Regular Article

Smoothed Particle Hydrodynamics for astrophysical flows

The dynamics of protostellar discs

G. Lodato[1,a] and P.J. Cossins[2]

[1] Dipartimento di Fisica, Università degli Studi di Milano - Via Celoria 16, Milano, Italy
[2] Department of Physics and Astronomy, University of Leicester - Leicester, LE1 7RH, UK

Abstract. In this paper we describe the application of the Smoothed Particle Hydrodynamics (SPH) method to the dynamics of protostellar discs, and in particular to the problem of characterising gravitational instabilities in such discs. Initially, we briefly describe the basic features of SPH, a Lagrangian mesh-free method of solving the equations of fluid dynamics which is widely used in astrophysics. We highlight those features of SPH that should be particularly kept in mind when simulating accretion discs —these include both physical and artificial viscosity— that directly affect the transport of angular momentum in the code. We then describe some recent advances in our understanding of the development and non-linear behaviour of gravitational instabilities in accretion discs, paying particular attention to the problem of numerical convergence within SPH simulations of self-gravitating discs.

1 Introduction

Smoothed Particle Hydrodynamics (SPH) [1–4] is a Lagrangian method of solving the equations of fluid dynamics which has been very widely used by the astrophysical community to simulate a large variety of different physical problems. These include large scale cosmological simulations [5–7], the dynamics of black holes in merging galaxies [8], the merger of compact objects such as neutron stars and white dwarfs [9–11], the dynamics of gas following the merger of supermassive black holes [12], the tidal disruption of stars by a massive black hole [13,14], galactic dynamics involving gas [15,16], star formation in molecular clouds [17,18] and the dynamics of protostellar discs and planet formation [19–22]. Additionally, it is also often used in other fields, such as engineering and geophysics [23].

The reason behind this large number of applications lies in several factors. On the practical side, it is a relatively easy-to-use method and some state-of-the-art codes are publicly available [5]. From a computational point of view, its mesh-free approach makes it ideally suited to deal with problems with no fixed boundaries (which are typical in astrophysics), and where a high spatial adaptivity is required. Until recently, when adaptive mesh refinement techniques [24] were developed for grid-based hydrodynamical codes, SPH offered the highest spatial resolution for treating astrophysical processes, which also require to cover a very large dynamical range. The absence of a grid, intrinsic to SPH, also implies that the numerical problem in SPH is free from any unphysical symmetries imposed by the grid, which are otherwise inherent in any grid-based method. The consequences of this property of SPH are extremely important. Indeed, as we will briefly discuss in this paper, it is possible to derive the whole discrete set of SPH equations from a variational principle, and as such an SPH simulation behaves as a Hamiltonian system. This implies that for every symmetry of the system there is a conserved property. In fact, a well-implemented SPH code will simultaneously conserve mass, momentum, angular momentum and energy exactly (to machine precision). We stress that these exceptional conservation properties are not shared by any other computational fluid dynamics method. Therefore, SPH is optimal in dealing with problems where global conservation laws are essential: for example, in a star formation calculation, if the code does not properly conserve the total angular momentum of the initial molecular cloud, the system might evolve to a compact, almost spherical configuration, rather than into a highly flattened, possibly fragmenting one. Due to its Lagrangian nature, SPH is also ideally suited to treat advection problems, which are also often encountered in astrophysics (as well as in geophysics and meteorology, for example).

Naturally, SPH also has its drawbacks. Firstly, it is a relatively young method, being developed in the late seventies. As such, it lacks the large body of optimization work that has characterized other fluid dynamics grid-based codes.

---

[a] e-mail: `giuseppe.lodato@unimi.it`

In particular, the treatment of shocks —that we will also describe in the paper— is done using an artificial viscosity term, while modern grid-based codes generally use higher-order shock-capturing algorithms, which have not yet been developed for SPH. This is probably the major flaw of SPH and might lead to artificial transport phenomena, which should be kept under control. Additionally, for the very same reason it is very good at handling advection problems, it is also not as good at handling mixing phenomena. Furthermore the treatment of physical boundaries, although certainly possible, is not a natural feature of SPH. Finally, it should also be mentioned that due to its relative ease of use, many apparent faults of SPH are not intrinsic to the method itself, but to specific implementations of it. In particular, many early implementation of SPH were not derived from a variational principle and thus gave up the most spectacular feature of SPH, that is its conservation properties.

Over the years, SPH has been supplemented by a large number of additional features, such as self-gravity [25], individual time-steps [26], that provide a significant speed up to the code, sink particles [26] to model accretion processes, physical viscosity [27–29], radiative transport [30], chemical enrichment [31], and nuclear reaction networks [32]. SPH can also be expanded to magneto-hydrodynamics [33–35] although with some unsolved issues [36], and can also be adapted to include both special and general relativistic effects [37–39].

In this paper, we will first provide a brief overview of the SPH method in sect. 2, referring the reader to other more specific reviews [40, 23, 38, 41, 42] for further details. We will focus our discussion on the specific problem of simulating protostellar disc dynamics with SPH. We will describe how one can use SPH to model these kind of systems, and what modifications to the basic implementations are needed to do that (sect. 2.5). We will then describe in sect. 3.1 some results on the transport properties of protostellar discs, and in particular on the locality of transport induced by gravitational instabilities in the disc. We will discuss (sect. 3.2) what are the resolution requirements for simulating self-gravitating accretion discs, a topic which has been recently the focus of some discussion. Finally, in sect. 4 we draw our conclusions.

2 Fundamentals of Smoothed Particle Hydrodynamics

2.1 SPH representation of a continuum field and its derivatives

The basic premise underlying the smoothed particle hydrodynamics method is that the continuous fields describing macroscopic fluid flow may be mapped smoothly onto a discrete set of fluid elements, or "particles". In the continuum, it is possible to isolate the value of a flow field f at any point $\mathbf{r}$ by taking the convolution of the field with a Dirac δ-function centred on $\mathbf{r}$, via the identity

$$f(\mathbf{r}) = \int_V f(\mathbf{r}') \, \delta(\mathbf{r} - \mathbf{r}') \, \mathrm{d}\mathbf{r}', \tag{1}$$

where $\mathbf{r}'$ is a dummy variable that ranges over the entire volume V. In order to discretise the fluid it is necessary first to broaden the δ-function into an appropriate smoothing kernel W, with characteristic width h (known as the smoothing length), subject to the following conditions;

$$\int_V W(\mathbf{r}, h) \, \mathrm{d}\mathbf{r}' = 1 \qquad \text{and} \qquad \lim_{h \to 0} W(\mathbf{r}, h) = \delta(\mathbf{r}), \tag{2}$$

namely that the volume integral over the smoothing kernel must be equal to that of the δ-function, and that the two functions become equal in the limit of small smoothing lengths. While in principle one can use non-spherical kernels [43], in what follows we will assume that $W(\mathbf{r}, h) \equiv W(|\mathbf{r}|, h)$. The next step is to discretize the integral in eq. (1) into a summation over a set of "particles" of mass m_j, respectively, by performing the substitution

$$\int_V \mathrm{d}\mathbf{r} \to \sum_j \frac{m_j}{\rho_j} \, . \tag{3}$$

Here $\rho_j = \rho(\mathbf{r}_j)$ is the fluid density at $\mathbf{r}_j$, m_j is the mass of the j-th particle and j ranges over all particles.

In this way, an estimate $f^*(\mathbf{r}_i)$ for the underlying continuum value of f at the position of the i-th particle may be obtained using

$$f^*(\mathbf{r}_i) = \sum_j \frac{m_j}{\rho_j} f(\mathbf{r}_j) \, W_{ij}, \tag{4}$$

where we have used the compact notation $W_{ij} = W(|\mathbf{r}_{ij}|, h)$, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. Equation (4) therefore represents the basis of the smoothed particle hydrodynamics method —smoothing the underlying continuum properties over a finite

volume discretised into an arbitrary number of particles. For clarity, it should be highlighted that these SPH particles are *not* representative of individual fluid molecules, but are simply a convenient way of discretising the fluid.

Let us now obtain an estimate ∇f^* for the gradient of the field $f(\mathbf{r})$. By considering eq. (4) we immediately see that the only quantity that depends on $\mathbf{r}$ on the right-hand side is the kernel function, and we therefore have

$$\nabla f^*(\mathbf{r}_i) = \sum_j \frac{m_j}{\rho_j} f(\mathbf{r}_j) \, \nabla_i W_{ij}. \tag{5}$$

Here ∇_i indicates the gradient with respect to the position of particle i[1]. Computationally, this is a very useful property indeed, as at no point does the underlying continuum value of ∇f need to be known *a priori*, whereas the gradient of the smoothing kernel will always be explicitly determined for any sensible choice of W. However, it should be noted at this point that eq. (5) is not a unique estimator for the spatial gradient, and moreover it is not necessarily the best one. For example, it does not guarantee that the gradient vanishes for a constant function. A better estimate for the gradient can be derived as follows. The discrete form of the normalization condition for the kernel (eq. (2)) is:

$$\sum_j \frac{m_j}{\rho_j} \, W_{ij} \approx 1, \qquad \text{and thus} \qquad \sum_j \frac{m_j}{\rho_j} \, \nabla_i W_{ij} \approx 0. \tag{6}$$

Thus, eq. (5) provides a good representation of constant functions only to the extent to which eqs. (6) are satisfied. Using eqs. (6) and identities such as $\nabla f(\mathbf{r}) = \nabla(f(\mathbf{r}).1) = 1.\nabla f(\mathbf{r}) + f(\mathbf{r})\nabla 1$ (see, for instance, [42, 40]) it is possible to construct a variety of estimators, some of which may be more useful than others. For example, an improved version of the gradient estimator which vanishes identically for constant functions is given by the following;

$$\nabla f^*(\mathbf{r}_i) = -\sum_j \frac{m_j}{\rho_j} f_{ij} \, \nabla_i W_{ij}, \tag{7}$$

where we have used the compact notation $f_{ij} = f(\mathbf{r}_i) - f(\mathbf{r}_j)$. In a similar manner it is possible to define estimators for a variety of vector calculus operators —for a more detailed discussion and derivations of these, see, for instance, [42, 38, 40]. Likewise, estimators for second derivatives can also be obtained, although the derivation is more complex [28, 23, 41]. The Laplacian of a scalar quantity can be estimated by:

$$\nabla^2 f^*(\mathbf{r}_i) = 2 \sum_j \frac{m_j}{\rho_j} f_{ij} \frac{F_{ij}}{|\mathbf{r}_{ij}|} \,, \tag{8}$$

where $\nabla_i W_{ij} = \hat{\mathbf{r}}_{ij} F_{ij}$ and $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|$, and similarly for vector quantities in three dimensions we have:

$$\nabla^2 \mathbf{f}^*(\mathbf{r}_i) = 2 \sum_j \frac{m_j}{\rho_j} \mathbf{f}_{ij} \frac{F_{ij}}{|\mathbf{r}_{ij}|} \,, \tag{9}$$

and

$$\nabla\big(\nabla \cdot \mathbf{f}^*(\mathbf{r}_i)\big) = \sum_j \frac{m_j}{\rho_j} \big[5(\mathbf{f}_{ij} \cdot \hat{\mathbf{r}}_{ij})\hat{\mathbf{r}}_{ij} - \mathbf{f}_{ij}\big] \frac{F_{ij}}{|\mathbf{r}_{ij}|} \,, \tag{10}$$

where again $\mathbf{f}_{ij} = \mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)$. Note that henceforth, for brevity, we shall drop the superscript * from estimated quantities.

At this stage it is worth considering the accuracy of the scheme just discussed. Inherent in the above is the notion that smoothed particle hydrodynamics provides only an *estimator* of the underlying flow properties, and clearly we would like this estimator to be as accurate as possible. A key source of error is the interpolation, the generalisation from a δ-function to a smoothing kernel of finite width. Careful analysis shows that for symmetric smoothing kernels this process is accurate to second order in the smoothing length h [38, 40, 44, 45], although it is possible to construct kernels that are accurate to arbitrary order by considering the relevant kernel moments [38, 40]. Generally speaking, however, higher-order kernels lead to problems elsewhere, such as potentially introducing negative density estimates in certain pathological situations [38, 40], and as such second order kernels are generally used. Notwithstanding this, higher-order kernels have been shown to be useful in resolving Kelvin-Helmholtz instabilities using SPH [46]. A commonly used smoothing kernel is the so-called cubic spline kernel, [42, 46, 38, 40, 47] which along with its spatial derivative is given by

$$W(\mathbf{r}, h) = \frac{W_0}{\pi h^\nu} \begin{Bmatrix} 1 - \dfrac{3}{2}x^2 + \dfrac{3}{4}x^3 \\ \dfrac{1}{4}(2-x)^3 \\ 0 \end{Bmatrix}, \quad \nabla W(\mathbf{r}, h) = \frac{W_0}{\pi h^{\nu+1}} \begin{Bmatrix} \dfrac{9}{4}x^2 - 3x \\ -\dfrac{3}{4}(2-x)^2 \\ 0 \end{Bmatrix}, \quad \text{where} \quad \begin{Bmatrix} 0 \le x < 1, \\ 1 \le x < 2, \\ x \ge 2, \end{Bmatrix}, \tag{11}$$

---

[1] Note that $W_{ij} = W_{ji}$ and that $\nabla_i W_{ij} = -\nabla_j W_{ij}$.

where $x = |\mathbf{r}|/h$ and $W_0 = (2\pi/3, 10/7, 1)$ in $\nu = (1, 2, 3)$ dimensions. An advantage of this kernel over others (such as the Gaussian kernel) is that it has *compact support*, in that it is zero for $|\mathbf{r}| > 2h$. This has the advantage that only particles within $2h$ of the kernel centre contribute to the summation for the estimator of any quantity, significantly reducing the computational effort required. Being second order in h, the error associated with using this kernel is clearly reduced by decreasing the smoothing length, *i.e.* by increasing the spatial resolution.

The second unavoidable source of error within the SPH framework is the discretisation error, *i.e.* that arising from the transition from an integral over a continuous volume to a finite sum over a series of discrete particles. In the case of a uniform distribution of particles within the smoothing kernel this error will reduce to a minimum, and therefore it essentially represents a sampling error within the volume of the kernel. For a non-uniform particle distribution (as will be the case in every practical SPH simulation) and a fixed number of particles this source of error is reduced by *increasing* h, so as to encompass more particles within the smoothing kernel and thus improve the numerical resolution.

These two sources of error within the SPH framework therefore work in opposition to each other —to improve both the spatial and the numerical resolution one not only has to decrease the kernel size but also to increase the number of particles within each kernel, leading to a much increased number of particles overall. In practice however, it has been found that an acceptable compromise is when the number of particles within the smoothing kernel (usually called the "neighbours") is in the range of 4–6, 12–20 and 40–60 in 1, 2 and 3 dimensions, respectively [48,46,40]. This generally provides satisfactory numerical sampling, with the total number of particles then adjusted to provide the required spatial resolution.

Finally, it is often implicitly assumed that implementations of SPH using a second-order kernel are themselves second order accurate. This may be somewhat disingenuous, as the kernel is only second order accurate when the sampling error is zero —otherwise errors of order h can arise, thereby lowering the accuracy of the scheme overall. Also this only applies to the *spatial* accuracy —many implementations of SPH that use a "second-order" Runge-Kutta-Fehlburg integrator to evolve the simulation forward in time are, strictly speaking, only first-order accurate. Nonetheless, in flow regions where the fluid properties are slowly varying these effects are small, and thus second-order accuracy is effectively achieved.

2.2 The equations of motion

Having established a mathematical basis for estimating fluid properties, it is now possible to consider how the equations of motion can be implemented using this scheme. For brevity, we shall firstly only examine the case where the smoothing length h is fixed and constant for all particles, that is where the spatial resolution is invariant.

SPH is a Lagrangian method, and thus follows the evolution of a given set of particles representing the fluid elements. Therefore, it automatically conserves total mass (unless particles are artificially removed from the simulation) and for equal mass particles the particle number density also satisfies the continuity equation. Actually, we do not need to solve continuity equation in order to get the fluid density, but starting from the particles' positions we can simply use eq. (4) to have an estimate of the density $\rho_i = \rho(\mathbf{r}_i)$ at the position of particle i,

$$\rho_i = \sum_j m_j W_{ij}. \tag{12}$$

In those cases where the self-gravity of the particles themselves is negligible, for the implementation discussed hereafter this is the *only* point at which the SPH estimator is required —all other fluid quantities are either evolved from initial conditions or can be determined from them and the density estimator.

It is interesting to show that this estimate of the fluid density does also satisfy the continuity equation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\rho\nabla \cdot \mathbf{v}, \tag{13}$$

where $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \mathbf{v} \cdot \nabla$ is the Lagrangian derivative with respect to time t. By taking the derivative of eq. (12) with respect to time t we obtain

$$\frac{\mathrm{d}\rho_i}{\mathrm{d}t} = \sum_j m_j \left(\mathbf{v}_i \cdot \nabla_i W_{ij} + \mathbf{v}_j \cdot \nabla_j W_{ij}\right) = \sum_j m_j \, \mathbf{v}_{ij} \cdot \nabla_i W_{ij}, \tag{14}$$

where, as before, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ and $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$ is the velocity of each particle. Here we note that the minus sign is introduced by the antisymmetry of the kernel gradient, such that $\nabla_j W_{ij} = -\nabla_i W_{ij}$. The right-hand side of the second equality in eq. (14) is simply an estimator for $-\rho_j\nabla_j \cdot \mathbf{v}_j$ [42,40], and thus the continuity equation is satisfied exactly, ensuring as expected the explicit conservation of mass.

Although there are various ways of deriving the momentum equations consistently with the SPH framework, a particularly appealing one is to use the Lagrangian formalism. As long as the discrete Lagrangian functional preserves the fundamental symmetries of the underlying continuous one, this confers the inherent advantages that the resulting SPH equations will automatically fulfill the requisite conservation law (through Noether's Theorem) and also that the only approximations made are in the discretisation of the Lagrangian itself.

We can therefore consider the discrete Lagrangian functional $\mathcal{L}$, which in the case of compressible flow under the influence of gravity is given by [49]

$$\mathcal{L} = \sum_j m_j \left(\frac{1}{2} \mathbf{v}_j \cdot \mathbf{v}_j - u_j(\mathbf{r}_j) - \frac{1}{2} \Phi_j(\mathbf{r}_j) \right), \tag{15}$$

where j ranges over *all* particles, u_j is the internal energy of the jth particle and Φ_j is its gravitational potential. Note that the factor of one half in the latter term arises due to the fact that the gravitational potential is determined pairwise —see [42] for a detailed derivation and explanation. From eq. (15) we can then obtain the equations of motion for particle i using the Euler-Lagrange equations,

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{v}_i} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = \mathbf{0}. \tag{16}$$

Clearly the only term to depend on the velocity is the kinetic energy term, with the result that

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{v}_i} \right) = \frac{\mathrm{d}}{\mathrm{d}t} (m_i \mathbf{v}_i) = m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}, \tag{17}$$

for constant mass particles and noting that since the particle velocities are independent, the derivative is zero unless $i = j$. To expand the second term in eq. (16) we require two further constraints, provided by the First Law of Thermodynamics and a definition of the gravitational potential. In the first instance, recall that for specific quantities the First Law can be written as $\mathrm{d}u = T\mathrm{d}s - P\mathrm{d}(1/\rho)$, where T is the temperature, $\mathrm{d}s$ is the change in specific entropy and P is the fluid pressure. In the isentropic case we may set $\mathrm{d}s = 0$, and thus $\mathrm{d}u = (P/\rho^2)\mathrm{d}\rho$. In order to obtain the gravitational potential, we have to keep in mind that we are interested in approximating the potential of a continuous field using a discrete set of particles. Thus, we should not simply consider the potential of the various particles, as if they were simple point masses. Within the SPH formalism we may obtain an estimate for the underlying gravitational potential Φ of the ith particle via

$$\Phi_i = G \sum_j m_j \, \phi_{ij}, \tag{18}$$

where G is the universal gravitational constant, $\phi_{ij} = \phi(\mathbf{r}_{ij}, \varepsilon)$ is known as the (gravitational) softening kernel and ε is the softening length —the characteristic width of the softening kernel analogous to the smoothing length. Generally, the softening and smoothing lengths are set to be equal, $\varepsilon = h$, although there is no formal requirement for this to be the case [50]. For a self-gravitating fluid, the softening kernel ϕ is determined from the smoothing kernel W via Poisson's equation,

$$\nabla^2 \Phi(\mathbf{r}) = 4\pi G \rho(\mathbf{r}). \tag{19}$$

For spherically symmetric smoothing kernels and using the estimators for ρ and Φ given in eqs. (12) and (18), we find

$$\frac{1}{4\pi r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi(r,h)}{\partial r} \right) = W(r,h), \tag{20}$$

and thus

$$\phi = 4\pi \left[-\frac{1}{r} \int^r r'^2 W(r') \mathrm{d}r' + \int^r r' W(r') \mathrm{d}r' \right] + \frac{C_1}{r^2} + \frac{C_2}{r}. \tag{21}$$

Here $r = |\mathbf{r}|$, and C_1, C_2 are integration constants such that for $r \geq 2\varepsilon$ we recover the standard Newtonian inverse square law, and the correct asymptotic behaviour ($\phi \to 0$ as $r \to \infty$) is observed. For the cubic spline kernel with $\varepsilon = h$, the corresponding softening kernel and its gradient may be found in [42].

We can now write the second term of eq. (16) as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = -\sum_j m_j \frac{P_j}{\rho_j^2} \frac{\partial \rho_j}{\partial \mathbf{r}_i} - \frac{G}{2} \sum_j m_j \frac{\partial \Phi_j}{\partial \mathbf{r}_i}. \tag{22}$$

The two spatial derivatives of ρ_j and Φ_j with respect to particle i are not straightforward, but after some algebra [42, 38,51] they can be shown to be as follows:

$$\frac{\partial \rho_j}{\partial \mathbf{r}_i} = \sum_k m_k \left[\delta_{ji} - \delta_{ki}\right] \nabla_i W_{jk}, \qquad (23)$$

and similarly

$$\frac{\partial \Phi_j}{\partial \mathbf{r}_i} = -\frac{G}{2} \sum_k m_k \left[\delta_{ij} - \delta_{ki}\right] \nabla_i \phi_{jk}. \qquad (24)$$

Finally, substituting these spatial derivatives into eq. (22) and thence back into the original Euler-Langrange equations (eq. (16)) we obtain the SPH equations of motion

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \nabla_i W_{ij} - G \sum_j m_j \nabla_i \phi_{ij}. \qquad (25)$$

Recall that the gradient of the smoothing kernel is antisymmetric in i, j, and in a similar manner the same is true of the softening kernel. This in turn means that eq. (25) is antisymmetric in i, j, $i.e.$ pairwise particle forces are equal and opposite. It is clear therefore that linear momentum is explicitly conserved. Furthermore, by taking the temporal derivative of the angular momentum $\mathbf{L} = \mathbf{r} \times m\mathbf{v}$ it can likewise be shown that angular momentum is also explicitly conserved [42,38,40].

Finally, we come to the energy equation. By considering the time derivative of the total energy E of the system, itself given by

$$E = \sum_i m_i \left(\frac{1}{2}\mathbf{v}_i \cdot \mathbf{v}_i + u_i + \frac{1}{2}\Phi_i\right), \qquad (26)$$

it is possible to show that the total energy is conserved exactly, in a precisely similar manner to that used for angular momentum conservation [42,38]. More usefully however, by considering the form of the First Law of Thermodynamics discussed above, the time derivative of the specific internal energy can be given by

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\rho_i^2} \frac{\mathrm{d}\rho_i}{\mathrm{d}t} = \frac{P_i}{\rho_i^2} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}, \qquad (27)$$

using eq. (14). This is the generally used method for advancing the internal energy of the particles (although not the only one, see, for instance, [38,40,52,53]), and along with eqs. (12) and (25) forms the basis of the Lagrangian smoothed particle hydrodynamics implementation. We have seen therefore that by using the Lagrangian formulation of the SPH algorithms, coupled with the fact that SPH particles are of constant mass, it is possible to derive a flow solver that is explicitly conservative of mass, linear and angular momentum and energy. There is therefore no additional source of error arising from the implementation itself.

However, the equation set is not yet complete, in that the particle pressure P has yet to be determined from the density and internal energy —an equation of state must be defined. It is noteworthy that up to now this has not been required, thereby allowing a considerable degree of latitude in the possible equations of state that can be implemented. For the simulations discussed later in this paper, and indeed for many purposes, the ideal gas equation of state is sufficient, such that the pressure for particle i is determined via $P_i = (\gamma - 1)\rho_i u_i$, where γ is the adiabatic index, or ratio of specific heats. Nonetheless, much effort has gone into implementing more accurate equations of state, to include opacity effects [54], radiative transfer [55–57] and others.

2.3 Adaptive resolution

A very generally applicable and powerful extension to the SPH algorithms developed above is to include adaptive resolution. In this case, we allow each particle i to have an individually varying smoothing length h_i. We can thus perform again the derivation of the Euler-Lagrange equations in this case, where the smoothing kernel $W(\mathbf{r}_{ij}, h_i)$ now has a non-zero derivative with respect to h_i. Note that, also in this case, the resulting equations of motion will naturally be characterized by the very same conservation properties as before, as long as the appropriate additional terms resulting from such dependencies are adequately taken into account.

The complete set of equations for adaptive spatial resolution, incorporating these correction terms, is given in the appendix. Here, we briefly introduce the relevant correction factors introduced by variable smoothing (and softening)

lengths. Equation (14) becomes [42, 38, 40][2]

$$\frac{d\rho_i}{dt} = \frac{1}{\hat{\Omega}_i} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}(h_i),$$ (28)

where

$$\hat{\Omega}_i = 1 - \frac{dh_i}{d\rho_i} \sum_j m_j \frac{\partial W_{ij}(h_i)}{\partial h_i}.$$ (29)

In a similar manner there is an additional correction term χ_i applied in the equations of motion to account for spatial variation of the softening kernel, given (for the case where $\varepsilon_i = h_i$) by [42,51] as

$$\chi_i = \frac{dh_i}{d\rho_i} \sum_j m_j \frac{\partial \phi_{ij}(h_i)}{\partial h_i}.$$ (30)

Note that in both eqs. (28) and (30) the derivative $dh/d\rho$ is needed, requiring the relationship between the density and the smoothing length to be known explicitly. This is generally achieved by fixing the mass within a smoothing kernel such that (in three dimensions)

$$\rho_i h_i^3 = \eta^3 m_i,$$ (31)

where η is a constant parameter that effectively sets the number of neighbours N_{neigh}, that is the number of particles within a smoothing sphere. With the choice $\eta = 1.2$ one gets $N_{\mathrm{neigh}} \approx 60$. To determine both the smoothing length and the density consistently it is therefore necessary to solve eqs. (12) and (31) concurrently, which may be achieved implicitly (and generally very rapidly) using the Newton-Raphson iteration method [42, 48, 51].

The choice described in eq. (31) has several advantages. Firstly, it automatically implies that regions of high density are characterized by a higher resolution, which is often desired. Secondly, by maintaining the mass within a smoothing kernel constant, it also implies that the sampling error discussed above is essentially uniform for all particles.

2.4 Artificial viscosity

Finally, in order to have a code that is reasonably widely applicable to realistic physical phenomena we need to relax the constraint that the flow be isentropic. Without this, key features such as shocks or viscous flows will be impossible to model. Shocks in particular present a challenge, as by their nature they inject entropy into the flow, moreover they do so over distances on the order of the mean free path of the fluid molecules —far less than a smoothing length. At the smoothing length resolution scale the flow therefore appears discontinuous in entropy, pressure, density and velocity, leading to large errors in SPH solvers designed for isentropic, smoothing varying flow conditions.

In order to model this, the general approach is therefore to *artificially* broaden the shock over a small number of smoothing lengths while ensuring that the Rankine-Hugoniot equations are satisfied on either side of this "shock" region, and to add entropy within by the use of an artificial viscosity term in the momentum equations. This is usually done by introducing an "artificial pressure" term, dependent on the magnitude of the divergence of the velocity field, in order to be effective in regions of strongly convergent (or if necessary, divergent) flow. Over a characteristic length scale h this gives a velocity difference $\mu = h \nabla \cdot \mathbf{v}$, which to first order is computed as [4]

$$\mu_{ij} = \frac{\bar{h}_{ij} \mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{|\mathbf{r}_{ij}|^2 + \epsilon \bar{h}_{ij}^2}.$$ (32)

Here the factor proportional to $\epsilon \ll 1$ is introduced to avoid a singularity for $\mathbf{r}_{ij} \to 0$ and the overbar represents pairwise particle mean values. By considering eq. (25), on dimensional grounds it can be seen that such viscous term $\Pi \nabla W$ appearing in the first (hydrodynamic) summation should be such that $\Pi \propto U^2/\rho$, where U is some characteristic velocity (for example, μ as defined above). Usually, two terms are introduced, one linear in the quantity μ and one quadratic, such that

$$\Pi_{ij} = \begin{cases} -\alpha_{\mathrm{SPH}} \dfrac{\bar{c}_{\mathrm{s},ij} \mu_{ij}}{\bar{\rho}_{ij}} + \beta_{\mathrm{SPH}} \dfrac{\mu_{ij}^2}{\bar{\rho}_{ij}} & \text{if } \mu_{ij} < 0, \\ 0 & \text{otherwise,} \end{cases}$$ (33)

and in the "standard" formulation of artificial viscosity, the coefficients α_{SPH} and β_{SPH} are set to 1 and 2, respectively.

---

[2] Note that [48] provide a mathematically equivalent and potentially computationally more efficient alternative formulation for $\hat{\Omega}_i$ for certain classes of smoothing kernel.

With this formulation viscous correction terms arise for both the momentum and energy equations, which are given as follows:

$$\frac{d\mathbf{v}_i}{dt}\bigg|_{\text{visc}} = -\sum_j m_j \Pi_{ij} \overline{\nabla_i W_{ij}}, \tag{34}$$

and

$$\frac{du_i}{dt}\bigg|_{\text{visc}} = \frac{1}{2}\sum_j m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \overline{\nabla_i W_{ij}}, \tag{35}$$

where $\overline{\nabla_i W_{ij}} = \left[\nabla_i W(\mathbf{r}_{ij}, h_i) + \nabla_i W(\mathbf{r}_{ij}, h_j)\right]/2$. Note that an alternative formulation, which in practice is only slightly different from the one discussed above, can be obtained by considering an analogy with Riemann solvers, where the two particles are considered as the "left" and "right" states of the Riemann problem [58,41]. We refer the reader to these papers for a discussion on the improvements obtained with such alternative implementations of the viscosity terms.

Ideally, the artificial viscosity should be negligible away from shocks, and indeed, it is for this reason that Π_{ij} is zero for divergent flows ($\mu_{ij} > 0$). Nonetheless, for convergent and shearing flows, the standard form of the artificial viscosity remains *non*-zero, and as such unwanted numerical dissipation is introduced into the flow. There are therefore a number of additional "switches" that can be implemented to control the artificial viscosity far from shocks. The Balsara switch [59] reduces the artificial viscosity term by a factor $f_{ij} = (\lambda_i + \lambda_j)/2$, with

$$\lambda_k = \frac{|\nabla \cdot \mathbf{v}_k|}{|\nabla \cdot \mathbf{v}_k| + |\nabla \times \mathbf{v}_k| + 0.0001 c_{s,k}/h_k}, \tag{36}$$

where the inclusion of the vorticity more accurately captures shearing and obliquely shocked flows. The Morris and Monaghan switch [60] additionally sets the coefficients α_{SPH} and β_{SPH} to be functions of time, decaying exponentially away from shocked regions. Finally, the Cullen-Dehnen switch [48] also implements a sophisticated shock-detection algorithm, to minimise as far as possible the artificial viscosity in unshocked regions.

2.5 Simulating an α-like viscosity

Although the use of artificial viscosity is required in current versions of SPH to resolve shocks and to avoid numerical inconsistencies such as particle interpenetration, it is also a key limitation of the method, because, as described above, it can introduce unwanted numerical dissipation and correspondingly artificial transport effects.

However, when one wants to simulate the evolution of viscous accretion discs, one might be interested in having some artificial viscosity terms, as long as they are able to reproduce the main features of a physical accretion disc viscosity. Clearly, this choice is not universal, and depends on the aims of the simulation. For example, in many cases one is interested in determining the fundamental physical mechanism responsible for the macroscopic disc viscosity, in the form of fluid instabilities, such as the magneto-rotational instability (MRI) or the gravitational instability described below. In such cases, one would clearly aim to keep any other dissipative terms as low as possible, since the aim is to measure the transport properties of an otherwise inviscid fluid.

On the other hand, in many other cases one might simply be interested in the evolution of an accretion disc, where the macroscopic disc viscosity is simply parameterized using the standard model of an α-like viscosity [61], in which the shear viscosity coefficient ν is given by

$$\nu = \alpha c_s H. \tag{37}$$

Here c_s is the sound speed, H is the disc thickness (equal to c_s/Ω for a Keplerian disc, where Ω is the angular velocity), and α is an unknown parameter, generally (although not necessarily) assumed to be constant throughout the disc. The magnitude of the macroscopic disc viscosity in astrophysically relevant accretion discs, as parameterized by α, can be obtained by comparing disc evolutionary models with known observed systems which involve a time-variable disc. Models of accretion discs in galactic binary systems, such as dwarf novae outbursts, indicate that $\alpha \approx 0.1$ [62]. Conversely, the viscous timescale for protostellar discs appears to be consistent with a much lower $\alpha \approx 0.01$ [63], and protostellar outbursts, such as FU Orionis objects —when modeled in a similar fashion to dwarf novae— would require an even lower $\alpha \approx 10^{-3}$ [64]. From a theoretical point of view, numerical simulations of the MRI or gravitational instabilities point to $\alpha \approx 10^{-3}$–10^{-2} (see also below). While clearly all these estimates strongly differ from case to case, one may wish to simulate an accretion disc with a given, fixed value of α. We describe here how this can be accomplished using SPH.

One simple approach would be to use the artificial viscosity term (eq. (33)). Indeed, in the case where artificial viscosity is applied for both approaching and receding particles and where ρ and c_s do not vary greatly within the smoothing kernel, it can be shown [65–67,23] using eqs. (9) and (10) that the linear term (proportional to α_{SPH}) in
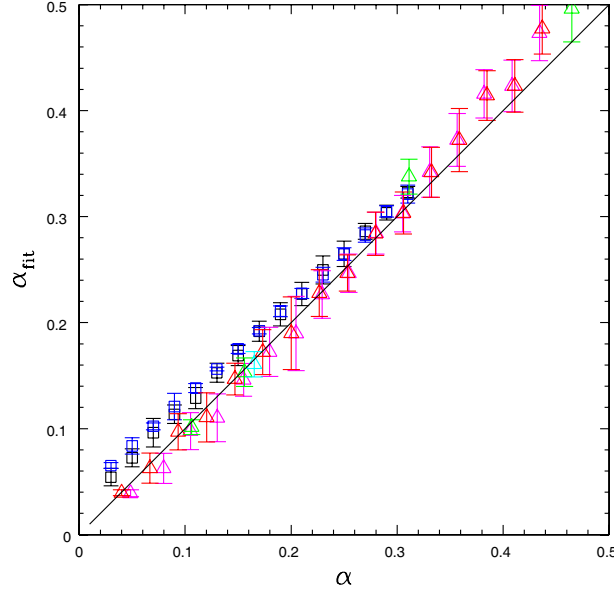
Fig. 1. Comparison between the value of α predicted from eq. (40) and the measured value obtained fitting the SPH data to the 1D evolution of Σ. Results are shown with triangles for the calculations where the disc viscosity has been simulated using the SPH artificial viscosity, whilst squares correspond to calculations where physical viscosity terms have been implemented directly (see text for details). All calculations are performed at a resolution of 2 million SPH particles, except for the green, cyan and yellow triangles that use 20 million particles. Error bars refer to the 1σ errors from the fitting procedure used to obtain α_{fit}. From [65]; see paper for details on the fitting procedure.

Π_{ij} (eq. (34)) corresponds to the viscous terms in the Navier-Stokes equations, where the shear and bulk viscosity coefficients ν^{AV} and ζ^{AV} are given by:

$$\nu^{\text{AV}} = \frac{1}{10}\alpha_{\text{SPH}}c_{\text{s}}h, \tag{38}$$

$$\zeta^{\text{AV}} = \frac{1}{6}\alpha_{\text{SPH}}c_{\text{s}}h. \tag{39}$$

We thus see that artificial shear viscosity, in this case, does mimic an α-like viscosity, where the equivalent α parameter is given by

$$\alpha^{\text{AV}} = \frac{1}{10}\alpha_{\text{SPH}}\frac{h}{H}. \tag{40}$$

By appropriately choosing α_{SPH}, and as long as the disc thickness is uniformly resolved ($h/H = \text{const}$), one is thus able to adjust the parameters in order to provide the required value of α. We show in fig. 1 an example of the calibration of the viscosity coefficient (taken from [65]) obtained using this method. The authors have performed a number of SPH simulations of thin accretion discs with different values of α_{SPH} and h/H and have then compared the evolution of the surface density Σ resulting from the simulation with the one expected for a viscous accretion disc, thus obtaining a best fit value for the α coefficient. Each data point in fig. 1 represents one simulation, where the x-axis shows the expected value of α based on eq. (40) and the y-axis shows the resulting best fit value based on the surface density evolution. The triangles refer to simulations where artificial viscosity is used to model an α-viscosity. One easily sees that, especially for low value of α, the agreement is very good.

This method thus provides an effective and easy way to model accretion discs in SPH, although it also has some drawbacks. First of all, we stress that it provides both a shear and a bulk component of viscosity. This might not be required, and in any event it should be taken into account when interpreting the results. For example, in the investigation of warp propagation in thin accretion discs [65], analytical non-linear models reproduce the simulation data only when bulk viscosity is taken into account. Secondly, the above analysis only refers to the α_{SPH} term of the artificial viscosity formulation. One might thus be tempted to remove altogether the β_{SPH} term in the artificial viscosity when simulating an accretion disc viscosity. However, this should only be done with great care, as the quadratic term in the artificial viscosity is needed in order to properly model flows with strong shocks or strongly converging flows, for which the linear term alone is not sufficient to prevent particle interpenetration. Paradoxically, removing the quadratic term (or using a too low value for β_{SPH}) might even lead to an enhanced viscous dissipation, due to the increased randomness of the particle distribution and to an artificially enhanced value of the divergence of the velocity [68].

There are also other ways to simulate a viscous flow in SPH. Another possibility is to simply find the SPH equivalent of Navier-Stokes equations directly, by appropriately evaluating the various components of the stress tensor. In this way, one has the freedom to choose the viscosity coefficients thereby removing the limitations imposed by artificial viscosity. The approach proposed by Flebbe *et al.* [27] is to use two first derivatives, that is to use the SPH technique to evaluate the first derivatives of the velocity to compute the stress tensor, and then again compute the SPH divergence of the stress tensor to include in the Navier-Stokes equations. Details on how to implement this method can be found in [65]. The black squares in fig. 1 show a calibration of the disc viscosity implemented through this method —and thus providing only a shear component to viscosity—, and where a small amount of artificial viscosity is also introduced to model shocks, using the Morris-Monaghan switch. As for the case when this is done through artificial viscosity, the agreement is very good, although the simulations appear to provide a slightly larger viscosity than expected, especially for low α. This can be attributed to the extra viscosity introduced artificially. However, the blue squares in fig. 1 show the results of simulations where artificial viscosity is removed altogether, showing an even larger viscosity, resulting once again from the spurious effects one might get when artificial viscosity is too low.

Finally, it is possible to implement directly Navier-Stokes equations using the SPH version of the relevant second order derivatives [28], expressed through eqs. (10).

2.6 Modeling discontinuities in SPH

The discussion of the two previous sections is only a particular case of a more general problem, intrinsic in SPH, as well as in other approximate methods that aim at solving the equations of hydrodynamics in differential form. Any such method is bound to take appropriate measures in cases where the flow variables become non differentiable, or even discontinuous. This obviously happens at shock fronts, where the fluid velocity is discontinuous, but not only there. Another classic example is represented by the equilibrium configuration that leads to the Kelvin-Helmholtz instability [69], in which case it is the internal energy that is discontinuous [70].

The general way to resolve discontinuous flows in SPH is to introduce artificial dissipation terms to smooth out discontinuities on a scale of the order of the smoothing length. Artificial viscosity is one such dissipation term, needed to resolve discontinuities in the velocity field. For contact discontinuities (such as the Kelvin-Helmholtz instability), an artificial conductivity term allows a proper modeling of the flow [70]. If one were to include MHD effects as well [33–35], then one would need to add an artificial resistivity term. Note that in order to evolve the density field in SPH one can either use eq. (12) (this is the standard practice) or eq. (14). While the former case ensures that the continuity equation is enforced in an integral form, the latter is the SPH representation of continuity equation in differential form. Thus, an SPH code that uses eq. (12) to compute densities can safely model density discontinuities, while if eq. (14) is used, one should also add an extra "artificial diffusivity" term.

Finally, we should mention that it is also possible to introduce Godunov-type Riemann solvers for SPH, which again should be able to properly describe shocks and contact discontinuities without the need to introduce artificial dissipation. This has been done [71,72], although unfortunately such Godunov-SPH is only rarely used in astrophysics.

3 Characterizing the gravitational instability in protostellar discs

3.1 On the locality of transport induced by the gravitational instability

The study of the process of star and planet formation has become one of the key areas of modern astrophysics. This is due in part to the spectacular series of discovery of hundreds of extra-solar planets and in part to recent advances in infrared and submillimetric astronomy (see the launch of the *Spitzer* and *Herschel* satellites, and the upcoming completion of the *ALMA* interferometer). From the theoretical point of view, we have evolved from the idealised models of the late 70s [73,74] to a more complex scenario [75] where numerical hydrodynamical and MHD simulations play a fundamental role.

During the collapse of a molecular cloud core, the angular momentum of the core prevents the gas from collapsing spherically onto the forming protostar, and leads to the formation and growth of a protostellar disc. The subsequent mass growth of the protostar occurs thanks to the redistribution of angular momentum occurring through torques in the disc, often modeled in terms of simple viscous torques. The actual physical origin of these torques is still unclear, and is most probably due to a combination of MHD [76], hydrodynamical and gravitational instabilities. Here, we will concentrate on the latter as a possible way to transport mass and angular momentum in protostellar discs.

During the envelope infall phase, as molecular gas is continuously fed onto the disc from the surrounding core, the gravitational field of the disc material itself becomes dynamically important. SPH has been widely used in this context to model the evolution of gravitationally unstable discs [77,78,19–21,79–85].

There are several dynamical effects related to the disc self-gravity. On the one hand, much like in galaxy discs [86], gravitational instabilities can arise and produce a prominent spiral structure. As we shall see below, this can happen

already for relatively light disc, where the total disc mass is only a fraction of the central object mass. On the other hand, if the disc mass becomes comparable to the central star (which might well happen in the earliest phases of the star formation process), the gravitational field of the disc can alter the overall centrifugal equilibrium of the disc, such that the angular velocity Ω will depart significantly from a Keplerian profile. Analytical models of such non-Keplerian discs have been proposed over the years [87–89], and massive discs have also been simulated numerically [20]. In the following, however, we will for simplicity consider mostly the case of relatively light discs, which are gravitationally unstable but are still characterised by a Keplerian rotation curve.

The dynamical importance of the disc self-gravity has long been known to be determined by the parameter Q, that determines the linear stability properties of the disc [90], given by

$$Q = \frac{c_{\rm s}\kappa}{\pi G \Sigma},$$ (41)

where Σ is the disc surface density and κ $(= \Omega$ for a Keplerian disc) is the epicyclic frequency. As the sound speed $c_{\rm s}$ is determined by the disc temperature, the balance between heating (both due to internal processes and to external source, as in the case of irradiation from the central star) and radiative cooling is also, indirectly, a major factor. Thus, for $Q \gtrsim 1$ (hot or light discs) the effect of self-gravity is negligible, whereas where $Q \lesssim 1$ (cold or massive discs) it dominates, and the disc becomes gravitationally unstable. It is easy to show that, for a Keplerian disc, the condition $Q \lesssim 1$ of gravitational instability can be cast in the equivalent form

$$\frac{M_{\rm disc}(R)}{M_\star} \gtrsim \frac{H}{R},$$ (42)

where $M_{\rm disc}(R) = \pi \Sigma R^2$ is of the order of the disc mass enclosed within radius R. We thus see that, in order to be unstable, it suffices that the disc mass is of order of the aspect ratio of the disc, $H/R \approx 0.1$, and thus even a light disc can be gravitationally unstable.

The non-linear evolution of the instability can lead to significantly different phenomenological outcomes, from relatively quiescent discs, where the instability saturates non linearly at a finite amplitude, to a runaway evolution of the system leading to disc fragmentation, a potential mechanism for the formation of stellar or planetary mass companions [91,92,54,93–98,81,82,99–102].

In this respect, it has long been recognised that the initiation and subsequent propagation of gravitationally induced spiral density waves within accretion discs feed back energy —through PdV work and shock heating— in the disc, and thus effectively increase the gas temperature. Thus, the combined effect of gravitationally induced heating and of radiative cooling can lead to a *self-regulation* process [103,87,104], where the disc evolves into a dynamic steady-state condition of marginal stability, characterised by a value of Q that is very close to unity, and the radiative cooling is balanced by the gravitationally induced heating. These gravitationally induced spiral density waves, which have been observed in a number of simulations [77,19–21,105–107], lead to energy and angular momentum being transported through the disc, and therefore, in turn, drive gas to be accreted on to the protostar [19–21]. Since it is thought that protostellar discs undergo a self-gravitating phase of a few 10^5 years the gravitational instability may therefore play a key role in the birth of stars.

Of particular interest is whether this energy and angular momentum transport occurs as a local process (and could therefore be parameterised as a pseudo-viscous process via the Shakura-Sunyaev α formalism), or whether long-range gravitational effects enable it to have global reach across the disc. In the former case, it is possible to show [108] that thermal equilibrium in the disc requires α to satisfy

$$\alpha = \frac{4}{9\gamma(\gamma - 1)\Omega t_{\rm cool}},$$ (43)

where $t_{\rm cool}$ is the cooling time scale of the disc. Furthermore, for a local, viscous process the torque exerted on the disc $\dot{\mathcal{L}}_\alpha$ is related to the power dissipated $\dot{\mathcal{E}}_\alpha$ via the Keplerian rotation rate Ω, such that

$$\dot{\mathcal{E}}_\alpha = \Omega \dot{\mathcal{L}}_\alpha.$$ (44)

A similar but not equal relation governs the case where potentially global effects are mediated through wave transport. Away from resonances, small-amplitude spiral waves which arise through the gravitational instability can be described within a WKB approach, and are governed via the dispersion relation,

$$(\omega - m\Omega)^2 = c_{\rm s}^2 k^2 - 2\pi G \Sigma |k| + \kappa^2,$$ (45)

where m, k are the radial and azimuthal wave numbers, respectively, and $\omega = m\Omega_{\rm p}$ is the angular frequency of the waves, with $\Omega_{\rm p}$ the wave pattern speed. From this the wave angular momentum and energy densities can be
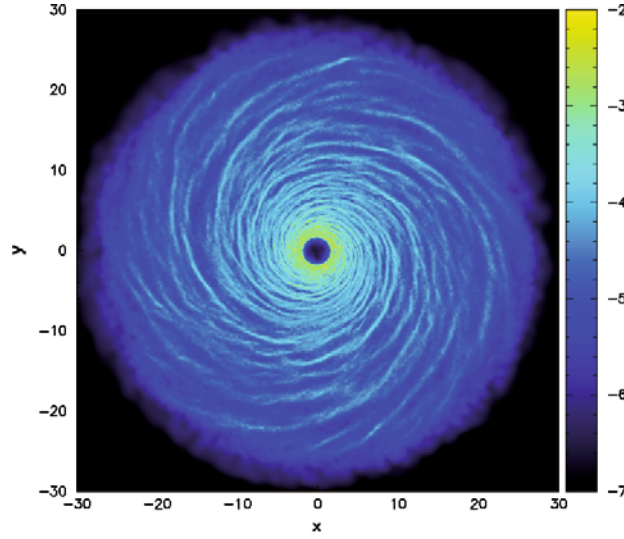
Fig. 2. Surface density plot for a self-gravitating disc with $q = 0.125$, clearly showing the presence of gravitationally induced spiral density waves. The simulation was dimensionless (see, for instance, [19–21]) but this is equivalent to a 0.125 $M_\odot$ disc about a 1 $M_\odot$ protostar, axes in AU and the logscale ranging from 10^{-2}–10^{-7} $M_\odot\mathrm{AU}^{-2}$. Taken from [21].

obtained [21,86,109], and in turn the wave-induced torque $\dot{\mathcal{L}}_\mathrm{w}$ and power dissipation $\dot{\mathcal{E}}_\mathrm{w}$ are found to be related via [21]

$$\dot{\mathcal{E}}_\mathrm{w} = \Omega_\mathrm{p}\dot{\mathcal{L}}_\mathrm{w}. \tag{46}$$

The transport properties of gravitationally induced waves are therefore determined not by the rotation rate of the disc material (cf. eq. (44)), but by the pattern speed of the density waves themselves. As these waves are excited or absorbed, the power exchanged with the background flow for a given stress is therefore significantly different than that dissipated by a viscous process that provides the same stress to the extent to which Ω_p is significantly different from Ω. The relative level of global *vs.* local transport can hence be quantified via the parameter ξ, where

$$\xi = \left| \frac{\Omega - \Omega_\mathrm{p}}{\Omega} \right|. \tag{47}$$

Alternatively, one can directly measure the stress provided by the waves in numerical simulations and compare the resulting value of the effective α with the one predicted from thermal balance in a viscous disc, eq. (43).

Simulations of self-gravitating circumstellar discs using SPH have been been important in studying the gravitational instability. Numerous simulations[3] at varying degrees of complexity have demonstrated that the relative cooling rate of the disc is crucial in determining whether it remains marginally stable to gravitational perturbations or fragments into bound objects. Various studies in both two and three dimensions and with varying equations of state have suggested that the fragmentation boundary occurs where $\beta = \Omega t_\mathrm{cool} \sim 3$–6 [20,54,77–80,93]. While recent results may call absolute values into question [110–112] (see below), the dependence of fragmentation on the cooling rate has been robustly demonstrated.

Transport through gravitational waves has also been simulated using SPH. Using simulations with a fixed value of β, [19,20] the value of α from eq. (43) has been compared to that generated from inter-particle gravitational and Reynolds stresses and found that for low disc to star mass ratios $q = M_\mathrm{disc}/M_\star$, wave transport is a predominantly *local* process, with the waves forming a quasi-stationary pattern on the disc face (as shown in fig. 2). However, as the disc becomes more massive relative to the central star, the wave patterns become more transient, and the transport properties vary accordingly.

Using a disc to star mass ratio q between 0.05 and 0.125, [21] ran simulations of protostellar discs where the cooling time scale t_cool was a fixed fraction of the dynamical time scale Ω^{-1}, such that $\Omega t_\mathrm{cool} = \beta$ was held constant with $4 \leq \beta \leq 10$. Analysis of the disc structure showed that while the cooling rate β does *not* influence the spectrum of wave numbers that are excited, it does affect that amplitude of the density perturbations, such that

$$\frac{\delta\Sigma}{\Sigma} \propto \beta^{-1/2}. \tag{48}$$

---

[3] Not all of which have used SPH —various grid codes have produced very similar results.
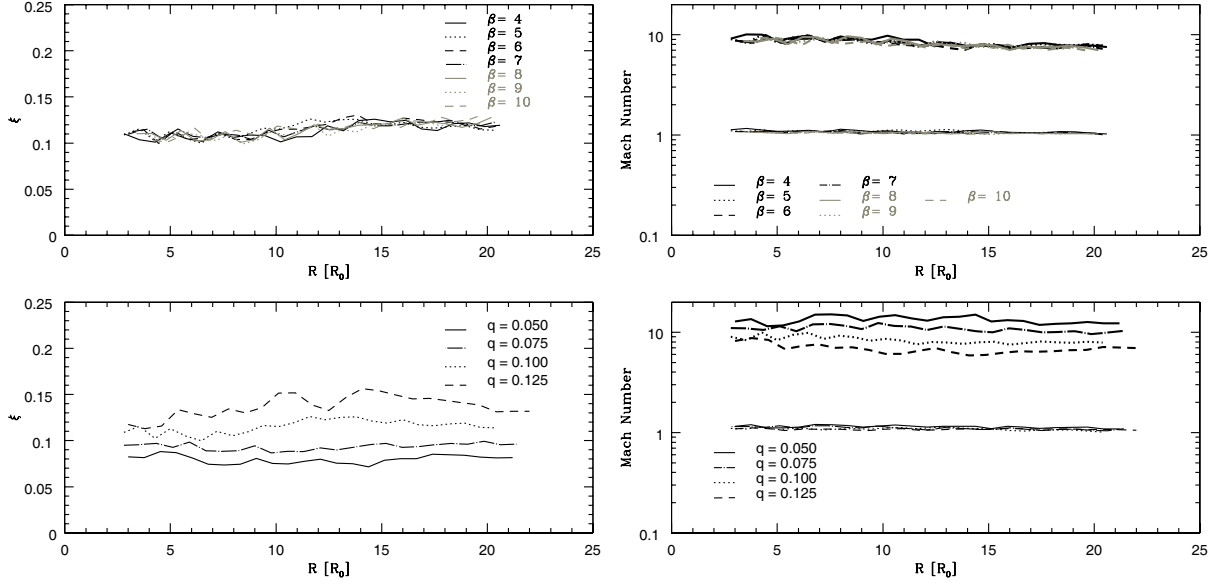
Fig. 3. Non-local transport parameter ξ (left) and Mach Numbers (right) for simulations at various values of mass ratio q and cooling parameter β —where β varies, $q = 0.1$ and where q varies, $\beta = 5$. For the Mach number plots, heavy lines denote the Mach number as measured in an inertial (static) frame, whereas light lines show the Doppler-shifted Mach number (measured in a frame co-rotating with the fluid). Taken from [21].

Thus as the cooling becomes more rapid (and thus as β *decreases*) the amplitude of the density perturbation increases. If the density perturbation exceeds a threshold value, that can be determined numerically, the disc fragments.

Similarly, it was found in [21] that the spectrum of the radial wave number k peaks strongly where $kH = 1$, a result that can be predicted from the dispersion relation eq. (45) but has now also been demonstrated numerically. This result is independent of both the cooling rate and the disc to star mass ratio. The Fourier analysis also showed that the azimuthal wave number m shows a dependence on the disc mass, such that for low q the spectrum peaks at high m (several spiral arms), while heavy discs ($q \approx 1$) are dominated by a small (≈ 2) number of spiral arms. Put together, the results for the radial and azimuthal wave numbers have allowed a spectrally averaged pattern speed to be determined, and thus in turn the non-local transport fraction ξ can be measured from the simulations. In agreement with [19,20], this shows that transport by gravitational waves is a predominantly local process for the systems modeled, with $\xi \approx 10\%$ for $q = 0.1$ and increasing with increasing disc to star mass ratio. This is shown as a function of radius in the left-hand panels of fig. 3, where the increase in non-locality is clearly seen with q. A corollary of this, seen from the form of eq. (47) is that the waves remain on average close to co-rotation, $\Omega_{\rm p} \approx \Omega$.

The right-hand panels of fig. 3 show a further interesting outcome resulting from the simulations of [21] —the wave Mach numbers. While the heavy lines show the values relative to an external inertial frame, the lighter lower lines give the Doppler-shifted Mach numbers $\mathcal{M}$, *i.e.* those relative to a frame co-rotating with the flow. These Doppler-shifted values are almost exactly unity, implying that the density waves excited by the gravitational instability are only weakly supersonic, and furthermore this result is invariant with either cooling rate or mass ratio. The gravitational instability therefore self-regulates so that not only is $Q \approx 1$, but also so that $\mathcal{M} \approx 1$. This result is intuitively reasonable in a quasi-steady disc —subsonic waves would not impart any net heat to the disc (any compression heating is balanced by the corresponding rarefaction cooling), whereas a strong shock would be highly dissipative, leading to rapid evolution. Furthermore, the only way a fluid element can remain on a circular orbit when passing through an oblique (spiral) shock is if that shock wave has a unit Mach number.

Actually, we can also show that the above-mentioned dependence of the factor ξ on the disc to star mass ratio q can be easily understood from the condition that these waves dissipate where they are almost sonic. In fact, using this sonic condition, we can rewrite eq. (47) as

$$\xi = \left| \frac{\Omega - \Omega_{\rm p}}{\Omega} \right| \approx \frac{c_{\rm s}}{v_\phi} \approx \frac{M_{\rm disc}}{M_\star}, \tag{49}$$

where $v_\phi = \Omega R$ is the azimuthal velocity of the disc, and the last equality holds for marginally stable discs ($Q \approx 1$). This trend can actually be seen in fig. 3 (bottom left panel).

3.2 Resolution requirements for measuring the fragmentation threshold in SPH simulations of self-gravitating discs

As described above, the non-linear evolution of the gravitational instability in cooling gaseous discs can lead to two different outcomes, depending on the cooling rate. If the cooling rate is small, or equivalently if the cooling timescale is significantly longer than the dynamical time —$\beta \gg 1$, using the notation introduced above— the amplitude of the perturbations excited by the instability saturates at a finite value, providing a significant source of transport of angular momentum in the disc (due to the non-axisymmetric spiral structure) and energy dissipation through shocks. In the opposite limit $\beta < \beta_{\mathrm{frag}} \approx 3$–6, the instability proceeds in a runaway fashion and the disc fragments into bound objects. The mechanism leading to fragmentation is fairly well understood. As described above, in a cooling disc, thermal saturation of the instability occurs when the amplitude of the density perturbation is large enough that the associated shock heating provided by the instability balances the cooling rate [21]. If the required density perturbation exceeds a critical value, the disc fragments.

Clearly, both regimes are of significant interest for the study of the dynamics of protostellar discs. If the disc does not fragment, the instability might be essential in determining the anomalous transport of angular momentum, generally described in terms of a local viscous model through the α-prescription [61]. In these cases, as discussed above, it is important to assess whether such a local model is appropriate for describing an intrinsically global phenomenon like the gravitational instability. On the other hand, if the disc fragments, this process can have implications for planet formation, and indeed some planet formation theories do appeal to the gravitational instability as a potential way to forming giant plants [113,114]. Determining the critical value of β that marks the transition between the two regimes is thus of significant interest.

Recently, some new results have cast doubts on the reliability of the methods used to determine this critical cooling time, and, in particular, on those obtained by SPH simulations. In particular, [110] have run an extensive set of simulations with different slopes of the disc surface density profile and different disc masses, and have considered the location at which the fragment appeared in their fragmenting runs. Their conclusion is that there is a positive correlation between the imposed value of β and the value of the "local disc mass", $m(R) = \Sigma(R)R^2/M_\star$, evaluated at the radius where the fragments first appear. Then, [111] have noted that, for a given disc set up (in terms of total disc mass, surface density profile, etc.) the critical β for fragmentation is an increasing function of the total number of SPH particles used, and —more importantly— they find no immediate evidence of convergence.

However, it has been shown [112] that both results can be readily explained as a consequence of limited numerical resolution, under the assumption that the resolution requirement for detecting fragmentation in a cooling disc is a function of the cooling time, *i.e.* simulations with long cooling times require higher numerical resolution. Here, we briefly summarize the issue.

The problem of defining the resolution required to simulate self-gravitating discs correctly with SPH has been considered several times over the years. As discussed above, spatial resolution in SPH is measured by the local value of the smoothing length h. This should be compared with the typical length scale of the system at hand, which in the case of discs is the pressure length scale $H = c_{\mathrm{s}}/\Omega$. Note that for a self-regulated Keplerian disc (which we will concentrate on) where $Q \approx 1$, this expression for the disc thickness is equivalent to $H = c_{\mathrm{s}}^2/\pi G \Sigma$, and these two length scales also correspond to the most unstable wavelength for discs subject to the gravitational instability. Indeed, most previous investigations [115,116] indicate a resolution criterion of the form

$$\frac{h}{H} \lesssim f, \tag{50}$$

where f is a number of order unity. It is interesting to evaluate h/H at the onset of the instability, that is in a marginally stable state where $Q \approx 1$ is roughly constant. This can be readily done [112] and gives:

$$\frac{h}{H} = \frac{\eta}{m(R)}\left(\frac{2q}{\pi^2 Q^2 N}\right)^{1/3} \approx \frac{5 \times 10^{-3}}{m(R)}\left(\frac{q}{0.1}\right)^{1/3}\left(\frac{N}{2.5 \times 10^5}\right)^{-1/3}, \tag{51}$$

where $q = M_{\mathrm{disc}}/M_\star$, N is the total number of SPH particles used and $\eta = 1.2$ is the numerical parameter that sets the number of SPH neighbours (see sect. 2). Note that in general $m(R)$ is an increasing function of radius, and thus the outer disc is generally better resolved (smaller h/H) than the inner disc. It is also useful to evaluate h/H at the disc outer boundary:

$$\left.\frac{h}{H}\right|_{\mathrm{out}} = 2\pi\eta\left(\frac{2}{q^2\pi^2 Q^2 N}\right)^{1/3} \approx 0.3\left(\frac{q}{0.1}\right)^{-2/3}\left(\frac{N}{2.5 \times 10^5}\right)^{-1/3}, \tag{52}$$

which holds when $\Sigma \propto R^{-1}$, which is the case for the simulations of [111].

We thus see that both the positive correlation between the critical β and $m(R)$ found by [110] and the positive correlation between the critical β and N found by [111] can be just simply explained as an inverse correlation between
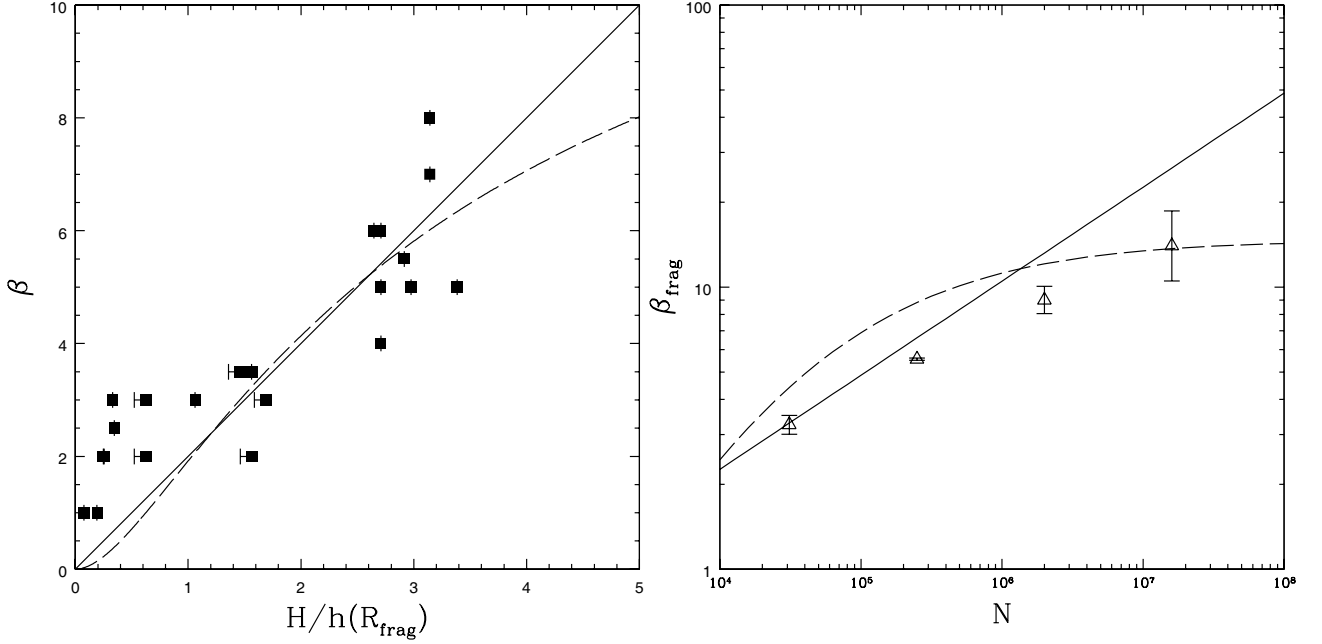
Fig. 4. Left panel: correlation between the imposed value of β and the quantity H/h computed at the location where the fragments appear. Right panel: correlation between the measured value of the critical cooling time for fragmentation β_{frag} and the number of SPH particles used N. Figures from [112], based on data by [110]. The solid lines indicate the best fit linear relation between β and H/h. For a discussion of the dashed lines, see [112].

β and h/H and can thus both be interpreted as an effect of resolution. The correlation between β and h/H, based on the data by [110] is shown in fig. 4 (left panel), while the one between β and N (based on [111]) is shown in fig. 4 (right panel). In particular, if we assume that fragmentation is suppressed due to insufficient resolution and that the resolution requirement becomes more stringent as β increases, we can interpret these results as follows: for a given β, the location at which the fragments first appear [110] marks the location in the disc at which the resolution requirement is first satisfied, that is when $\beta = \beta_{\mathrm{res}}(h/H)$; for a given number of particles, if we are just interested in fragmentation regardless of the location where this happens, the measured critical value of β marks the cooling time for which the resolution criterion is satisfied at the disc outer edge.

We can thus use the data of [110] to get an empirical determination of such cooling dependent resolution requirement. Clearly, due to the scatter in the data in the left panel of fig. 4 and to the limited range of h/H probed by the simulations, a number of different functional relations between β_{res} and h/H can be fitted to the data. We refer the reader to [112] for a more extensive discussion of these. Here, we concentrate on the simplest case of a linear relation. In this case, the best fit to the data (shown as a solid line in fig. 4, left panel) yields

$$\beta_{\mathrm{res}} = 2 \left(\frac{h}{H} \right)^{-1} . \tag{53}$$

If we evaluate the above empirically determined criterion at the disc outer edge, we can thus obtain a criterion for resolving fragmentation in a cooling self-gravitating disc at any radius;

$$\beta \le \beta_{\mathrm{res}} \left(\left. \frac{h}{H} \right|_{\mathrm{out}} \right) \approx 6.6 \left(\frac{q}{0.1} \right)^{2/3} \left(\frac{N}{2.5 \times 10^5} \right)^{1/3} , \tag{54}$$

which is shown with a solid line in the right panel of fig. 4 —one can easily see that this relation is also consistent with the data. Actually, we also note that for the largest number of particles, the measured critical cooling time lies below the predicted line, indicating that there are non-fragmenting simulations in a region of the parameter space where the resolution requirement above is satisfied. This can thus be interpreted as a tentative evidence that convergence is being reached for $N \approx 2$ millions and the critical cooling time for fragmentation would thus lie in the region $\beta \approx 10$–15.

We need to discuss two further issues at this point; a) what is the physical origin of the resolution criterion, and b) which of the previous results are affected by poor resolution, excluding the fragmentation threshold?

A clue to answering the first question comes from the following argument. It is well known [108, 93] that in thermal equilibrium viscous (Keplerian) discs satisfy eq. (43), which indicates that the α viscosity coefficient is inversely

proportional to β. On the other hand, we also know that the equivalent α-coefficient corresponding to artificial viscosity in SPH (eq. (40)) is linearly proportional to h/H. Thus, eq. (53) corresponds to a region of the parameter space for which artificial viscosity provides a fixed fraction of the overall energy budget in the disc, and the resolution criterion thus simply requires that in order to resolve fragmentation, the artificial viscosity should contribute less than this fraction of the energy budget. This looks as a fair requirement —indeed, if artificial viscous heating completely offsets the imposed cooling, clearly no gravitational perturbations will be able to grow. Putting the relevant numbers in, we obtain that our resolution criterion eq. (53) corresponds to cases where artificial viscosity contributes to 5 per cent of the overall energy budget, thus implying that the condition is rather stringent. Alternative physical explanations can be found in [112].

Finally, we discuss what the implications of the above results for our understanding of the dynamics of gravitationally unstable protostellar discs are. First and foremost, [111] have demonstrated that determining the critical value of the cooling parameter β might have been affected by poor resolution. On the other hand, we have just shown that, if indeed convergence has been reached at 2 million particles, this would only produce a modest change, of order of a factor 2, on the value of β_{frag}. Still, such a change might affect the location where fragmentation may be predicted to occur in protostellar discs. However, in the regime of ice cooling appropriate to the outer parts of protostellar discs, this change would bring in the minimum radius for fragmentation by a factor $\sim 2^{2/9} \approx 1.2$ [100, 79]. Although this change is quite modest, it may be important to the discussion of whether recently imaged planets (*e.g.*, around HR 8799, [117], around β Pic, [118] or around Fomahault, [119]) could have been formed by gravitational instability. Note, however, that —for a given critical cooling time— the uncertainty in determining the fragmentation radius due to uncertainties in the relevant opacities, the detailed vertical structure of the disc and the effects of magnetic fields are probably larger than the modest change discussed above.

More in general, we might ask which of the properties of non-fragmenting self-gravitating accretion discs are going to be significantly affected by poor resolution, in the sense expressed by eq. (53). Probably the most important property, apart from fragmentation, is related to the stress induced by the instability, and in particular its magnitude and the locality of the associated dissipation (which, from the discussion above, are in turn related to the power spectrum of the modes excited by the instability). Here the question is whether the typical wavelength of the disturbances is a significant fraction of the radius R. The studies discussed in sect. 3.1 (and in particular [21]) have shown that the spectrum of the disturbances peaks at a wavelength $1/k \approx H$ and becomes negligible for $1/k \gtrsim 3H$. In this respect, the relevant resolution requirement is the generally less restrictive condition $h/H \lesssim 1$ already discussed in [116]. Furthermore, not resolving the smallest scales would only prevent the disc from developing small-scale structures, which would not in any case contribute to global transport.

Let us now discuss whether the evaluation of the stress from such simulations is affected by eq. (53). Clearly, *for a given disturbance*, regardless of the way it has been excited, a correct evaluation of the resulting stress only requires that the typical wavelengths of the disturbance are resolved. However, here the question is whether the disc response to an external boundary condition (in this case, to an externally imposed cooling rate) depends on resolution. The condition of thermal equilibrium relates in a unique way the external cooling to the magnitude of the stress induced by the instability (eq. (43)), under the hypothesis that the only heating source in the disc is provided by the instability itself, and that this heat is dissipated locally. Here, we have shown that the new empirically determined resolution condition corresponds to the case where artificial viscosity only provides 5 per cent of the heating necessary for thermal balance. Therefore, unless the effects of artificial viscosity are much larger than expected, one would not expect a change in the stress above the 5 per cent level. In this respect, the effect of poor resolution is just to suppress the development of small-scale disturbances, and thus reduce the peak value of the density perturbation, while keeping the overall average value of the stress essentially unchanged.

4 Conclusions

In this paper we have described some applications of the smoothed particle hydrodynamics method to the dynamics of protostellar discs. We have first described some important features of the method that make it especially well suited to treat several astrophysical phenomena, in particular its remarkable conservation properties, which can be elegantly derived from a Lagrangian approach. We have also discussed the extent to which SPH can be used to simulate accretion discs. We have paid special attention to the issue of numerical modeling of dissipative effects in SPH and in particular to viscous effects. We have discussed how artificial viscosity, which is essential in order to model shocks in SPH correctly, can also be effectively used to mimic an α-like viscosity for an accretion disc [65], although with some caveats.

Numerical simulations using SPH have been essential in improving our understanding of the evolution of such discs and in particular the non-linear evolution of gravitational instabilities. Such evolution might lead to two possible outcomes; for cooling times much longer than the dynamical time the instability can saturate due to thermal effects at a finite amplitude and provide through the resulting spiral structure a significant source of angular momentum transport, allowing accretion of the protostellar nebula onto the forming star. Conversely, for short cooling times the

saturation mechanism is not effective and the instability proceeds in a runaway fashion, leading to the formation of gravitationally bound satellites (giant planets, or low-mass stellar companions) of the central star. Here, we have principally described results concerning the non-fragmenting case, where the disc evolves toward a self-regulated state of marginal stability, and where the energy losses due to cooling are balanced by the shock heating from the instability. We have described how, in this self-regulated state, the Mach number of the spiral disturbances with respect to the local background flow is found to be very close to sonic. We have thus shown how this condition implies that for light discs ($M_{\mathrm{disc}} \ll M_\star$) the transport induced by the instability is essentially a local process, amenable to description in terms of a simple effective viscosity prescription [100]. On the other hand, when the disc mass is a significant fraction of the central object mass, the above description fails and transport becomes dominated by global spiral modes.

A recent important numerical result in this context is related to the determination of the exact threshold — in terms of the cooling time— that separates the two regimes of self-regulated (non-fragmenting) and fragmenting discs. Previous analyses that used SPH [78] have set this threshold to $\approx 6\Omega^{-1}$. However, recently [111] it has been shown that such results have not converged numerically, and that the threshold cooling time for fragmentation keeps increasing with increasing resolution, with no apparent convergence. We have then discussed how these results [112], together with other related results concerning the location within the disc where the fragments appear [110] are in fact a manifestation of a single effect, namely that the resolution requirement for cooling self-gravitating discs is a function of the cooling time. Based on this analysis, we conclude that while indeed previous numerical determinations of the threshold for fragmentation are affected by poor resolution, convergence is tentatively reached at a resolution of roughly 2 million particles, indicating that the "true" value of the threshold cooling time for fragmentation lies in the range 10–$15\Omega^{-1}$, roughly a factor 2 larger than previously thought. On the other hand, we argue that those results pertaining to the locality of transport induced by the instability in non-fragmenting discs, and the magnitude of the relative stress, should be free from these resolution effects.

Appendix A.

For completeness, the equations for an SPH implementation in three dimensions that is fully conservative of mass, linear and angular momentum and total energy and includes variable spatial resolution, gravity and an artificial viscosity for capturing non-isentropic effects such as shocks, are given in full below. As previously, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$, $W_{ij}(h_i) = W(\mathbf{r}_{ij}, h_i)$ and $\phi_{ij}(\varepsilon_i) = \phi(\mathbf{r}_{ij}, \varepsilon_i)$.

The fluid density ρ and particle smoothing length h are determined through an implicit solution of the following equations, which guarantee conservation of mass

$$\rho_i = \sum_j m_j W_{ij}(h_i), \tag{A.1}$$

and

$$\rho_i h_i^3 = m_i \eta^3, \tag{A.2}$$

where m_i is the particle mass and $\eta \approx 1.2$ is a dimensionless parameter linked to the mass within a defined smoothing kernel W. In a similar manner the gravitational potential Φ is given by

$$\Phi_i = G \sum_j m_j \phi_{ij}(\varepsilon_i), \tag{A.3}$$

where $\varepsilon_i = \varepsilon_i(h_i)$ is the gravitational softening length of the particle, frequently set equal to the smoothing length, and as before the softening kernel ϕ_{ij} is determined from the smoothing kernel via Poisson's equation, such that

$$\phi(r, \varepsilon) = 4\pi \left[-\frac{1}{r} \int^r r'^2 W(r', h) \mathrm{d}r' + \int^r r' W(r', h) \mathrm{d}r' \right] + \frac{C_1}{r^2} + \frac{C_2}{r}. \tag{A.4}$$

Here C_1 is such that the Newtonian potential is recovered for $r \geq 2\varepsilon$ and C_2 ensures that $\phi \to 0$ as $r \to \infty$.

Conservation of momentum is given by the following equation;

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = &-\sum_j m_j \left(\left[\frac{P_i}{\hat{\Omega}_i \rho_i^2} + \frac{\Pi_{ij}}{2} + \frac{G}{2}\frac{\chi_i}{\hat{\Omega}_i} \right] \nabla_i W_{ij}(h_i) + \frac{G}{2}\nabla_i \phi_{ij}(\varepsilon_i) \right) \\
&-\sum_j m_j \left(\left[\frac{P_j}{\hat{\Omega}_j \rho_j^2} + \frac{\Pi_{ij}}{2} + \frac{G}{2}\frac{\chi_j}{\hat{\Omega}_j} \right] \nabla_i W_{ij}(h_j) + \frac{G}{2}\nabla_i \phi_{ij}(\varepsilon_j) \right).
\end{aligned} \tag{A.5}$$

In each summation, the first term concerns accelerations due to purely hydrodynamic (pressure P) forces, the second one gives the acceleration due to artificial viscosity Π_{ij} (see main text for a simple implementation), the third one is a correction term to account for a variable gravitational softening length and the fourth one is the gravitational acceleration. The correction factors $\hat{\Omega}$ and χ for variation in the smoothing and softening lengths, respectively, are given by

$$\hat{\Omega}_i = 1 + \frac{h_i}{3\rho_i} \sum_j m_j \frac{\partial W_{ij}(h_i)}{\partial h_i}, \tag{A.6}$$

and

$$\chi_i = -\frac{h_i}{3\rho_i} \frac{\mathrm{d}\varepsilon_i}{\mathrm{d}h_i} \sum_j m_j \frac{\partial \phi_{ij}(\varepsilon_i)}{\partial \varepsilon_i}. \tag{A.7}$$

Conservation of total energy is ensured if the specific internal energy u is determined using

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\hat{\Omega}_i \rho_i^2} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}(h_i) + \frac{1}{2} \sum_j m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \overline{\nabla_i W_{ij}}, \tag{A.8}$$

where, as before, $\overline{\nabla_i W_{ij}} = \left[\nabla_i W(\mathbf{r}_{ij}, h_i) + \nabla_i W(\mathbf{r}_{ij}, h_j) \right]/2$. Here again, the first term accounts for changes in internal energy due to purely hydrodynamic effects ($P\mathrm{d}V$ work) and the second term accounts for viscous heating.

Finally, to close the set of equations an equation of state is required. Here there is a degree of latitude —an oft-used case is the ideal gas equation of state with ratio of specific heats γ, in the form

$$P_i = \rho_i u_i (\gamma - 1). \tag{A.9}$$

References

1. L.B. Lucy, Astron. J. **82**, 1013 (1977).
2. R.A. Gingold, J.J. Monaghan, Mon. Not. R. Astron. Soc. **181**, 375 (1977).
3. W. Benz, in *The Numerical Modeling of Nonlinear Stellar Pulsations*, edited by J. Buchler (Kluwer, Dordrecht, 1990).
4. J.J. Monaghan, Annu. Rev. Astron. Astrophys. **30**, 543 (1992).
5. V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).
6. V. Springel, S.D.M. White, A. Jenkins, C.S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly *et al.*, Nature **435**, 629 (2005).
7. S. Borgani, G. Murante, V. Springel, A. Diaferio, K. Dolag, L. Moscardini, G. Tormen, L. Tornatore, P. Tozzi, Mon. Not. R. Astron. Soc. **348**, 1078 (2004).
8. S. Callegari, L. Mayer, S. Kazantzidis, M. Colpi, F. Governato, T. Quinn, J. Wadsley, Astrophys. J. Lett. **696**, L89 (2009).
9. S. Rosswog, Astrophys. J. **634**, 1202 (2005).
10. D.J. Price, S. Rosswog, Science **312**, 719 (2006).
11. S. Rosswog, E. Ramirez-Ruiz, W.R. Hix, Astrophys. J. **695**, 404 (2009).
12. E.M. Rossi, G. Lodato, P.J. Armitage, J.E. Pringle, A.R. King, Mon. Not. R. Astron. Soc. **401**, 2021 (2010).
13. C.R. Evans, C.S. Kochanek, Astrophys. J. **346**, L13 (1989).
14. G. Lodato, A.R. King, J.E. Pringle, Mon. Not. R. Astron. Soc. **392**, 332 (2009).
15. C.L. Dobbs, I.A. Bonnell, Mon. Not. R. Astron. Soc. **385**, 1893 (2008).
16. C.L. Dobbs, D.J. Price, Mon. Not. R. Astron. Soc. **383**, 497 (2008).
17. M.R. Bate, I.A. Bonnell, V. Bromm, Mon. Not. R. Astron. Soc. **339**, 577 (2003).
18. M.R. Bate, Mon. Not. R. Astron. Soc. **404**, L79 (2010).
19. G. Lodato, W.K.M. Rice, Mon. Not. R. Astron. Soc. **351**, 630 (2004).
20. G. Lodato, W.K.M. Rice, Mon. Not. R. Astron. Soc. **358**, 1489 (2005).
21. P. Cossins, G. Lodato, C.J. Clarke, Mon. Not. R. Astron. Soc. **393**, 1157 (2009).
22. L. Mayer, G. Lufkin, T. Quinn, J. Wadsley, Astrophys. J. **661**, L77 (2007).
23. J.J. Monaghan, Rep. Prog. Phys. **68**, 1703 (2005).
24. M.J. Berger, P. Colella, J. Comput. Phys. **82**, 64 (1989).
25. L. Hernquist, N. Katz, Astrophys. J. Suppl. **70**, 419 (1989).
26. M.R. Bate, I.A. Bonnell, N.M. Price, Mon. Not. R. Astron. Soc. **277**, 362 (1995).
27. O. Flebbe, S. Muenzel, H. Herold, H. Riffert, H. Ruder, Astrophys. J. **431**, 754 (1994).
28. P. Español, M. Revenga, Phys. Rev. E **67**, 026705 (2003).
29. D. Sijacki, V. Springel, Mon. Not. R. Astron. Soc. **371**, 1025 (2006).
30. S.C. Whitehouse, M.R. Bate, J.J. Monaghan, Mon. Not. R. Astron. Soc. **364**, 1367 (2005).
31. L. Tornatore, S. Borgani, F. Matteucci, S. Recchi, P. Tozzi, Mon. Not. R. Astron. Soc. **349**, L19 (2004).
32. W. Benz, F. Thielemann, J.G. Hills, Astrophys. J. **342**, 986 (1989).
33. D.J. Price, J.J. Monaghan, Mon. Not. R. Astron. Soc. **348**, 123 (2004).
34. D.J. Price, J.J. Monaghan, Mon. Not. R. Astron. Soc. **348**, 139 (2004).

35. D.J. Price, J.J. Monaghan, Mon. Not. R. Astron. Soc. **364**, 384 (2005).
36. D.J. Price, Mon. Not. R. Astron. Soc. **401**, 1475 (2010).
37. S. Rosswog, J. Comput. Phys. **229**, 8591 (2010).
38. S. Rosswog, New Astron. Rev. **53**, 78 (2009).
39. J.J. Monaghan, D.J. Price, Mon. Not. R. Astron. Soc. **328**, 381 (2001).
40. D. Price, PhD Thesis, ArXiv Astrophysics e-prints **0507472** (2005).
41. D.J. Price, ArXiv e-prints (2010) `1012.1885`.
42. P.J. Cossins, arXiv:1007.1245 (2010).
43. M.S. Fulbright, W. Benz, M.B. Davies, Astrophys. J. **440**, 254 (1995).
44. D.A. Fulk, D.W. Quinn, J. Comput. Phys. **126**, 165 (1996).
45. J.J. Monaghan, Comput. Phys. Rep. **3**, 71 (1985) ISSN 0167-7977.
46. J.I. Read, T. Hayfield, O. Agertz, Mon. Not. R. Astron. Soc. **405**, 1513 (2010).
47. J.J. Monaghan, J.C. Lattanzio, Astron. Astrophys. **149**, 135 (1985).
48. L. Cullen, W. Dehnen, Mon. Not. R. Astron. Soc. **408**, 669 (2010).
49. C. Eckart, Phys. Fluids **3**, 421 (1960).
50. W. Dehnen, Mon. Not. R. Astron. Soc. **324**, 273 (2001).
51. D.J. Price, J.J. Monaghan, Mon. Not. R. Astron. Soc. **374**, 1347 (2007).
52. V. Springel, L. Hernquist, Mon. Not. R. Astron. Soc. **333**, 649 (2002).
53. V. Springel, N. Yoshida, S.D.M. White, New Astron. Rev. **6**, 79 (2001).
54. B.M. Johnson, C.F. Gammie, Astrophys. J. **597**, 131 (2003).
55. D. Forgan, K. Rice, D. Stamatellos, A. Whitworth, Mon. Not. R. Astron. Soc. **394**, 882 (2009).
56. M. Petkova, V. Springel, Mon. Not. R. Astron. Soc. **396**, 1383 (2009).
57. S. Nayakshin, S. Cha, A. Hobbs, Mon. Not. R. Astron. Soc. **397**, 1314 (2009).
58. J.J. Monaghan, J. Comput. Phys. **136**, 298 (1997).
59. D.S. Balsara, PhD Thesis, Illinois University, Urbana-Champaign (1991).
60. J.P. Morris, J.J. Monaghan, J. Comput. Phys. **136**, 41 (1997).
61. N.I. Shakura, R.A. Sunyaev, Astron. Astrophys. **24**, 337 (1973).
62. A.R. King, J.E. Pringle, M. Livio, Mon. Not. R. Astron. Soc. **376**, 1740 (2007).
63. L. Hartmann, N. Calvet, E. Gullbring, P. D'Alessio, Astrophys. J. **495**, 385 (1998).
64. G. Lodato, C.J. Clarke, Mon. Not. R. Astron. Soc. **353**, 841 (2004).
65. G. Lodato, D.J. Price, Mon. Not. R. Astron. Soc. **405**, 1212 (2010).
66. P. Artymowicz, S.H. Lubow, Astrophys. J. **421**, 651 (1994).
67. J.R. Murray, Mon. Not. R. Astron. Soc. **279**, 402 (1996).
68. D.J. Price, C. Federrath, Mon. Not. R. Astron. Soc. **406**, 1659 (2010).
69. O. Agertz, B. Moore, J. Stadel, D. Potter, F. Miniati, J. Read, L. Mayer, A. Gawryszczak, A. Kravtsov, Å. Nordlund *et al.*, Mon. Not. R. Astron. Soc. **380**, 963 (2007).
70. D.J. Price, J. Comput. Phys. **227**, 10040 (2008).
71. S. Inutsuka, J. Comput. Phys. **179**, 238 (2002).
72. S.H. Cha, A.P. Whitworth, Mon. Not. R. Astron. Soc. **340**, 73 (2003).
73. F. Shu, Astrophys. J. **214**, 488 (1977).
74. F. Shu, F.C. Adams, S. Lizano, Annu. Rev. Astron. Astrophys. **25**, 23 (1987).
75. C.F. McKee, E.C. Ostriker, Annu. Rev. Astron. Astrophys. **45**, 565 (2007).
76. S.A. Balbus, Annu. Rev. Astron. Astrophys. **41**, 555 (2003).
77. W.K.M. Rice, P.J. Armitage, M.R. Bate, I.A. Bonnell, Mon. Not. R. Astron. Soc. **339**, 1025 (2003).
78. W.K.M. Rice, G. Lodato, P.J. Armitage, Mon. Not. R. Astron. Soc. **364**, L56 (2005).
79. P. Cossins, G. Lodato, C. Clarke, Mon. Not. R. Astron. Soc. **401**, 2587 (2010).
80. D. Stamatellos, D.A. Hubber, A.P. Whitworth, Mon. Not. R. Astron. Soc. **382**, L30 (2007).
81. D. Stamatellos, A.P. Whitworth, Astron. Astrophys. **480**, 879 (2008).
82. D. Stamatellos, A.P. Whitworth, Mon. Not. R. Astron. Soc. **392**, 413 (2009).
83. D. Stamatellos, A.P. Whitworth, Mon. Not. R. Astron. Soc. **400**, 1563 (2009).
84. S. Walch, A. Burkert, A. Whitworth, T. Naab, M. Gritschneder, Mon. Not. R. Astron. Soc. **400**, 13 (2009).
85. S. Walch, T. Naab, A. Whitworth, A. Burkert, M. Gritschneder, Mon. Not. R. Astron. Soc. **402**, 2253 (2010).
86. G. Bertin, *Dynamics of Galaxies* (Cambridge University Press, 2000).
87. G. Bertin, Astrophys. J. **478**, L71 (1997).
88. G. Bertin, G. Lodato, Astron. Astrophys. **350**, 694 (1999).
89. M. Krumholz, A. Burkert, Astrophys. J. **724**, 895 (2010).
90. A. Toomre, Astrophys. J. **139**, 1217 (1964).
91. W.K.M. Rice, G. Lodato, P.J. Armitage, Mon. Not. R. Astron. Soc. **364**, L56 (2005).
92. W.K.M. Rice, P.J. Armitage, I.A. Bonnell, M.R. Bate, S.V. Jeffers, S.G. Vine, Mon. Not. R. Astron. Soc. **346**, L36 (2003).
93. C.F. Gammie, Astrophys. J. **553**, 174 (2001).
94. A.P. Boss, Science **276**, 1836 (1997).
95. A.P. Boss, Nature **393**, 141 (1998).
96. A.P. Boss, Astrophys. J. **536**, L101 (2000).

97. A.C. Boley, Astrophys. J. Lett. **695**, L53 (2009) 0902.3999.
98. A.C. Boley, T. Hayfield, L. Mayer, R.H. Durisen, Icarus **207**, 509 (2010).
99. R. Rafikov, Astrophys. J. **621**, 69 (2005).
100. C.J. Clarke, Mon. Not. R. Astron. Soc. **396**, 1066 (2009).
101. E.I. Vorobyov, S. Basu, Astrophys. J. **714**, L133 (2010).
102. E.I. Vorobyov, S. Basu, Astrophys. J. **719**, 1896 (2010).
103. B. Paczynski, Acta Astron. **28**, 91 (1978).
104. G. Lodato, Riv. Nuovo Cimento **30**, 293 (2007).
105. B.K. Pickett, A.C. Mejía, R.H. Durisen, P.M. Cassen, D.K. Berry, R.P. Link, Astrophys. J. **590**, 1060 (2003).
106. A.C. Mejia, R.H. Durisen, M.K. Pickett, K. Cai, Astrophys. J. **619**, 1098 (2005).
107. A.C. Boley, A.C. Mejía, R.H. Durisen, K. Cai, M.K. Pickett, P. D'Alessio, Astrophys. J. **651**, 517 (2006).
108. J.E. Pringle, Annu. Rev. Astron. Astrophys. **19**, 137 (1981).
109. Z. Fan, Y.Q. Lou, Mon. Not. R. Astron. Soc. **307**, 645 (1999).
110. F. Meru, M.R. Bate, Mon. Not. R. Astron. Soc. **410**, 559 (2011).
111. F. Meru, M.R. Bate, Mon. Not. R. Astron. Soc. **411**, L1 (2011).
112. G. Lodato, C.C. Clarke, ArXiv e-prints (2011) 1101.2448.
113. A.P. Boss, Astrophys. J. **610**, 456 (2004).
114. A.P. Boss, Astrophys. J. **641**, 1148 (2006).
115. M.R. Bate, A. Burkert, Mon. Not. R. Astron. Soc. **288**, 1060 (1997).
116. A.F. Nelson, Mon. Not. R. Astron. Soc. **373**, 1039 (2006).
117. C. Marois, B. Macintosh, T. Barman, B. Zuckerman, I. Song, J. Patience, D. Lafrenière, R. Doyon, Science **322**, 1348 (2008).
118. A. Lagrange, D. Gratadour, G. Chauvin, T. Fusco, D. Ehrenreich, D. Mouillet, G. Rousset, D. Rouan, F. Allard, É. Gendron *et al.*, Astron. Astrophys. **493**, L21 (2009).
119. P. Kalas, J.R. Graham, E. Chiang, M.P. Fitzgerald, M. Clampin, E.S. Kite, K. Stapelfeldt, C. Marois, J. Krist, Science **322**, 1345 (2008).