Collisionless Stellar Hydrodynamics as an Efficient Alternative to N-body Methods

Nigel L. Mitchell\textsuperscript{1*}, Eduard I. Vorobyov\textsuperscript{1,2}, Gerhard Hensler\textsuperscript{1}
\textsuperscript{1}University of Vienna, Dept. of Astrophysics, 1180 Vienna, Austria
\textsuperscript{2}Institute of Physics, Southern Federal University, Rostov-on-Don, 344090, Russia
3 May 2014

ABSTRACT
The dominant constituents of the Universe’s matter are believed to be collisionless in nature and thus their modelling in any self-consistent simulation is extremely important. For simulations that deal only with dark matter or stellar systems, the conventional N-body technique is fast, memory efficient, and relatively simple to implement. However when extending simulations to include the effects of gas physics, mesh codes are at a distinct disadvantage compared to SPH codes. Whereas implementing the N-body approach into SPH codes is fairly trivial, the particle-mesh technique used in mesh codes to couple collisionless stars and dark matter to the gas on the mesh, has a series of significant scientific and technical limitations. These include spurious entropy generation resulting from discreteness effects, poor load balancing and increased communication overhead which spoil the excellent scaling in massively parallel grid codes.

In this paper we propose the use of the collisionless Boltzmann moment equations as a means to model the collisionless material as a fluid on the mesh, implementing it into the massively parallel FLASH AMR code. This approach which we term “collisionless stellar hydrodynamics” enables us to do away with the particle-mesh approach and since the parallelisation scheme is identical to that used for the hydrodynamics, it preserves the excellent scaling of the FLASH code already demonstrated on peta-flop machines.

We find that the classic hydrodynamic equations and the Boltzmann moment equations can be reconciled under specific conditions, allowing us to generate analytic solutions for collisionless systems using conventional test problems. We confirm the validity of our approach using a suite of demanding test problems, including the use of a modified Sod shock test. By deriving the relevant eigenvalues and eigenvectors of the Boltzmann moment equations, we are able to use high order accurate characteristic tracing methods with Riemann solvers to generate numerical solutions which show excellent agreement with our analytic solutions. We conclude by demonstrating the ability of our code to model complex phenomena by simulating the evolution of a two armed spiral galaxy whose properties agree with those predicted by the swing amplification theory.

Key words: methods : numerical — hydrodynamics

1 INTRODUCTION
Galaxies are complex systems consisting of a multitude of components that often require different approaches for both analytical and numerical study. The gaseous component can be described with the classic hydrodynamics equations (CHE) thanks to frequent collisions between constituent particles that act to isotropise the local pressure. On the other hand, stellar and dark matter particles lack collisions which effectively means the local pressure is anisotropic and should be described by a stress tensor. Although this modification is rather straightforward to implement in the CHE (see e.g. Samland et al.\textsuperscript{1997}, Vorobyov & Theis\textsuperscript{2006}), this approach has mostly been applied to analytical and semi-analytical studies of growing instabilities in stellar systems (e.g. Binney & Tremaine\textsuperscript{1987}).

When it comes to numerical simulations of collisionless galactic components, N-body methods have often been the...
method of choice (e.g. Athanassoula 1984). One owes this to the relative ease with which they can be implemented and one can start modelling three-dimensional collisionless systems - often with rather small numbers of particles. In particular, large dark-matter-only N-body simulations (e.g. the Millennium simulation, Springel 2005) have led to significant advances in our understanding of cosmological structure formation, allowing some of the founding tenets of the current paradigm to be tested.

As simulations have grown in complexity, incorporating the effects of baryonic matter, it has become necessary for theorists to develop a wider range of techniques to model the gaseous component. For this, there exist two dominant approaches; that of Smooth Particle Hydrodynamics (SPH) and alternatively, grid codes. Whilst both approaches agree for simple tests where analytic solutions are available, discrepancies are still often found when modelling more complex phenomena. SPH is very robust and memory efficient, allowing large cosmological runs to self-consistently model the effects of gas physics and collisionless matter (see for example the OWLS and GIMIC simulations; Schaye et al. 2004; Crain et al. 2009). However, due to the higher shock capturing powers and more flexible refinement criteria of mesh codes, along with the tendency of SPH to suppress the formation of hydrodynamic instabilities and turbulent mixing (Agertz et al. 2007; Mitchell et al. 2009), grid codes present a powerful alternative. In particular, grid based schemes are better suited to modelling small-scale physical processes such as heat conduction and radiative transfer.

Unfortunately, when coupling the collisionless matter to the collisionally dominated baryonic matter, it begins to become clear that the N-body approach is not always ideally suited to grid codes. In SPH the gas is discretised into particles of a given mass, allowing the effects of gravity to be calculated using an N-body method in a very convenient self-consistent manner for all components. Grid codes however require the mapping of the collisionless particle masses to the mesh, from which the global density field can be used to obtain the gravitational potential. This can be done using the multigrid technique (Fryxell et al. 2000) with either a nested grid or Adaptive Mesh Refinement (AMR). Alternatively a Fourier Transform or a hybrid of the two can accelerate this process. Once the gravitational acceleration is computed on the grid, it is then interpolated back to the particles. These are updated using a leapfrog method. This particle-mesh approach, a mapping of discrete particle properties to and from the mesh, represents a series of scientific and technical limitations to grid codes which degrade the physical reliability of their results and hampers the scalability of grid based simulations. These include:

- **Spurious generation of entropy** - The particle-mesh technique allows numerical noise to be introduced if insufficient particles are present in a given resolution element or “cell.” It has been argued by Springel (2010) that in regions where the mesh is over-sampled relative to the number density of particles, then some cells may receive mass mapped from particles whilst their neighbours may not. This can lead to strong variations in the local density field when realistically the field should be smooth and continuous. This can drive spurious weak shocks and turbulence on small scales, generating an unphysical entropy excess. Springel (2010) cites this as the origin of the higher entropy cores seen in galaxy cluster simulations when run using grid codes, compared to those run with SPH codes which have cuspy low entropy cores (Frenk et al. 1999; Mitchell et al. 2009).

- As actually these particles should represent a continuous mass field, this effect can be limited by mapping the discrete mass of a given particle over a larger region using some pre-defined kernel function (for example the “Cloud-in-cell” technique; Harlow, Ellison & Reid 1964), this however leads to more technical problems.
  - **Increased communication overhead** - In massively parallel programs, a large amount of effort is often put into minimising communication between different nodes over the network. Such communication is relatively time consuming, with latencies (the time taken to establish and terminate a communication) many orders of magnitude greater than the clock speed of the processor. As such, many calculations could be performed whilst waiting for a given group of processes to communicate, and if there is a list of different processes to communicate with, the time adds up. Therefore any common communication scheme and grouping of messages, along with a reduction in the number of different nodes to which we must communicate with, presents a significant speed up in the code.

When particle masses are mapped using a kernel, there will inevitably be some overlap with neighbouring cells which may or may not reside on different nodes. Although communication over the network is natural for any parallel simulation, it is lower when dealing with a system which advocates only fluxes through cell surfaces instead of a system which can map mass to any neighbouring cell, whether it is directly adjacent to a face or merely in contact with an edge or vertex of the cell. This is demonstrated in figure 1 in which a schematic layout of two dimensional (2D) block of cells (as used in the FLASH code) is shown along with the dashed outline of the neighbouring blocks which surround it. The internal cells in the block are shown in red and a surrounding layer four guardcells deep contains boundary data copied from neighbouring blocks. Grey cells are guardcells which do not need to be updated or communicated for the given scheme, whilst green cells are guardcells which need to be updated or exchanged with neighbours. A 2D hydrodynamic simulation (top panel) requires the communication of boundary data to compute the four fluxes through its cell faces with the grey corner guardcells not being used. The bottom panel however, shows how mapping of particle mass over a kernel of finite size can lead to particle properties being mapped to any of the surrounding guardcells. This data then needs to be communicated to the internal cells of the blocks to which they correspond. Thus in 2D the communication overhead is doubled, whilst in 3D it can increase by a factor of 3.3 (6 face fluxes / 20 edge and corner cells to which mass may be mapped). Sophisticated space filling curves can help optimise the distribution of blocks (or individual cells) across nodes so that as many adjacent neighbours as possible reside on the same processor, minimising inter-node communication. However increased overhead remains unavoidable, especially as we always remain bottle-necked by the slowest processor if the simulation is to proceed in lock-step.

- **Poor load balancing** - The use of particles in conjunction with a mesh also presents a load-balancing issue as many forms of astrophysical structure formation result in the nat-
ural concentration of the majority of mass in dense compact regions. This leads to too many particles accumulating in one or two cells whilst the remaining cells have few or none, even if the initial particle distribution is fairly uniform. Ultimately, the mass field ends up poorly sampled in the under-dense regions of the simulation whilst the mapping of properties to and from the mesh is performed by a limited number of cells within the densest regions. As already mentioned, the hydrodynamics requires that as many neighbouring cells as possible be stored on the same processor. Therefore these heavily occupied dense cells are all stored on the same few processors leading to extremely poor load balancing. Although distributing these dense blocks more evenly would improve the load balancing for the particle-mesh algorithm, it would ruin the load balancing for the hydrodynamics, creating an impossible system to optimise.

Fortunately the use of the collisionless Boltzmann moment equations allows us to overcome all of these limitations since it allows the stellar and dark matter components to be represented as a collisionless fluid in a near identical manner to the standard baryonic gas. This removes the need to develop a separate parallelisation scheme for the particle-mesh approach and allows us to use the same well parallelised hydrodynamics scheme. Instead of communicating guardcell data for the gas and stellar material separately, the data can now be grouped and communicated all at once, minimising the latency overhead and maximising bandwidth usage. Ideal load-balancing is also achieved as the time taken to calculate the inter-cell fluxes is independent of the amount of collisionless mass within a cell. Thus processors will not be left idle whilst others calculate, provided that the cells are evenly distributed across processors. Given that the hydrodynamics in the FLASH AMR code we use has been shown to scale well for tens of thousands of processors (Antypas 2006), then by using the same parallelisation scheme for the collisionless material, we naturally preserve this scaling. In an era of increasingly large and complex simulations, this scalability is vital in keeping performance in-line with scientific requirements.

Most importantly, as the density of the collisionless material is now represented in each cell continuously, this removes the potential for spurious generation of entropy through discreteness effects introduced when mapping particle properties to the mesh. It also removes limitations involved when converting gas into stars of discrete masses - problematic if cells contain insufficient mass but yet have densities that satisfy star formation criteria.

We present in this paper our new implementation of the Boltzmann moment equations which allows collisionless material to be modelled as a collisionless fluid on the mesh. A major application of this technique will be in modelling the behaviour of large collections of stars, such as those in a galaxy, as a collisionless fluid in much the same way that we model gas hydrodynamics in grid codes. We will therefore refer to it from here on as “collisionless stellar hydrodynamics.” We begin in § 2 with a derivation of the zeroth, first and second order moment equations of the collisionless Boltzmann equation and then proceed in § 3 to outline the FLASH hydrodynamic code (§ 3.1) and the different numerical schemes we use to numerically integrate the equations. These include both a simpler Riemann solver free technique

---

**Figure 1.** Diagrammatic representation of a block of cells and the required communication overhead for a flux based scheme (top panel) and a particle-mesh scheme (lower panel). The internal cells of the block are shown in red. Surrounding guardcells which contain boundary data mapped from the internal cells of neighbouring blocks (visible as dashed squares) are shown in grey. Those guardcells which need to be updated after every step for the given scheme are highlighted in green. In the top panel, only flux data through the cell faces needs calculating and therefore only guardcell data directly adjacent to these surfaces needs to be communicated. Thus in 2D, only four of the eight neighbouring blocks need to be communicated with. However, the particle-mesh algorithm (bottom panel) which maps particle properties to the mesh, over a finite region with some Kernel, allows data to be mapped into all of the guardcells. Therefore it is necessary to perform a full guardcells exchange and to update the corresponding internal cells of all neighbouring blocks. This results in a much greater communication overhead.
and a more sophisticated Riemann solver scheme with characteristic tracing (§4.2).

In §4.1 we outline our discovery of how standard hydrodynamic test problems for which analytic solutions are known, can be modified under special conditions to reproduce the results of a collisionless fluid. This enables us to use the Sod-shock-tube test problem to test our results, the findings of which are given in §4.1. We continue our tests with both a pressure-free spherical collapse test (§4.2) and an anisotropic collapse with non-isotropic dispersion tensor in §4.3. As a more complex test of our collisionless stellar hydrodynamics, in §4.4 we track the growth of a two armed spiral in a stellar disc, contrasting the resulting structure with the predictions of the swing amplification theory.

We then conclude our findings with an overview of our results, a discussion of the runtime performance of our code and the future implications of our work in §5.

2 BOLTZMANN MOMENT EQUATIONS

The dynamics of collisionless particles are fundamentally described by the collisionless Boltzmann equation for the distribution function of particles in the six dimensional position-velocity phase space $f(t, x, v)$

$$\frac{\partial f}{\partial t} + v_i \cdot \frac{\partial f}{\partial x_i} + a_i \cdot \frac{\partial f}{\partial v_i} = 0,$$  \hspace{.5cm} (1)

where $a_i \equiv dv_i/dt$ is the acceleration due to gravity and curvilinear terms (in non-Cartesian coordinate systems). In this paper, we use Cartesian coordinates and hence $i \equiv (x, y, z)$. Equation (1) is a six-dimensional, time-dependent partial differential equation and although its direct numerical solution is possible (see, e.g. Yoshikawa et al. 2012) high resolution simulations are prohibitively expensive even for the most powerful supercomputers. N-body methods circumvent this problem by solving a system of ordinary differential equations for the dynamics of individual particles, which on galactic scales often represent a cluster of objects rather than individual physical entities. In contrast to this discrete mechanics approach, a continuous mechanics approach is based on taking the first three velocity moments of equation (1). More specifically, equation (1) is multiplied by ten quantities $m, m v_i, v_i v_j, i \equiv (x, y, z)$, and integrated over the velocity space $d^3v \equiv dv_x dv_y dv_z$ to obtain partial differential equations for the volume density $\rho = \int m f d^3v$, the mean velocity $u = \rho^{-1} \int m v d^3v$, and velocity dispersions $\sigma_{ij} = \rho^{-1} \int m f (v_i - u_i)(v_j - u_j) d^3v$ of collisionless particles.

The resulting continuity, momentum, and velocity dispersion equations, named Boltzmann moment equations (BME), are as follows,

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i} (\rho u_i) = 0,$$  \hspace{.5cm} (2)

$$\frac{\partial}{\partial t} (\rho u_i) + \frac{\partial}{\partial x_j} (\rho u_i u_j + \Pi_{ij}) - \rho a_i = 0,$$  \hspace{.5cm} (3)

$$\frac{\partial E_{ij}}{\partial t} + \frac{\partial}{\partial x_k} (E_{ij} u_k + \rho \sigma_{jk} u_i + \rho \sigma_{ik} u_j)$$

$$- \rho u_i a_j - \rho u_j a_i = 0.$$  \hspace{.5cm} (4)

Here, $E_{ij} = \Pi_{ij} + \rho u_i u_j$ and $\Pi_{ij} = \rho \sigma_{ij}^2$ is the symmetric velocity dispersion tensor. All indices obey the usual Einstein summation rule. Equations (2-4) are closed by using the usual zero-heat-flux approximation and setting the third-order moments to zero, $Q_{ijk} = \rho^{-1} \int f (v_i - u_i)(v_j - u_j)(v_k - u_k) d^3v = 0$. We discuss the applicability of this approximation in Appendix B.

Equations (2) and (3) are in fact the Jeans equations (e.g. Binney & Tremaine 1987), from which the classical hydrodynamics equation are derived assuming a diagonal and isotropic form of the stress tensor $\Pi_{ij}$ and introducing the gas pressure $P = 1/3 \sum \rho \sigma_{ii}^2$ and internal energy per unit volume $\varepsilon = 1/2 \sum \rho \sigma_{ij}^2$, where $i$ and $j$ are the number of translational and total (translational plus internal) degrees of freedom. In the Boltzmann moment equations (BME), no simplifying assumptions regarding the form of the stress tensor are made. Due to this reason, every component of $\Pi_{ij}$ needs to be evolved in time separately. Equation (4) does that for the quantity $E_{ij} = \Pi_{ij} + \rho u_i u_j$, which can be regarded as an analogue to the total energy $\varepsilon + 1/2 \rho |u|^2$ in the CHE. In total, one has 10 equations for ten unknown quantities: $\rho, u_i$, and $\rho \sigma_{ij}^2$. Note that equations (2-4) can be convolved to the usual CHE if an isotropic and diagonal form of $\Pi_{ij}$ is assumed (see Appendix A).

The fundamental similarity of the BME with the CHE makes it straightforward to couple collisionless (stars, dark matter) and collisional (gas, dust) galactic subsystems in one grid-based code. Below, we provide a detailed implementation of two numerical solvers which can accurately follow the behaviour of a collisionless fluid in the FLASH AMR code.

3 NUMERICAL SCHEMES

A full solution of the collisionless Boltzmann equation requires a detailed knowledge of the velocity phase space distribution function for a collisionless fluid. Unfortunately no known solution is currently available and therefore the first scheme we implement is one which neither requires any detailed knowledge of the systems equation of state, nor any consideration of the behaviour of the different entropy and pressure waves around cell interfaces as would be used by more sophisticated Riemann solver schemes. For this purpose we adopt the unsplit solver of Kurganov & Tadmor (2000) for hyperbolic partial differential equations (KT scheme hereafter). The scheme is second order accurate in both space and time using a simplistic Runge-Kutta midpoint technique to ensure that the solution is second order accurate in time. This removes the need to deconvolve the moment equations into their separate eigenvectors within the Riemann interaction fan. Despite being Riemann solver free, leading to an averaging over interactions within the Riemann fan, when modelling a conventional ideal gas (Kurganov & Tadmor (2000) have shown that it possess very low numerical diffusion and can accurately capture sharp shocks and contact discontinuities. This makes it an ideal means of test the reliability of higher order more sophisticated schemes and as it is also highly extensible, higher order moments of the Boltzmann equation can be readily incorporated.

The second scheme we implement involves a more accurate modified state reconstruction using a new set of eigen-
vectors so that Riemann solvers can be directly applied. Unlike the KT scheme, the Riemann solver approach achieves second order accuracy in time by reconstructing the time averaged fluid properties (primitive or characteristic) using more detailed knowledge of the eigenvector structure of the fluid equations. This grants the shock superior resolving powers, allowing sharper changes in the fluid variables to be tracked with less numerical diffusion. The scheme is also found to be relatively efficient compared to more dissipative schemes. As far as the authors are aware, this represents the first time that direct solution of the Riemann problem for an anisotropic collisionless fluid has been achieved, with previous efforts focussing on either the use of less complex schemes such as the aforementioned KT scheme or alternatively more complete direct integration of the collisionless Boltzmann equation which represents a more significant computational overhead.

3.1 The FLASH Code

FLASH was developed at the Alliance Centre for Astrophysical Thermonuclear Flashes as part of a DOE grant. Originally intended for the study of X-ray bursts and supernovae, it has since been adapted for many astrophysical conditions and now includes modules for relativistic hydrodynamics, thermal conduction, radiative cooling, magnetohydrodynamics, thermonuclear burning, self-gravity and particle dynamics via a particle-mesh approach. In this study we use FLASH version 3.2.

The code is massively parallel, scaling well over more than ten thousand of processors (Antypas 2006), and is highly modular enabling several different solvers for the hydrodynamics and stellar dynamics to be introduced.

The code features an Oct-tree structure with the simulation being divided into “blocks”, each of which contains a fixed number of cells; \(N_{\text{cells}} = 16\) in each dimension for our simulations. In addition to these cells, an additional four “guardcells” either side of the internal cells are required, which contain a copy of the properties of the neighbouring blocks in-case they are stored on a different node.

The code is capable of adaptive mesh refinement, increasing the resolution in regions where a user defined refinement criteria is satisfied and conversely removing refinement in regions where excessive refinement is both unnecessary and wasteful. Refinement is achieved by sub-dividing each “parent” block into \(2^{NDIM}\) “child” blocks. Those at the highest level of refinement for their region of space are termed “leaf blocks.” The code enforces a maximum jump of two between levels of refinement. At refinement level \(l\), a fully refined AMR grid will have a total of \(N_{\text{cells}}^{(2^l-1)}\) cells on a side.

To identify regions of rapid flow change, FLASH’s refinement and de-refinement criteria can incorporate the adapted Löhner (1987) error estimator. This calculates the modified second derivative of the desired variable, normalised by the average of its gradient over one cell. With this applied to the density as is common place, we impose maximum and minimum levels of refinement; \(l_{\text{max}}\) and \(l_{\text{min}}\).

3.2 The Kurganov and Tadmor Scheme

We implement into the FLASH AMR code the semi-implicit numerical advection scheme of Kurganov & Tadmor (2000). We briefly outline the scheme below but refer readers to their paper for a more in-depth discussion. In conjunction with the semi-implicit scheme, we use a Runge-Kutta integration scheme for the temporal integration using the following time derivative for the fluid variables, \(u\) (eqn 4.2 in Kurganov & Tadmor 2000),

\[
\frac{d}{dt} u_i(t) = -\frac{H_{i+1/2}(t) - H_{i-1/2}(t)}{\Delta x},
\]

where the numerical fluxes are defined as,

\[
H_{i+1/2} = \frac{f(u_{i+1/2}(t)) + f(u_{i+1/2}(t))}{2}
- \frac{a_{i+1/2}}{2} \left( u_{i+1/2} - u_{i+1/2} \right).
\]

Here \(u_{i-1/2}\) and \(u_{i+1/2}\) represent the fluid variable values at the left and right side of cell \(i\) with the indices + and − representing the interpolated values to the left and right sides of the respective cell interfaces. The fluxes \(f(u)\) are those taken from the derived BME 2, 3 and 4. The quantity \(a\) represents the “spectral frequency” of the system;

\[
a_{i+1/2} = \max \left\{ \rho \left( \frac{\partial f}{\partial u}(u_{i+1/2}) \right), \rho \left( \frac{\partial f}{\partial u}(u_{i+1/2}) \right) \right\}.
\]

As this is the speed with which information is propagated, it is unsurprising that this equates to the effective sound speed of the collisionless fluid. This can be computed component wise with no loss of numerical accuracy, or alternatively as we show in § 4 that it is possible to reconcile the BME with the CHE, we can use the conventional sound speed formula (8) in a directionally dependant manner. This makes use of a value for \(\gamma = 3\) and the dispersion tensor \(\Pi_{jk}\) which can be treated as a directionally dependant pressure \(P_{jk} = \Pi_{jk} = \rho \sigma^2_{jk}\).

\[
a_{jk} = \sqrt{\frac{\gamma \Pi_{jk}}{\rho}} = \sqrt{\gamma \rho \sigma^2_{jk}}.
\]

The scheme can easily be extended to arbitrary orders of temporal and spatial accuracy (see Kurganov & Levy 2004), however we choose to implement a second order accurate Runge-Kutta solver, providing a good compromise between accuracy and computational cost. Higher order schemes would require additional guardcell updates at the many midpoints, increasing communication overhead, whereas the second order accurate scheme can perform the update using just the information in locally stored guardcells.

The only modification which needs to be made in order for the scheme to function well with the adaptive mesh, is that flux conservation be checked at refinement boundaries. Fluxes calculated for higher resolution cells are considered to be more accurate than those at lower levels of refinement. Therefore the mean flux from the four higher resolution cells on a refinement boundary is calculated and passed to the neighbouring block at the coarser level of refinement to be used in place of the locally calculated flux.

© 0000 RAS, MNRAS 000, 000–000
3.3 Accurate Solution of the Riemann Problem

We implement our Riemann solver based collisionless stellar hydrodynamics module into FLASH using the MUSCL-Hancock dimensionally unsplit scheme which has already been successfully implemented for the CHE including Magneto-hydrodynamics (Lee & Deane 2009). We adopt an unsplit approach over a less memory intensive split version (e.g. PPM of Colella & Woodward 1984) given its superior ability to preserve flow symmetries. This is necessary given the highly non-linear nature of the BME which would otherwise quickly exaggerate any anisotropies. As an aside, we also note that we have significantly reduced the memory consumption of the default MUSCL-Hancock scheme provided with FLASH through more efficient memory management, ensuring that the increased memory overhead of using the unsplit scheme is equivalent to only a single extra AMR block. The results for the reduced memory version are of course bit-wise identical to the default release.

Our collisionless scheme differs from the standard hydrodynamics scheme primarily through our determination of the new eigenvalues and eigenvectors which are required for the characteristic tracing step. This allows for the calculation of the time averaged fluid properties which can then be used to calculate the second order accurate fluxes using a conventional Riemann solver such as the approximate HLLC solver (see, e.g. Toro 1996; Toro, Spruce & Speares 1994; Quirk 1994) or more accurate Roe scheme (Roe 1981).

We begin by expressing the zeroth, first and second order moments of the Boltzmann equation (see equations (2) and (3) in compact matrix notation form (neglecting gravity for the moment),

$$U_t + F(U)_x + G(U)_y + H(U)_z = 0,$$

with the conserved advection quantities \( U \) and their corresponding fluxes \( F(U) \), \( G(U) \) and \( H(U) \) in the respective \( x \), \( y \) and \( z \)-directions being,

$$U = \begin{bmatrix} \rho \\ \rho v_x \\ \rho v_y \\ \rho v_z \\ E_{xx} \\ E_{xy} \\ E_{yz} \\ E_{zz} \end{bmatrix}, \quad F(U) = \begin{bmatrix} \rho v_x \\ \rho v_x v_x + \sigma_{xx}^2 \\ \rho v_x v_y \\ \rho v_x v_z \\ E_{xx} v_x + 2\sigma_{xx}^2 v_x \\ E_{xy} v_x \\ E_{xz} v_x \\ E_{zz} v_x \end{bmatrix},$$

$$G(U) = \begin{bmatrix} \rho v_y \\ \rho v_x v_y \\ \rho v_y v_y + \sigma_{yy}^2 \\ \rho v_y v_z \\ E_{xy} v_y \\ E_{yy} v_y \\ E_{yz} v_y + 2\sigma_{yy}^2 v_y \\ E_{zz} v_y + 2\sigma_{yy}^2 v_y \end{bmatrix}, \quad H(U) = \begin{bmatrix} \rho v_z \\ \rho v_x v_z \\ \rho v_y v_z \\ \rho v_z v_z + \sigma_{zz}^2 \\ E_{xz} v_z \\ E_{yz} v_z + 2\sigma_{zz}^2 v_z \\ E_{zz} v_z + 2\sigma_{zz}^2 v_z \end{bmatrix}$$

where \( E_{jk} = \rho v_j^2 + \sigma_{jk}^2 \). For simplicity, in this paper we assume a diagonalised velocity dispersion tensor in the above flux vectors, neglecting off diagonal terms, i.e. \( \sigma_{jk} = 0 \) \( \forall j \neq k \). As the dimensions of the matrix and therefore complexity of the problem grow quickly with the increasing order of moments that we consider, we leave the tracking of the off diagonal terms and higher order moments to our next paper.

By taking the conservative equations and expanding the derivatives, it is possible to express equivalent advection equations for the primitive quantities; density, bulk velocity and velocity dispersion:

$$W = \begin{bmatrix} \rho \\ v_x \\ v_y \\ \rho \sigma_{xx}^2 \\ \rho \sigma_{yy}^2 \\ \rho \sigma_{zz}^2 \end{bmatrix}, \quad \tilde{F}(W) = \begin{bmatrix} \rho v_x \\ v_x v_x + \sigma_{xx}^2 \\ v_x v_y \\ \rho v_x v_y + \gamma \sigma_{yy}^2 v_x \\ v_x v_z + \gamma \sigma_{zz}^2 v_x \\ \rho v_x v_z + \gamma \sigma_{zz}^2 v_x \end{bmatrix},$$

$$G(W) = \begin{bmatrix} \rho v_y \\ v_y v_y + \sigma_{yy}^2 \\ v_y v_z + \gamma \sigma_{zz}^2 v_y \\ \rho v_y v_z + \gamma \sigma_{zz}^2 v_y \\ \rho v_z v_z + \gamma \sigma_{zz}^2 v_z \end{bmatrix}, \quad H(W) = \begin{bmatrix} \rho v_z \\ v_x v_z + \gamma \sigma_{zz}^2 v_x \\ v_y v_z + \gamma \sigma_{zz}^2 v_y \\ \rho v_z v_z + \gamma \sigma_{zz}^2 v_z \end{bmatrix}$$

where \( \gamma = 3 \) (see § 4).

For brevity we now consider just the \( x \)-directional flux vectors in the conserved quantities, \( F(U) \) and primitive variables, \( \tilde{F}(W) \). As the components of the primitive flux vectors \( f_i \) are each functions of the individual primitive quantities \( u_i \), the Jacobian of the system in the \( x \)-direction, \( \tilde{A}(W)_x \), can be expressed in terms of the partial derivatives of \( \tilde{F}(W) \) with respect to \( W \):

$$\tilde{A}(W) = \partial \tilde{F}/\partial W = \begin{bmatrix} \partial \tilde{f}_1/\partial w_1 & \cdots & \partial \tilde{f}_3/\partial w_1 \\ \vdots & \ddots & \vdots \\ \partial \tilde{f}_m/\partial w_1 & \cdots & \partial \tilde{f}_n/\partial w_n \end{bmatrix}.$$

The eigenvalues \( \lambda_i \) of the system are then obtained by solving for the roots of the polynomial,

$$|\tilde{A} - \lambda I| = 0,$$

where \( I \) is the identity matrix. Physically, the eigenvalues represent the speeds of signal propagation within the system and are thus important for determining the extent to which fluid bounding a cell interface affects the flux through it over the time step. Unlike a collisionally dominated fluid which has five eigenvalues, the (diagonalised) collisionless fluid has seven to account for the anisotropy in the dispersion tensor (For a non-diagonalised version, this increases to ten). We determine the eigenvalues for the \( x \)-direction to be

$$\lambda = \left[ v_x - a_{xx}, v_x, v_x, v_x, v_x, a_{xx}, v_x, v_x, v_x \right]^T,$$

with \( a_{xx} \) being the sound speed in the \( x \)-direction. The directionally dependant sound speed, \( a_{jk} \), is in close analogy to that of a standard collisionally dominated gas,

$$a_{jk} = \sqrt{\frac{\gamma P_{jk}}{\rho}} = \sqrt{\gamma \sigma_{jk}^2 \rho}.$$

The right eigenvectors of the Jacobian, \( \tilde{R}^{(i)} \), are those that obey the relation \( \tilde{A} \tilde{R}^{(i)} = \lambda_i \tilde{R}^{(i)} \) whilst the left eigenvectors of the Jacobian, \( \tilde{L}^{(i)} \), are those which satisfy \( \tilde{L}^{(i)} \tilde{A} = \lambda_i \tilde{L}^{(i)} \). Using the Jacobian for the primitive variable matrix equations,
\[ \mathbf{A}(\mathbf{W})_x = \begin{bmatrix} v_x & 0 & 0 & 0 & 0 & 0 \\ 0 & v_x & 0 & 0 & 0 & 0 \\ 0 & 0 & v_x & 0 & 0 & 0 \\ 0 & 0 & 0 & v_x & 0 & 0 \\ 0 & \rho a_{xx}^2 & 0 & 0 & v_x & 0 \\ 0 & 0 & 0 & 0 & 0 & v_x \end{bmatrix}, \quad (16) \]

and the equivalent eigenvalues \( \lambda_i \), the eigenvector matrices are found to be:

\[ \mathbf{R}_x = \begin{bmatrix} \rho & 0 & 0 & 1 & 0 & 0 \\ -a_{xx} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\rho} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\rho} & 0 \\ \rho a_{xx}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\rho} & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (17) \]

\[ \mathbf{L}_x = \begin{bmatrix} 0 & \frac{1}{a_{xx}} & 0 & 0 & \frac{1}{a_{xx}} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho \\ 0 & 0 & -\rho & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (18) \]

With the general form of the eigenvalues and corresponding eigenvector calculated, we can implement them into the MUSCL-Hancock scheme in-line with the approach taken by \cite{Stone2008} which we outline below. First we compute the left, right and central differences of the primitive variables using the cell centred values within each cell, where \( i \) is the cell index.

\[ \delta w_{L,i} = w_i - w_{i-1}, \quad (19) \]

\[ \delta w_{C,i} = \frac{(w_{i+1} - w_{i-1})}{2}, \quad (20) \]

\[ \delta w_{R,i} = w_{i+1} - w_i. \quad (21) \]

We then map them onto the characteristic variables using the left eigenvectors calculated using \( \mathbf{L}(\mathbf{w}) \), where \( \mathbf{L}(\mathbf{w}) \) is the matrix corresponding to the \( i^{th} \) row of the left eigenvector matrix, and the primitive variables \( w_i \) are defined at the cell centres,

\[ \delta \mathbf{u}_{L,i} = \mathbf{L}(\mathbf{w}_i) \cdot \delta \mathbf{w}_{L,i}, \quad (22) \]

\[ \delta \mathbf{u}_{C,i} = \mathbf{L}(\mathbf{w}_i) \cdot \delta \mathbf{w}_{C,i}, \quad (23) \]

\[ \delta \mathbf{u}_{R,i} = \mathbf{L}(\mathbf{w}_i) \cdot \delta \mathbf{w}_{R,i}. \quad (24) \]

As the scheme is higher than first order, the reconstruction of the time averaged variables at the cell interfaces, can lead to the formation of new non-physical local maxima or minimum. To prevent their formation which otherwise lead to spurious post shock oscillations, we apply monotonicity constraints to the characteristic differences, ensuring that the time averaged variables are total variation diminishing (TVD, see \cite{LeVeque2002}),

\[ \delta \mathbf{u}_i^m = \text{sgn}(\delta \mathbf{u}_{C,i}) \min(2|\delta \mathbf{u}_{L,i}|, |\delta \mathbf{u}_{C,i}|, 2|\delta \mathbf{u}_{R,i}|). \quad (25) \]

The monotonised differences are then projected back onto the primitive variables, where \( \tilde{\mathbf{R}}(\mathbf{w}_i) \) is the matrix corresponding to the \( i^{th} \) column of the right eigenvector matrix \( \mathbf{R} \),

\[ \delta \mathbf{w}_i^m = \delta \mathbf{u}_i^m \cdot \tilde{\mathbf{R}}(\mathbf{w}_i). \quad (26) \]

We now calculate the values of the primitives at the left, \( i - 1/2 \), and right, \( i + 1/2 \), cell interfaces using the primitive variable monotonised differences,

\[ \tilde{w}_{L,i+1/2} = w_i + \left[ \frac{1}{2} - \max(\lambda_i^M, 0) \right] \frac{\delta t}{2 \delta x} \delta \mathbf{w}_i^m, \quad (27) \]

\[ \tilde{w}_{R,i-1/2} = w_i - \left[ \frac{1}{2} - \max(\lambda_i^0, 0) \right] \frac{\delta t}{2 \delta x} \delta \mathbf{w}_i^m, \quad (28) \]

where \( \lambda_i^M \) and \( \lambda_i^0 \) are the largest and smallest eigenvalues based upon the cell centred data.

We can now perform the characteristic tracing step, removing from the quantities calculated in equations \( 27 \) and \( 28 \), contributions from each set of waves that travelled towards the interface but did not reach it during the forward half-time-step \( \delta t/2 \) using \cite{Colella1984, Colella1990}.

\[ \tilde{w}_{L,i+1/2} = \tilde{w}_{L,i+1/2} + \frac{\delta t}{2 \delta x} \sum_{\lambda_i^M \leq 0} \left( \lambda_i^M - \lambda_i^0 \right) \mathbf{L}^\alpha \cdot \delta \mathbf{w}_i^m \tilde{R}_i^\alpha, \quad (29) \]

\[ \tilde{w}_{R,i-1/2} = \tilde{w}_{R,i-1/2} + \frac{\delta t}{2 \delta x} \sum_{\lambda_i^0 < \lambda_i^L \leq 0} \left( \lambda_i^L - \lambda_i^0 \right) \mathbf{L}^\alpha \cdot \delta \mathbf{w}_i^m \tilde{R}_i^\alpha. \quad (30) \]

With time averaged primitive values calculated for each cell face, Godunov’s fluxes can now be computed and applied using a conventional Riemann solver technique such as the accurate Roe scheme. If using an approximate Riemann solver such as the HLL scheme, then due to the averaging of the intermediate eigenstates between the fastest left and right waves, it is necessary to add the additional terms \( 31 \) and \( 32 \) shown below, to equations \( 29 \) and \( 30 \) respectively.

\[ \tilde{w}_{L,i+1/2} = \tilde{w}_{L,i+1/2} - \frac{\delta t}{2 \delta x} \sum_{\lambda_i^L \leq 0} \left( \lambda_i^L - \lambda_i^M \right) \mathbf{L}^\alpha \cdot \delta \mathbf{w}_i^m \tilde{R}_i^\alpha, \quad (31) \]

\[ \tilde{w}_{R,i-1/2} = \tilde{w}_{R,i-1/2} - \frac{\delta t}{2 \delta x} \sum_{\lambda_i^0 \leq \lambda_i^L} \left( \lambda_i^L - \lambda_i^0 \right) \mathbf{L}^\alpha \cdot \delta \mathbf{w}_i^m \tilde{R}_i^\alpha. \quad (32) \]

These remove the effect of waves which propagate away from the cell interface which would otherwise be included, allowing for the scheme to achieve second order accuracy.

Through the use of operator splitting, we can deal with the addition of extra source and sink terms, including gravity, separately after the collisionless stellar hydrodynamics step. To ensure that the gravitational acceleration term is kept second order accurate in time, we store the acceleration from the previous time step and interpolate forward to account for temporal variation in the gravitational field. Additional source and sink terms can include star formation and supernova, although we only consider gravity in this paper.
4 TESTS

Notwithstanding a formal similarity of the CHE and BME, testing the latter is nontrivial as it requires knowing analytical solutions for the dynamics of collisionless systems. Fortunately, some of the test problems that are used to benchmark the performance of hydro codes can be adapted to test the collisionless systems. This is because equations (2)-(4) can be made formally identical to those of classical hydrodynamics in special cases. Indeed, for a one-dimensional flow of a non-gravitating collisionless fluid along the x-direction, the BME become

\[ \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (\rho u_x) = 0, \]  
\[ \frac{\partial}{\partial t} \rho u_x + \frac{\partial}{\partial x} (\rho u_x u_x + \rho \sigma_{xx}) = 0, \]  
\[ \frac{\partial E_{xx}}{\partial t} + \frac{\partial}{\partial x} \left[ (E_{xx} + 2\rho \sigma_{xx}) u_x \right] = 0. \]

If we now assume that \( \rho \sigma_{xx}^2 \) is equal to the gas pressure \( P \) in classic hydrodynamics and \( \varepsilon = \frac{1}{2} \rho \sigma_{xx}^2 \), then \( P = (\gamma - 1)\varepsilon = 2\varepsilon \), i.e. the ratio of specific heats \( \gamma \) is equal to 3, then equations (33)-(35) turn into the following form

\[ \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (\rho u_x) = 0, \]  
\[ \frac{\partial}{\partial t} \rho u_x + \frac{\partial}{\partial x} (\rho u_x u_x + \rho P) = 0, \]  
\[ \frac{\partial}{\partial t} \left( \varepsilon + \frac{1}{2} \rho u_x^2 \right) + \frac{\partial}{\partial x} \left[ \left( \varepsilon + \frac{1}{2} \rho u_x^2 + P \right) u_x \right] = 0, \]

which are exactly one-dimensional classic hydrodynamics equations. The fact that the BME for a one-dimensional flow turns into the CHE for \( \gamma = 3 \) is not a coincidence but is related to the fact that a one-dimensional flow of a collisionless fluid has only one translational degree of freedom \( i = 1 \), unlike a collisional fluid which can still be characterised by \( i = 3 \) due to frequent collisions that equipartition the kinetic energy of particles between the three translational degrees of freedom. This means that in order to properly describe the one-dimensional flow of a collisionless fluid using the classical hydrodynamics approach, one needs to assume that the fluid has only one translational degree of freedom, i.e. \( \gamma = (i + 2)/i = 3 \). It is straightforward to show that for a two-dimensional flow of a collisionless fluid, one needs to consider a specific case with \( \rho \sigma_{xx}^2 = P \) and \( \rho \sigma_{yy}^2 = P \) and also \( \gamma = 2 \), i.e. a flow of a collisional fluid with two degrees of freedom.

4.1 Sod shock tube

The above analysis allows us to use the standard Sod shock tube problem, often employed to benchmark collisional hydrocodes. To generate an analytic solution to a one-dimensional discontinuous state the x-coordinate, we choose \( \gamma = 3 \) and use the standard solution procedure (Sod 1978). For a two-dimensional discontinuous state in the \( x-y \) plane at an arbitrary angle \( \alpha \) to the x-direction, an analytic solution is obtained by choosing \( \gamma = 2 \).

In the 1D case, we run the test along the x-coordinate and set \( \rho \sigma_{xx}^2 = \rho \sigma_{yy}^2 = P \) at \( x \in [0 - 0.5] \) to 1.0, while at \( x \in [0.5 - 1.0] \) the x-component of the stress tensor is set to 0.1 and the stellar density is 0.125. The numerical resolution is 200 grid zones and the cells are equidistantly spaced. In the 2D case, we align the initial fluid discontinuity along the \([1,0,0] \) plane and use the same densities and velocity dispersions as were set for the 1D setup setting \( \rho \sigma_{xx}^2 = \rho \sigma_{yy}^2 \) in order to yield agreement with the results for an isotropic collisionally dominated gas. This yields a shock wave which propagates at \( 45^\circ \) to the \( x \) and \( y \) axes. Again we use 200 cells in the \( x \) and \( y \) directions and note that the setup can easily be extended to three dimensions although for brevity we do not show these here.

Figures 2 and 3 show the results of our tests for the 1D and 2D setups, respectively, at \( t=0.1 \) using the HLLC Rie-
The latter two quantities may be regarded as analogues to the \( x \)-component of the stress tensor \( \sigma_{xx} \) and \( \sigma_{xx}/(\gamma - 1) \). The latter two quantities may be regarded as analogues to the gas pressure \( P \) and internal energy density \( \varepsilon \) in classical hydrodynamics. In the 1D case, the plotted velocity is \( u_x \), whereas in the 2D case we plot the absolute value of velocity \( (u_x^2 + u_y^2)^{1/2} \) along the diagonal \( x = y \), i.e., along the direction of the shock wave propagation. It is evident that the Riemann-solver-based numerical scheme can trace the shock waves (propagating to the right) very well, usually within one to two grid zones. Contact discontinuities (propagating to the left) are smeared to a greater extent, usually by 5-6 grid zones. Nevertheless, the position of the shock waves and contact discontinuities along with the values of the flow variables are reproduced quite accurately.

In a conventional collisional shock, the energy of the macroscopic motion of the gas is transferred through collisions into random isotropic kinetic motions of the particles (thermal energy). In a collisionless fluid, the energy of macroscopic motion is transferred into an effective increase in the velocity dispersion as stellar streams with different velocities mix at the shock front.

Open circles in Figure 4 present the results of the 1D Sod shock tube test using the KT scheme. It is evident that this scheme does not provide as good an agreement with the analytic solution as the HLLC Riemann solver, particularly for the velocity dispersion (bottom-right panel). In the KT scheme, the shock fronts and contact discontinuities are smeared out over roughly twice as many grid zones as in the HLLC scheme. However encouragingly, the positions of the shocks and discontinuities and also the values of the flow variables are still accurately reproduced. All in all, the HLLC scheme shows an undoubtedly superior performance though at a higher CPU cost, with the KT scheme typically taking around 22% less time to run and a near identical amount of memory.

4.2 Collapse of a pressure-free sphere

The gravitational collapse of a pressure-free sphere is used to assess the code’s ability to accurately treat converging spherical flows on the Cartesian grid. This test is also useful for estimating the performance of the gravitational potential solver on dynamical problems for which one has to use a finite-difference form of the gravitational potential gradient. Since the test setup involves no stress tensor, the resulting equations are identical in the collisionless and collisional fluid dynamics cases and one can use an analytic solution describing the collapse of every mass shell in the limit of an infinite sphere radius (Hunter 1962).

To run this test, we set a cold homogeneous sphere of unit radius and density (for convenience, the gravitational constant is also set to unity) and let it collapse under its own gravity. We use a block size of \( 16^3 \) cells and six levels of refinement leading to an effective grid size of \( 512^3 \) cells. We adopt isolated gravitational boundary conditions. Unfortunately, we must consider a cloud of finite radius in Cartesian geometry with a sharp outer boundary to preserve the cloud sphericity. As a result, a rarefaction wave develops after the onset of the collapse, propagating towards the coordinate origin and necessitating complicated corrections to the analytic solution.

Figure 5 compares the results of our numerical simulation with the “uncorrected” analytic solution of Hunter (1962) (solid line) at 0.985 free fall timescales, when the initial density has increased by nearly three orders of magnitude. The majority of cells within the sphere lie upon the analytic solution indicating a good homologous collapse with only a small peak above the analytic solution forming within the very central few cells. In addition to this small peak, the initially sharp boundary of the cloud is smeared out over several cells as a result of the rarefaction developing. The analytic radii of the cloud however shows reasonable agreement with the radii at which the half peak density is reached. In general, our results are found to be in strong agreement with those of other authors applying this test problem to the classic hydrodynamics (e.g., Stone & Norman 1992).
4.3 Collapse of a sphere with anisotropic stress tensor

As a third test, to demonstrate that the code can model the formation of anisotropies, we initialise two copies of the same sphere as in §4.2 with a non-negligible dispersion tensor. In one we initially set an isotropic dispersion tensor whilst in the second we initially weight the different diagonal terms $\sigma_{xx}^2 : \sigma_{yy}^2 : \sigma_{zz}^2$ with the ratio of 1 : 2 : 3. We set the system up so that the initial maximum effective dispersion energy is only 25% of that needed to support the sphere against gravitational collapse. This allows the sphere to collapse down from its initially symmetric configuration and deform into an ellipsoid. The velocity dispersion is then free to change as the simulation evolves.

We plot in figure 6 the density contour for $\rho = 0.9\rho_{max}$ at regular time intervals for both the cloud with an isotropic dispersion tensor (left hand panel) and an anisotropic dispersion tensor (right hand panel). The isotropic dispersion tensor leads to a collapse which is isotropic as expected whilst in the anisotropic case, the collapse proceeds quickest in the direction with the lowest dispersion term. This leads to a flattening of the sphere into an ellipse.

In figure 7 we plot the ellipticity of the collapsing anisotropic cloud as a function of time. The ellipticity is determined to be the ratio between the minor and major axes of the ellipsoid over a specified iso-density surface. The solid line corresponds to the ellipticity of the iso-density surface of $0.9\rho_{max}$ and the dashed line is the ellipticity for $0.5\rho_{max}$.

Initially, since the velocity dispersion is lowest in the $x$-direction, material along that axis collapses fastest, with the ellipticity increasing as the density builds. This leads to a greater effective pressure, $P_{xx} \equiv \rho \sigma_{xx}^2$, from the velocity dispersion. Eventually the core collapses to a density and pressure sufficient to stop further collapse. At this point the sphere begins to isotropise as material from the other two axial directions catches up and falls onto the core. This collapse of material onto the core, drives up the pressure and material is forced out around $t \approx 0.9t_{ff}$ along the $x$-direction where least resistance is offered. This again increases the ellipticity and results in the major and minor axes of the ellipse switching places as the $x$-axis becomes the major instead of the minor axis and vice versa for the $z$-axis. The system then oscillates between a high degree of ellipticity and near spherical symmetry as the different axes of the ellipsoid collapse and rebound at different times and speeds. These oscillations repeat over many free fall timescales. Material at a lower density follows the same cyclical pattern of varying ellipticity but with less pronounced maxima and minima. As a result of the non-linear nature of the collapse and its relative separation from the violent oscillations within the core, transitions at lower density are also smoother.

4.4 Spiral instability in a stellar disc

The final test problem describes the growth of a spiral structure in a gravitationally unstable stellar disc. The stability analysis of a thin stellar disc to a local axisymmetric perturbation states that the disc is gravitationally unstable if (Toomre 1964)

$$Q_T = \frac{\sigma_{vz} \kappa}{3.366G\Sigma} < Q_c = 1.0,$$  \hspace{1cm} (39) 

where $\sigma_{vz}$ is the radial stellar velocity dispersion, $\kappa$ is the epicycle frequency, and $\Sigma$ is the stellar surface density. A finite disc thickness and non-axisymmetric perturbations may increase the critical Toomre parameter $Q_c$ by a factor of unity.

The initial setup consists of a self-gravitating, rotating stellar disc submerged in a fixed dark matter (DM) halo. The latter is described by the quasi-isothermal density profile of the form

$$\rho_{DM} = \frac{\rho_0}{1 + (\varpi/r_0)^2}.$$  \hspace{1cm} (40) 

where $\varpi$ is the galactocentric distance. The central density...
$\rho_0 = 1.24 \times 10^{-2} M_\odot \text{ pc}^{-3}$ and the characteristic scale length of the quasi-isothermal halo $r_0 = 3.2$ kpc are calculated based on the assumed DM halo mass $M_{\text{DM}} = 2 \times 10^{11} M_\odot$ and halo radius, $r_h = 124.7$ kpc, using relations provided in Vorobyov et al. (2012).

The computational box spans 30 kpc in each coordinate direction ($x, y, z$). The numerical procedure for generating a rotationally supported stellar disc in the combined gravitational potential of stars and DM halo is described in detail in Vorobyov et al. (2012). The rotation curve, the radial gas surface density profile and the radial profile of the Toomre Q parameter $Q_T$ of the initial stellar disc are plotted in Fig. 8 with the solid, dashed, and dot-dashed lines, respectively. Evidently, the disc is initially gravitationally unstable.

Figure 9 presents a series of disc images at several successive times since the beginning of numerical simulations. Usually, an initial seed perturbation is required to drive the system out of equilibrium and initiate the growth of a spiral structure. However, in our case, the initial perturbation is introduced due to a re-map of the initial equilibrium configuration from the cylindrical coordinates (Vorobyov et al. 2012) onto the Cartesian coordinates (the FLASH code). The growth of a two-armed spiral mode is apparent in the figure.

To quantify the growth rate of spiral modes in our model, we employ global Fourier amplitudes defined as

$$A_m(t) = \frac{1}{M_\text{d}} \left| \int_{-X}^{+X} \int_{-Y}^{+Y} \Sigma(x, y, t) e^{im\phi} \, dx \, dy \right|, \quad (41)$$

where $m$ is the spiral mode, $\Sigma(x, y, t)$ is the surface density of the stellar disc, $\phi = \arctan(y/x)$ is the polar angle, and $M_\text{d}$ is the total mass of the stellar disc. The amplitudes are calculated on a square region with size $(-X : X, -Y : Y) = (-2 : 2, -2 : 2)$ kpc, which is centred on the coordinate origin and encompasses the growing spiral. The global Fourier amplitudes can be regarded as the amplitude of a spiral density perturbation relative to the surface density of the axisymmetric disc.

Figure 10 shows the temporal evolution of the first six global Fourier amplitudes (in log units) in our model stellar disc. Initially all amplitudes are negligibly small except for that corresponding to the $m = 4$ mode. This is a numerical artifact and arises due to the finite resolution within the core. Similar to the results of Tasker et al. (2008) in which they model a static King-profile that is initially in hydrostatic equilibrium, the finite resolution within the core leads to an under-resolved gravitational potential. This results in a slight relaxation and expansion of the very central region. An outward flow of material results through the faces of the central most cell and due to the fact that our disk is
constructed so as to promote the growth of gravitational instabilities, this generates a small yet non-negligible perturbation. Although similar behaviour may occur in all mesh codes, those using Cartesian geometries are particularly susceptible given their strong natural bias for flows along the axes of the mesh. Despite being present from the very onset of the simulation, the $m = 4$ mode slowly decays and does not lead to the growth of a spurious four armed spiral given the dominance of the $m = 2$ mode at late times by around two orders of magnitude.

The $m = 2$ mode demonstrates the fastest growth rate. It saturates at around $t \approx 370$ Myr and declines slowly in the subsequent evolution due to continuing disc heating. The maximum amplitude of the $m = 2$ mode is about $10^{-1.73}$, which suggests a non-axisymmetric density perturbation of about 2% relative to the underlying axisymmetric disc.

An appealing physical interpretation for the growth of a spiral structure in our model disc is swing amplification. Amplification occurs when any leading spiral disturbance, e.g., introduced by the initial seed perturbation, unwinds into a trailing one due to differential rotation (e.g., Toomre 1981; Athanassoula 1986; Vorobyov & Theis 2006, 2008). In addition, swing amplification needs a feedback mechanism that can constantly feed a disc with leading spiral disturbances. Trailing short-wavelength disturbances propagating through the disc centre and emerging on the other side as leading ones provide a feedback for the swing amplifier (Toomre 1981).

We can check if the growth mechanism of our spiral structure is consistent with the predictions of the swing amplification theory. For our spiral pattern to be triggered by the swing amplifier, there should be no inner Lindblad resonance (ILR) for the $m = 2$ mode. Otherwise, the corresponding trailing disturbances will damp at the ILR, thus failing to pass through the disc centre and promoting the growth of the $m = 2$ mode. Figure 11 shows the radial profiles of $\Omega$ and $\Omega \pm \kappa/m$ at $t = 330$ Myr, where $\Omega$ is the angular velocity of the stellar disc and $\kappa$ is the epicycle frequency. Both quantities are the azimuth averages. In particular, the solid and dashed lines show the corresponding values for the $m = 2$ and $m = 3$ modes, respectively. The dash-dot-dot-dotted and dash-dotted lines show the angular velocity $\Omega_p$ of the global spiral pattern measured at two distinct positions in the disc: 2.0 kpc and 4.0 kpc. In theory, $\Omega_p$ should be independent of radius, though in practice it is always slightly faster in the inner regions than in the outer ones due to gradual winding of a spiral pattern. The radial position of Lindblad resonances are determined as $m(\Omega - \Omega_p) = \pm \kappa$, i.e., as the radial distance where the forcing frequency of the spiral pattern coincides with the epicycle frequency of the stars. In particular, the inner and outer Lindblad resonances correspond to the plus and minus signs, respectively.

Figure 11 demonstrates that there is clearly no ILR for the $m = 2$ mode, whereas there is one marginally for the $m = 3$ mode (and thus for any higher-order mode). Indeed, the dash-dotted and dash-dot-dot-dotted lines never intersect the $\Omega - \kappa/2$ curve, whereas they do intersect the $\Omega - \kappa/3$ curve at $r < 1$ kpc, thus damping $m = 3$ disturbances that try to pass through the disc center. Although slight growth over time of the $m = 3$ mode is observed in figure 10 it is damped relative to the $m = 2$ mode by over two orders of magnitude. Higher order modes for which a clearer ILR will be present are damped to a much greater extent. This explains why the $m = 2$ mode (and not any higher-order mode) ultimately dominates the spiral pattern.

Another consistency check on our model is the position of the outer Lindblad resonance (OLR). Spiral disturbances cannot propagate through the OLR, which effectively limits the radial extent of a spiral pattern. Figure 11 indicates that the OLR for the $m = 2$ mode is located approximately at 4 kpc and this value is in agreement with the radial extent of the spiral pattern in Fig. 3. We conclude that our numerical results are in general agreement with the predictions of the swing amplification theory.

5 DISCUSSION AND SUMMARY

In this paper we have outlined a series of important limitations that are encountered in mesh codes when using the particle-mesh technique. Used to consistently integrate collisionless stellar or dark matter components into simulations which model gas physics on an Eulerian mesh, the particle-mesh technique has the advantage of being relatively straightforward to implement but is unfortunately subject to several key limitations. From a technical perspective these include poor load balancing and increased communication overhead. From a physical point of view, discontinuity effects incurred when discrete particle properties are mapped to the mesh, are believed to result in spuriously entropy generation. In galaxy formation simulations, this can have a significant influence on the reservoir of cold gas available to form stars, potentially changing the entire dynamics of the system.

To overcome these limitations we have proposed the use of the collisionless Boltzmann moment equations as a powerful alternative. Such an approach allows us to model the collective properties of collisionless objects such as stars as a fluid, instead of relying upon the traditional N-body approach. In this paper we refer to this approach as “colli-
Through the use of a spherical pressure-free collapse problem, we confirm that the code can maintain good flow symmetry in convergent flows in the absence of any anisotropy in the dispersion tensor. Agreement with the peak density within the collapsed sphere also confirms the validity of the Poisson solver used in FLASH. Through the inclusion of a non-negligible anisotropic dispersion tensor, we extend the spherical collapse to observe the behaviour of an initially homologous spherical cloud as it collapses and deforms into an elliptical profile. Although initial behaviour of the system occurs in-line with expectations, the long term behaviour is more complicated with repeated transitions between an elliptical and spherical profile, along with switching of the major and minor axes after the initial core implosion. Although the inclusion of gas physics may help to isotropise the internal structure of the cloud and damp those oscillations, they will nevertheless remain important to the dynamics outside of the core. We will explore in more detail interactions between the collisionally dominated gas physics and the stellar material in future papers. In particular, we will extend our scheme to include off-diagonal terms and a more thorough exploration of the effects of higher order terms. Although powerful in its current form, the addition of off-diagonal terms in the velocity dispersion tensor will allow us to directly measure the vertex deviation, a key observable used for probing galactic structure. To date observers have made extensive measurements of our own Milky Way’s vertex deviations which can hide detailed information about the kinematic properties of the bar and spiral arms. However noise in particle based simulations has made accurate theoretical predictions difficult, something which our new collisionless stellar hydrodynamics code can allow us to overcome. Having already confirmed the applicability of the collisionless stellar hydrodynamics code to the formation of spiral structure, we have found excellent agreement between our numerical simulations of the growing spiral pattern and the predictions of the swing amplification theory, with the correct relation between the disc size and the outer Lindblad resonances being observed.

As a final note, we find the time taken in FLASH for the Riemann solver based collisionless stellar hydrodynamics routine is only 7% greater than that for the original classic hydrodynamic scheme, allowing the collisionless stellar hydrodynamics to scale linearly with the classic hydrodynamics. We also find that instead of doubling the overall amount of time spent communicating when both gas and collisionless stellar hydrodynamics are included, our grouping of communications results in a net increase of only 1.2% to the overall communication time compared to that of just a single phase hydrodynamic scheme. Thus we conclude by confirming that our new collisionless stellar hydrodynamic approach using the Boltzmann moment equations both preserves the excellent scaling previously demonstrated in FLASH as well as enhancing the level of detail we can expect to be able to extract from our results, with none of the discreteness effects observed for particle-mesh techniques.

ACKNOWLEDGEMENTS

The authors acknowledge the generous support of the Austrian FWF for project numbers P21097-N16 (Nigel L.
APPENDIX A: CONVOLVING THE BOLTZMANN MOMENT EQUATIONS TO THE CLASSIC HYDRODYNAMICS FORM

On general grounds one can argue that the Boltzmann moment equations (2)-(4) should convolve to the classic hydrodynamics equations (CHE) if the stress tensor $\Pi_{ij}$ is diagonal and isotropic. In this section we demonstrate this by assuming that $\Pi_{ij} = 0$ for $i \neq j$ and $\Pi_{xx} = \Pi_{yy} = \Pi_{zz}$, implying that $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma$. The equation of continuity (2) is identical to that of the CHE and the momenta equations (3)-(4) should convolve to the classic hydrodynamics equation (5) if the stress tensor $\Pi_{ij}$ is diagonal and isotropic. In this section we demonstrate this by assuming that $\Pi_{ij} = 0$ for $i \neq j$ and $\Pi_{xx} = \Pi_{yy} = \Pi_{zz}$, implying that $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma$. The equation of continuity (2) is identical to that of the CHE and the momenta equations (3)-(4) should convolve to the classic hydrodynamics equation (5) if the stress tensor $\Pi_{ij}$ is diagonal and isotropic.

After summing up equations (A1)-(A3) we obtain

\[
\begin{align*}
\frac{\partial E_{xx}}{\partial t} + \frac{\partial}{\partial x_k} \left( E_{xx} u_k + 2 \rho \sigma_{xx} u_k^2 \right) - 2 \rho u_k a_k = 0 , \\
\frac{\partial E_{yy}}{\partial t} + \frac{\partial}{\partial y_k} \left( E_{yy} u_k + 2 \rho \sigma_{yy} u_k^2 \right) - 2 \rho u_y a_y = 0 , \\
\frac{\partial E_{zz}}{\partial t} + \frac{\partial}{\partial z_k} \left( E_{zz} u_k + 2 \rho \sigma_{zz} u_k^2 \right) - 2 \rho u_z a_z = 0 .
\end{align*}
\]

Then, we obtain

\[
\begin{align*}
\frac{\partial}{\partial t} \left( \frac{3}{2} \rho \sigma^2 \right) + \frac{1}{2} \rho |u|^2 &+ \frac{\partial}{\partial x_k} \left[ u_k \left( \frac{3}{2} \rho \sigma^2 + \frac{1}{2} \rho |u|^2 \right) + \rho \sigma_{ik}^2 u_i \right] \\
- \rho u_i a_i = 0 ,
\end{align*}
\]

where we have divided the resulting equation by a factor of 2 and have taken into account that $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma$ and $|u|^2 = u_x^2 + u_y^2 + u_z^2$. In the classical hydrodynamics the gas pressure relates to the internal energy density via the relation $P = \varepsilon (\gamma - 1)$, where $\gamma = 5/3$ for a flow with three translational degrees of freedom, meaning that $\frac{4}{3} \rho \sigma^2 = \varepsilon$ in the above equation. The final step is to note extract the diagonal part from the symmetric tensor $\rho \sigma_{ik}^2$ by introducing the viscous stress tensor $\pi_k$. \[
\rho \sigma_{ik}^2 = P \delta_{ik} - (P \delta_{ik} - \rho \sigma_{ik}^2) = P \delta_{ik} - \pi_{ik} .
\]

With this final transformation, equation (A4) turns into the usual hydrodynamics equation for the total energy per unit volume $E_T$ of a viscous fluid.

\[
\frac{\partial E_T}{\partial t} + \frac{\partial}{\partial x_k} \left[ u_k (E_T + P) - \pi_{ik} u_i \right] - \rho u_i a_i = 0 .
\]
