for  $\nabla \cdot \mathbf{v}$  and green the improved estimator. In blue we show  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\mathbf{A} - \mathbf{V}^2)$ . The **middle** plot show the ratio  $R$  (equation 3.57). The **Bottom** plot shows the strengths of the velocity divergence and shear in the velocity gradient.

increase viscosity leading to particles moving out of equilibrium. This “slight noise” can occur from higher order terms in the Taylor expansion of equation (3.46) which are not accounted for. This will further enhance the viscosity, leading to a break down of the disc. The noise limiter (see section 3.4.4) is able to detect that the signal in  $D_t(\nabla \cdot \mathbf{v})$  is not caused by true convergence in the velocity field, but instead is caused by anisotropies in the particle distribution. As no shocks are detected, the shear term dominates the limiter leading to a substantial reduction in the artificial viscosity. This is a vast improvement of the previously best result of the M&M method with the Balsara switch.

The values of  $D_t(\nabla \cdot \mathbf{v})$  for three of the estimators are shown at  $t = 1$  (which is  $1/6^{\text{th}}$  of a rotation period) in Fig. 3.19. The estimate  $D_t(\nabla \cdot \mathbf{v}) = \delta \nabla \cdot \mathbf{v} / \delta t$  (with the naive estimate for  $\nabla \cdot \mathbf{v}$ ) is significantly larger than the other two estimators as the correction for the anisotropy in the particle positions is not applied. We also



**Figure 3.20:** *Rotation curve for the Softened disc test.*

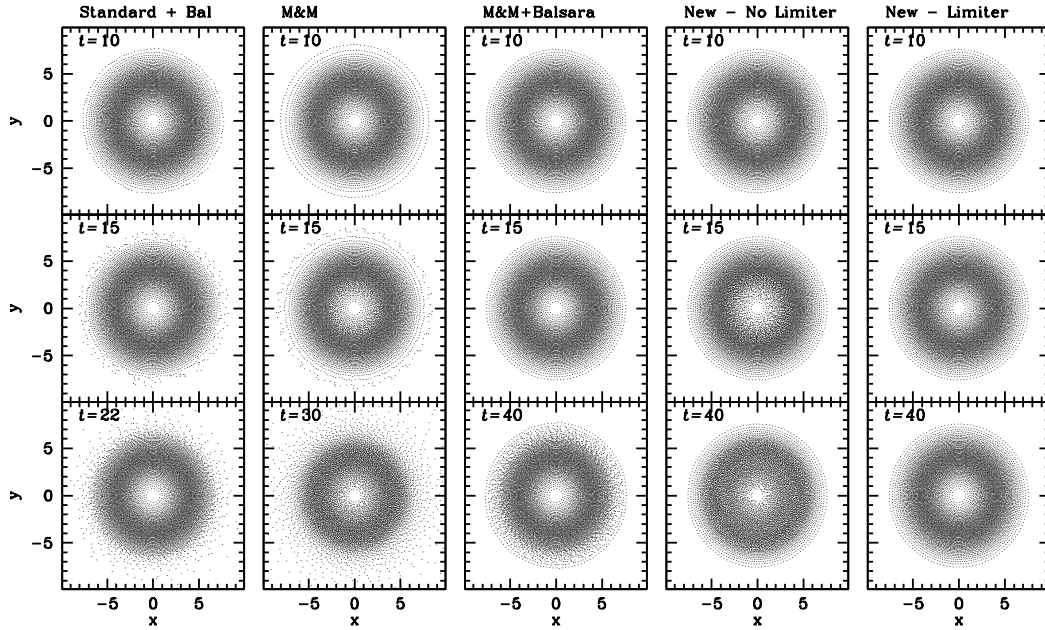
find that in the inner parts of the disc, the estimate  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2)$  has a much smaller spread than the estimate  $D_t(\nabla \cdot \mathbf{v}) = \delta \text{tr}(\underline{\mathbf{V}})/\delta t$ . This indicates that the estimate  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2)$  may be less susceptible to numerical noise in shearing flows.

Fig. 3.19 also shows the ratio  $R$  (equation 3.57) based on the improved estimate for  $\nabla \cdot \mathbf{v}$  (equation 3.49). We see that around  $r = 9$  the ratio does reach  $R = -1$ , indicating a real convergence. However, as the estimate for  $D_t(\nabla \cdot \mathbf{v})$  is so small compared to the shear in this region, the limiter reduces the viscosity such that it has a negligible effect on the simulation.

### 3.6.4 Softened Disc

We may investigate the effect of artificial viscosity on discs with more complex rotation curves. A problem with a viscosity scheme based on  $\nabla \cdot \mathbf{a}$  is that in discs with a rising rotation curve the viscosity will be switched on (this was demonstrated in section 3.3.3). We again use a Gaussian ring as described in the last section, however this time we centre the disc on  $a = 4$  (see equation 3.64) and apply a softening of  $e = 3$ . This means that part of the disc is strongly softened and therefore the rotation curve in the inner parts rises as  $r$  increases, as shown in Fig. 3.20. The period at the maximum density is  $P = 2.2$ .

Fig. 3.21 shows the particle distributions at various times for the viscosity schemes. We observe that with standard SPH (with the Balsara limiter) and the M&M method, the particles are perturbed from the initial distribution due to the application of artificial viscosity. The M&M method with the Balsara limiter performs well in this test, keeping the viscosity low. However, we notice some effects of



**Figure 3.21:** Comparing particle distributions for the softened disc test for

viscosity in the outer regions of the disc. The new scheme without the limiter also shows some slight effects of viscosity. However, when the limiter is applied these slight effects are prevented.

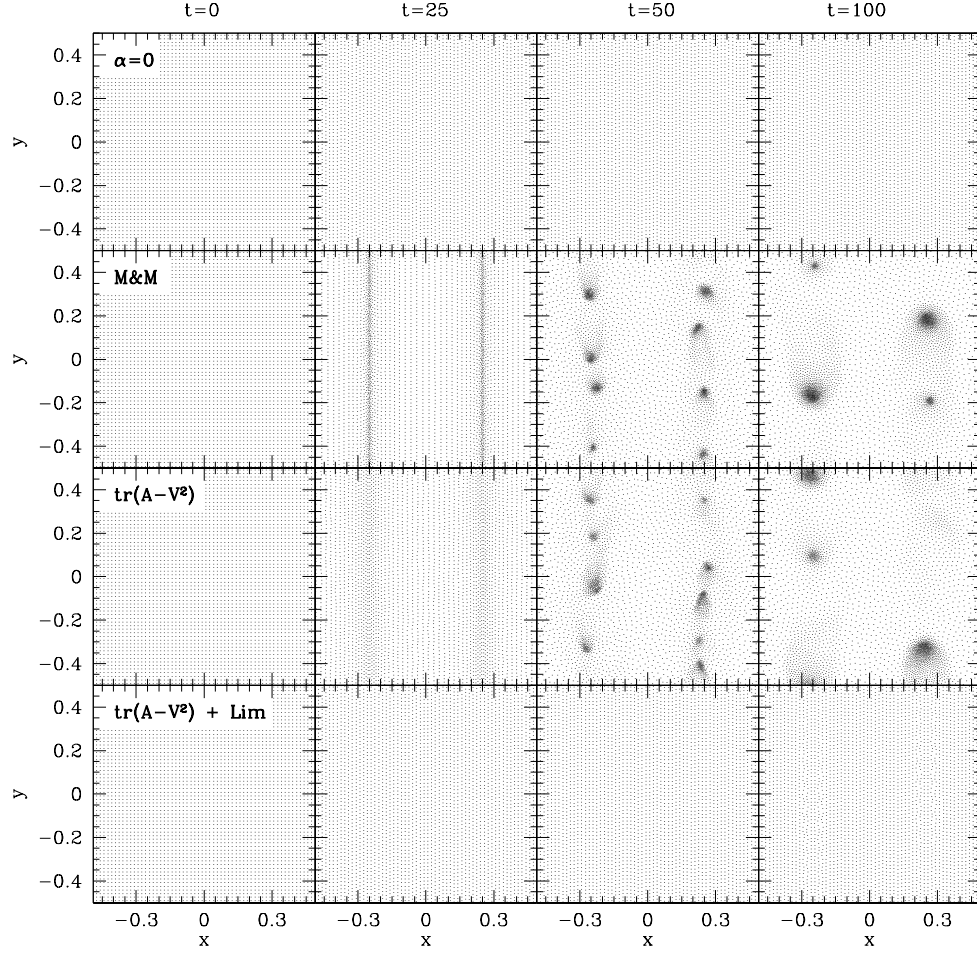
### 3.6.5 2D Shear Flows

This test looks at the response of the viscosity parameter to shearing flows. It has been suggested by Imaeda & Inutsuka (2002) that SPH is unstable to shear flows, with large density errors emerging after around a dynamical time. Further investigation by Price (2004) and Monaghan (2006) failed to show this instability. Here we investigate not only the stability of SPH in shear flows but also the effect of artificial viscosity on the scheme.

We firstly repeat the simulations of Imaeda & Inutsuka (2002), Price (2004) and Monaghan (2006). We use a uniform 2D grid of  $60 \times 60$  particles in a periodic box of size  $-0.5 \leq x \leq 0.5$  and  $-0.5 \leq y \leq 0.5$ . The velocity is set such that  $v_x = 0$  and  $v_y = \sin(2\pi x)$ , with a sound speed of  $c_s = 0.01$ .

The top plot in Fig. 3.22 shows the particle positions at various times when no artificial viscosity is used. We find that the particles remain well ordered after 100 crossings of the shear wave. However, increasing the viscosity leads to particles

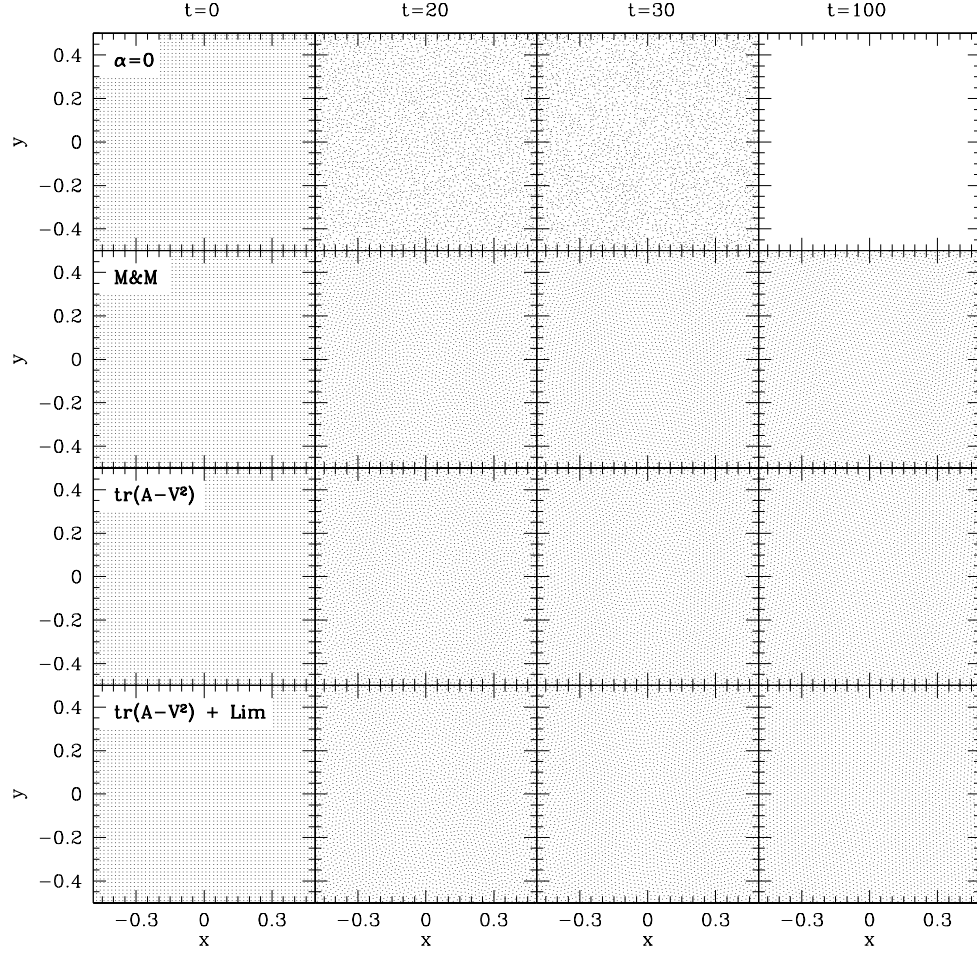




**Figure 3.22:** Particle positions for a 2D sinusoidal shear flow with amplitude of unity and sound speed of  $c_s = 0.01$ . The **top** row has no viscosity, **2<sup>nd</sup>** row is M&M viscosity, **3<sup>rd</sup>** row is the new scheme with  $D_t(\nabla \cdot \mathbf{v}) = \delta \text{tr}(\underline{\mathbf{V}})/\delta t$  and the **bottom** row is the new scheme with  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2)$ .

moving towards the peaking in the velocity profile. This can be seen in the plots for the M&M scheme and the new scheme (using  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\underline{\mathbf{A}} - \underline{\mathbf{V}}^2)$ ) at  $t = 25$ . This eventually leads to dense clumps of particles after 100 wave crossings. It should be noted that this only occurs when artificial viscosity is applied, we do not see these errors in the absence of artificial viscosity.

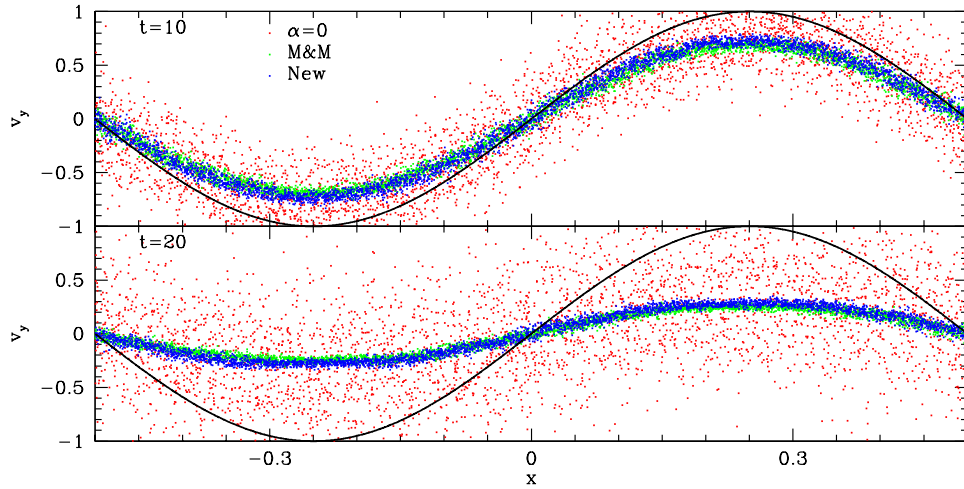
The effect of viscosity on the simulations can be readily understood. As particles shear past each other there are small density errors due to the anisotropy of particle field. These small density errors can cause small pressure variations which drive a small acceleration in the  $x$  direction. In the absence of artificial viscosity these density and pressure errors oscillate with small amplitudes which do not grow. Arti-



**Figure 3.23:** Particle positions for a 2D sinusoidal shear flow with amplitude of unity and sound speed of  $c_s = 1$ . The **top** row has no viscosity (note that the code crashed around  $t = 50$ ), **2<sup>nd</sup>** row is M&M viscosity, **3<sup>rd</sup>** row is the new scheme with  $D_t(\nabla \cdot \mathbf{v}) = \delta \text{tr}(\mathbf{V})/\delta t$  and the **bottom** row is the new scheme with  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(\mathbf{A} - \mathbf{V}^2)$ .

ficial viscosity tries to suppress this small motion in the  $x$  direction which generates heat, further increasing the pressure. This drives larger accelerations and therefore larger  $x$  velocities. This dissipates more heat leading to a run away process causing particles to be strongly accelerated in the  $x$  direction.

When using the noise limiter we find a substantial improvement in the results. The viscosity is greatly reduced which prevents the dense clumps forming. However, we do find that if the simulation is allowed to proceed for longer dense clumps form by  $t = 150$ . This is because although the viscosity is strongly limited it is not completely eliminated. Therefore the viscosity induced errors still grow at a much slower rate.



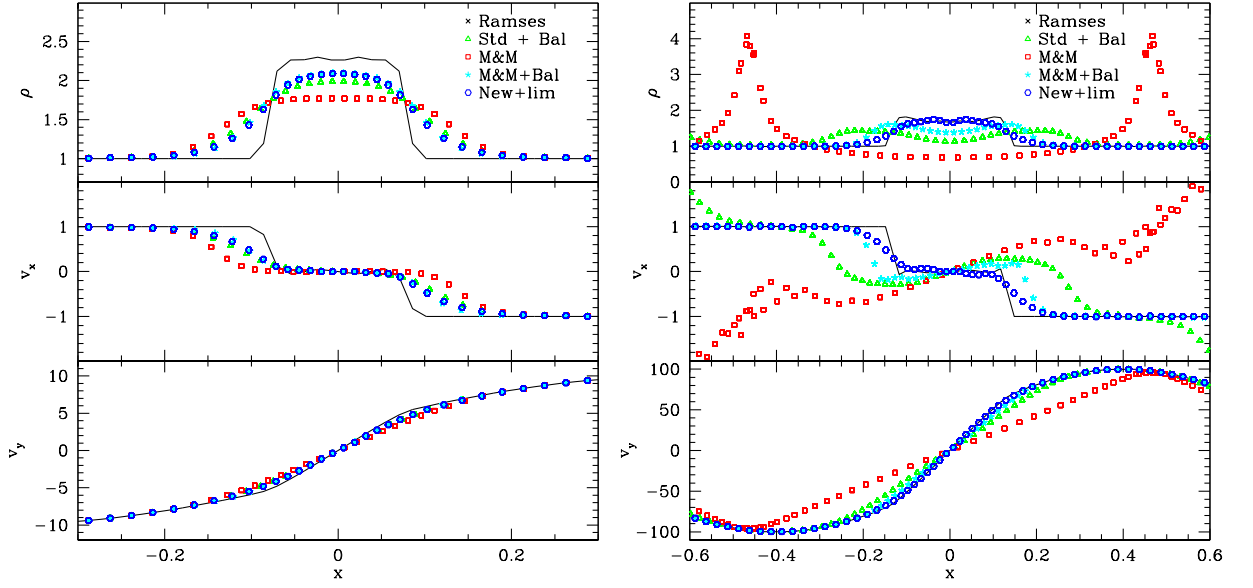
**Figure 3.24:** Particle velocities for a 2D sinusoidal shear flow with amplitude of unity and sound speed of  $c_s = 1$ .

We may also investigate the shear flow when a higher sound speed is used. We repeat the experiment with  $c_s = 1$  and show the results in Fig. 3.23. When no artificial viscosity is used the particle distribution quickly becomes noisy, as does the velocity profile (see Fig. 3.24). However, when using any of viscosity schemes tested the particles remained ordered. However, as we see from Fig. 3.24 the shear flow is damped away.

The results are similar to that of Imaeda & Inutsuka (2002) who find that long term shear flows in SPH are unstable and lead to noise in both the particle distribution and velocity field. However, we find that there are two instabilities, a viscous instability and a fluid instability. The timescale of the fluid instability decreases with increasing sound speed and can be suppressed with a low sound speed (as was the case with Price ,2004 and Monaghan, 2006). We also show that with higher sound speeds the fluid instability grows quickly leading to noise in the particle and velocity fields. This noise may be suppressed by the application of artificial viscosity, at the expense of damping the shear wave.

### 3.6.6 A shearing shock

This test combines a shock with a perpendicular shear and presents a difficult test for any SPH scheme. We use periodic boundary conditions and start from positions



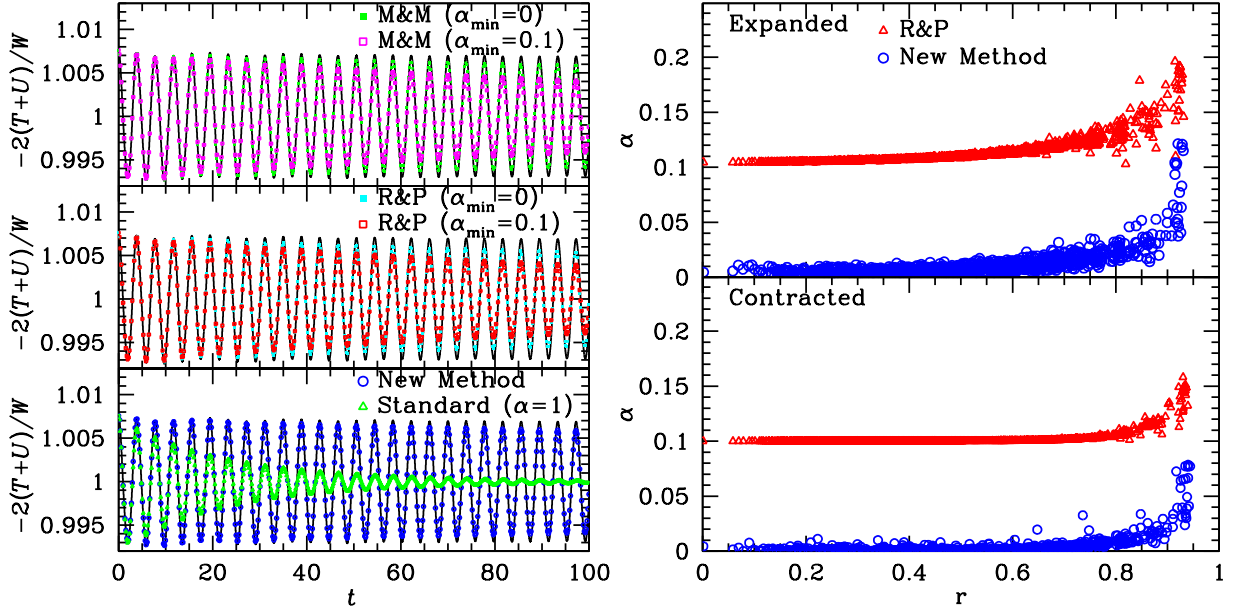
**Figure 3.25:** *Shear shock test with  $s = 10v = 10c$  (left) and  $s = 100v = 100c$  (right) showing the density and velocity for the various viscosity schemes.*

on a FCC grid and velocities

$$v_x = -v \operatorname{sign}(x), \quad v_y = s \sin(\pi x), \quad \text{and} \quad v_z = 0. \quad (3.66)$$

In Fig. 3.25, we present results for various SPH simulations as well as a grid-code run for  $s = 10v = 10c$  and  $s = 100v = 100c$ . The M&M method produces a large viscosity due to the contamination of  $\nabla \cdot \mathbf{v}$  by the shear field, leading to spurious results, especially with  $s = 10v$ . Using the Balsara limiter with either M&M or Standard SPH results in much improved results, however the shock is somewhat over-smoothed especially with Standard SPH. The new scheme is able to limit the viscosity to the correct level allowing good capturing of the shock and retaining particle order in the post-shock region. The results with  $s = 10v$  show a slight over-smoothing of the shock, indicating that the limiter is slightly too weak in this test.

This is a tricky problem for SPH due to shear induced artificial viscosity. Fig. 3.25 demonstrates the need for a good shear limiter, the viscosity must be reduced such that shear does not hamper the simulation whilst still retaining the ability to capture shocks correctly. For this reason this test and the Keplerian disc are important tests when trying to limit the viscosity in shearing flows. This problem can be made



**Figure 3.26:** *Left:* virial ratio plotted versus time for SPH models of a radially oscillating polytrope which initially was in its fundamental radial eigenmode with relative radial amplitude of 0.01 and period 3.89. The solid curves are for a SPH model without any artificial viscosity. *Right:* the viscosity parameter  $\alpha$  at  $t = 97$  (maximum contraction) and  $t = 99$  (maximum expansion) for every 100th particle. The new method keeps viscosity lower at the edge of the polytrope.

harder by increasing the ratio  $s/v$  of shear to shock amplitude, which will over limit the viscosity preventing the capturing of the shock.

### 3.6.7 An Oscillating Polytropic Sphere

The aim of this test is to examine the behaviour of our new viscosity method in situations where the gas self-gravity is important. We set up a polytropic sphere of  $10^5$  particles and induce oscillations in its fundamental mode (e.g. Cox, 1980, see appendix C) with relative amplitude of 0.01 in radius and a period of  $P = 3.8$ .

In the absence of viscosity we expect the radial oscillations to continue with the initial amplitude and period over a many oscillations as SPH is a dissipationless scheme. However, as with any numerical method we may experience a small amount of numerical dissipation due to other factors such as the type of integrator used. Nonetheless, such effects should be small compared to the dissipation caused by artificial viscosity. Since the size of the radial perturbations increases with radius, we expect the oscillations to be small at the centre of the polytrope and therefore

our new method to keep the viscosity low there. However, at the edge the size of the oscillations are more significant, and we may see an increase in viscosity at this point.

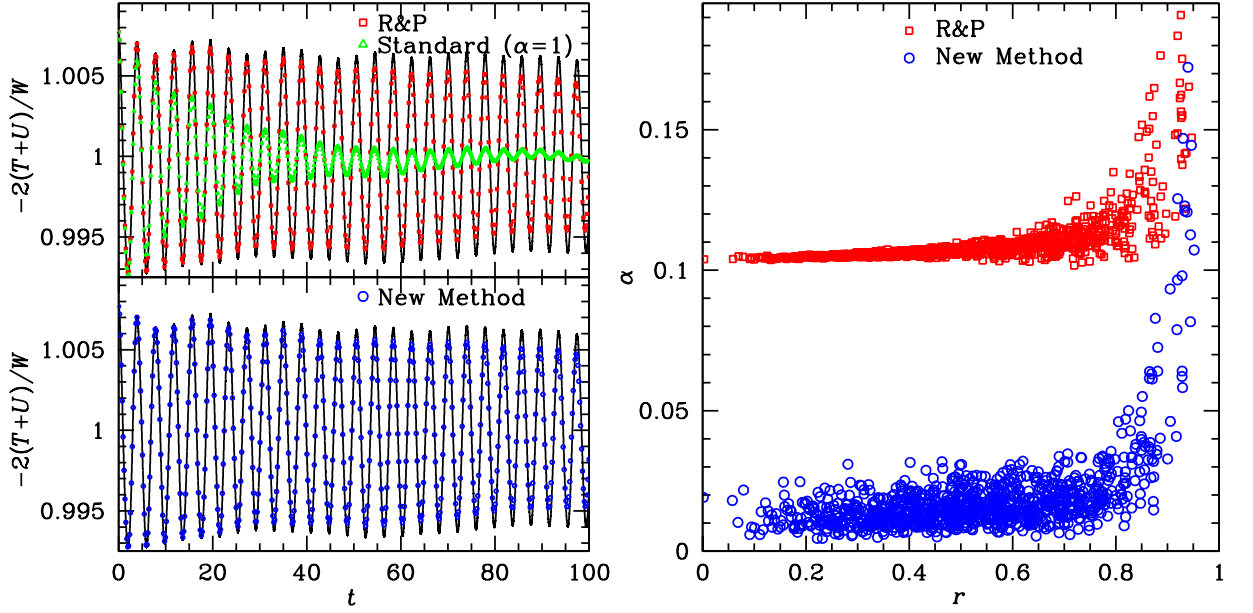
In order to track the oscillations, we monitor in Fig. 3.26 the time evolution of the virial ratio  $-2(T + U)/W$  where  $T$ ,  $U$ , and  $W$ , are the kinetic, internal, and the gravitational energies, respectively. At maximum contraction the virial ratio is at its peak because the sphere, which is about to expand, has too much internal energy compared to gravitational energy. Similarly, at maximum expansion the virial ratio is minimal as the sphere's internal energy is not high enough to balance gravity, and so it starts to contract. With no artificial viscosity (solid curves in Fig. 3.26) the wave remains at an almost constant amplitude barring a slight initial drop after the first oscillation. The period averaged over 25 oscillations is  $P = 3.89$ , only slightly larger than the expected value. The reason for this discrepancy is most likely the unavoidable deviation of the (finite-resolution) SPH model from a perfect polytropic sphere. This deviation also means that our SPH model is not exhibiting a pure eigenmode, but in addition contains some higher-order modes at low amplitudes, resulting in some beating between them.

The M&M and R&P methods both show a slow but continuous decay of the oscillations, though the period is hardly affected: after 25 oscillations damped and undamped models are still in phase. This damping can be blamed largely on the finite  $\alpha_{\min}$ , although both methods with  $\alpha_{\min} = 0$  still show noticeable damping. This demonstrates the inability of the M&M source term to distinguish shocks from converging flows. Note that the standard SPH damps the oscillation away ten times faster than the M&M and R&P methods. Our new method, on the other hand, hardly damps the oscillations at all, even though the viscosity in the outermost regions becomes almost comparable to that for the M&M method (right panel). This is because the viscosity is kept very small throughout the inner regions of the sphere, demonstrating the increased ability to distinguish between converging flows, such as this, and shocks.

### 3.6.8 Oscillating Polytrope Orbiting a Point Mass

To determine how the various schemes perform in the presence of a strong external acceleration field, we place the oscillating polytropic sphere on a circular orbit 20 times the radius of the sphere, around a point mass 100 times the mass of the sphere (corresponding to a period of 56 time units). With this choice, the tidal



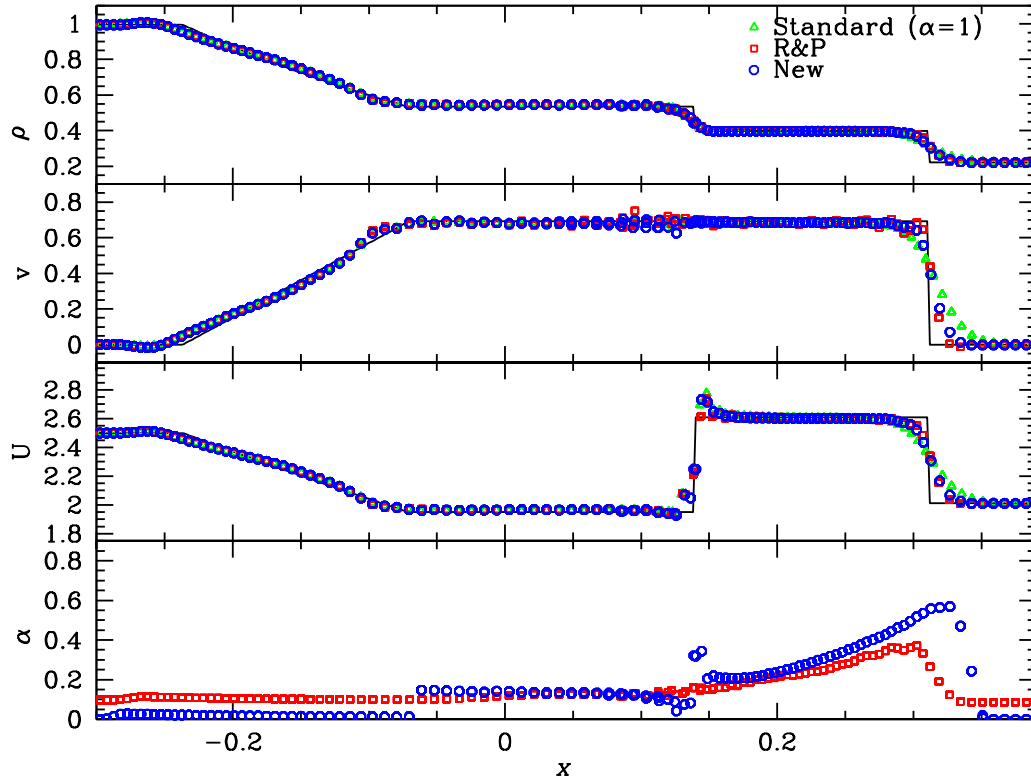


**Figure 3.27:** *Left:* virial ratio plotted versus time for SPH models of a radially oscillating polytrope in circular orbit around a point mass of 100 times its mass and with orbital radius 20 times its radius (the kinetic and potential energies are corrected for the contributions from the orbit). The solid curves are for a SPH model without any artificial viscosity. *Right:* The viscosity parameter for every 100th particle at  $t = 100$ .

radius is approximately four times closer to the point mass than the gas sphere, implying that tides are strong but not enough to break up the polytrope. However, the accelerations due to the orbit are much larger than those due to the polytropic oscillations, and therefore this is a tough test for any numerical hydrodynamical scheme. In particular, Eulerian grid codes generally have severe problems.<sup>13</sup>

The time evolution of the virial ratio and the viscosity parameter  $\alpha$  are shown in Fig. 3.27. First note that the undamped simulations (solid curves) behave differently from the isolated case, exhibiting variations and a slight decay, both of which are most likely caused by the tidal field. As to be expected for any Lagrangian scheme, both methods perform very similar to the isolated case, because neither  $\nabla \cdot \mathbf{v}$  nor  $D_t(\nabla \cdot \mathbf{v})$  are affected by the external acceleration field.

<sup>13</sup>Using co-rotating coordinates is not an option if one wants to allow for tidal evolution of the orbital period. Moreover, in more general problems, i.e. elliptic orbits, such methods are unavailable



**Figure 3.28:** Comparison of new method and R&P's method for the standard Sod (1978) shock tube test with the analytic solution (solid).

## 3.7 Shock capturing tests

In this section, we subject our method to situations where artificial viscosity is required, mainly high-Mach number shocks. Our aim is to demonstrate that it performs at least as well as the M&M method.

### 3.7.1 Sod shock tube test

The Sod (1978) shock tube test is a standard test for any shock capturing method. This problem consists of an initial discontinuity in pressure and density leading to the production of a rarefaction wave, contact discontinuity and shock wave which forms from the steepening of a subsonic wave. The whole system is subsonic with a maximum Mach number of  $\mathcal{M} \approx 0.63$  in the pre-shock region.

We perform the test in a 3D box of dimensions  $(\pm 1, \pm 0.1, \pm 0.1)$  with 195712 particles. The domain is initially split into two equal regions at  $x = 0$ . The left region has density, pressure and velocity given by  $(0.993, 0.993, 0)$  and the right



region has  $(0.221, 0.178, 0)$ . Therefore in the high density region there are 200 layers of particles in the  $x$  direction.

The density, energy, velocity and viscosity for the R&P switch and our method are shown in Fig. 3.28. As for the 1D ram test (§3.4.7), our new method switches on viscosity already in the pre-shock region peaking about one smoothing length before the actual shock front (which travels to the right in Fig. 3.28). The R&P switch turns on viscosity later, lagging our method by about four particle separations. As a consequence, the transition of the fluid values across the shock front is smoother with our method than with the R&P method.

Note that the irregularities around the contact discontinuity at  $x = 0.138$  common to all schemes tested are not related to artificial viscosity (the irregularities in  $\alpha$  at that point could be removed by choosing non-zero initial  $\alpha$  at the initial discontinuity); they can be alleviated by artificial conductivity (Price, 2004, 2008).

For the new scheme we use the estimate  $D_t(\nabla \cdot \mathbf{v}) = \text{tr}(A - V^2)$  with the limiter described in section 3.4.4. However, we find that all estimates of  $D_t(\nabla \cdot \mathbf{v})$  produce good results in this test. Furthermore, the limiter has no effect on the results as the scheme is detecting clear convergence ( $R = -1$  around the shock front) and there is no shear flow.

### 3.7.2 1D Ram $\mathcal{M} = 50$ Shock

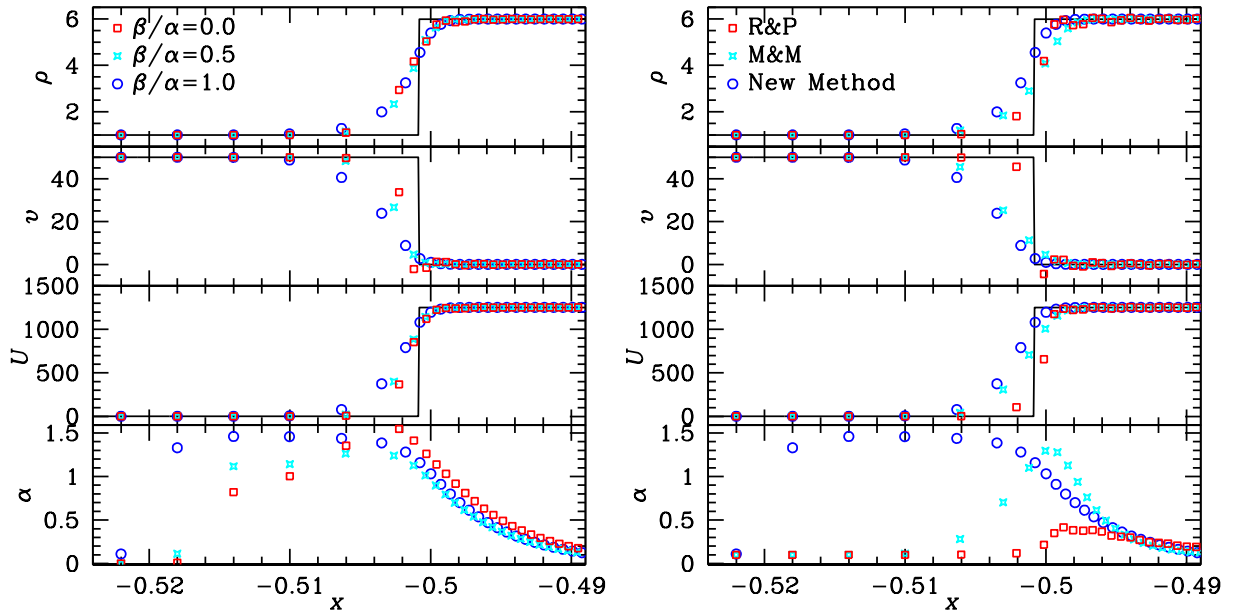
Thus far we have demonstrated that the new switch is able to resolve weak shocks and converging flows with greater accuracy than the R&P method. Now we demonstrate that the new method is also able to handle high Mach number shocks where strong dissipation is required.

In order to see how the  $\beta$  term affects a simple 1D strong shock we perform a ram shock with a Mach number of  $\mathcal{M} = 50$ . We then compare the effects of the  $\beta$  term on the new viscosity method<sup>14</sup>, as shown in Fig. 3.29. All three values of  $\beta/\alpha$  allow the viscosity to reach the same peak value, however the higher  $\beta/\alpha$  the further in front of the shock  $\alpha$  increases. This is due to the shock being smoothed over a larger region with the higher  $\beta$  term, allowing the shock to be seen earlier by the viscosity method. Although the peak in  $\alpha$  is the same, setting  $\beta/\alpha = 0$  tends to undersmooth the shock compared to  $\beta/\alpha = 1$ .

In Fig. 3.29 we also compare the R&P and M&M methods to the new scheme

---

<sup>14</sup>Note that in all tests so far we have used  $\beta/\alpha = 2$  for standard SPH, M&M and R&P. For the new method all tests so far have used  $\beta/\alpha = 1$ .



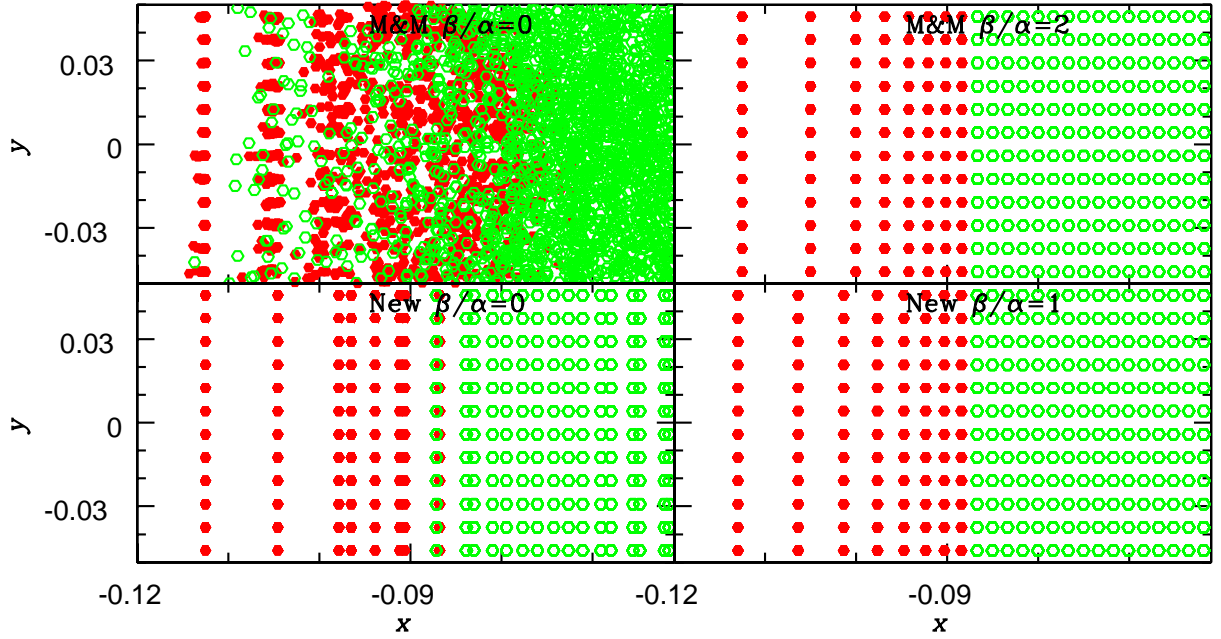
**Figure 3.29:** 1D ram shock with  $\Delta v = 50$ . **Left:** Comparing the effects of  $\beta/\alpha$  for the new viscosity method. **Right:** Comparing the M&M, R&P and the new method.

( $\beta/\alpha = 1$ ). The R&P method, which uses equation (3.15) with  $\alpha_{\max} = 2$ , is implemented in some contemporary SPH codes and has been used in our tests so far. We find that the R&P method fails this test as  $\alpha$  remains too low and as a consequence the velocity discontinuity is not correctly smoothed and some post-shock ringing occurs. However we find that the original M&M method works well (stars in Fig. 3.29) as there is no limit on the asymptotic value. As this is only integrated for a short time, an asymptotic value of  $\alpha_s \gg 2$  is needed to allow the viscosity to reach the required amount. Our new method works about as well as the original M&M scheme, with  $\alpha$  reaching the same level, though our scheme detects the coming shock much earlier:  $\alpha$  is ahead of the original M&M method by about four particle separations.

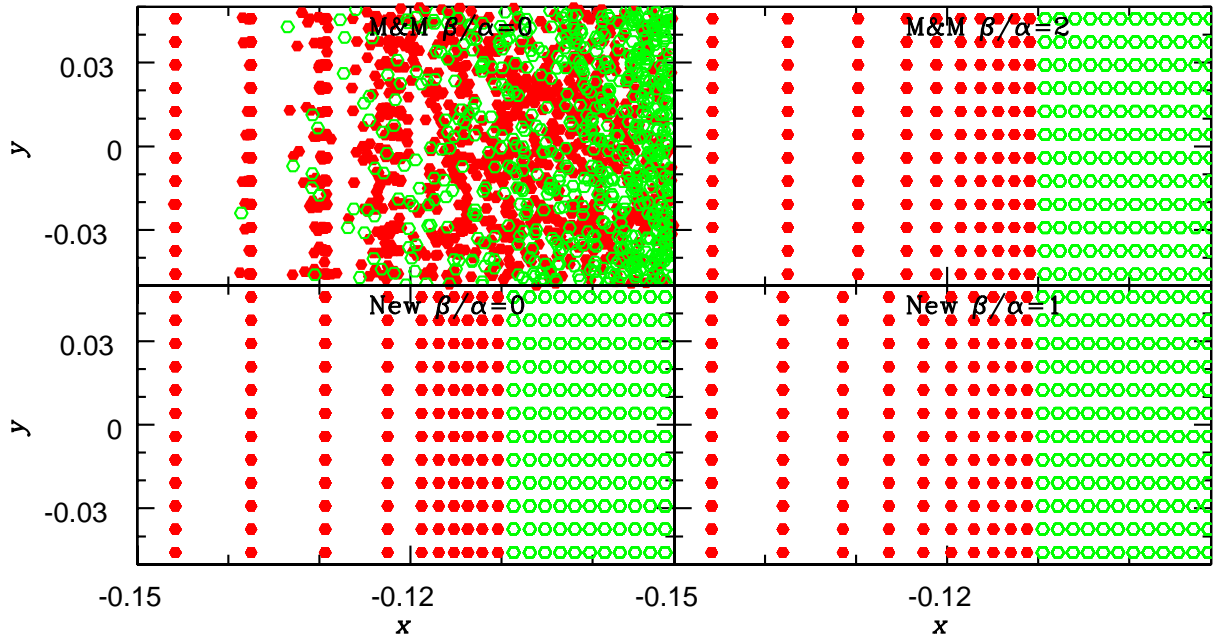
### 3.7.3 Particle Penetration

Whilst the main role of artificial viscosity is to resolve shocks by transferring entropy, a secondary but vital role is to prevent particle penetration by reducing the relative velocity of approaching particle pairs. In order to prevent particle penetration, this reduction must occur sufficiently fast, requiring an appropriate amount of viscosity.

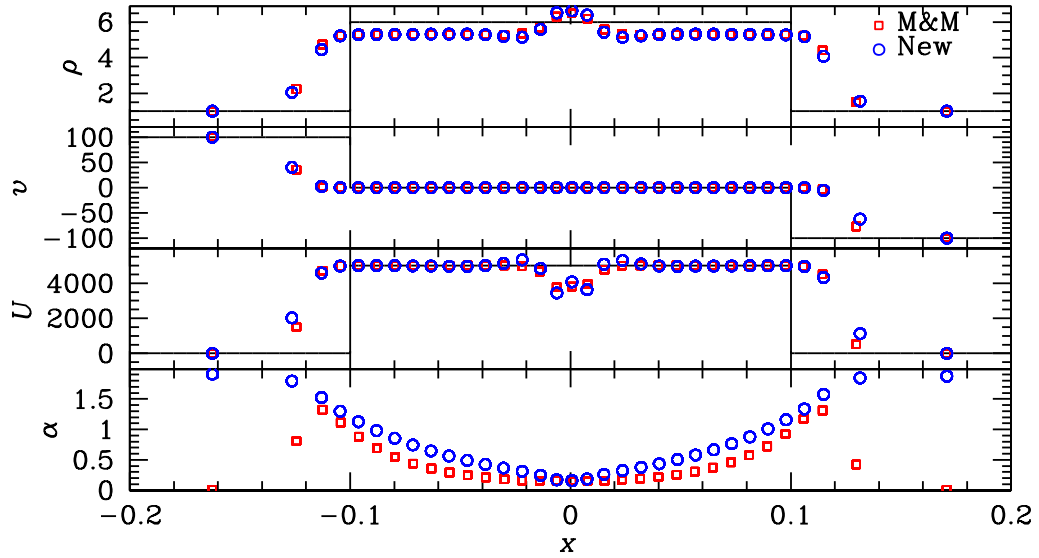
Bate (1995) performed many tests to determine the value of the parameters  $\alpha$  and  $\beta$  required to prevent particle penetration in ram shock test of various Mach numbers. In these tests, the particle configuration plays an important role. For an



**Figure 3.30:** Particle positions in the  $x$ - $y$  plane of 3D simulations of a  $M = 20$  ram shock along the  $x$  direction at  $t = 0.02$ . Particles are coloured red if there initial positions was  $x_0 < -0.045$  and green if  $x_0 > -0.045$



**Figure 3.31:** Particle positions in the  $x$ - $y$  plane of 3D simulations of a  $M = 100$  ram shock along the  $x$  direction at  $t = 0.005$ . Particles are coloured red if there initial positions was  $x_0 < -0.57$  and green if  $x_0 > -0.57$



**Figure 3.32:** *Density, velocity, internal energy and viscosity profiles for 3D simulation of a  $\mathcal{M} = 100$  ram shock at  $t = 0.005$ . The M&M switch had  $\beta/\alpha = 2$  whilst the new scheme had  $\beta/\alpha = 1$ . Note that not all particles are plotted.*

offset grid, Bate found that no values for the artificial viscosity can prevent particle penetration for shocks with  $\mathcal{M} \gtrsim 5$ . For particles arranged in densest packing (face-centred-cubic grid, FCC), the SPH equilibrium for homogeneous 3D fluids, or cubic grids, Bate found that appropriate values for the viscosity parameters can prevent particle penetration for shocks up to  $\mathcal{M} = 8$ . Most SPH practitioners opt for a value of  $\beta = 2\alpha$  (Morris & Monaghan, 1997).

To determine the correct value of  $\beta$  required for the new scheme, we perform high resolution 3D runs of ram shocks with  $\mathcal{M} = 20$  and  $\mathcal{M} = 100$ . Various values for  $\beta/\alpha$  were used to determine which values prevent particle penetration and allow good smoothing of the shock.

In Fig. 3.30 we plot the  $x$  and  $y$  positions (for all values of  $z$ ) of particles near a  $\mathcal{M} = 20$  shock front. We compare the effect of the  $\beta$  term for the M&M switch and the new scheme. Fig 3.31 shows the same results for a  $\mathcal{M} = 100$  shock. We observe that both viscosity methods suffer from particle penetration when  $\beta/\alpha = 0$  for the  $\mathcal{M} = 20$  shock. The new method however does significantly better as particles remain ordered around the shock with only one layer of particles inter-penetrating. The M&M method causes particle positions around the shock to become highly disordered with many layers of particles inter-penetrating. We also find that with the  $\mathcal{M} = 100$  shock there is no particle penetration with the new scheme, even with

$\beta/\alpha = 0$ . The improvement by the new scheme is likely caused by the viscosity triggering earlier and being increased faster via the jumping method. We find that the suggested value of  $\beta/\alpha = 2$  for the M&M switch is sufficient to stop particle penetration, in fact we found that in these two cases a value of  $\beta/\alpha = 1$  was also sufficient. The new scheme also prevents particle penetration when using  $\beta/\alpha = 1$ .

Fig. 3.32 compares the density, velocity, internal energy and viscosity profiles for the  $\mathcal{M} = 100$  shock. The new scheme used  $\beta/\alpha = 1$  whilst the M&M scheme used the suggested  $\beta/\alpha = 2$ . We find good agreement between the two methods, the shock is smoothed over the same number of particles and the viscosity parameter reaches approximately the same peak value. The new scheme is able to detect the shock earlier, indicated by the earlier increase in the viscosity parameter compared to the M&M method. When compared to the analytical solution we find that the shock has propagated further than expected and that the post-shock density is under estimated. It should be noted that the reason for this is clearly not viscosity related as both schemes cause the shock to be located in the same position. This problem seems to be an SPH inherent problem, simulations of strong shocks by Price & Monaghan (2004)<sup>15</sup> and Bate (1995) also show a similar effect, the shock seems to propagate slightly too quickly for higher Mach number shocks. Furthermore we find that this overshoot does not occur in 1D (as indicated by Fig. 3.29) or when 3D glass initial conditions are used. Therefore it seems likely that the overshoot is caused by taking a FCC lattice for the initial conditions.

### 3.7.4 Evrard Test

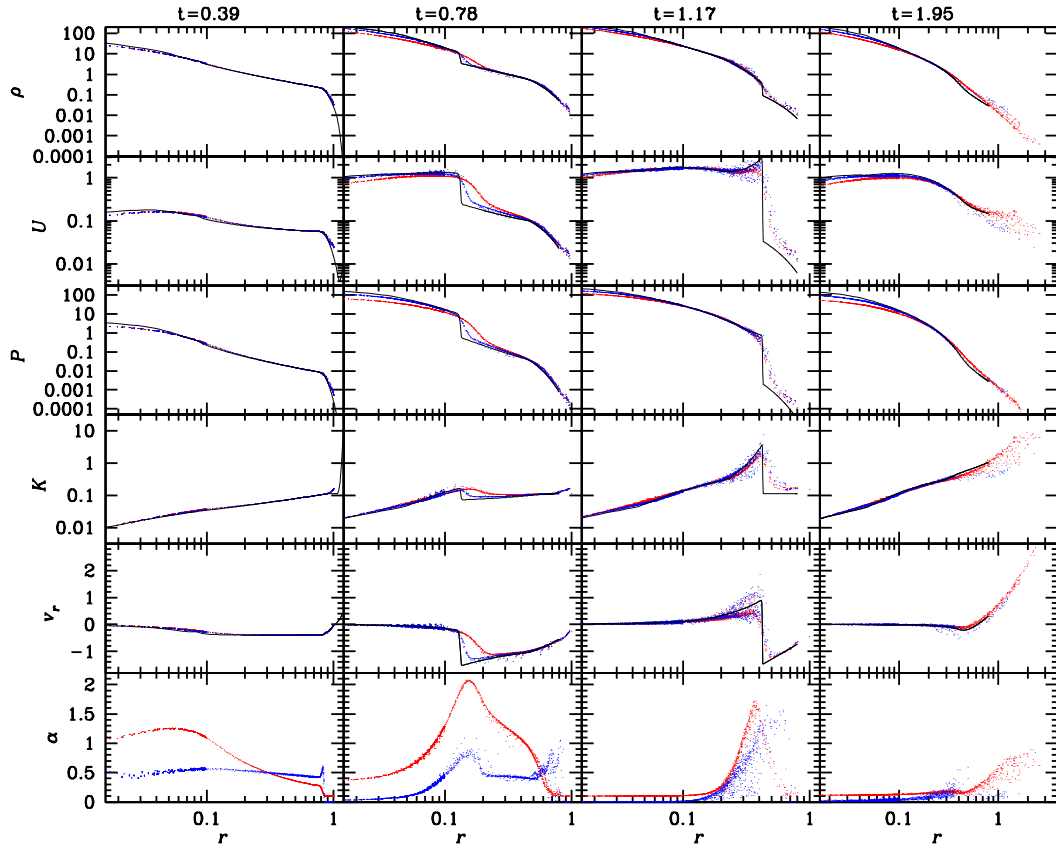
In this test the inward gravitational pull of a gas cloud exceeds its outward pressure force causing the cloud to collapse under its own self-gravity. The initial conditions consist of a gas sphere with density profile (Evrard, 1988)

$$\rho(r) = \frac{M}{2\pi R^2} \frac{1}{r} \quad (3.67)$$

for  $r < R$  and  $\rho = 0$  for  $r > R$ . Initially the gas is at rest and has constant specific internal energy  $u = 0.05 GM/R$ , which corresponds to a virial ratio  $-2U/W = 0.075 \ll 1$ . The initial gravitational inward pull is the same at each radius, while the pressure forces decline towards the outer parts, leading to collapse and, as a consequence, the formation of a shock front at  $t \approx 0.8$ . This shock steepens and

---

<sup>15</sup>see figure 4



**Figure 3.33:** *The Evrard test (see text for the initial setup): shown are various physical quantities and  $\alpha$  at different times for SPH simulations with  $N = 10^5$  particles using our new viscosity scheme, including limiter (blue) or the M&M method (red). Also shown (black) are the results from an adaptive mesh grid code (*Ramses* with up to 262144 cells). Not every particle or grid cell is plotted.*

by  $t \approx 1.2$  evolves into a strong shock at  $r = 0.2$ , propagating outwards as more incoming material joins the jam. Even though the problem is initially spherically symmetric, the SPH realisation of initial conditions cannot be exactly spherically symmetric and the system may well evolve away from sphericity, for instance driven by dynamical instabilities.

In our tests we use a unit system such that  $G = R = M = 1$  and represent the cloud by 100280 SPH particles, initially placed in densest sphere packing (face-centred-cubic grid) and radially stretched to match the density profile. Fig. 3.33 compares the simulation results for our method, the original M&M method, and a 1D calculation by Steinmetz & Mueller (1993) using the piece-wise parabolic method (PPM).

At early times ( $t = 0.39$ , left column) the results from all three methods match

Simulation	M&M	New naive	New $\text{tr}(V)$ + Lim	New $\text{tr}(A - V^2)$ No Lim	New $\text{tr}(A - V^2)$ + Lim
2D Kep disc at $t = 10$	1	2.24	1.08	2.61	0.78
Polytrope Osc at $t = 100$	1	1.24	1.42	1.60	1.38
Poly + Point at $t = 100$	1	1.28	1.50	1.63	1.44
3D sod	1	0.93	1.40	1.70	1.45
3D Ram $\mathcal{M} = 20$	1	1.64	3.35	1.62	1.77
Evrard Test	1	0.93	1.38	1.46	1.17

**Table 3.2:** A comparison of the computational cost for various viscosity methods relative to the M&M method.

very well, but the M&M scheme already shows a large viscosity. At later times a shock forms (at  $r \approx 0.13$  by  $t = 0.78$ ), which moves outwards until it reaches the end of the sphere, when a significant fraction of the gas still has outwards velocities (by  $t = 1.95$ ). The most obvious difference between the two SPH schemes is the amount of (artificial) dissipation: the M&M method is much more viscous, reaching  $\alpha \approx 2$ , resulting in significant over-smoothing of the shock front by  $t = 0.78$  accompanied by unphysical pre-shock heating as visible in the entropy ( $K$ ) profile. Our new scheme performs much better, and agrees well with the 1D calculation, in particular in the inner (post-shock) regions. Note that with our new method  $\alpha$  peaks well before the shock arrives (at  $t = 1.17$ ), while for the M&M method the peak in  $\alpha$  appears actually slightly after the shock.

We should note that we found this to be a valuable test as early versions of our scheme tended to be far too viscous, while our final version passes this test ahead of the M&M switch. Standard SPH (not shown in the figure) shows similar results, though the shock at  $t = 0.78$  appears less smoothed than with the M&M method but more smoothed than with the new scheme.

### 3.8 Summary of New Viscosity Method

The new method, described in §3.4, improves upon the method of Morris & Monaghan (1997) in four important ways.

- We no longer use a non-zero lower viscosity level  $\alpha_{\min}$ , and instead allow  $\alpha_i = 0$  away from shocks, effectively modelling the fluid as inviscid.
- We use  $D_t(\nabla \cdot \mathbf{v}) < 0$  rather than  $\nabla \cdot \mathbf{v} < 0$  as shock indicator. This allows us to distinguish pre-shock from post-shock regions. It also gives a much better



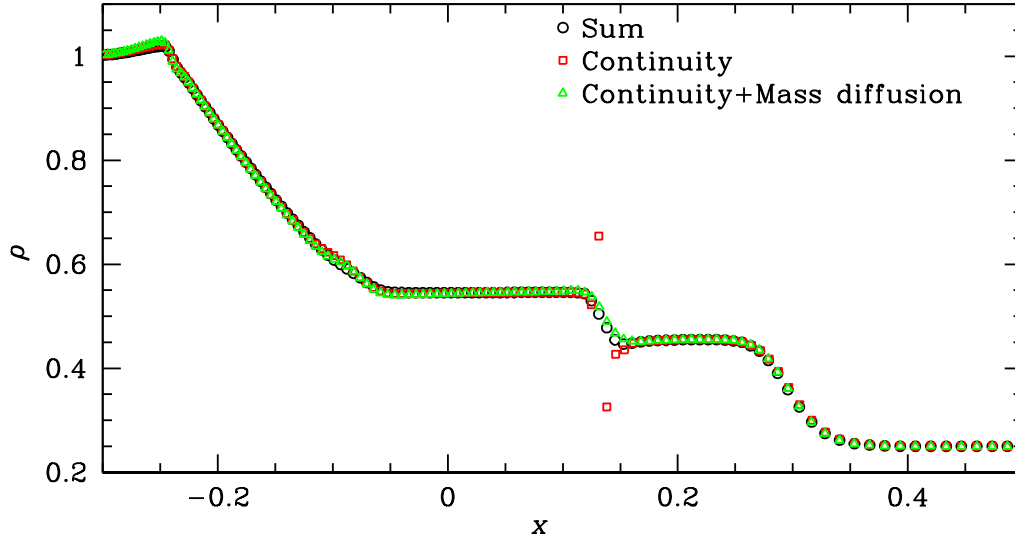
discrimination between converging flows and weak shocks.

- Rather than increasing  $\alpha_i$  by integrating a differential equation, we set it directly to an appropriate local value  $\alpha_{\text{loc}}$ . Together with the improved shock indicator, this guarantees that the viscosity peaks well *before* an incoming shock front.
- We use an improved estimate for  $\nabla \cdot \mathbf{v}$  and  $D_t(\nabla \cdot \mathbf{v})$  and employ a limiter to avoid viscosity being driven by shear-induced errors.

Together these novelties result in a much improved artificial viscosity method. In §3.7, we demonstrate convincingly that with this new technique SPH can resolve strong shocks at least as well, if not better, than any previous SPH technique for artificial viscosity, while §3.6 shows that it is far superior in the regime of convergent flows and weak shocks. In particular, with our scheme SPH can model adiabatic oscillations, such as sound waves or stellar pulsations, undamped over many periods. This was not possible with any previous SPH implementation, which all required a significant increase in resolution just to suppress adverse effects of artificial viscosity. Thus, our new method allows one to choose the resolution based on physical rather than numerical considerations. We also demonstrate in §3.5 that our new scheme is able to detect and damp away noise in the velocity field without the need for a minimum value for the viscosity.

Table 3.2 shows the computational cost of the new scheme compared to the M&M method. We find that the new scheme using the most accurate estimate for  $D_t(\nabla \cdot \mathbf{v})$  and the limiter is on average  $\approx 1.4$  times slower than the M&M scheme. However, little work has been done in optimising these algorithms, therefore it is likely that an improvement on this can be made. Whilst this may seem expensive, we point out that for most problems the naive estimate for  $\nabla \cdot \mathbf{v}$  will suffice to estimate  $D_t(\nabla \cdot \mathbf{v})$ . Also as a proxy for the shear component of the velocity gradient, one may use the curl of the velocity field instead. This would avoid matrix computations, reducing the computational cost. However, in simulations of strong shear flow, such as the Keplerian ring test, the higher order estimates are required with the limiter to minimise adverse effects of artificial viscosity. Furthermore the knowledge of the velocity gradient matrix  $\nabla \mathbf{v}$  is useful for implementing a physical viscosity and the matrix  $\underline{T}$  is useful for correcting errors in other quantities such as the momentum equation (Read et al., 2010).





**Figure 3.34:** *The effect of using the summation equation for the density vs. the continuity equation for the density on the Sod shock tube test. We also compare the effect of artificial mass diffusion.*

### 3.9 Missing Discontinuities - Artificial Mass Diffusion

In numerical methods based on Eulerian grids, a distinction is made between finite difference and finite volume methods. The difference between the two methods is which form of the Euler equations are discretised, the differential or integral form. The main practical difference is that with the differential form it is assumed that the fluid variables may be differentiated, however this is not the case at discontinuities. In the integral form no assumption is made about the ability to differentiate fluid variables.

The SPH analogy to this can be seen by considering the density equation. Using the summation equation is equivalent to taking the integral form of the continuity equation as we are simply discretising the convolution integral of the density with the kernel. The SPH continuity equation is the differential form as we use the SPH gradient operator to calculate  $\nabla \cdot \mathbf{v}$  which is used to calculate  $d\rho/dt$ . By comparing the density summation equation and the continuity equation for SPH, Price (2008) found that the two forms are not equivalent,

$$\frac{d}{dt} \sum_j m_j W_{ij} = \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla W_{ij} + \int [\rho' \mathbf{v}' W] \cdot d\mathbf{s}. \quad (3.68)$$

The integral form of the continuity equation contains not only the differential form of the continuity equation but also an extra surface integral term. This surface integral vanishes in most cases but does not at discontinuities. Therefore if we only take the differential form we lose the information contained by the surface integral at density discontinuities. This can be observed as a blip in the density profile at a contact discontinuity<sup>16</sup>. The effect this has on a simulation can be seen in Fig. 3.34. At the contact discontinuity there is a blip in the density profile when the continuity equation is used. The continuity equation has assumed that the density is differentiable which is not the case at the contact discontinuity. However, the summation equation takes these surface terms into account by smoothing the density. This type of problem is essentially the same problem that occurs at shocks. The density changes *almost* discontinuously over the mean free path of the fluid, which the SPH scheme can not resolve. Therefore the density appears discontinuous, and so the continuity equation loses information as it assumes that the density is always differentiable. This assumption is presumably true in nature as there are no discontinuities in nature. In SPH the standard approach is to smooth away the discontinuity as happens in nature only over a much smaller scale. In the case of shocks this is done by artificial viscosity and in most SPH implementations which use the density summation the density is automatically smoothed by the kernel. However, with the continuity equation this smoothing is not done automatically as the Euler equations model macroscopic processes on a scale much larger than the mean free path. This is the same problem that occurs with shocks, the Euler equations do not directly model the processes that allow entropy to be produced that occur at the mean free path scale. However, the solution to the problem is the same, we simply need to remove discontinuities from the SPH scheme.

A simple correction to this was suggested by (Monaghan, 1997) to force smoothing to occur when using the continuity equation

$$\left. \frac{d\rho}{dt} \right|_{\text{diss}} = - \sum_j m_j \frac{\alpha_\rho (C_i + C_j - \mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij})}{\bar{\rho}_{ij}} (\rho_i - \rho_j) \hat{\mathbf{r}}_{ij} \cdot \frac{\partial W_{ij}}{\partial \mathbf{r}_i}. \quad (3.69)$$

The results of this artificial mass diffusion can be seen in Fig. 3.34. The blip that was present when using the continuity equation is removed by applying this smoothing. However, as the density summation equation does this automatically without the

---

<sup>16</sup>The viscosity acts to smooth out the density automatically at shock fronts and so the problem is not observed at shocks.

need for any artificial terms, this method of evolving the continuity equation is rarely used.

### 3.10 Artificial Thermal Conductivity

The SPH equations for momentum and energy are derived using a differential form of the continuity equation. Therefore the velocity and internal energy will experience similar problems at shock fronts as the density does at contact discontinuities when using the continuity equation. The artificial viscosity takes care of the problem with the velocity as it acts to remove discontinuities in velocity meaning that the velocity is no longer discontinuous. However, the internal energy will experience a blip at the contact discontinuity, this is the well known “wall heating” effect. This occurs as we are taking the differential form of the Euler equations leading to a loss of information at discontinuities.

The effect of information loss at discontinuities is most apparent for the internal energy. The reason for this is that most SPH implementations use the density summation equation (which is the integral form) and so no information is lost. The summation equation also naturally smooths out any discontinuities in the density field, alleviating the problem completely. Most implementations also include some form of artificial dissipation which acts to smooth out discontinuities in the velocity. However no treatment is usually taken for the internal energy and it is not possible to derive an integral form of this equation. The reason for not including artificial thermal conductivity is that often many problems can still be solved with good accuracy without any thermal conductivity. This is not the case with the omission of artificial viscosity, which leads to problems of particle penetration in shocks. It was also unknown that discontinuities in the SPH fluid properties causes information to be lost when using the differential forms of the Euler equation. It was not until problems with Kelvin-Helmholtz type simulations were found that the need for conductivity really came to light (Agertz et al., 2007; Price, 2008).

There have been a few suggestions for the form thermal conductivity should take, each of which are described below.

#### 3.10.1 Monaghan (1997) Conductivity

Monaghan (1997) made a comparison of the form of artificial viscosity applied by

SPH and that inherently built into a Riemann solver. Based on this analysis it was discovered that Riemann solvers also have a built in mass diffusion and thermal conductivity. The SPH equivalent to this is given by

$$\left. \frac{du}{dt} \right|_{\text{diss}} = \begin{cases} - \sum_j m_j \frac{\alpha^u v_{\text{sig}}}{\bar{\rho}_{ij}} (u_i - u_j) \hat{\mathbf{r}}_{ij} \cdot \overline{\nabla W_{ij}} & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.70)$$

where the signal velocity is given by

$$v_{\text{sig}} = C_i + C_j - \mathbf{v}_{ij} \cdot \hat{\mathbf{r}}_{ij}. \quad (3.71)$$

### 3.10.2 Rosswog & Price (2007) Conductivity

Rosswog & Price (2007) allowed each particle to have its own individual parameter in the thermal conductivity equation

$$\left. \frac{du}{dt} \right|_{\text{diss}} = - \sum_j m_j \frac{\alpha^u v_{\text{sig}}}{\bar{\rho}_{ij}} (u_i - u_j) |\overline{\nabla W_{ij}}|. \quad (3.72)$$

The parameter is evolved in a similar way to the viscosity parameter

$$\frac{d\alpha_i}{dt} = -\frac{\alpha_i}{\tau} + 0.1 h_i |\nabla^2 u_i|, \quad (3.73)$$

where the Brookshaw type second derivative is used (Brookshaw, 1985)

$$\nabla^2 u_i = 2 \sum_j m_j \frac{u_i - u_j}{\rho_j} \frac{|\overline{\nabla W_{ij}}|}{x_{ij}}. \quad (3.74)$$

This method is very similar to the method proposed by Monaghan (1997), the difference being that this method allows for a time-varying conductivity and does not take into account the sign of  $\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}$ .

### 3.10.3 Price (2008) Conductivity

Price (2008) based his conductivity term by trying to apply it to regions where the smoothing in density and discontinuity in internal energy cause the pressure to become discontinuous. Therefore his switch was based on the local change in

pressure

$$\left. \frac{du}{dt} \right|_{\text{diss}} = - \sum_j \frac{m_j}{\bar{\rho}_{ij}} \sqrt{\frac{|P_i - P_j|}{\bar{\rho}_{ij}}} \alpha^u (u_i - u_j) \hat{\mathbf{r}}_{ij} \cdot \nabla W_{ij}. \quad (3.75)$$

This smoothes out differences in pressure which are often seen at contact discontinuities. These occur as the density is smoothed by the kernel but the internal energy remains discontinuous, leading to anomalies in the pressure when computed from an equation of state.

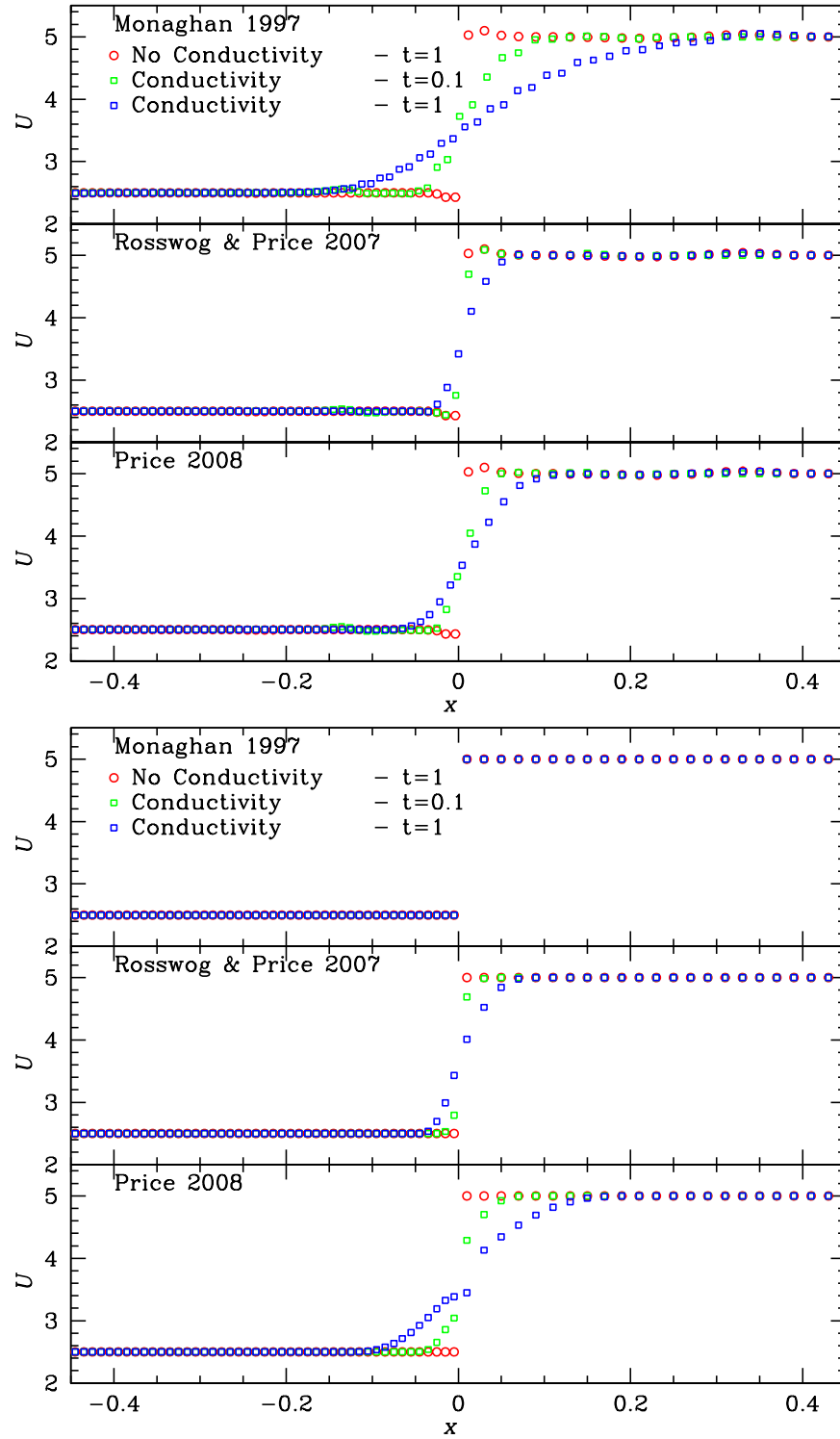
### 3.10.4 Problems With Existing Methods

All of the above methods have a significant draw back, they will all conduct thermal energy whenever there is a difference in energy between two particles (with Price, 2008, a pressure difference is also required, however this is the case in rarefaction waves). Suppose we have a region of constant pressure which contains a density discontinuity, a typical situation in the Kelvin-Helmholtz instabilities. The density discontinuity gives rise to an internal energy discontinuity at the interface. The force between particles at the interface should be exactly zero as there is no pressure gradient in this region. Due to errors in the SPH momentum equation, the particles are accelerated causing particles with low energy to further decrease their energies and particles with high energies to increase their energies. Note that this effect does not occur when the SPH accelerations are forced to be zero (see bottom plot of Fig. 3.35 ), the energies are unchanged. This is because the internal energies only change when there is a non zero  $\nabla \cdot \mathbf{v}$  as indicated by equation (2.2c).

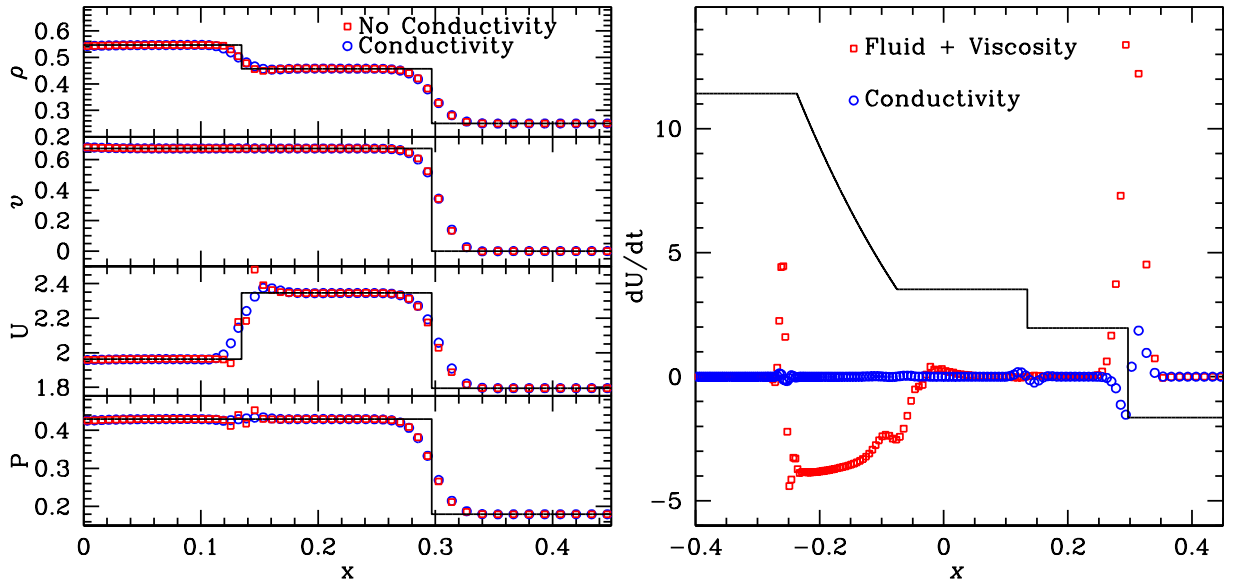
This type of discontinuity in the initial conditions is a poor choice for SPH as it is well known SPH is not well suited discontinuities, as we have discussed previously. Furthermore, as the density is smoothed by the kernel as soon as the simulation starts, this leads to a blip in the pressure, further exacerbating the problem. As well as smoothing the initial conditions to help alleviate this problem, thermal conductivity can be used to smooth away this discontinuity.

In general it is not enough to smooth the initial conditions as these types of discontinuities may form from other fluid interactions well after the simulation has started. In this case it will not be possible to smooth away the internal energy discontinuity unless a thermal conductivity is used.

The various methods outlined above all over smooth this discontinuity leading



**Figure 3.35:** *The effect of the conductivity methods of Monaghan (1997), Rosswog & Price (2007) and Price (2008) on a standing density discontinuity. The top plot shows standard SPH, the bottom plot shows a scheme where the accelerations are forced to zero.*



**Figure 3.36:** *Left:* The shock and contact for the Sod shock tube shown at  $t = 0.2$ . *Right:* The contributions from the fluid equations, viscosity and conductivity to  $du/dt$ .

to a loss of the contact. This is illustrated by Fig. 3.35 which shows how the energy profile changes across the discontinuity as the simulation progresses. The conductivity constantly smooths away the internal energy over much more than two smoothing lengths. This is particularly severe for the Monaghan (1997) method but also occurs in the other methods. However the Monaghan (1997) approach has the advantage that if the SPH acceleration are correct (i.e. zero) then no conductivity is applied.

### 3.10.5 Where Should Conductivity be Applied?

We may consider the 1D Sod shock tube test, which contains a rarefaction wave, shock wave and contact wave, to determine regions of the flow that require thermal conductivity. Internal energy is only changed whenever there is a velocity divergence (or convergence), and whenever artificial viscosity is applied. As a consequence the 1D shock tube only has a changing internal energy at shock fronts and in the rarefaction wave. There is no change in the internal energy at a contact discontinuity (see Fig. 3.36). Therefore the blip at the contact must occur during the formation of the wave when all three wave states occupied the same space. Furthermore, as the velocity across the contact is constant this implies that once the contact has formed

the same particles make up the contact, i.e. the particles do not move through the contact as they do in the shock.

This is also in agreement with the results of the standing contact wave. Conductivity only needs to be applied when there is motion caused by SPH forces which causes a change in internal energy through  $\nabla \cdot v$ . If the SPH forces are turned off (as they should be for a contact wave) there is no change in internal energy across the contact and therefore no need for thermal conductivity (although the initial conditions should smooth the contact). Therefore this would suggest that we only need to apply conductivity at shocks and not at contact discontinuities. However, this may be misleading as we also require that small blips in pressure be smoothed away, these can also occur at contact discontinuities where a shock is not present.

The Price (2008) method turns on conductivity by looking for differences in pressure. This not only causes conductivity to be turned on at shocks and contacts, it is also active in rarefaction waves (see Fig. 3.36).

### 3.10.6 Towards an Improved Conductivity Scheme

We have highlighted two problems with the widely used conductivity scheme as described by Price (2008). Firstly, at contact discontinuities the energy is continuously smoothed leading to a loss in sharpness of the contact wave. Secondly we have found that conductivity only needs to be applied at shocks and possibly at contact discontinuities, but is not required at rarefaction waves. Also in astrophysical simulations we may well have pressure gradients which are supporting gravitational forces. In this case thermal conductivity is again not required but the Price (2008) method would introduce a conductivity.

Furthermore the strength of conductivity, as controlled by the parameter  $\alpha^u$  must be set at the start of the simulation. Ideally we would like a scheme where the strength is adapted to the situation, in a similar way to the viscosity methods described previously. Therefore we can propose two alternative schemes.

#### A Modified Price Conductivity

A simple method to adjust the strength of conductivity is to set

$$\alpha^u = f_s \alpha^{\text{vis}} \quad (3.76)$$



whilst still using the conductivity given by Price (2008) (see equation 3.75). The parameter  $f_s$  is a constant that sets the scaling between the strength of the conductivity and strength of artificial viscosity. This allows conductivity to increase according to the strength of the shock and turns off the conductivity in rarefaction waves. It also allows the smoothing of the contact as the particles at the interface initially receive a small velocity divergence. As the velocity is damped away the viscosity will decrease, thus decreasing the conductivity. This will allow the preservation of the contact wave. We label this scheme “Price TV” (TV stands for time-varying).

### A Viscosity Inspired Conductivity

As we have seen previously, thermal conductivity does not need to be applied at contact discontinuities which are already smoothed by the SPH scheme. Rather it needs to be applied in shock situations, and this is done already by the method proposed by Price (2008) (see Fig. 3.35). As this is also the case for artificial viscosity we will consider a viscosity inspired conductivity switch.

To construct such a switch we can make comparisons to the viscosity switch and try to apply the same methods used to construct the conductivity switch. The viscosity used a functional form given by

$$\left. \frac{du}{dt} \right|_{AV} = \frac{1}{2} \sum_j m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \nabla W_{ij}. \quad (3.77)$$

This causes the velocity to be smoothed between particle pairs generating heat energy. Therefore we can construct a similar term for conductivity

$$\left. \frac{du}{dt} \right|_{\text{cond}} = \sum_j m_j \Pi_{ij}^{\text{cond}} u_{ij} \hat{\mathbf{r}}_{ij} \cdot \nabla W_{ij}. \quad (3.78)$$

This causes the energy between two particles to be smoothed by attempting to minimise the difference in energies between the two particles. If the particles are of the same energy then there is no dissipation. This is similar to the method proposed by Monaghan (1997) except that we have replaced  $\alpha^u v_{\text{sig}} / \bar{\rho}_{ij} \rightarrow \Pi_{ij}^{\text{cond}}$ .

In both viscosity and conductivity the strength of the effect is controlled by  $\Pi_{ij}$ . In the case of viscosity,  $\Pi_{ij} = 0$  whenever particles are receding from each other, otherwise the strength depends on the local velocity difference along the line joining

the two particles. Therefore constructing a conductivity method essentially comes down to finding the correct functional form for  $\Pi_{ij}$ . We are however constrained by the units it must take, in this case  $[m^{-1}][l^4][t^{-1}]$ .

For artificial viscosity the following form was taken

$$\Pi_{ij} = -\bar{\alpha}_{ij} \frac{\bar{c}_{ij} \bar{h}_{ij}}{\bar{\rho}_{ij}} \frac{\mathbf{v}_{ij} \cdot \mathbf{x}_{ij}}{r_{ij}^2}. \quad (3.79)$$

This form is useful as it is resolution dependent, a property that we will require for the conductivity switch. Also it detects if the position vectors of two particles are approaching or receding from each other, i.e. how the vector joining the particles  $\mathbf{x}_{ij}$  is changing with time along itself, that is  $\mathbf{v}_{ij} \cdot \mathbf{x}_{ij} = \mathbf{x}_{ij} \cdot d\mathbf{x}_{ij}/dt$ . This is particularly useful as it only activates viscosity for regions where the flow is converging. We can construct a similar term for conductivity as

$$\Pi_{ij} = \begin{cases} -\bar{\alpha}_{ij}^u \frac{\bar{h}_{ij}}{\bar{\rho}_{ij}} \frac{\mathbf{v}_{ij} \cdot \mathbf{x}_{ij}}{r_{ij}^2} & \mathbf{v}_{ij} \cdot \mathbf{x}_{ij} < 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.80)$$

We may also extend this method (referred to as “CONDV”) to a time-varying conductivity by setting  $\alpha^u = f_s \alpha^{vis}$ .

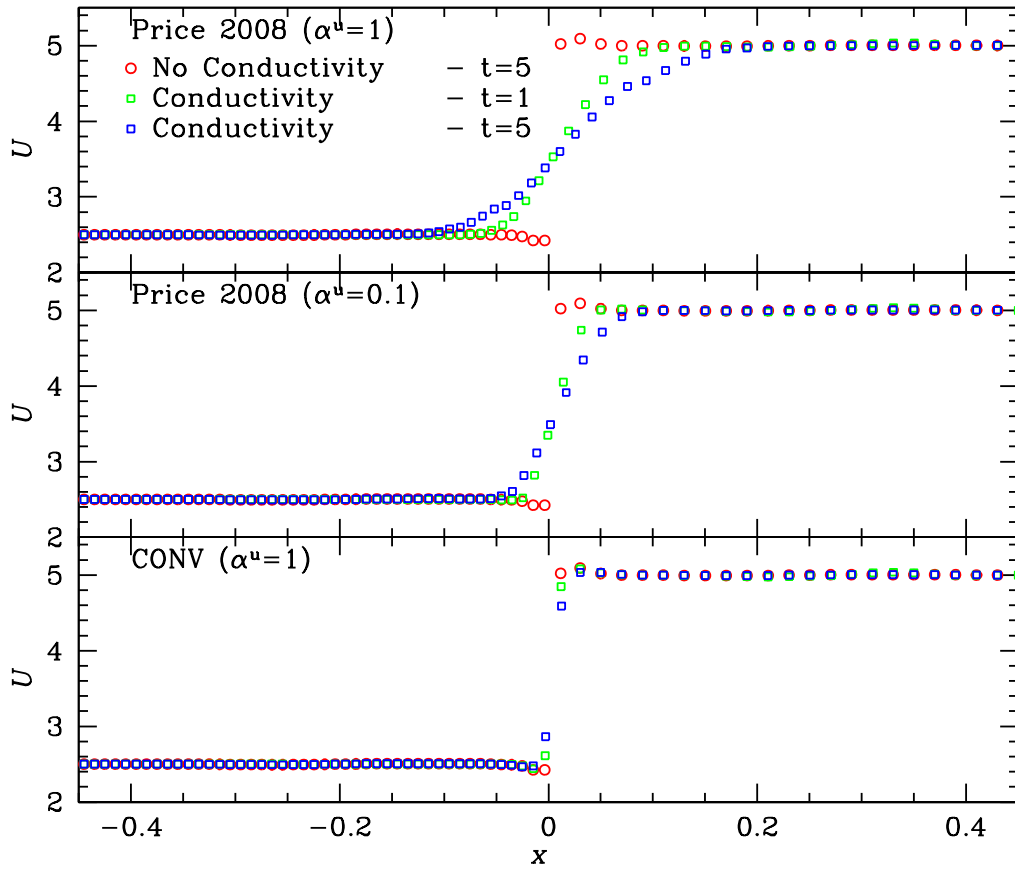
### 3.10.7 Tests With Fixed Strength

In order to determine if the new form of conductivity (“CONDV”) is useful, we can perform a series of tests using a fixed strength of conductivity. That is we set  $\alpha^u$ , equal to some constant at the start of the simulation, in a similar way one would for a fixed artificial viscosity.

#### Density Discontinuity

Fig. 3.37 compares the effect of conductivity on the density discontinuity. The top two plots show the effect of altering  $\alpha^u$  for the Price (2008) scheme. As expected reducing the size of  $\alpha^u$  decreases the region over which the contact is smoothed. Therefore a scheme based on the method of Price (2008) which is able to reduce  $\alpha^u$  on the fly will reduce the over-smoothing of the internal energy differences.

The bottom plots shows the “CONDV” scheme as described by equations (3.77) and (3.79). We used  $\alpha^u = 1$  for all particles and the viscosity of the previous section

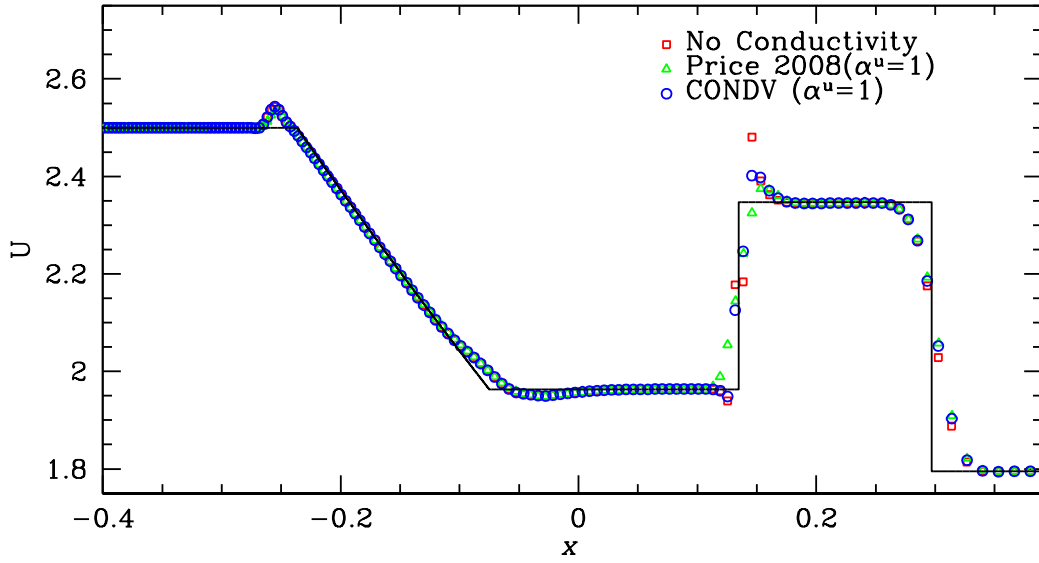


**Figure 3.37:** Comparing effect of Price (2008) conductivity and the scheme proposed by equations (3.77) and (3.79) for a density discontinuity.

was used. This scheme provides only minimal smoothing of the contact, leading to an internal energy profile which is smoothed over only two particles, which in this case is smaller than the size of the smoothing sphere. The reason for this is that the viscosity damps the velocities across the contact, leading to small values of  $\mathbf{v}_{ij} \cdot \mathbf{x}_{ij}$  which suppresses  $\Pi_{ij}$ . Therefore the two approaches described above act to minimise the conductivity using two different methods. The “CONDV” method relies on  $\mathbf{v}_{ij} \cdot \mathbf{x}_{ij}$  approaching zero where conductivity is not required. The “Price TV” method relies on the parameter  $\alpha^u$  approaching zero where conductivity is not required, which is determined by the viscosity method.

### Sod Shock Tube

Fig. 3.38 compares the effect of conductivity on the sod shock tube problem. When conductivity is not applied we observe that there is a spike in the internal energy



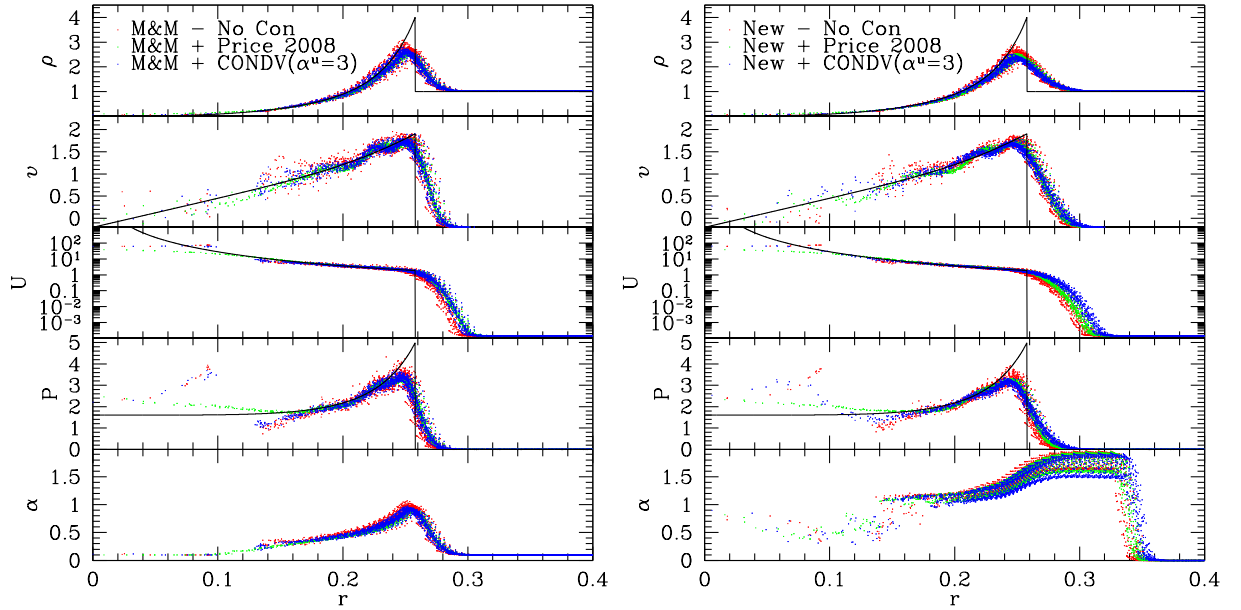
**Figure 3.38:** Comparing effect of Price (2008) conductivity and the scheme proposed by equations (3.77) and (3.79) for the Sod shock tube test.

profile. However applying conductivity (with both methods) smoothes away this spike. The Price (2008) method smooths the contact over four particles where as the “CONDV” method only smoothes over two. The “CONDV” method therefore does not quite provide enough smoothing as ideally we would like the discontinuity smoothed over four particles. However, this can be adjusted by changing the strength of  $\alpha^u$ . Furthermore both methods still lead to a slight increase in the internal energy compared to the analytical result to the immediate right of the contact wave. Again it may be possible to reduce this by both correctly smoothing initial conditions and adjusting the strength of  $\alpha^u$ .

### Sedov Blast Wave

This tests involves concentrating a large amount of energy in a small region to simulate an explosion. This is a particularly difficult test for SPH as very high resolution is required to obtain accurate results. Furthermore, due to the large internal energy discontinuity of the point-like explosion, it is an excellent test for the conductivity scheme. We construct the initial conditions in 2D by creating a  $50 \times 50$  square grid of particles with constant density and internal energy. We then add internal energy given by  $u = E/(\pi h^2 \rho)$  (where  $E = 1$ ) to all particles with  $r < h$ .

Fig. 3.39 shows firstly the differences between the M&M viscosity scheme and the



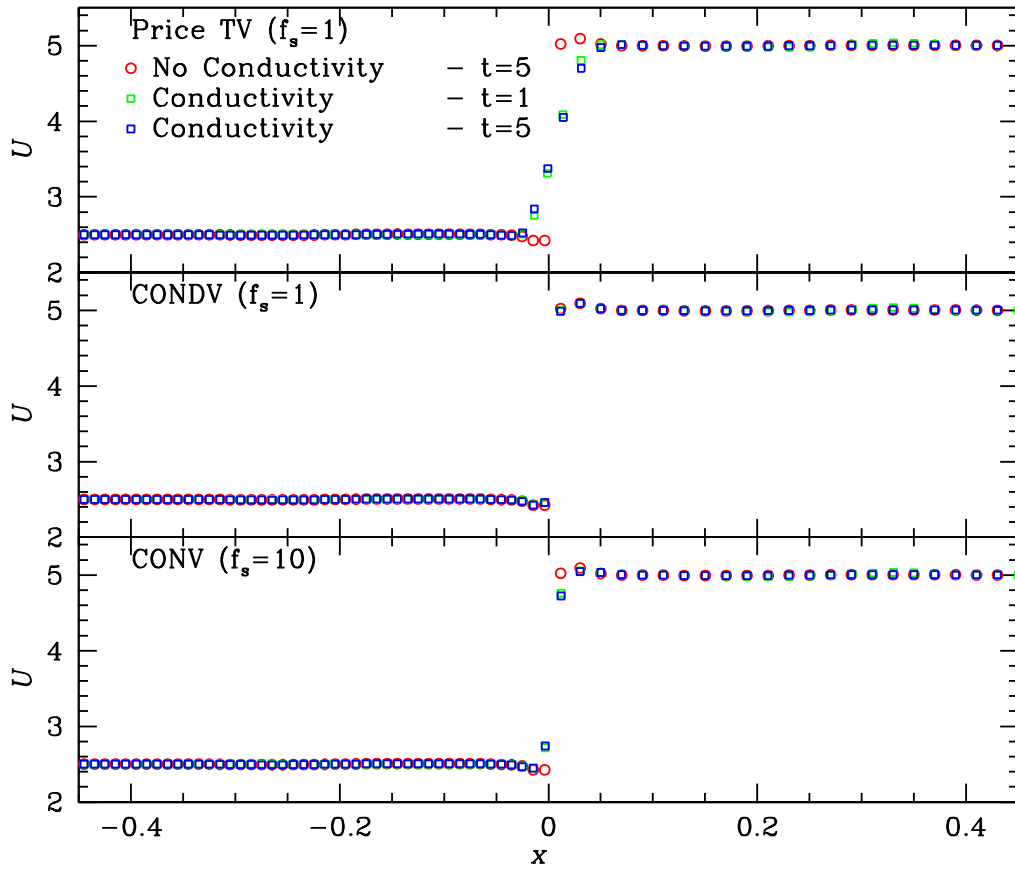
**Figure 3.39:** Comparing effect of Price (2008) conductivity (*left*) and the scheme proposed by equations (3.77) and (3.79) (*right*) for the Sedov blast wave.

new viscosity scheme. We see that in this case there is only a slight reduction in the noise in the profiles with the new viscosity scheme, despite the viscosity parameter being much higher in the new scheme. We again find that the new scheme allows the viscosity to peak in front of the shock rather than after the shock, as is the case for the M&M method.

Fig. 3.39 also compares the differences between conductivity schemes on the 2D Sedov blast wave. We compare no conductivity, the “CONDV” scheme (using  $\alpha^u = 3$ ) and the Price (2008) scheme. Both conductivity schemes show a significant improvement in reducing the noise of the simulation, particularly in the post shock region. In this case the “CONDV” scheme tends to over-smooth the internal energy profile slightly. This is due to the choice of  $\alpha^u$ , a lower value removes this over-smoothing. However, doing this leads to more noise creeping back into the simulation as less smoothing is applied in all regions. Therefore it is likely that with a time-varying  $\alpha^u$  this problem can be addressed.

### 3.10.8 Strength of Conductivity Required

The tests of the previous section have demonstrated the need for a time dependent conductivity parameter  $\alpha^u$ , similar to that of time dependent viscosity. As the



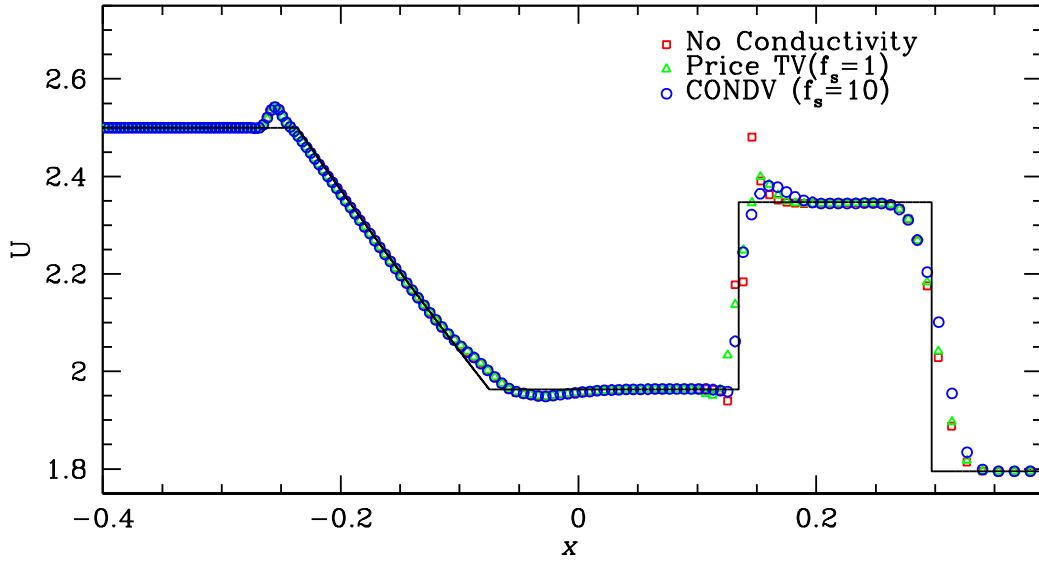
**Figure 3.40:** Comparing effect of Price (2008) conductivity and the scheme proposed by equations (3.77) and (3.79) for a density discontinuity.

conductivity should be triggered at the shock front, we may use the same method for determining  $\alpha^u$  as the viscosity method. However, we may need to change the scaling slightly.

### Density Discontinuity

Fig. 3.40 shows how allowing varying  $\alpha^u$  can affect the density discontinuity. The “Price TV” method is able to prevent the contact becoming over smoothed. The energy profile is only smoothed over 4 particles at both  $t = 1$  and  $t = 5$ . This is because the viscosity is only large initially to smooth away the unphysical velocity kicks, after this the viscosity is reduced, also reducing the conductivity. At this point the velocities are approximately zero and so there is (almost) no conductivity, hence there is no further evolution to the energy profile.

The “CONDV” method on the other hand produces almost no conductivity as



**Figure 3.41:** Comparing effect of Price (2008) conductivity and the scheme proposed by equations (3.77) and (3.79) for a density discontinuity.

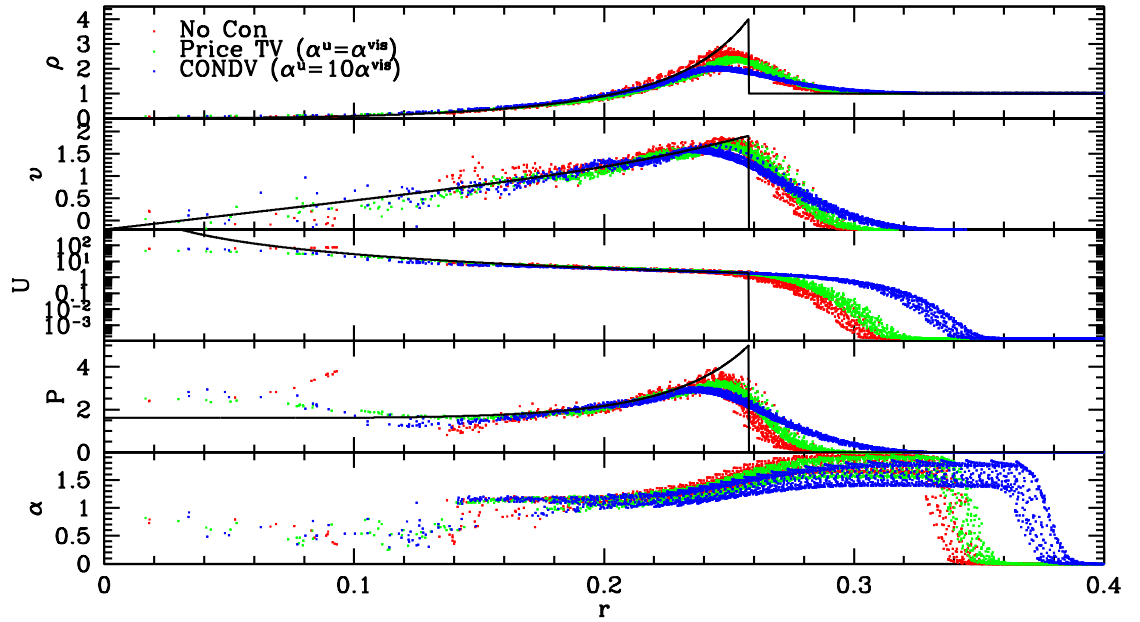
$\mathbf{v}_{ij} \cdot \mathbf{x}_{ij}$  is always small, as is  $\alpha^{\text{vis}}$ . Therefore the conductivity is always small; even with  $f_s = 10$  the internal energy is only very slightly smoothed over two particles.

### Sod Shock Tube

Fig. 3.41 shows how allowing varying  $\alpha^u$  can affect the Sod shock tube. Both the “Price TV” method (with  $f_s = 1$ ) and the “CONDV” method (with  $f_s = 10$ ) are able to smooth away the energy spike. Both methods smooth the internal energy over four particles. However, we still see that there is a slight increase in internal energy to the right of the contact wave.

### 2D Sedov Blast Wave

Fig. 3.42 shows how allowing varying  $\alpha^u$  can affect the Sod shock tube. Both the “Price TV” method (with  $f_s = 1$ ) and the “CONDV” method (with  $f_s = 10$ ) are able to reduce the noise compared to using no conductivity. However, we find that with the “CONDV” method the internal energy is severely smoothed ahead of the shock compared to the run without. Also the peak in density is lowered and occurs slightly after the shock. It seems that “CONDV” with  $f_s = 10$  is too strong for these tests but not quite strong enough for density discontinuities. The “Price TV” method on the other hand only slightly increases the smoothing in the internal



**Figure 3.42:** Comparing effect of a time-varying conductivity on the Sedov blast shock.

energy profile and is able to stabilise the post shock pressures and velocities.

### KH Instability

The Kelvin Helmholtz instability (KHI) arises due to a velocity shear between two media. At the boundary between two media a small disturbance is applied, for the purposes of simulating this phenomena the disturbance is assumed to be sinusoidal. The origin of the instability is due to Bernoulli's principle, fluid flow is faster at the disturbance point giving rise to a pressure difference, causing the disturbance to grow. A full treatment of the KHI is given by Chandrasekhar (1961) and Drazin & Reid (1981).

The ability of SPH to reproduce the KHI has been somewhat of a controversial topic. So called “traditional” SPH has been unable to reproduce the KHI, whereas finite difference methods have been able to handle the KHI with no difficulties. Agertz et al. (2007) suggested that there was a fundamental problem with SPH that prevented the reproduction of the KHI. They suggested that spurious pressure forces at the interface of the two fluids (of differing density) prevented the fluid mixing, thus suppressing the instability. The reason for the spurious pressure forces are partly due to the inconsistency of the SPH initial conditions. The density at the interface



is smoothed but the internal energy is not smoothed and is still discontinuous. Therefore the pressure across the interface of the two fluids will not be constant. Price (2008) demonstrated that the “standard” SPH implementation was incomplete and artificial conductivity should be included. Conductivity acts to smooth the internal energy at discontinuities and therefore removes the spurious increases in pressure.

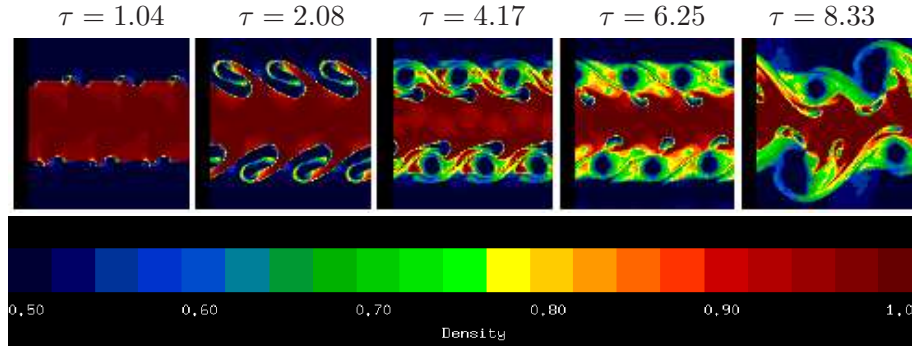
We set up the problem in 2D using equal mass particles in a periodic box of size  $(\pm 0.5, \pm 0.25)$ . The density is set such that  $\rho = 1$  for  $|y| < 0.125$  and  $\rho = 0.5$  for  $|y| > 0.125$ . The particle spacing in the densest region is  $\delta x = 1/900$  giving a total of 303624 particles. The regions are initially in pressure equilibrium with  $P = 2.5$ . We use an ideal gas equation of state with  $\gamma = 5/3$  and the internal energy is set using the smoothed density to ensure the pressure is constant across the contact. We set a shear flow in the  $x$  direction such that  $v_x = 0.5\mathcal{M}$  for  $|y| < 0.125$  and  $v_x = -0.5\mathcal{M}$  for  $|y| > 0.125$ . We then seed the instability by applying a small perturbation in the  $y$  direction such that  $v_y = A \sin(2\pi(x+0.5)/\lambda)$  where  $A = 0.025$  and  $\lambda = 1/6$ . The KH timescale is given by

$$\tau_{\text{KH}} = \frac{\lambda}{|v_{x,1} - v_{x,2}|} \frac{\rho_1 + \rho_2}{\sqrt{\rho_1 \rho_2}}, \quad (3.81)$$

which for the setup described above gives  $\tau_{\text{KH}} = 0.144$ . Fig. 3.43 shows the results of a simulation with **Ramses** using a fixed  $512 \times 512$  grid. Each grid cell had a size of  $9.8 \times 10^{-4}$  whereas in the SPH simulation the smoothing length was  $2.1 \times 10^{-3}$  and  $3.1 \times 10^{-3}$  in the high and low density region respectively. Therefore the resolution in the grid code is slightly higher. We find that in the grid code the  $\lambda = 1/3$  mode has become prominent by  $\tau = 2$  and the  $\lambda = 1/2$  mode can be observed by  $\tau = 8$ .

Fig. 3.44 compares the effect of thermal conductivity on the KH instability. All seven runs show similar growth times of the velocity perturbations with the growth in the  $\lambda = 1/3$  mode occurring around  $\tau = 2$  and the growth in the  $\lambda = 1/2$  mode occurring by around  $\tau = 6$ , this is in good agreement with the findings of Price (2008) and that of the grid code.

The top three rows show runs without any thermal conductivity. We see that although there is growth of the velocity perturbation the fluid does not appear to mix, leading to blobs and bubbles. However, with the new viscosity scheme the instability is able to grow slightly better, distinct rolls can be seen starting to form, although they do not appear smooth.

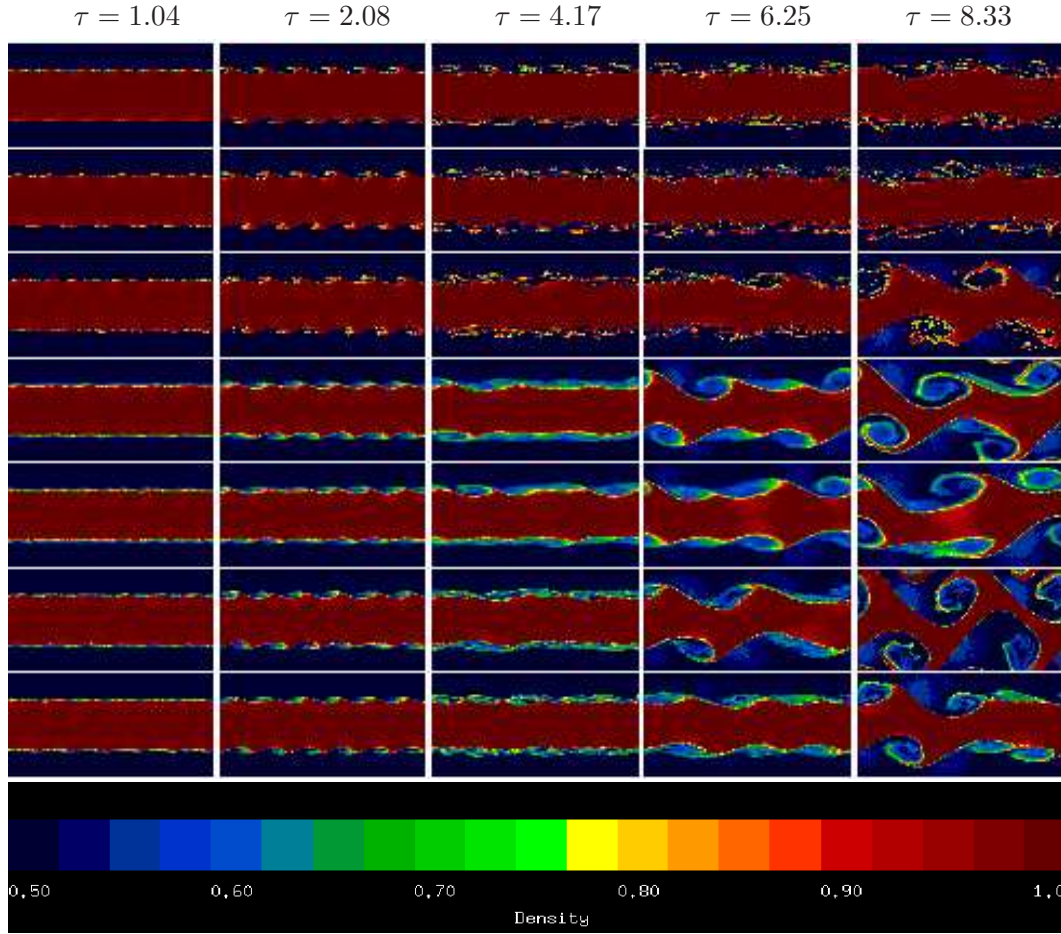


**Figure 3.43:** The KH instability using the grid code *Ramses* with a fixed resolution of 262144 grid cells.

The 4<sup>th</sup> and 5<sup>th</sup> rows show the M&M and new viscosity schemes with Price (2008) conductivity. We see an immediate improvement as the fluid is able to mix. The characteristic Kelvin “cats-eyes” can be observed at  $\tau = 8$  for the  $\lambda = 1/2$  mode for both viscosity methods. The bottom two rows show the new viscosity scheme with an adaptive  $\alpha^u$ ; the sixth row using the “CONDV” scheme with  $\alpha^u = 10\alpha^{\text{vis}}$  whereas the bottom row uses the “PRICE-TV” scheme with  $\alpha^u = \alpha^{\text{vis}}$  (both with the new viscosity scheme). We see that with both alternative methods of conductivity the “cats-eye” features are well resolved, showing similar properties to the Price scheme (row 5). We notice that the  $\lambda = 1/3$  mode is resolved best with the new viscosity scheme and “PRICE-TV” conductivity (seventh row). Clearly defined rolls can be observed, similar to the *Ramses* simulation, only slightly smaller. We also see the beginning of secondary rolls which are also seen in the *Ramses* simulation.

### 3.11 Conclusions

In this chapter we have discussed in detail the need for dissipative processes in the conservative SPH scheme. The most common form of artificial dissipation in SPH is artificial viscosity, which has been widely used in SPH for many years. The aim of artificial viscosity is to allow the generation of shocks by transferring entropy between SPH particles. This is done by removing kinetic energy and converting it into thermal energy in order to mimic the collisional behaviour of a fluid that occurs on the mean free path scale. A secondary but important aim of artificial viscosity is to prevent particle penetration and suppress numerical noise. Unfortunately, there is no method to derive the functional form of this dissipation from a Lagrangian



**Figure 3.44:** *Effect of artificial viscosity and thermal conductivity on KH test. Top shows  $\alpha = 1$  with no conductivity. 2<sup>nd</sup> and 3<sup>rd</sup> rows show M&M and the new scheme with no conductivity. 4<sup>th</sup> and 5<sup>th</sup> rows show M&M and the new scheme with Price conductivity. 6<sup>th</sup> and 7<sup>th</sup> rows show the new scheme with the CONDV scheme with  $\alpha^u = 10\alpha^{vis}$  and the PRICE-TV scheme with  $\alpha^u = \alpha^{vis}$*

approach. However, the form most commonly used in SPH (Monaghan, 1992) has been shown to be the SPH version of the linear viscous stress tensor in the Navier Stokes equation (Meglicki et al., 1993).

In grid methods, artificial viscosity is usually implicit in the Riemann solver (Monaghan, 1997) and interpolation methods are employed to effectively suppress artificial viscosity away from shocks. Most SPH simulations to date hardly use such precautions, and artificial viscosity is applied whenever a particle and its neighbour approach each other. The strength of the damping applied is controlled by the parameter  $\alpha$ , which is set at the start of the simulation and fixed for the duration

of the simulation. However this is problematic as it means that viscosity is always turned on with the same strength for converging flows, weak shocks and strong shocks, even though it is known that viscosity is not required for converging flows and a lower strength of viscosity can be used for weak shocks. As a consequence, adiabatic oscillations are damped and shear-flows decelerated.

To address this problem Morris & Monaghan (1997) suggested each particle be assigned its own parameter  $\alpha_i$  which controlled the strength of the viscosity applied. This reduced the default amount of artificial viscosity by an order of magnitude compared to standard SPH practice. In this method, explained in detail in §3.2.3, individual artificial viscosities  $\alpha_i$  are adapted by integrating a differential equation. This allows strong shocks to have more damping than weak shocks and has been shown to give good shock capturing in strong shocks (e.g. Morris & Monaghan, 1997). Despite this significant step forward, this method is only now gaining popularity, with many SPH simulations still using a fixed viscosity.

Though constituting a major improvement, this method is still unsatisfactory because it still damps adiabatic oscillations and over smooths weak as shocks we argued in §3.4 and demonstrated in §3.6. This problem is compounded by the need for a minimum value of  $\alpha_i$  to be present for all particles, leading to a minimum persistent damping that is always present. Removing this floor leads to ringing in the post shock region and in some cases difficulties to resolve the shock.

The new method, described in §3.4, improves upon the method of Morris & Monaghan (1997) allowing SPH to resolve not only strong shocks (§3.7) but also convergent flows and weak shocks (§3.6). This is done by changing the shock indicator from  $\nabla \cdot \mathbf{v}$  to  $D_t(\nabla \cdot \mathbf{v})$  which allows the viscosity to be increased before the shock arrives. This, combined with  $\alpha$  jumping to the correct value rather than integrating, allows shocks to be well resolved. Furthermore, the new indicator is smaller than  $\nabla \cdot \mathbf{v}$  for weakly converging flows (see table 3.1). This combined with  $\alpha_{\min} = 0$  allows weakly converging flows and sound waves to remain undamped after many periods. However, the new scheme still has problems of over-damping in shear flows. This is because the estimate for  $\nabla \cdot \mathbf{v}$  has a small errors due to the anisotropy of the particle distribution. We attempt to eliminate this problem by using a Balsara like limiter, which reduces the viscosity applied in strong shear flows in the absence of a shock.

Another dissipative process required in a conservative SPH scheme is thermal conductivity. This is similar to artificial viscosity in that it smooths away discon-

tinuities in internal energy. Price (2008) showed that without thermal conductivity the pressure field at contact discontinuities can have large errors, leading to a force separating particles at the boundaries inhibiting mixing. However, the method proposed by Price suffers from similar problems to that of traditional artificial viscosity. The strength of the conductivity is characterised by a parameter which is fixed at the start of a simulation. Therefore resolved discontinuities are further smoothed, leading to a loss of resolution.

A possible solution to this problem, as investigated in section 3.10.6, is to allow a time-varying conductivity parameter, similar to that of artificial viscosity. The results of such a scheme were shown to have some limited success in that the diffusion of contact waves was suppressed whilst still applying conductivity where required. However, more testing of such schemes are required.

In more recent years, there has been a move away from these somewhat ad-hoc methods towards schemes such as Godunov-SPH (Cha, 2002), Lagrangian grid codes (such as Springel, 2009), and non-conservative SPH schemes (Read et al., 2010; Abel, 2010). Godunov-SPH and Springel’s Lagrangian-grid hybrid method require no additional dissipation terms as they are based on Riemann solvers. These schemes have been shown to resolve both strong and weak shocks, and are able to resolve mixing. Non-conservative SPH schemes such as (Read et al., 2010) are also able to capture phenomena such as KH instabilities without the need for thermal conductivity. They do however still rely on artificial viscosity to resolve shocks and may also need some thermal conductivity; further testing of the method, which is still in its early stages, is required.

Together with possible improvements to the SPH method itself, such as artificial conductivity (e.g. Price, 2004), better density estimates (e.g. Read et al., 2010), or reduced errors, our new artificial viscosity treatment should eventually result in a SPH method that can compete with any other technique, even in domains where SPH is traditionally thought to be inferior.

---

# 4

## SPH Stability

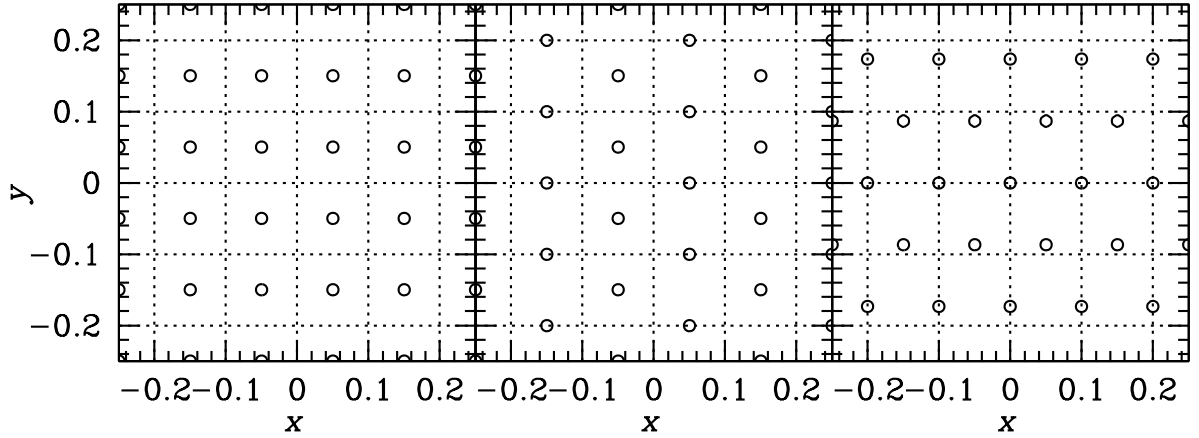
### 4.1 Introduction

As with the majority of numerical codes, it is vital to perform test calculations to ensure that the code functions as desired. The testing phase of a codes development is not only important to remove coding errors (such as forgetting to reset the acceleration to zero before recalculation), it is required to determine the regimes in which the code performs well. Any areas of the code that performs poorly may be identified and possibly remedied by producing improved algorithms or using tricks/approximations to avoid the problem (such as smoothing initial conditions). This not only prevents one from using the code in situations where the code is known not to perform well, it allows an insight into the SPH method in general and where improvements can be made.

The test problems should ideally be simple problems to set up where there are well known analytical solutions. Failing a simple analytical solution one may also look to test problems where a solution can be found using direct or iterative methods. Failing this, some tests without analytic solutions can be compared with other fluid dynamics methods such as grid codes and 1D Lagrangian methods. This allows the method to be compared with alternative SPH methods, grid based codes or any alternative computational fluid method for both speed and accuracy.

We need to ensure that our SPH code is able to capture shocks, allow the prop-





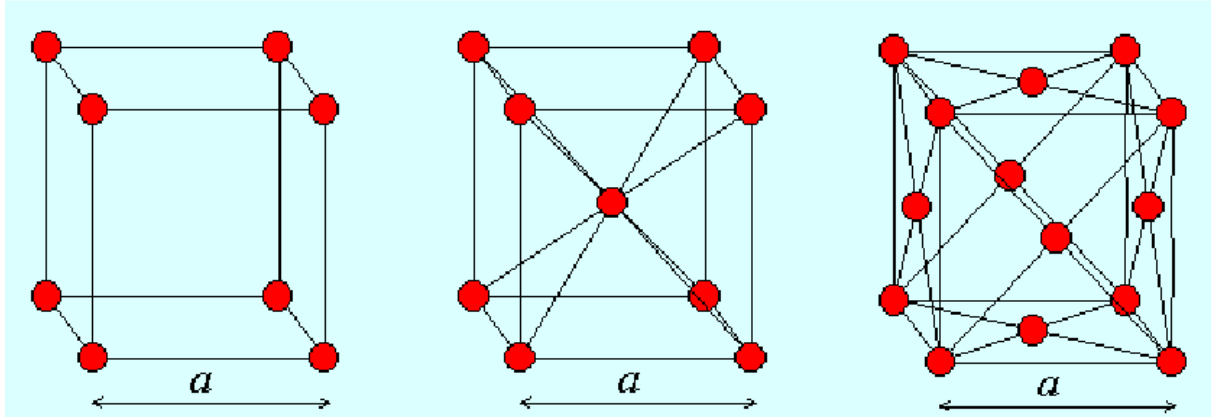
**Figure 4.1:** 2D ordered initial conditions using a grid (*left*), offset grid (*middle*) and hexagonal packing (*right*).

agation of small amplitude sound waves (both of which were dealt with in chapter 3) and allow reproduction of physical instabilities whilst suppressing numerical instabilities.

## 4.2 Initial Conditions in SPH

An often overlooked but critical aspect of SPH is the construction of initial conditions, particularly the initial particle placement. The only constraint on particle placement is to satisfy an initial density profile.

Constant density profiles may be obtained by setting particles up in any number of lattice arrangements. This involves placing particles on vertices or faces of cubes (or squares in 2D and lines in 1D). In 1D particles are simply placed equi-distance from each other, thus creating a constant density profile. In higher dimensions particles are packed onto tiles and these tiles are evenly spaced to fill the domain. In 2D these tiles could be obtained by placing particles on the vertices of squares which are then repeated to fill the domain (see Fig. 4.1). Alternatively one could offset one half of the square in order to produce a filling pattern as shown in the middle panel of Fig. 4.1. One could also use a hexagonal tiling structure (see right panel Fig. 4.1) but these are tricky to construct whilst maintaining periodic boundary conditions as the width of the cell is two units whilst the height of the cell



**Figure 4.2:** 3D cells used for ordered initial conditions. *Left* shows a simple cubic lattice, *middle* shows a body centred cubic lattice and *right* a face centred cubic lattice.

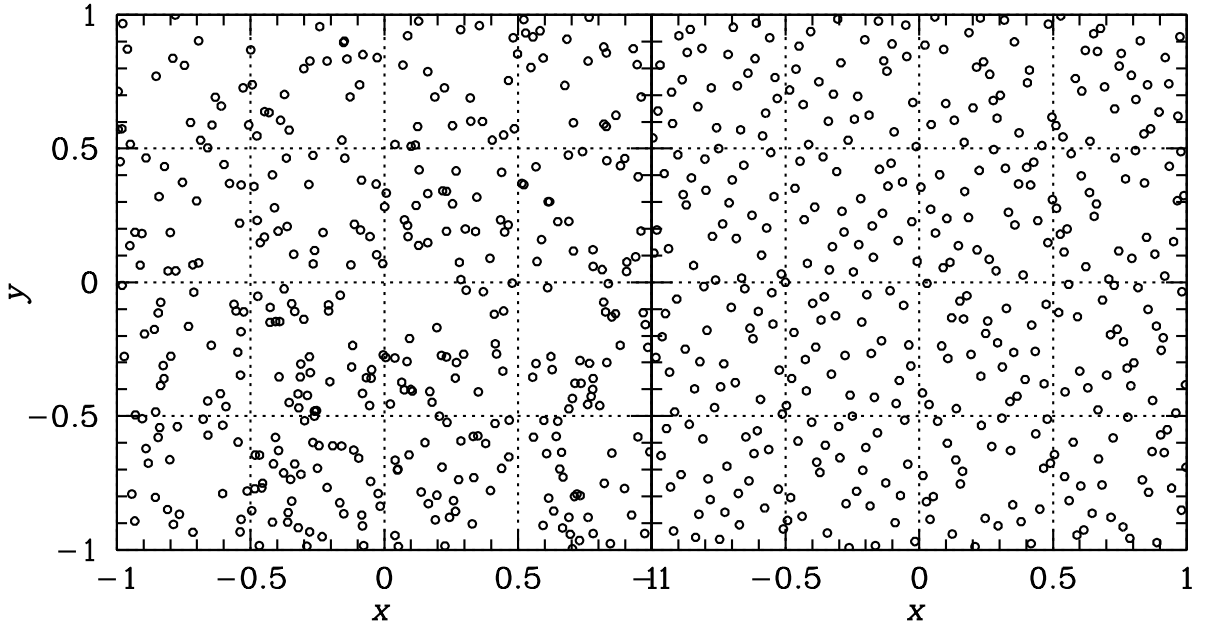
is  $\sqrt{3}$  units. In 3D there are many more possible packing configurations.<sup>1</sup> There are 3 types of cubic lattice which are simple cubic (SC), body centred cubic (BCC) and face centred cubic (FCC). Of these the FCC has the densest packing arrangement (see Fig. 4.2) with 4 particles per cell compared to one and two for SC and BCC respectively.

Ordered particle configurations (those based on lattices) may also be obtained for non constant density profiles. In 1D this essentially means setting the interparticle separation as  $dx \propto 1/\rho$ . In higher dimensions this is equivalent to placing particles on squares/cubes (for example in a face centred cubic lattice) of constant size and then either compressing or expanding the size of the square/cube to fit the required density. Interpolation may be used to allow smooth changes in density, such as those found in the radial density profile of a star.

A widely used alternative to ordered particle packing is to place particles randomly in space. To achieve a uniform density profile one may use a uniform random number generator. As no computer generated sequence of numbers can be truly random (as they are generated by a predefined algorithm), one may use pseudo-random numbers which approximate a truly random sequence. This method usually works by providing a seed, such as the clock time, which generates a sequence. The better the algorithm for producing this sequence the more random the sequence will appear to statistical tests. However, in most applications of SPH, the pseudo-random

<sup>1</sup>There are 7 crystal systems, each of which has many packing arrangements, giving a total of 14 Bravais lattices (an infinite set of points generated by a set of discrete translation operations described by:  $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ ).



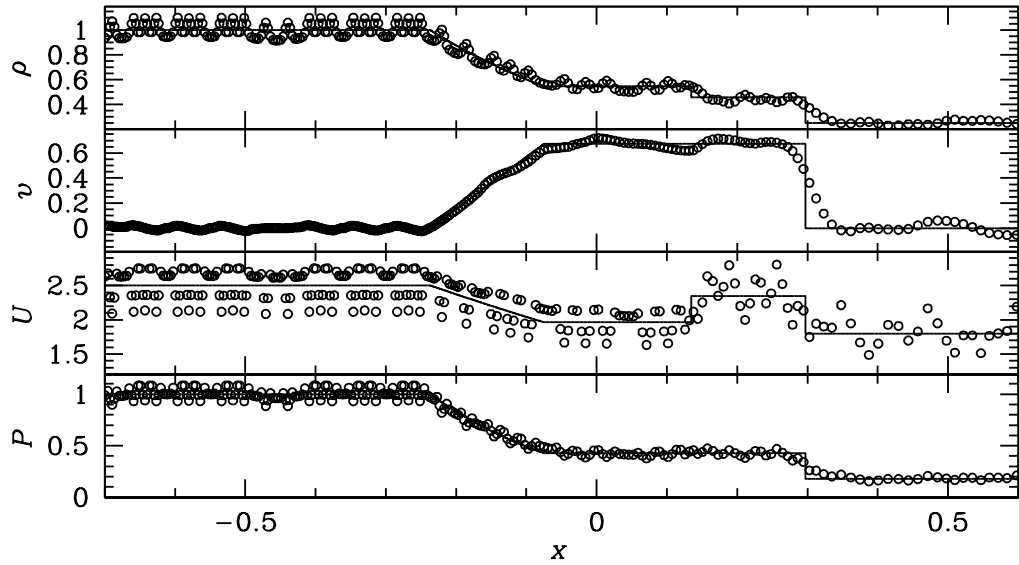


**Figure 4.3:** 2D random initial conditions using a pseudo-random generator (*left*) and a Sobol sequence (*right*).

sequence can be considered as a truly random sequence for the generation of initial conditions.

Alternatively one may use quasi-random number generators, these are number generators that more evenly fill the domain. Whilst they also give a uniform distribution, the placement of the points differs somewhat from that of a pseudo-random sequence (see Fig. 4.3). In pseudo-random numbers the probability of a point being in the interval  $[0, 1/2)$  is that same as for the interval  $(1/2, 0]$ . However, with quasi-random numbers this is not the case, each point knows where the previous point went. This allows one to construct a more even distribution of points. An example of this is the Sobol (1967) sequence.

The procedure for setting up density profiles based on random particle configuration is similar to that of ordered configurations. Particles are placed randomly with a non uniform distribution in space. The non uniform distribution allows more particles to be placed in higher density regions.



**Figure 4.4:** *The Sod shock tube test with Sobol initial conditions showing density, velocity, internal energy and pressure at  $t = 0.2$ .*

#### 4.2.1 Sobol Initial Conditions for the 1D Sod Shock Tube

The importance of avoiding randomness in SPH particle distributions was demonstrated in section 2.6 where the bias and variance for various distributions was shown. Whilst there may be some noise in the the density estimate with random distributions, the results say nothing about the ability of SPH to evolve these distributions.

To illustrate the effect initial conditions can have on a simulation, we perform the Sod shock tube test in 1D with initial conditions generated using a Sobol sequence. This is particularly useful in SPH as one should not place particles on top of each other in the initial conditions. However, one may still want to randomly place particles but in such a way as to avoid clumps and voids of particles.

Fig. 4.4 shows that the results with Sobol initial conditions are very noisy, especially in the internal energy and pressure profiles. The velocity profile is the least noisy of the profiles, possibly because artificial viscosity acts to remove small perturbations in the velocity profile. Although the overall features of the results are correct, the accuracy and sharpness of the results is lost due to the irregularity of the initial conditions.

## 4.3 Stability of Initial Condition

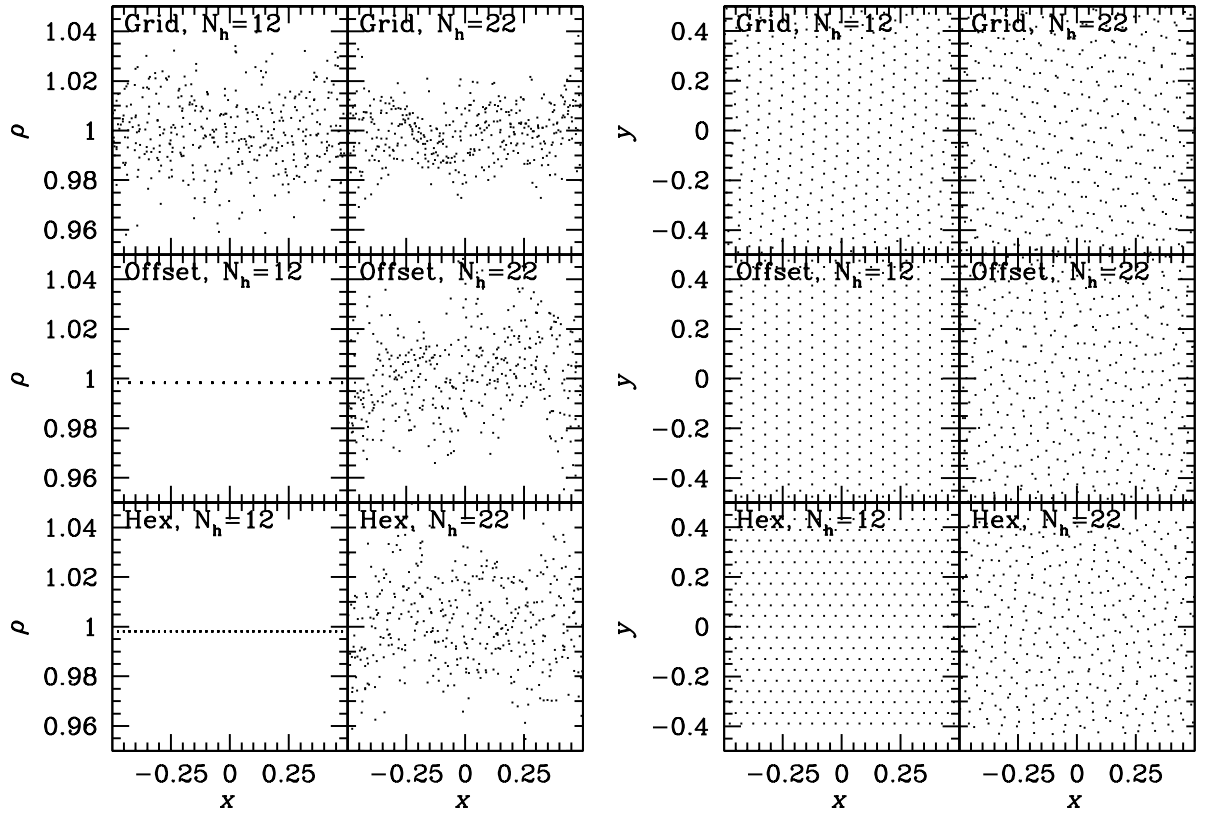
The stability of initial conditions to the onset of numerical noise can be determined by observing the long term evolution of an initial particle distribution. Ideally our SPH scheme should be stable to numerical noise and therefore the particle configuration should remain unchanged. However, the results may depend strongly on the initial particle setup.

Particles are placed such that they have a constant uniform density, the flow variables are set such that all velocities are zero and the sound speed is unity. The simulations are performed with periodic boundary conditions and particles are free to move across the boundary. An isothermal equation of state is used so that viscosity will remove unwanted kinetic energy from the system rather than heating the gas and increasing the pressure. The particles are placed on 2D grids in three possible configurations, a square grid, offset grid or a hexagonal grid, as shown in Fig. 4.1. The hexagonal grid requires careful treatment of the periodic boundaries as the height and width of a hexagonal cell is not the same. Therefore the periodic boundary is placed at  $(\pm 0.5, \pm \sqrt{3}/4)$  for the hexagonal grid.

As there is no initial flow or pressure gradients the system is in perfect equilibrium and should remain unchanged. Small perturbations caused by numerical noise may require the application of artificial viscosity to ensure that particles are initially in equilibrium. Therefore we test the long term evolution of this system both with and without artificial viscosity using both undamped and damped initial conditions. The stability can be determined by observing the spread in the density across the simulation domain, stable initial conditions should preserve a constant uniform density where as unstable conditions will lead to regions of over and under density. Also as we are performing the tests in 2D it is easy to observe the distribution of particles in the  $x - y$  plane and observe the change in the particle distribution.

### 4.3.1 Undamped Initial Conditions

We begin by testing the stability of undamped initial conditions. Particles are placed in a particular distribution and then the simulation is allowed to proceed without any prior relaxation procedure.

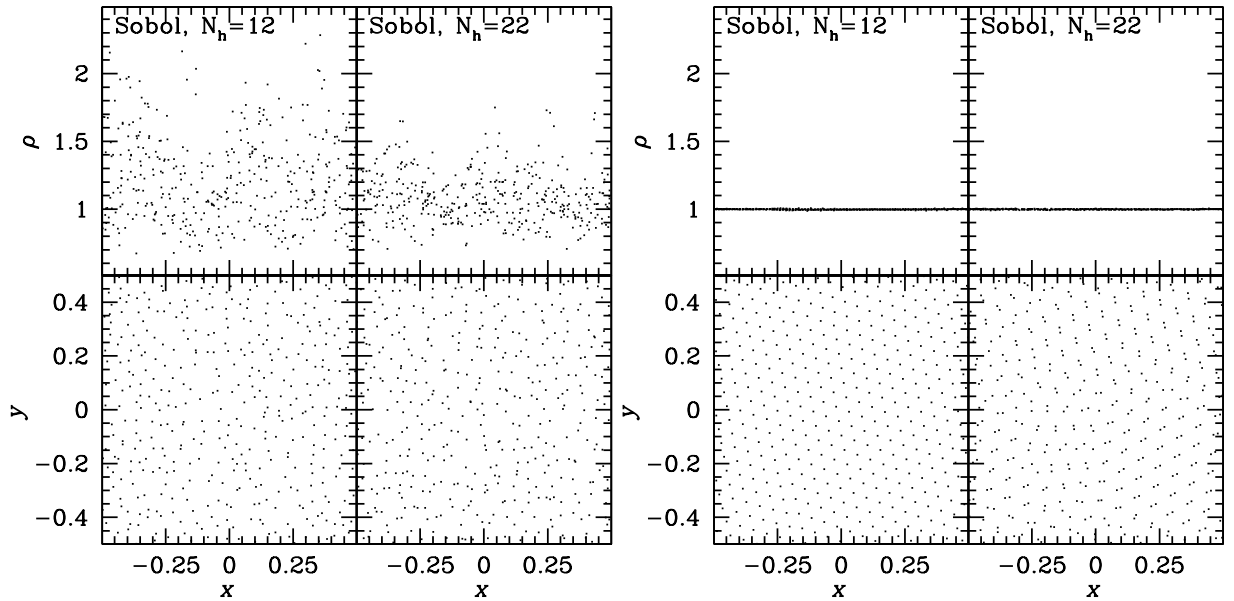


**Figure 4.5:** *Left:* Comparing density at  $t = 100$  for 2D square, offset and hexagonal 2D grids with  $N_h = 12$  and  $N_h = 22$ . *Right* Particle positions at  $t = 100$ . No artificial damping was used.

### No Viscosity

Fig. 4.5 shows how the type of grid used can affect the long term evolution of a stable system. We observe that with offset and hexagonal grids using 12 SPH neighbours, the system remains perfectly stable until at least  $t = 100$ , which is 100 sound crossing times. However, with square grid initial conditions we observe that the particles move off the grid causing density fluctuations of the order of a few percent.

With 22 SPH neighbours we find that in all three grids particles tend to cluster together. However, the density is relatively unaffected with errors of only a few percent. The particles tend to cluster into pairs with the force between clumped pairs approaching zero as the first derivative of the kernel approaches zero for closely spaced particles. Furthermore, as all neighbouring sites contain a pair of particles,



**Figure 4.6:** A run using initial conditions generated by a Sobol sequence. The **left** plots use no artificial viscosity whilst the **right** plot uses using  $\alpha = 1$ , both are shown at  $t = 100$ .

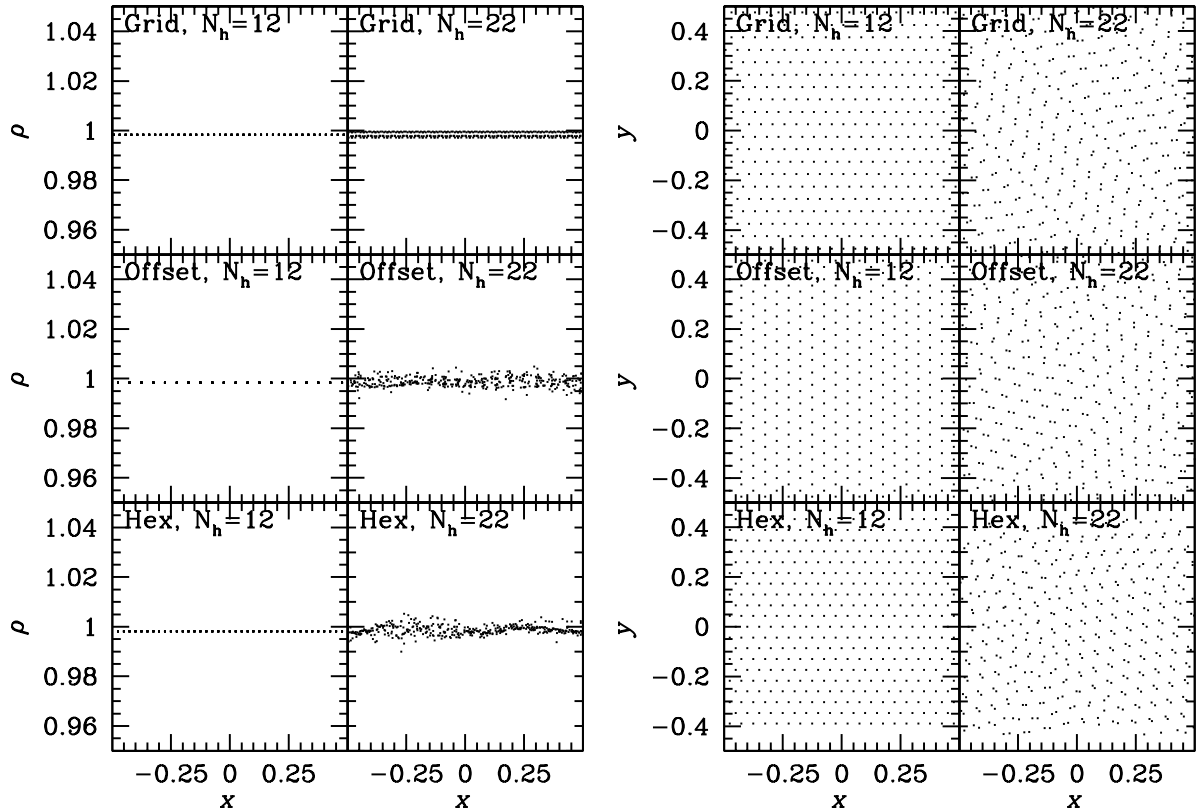
the systems acts as though each site has a single particle with twice the mass. The number of *effective* neighbours is therefore halved, although the mass inside a smoothing sphere and the smoothing length remains the same. Therefore the effective resolution has been decreased as the domain has fewer points sampled.

We also investigate the stability of initial conditions created using Sobol sequences. The left panel of Fig. 4.6 shows the evolution of the Sobol distribution with no artificial viscosity. We find that the errors in the density are considerable, around 100% for  $N_h = 12$  and 50% with  $N_h = 22$ .

### Viscosity

We test the affect of damping the initial conditions by applying a constant artificial viscosity of  $\alpha = 1$ . This is a common approach to construct stable initial conditions in SPH, particles are placed into a system that is close to equilibrium and a damping is applied to remove any numerical noise from the system.

The particle configurations for the 2D grids are shown after  $t = 100$  in Fig. 4.7. In the offset and hexagonal grids with  $N_h = 12$ , applying artificial viscosity has no affect on the results as the system is already stable without viscosity. The density remains accurate to floating point precision and the particles remain in their initial



**Figure 4.7:** *Left:* Comparing density at  $t = 100$  for 2D square, offset and hexagonal grids with  $N_h = 12$  and  $N_h = 22$ . *Right* Particle positions at  $t = 100$ . An artificial viscosity of  $\alpha = 1$  was used.

positions. Applying artificial viscosity to the square grid causes the particles to evolve into a configuration similar to a hexagonal grid. The application of viscosity also significantly reduced the density errors observed in the undamped case.

The right panel of Fig. 4.6 also highlights a significant difference when artificial viscosity is used for the Sobol initial conditions. We find that with  $N_h = 12$  particles become more ordered and approach an offset grid configuration. With  $N_h = 22$  we see clumps of particles but again find that the individual clumps are well separated and density errors are low.

### 4.3.2 Damped Initial Conditions

The previous section showed that whilst the square grid initial conditions are unstable without viscosity, when viscosity is applied the particles rearrange themselves

into a new configuration. We may determine if this new particle configuration is stable. We take the final snapshot ( $t = 100$ ) of the square grid when a damping has been applied and use it as initial conditions for a simulation without any damping. At this point the error in the density is very small and particle accelerations are small (less than  $10^{-5}$  in code units). Before the new particle configuration is allowed to evolve we reset the velocities and accelerations of all particles to zero. This is done so that we only observe the evolution of the particle positions and reduce the effects of previously having artificial viscosity imposed on the system.

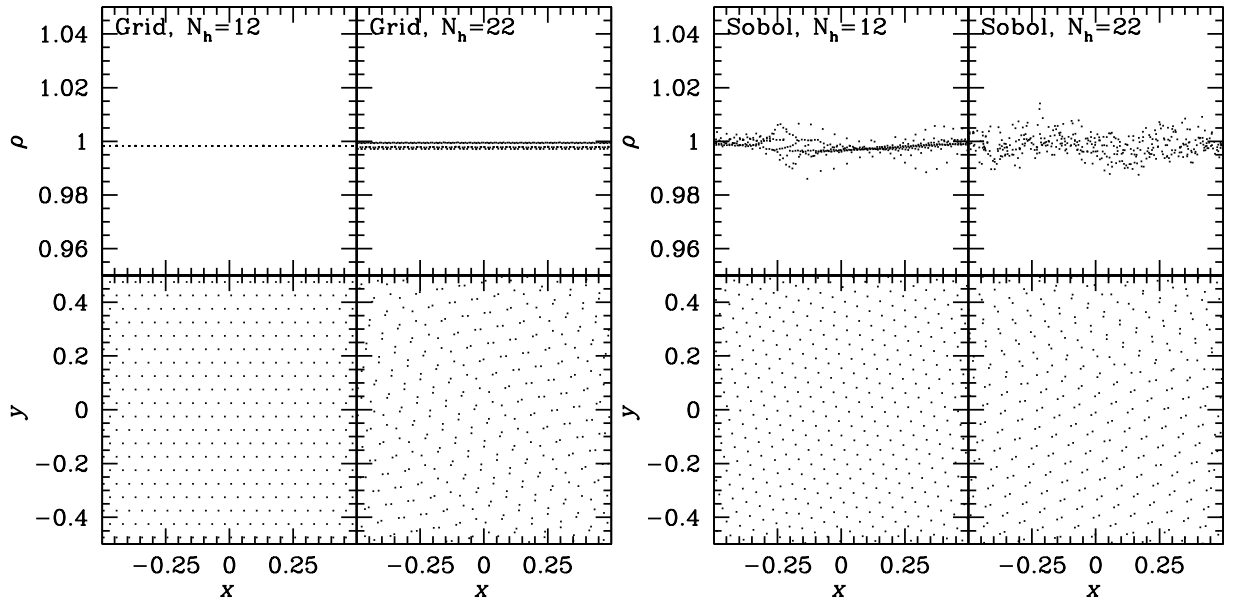
Fig. 4.8 shows hows the damped system evolves without further artificial damping. We find that the system remains stable in its new configuration. With  $N_h = 12$  the particles remain on a lattice similar to the hexagonal grid. With  $N_h = 22$  particles remain clumped, however the density errors do not become any larger. Furthermore the particles did not undergo any further clumping, and the individual clumps remained separated. We also observe a similar result when using a Sobol distribution, with  $N_h = 12$  the particles remain on a lattice whilst with  $N_h = 22$  the clumps remain but each clump is well separated. The density errors for the Sobol sequence do grow slightly, possibly because the initial conditions have not quite reached an equilibrium, and therefore they become slightly unstable.

### 4.3.3 Setting up Initial Condition

The results of the previous section demonstrate the important role that initial conditions play in both the stability and accuracy of results. Random initial conditions allow for the general features of results to be correct but lack the accuracy due to being swamped by numerical noise. We also find that the stability of random initial conditions is poor when no artificial viscosity is used; this can become problematic if a simulation requires low artificial viscosity to avoid damping, such as shearing flows and the evolution of sound waves (see sections 3.6.5, 3.6.3 and 3.6.1). Regular initial conditions can also suffer from poor stability if the particles are not packed in the tightest possible configuration, as is the case for the square grid initial conditions.

Artificial viscosity can help to damp away any small amplitude perturbations aiding the settling of the system into a stable state, and is often used for such purposes. This method of damping initial conditions should be used wherever possible before the main simulation proceeds, especially in situations where we require the modelling of small amplitude processes, such as oscillations of stars.

For initial conditions that require an unbalanced force, for example in the grav-



**Figure 4.8:** A grid (*left*) and a Sobol distribution (*right*) that has initially been damped into equilibrium using a viscosity of  $\alpha = 1$ . The viscosity is then turned off and the system is evolved until  $t = 100$ .

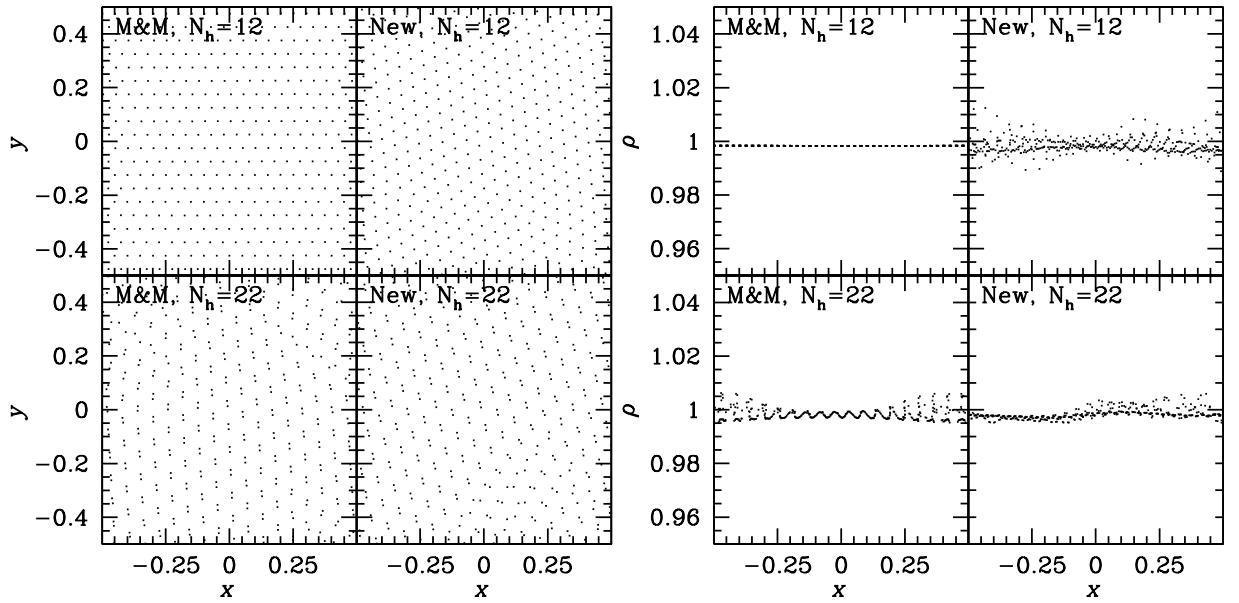
itational collapse of a cold cloud, it may be difficult or impossible to damp initial conditions into equilibrium. Furthermore, as these systems are not in equilibrium initially the dampening of the initial conditions is not as vital. In this case it is important to avoid randomness in the initial conditions and pack particles as tightly as possible such that they are as close to their equilibrium configuration as possible.

We find that the numerical dissipation schemes outlined in chapter 3 are able to prevent the particles becoming disordered when using  $N_h = 12$  (see Fig. 4.9). With  $N_h = 22$  we find that clumping still occurs but the errors in the density are kept to less than one percent.

## 4.4 Small Amplitude Perturbations and Stability

An important test for any code is numerical stability; in the context of SPH, this means testing the stability of the code under small amplitude perturbations to ensure that the amplitude of the perturbations do not grow. In order to test this we of course need to pick scenarios where physical instabilities are not present which would naturally cause the growth of the perturbations, as in the Kelvin-Helmholtz



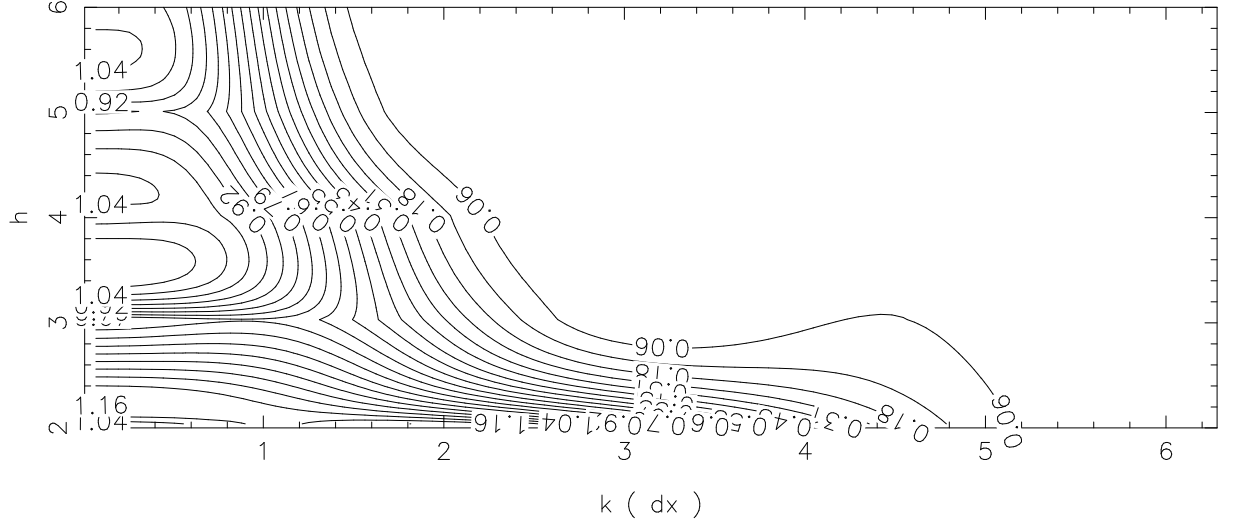


**Figure 4.9:** Comparing the new viscosity scheme of section 3.4 to the M&M method for the 2D square grid with  $N_h = 12$  and  $N_h = 22$ .

instability.

A well known SPH numerical instability is the so called clumping instability, this is where particles cluster together in a very small radius for no physical reason, as was observed in the previous section (for example see Fig. 4.5). When this happens the density of the gas can be relatively unaffected with errors of less than 5%. A good indicator of clumping is the quantity  $q_{\min} = r_{\min}/h$ . As the density can remain roughly constant, the smoothing length also remains roughly constant; the reason for this being that only around 3 – 5 particles cluster together and the number of neighbours used is typically 40 – 60. Therefore as particles clump  $q_{\min} \rightarrow 0$ .

Although clumping may not seem much of a problem given that the density can remain fairly accurate, the clustering of particles reduces the effective resolution as the system has more than one interpolation point in a small region of space. This essentially means that the number of effective particles in the simulation has been reduced by a factor of 2 – 3, whilst the computational cost remains the same. Furthermore clumping instability can drive other unphysical instabilities leading to a loss of accuracy of the results.



**Figure 4.10:** Contour plot of numerical sound speed  $c_{\text{num}}^2$  in the  $k - h$  plane using the Morris (1996) analysis (which excludes artificial viscosity). Red regions show where  $c_{\text{num}}^2 < 0$  leading to a growth in the perturbations.

#### 4.4.1 Stability Analysis

Morris (1996) performed a stability analysis on the SPH equations assuming small amplitude perturbations to the particle positions (see also Cha, 2002; Price, 2004; Read et al., 2010). Consider a uniform medium where all particles are initially placed in some distribution and have equal mass  $m$ . We then perturb the particles according to

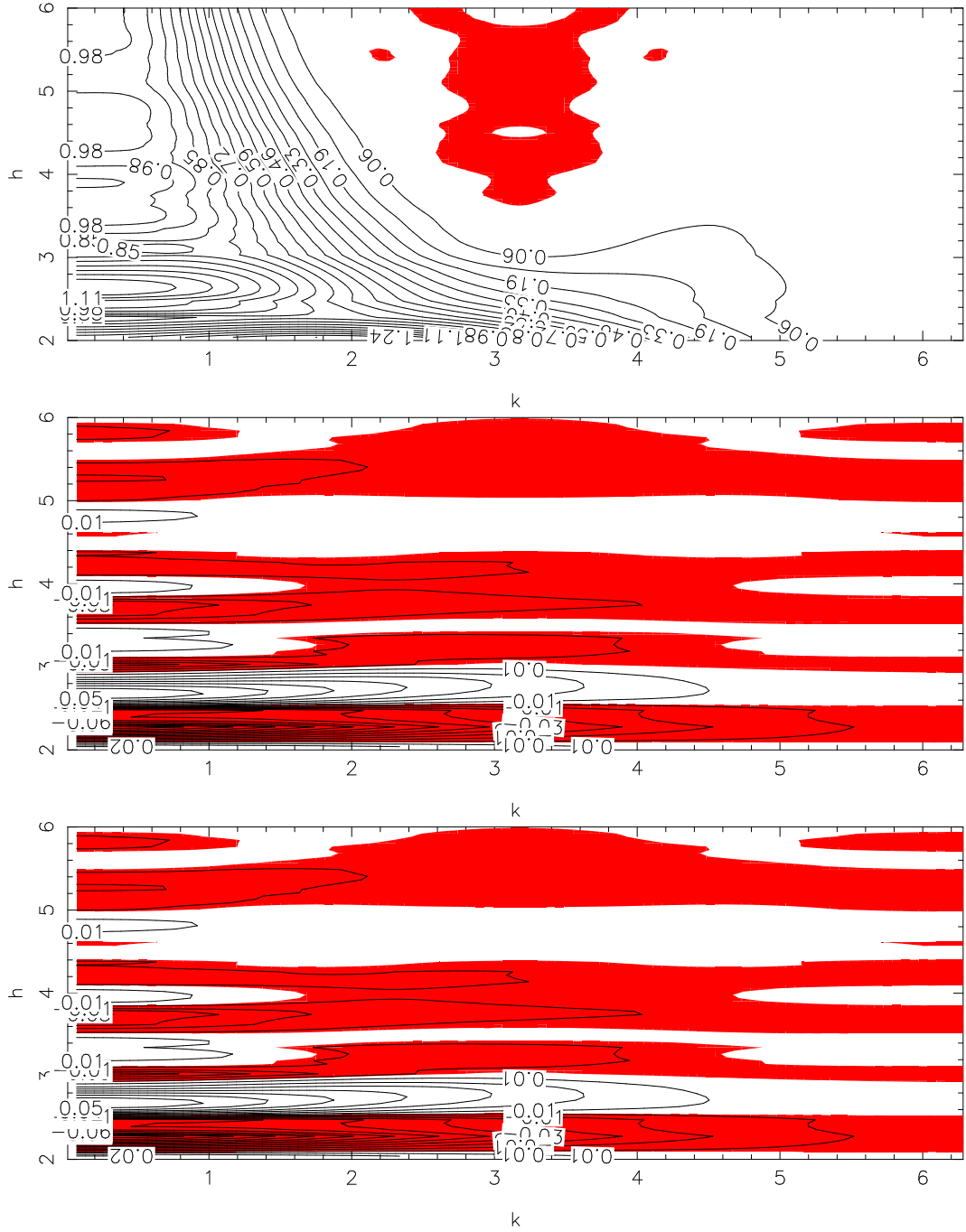
$$\mathbf{x}'_a = \mathbf{x}_a + \mathbf{A}e^{i\phi} \quad (4.1a)$$

$$\mathbf{v}'_a = \mathbf{V}e^{i\phi} \quad (4.1b)$$

$$\rho'_a = \rho_0 + D e^{i\phi}, \quad (4.1c)$$

where the vector  $\mathbf{A} = (X, Y, Z)$  determines the amplitude of the perturbation in each coordinate direction, and  $\phi_a = \mathbf{k} \cdot \mathbf{x} - \omega t$ . The perturbation to the velocity can be computed as

$$\mathbf{V} = -\mathbf{A}i\omega e^{i\phi} \quad (4.2)$$



**Figure 4.11:** Contour plot of numerical sound speed  $c_{\text{num}}^2$  in the  $k-h$  plane using the Morris (1996) 3D analysis (which excludes artificial viscosity). Red regions show where  $c_{\text{num}}^2 < 0$  leading to a growth in the perturbations. The perturbation was applied along the  $\mathbf{k} = (1, 0, 0)$  direction. The top plot shows the longitudinal wave, the middle and bottom plots show the two transverse waves.

and perturbation density as

$$D = m \sum_j (1 - e^{i(\phi_j - \phi_i)}) \mathbf{A} \cdot \nabla W_{ij}. \quad (4.3)$$

By deriving the perturbed form of the momentum equation it is possible to obtain the dispersion relationship for the perturbations given above. If we assume that the particles are placed on either SC or FCC lattices (or any distribution that ensures that particles are evenly spaced) we can use the identities

$$\sum \cos(\phi_i - \phi_j) \nabla W = 0 \quad (4.4a)$$

$$\sum \sin(\phi_i - \phi_j) \underline{H}(W_{ij}) = 0. \quad (4.4b)$$

where  $\underline{H}$  is the Hessian matrix of  $W$ . This gives the dispersion relation as (Read et al., 2010)

$$\omega^2 \mathbf{A} = \left( \frac{2Pm}{\rho} \sum_j \underline{H}(W_{ij}) (1 - \cos(\mathbf{k} \cdot \mathbf{x})) + (\gamma - 2) \frac{m^2 P}{\rho^3} (\mathbf{q} \otimes \mathbf{q}) \right) \cdot \mathbf{A}, \quad (4.5)$$

where the vector

$$\mathbf{q} = \sum_j \sin(\mathbf{k} \cdot \mathbf{x}_0) \nabla W_{ij}. \quad (4.6)$$

There is no imaginary part to the dispersion relationship as the terms that contribute are both odd (see equation 4.4), thus  $\omega^2$  is always real. However, if  $\omega^2 < 0$  then  $\omega$  will have an imaginary component which will lead to a growth of the perturbation. This analysis is valid for a superposition of many wave modes as for each wave mode there is only one associated frequency. Therefore the dispersion relationship holds for each wave number regardless of the number of superpositions applied. Using Morris's analysis we may plot the numerical sound speed in  $k-h$  space for a 1D system of equally spaced particles (see Fig. 4.10) which allows us to determine the stability of the system. We find that in 1D, SPH is stable to all wavelength perturbations as the numerical sound speed remains real. Therefore, to first order, the perturbations applied will not grow.

In 3D the stability properties can be found by assuming the perturbation is applied along a particular direction. The stability analysis gives three dispersion relations, one parallel to the perturbation direction corresponding to a longitudinal

wave, and two perpendicular to the perturbation direction corresponding to two transverse waves. We set up a simple cubic grid with grid size unity, meaning that the nearest neighbour is also unity. Fig. 4.11 shows the longitudinal and two transverse waves if the perturbation is applied along the  $\mathbf{k} = (1, 0, 0)$  direction to a simple cubic grid.

An interesting point to note is that the longitudinal stability differs from that of the 1D case. In the 3D case there is a region around  $k = \pi$  which is unstable to longitudinal perturbations. We also observe the so called banding instability (Read et al., 2010) in the two transverse modes. Although the transverse mode instabilities can cause problems, it is thought that the major role in preventing mixing is the clumping instability (Read et al., 2010). The prevention of mixing makes it difficult for SPH to resolve the Kelvin-Helmholtz instability. The stability analysis above is often applied for the case of a simple cubic<sup>2</sup> grid, however, the tightest packing arrangement in 3D is the FCC lattice. The stability analysis above is also valid for a FCC lattice, therefore we can obtain the equivalent stability properties for a FCC lattice. The FCC lattice size is set to twice the SC lattice to ensure that the nearest neighbour is still at  $(1, 0, 0)$ .

The results, shown in Fig. 4.12, exhibit slightly different stability properties for the FCC lattice, particularly in the transverse modes. The two instability bands that were present in the SC lattice between  $h \approx 3$  and  $h \approx 4.5$  have become a single band extending from  $h \approx 3.5$  to  $h \approx 4$ . This gives a larger region where the instability is not present in the FCC lattice. Therefore we may expect that the FCC lattice to be more stable to the banding instability.

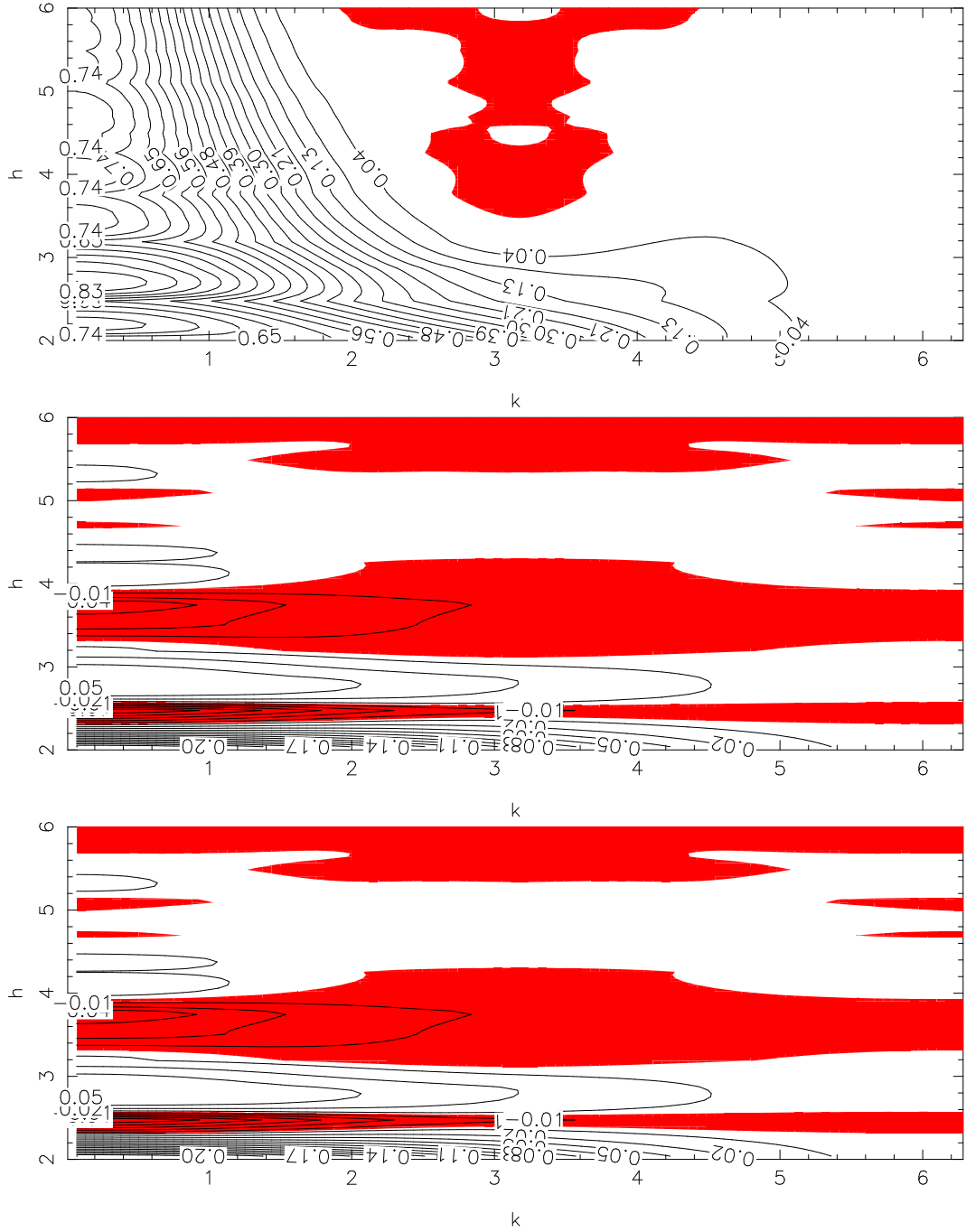
It must be noted that the stability analysis given in Fig. 4.11 and 4.12 is only for waves applied in the  $\hat{\mathbf{k}} = (1, 0, 0)$  direction. The stability properties may vary for perturbations applied in different directions (see also Read et al., 2010).

#### 4.4.2 Stability Analysis With Viscosity

Further work by Morris (1996); Cha (2002) showed that when viscosity is accounted for in the SPH acceleration equations an extra term is added to the dispersion

---

<sup>2</sup>The original Morris (1996) analysis did assume a simple cubic lattice. However, equation (4.5) only assumes that particles are distributed such that equation 4.4 holds.



relation. In 3D the dispersion relation is

$$\omega_{\text{AV}}^2 \mathbf{A} = \omega^2 \mathbf{A} + i\omega_{\text{AV}} \frac{\alpha m h c}{\rho} \sum_j \frac{1 - \cos(\mathbf{k} \cdot \mathbf{r}_{ij})}{|\mathbf{r}_i - \mathbf{r}_j|} \hat{\mathbf{r}} \cdot \mathbf{A} \nabla W_{ij}.$$

In 1D with particle spacing of unity the solution is

$$\omega_{\text{AV}} = \frac{-ib \pm \sqrt{4\omega^2 - b^2}}{2}, \quad (4.7)$$

where

$$b = \frac{\alpha m h c}{\rho \Delta x} \sum_j \frac{1 - \cos(k \Delta x j)}{j} \nabla W_{ij}. \quad (4.8)$$

We can decouple the real and imaginary parts of the frequency such that  $\omega_{\text{AV}} = \omega_R + i\omega_I$  and therefore the perturbation can be written as  $\delta x = A_k e^{-i\omega_R t} e^{\omega_I t}$ . The first exponential describes the sinusoidal oscillation of the perturbation whilst the second exponential term describes the exponential growth or decay of the wave; if  $\omega_I$  is positive the perturbation will grow exponentially. From evaluating this equation we find that  $\omega_I$  is always positive provided that  $\omega \neq 0$  (regardless of the  $\pm$  sign of the root) and therefore we always get exponential growth of the perturbation for all values of  $h - k$  space. This implies that SPH is unstable when artificial viscosity is used. However, this analysis is somewhat misleading because it assumed that both approaching and receding pairs took place in the viscosity interaction. This of course means that for receding pairs viscosity acted as an artificial attractive force, pulling particles together. In practice artificial viscosity is only used for approaching pairs of particles where it acts as a repulsive force preventing the particles approaching too closely. Therefore we would expect artificial viscosity to prevent clumping rather, as this analysis suggested, than promoting clumping.

More recently Read et al. (2010) performed a full 3D stability analysis on SPH. They found that when using the cubic spline kernel, SPH can be unstable to longitudinal waves for certain neighbour numbers, which causes the clumping instability. They also discovered that SPH is unstable to transverse waves, which they call the banding instability, for a large range of neighbour number. In order to overcome this they suggest using a high order cored triangle kernel with 442 neighbours (as a rough estimate this corresponds to 58 in 2D and 7 in 1D)<sup>3</sup>. This choice suppresses

---

<sup>3</sup>given that  $\rho h^\nu = \mu = m_i N_h$  we compute  $h$  assuming  $m_i = 1$  and  $\rho = 1$  for the 3D case. This

both the transverse and longitudinal waves and gives a small leading error (which the authors call the  $E_0$  error).

### 4.4.3 Numerical Results

We can confirm the Morris (1996) stability analysis by performing a set of SPH simulations in the  $h - k$  plane where we apply small perturbations (equation 4.1) to an initially uniform and static lattice. Ideally one would like to measure the numerical sound speed and produce a contour plot for direct comparison. However, doing this for SPH data is not straightforward. Firstly the SPH method will not be able to resolve waves with  $k > \pi$ . The reason for this is in order to resolve the wave there needs to be at least two SPH particles per wave giving the minimum wavelength as  $\lambda_{\min} = 2\delta x$ . This translates to a maximum wave number of  $k_{\max} = \pi$ . Furthermore the data must be Fourier transformed to determine the real and imaginary part of the frequency at a particular wave number. There may be power developing at other wave modes than the applied perturbation mode due to the non-linear nature of SPH, making the presentation of the results difficult. A more straightforward approach is to perform the simulation and measure the minimum distance between two particles. If this minimum distance decreases during the simulation then particles must have moved together. A simulation is defined to have clumped if by the end of the simulation the minimum particle separation is 20% of the unperturbed particle separation. If the minimum separation is between 20% – 80% of the unperturbed separation a glass has formed. Systems that return to at least 80% of the original separation are deemed to be stable to these perturbations.

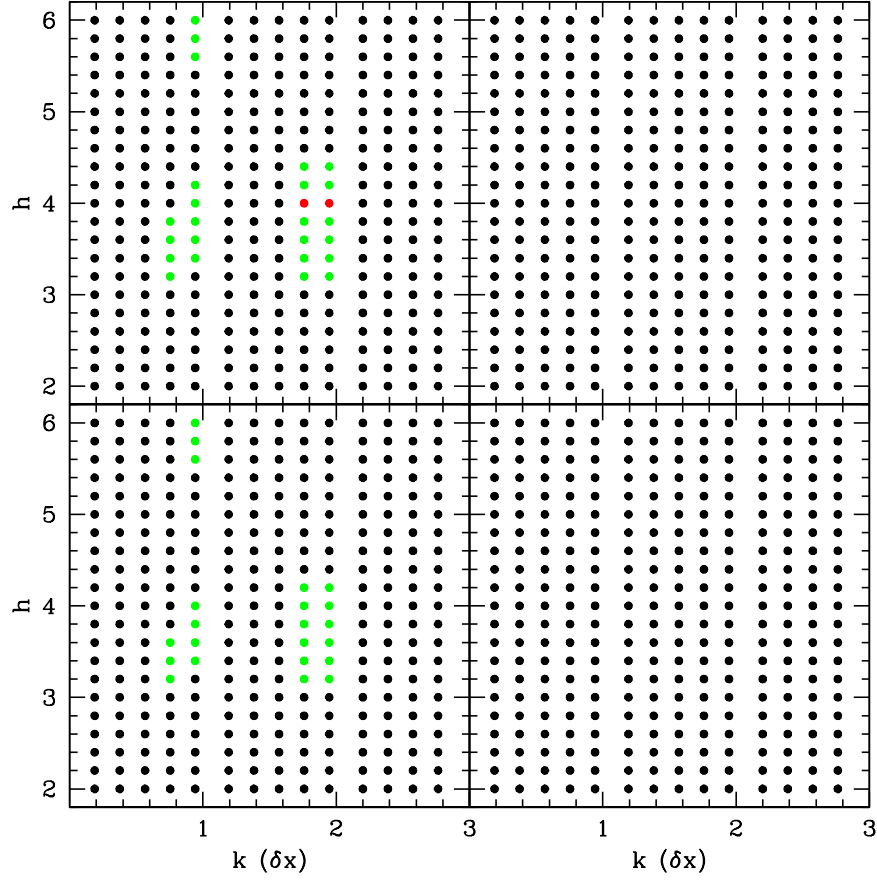
Care must be taken when selecting the wave number as we require that the perturbation wave fits exactly within the periodic box of size  $x = (-0.5\delta\ell, 0.5\delta\ell)$ ,  $y = (-0.5\delta\ell, 0.5\delta\ell)$ ,  $z = (-0.5\delta\ell, 0.5\delta\ell)$ , and there are only a discrete set of wave numbers that are able to do so. The wave number allowed also depends on the direction of the perturbation wave,  $\hat{\mathbf{k}}$ .

To set the wave number ( $k$ ) used we firstly set the wave number in code units as  $k' = k/\delta x$  where  $\delta x$  is the cell separation and  $k$  is the requested wave number. We then calculate the number of whole waves that fit inside the periodic box as  $N_{\text{waves}} = \text{int}\{k'/(2\pi)\ell\}$  where  $\ell = \sqrt{\delta\ell^2 \hat{\mathbf{k}} \cdot \hat{\mathbf{k}}}$ . We then calculate the wavelength of the wave as  $\lambda' = \ell/N_{\text{waves}}$  and therefore the actual wavenumber used is  $k'_{\text{used}} = 2\pi/\lambda'$ .

---

value is then used to estimate  $N_h$  for the 1D and 2D case. The exact scaling depends on the type of lattices used.





**Figure 4.13:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black). **Left** plots are without artificial viscosity, **right** plots are with a fixed  $\alpha = 1$  viscosity. The **top** plots are with an isothermal equation of state, the **bottom** plots are with an adiabatic equation of state.

We then convert back to find the wavenumber used in units of cell separation as  $k_{\text{used}} = k'_{\text{used}} \delta x$ . This gives us the wave number  $k$  in units of cell separation of the applied perturbation in the direction  $\hat{\mathbf{k}}$ .

## 1D Results

We perform the numerical simulations in 1D by taking particles with a mass, separation and sound speed of unity and perturb both the position and velocity of the particles according to equation (4.1) with a perturbation amplitude of  $A = 0.01$ . We test both isothermal and adiabatic equation of state and compare the effects of using artificial viscosity. The simulation is run for 10 sound crossing times, at which point we take the minimum separation of particles to determine if clumping

has occurred.

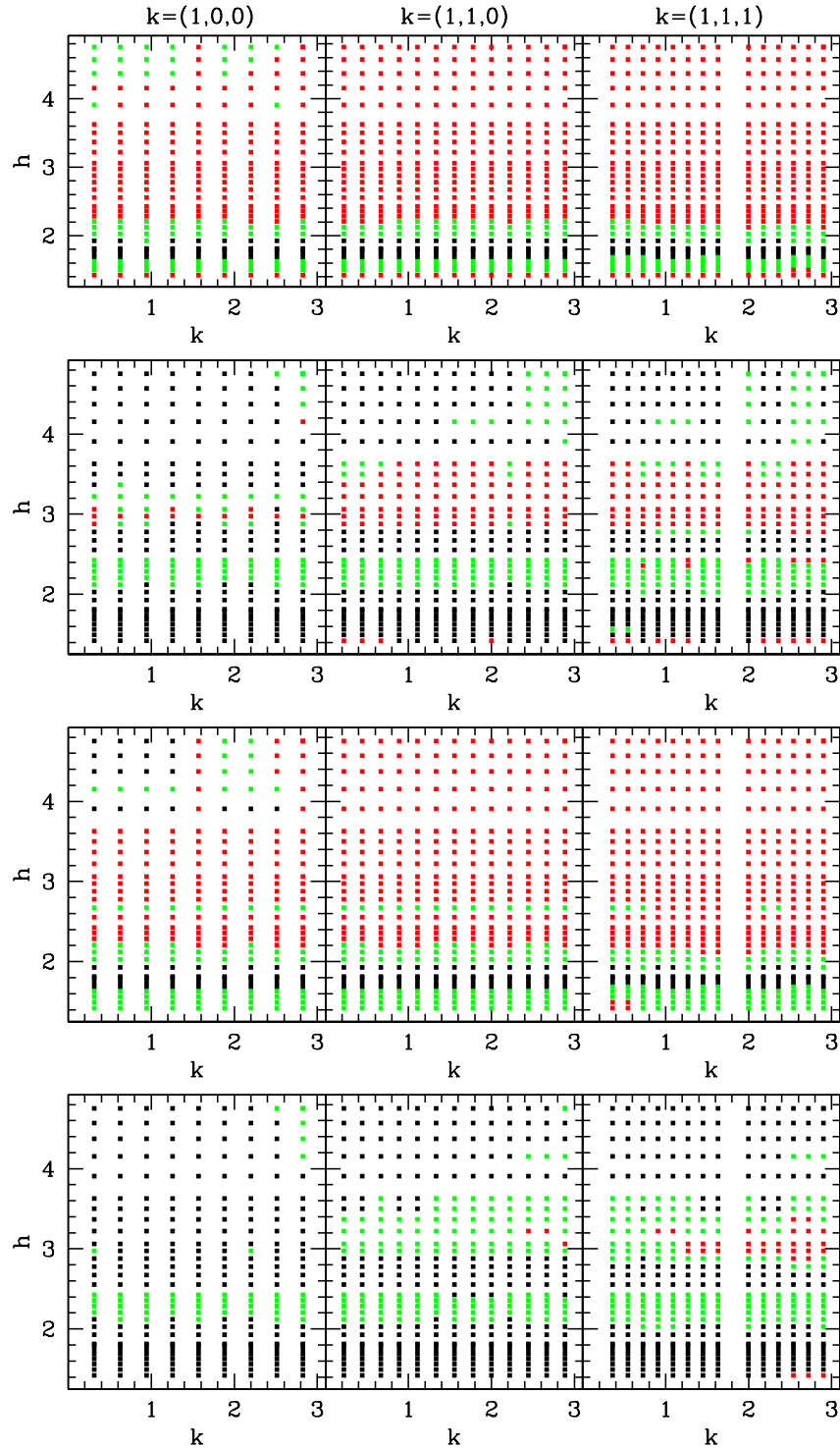
Fig. 4.13 shows the regions of  $h - k$  space which are unstable to clumping for the standard SPH implementation in 1D using the cubic spline kernel. Without viscosity the results agree fairly well with the Morris (1996) analysis, most of the simulations remained stable. However, there were a few simulations where particle clumping occurred which were not detected by the analysis of the previous section. It is possible that this clumping is caused by non-linear effects which were neglected in the analysis, causing a growth in the perturbations.

The results of the viscosity runs disagree with the analysis performed by Morris (1996) and Cha (2002), which suggested that artificial viscosity would promote a region of clumping for wavelengths around  $k = \pi$ . The results of the simulations did not find this, they in fact show that the scheme is more stable with artificial viscosity. Simulations that underwent clumping without viscosity were found to be stable when artificial viscosity was applied. Furthermore the simulations that previously evolved into a glass now return to their initial conditions. A possible explanation for the discrepancy with the expected results is due to taking both approaching and receding pairs in the analysis, meaning that the receding particles would experience an unphysical tension force attempting to pull them together. However, we also performed SPH simulations where both approaching and receding particle pairs were taken for the viscosity and we found them to have the same stability properties as those shown in Fig. 4.13. This indicates that the assumption in taking approaching and receding particle pairs may not be responsible for the discrepancy in the analytical results and the simulation results.

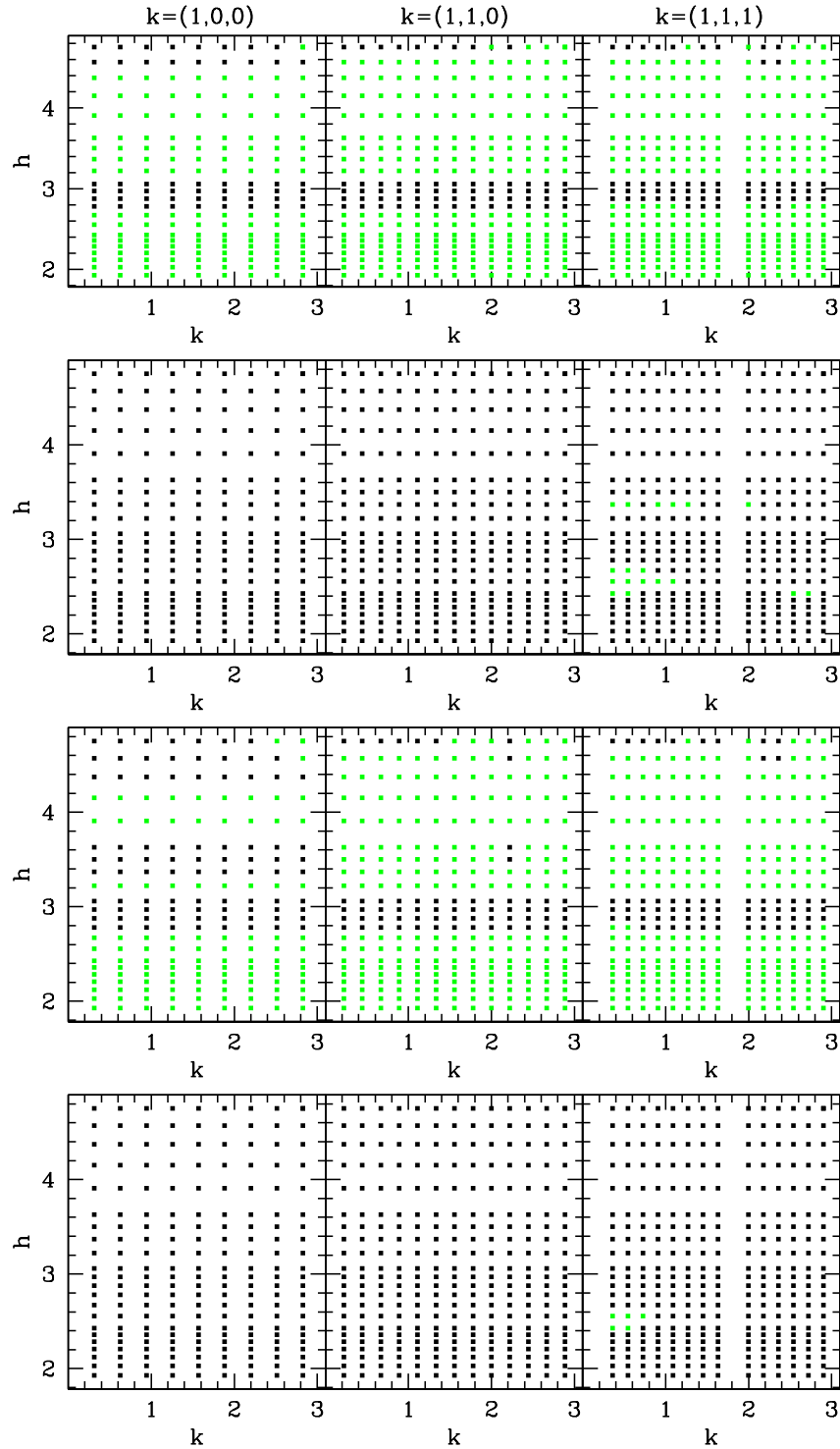
### 3D Simple Cubic Lattice Results

As a direct comparison of the analysis carried out by Read et al. (2010), we can perform 3D simulations to determine the stability properties of a simple cubic (SC) grid which is subject to perturbations. The cell separation was set to  $d_c = 0.1$  and the perturbation amplitude was set to 1% of the cell separation, that is  $\mathbf{A} = 0.001\mathbf{k}$  (see equation 4.1). We perturbed the particle positions in three different directions,  $\mathbf{k} = (1, 0, 0), (1, 1, 0), (1, 1, 1)$ . The velocity perturbations were applied in the same direction as the position perturbations. The number of neighbours for a SC lattice can be computed from  $h$  using the formula

$$N_h = \frac{4\pi}{3}h^3. \quad (4.9)$$



**Figure 4.14:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black) for a SC lattice using a cubic spline in 3D. **Top row** Isothermal equation of state with no viscosity. **2nd row** Isothermal equation of state with  $\alpha = 1$ . **3rd row** Adiabatic equation of state with no viscosity. **Bottom row** Adiabatic equation of state with  $\alpha = 1$ .



**Figure 4.15:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black) for a SC lattice using a HOCT kernel in 3D. **Top row** Isothermal equation of state with no viscosity. **2nd row** Isothermal equation of state with  $\alpha = 1$ . **3rd row** Adiabatic equation of state with no viscosity. **Bottom row** Adiabatic equation of state with  $\alpha = 1$ .

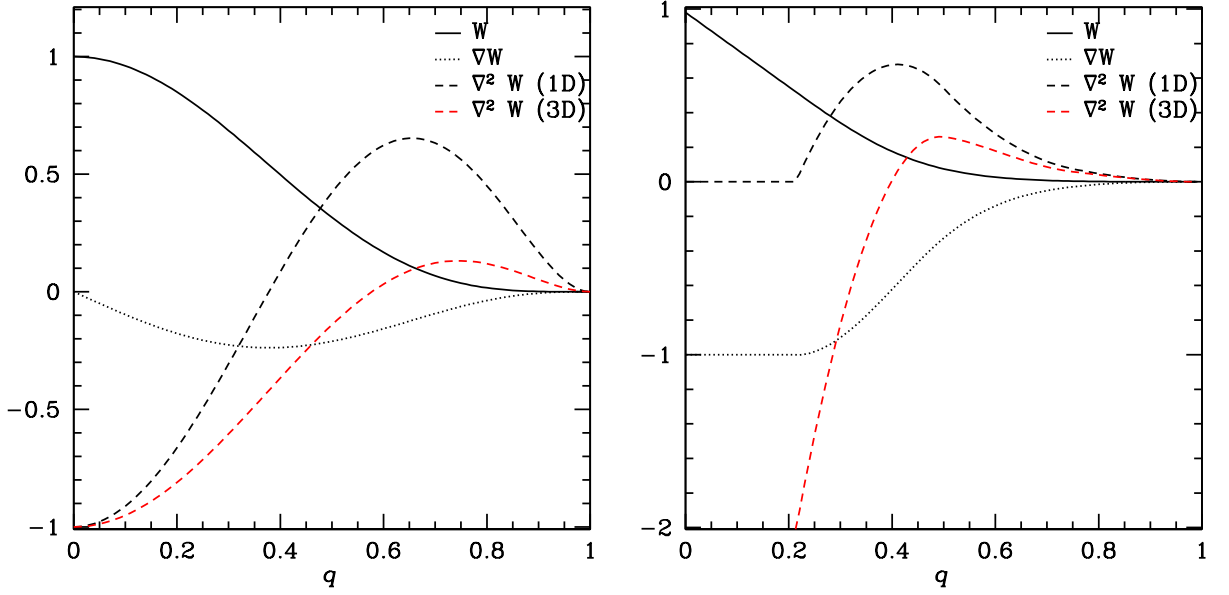
Fig. 4.14 shows the stability properties of the SC lattice with the cubic spline kernel. We observe that with no artificial viscosity the system is unstable to all wavelengths for smoothing lengths  $h > 2.2$  ( $N_h > 44$ ). When artificial viscosity is applied the stability properties improve for regions  $h > 2.2$ , especially for the adiabatic equation of state where there are only a few simulations that become unstable. Therefore the results loosely agree with the stability analysis of Read et al. (2010) in that we find that there are regions of instability which depend only on the smoothing length. However we find that the stability properties without artificial viscosity are much worse than expected by Read et al. (2010) as we find that as  $h$  increases the simulations do not become stable again. We should also note that the stability analysis of the previous section assumed a fixed smoothing length. However, the SPH simulations used a smoothing length which was allowed to vary according to the method outlined in section 2.4.2. However, we find that the density only changes by around 1% which leads to an even smaller change in  $h$  via equation (2.11).

We can also repeat the experiment with different SPH kernels. A similar kernel to the cubic spline is the Ferrers  $N = 4$  kernel as described by the equation below and as shown in the left panel of Fig. 4.16.

$$W(q) = f(1 - q^2)^4. \quad (4.10)$$

The shape of the kernel is similar to that of the cubic spline, the noticeable difference is that the maxima of the 1st derivative occurs farther from the centre in the Ferrers kernel, at  $q = 0.378$  rather than at  $q = 1/3$  as for the cubic spline. The implication of this for the Ferrers kernel is that particles farther from the kernel centre will experience a reduction in force as two particles approach. Therefore one may expect the Ferrers kernel to be stable to clumping for a smaller range of  $h$  as there is a smaller region where approaching particles can repel each other. We also note that the maximum of the second derivative of the kernel occurs slightly farther from the centre with the Ferrers kernel  $q = 0.75$  in 3D compared to  $q = 0.71$  for the cubic spline.

An alternative kernel suggested by Read et al. (2010) is the high order cored triangle (HOCT) kernel as described by the equation below and as shown in the



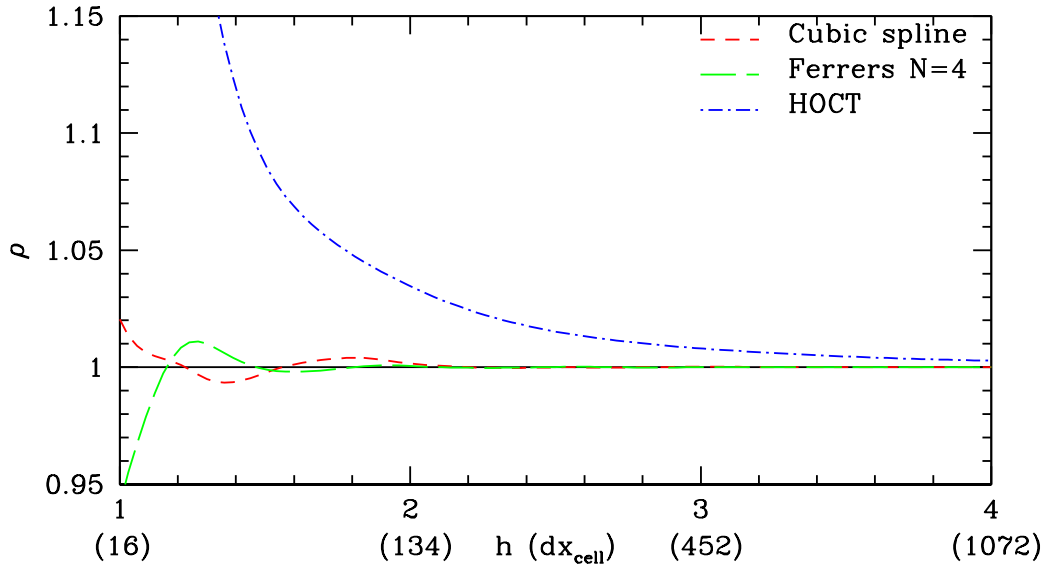
**Figure 4.16:** The left panel shows Ferrers  $N=4$  kernel and its first and second derivative. Note that the maxima in the 1st and 2nd derivative are at  $q = 0.378$  ( $N_h = 110$ ) and  $q = 0.71$  ( $N_h = 17$ ) respectively. The right panel shows the HOCT kernel (Read et al., 2010).

right panel of Fig. 4.16.

$$W(q) = \frac{N}{h^\nu} \begin{cases} -2.15q + 0.98 & 0 < q \leq 0.214, \\ (1-q)^4 + 3.2(0.75-q)^4 - 18.8(0.5-q)^4 & 0.214 < q \leq 0.5, \\ (1-q)^4 + 3.2(0.75-q)^4 & 0.5 < q \leq 0.75, \\ (1-q)^4 & 0.75 < q \leq 1. \end{cases} \quad (4.11)$$

The parameter  $N$  is a normalisation constant with values  $N = (2.07, 3.12, 6.52)$  for 1D, 2D and 3D respectively. This kernel is constructed such that the first derivative becomes constant in the inner parts of the kernel. The aim of this is to stop the reduction in force experienced by closely approaching particles, which is one of the draw backs of the cubic spline and Ferrers kernels. However, it should be noted that whilst in 1D the 2<sup>nd</sup> derivative of the kernel is zero inside  $q < 0.214$ , in 3D the 2<sup>nd</sup> derivative diverges as  $q \rightarrow 0$ .

Each kernel has differing density estimation errors. The SPH density estimate (equation 2.10) for a uniform FCC lattice is shown in Fig. 4.17 for the cubic spline, Ferrers and HOCT kernels. We observe that the density estimate is much more



**Figure 4.17:** The SPH density estimation (equation 2.10) of a uniform 3D FCC lattice (with cell size equal to unity) as a function of smoothing length for the cubic spline kernel (2.9), the Ferrers  $N = 4$  kernel (4.10) and the HOCT kernel (4.11). The corresponding neighbour number is shown in parenthesis.

accurate for the cubic spline and Ferrers kernel than the HOCT. Using<sup>4</sup>  $N_h = 40$  the density is accurate to around 1% with both the cubic spline and Ferrers kernel. However to attain this accuracy with the HOCT kernel one needs to use around  $N_h = 450$ . This can greatly increase the computational cost of the simulation. Furthermore the simulation will appear more smoothed as the properties of each particle are essentially smoothed over a larger region. Therefore in order to achieve the same resolution as with  $N_h = 40$ , many more particles may be required, further increasing the computational cost.

Fig. 4.15 shows the stability results in  $h - k$  space for the SC lattice using the HOCT kernel. We observe that clumping is avoided for all wave-numbers and smoothing lengths regardless of viscosity applied or equation of state used. This again disagrees somewhat with the analysis of Read et al. (2010) who find that there are bands of stable and unstable regions. Whilst we do find a banding of increased stability around  $h = 3$  in the cases without artificial viscosity, we find that the regions away from this are not unstable to clumping but instead form a glass structure.

<sup>4</sup>The number of neighbours can be computed from  $q$  using the formula  $N_h = \frac{8\pi}{3\sqrt{2}} \frac{1}{q^3}$  for a FCC grid in 3D.

Equation of State	Viscosity	Kernel	Unstable Region
Isothermal	$\alpha = 0$	Cubic Spline	$N_h \gtrsim 44$
Isothermal	$\alpha = 1$	Cubic Spline	$92 \lesssim N_h \lesssim 164$
Adiabatic	$\alpha = 0$	Cubic Spline	$N_h \gtrsim 44$
Adiabatic	$\alpha = 1$	Cubic Spline	$92 \lesssim N_h \lesssim 164$
Isothermal	$\alpha = 0$	HOCT	-
Isothermal	$\alpha = 1$	HOCT	-
Adiabatic	$\alpha = 0$	HOCT	-
Adiabatic	$\alpha = 1$	HOCT	-

**Table 4.1:** Numerical stability results for perturbations applied according to equation 4.1 for the SC lattice.

### 3D Face Centred Cubic Lattice Results

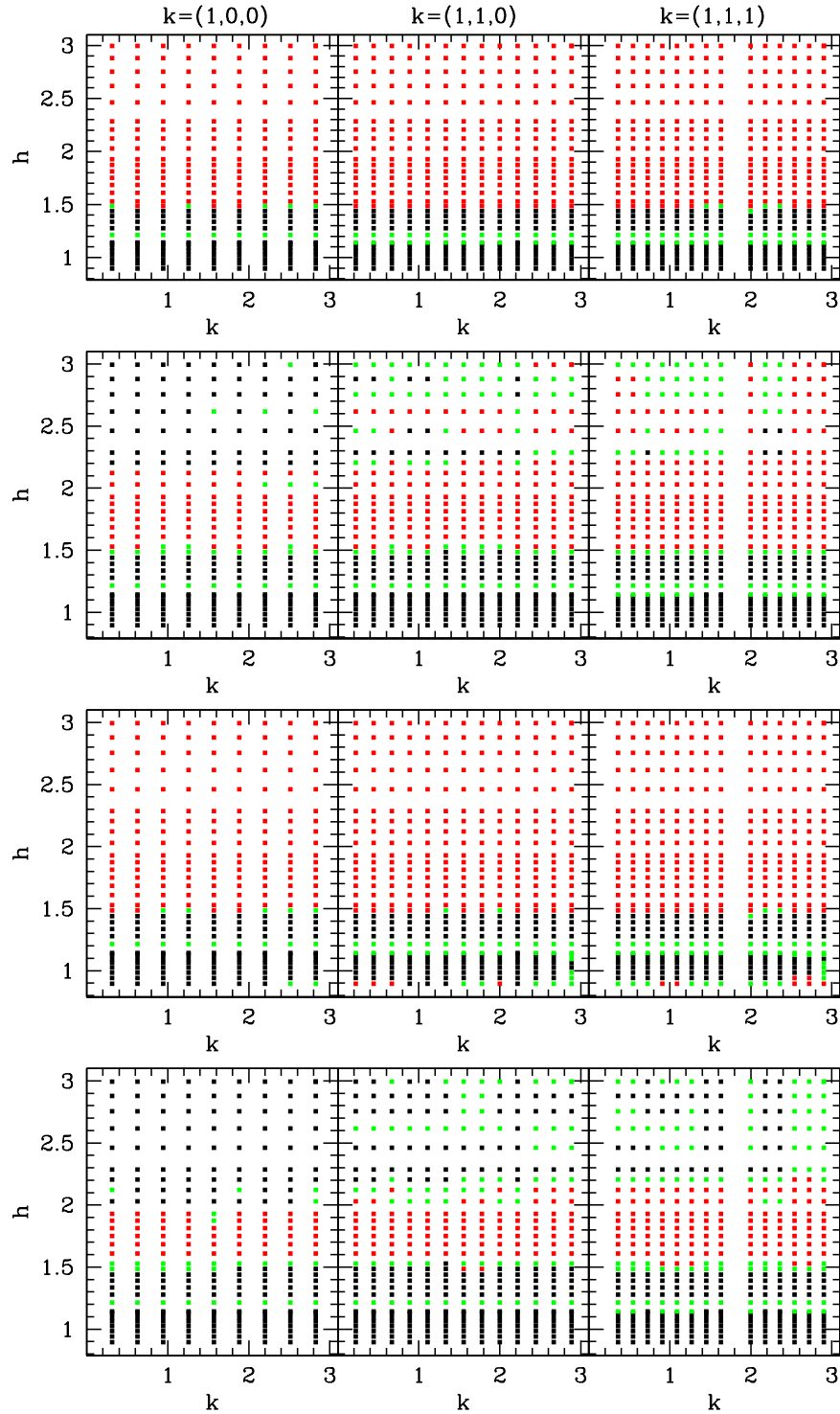
We can perform the stability simulations in 3D using an FCC lattice. The cell separation was set to  $d_c = 0.1$  and the perturbation amplitude was set to 1% of the cell separation, that is  $\mathbf{A} = 0.001\mathbf{k}$  (see equation 4.1). We perturbed the particle positions in three different directions,  $\mathbf{k} = (1, 0, 0), (1, 1, 0), (1, 1, 1)$ . The velocity perturbations were applied in the same direction as the position perturbations. The number of neighbours for an FCC can be computed from  $h$  using the formula

$$N_h = \frac{16\pi}{3}h^3. \quad (4.12)$$

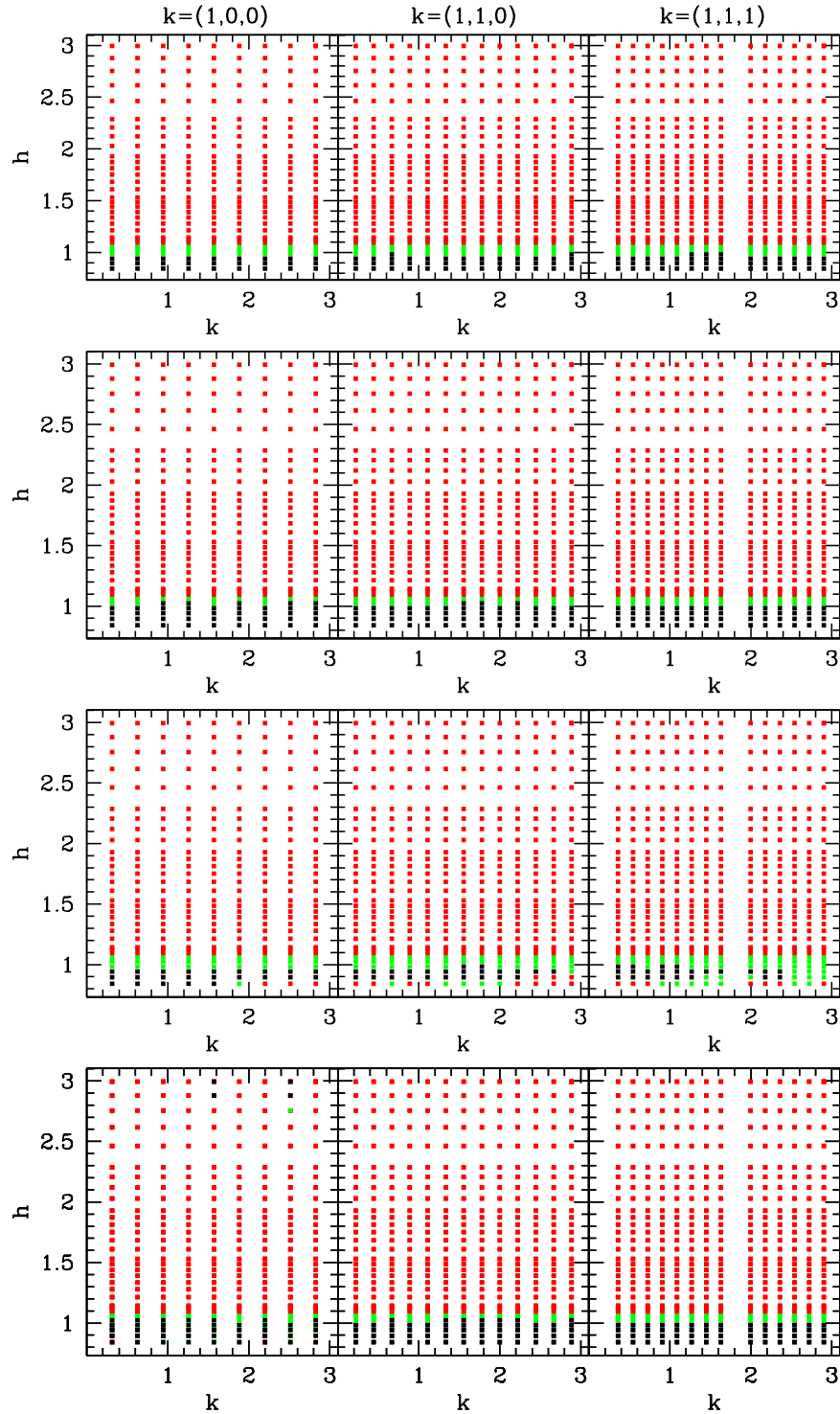
Fig. 4.18 shows the results of perturbations carried out using the standard SPH implementation (see Chapter 2) and the cubic spline kernel. The equation of state used and the direction of the perturbation have very little effect on the stability of the system. However, we find viscosity to be important in the stability with viscosity acting to prevent the system becoming clumped, especially in the regime of  $h > 2$  ( $N_h > 134$ ). This disagrees with the stability analysis performed by Morris (1996) and Cha (2002), however, intuitively one would expect viscosity to prevent clumping as it acts as a repulsive force as particle approach each other. The system is unstable for  $h = 1.5 \rightarrow 2.4$  ( $N_h = 56 \rightarrow 231$ ) for all wavelengths tested regardless of artificial viscosity.

Figure 4.19 shows the effect of using the Ferrers kernel. We find that the simulations are unstable to all wavelengths for  $h > 1.2$  ( $N_h > 28$ ). This corresponds to a nearest neighbour distance of  $q = 0.589$ , which is between the peaks in the first and second derivative. Artificial viscosity has almost no stabilising effect on the results.

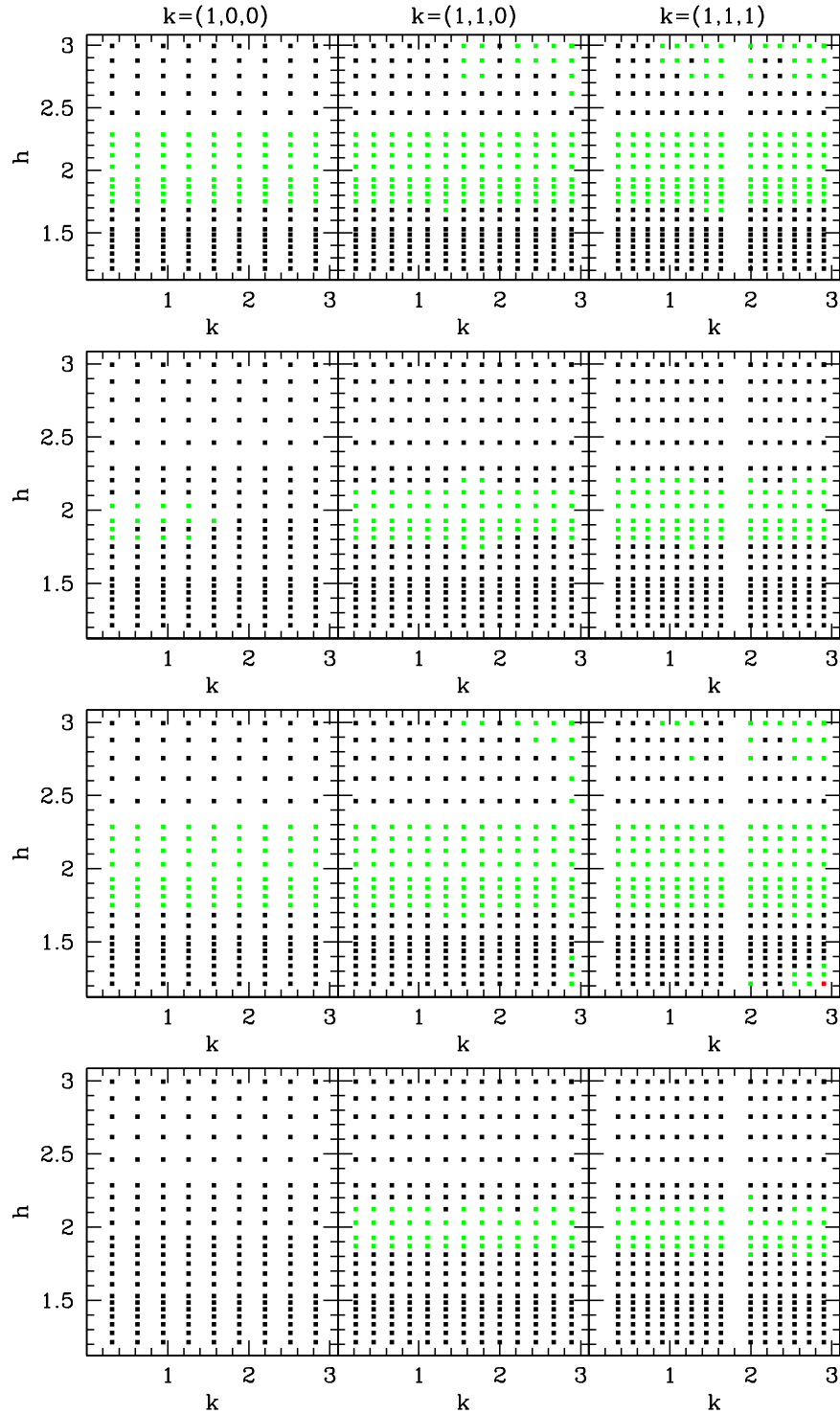




**Figure 4.18:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black) for a FCC lattice using a cubic spline in 3D. **Top row** Isothermal equation of state with no viscosity. **2nd row** Isothermal equation of state with  $\alpha = 1$ . **3rd row** Adiabatic equation of state with no viscosity. **Bottom row** Adiabatic equation of state with  $\alpha = 1$ .



**Figure 4.19:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black) for a FCC lattice using a Ferrers kernel in 3D. **Top row** Isothermal equation of state with no viscosity. **2nd row** Isothermal equation of state with  $\alpha = 1$ . **3rd row** Adiabatic equation of state with no viscosity. **Bottom row** Adiabatic equation of state with  $\alpha = 1$ .



**Figure 4.20:** A map displaying which simulations have undergone particle clumping (red), possible particle clumping (green) and are unaffected (black) for a FCC lattice using the HOCT kernel in 3D. **Top row** Isothermal equation of state with no viscosity. **2nd row** Isothermal equation of state with  $\alpha = 1$ . **3rd row** Adiabatic equation of state with no viscosity. **Bottom row** Adiabatic equation of state with  $\alpha = 1$ .

Equation of State	Viscosity	Kernel	Unstable Region
Isothermal	$\alpha = 0$	Cubic Spline	$N_h \gtrsim 56$
Isothermal	$\alpha = 1$	Cubic Spline	$N_h \gtrsim 56$
Adiabatic	$\alpha = 0$	Cubic Spline	$N_h \gtrsim 56$
Adiabatic	$\alpha = 1$	Cubic Spline	$56 \lesssim N_h \lesssim 231$
Isothermal	$\alpha = 0$	Ferrers	$N_h \gtrsim 28$
Isothermal	$\alpha = 1$	Ferrers	$N_h \gtrsim 28$
Adiabatic	$\alpha = 0$	Ferrers	$N_h \gtrsim 28$
Adiabatic	$\alpha = 1$	Ferrers	$N_h \gtrsim 28$
Isothermal	$\alpha = 0$	HOCT	-
Isothermal	$\alpha = 1$	HOCT	-
Adiabatic	$\alpha = 0$	HOCT	-
Adiabatic	$\alpha = 1$	HOCT	-

**Table 4.2:** Numerical stability results for perturbations applied according to equation 4.1 for the FCC lattice.

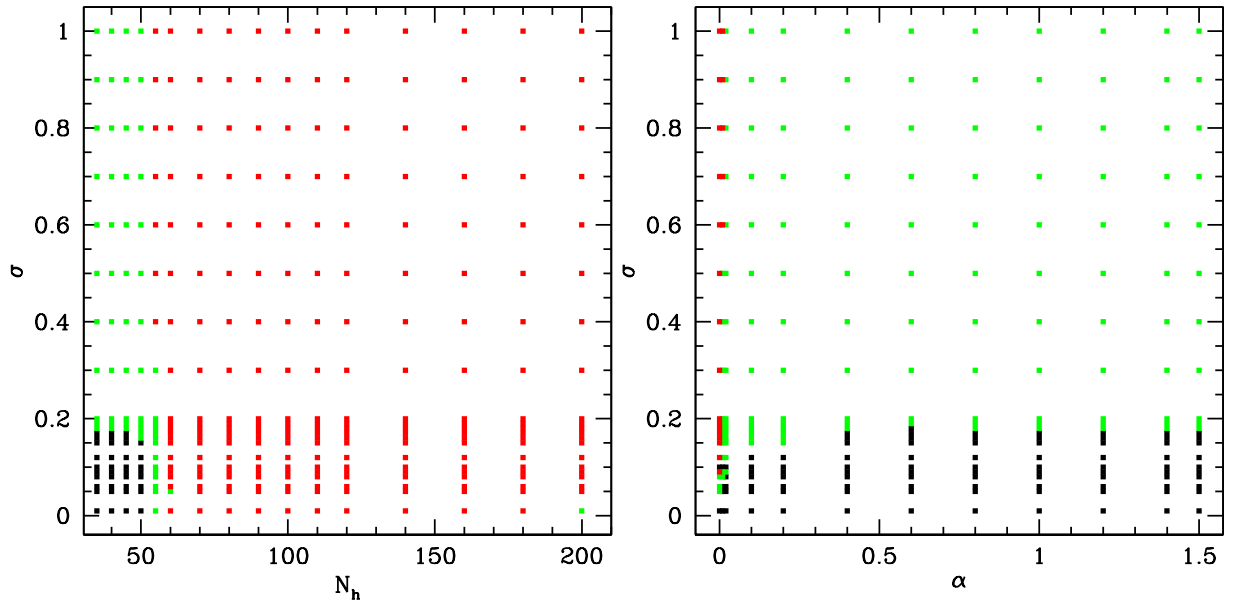
Figure 4.20 shows the effect of using the HOCT kernel. We see that the simulations remain remarkably stable, there was only one simulation where clumping occurred ( $\mathbf{k} = (1, 1, 1)$ , adiabatic with no viscosity). The region which was responsible for clumping in the cubic spline ( $h = 1.5 \rightarrow 2$ ) is again the least stable region, the particles do not return to their unperturbed positions. However, they are prevented from clumping by the cored kernel and eventually form a glass.

The results of both the cubic spline and HOCT kernel are very similar for the SC lattice and the FCC lattice. In the cubic spline case there is an instability band above  $N_h \approx 50$ . The application of artificial viscosity allows the system to become stable when using  $N_h \gtrsim 130$  for the SC lattice and  $N_h \gtrsim 200$  for the FCC lattice. The HOCT kernel keeps both the FCC and SC lattice stable to this type of perturbation. In both types of lattice there is a band where the system does not return to its unperturbed state, but forms a glass. However, the size of the perturbations do not grow and the system avoids the clumping instability.

#### 4.4.4 3D Gaussian Position Perturbations

The analysis carried out by Morris (1996) and the simulations of the previous section assume a sinusoidal perturbation applied in a particular direction. However, we can also investigate the effect of Gaussian perturbations on the stability of the SPH scheme.

Particles are given a perturbation in all three directions drawn from three in-

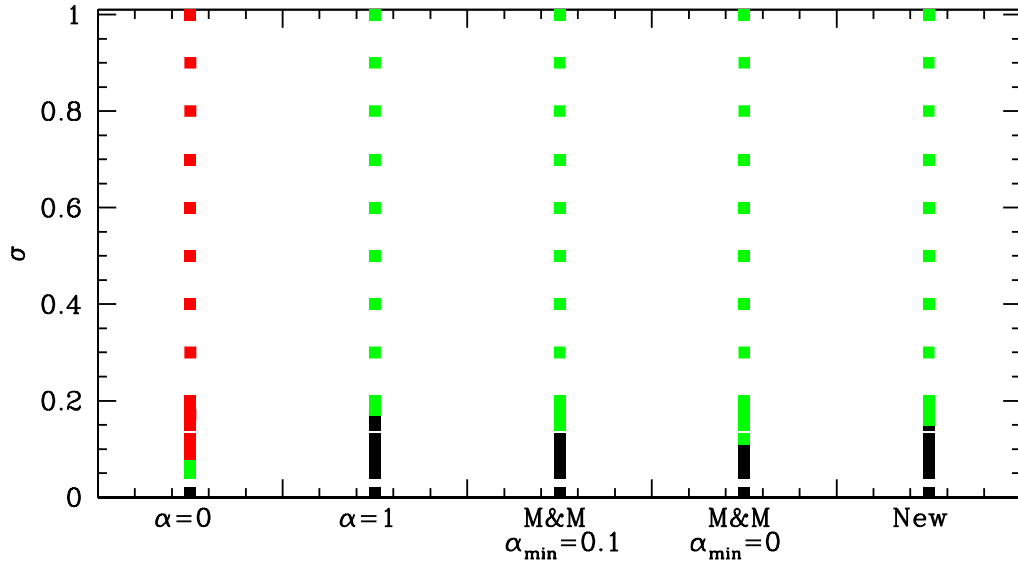


**Figure 4.21:** *Stability properties of an FCC lattice subject to 3D Gaussian perturbations. **Left:** The effects of neighbour number on stability ( $\alpha = 1$ ). **Right:** The effects of viscosity on stability ( $N_h = 40$ ).*

dependent random sequences with a Gaussian distribution centred on zero. The variance of the Gaussian therefore determines how likely it is that large perturbations are present. The variance,  $\sigma$ , is measured in units of the initial minimum particle separation. Therefore  $\sigma = 1$  means there is only a 3% chance that a particle will not suffer an offset as large as the initial separation along at least one direction<sup>5</sup>. In order to allow a fair test between all runs the same seed for the random perturbations is used for each simulation, therefore changing  $\sigma$  merely alters the amplitude of the offset, the normalised distributions are the same for all initial conditions.

There are three possible outcomes to this simulation which we can detect by monitoring the minimum particle separation. If the particle separation returns to the initial unperturbed value then the system has returned to the unperturbed state. If the minimum distance decreases to a stable value lower than the unperturbed value (between 80% and 20%) then the system has evolved into a glass, whereas a minimum value approaching zero means that the system has become unstable and particles have clumped. The affects of artificial viscosity and neighbour number can be tested.

<sup>5</sup>The probability of not suffering an offset as large as  $\sigma$  in a single direction is 0.32. Therefore, the probability of not suffering an offset of at least  $\sigma$  in at least 1 direction is  $0.32^3 \rightarrow 3.2\%$



**Figure 4.22:** *Effect of viscosity method on the stability of Gaussian random perturbations using  $N_h = 40$ . Black points indicate the system returned to its equilibrium state, green points indicate a glass was formed, red points indicate clumping.*

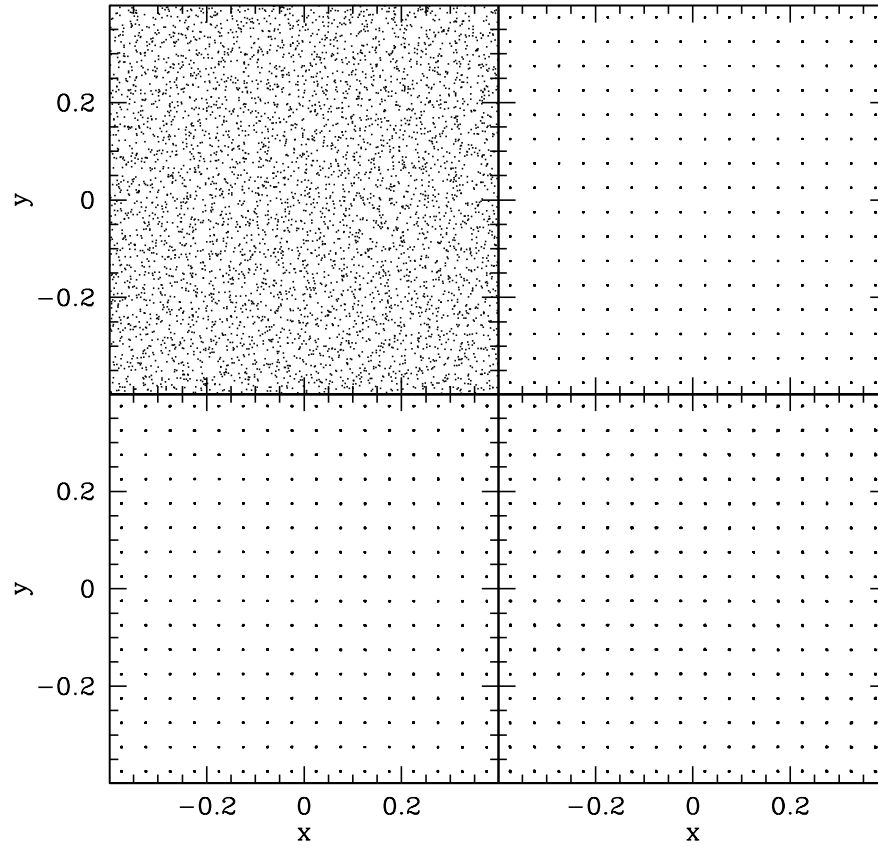
### Neighbours

To test the affect the number of neighbours has on the stability we use a fixed artificial viscosity of  $\alpha = 1$ . Figure 4.21 shows the stable regions in the perturbation-neighbour plane. With around 54 or more neighbours the system is unstable regardless of the size of the perturbations (we tested as low as  $\sigma = 0.01$ ). The region of optimum stability is in the 35-50 range.

### Viscosity and Noise

We saw from the previous section that viscosity acted to stabilise the system, preventing clumping in some cases. However, we did not investigate the strength of viscosity required to prevent clumping. To test the strength of artificial viscosity required we use  $N_h = 40$ , which we found to be stable with a fixed viscosity of  $\alpha = 1$ . We adjust the size of the perturbation and observe the effect the value of artificial viscosity has on the stability of the results.

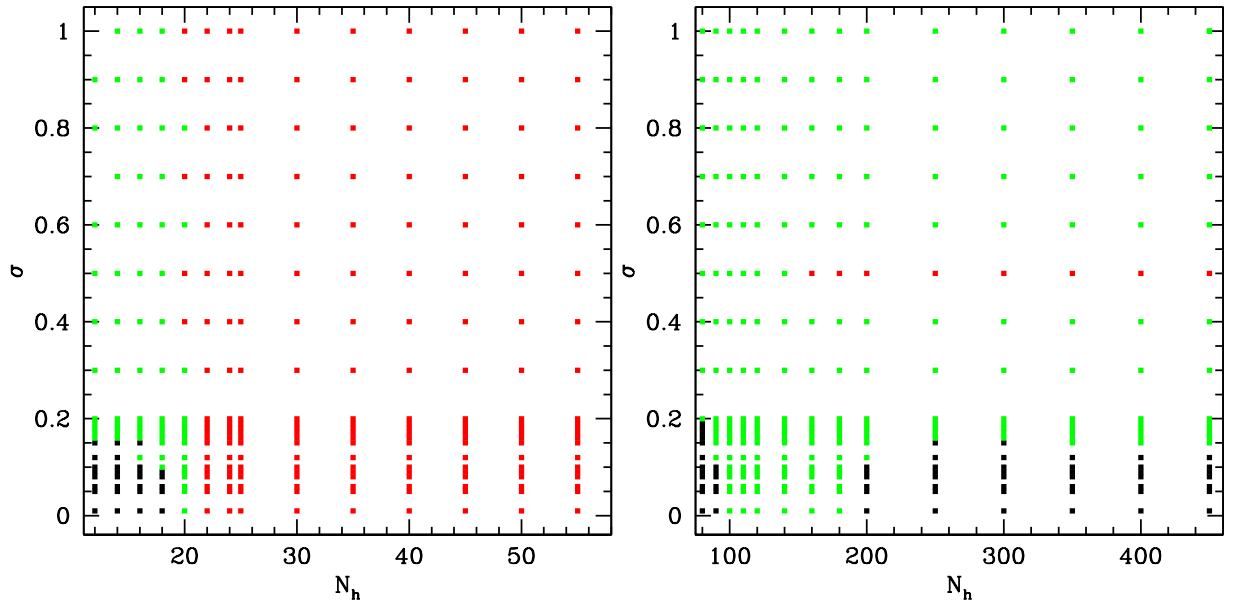
The effect of viscosity can be seen in the right panel of Fig. 4.21. We observe that as the amplitude of the perturbations increases the system moves from stable to glass for a given viscosity. The transitions from stable to glass occurs almost independently of the strength of the artificial viscosity used and depends strongly on the perturbation amplitude. In order to stabilise the perturbations only a minimal



**Figure 4.23:** Particle plots for  $\alpha = 0.0$  (*top left*),  $\alpha = 1.0$  (*top right*), M&M switch (*bottom left*) and the new scheme (*bottom right*) for Gaussian perturbations of amplitude  $\sigma = 0.1$ .

amount of viscosity is required. Perturbations smaller than  $\sigma = 0.1$  are always stable provided some numerical dissipation is present (we tested as low as  $\alpha = 0.01$ ). Once the perturbation size increases beyond  $\sigma = 0.2$  no value of viscosity can restore the system to its FCC lattice configuration. However, the system is able to remain stable by forming a glass.

We may also investigate the affect of artificial viscosity switches (see chapter 3) on the stability properties. Fig. 4.22 shows how the stability changes with initial perturbation size for a variety of viscosity schemes with  $N_h = 40$ . In the absence of viscosity we see that the system is unstable for perturbations above  $\sigma = 0.1$ . However we find that all other viscosity methods give the same stability properties as a fixed  $\alpha = 1$  viscosity. Most notably this includes schemes where  $\alpha_{\min} = 0$ , demonstrating that a minimum viscosity is not required to prevent the growth of numerical



**Figure 4.24:** *Effect of neighbour number on the stability of Gaussian random perturbations using  $\alpha = 1$ . Black points indicate the system returned to its equilibrium state, green points indicate a glass was formed, red points indicate clumping. **Left** shows the Ferrers kernel, **right** shows the HOCT kernel.*

noise, as suggested by many authors, including Morris & Monaghan (1997). Fig. 3.8 shows how the minimum value of  $q$  changes during the course of the simulation for perturbations of size  $\sigma = 0.1$  and  $\sigma = 1.0$  for the various viscosity schemes. For  $\sigma = 0.1$  we observe that the minimum distance decreases only in the case of no viscosity, all other methods allow the perturbations to be damped away. The scheme returns to its initial unperturbed state as shown by the particle plots in Fig. 4.23. With  $\sigma = 1$  we observe that with viscosity the minimum distance increases but does not return to its unperturbed value of  $q = 0.5$ . Instead particle clumping is avoided by the formation of a glass. In the absence of artificial viscosity there is no mechanism to prevent particles becoming arbitrarily close, therefore the minimum distance remains low indicating particle clumping. We also point out that the time taken for the minimum distance to increase is almost independent of the strength of the viscosity. Larger viscosities cause the system to return to a stable state only slightly faster than smaller viscosities. However, the transition to a more stable state occurs more smoothly for larger viscosities, as can be seen in the  $\sigma = 1$  results. This is because a smaller viscosity allows particles more freedom to oscillate.



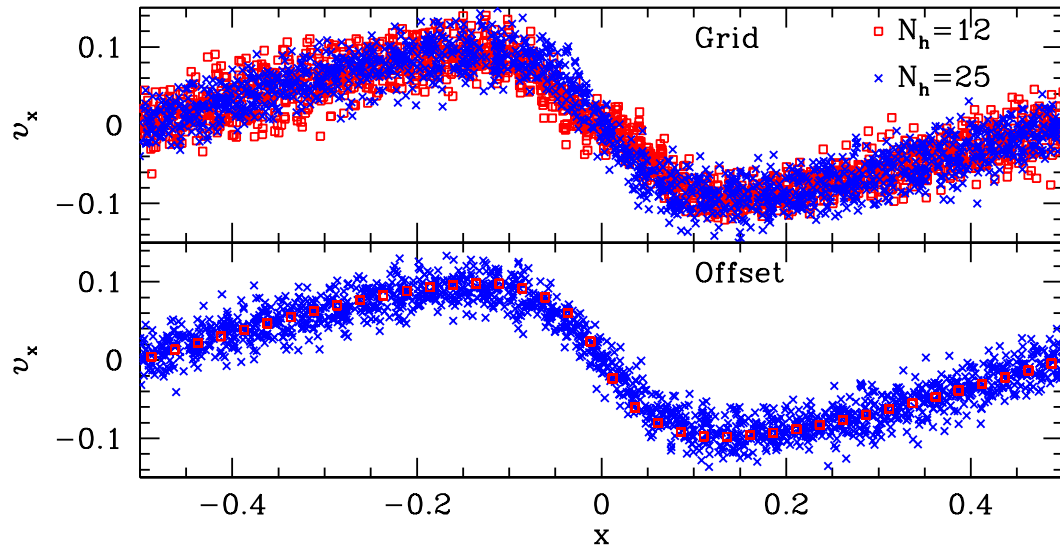
#### 4.4.5 Effect of Kernel on Particle Clumping

We may briefly summarise the finding of the stability experiments as follows. In order to prevent particle clumping the velocity of approaching pairs must be reversed. Therefore it can be readily seen that artificial viscosity must act to prevent clumping as viscosity acts to prevent particle penetration. This therefore easily explains the increased stability for a given neighbour number caused by allowing artificial viscosity.

An interesting point to note is the region where stability changes is the same in both the Gaussian perturbations and the wave perturbations for the cubic spline. The simulations become unstable for  $N_h > 60$  ( $q \approx 0.46$ ,  $h \approx 1.5$ ) which is just outside the peak in the first derivative of the kernel located at  $q = 1/3$ . Furthermore we find that for the Ferrers kernel the system becomes unstable at  $N_h > 22$  ( $q \approx 0.642$ ,  $h \approx 1.1$ ) (see Fig. 4.24 and Fig. 4.19) which is also just outside the peak in the first derivative of the kernel located at  $q = 0.378$ . One may have naively expected the boundary to be at the peak in the first derivative ( $q = 1/3$ ,  $N_h = 160$  for the cubic spline and  $q = 0.378$ ,  $N_h = 110$  for the Ferrers kernel) as this is the location of the maxima in the first derivative. At this point as particles move closer together the force between them drops causing particles to clump. However, the analysis of Swegle et al. (1995) suggest a criterion of  $W'' > 0$  for stability which occurs further from the kernel centre than the first derivative peak; around  $q \sim 0.5$  for the cubic spline and  $q \sim 0.6$  for the Ferrers kernel. The HOCT kernel is very stable as the force never diminishes as two particles approach one another.

### 4.5 2D Sound-wave Steepening

To illustrate how initial conditions in SPH can have a dramatic effect on the outcome of a simulation, we look at the steepening of a 2D sound wave. We take a 2D box of size  $(-0.5, 0.5), (-0.5, 0.5)$  with  $40 \times 40$  particles. The sound speed was set to  $c = 1$  and the particles were given velocities  $v_x = 0.1 \sin(2\pi x)$ ,  $v_y = 0$ . The viscosity scheme described in section 3.4 is used. In 1D this viscosity scheme is able to suppress viscosity until the wave steepens, at which point a small amount of viscosity is applied allowing the wave to dissipate entropy and steepen (see section 3.6.1 and Fig. 3.10 for more details). We perform this simulation using both the grid and offset grid initial conditions described in section 4.2 with both  $N_h = 12$  and  $N_h = 25$ . Fig. 4.25 shows the results with the different initial conditions and



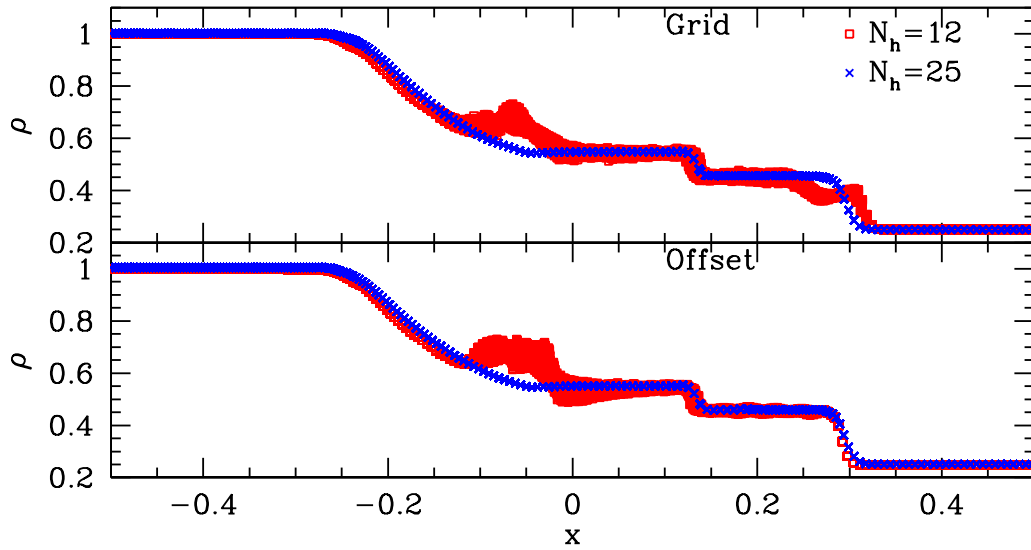
**Figure 4.25:** *The effect of the initial particle distribution and neighbour number on the steepening of a 2D sound wave.*

neighbour numbers after  $t = 2.5$ . At this point the wave is just beginning to steepen. We find that with the grid initial conditions the velocity wave becomes extremely noisy. This is because particle noise has crept into the simulation, resulting in noisy density and velocity profiles. With the offset grid initial conditions we find that using  $N_h = 12$  is able to suppress this particle noise and is able to resolve the steepening of the wave well. With  $N_h = 25$  the particles tend to clump together, again causing noise in the velocity wave.

## 4.6 2D Sod Shock Tube

We may also investigate how the particle distribution and neighbour number play an important role in the evolution of the 2D Sod shock tube test. We use 2D periodic boundaries of size  $(-1, 1)$ ,  $(-0.2, 0.2)$  with a total of 45,000 particles. Fig. 4.26 shows that there is little difference between the grid and offset grid initial conditions. However, there is significant difference between using  $N_h = 12$  and  $N_h = 25$ . With  $N_h = 12$  the density profile is very noisy and contains unphysical kinks and dips. However, with  $N_h = 25$  the density profile is accurate and there is little noise.

Fig. 4.27 shows the particle distribution around the contact wave. With  $N_h = 12$  the particles remain well separated but are no longer in their initial configuration, a glass has formed. Due to the density profile across the contact wave being somewhat



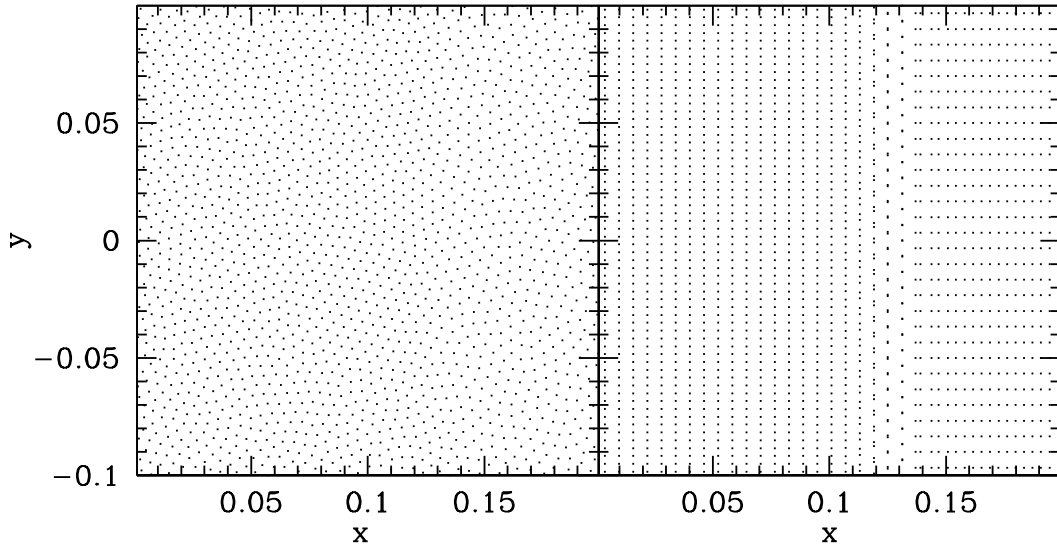
**Figure 4.26:** *The effect of the initial particle distribution and neighbour number on the density profile of the 2D Sod shock tube.*

smudged, and that there is only a small density ratio across the contact, it is difficult to see a sharp density contrast in the particle positions. With  $N_h = 25$  the particles either side of the contact wave remains in the initial configuration. However, we notice that at the contact wave front particles have clumped in the  $y$  direction.

## 4.7 Conclusion

In this chapter we looked at how initial conditions can play an important role in the stability and accuracy of SPH simulations. We demonstrated that particles set up using a random sequence often produces results which are noisy. This noise can be reduced by firstly settling the particles into a glass by allowing the distribution to run with artificial dissipation.

We then investigated the well known clumping instability, where around 2-5 particles clump together for no physical reason. Analysis by Morris (1996); Cha (2002); Read et al. (2010) suggested that there certain neighbour numbers which excite the clumping instability. The exact neighbour number where this occurs depends on the kernel used (Read et al., 2010). We performed a numerical verification of this analysis by perturbing uniform lattices and monitoring the minimum particle separation. The numerical results agreed to some degree with the theoretical analysis, although the exact neighbour number at which the instability occurred differed. We



**Figure 4.27:** *The effect of neighbour number on the particle distribution around the contact wave. Left shows  $N_h = 12$ , right shows  $N_h = 25$ .*

found that for the widely used cubic kernel, SPH is unstable to perturbations when using  $N_h \gtrsim 50 \rightarrow 60$ . However, we found that the stability properties are vastly improved when using even a small amount of artificial viscosity (see tables 4.1 and 4.2). We also demonstrated that the HOCT kernel is much more stable than the analysis by Read et al. (2010) suggested, which opens the possibility of using less neighbours, decreasing the computational cost and increasing the resolution.

Finally we demonstrated the affect neighbour number can have on the results of a simulation. We find that in 2D simulations that require the suppression of artificial viscosity a lower neighbour number is required to suppress particle clumping. However, in simulations where shocks are important we find that more neighbours are required to provide enough smoothing of the shock front.

---

# 5

## Stellar Disruption

### 5.1 Introduction

Stellar disruptions may occur when a star passing close to a super-massive black hole (SMBH) becomes tidally disrupted, leading to a partial or complete break up of the star. This scenario was originally suggested (Lynden-Bell, 1969; Hills, 1975, 1978) as a mechanism to fuel active galactic nuclei (AGN). However, it was quickly realised that the stellar concentration required to allow tidal disruption to fuel typical AGN luminosities ( $L \sim 10^{43} - 10^{44} \text{ erg s}^{-1}$ ) would induce other processes such as stellar collisions and supernova (Young et al., 1977; Frank, 1978).

However, the stellar disruption mechanism was later used (Rees, 1988) as a possible mechanism to explain observed X-ray flares from inactive galaxies. By considering the energy of the mass disrupted from the star, Rees (1988) was able to derive the luminosity of the flare as a function of time, finding<sup>1</sup> that  $L \propto t^{-5/3}$ . Observations of flares such as the one observed by Komossa & Bade (1999), are consistent with a characteristic light curve of  $t^{-5/3}$ , although the light curve is not well sampled. The main characteristic of the outbursts were a short duration X-ray peak of  $L_x \gtrsim 10^{42-43} \text{ erg s}^{-1}$ , with no change in optical luminosity. They also detected an optical spectrum with no signs of permanent Seyfert activity, implying that the galaxy is not permanently active.

---

<sup>1</sup>In his paper, Rees (1988) gives the luminosity as  $L \propto t^{-5/2}$ . This is a mathematical error and was corrected by Phinney (1989).

In order to determine the origin of the flare, Komossa & Bade (1999) considered several outburst scenarios. They concluded that

- Supernovae are not bright enough, the peak luminosity is only  $L_x \approx 10^{35} \text{ erg s}^{-1}$ , with the brightest peak reaching  $\approx 10^{41} \text{ erg s}^{-1}$ .
- Gravitational lensing would produce the same magnification in optical as well as X-rays. As there is no optical variability gravitational lensing is unlikely.
- No Gamma-Ray Burst (GRB) was detected around the time of the X-ray outburst, making it unlikely to be an X-ray afterglow of a GRB. However it is possible that the GRB was not detected.
- An accretion disc instability with a central black hole of mass  $\sim 10^4 - 10^5 M_\odot$  could account for the peak in the luminosity.
- Tidal disruption of a star with black hole of at least  $\sim 10^4 - 10^5 M_\odot$  could explain both the high peak luminosity and the light-curve profile.

They conclude that tidal disruption is best candidate, although much of the theoretical details are still not understood.

SPH simulations (Evans & Kochanek, 1989; Lodato et al., 2009) of polytropes on parabolic orbits that reach their tidal disruption radius at pericentre (see section 5.2.1) have shown that the light curve is proportional to  $t^{-5/3}$  only at late times. At earlier times it was found that the light curve is shallower than this and the deviations are stronger for more concentrated polytropes (i.e. lower polytropic index  $\gamma$ ). However, thus far only the affect of polytropic index has been investigated with SPH. There are many more parameters which can be investigated such as pericentric distance, eccentricity, spin and mass ratio. This leaves a large parameter space which can be explored with SPH simulations. In this chapter we will investigate how these five parameters affect the disruption of stars which encounter a SMBH.

## 5.2 Theory

### 5.2.1 Tidal Disruption

In this section we give an overview of the theoretical arguments of Rees (1988), from which the predicted light curve of  $L \propto t^{-5/3}$  is obtained. We assume that the star is

initially far from the black hole and is in hydrostatic equilibrium. We also assume that the centre of mass of the star is on a parabolic orbit around the black hole. Therefore the gas moves in Keplerian orbits around the black hole with each fluid element orbiting with slightly different eccentricities than the centre of mass of the star (Lodato et al., 2009).

As the star moves on its orbit around the SMBH the edges of the star will experience slightly different forces due to the differences in distance from the SMBH. If this difference in force is larger than the self-gravity force which binds the star, fluid elements on the surface of the star are able to escape the star's self-gravity. At this point the star begins to break up. We can derive an approximate expression for the distance at which the star begins to break up if we consider a rigid satellite of mass  $M_*$  and radius  $R_*$ , orbiting a black hole of mass  $M_{\text{BH}}$  at a distance  $R$ . The force acting on a mass element  $m$  at the edge of the satellite due to its self-gravity is

$$F_{\text{SG}} = \frac{GM_* m}{R_*^2}. \quad (5.1)$$

The gravitational attraction due to the black hole on the satellite at  $R \pm R_*$  is

$$F_{\text{BH}} = \frac{GM_{\text{BH}} m}{(R \pm R_*)^2} = \frac{GM_{\text{BH}} m}{R^2(1 \pm R_*/R)^2} \approx \frac{GM_{\text{BH}} m}{R^2} \mp \frac{2GM_{\text{BH}} m}{R^2} \frac{R_*}{R}. \quad (5.2)$$

Therefore, the difference in force felt at the edge of the satellite compared to the centre of the satellite, i.e. the tidal force, is

$$F_T = \frac{2GM_{\text{BH}} m}{R^2} \frac{R_*}{R}. \quad (5.3)$$

When the tidal force exceeds the self-gravity force, the satellite is pulled apart. The radius  $R_T$  at which this occurs is

$$R_T \propto q^{1/3} R_*, \quad (5.4)$$

where the mass ratio is  $q = M_{\text{BH}}/M_*$ , and the constant of proportionality is of order unity<sup>2</sup>. The proportional symbol accounts for the ability of a real star to deform slightly before break up. Therefore the approximations used to derive this expression are not strictly true. However, this still gives a good estimate of where

---

<sup>2</sup>When we refer to this equation we assume a constant of proportionality of unity

the tidal disruption radius occurs.

The edges of the star experience a spread in energy caused by the difference in the black hole potential at each fluid element in the star, which is maximal at pericentre. The spread in energy  $\Delta E$ , at pericentre  $R_p$ , to first order is

$$\Delta E = \frac{GM_{\text{BH}}}{R_p^2} R_\star. \quad (5.5)$$

Therefore the star obtains an energy distribution between  $-\Delta E$  and  $\Delta E$ . As the star was initially symmetric this implies that half of the star's material is on an orbit which is bound to the SMBH, and half of the material is on an unbound orbit.

However, this alone is not sufficient to determine what happens to the gas. A star far from the tidal radius will not experience a break up, despite half of the mass being bound to the black hole. This is because the fluid elements in the star are also bound to the centre of mass of the star, which is on a parabolic orbit. Providing the gas is not unbound from the star (for example this occurs for stars well outside the tidal radius), the gas bound to the black hole is dragged on a parabolic orbit with the centre of mass of the star and is therefore not accreted. If the gas becomes unbound from the star then its fate depends on the energy with respect to the black hole, if it is bound to the black hole it is assumed that it will be accreted, otherwise it will escape on a hyperbolic orbit. The distance at which the star can become unbound is of the order of the tidal radius, as given by equation (5.4).

### 5.2.2 The Light Curve

If we now assume, for simplicity, that the star passes the black hole at the tidal disruption radius given by equation (5.4) and is completely unbound. The energy distribution is symmetric about zero as half of the star is further from the black hole than the centre of mass and half is closer. We assume that once the bound material returns to pericentre the gas loses its energy and angular momentum on a timescale much shorter than the return time. Therefore we may consider the gas to be suddenly accreted on to the black hole as soon as it returns to pericentre (Evans & Kochanek, 1989; Lodato et al., 2009). Of course in reality the material must first enter an orbit, forming an accretion torus. Mass is then propagated inwards and angular momentum transported outwards by viscous effects, causing mass to be accreted causing the flare.

The bound mass will be on elliptical orbits with a range of specific orbital energies



with respect to the black hole  $E_{\text{BH}}$  (see equation 5.11). The time taken for mass of energy  $E_{\text{BH}}$  to return to pericentre is given by

$$T = 2\pi \frac{GM_{\text{BH}}}{(2E_{\text{BH}})^{3/2}}. \quad (5.6)$$

Therefore the higher energy material returns faster than the low energy material. The rate of mass returning to pericentre (where it is assumed to be accreted) is simply

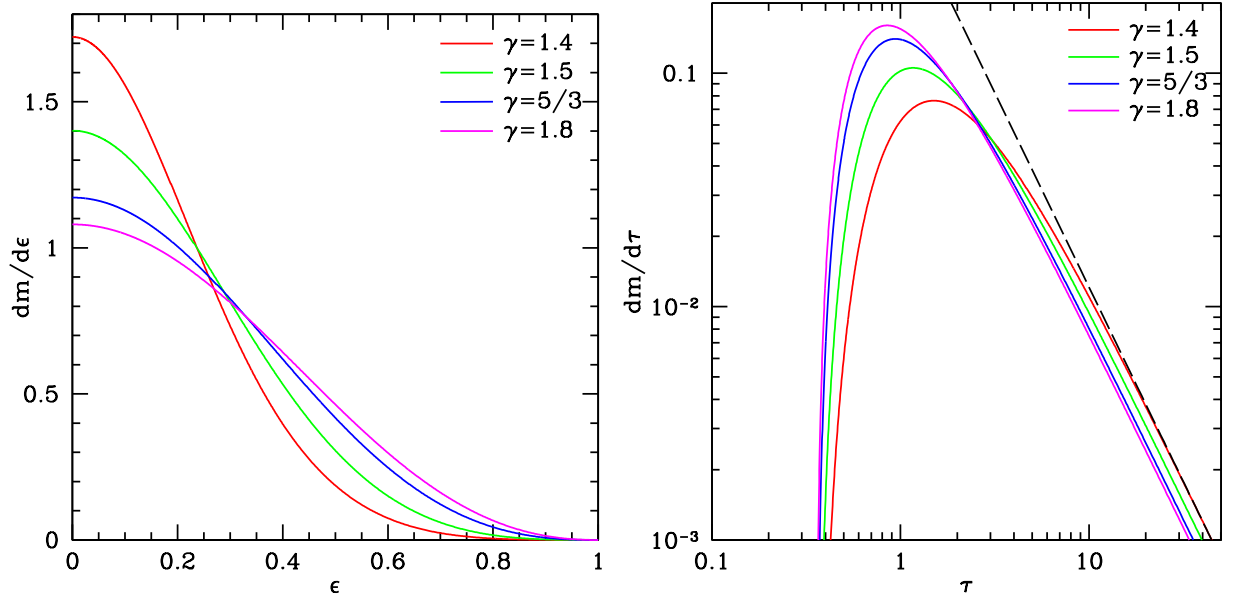
$$\frac{dM}{dT} = \frac{dM}{dE} \frac{dE}{dT} = \frac{(2\pi GM_{\text{BH}})^{2/3}}{2} \frac{dM}{dE} T^{-5/3}. \quad (5.7)$$

Therefore the light curve obtained is  $\propto t^{-5/3}$  only when the energy distribution is uniform, as was assumed by Rees (1988). Indeed as Lodato et al. (2009) demonstrate, the mass distribution is not uniform for polytropic spheres, and therefore the light curve is expected to differ from  $t^{-5/3}$ . Furthermore, even if the mass distribution in the polytrope was uniform, as the star approaches the black hole and is perturbed, the hydrostatic equilibrium inside the star is disrupted. This causes pressure forces to act which redistribute material inside the star, changing the mass-energy distribution.

We may compute the theoretical mass distribution with specific energy,  $dM/dE$ , which is shown in Fig. 5.1. We assume that the polytrope is shattered which essentially means that the particles of gas that make up the polytrope are considered to no longer interact gravitationally with each other. Therefore the particles remain on their individual orbits around the black hole. As discussed previously this is unlikely, however it will enable us to approximate the mass distribution therefore allowing us to make a prediction about the light curve shape. We observe that the slope of the light curve only approaches  $t^{-5/3}$  at late times, with the light curve initially decaying more slowly (Fig. 5.1). We also observe that the more centrally condensed polytropes (lower  $\gamma$ ) have a lower peak with a longer rise time.

From equations (5.6) and (5.7) we see that the energy  $E_{\text{BH}}$  of the particles gives the return time of the gas and the mass distribution  $dM/dE$  gives the accretion rate (and therefore the luminosity). Finally we define the “penetration factor” which determines how strong the tidal disruption is

$$\beta = \frac{R_p}{R_T}. \quad (5.8)$$



**Figure 5.1:** *Left:* We obtain numerically  $dm/d\epsilon = dm/dx = 2\pi \int_x^1 \hat{\rho}(x')x'dx'$ , where we use stellar radial coordinate  $x = r/R_*$ , dimensionless density  $\hat{\rho} = \rho/\rho_0$ , dimensionless energy  $\epsilon = -E_{\text{BH}}/\Delta E$  and fiducial density  $\rho_0 = M_*/R_*^3$ . *Right:* The mass accretion rate with time compared to a  $\tau^{-5/3}$  power law where  $\tau = T/T_0$  and  $T_0 = 2\pi\sqrt{R_p^3/GM_{\text{BH}}}$ .

All previous work (e.g. Evans & Kochanek, 1989; Lodato et al., 2009) has only considered parabolic orbits ( $e = 1$ ) where the pericentre distance is equal to the tidal disruption radius ( $\beta = 1$ ). Here we extend this to a range of parabolic orbits and also look at hyperbolic and elliptic orbits.

### 5.3 Initial Conditions

We model the star as a  $10^5$  particle polytrope with a mass and radius of unity (in code units). The polytrope is initially allowed to settle into equilibrium by using an isentropic equation of state (see equation 2.42) with a fixed viscosity of  $\alpha = 2$ ,  $\beta = 4$ . This somewhat large viscosity allows small perturbations in the particles velocities to be quickly damped away. The energy lost by reducing particle velocities is not converted into thermal energy, rather it is lost from the system. This ensures that the polytrope has a constant entropy throughout and remains as close to the original polytropic model as possible. Once the kinetic energy of the polytrope has reached an acceptable level ( $\sim 10^{-7}$ ) the internal energy is set using the density and

entropy.

The star is then placed on a parabolic orbit. We initially place the star far from pericentre,  $r_0 = (10R_p, 0, 0)$ . Using the pericentre velocity which is tangential to the direction of motion  $v_p = v_t = 2GM_{\text{BH}}/R_p$  and conservation of energy and angular momentum<sup>3</sup>, we may determine the initial velocity; this is given by

$$v_{y,i} = \left( \frac{2GM_{\text{BH}}}{r_0} \frac{R_p}{r_0} \right)^{1/2}, \quad (5.9a)$$

$$v_{x,i} = \left( \frac{2GM_{\text{BH}}}{r_0} \left[ 1 - \frac{R_p}{r_0} \right] \right)^{1/2}. \quad (5.9b)$$

We use a polytrope with adiabatic index  $\gamma = 5/3$  and represent the black hole with a point mass of  $M_{\text{BH}} = 10^6 M_*$  initially<sup>4</sup> centred on  $(0, 0, 0)$ . We use an ideal-gas equation of state (see equation 2.43,  $\gamma = 5/3$ ).

We can determine the energy of the gas with respect to the most bound gas particle. This allows us to determine if any of the gas is still bound together gravitationally. The energy of the gas is computed as

$$E_{\text{gas}} = \frac{1}{2} v_{\text{gas}}^2 + u + \phi_{\text{gas}}, \quad (5.10)$$

where  $v_{\text{gas}}$  is the velocity of the gas with respect to the most bound particle,  $u$  is the specific internal energy of the gas and  $\phi_{\text{gas}}$  is the potential due to the self-gravity of the gas. The most bound gas particle is the one with the lowest (most negative) internal plus potential energy (from just the gas particles, i.e.  $u + \phi_{\text{gas}}$ ). We then assume that this particle is the centre of the polytrope and calculate the velocity of each particle with respect to the centre particle. This, combined with the internal and potential energy gives the total energy  $E_{\text{gas}}$  of each particle in the polytrope. If any particles have  $E_{\text{gas}} > 0$  they are unbound from the polytrope and therefore may be accreted by the black hole. We may determine which particles are accreted by considering the specific energy of the gas particles with respect to the black hole; that is

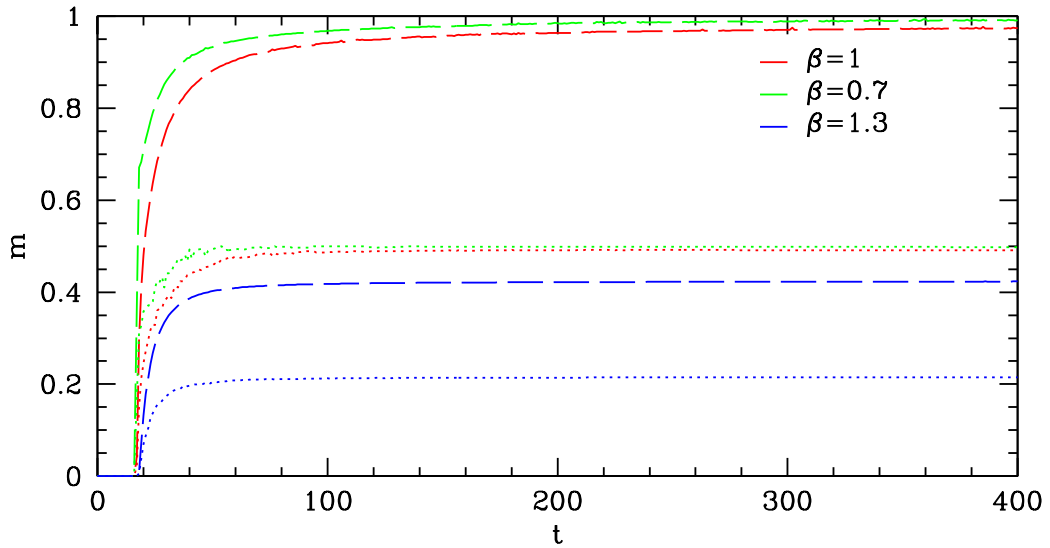
$$E_{\text{BH}} = \frac{1}{2} (v - v_{\text{BH}})^2 - \frac{GM_{\text{BH}}}{|r - r_{\text{BH}}|}, \quad (5.11)$$

where  $v$  is the velocity of the gas. To compute  $dM/dE$  we firstly bin the gas particles

---

<sup>3</sup>Conservation of energy  $\frac{1}{2}v^2 - \frac{GM}{r} = 0$  and angular momentum  $L = rv_t$

<sup>4</sup>The point mass is free to move, thus conserving momentum.



**Figure 5.2:** Dashed line shows mass loss from the star as a function of time for three pericentre encounters  $\beta = 0.7$ ,  $\beta = 1$  and  $\beta = 1.3$ . The dotted line shows the mass of gas which is both unbound from the star and bound to the black hole, therefore showing the amount of mass that will be accreted. The time is given in code units where one time unit is equal to  $1588s$  if the mass and radii are in solar units and  $G = 1R_{\odot}^3M_{\odot}^{-1}T^{-2}$

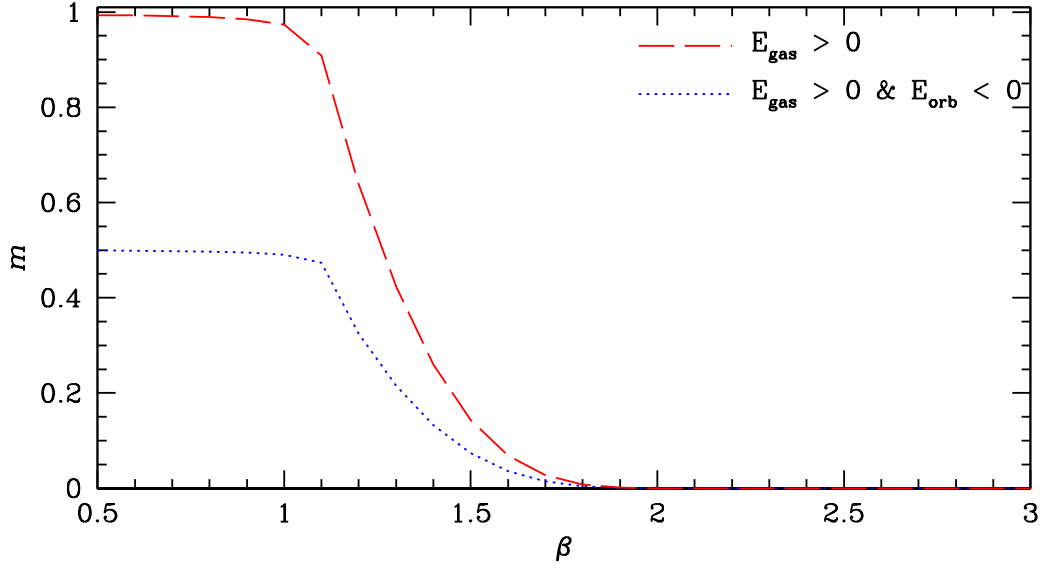
according to their  $E_{\text{BH}}$ . We then compute the mass in each bin and divide it by the energy range of the bin. This allows us to compute the mass accretion rate using equations (5.6) and (5.7).

## 5.4 Varying Pericentric Distance

As a starting point we firstly repeat the experiment of Evans & Kochanek (1989) (see also Lodato et al., 2009). We also investigate the affect of pericentric distance on the disruption of the star.

### 5.4.1 Mass Loss from the Star

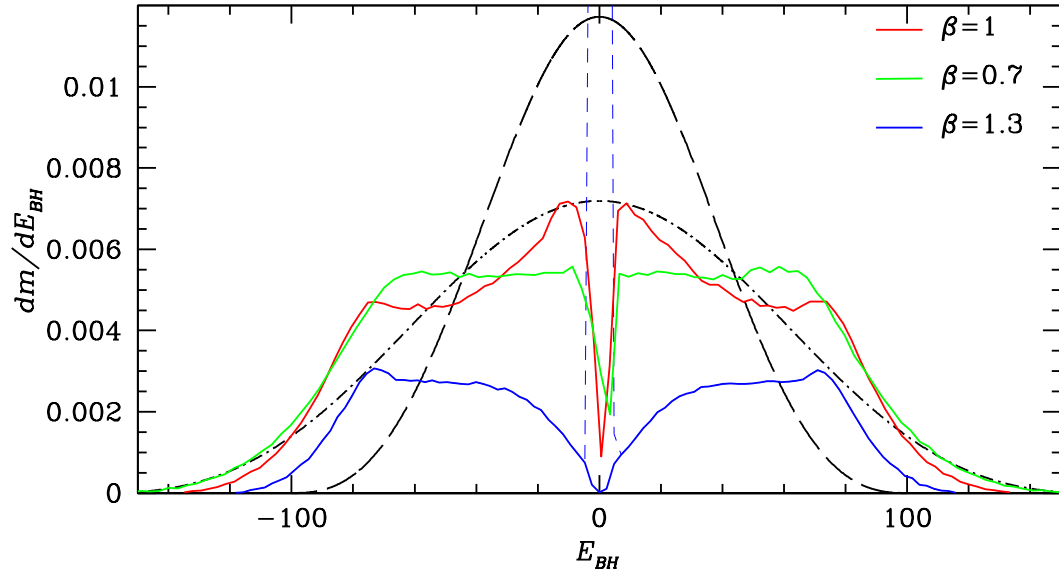
Fig. 5.2 shows how the mass lost from the star and mass accreted by the black hole changes over the course of the simulation for three pericentre encounters ( $\beta = 0.7$ ,  $\beta = 1$  and  $\beta = 1.3$ ). We find that for closer encounters (i.e.  $\beta = 0.7$ ) the star is disrupted earlier. This is as expected as the star reaches the tidal disruption radius quicker for closer pericentre encounters. After  $t = 100$  the stars in the  $\beta = 0.7$  and  $\beta = 1.0$  encounters slowly lose more mass due to the redistribution of energy via



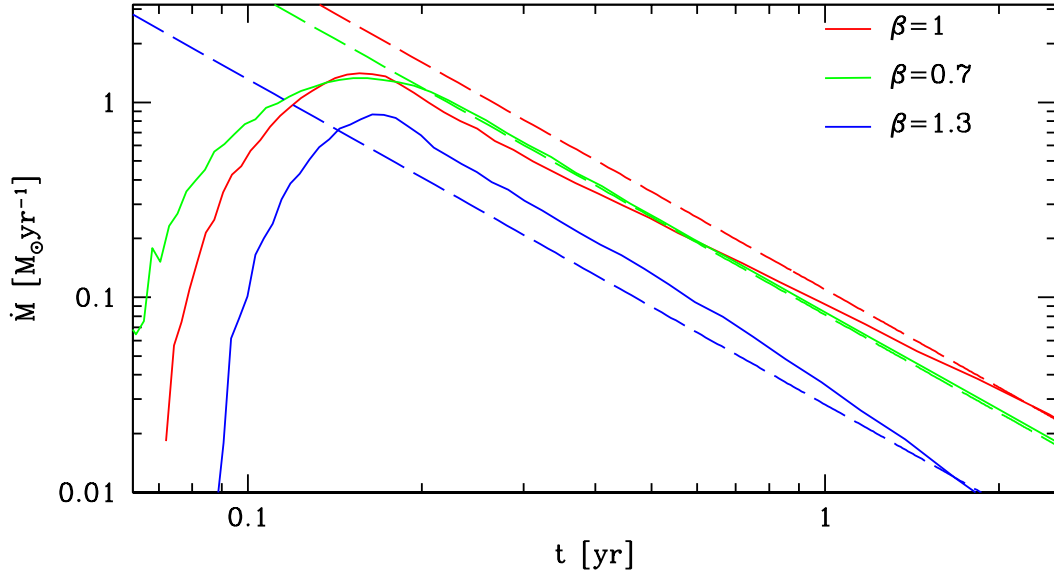
**Figure 5.3:** The mass unbound from the star (red dashed line) and mass accreted by black hole (blue dotted line) as a function of pericentre distance.

shocks. However, in the  $\beta = 1.3$  encounter the star loses a significant part of its mass (around 40%) but quickly stabilises to prevent further mass loss. The amount of matter accreted by the black hole (dotted line in Fig. 5.2) is half of the mass unbound from the star. This is because the star was initially on a parabolic orbit with half of the mass bound to the black hole and half unbound. This gas will return to the black hole on elliptical orbits in a time given by equation (5.6), at which point it is quickly accreted giving rise to a flare.

The final mass unbound from the star (at  $t = 400$ ) for various pericentre distances are shown in Fig. 5.3. The mass unbound from the star remains roughly constant at around 99% of the total star mass for  $\beta < 1$ . At this point the mass unbound decreases sharply until all of the mass remains bound to the star when  $\beta \geq 2$ . Similarly the mass accreted by the black hole drops sharply in the range  $\beta = 1 \rightarrow 2$ . Therefore as pericentric distance is increased, the star becomes less tidally affected until around  $\beta = 2$ . At  $\beta = 2$  we find that the outer part of the star does become extended and the internal structure is clearly affected by the tidal encounter. However the mass remains bound to the star and is therefore not accreted by the black hole. This is in good agreement with equation (5.4), suggesting that a star is tidally disrupted in the range  $\beta \lesssim 1 \rightarrow 2$ .



**Figure 5.4:** The distribution of specific energies against energy with respect to the black hole. The dashed line shows the predicted distribution from a  $\gamma = 5/3$  star, the dot-dashed line shows the prediction from a homologously expanded star (by factor of 1.63). In these units, the expected spread in energy (equation 5.5) is  $\Delta E = 100$



**Figure 5.5:** Mass accretion rate as a function of time. The units refer to a star with  $M_* = M_\odot$  and  $R_* = R_\odot$ . The light curves differ depending on pericentric distance.

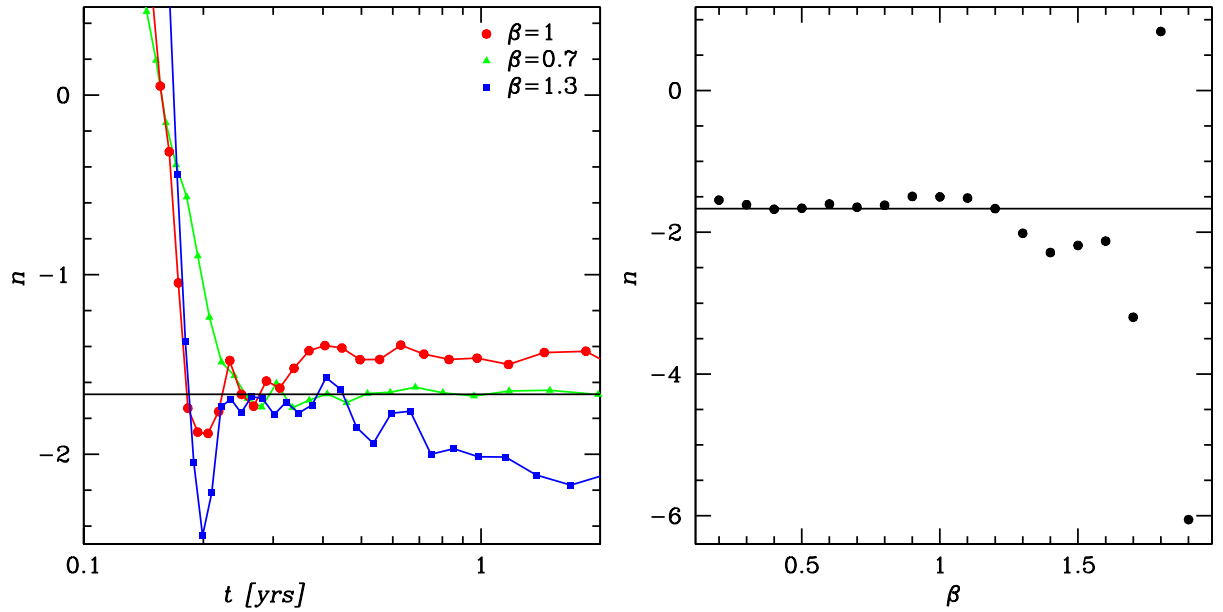
### 5.4.2 Specific Energy Distribution

The distribution of specific energies is shown in Fig. 5.4, where we compare the SPH results of three encounters with different pericentre distances ( $\beta = 0.7, 1, 1.3$ ) to the analytical result (for  $\beta = 1$ ). The mass distribution will determine how quickly material returns to pericentre where it is assumed to be accreted instantly.

The first thing to note is that not all of the mass distribution is accreted. Particles with energies greater than zero are not bound to the black hole and instead escape on hyperbolic orbits. Furthermore particles near the centre of the distribution, which are energetically bound to the black hole, are not accreted as they are still bound to the star. This gas will not return to the black hole as it is dragged by the star which is on a parabolic orbit. This results in a dip in the mass distribution around the original orbital energy of the star. The dotted line for  $\beta = 1.3$  highlights the difference in the mass distribution if we include the gas which is bound to the star, we observe a large amount of gas at low energy. It is important to discard this material when considering the light curve as it will artificially enhance the tail of the light curve.

For encounters with  $\beta = 0.7$  and  $\beta = 1.0$ , the mass of gas unbound from the star is still increasing asymptotically towards unity at  $t = 400$  (see Fig. 5.2). This is due to shocks in the remaining star transferring energy leading to further unbinding. Therefore we expect the star to become completely unbound by the time the bound material returns ( $t \sim 2000$  for  $E_{\text{BH}} \sim 100$ ). However, this is not the case for  $\beta = 1.3$ , at  $t = 400$  there is clearly no more mass loss from the star implying an attenuation to the light curve at late times. We can take this into account when computing the light curve in the following section. However we find that the effect on the light curve only occurs at very late times (i.e. around  $t \sim 10$ ) which is outside the range shown in Fig. 5.5.

The SPH results for  $\beta = 1$  agree well with that of Lodato et al. (2009). The distribution exhibits wings which are at higher energies than predicted by the model. This is because the star has expanded due to the tidal interaction which is not taken into account when computing the analytical prediction. The analytical prediction assumes that the star approaches the centre of mass unaffected by the black hole until it reaches pericentre, at which point it is shattered with particles remaining in the same positions, although no longer bound together. However, this is not the case, the star will expand as it approaches pericentre. If we take the expansion of the star into account we find that the total spread in energy is similar to that



**Figure 5.6:** *Left:* Light curve power law slope plotted against time for three encounters. *Right:* Light curve slope against pericentre distance plotted at  $t = 1$  yr.

predicted (the magnitude of the spread in energy is simply due to the size of the star). However, we also find that there is still more mass at higher energies than predicted as can be observed by the presence of the wings. This is caused by the transfer of energy through shocks in the tidal tails.

Fig. 5.2 also demonstrates the effect of pericentre distance on the shape of the light curve. The closer pericentre encounters (i.e.  $\beta = 0.7$ ) show an almost uniform distribution of mass at all energies. As the interaction weakens we find that the energy distribution becomes more centrally condensed indicating that the star is less affected by the encounter.

### 5.4.3 The Light Curve

The effect of changing the specific energy profile is to change the slope of the light curve decay, in accordance to equation (5.7).  $\beta = 1$  produces a light curve shallower than  $t^{-5/3}$  whereas  $\beta = 1.3$  produces a light curve steeper than  $t^{-5/3}$ . This is demonstrated in Fig. 5.6 where we plot<sup>5</sup> the instantaneous power against time for the three encounters  $\beta = 0.7, 1, 1.3$ . We see that after the initial peak the  $\beta = 0.7$

<sup>5</sup>Here we assumed units of solar mass and radii for the star, although the power would be the same regardless of units



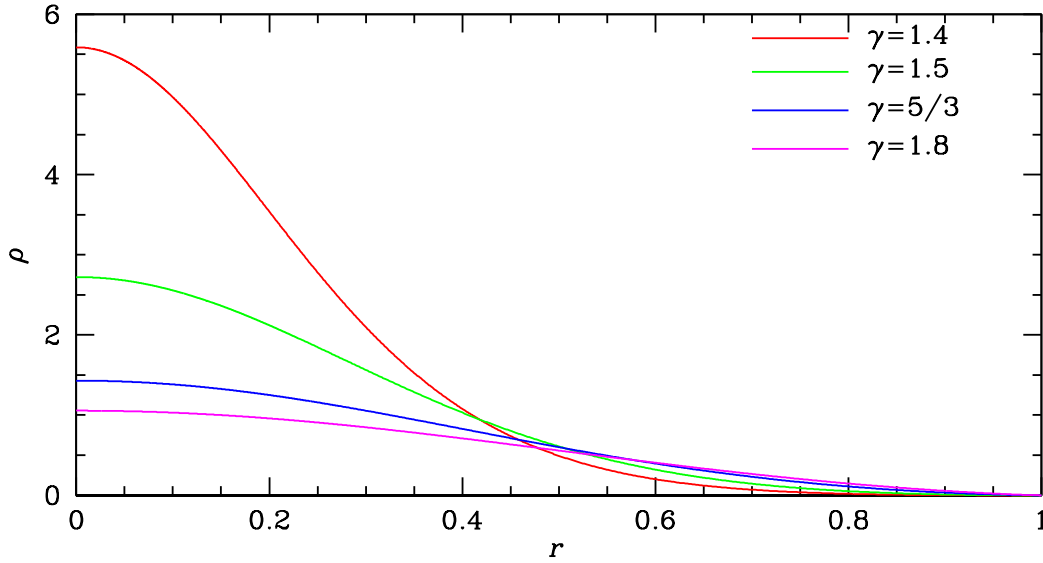
encounter quickly reaches a uniform  $t^{-5/3}$  power law. The  $\beta = 1$  encounter quickly settles to a fixed slope which is shallower than the expected  $t^{-5/3}$ . The light curve for  $\beta = 1.3$  is still changing towards the tail, gradually becoming steeper. This indicates that the amount of mass accreted at late times drops off much faster than a uniform mass distribution. This is because of the dip in the mass distribution curve (see Fig. 5.4) caused by part of the star remaining intact. The right panel of Fig. 5.6 shows that as  $\beta$  increases towards unity the power law becomes progressively shallower until around  $\beta = 1.1$ , at which point the power law becomes progressively steeper.

The light curves for the three encounters (shown in Fig. 5.5) also display different rise times, with lower values of  $\beta$  rising faster. This is because for smaller values of  $\beta$  there is more material with higher energies (see Fig. 5.4) which through equation (5.6) leads to a shorter return time. We find that the peak in the light curve occurs at the same time in all three methods, this corresponds to the peak of the wing in the specific energy distribution, which occurs at the same energy in all three cases (see Fig. 5.4). However we find that the peak accretion rate is lower for  $\beta = 1.3$  than the other two cases as less of the star is disrupted and therefore less mass accreted.

The difference in the mass distributions and therefore the light curves is simply due to the strength of the tidal interaction. For  $\beta = 1.3$  the star is only marginally affected by tides, in particular most of the core of the star remains intact. This leaves the specific energy distribution strongly peaked at the centre ( $E_{\text{BH}} = 0$ ). However, as the gas is still bound to the star the gas is not accreted by the black hole, only the tidal tails are accreted. For  $\beta = 0.7$  the star is strongly affected by tides as it is inside the tidal radius for longer, leading to a full disruption of the core of the star. As  $\beta$  increases the core of the star is able to remain stable and therefore produces a more peaked specific energy distribution.

## 5.5 Varying Polytropic Index of the Star

Now we investigate the affect of changing the polytropic index,  $\gamma$ . By changing the polytropic index we change the mass-radius and density-radius relationship of the polytrope. Increasing the polytropic index amounts to making the star less centrally condensed (see Fig. 5.7). A polytrope with  $\gamma = 1.4$  is considered to be a good model for sun like stars,  $\gamma = 5/3$  polytropes are usually used to model giant stars whilst  $\gamma = 2$  polytropes are used to model neutron stars.

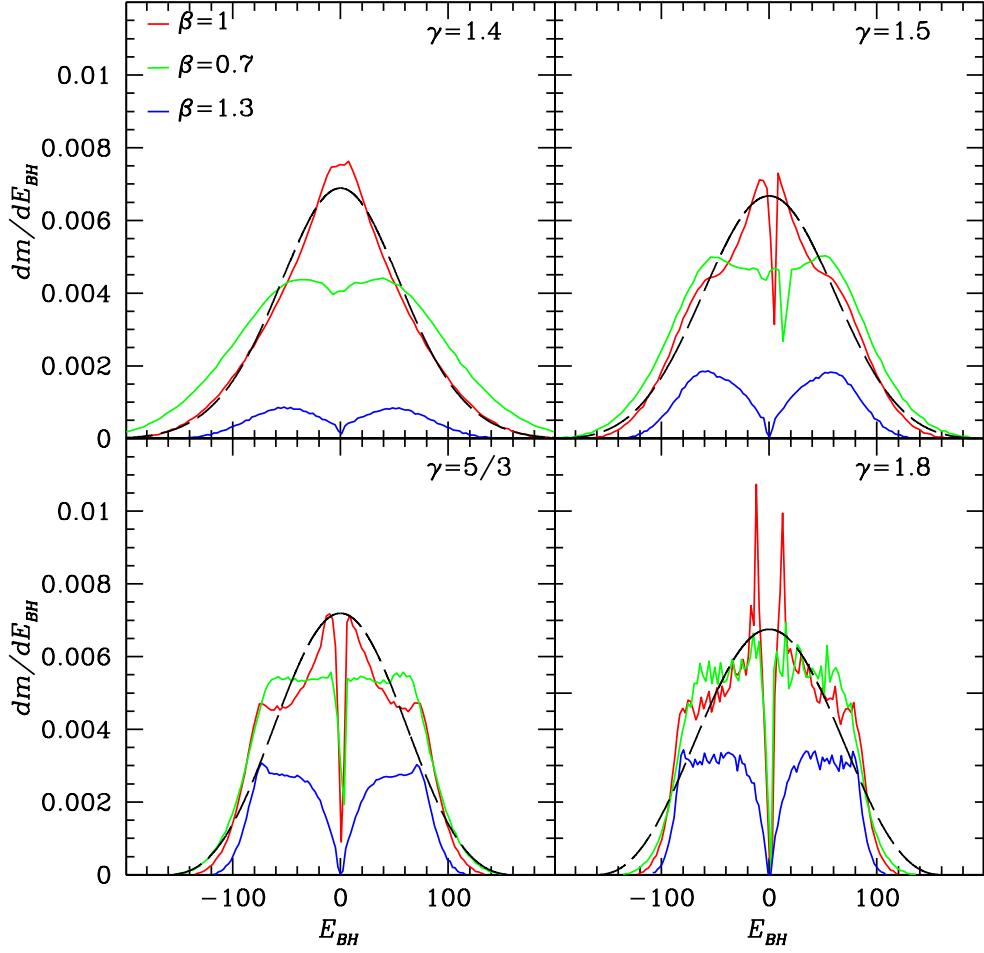


**Figure 5.7:** *Density-radius relation for various polytropic models.*

Fig. 5.1 shows the expected specific energy distribution and expected light curve for a  $\beta = 1$  encounter for various polytropic models, assuming the polytrope's density is unaffected by the encounter. We expect the type of polytrope used to affect both the peak of the flare and the rise time with less centrally condensed models having a higher peak. This is due to models with higher  $\gamma$  having more mass in the outer parts of the polytrope and therefore a higher energy when disrupted. We also expect the peak to occur earlier with less centrally condensed models for the same reason.

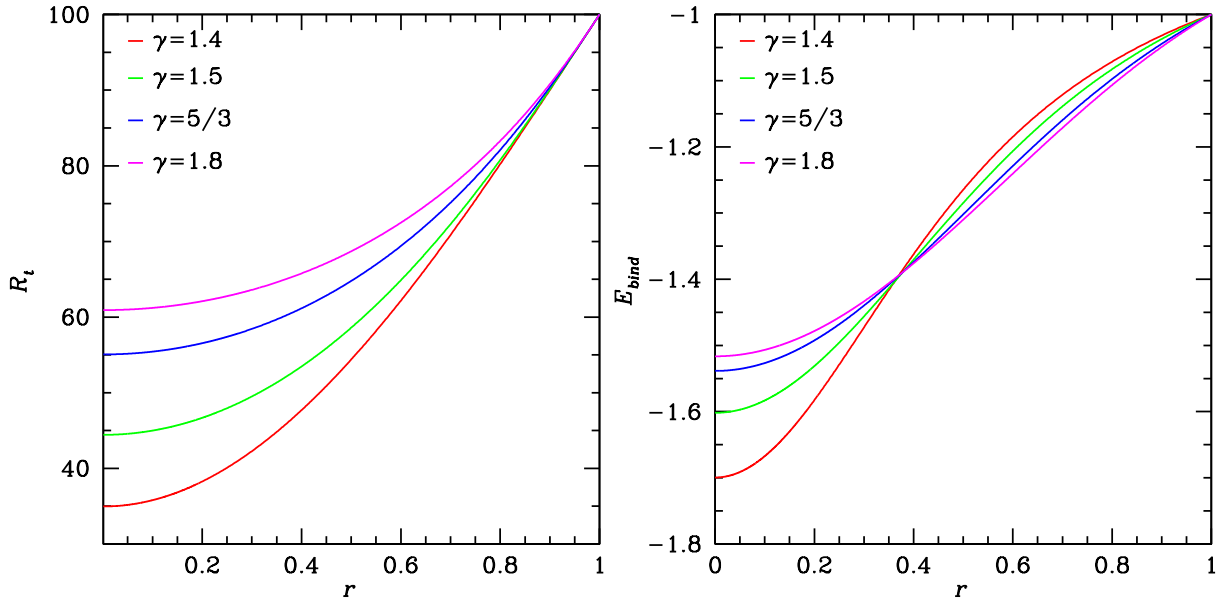
### 5.5.1 The Specific Energy Distribution

The mass distribution is shown in Fig. 5.8 for the three encounters considered in the previous section for four polytropic models  $\gamma = 1.4, 1.5, 5/3$  and  $1.8$ . We clearly see a change in the behaviour of the distribution with increasing  $\gamma$ . Low values of  $\gamma$ , which represent more centrally condensed stars, show a distribution which is more centrally peaked, with less prominent wings. For  $\gamma = 1.4$  we find that the homologously expanded model fits the SPH data particularly well with only slight deviations. This indicates that the star becomes expanded but there is little change to the internal structure. As  $\gamma$  increases the appearance of wings becomes more noticeable indicating that the star is not only expanded, the internal structure is also altered. This is particularly noticeable for  $\gamma = 1.8$ , where the mass distribution is almost uniform, even in the  $\beta = 1.3$  encounter.



**Figure 5.8:** *Distribution of energies for various polytropic models. The analytical solutions (dashed lines) assume the polytropes are homologously expanded by  $\xi = 2.5, 2.1, 1.63, 1.6$  for each polytropic model  $\gamma = 1.4, 1.5, 5/3, 1.8$  respectively.*

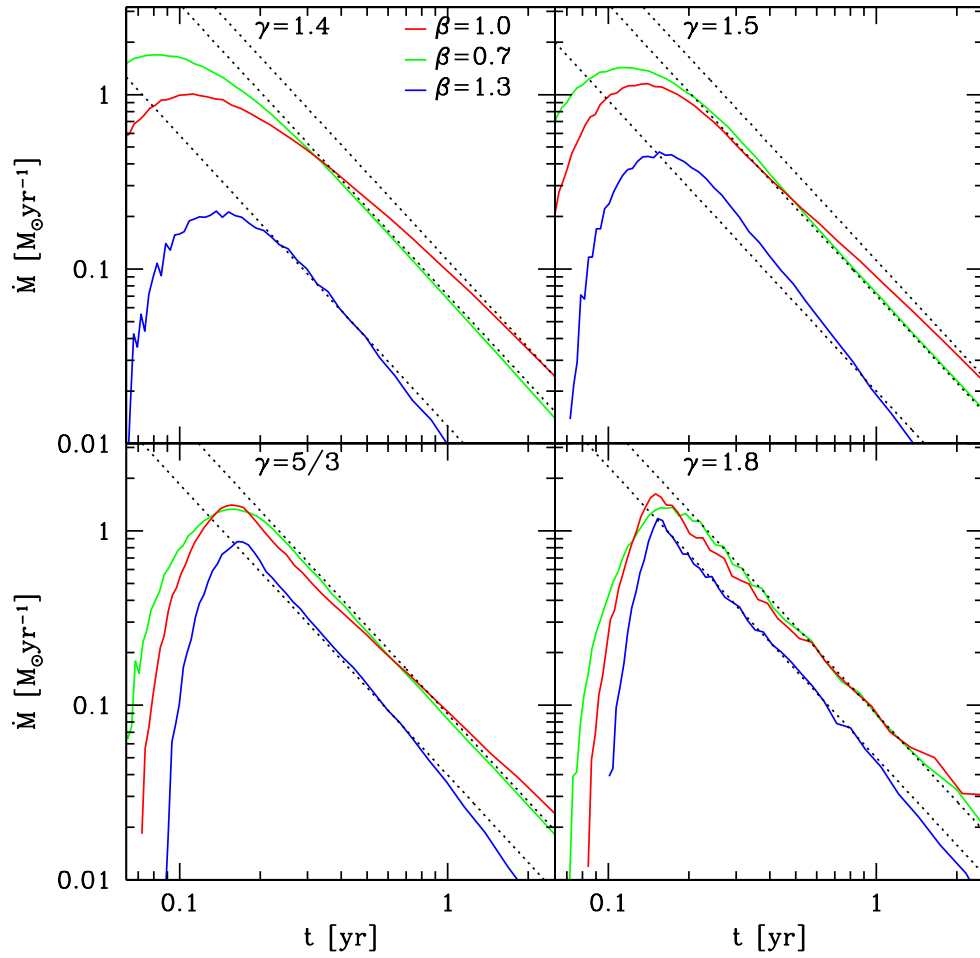
We can understand this trend if we consider how strongly material at different radii are bound to the centre of the star. One measure of this is the binding energy shown in the right panel of Fig. 5.9. This is simply found assuming that the gas particles are stationary, giving the binding energy as  $E_{\text{bind}} = \phi + u$ . However, another useful measure is to compute the tidal disruption radius for points inside the star by applying equation (5.4) by replacing  $R_*$  with  $r$  and  $M_*$  with  $m(r)$ . Whilst this is a somewhat simplistic measure, it does give an indication of how strong the tidal interaction has to be to affect different regions of the star. This measure will not account for the effect mass loss in outer regions has on inner regions, or the effect of shocks in redistributing energy.



**Figure 5.9:** *Left:* The tidal radius of a particle as a function of radius from the centre of the star for various polytropic models. *Right:* The binding energy of gas as a function of radius from the centre of the star for various polytropic models.

Fig. 5.9 shows that as the polytropic index increases the disruption radius of particles close to the centre increases, as does the binding energy (i.e. the gas is less bound). This indicates that the core of the star is more affected by tides as  $\gamma$  increases. Therefore we would expect the energy distribution to become more uniform with higher  $\gamma$  as more of the star is fully disrupted. This simple prediction agrees well with the SPH results that show that polytropic models with higher  $\gamma$  give a more uniform distribution for lower  $\beta$  (e.g.  $\gamma = 1.8$  and  $\beta = 1.3$  compared to  $\gamma = 1.4$  and  $\beta = 1$ ). In models with low  $\gamma$  the core is not destroyed via shocks rather it is expanded, therefore exhibiting a centrally peaked mass distribution. However, this expansion is still able to unbind material from the star with the same effect (in fact we find slightly more material is unbound from the star with  $\gamma = 1.4$  than  $\gamma = 5/3$  for the  $\beta = 1$  encounter). This result is also in agreement with that of Lodato et al. (2009) who find that the amount of mass shocked in centrally condensed stars is less than that for uniform (higher  $\gamma$ ) stars.

As the pericentre is decreased, all polytropic models are more strongly affected by tides, in particular the core of the star becomes more affected by tides. This leads to a progressively more uniform distribution rather than a gentle stretching of the



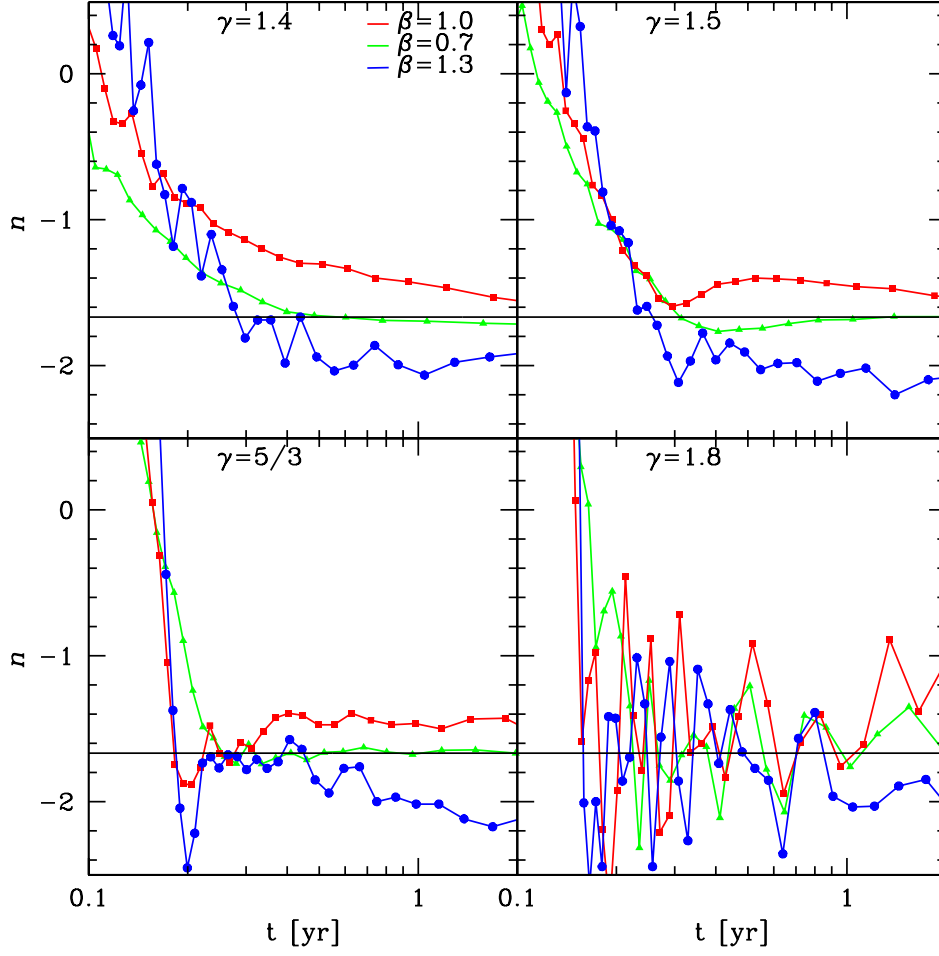
**Figure 5.10:** *Light curves for four polytropic models ( $\gamma = 1.4, 1.5, 5/3, 1.8$ ) for three pericentre encounters  $\beta = 0.7, 1, 1.3$ .*

star. Therefore we can say that the energy distribution becomes more uniform as  $\beta$  decreases and  $\gamma$  increases. This is because both changes cause the core of the star to experience tidal forces which are closer to (or greater than) the star's self-gravity.

### 5.5.2 Light Curves

#### Decay Slope

The light curves for various  $\gamma$  and  $\beta$  are shown in Fig. 5.10. We firstly observe that for  $\beta = 0.7$  the decay of the light curve at late times goes as  $t^{-5/3}$  for all four polytropic models. This suggests that all polytropic models are strongly disrupted leading to a uniform distribution of mass with energy, particularly at low energies.



**Figure 5.11:** *Instantaneous power for various polytropic models ( $\gamma = 1.4, 1.5, 5/3, 1.8$ ) for three pericentre encounters  $\beta = 0.7, 1, 1.3$ .*

This is in agreement with the mass distributions shown in Fig. 5.8. We also find that the time taken to reach this  $t^{-5/3}$  slope decreases with increasing polytropic index (see Fig. 5.11), again suggesting that the core of the star is more strongly disrupted in less centrally concentrated models leading to more uniform mass distributions.

The light curves also suggest that at late times the power law slopes are almost independent of polytropic model with all models approaching a  $t^{-1.55}$  power law slope for  $\beta = 1.0$  and a  $t^{-2}$  power law for  $\beta = 1.3$ . However, the time taken to reach the final slope does depend on the polytropic model with more condensed stars taking longer.

### Rise Time

Changing the polytropic model and encounter parameter not only change the light curves decay slope, they also affect both the rise time and the peak accretion rate. The light curves show that the rise and peak occur earlier for more centrally condensed stars (smaller  $\gamma$ ).

This occurs because the outer parts of the star have a lower (less negative) binding energy in more centrally condensed the stars (see Fig. 5.9). Therefore this material is ejected with a higher energy leading to the larger spread in energy (see Fig. 5.8) and so there is more material with a shorter return time. The rise is also steeper with stars with larger  $\gamma$  due to the outer parts of the star having less mass at higher binding energy but having a more uniform distribution in the energy range  $E_{\text{BH}} = 0 \rightarrow 80$ . This means that the spread in energies is smaller but there is more mass at a given energy, leading to light curves with late, steep rises.

### Light curve Peak

The peak accretion rate of the light curve is strongly dependent on both the polytropic model used and the pericentre of the encounter. For weak encounters, i.e. those just outside the tidal radius, the peak of the light curve is lower for more centrally condensed stars.

This is because in weak encounters most of the mass will remain in the core of the star, with less mass at high energies. Therefore most of the mass is returned at late times, or in some cases not at all as it is still bound to the star and therefore carried away on a parabolic orbit. This means that more uniform stars that have less mass in the core and more material in the outer parts, end up with more mass at higher energies. This leads to more uniform stars having higher peaks and more centrally condensed stars having a lower peak in weak encounters.

As the interaction gets stronger the central regions of the star are more strongly disrupted leaving more material at higher energies. This leads to a stronger peak in more centrally condensed stars as the mass is spread over a smaller energy range than in more uniform stars. Therefore the strongest peaks occur for centrally condensed stars in strong interaction or for more uniform (higher  $\gamma$ ) stars in weak interactions.

## 5.6 Spinning Star

We investigate the affect a spinning star has on the results. The star is spun up slowly such that it has a fixed angular velocity. For each angular velocity we extract the spun up star and evolve it on its own in order to allow it to settle into an equilibrium. We then place the star in a parabolic orbit around the SMBH with varying pericentre distance. We set up the orbit such that we investigate both prograde and retrograde orbits.

In order to understand the affect that spin can have on the interaction we used a relatively large spin,  $\omega = 0.4$  in code units. In our code units, assuming the star has  $M = M_{\odot}$  and  $R = R_{\odot}$  the sun has a spin of  $\omega = 4.6 \times 10^{-3}$  and the break up spin of the star is approximately  $\omega_{\max} \approx 1$ . Whilst the spin used may be unrealistic<sup>6</sup> for the majority of stars it allows us to determine the qualitative affect spin has on the interaction although the magnitude of the effect may be much smaller for a real system<sup>7</sup>.

Fig. 5.12 compares the effect on the energy distribution of the prograde spinning star and non-spinning star. The results show that the energy distributions becomes flatter towards the centre and extends over a larger range of  $E_{\text{BH}}$ . In the retrograde case (Fig. 5.13) we find the opposite occurs, the peak at the centre becomes stronger and the distribution extends over a smaller range in  $E_{\text{BH}}$ . In other words the retrograde star is more tightly bound whereas the prograde star is more loosely bound. The reason for this difference can be explained by considering how the stars spin is changed during the interaction. In the prograde case the interaction causes the spin of the star to be increased, whereas in the retrograde case the interaction tries to reverse the spin of the star, causing an overall slowing of the spin.

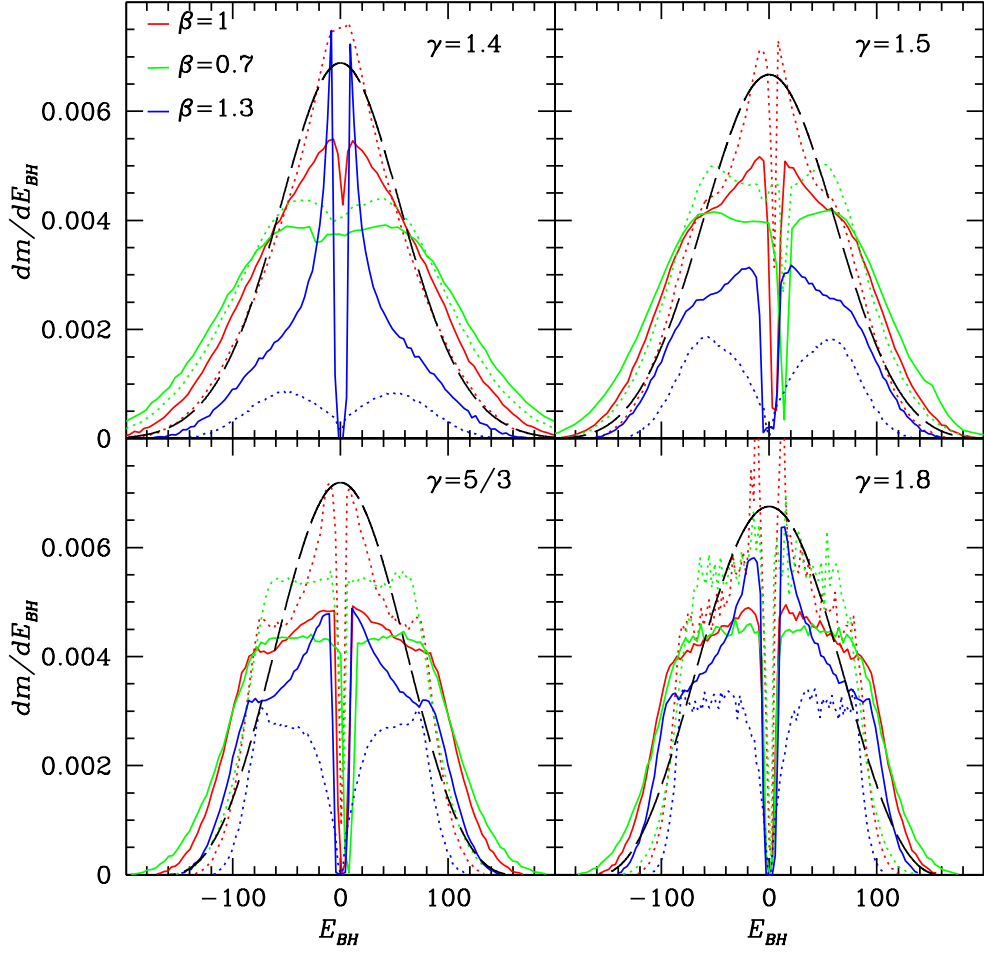
A higher spin in the forward direction means that particles closer to the black hole than the centre of mass have their absolute velocities reduced, whereas particles further from the black hole than the centre of mass have their absolute velocities increased. This means that all particles have a higher energy (in absolute terms) with respect to the black hole, causing a widening of the energy distribution. In the retrograde case the opposite is true and therefore the energy distribution is reduced.

Fig. 5.14 and Fig. 5.15 shows the effect of spin on the light curves for the prograde and retrograde case respectively. As one would expect from the results of the energy

<sup>6</sup>Although Vega is thought to be rotating at 93% of its break up speed (Peterson et al., 2006).

<sup>7</sup>We found that for solar type spins the effect on the results were indistinguishable from the non-spinning stars.

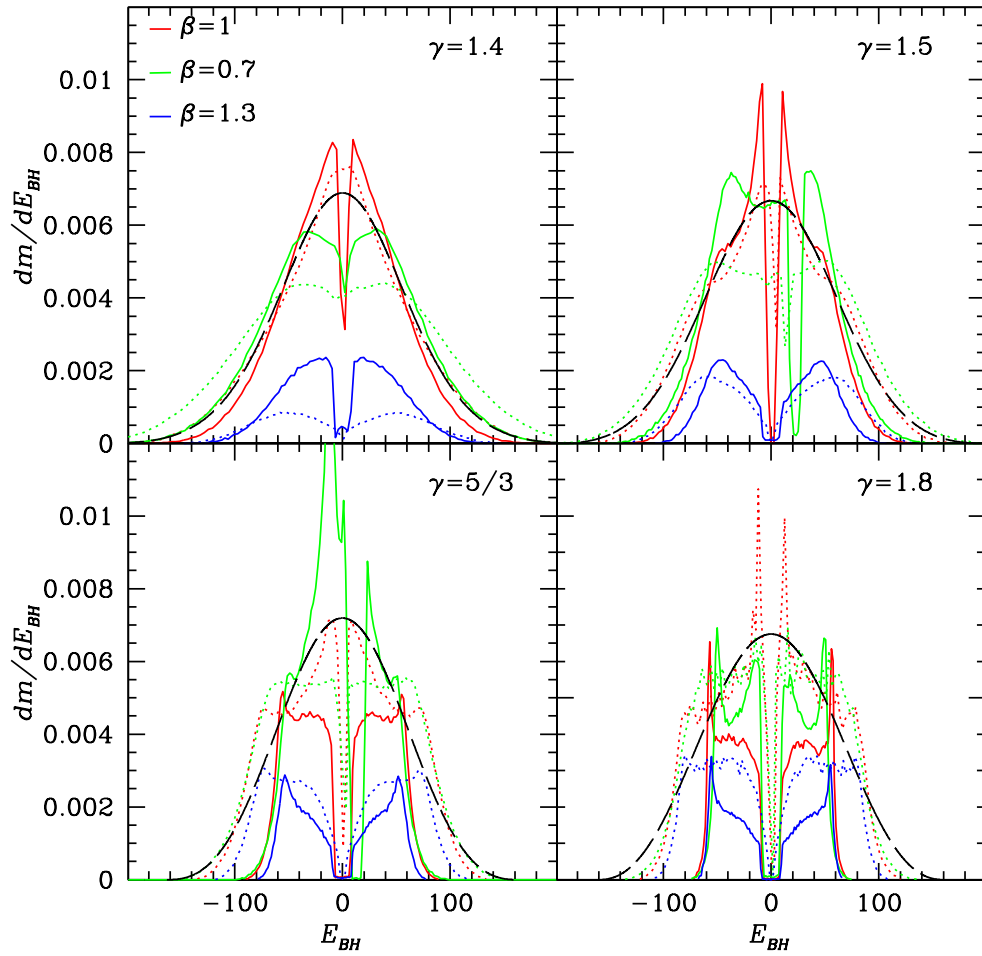




**Figure 5.12:** Effect of star spin on the energy distribution for a **prograde** spin. The spin was set to  $\omega = 0.4$ . The dotted lines show the mass distribution for the non-spinning case.

distribution, we find that for prograde spin the peak in the light curve is higher and occurs earlier than the retrograde case. This is because the prograde case has more mass at high energies leading to a faster return time for the gas. The slope of the light curves are unaffected by the star's spin for  $\beta = 0.7$ . The star is still fully disrupted leaving a uniform mass distribution, leading to a  $t^{-5/3}$  light curve.

For  $\beta = 1.3$  the slope of the light curves appear as though the interaction distance was closer for the prograde case, i.e. the effective  $\beta$  is reduced. This can be seen for the  $\gamma = 5/3$  case by looking at how the power law index varies with  $\beta$  in the non-spinning case (see Fig. 5.6). For the non-spinning case the light curve slope approaches  $n = -2$ , whilst the spinning case approaches  $n = -1.6$ . This is about



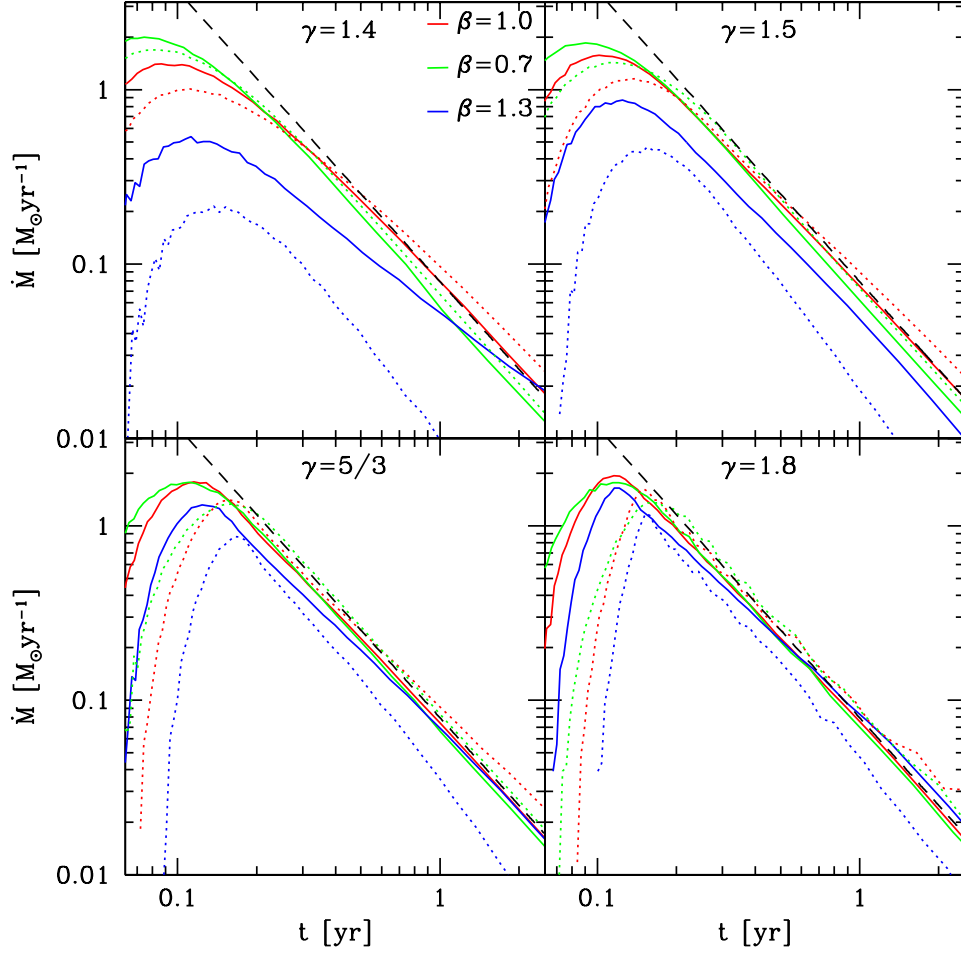
**Figure 5.13:** *Effect of star spin on the energy distribution for a **retrograde** spin. The spin was set to  $\omega = 0.4$ . The dotted lines show the mass distribution for the non-spinning case.*

the same slope as a non-spinning star with encounter parameter  $\beta = 1.2$ , therefore the spin of the star effectively strengthens the encounter.

The opposite is true for the retrograde star, the encounter has become weaker and so it appears as if the encounter was from a non-spinning star with larger  $\beta$ .

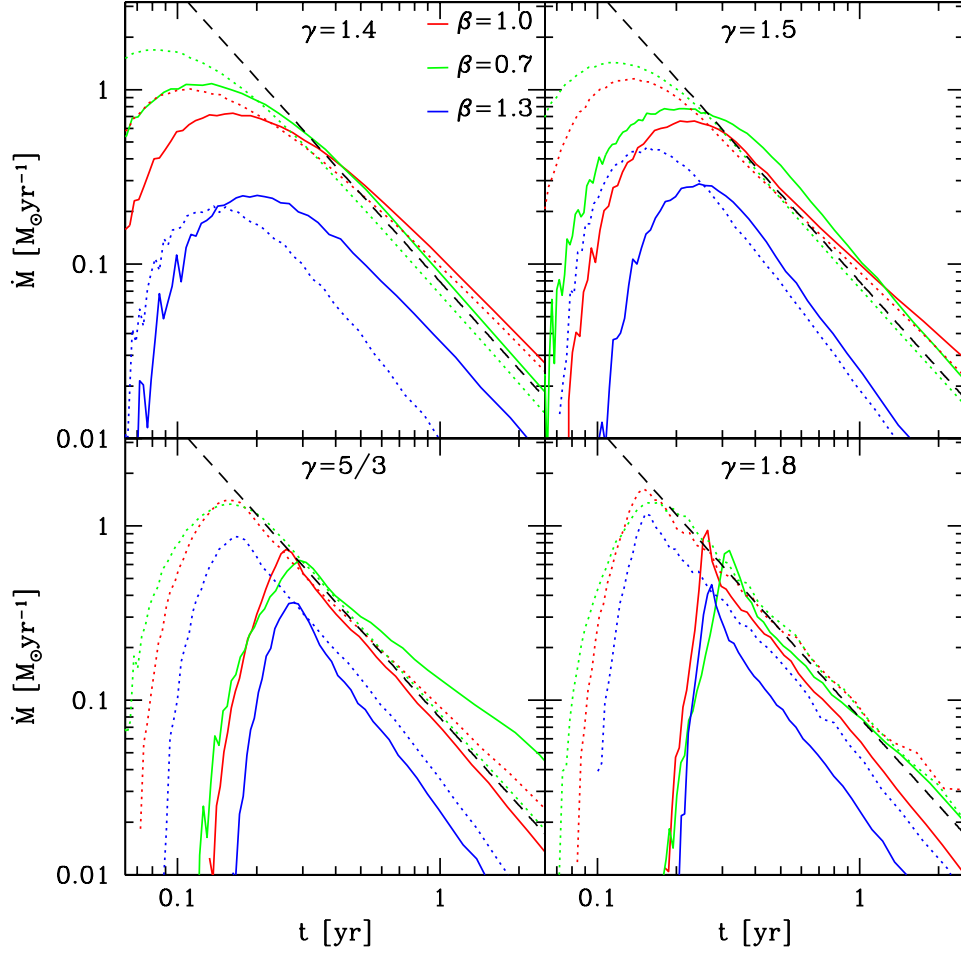
## 5.7 Mass Ratio

Thus far we have only considered interactions where the black hole to star mass ratio is  $10^6$ . This leads to a specific energy distribution which we may then convert to a light curve by assuming that  $M_{\text{unit}} = M_{\odot}$ . We could have assumed instead that



**Figure 5.14:** *Effect of star spin on the light curves for **prograde** spin. The spin was set to  $\omega = 0.4$ . The dotted lines show the light curves for the non-spinning case.*

the unit of mass was say  $M_{\text{unit}} = 10M_{\odot}$ , in which case the peak of the light curve would be shifted in both time and mass accretion rate. However, the overall shape (and in particular the decay rate) of the light curve would be unaffected. In other words, the shape of the light curve is unaffected by the unit scale we choose and is determined only by the details of the simulation such as the mass ratio, polytropic index and encounter radius. We now consider the affect of mass ratio on the tidal disruption of a  $\gamma = 5/3$  star by varying the mass of the black hole whilst keeping  $\beta = 1$ . Consequently the pericentre distance for each encounter differed according to equation (5.4). In this analysis it is important to consider the location of the tidal disruption radius compared to the Schwartzchild radius of a black hole. If the tidal disruption radius is inside the Schwartzchild radius then the flare will not be



**Figure 5.15:** *Effect of star spin on the light curves for **retrograde** spin. The spin was set to  $\omega = 0.4$ . The dotted lines show the light curves for the non-spinning case.*

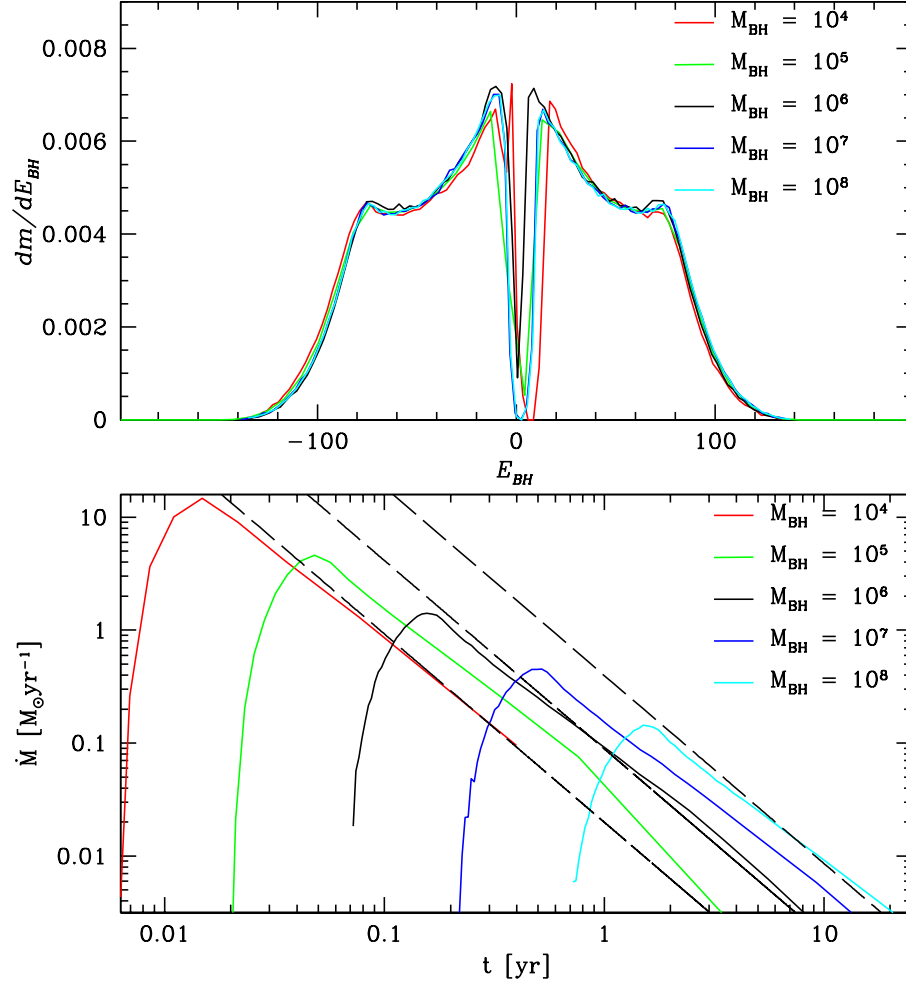
observed. By equating the Schwartzchild radius, given by

$$R_s = \frac{2GM_{\text{BH}}}{c^2}, \quad (5.12)$$

to the tidal disruption radius (see equation 5.4) we find the maximum black hole mass at which a flare can be observed is

$$M_{\text{BH}} = \left[ \frac{R_* c^2}{2GM_*^{1/3}} \right]^{3/2} \quad (5.13)$$

Table 5.1 shows the maximum mass of black hole at which typical stars are disrupted outside the Schwartzchild radius. We see that the maximum mass is  $\sim 10^8 M_\odot$  for



**Figure 5.16:** *Effect of black hole mass on the energy distribution (top) and the light curves (bottom). The energies in the top plot were scaled by a factor  $\Delta E_{10^6}/\Delta E$  for ease of comparison.*

main sequence stars and  $\sim 10^{10} \rightarrow 10^{11} M_{\odot}$  for giant stars.

Figure 5.16 shows the effect of mass ratio on the specific energy distribution. In these calculations we assumed a mass unit of a solar mass and a length scale of a solar radii. Thus the star had a mass  $M_{\odot}$  and radius of  $R_{\odot}$  and the black hole masses were given in terms of solar mass. We scaled the energy by a factor  $\Delta E_{10^6}/\Delta E = (10^6/M_{\text{BH}})^{1/3}$  in order to allow comparisons of the specific energy distribution for each black hole mass. We observe that the shape of the energy distribution is unaffected by changing the mass ratio and therefore the overall shape of the light curve is also unaffected. However, the positions of the light curves in  $\dot{M} - t$  space is altered by altering the mass of the black hole. Less massive black

Star type	Example Star	Mass ( $M_\odot$ )	Radius ( $R_\odot$ )	Max BH Mass ( $M_\odot$ )
Main Sequence	Sun	1	1	$\sim 10^8$
Red Giant	Betelgeuse	18	736	$\sim 10^{11}$
Blue Giant	Rigel	17	78	$\sim 10^{10}$

**Table 5.1:** Table comparing the maximum black hole mass at which stellar disruption can be observed for a selection of stars.

holes shift the light curve upwards and to the left resulting in a faster accretion rate peaking at an earlier time. As the luminosity depends on the mass accretion rate as

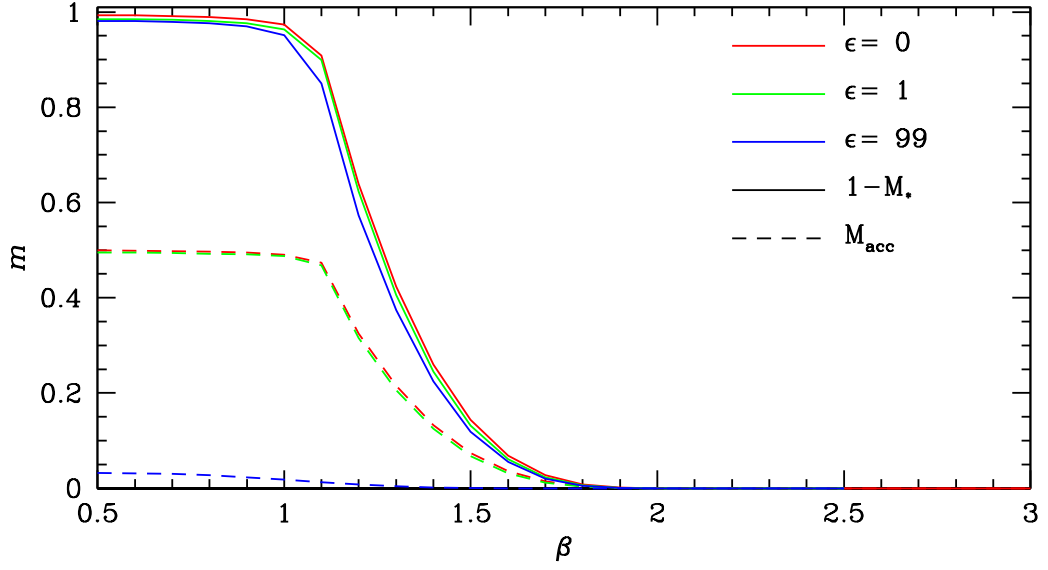
$$L_{\text{acc}} = \epsilon \frac{GM_{\text{BH}}\dot{M}}{R_s} = \frac{1}{2}\epsilon c^2 \dot{M}, \quad (5.14)$$

the luminosity of the flare is increased with decreasing black hole mass. This somewhat counter-intuitive but can be explained simply because the stars mass is unchanged and the orbit is smaller (as we preserve  $\beta = 1$ .) Therefore to a first order approximation the same amount of mass is spread out over a shorter orbit, leading to more mass returning at a given time. This creates the brighter light curve with an earlier peak. This analysis demonstrates that the mass ratio plays no part in the mechanics of the break up although it can affect the how easy it is to observe the flares.

## 5.8 Hyperbolic Orbits

Thus far we have only considered parabolic orbits. However, it is unlikely that an approaching star will be on an exactly parabolic orbit. Therefore we explore hyperbolic orbits. We may investigate hyperbolic orbits which are close to parabolic (i.e. small energies) and orbits which are far from parabolic. We expect that increasing the energy of the orbits will push the centre point of the energy distribution to higher energies, therefore reducing the accretion rate. This is simply because the centre of mass of the star will be on an orbit with  $E > 0$ . However, changes in orbit may also affect the dynamics of the break up of the star, which may in turn affect the accretion rate onto the black hole. We use a  $\gamma = 5/3$  star and vary the pericentre distance of the orbit.

The specific energy of an orbit  $\epsilon$  is related to the eccentricity and specific relative



**Figure 5.17:** Mass unbound from the star (*solid line*) and accreted mass (*dashed line*) for two hyperbolic orbits compared to a parabolic orbit.

angular momentum,  $h = rv$  via

$$\epsilon = (e^2 - 1) \frac{(GM_{\text{BH}})^2}{2h^2}. \quad (5.15)$$

Therefore in parabolic orbits where  $e = 1$ , the specific energy is zero. Elliptical orbits with  $e < 1$  have a negative specific energy whilst hyperbolic orbits ( $e > 1$ ) have a positive specific energy. The velocity of an orbit is related to the specific orbital energy by

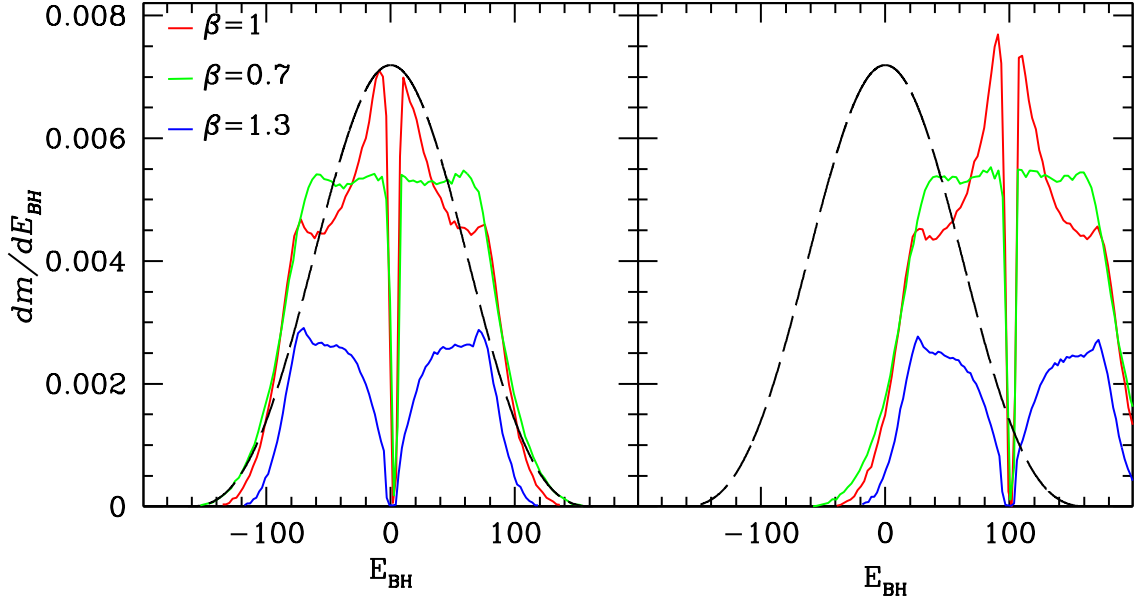
$$\epsilon = \frac{v^2}{2} - \frac{GM_{\text{BH}}}{r}. \quad (5.16)$$

Again using conservation of energy and angular momentum we determine the initial velocity of the hyperbolic orbit at  $(r_0, 0)$  as

$$v_x = \sqrt{2\epsilon \left(1 - \left(\frac{R_p}{r_0}\right)^2\right) + \frac{2M_{\text{BH}}}{r_0} \left(1 - \frac{R_p}{r_0}\right)} \quad (5.17a)$$

$$v_y = \frac{R_p}{r_0} \sqrt{2\epsilon + \frac{2M_{\text{BH}}}{R_p}}. \quad (5.17b)$$

The initial separation was set to ten times the pericentre distance. We consider two



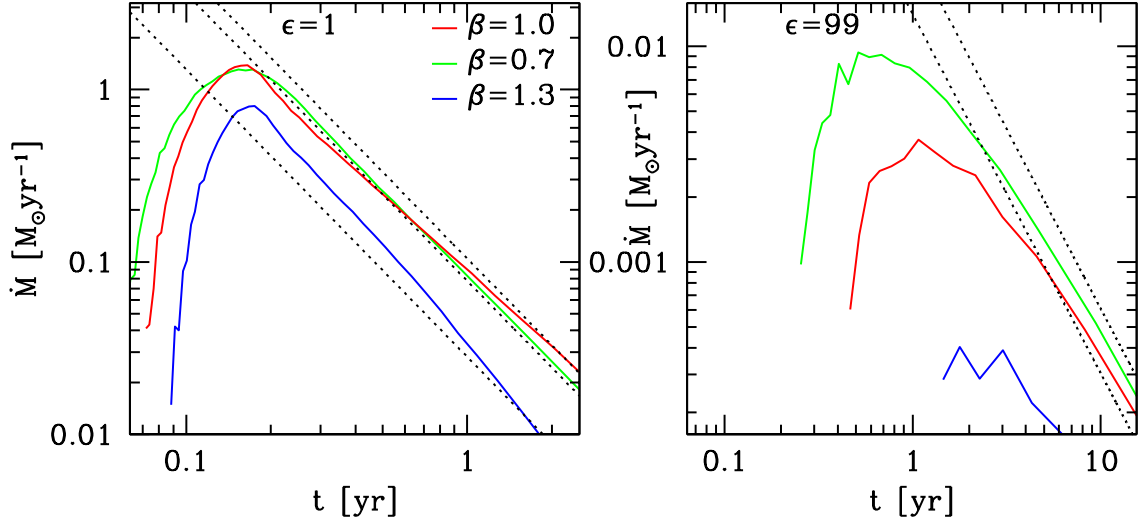
**Figure 5.18:** *Distribution of energies for two hyperbolic orbits using various pericentre distances. **Left** shows with  $\epsilon = 1$ , **right** with  $\epsilon = 99$ .*

hyperbolic orbits with  $\epsilon = 1$  and  $\epsilon = 99$  (in code units where  $M_\odot = R_\odot = G = 1$ ). The  $\epsilon = 99$  case is particularly interesting as it is just below the theoretical width of the specific energy distribution as predicted by equation (5.5). Therefore we expect only the tail of the specific energy distribution to be accreted.

Fig. 5.17 shows how the mass of the star changes with pericentre distance for a range of orbital energies. We see that for a given pericentre distance the type of orbit the star is on makes very little difference to the mass unbound from the star. The hyperbolic orbits unbind slightly less mass than the parabolic orbits but the transition from fully unbound to fully bound occurs over the same range of  $\beta$ . We see that the main difference is in the mass accreted by the black hole. In the parabolic orbit half of the mass unbound is accreted by the black hole. In the  $\epsilon = 99$  orbit this is not the case. The reason for this is that in a  $\epsilon = 99$  orbit most of the mass of the star is on an unbound orbit from the black hole as the spread in energy is only  $\Delta E \sim 100$  (see eq. 5.5). Therefore most of the mass unbound from the star escapes on hyperbolic orbits.

Figure 5.18 compares the energy distributions of the hyperbolic orbits for different pericentre encounters. We find that for the hyperbolic encounters the shape of the energy distribution is unchanged by the energy of the orbit. However, the centre





**Figure 5.19:** *Light curves for two hyperbolic orbits using various pericentre distances. Left shows with  $\epsilon = 1$ , right with  $\epsilon = 99$ .*

point of the distribution is shifted as one would expect. This affects the amount of material accreted by the black hole as only gas with  $E_{\text{BH}} < 0$  may be accreted, the rest of the gas is ejected on hyperbolic orbits. The return time for the mass is also affected as the return time directly depends on  $E_{\text{BH}}$  via equation (5.6).

The consequences on the light curve may be seen in Fig. 5.19. We observe that for  $\epsilon = 1$  the light curve agrees well with  $t^{-5/3}$  for the strongest encounters (i.e.  $\beta = 0.7$ ). This is because the specific energy distribution is flat in the range  $E_{\text{BH}} < 0$ , which was assumed in the derivation of the light curve (see equation 5.7). There is also good agreement between  $\epsilon = 1$  and the parabolic light curves for  $\beta = 1$  and  $\beta = 1.3$ . This is because the specific energy distribution is only slightly shifted compared to the parabolic case. The only noticeable effect is that the light curve rises and peaks slightly later in the  $\epsilon = 1$  case compared to the parabolic case. This is again due to the shift in the light curve towards higher (less negative) energies.

The light curve for the  $\epsilon = 99$  case is significantly different from the parabolic case. The peak in the light curve not only occurs much later, the peak is around two orders of magnitude lower than the parabolic case. The peak in the  $\epsilon = 99$  has the same accretion rate as the tail at  $t \sim 1$  in the parabolic case. This means that the chance of observing light curves produced by highly hyperbolic orbits is much lower than orbits that are parabolic or just hyperbolic.

We can determine how the specific orbital energy is related to the eccentricity

by rearranging equations (5.15) and (5.16)

$$e = \frac{R_\odot}{GM_\odot} 2\epsilon \frac{M_\odot}{M_{\text{BH}}} \frac{R_*}{R_\odot} \beta q^{1/3} + 1 \quad (5.18)$$

As we only consider black holes with  $M_{\text{BH}} = 10^6 M_\odot$  and stars with  $M_* = M_\odot$  and  $R_* = R_\odot$  in units where  $M_\odot = R_\odot = G = 1$ , this equation can be written as

$$e = 0.0002 \epsilon \beta + 1 \quad (5.19)$$

Therefore for the  $\epsilon = 1$  case the eccentricity varies between  $e = 1.0001$  ( $\beta = 0.5$ ) and  $e = 1.0006$  ( $\beta = 3$ ). For  $\epsilon = 99$  the eccentricity varies between  $e = 1.0099$  ( $\beta = 0.5$ ) and  $e = 1.0594$  ( $\beta = 3$ ).

We can determine the typical eccentricity of a bulge star by considering their typical velocities. Gültekin et al. (2009) give the mass-dispersion relation for stars as

$$\sigma^2 = 5.9 \times 10^6 \left( \frac{M_{\text{BH}}}{M_\odot} \right)^{25/53} \text{m}^2 \text{s}^{-2} \quad (5.20)$$

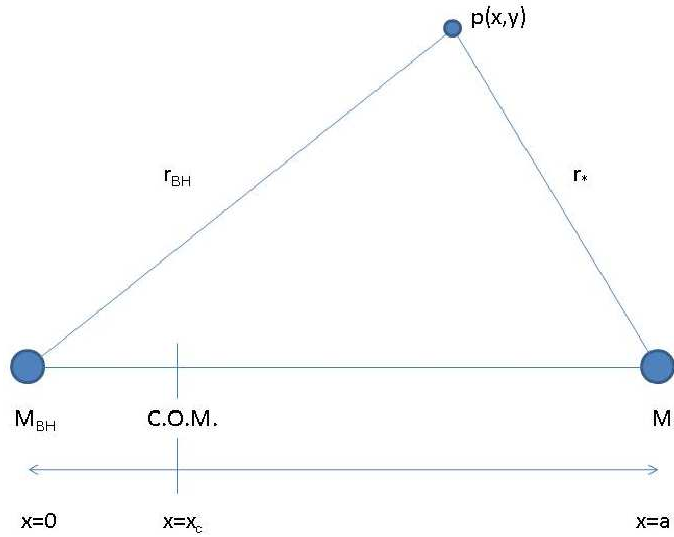
If we assume that the star falls in from infinity, the specific orbital energy is related the velocity dispersion via  $\epsilon = 3/2 \sigma^2$ . By substituting this into equation (5.18) we obtain

$$e = 9.2 \times 10^{-5} \left( \frac{M_{\text{BH}}}{M_\odot} \right)^{-28/53} \frac{R_*}{R_\odot} \beta q^{1/3} + 1 \quad (5.21)$$

Therefore, a solar type star orbiting a  $M_{\text{BH}} = 10^6 M_\odot$  black hole would typically have an eccentricity of  $e = 1.000006256$  (for  $\beta = 1$ ). The velocity dispersion for a  $M_{\text{BH}} = 10^6 M_\odot$  black hole is  $\sigma = 63 \text{km s}^{-1}$ , which corresponds to a specific orbital energy of  $\epsilon = 0.03$  in code units<sup>8</sup>. For the star to be on a hyperbolic orbit around a  $M_{\text{BH}} = 10^6 M_\odot$  black hole with  $\epsilon = 1$  or  $\epsilon = 99$  (in code units), the star must be travelling with a velocity in excess of  $\sim 350 \text{ km s}^{-1}$  or  $\sim 3500 \text{ km s}^{-1}$  respectively. Therefore it is unlikely that many stars will travel with this velocity based on the velocity dispersion. However, some stars may be able travel with velocities in excess of  $\sim 1000 \text{ km s}^{-1}$  due to binary encounters ejecting stars (Hills, 1988; Brown et al., 2005) or supernova kicks.

---

<sup>8</sup>Assuming units of  $M_\odot = R_\odot = G = 1$  gives the scaling  $1 \text{m}^2 \text{s}^{-2} = 5.2 \times 10^{-12} \text{R}_\odot^2 \text{T}^{-2}$  where the time unit is  $1 \text{T} = 1588 \text{ s}$



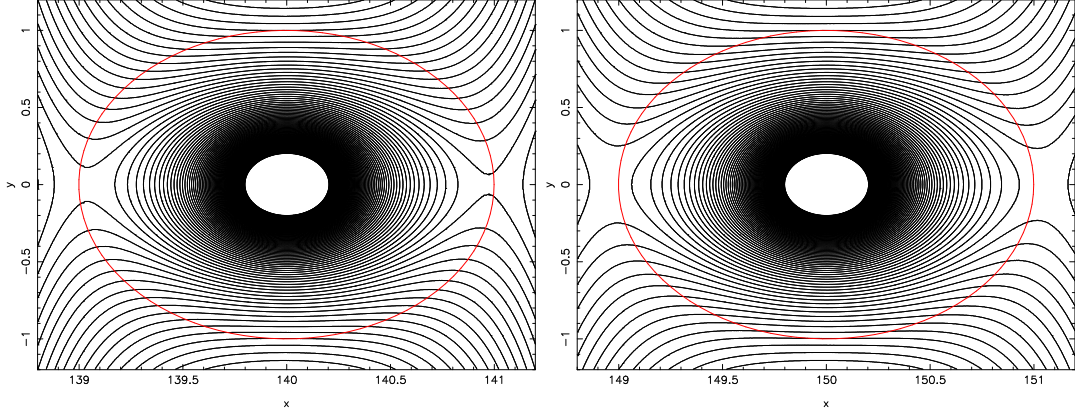
**Figure 5.20:** *Diagram showing black hole and star orbiting the centre of mass of the system.*

## 5.9 Elliptical Orbits

Thus far we have only considered open orbits, however, elliptical orbits are also of interest. This is because in open orbits the star will only pass through the pericentre once, therefore if the star is not significantly disrupted at pericentre it will remain bound. However, in an elliptical orbit the star will make many pericentre passages therefore the small perturbations induced by a single pericentre passage may not have settled down by the time the next pericentre passage occurs. This may lead to the star eventually becoming unbound after several pericentre passages. Furthermore, as the star is subjected to the tidal forces for longer, the star may become more stretched than in the parabolic case. This would allow tidal disruption events to occur farther from the black hole.

The tidal disruption of stars on elliptical orbits are of interest in relation to the “S-Stars”, which are stars within the central arc-second of the galactic centre. These stars are thought to be main sequence stars which will eventually become giant stars. When these stars become giants they will expand, with the outer layers possibly entering the tidal disruption radius. Therefore looking at how stars on elliptical orbits are affected tidal disruption events can help determine the possible fate of the “S-stars”.

We can determine the theoretical tidal radius for the case of a circular orbit by considering the Roche potential in the rotating frame (see Fig. 5.20); this is given



**Figure 5.21:** *Roche potential around the star orbiting a  $M_{\text{BH}} = 10^6 M_{\odot}$  black hole with  $\beta = 1.4$  (left) and  $\beta = 1.5$  (right). The red line shows the radius of the star.*

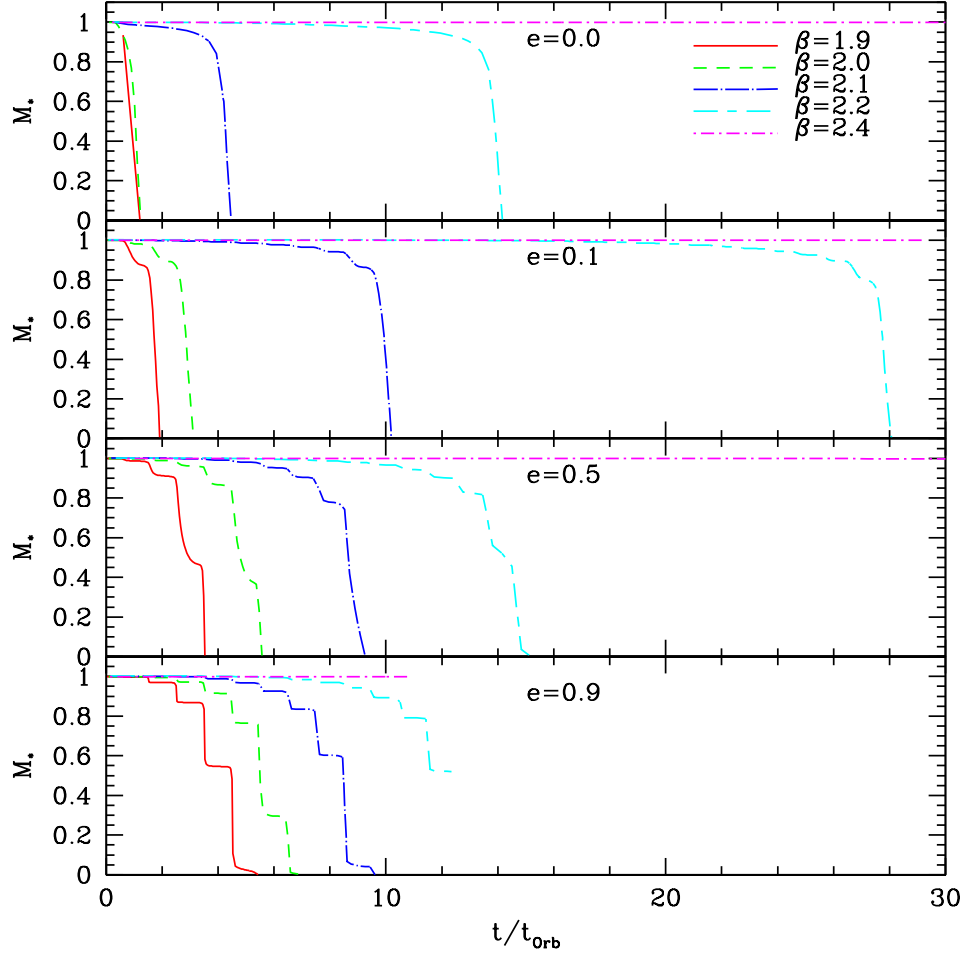
by

$$\phi_p = -\frac{GM_{\text{BH}}}{r_{\text{BH}}} - \frac{GM_*}{r_*} - \frac{\omega^2}{2} [(x - x_c)^2 + y^2] \quad (5.22)$$

where  $r_{\text{BH}}^2 = x^2 + y^2$ ,  $r_* = (x - a)^2 + y^2$  and  $\omega^2 = G(M_{\text{BH}} + M_*)/a^3$ . The potential is computed for a solar type star orbiting a  $M_{\text{BH}} = 10^6 M_{\odot}$  black hole at  $\beta = 1.4$  and  $\beta = 1.5$  as shown in Fig. 5.21. We see that in the  $\beta = 1.4$  orbit the outer edge of the star is outside of the Roche lobe defined by the L1 point. With  $\beta = 1.4$  the star is contained within the L1 point, indicating that it is stable.

Here we will consider a range of elliptical orbits in order to determine the specific trends that altering the eccentricity can have on the disruption encounter. It is interesting to note that for parabolic encounters with  $\beta \geq 1.9$  the star remained completely undisrupted. However, in parabolic encounters the star only experienced a single pericentre encounter. With elliptical orbits the star may reach the pericentre distance a number of times, therefore the star is repeatedly perturbed, which may have an affect on the overall stability.

As we are interested in the long term evolution of the star we remove any particles that become unbound from the star, this is determined by the analysis in section 5.3 (in particular equation 5.10). If a particle has  $E_{\text{gas}} > 0$  then it is deleted from the simulation. We found this necessary to prevent particles clustering around the black hole, requiring small time steps and to prevent particles gaining large accelerations. The gas was not accreted to the black hole and therefore the black hole did not grow. However, this will not affect the simulation as the black hole is  $10^6$  times



**Figure 5.22:** The star mass as a function of time for various elliptical orbits. The pericentre of the orbits occur at half integers in time.

larger than the total mass of gas in the simulation.

We model the star using a  $\gamma = 5/3$  polytrope initially at apo-centre, located at  $(r_a M_{\text{BH}}/(M_{\text{BH}} + M_*), 0)$  where  $r_a = a(1 + e)$ . The semi major axis,  $a$ , is related to the penetration factor  $\beta$  via

$$\beta = \frac{R_p}{R_T} = (1 - e) \frac{a}{R_T}. \quad (5.23)$$

The initial velocity is set to  $(0, v M_{\text{BH}}/(M_{\text{BH}} + M_*))$  where

$$v = \sqrt{\frac{M_{\text{BH}} + M_*}{a} \frac{1 - e}{1 + e}} \quad (5.24)$$

is the relative orbital velocity. The black hole is initially located at  $(-r_a M_*/(M_{\text{BH}} + M_*), 0)$  with initial velocity  $(0, -v M_*/(M_{\text{BH}} + M_*))$

Fig. 5.22 shows how the mass of the star decreases with time for a range of eccentricities and pericentre distance. We firstly note that in the circular orbit the tidal limit occurs at  $\beta \approx 2.4$ , which is somewhat larger than  $\beta = 1.5$  as suggested by Fig. 5.21. However, this does not take into account that the star is tidally stretched by the black hole. This stretching allows the star to be broken up farther from the black hole than one would expect from simple analysis. This is because as the star stretches the tidal disruption radius is increased due to the increase in  $R_*$  (see equation 5.4). Therefore not only has the star increased in size, the Roche lobe has also decreased in size, allowing the break up of the star. Indeed Roche derived the tidal limit as  $\beta \approx 2.44$  for a liquid satellite. The effect is not as severe in the parabolic case as the star only makes one pericentre passage, and therefore the star is only stretched for a short period of time, after which it may settle back into equilibrium. When the star is in a circular orbit, it is continuously being stretched and can therefore be disrupted farther out.

We find that with high eccentricities the star is slightly disrupted with each pericentre passage with part of the star becoming unbound. This is especially noticeable with the  $e = 0.9$  case, where the mass is reduced in discrete steps. With smaller eccentricities the mass is reduced much more sharply. These differences indicate that with smaller eccentricities the pericentre encounter is much more severe.

We also find that the number of orbits the star can make before it disrupts depends on *both* the pericentre distance and the eccentricity. For orbits where the pericentre distance is close to the disruption distance (e.g.  $\beta = 1.9$ ) increasing the eccentricity increases the number of orbits it takes for the polytope to disrupt and in the limit where  $e \rightarrow 1$  (i.e. parabolic orbit) the star would not disrupt. This is because increasing the eccentricity reduces the amount of time the star spends close to the black hole. Therefore the amount of time over which the star is strongly perturbed is reduced in more eccentric orbits. For orbits that are further from the disruption radius (e.g.  $\beta = 2.2$ ), increasing the eccentricity decreases the number of orbits it takes for the star to become disrupted.

### 5.9.1 The Fate of S2

Based on the experiments of the previous section we may determine the fate of the S2 star. Eisenhauer et al. (2003) find the orbital period of S2 as 15.56 years with

eccentricity 0.881. Combining this with the mass of Sgr A\* determined from a multi-star fit from Gillessen et al. (2009) as  $M_{\text{BH}} = 4.3 \times 10^6 M_{\odot}$  this gives a semi major axis of  $a = 4.9 \times 10^{-3} \text{pc} = 2.2 \times 10^5 R_{\odot}$  and pericentre distance of  $R_p = 2.6 \times 10^4 R_{\odot}$ . We may combine equations (5.8) and (5.4) to obtain an expression for the disruption parameter

$$\beta = \left( \frac{M_*}{R_*^3} \right)^{1/3} \left( \frac{R_p^3}{M_{\text{BH}}} \right)^{1/3}. \quad (5.25)$$

Using the parameters found by Eisenhauer et al. (2003) and Gillessen et al. (2009) we obtain

$$\beta = 1.6 \times 10^2 \left( \frac{M_*}{R_*^3} \right)^{1/3}, \quad (5.26)$$

where  $M_*$  and  $R_*$  are in solar mass and radii.

We may now look at the fate of the S2 star by assuming typical mass and radius estimates. Supposing that S2 is a solar type star with  $M_* = M_{\odot}$  and  $R_* = R_{\odot}$  we obtain the impact parameter  $\beta = 1.6 \times 10^2$ , and therefore the star is clearly stable. A solar type star will eventually pass into a red giant phase, where its radius may increase by a factor of around  $\sim 200 \rightarrow 1000$  (Joss et al., 1987). Therefore again assuming a solar type star of  $M_* = M_{\odot}$  in a red giant phase with either  $R_* = 200 R_{\odot}$  or  $R_* = 1000 R_{\odot}$  gives a penetration factor of  $\beta = 0.8$  and  $\beta = 0.16$  respectively. Therefore S2 will be strongly affected by tides in both cases.

This analysis suggests that the star will be completely disrupted with in one full orbit (as suggested by results of section 5.9). Furthermore, the results of section 5.4 suggest that for strong encounters (i.e.  $\beta < 1$ ) the star is fully disrupted leading to a uniform mass distribution; hence producing a  $t^{-5/3}$  light curve. However, the transition from main sequence star to red giant will occur with a gradual expansion of the star. When this happens, the penetration factor  $\beta$  will gradually decrease as  $R_*$  increases. Therefore there will be some mass skimmed off the surface of the star at each pericentre passage.

---

# 6

## Conclusions

In chapter 1 we highlighted the need for numerical methods in astrophysics, in particular a method for the simulation of a self-gravitating fluid. Here we give an overview of the results of this study and indicate how this work may be extended in the future.

### 6.1 Summary

In chapter 2 we gave a review Smoothed Particle Hydrodynamics (SPH), a numerical method that is widely used in astrophysics (and other fields) to model fluid flow. SPH is a Lagrangian method which uses interpolation points, often referred to as particles, to represent the fluid. At the heart of the SPH method lies the density estimator which computes the density based on the relative positions of the surrounding particles. The particles are free to move throughout the fluid ensuring that regions of higher density automatically have more particles. This density estimator, combined with a particle Lagrangian which represents the fluid allows one to solve the Euler equations in Lagrangian form. This approach, suggested by Springel & Hernquist (2002) produces a unique SPH scheme which conserves energy, entropy, momentum and angular momentum by construction. In order to be useful for astrophysics the SPH method can be coupled with an N-body solver allowing one to model self-gravitating fluids such as stars.



In chapter 3 we reviewed the current method for handling shocks in SPH using artificial viscosity. Artificial viscosity is a sub-resolution method used to mimic dissipation and entropy generation processes that occur at shock fronts. We found that the best available method proposed thus far is that of Morris & Monaghan (1997) in which each particle is afforded its own artificial viscosity parameter,  $\alpha$ , which is then evolved according to a differential equation. The main problem with this method is the non-zero minimum viscosity, meaning that dissipation occurs away from shocks. We also find that the source term is unable to distinguish pre and post shock regions, this combined with the slow integration of a differential equation often means that  $\alpha$  peaks after the shock. We improve on this method by allowing  $\alpha = 0$  away from shocks, therefore modelling the fluid as inviscid away from shocks. We also use a higher order derivative,  $D_t(\nabla \cdot \mathbf{v})$ , which is able to distinguish between pre and post shock regions. This combined with allowing  $\alpha$  to jump to the required value leads to the viscosity peaking before the shock arrives. We find that the new method is able to model sound waves over many periods with hardly any dissipation, something which was not possible until now. We also find that the new method is as good as any other method in resolving strong shocks and preventing particle disorder.

In chapter 4 we focused on the stability of SPH. We looked at initial conditions and found that randomness in particle initial conditions should be avoided. We then examined the stability analysis performed by Morris (1996), Cha (2002), Price (2004) and Read et al. (2010) and performed a numerical comparison to their analysis. We find that the numerical results agree with the analysis in that we find bands of instability in  $h - k$  space. We find, as predicted by Read et al. (2010), the HOCT kernel is able to prevent particle clumping, although we find that the number of neighbours required to do this is less than suggested by their analysis. However, we find that the bias in the density is quite high for typical neighbour numbers of  $N_h = 40 \rightarrow 100$ , instead one has to go to around  $N_h \sim 400$  to achieve the same accuracy of the cubic spline kernel with  $N_h = 40$ . This of course will greatly increase the running time of any simulation. Furthermore, in order to achieve the same resolution the total number of particles must also be increased, again increasing the running cost.

Chapter 5 extends the stellar disruption work of Evans & Kochanek (1989) and Lodato et al. (2009). Stellar disruption can occur when a star passing close to a super massive black hole experiences a large tidal force. The disrupted material is

then accreted by the black hole giving rise to a light curve. This mechanism was suggested by Rees (1988) as a way to explain observed X-ray flares from inactive galaxies. We looked at the affect of altering the encounter parameter on the disruption of the star and on the light curve produced by the accreted material. We find that for encounters on a parabolic orbit that pass within the theoretical tidal radius, the star is broken up with half of the material being accreted by the black hole. Orbits further from the black hole may be partially disrupted within a small range of pericentre distances, leading to some mass being accreted by the black hole. We conclude that the pericentre distance plays an important role on the decay of the light curve with close encounters producing a  $t^{-5/3}$  decay (as predicted by Rees, 1988), and more distant encounters producing light curves with different decay rates.

For typical stellar spins there is almost no effect on the light curve but for rapidly rotating stars, such as Wolf-Rayet stars the disruption can be significantly effected, depending on the orientation of the spin to the orbit. We looked at the affect of orbit on the break up of the star. We found that in elliptical orbits the star can be broken up further from the black hole than in the parabolic case. The reason for this is in elliptical orbits the star is being continuously stretched, which moves the critical tidal disruption radius farther from the black hole. Finally we looked at the implications of this study for the S2 star which is a main sequence star orbiting Sgr A\* with a period of around fifteen years. We found that once the star becomes a giant it will suffer a tidal disruption event with the material being accreted by the black hole. However, it is likely that material will be striped off the surface of the star whilst it is still expanding.

## 6.2 Future Work

Over the course of this study we found several areas in which the SPH algorithm can be improved upon. One such area was artificial viscosity, which was addressed in chapter 3. However, we also found shear flows to be particularly difficult to model in SPH, especially in situations where we require the suppression of artificial viscosity. This is because the particles quickly become anisotropic which leads to noise in the SPH quantities. The method suggested in chapter 3 employed a limiter, similar to that proposed by Balsara (1995) to reduce the viscosity when there is a large shear flow and no shock. Whilst this worked well for the test cases considered, it is likely that improvements can be made in this area. Furthermore, SPH has

problems dealing with shearing flows in general due to noisy estimates cause by particle anisotropies. This can cause severe problem in modelling discs, as pointed out by Imaeda & Inutsuka (2002). Although the authors suggested a fixed for this, their method is computationally expensive.

The work done on the SPH viscosity scheme presented in Chapter 3 should prove valuable in many areas of astrophysics. For example, the modelling of turbulent shells falling towards a black hole which are then accreted (Hobbs et al., 2010). Turbulence is difficult to model as it involves weak shocks of varying strength. Therefore a fixed viscosity is far from ideal as it will certainly over damp weak shocks and under damp the stronger shocks. The new scheme can also be applied to situations where polytrope oscillations are important, for example one could extend the work of chapter 5 to look at the disruption of pulsating stars.

Another area of SPH which needs improvement is the modelling of mixing, for example in KH simulations. There are two different approaches to this problem, both of which seem to have helped resolve this problem. The first was suggested by Price (2008) in which a thermal conductivity is applied to smooth away sharp discontinuities in thermal energy. This helps remove the pressure anomalies that occur at contact discontinuities which tend to separate layers of particles. In chapter 3 we suggested a scheme in which conductivity could be applied such that it is only turned on where required. This method was relatively successful in modelling KH instabilities but clearly more work is required in this area.

An alternative method to resolve this mixing problem was suggested by Read et al. (2010) in which a few changes to the SPH scheme allowed significant improvement in the modelling of mixing. These changes most notably included a density estimate suggested by Ritchie & Thomas (2001) and a High Order Cored Triangle kernel. However, the draw back of this method is that a high number of neighbours are required ( $N_h = 442$ ) which can dramatically increase the running cost of the simulation. Further advances to this scheme (Read et al, in prep) abandon the conservative formulation of SPH as described in chapter 2 by using a momentum equation which does not conserve pair-wise momentum by construction.

Whilst both of these schemes are showing significant improvement in the ability of SPH to handle mixing, there are still some draw backs. In particular modelling high density contrast KH instabilities are still difficult. It also maybe possible that a combination of an improved SPH scheme (such as Read et al., 2010) and thermal conductivity is required to fully resolve the mixing problem.

The work done on the stellar disruption could be extended by explicitly modelling the “S stars”. One could investigate the effect of a star moving from a main sequence to a red giant and look at how the accretion onto the black hole changes during this time. In order to do this one would need a stellar evolution model which correctly expands the polytrope with time to simulate the evolution from main sequence to giant star. This would be particularly interesting because as the star expands, outer layers of the star may be stripped by tidal force. Furthermore, it would be interesting to trace the dynamics of the gas, which in the elliptical orbit case may be accreted very quickly. For example the gas may be on elliptical orbits which pass inside the Schwartzchild radius of the black hole, giving rise to several bright, short lived flares. In order to do this one would need a realistic model for black hole accretion, as well as a SPH scheme which can effectively handle the shear flows found in accretion discs. Such a scheme would have to keep artificial dissipation low (as is the case for the viscosity scheme outlined in chapter 3) as well as having a physical viscosity in order to transport material throughout the disc.

One could also use more realistic models for the stars instead of polytropes, for example one could create a composite polytrope to model the degenerate cores and convective envelopes of giant stars.

---

# Appendices



## Comparison of Viscosity Magnitudes

### A.1 Homologous Flow

Suppose we have a homologous flow given by

$$\mathbf{v} = -b\mathbf{x}. \tag{A.1}$$

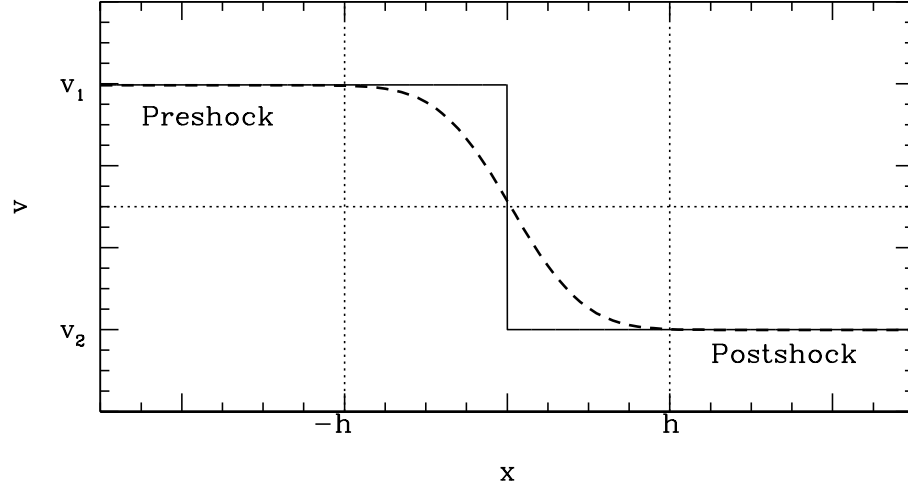
The velocity divergence is given by  $\nabla \cdot \mathbf{v} = -3b$ , thus the Morris & Monaghan (1997) source term is  $S_v = 3b$ . Taking the time derivative of this gives  $D_t(\nabla \cdot \mathbf{v}) = 0$ , therefore the new viscosity scheme keeps viscosity turned off in this situation. Furthermore the acceleration of the flow is  $\mathbf{a} = -b\mathbf{v} = b^2\mathbf{x}$ . Therefore the acceleration divergence is  $\nabla \cdot \mathbf{a} = 3b^2$ .

### A.2 Sound wave

Suppose we have a sound wave given by

$$\mathbf{v} = v_a \sin(k|\mathbf{x} - ct|) \tag{A.2}$$

where the velocity amplitude  $v_s \ll c$  and wave number  $k < h^{-1}$ . This is an example of a well resolved, weakly convergent flow. The velocity divergence is given by  $\nabla \cdot \mathbf{v} = v_a k \cos(k|\mathbf{x} - ct|)$ . Therefore the maximum of the Morris & Monaghan (1997)



**Figure A.1:** *The idealised smoothing of a strong shock over a smoothing length.*

source term is  $S = kv_a$ . Taking the time derivative of this gives  $D_t(\nabla \cdot \mathbf{v}) = \mathbf{v}k^2c$ , giving a source term as  $S = \mathbf{v}k^2c \times h/v_{\text{sig}} \approx \mathbf{v}k^2h$ . This varies from the M&M source by a factor  $kh \rightarrow 0$  for well resolved waves. Also the acceleration of this sound wave can be found by differentiating equation A.2 as  $\mathbf{a} = -v_akc \cos(k|\mathbf{x} - ct|)$ . Thus the acceleration divergence is  $\nabla \cdot \mathbf{a} = \mathbf{v}k^2c$  giving the source term (3.29) as  $S_a = fhk^2\mathbf{v}$ . This again varies from the M&M source by a factor  $kh \rightarrow 0$  for well resolved waves.

### A.3 Strong Shock

Suppose we have a strong shock such that  $\delta v \gg c$ . In a strong shock the following shock relations are valid

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} = \frac{\gamma - 1}{\gamma + 1} \quad (\text{A.3})$$

$$P_2 = \rho_1 v_1^2 \frac{2}{\gamma + 1} \quad (\text{A.4})$$

where the subscripts <sub>1</sub> and <sub>2</sub> refer to the pre-shock and post-shock region respectively. The pressure in the pre-shock region  $P_1$  can be shown to be negligible. If we assume that the shock is smoothed over a kernel width we may approximate the velocity as

(see figure A.1)

$$\mathbf{v} = \begin{cases} v_1 & x < -h \\ v_1 - (v_1 - v_2)W\left(\frac{|\mathbf{x}-h|}{h}\right) & -h < x < h \\ v_2 & x > h. \end{cases} \quad (\text{A.5})$$

The velocity divergence across the shock is

$$\nabla \cdot \mathbf{v} = -(v_1 - v_2) \cdot \nabla W\left(\frac{|\mathbf{x}-h|}{h}\right). \quad (\text{A.6})$$

Assuming a Gaussian kernel this gives

$$\nabla \cdot \mathbf{v} = -\delta v W_0 \frac{|\mathbf{x}-h|}{h^2}. \quad (\text{A.7})$$

We may determine the strength at the centre of the shock (i.e.  $x = 0$ )

$$\nabla \cdot \mathbf{v} \approx -W_0 \frac{\delta v}{h} \quad (\text{A.8})$$

where  $W_0$  is a constant which depends on the kernel. Thus the M&M source term is  $S \sim \delta v/h$ . We can determine the rate of change of this by approximating the time over which the shock acts

$$\frac{d}{dt}(\nabla \cdot \mathbf{v}) \approx \frac{\nabla \cdot \mathbf{v}}{\delta t} \approx \frac{v_s \nabla \cdot \mathbf{v}}{\ell h} \quad (\text{A.9})$$

where  $\ell$  is the number of smoothing lengths the shock is smoothed over. Therefore

$$\frac{d}{dt}(\nabla \cdot \mathbf{v}) \approx \frac{W_0 (\delta v)^2}{\ell h^2}. \quad (\text{A.10})$$

The acceleration in a strong shock is given by

$$\mathbf{a} = -\frac{1}{\rho} \nabla P. \quad (\text{A.11})$$

Assuming the pressure is smoothed over a kernel width, the pressure across the shock can be approximated by

$$P = P_2 W\left(\frac{|\mathbf{x}+h|}{h}\right) = \rho_1 v_1^2 \frac{2}{\gamma+1} W\left(\frac{|\mathbf{x}+h|}{h}\right) \quad (\text{A.12})$$



Using the shock relations and that  $\delta v = v_1 - v_2$  we find

$$v_1^2 = \left( \frac{\gamma + 1}{2} \right)^2 \delta v^2 \quad (\text{A.13})$$

Therefore

$$P = \rho_1 \delta v^2 \frac{\gamma + 1}{2} W \left( \frac{|\mathbf{x} + h|}{h} \right) \quad (\text{A.14})$$

Again assuming a Gaussian kernel this gives the acceleration divergence as

$$\nabla \cdot \mathbf{a} = -\frac{\rho_1}{\rho} \frac{\gamma + 1}{2} \delta v^2 \nabla^2 W \quad (\text{A.15a})$$

$$= -\frac{\rho_1}{\rho} \frac{\gamma + 1}{2} \delta v^2 2W_0 \left( \frac{|\mathbf{x} + h|}{h^2} \right)^2 \quad (\text{A.15b})$$

Again evaluating at the centre of the shock, and assuming the shock is smoothed symmetrically such that  $\rho = \frac{1}{2}(\rho_1 + \rho_2) = \gamma/(\gamma - 1)$  we obtain

$$\nabla \cdot \mathbf{a} \approx -2W_0 \frac{(\gamma - 1)(\gamma + 1)}{2\gamma} \left( \frac{\delta v}{h} \right)^2 \quad (\text{A.16})$$

Thus the new viscosity trigger is

$$A \approx 2W_0 \frac{(\gamma - 1)(\gamma + 1)}{2\gamma} \left( \frac{\delta v}{h} \right)^2 \quad (\text{A.17})$$

---

# B

## Viscosity in the Continuum Limit

Meglicki et al. (1993) showed that in the continuum limit the SPH viscous forces take the same functional form as that in the Navier-Stokes equation. For completeness the derivation is given below.

The linear term of SPH viscosity is given by

$$F = \sum_j m_j \frac{\alpha \bar{c}_{ij}}{\bar{\rho}_{ij}} \frac{h \mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{r_{ij}^2} \nabla W_{ij}. \quad (\text{B.1})$$

Replacing the sum over mass with an integral over density we obtain

$$F = \int \rho_j \frac{\alpha \bar{c}_{ij}}{\bar{\rho}_{ij}} \frac{h \mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{r_{ij}^2} \nabla W_{ij} d^3 z, \quad (\text{B.2})$$

where  $\mathbf{z} = (\mathbf{x}_i - \mathbf{x}_j)/h$ . We may Taylor expand around particle  $j$  such that

$$f_j = f(\mathbf{x}_i - h\mathbf{z}) = \sum_k \frac{(-h)^k}{k!} \mathbf{z} \cdot \nabla^k f(\mathbf{x}_i) = f_i - h\mathbf{z} \frac{\partial f_i}{\partial \mathbf{x}}. \quad (\text{B.3})$$

Using the notation  $-h\mathbf{z} = \mathbf{x}_j - \mathbf{x}_i = \Delta \mathbf{x}$  we may expand all terms in  $j$  using

$$f_j = f_i + \Delta \mathbf{x} \frac{\partial f_i}{\partial \mathbf{x}} \quad (\text{B.4})$$

$$c_j = c_i + \Delta x \frac{\partial c_i}{\partial \mathbf{x}} \quad (\text{B.5a})$$

$$\rho_j = \rho_i + \Delta x \frac{\partial \rho_i}{\partial \mathbf{x}} \quad (\text{B.5b})$$

$$\frac{1}{\rho_i + \rho_j} = \frac{1}{2\rho_i} - \frac{1}{4\rho_i^2} \Delta x \frac{\partial \rho_i}{\partial \mathbf{x}} \quad (\text{B.5c})$$

$$\mathbf{v}_j = -\Delta x^\alpha \frac{\partial v_i}{\partial \mathbf{x}^\alpha} - \frac{1}{2} \Delta x^\alpha \Delta x^\beta \frac{\partial^2 v_i}{\partial \mathbf{x}^\alpha \partial \mathbf{x}^\beta} \quad (\text{B.5d})$$

$$\mathbf{r}_{ij} = -\Delta \mathbf{x}. \quad (\text{B.5e})$$

This gives the viscous force on particle  $i$  as

$$F^k = ah \int \left[ \rho + \Delta x^l \frac{\partial \rho}{\partial \mathbf{x}^l} \right] \left[ 2c + \Delta x^m \frac{\partial c}{\partial \mathbf{x}^m} \right] \left[ \frac{1}{\rho} - \Delta x^n \frac{1}{4\rho^2} \frac{\partial \rho}{\partial \mathbf{x}^n} \right] \left[ \Delta x^p \frac{\partial v}{\partial \mathbf{x}^p} + \frac{1}{2} \Delta x^p \Delta x^q \frac{\partial^2 v}{\partial \mathbf{x}^p \partial \mathbf{x}^q} \right] \Delta x^r \frac{1}{r^3} \frac{dW}{dr} \Delta x^k d^3x. \quad (\text{B.6})$$

Terms that are odd in  $\Delta x$  cause the integral to vanish. Therefore we may simplify this expression to

$$F^k = \frac{\alpha h}{2} \left[ \frac{\partial c}{\partial x^p} \frac{\partial v^r}{\partial x^q} + \frac{c}{\rho} \frac{\partial \rho}{\partial x^p} \frac{\partial v^r}{\partial x^q} + c \frac{\partial^2 v^r}{\partial x^p \partial x^q} \right] \int \Delta x^p \Delta x^q \Delta x^r \Delta x^k \frac{1}{r^3} \frac{dW}{dr} d^3x. \quad (\text{B.7})$$

As the integral is a fourth order isotropic tensor<sup>1</sup> we may express the viscous force as

$$F^k = \frac{\alpha h \kappa}{2\rho} \left( c\rho \frac{\partial^2 v^k}{\partial x^p \partial x^p} + \frac{\partial c\rho}{\partial x^p} \frac{\partial v^k}{\partial x^p} + c\rho \frac{\partial^2 v^p}{\partial x^p \partial x^k} + \frac{\partial c\rho}{\partial x^p} \frac{\partial v^p}{\partial x^k} + c\rho \frac{\partial^2 v^q}{\partial x^k \partial x^q} + \frac{\partial c\rho}{\partial x^k} \frac{\partial v^q}{\partial x^q} \right). \quad (\text{B.8})$$

By defining

$$S_{ij} = \frac{\partial v^i}{\partial x^j} + \frac{\partial v^j}{\partial x^i}, \quad (\text{B.9})$$

---

<sup>1</sup>An tensor is isotropic if its components do not vary if rotated. A fourth order tensor may be expressed as  $\eta_{ijkl} = c_1 \delta_{ij} \delta_{kl} + c_2 \delta_{ik} \delta_{jl} + c_3 \delta_{il} \delta_{jk}$

we can write the viscous force as

$$F^k = \frac{\alpha h \kappa}{2\rho} [\nabla \cdot (c\rho S) + \nabla(c\rho \nabla \cdot v)] \quad (\text{B.10})$$

which may be expressed as

$$F^k = \frac{\alpha h \kappa}{2\rho} \left[ \frac{\partial}{\partial x^p} \left[ c\rho \frac{\partial v^k}{\partial x^p} + c\rho \frac{\partial v^p}{\partial x^k} + \delta_{pk} c\rho \frac{\partial v^q}{\partial x^q} \right] \right]. \quad (\text{B.11})$$

Comparing this to the viscous terms of the Navier-Stokes equation

$$F = \frac{1}{\rho} \left[ \frac{\partial}{\partial x_i} \left( \eta \left( \frac{\partial v^j}{\partial x^i} + \frac{\partial v^i}{\partial x^j} - \frac{2}{3} \delta_{ij} \nabla \cdot v \right) \right) + \xi \delta_{ij} \nabla \cdot v \right] \quad (\text{B.12})$$

it is clear that

$$\eta = \frac{\alpha h \kappa}{2} c\rho \quad (\text{B.13})$$

and

$$\xi = \frac{5}{3} \eta \quad (\text{B.14})$$



# Polytrope Oscillations

## C.1 Stable Polytropic Setup

In this section we describe the setup of a stable polytropic sphere. A polytropic sphere is described by the Lane-Emden equation, which describes the hydrostatic equilibrium for a self-gravitating, polytropic fluid which is spherically symmetric. The Lane-Emden equation is

$$\frac{d^2\theta}{d\xi^2} = \frac{d\theta}{d\xi} - \theta^n, \quad (\text{C.1})$$

where  $\theta$  is the dimensionless density and  $\xi$  is the dimensionless radius.  $n$  is the polytropic index and is related to the polytropic gamma as  $n = 1/(\Gamma - 1)$ . It should be noted that the polytropic index  $\Gamma$  is not necessarily equal to the adiabatic index  $\gamma$  which is given by the ratio of specific heats. The polytropic index  $\Gamma$  gives a power law describing how a change in density with radius gives rise to a change in pressure with radius, that is

$$P = K\rho^\Gamma. \quad (\text{C.2})$$

The polytropic constant  $K$  is assumed to be constant with radius. A polytropic fluid, which is a type of barotropic fluid (pressure only depends on density), may have this equation of state where again the polytropic index is not equal to the

adiabatic index. An example of this is an isothermal ideal gas, the polytropic index is  $\Gamma = 1$  which differs from the ratio of specific heats  $\gamma$ .

The Lane-Emden equation may be solved numerically as

$$\theta_{i+1} = \theta_i + \Delta\xi \left. \frac{d\theta}{d\xi} \right|_i \quad (\text{C.3})$$

$$\left. \frac{d\theta}{d\xi} \right|_{i+1} = \left. \frac{d\theta}{d\xi} \right|_{i+1} + \left. \frac{d^2\theta}{d\xi^2} \right|_i \Delta\xi \quad (\text{C.4})$$

by starting at an initial state  $i = 0$  corresponding to the centre of the polytrope one integrates to the outer edge of the polytrope where  $\theta = 0$ . The boundary conditions are  $\frac{d\theta}{d\xi} = 0$  and  $\theta = 1$  at  $\xi = 0$ . The second derivative of the density can be found from the Lane-Emden equation C.1 as the variables at  $i$  are known. Once this equation is solved one may transform  $\theta$  and  $\xi$  into scaled quantities. A scale factor relating the scaled radius and  $\xi$  is given as

$$\alpha = \frac{R}{\xi_{max}}. \quad (\text{C.5})$$

The core density may be computed using the average density as

$$\rho_c = -\frac{3M}{4\pi R^3} \left( 3 \left| \frac{1}{\xi} \frac{d\theta}{d\xi} \right|_{\xi=\xi_{max}} \right)^{-1} \quad (\text{C.6})$$

and the entropy constant is

$$K = \frac{\alpha^2 4\pi G \rho_c^{(n+1)/n}}{n+1}. \quad (\text{C.7})$$

This allows us to compute the mass-radius relation, the scaled density, the potential and the pressure gradient as

$$m(r) = 4\pi\alpha^3 \rho_c \xi^2 \frac{d\theta}{d\xi}, \quad (\text{C.8})$$

$$\rho = \rho_c \theta^n, \quad (\text{C.9})$$

$$\phi = -\frac{\Gamma K \rho^{\Gamma-1}}{\Gamma-1} - 1, \quad (\text{C.10})$$

$$\frac{dP}{dr} = -\rho m/r^2. \quad (\text{C.11})$$

The pressure is computed using equation C.2. To construct initial conditions for

the purposes of SPH we firstly set up a 3D uniform grid by constructing a set of face-centred-cubic close packed boxes. Particles outside of the maximum of the radius of the polytrope are deleted. The polytropic equations are solved and the particles are radially stretched in order to achieve the correct density profile. Using this setup means that the particles will not be exactly in equilibrium. Equilibrium may be achieved by integrating the system with an polytropic equation of state and a high viscosity (say  $\alpha = 2$ ). The dissipated kinetic energy will be removed from the system.

## C.2 Applying a Perturbation

The perturbation of a polytrope is discussed in detail by Cox (1980), here we will give an overview. We may apply a small radial perturbation by deriving and solving the linear adiabatic wave equation. We apply a small perturbation of the form

$$\eta(r, t) = \frac{\delta r}{r} = \epsilon(r)e^{i\omega t}. \quad (\text{C.12})$$

In order to apply the perturbation we must find the solution to the eigenfunction  $\epsilon(r)$ . Firstly we derive the perturbation equations for mass, momentum and energy conservation.

### C.2.1 Mass Conservation

The unperturbed equation is

$$\frac{\partial r}{\partial m} = \frac{1}{4\pi\rho(r)r^2}. \quad (\text{C.13})$$

If we apply a small perturbation this equation becomes

$$\delta \left( \frac{\partial r}{\partial m} \right) = \delta \left( \frac{1}{4\pi\rho(r)r^2} \right) \quad (\text{C.14})$$

$$\epsilon \frac{\partial r}{\partial m} + r \frac{\partial \epsilon}{\partial m} = -\frac{1}{4\pi\rho(r)r^2} \left( \frac{\delta \rho}{\rho} + \frac{\delta r}{r} \right). \quad (\text{C.15})$$

By multiplying through by  $\frac{\partial m}{\partial r}$  and using  $\frac{\delta r}{r} = \eta$  we obtain

$$\frac{\delta \rho}{\rho} = -3\eta - r \frac{\partial \eta}{\partial r}. \quad (\text{C.16})$$

### C.2.2 Energy Equation

We use the equation of state to derive the perturbation for the pressure. The unperturbed equation is

$$P = K\rho^\Gamma. \quad (\text{C.17})$$

Applying a perturbation and dividing by the unperturbed equation we obtain

$$\frac{\delta P}{P} = \Gamma \frac{\delta \rho}{\rho}. \quad (\text{C.18})$$

### C.2.3 Momentum Conservation

The unperturbed momentum equation is given by

$$\frac{d^2 r}{dt^2} = -\frac{1}{\rho} \nabla P + f. \quad (\text{C.19})$$

If we apply a perturbation to this equation we obtain

$$\frac{d^2 \delta r}{dt^2} = -\delta \left( \frac{1}{\rho} \nabla P \right) + \delta f. \quad (\text{C.20})$$

By replacing the radial gradients, with mass gradients (using equation C.13) we obtain

$$\frac{d^2 \delta r}{dt^2} = -4\pi r^2 \left( \frac{\partial \delta P}{\partial m} + \frac{2\delta r}{r} \frac{\partial P}{\partial m} \right) + Gm(r) \delta \left( \frac{1}{r^2} \right). \quad (\text{C.21})$$

Tidying up and replacing  $\delta r/r = \eta$  we obtain

$$\frac{d^2 \delta r}{dt^2} = -4\pi r^2 \left( \frac{\partial \delta P}{\partial m} + 2\eta \frac{\partial P}{\partial m} \right) + \frac{2Gm(r)}{r^2} \eta. \quad (\text{C.22})$$

### C.2.4 Linear Adiabatic Wave Equation (LAWWE)

We will now derive the LAWWE which will describe the spatial part of the perturbation to be applied. By taking the perturbed momentum equation C.22 we may substitute



$\delta P$  using equation C.18 and replace  $Gm/r^2 = -4\pi r^2 \partial P / \partial \rho$ ; this gives

$$\frac{d^2 \delta r}{dt^2} = -4\pi r^2 \left( 4\eta \frac{\partial P}{\partial m} + \frac{\partial}{\partial m} \left( P \Gamma \frac{\delta \rho}{\rho} \right) \right) \quad (\text{C.23})$$

$$= -4\pi r^2 \left( 4\eta \frac{\partial P}{\partial m} + \frac{\partial}{\partial m} (-3P\Gamma\eta) + \frac{\partial}{\partial m} (-\Gamma P r \eta') \right), \quad (\text{C.24})$$

where we have used equation C.16. If we expand this out we obtain

$$\frac{d^2 \delta r}{dt^2} = 4\pi r^2 \left( \eta(3\Gamma - 4) \frac{\partial P}{\partial m} + 12\pi \Gamma r^2 \eta' + 4\pi r^2 \frac{\partial}{\partial m} (4\pi \Gamma \rho P r^3 \eta') \right). \quad (\text{C.25})$$

We combine the 2nd and 3rd term using the identity below

$$\frac{4\pi \Gamma}{r} \left( 3r^3 \eta' + r^3 \frac{\partial}{\partial m} (4\pi \rho P r^3 \eta') \right) = \frac{4\pi \Gamma}{r} \frac{\partial}{\partial m} (4\pi \rho P r^3 \eta' r^3) \quad (\text{C.26})$$

thus we may write the perturbation as

$$\frac{d^2 \delta r}{dt^2} = 4\pi r^2 \eta(3\Gamma - 4) \frac{\partial P}{\partial m} + \frac{4\pi \Gamma}{r} \frac{\partial}{\partial m} (4\pi \rho P r^3 \eta' r^3). \quad (\text{C.27})$$

Using the identity C.13 to rewrite this expression in terms of radial gradients we obtain

$$\frac{d^2 \delta r}{dt^2} = \frac{1}{\rho} \eta(3\Gamma - 4) \frac{\partial P}{\partial r} + \frac{1}{\rho r^3} \frac{\partial}{\partial r} (\Gamma P r^4 \eta'). \quad (\text{C.28})$$

We may use the definition of  $\delta r$  to obtain

$$\frac{d^2 \delta r}{dt^2} = -\omega^2 r \epsilon e^{i\omega t} = -\omega^2 r \eta. \quad (\text{C.29})$$

Thus we may write the spatial part of the perturbation as

$$\frac{\partial}{\partial r} (\Gamma P r^4 \epsilon') + \epsilon \left( \omega^2 \rho r^4 + r^3 (3\Gamma - 4) \frac{\partial P}{\partial r} \right) = 0. \quad (\text{C.30})$$

The equation is the LAWE equation and describes how the spatial part of the perturbation  $\epsilon$  and its derivatives are related to pressure gradient in the star and  $\omega$ , the frequency of the radial perturbation.

## C.3 Solutions to LAWE

The solution to this equation is similar to that of the Lane-Emden equation, however, we need to find the two boundary conditions.

### C.3.1 Boundary Conditions

Unlike the Lane-Emden equation which contains two boundary conditions at the same point in space (at  $\xi = 0$ ), the boundary conditions for the LAWE equation contains only one boundary condition at the centre, the other is at the surface. The centre boundary conditions requires that  $\epsilon' = 0$ , and is required to ensure that the perturbation  $\delta r$  does not become singular at the centre. The surface boundary condition comes from the requirement that the pressure (and density) drop to zero at the surface, therefore the perturbation in these quantities must also drop to zero.

To obtain the surface condition we rewrite the LAWE equation in dimensionless form, this is first done by expanding out the LAWE equation and introducing the variables  $\lambda = (1/P)\partial P/\partial r$  and  $r = xR$

$$\epsilon'' = \frac{\lambda}{x} \left[ \left( x - \frac{4}{\lambda R} \right) \epsilon' - \frac{\epsilon}{\Gamma} \left( \frac{\omega^2 x \rho}{\lambda P} - \frac{3\Gamma - 4}{R} \right) \right]. \quad (\text{C.31})$$

At the surface of the star  $P \rightarrow 0$ , therefore  $\lambda \rightarrow \infty$ . Therefore, in order for  $\epsilon''$  to have a finite value the bracketed function must approach zero. Also noting that the term  $4/(\lambda R) \rightarrow 0$ , then we may express the condition as

$$\epsilon' x - \frac{\epsilon}{\Gamma} \left( \frac{\omega^2 x \rho}{\lambda P} - \frac{3\Gamma - 4}{R} \right) = 0. \quad (\text{C.32})$$

Using the identity

$$\frac{\rho}{\lambda P} = \frac{P\rho}{P} \left( \frac{\partial P}{\partial r} \right)^{-1} = \frac{1}{g} \quad (\text{C.33})$$

and that at the surface of the polytrope  $x = R = g = 1$ , and normalising such that  $\epsilon_R = 1$ , we obtain the surface boundary condition as

$$\epsilon' = \frac{1}{\Gamma} (\omega^2 - (3\Gamma - 4)). \quad (\text{C.34})$$

### C.3.2 The Solution

If we only had to satisfy a single boundary condition the solution of the LAWE equation would be simple. If we had to satisfy the central boundary condition we would set  $\epsilon' = 0$  and integrate outwards starting at the centre.  $\epsilon$  and  $\omega$  would be set to some arbitrary value we could pick, in other words a solution would be possible for all frequencies. If we integrated inwards from the surface we would set  $\epsilon = 1$  and  $\epsilon'$  using a given value of  $\omega$  and equation C.34. Again a solution would be possible for all frequencies.

However, as we need to satisfy both boundary conditions then  $\omega$  and  $\epsilon_{r=0}$  must be regarded as trial parameters, and so only a discrete set of eigenfrequencies will be possible solutions. Therefore  $\epsilon$  will only take a finite set of eigenvectors. To solve this equation we may use a shooting method as described in numerical recipes (Press et al., 2007). As we are interested in the fundamental mode, we only solve for the fundamental frequency, this will always be the lowest frequency that satisfies the LAWE equation and both boundary conditions.

Once we have solved the LAWE equation we may use  $\epsilon$  to apply a perturbation of the form C.12.

---

## References

- Abel T., 2010, ArXiv e-prints
- Agertz O., Moore B., Stadel J., Potter D., Miniati F., Read J., Mayer L., Gawryszczak A., Kravtsov A., Nordlund Å., Pearce F., Quilis V., Rudd D., Springel V., Stone J., Tasker E., Teyssier R., Wadsley J., Walder R., 2007, MNRAS, 380, 963
- Attwood R. E., Goodwin S. P., Whitworth A. P., 2007, A&A, 464, 447
- Balsara D., 1995, Journal of Computational Physics, 121, 357
- Barnes J., Hut P., 1986, Nature, 324, 446
- Bastien P., Cha S., Viau S., 2004, in G. Garcia-Segura, G. Tenorio-Tagle, J. Franco, & H. W. Yorke ed., Revista Mexicana de Astronomia y Astrofisica Conference Series Vol. 22 of Revista Mexicana de Astronomia y Astrofisica Conference Series, SPH with radiative transfer: method and applications. pp 144–147
- Bate M., 1995, PhD Thesis, University of Cambridge, 0
- Brookshaw L., 1985, Proceedings of the Astronomical Society of Australia, 6, 207
- Brown W. R., Geller M. J., Kenyon S. J., Kurtz M. J., 2005, ApJ, 622, L33
- Cartwright A., Stamatellos D., Whitworth A. P., 2009, MNRAS, 395, 2373
- Cha S.-H., 2002, PhD Thesis, Cardiff University, 0
- Cha S.-H., Whitworth A. P., 2003, MNRAS, 340, 73
- Chandrasekhar S., 1961, Hydrodynamic and hydromagnetic stability. International Series of Monographs on Physics, Oxford: Clarendon, 1961
- Colella P., 1990, Journal of Computational Physics, 87, 171
- Cox J. P., 1980, Theory of Stellar Pulsation, 1st edn. Published by Princeton University Press, Princeton, NJ, USA

- 
- Das R., Cleary P. W., 2009, in A. Korsunsky ed., American Institute of Physics Conference Series Vol. 1138 of American Institute of Physics Conference Series, Simulating Brittle Fracture of Rocks using Smoothed Particle Hydrodynamics. pp 1–12
- Dehnen W., 2000, *ApJ*, 536, L39
- Dehnen W., 2002, *Journal of Computational Physics*, 179, 27
- Dolag K., Bartelmann M., Lesch H., 1999, *A&A*, 348, 351
- Dolag K., Vazza F., Brunetti G., Tormen G., 2005, *MNRAS*, 364, 753
- Drazin P. G., Reid W. H., 1981, NASA STI/Recon Technical Report A, 82, 17950
- Eisenhauer F., Schödel R., Genzel R., Ott T., Tecza M., Abuter R., Eckart A., Alexander T., 2003, *ApJ*, 597, L121
- Evans C. R., Kochanek C. S., 1989, *ApJ*, 346, L13
- Evrard A. E., 1988, *MNRAS*, 235, 911
- Frank J., 1978, *MNRAS*, 184, 87
- Frenk C., White S., Bode P., Bond J., Bryan G., Cen R., Couchman H., Evrard A., Gnedin N., Jenkins A., Khokhlov A., Klypin A., Navarro J., Norman M., Ostriker J., Owen J., Pearce F., Pen U.-L., Steinmetz M., 1999, *ApJ*, 525, 554
- Gillessen S., Eisenhauer F., Fritz T. K., Bartko H., Dodds-Eden K., Pfuhl O., Ott T., Genzel R., 2009, *ApJ*, 707, L114
- Gingold R. A., Monaghan J. J., 1977, *MNRAS*, 181, 375
- Gingold R. A., Monaghan J. J., 1983, *MNRAS*, 204, 715
- Godunov S., 1959, *Mat. Sb.*, 47, 271
- Goldreich P., Narayan R., 1985, *MNRAS*, 213, 7P
- Greengard L., Rokhlin V., 1997, *Journal of Computational Physics*, 135, 280
- Gültekin K., Richstone D. O., Gebhardt K., Lauer T. R., Tremaine S., Aller M. C., Bender R., Dressler A., Faber S. M., Filippenko A. V., Green R., Ho L. C., Kormendy J., Magorrian J., Pinkney J., Siopis C., 2009, *ApJ*, 698, 198

- 
- Hernquist L., 1993, ApJ, 404, 717
- Hernquist L., Katz N., 1989, Astrophysics Journal, Supplement, 70, 419
- Hills J. G., 1975, Nature, 254, 295
- Hills J. G., 1978, MNRAS, 182, 517
- Hills J. G., 1988, Nature, 331, 687
- Hobbs A., Nayakshin S., Power C., King A., 2010, ArXiv e-prints
- Imaeda Y., Inutsuka S.-i., 2002, ApJ, 569, 501
- Inutsuka S.-I., 2002, Journal of Computational Physics, 179, 238
- Joss P. C., Rappaport S., Lewis W., 1987, ApJ, 319, 180
- Jutzi M., Benz W., Michel P., 2008, Icarus, 198, 242
- Kermani M.J. andX Gerber A., Stockie J., 2003, The 4th Conference of Iranian AeroSpace Society, 569, 501
- Komossa S., Bade N., 1999, in Poutanen J., Svensson R., eds, High Energy Processes in Accreting Black Holes Vol. 161 of Astronomical Society of the Pacific Conference Series, The Giant X-ray Outburst in NGC 5905 - a Tidal Disruption Event?. pp 234–+
- Li P., Johnston H., Krasny R., 2009, J. Comput. Phys., 228, 3858
- Lodato G., King A. R., Pringle J. E., 2009, MNRAS, 392, 332
- Lucy L. B., 1977, AJ, 82, 1013
- Lynden-Bell D., 1969, Nature, 223, 690
- Lynden-Bell D., Pringle J. E., 1974, MNRAS, 168, 603
- Lyubarskij Y. E., Postnov K. A., Prokhorov M. E., 1994, MNRAS, 266, 583
- Maddison S. T., Murray J. R., Monaghan J. J., 1996, Publications of the Astronomical Society of Australia, 13, 66
- Meglicki Z., Wickramasinghe D., Bicknell G. V., 1993, MNRAS, 264, 691

- 
- Monaghan J. J., 1992, *Annual Review of Astron and Astrophysics*, 30, 543
- Monaghan J. J., 1997, *Journal of Computational Physics*, 136, 298
- Monaghan J. J., 2005, *Reports of Progress in Physics*, 68, 1703
- Monaghan J. J., 2006, *MNRAS*, 365, 199
- Morris J. P., 1996, *Publications of the Astronomical Society of Australia*, 13, 97
- Morris J. P., Monaghan J. J., 1997, *Journal of Computational Physics*, 136, 41
- Murray J. R., 1998, *MNRAS*, 297, 323
- Papaloizou J. C. B., Pringle J. E., 1984, *MNRAS*, 208, 721
- Papaloizou J. C. B., Pringle J. E., 1985, *MNRAS*, 213, 799
- Peterson D. M., Hummel C. A., Pauls T. A., Armstrong J. T., Benson J. A., Gilbreath G. C., Hindsley R. B., Hutter D. J., Johnston K. J., Mozurkewich D., Schmitt H. R., 2006, *Nature*, 440, 896
- Phinney E. S., 1989, in Morris M., ed., *The Center of the Galaxy Vol. 136 of IAU Symposium, Manifestations of a Massive Black Hole in the Galactic Center.* pp 543–+
- Press W., Teukolsky S., Vetterling W., Flannery B., 2007, *Numerical Recipes in C*, 3rd edn. Cambridge University Press, Cambridge, UK
- Price D. J., 2004, Thesis
- Price D. J., 2008, *Journal of Computational Physics*, 227, 10040
- Price D. J., Monaghan J. J., 2004, *MNRAS*, 348, 139
- Pringle G. J., 1995, in *PPSC Comparison of an  $o(n)$  and an  $o(n \log n)$   $n$ -body solver.* pp 337–342
- Quinn T., Katz N., Stadel J., Lake G., 1997, *ArXiv Astrophysics e-prints*
- Read J. I., Hayfield T., Agertz O., 2010, *MNRAS*, pp 767–+
- Rees M. J., 1988, *Nature*, 333, 523

- 
- Ritchie B. W., Thomas P. A., 2001, MNRAS, 323, 743
- Rosswog S., Davies M. B., Thielemann F.-K., Piran T., 2000, A&A, 360, 171
- Rosswog S., Price D., 2007, MNRAS, 379, 915
- Silverman B. W., 1986, Density estimation for statistics and data analysis
- Sobol I. M., 1967, USSR Computational Mathematics and Mathematical Physics, 7, 86
- Sod G. A., 1978, Journal of Computational Physics, 27, 1
- Springel V., 2005, MNRAS, 364, 1105
- Springel V., 2009, ArXiv e-prints
- Springel V., Hernquist L., 2002, MNRAS, 333, 649
- Stamatellos D., Whitworth A. P., 2005, A&A, 439, 153
- Steinmetz M., Mueller E., 1993, A&A, 268, 391
- Swegle J., D. H., S. A., 1995, Journal of Computational Physics, 116, 123
- Tasker E. J., Brunino R., Mitchell N. L., Michielsen D., Hopton S., Pearce F. R., Bryan G. L., Theuns T., 2008, MNRAS, 390, 1267
- Teyssier R., 2002, A&A, 385, 337
- Toro E., 1999, Riemann Solvers and Numerical Methods for Fluid Dynamics 2nd Ed. Springer, Berlin
- van Leer B., 1979, Journal of Computational Physics, 32, 101
- Wetzstein M., Nelson A. F., Naab T., Burkert A., 2008, ArXiv e-prints
- Whitehouse S. C., Bate M. R., 2004, MNRAS, 353, 1078
- Young P. J., Shields G. A., Wheeler J. C., 1977, ApJ, 212, 367