2

Numerical methods:
simulations with smoothed particle hydrodynamics

"...and one has to resort to the indignity of numerical simulations to settle even the simplest questions... "
– Philip Anderson, Nobel Lecture, 1977

2.1   Introduction

In order to construct synthetic Galactic observations we first need to generate the physical conditions from which to build them. Numerical simulations presented in this thesis are performed using the *smoothed particle hydrodynamical* technique, or SPH, where the fluid is decomposed into discretised packets, or *particles*. This chapter presents an overview of the SPH method and codes used throughout this thesis. We begin with a discussion of the basic principles, including some simple tests of the codes, in Section 2.2. We then discuss the additions to allow for the study of ISM scale physics in Section 2.3. This includes additions to the standard energy evolution to allow for ISM heating and cooling, and the various chemical processes needed to track the Galactic molecular content. Two existing SPH codes are used in this study, the structure of each and the developments made for the work in this thesis are discussed in Section 2.2.12.

These codes and techniques will be utilised in simulations using both fixed analytic stellar potentials and live stellar systems, presented in Chapters 3, 4 and 5.

2.2 Smoothed Particle Hydrodynamics

The problem of simulating fluid motion is by no means a new one. Clearly the fluid cannot be simulated on a point mass level, molecule-by-molecule, so some large scale discretisation must be used. The two main techniques are to either evolve the flow of packets of the fluid in a Lagrangian prescription, or to measure the flow of fluid though a static or adaptive grid in a Eulerian sense. These methods include Eulerian grid based method such as adaptive mesh refinement, or AMR, codes (e.g. FLASH, Fryxell et al. 2000; RAMSES, Teyssier 2002; ENZO, Bryan et al. 2014), Lagrangian particle based methods such as smoothed particle hydrodynamics, or SPH, codes (e.g. GADGET, Springel et al. 2001; GASOLINE, Wadsley et al. 2004; SEREN, Hubber et al. 2011) or some intermix of the two such as moving mesh codes (e.g. AREPO: Springel 2010a). While each has its own merits, no clear consensus has been reached as to whether there is a single method that outstrips the others in every regard, though numerous studies have focused on comparing the different techniques (e.g. Agertz et al. 2007, Tasker et al. 2008, Price & Federrath 2010), with a large number specifically focussing on comparisons in a cosmological context (e.g. Frenk et al. 1999, Thacker et al. 2000, O'Shea et al. 2005).

In this thesis simulations are performed using the SPH method, specifically using the codes PHANTOM (Price & Federrath 2010) and SPHNG (Bate et al. 1995) the details of which will be discussed in greater detail later in this chapter. The SPH method was first formulated by Lucy (1977) and Gingold & Monaghan (1977). The crux of the method is to discretise a fluid into a finite number of mass elements with variable density, i.e. volume. The density of a particle is calculated by interpolation between neighbouring particles, normalised by some smoothing function known as the smoothing kernel function, $W(r)$, that decreases in magnitude with increasing distance from the particle of interest.

In the 40 years since its conception the SPH method has evolved considerably. While today it is arguably most well know for its application in the field of cosmology it has been applied to numerous media both in and out of astrophysics (e.g. the study of accretion discs, planet and star formation). As such there are numerous additional physics that have been added to study different problems, including magnetic fields, radiative transfer and relativity. Further information on these improvements, as well as the basics of SPH can be found in the reviews of Price (2012a), Rosswog (2009), Monaghan (1992, 2005) and (Springel 2010b). As with any numerical method, there are key benefits and detriments of SPH. Some, but certainly not all, of these are highlighted below in particular in comparison to grid-based methods.

Advantages

- SPH is usually boundless, so no matter is lost or forced back into the simulation at domain boundaries thereby inherently conserving the mass of the system.

- The code spends its time evolving only the regions with a non-zero density field, so no time is wasted modelling empty space as in grid codes.

- The adaptive resolution (both spatially and temporally) are relatively easy to implement, i.e.

no need to re-create/adapt meshes as in grid-based codes.

- The fluid evolution history is intrinsically simple to trace due to the particle-like nature. This would require the inclusion of tracer particles in grid-based codes to follow the fluid flow.

- SPH is comparatively simple and easy to model complex physics and geometries in 3D, due to free/moving material boundaries.

- Particle nature makes coupling to *N*-body or self-gravity physics relatively straight-forward.

- The distribution of mass between particles ensures exact conservation of mass, as the mass of each particle is constant throughout time.

Disadvantages

- Need to build and constantly update neighbour lists (by link-lists or binary trees) in order to evaluate particle summations.

- The initial conditions can be influential on the eventual outcome. Need to decide on whether to set particles on a cubic, hexagonal or random lattice arrangement initially.

- Resolution is limited by particle number, which is fixed at the start of the simulation, whereas in theory a grid can be sub-divided indefinitely.

- Radiative transfer and magnetohydrodynamics can be more difficult to implement than the cell structured nature of grid-based codes.

2.2.1 Equations of fluid dynamics

Throughout this chapter we will be referring to numerous standard formulae to derive and explain the SPH equations. We include these here briefly for reference before continuing. SPH is a Lagrangian fluid formulation by design, where the Lagrangian itself takes the classical form

$$L = T - V = \int \left(\frac{1}{2} \rho v^2 - \rho u \right) d\vec{r}, \tag{2.1}$$

which is simply the difference between the kinetic and thermal potential energy, *T* and *V* respectfully (neglecting gravity for now), where ρ, u, v and $\vec{r}$ are the density, internal energy, velocity and position of a fluid element. We can minimise the action of the Lagrangian to give the Euler-Lagrange equations of fluid dynamics

$$\frac{\partial L}{\partial \vec{r}} - \frac{d}{dt} \frac{\partial L}{\partial \vec{v}} = 0, \tag{2.2}$$

which can be used with an appropriate Lagrangian to obtain the equations of motion (EoM) of the fluid system. We will also be making use of the material, or Lagrangian, derivative which is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \vec{\nabla}, \tag{2.3}$$

where the first derivative is the local rate of change (i.e. the Eulerian derivative), and the second the convective derivative. We will also utilise the Navier-Stokes equations for fluid flow (a.k.a. the Euler equations) given by

$$\frac{D\vec{v}}{Dt} + \frac{\vec{\nabla}P}{\rho} + \vec{f}_{visc} + \vec{f}_{ext} = 0, \tag{2.4}$$

where $\vec{f}_{visc}$ contains all the physical viscosity information and may also contain extra magneto-hydrodynamical (MHD) terms. For gravitational external forces $\vec{f}_{ext} = -\vec{\nabla}\Phi$, where Φ is the external potential. In the work presented here Φ comes from galactic potentials we impose, which are related to the density distribution by Poisson's equation, $\nabla^2\Phi = \rho_{ext}(\vec{r})4\pi G$. The material derivative is effectively Newton's second law for a fluid. Built into any numerical fluid simulation should be the continuity equation, which ensures mass conservation in the system

$$\frac{\partial\rho}{\partial t} + \vec{\nabla} \cdot (\rho\vec{v}) = \frac{\partial\rho}{\partial t} + \vec{v} \cdot \vec{\nabla}\rho + \rho\vec{\nabla} \cdot \vec{v} = \frac{D\rho}{Dt} + \rho\vec{\nabla} \cdot \vec{v} = 0. \tag{2.5}$$

And finally we will also use the first law of thermodynamics. For an adiabatic ($dQ = 0$) equation of state (EoS) we have $dU = dQ - PdV = Pd\rho/\rho^2$, giving the rate of change of internal energy as

$$\frac{Du}{Dt} = \frac{P}{\rho^2}\frac{D\rho}{Dt}. \tag{2.6}$$

where specific internal energy is $u = U/m$. Between these formulae we have a framework to evolve a fluid system over time, tracing changes in velocity, internal energy and position ($d\vec{r}/dt = \vec{v}$) while maintaining mass conservation by satisfying Equation 2.5. We can also calculate the total energy simply by the addition of the internal to the kinetic energy, $e = u + v^2/2$.

2.2.2 The SPH kernel

The kernel function is a key parameter of SPH, and defines how much we care about particle neighbours when calculating fluid properties, akin to a window function. It effectively puts the "smooth" in SPH, and makes sure particle properties are smoothly interpolated from neighbouring particles. The kernel is defined by some scale length/smoothing length/kernel support radius, h, which determines the rate of radial decay. Two basic properties of an appropriate kernel are that it is correctly normalised

$$\int W(|\vec{r} - \vec{r}'|, h)d^3\vec{r}' = 1, \tag{2.7}$$

and that it tends to a delta function as the kernel support radius tends to 0

$$\lim_{h\to 0} W(|\vec{r} - \vec{r}'|, h) = \delta(\vec{r} - \vec{r}'). \tag{2.8}$$

The kernel is chosen to be spherically symmetric so that the system is independent of rotation, and to be only a function of separation, rather than actual particle position. We can first define a property of the system, $A(\vec{r})$, using the delta function by

$$A(\vec{r}) = \int A\delta(\vec{r} - \vec{r}')d^3\vec{r}'. \tag{2.9}$$

We can then approximate this expression with our kernel function to give the integral interpolation approximation

$$A(\vec{r}) = \int A(\vec{r}\,')W(|\vec{r} - \vec{r}\,'|, h)d^3\vec{r}\,', \tag{2.10}$$

where the parameter A is integrated over all other fluid elements at positions $\vec{r}\,'$. This would produce $A(\vec{r})$ exactly when the kernel is the delta function. For practical purposes the above integral is formulated into a summation over a set of interpolation points throughout the medium, the SPH particles. This then means we can estimate A for some particle at $\vec{r}$ by a weighted sum of that same A evaluated at every other particle at $\vec{r}\,'$. We can replace the volume integral for a mass integral, as in the usual construction of SPH we know the mass of each particle, so that Equation 2.10 becomes

$$A(\vec{r}) = \int \frac{A(\vec{r}\,')}{\rho(\vec{r}\,')}W(\vec{r} - \vec{r}\,', h)\rho(\vec{r}\,')d^3 r' = \int \frac{A(\vec{r}\,')}{\rho(\vec{r}\,')}W(\vec{r} - \vec{r}\,', h)dm \tag{2.11}$$

we can then form a discretised version, by a summation interpolation using particles as our interpolation points. This gives a sum over b particles at positions $\vec{r}\,'$

$$A(\vec{r}) = \sum_b m_b \frac{A_b}{\rho_b} W(\vec{r} - \vec{r_b}, h) \tag{2.12}$$

where m, ρ are the particles' mass and density respectively. These leads to a way of calculating some property of the system, A, at any given point by summation over discretised fluid elements. For instance, we can find the density simply by substituting $A(\vec{r}) \rightarrow \rho(\vec{r})$ to give the standard SPH density summation equation

$$\rho(\vec{r}) = \sum_b m_b W(\vec{r} - \vec{r_b}, h). \tag{2.13}$$

In replacing the integral with summation expressions (Equations 2.10 and 2.12) with have introduced some discretisation/sampling error of the order h^2 (Monaghan 1992, 2005), which depends on particle disorder and is reduced by increasing the number of particles (Price 2005). There is also an error associated with the initial introduction of the integral interpolation Equation 2.10 (Monaghan 2005; Liu & Liu 2010) which is also of order h^2, which is reduced by decreasing the smoothing length (evident by Equation 2.8). This would lead to the conclusion of using a high number of particles that are well ordered and a kernel with a small smoothing length to reduce errors. However this would increase computational time-scales, requiring a give-and-take approach between reducing errors and optimising calculation times (Cossins 2010).

By using Equation 2.13 we have the capability to calculate the density of the fluid at any arbitrary point by summing over the kernels of the neighbouring particles. This is illustrated by the cartoon in Figure 2.1. We shall show later in this chapter that this approach can be used to calculate other properties of the system, specifically the forces and internal energies of the particles. But first we turn our attention to the actual form of the kernel function.

Ideally an appropriate kernel must have several properties, in addition to those of Equations 2.7 and 2.8. The kernel should be "bell-shaped" (Lucy 1977) so that the kernel smoothly decreases with distance but also flattens near the centre to avoid steep gradients for particles in close proxim-
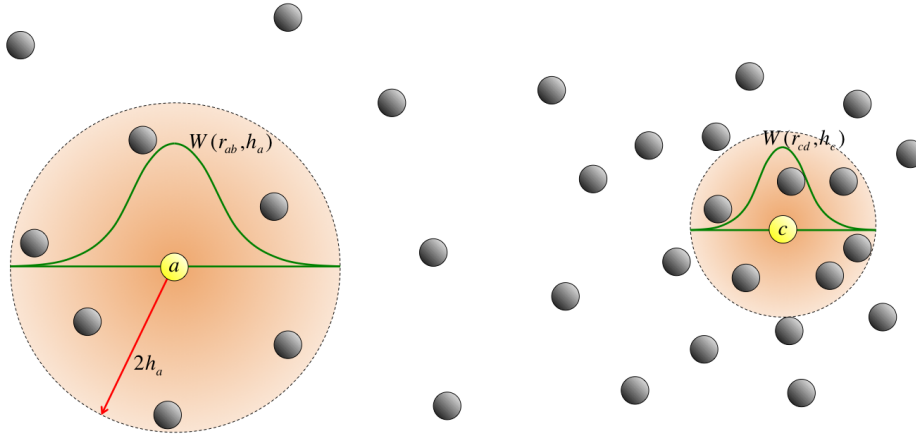
Figure 2.1: Simple schematic of how to visualise SPH particles. Density is calculated by summing over neighbours, with masses weighted by some smoothing kernel, $W(r, h)$, that decays with distance from the particle of interest. The compact support shown here is $2h$, specifically that of the cubic spline kernel. This cartoon shows particles with individual smoothing lengths so that each particle has approximately equal neighbours.

ity. The kernel should also be smoothly differentiable (at least singularly) and be an odd function and ≥ 0 in all space. Given these pre-requisites the simplest choice is a Gaussian kernel function which has the advantage of being smooth for any order of differentiation (employed by Gingold & Monaghan 1977). However as it is non-zero at all radii we would require the summation over all neighbours. Instead we can choose to limit our summation to a kernel of "compact support", i.e. one that drops to 0 outside some radius, limiting ourselves to a finite number of particles and the kernel to a finite volume (e.g. a sphere of radius $2h$ in Fig. 2.1). A common choice is the cubic spline (Monaghan & Lattanzio 1985) with compact support inside $2h$, used by default in both PHANTOM and SPHNG. This kernel takes the form

$$W(\vec{r}, h) = \frac{\sigma(\nu_D)}{h^{\nu_D}} \begin{cases} 1 - 1.5q^2 + 0.75q^3 & 0 \leq q < 1 \\ 0.25(2 - q)^3 & 1 \leq q < 2 \\ 0 & 2 < q \end{cases} \tag{2.14}$$

where $q = |r|/h$, ν_D is the number of spatial dimensions and $\sigma = (2/3, 10/7\pi, 1/\pi)$ for 1, 2 and 3D, ensuring the correct normalisation (Equation 2.8). The number of terms can be increased to form quartic, $O(q^4)$ or quintic splines, $O(q^5)$, though at the expense of increased computation time. Plots of a selection of kernels and their derivatives are shown in Figure 2.2. Here we can see the commonly used cubic spline and its derivative has the same overall shape as other kernels, but the second derivative is discontinuous. Note that the second derivative is not used in this study, and comes into play when using velocity dependent forces (e.g. MHD). Now we have a suitable kernel we must find some appropriate value for the smoothing length, which we will address next.
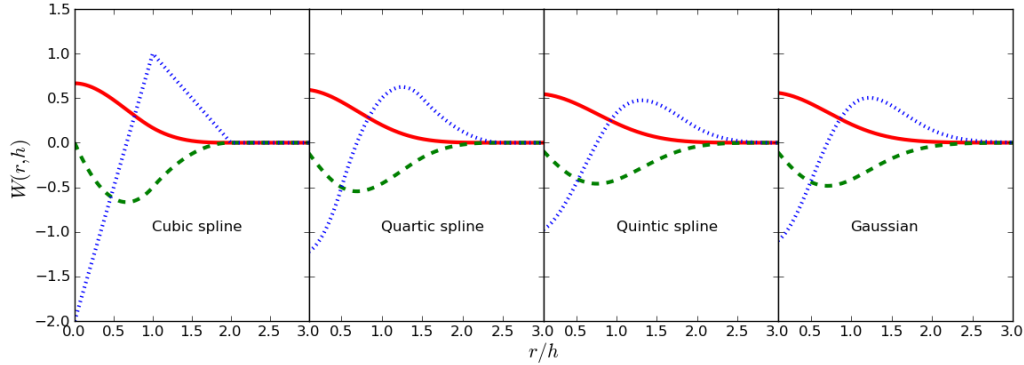
Figure 2.2: Various SPH kernels in red, their first and second spatial derivatives (dashed and dotted lines). The cubic spline is the simplest, but experiences discontinuities at higher derivatives, while the Gaussian is the most robust it has the undesirable feature of being non-zero in all space.

2.2.3 Mass equation of SPH

As we have shown above the ath particle density is calculated in SPH formalism by setting $A_a = \rho_a$ in the interpolation summation to give a density equation of the form

$$\rho_a = \sum_b m_b W_{ab} \tag{2.15}$$

where we will be using the shorthand notation of $\vec{r}_{ab} = \vec{r}_a - \vec{r}_b$ and $W_{ab} = W(|\vec{r}_{ab}|, h)$ throughout the remainder of this chapter. The density summation shows that the dimensions of the kernel are [1/distance^3] in 3D, hence the h^{ν_D} factor in Equation 2.14. The density of each particle is not constant, but rather the mass is (i.e. the "size" of the particles the variable).

Any density formulation we construct should inherently satisfy the continuity equation (Equation 2.5), ensuring that the rate of change of material in a system is equal to the rate it flows out of the surface. First addressing the left hand side of the continuity equation, in material derivative form, we obtain

$$\begin{aligned} \frac{D\rho_a}{Dt} &= \sum_b m_b \frac{D}{Dt} W_{ab} = \sum_b m_b \frac{\partial W_{ab}}{\partial r_{ab}} \frac{Dr_{ab}}{Dt} \\ &= \sum_b m_b \frac{\partial W_{ab}}{r_{ab}} \hat{e}_{ab} \cdot \vec{v}_{ab} = \sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab} \end{aligned} \tag{2.16}$$

where we have substituted for the particle velocity, $\vec{v}_{ab} = D\vec{r}_{ab}/Dt$. This could be used to evolve the density of the particles in the system, rather than evolving the integration of an additional quantity (as we will with v, and u). However we would be introducing additional errors from the adopted integration scheme and it is more practical to calculate the density via direct summation. This can be done in the same loop over particles as all the other summation formulae (discussed in the following subsections) at little additional cost. Now evaluating the right hand side of the

continuity equation we find that

$$\rho_a \vec{\nabla}_a \cdot \vec{v}_a = \vec{\nabla}_a \cdot (\rho_a \vec{v}_a) - \vec{v}_a \cdot \vec{\nabla}_a \rho_a$$

$$= \vec{\nabla}_a \cdot \left(\sum_b \frac{m_b}{\rho_b} (\rho_b \vec{v}_b) W_{ab} \right) - \vec{v}_a \cdot \vec{\nabla}_a \left(\sum_b \frac{m_b}{\rho_b} (\rho_b) W_{ab} \right). \tag{2.17}$$

Where we have used the SPH summation equation with the variables $\vec{A}_a = \rho_a \vec{v}_a$ and $A_a = \rho_a$. We can move the gradients inside the summations as they act only upon the properties of a. This gives (recalling that $a \cdot b = b \cdot a$)

$$\rho_a \vec{\nabla}_a \cdot \vec{v}_a = \sum_b m_b \vec{v}_b \cdot \vec{\nabla}_a W_{ab} - \sum_b m_b \vec{v}_a \cdot \vec{\nabla}_a W_{ab}$$

$$= - \sum_b m_b (\vec{v}_a - \vec{v}_b) \cdot \vec{\nabla}_a W_{ab} \equiv - \sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}. \tag{2.18}$$

The SPH continuity equation is then simply the sum of equations 2.16 and 2.18

$$\frac{D\rho_a}{Dt} + \rho_a \vec{\nabla}_a \cdot \vec{v}_a = \sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab} - \sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab} = 0 \tag{2.19}$$

thus satisfying the continuity equation. This is somewhat of a mute point however, as by construction SPH *should* conserve mass due to each particle having a fixed mass. The process of re-working the format of the SPH equation, i.e. "putting the density inside the operator", is a good way of ensuring symmetric functions (Monaghan 1992 refers to this as the second golden rule of SPH).

So far we have limited our discussion to particles with fixed smoothing lengths. If we were to give each particle its own individual smoothing length then we allow for an additional adaptability. This does however negate some approximations in the derivations above as we can no longer neglect the gradients in the smoothing lengths. An additional normalisation factor must be added, Ω_a, to the Equation 2.16 above

$$\frac{D\rho_a}{Dt} = \frac{1}{\Omega_a} \sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}(h_a). \tag{2.20}$$

the origin of which, and its affect on the standard SPH rate equations is the subject of Appendix A. For the purpose of this chapter we simply quote the resulting rate equation with the additional "grad-h" term where W_{ab} has been replaced by $W_{ab}(h_a)$. The smoothing gradient factor can be shown to be

$$\Omega_a = 1 - \frac{\partial h_a}{\partial \rho_a} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a} \tag{2.21}$$

which is also shown in Appendix A.

2.2.4 Momentum equation of SPH

The fluid mechanics momentum equation is summarised by the Navier-Stokes equations (or Euler equations with no viscous forces) given by Equation 2.4. Ignoring the viscosity and the external forces for now, the EoM of the system becomes[1]

$$\frac{D\vec{v}}{Dt} = -\frac{\vec{\nabla}P}{\rho}. \tag{2.22}$$

In order to find the rate of change of momentum we need find an expression for $\vec{\nabla}P/\rho$, the force for a purely inviscid fluid under no external forces. We will use the second golden rule of SPH again and put the density inside the operators. First we will re-arrange the pressure gradient to give

$$\frac{\vec{\nabla}_a P_a}{\rho_a} = \vec{\nabla}_a\left(\frac{P_a}{\rho_a}\right) + \frac{P_a}{\rho_a^2}\vec{\nabla}_a\rho_a \tag{2.23}$$

and substituting in the relevant SPH summation equations

$$\frac{\vec{\nabla}_a P_a}{\rho_a} = \vec{\nabla}_a\left(\sum_b \frac{m_b}{\rho_b}\left(\frac{P_b}{\rho_b}\right)W_{ab}\right) + \frac{P_a}{\rho_a^2}\vec{\nabla}_a\left(\sum_b \frac{m_b}{\rho_b}(\rho_b)W_{ab}\right) \tag{2.24}$$

which re-arranges to give:

$$\frac{\vec{\nabla}_a P_a}{\rho_a} = \sum_b m_b \frac{P_b}{\rho_b^2}\vec{\nabla}_a W_{ab} + \sum_b m_b \frac{P_a}{\rho_a^2}\vec{\nabla}_a W_{ab}. \tag{2.25}$$

The basic momentum equation is then simply

$$\frac{D\vec{v}_a}{Dt} = -\sum_b m_b\left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2}\right)\vec{\nabla}_a W_{ab}. \tag{2.26}$$

If we had not placed the density inside the operator then we would not have this antisymmetric form (noting that $\nabla_a W_{ab} = -\nabla_b W_{ba}$). This momentum equation fulfils the conservation of momentum, and Newton's third law, seen upon the swapping of indexes a and b and checking the forces between them are antisymmetric ($F_{ab} = -F_{ba}$). For individual smoothing lengths ("grad-h" formalism) we again just quote the solution here

$$\frac{D\vec{v}_a}{Dt} = -\sum_b m_b\left(\frac{P_a}{\Omega_a\rho_a^2}\vec{\nabla}_a W_{ab}(h_a) + \frac{P_b}{\Omega_b\rho_b^2}\vec{\nabla}_a W_{ab}(h_b)\right), \tag{2.27}$$

and direct the reader to Appendix A for a brief derivation of this form, and also evidence that the momentum equation can also be derived from the Lagrangian and the Euler-Lagrange equations rather than the Navier-Stokes equation.

---

[1]While we could simply use the discretisation equation and set $A_b = P_b$ to give $\rho_a D\vec{v}_a/Dt = -\sum_b m_b(P_b/\rho_b)\vec{\nabla}_a W_{ab}$, this does not conserve momentum. This is seen by calculating the forces from a on b, using the anti-symmetric identity $\nabla_a W_{ab} = -\nabla_b W_{ba}$ and seeing that $F_{ab} \neq -F_{ba}$, i.e. conflicting with Newton's third law (see Rosswog 2009).

2.2.5 Energy equation of SPH

The final basic SPH equation required for the evolution of a fluid system is the energy equation. We start from the first law of thermodynamics, $dU_a = P_a dV_a$. Working in per-unit-mass units we can substitute $dV_a = d(1/\rho_a) = 1/\rho_a^2 \times d\rho_a$. Taking these rates as a function of time and using D/Dt for Lagrangain dynamics we have

$$\frac{DU_a}{Dt} = \frac{P_a}{\rho_a^2}\frac{D\rho_a}{Dt}.$$ (2.28)

We then use the continuity equation to substitute the change in density, giving the energy equation

$$\frac{DU_a}{Dt} = \frac{P_a}{\rho_a^2}\sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}.$$ (2.29)

Monaghan (1992) uses the same method as for the above formulations to derive a similar expression instead involving P_b and ρ_b,

$$\frac{DU_a}{Dt} = \frac{1}{2}\sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2}\right)\vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}.$$ (2.30)

which conserves energy exactly. This form is less often used in practice however (e.g. Hubber et al. 2011, Rosswog & Price 2007) as it has the unfortunate side effect of producing negative energies if there are significant local pressure variations, caused by the P_b term in the energy calculation of a (Benz 1990; Wadsley et al. 2004). We can also include the effect of individual smoothing lengths (see Appendix A),

$$\frac{DU_a}{Dt} = \frac{1}{\Omega_a}\frac{P_a}{\rho_a^2}\sum_b m_b \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}(h_a)$$ (2.31)

which is equivalent to $P_a/\Omega_a\rho_a^2 \times D\rho_a/Dt$

2.2.6 The equation of state

In order to evaluate the SPH rate equations we need to set an equation of state to provide the pressure of the ISM particles as a function of density. Our primary interest is the abundance of molecular material which, as will become clear in the following sections, is a strong function of temperature. As such we must evolve the thermal energy of the particles alongside the kinematic quantities. We utilise an adiabatic equation of state, which gives the pressure as a function of internal energy and density of the particles

$$P = (\gamma - 1)\rho u$$ (2.32)

where γ is the adiabatic index, given by the ratio of specific heats at constant pressure and volume, which for a monatomic gas is $5/3 \approx 1.67$, which is the case for the majority of the ISM (but is $5/7$ for diatomic molecules such as CO and H_2). If we were not storing thermal energy we could use a

general polytropic form without thermal energy, $P = K\rho^\gamma$, where K is some constant, though this would result in a phase diagram that is a poor representation of the ISM[2], where complex heating and cooling mechanisms determine the evolution of u (see Section 2.3.1 and Figure 2.12). We also use an isothermal EoS for testing purposes, where $P = K\rho$.

When we are discussing artificial dissipation in a later section we will be referring to the adiabatic sound speed of the gas. This is calculated as

$$c_s = \sqrt{\frac{\partial P}{\partial \rho}} = \sqrt{\frac{\gamma P}{\rho}} = \sqrt{\frac{k_B T}{m_p \mu}} \tag{2.33}$$

where μ is the mean molecular weight of the material in question (see Section 2.3 for a brief description in relation to the ISM). We can then also calculate the thermal temperature from $T = P\mu/R\rho$, i.e. $T = \mu u(\gamma - 1)/R$, where R is the gas constant.

2.2.7 The density and smoothing length in "grad-h" SPH

As a rule of thumb we want the kernel to contract in regions of high density and keep the number of neighbours per particle approximately constant, making sure the resolution is consistent between dense and diffuse regions of particles. The particles in the denser regions require smaller smoothing lengths compared to those in the more diffuse medium, maintaining the same number of neighbours (6 in the case of Fig. 2.1). If fixed smoothing lengths are used then some sacrifice would need to be made between either over-resolving the very diffuse media, which are usually of minimal importance, or under-resolving the high density regions which is often the location of the more complex and interesting physics.

A natural choice would be to choose a smoothing length that is analogous to the length scale which defines the density of a SPH particle, $\rho \propto m/h^{\nu_D}$, where ν_D is the number of spatial dimensions (Gingold & Monaghan 1982; Price 2012a). This gives a simple equation relating the density and smoothing length of each individual particle

$$h_a = \eta \left(\frac{m_a}{\rho_a} \right)^{1/\nu_D}. \tag{2.34}$$

The η factor is chosen to roughly give a number of neighbours, and specifies the smoothing length scale, which can be calculated by

$$N_{neigh} = \frac{4}{3}\pi \left(\zeta \eta \right)^3 \tag{2.35}$$

in 3D[3] where ζ is the compact support of the kernel, which is $2h$ for the cubic spline (i.e. the finite radius of the smoothing sphere). We could define the number of neighbours explicitly rather than η, but there are numerous pit-falls when constraining the N_{neigh} factor throughout a simulation (see Price 2012a for a discussion). Good values of η have found to be between 1.2-1.5 (Rosswog & Price 2007), and we adopt $\eta = 1.2$ in all simulations presented here corresponding to $N_{neigh} = 58$

---

[2]This would correspond to an isentropic flow, applicable in the absence of shocks or external energy sources (Springel 2010b).

[3]2D and 1D forms are similarly given by $\pi(\zeta\eta)^2$ and $2\zeta\eta$ respectively (Price 2012a) .

in 3D.

As we have now defined h in terms of ρ, and ρ in terms of h we can use these relations to iteratively solve for the smoothing lengths by ensuring that Equation 2.15 and Equation 2.34 are equivalent, i.e. $\rho_{\text{sum}} = \rho(h_a)$ to some tolerance. To do so it is common to use some solver to find solutions (i.e. minimising the function) of the non-linear equation

$$f(h_a) = \rho(h_a) - \rho_{\text{sum}} \tag{2.36}$$

which can be done by the Newton-Raphson method (Gingold & Monaghan 1982; Price 2012a), which requires the derivative of the density difference function, $f'(h_a)$,

$$
\begin{aligned}
\frac{\partial f(h_a)}{\partial h_a} &= \frac{\partial \rho(h_a)}{\partial h_a} - \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a} \\
&= \frac{\partial \rho_a}{\partial h_a} \left[1 - \frac{\partial h_a}{\partial \rho_a} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a} \right] = \frac{\partial \rho_a}{\partial h_a} \Omega_a
\end{aligned}
\tag{2.37}
$$

where $\partial \rho_a / \partial h_a$ can be easily computed from the analytic expression for ρ_a. This results in the iteration equation

$$h_{a,n+1} = h_{a,n} \left(1 + \frac{\rho(h_{a,n}) - \rho_{sum}}{3\Omega_a \rho(h_{a,n})} \right) \tag{2.38}$$

which can be iterated until some convergence is reached. We have used the definition of the smoothing gradient parameter from Equation 2.21, the calculation of which is already done in the force and energy rate equations, reducing the overall computational cost of the density calculation. The cost can be reduced further by making a sensible initial estimate for the smoothing length iterations using a predictor step for h and $dh_i/dt = dh_i/d\rho_i \times d\rho_i/dt = -h_i/3\rho_i \times d\rho_i/dt$ (Price & Monaghan 2007).

2.2.8 Evolving the system

Once the forces have been calculated from the SPH summation equations, giving $\dot{v}_{x,y,z}(t), \dot{u}(t)$, they are used to evaluate new positions (x, y, z), velocities (x, y, z), and energies, as well as updating h if required. We effectively now need to solve a set of ordinary differential equations in order to find the change in position and velocity from $\dot{v}$ and $\dot{u}$.

Each of the codes we utilise uses a different integrator. In PHANTOM a second order "leapfrog" is used, with a specific "kick-drift-kick" formulation (Springel 2005; Monaghan 2005). The initial force "kick" is the applied to the velocity for half the initial timestep, and the particle is allowed to "drift" at this speed for the full timestep. The force is then re-evaluated at the post-drift position and is used to provide an additional velocity kick to update the velocity to the full timestep. This integrator has been shown to be stable considering its low order, displaying greater integration stability and conservation properties over its "drift-kick-drift" and even higher order Runge-Kutta contemporaries (Springel 2005; Rosswog 2009).

SPHNG uses the Runge-Kutta-Fehlberg integrator, specifically the RKF1(2) integrator, which is in essence a second order method imbedded within a first order method (Fehlberg 1985, see

Cossins 2010 for a discussion). The construction of a first and second order estimate for the updated variables means we can control the error in the integration explicitly, though we would have to use the first order estimate in the actual integration. This issue can be mitigated by simply enforcing a very small tolerance on the first-second order difference, effectively making RKF1(2) a second order method with controllable accuracy.

2.2.9 Timestepping

Regardless of the actual integration scheme, we need some way of sensibly deciding the timestep on which to advance the properties of the system. In the most basic case a timestep can be imposed that is the time between the creation of dump files, Δt_0 (e.g. every 1/100th of the total simulation run-time). However, this can easily be much larger than the dynamical time-scale of changes in system variables, resulting in integration steps that do not correctly encompass the physics of the system. The size of a timestep can then be chosen by use of the Courant condition which relates the ratio of the spatial resolution Δx to the time resolution Δt and the velocity of the simulated particles by $C_{CFL} = v\Delta t/\Delta x$, which ensures information speed does not exceed the physical speed of material in the simulation (Courant et al. 1928). C_{CFL} is the dimensionless Courant number, and has been constrained from numerical studies to be from 0.25-1 in order to satisfy convergence of the time integration. We take a minimum of this timing criteria over all particles in the simulation to obtain the global Courant timestep as

$$\Delta t_{CFL} = C_{CFL} \min \left(\frac{h}{|v_{sig}|} \right) \tag{2.39}$$

where v_{sig} is the same signal velocity as that defined in Section 2.2.11. We use $C_{CFL} = 0.3$ in the simulations in this thesis. We also include a forcing timestep, which is calculated from the ratio of the magnitude of the forces to the smoothing length

$$\Delta t_f = C_f \min \sqrt{\frac{h}{|\vec{f}|}} \tag{2.40}$$

where we use $C_f = 0.25$. Finally there is a time scale for astrophysical cooling, which will be discussed later,

$$\Delta t_{cool} = C_{cool} \min \left| \frac{U}{\dot{U}_{ISM}} \right| \tag{2.41}$$

where $C_{cool} = 0.3$ as in Glover & Mac Low (2007). This is normally grouped with the forcing timestep condition. The hydrodynamical time is dictated by the Courant timestep, whereas the forcing and cooling times are used to advance the timestep with the same initial SPH force, i.e. external and cooling forces are subcycled inside the main steps which is where neighbour calculations are required.

Individual timesteps

In astrophysical simulations time scales of importance can be very large when assessing gravitational effects, or small when resolving molecular/shock effects. It then becomes prudent to allow each particle its own individual timestep to avoid evolving all particles in the calculation on integration times of the slowest particles. In these cases particles are binned up into groups of $2^n \Delta t_{min}$ and each time bin is evolved separately (Rosswog 2009; Hubber et al. 2011). For example these timesteps can be of similar form to those above, but the actual evolution of the particles is done in bins of particles with similar magnitude timesteps, making sure the timesteps of all particles in the simulation coincide at some point, i.e. they should all be synchronisable at any given point in the calculation (Hernquist & Katz 1989). Care must obviously be take to ensure that a particle is correctly aligned with the others in the simulation before it is moved to a different timestepping bin as the individual timestep increases or decreases (Hubber et al. 2011).

2.2.10 Neighbour finding

Now we have a good grounding in the numerical recipes behind SPH, and a method of advancing the particle properties, all that remains is some way of knowing the relevant particles in the SPH summation equations, the "nearest neighbours" (i.e. the relevant b's for which to calculate the properties of a). The most obvious way to do so would be to simply loop over all particles, but this would be a $O(N^2)$ process, a very computationally expensive scaling. By using a kernel of compact support we have limited ourselves to a small number of neighbours (usually in the range 50-100) but as the particles move around the identity of these neighbours will change, so some method of re-populating these neighbour-lists is required. Two such options that are used frequently in SPH codes are "Link lists" and "Tree codes", both offering an improvement on the basic $O(N^2)$ neighbour search (Liu & Liu 2003).

Link lists

A simple neighbour searching scheme is to use a link list method. All SPH particles are binned into a grid that covers the entire computational domain, where grid cells have a size of $2h$ when using the cubic spline. Particles in a cell only then search their own and adjacent cells for neighbours, resulting in a much smaller search domain, only 27 cells in 3D (Domínguez et al. 2011). Particles are also chained to each other via some link array allowing for fast looping through all particles in each cell (Liu & Liu 2003).

This method can be of order $O(N)$ in some cases, but can be less efficient if there are large fluctuations in smoothing lengths, as it could result in large numbers of particles in each of the grid cells increasing the time required to walk through the neighbouring cells (Hernquist & Katz 1989). Link lists are used in PHANTOM where the cell width is set to $2 \sum_a h_a / N$ and we utilise a cylindrical cell structure for our galactic disc simulations.

Trees

The other common neighbour finding method is to use a hierarchical tree-like structure, such as that of Barnes & Hut (1986). The basic method involves building up an octree of the simulation domain (i.e. first splitting the simulation into a 2x2x2 octal) which is then subdivided into further cells if they contain any particles. This is done until the child cells at the bottom of the tree contain 1 or 0 particles (the "leaves" of the tree, where the tree is actually an inverted tree). For nearest neighbour calculations, such as finding which particles are required for the SPH summations, a search is done by moving back up the tree from the particle of interest to find which particle nodes (a point where branches separate) are within the region of interest. For instance, whether a neighbouring particle is within the kernels compact support radius, $r \le 2h$ when using a cubic spline. The neighbour search then descends down the branches of that node, checking whether the leaf particles are within the compact support region, thus building a neighbour list for a specific particle while retaining the tree for use with the other particles.

The use of a tree has the added advantage of being able to be used for the neighbour finding and gravitational force calculation (Hernquist & Katz 1989), and is used for calculation of *N*-body gravitational forces in SPHNG. When the calculating the gravitational forces the tree is walked through, instead of only using the nearest neighbours the gravitational calculation must also take into effect the material outside the kernel support radius. As the tree is traversed a distance criterion is calculated on each node. If ratio of the size of the node to the distance from the particle of interest is greater than some tolerance then the branches of the node are then traversed. However, if less than this tolerance then the material encompassed by the node is lumped together to form a single larger body from which the gravitational attraction is calculated using the appropriate centre of mass and node mass which is stored when the tree is created (Hernquist & Katz 1989). The gravitational forces must also be softened to avoid extreme accelerations of particles during close-encounters; this is soften on a scale of the smoothing kernel (Bate & Burkert 1997). The tree in SPHNG is described in greater detail in Benz (1988) and Benz et al. (1990), which was implemented using the version from Press (1986). Tree based neighbour finding is usually an order $O(N \log N)$ process and can be pushed down to $O(N)$ by clever symmetrising and storing of cell-cell forces (Hernquist & Katz 1989; Dehnen 2002).

2.2.11   Artificial dissipation

Artificial viscosity

One of the major advantages of SPH is that it is completely dissipation free by construction. This can have adverse properties when attempting to model shock-based problems. On large scales the rapid change in system properties (be it density, velocity, pressure, internal energy etc.) appear as discontinuities in the fluid flow. While in reality this discontinuity is in fact smooth it would require a simulation resolution on macroscopically small scales to correctly capture. The smooth gradients inherent in the SPH formulation do a poor job in such regions where steep gradients are effectively discontinuities on macroscopic scales.

Von Neumann & Richtmyer (1950) suggested that this problem could be over-come by the addition of some artificial dissipation, or viscosity. The addition of an artificial viscosity (AV) smoothes out gradients near these discontinuities, acting similarly to physical viscosities. Such an artificial viscosity will dissipate kinetic energy into heat, generating the required entropy increase, and will broaden the shock to a resolvable scale rather than a discontinuous one, allowing for shock capturing. The AV does not in essence reproduce a physical process but rather smooth out a shock front that would otherwise be discontinuous on the length scales relevant to the calculation.

Von Neumann & Richtmyer (1950) formulate this viscosity into a pressure of the form $P_\nu = -A\rho c_s l(\vec{\nabla}.\vec{v}) + B\rho l^2(\vec{\nabla}.\vec{v})^2$ for some constants A and B over some resolvable length scale l (h in SPH Rosswog 2009). The A term is analogous to a classical bulk viscosity (as is α described shortly) and B the Neumann-Richtmyer quadratic term (as is β), and both only activate in converging flows ($\nabla . v < 0$). Using AV's of this form had drawbacks however, resulting in either over-dissipation or post-shock oscillations (Monaghan & Gingold 1983). This approach was refined by Monaghan & Gingold (1983) who suggested additional pressure term (actually of dimensions P/ρ^2) to be included in the standard momentum equation, Equation 2.26, of the form

$$\Pi_{ab} = \frac{-\alpha\bar{c}_{ab}\mu_{ab} + \beta\mu_{ab}^2}{\bar{\rho}_{ab}} \tag{2.42}$$

where

$$\mu_{ab} = \frac{h_{ab}\vec{v}_{ab} . \vec{r}_{ab}}{r_{ab}^2 + \epsilon\bar{h}_{ab}^2} \tag{2.43}$$

where α, β, ϵ are constants to be set. The over-lined terms indicate averaged quantities between two particles; $\bar{\rho}_{ab} = [\rho_a + \rho_b]/2$ and $\bar{h}_{ab} = [h_a + h_b]/2$ where the velocity term is simply $\vec{v}_{ab} = \vec{v}_a - \vec{v}_b$. Then $\vec{r}_{ab} . \vec{v}_{ab}$ will be < 0 if particles are approaching, i.e. when a shock may occur, and is used as a criterion for activating the viscosity (i.e. $\Pi_{ab} = 0$ if $\vec{r}_{ab} . \vec{v}_{ab} > 0$). This form is effectively a combination of Neumann-Richtmyer (β) and bulk (α) viscous terms. The β term was not originally present in the Monaghan & Gingold (1983) formalism but was added to prevent particle penetration and correctly model strong shocks (Monaghan 1992). The standard values adopted are $\alpha = 1, \beta = 2\alpha$ (Monaghan 1992), usually providing adequate dissipation at a shock while $\epsilon = 0.01$ protects against small separation singularities (as $r_{ab} \rightarrow 0$). There is also the capacity to set individual values of α to the particles, to avoid dissipation where it may not be needed such as converging flows in the absence of shocks.

This results in a full momentum equation of the form

$$\frac{D\vec{v}_a}{Dt} = -\sum_b m_b\left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2} + \Pi_{ab}\right)\vec{\nabla}_a W_{ab}, \tag{2.44}$$

where the artificial viscous force produces a repulsive force when particles move towards each other. The dissipated kinetic energy must be re-assigned as thermal energy to satisfy energy conservation. To this end the standard SPH energy equation gains the contribution

$$\frac{DU_{a,AV}}{Dt} = \frac{1}{2}\sum_b m_b\Pi_{ab}\vec{v}_{ab} . \vec{\nabla}_a W_{ab} \tag{2.45}$$
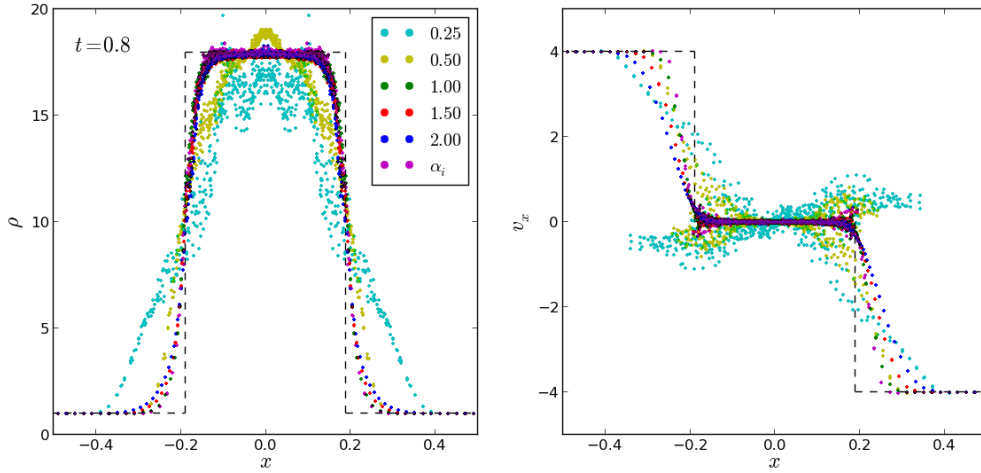
Figure 2.3: An isothermal colliding flows test using PHANTOM in 3D where gas flows are moving at $v = \pm 4$ initially. The density and velocity profiles are shown after 0.8 dimensionless time units. Differences in shock capturing can be with different AV parameters, shown in different colours. The β term is fixed to 2α when α is fixed, and 2 when variable (pink points). The analytical solutions are shown by dashed lines.

giving a full energy equation

$$\frac{DU_a}{Dt} = \sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{\Pi_{ab}}{2} \right) \vec{v}_{ab} \cdot \vec{\nabla}_a W_{ab}. \tag{2.46}$$

This is the "classical" AV formulation, and is the standard in the SPHNG code (with additional "grad-h" terms). The α term can actually be related directly to the coefficients of physical viscosity[4]. The shear viscosity parameter can be shown to be equivalent to $\eta = \alpha h c_s / 10$ and bulk parameter to $\zeta = 5\eta/3$ (Monaghan 2005; Lodato & Price 2010; Price 2012b).

The addition of AV is illustrated by the colliding flows test shown in Fig. 2.3. Here we have set up an isothermal shock tube test in 3D in a box of dimensions $8 \times 1 \times 1$ using approximately 160000 particles initially arranged on a hexagonal lattice with velocities $v = \pm 4$ with flows directed towards each other either side of $x = 0$. While better analysis of AV forces is seen in 1D, neither of the codes we use in this thesis have the capacity for anything but 3D. This results in a strong shock propagating from $x = 0$ manifesting as a plateau in ρ and v_x, shown by the analytic solution (dashed line). The different coloured points refer to different values for the AV coefficients and $\beta = 2\alpha$ for all tests apart from the pink points where α is variable and we have fixed $\beta = 2$. It is clear that $\alpha = 1$ to 2 do a good job at capturing the shock in ρ and v_x whereas lower values do a poorer job. The $\alpha = 2$ case applies slightly too much dissipation, which can be clearly seen in the pre-shock regions of the v_x panel. A test was also performed where individual values of α are assigned to each particle which allow for dissipation only where necessary utilising the switch of Morris & Monaghan (1997) that is a function of sound speed and smoothing length of the particles

---

[4]The shear viscosity develops when a fluid flow when it passes some boundary moving at a different velocity, know as a shear flow. Bulk viscosity on the other hand is the manifest as the viscous friction experienced by a fluid expansion or compression in the absence of shear flow. The η and ζ terms enter into the viscous Euler equations, quantifying shear and bulk viscosity respectively.

(Lodato & Price 2010). This differs little compared to the fixed case, though it does apply the least dissipation in the region not yet experiencing a shock (v_x panel). This highlights that a variable α can be useful in that it applies less dissipation where it is not needed.

An alternative formulation of AV was proposed by Monaghan (1997), where the authors use solutions analogous to Riemann solvers[5]. The original purpose of which was to enable the capture of relativistic shocks. We present this form of AV in a similar form to the classical $\alpha - \beta$ version using the parameter Q_{ab} with the same dimensions as Π_{ab} such that the AV contribution to the momentum equation is given by

$$\frac{D\vec{v}_{a,AV}}{Dt} = -\sum_b m_b Q_{ab} \nabla_a \overline{W}_{ab} \tag{2.47}$$

where Q_{ab} is given by

$$Q_{ab} = \frac{\alpha^{AV} v_{sig} |\vec{v}_{ab} \cdot \hat{r}_{ab}|}{\overline{\rho}_{ab}} \tag{2.48}$$

when $\vec{v}_{ab} \cdot \hat{r}_{ab} \leq 0$ and 0 otherwise to ensure converging flows. We use the shorthand notation of $\overline{W}_{ab} = [W(r_{ab}, h_a) + W(r_{ab}, h_b)]/2$ to average kernel weights, and can include additional Ω denominators for the kernels to correct for variable smoothing lengths. This formulation contains a signal velocity term, v_{sig}, which is the averaged signal speed between two particles. The exact choice of signal velocity differs between users, but a general form is given in Price (2012a) as

$$v_{sig} = \frac{1}{2}\left[c_{s,a} + c_{s,b} - \beta^{AV} \vec{v}_{ab} \cdot \hat{r}_{ab}\right] \tag{2.49}$$

for $\vec{v}_{ab} \cdot \hat{r}_{ab} \leq 0$, and is 0 otherwise, where c_s is sound speed of a given particle. While some forms in the literature fix $\beta^{AV} = 1$ (Rosswog & Price 2007; Hubber et al. 2011) or some other value (Springel 2010b) the general form is that of a sound speed term in combination with a velocity term. The formalism is similar to the previous notation, with linear and quadratic terms in velocity projection ($\vec{v}_{ab} \cdot \hat{r}_{ab}$) scaled by α^{AV} and β^{AV} terms respectively. This can then be similarly applied to the thermal energy equation to give

$$\frac{DU_{a,AV}}{Dt} = \frac{1}{2}\sum_b m_b Q_{ab} \vec{v}_{ab} \cdot \vec{\nabla}_a \overline{W}_{ab} \tag{2.50}$$

which can be added to the standard energy equation.

Artificial thermal conductivity

It has been shown that in some cases the AV alone is not enough to resolve certain physical phenomena. There is a need to add additional dissipation terms for each variable of the system. In adding AV we have corrected for discontinuities in the velocity distribution, but it is also necessary to add additional dissipative energy terms. This problem was noted by Agertz et al. (2007) whose calculations of Kelvin-Helmholtz instabilities showed a lack of characteristic rolling fea-

---

[5]A Riemann solver is (unsurprisingly) designed to solve Riemann problems which include shock dynamics and other discontinuities coupled to the Euler equations (e.g. Toro 1992).
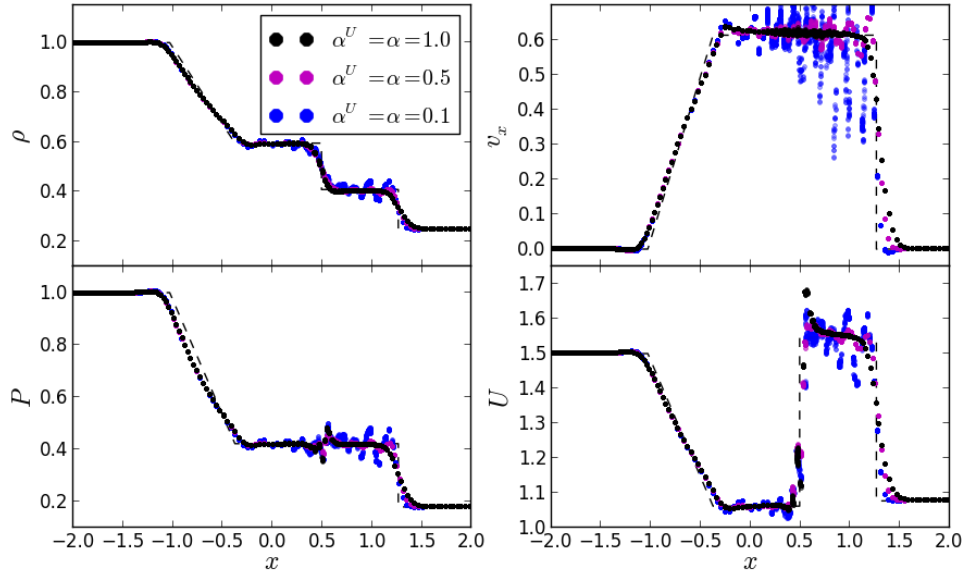
Figure 2.4: Adiabatic "sod" shock tube problem test in 3D using PHANTOM with differing values of α, β and α^u, indicated by different colours. The fiducial values are shown in black. Analytical solutions to the density, velocity, energy and pressure over-plotted as dashed lines. Low values of α and β result in a failure to capture many aspects of the shock structure and sinusoidal oscillations
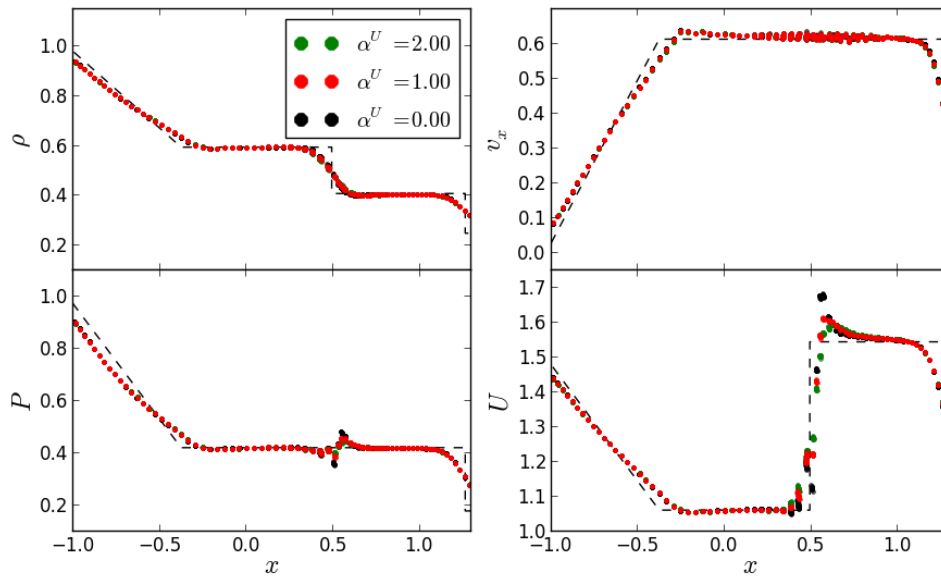

Figure 2.5: Same as Fig. 2.4 but keeping α and β fixed to 1.0 and 2.0 respectively while only varying α^U. The main effect of α^U is to reduce the energy (and consequentially P) spike at $x = 0.5$ in the figure. α^U is fixed to 1 when used. The x-axis shows a narrower range than Fig. 2.4 as the points are indistinguishable outside of the range shown.
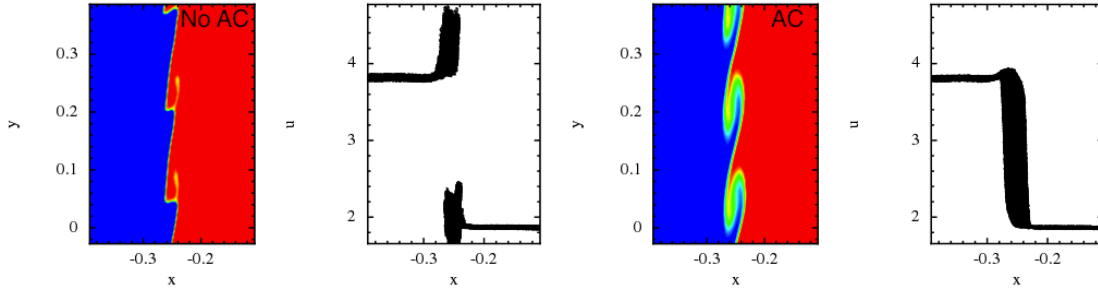
Figure 2.6: Kelvin-Helmholtz instability test without and with artificial thermal dissipation (left and right panel pairs respectively). The lack of heat dissipation across the density boundary is clearly seen in the run without artificial conductivity, which also manifests as a strong discontinuity in internal energy of the particles. With the addition of artificial thermal conductivity the fluid dissipates energy smoothly across the contact region, and displays a smooth energy transition. The test was set up using an initial 2:1 density contrast and 3.8 million particles in 3D using PHANTOMwith α^u fixed to 1 when used.

tures at the high-low density contact region. Price (2008) proposed an additional artificial thermal conductivity, AC, to address this problem, where the energy equation is augmented by

$$\frac{DU_{a,AC}}{Dt} = -\sum_b m_b \frac{\alpha^U v_{sig}^U u_{ab}}{\overline{\rho}_{ab}} \hat{r}_{ab} \cdot \vec{\nabla}_a \overline{W}_{ab} \qquad (2.51)$$

where α^U is the dimensionless artificial thermal conductivity parameter that smoothes out gradients in internal energy between particles, and $u_{ab} = u_a - u_b$ (see also Valcke et al. 2010). There is an additional signal velocity which can be different to that used in standard artificial viscosity. Two suggestions in the literature are either a pressure difference, $v_{sig}^U = \sqrt{|P_a - P_b|/\hat{\rho}_{ab}}$ from Price (2008) or velocity projection, $v_{sig}^U = |\vec{v}_{ab} \cdot \hat{r}_{ab}|$ from Wadsley et al. (2008). The velocity projection form is used in PHANTOM.

Illustrations of the importance of AC are shown in Figures 2.4, 2.5 and 2.6. In Figures 2.4 and 2.5 we show an adiabatic/Sod shock tube test, where gas is initially stationary but with large discontinuities in the density and energy distributions on either side of $x = 0$. The gas is then allowed to expand, causing a shock wave to propagate through the tube. Once again this test is in 3D, and the analytic solutions for ρ, v_x, P and U are shown by the dashed lines. $\beta=2\alpha$ for all the tests shown. Figure 2.4 shows that the change in the standard α and β viscosity parameters result in poor capturing of the shock, magenta and blue points, as in Figure 2.3.

In Figure 2.5 the effect of AC is included, where the black points are the same as those in the previous figure and represent $\alpha^U = 0$. Increasing α^U to 1 or 2 has a much more subtle effect than α, but can be seen to reduce the "blip" in U and P at the contact discontinuity. The points for $\alpha^U = 2$ are beginning to over-smooth the shock, and if increased further will severely dissipate the shock front discontinuity.

The importance of AC is much more evident from the Kelvin-Helmholtz test shown in Figure 2.6, where the gas is set in two parallel streams of material with a 2:1 density contrast. The left panels show the test with no AC, and the right with AC added and $\alpha^U = 1$. In the test with
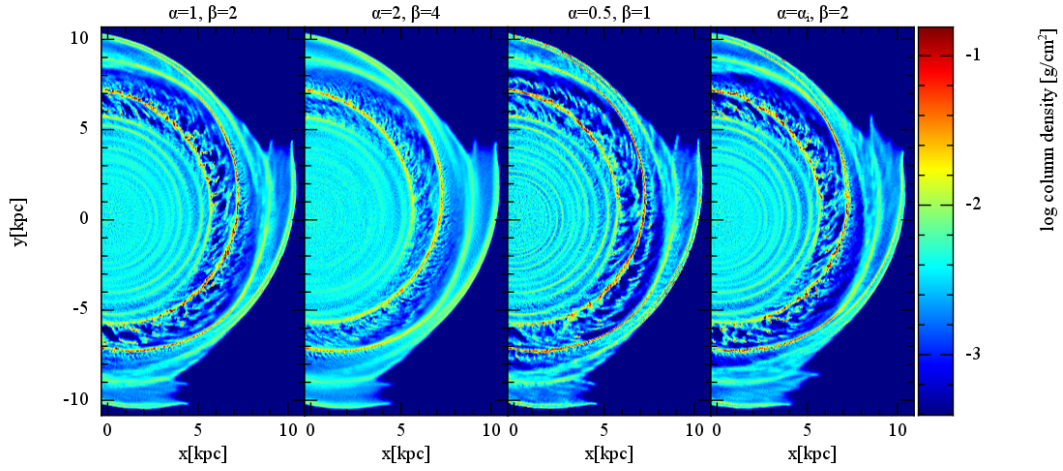
Figure 2.7: Top-down column densities of disc galaxies after 330Myrs of evolution with different AV parameters. A simulation with the fiducial values of $\alpha = 1$ and $\beta = 2$ is shown in the first panel, with double and half these values in the second and third panels. The fourth panel shows a simulation with $\beta = 2$ and individual α parameters for each particle in the range $0 < \alpha_i < 1$.

active AC the thermal energy smoothly traverses the contact discontinuity. The run with no AC keeps the different flows segregated throughout the test, with a clear discontinuity in the thermal energy.

The impact of different artificial viscosity parameters in our simulations is shown in Figures 2.7 and 2.8. Here we show a simple galactic disc simulation with a four armed spiral pattern (details of which will be discussed in Chapter 3) after 330Myrs of evolution. The four different simulations show our standard parameters, $\alpha = 1$, $\beta = 2$, double and half these values, and a simulation where each particle has its own individual α_i. The top-down column density plot in Fig. 2.7 shows some differences with AV parameters. It appears that increasing the strength of the AV causes the dilution of inter-arm structures, and decreasing the AV enhances them. The calculation with variable viscosity appears similar to the standard values. Figure 2.8 clarifies this difference somewhat by showing the temperature profile of the gas as a function of density. Here we see that the calculations with higher or lower than the standard AV strength do not produce the coldest regions present when using the fiducial values. The variable viscosity calculation has a temperature distribution near identical to that of the standard values, though this could be a result of both having $\beta = 2$. The main conclusion to draw from these figures is that there is not a large difference in morphology and thermal properties when using the standard, weaker, or adaptive AV parameters but using stronger than average values can cause artificial smoothing of morphological features and a less populated cold phase ISM.

Similar tests to those shown for AV were performed for the α^U parameter, with values 0, 0.5, 1 and 2. Both the top-down density distributions and the thermal properties of the gas were effectively the same for the different values of α^U. The only minor difference was a reduced population of the hottest ISM gas with the highest strength AC, $\alpha^U = 2$. The gas in the simulations presented here is unlikely to have strong thermal discontinuities as present in the Kelvin-Helmoltz
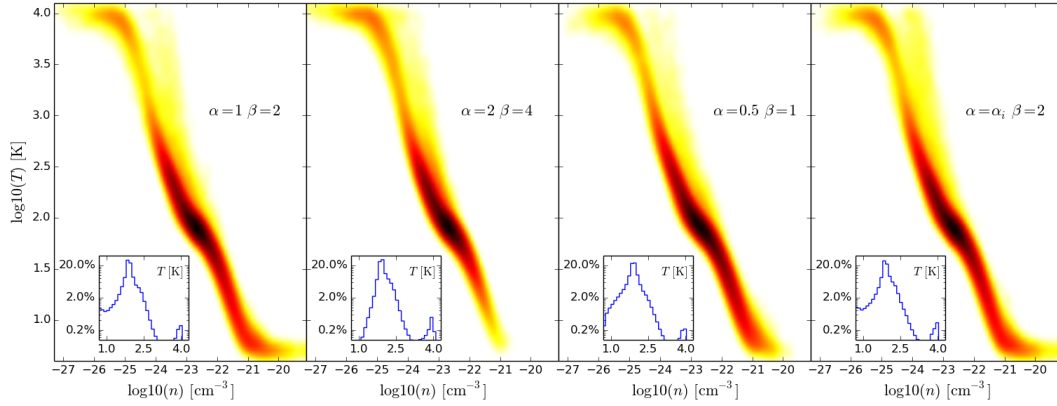
Figure 2.8: Temperature-density distribution from the same four simulations shown in Figure 2.7. The small insert shows a histogram of T for each case. The standard AV parameters allow for the development of cold ISM component, where CO is easily maintained. The impact of individual α coefficients is marginal.

test, as any dramatic thermal change should be smoothed out by astrophysical cooling and strongly coupled to density gradients which should be smoothed by standard AV. If there were some strong thermal source that caused large temperature gradients in the gas then the choice of α^U could become more important. This could play a role in calculations including supernovae feedback, where the resulting instantaneous thermal shockwave could produce strong thermal discontinuities.

The standard literature values of $\alpha = 1$, $\beta = 2$ and $\alpha^U = 1$ appear suitable for capturing the shock fronts in the tests shown here, to the capabilities of SPH. Lower values of α and β result in a dramatic inability to reproduce shock features in thermodynamical quantities, as shown by the colliding flows and sod tube tests ins Figures 2.3 and 2.4. Stronger values can be seen to be beginning to add too much dissipation to the shocked front, evident from Figures 2.7 and 2.8. α^U had a much weaker effect, but is seen to reduce artefacts at the contact region in the sod tube test (Fig. 2.5). We therefore adopt $\alpha = 1$, $\beta = 2$ and $\alpha^U = 1$ throughout the remainder of this thesis. Some improvement can be found in using individual values of α to avoid unwanted dissipation, but the results shown in Figures 2.7 and 2.8 shows the effect of this is unnoticeable for the scales investigated here.

2.2.12 SPH code specifics

Here we provide a brief description of the SPH codes used in this thesis. The main differences/similarities between the two codes are summarised in Table 2.1.

PHANTOM

The PHANTOM code is built specifically for non-gravitating problems, at high resolutions with a low-memory footprint, and has been extensively adapted for MHD computations (Price & Federrath 2010; Lodato & Price 2010; Tricco & Price 2012). The particles have individual smoothing lengths and timesteps, the system is evolved using a leapfrog integrator, and neighbours found

Characteristic	PHANTOM	SPHNG
Neighbour finding	Link list	Binary tree
Time-step integrator	2nd order leapfrog	2nd order Runge-Kutta-Fehlberg
Gravity	N/A	Binary tree (or Grape board)
Dissipation	Adapted Monaghan 97 + Conductivity	Standard α-β
Kernel	Cubic spline	Cubic spline
N_{neigh} (3D)	58	58
Density evolution	Newton-Raphson iteration	Newton-Raphson iteration

Table 2.1: Summary of the main aspects of the two SPH codes used in this thesis.

using linked-lists for reducing computational effort with OpenMP and MPI parallelisation. Due to a unique arrangement of the SPH equations, and separation of the average density terms in Equation 2.47, the density sums need only be calculated once on each timestep. This speed comes at a cost, with no current implementation of particle self-gravity (which would require a much slower tree-like neighbour finding algorithm).

ISM chemistry, cooling and simple galactic potentials were only recently incorporated into PHANTOM (Dobbs 2011a), but have not been well tested. Here we extensively tested the code, leading to corrections in the particle timestepping. The main improvement was the inclusion of a suite of gravitational potentials for use with galactic disc simulations, and are discussed in Chapter 3.

SPHNG

The second SPH code utilised in this thesis is SPHNG ("SPH-Next-Generation"). An older code than PHANTOM, SPHNG is based on the original version of Benz et al. (1990) and has been substantially modified since its creation. Notable improvements include the addition of accreting sink particles (Bate et al. 1995), magnetic fields (Price & Monaghan 2004) and radiative transfer (Whitehouse et al. 2005), as well as the standard variable smoothing lengths, individual timesteps and parallelisation using both OpenMP and MPI. Most importantly is that SPHNG allows for the calculation of gravitational forces. This allows for inclusion of point mass gravitating particles that are used to represent the Galactic stellar component, the subject of Chapter 5.

The code has already been used extensively for galactic scale ISM modelling. These studies include the effect of ISM cooling and chemistry (Dobbs et al. 2008), self-gravity (Dobbs 2008) and stellar feedback (Dobbs et al. 2011).

Brief comparison between codes

Figure 2.9 shows simulations of simple, low-resolution galactic discs using the SPHNG (top: blue) and PHANTOM (bottom: red) codes. These snapshots are at 472 Myrs of evolution[6] and show the top down particle distribution, the temperature-density profile and the CO abundance (the calculation

---

[6]The evolution times frequently used in this thesis of 236, 354 and 472 Myrs correspond to 5, 7.5 and 10 code units, determined from the gravitational constant, the distance and mass code units adopted; $u_m = 1 \times 10^5 M_\odot$ and $u_d = 100$ pc, giving a time unit of $u_t = \sqrt{u_d^3/Gu_m} = 47.2$ Myrs.
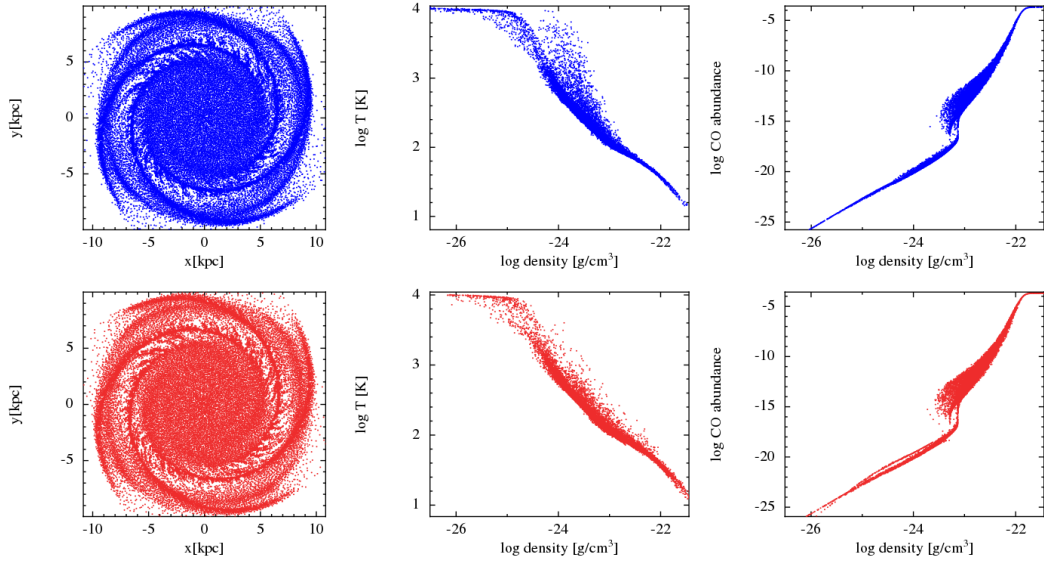
Figure 2.9: Simulations of gas in a disc galaxy constituting of 40000 particles with a total mass of $4 \times 10^9 M_\odot$. The top row is from a simulation in PHANTOM, and the bottom in SPHNG, both after 472 Myrs of evolution. The position, T and CO abundance of the SPH particles trace the same global flow in codes, with minor differences resulting from inherent code differences listed in Section 2.2.12. The chemical network are described in detail in Section 2.3.

of which is discussed in the next section). The simulation is of only 40000 particles, embedded in a simple disc potential constructed to reproduce the flat Galactic rotation curve including a a four armed spiral potential.

Comparing global quantities of the simulations from each code we see that they agree reasonably well considering the different architecture. The position of the gas traces that of the spiral arms, with small scale spurs peeling away at around R=5kpc. The temperature of the particles behaves the same with density for both codes. There is a small population of particles that reach higher temperatures in the range $-24 < \log_{10}\rho < -23$ in the PHANTOM run, though this only amounts to 1% of the particles. The associated phase diagram shows that the gas has followed the warm ISM track into the unstable region (left, Fig. 2.12), implying the PHANTOM calculations allow gas to maintain its warm nature longer in the thermally unstable region before rapidly cooling into the cold ISM region compared to SPHNG. This could easily be the result of the different AV formalism, or the different formulation of the SPH energy equation. The chemical evolution is similar for both codes, showing only minor differences. The median values[7] of the thermal energy, CO, H_2 and H I fractions agree between 0.5-10%, with CO giving the greatest discrepancy between the codes. This CO is sub-cycled numerous times, and is an extremely sensitive function of density, making slight differences in density evolution between the codes lead to moderate differences in the CO abundance of individual particles.

---

[7]As the abundances vary on log scales their means are extremely sensitive to single particles having a slightly increased density, which can differ easily between codes due to the different architecture.

2.3 ISM specific physics

The physics so far discussed only covers hydrodynamical and gravitational forces. For the use of SPH on galactic scales we must include additional physical effects, predominantly ISM heating and cooling mechanisms. The various heating and cooling mechanisms have been adapted for Galactic scale use in Dobbs et al. (2008), and will only briefly be discussed in Section 2.3.1. Of paramount importance to the work presented here is the molecular content of the ISM. Rather than assuming some constant molecular gas fraction that would linearly scale with gas density, we evolve the gas content of each SPH particle individually. This is discussed in full in Section 2.3.2.

The various heating, cooling and chemical processes require a measurement of temperature, determined by the equation $T = \mu u(\gamma - 1)/R$. This requires some value of the mean molecular weight, defined by the ratio of mass density to number density (M/n) of all species of interest, j. We can use the atomic weight, A_j, of each species to make the calculation simply a sum of number ratios of each species, χ, a.k.a. the abundance of each species[8] with respect to H I, given by n_j/n_{HI}. The mean molecular weight is then

$$\mu = \frac{\sum_j M_j}{\sum_j n_j} = \frac{\sum_j n_j A_j}{\sum_j n_j} = \frac{\sum_j \chi_j A_j}{\sum_j \chi_j} = \frac{\sum_j q_j}{\sum_j \chi_j} \qquad (2.52)$$

where we have divided through by the number density, and refer to the ratio by mass of each species as $q_j = \chi_j A_j$. In the ISM there are a wealth of different species, but only a handful are of significant abundance to contribute to the calculation above, specifically H I, He and H_2 where each has weights of $A_{\mathrm{HI}} = 1$, $A_{H_2} = 2$ and $A_{\mathrm{He}} = 4$. The abundance of H_2 is evolved in our calculations, so we know this at any point in the simulation, and as it is composed of hydrogen we can formulate the H I abundance simply as $\chi_{\mathrm{HI}} = 1 - 2\chi_{H_2}$. Helium is of little importance to the simulations presented in this thesis, so we have fixed the abundance at $\chi_{\mathrm{He}} = 0.1$ (Glover & Mac Low 2007). This gives the following equation for the mean molecular weight

$$\mu_{ISM} \approx \frac{q_{\mathrm{HI}} + q_{H_2} + q_{\mathrm{He}}}{\chi_{\mathrm{HI}} + \chi_{H_2} + \chi_{\mathrm{He}}} = \frac{\chi_{\mathrm{HI}} + 2\chi_{H_2} + 4\chi_{\mathrm{He}}}{\chi_{\mathrm{HI}} + \chi_{H_2} + \chi_{\mathrm{He}}}. \qquad (2.53)$$

In the case where the gas is entirely atomic we find $\mu_{ISM} = 1.27$ and if the hydrogen is entirely in a molecular state then $\mu_{ISM} = 2.33$.

Several processes involved in the chemistry and cooling involve the heating/destructive effects of the local radiation field. Here we need to take into account the attenuation/extinction effects from the column density of the surrounding material that effectively shield the target atom/molecule from the incident photons. This manifests as an additional multiplicity factor that reduces the reaction rate as a function of the optical depth of the surrounding medium, simply of the form[9] $f = e^{-\tau_\lambda}$. This can be through individual line absorption by certain species, such as in H_2 self shielding, or through continuous absorption by ISM dust grains. We can relate the opacity

---

[8]The number ratios and mass ratios are χ and q respectively, where for the species of interest we have $q_{H2} = 2\chi_{H2}$ and $q_{He} = 4\chi_{He}$.

[9]This is solution of the radiative transfer equation in the absorbing only case, i.e. Beer's Law, and will the subject of further discussion in Chapter 5

to a parameter called the "visual extinction", A_v, by $\tau = \gamma A_v$, which is the difference in magnitudes between the shielded and unshielded cases. The ratio of visual extinction has been seen to be a constant ratio of the colour excess, $E(B - V) = A_B - A_V$, (Draine & Bertoldi 1996; Bergin et al. 2004), by amounts of either

$$R_V = \frac{A_V}{E(B - V)} \approx \begin{cases} 3.1 & \text{In diffuse ISM} \\ 5.0 & \text{In dense clouds} \end{cases} \tag{2.54}$$

Calculating the visual extinction is no easy task, but there exists a simple conversion between the column density of the gas and the extinction using a visual extinction conversion factor, A_V^{CF}, via

$$A_V^{CF} = A_V / N_{col} = 5.348 \times 10^{-22} \text{cm}^{-2} \tag{2.55}$$

where A_V^{CF} is the extinction per unit column density (N_{col}) for which we will use a constant value throughout the work presented in this thesis unless stated otherwise. This relation was found by measuring the colour excess as a function of column density in various Galactic sources and noticing the ratio between them was roughly constant, with a slope $E(B-V)/N_{tot} = A_V^{CF} R_v$ (Bohlin et al. 1978). The value above is the standard value from the work of Bohlin et al. (1978) but there are several values in the literature. A statistical analysis by Güver & Özel (2009) find a value 20% greater than this standard, however, as we will show later in this section, A_V^{CF} must change by orders of magnitude to have a noticeable effect in our calculations. This extinction will come into play in dust shielding factors in photon based reactions in dense regions. In several rates the UV radiation field will be attenuated by a dust shielding factor, e.g. $G_{dust} = f_{dust}G_0$ where $f_{dust} = e^{-\gamma_X A_v^{CF} N_{HI}}$. The values of γ_X adapt this relation slightly for specific target species (Glover et al. 2010), and take values of 2.5 for cooling, H I and CO chemistry and 3.74 for H_2 dust shielding (which has additional self shielding attenuation).

In order to utilise Equation 2.55 we require a measurement of column density, which is the density of material integrated spatially along the line of sight between two points. The column density is somewhat difficult to compute while keeping the code relatively simple and streamlined. A full treatment of column densities would require large neighbour calculations, scaling as $O(N^{5/3})$ compared to the $O(N \log N)$ for gravity or $O(N)$ for nearest neighbour finding (Glover & Mac Low 2007). Instead we assume the column density is simply the local density times by some distance measurement, l_i

$$N = \int n \mathrm{d}s \approx n l_i \tag{2.56}$$

where we have approximated the distance of integration by a distance relevant to that of the chemistry in question. The smallest is the general distance scale used for cooling and heating rates, specifically PAH and photoelectric effects, where we adopt $l_{cool} = 10$pc. The H I ionisation is shielded on much larger scale, a value of $l_{HI} = 100$pc, due to its chemistry mostly evolving in sparser regions. Finally there is the length scale used in the evolution of H_2 and CO, which evolves in much denser regions than H I, hence has a smaller column density scale length $l_{ph} = 35$pc. The latter of these comes from the typical distance to a B0 star, due to their large luminosity and higher abundance than the more luminous O stars. This is the same method as adopted by Dobbs (2008),
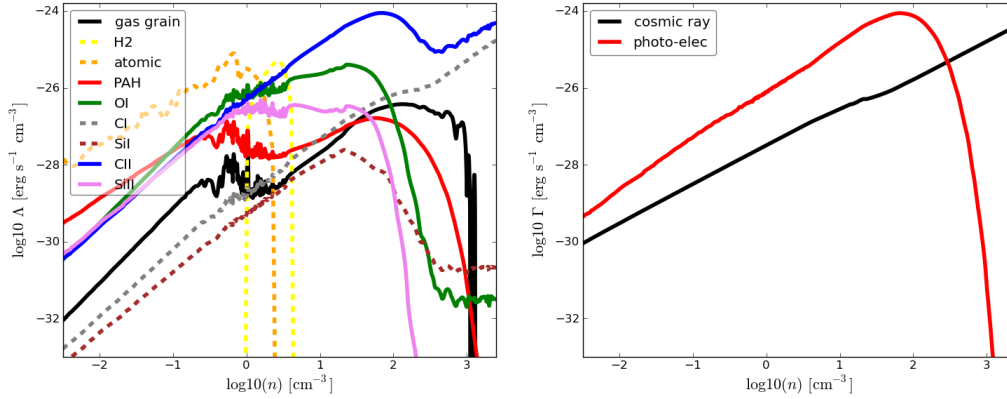
Figure 2.10: Cooling (left) and heating (right) functions discussed in the main text and utilised in our calculations. We have included CI and SiI for comparison using abundances of 1% of their ionised counterparts, but these are absent in our simulations. The pressure-density profile used is similar to that shown in Fig. 2.12.

who investigate values of l_{ph} from 15-100pc, finding the amount of molecular gas is only weakly coupled to this distance measurement.

2.3.1 Cooling and heating functions

The thermal evolution of the ISM can be separated into two distinct components,

$$\dot{U}_{ISM} = \Gamma_{ISM} + \Lambda_{ISM}, \qquad (2.57)$$

which are added into the standard SPH energy equation. We use the convention that Λ is energy loss (cooling) and Γ is energy gain (heating). The various heating and cooling processes contributing to Λ and Γ are taken from the work of Glover & Mac Low (2007). In order for an energy scheme to be relevant to the calculation there are numerous criteria that must be satisfied. The species involved in the process must be abundant enough to ensure frequent collisions. The energy required for the reaction must be of the order of the kinetic energy of the gas, and the probability of the reaction (e.g. the Einstein coefficients) must be large enough for the process to be sufficiently frequent. The various heating and cooling effects relevant to our calculations are shown as a function of density in Figure 2.10, and are described individually briefly below. Due to the complex nature of many of these rates, the cooling and heating is tabulated into numerous arrays as a function of temperature at the beginning of the calculation. Rates that are collisional in origin have a density dependance of n^2, whereas those caused by direct photon interaction scale as n.

The various heating and cooling processes are listed in Table 2.2, along with their corresponding references. We refer the reader to these references for details and formulation of the heating and cooling processes, and will only briefly discuss them here. ISM cooling comes from a variety of mechanisms including H I collisional excitation, H$_2$ vibro-rotational collisional excitation, fine structure excitation of C II, Si II and O I, heat transfer between gas and dust grains, and
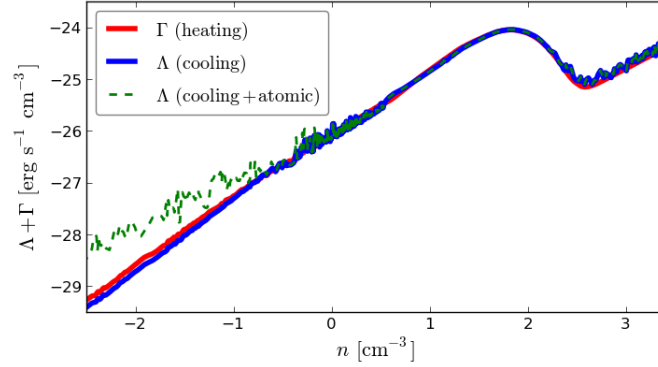
Figure 2.11: Combination of the heating and cooling functions shown in Fig. 2.10. The atomic (H I) cooling is shown for comparison but only dominates for a small number of particles in each simulation. Si I and C I are not included in these total rates. The resulting heating and cooling functions are of the same order of magnitude as a function of density.

the recombination of free electrons with PAH and dust grain surfaces. The dominant of these processes are the fine-structure cooling lines in the mid to high density regime, with recombination cooling becoming more important at lower densities (Fig. 2.10). Atomic and molecular hydrogen cooling, while strong, only becomes significant in a small fraction of the ISM. These are shown as the dashed lines in Fig. 2.10. We neglect the cooling effects of CO, but this is compensated for by maintaining a constant fraction of C II for cooling purposes, and the fact that the C II and CO cooling functions are similar (Glover & Jappsen 2007).

ISM heating is provided by two mechanisms; photo-electric heating on dust grains, large molecules and PAH's, and heating by cosmic-rays (right panel of Fig. 2.10). The photo-electric heating is the stronger of the two, except in the densest regions where dust becomes significantly shielded by the high column density ISM (Bergin et al. 2004). Cosmic-ray heating provides 20eV per reaction and takes into account heating of all ISM species, though is predominantly the ionisation of H I.

Process	Description	Reference
H I (atomic) cooling	Electron collisional excitation/ resonance line emission	Sutherland & Dopita (1993)
H$_2$ (molecular) cooling	Vibrational/rotational excitation cooling by collisions with H I, He and H$_2$	Le Bourlot et al. (1999)
Fine structure cooling	C II, Si II and O I collisions with H I, H$_2$, free e$^-$ and H II	Glover & Jappsen (2007)
Recombination cooling	Free e$^-$ recombining with ionised gas on PAH and dust grain surfaces	Wolfire et al. (2003)
Gas-grain cooling	Dust-gas collisional heat transfer	Hollenbach & McKee (1989)
Cosmic-ray heating	Temperature independent 20eV cosmic-ray photons	Goldsmith & Langer (1978)
Photo-electric heating	UV e$^-$ excitation from dust and PAH	Wolfire et al. (2003)

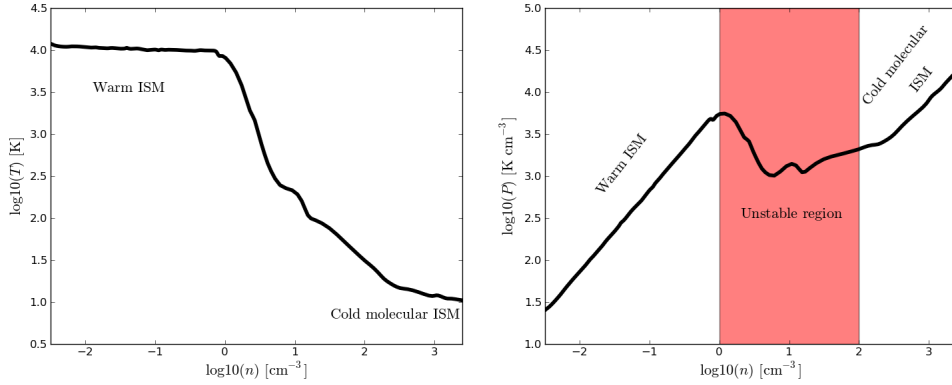Table 2.2: Heating and cooling processes present in our ISM calculations.

Figure 2.12: Thermodynamical properties in a simple disc galaxy simulation. Particle properties have been averaged in density space over the 1 million particles to clarify features. The temperature of the gas plateaus around 10,000K, corresponding to the warm neutral/ionised ISM component. A two-phase ISM is clearly seen in the P-ρ diagram, separated by an unstable region. A small fraction of the gas follows the warm track higher with increasing density and drops down further later in the unstable region.

The total heating and cooling rates are shown in Figure 2.11. We also show the effect of the atomic cooling function, which dominates the hot medium, but this is only a small fraction of the simulation particles (green dashed line). The heating and cooling is of comparable strength in most places (as can be seen by eye). In the region of thermal instability near 1 atom cm^{-3} there is a lot of variation in the cooling rates (this has been smoothed over in Figure 2.11 for clarity). This is due to the splitting of the population between gas that falls into the pressure well and gas that continues to travel up the warm branch of the phase-diagram before falling in, continuing to cool.

The resulting thermodynamic properties of the ISM gas in a "standard" simulation are shown in Figure 2.12. The temperature and pressure profiles as a function of number density are shown for gas in a disc subject to the multitude of heating and cooling effects outlined above. Properties have been binned in density space for clarity, as there is a large variation in the mid-density range. The temperature profile in the left panel shows a clear plateau around 10000K, where the gas condenses almost isothermally up to a density of 1 atom cm^{-3}. Here the gas begins to cool and experiences a drop in pressure, entering the thermally unstable region (the shaded area in the right panel of Fig. 2.12). The gas can continue to cool and contract, becoming molecular in the process. These phase curves are a direct result of the heating and cooling functions, and show a good match to others in the literature (e.g. Field et al. 1969, Wolfire et al. 1995, Liszt 2002) showing a clear distinction between warm and cool phases (see Section 1.4).

2.3.2 Chemistry

One of the most important aspects of the calculations is the tracing the chemical evolution of the gas. As hydrogen is the most abundant gas in the ISM, the first order chemical processes of this gas are some of the most important in the ISM. Hydrogen is allowed to dissociate over time (and to recombine), and to become molecular at higher densities. In order to create molecular

emission maps we must also include a prescription for creating CO. This process is much more complex than the other reactions as it is dependent on numerous intermediate species between the atomic stage and the creation of CO. The ionisation processes alone for the species relevant to CO formation encompass a staggering amount of different processes (Glover & Jappsen 2007; De Becker 2013). If we were to attempt to include only molecular hydrogen and carbon monoxide, then the data in the UMIST database (McElroy et al. 2013) suggests you would need to encompass approximately 800 and 400 reactions respectively (assuming that database is complete!). This is clearly not computationally feasible, especially in the case of a galactic scale simulation where our time and spatial dimensions far exceed those important chemically. Instead it is prudent to limit ourselves to only those reactions deemed most important in each case, sacrificing some of the accuracy for computational simplicity.

The chemical evolution of each species (i.e. the evolution of the number density, n_X) is encompassed by a single ordinary differential equation of the form

$$\frac{dn_X}{dt} = C_X - D_X n_X, \tag{2.58}$$

so that the density of species X at the next timestep, $t + \Delta t$, is

$$n_X(t + \Delta t) = n_X(t) + \frac{dn_X}{dt}\Delta t \tag{2.59}$$

including a creation coefficient, C_X, and a destruction coefficient, D_X, unique to the species and often a complex function of density, temperature, and the abundance of other species. For example, the rate of H_2 formation will depend heavily of the abundance of atomic hydrogen. The exact form of our H I, H_2 and CO rate equations are discussed in the following sub-sections. Each particle in our calculations carries with it a chemistry array of the form $\vec{\chi}_a = (\chi_{HI}, \chi_{H_2}, \chi_{CO})|_a$, which is $(1, 0, 0)$ initially. The chemistry arrays keep track of 3 distinct parameters, the H I ionisation fraction, the H_2 ratio and CO abundance which have the values of $(0, 0, 0)$ and $(1, 0.5, \chi_{CII})$ in the low and high density extremes respectively. The electron and proton abundances are also tracked, but χ_{pr} is simply $\chi_{pr} = 1 - \chi_{HI}$ and $\chi_{el} = 1 - \chi_{HI} + \chi_c$ where χ_c is the constant free electron abundance, a result of the ionisation of species other than H I.

The regime of negative abundances should be avoided at all times, i.e. the right hand side of Equation 2.59 should always be > 0. To ensure this we must include some chemical timestepping criteria (Glover & Mac Low 2007; Dobbs et al. 2008). This can be calculated from $dt_{dest} = -n_X/(C_X - D_X n_X)$. The sub-stepping time frame is then taken to be 10% of the time taken to completely destroy the species of interest ($dt_{chem} = 0.1 dt_{dest}$), ensuring negative abundances are avoided. If creation is occurring, then the chemistry is sub-cycled on some fixed scale ($dt_{chem} = dt_{hydro}/200$). The sub-stepping used to avoid negative abundances is not applied to the H I chemistry, as the abundance never reaches into this range in our simulations. Some sub-stepping is still applied however, as the chemistry is at least evolved on the cooling time-scale (Equation 2.41). The CO evolution is done inside of the H_2 evolution, as the two are intrinsically linked, and the CO evolution is also allowed to sub-step further if required.

As the chemistry is evolved as in Equation 2.58 there is expected to be some loss of accuracy
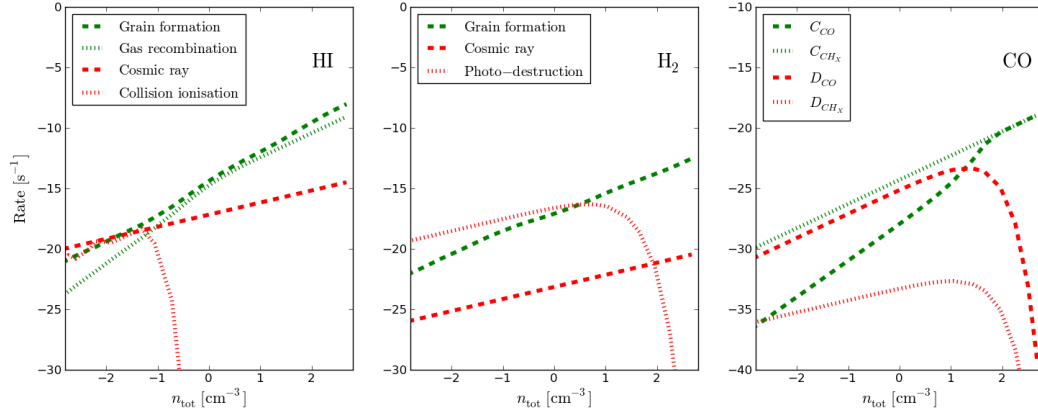
Figure 2.13: The individual components in the creation and destruction terms for H I (left), H$_2$ (middle) and CO chemistry (right). The density-temperature profile is that used in Figure 2.12 where the rates are calculated at fixed abundances of $\vec{\chi} = (0.94, 1 \times 10^{-6}, 1 \times 10^{-15})$.

over large integration time-scales (as opposed to the 2nd order integrations used to evolve the SPH quantities). However, by evolving the chemistry on the cooling time-scale, as well as including additional chemistry sub-stepping if required, we hope to minimise this. Any sources of error involved in the order of integration are also considered to be minimal compared to the simplistic nature of the chemistry itself.

Three figures will be referenced when referring to the chemical evolution. Figure 2.13 shows the various components contributing to the various creation and destruction rates of each species and Figure 2.14 shows the resulting evolution tracks for each species. The latter was created using a simple 1D code used to test the chemistry, that evolves abundances alone without various thermodynamical properties, but with a fixed *P-T* profile representing the general behaviour of the ISM cooling function (as in Fig. 2.12). Figure 2.15 is similar to Figure 2.14 but shows the effect of changing various parameters important for the chemistry of H$_2$ and CO. All the various chemical reactions are listed in Table 2.3 and the multitude of required parameters are included in Table 2.4.

HI chemistry

At the most basic chemical level we must include the effect of the ionisation of neutral hydrogen gas, creating an abundance of H II and free electrons. The various atomic hydrogen processes have been coded by Glover & Mac Low (2007) and included in our codes (see Table 1 of their paper). The H I chemistry coefficients included are

$$C_{\text{HI}} = k_{rec}n_e n_p + k_{gr}n_p n \tag{2.60}$$

and

$$D_{\text{HI}} = \zeta_{CR} + k_{ci}n_e \tag{2.61}$$

where we include the effects of gas-phase recombination, k_{rec}, formation on grain surfaces, k_{gr}, cosmic ray ionisation, ζ_{CR} and free electron collisional ionisation, k_{ci}. The gas-phase recombina-

tion rate of hydrogen, $H\,II + e^- \rightarrow H\,I + \gamma$, is that of Ferland et al. (1992) who produce a grid of temperature based recombination rates and approximate temperature dependent functions. The exact form is from Glover & Jappsen (2007), given as

$$k_{rec} = 2.753 \times 10^{-14}(315614/T)^{3/2}(1 + (115188/T)^{0.407})^{-2.242} \tag{2.62}$$

which is purely temperature dependent, but the actual recombination rate is a function of n_e and n_{HII} which are heavily density dependent. The gas phase reaction is often an order of magnitude lower than the grain phase, though will exceed the grain phase in the highest density regions. The grain phase recombination, $H\,II + e^- + grain \rightarrow H\,I + grain$, is given in Weingartner & Draine (2001) as;

$$k_{gr} = \frac{1.22 \times 10^{-13}}{1 + 8.074 \times 10^{-6}\psi^{1.378}(1 + 508.7T^{0.01586}\psi^{-0.4723-1.102\times10^{-5}\ln T})} \tag{2.63}$$

with the effect of grain charging is characterised by ψ parameter, given by $\psi = G_{dust}\sqrt{T}/n_e$. The ionisation of $H\,I$ comes from a combination of cosmic ray ionisation and ionisation by collisions with free electrons, $H\,I + e^- \rightarrow H\,II + 2e^-$, from Abel et al. (1997) at a rate

$$k_{ci} = \exp\left(\sum_{i=1}^{8} c_i(\ln T)^i\right) \tag{2.64}$$

where the c_i parameters were fit to experimental data. This rate decays extremely fast with increasing density (and decreasing temperature), as seen in Fig. 2.13, and so will only add to additional ionisation already caused by cosmic rays. The evolution track of $H\,I$ ionisation is shown in Fig. 2.14 (left panels). The figure shows that the ionisation is of little importance in the cold phase of the ISM ($n > cm^{-3}$) and has a near linear dependance on density in the warm phase. The maximum ionisation fractions seen in our calculations are approximately 15%. While seemingly low, it is not surprising considering we are not effectively modelling the hot ISM component.

An important note is that we are talking about the hydrogen ionisation fraction above. The actual abundance of atomic hydrogen is the amount of unionised hydrogen that is not locked in molecular form, $\chi_{HI}(1 - \chi_{H_2})$.

H_2 chemistry

The chemistry of H_2 used here is taken from Bergin et al. (2004) and Draine & Bertoldi (1996) based on the work of Hollenbach et al. (1971), and used by Glover & Mac Low (2007) and Dobbs et al. (2008). The creation and destruction terms are given by

$$C_{H_2} = R_{gr}(T)n_{tot}n_{HI} \tag{2.65}$$

and

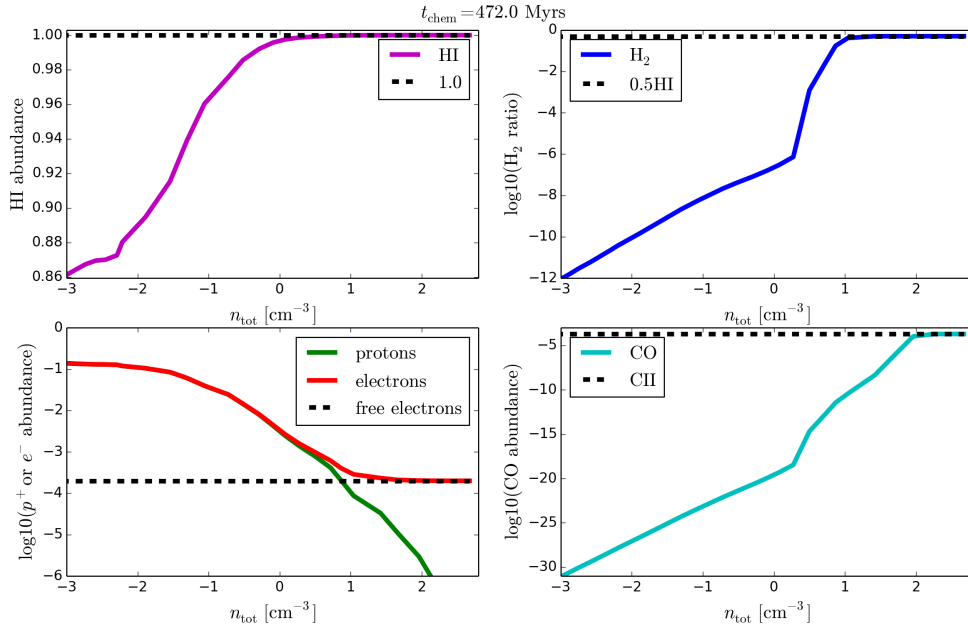$$D_{H_2} = \zeta_{CR} + \zeta_{H_2}(N_{H_2}, N_{tot}). \tag{2.66}$$

Figure 2.14: Chemical evolution for a simple 1D code with a fixed *P-T* profile (Figure 2.12). Moelcular processes are a strong function of total gas density, with abundances displaying a sharp increase in the region of thermal instability of the ISM phase diagram. The ionisation fraction of H I also scales with density, reaching a maximum ionisation fraction of 15% in our calculations. The high density limits for the H I, H_2 and CO abundances of 1, 0.5 and χ_{CII} are shown as dashed lines.

where H_2 is formed on grain surfaces and is destroyed by photo-dissociation and cosmic rays. The formation of H_2 on grain surfaces occurs at a rate of

$$R_{gr}(T) = R_{gr}(T_\circ)S \sqrt{T} \; \text{cm}^3\text{s}^{-1}\text{K}^{-0.5}, \tag{2.67}$$

where $R_{gr}(T_\circ) = 2.2 \times 10^{-18}\text{cm}^3\text{s}^{-1}$ and the grain formation efficiency is $S \approx 0.3$. The efficiency factor should be a function of gas and grain temperatures, but we use a fixed value for simplicity (Dobbs et al. 2008). The destruction of H_2 is a result of cosmic ray photo-ionisation, ζ_{CR}, and a local photo-destruction term, ζ_{H_2}, which is a function of the column density and the visual extinction, A_v. The term takes the form as given in Draine & Bertoldi (1996)

$$\zeta_{H_2}(N_{H_2}, N_{\text{tot}}, A_v) = f_{\text{shield}}(N_{H_2})f_{\text{dust}}(N_{tot})\zeta_{H_2}(0) \tag{2.68}$$

where the constant raw/unshielded photo-dissociation rate is $\zeta_{H_2}(0)$ (with a UV field strength of $G_\circ$ built in). The additional terms are the dust shielding/attenuation, f_{dust} and the self-shielding, f_{shield} (i.e. shielding by absorption of other H_2 molecules). The dust shielding factor is similar to that used in the cooling, specifically given by

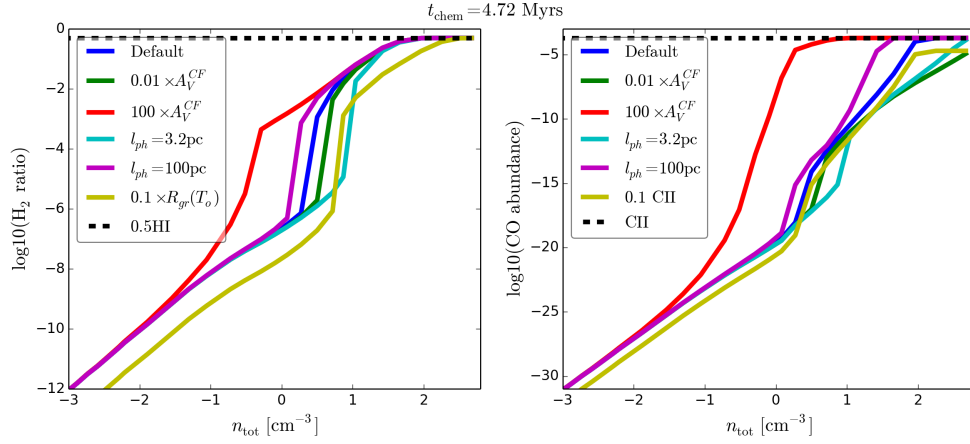$$f_{\text{dust}} = e^{-\tau_{d,1000}} \tag{2.69}$$

Figure 2.15: Similar evolution tracks as Figure 2.14 but showing the effect of different values of A_V^{CF}, l_{ph}, $R_{gr}(T_\circ)$ and C II on the formation of H_2 and CO.

where the optical depth for shielding is taken at 1000Å. This is given by the $\tau_{d,1000} = \gamma_{H_2} A_V$ where the visual extinction can be calculated from the column density of the gas and $\gamma_{H_2} = 3.74$ (Glover et al. 2010) for the diffuse ISM where $R_V \approx 3.1$, but could grow much greater in the dense, $R_V \approx 5$, ISM. The self shielding factor is given by

$$f_{\text{shield}} = \frac{0.965}{(1 + x/b_5)^2} + \frac{0.035}{\sqrt{1 + x}} \exp\left[-8.5 \times 10^{-4} \sqrt{1 + x}\right] \qquad (2.70)$$

where $x = N_{H_2}/5 \times 10^{14}$ cm^{-2} and b_5 is the Doppler broadening width of the line absorption in units of 10^5 cms^{-1}, where we adopt 3km s^{-1} (Lee et al. 1996), and approximate column densities by $N_X = n_X l$.

The cosmic ray rate is only significant in regions of high column density gas where H_2 is well shielded against the surrounding UV field (middle, Fig. 2.13). The grain formation rate is predominantly a linear function of density, with the $\sqrt{T}$ dependance only having a marginal effect. The majority of parameters included in the above equations do not have a very strong effect on the global H_2 production, illustrated by Figure 2.15. The impact of varying l_{ph} and A_V^{CF} is only noticeable in the mid-density regime. Changes in the extinction conversion factor, A_V^{CF}, need to be of around two orders of magnitude to see a noticeable difference H_2 production. Fortunately the parameter with the greatest uncertainty, the distance for column density calculations, shows to only impact the density range of 1cm^{-3} < n < 10cm^{-3}. Even then the change is only minor considering the distance is changed from 3.2-100pc.

CO chemistry

The chemistry underlying the production of CO is somewhat more complicated than that of H_2 due to the wealth of intermediate species between the atomic and final molecular stage of CO. Previous studies rarely attempt to model all the species involved in CO formation/destruction due to the cripplingly slow speed it which it would take to evolve all these species, and their ions.

Instead works in the literature focus on what they deem to be the most important rates for the problem at hand. In Glover et al. (2010) the authors compare some of these different approaches including (ordered by complexity) Glover et al. (2010), Keto & Caselli (2008), Nelson & Langer (1999) and Nelson & Langer (1997), ranging from modelling 218 to 4 reactions. The authors find that more reactants doesn't necessarily imply better accuracy, and that if the primary concern is the distribution of galactic CO then even the simplest model does a good job compared to the those with much greater complexity. In light of this we utilise the simple model of Nelson & Langer (1997)[10]. In this model C II is converted to CO through the production of some intermediate hydrocarbon step, denoted CH_X, resulting from an initial reaction of C II with H_2 (at a rate k_0). An intermediate stage of the neutralisation of CH_2^+ to CH_X (encompassing CH_2 and CH) is not modelled, and is assumed to occur on timescales much smaller than anything else in the CO formation process. The CO and CH_X are subjected to photodestruction which is several orders of magnitude stronger than cosmic ray ionisation in most regions. The creation and destruction rates of CO are

$$C_{CO} = k_0 n_{H_2} n_{CII} \beta \qquad (2.71)$$

and

$$D_{CO} = \zeta_{CO}(N_{tot}), \qquad (2.72)$$

where β quantifies the efficiency of the reaction $CH_X + O\,I \rightarrow CO + H_X$ (at a rate k_1) over the photodestruction of CH_X. This is represented by

$$\beta = \frac{k_1 n_{OI}}{k_1 n_{OI} + \zeta_{CH_X}(N_{tot})}. \qquad (2.73)$$

The abundances of O I and C II are needed to quantify the production of CO, however as we have already mentioned these are not species we follow in our calculations. We make the basic assumption that O I and C II are either in their original forms or locked into CO, i.e. $n_{CII}(t) = n_{CII}(0) - n_{CO}(t)$ and $n_{OI}(t) = n_{OI}(0) - n_{CO}(t)$. The two separate photo-destruction rates are given by

$$\zeta_{CH_X}(N_{tot}) = G_{dust} \zeta_{CH_X}(0) \qquad (2.74)$$

and

$$\zeta_{CO}(N_{tot}) = G_{dust} \zeta_{CO}(0) \qquad (2.75)$$

where $\zeta_{CH_X}(0) = 5 \times 10^{-10} s^{-1}$ and $\zeta_{CO}(0) = 1 \times 10^{-10} s^{-1}$ and the UV field coupled to the dust attenuation is similar to previous sections, $G_{dust} = G_\circ f_{dust} = G_\circ e^{-\tau_{UV}}$. The visual extinction is grouped together for both processes and calculated using the total column density as $\tau_{UV} = \gamma_{CO,CH_X} A_V$ where $\gamma_{CO,CH_X} = 2.5$ (Nelson & Langer 1997). There have been a few simplifications in this model, not including the simplification of a limited number of tracked species and reactions. No self-shielding of CO, or shielding by H_2 is included (f_{H_2}, f_{CO}), which will inhibit the destruction of CO in high density regions. Tabulated forms of these shielding factors are given by (Lee et al.

---

[10]We actually use the formulation from Glover & Clark (2012) as it appears there are a couple of minor typos in the equations of Nelson & Langer (1997) that allow CO formation in the complete absence of H_2. Though this change made only very minor difference to the CO abundance, and only at the highest densities.

1996), but we maintain standard form from Nelson & Langer (1997) as above. We also do not include cosmic ray ionisation in the CO chemistry as the photodissociation rate is many orders of magnitude greater. Even when using 1000 times our fiducial value of ζ_{CR} there was no visible effect on the evolution track of CO in shown in Fig. 2.15. There is also the assumption that the dust shielding factors (γ_X) are the same for all reactions in the CO chemistry. In the substantially more complicated chemical network of Nelson & Langer (1999) the authors use separate factors of $\gamma_{CO} = 3$ and $\gamma_{CH_X} = 1.5$. Individual coefficients for the CH_X components are given in Glover et al. (2010) as $\gamma_{CH} = 1.2\text{-}2.8$ and $\gamma_{CH_2} = 1.7\text{-}2.3$ depending on whether the reaction is dissociative (former) or ionising (later), though the incorporation of these would require a much more sophisticated network. This additional shielding is of little importance in the simulations shown here, as once present in the coldest regions it is very stable, there is no efficient heating mechanism to break it apart once there. Thus any additional shielding to the photo-dissociation is not needed, as the CO is already saturated. If we were to include additional feedback mechanisms then additional shielding may be required.

The behaviour of CO with varying A_V^{CF} or l_{ph} is similar to H_2 in Figure 2.15. The broad evolution is insensitive to small variations in either parameter. However, with either l_{ph} as low as 3.2pc or an extinction factor two orders of magnitude lower than fiducial value then C II is no longer fully saturated into CO in the range of densities modelled in our calculations.

Reaction	Description	Reference
H II + e$^-$ + grain → H I + grain	Grain surface formation	Weingartner & Draine (2001)
H II + e$^-$ → H I + γ	Gas-phase recombination	Ferland et al. (1992)
H I + e$^-$ → H II + 2e$^-$	e$^-$ collisional ionisation	Abel et al. (1997)
H I + c.r. → +H II + e$^-$	Cosmic ray ionisation	Glover & Mac Low (2007)
H I + H I + grain → H$_2$ + grain	Grain surface formation	Bergin et al. (2004)
H$_2$ + γ → 2H I	UV photodissociation	Draine & Bertoldi (1996)
H$_2$ + c.r. → H$_2^+$ + e$^-$	Cosmic ray ionisation	Bergin et al. (2004)
C II + H$_2$ → CH$_2^+$ + γ	Radiative association	Nelson & Langer (1997)
CH$_2^+$ + various → CH$_X$ + various	Rapid neutralisation*	-
CH$_X$ + O I → CO + H$_X$	Gas phase formation	-
CH$_X$ + γ → C + H$_X$	UV photodissociation	-
CO + γ → C I + O I	UV photodissociation†	-

Table 2.3: Processes present in our chemical model focussed on tracing the evolution of H I, H$_2$ and CO with their relevant references.

* Process is intermediate and is assumed rather than fully represented.

† C I is not present in our chemistry, but is assumed to very rapidly photoionise to C II.

Term	Description	Value
l_{cool}	Distance measurement for cooling	10 pc
$l_{\text{H\,I}}$	Distance measurement H I chemistry	100 pc
l_{ph}	Distance measurement for H_2 and CO chemistry (B star)	35 pc
$R_{gr}(T_o)$	H_2 grain formation rate at $T_o = 100K$	$6 \times 10^{-18} \text{cm}^3\text{s}^{-1}$
k_0	CH_X intermediate species formation rate from C II	$5 \times 10^{-16} \text{ cm}^3\text{s}^{-1}$
k_1	CO formation rate from O I + CH_X	$5 \times 10^{-10} \text{ cm}^3\text{s}^{-1}$
$\zeta_{\text{H}_2}(0)$	Unshielded H_2 photodissociation rate	$4.17 \times 10^{-11} \text{ s}^{-1}$
$\zeta_{\text{CH}_X}(0)$	Unshielded CH_X photodissociation rate	$5.00 \times 10^{-10} \text{ s}^{-1}$
$\zeta_{\text{CO}}(0)$	Unshielded CO photodissociation rate	$1.00 \times 10^{-10} \text{ s}^{-1}$
ζ_{CR}	Cosmic ray ionisation rate	$1.00 \times 10^{-17} \text{ s}^{-1}$
γ_{H_2}	Dust shielding factor of H_2 photodissociation	3.74
$\gamma_{\text{CO,CH}_X}$	Dust shielding factor of CO & CH_X photodissociation	2.50
b_5	Doppler broadening factor for H_2	3 km s^{-1}
T_{dust}	Dust temperature for heating/cooling	10 K
A_V^{CF}	Visual extinction conversion factor (A_V/N_{tot})	$5.348 \times 10^{-22} \text{ cm}^{-2}$
G_o	Strength of the UV radiation field in Habing units	1.56
χ_{He}	He abundance	1.0×10^{-1}
χ_{el}	free e^- abundance	2.0×10^{-4}
χ_{CII}	CII abundance	2.0×10^{-4}
χ_{SiII}	SiII abundance	3.0×10^{-5}
χ_{OI}	OI abundance	4.5×10^{-4}
χ_{CI}	CI abundance	0.0
χ_{SiI}	SiI abundance	0.0

Table 2.4: Various parameters and their adopted values for the various chemistry and cooling routines used in this thesis, unless otherwise specified.

2.4 Chapter summary

In this chapter the we have outlined the main computational details of our calculations presented in the remainder of this thesis. In Section 2.2 we discussed the primary workhorse of our investigation, smoothed particle hydrodynamics, used to simulate the evolution of the ISM on a galactic scale. The SPH density formulation was constructed, and momentum and energy rate equations we derived. These including the effect of hydrodynamical and artificially dissipative forces, taking the general forms

$$Dv/Dt = f_{hydro} + f_{AV} + f_{ext}, \tag{2.76}$$

$$DU/Dt = \dot{U}_{hydro} + \dot{U}_{AV} + \dot{U}_{AC} + \dot{U}_{ISM} \tag{2.77}$$

where f_{ext} will be discussed further Chapters 3 and 5. In the former case this is due to purely analytical, smooth potentials, while in the later this is a combination of analytical potentials and the evolution of an N-body stellar component. The importance of artificial dissipation was also discussed (f_{AV}, $\dot{U}_{AV}$, $\dot{U}_{AC}$), required to correctly capture shocks in the converging flows and contact discontinuities, illustrated with a few well known test cases. The two SPH codes utilised in this work were briefly discussed, and a few test calculations were presented to illustrate correct behaviour of the codes in a galactic context.

Section 2.3 presents the main adaptations to required to track the molecular content of the ISM. The thermal profile is reproduced by the inclusion of the $\dot{U}_{ISM}$ term. This includes the various heating and cooling mechanisms of importance to galactic scales, and is discussed briefly in Section 2.3.1. A simple chemical network is also included that allows the tracing of the abundances of H I, H_2 and most importantly CO. The chemical processes are outlined in Section 2.3.2 including the effects of grain formation, gas phase formation, cosmic ray ionisation and photo-destruction. While seemingly rudimentary, the chemistry includes the basic processes required to track the global distribution of molecular gas. If smaller scale structures such as individual clouds were the subject of investigation then the chemical network, and possibly cooling processes, would need to be substantially more complex.

To actually construct our synthetic observations we must use another numerical technique to calculate the emission from the ISM gas. A brief discussion of the theory of radiative transfer, and the torus code used is included in Chapter 4.