Smoothed Particle Hydrodynamics in Astrophysics

Volker Springel

Max-Planck-Institut für Astrophysik, D-85741 Garching, Germany;
email: volker@mpa-garching.mpg.de

Annu. Rev. Astron. Astrophys. 2010. 48:391–430

First published online as a Review in Advance on May 24, 2010

The *Annual Review of Astronomy and Astrophysics* is online at astro.annualreviews.org

This article's doi:
10.1146/annurev-astro-081309-130914

0066-4146/10/0922-0391$20.00

Key Words

conservation laws, fluid particles, gas dynamics, numerical convergence, numerical simulations, structure formation

Abstract

This review discusses smoothed particle hydrodynamics (SPH) in the astrophysical context, with a focus on inviscid gas dynamics. The particle-based SPH technique allows an intuitive and simple formulation of hydrodynamics that has excellent conservation properties and can be coupled to self-gravity with high accuracy. The Lagrangian character of SPH allows it to automatically adjust its resolution to the clumping of matter, a property that makes the scheme ideal for many application areas in astrophysics, where often a large dynamic range in density is encountered. We discuss the derivation of the basic SPH equations in their modern formulation, and give an overview about extensions of SPH developed to treat physics such as radiative transfer, thermal conduction, relativistic dynamics, or magnetic fields. We also briefly describe some of the most important applications areas of SPH in astrophysical research. Finally, we provide a critical discussion of the accuracy of SPH for different hydrodynamical problems, including measurements of its convergence rate for important classes of problems.

1. INTRODUCTION

Smoothed particle hydrodynamics (SPH) is a technique for approximating the continuum dynamics of fluids through the use of particles, which may also be viewed as interpolation points. SPH was originally developed in astrophysics, as introduced by Lucy (1977) and Gingold & Monaghan (1977) some 30 years ago. Since then it has also found widespread use in other areas of science and engineering. In this review, I discuss SPH in its modern form, based on a formulation derived from variational principals, giving SPH very good conservation properties and making its derivation largely free of ad hoc choices that needed to be made in older versions of SPH.

A few excellent reviews have discussed SPH previously (e.g., Monaghan 1992, 2005; Dolag et al. 2008; Rosswog 2009); hence, we concentrate primarily on recent developments and on a critical discussion of SPH's advantages and disadvantages, rather than on giving a full historical account of the most important literature on SPH. Also, we generally restrict the discussion of SPH to inviscid ideal gases, which is the relevant case for the most common applications in astrophysics, especially in cosmology. Only in passing we comment on some other important uses of SPH in astronomy, where fluids quite different from an ideal gas are modeled, e.g., in planet formation. Fully outside the scope of this review are the many successful applications of SPH-based techniques in fields such as geophysics and engineering. For example, free-surface flows tend to be very difficult to model with Eulerian methods, whereas this is comparatively easy with SPH. As a result, there are many applications of SPH to problems such as dam-braking, avalanches, and the like, which we however do not discuss here.

Numerical simulations have become an important tool in astrophysical research. For example, cosmological simulations of structure formation within the ΛCDM model have been instrumental to understanding the nonlinear outcome of the initial conditions predicted by the theory of inflation. By now, techniques to simulate collisionless dark matter through the particle-based N-body method (Hockney & Eastwood 1981) have fully matured and are comparatively well understood. However, to represent the collisional baryons as well, the hydrodynamical fluid equations need to be solved, which represents a much harder problem than the dark matter dynamics. On top of the more complicated gas dynamics, additional physics, like radiation fields, magnetic fields, or nonthermal particle components need to be numerically followed as well to produce realistic models of the formation and evolution of galaxies, stars, or planets. There is therefore ample need for robust, accurate, and efficient hydrodynamical discretization techniques in astrophysics.

The principal idea of SPH is to treat hydrodynamics in a completely mesh-free fashion, in terms of a set of sampling particles. Hydrodynamical equations of motion are then derived for these particles, yielding a quite simple and intuitive formulation of gas dynamics. Moreover, it turns out that the particle representation of SPH has excellent conservation properties. Energy, linear momentum, angular momentum, mass, and entropy (if no artificial viscosity operates) are simultaneously conserved. In addition, there are no advection errors in SPH, and the scheme is fully Galilean invariant, unlike alternative mesh-based Eulerian techniques. Due to its Lagrangian character, the local resolution of SPH follows the mass flow automatically, a property that is extremely convenient in representing the large density contrasts often encountered in astrophysical problems. Together with the ease with which SPH can be combined with accurate treatments of self-gravity (suitable and efficient gravity solvers for particles can be conveniently taken from a cosmological N-body code developed for the representation of dark matter), this has made the method very popular for studying a wide array of problems in astrophysics, ranging from cosmological structure growth driven by gravitational instability to studies of the collisions of protoplanets. Furthermore, additional subresolution treatments of unresolved physical processes (such as star formation in galaxy-scale simulations) can be intuitively added at the particle level in SPH.

In this review, we first give, in Section 2, a derivation of what has become the standard formulation of SPH for ideal gases, including also a description of the role of artificial viscosity. We then summarize some extensions of SPH to include additional physics such as magnetic fields or thermal conduction in Section 3, followed by a brief description of the different application areas of SPH in astronomy in Section 4. Despite the popularity of SPH, there have been few systematic studies of the accuracy of SPH when compared with the traditional Eulerian approaches. We therefore include a discussion of the convergence, consistency, and stability of standard SPH in Section 5, based on some of our own tests. Finally, we give a discussion of potential future directions of SPH development in Section 6 and our conclusions in Section 7.

2. BASIC FORMULATION OF SMOOTHED PARTICLE HYDRODYNAMICS FOR IDEAL GASES

2.1. Kernel Interpolants

At the heart of SPH lie so-called kernel interpolants, which are discussed in detail by Gingold & Monaghan (1982). In particular, we use a kernel summation interpolant for estimating the density, which then determines the rest of the basic SPH equations through the variational formalism.

For any field $F(\mathbf{r})$, we may define a smoothed interpolated version, $F_s(\mathbf{r})$, through a convolution with a kernel $W(\mathbf{r}, h)$:

$$F_s(\mathbf{r}) = \int F(\mathbf{r})W(\mathbf{r}-\mathbf{r}', h)\mathrm{d}\mathbf{r}'. \tag{1}$$

Here, h describes the characteristic width of the kernel, which is normalized to unity and approximates a Dirac δ-function in the limit $h \to 0$. We further require that the kernel be symmetric and sufficiently smooth to make it differentiable at least twice. One possibility for W is a Gaussian, which was in fact used by Gingold & Monaghan (1977). However, most current SPH implementations are based on kernels with a finite support. Usually a cubic spline is adopted with $W(\mathbf{r}, h) = w(\frac{|\mathbf{r}|}{2h})$, and

$$w_{3\mathrm{D}}(q) = \frac{8}{\pi}\begin{cases} 1-6q^2+6q^3, & 0 \le q \le \frac{1}{2}, \\ 2(1-q)^3, & \frac{1}{2} < q \le 1, \\ 0, & q > 1, \end{cases} \tag{2}$$

in 3D normalization. This kernel belongs to a broader class of interpolation and smoothing kernels (Schoenberg 1969, Hockney & Eastwood 1981, Monaghan 1985). Note that in the above most commonly used definition of the smoothing length h, the kernel drops to zero at a distance of $r = 2h$. Through Taylor expansion, it is easy to see that the kernel interpolant is at least second-order accurate due to the symmetry of the kernel.

Suppose now we know the field at a set of points $\mathbf{r}_i$; that is, $F_i = F(\mathbf{r}_i)$. The points have an associated mass m_i and density ρ_i, such that $\Delta\mathbf{r}_i \sim m_i/\rho_i$ is their associated finite volume element. Provided the points sufficiently densely sample the kernel volume, we can approximate the integral in Equation 1 with the sum

$$F_s(\mathbf{r}) \simeq \sum_j \frac{m_j}{\rho_j} F_j W(\mathbf{r}-\mathbf{r}_j, h). \tag{3}$$

This is effectively a Monte-Carlo integration, except that thanks to the comparatively regular distribution of points encountered in practice, the accuracy is considerably better than for a random distribution of the sampling points. In particular, for points in one dimension with equal spacing d, one can show that for $h = d$ the sum of Equation 3 provides a second-order accurate

approximation to the real underlying function. Unfortunately, for the irregular yet somewhat ordered particle configurations encountered in real applications, a formal error analysis is not straightforward. It is clear, however, that at the very least one should have $h \geq d$, which translates to a minimum of ∼33 neighbors in three dimensions.

Importantly, we see that the estimate for $F_s(\mathbf{r})$ is defined everywhere (not only at the underlying points) and is differentiable thanks to the differentiability of the kernel, albeit with a considerably higher interpolation error for the derivative. Moreover, if we set $F(\mathbf{r}) = \rho(\mathbf{r})$, we obtain

$$\rho_s(\mathbf{r}) \simeq \sum_j m_j W(\mathbf{r} - \mathbf{r}_j, h), \tag{4}$$

yielding a density estimate based just on the particle coordinates and their masses. In general, the smoothing length can be made variable in space, $h = h(\mathbf{r}, t)$, to account for variations in the sampling density. This adaptivity is one of the key advantages of SPH and is essentially always used in practice. There are two options to introduce the variability of h into Equation 4. One is by adopting $W[\mathbf{r} - \mathbf{r}_j, h(\mathbf{r})]$ as kernel, which corresponds to the scatter approach (Hernquist & Katz 1989). It has the advantage that the volume integral of the smoothed field recovers the total mass, $\int \rho_s(\mathbf{r})\mathrm{d}\mathbf{r} = \sum_i m_i$. However, the so-called gather approach, where we use $W[\mathbf{r} - \mathbf{r}_j, h(\mathbf{r}_i)]$ as kernel in Equation 4, requires only knowledge of the smoothing length h_i for estimating the density of particle i, which leads to computationally convenient expressions when the variation of the smoothing length is consistently included in the SPH equations of motion. Because the density is only needed at the coordinates of the particles and the total mass is conserved anyway (because it is tied to the particles), it is not important that the volume integral of the gather form of $\rho_s(\mathbf{r})$ exactly equals the total mass.

In the following, we drop the subscript s for indicating the smoothed field, and adopt as the SPH estimate of the density of particle i the expression

$$\rho_i = \sum_{j=1}^{N} m_j W(\mathbf{r}_i - \mathbf{r}_j, h_i). \tag{5}$$

It is clear now why kernels with a finite support are preferred. They allow the summation to be restricted to the N_{ngb} neighbors that lie within the spherical region of radius $2h$ around the target point $\mathbf{r}_i$, corresponding to a computational cost of order $\mathcal{O}(N_{\mathrm{ngb}}\, N)$ for the full density estimate. Normally this number N_{ngb} of neighbors within the support of the kernel is approximately (or exactly) kept constant by choosing the h_i appropriately. N_{ngb}, hence, represents an important parameter of the SPH method and needs to be made large enough to provide sufficient sampling of the kernel volumes. Kernels like the Gaussian, however, would require a summation over all particles N for every target particle, resulting in an $\mathcal{O}(N^2)$ scaling of the computational cost.

If SPH were really a Monte-Carlo method, the accuracy expected from the interpolation errors of the density estimate would be rather problematic. But the errors are much smaller because the particles do not sample the fluid in a Poissonian fashion. Instead, their distances tend to equilibrate due to the pressure forces, which makes the interpolation errors much smaller. Yet, they remain a significant source of error in SPH and are ultimately the primary origin of the noise inherent in SPH results.

Even though we have based most of the above discussion on the density, the general kernel interpolation technique can also be applied to other fields and to the construction of differential operators. For example, we may write down a smoothed velocity field and take its derivative to

estimate the local velocity divergence, yielding the following:

$$(\nabla \cdot \mathbf{v})_i = \sum_j \frac{m_j}{\rho_j} \mathbf{v}_j \cdot \nabla_i W(\mathbf{r}_i - \mathbf{r}_j, h). \tag{6}$$

However, an alternative estimate can be obtained by considering the identity $\rho\nabla\cdot\mathbf{v} = \nabla(\rho\mathbf{v}) - \mathbf{v}\cdot\nabla\rho$ and computing kernel estimates for the two terms on the right-hand side independently. Their difference then yields

$$(\nabla \cdot \mathbf{v})_i = \frac{1}{\rho_i} \sum_j m_j (\mathbf{v}_j - \mathbf{v}_i) \cdot \nabla_i W(\mathbf{r}_i - \mathbf{r}_j, h). \tag{7}$$

This pair-wise formulation turns out to be more accurate in practice. In particular, it has the advantage of always providing a vanishing velocity divergence if all particle velocities are equal.

2.2. Variational Derivation

The Euler equations for inviscid gas dynamics in Lagrangian (comoving) form are given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\nabla \cdot \mathbf{v} = 0, \tag{8}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \frac{\nabla P}{\rho} = 0, \tag{9}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} + P\nabla \cdot \mathbf{v} = 0, \tag{10}$$

where $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \mathbf{v}\cdot\nabla$ is the convective derivative. This system of partial differential equations expresses conservation of mass, momentum, and energy. Eckart (1960) has shown that the Euler equations for an inviscid ideal gas follow from the Lagrangian

$$L = \int \rho \left(\frac{\mathbf{v}^2}{2} - u \right) \mathrm{d}V\,. \tag{11}$$

This opens up an interesting route for obtaining discretized equations of motion for gas dynamics. Instead of working with the continuum equations directly and trying to heuristically work out a set of accurate difference formulas, one can discretize the Lagrangian and then derive SPH equations of motion by applying the variational principles of classical mechanics, an approach first proposed by Gingold & Monaghan (1982). Using a Lagrangian also immediately guarantees certain conservation laws and retains the geometric structure imposed by Hamiltonian dynamics on phase space.

We here follow this elegant idea, which was first worked out by Springel & Hernquist (2002), with a consistent accounting of variable smoothing lengths. We start by discretizing the Lagrangian in terms of fluid particles of mass m_i, yielding

$$L_{\mathrm{SPH}} = \sum_i \left(\frac{1}{2} m_i \mathbf{v}_i^2 - m_i u_i \right), \tag{12}$$

where it has been assumed that the thermal energy per unit mass of a particle can be expressed through an entropic function A_i of the particle, which simply labels its specific thermodynamic entropy. The pressure of the particles is

$$P_i = A_i \rho_i^\gamma = (\gamma - 1)\rho_i u_i, \tag{13}$$

where γ is the adiabatic index. Note that for isentropic flow (that is, in the absence of shocks, and without mixing or thermal conduction), we expect the A_i to be constant. We, hence, define u_i, the thermal energy per unit mass, in terms of the density estimate as

$$u_i(\rho_i) = A_i \frac{\rho_i^{\gamma-1}}{\gamma - 1}. \tag{14}$$

This raises the question of how the smoothing lengths h_i needed for estimating ρ_i should be determined. As we discussed above, we would like to ensure adaptive kernel sizes, meaning that the number of points in the kernel should be approximately constant. In much of the older SPH literature, the number of neighbors was allowed to vary within some (small) range around a target number. Sometimes the smoothing length itself was evolved with a differential equation in time, exploiting the continuity relation and the expectation that ρh^3 should be approximately constant (e.g., Steinmetz & Mueller 1993). In case the number of neighbors inside the kernel happened to fall outside the allowed range, h was suitably readjusted. However, Nelson & Papaloizou (1994) pointed out that for smoothing lengths varied in this way, the energy is not conserved correctly. They showed that the errors could be made smaller by keeping the number of neighbors exactly constant, and they also derived leading order correction terms (which became known as ∇h terms) for the classic SPH equations of motion that could reduce them still further. In the modern formulation discussed below, these ∇h terms do not occur; they are implicitly included at all orders.

The central trick making this possible is to require that the mass in the kernel volume should be constant, e.g.,

$$\rho_i h_i^3 = \text{const} \tag{15}$$

for three dimensions. Because $\rho_i = \rho_i(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N, h_i)$ is only a function of the particle coordinates and of h_i, this equation implicitly defines the function $h_i = h_i(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ in terms of the particle coordinates.

We can then proceed to derive the equations of motion from

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\mathbf{r}}_i} - \frac{\partial L}{\partial \mathbf{r}_i} = 0. \tag{16}$$

This first gives

$$m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_{j=1}^{N} m_j \frac{P_j}{\rho_j^2} \frac{\partial \rho_j}{\partial \mathbf{r}_i}, \tag{17}$$

where the derivative $\partial\rho_j/\partial\mathbf{r}_i$ stands for the total variation of the density with respect to the coordinate $\mathbf{r}_i$, including any variation of h_j this may entail. We can then write

$$\frac{\partial \rho_j}{\partial \mathbf{r}_i} = \nabla_i \rho_j + \frac{\partial \rho_j}{\partial h_j}\frac{\partial h_j}{\partial \mathbf{r}_i}, \tag{18}$$

where the smoothing length is kept constant in the first derivative on the right-hand side (in our notation, the Nabla operator $\nabla_i = \partial/\partial\mathbf{r}_i$ means differentiation with respect to $\mathbf{r}_i$ holding the smoothing lengths constant). However, differentiation of $\rho_j h_j^3 = \text{const}$ with respect to $\mathbf{r}_i$ yields

$$\frac{\partial \rho_j}{\partial h_j}\frac{\partial h_j}{\partial \mathbf{r}_i}\left[1 + \frac{3\rho_j}{h_j}\left(\frac{\partial \rho_j}{\partial h_j}\right)^{-1}\right] = -\nabla_i \rho_j. \tag{19}$$

Combining Equations 18 and 19, we then find

$$\frac{\partial \rho_j}{\partial \mathbf{r}_i} = \left(1 + \frac{h_j}{3\rho_j}\frac{\partial \rho_j}{\partial h_j}\right)^{-1} \nabla_i \rho_j. \tag{20}$$

Using

$$\nabla_i \rho_j = m_i \nabla_i W_{ij}(h_j) + \delta_{ij} \sum_{k=1}^{N} m_k \nabla_i W_{ki}(h_i), \tag{21}$$

we finally obtain the equations of motion

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_{j=1}^{N} m_j \left[f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right], \tag{22}$$

where the f_i are defined by

$$f_i = \left[1 + \frac{h_i}{3\rho_i} \frac{\partial \rho_i}{\partial h_i} \right]^{-1}, \tag{23}$$

and the abbreviation $W_{ij}(h) = W(|\mathbf{r}_i - \mathbf{r}_j|, h)$ has been used. Note that the correction factors f_i can be easily calculated alongside the density estimate; all that is required is an additional summation to get $\partial \rho_i / \partial \mathbf{r}_i$ for each particle. This quantity is in fact also useful to get the correct smoothing radii by iteratively solving $\rho_i h_i^3 = \text{const}$ with a Newton-Raphson iteration.

The equations of motion (Equation 22) for inviscid hydrodynamics are remarkably simple. In essence, we have transformed a complicated system of partial differential equations into a much simpler set of ordinary differential equations. Furthermore, we only have to solve the momentum equation explicitly. The mass conservation equation as well as the total energy equation (and, hence, the thermal energy equation) are already taken care of, because the particle masses and their specific entropies stay constant for reversible gas dynamics. However, later we introduce an artificial viscosity that is needed to allow a treatment of shocks. This introduces additional terms in the equations of motion and requires the time integration of one thermodynamic quantity per particle, which can be chosen as either entropy or thermal energy. Indeed, Monaghan (2002) pointed out that the above formulation can also be equivalently expressed in terms of thermal energy instead of entropy. This follows by taking the time derivative of Equation 14, which first yields

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\rho_i} \sum_j \mathbf{v}_j \cdot \frac{\partial \rho_i}{\partial \mathbf{r}_j}. \tag{24}$$

Using Equations 20 and 21 then gives the evolution of the thermal energy as

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = f_i \frac{P_i}{\rho_i} \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla W_{ij}(h_i), \tag{25}$$

which needs to be integrated along the equation of motion if one wants to use the thermal energy as an independent thermodynamic variable. There is no difference, however, between using the entropy or the energy; the two are completely equivalent in the variational formulation. This also solves the old problem pointed out by Hernquist (1993): that the classic SPH equations did not properly conserve energy when the entropy was integrated, and vice versa. Arguably, it is numerically advantageous to integrate the entropy though, as this is computationally cheaper and eliminates time integration errors in solving Equation 25.

Note that the above formulation readily fulfills the conservation laws of energy, momentum, and angular momentum. This can be shown based on the discretized form of the equations, but it is also manifest due to the symmetries of the Lagrangian that was used as a starting point. The absence of an explicit time dependence gives the energy conservation, the translational invariance implies momentum conservation, and the rotational invariance gives angular momentum conservation.

Other derivations of the SPH equations based on constructing kernel interpolated versions of differential operators and applying them directly to the Euler equations are also possible (see, e.g.,

Monaghan 1992). However, these derivations of "classic SPH" are not unique in the sense that one is left with several different possibilities for the equations and certain ad hoc symmetrizations need to be introduced. The choice for a particular formulation then needs to rely on experimentally comparing the performance of many different variants (Thacker et al. 2000).

2.3. Artificial Viscosity

Even when starting from perfectly smooth initial conditions, the gas dynamics described by the Euler equations may readily produce true discontinuities in the form of shock waves and contact discontinuities (Landau & Lifshitz 1959). At such fronts the differential form of the Euler equations breaks down, and their integral form (equivalent to the conservation laws) needs to be used. At a shock front, this yields the Rankine-Hugoniot jump conditions that relate the upstream and downstream states of the fluid. These relations show that the specific entropy of the gas always increases at a shock front, implying that in the shock layer itself the gas dynamics can no longer be described as inviscid. In turn, this also implies that the discretized SPH equations we derived above cannot correctly describe a shock for the simple reason that they keep the entropy strictly constant.

One thus must allow for a modification of the dynamics at shocks and somehow introduce the necessary dissipation. This is usually accomplished in SPH by an artificial viscosity. Its purpose is to dissipate kinetic energy into heat and to produce entropy in the process. The usual approach is to parameterize the artificial viscosity in terms of a friction force that damps the relative motion of particles. Through the viscosity, the shock is broadened into a resolvable layer, something that makes a description of the dynamics everywhere in terms of the differential form possible. It may seem a daunting task, though, to somehow tune the strength of the artificial viscosity such that just the right amount of entropy is generated in a shock. Fortunately, this is relatively unproblematic. Provided the viscosity is introduced into the dynamics in a conservative fashion, the conservation laws themselves ensure that the right amount of dissipation occurs at a shock front.

What is more problematic is to devise the viscosity such that it is only active when there is really a shock present. If it also operates outside of shocks, even if only at a weak level, the dynamics may begin to deviate from that of an ideal gas.

The viscous force is most often added to the equation of motion as

$$\left.\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}\right|_{\mathrm{visc}} = -\sum_{j=1}^{N} m_j \Pi_{ij} \nabla_i \overline{W}_{ij}, \tag{26}$$

where

$$\overline{W}_{ij} = \frac{1}{2}[W_{ij}(h_i) + W_{ij}(h_j)] \tag{27}$$

denotes a symmetrized kernel, which some researchers prefer to define as $\overline{W}_{ij} = W_{ij}([h_i + h_j]/2)$. Provided the viscosity factor Π_{ij} is symmetric in i and j, the viscous force between any pair of interacting particles will be antisymmetric and along the line joining the particles. Hence, linear momentum and angular momentum are still preserved. In order to conserve total energy, we need to compensate the work done against the viscous force in the thermal reservoir, described either in terms of entropy,

$$\left.\frac{\mathrm{d}A_i}{\mathrm{d}t}\right|_{\mathrm{visc}} = \frac{1}{2}\frac{\gamma-1}{\rho_i^{\gamma-1}} \sum_{j=1}^{N} m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \nabla_i \overline{W}_{ij}, \tag{28}$$

or in terms of thermal energy per unit mass,

$$\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{visc}} = \frac{1}{2}\sum_{j=1}^{N} m_j \Pi_{ij}\mathbf{v}_{ij}\cdot\nabla_i\overline{W}_{ij}, \tag{29}$$

where $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. There is substantial freedom in the detailed parameterization of the viscosity Π_{ij}. The most commonly used formulation is an improved version of the viscosity introduced by Monaghan & Gingold (1983),

$$\Pi_{ij} = \begin{cases} [-\alpha c_{ij}\mu_{ij} + \beta\mu_{ij}^2]/\rho_{ij} & \text{if } \mathbf{v}_{ij}\cdot\mathbf{r}_{ij} < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{30}$$

with

$$\mu_{ij} = \frac{h_{ij}\mathbf{v}_{ij}\cdot\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|^2 + \varepsilon h_{ij}^2}. \tag{31}$$

Here, h_{ij} and ρ_{ij} denote arithmetic means of the corresponding quantities for the two particles i and j, with c_{ij} giving the mean sound speed. The strength of the viscosity is regulated by the parameters α and β, with typical values in the range of $\alpha \simeq 0.5 - 1.0$ and the frequent choice of $\beta = 2\alpha$. The parameter $\varepsilon \simeq 0.01$ is introduced to protect against singularities if two particles happen to get very close.

In this form, the artificial viscosity is basically a combination of a bulk and a von Neumann-Richtmyer viscosity. Historically, the quadratic term in μ_{ij} has been added to the original Monaghan-Gingold form to prevent particle penetration in high Mach number shocks. Note that the viscosity only acts for particles that rapidly approach each other; hence, the entropy production is always positive definite. Also, the viscosity vanishes for solid-body rotation, but not for pure shear flows. To cure this problem in shear flows, Balsara (1995) suggested adding a correction factor to the viscosity, reducing its strength when the shear is strong. This can be achieved by multiplying Π_{ij} with a prefactor $(f_i^{\mathrm{AV}} + f_j^{\mathrm{AV}})/2$, where the factors

$$f_i^{\mathrm{AV}} = \frac{|\nabla\cdot\mathbf{v}|_i}{|\nabla\cdot\mathbf{v}|_i + |\nabla\times\mathbf{v}|_i} \tag{32}$$

are meant to measure the rate of the local compression in relation to the strength of the local shear (estimated with formulas such as Equation 7). This Balsara switch has often been successfully used in multidimensional flows and is enabled as default in many SPH codes. We note, however, that it may be problematic sometimes in cases where shocks and shear occur together, e.g., in oblique shocks in differentially rotating disks.

In some studies, alternative forms of viscosity have been tested. For example, Monaghan (1997) proposed a modified form of the viscosity based on an analogy to the Riemann problem, which can be written as

$$\Pi_{ij} = -\frac{\alpha}{2}\frac{v_{ij}^{\mathrm{sig}} w_{ij}}{\rho_{ij}}, \tag{33}$$

where $v_{ij}^{\mathrm{sig}} = c_i + c_j - 3w_{ij}$ is an estimate of the signal velocity between two particles i and j, and $w_{ij} = \mathbf{v}_{ij}\cdot\mathbf{r}_{ij}/|\mathbf{r}_{ij}|$ is the relative velocity projected onto the separation vector. This viscosity is identical to Equation 30 if one sets $\beta = 3/2\,\alpha$ and replaces w_{ij} with μ_{ij}. The main difference between the two viscosities lies therefore in the additional factor of h_{ij}/r_{ij} that μ_{ij} carries with respect to w_{ij}. In Equations 30 and 31, this factor weights the viscous force toward particle pairs with small separations. In fact, after multiplying with the kernel derivative, this weighting is strong enough to make the viscous force in Equation 31 diverge as $\propto 1/r_{ij}$ for small pair separations up to the limit set by the εh_{ij}^2 term.

Lombardi et al. (1999) have systematically tested different parameterization of viscosity, but in general, the standard form was found to work best. Recently, the use of a tensor artificial viscosity was conjectured by Owen (2004) as part of an attempt to optimize the spatial resolution of SPH. However, a disadvantage of this scheme is that it breaks the strict conservation of angular momentum.

In attempting to reduce the numerical viscosity of SPH in regions away from shocks, several studies have instead advanced the idea of keeping the functional form of the artificial viscosity, but making the viscosity strength parameter α variable in time. Such a scheme was first suggested by Morris (1997), and it was successfully applied for studying astrophysical turbulence more faithfully in SPH (Dolag et al. 2005b) and to follow neutron star mergers (Rosswog et al. 1999, Rosswog 2005). Adopting $\beta = 2\alpha$, one may evolve the parameter α individually for each particle with an equation such as

$$\frac{\mathrm{d}\alpha_i}{\mathrm{d}t} = -\frac{\alpha_i - \alpha_{\mathrm{max}}}{\tau_i} + S_i, \tag{34}$$

where S_i is some source function meant to ramp up the viscosity rapidly if a shock is detected, while the first term lets the viscosity exponentially decay again to a prescribed minimum value α_{min} on a timescale τ_i. So far, simple source functions like $S_i = \mathrm{max}[-(\nabla \cdot \mathbf{v})_i, 0]$ and timescales $\tau_i \simeq h_i/c_i$ have been explored, and the viscosity, α_i, has often also been prevented from becoming higher than some prescribed maximum value α_{max}. It is clear that the success of such a variable α scheme depends critically on an appropriate source function. The form above can still not distinguish purely adiabatic compression from that in a shock, so is not completely free of creating unwanted viscosity.

2.4. Coupling to Self-Gravity

Self-gravity is extremely important in many astrophysical flows, quite in contrast to other areas of computational fluid dynamics, where only external gravitational fields play a role. It is noteworthy that Eulerian mesh-based approaches do not manifestly conserve total energy if self-gravity is included (Müller & Steinmetz 1995, Springel 2010), but it can be easily and accurately incorporated in SPH.

Arguably the best approach to account for gravity is to include it directly into the discretized SPH Lagrangian, which has the advantage of also allowing a consistent treatment of variable gravitational softening lengths. This was first conducted by Price & Monaghan (2007) in the context of adaptive resolution N-body methods for collisionless dynamics (see also Bagla & Khandai 2009).

Let $\Phi(\mathbf{r}) = G\sum_i m_i \phi(\mathbf{r} - \mathbf{r}_i, \varepsilon_i)$ be the gravitational field described by the SPH point set, where ε_i is the gravitational softening length of particle i. We then define the total gravitational self-energy of the system of SPH particles as

$$E_{\mathrm{pot}} = \frac{1}{2}\sum_i m_i \Phi(\mathbf{r}_i) = \frac{G}{2}\sum_{ij} m_i m_j \phi(r_{ij}, \varepsilon_j). \tag{35}$$

For SPH including self-gravity, the Lagrangian then becomes

$$L_{\mathrm{SPH}} = \sum_i \left(\frac{1}{2} m_i \mathbf{v}_i^2 - m_i u_i\right) - \frac{G}{2}\sum_{ij} m_i m_j \phi(r_{ij}, \varepsilon_j). \tag{36}$$

As a result, the equation of motion acquires an additional term due to the gravitational forces, given by:

$$
\begin{aligned}
m_i \mathbf{a}_i^{\mathrm{grav}} &= -\frac{\partial E_{\mathrm{pot}}}{\partial \mathbf{r}_i} \\
&= -\sum_j G m_i m_j \frac{\mathbf{r}_{ij}}{r_{ij}} \frac{[\phi'(r_{ij}, \varepsilon_i) + \phi'(r_{ij}, \varepsilon_j)]}{2} \\
&\quad -\frac{1}{2} \sum_{jk} G m_j m_k \frac{\partial \phi(r_{jk}, \varepsilon_j)}{\partial \varepsilon} \frac{\partial \varepsilon_j}{\partial \mathbf{r}_i},
\end{aligned}
\tag{37}
$$

where $\phi'(r, \varepsilon) = \partial\phi/\partial r$.

The first sum on the right-hand side describes the ordinary gravitational force, where the interaction is symmetrized by averaging the forces in case the softening lengths between an interacting pair are different. The second sum gives an additional force component that arises when the gravitational softening lengths are a function of the particle coordinates themselves, that is, if one invokes adaptive gravitational softening. This term has to be included to make the system properly conservative when spatially adaptive gravitational softening lengths are used (Price & Monaghan 2007).

In most cosmological SPH codes of structure formation, this has usually not been done thus far, and a fixed gravitational softening was used for collisionless dark matter particles and SPH particles alike. However, especially in the context of gravitational fragmentation problems, it appears natural, and also indicated by accuracy considerations (Bate & Burkert 1997), to tie the gravitational softening length to the SPH smoothing length, even though some studies caution against this strategy (Williams, Churches & Nelson 2004). Adopting $\varepsilon_i = h_i$ and determining the smoothing lengths as described above, we can calculate the $\partial\varepsilon_j/\partial\mathbf{r}_i = \partial h_j/\partial\mathbf{r}_i$ term by means of Equations 19 and 21. Defining the quantities

$$
\eta_j = \frac{h_j}{3\rho_j} f_j \sum_k m_k \frac{\partial \phi(\mathbf{r}_{jk}, h_j)}{\partial h},
\tag{38}
$$

where the factors f_i are defined as in Equation 23, we can then write the gravitational acceleration compactly as

$$
\begin{aligned}
\left.\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}\right|_{\mathrm{grav}} &= -G \sum_j m_j \frac{\mathbf{r}_{ij}}{r_{ij}} \frac{[\phi'(r_{ij}, h_i) + \phi'(r_{ij}, h_j)]}{2} \\
&\quad + \frac{G}{2} \sum_j m_j [\eta_i \nabla_i W_{ij}(h_i) + \eta_j \nabla_i W_{ij}(h_j)].
\end{aligned}
\tag{39}
$$

This acceleration has to be added to the ordinary SPH equations of motion (Equation 22) due to the pressure forces, which arise from the first part of the Lagrangian (Equation 36). Note that the factors η_j can be conveniently calculated alongside the SPH density estimate. The calculation of the gravitational correction force, that is, the second sum in Equation 39, can then be done together with the usual calculation of the SPH pressure forces. Unlike in the primary gravitational force, here only the nearest neighbors contribute.

We note that it appears natural to relate the functional form of the gravitational softening to the shape of the SPH smoothing kernel, even though this is not strictly necessary. If this is done, $\phi(r, h)$ is determined as a solution to Poisson's equation,

$$
\nabla^2 \phi(r, h) = 4\pi W(r, h).
\tag{40}
$$

The explicit functional form for $\phi(r, h)$ resulting from this identification for the cubic spline kernel is also often employed in collisionless N-body codes and can be found, for example, in the appendix of Springel, Yoshida & White (2001).

2.5. Implementation Aspects

The actual use of the discretized SPH equations in simulation models requires a time-integration scheme. The Hamiltonian character of SPH allows in principle the use of symplectic integration schemes (Hairer, Lubich & Wanner 2002; Springel 2005), such as the leapfrog, which respects the geometric phase-space constraints imposed by the conservation laws and can prevent the build-up of secular integration errors with time. However, because most hydrodynamical problems of interest are not reversible anyway, this aspect is less important than in dissipation-free collisionless dynamics. Hence, any second-order accurate time integration scheme, like a simple Runge-Kutta or predictor-corrector scheme, may equally well be used. It is common practice to use individual time steps for the SPH particles, often in a block-structured scheme with particles arranged in a power-of-two hierarchy of time steps (Hernquist & Katz 1989). This greatly increases the efficiency of calculations in systems with large dynamic range in timescale, but can make the optimum choice of time steps tricky. For example, the standard Courant time step criterion of the form

$$\Delta t_i = C_{\rm CFL} \frac{h_i}{c_i}, \tag{41}$$

usually used in SPH, where $C_{\rm CFL} \sim 0.1 - 0.3$ is a dimensionless parameter, may not guarantee fine enough time-stepping ahead of a blast wave propagating into very cold gas. This problem can be avoided with an improved method for determining the sizes of individual time steps (Saitoh & Makino 2009).

If self-gravity is included, one can draw from the algorithms employed in collisionless N-body dynamics to calculate the gravitational forces efficiently, such as tree-methods (Barnes & Hut 1986) or mesh-based gravity solvers. The tree approach, which provides for a hierarchical grouping of the particles, can also be used to address the primary algorithmic requirement to write an efficient SPH code, namely the need to efficiently find the neighboring particles inside the smoothing kernel of a particle. If this is done naively, by computing the distance to all other particles, a prohibitive $\mathcal{O}(N^2)$ scaling of the computational cost results. Using a range-searching technique together with the tree, this cost can be reduced to $\mathcal{O}(N_{\rm ngb} N \log N)$, independent of the particle clustering. To this end, a special walk of the tree is carried out for neighbor searching in which a tree node is only opened if there is a spatial overlap between the tree node and the smoothing kernel of the target particle; otherwise the corresponding branch of the tree can be immediately discarded.

These convenient properties of the tree algorithm have been exploited for the development of a number of efficient SPH codes in astronomy; several of them are publicly available and parallelized for distributed memory machines. Among them are TreeSPH (Hernquist & Katz 1989; Katz, Weinberg & Hernquist 1996), HYDRA (Pearce & Couchman 1997), GADGET (Springel, Yoshida & White 2001; Springel 2005), GASOLINE (Wadsley, Stadel & Quinn 2004), MAGMA (Rosswog & Price 2007), and VINE (Wetzstein et al. 2009).

3. EXTENSIONS OF SMOOTHED PARTICLE HYDRODYNAMICS

For many astrophysical applications, additional physical processes in the gas phase besides inviscid hydrodynamics need to be modeled. This includes, for example, magnetic fields, transport

processes such as thermal conduction or physical viscosity, or radiative transfer. SPH also needs to be modified for the treatment of fluids that move at relativistic speeds or in relativistically deep potentials. Below we give a brief overview of some of the extensions of SPH that have been developed to study this physics.

3.1. Magnetic Fields

Magnetic fields are ubiquitous in astrophysical plasmas, where the conductivity can often be approximated as being effectively infinite. In this limit one aims to simulate ideal, nonresistive magnetohydrodynamics (MHD), which is thought to be potentially very important in many situations, in particular in star formation, cosmological structure formation, and accretion disks. The equations of ideal MHD are composed of the induction equation,

$$\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} = (\mathbf{B}\cdot\nabla)\mathbf{v} - \mathbf{B}(\nabla\cdot\mathbf{v}), \tag{42}$$

and the magnetic Lorentz force. The latter can be obtained from the Maxwell stress tensor,

$$\mathcal{M}_{ij} = \frac{1}{4\pi}\left(\mathbf{B}_i\mathbf{B}_j - \frac{1}{2}\mathbf{B}^2\delta_{ij}\right), \tag{43}$$

as

$$\mathbf{F}_i = \frac{\partial\mathcal{M}_{ij}}{\partial x_j}. \tag{44}$$

Working with the stress tensor is advantageous for deriving equations of motion that are discretized in a symmetric fashion. The magnetic force $\mathbf{F}$ then has to be added to the usual forces from gas pressure and the gravitational field.

First implementations of magnetic forces in SPH date back to Gingold & Monaghan (1977), soon followed by full implementations of MHD in SPH (Phillips & Monaghan 1985). However, a significant problem with MHD in SPH has become apparent early on. The constraint equation $\nabla\cdot\mathbf{B} = 0$, which is maintained by the continuum form of the ideal MHD equations, is in general not preserved by discretized versions of the equations. Those tend to build up $\nabla\cdot\mathbf{B} \neq 0$ errors over time, corresponding to unphysical magnetic monopole sources. In contrast, in the most modern mesh-based approaches to MHD, so-called constrained transport schemes (Evans & Hawley 1988) are able to accurately evolve the discretized magnetic field while keeping a vanishing divergence of the field.

Much of the recent research on developing improved SPH realizations of MHD has therefore concentrated on constructing formulations that either eliminate the $\nabla\cdot\mathbf{B} \neq 0$ error or at least keep it reasonably small (Dolag, Bartelmann & Lesch 1999; Price & Monaghan 2004a,b, 2005; Dolag & Stasyszyn 2009; Price 2010). To this extent, different approaches have been investigated in the literature, involving periodic field cleaning techniques (e.g., Dedner et al. 2002), formulations of the equations that "diffuse away" the $\nabla\cdot\mathbf{B} \neq 0$ terms if they should arise, or the use of the Euler potentials or the vector potential. The latter may seem like the most obvious solution, because deriving the magnetic field as $\mathbf{B} = \nabla\times\mathbf{A}$ from the vector potential $\mathbf{A}$ will automatically create a divergence-free field. However, Price (2010) explored the use of the vector potential in SPH and concluded that there are substantial instabilities in this approach, rendering it essentially useless in practice.

Another seemingly clean solution lies in the so-called Euler potentials (also known as Clebsch variables). One may write the magnetic field as the cross product $\mathbf{B} = \nabla\alpha\times\nabla\beta$ of the gradients of two scalar fields α and β. In ideal MHD, where the magnetic flux is locked into the flow, one obtains the correct field evolution by simply advecting these Euler potentials α and β with the

motion of the gas. This suggests a deceptively simple approach to ideal MHD; just construct the fields α and β for a given magnetic field, and then move these scalars along with the gas. Although this has been shown to work reasonably well in a number of simple test problems and also has been used in some real-world applications (Price & Bate 2007, Kotarba et al. 2009), it likely has significant problems in general MHD dynamics, as a result of the noise in SPH and the inability of this scheme to account for any magnetic reconnection. Furthermore, Brandenburg (2010) points out that the use of Euler potentials does not converge to a proper solution in hydromagnetic turbulence.

Present formulations of MHD in SPH are, hence, back to discretizing the classic magnetic induction equation. For example, Dolag & Stasyszyn (2009) give an implementation of SPH in the GADGET code where the magnetic induction equation is adopted as

$$\frac{\mathrm{d}\mathbf{B}_i}{\mathrm{d}t} = \frac{f_i}{\rho_i}\sum_j m_j\{\mathbf{B}_i[\mathbf{v}_{ij}\cdot\nabla_i W_{ij}(h_i)] - \mathbf{v}_{ij}[\mathbf{B}_i\cdot\nabla_i W_{ij}(h_i)]\} \tag{45}$$

and the acceleration due to magnetic forces as

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = \sum_j m_j\left[f_i\frac{\mathcal{M}_i}{\rho_i^2}\cdot\nabla_i W_{ij}(h_i) + f_j\frac{\mathcal{M}_j}{\rho_j^2}\cdot\nabla_i W_{ij}(h_j)\right], \tag{46}$$

where the f_i are the correction factors of Equation 23, and $\mathcal{M}_i$ is the stress tensor of particle i. Combined with field cleaning techniques (Børve, Omang & Trulsen 2001) and artificial magnetic dissipation to keep $\nabla\cdot\mathbf{B} \neq 0$ errors under control, this leads to quite accurate results for many standard tests of MHD, like magnetic shock tubes or the Orszag-Tang vortex. A similar implementation of MHD in SPH is given in the independent MAGMA code by Rosswog & Price (2007).

3.2. Thermal Conduction

Diffusion processes governed by variants of the equation

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = D\nabla^2 Q, \tag{47}$$

where Q is some conserved scalar field and D is a diffusion constant, require a discretization of the Laplace operator in SPH. This could be done in principle by differentiating a kernel-interpolated version of Q twice. However, such a discretization turns out to be quite sensitive to the local particle distribution, or in other words, it is fairly noisy. It is much better to use an approximation of the ∇^2 operator proposed first by Brookshaw (1985), in the form

$$\nabla^2 Q\big|_i = -2\sum_j \frac{m_j}{\rho_j}\frac{Q_j - Q_i}{r_{ij}^2}\mathbf{r}_{ij}\cdot\nabla_i\overline{W}_{ij}. \tag{48}$$

Cleary & Monaghan (1999) and Jubelgas, Springel & Dolag (2004) have used this to construct implementations of thermal conduction in SPH that allow for spatially variable conductivities. The heat conduction equation,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{1}{\rho}\nabla(\kappa\nabla T), \tag{49}$$

where κ is the heat conductivity, can then be discretized in SPH as follows:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_j \frac{m_j}{\rho_i\rho_j}\frac{(\kappa_i+\kappa_j)(T_j - T_i)}{r_{ij}^2}\mathbf{r}_{ij}\cdot\nabla_i\overline{W}_{ij}. \tag{50}$$

In this form, the energy exchange between two particles is balanced on a pairwise basis, and heat always flows from higher to lower temperature. Also, it is easy to see that the total entropy increases in the process. This formulation has been used, for example, to study the influence of *Spitzer* conductivity due to electron transport on the thermal structure of the plasma in massive galaxy clusters (Dolag et al. 2004).

3.3. Physical Viscosity

If thermal conduction is not strongly suppressed by magnetic fields in hot plasmas, then one also expects a residual physical viscosity (a pure shear viscosity in this case). In general, real gases may possess both physical bulk and physical shear viscosity. They are then correctly described by the Navier-Stokes equations and not the Euler equations of inviscid gas dynamics that we have discussed thus far. The stress tensor of the gas can be written as

$$\sigma_{ij} = \eta \left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i} - \frac{2}{3}\delta_{ij} \frac{\partial v_k}{\partial r_k} \right) + \zeta\, \delta_{ij} \frac{\partial v_k}{\partial r_k}, \tag{51}$$

where η and ζ are the coefficients of shear and bulk viscosity, respectively. The Navier-Stokes equation including gravity is then given by

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{\nabla P}{\rho} - \nabla\Phi + \frac{1}{\rho}\nabla\boldsymbol{\sigma}. \tag{52}$$

Sijacki & Springel (2006) have suggested an SPH discretization of this equation, which estimates the shear viscosity of particle i based on

$$\left.\frac{\mathrm{d}v_\alpha}{\mathrm{d}t}\right|_{i,\,\mathrm{shear}} = \sum_j m_j \left[\frac{\eta_i \sigma_{\alpha\beta}|_i}{\rho_i^2} [\nabla_i W_{ij}(h_i)]|_\beta + \frac{\eta_j \sigma_{\alpha\beta}|_j}{\rho_j^2} [\nabla_i W_{ij}(h_j)]|_\beta \right], \tag{53}$$

and the bulk viscosity as

$$\left.\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right|_{i,\,\mathrm{bulk}} = \sum_j m_j \left[\frac{\zeta_i \nabla\cdot\mathbf{v}_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + \frac{\zeta_j \nabla\cdot\mathbf{v}_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right]. \tag{54}$$

These viscous forces are antisymmetric, and together with a positive definite entropy evolution,

$$\frac{\mathrm{d}A_i}{\mathrm{d}t} = \frac{\gamma-1}{\rho_i^{\gamma-1}} \left[\frac{1}{2}\frac{\eta_i}{\rho_i}\boldsymbol{\sigma}_i^2 + \frac{\zeta_i}{\rho_i}(\nabla\cdot\mathbf{v})^2 \right], \tag{55}$$

they conserve total energy. Here, the derivatives of the velocity can be estimated with pair-wise formulations as in Equation 7.

In some sense, it is actually simpler for SPH to solve the Navier-Stokes equation than the Euler equations. As we discuss in more detail in Section 5.4, this is because the inclusion of an artificial viscosity in SPH makes the scheme problematic for purely inviscid flow, because this typically introduces a certain level of spurious numerical viscosity also outside of shocks. If the gas has, however, a sizable amount of intrinsic physical viscosity anyway, it becomes much easier to correctly represent the fluid, provided the physical viscosity is larger than the unavoidable numerical viscosity.

3.4. Radiative Transfer

In many astrophysical problems, like the reionization of the Universe or in first star formation, one would like to self-consistently follow the coupled system of hydrodynamical and radiative

transfer equations, allowing for a proper treatment of the backreaction of the radiation on the fluid, and vice versa. Unfortunately, the radiative transfer equation is in itself a high-dimensional partial differential equation that is extremely challenging to solve in its full generality. The task is to follow the evolution of the specific intensity field,

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \mathbf{n}\cdot\nabla I_\nu = -\kappa_\nu I_\nu + j_\nu, \tag{56}$$

which is not only a function of spatial coordinates, but also of direction $\mathbf{n}$ and of frequency ν. Here j_ν is a source function and κ_ν is an absorption coefficient, which provide an implicit coupling to the hydrodynamics. Despite the formidable challenge to develop efficient numerical schemes for the approximate solution of the radiation-hydrodynamical set of equations, the development of such codes based on SPH-based techniques has really flourished in recent years, where several new schemes have been proposed.

Perhaps the simplest and also oldest approaches are based on flux-limited diffusion approximations to radiative transfer in SPH (Whitehouse & Bate 2004, 2006; Whitehouse, Bate & Monaghan 2005; Fryer, Rockefeller & Warren 2006; Viau, Bastien & Cha 2006; Forgan et al. 2009). A refinement of this strategy has been proposed by Petkova & Springel (2009), based on the optically thin variable tensor approximations of Gnedin & Abel (2001). In this approach, the radiative transfer equation is simplified in moment-based form, allowing the radiative transfer to be described in terms of the local energy density J_ν of the radiation,

$$J_\nu = \frac{1}{4\pi}\int I_\nu \, \mathrm{d}\Omega. \tag{57}$$

This radiation energy density is then transported with an anisotropic diffusion equation,

$$\frac{1}{c}\frac{\partial J_\nu}{\partial t} = \frac{\partial}{\partial r_j}\left(\frac{1}{\kappa_\nu}\frac{\partial J_\nu h^{ij}}{\partial r_i}\right) - \kappa_\nu J_\nu + j_\nu, \tag{58}$$

where the matrix h^{ij} is the so-called Eddington tensor, which is symmetric and normalized to unit trace. The Eddington tensor encodes information about the angle dependence of the local radiation field, and in which direction, if any, it prefers to diffuse in case the local radiation field is highly anisotropic. In the ansatz of Gnedin & Abel (2001), the Eddington tensors are simply estimated in an optically thin approximation as a $1/r^2$-weighted sum over all sources, which can be calculated efficiently for an arbitrary number of sources with techniques familiar from the calculation of gravitational fields.

Petkova & Springel (2009) derive an SPH discretization of the transport part of this radiative transfer approximation (described by the first term on the right-hand side) as

$$\frac{\partial N_i}{\partial t} = \sum_j w_{ij}(N_j - N_i), \tag{59}$$

where the factors w_{ij} are given by

$$w_{ij} = \frac{2cm_{ij}}{\kappa_{ij}\rho_{ij}}\frac{\mathbf{r}_{ij}^{\mathrm{T}}\tilde{\mathbf{h}}_{ij}\nabla_i\overline{W}_{ij}}{\mathbf{r}_{ij}^2}, \tag{60}$$

and $N_i = c^2 J_\nu/(h_{\mathrm{Planck}}^4 \nu^3)$ is the photon number associated with particle i. The tensor,

$$\tilde{\mathbf{h}} = \frac{5}{2}\mathbf{h} - \frac{1}{2}\mathrm{Tr}(\mathbf{h}), \tag{61}$$

is a modified Eddington tensor such that the SPH discretization in Equation 60 corresponds to the correct anisotropic diffusion operator. The exchange described by Equation 59 is photon

conserving, and because the matrix w_{ij} is symmetric and positive definite, reasonably fast iterative conjugate gradient solvers can be used to integrate the diffusion problem implicitly in time in a stable fashion.

A completely different approach to combine radiative transfer with SPH has been described by Pawlik & Schaye (2008), who transport the radiation in terms of emission and reception cones, from particle to particle. This effectively corresponds to a coarse discretization of the solid angle around each particle. Yet another scheme is given by Nayakshin, Cha & Hobbs (2009), who implemented a Monte Carlo implementation of photon transport in SPH. Here, the photons are directly implemented as virtual particles, thereby providing a Monte Carlo sampling of the transport equation. Although this method suffers from the usual Monte Carlo noise limitations, only few simplifying assumptions in the radiative transfer equation itself have to be made, and it can therefore be made increasingly more accurate simply by using more Monte-Carlo photons.

Finally, Altay, Croft & Pelupessy (2008) suggested a ray-tracing scheme within SPH, which essentially implements ideas that are known from so-called long- and short-characteristics methods for radiative transfer around point sources in mesh codes. These approaches have the advantage of being highly accurate, but their efficiency rapidly declines with an increasing number of sources.

3.5. Relativistic Dynamics

It is also possible to derive SPH equations for relativistic dynamics from a variational principle (Monaghan & Price 2001), both for special and general relativistic dynamics. This is more elegant than alternative derivations and avoids some problems inherent in other approaches to relativistic dynamics in SPH (Monaghan 1992; Laguna, Miller & Zurek 1993; Faber & Rasio 2000; Siegler & Riffert 2000; Ayal et al. 2001).

The variational method requires discretizing the Lagrangian,

$$L = -\int T_{\mu\nu} U^{\nu} U^{\nu} \, \mathrm{d}V \,, \tag{62}$$

where U^{μ} is the four-velocity, and $T_{\mu\nu} = (P + e) U_{\nu} U_{\nu} + P \, \eta_{\mu\nu}$ is the energy momentum tensor of a perfect fluid with pressure P and rest-frame energy density

$$e = n \, m_0 c^2 \left(1 + \frac{u}{c^2} \right). \tag{63}$$

Here a $(-1, 1, 1, 1)$ signature for the flat metric tensor $\eta^{\mu\nu}$ is used. n is the rest frame number density of particles of mean molecular weight m_0, and u is the rest-frame thermal energy per unit mass. In the following, we set $c = m_0 = 1$.

Rosswog (2009) gives a detailed derivation of the resulting equations of motion when the discretization is done in terms of fluid parcels with a constant number ν_i of baryons in SPH-particle i and when variable smoothing lengths are consistently included. In the special-relativistic case, he derives the equations of motion as

$$\frac{\mathrm{d}\mathbf{s}_i}{\mathrm{d}t} = -\sum_j \nu_j \left[f_i \frac{P_i}{N_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{N_j^2} \nabla_i W_{ij}(h_j) \right], \tag{64}$$

where the generalized momentum $\mathbf{s}_i$ of particle i is given by

$$\mathbf{s}_i = \gamma_i \mathbf{v}_i \left(1 + u_i + \frac{P_i}{n_i} \right), \tag{65}$$

and γ_i is the particle's Lorentz factor. The baryon number densities N_i in the computing frame are estimated just as in Equation 4, except for the replacements $\rho_i \rightarrow N_i$ and $m_j \rightarrow \nu_j$. Similarly, the correction factors f_i are defined as in Equation 23 with the same replacements.

One also needs to complement the equation of motion with an energy equation of the form

$$\frac{\mathrm{d}\varepsilon_i}{\mathrm{d}t} = -\sum_j \nu_j \left[\frac{f_i P_i}{N_i^2} \mathbf{v}_i \cdot \nabla W_{ij}(h_i) + \frac{f_j P_j}{N_j^2} \mathbf{v}_j \cdot \nabla W_{ij}(h_j) \right], \tag{66}$$

where

$$\varepsilon_i = \mathbf{v}_i \cdot \mathbf{s}_i + \frac{1 + u_i}{\gamma_i} \tag{67}$$

is a suitably defined relativistic energy variable. These equations conserve the total canonical momentum as well as the angular momentum. A slight technical complication arises for this choice of variables due to the need to recover the primitive fluid variables after each time step from the updated integration variables $\mathbf{s}_i$ and ε_i, which requires finding the root of a nonlinear equation.

4. APPLICATIONS OF SMOOTHED PARTICLE HYDRODYNAMICS

The versatility and simplicity of SPH have led to a wide range of applications in astronomy, essentially in every field where theoretical research with hydrodynamical simulations is carried out. We here provide a brief, necessarily incomplete overview of some of the most prominent topics that have been studied with SPH.

4.1. Cosmological Structure Formation

Among the most important successes of SPH in cosmology are simulations that have clarified the origin of the Lyman-α forest in the absorption spectra to distant quasars (e.g., Hernquist et al. 1996, Davé et al. 1999), which have been instrumental for testing and interpreting the cold dark matter cosmology. Modern versions of such calculations use detailed chemical models to study the enrichment of the intragalactic medium (Oppenheimer & Davé 2006), and the nature of the so-called warm-hot IGM, which presumably contains a large fraction of all cosmic baryons (Davé et al. 2001).

Cosmological simulations with hydrodynamics are also used to study galaxy formation in its full cosmological context, directly from the primordial fluctuation spectrum predicted by inflationary theory. This requires the inclusion of radiative cooling and subresolution models for the treatment of star formation and associated energy feedback processes from supernovae explosions or galactic winds. A large variety of such models have been proposed (e.g., Springel & Hernquist 2003); some also involve rather substantial changes of SPH, for example, the decoupled version of SPH by Marri & White (2003) for the treatment of multiphase structure in the ISM. Despite the uncertainties such modeling involves and the huge challenge to numerical resolution this problem entails, important theoretical results on the clustering of galaxies, the evolution of the cosmic star-formation rate, and the efficiency of galaxy formation as a function of halo mass have been reached.

One important goal of such simulations has been the formation of disk galaxies in a cosmological context, which has however proven to represent a significant challenge. In recent years, substantial progress in this area has been achieved, however, where some simulated galaxies have become quite close now to the morphology of real spiral galaxies (Governato et al. 2007, Scannapieco et al. 2008). Still, the ratio of the stellar bulge to disk components found in simulations is typically much higher than in observations, and the physical or numerical origin for this discrepancy is debated.

The situation is better for clusters of galaxies, where the most recent SPH simulations have been quite successful in reproducing the primary scaling relations that are observed (Borgani et al. 2004; Puchwein, Sijacki & Springel 2008; McCarthy et al. 2010). This comes after a long struggle in the literature with the cooling-flow problem, namely the fact that simulated clusters tend to efficiently cool out much of their gas at the center, an effect that is not present in observations in the predicted strength. The modern solution is to attribute this to a nongravitational heat source in the center of galaxies, which is thought to arise from active galactic nuclei. This important feedback channel has recently been incorporated in SPH simulations of galaxy clusters (Sijacki et al. 2008).

SPH simulations of cosmic large-scale structure and galaxy clusters are also regularly used to study the Sunyaev-Zel'dovich effect (da Silva et al. 2000), the buildup of magnetic fields (Dolag, Bartelmann & Lesch 1999, 2002; Dolag et al. 2005a), or even the production of nonthermal particle populations in the form of cosmic rays (Pfrommer et al. 2007, Jubelgas et al. 2008). The latter also required the development of a technique to detect shock waves on the fly in SPH and to measure their Mach number (Pfrommer et al. 2006).

4.2. Galaxy Mergers

The hierarchical theory of galaxy formation predicts frequent mergers of galaxies, leading to the buildup of ever larger systems. Such galaxy interactions are also prominently observed in many systems, such as the Antennae Galaxies NGC 4038/4039. The merger hypothesis (Toomre & Toomre 1972) suggests that the coalescence of two spiral galaxies leads to an elliptical remnant galaxy, thereby playing a central role in explaining Hubble's tuning fork diagram for the morphology of galaxies. SPH simulations of merging galaxies have been instrumental in understanding this process.

In pioneering work by Hernquist (1989), Barnes & Hernquist (1991), and Mihos & Hernquist (1994, 1996), the occurrence of central nuclear starbursts during mergers was studied in detail. Recently, the growth of supermassive black holes has been added to the simulations (Springel, Di Matteo & Hernquist 2005), making it possible to study the coevolution of black holes and stellar bulges in galaxies. Di Matteo, Springel & Hernquist (2005) demonstrated that the energy output associated with accretion regulates the black hole growth and establishes the tight observed relation between black hole masses and bulge velocity dispersions. Subsequently, these simulations have been used to develop comprehensive models of spheroid formation and for interpreting the evolution and properties of the cosmic quasar population (Hopkins et al. 2005, 2006).

4.3. Star Formation and Stellar Encounters

On smaller scales, many SPH-based simulations have studied the fragmentation of molecular clouds, star formation, and the initial mass function (Bate 1998; Klessen, Burkert & Bate 1998; Bate & Bonnell 2005; Smith, Clark & Bonnell 2009). This includes also simulations of the formation of the first stars in the Universe (Bromm, Coppi & Larson 2002; Yoshida et al. 2006; Clark, Glover & Klessen 2008), which are thought to be very special objects.

In this area, important numerical issues related to SPH have been examined as well. Bate & Burkert (1997) show that the Jeans mass needs to be resolved to guarantee reliable results for gravitational fragmentation, a result that was strengthened by Whitworth (1998), who showed that, provided $h \sim \varepsilon$ and the Jeans mass is resolved, only physical fragmentation should occur. Special on-the-fly particle splitting methods have been developed in order to guarantee that the Jeans mass stays resolved over the course of a simulation (Kitsionas & Whitworth 2002, 2007).

Also, the ability of SPH to represent driven isothermal turbulence has been studied (Klessen, Heitsch & Mac Low 2000).

Pioneering work on stellar collisions with SPH has been performed by Benz & Hills (1987) and Benz et al. (1990). Some of the most recent work in this area studied quite sophisticated problems, such as relativistic neutron star mergers with an approximate treatment of general relativity and sophisticated nuclear matter equations of state (Rosswog, Ramirez-Ruiz & Davies 2003; Oechslin, Janka & Marek 2007), or the triggering of subluminous supernovae type Ia through collisions of white dwarfs (Pakmor et al. 2010).

4.4. Planet Formation and Accretion Disks

SPH have been successfully employed to study giant planet formation through the fragmentation of protoplanetary disks (Mayer et al. 2002, 2004) and to explore the interaction of high- and low-mass planets embedded in protoplanetary discs (Bate et al. 2003, Lufkin et al. 2004). A number of works have also used SPH to model accretion disks (Simpson 1995), including studies of spiral shocks in 3D disks (Yukawa, Boffin & Matsuda 1997), accretion of gas onto black holes in viscous accretion disks (Lanzafame, Molteni & Chakrabarti 1998), or cataclysmic variable systems (Wood et al. 2005).

Another interesting application of SPH lies in simulations of materials with different equations of state, corresponding to rocky and icy materials (Benz & Asphaug 1999). This allows the prediction of the outcome of collisions of bodies with sizes from the scale of centimeters to hundreds of kilometers, which has important implications for the fate of small fragments in the Solar System or in protoplanetary disks. Finally, such simulation techniques allow calculations of the collision of protoplanets, culminating in numerical simulations that showed how the Moon might have formed by an impact between protoearth and an object a tenth of its mass (Benz, Slattery & Cameron 1986).

5. CONVERGENCE, CONSISTENCY, AND STABILITY OF SMOOTHED PARTICLE HYDRODYNAMICS

There have been a few code comparisons in the literature between SPH and Eulerian hydrodynamics (Frenk et al. 1999, O'Shea et al. 2005, Tasker et al. 2008), but very few formal studies of the accuracy of SPH have been carried out. Even most code papers on SPH report only circumstantial evidence for SPH's accuracy. What is especially missing are rigorous studies of the convergence rate of SPH toward known analytic solutions, which is ultimately one of the most sensitive tests of the accuracy of a numerical method. For example, this may involve measuring the error in the result of an SPH calculation in terms of an L1 error norm relative to a known solution, as is often done in tests of Eulerian hydrodynamics codes. There is no a priori reason why SPH should not be subjected to equally sensitive tests to establish whether the error becomes smaller with increasing resolution (convergence) and whether the convergence occurs toward the correct physical solution (consistency). Such tests can also provide a good basis to compare the efficiency of different numerical approaches for a particular type of problem with each other.

In this Section, we discuss a number of tests of SPH in one and two dimensions in order to provide a basic characterization of the accuracy of "standard SPH" as outlined in Section 2. Our discussion is, in particular, meant to critically address some of the weaknesses of standard SPH, both to clarify the origin of the inaccuracies and to guide the ongoing search for improvements in the approach. We note that there is already a large body of literature with suggestions for improvements of standard SPH, ranging from minor modifications, say in the parameterization

of the artificial viscosity, to more radical changes, such as outfitting SPH with a Riemann solver or replacing the kernel-interpolation technique with a density estimate based on a Voronoi tessellation. We shall summarize some of these attempts in Section 6, but refrain from testing them in detail here.

We note that for all test results reported below, it is well possible that small modifications in the numerical parameters of the code that was used (GADGET2, Springel 2005) may lead to slightly improved results. However, we expect that this may only reduce the error by a constant factor as a function of resolution, but is unlikely to significantly improve the order or convergence. The former is very helpful of course, but ultimately represents only cosmetic improvements of the results. What is fundamentally much more important is the latter, the order of convergence of a numerical scheme.

5.1. One-Dimensional Sound Waves

Based on analytic reasoning, Rasio (2000) argued that convergence of SPH for 1D sound waves requires increasing both the number of smoothing neighbors and the number of particles, with the latter increasing faster such that the smoothing length and, hence, the smallest resolved scale decreases. In fact, he showed that in this limit the dispersion relation of sound waves in one dimension is correctly reproduced by SPH for all wavelengths. He also pointed out that if the number of neighbors is kept constant and only the number of particles is increased, then one may converge to an incorrect physical limit, implying that SPH is inconsistent in this case.

We note that in practice the requirement of consistency may not be crucially important, provided the numerical result converges to a solution that is close enough to the correct physical solution. For example, a common error in SPH for a low number of neighbors is that the density estimate carries a small bias of up to a percent or so. Although this automatically means that the exact physical value of the density is systematically missed by a small amount, in the majority of astrophysical applications of SPH, the resulting error will be subdominant compared to other errors or approximations in the physical modeling, and is therefore not really of concern. In our convergence tests below, we therefore stick with the common practice of keeping the number of smoothing neighbors constant. Note that because we compare to known analytic solutions, possible errors in the dispersion relation will be picked up by the error measures anyway.

We begin with the elementary test of a simple acoustic wave that travels through a periodic box. To avoid any wave steepening, we consider a very small wave amplitude of $\Delta\rho/\rho = 10^{-6}$. The pressure of the gas is set to $P = 3/5$ at unit density, such that the adiabatic sound speed is $c_s = 1$ for a gas with $\gamma = 5/3$. We let the wave travel once through the box of unit length and compare the velocity profile of the final result at time $t = 1.0$ with the initial conditions in terms of an L1 error norm. We define the latter as

$$\mathrm{L1} = \frac{1}{N}\sum_i |v_i - v(x_i)|, \tag{68}$$

where N is the number of SPH particles, v_i is the numerical solution for the velocity of particle i, and $v(x_i)$ is the expected analytic solution for the problem, which is here identical to the initial conditions. Note that we deliberately pick the velocity as error measure, because the density estimate can be subject to a bias that would then completely dominate the error norm. The velocity information on the other hand tells us more faithfully how the wave has moved and whether it properly returned to its original state.

In **Figure 1**, we show results for the L1 error norm as a function of the number N of equal-mass particles used to sample the domain, for different numbers of smoothing neighbors. Interestingly,

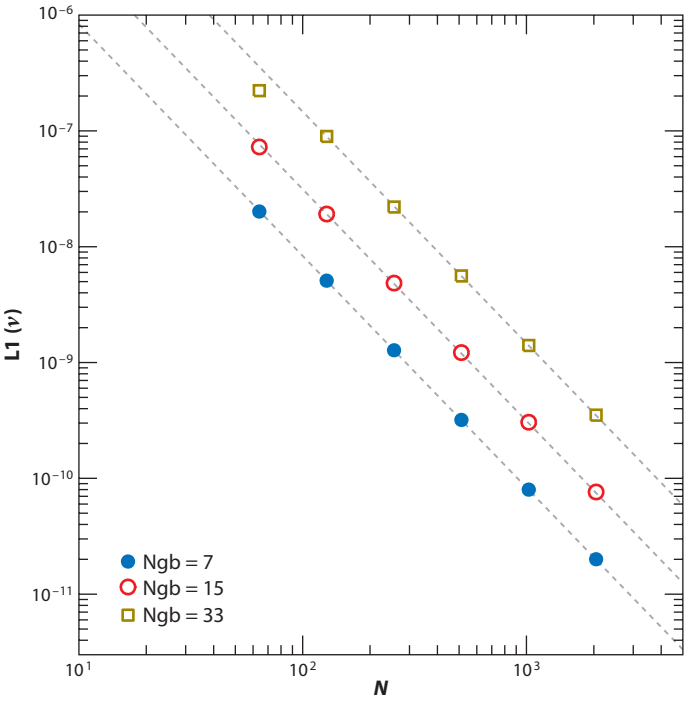


Figure 1

Convergence rate for a traveling sound wave in 1D, calculated in smoothed particle hydrodynamics without artificial viscosity. The solution (*symbols*) converges with second-order accuracy as L1 $\propto N^{-2}$ (this power law is shown with *gray dashed lines*), for different numbers of neighbors.

the results show second-order convergence of the code, with L1 $\propto N^{-2}$, as expected in a second-order accurate scheme for smooth solutions without discontinuities. This convergence rate is the same as the one obtained for this problem with standard state-of-the-art second-order Eulerian methods (e.g., Stone et al. 2008, Springel 2010), which is reassuring. Note that in this test a larger number of neighbors simply reduces the effective spatial resolution but does not lead to an advantage in the L1 velocity norm. Nevertheless, the density is more accurately reproduced for a larger number of neighbors. Because the sound speed depends only on temperature (which is set as part of the initial conditions), the density error apparently does not appreciably affect the travel speed of acoustic waves in this test.

5.2. One-Dimensional Riemann Problems

Let us now consider a few Riemann problems in one dimension. Their initial conditions are characterized by two piece-wise constant states that meet discontinuously at $x = 0.5$ at time $t = 0$. The subsequent evolution then gives rise to a set of self-similar waves containing always one contact wave, which is sandwiched on the left and the right by either a shock wave or a rarefaction wave. The ability to correctly reproduce the nonlinear outcome of arbitrary Riemann problems is the backbone of any hydrodynamical method.

In **Figure 2**, we show SPH results for three different Riemann-type problems, with initial conditions characterized by triples of density, pressure, and velocity for either side, as listed in **Table 1**. For all these problems (which are taken from Toro 1997), $\gamma = 1.4$ has been assumed. Problem 1 gives rise to a comparatively weak shock, which has been studied in very similar form (but with a different value of $P_2 = 0.1795$) in a number of previous tests of SPH (Hernquist & Katz 1989; Rasio & Shapiro 1991; Wadsley, Stadel & Quinn 2004; Springel 2005; among others).

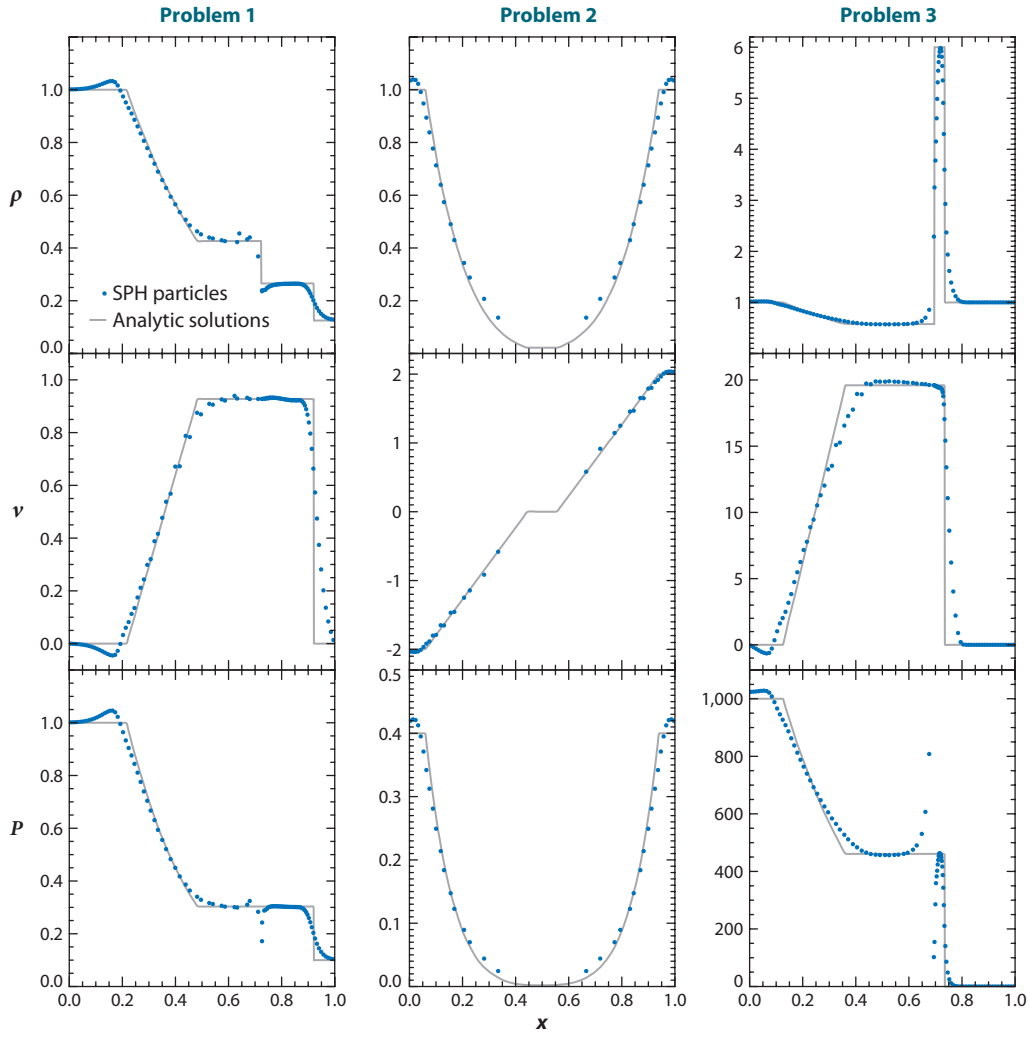


Figure 2

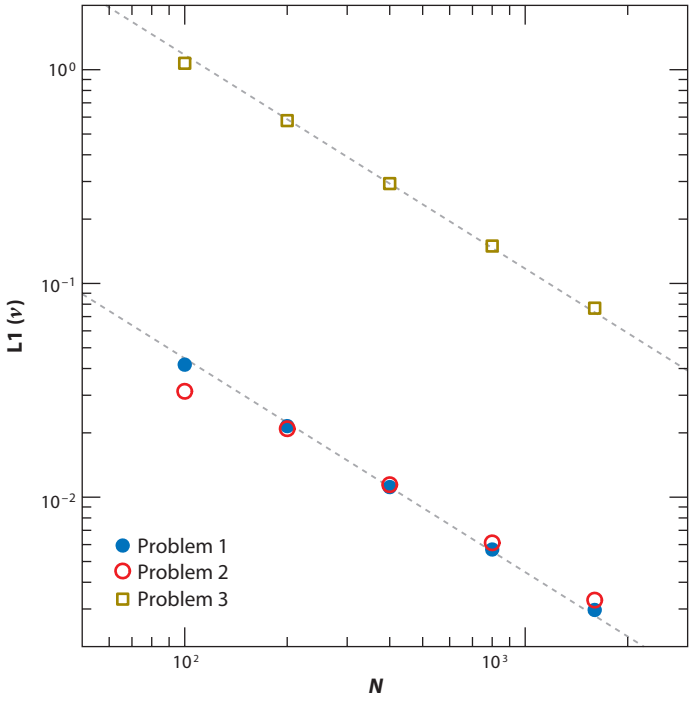
Different 1D Riemann problems, calculated with a resolution of 100 points in the unit domain, and 7 smoothing neighbors. The three columns show results for the initial conditions of problems 1, 2, and 3 as specified in **Table 1**. Symbols represent the smoothed particle hydrodynamics (SPH) particles, and solid lines show the exact solutions for density, velocity, and pressure, from top to bottom.

Table 1 Parameters of the 1D Riemann problems (for $\gamma = 1.4$) examined here

	ρ_1	P_1	v_1	ρ_2	P_2	v_2
Problem 1	1.0	1.0	0.0	0.125	0.1	0.0
Problem 2	1.0	0.4	−2.0	1.0	0.4	2.0
Problem 3	1.0	1000.0	0.0	1.0	0.01	0.0

Figure 3

Convergence rate of 1D Riemann problems (as labeled), calculated with a resolution of 100 points in the unit domain, using 7 smoothing neighbors. The gray dashed lines indicate a L1 $\propto N^{-1}$ scaling of the error.



In the second problem, the gas is suddenly ripped apart with large supersonic velocity, giving rise to a pair of strong rarefaction waves. Finally, problem 3 is again a Sod shock-type problem but involves an extremely strong shock.

The results shown in **Figure 2** are based on 100 points with initially equal spacing in the unit domain, 7 smoothing neighbors, and a standard artificial viscosity setting. All problems are treated qualitatively correctly by SPH, with some inaccuracies at the contact discontinuities. Characteristically, shocks and contact discontinuities are broadened over 2 to 3 smoothing lengths, and rarefaction waves show a small over- and underestimate at their high- and low-density sides, respectively. Also, there is a pressure blip seen at the contact discontinuity. However, the properties of the postshock flow are correct, and the artificial viscosity has successfully suppressed all postshock oscillations. Also, the errors become progressively smaller as the resolution is increased. This is seen explicitly in **Figure 3**, where we show results of the L1 error norm for the velocity as a function of resolution. The error declines as L1 $\propto N^{-1}$, which is expected due to the reduced order of the scheme at the discontinuities. The same convergence rate for the shock-problem is obtained with Eulerian approaches, as they too exhibit only first-order accuracy around discontinuities in the solutions.

Often the discussion of the numerical accuracy of shocks focuses on the sharpness with which they are represented, and on that basis SPH has frequently been portrayed as being inferior in comparison with Eulerian methods. However, it should be noted that in both approaches the numerical shock width is always many orders of magnitude larger than the true width of the physical shock layer. What is much more relevant than the width is therefore that the properties of the postshock flow are correct, which is the case in SPH. Also, the numerical thickness of the shocks may be reduced arbitrarily by using more particles and is, hence, entirely a matter of resolution. We also point out that contact discontinuities can often become quite broad in Eulerian methods; in fact, their width increases with growing advection speed, unlike in the Galilean-invariant SPH.

As far as 1D hydrodynamics is concerned, SPH can hence be characterized as being quite accurate and, in particular, its convergence rate appears competitive with second-order Eulerian schemes.

5.3. Two-Dimensional Shock Waves

Multidimensional hydrodynamics adds much additional complexity, such as shear flows, fluid instabilities, and turbulence. We here first briefly examine whether a problem with 1D symmetry, again a Riemann problem, can be equally well represented with SPH in multiple dimensions. For definiteness, we study again the strong shock of problem 3 from the previous section, but this time with the y dimension added to the initial setup. This is taken as a regular Cartesian $N \times N$ particle grid in the unit square, with periodic boundaries in the y direction and reflective boundaries in the x direction.

In **Figure 4*a***, we show the x velocities of all particles of a run at 100×100 resolution and compare them to the analytic solution. Relative to the corresponding 1D result shown in the middle-right panel of **Figure 2**, the primary difference is a considerably increased noise in the particle velocities. This noise is a generic feature of multidimensional flows simulated with SPH. It can be reduced by an enlarged artificial viscosity or a larger number of neighbors, but typically tends to be much larger than in 1D calculations. Presumably, this is just a consequence of the higher degree of freedom in the particle motion.

Unsurprisingly, the noise has a negative impact on the convergence rate of planar shocks in multidimensional SPH. This is shown for the same problem in **Figure 4**, where we measure $\mathrm{L1} \propto N^{-0.7}$ instead of the ideal $\mathrm{L1} \propto N^{-1.0}$. It also does not appear to help to first bin the particles and then to compute the L1 error as a difference between the averaged and the analytical result. Although this reduces the absolute size of the error substantially and is certainly warranted to eliminate the intrinsic noise, it does not affect the convergence rate itself (see **Figure 4**). Nevertheless, the noise does not destroy the principal correctness of the solution obtained with SPH for multidimensional planar Riemann problems.

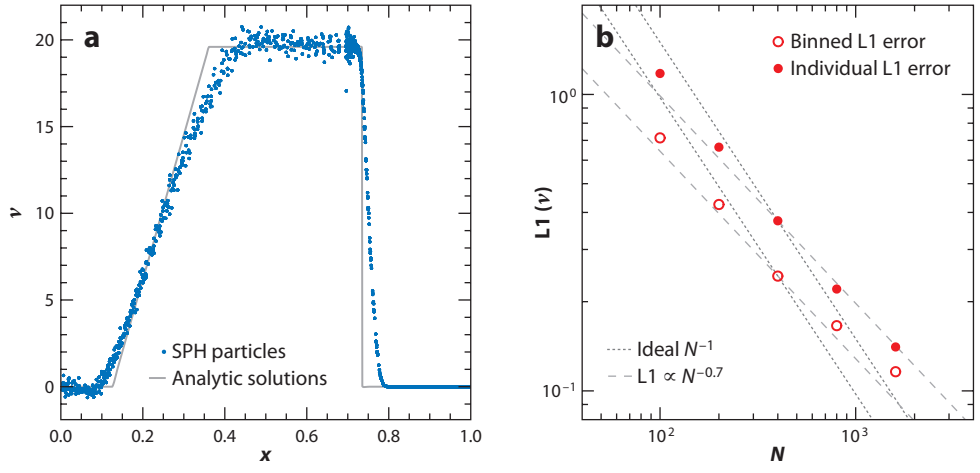


Figure 4

Velocity profile and convergence rate of a 2D strong shock problem. (*a*) Blue circles show the velocities of all smoothed particle hydrodynamics (SPH) particles, compared to the analytic result (*solid gray line*). (*b*) Open red circles show the L1 error for a binned result (using N bins in the x direction in each case); filled red circles give the error when individual particles are considered. In this 2D problem, the ideal N^{-1} convergence rate (*dotted gray lines*) of 1D simulations is reduced to approximately $\mathrm{L1} \propto N^{-0.7}$ (*dashed gray lines*).

5.4. Shear Flows

We now a discuss a yet more demanding 2D problem, the vortex test of Gresho (Gresho & Chan 1990, Liska & Wendroff 2003). It consists of a triangular azimuthal velocity profile,

$$v_\phi(r) = \begin{cases} 5r & \text{for} \quad 0 \le r < 0.2 \\ 2 - 5r & \text{for} \quad 0.2 \le r < 0.4 \\ 0 & \text{for} \quad r \ge 0.4 \end{cases}, \tag{69}$$

in a gas of constant density equal to $\rho = 1$ and an adiabatic index of $\gamma = 5/3$. By adopting the pressure profile,

$$P(r) = \begin{cases} 5 + 25/2r^2 & \text{for} \quad 0 \le r < 0.2 \\ 9 + 25/2r^2 - & \\ 20r + 4\ln(r/0.2) & \text{for} \quad 0.2 \le r < 0.4 \\ 3 + 4\ln 2 & \text{for} \quad r \ge 0.4 \end{cases}, \tag{70}$$

the centrifugal force is balanced by the pressure gradient, and the vortex becomes independent of time.

In **Figure 5**, we compare the results for the azimuthal velocity profiles at time $t = 1.0$ for three different runs, carried out with 80×80 particles in the unit domain for different settings of the artificial viscosity. **Figure 5*a*** shows the outcome for a standard viscosity of $\alpha = 1.0$, **Figure 5*b*** shows $\alpha = 0.05$, and **Figure 5*c*** does not use any artificial viscosity at all. We see that in all three cases, substantial noise in the velocity profile develops, but it is clearly largest in the simulation without viscosity. However, one can see that the average velocity profile of the run without viscosity is actually closest to the expected stationary solution (*blue lines*), whereas the standard viscosity run already shows a reduced angular frequency in the solid-body part of the rotation in the inner part of the vortex. Despite the use of the Balsara switch in this problem, the velocity noise, and as a consequence the noisy estimates of divergence and curl, has produced enough residual viscosity to lead to appreciable angular momentum transport.

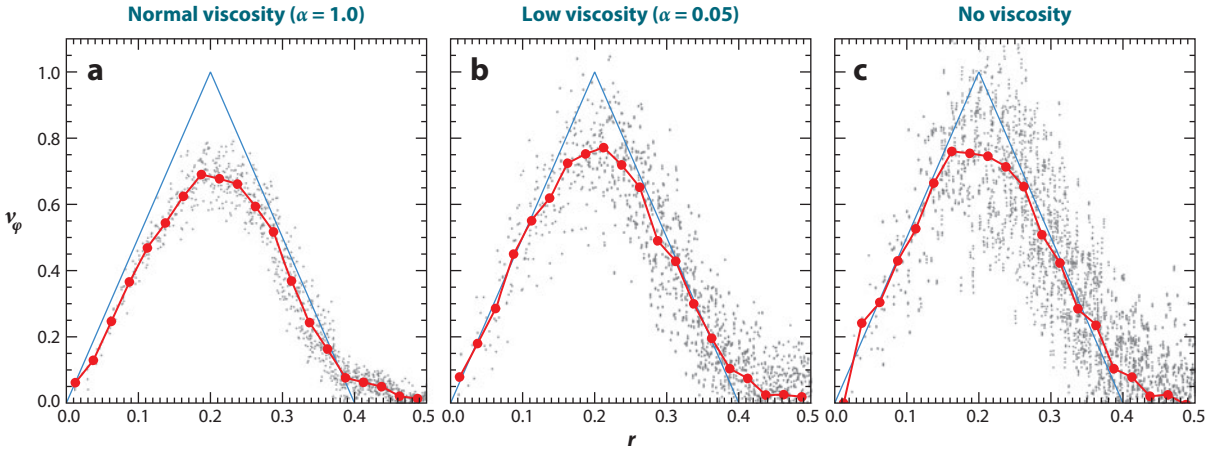


Figure 5

Velocity profile in the Gresho vortex test at time $t = 1.0$, at a base resolution of 80×80, for different viscosity settings. Here, low viscosity corresponds to $\alpha = 0.05$, and normal viscosity to $\alpha = 1.0$, with the Balsara switch enabled in both cases. The small gray dots are the azimuthal velocities of individual smoothed particle hydrodynamics particles, whereas the filled red circles show binned results. The blue triangular-shaped profile is the analytic (stationary) solution.

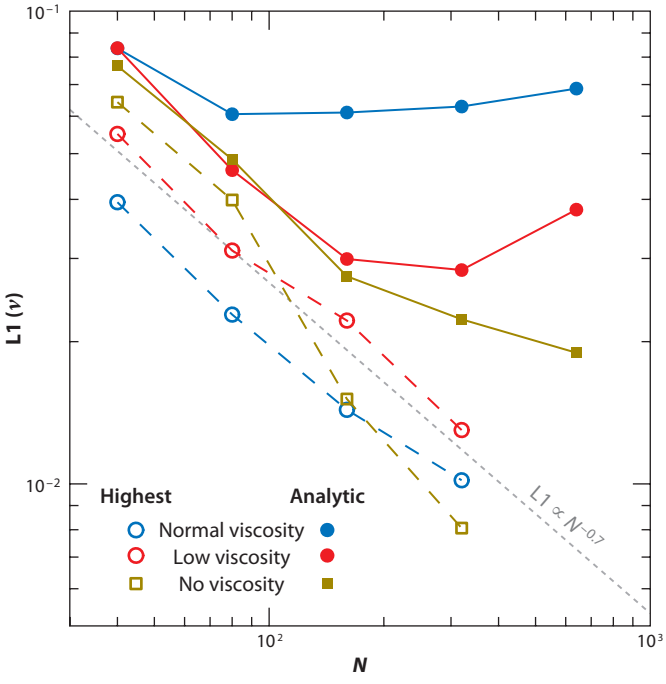


Figure 6

Convergence rate of the Gresho problem, in terms of the binned azimuthal velocity, for different viscosity settings. Here the open symbols measure the L1 error toward the highest resolution result, whereas the filled symbols measure the difference with respect to the analytic solution. The dotted line indicates a $L1 \propto N^{-0.7}$ convergence.

This is corroborated by the results of convergence tests for this problem. In **Figure 6**, we consider the L1 error norm of the binned azimuthal velocity profile (based on $N/2$ bins, linearly placed in the radial range of 0 to 0.5). Interestingly, the two runs with nonvanishing viscosity do not converge to the analytic solution with higher resolution. Instead, they converge to a different solution. If we approximately identify this solution as the highest resolution result in each case, then we see that the lower resolution calculations converge to it with $L1 \propto N^{-0.7}$. What is important to note is that the effects of the artificial viscosity on the shear flow do not diminish with higher resolution. Instead, the solutions behave as if one had simulated a fluid with some residual shear viscosity instead. Only the run without viscosity appears to converge to the correct physical solution, albeit at a very low rate for high N. In any case, it appears that the convergence rate of SPH for this problem is considerably worse than the $L1 \propto N^{-1.4}$ measured for a moving-mesh code and Eulerian codes by Springel (2010).

5.5. Contact Discontinuities and Fluid Instabilities

An issue that has created considerable attention recently is the question of whether SPH can properly resolve fluid instabilities, such as the Kelvin-Helmholtz (KH) instability in shear flows. For a setup with equal particle masses and a sharp initial density contrast of 1:2 at a contact discontinuity, Agertz et al. (2007) did not observe any growth of the instability, whereas for a vanishing density jump the fluids would start to mix. Agertz et al. (2007) attributed these problems to substantial errors in the pressure gradient estimates at the contact discontinuity. A number of recent studies have addressed this problem as well, proposing modifications of SPH designed to improve the KH results. In particular, Price (2008) has suggested adding an artificial heat conduction to smooth out the phase boundary, which indeed improved the results. A different approach was followed by Read, Hayfield & Agertz (2009), who employed a different kernel, a much enlarged number of

neighbors, and a modified density estimation formula to obtain a better representation of mixing of different phases in shear layers. We do not examine these modifications here but want to shine more light onto the nature of the inaccuracies of standard SPH for the KH problem.

We adopt the KH test parameters of Springel (2010), that is, a central phase in the region $|y - 0.5| < 0.25$ is given density $\rho_2 = 2$ and velocity $v_x = 0.5$, whereas the rest of the gas has density $\rho_1 = 1$ and velocity $v_x = -0.5$, all at the same pressure of $P = 2.5$ with $\gamma = 5/3$. In order to soften the transition at the two interfaces, we follow Robertson et al. (2010) and impose a smooth transition in the vertical profiles of density $\rho(y)$, shear velocity $v_x(y)$, and specific entropy $A(y) = P/\rho(y)^\gamma$. For example, the initial density structure is adopted as

$$\rho(y) = \rho_1 + \frac{\rho_2 - \rho_1}{\left[1 + \exp(-2(y - 0.25)/\sigma)\right]\left[1 + \exp(2(y - 0.75)/\sigma)\right]}, \tag{71}$$

with $\sigma = 0.025$, and similar adoptions are made for the other profiles. The transition removes the sharp discontinuities from the initial conditions, such that they can in principle be fully resolved by the numerical scheme. This regularization makes the problem well posed for convergence studies, which is particularly important for the KH instability, where smallest scales grow fastest.

For simplicity, we consider a Cartesian particle grid in the unit domain with periodic boundaries and trigger the instability by applying the velocity perturbation throughout the y-domain, as

$$v_y(x) = v_0 \sin(k\,x), \tag{72}$$

with an initial amplitude of $v_0 = 0.01$ and a wave number of $k = 2 \times (2\pi)/L$. We can easily study the growth of this velocity mode by measuring its amplitude through a Fourier transform of the 2D v_y velocity field.

In **Figure 7**, we show density maps of the evolved fields at time $t = 2.0$ for different SPH simulations, using 27 neighbors and no artificial viscosity. As can be seen, the KH instability

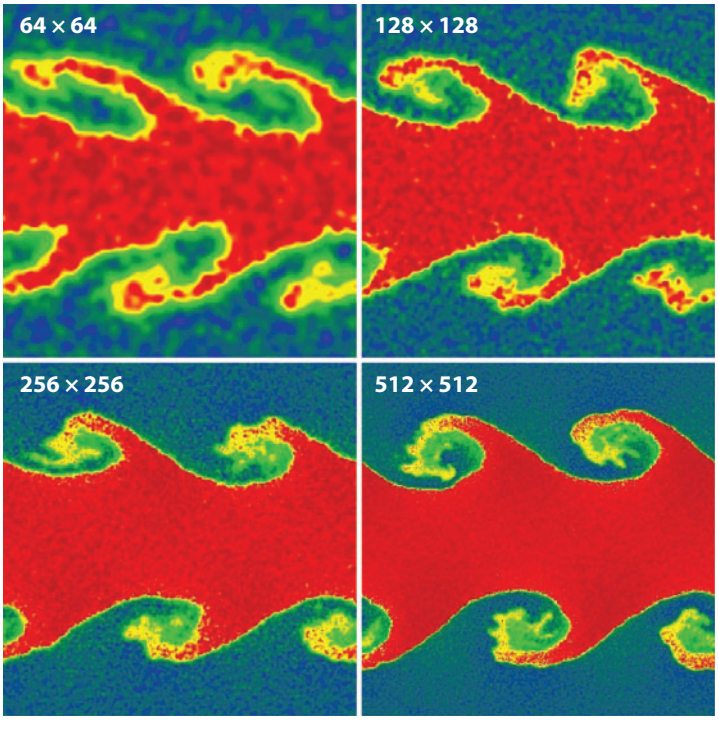


Figure 7

Density fields of the Kelvin-Helmholtz instability test at $t = 2.0$, simulated with smoothed particle hydrodynamics for different resolutions, as labeled.

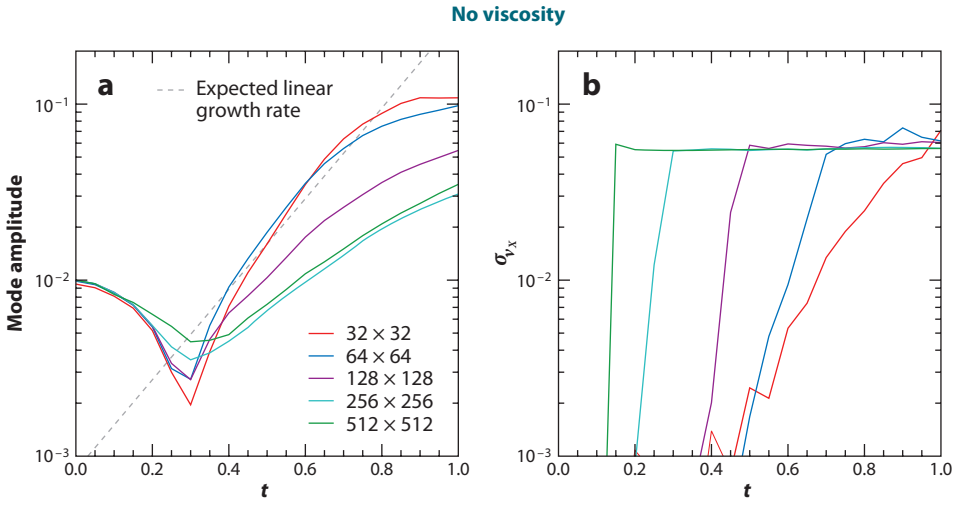


Figure 8

Growth rate of Kelvin-Helmholtz instabilities in smoothed particle hydrodynamics (SPH) test simulations for different resolutions as labeled. (*a*) The amplitude of the initially excited velocity mode, compared to the expected linear growth rate (*dashed line*). (*b*) The velocity dispersion of the SPH particles in the x direction as a function of time for a thin layer close to the midplane of the box. This dispersion is an indication of the level of velocity noise in the calculation.

clearly develops and produces the characteristic wave-like features. The density field is visibly noisy though, as a result of the absence of artificial viscosity. If the latter is added, the noise is substantially reduced, but the KH billows look much more anemic and have a reduced amplitude.

An interesting question is whether the KH instability actually grows with the right rate in these SPH calculations. This is examined in **Figure 8**, where the amplitude of the excited v_y mode of the velocity field is shown as a function of time for the case without viscosity. The growth is compared to the expected exponential growth rate $v_y \propto \exp(t/\tau_{\rm KH})$, shown as a dashed line, where

$$\tau_{\rm KH} = \frac{\rho_1 + \rho_2}{|v_2 - v_1| k \sqrt{\rho_1 \rho_2}} \tag{73}$$

is the KH growth timescale for an inviscid gas. After an initial transient phase (which is expected because the initial perturbation was not set up self-consistently), the two low-resolution calculations with 32^2 and 64^2 particles do actually reproduce this growth rate quite accurately for some time, but then the growth rate suddenly slows down considerably. Curiously, the higher resolution simulations never quite reach the expected growth rate. So what is going on here?

A hint is obtained by the results shown in **Figure 8*b***, where the v_x velocity dispersion of the particles in a narrow strip around $y \simeq 0.5$ is shown as a function of time. We, in principle, expect this to remain close to zero until very late in the evolution. However, what we actually observe is a sudden increase of this dispersion, when the regularity and coherence of the initial Cartesian particle grid are finally lost. At this moment, SPH develops its velocity noise, which is characteristic even of smooth flows. Interestingly, we observe that the development of this velocity noise closely coincides in time with the termination of the correct linear growth rate. The natural interpretation is that as long as the particles are still quite regularly ordered, the analytic linear KH growth rate is calculated accurately by SPH, but at late times the developing noise introduces a substantial degradation of the accuracy that reduces the proper growth rate.

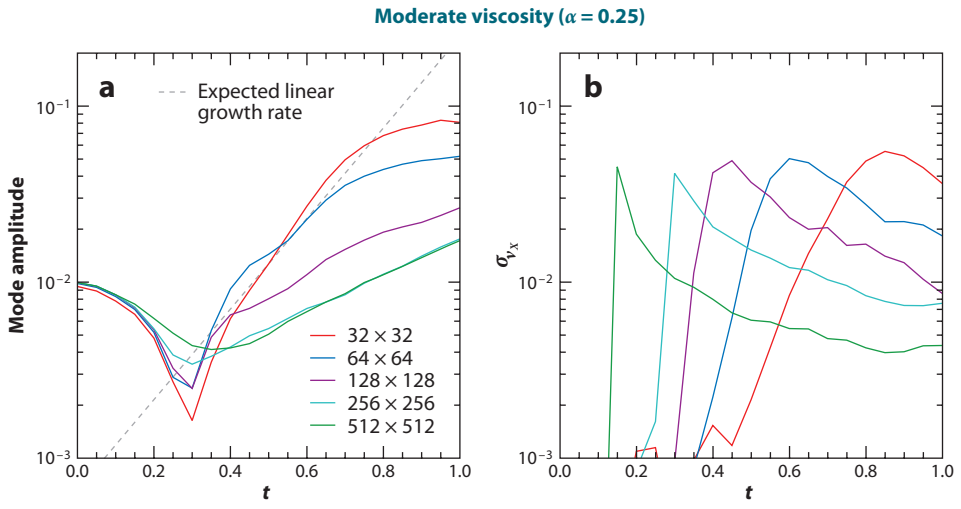


Figure 9

The same as in **Figure 8**, but for simulations that include an artificial viscosity of moderate strength with $\alpha = 0.25$.

This suggests that perhaps the use of the right amount of artificial viscosity may yield an accurate growth rate even at late times if it prevents the detrimental noise effects. In **Figure 9**, the equivalent results are shown with artificial viscosity included. Here, the behavior is rather similar initially. As long as the regular order of the particles is still approximately maintained, the artificial viscosity is not changing anything as it is reduced by the Balsara switch to a very low level. Once the initial order decays, the velocity dispersion suddenly increases, first to the same level that one would have without viscosity. Only after some time, the viscosity damps out this noise, but this comes at the price of also damping the growth rate even further, because the fluid behaves now in a slightly viscous fashion.

We conclude that whereas KH fluid instabilities do occur in SPH, calculating them with high accuracy is a challenge due to the significant noise present in the multidimensional velocity field. Getting rid of this noise with artificial viscosity tends to reduce the KH growth rate below the analytic expectation, making it hard to achieve truly inviscid behavior. More work on this problem is therefore clearly warranted in the future.

5.6. Surface Tension

In the standard energy- and entropy-conserving formulation of SPH, two phases of particles with different specific entropies tend to avoid mixing simply because this is energetically disfavored. To understand the origin of this effect, let us consider a simple virtual experiment where a box of volume V is filled in one half with gas of density ρ_1 and in the other half with gas of density ρ_2. We assume an equal pressure P, such that the SPH particles in the different phases will have specific entropies $A_1 = P/\rho_1^\gamma$ and $A_2 = P/\rho_2^\gamma$. Then the total thermal energy in the box will be $E_{\rm therm} = u_1 M_1 + u_2 M_2 = PV/(\gamma - 1)$. The numerical estimate for a particular SPH realization may be slightly different from this due to intermediate density values in the transition layer, but this effect can be made arbitrarily small if a large number of particles is used. Now imagine that we rearrange the particles by homogeneously spreading them throughout the volume, in

some regular fashion (say as two interleaved grids), but keeping their initial entropies. Provided the SPH smoothing lengths are large enough, each particle will then estimate the mean density $\bar{\rho} = (\rho_1+\rho_2)/2$ as its new density throughout the volume. The new thermal energy per unit mass for particles of species 1 will then be $u'_1 = A_1\bar{\rho}^{\gamma-1}/(\gamma-1)$, and this is similar for particles of species 2. As a result, the new thermal energy becomes $E'_{\rm therm} = [PV/(\gamma-1)][(\rho_1+\rho_2)/2]^{\gamma-1}(\rho_1^{1-\gamma}+\rho_2^{1-\gamma})/2$, which is larger than the original energy if $\rho_1 \neq \rho_2$. Clearly then, the mixing of the particle set in this fashion is energetically forbidden, and an energy conserving code will resist it. This resistance appears as an artificial surface tension term in SPH.

We note that the mixing can be accomplished at constant thermal energy, but this requires that the entropies of the particles and, hence, their temperature estimates be made equal. The final entropy is then $\bar{A} = P/\bar{\rho}^{\gamma}$, which corresponds to a total thermodynamic entropy that is larger than in the unmixed state. Mixing the phases in this irreversible fashion hence requires creating the relevant amount of mixing entropy, but for this no source is foreseen in the entropy-conserving formulation of standard SPH. Including artificial thermal conduction (as in Price 2008) is one interesting approach to address this problem, as this process equilibrates the temperatures while conserving energy and increasing the entropy.

The presence of some level of surface tension in SPH can be demonstrated experimentally through simple settling tests (Hess & Springel 2010). For example, one possibility is to set up an overdense spherical region and let it relax to an equilibrium distribution (this can be done by keeping the particle entropies fixed and by adding an artificial decay of the velocities). Then the pressure inside of the sphere in the final relaxed state is found to be slightly higher than outside, as it needs to be to balance the surface tension. In fact, according to the Young-Laplace equation, the expected pressure difference is

$$\Delta P = \sigma\left(\frac{1}{R_1}+\frac{1}{R_2}\right), \tag{74}$$

where σ is the surface tension, and R_1 and R_2 are the two principal radii of curvature of the surface. In **Figure 10**, we show the outcome of such an experiment in 2D, for densities $\rho_2 = 2$ and $\rho_1 = 1$ realized with equal mass particles, a pressure of $P = 2.5$, and 13 smoothing neighbors. Independent of whether the high- or low-density phase is arranged to be inside the sphere, the inner region shows a small pressure difference relative to the outer region, which is $\Delta P \sim 0.002$ at an effective resolution of $N^2 = 128^2$ for the lower density phase. For a radius of $R_1 = 0.4$ of the sphere, this then implies a surface tension of $\sigma = 0.0025$. We have confirmed that this surface tension varies with 1D resolution as $\sigma \propto 1/N$; that is, it declines with higher spatial resolution. However, it increases for higher density ratios and for a larger number of smoothing neighbors.

One consequence of this tension is that SPH in principle may support capillary waves at interfaces, with dispersion relation $\omega^2 = \sigma k^3/(\rho_1+\rho_2)$. In the context of the KH problem considered above, we note that surface tension may also modify the growth rate of small wavelength perturbations. In the presence of surface tension, the KH growth timescale becomes

$$t_{\rm KH} = \left[\frac{k^2\rho_1\rho_2(v_2-v_1)^2}{(\rho_1+\rho_2)^2} - \frac{\sigma\, k^3}{\rho_1+\rho_2}\right]^{-1/2}, \tag{75}$$

and waves with wavelength

$$\lambda < \lambda_{\rm crit} = 2\pi\,\frac{\rho_1+\rho_2}{\rho_1\rho_2}\,\frac{\sigma}{(v_2-v_1)^2} \tag{76}$$

will not grow at all. Using the value measured for σ for the numerical setup we considered above, we obtain $\lambda_{\rm crit}/d \simeq 3.0/(v_2-v_1)^2$, where d is the mean particle separation. For a shear of $v_2 - v_1 = 1$, we hence expect only waves with wavelengths up to a couple of particle separations

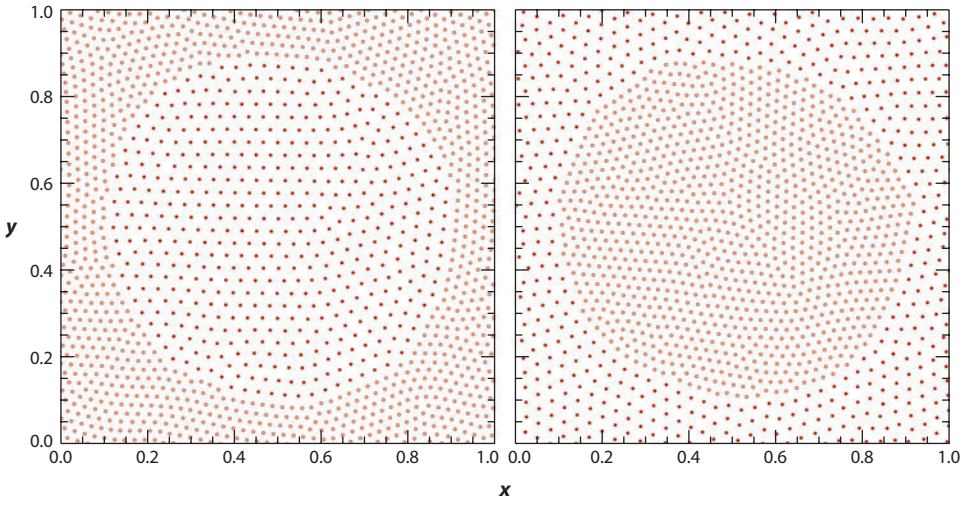


Figure 10

Equilibrium configurations of particles in 2D for two phases of different entropy, corresponding to a jump in density by a factor of 2. Both in the left and in the right cases, the final pressure in the inside of the sphere is slightly higher than the one on the outside. This pressure difference offsets the spurious surface tension present at the interface.

to be suppressed. In particular, this effect should not have influenced the KH tests carried out in the previous section. However, for small shear, the effect can be much more of a problem, and here may stabilize SPH against the KH instability.

5.7. The Tensile Instability

First described by Swegle (1995), the clumping or tensile instability is a well-known nuisance in SPH that occurs if a large number of smoothing neighbors is used. In this case, it can happen that the net force between a close particle pair is not repulsive but attractive simply because the kernel gradient tends to become shallow for close separations such that the pair may be further compressed by other surrounding particles. As a result, particles may clump together, thereby reducing the effective spatial resolution available for the calculation.

Settling tests, in which an initially random distribution of particles with equal and fixed entropies is evolved toward an equilibrium state under the influence of a friction force, are a quite sensitive tool to reveal the presence of this instability. We show an example of the outcome of such a test in **Figure 11**, where 50^2 particles were randomly placed in a periodic 2D box of unit length and then evolved until a good equilibrium of the pressure forces was reached, corresponding to a low-energy state of the system. **Figure 11*a*** shows the result for 13 smoothing neighbors, **Figure 11*b*** shows the result for 27 neighbors, and **Figure 11*c*** shows the result for 55 neighbors. Although for a low number of neighbors a nicely regular, hexagonal particle distribution is obtained, for a large number of smoothing neighbors many particle clumps are formed, strongly reducing the number of independent sampling points for the fluid.

Taken at face value, this seems to make attempts to use a large number of neighbors to increase the accuracy of SPH calculations a futile exercise. However, we note that the clumping instability is essentially never observed in this extreme form in real applications of SPH. This is because the dissipation added in the settling test is not present in a real application, and there the dynamics is

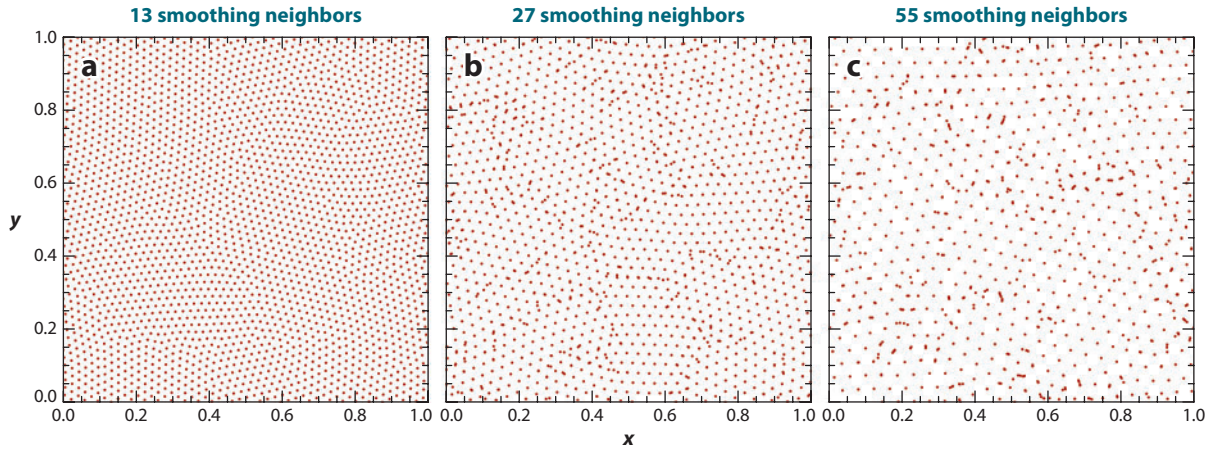


Figure 11

Clumping instability in a 2D settling test, using (*a*) 13, (*b*) 27, or (*c*) 55 smoothing neighbors. In all three cases, the number of particles used (50^2) is the same. It is seen that for a large number of smoothing neighbors, particles clump into groups, thereby reducing the effective spatial resolution of the scheme.

usually also not cold enough to create many artificial SPH particle clumps or string-like configurations. Nevertheless, a number of suggestions have been made to fix the tensile instability through modifications of the SPH formalism. Steinmetz (1996) simply invokes an artificially steepened first derivative of the kernel to obtain a larger repulsive pressure force at small separations. Monaghan (2000) proposes instead an additional artificial pressure, which has the advantage of maintaining the conservative properties of SPH, but it introduces some small errors in the dispersion relation. Finally, Read, Hayfield & Agertz (2009) suggest the use of a peaked kernel, thereby ensuring large repulsive pressure forces at small separations at the price of a lower order of the density estimate.

6. ALTERNATIVE FORMULATIONS OF SMOOTHED PARTICLE HYDRODYNAMICS AND FUTURE DIRECTIONS

There are many avenues for modifications of the standard SPH method, outlined in Section 2, even though this usually implies giving up the elegance of the Lagrangian derivation in favor of a more heuristic construction of the discretization scheme. However, what ultimately matters in the end is the achieved accuracy, and hence there is certainly a lot of merit in investigating such alternative schemes. We here briefly mention some of the proposed modifications.

6.1. Modifications of the Kernel Estimation Scheme

Changing the basic SPH kernel shape has been investigated in a number of studies (Monaghan 1985; Fulk 1996; Cabezón, García-Senz & Relaño 2008), but generally no kernel has been found that performed significantly better than the cubic spline. In particular, higher order kernels may both be positive and negative, which invokes numerous stability problems.

A few studies (Shapiro et al. 1996, Owen et al. 1998) proposed anisotropic kernels that adapt to local flow features. Together with the use of a tensorial artificial viscosity, this promises an optimum use of the available spatial resolution and, hence, ultimately higher accuracy for a given number of particles.

In the context of galaxy-formation simulations and the so-called overcooling problem, it has been pointed out that SPH's density estimate in the vicinity of contact discontinuities may be problematic (Pearce et al. 1999). This led Ritchie & Thomas (2001) to suggest a modified density estimate specifically designed to work better across contact discontinuities. To this extent, particles are weighted with their thermal energies in the density estimate, which works well across contacts. In particular, this removes the pressure blip that is often found in contact waves that develop in Riemann problems. This idea has also been recently reactivated by Read, Hayfield & Agertz (2009) and combined with the suggestion of a much larger number of neighbors and a modified shape of the smoothing kernel.

A more radical approach has been proposed by Børve, Omang & Trulsen (2001), who introduced the so-called regularized SPH formalism. In it the discretization errors of the ordinary kernel estimation are reduced by introducing a new concept of intermediate interpolation cells. They are also used for periodic redefinitions of the whole particle distribution to obtain a better numerical description of the dynamics, in particular of MHD-related problems.

Another attempt to reduce the errors associated with kernel sums is given by Imaeda & Inutsuka (2002) in terms of their "consistent velocity method." The central idea of this scheme is to achieve consistency between the continuity equation and the density estimate. In some sense, this makes it reminiscent of the simpler idea of "XSPH," originally introduced by Monaghan (1989) and later derived in more accurate form from a Lagrangian (Monaghan 2002). In this scheme, the particles are moved with a SPH-interpolated version of the velocity and not with the individual particle velocities. This has been used in some simulations to reduce problems of particle interpenetration. Finally, a higher order formulation of SPH has also been proposed (Oger et al. 2007). It has been shown to eliminate the tensile instability, but at the price of somewhat larger errors in the gradient estimates.

6.2. Improved Artificial Viscosity

The idea to use a time-variable coefficient for the parameterization of the artificial viscosity is in general very interesting. It can help to restrict the viscosity to regions where it is really needed (that is, in shocks), though it should be reduced to very small levels everywhere else, ideally to the lowest level still consistent with keeping the velocity noise of the scheme under control. There is clearly still room for improvement of the switches used at present to increase or decrease the viscosity, as discussed in Section 2.3.

Attempts to outfit SPH with a Riemann solver go a step further (e.g., Inutsuka 2002, Cha & Whitworth 2003), making SPH more similar to the Godunov approach used in Eulerian hydrodynamics. If successful, this may in principle eliminate the need for an artificial viscosity entirely, which would be a very welcome feature. So far, these versions of Godunov-based SPH have not found widespread application though, presumably an indication that still more work is required to make them equally robust as the standard SPH formulation in all situations.

6.3. Modeling Mixing

As discussed earlier, the standard form of SPH has no mechanism to equilibrate the specific entropies of neighboring particles, a property that does not allow different phases to mix into a perfectly homogeneous phase where all particles have the same thermodynamic properties. As seen above, this even causes an artificial surface tension effect. It has been suggested that the issue of mixing is ultimately the primary cause for the differences seen in the central entropy profiles of galaxy clusters simulated with SPH and mesh-based Eulerian techniques (Mitchell et al. 2009).

Recently, there have been a few suggestions to treat mixing in SPH by invoking artificial heat conduction. Price (2008) has shown that this improves the results for KH instability tests when there are sharp discontinuities in the initial conditions. Also, in a study by Rosswog & Price (2007), it was found that this improves the accuracy of very strong shocks in Sedov-Taylor blast waves. One problem is to parameterize the heat conductivity such that it only affects the calculations where it is really needed and not anywhere else. The parameterization by Price (2008), for example, also operates for static contact discontinuities, where it is not clear that any mixing should occur in such a case. A more physical parameterization of the diffusive mixing has been suggested by Wadsley, Veeravalli & Couchman (2008), who use an estimate of the local velocity shear to parameterize the mixing rate.

Yet another approach to mixing has been recently proposed by Read, Hayfield & Agertz (2009), who do not attempt to make the thermodynamic properties of a mixed fluid equal at the particle level. Instead, the number of smoothing neighbors is drastically enlarged by an order of magnitude, such that resolution elements of the mixed phases are sampled by a very large number of particles.

6.4. Alternative Fluid Particle Models

One possibility to substantially modify SPH lies in replacing the kernel interpolation technique of SPH by a different technique to estimate densities. This idea has been explored in the fluid particle model of Hess & Springel (2010), where an auxiliary Voronoi tessellation was used to define a density estimate based on the mass of a particle divided by its associated Voronoi volume. The latter is simply the region of space that lies closer to the point than to any other particle. The same Lagrangian as in Equation 12 can then be used to derive the equations of motion of the new particle-based scheme, where now no smoothing kernel is necessary. This highlights that the kernel interpolation technique in principle only really enters in the density estimate of SPH. If the density estimate is replaced with another construction, different particle-based hydrodynamical schemes can be constructed that are nevertheless similar in spirit to SPH. The Voronoi-based scheme of Hess & Springel (2010) does not show the surface tension effect and has the advantage of leading to a consistent partitioning of the volume. It may also be more amenable to Riemann-solver-based approaches to avoid the need of artificial viscosity, because an unambiguous interaction area between two particles (namely the area of the facet shared by the two Voronoi cells) is given. However, it is not clear whether it can significantly reduce the noise inherent in multidimensional SPH calculations.

A considerable step further is the moving-mesh technique devised by Springel (2010). This also exploits a Voronoi mesh, but solves the fluid equations with the finite-volume approach of the traditional Eulerian Godunov approach, where a Riemann solver is used to work out fluxes across boundaries. This is really a Eulerian scheme, but adjusted to work with particle-based fluid cells. In fact, as Springel (2010) discusses, this approach is identical to the well-known Eulerian MUSCL scheme (van Leer 1984) when the mesh-generating points are kept fixed. If they are allowed to move with the flow, the scheme behaves in a Lagrangian way similarly to SPH. Nevertheless, in this mode the conceptual differences with respect to SPH also become most apparent. The first difference is that the errors in the discrete kernel sum are avoided through the use of an explicit second-order accurate spatial reconstruction. The second difference is that the cells/particles may exchange not only momentum, but also mass and specific entropy, allowing already for the mixing that may happen in multidimensional flow.

7. CONCLUSIONS

Smoothed particle hydrodynamics is a remarkably versatile approach to model gas dynamics in astrophysical simulations. The ease with which it can provide a large dynamic range in spatial resolution and density, as well as an automatically adaptive resolution, is unmatched in Eulerian methods. At the same time, SPH has excellent conservation properties, not only for energy and linear momentum, but also for angular momentum. The latter is not automatically guaranteed in Eulerian codes, even though it is usually fulfilled at an acceptable level for well-resolved flows. When coupled to self-gravity, SPH conserves the total energy exactly, which is again not manifestly true in mesh-based approaches to hydrodynamics. Finally, SPH is Galilean invariant and free of any errors from advection alone, which is another significant advantage compared to Eulerian approaches.

Thanks to its completely mesh-free nature, SPH can easily deal with complicated geometric settings and large regions of space that are completely devoid of particles. Implementations of SPH in a numerical code tend to be comparatively simple and transparent. For example, it is readily possible to include passively advected scalars in SPH (for example, chemical composition) in a straightforward and simple way. At the same time, the scheme is characterized by remarkable robustness. For example, negative densities or negative temperatures, sometimes a problem in mesh-based codes, cannot occur in SPH by construction. Although shock waves are broadened in SPH, the properties of the postshock flow are correct. Also, contact discontinuities can even be narrower than in mesh-based codes.

All of these features make SPH a very interesting method for many astrophysical problems. Indeed, a substantial fraction of the simulation progress made in the past two decades on understanding galaxy formation, star formation, and planet formation has become possible thanks to SPH.

The main disadvantage of SPH is clearly its limited accuracy in multidimensional flows. One source of noise originates in the approximation of local kernel interpolants through discrete sums over a small set of nearest neighbors. While in 1D the consequences of this noise tend to be quite benign, particle motion in multiple dimensions has a much higher degree of freedom. Here the mutually repulsive forces of pressurized neighboring particle pairs do not easily cancel in all dimensions simultaneously, especially not given the errors of the discretized kernel interpolants. As a result, some jitter in the particle motions readily develops, giving rise to velocity noise up to a few percent of the local sound speed. This noise in SPH is likely also a primary cause for the mixed results obtained thus far for MHD techniques implemented on top of SPH despite the considerable effort invested to make them work accurately.

In some sense it may seem surprising that SPH still models acceptable fluid behavior despite the presence of this comparatively large noise. However, because SPH accurately respects the conservation laws of fluid dynamics as described by the Lagrangian, the local fluctuations tend to average out, enforcing the correct large-scale motion of the fluid. Fulfilling the conservation laws is thus clearly more important than a high order of the underlying scheme. However, as a consequence of this noise, individual SPH particles cannot be readily interpreted as precise tracers of the local state of the fluid at their representative location. Instead one needs to carry out some sort of averaging or binning procedure first. In this sense, SPH then indeed exhibits a kind of Monte Carlo character, as manifested also in its comparatively slow convergence rate for multidimensional flow.

The noise inherent in SPH can be reduced by using a larger artificial viscosity and/or a higher number of smoothing neighbors. Both approaches do work at some level, but they are not without caveats and, hence, need to be used with caution. The standard artificial viscosity makes the

simulated gas slightly viscous, which can affect the calculated solutions. The results will then in fact not converge to the solution expected for an inviscid gas, but to those of a slightly viscous gas, which is really described by the appropriate Navier-Stokes equations and not the Euler equations. However, improved schemes for artificial viscosity, such as a time-dependent artificial viscosity with judiciously chosen viscosity triggers, can improve on this substantially. Simply using a larger number of smoothing neighbors is computationally more costly and invokes the danger of suffering from artificial particle clumping. Here, modified kernel shapes and different viscosity prescriptions may provide a satisfactory solution.

In the future, it will be of primary importance to make further progress in understanding and improving the accuracy properties of SPH in order to stay competitive with the recently matured adaptive-mesh-refining and moving-mesh methodologies. If this can be achieved, the SPH technique is bound to remain one of the primary workhorses in computational astrophysics.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I apologize in advance to all researchers whose work could not be cited due to space limitations. I would like to thank Lars Hernquist, Simon White, Stephan Rosswog, and Klaus Dolag for very helpful comments on this review.

LITERATURE CITED

Agertz O, Moore B, Stadel J, Potter D, Miniati F, et al. 2007. *MNRAS* 380:963–78
Altay G, Croft RAC, Pelupessy I. 2008. *MNRAS* 386:1931–46
Ayal S, Piran T, Oechslin R, Davies MB, Rosswog S. 2001. *Ap. J.* 550:846–59
Bagla JS, Khandai N. 2009. *MNRAS* 396:2211–27
Balsara DS. 1995. *J. Comput. Phys.* 121:357–72
Barnes J, Hut P. 1986. *Nature* 324:446–49
Barnes JE, Hernquist LE. 1991. *Ap. J. Lett.* 370:L65–68
Bate MR. 1998. *Ap. J. Lett.* 508:L95–98
Bate MR, Bonnell IA. 2005. *MNRAS* 356:1201–21
Bate MR, Burkert A. 1997. *MNRAS* 288:1060–72
Bate MR, Lubow SH, Ogilvie GI, Miller KA. 2003. *MNRAS* 341:213–29
Benz W, Asphaug E. 1999. *Icarus* 142:5–20
Benz W, Cameron AGW, Press WH, Bowers RL. 1990. *Ap. J.* 348:647–67
Benz W, Hills JG. 1987. *Ap. J.* 323:614–28
Benz W, Slattery WL, Cameron AGW. 1986. *Icarus* 66:515–35
Borgani S, Murante G, Springel V, Diaferio A, Dolag K, et al. 2004. *MNRAS* 348:1078–96
Børve S, Omang M, Trulsen J. 2001. *Ap. J.* 561:82–93
Brandenburg A. 2010. *MNRAS* 401:347–54
Bromm V, Coppi PS, Larson RB. 2002. *Ap. J.* 564:23–51
Brookshaw L. 1985. *Proc. Astron. Soc. Aust.* 6:207–10
Cabezón RM, García-Senz D, Relaño A. 2008. *J. Comput. Phys.* 227:8523–40
Cha S, Whitworth AP. 2003. *MNRAS* 340:73–90
Clark PC, Glover SCO, Klessen RS. 2008. *Ap. J.* 672:757–64
Cleary PW, Monaghan JJ. 1999. *J. Comput. Phys.* 148:227–64

da Silva AC, Barbosa D, Liddle AR, Thomas PA. 2000. *MNRAS* 317:37–44
Davé R, Cen R, Ostriker JP, Bryan GL, Hernquist L, et al. 2001. *Ap. J.* 552:473–83
Davé R, Hernquist L, Katz N, Weinberg DH. 1999. *Ap. J.* 511:521–45
Dedner A, Kemm F, Kröner D, Munz C-D, Schnitzer T, Wesenberg M. 2002. *J. Comput. Phys.* 175:645–73
Di Matteo T, Springel V, Hernquist L. 2005. *Nature* 433:604–7
Dolag K, Bartelmann M, Lesch H. 1999. *Astron. Astrophys.* 348:351–63
Dolag K, Bartelmann M, Lesch H. 2002. *Astron. Astrophys.* 387:383–95
Dolag K, Borgani S, Schindler S, Diaferio A, Bykov AM. 2008. *Space Sci. Rev.* 134:229–68
Dolag K, Grasso D, Springel V, Tkachev I. 2005a. *J. Cosmol. Astropart. Phys.* 1:9
Dolag K, Jubelgas M, Springel V, Borgani S, Rasia E. 2004. *Ap. J. Lett.* 606:L97–100
Dolag K, Stasyszyn F. 2009. *MNRAS* 398:1678–97
Dolag K, Vazza F, Brunetti G, Tormen G. 2005b. *MNRAS* 364:753–72
Eckart C. 1960. *Phys. Fluids* 3:421–27
Evans CR, Hawley JF. 1988. *Ap. J.* 332:659–77
Faber JA, Rasio FA. 2000. *Phys. Rev. D* 62:064012
Forgan D, Rice K, Stamatellos D, Whitworth A. 2009. *MNRAS* 394:882–91
Frenk CS, White SDM, Bode P, Bond JR, Bryan GL, et al. 1999. *Ap. J.* 525:554–82
Fryer CL, Rockefeller G, Warren MS. 2006. *Ap. J.* 643:292–305
Fulk D. 1996. *J. Comput. Phys.* 126:165–80
Gingold RA, Monaghan JJ. 1977. *MNRAS* 181:375–89
Gingold RA, Monaghan JJ. 1982. *J. Comput. Phys.* 46:429–53
Gnedin NY, Abel T. 2001. *New Astron.* 6:437–55
Governato F, Willman B, Mayer L, Brooks A, Stinson G, et al. 2007. *MNRAS* 374:1479–94
Gresho PM, Chan ST. 1990. *Int. J. Numer. Methods Fluids* 11:621–59
Hairer E, Lubich C, Wanner G. 2002. *Geometric Numerical Integration. Springer Ser. Comput. Math.*, Vol. 31. Berlin: Springer-Verlag
Hernquist L. 1989. *Nature* 340:687–91
Hernquist L. 1993. *Ap. J.* 404:717–22
Hernquist L, Katz N. 1989. *Ap. J. Suppl.* 70:419–46
Hernquist L, Katz N, Weinberg DH, Miralda-Escudé J. 1996. *Ap. J. Lett.* 457:L51
Hess S, Springel V. 2010. *MNRAS* In press (DOI: 10.111/j.1365-2966.2010.16892.x)
Hockney RW, Eastwood JW. 1981. *Computer Simulation Using Particles.* New York: McGraw-Hill
Hopkins PF, Hernquist L, Cox TJ, Di Matteo T, Robertson B, Springel V. 2006. *Ap. J. Suppl.* 163:1–49
Hopkins PF, Hernquist L, Martini P, Cox TJ, Robertson B, et al. 2005. *Ap. J. Lett.* 625:L71–74
Imaeda Y, Inutsuka S. 2002. *Ap. J.* 569:501–18
Inutsuka S. 2002. *J. Comput. Phys.* 179:238–67
Jubelgas M, Springel V, Dolag K. 2004. *MNRAS* 351:423–35
Jubelgas M, Springel V, Enßlin T, Pfrommer C. 2008. *Astron. Astrophys.* 481:33–63
Katz N, Weinberg DH, Hernquist L. 1996. *Ap. J. Suppl.* 105:19
Kitsionas S, Whitworth AP. 2002. *MNRAS* 330:129–36
Kitsionas S, Whitworth AP. 2007. *MNRAS* 378:507–24
Klessen RS, Burkert A, Bate MR. 1998. *Ap. J. Lett.* 501:L205
Klessen RS, Heitsch F, Mac Low M. 2000. *Ap. J.* 535:887–906
Kotarba H, Lesch H, Dolag K, Naab T, Johansson PH, Stasyszyn FA. 2009. *MNRAS* 397:733–47
Laguna P, Miller WA, Zurek WH. 1993. *Ap. J.* 404:678–85
Landau LD, Lifshitz EM. 1959. *Fluid Mechanics*. Oxford: Pergamon
Lanzafame G, Molteni D, Chakrabarti SK. 1998. *MNRAS* 299:799–804
Liska R, Wendroff B. 2003. *SIAM J. Sci. Comput.* 25:995–1017
Lombardi JC, Sills A, Rasio FA, Shapiro SL. 1999. *J. Comput. Phys.* 152:687–735
Lucy LB. 1977. *Astron. J.* 82:1013–24
Lufkin G, Quinn T, Wadsley J, Stadel J, Governato F. 2004. *MNRAS* 347:421–29
Marri S, White SDM. 2003. *MNRAS* 345:561–74
Mayer L, Quinn T, Wadsley J, Stadel J. 2002. *Science* 298:1756–59

Mayer L, Quinn T, Wadsley J, Stadel J. 2004. *Ap. J.* 609:1045–64
McCarthy IG, Schaye J, Ponman TJ, Bower RG, Booth CM, et al. 2010. *MNRAS* In press (DOI: 10.111/j.1365-2966.2010.16750.x)
Mihos JC, Hernquist L. 1994. *Ap. J. Lett.* 431:L9–12
Mihos JC, Hernquist L. 1996. *Ap. J.* 464:641
Mitchell NL, McCarthy IG, Bower RG, Theuns T, Crain RA. 2009. *MNRAS* 395:180–96
Monaghan JJ. 1985. *Comput. Phys. Rep.* 3:71–124
Monaghan JJ. 1989. *J. Comput. Phys.* 82:1–15
Monaghan JJ. 1992. *Annu. Rev. Astron. Astrophys.* 30:543–74
Monaghan JJ. 1997. *J. Comput. Phys.* 136:298–307
Monaghan JJ. 2000. *J. Comput. Phys.* 159:290–311
Monaghan JJ. 2002. *MNRAS* 335:843–52
Monaghan JJ. 2005. *Rep. Prog. Phys.* 68:1703–59
Monaghan JJ, Gingold RA. 1983. *J. Comput. Phys.* 52:374
Monaghan JJ, Price DJ. 2001. *MNRAS* 328:381–92
Morris J. 1997. *J. Comput. Phys.* 136:41–50
Müller E, Steinmetz M. 1995. *Comput. Phys. Commun.* 89:45–58
Nayakshin S, Cha S, Hobbs A. 2009. *MNRAS* 397:1314–25
Nelson RP, Papaloizou JCB. 1994. *MNRAS* 270:1
Oechslin R, Janka H, Marek A. 2007. *Astron. Astrophys.* 467:395–409
Oger G, Doring M, Alessandrini B, Ferrant P. 2007. *J. Comput. Phys.* 225:1472–92
Oppenheimer BD, Dave R. 2006. *MNRAS* 373:1265–92
O' Shea BW, Nagamine K, Springel V, Hernquist L, Norman ML. 2005. *Ap. J. Suppl.* 160:1–27
Owen JM. 2004. *J. Comput. Phys.* 201:601–29
Owen JM, Villumsen JV, Shapiro PR, Martel H. 1998. *Ap. J. Suppl.* 116:155
Pakmor R, Kromer M, Roepke FK, Sim SA, Ruiter AJ, Hillebrandt W. 2010. *Nature* 463:61–64
Pawlik AH, Schaye J. 2008. *MNRAS* 389:651–77
Pearce FR, Couchman HMP. 1997. *New Astron.* 2:411–27
Pearce FR, Jenkins A, Frenk CS, Colberg JM, White SDM, et al. 1999. *Ap. J. Lett.* 521:L99–102
Petkova M, Springel V. 2009. *MNRAS* 396:1383–403
Pfrommer C, Enßlin TA, Springel V, Jubelgas M, Dolag K. 2007. *MNRAS* 378:385–408
Pfrommer C, Springel V, Enßlin TA, Jubelgas M. 2006. *MNRAS* 367:113–31
Phillips GJ, Monaghan JJ. 1985. *MNRAS* 216:883–95
Price DJ. 2008. *J. Comput. Phys.* 227:10040–57
Price DJ. 2010. *MNRAS* 364:384–406
Price DJ, Bate MR. 2007. *MNRAS* 377:77–90
Price DJ, Monaghan JJ. 2004a. *MNRAS* 348:123–38
Price DJ, Monaghan JJ. 2004b. *MNRAS* 348:139–52
Price DJ, Monaghan JJ. 2005. *MNRAS* 364:384–406
Price DJ, Monaghan JJ. 2007. *MNRAS* 374:1347–58
Puchwein E, Sijacki D, Springel V. 2008. *Ap. J. Lett.* 687:L53–56
Rasio FA. 2000. *Progr. Theor. Phys. Suppl.* 138:609–21
Rasio FA, Shapiro SL. 1991. *Ap. J.* 377:559
Read JI, Hayfield T, Agertz O. 2010. *MNRAS* In press (DOI: 10.111/j.1365-2966.2010.16577.x)
Ritchie BW, Thomas PA. 2001. *MNRAS* 323:743–56
Robertson BE, Kravtsov AV, Gnedin NY, Abel T, Rudd DH. 2010. *MNRAS* 401:2463–2476
Rosswog S. 2005. *Ap. J.* 634:1202–13
Rosswog S. 2009. *New Astron. Rev.* 53:78–104
Rosswog S, Liebendörfer M, Thielemann FK, Davies MB, Benz W, Piran T. 1999. *Astron. Astrophys.* 341:499–526
Rosswog S, Ramirez-Ruiz E, Melvyn B. 2003. *MNRAS* 345:1077–90
Rosswog S, Price D. 2007. *MNRAS* 379:915–31
Saitoh TR, Makino J. 2009. *Ap. J. Lett.* 697:L99–102

Scannapieco C, Tissera PB, White SDM, Springel V. 2008. *MNRAS* 389:1137–49
Schoenberg I. 1969. *J. Approx. Theory* 2:167–206
Shapiro PR, Martel H, Villumsen JV, Owen JM. 1996. *Ap. J. Suppl.* 103:269
Siegler S, Riffert H. 2000. *Ap. J.* 531:1053–66
Sijacki D, Pfrommer C, Springel V, Enßlin TA. 2008. *MNRAS* 387:1403–15
Sijacki D, Springel V. 2006. *MNRAS* 371:1025–46
Simpson JC. 1995. *Ap. J.* 448:822
Smith RJ, Clark PC, Bonnell IA. 2009. *MNRAS* 396:830–41
Springel V. 2005. *MNRAS* 364:1105–34
Springel V. 2010. *MNRAS* 401:791–851
Springel V, Di Matteo T, Hernquist L. 2005. *MNRAS* 361:776–94
Springel V, Hernquist L. 2002. *MNRAS* 333:649–64
Springel V, Hernquist L. 2003. *MNRAS* 339:289–311
Springel V, Yoshida N, White SDM. 2001. *New Astron.* 6:79–117
Steinmetz M. 1996. *MNRAS* 278:1005–17
Steinmetz M, Mueller E. 1993. *Astron. Astrophys.* 268:391–410
Stone JM, Gardiner TA, Teuben P, Hawley JF, Simon JB. 2008. *Ap. J. Suppl.* 178:137–77
Swegle J. 1995. *J. Comput. Phys.* 116:123–34
Tasker EJ, Brunino R, Mitchell NL, Michielsen D, Hopton S, et al. 2008. *MNRAS* 390:1267–81
Thacker RJ, Tittley ER, Pearce FR, Couchman HMP, Thomas PA. 2000. *MNRAS* 319:619–48
Toomre A, Toomre J. 1972. *Ap. J.* 178:623–66
Toro E. 1997. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Berlin: Springer-Verlag
van Leer B. 1984. *SIAM J. Sci. Stat. Comput.* 5:1–20
Viau S, Bastien P, Cha S. 2006. *Ap. J.* 639:559–70
Wadsley JW, Stadel J, Quinn T. 2004. *New Astron.* 9:137–58
Wadsley JW, Veeravalli G, Couchman HMP. 2008. *MNRAS* 387:427–38
Wetzstein M, Nelson AF, Naab T, Burkert A. 2009. *Ap. J. Suppl.* 184:298–325
Whitehouse SC, Bate MR. 2004. *MNRAS* 353:1078–94
Whitehouse SC, Bate MR. 2006. *MNRAS* 367:32–38
Whitehouse SC, Bate MR, Monaghan JJ. 2005. *MNRAS* 364:1367–77
Whitworth AP. 1998. *MNRAS* 296:442–44
Williams PR, Churches DK, Nelson AH. 2004. *Ap. J.* 607:1–19
Wood MA, Robertson JR, Simpson JC, Kawaler SD, O'Brien MS, et al. 2005. *Ap. J.* 634:570–84
Yoshida N, Omukai K, Hernquist L, Abel T. 2006. *Ap. J.* 652:6–25
Yukawa H, Boffin HMJ, Matsuda T. 1997. *MNRAS* 292:321

Annual Review of Astronomy and Astrophysics

Volume 48, 2010

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Astronomy and Astrophysics* articles may be found at http://astro.annualreviews.org/errata.shtml