Numerical Methods for Simulating and Understanding the Universe

JOSHUA JAMES BORROW

Numerical Methods for Simulating and Understanding the Universe

Joshua James Borrow

Abstract: Only within the past century have we discovered the existence of external galaxies outside our own Milky Way. The study of the formation and evolution of galaxies is now an entire field unto itself, with a key part of this field being the direct numerical simulation of galaxy formation. These simulations naturally depend heavily on an assortment of numerical methods, from fluid solvers to detailed prescriptions for metal evolution in stars. This thesis explores existing numerical methods and develops novel methods for simulating astrophysical fluids and analysing the baryon cycle in galaxies. We first show how the commonly used Pressure-SPH (Smoothed Particle Hydrodynamics) method leads to large integration errors when coupled to galaxy formation physics, before developing a new SPH method called Sphenix that does not suffer from the same errors. Sphenix is based on Density-SPH, and employs a novel artificial conduction scheme to reduce errors at contact discontinuities. Sphenix is then shown to solve a number of challenging problems for SPH, including vorticity conservation and fluid mixing, thanks to its conduction and viscosity schemes. Finally, we develop two new numerical schemes to study the baryon cycle in the Simba simulations. The spread metric is used to show that matter can be transported huge distances ($\gg 10$ Mpc) by redshift $z = 0$, primarily due to AGN feedback. By comparing the Lagrangian region that gas resides in at the initial state of the simulation to its resident halo at $z = 0$ we show how matter can be transported between bound haloes at the end of the simulation. Notably, we show that 5-10% of the baryonic mass in a typical Milky Way mass halo originated in the region defined by the dark matter of another halo, leading to potential difficulties for so-called 'zoom-in' simulations.

Numerical Methods for Simulating and Understanding the Universe

Joshua James Borrow

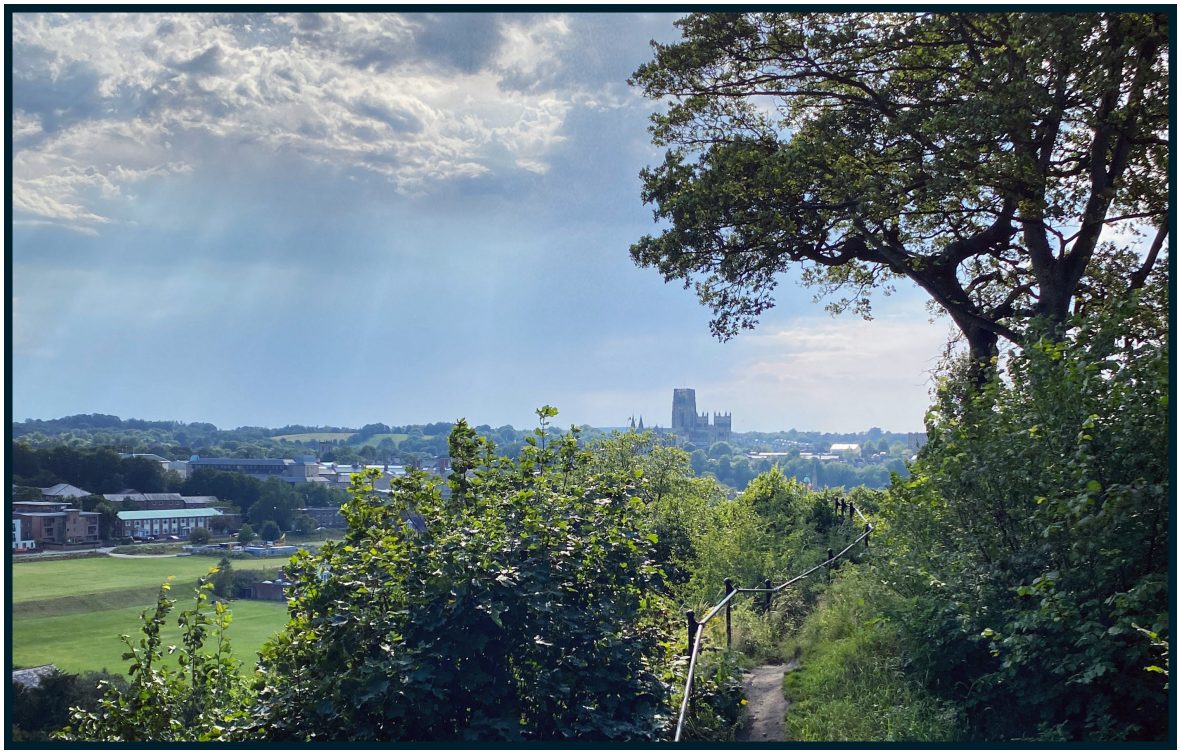
A thesis presented for the degree of
Doctor of Philosophy

Institute for Computational Cosmology
Department of Physics
Durham University
United Kingdom

July 2021

Dedicated to

Durham, the city that stole my heart and became my home.



Contents

Declaration

The work in this thesis is based on research carried out in the Institute for Computational Cosmology, Department of Physics, Durham University, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

The content presented in Chapter 2 and Chapter 4 have been published in:

Borrow, J., Schaller, M., and Bower, R. G., "Inconsistencies arising from the coupling of galaxy formation sub-grid models to pressure-smoothed particle hydrodynamics", Monthly Notices of the Royal Astronomical Society, vol. 505, no. 2, pp 2316-2327, 2021.

Borrow, J., Anglés-Alcázar, D., and Davé, R., "Cosmological baryon transfer in the SIMBA simulations", Monthly Notices of the Royal Astronomical Society, vol. 491, no. 4, pp. 6102–6119, 2020.

The content presented in Chapter 3 has been resubmitted to MNRAS after a round of peer review as:

Borrow, J., Schaller M., Bower R.G., and Schaye, J., "Sphenix: Smoothed Particle Hydrodynamics for the next generation of galaxy formation simulations".

The author of the thesis was primarily responsible for all aspects of these publications. The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

To Richard, thank you for always being enthusiastic, believing in me, and giving me the freedom that I needed. I never expected this PhD to be such a journey, and although I think it's fair to say that we haven't always agreed on everything, there is nobody I would rather have had guiding me than you. I would never have guessed that when I walked into your office seven years ago that this would be where our path would lead. Thank you for everything.

To Matthieu, I can honestly say that you are the only reason that I made it through this process. Thank you for your advice, support, and friendship, especially this past year. Your constant (admittedly sometimes frustrating) attention to detail has made me a much better scientist, and you taught me that science doesn't have to be a competition.

To my co-authors: Joop, thank you for always being right, even when I don't want you to be, and never backing down; Daniel, thanks to your pursuance of excellence, our paper was a thousand times better than it was when I first thought it was 'done'; Romeel, the respect you gave me from day one boosted my confidence more than any award or honour ever could; Bert, thank you for teaching me so much and never giving in to my hand-wavy arguments; and Ash, thank you for all of your insightful comments and reminding me that not all excellent scientists fit into the same mould.

It has been an honour to be part of the SWIFT, EAGLE-XL, and COLIBRE teams, and the people who make them up are some of the finest engineers and scientists I have ever had the opportunity to meet. Thank you to all of them for their advice, hard work, and dedication.

None of my work would have been possible without the COSMA team, so thank you Lydia, Alastair, John, Richard, and Peter.

I have to thank Till Sawala for helping me get my first opportunity in science.

To my friends:

Stu, Behzad, Victor, Aidan, and Rachel, thank you for sticking around during the pandemic and supporting me through the job season and the final years of my PhD.

Chris and Christian, my forever office mates. Thank you for putting up with my obsession of taking over the office, and especially the coffee grinder. I miss you dearly and wish we could be back for the end.

Alex, Andrew, Anna, Arnau, Azi, Calvin, Carol, Christoph, Ellen, Jack, Jacob, Jaime, Jake, James, Joaquin, Kyle, Louise, Oliver, Piotr, Stefan, TK, Toms C and R, Ugne, and Vicky (and probably some people I forgot, I'm sorry); thank you for making the ICC such a welcoming place, I will miss you all.

With this last year being spent mainly at home, with little company, I must say thank you to the music that kept me going: Ariana Grande, Bleachers, Bombay Bicycle Club, Carly Rae Jepsen, CHVRCHES, easy life, Florence + The Machine, Foo Fighters, HONNE, Jack River, Litany, Little Comets, Maggie Rogers, Olivia Rodrigo, Sam Fender, Sea Girls, Taylor Swift, The Kooks, The Wombats, Wolf Alice, and many others. Additionally, I must thank Jimmy Broadbent whose videos dug me out of a real isolation hole, and gave me a new hobby.

The work in Chapter 4 was initiated as a project for the Kavli Summer Program in Astrophysics held at the Center for Computational Astrophysics of the Flatiron Institute in 2018. The program was co-funded by the Kavli Foundation and the Simons Foundation. This work was also supported by collaborative visits funded by the Cosmology and Astroparticle Student and Postdoc Exchange Network (CASPEN), financially supported by the Flatiron Institute and the Simons Foundation. I thank them for their generous support.

I was supported by STFC studentship ST/R504725/1. I was additionally supported by Intel for a number of conferences that were vital to the successful completion of this thesis.

This thesis used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

This thesis made use of the following software and libraries:

- SWIFT (Schaller et al., 2018)
- GIZMO (Hopkins, 2017)
- GADGET (Springel, 2005)
- `python` (van Rossum & Drake Jr, 1995), with the following libraries
  - `numpy` (Harris et al., 2020)
  - `scipy` (SciPy 1.0 Contributors et al., 2020)
  - `numba` (Lam et al., 2015)
  - `matplotlib` (Hunter, 2007)
  - `swiftsimio` (Borrow & Borrisov, 2020; Borrow & Kelly, 2021)
  - `py-sphviewer` (Benitez-Llambay, 2015)
  - `yt` (Turk et al., 2011)
- VELOCIRAPTOR (Elahi et al., 2019)
- The Amiga Halo Finder (AHF) (Gill et al., 2004; Knollmann & Knebe, 2009)

Chapter 1

Introduction

1.1 Introduction to the Thesis

The past century has seen a revolution in our understanding of the Universe. Only in Hubble (1929) was the concept that the Universe may be expanding discovered observationally, and our concept that there may be galaxies outside our own developed. In the intervening time a huge body of knowledge has been created around galaxies, their formation, and their evolution.

Another step change in our understanding came when forward modelling of galaxy evolution became feasible, as recently as the turn of the 21st Century thanks to massive increases in the available computing resource (see e.g. Springel & Hernquist, 2003). Just two decades later there is a whole industry of cosmological galaxy formation simulations that aim to recreate the history of the universe over cosmic time through direct simulation (Vogelsberger et al., 2014; Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019).

At the core of these simulations are the numerical methods used to simulate fluids, gravity on large scales, and the core physics powering the evolution of galaxies such as star formation. The analysis of these simulations also requires the development of novel methods and algorithms. Such methods are as (if not more) important than the increase in computing power, and the staggering pace of their development over the past few decades has allowed increases in fidelity previously believed impossible (Hopkins et al., 2018; Marinacci et al., 2019).

In this thesis, we will explore these numerical tools in detail, finding errors in methods that were previously hidden, developing new tools for simulating astronomical fluids, and creating novel methods for analysing simulations to understand the baryon cycle in galaxies.

1.2 Introduction to Numerical Methods

There are many problems fundamental to our understanding of the Universe that elude analytical solutions, that can be written down explicitly in algebraic form, and as such must be solved approximately. Such problems span from the values of irrational numbers like $\sqrt{2}$, through the motion of a pendulum, all the way to extremely complex problems like the interaction of partially charged oil with turbine blades. To solve these problems we can employ simplifications to allow the production of a closed-form solution (for instance the small angle approximation in the case of the pendulum), or we can formulate adjacent representations that can be solved numerically through the use of repeated arithmetic operations.

Before we take a journey on the origin and history of numerical methods and their application to astrophysics, let us take a brief view on the necessity of numerical methods. Consider the function:

$$f(x) = e^{e^x} \tag{1.2.1}$$

This function is well behaved in all typical senses, it is smooth, and has a well defined derivative; however, attempts to integrate it will frustrate even the most hardy mathematicians. Numerically, though, this function can be integrated easily, thanks to its smooth nature. We can use a left-hand Riemann sum (see Fig. 1.1) to evaluate the definite integral

$$I = \int_0^1 e^{e^x}\ \mathrm{dx}. \tag{1.2.2}$$

The left-hand Riemann sum is calculated using the following algorithm:

- Take your domain $[a,b]$ (in our case this is $[0,1]$), and split it into n subdivisions of width $\Delta x = (b-a)/n$.
- Evaluate the function f at these positions, $x = (a, a+\Delta x, a+2\Delta x, ..., b-\Delta x)$.
- The integral can be approximated by the sum of these values multiplied by the width.

This can be written as

$$I \approx \sum_{k=0}^{n-1} f(a+k\Delta x)\cdot\Delta x. \tag{1.2.3}$$

With just a few rectangles, we have managed to take a problem that was not solve-able using traditional methods and 'brute forced' it with numerical analysis. As the method is so simple, it is easy to perform even by hand; a similar method was used by Ancient Babylonian astronomers to track the motion of celestial bodies (Ossendrijver, 2016).

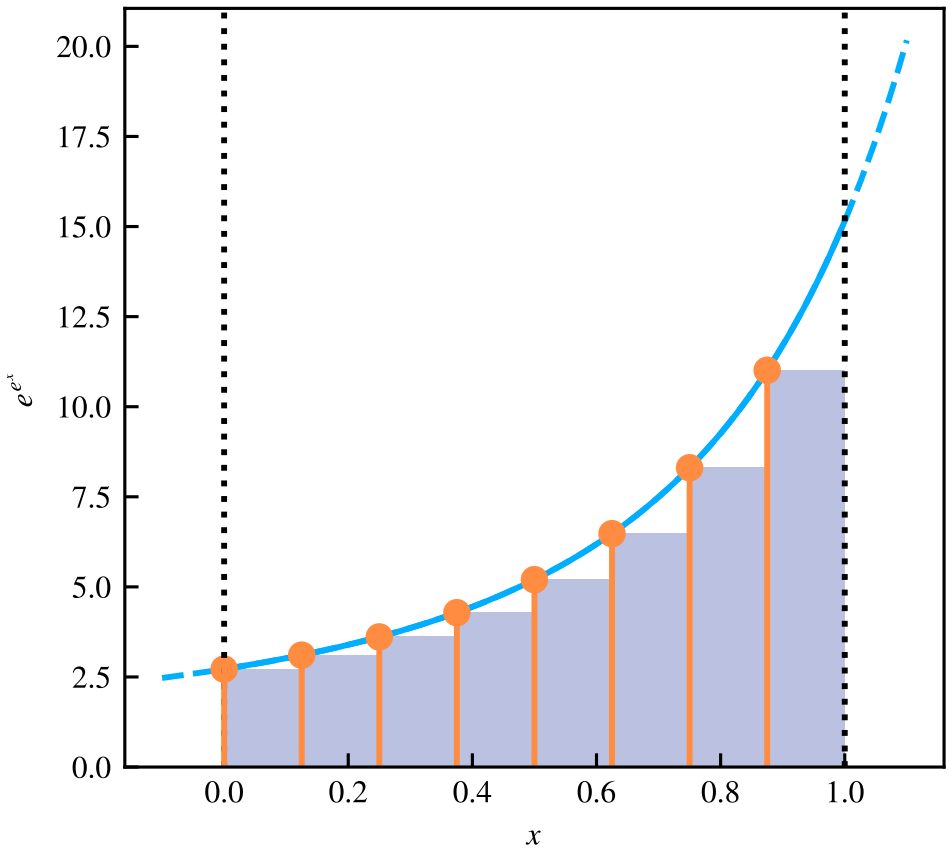


Figure 1.1: A demonstration of the left hand Riemann sum method for computing a numerical integral of Eqn. 1.2.1 on the domain $[0,1]$. The function is shown as a blue line, between the limits denoted by black dotted lines. The orange lines and points show where the function is evaluated; in this case, they lie along the left-hand edge of the purple rectangles making up the approximation to the integral.

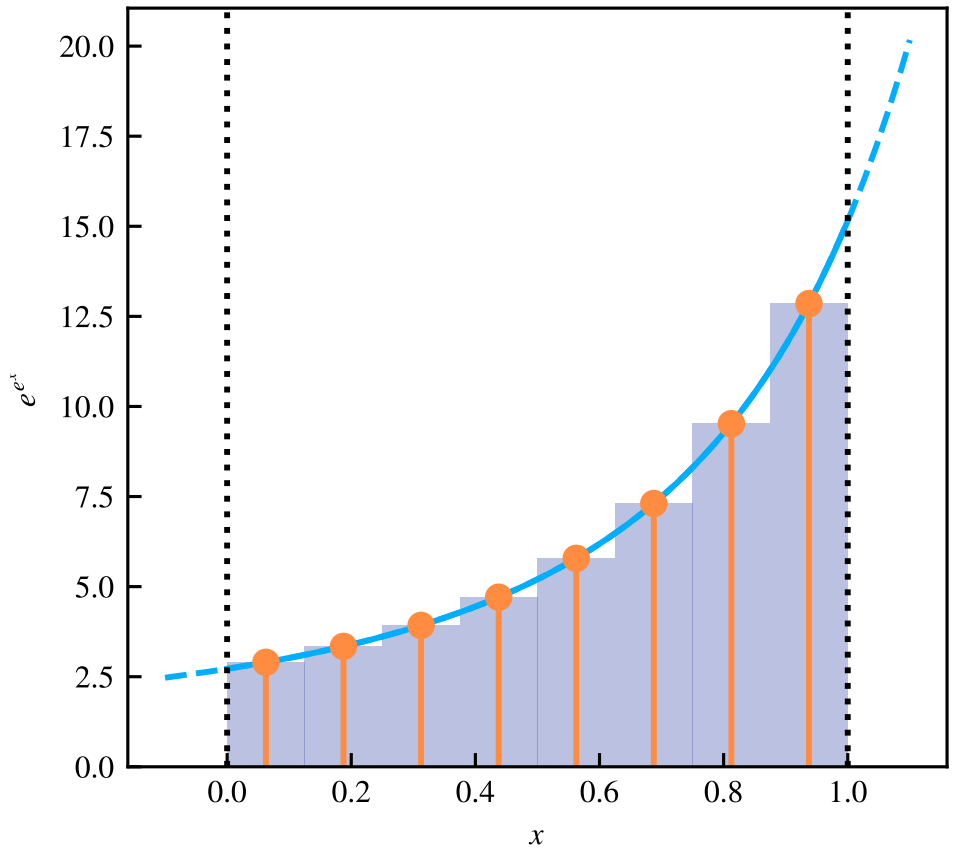


Figure 1.2: A demonstration of the midpoint rule method for computing a numerical integral of Eqn. 1.2.1 on the domain $[0,1]$, with the same colour scheme as Fig. 1.1. Note here how, due to the use of the midpoint, the rectangles show a marked improvement in their reconstruction of the function. From this figure it is also clear to see that this method would be able to capture the integral of a linear function exactly.

From Fig. 1.1, it is quite clear that there is an issue with our method; by choosing the left-hand edge of each rectangle, we have ensured that we will always underestimate the value of an integral for monotonically increasing functions, and overestimate it for monotonically decreasing functions (even if those functions are straight lines), and the same (albeit inverted) issue holds when we use the right hand edge.

A (seemingly) minor improvement to this basic method is to calculate the height of the rectangle in the middle of the interval, so:

$$I \approx \sum_{k=0}^{n-1} f\left[a + \left(k + \frac{1}{2}\right)\Delta x\right]\Delta x. \tag{1.2.4}$$

This method, shown in Fig. 1.2, is known as the Midpoint rule, and clearly shows some level of improvement over the method that uses the left-hand edge. To further investigate the level of improvement in the approximation of the integral, we can introduce the concept of convergence and error.

We measure the percentage error of the approximation of a numerical method $\tilde{I}$ relative to a reference

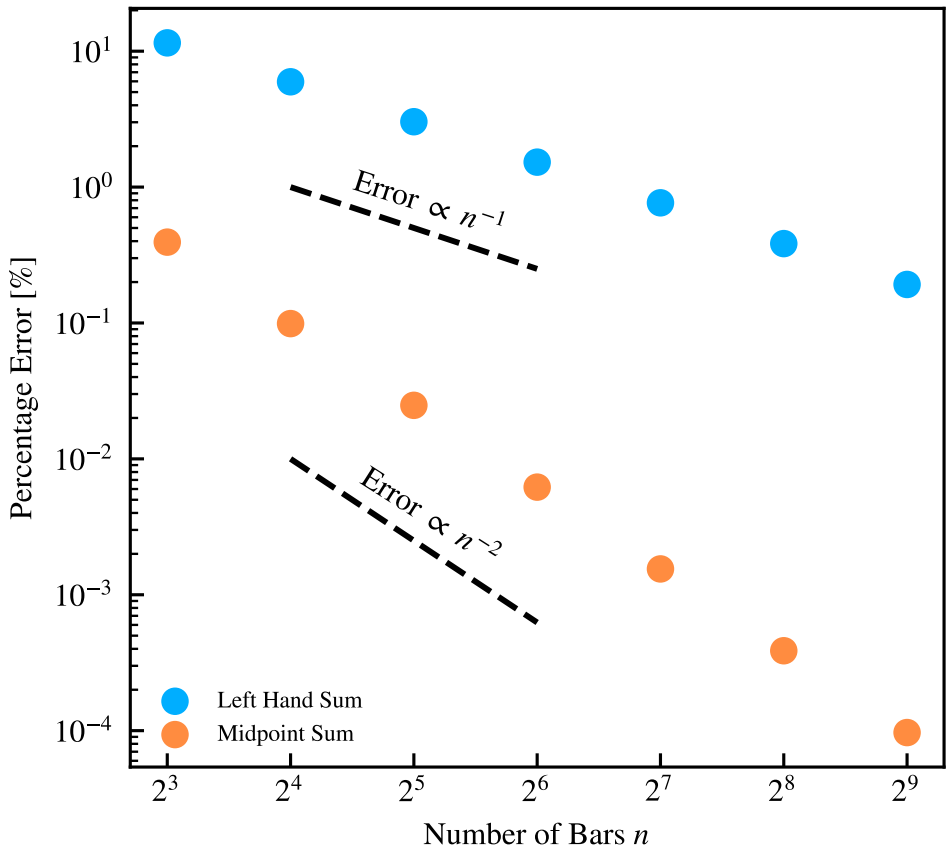


Figure 1.3: A comparison of the accuracy of the left hand Riemann sum method (blue; Eqn. 1.2.3) and the midpoint rule (orange; Eqn. 1.2.4) as a function of the number of intervals (or rectangular bars) n used. This demonstrates the faster convergence achieved with the midpoint method, with the convergence rates shown as black dashed lines.

value $\mathcal{I}$ as

$$E_{\%} = 100 \cdot \frac{|\tilde{I} - \mathcal{I}|}{|\mathcal{I}|} \tag{1.2.5}$$

In Fig. 1.3, we use a highly accurate, adaptive, numerical integrator to provide the reference value[1]. In cases where some experimental or analytic value is known, it is typical to use those; here, as the integral does not have a closed form solution, we are forced to use another highly accurate numerical method.

If we look at how the percentage error scales as we increase the number of sub-divisions n, as is shown in Fig. 1.3, the first thing to note is the dramatically lower error (sub-percent versus 10s of percent) that the midpoint sum returns. In addition, we see a different convergence rate for the two methods; the left hand sum (blue) shows an error that decreases linearly with the number of bars used (known as first-order convergence), whilst the midpoint sum is able to reduce its error twice as quickly, with the error decreasing as the square of the number of bars (second-order convergence).

This problem demonstrates how a small change in a numerical method can produce both a dramatically different error structure and emergent behaviour[2]. Such complexity out of a simple method additionally motivates the study of the numerical tools themselves.

1.3 History of Numerical Methods

1.3.1 Pre-computer numerical analysis

In the current era the development of numerical methods and their application in algorithms and computer simulations are closely linked, but this was not always the case. The history of numerical methods has been traced as far back as 1800 BC, with the Babylonian clay tablet YBC 7289 using numerical analysis to approximate the value of the square root of two to six decimal digits (Fowler & Robson, 1998). Other historical artefacts from around a similar time, such as the Rhind papyrus (dated to 1650 BC) contain numerical tables and root-finding procedures (Clagett, 1989).

Numerical methods then developed over the coming millennia, with a notable case being the approximation of π by Archimedes in 250 BC. This differed from previous estimates, that used approximations based on measurements of a square and a circle, through the use of an iterative method. This method, described in Proposition 3 of Book VI of Euclid's Elements, allowed for an approximation

[1] We use `scipy.integrate.quad` here using the default parameter values.

[2] Of course, there are further more complex numerical integration techniques, that employ the use of local derivatives and adaptive bin-widths to integrate functions to a pre-specified level of accuracy.

of pi to within an accuracy of 0.02%. The discovery and approximation of logarithms through the use of geometric progressions by Bürgi and Napier in the 16th Century cemented the use of numerical methods in day-to-day calculations (Hobson, 1914).

From the development of finite difference methods by Taylor, Newton, Raphson, and others (with Newton's method first published in Caswell & Wallis, 1685), to the numerical solutions to ordinary differential equations with the Euler method, numerical methods were placed at the core of mathematics from the 17th Century onwards. These methods were usually designed alongside physical or astronomical need. For instance, Taylor series were first used to investigate movement of vibrating strings and astronomical refraction and the finite difference method developed by Newton being applied to numerically integrate orbits of comets (Book III, Lemma VI of Philosophiæ Naturalis Principia Mathematica).

At the turn of the 20th Century, numerical analysis became its own differentiated branch of mathematics, with the founding of the first institutes in Germany and Italy devoted to the study and development of numerical methods.

1.3.2 Courant and the CFL condition

For those interested in performing numerical simulations, with the governing equations usually given as partial differential equations (PDEs), a key moment in history is the development of the so-called Courant-Friedrichs-Levy (CFL) condition in 1928 (Courant et al., 1928). The CFL condition ensures the stability of the integration scheme, and as such is a necessary (but not sufficient, as we will see in Chapter 2) condition for accurate results with numerical schemes. There is a more detailed discussion of the CFL condition in §2.5.3.

To briefly introduce the concept of the CFL condition, let us consider that we wish to integrate the following partial differential equation (in one dimension),

$$\frac{\partial \rho}{\partial t} = -v\frac{\partial \rho}{\partial x}, \tag{1.3.1}$$

with ρ the gas density, t the time, x the cartesian spatial co-ordinate, and v the fluid velocity[3]. The CFL condition then states that the spatial and temporal discretisation (Δx and Δt respectively) must obey

$$C = v\frac{\Delta t}{\Delta x} \leq \mathrm{C_{CFL}} \tag{1.3.2}$$

[3] You may find an analogy to this specific choice of equation in Eqn. 1.5.1.

with C known as the Courant number. In most schemes the spatial extent is pre-determined, for instance if using a fixed grid (or in SPH, see later, the inter-particle spacing), and as such the pre-specified $C_{CFL} \leq 1$ is usually used for a time-step condition such that

$$\Delta t \leq C_{CFL} \frac{\Delta x}{v} \tag{1.3.3}$$

is used to specify the maximum allowed time discretisation for that element.

1.3.3 The First Numerical Analysis on Computers

With the first electro(mecha)nical computers came the first opportunity to begin larger-scale simulations. ENIAC (Electronic Numerical Integrator and Computer) was the first programmable, general purpose, electronic computer, and was almost immediately used by Von Neumann and others for simulating neutron diffusion using the Monte Carlo method[4]. The Monte Carlo method is uniquely suited to arithmetic applications on electronic computers due to its heavy use of (pseudo-)random numbers (Metropolis, 1987). The success of the project, part of the development of the Hydrogen bomb, cemented both the use of numerical methods and electronic computers in engineering and physics.

1.4 Numerical Methods in Astrophysics

Before mechanical and electrical computers became commonplace, computing the interaction of many bodies, such as their motion under their own gravity, either required an impossible number of calculations, or the use of innovative techniques. For instance, to calculate the exact gravitational force on each of n self-interacting particles (known as the n-body problem), given by Newton's equation,

$$\mathbf{F} = -G\frac{m_a m_b}{r^3}\mathbf{r} \tag{1.4.1}$$

(with G Newton's gravitational constant, m_a and m_b the masses of two interacting bodies, and $\mathbf{r}$ the cartesian vector between them) requires a contribution from every other particle in the volume. This means that $n(n-1)$ contributions (ignoring symmetry) are required, which quickly becomes infeasible to perform by hand, or even with a modern computer for a modest number of particles.

In Holmberg (1941), the author finds an analogy to the gravitational force between bodies in the inverse square law of light, where the intensity of light drops off with the distance squared, the same

[4]The term *Monte Carlo* was also coined at the same time by Ulam, whose uncle would frequently gamble at the casinos in Monte Carlo, Monaco.

as gravity. Holmberg set up two 'nebulae' made up of 37 lightbulbs each (presumably, in a dark room) and used a light meter to integrate each of the bulbs forward in time, using the light intensity at the position of each bulb as a proxy for the gravitational field. This is widely accepted as the first attempt at an n-body simulation, now used ubiquitously throughout extragalactic astronomy.

Modern-day approximate numerical methods for solving the n-body problem typically only require a number of interactions proportional to $n \log n$ (e.g. the Barnes-Hut tree method Barnes & Hut, 1986) or even just n (the Fast Multipole Method Greengard & Rokhlin, 1987; Dehnen, 2014), rather than the n^2 required for direct summation of forces. These methods achieve such high efficiency by grouping together mass at large distances, making use of the $1/r^2$ dependence in the gravitational force.

1.5 Numerical Hydrodynamics in Astrophysics

Attempts to study fluid flows numerically began with the work of Lewis Fry Richardson, who, like many of his time, was simultaneously a mathematician, physicist, meteorologist, and even psychologist. Weather Prediction by Numerical Process (originally 1922, latest edition Richardson, 2007) is one of the first Computational Fluid Dynamics (CFD) texts, and describes what would now be termed a halo cell-based data parallelism approach to solving differential equations on a large grid - but the individual cells are represented by real people performing the numerical calculations, rather than machines.

All modern CFD focuses on solving the Navier-Stokes equations in either their original or simplified forms. The Navier-Stokes equations are famously difficult to solve, even numerically; whether or not smooth solutions always exist in three dimensions is still unknown. In astrophysics it is common to solve the compressible Euler equations that approximate the Navier-Stokes equations by assuming an adiabatic and inviscid flow. The Euler equations are:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{1.5.1}$$

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + \nabla P = \mathbf{0} \tag{1.5.2}$$

$$\frac{\partial u}{\partial t} + [(u + P)\mathbf{v}] = 0 \tag{1.5.3}$$

where ρ is the fluid density, $\mathbf{v}$ is the cartesian velocity, u is the specific internal energy, P is the pressure, and $\otimes$ denotes the outer product. These three equations, known as the continuity, momentum, and energy equations, form the basis of the vast majority of modern numerical methods for integrating astrophysical fluids. As we will see later, the underlying assumption that the fluids are inviscid and adiabatic will need to be challenged (in Chapter 3).

There have been many approaches to solving the equations of motion for a collisional fluid in a cosmological context over the years, from simple first order fixed-grid (Cen, 1992) to high-order discontinuous Galerkin schemes solved in an adaptive environment (Guillet et al., 2019). With the unpredictable and highly dynamic nature of astrophysical fluids, though, the concept of a meshless method is inviting. These methods use particle tracers of the field that are free to move, and the fluid equations are solved through interactions between particles.

The first meshless method for solving the fluid equations was proposed independently by Gingold & Monaghan (1977) and Lucy (1977) and is known as Smoothed Particle Hydrodynamics (SPH). This method was first applied to the gas dynamics within stars, but rapidly became used throughout the literature in modified forms (see e.g. Benz (1988) for a review). In SPH, fluid elements (particles) carry momentum, energy, and a mass that is fixed throughout the simulation, with density contrasts modelled through their clustering. Chapter 2 gives a more detailed overview of the fundamentals of SPH, with Chapter 3 investigating the governing equations in detail.

SPH is notably a Lagrangian method, with the governing equations in Chapter 3 derived explicitly from a Lagrangian. In a Lagrangian scheme, the movement of the fluid elements (particles of a fixed mass) is explicitly tied to the fluid flow, with two key benefits: advection of even complex fluid structures can be integrated exactly without excess diffusivity, and the scheme is naturally adaptive with higher density regions resolved with more particles.

1.6 Cosmological Galaxy Formation Simulations

Cosmological galaxy formation simulations take a (often representative) volume of universe and track the evolution of various matter components over cosmic time, usually from redshift $z \approx 100$ to $z = 0$ (from just after the big bang to the present day). They aim to reproduce the observed abundances and properties of galaxies in the real Universe to better understand the physics that governs galaxy evolution. The earliest simulations included only dark matter acting under gravity (see e.g. Frenk et al., 1988; Springel et al., 2005b), which remains an important approach to this day because such simulations are computationally efficient and can model very large volumes required for, e.g., dark energy studies (Knabenhans et al., 2021). These ‘dark matter only’ simulations, naturally, only track the evolution of the dark component of the universe and as such do not track the evolution and formation of galaxies, only dark matter haloes. To understand the formation of galaxies by leveraging the formation history of the dark matter component, Semi-Analytic Models (SAMs Frenk et al., 1990; Kauffmann, 1996; Somerville & Primack, 1998) were developed to populate haloes with galaxies

(see e.g. Porter et al., 2014; Henriques et al., 2015; Somerville et al., 2015; Lacey et al., 2016; Croton et al., 2016, for modern examples of SAM frameworks).

Over the past three decades, the inclusion of hydrodynamics in (cosmological) galaxy formation simulations has become commonplace (Hernquist & Katz, 1989; Evrard et al., 1994; Teyssier, 2002; Springel & Hernquist, 2002; Springel, 2005; Dolag et al., 2009), with SPH a commonly chosen technique for evolving cosmological fluids. SPH is prized for its adaptivity, conservation properties, and stability and is still used in state-of-the-art simulations by many groups today (Schaye et al., 2015; Teklu et al., 2015; McCarthy et al., 2017; Tremmel et al., 2017; Cui et al., 2019; Steinwandel et al., 2020); see Vogelsberger et al. (2020) for a recent overview of cosmological simulations.

Cosmological simulations typically include so-called 'sub-grid' physics that aims to represent underlying physics that is below the (mass) resolution limit (which is usually around 10^{3-7} $M_{\odot}$; Schaye et al., 2010; Vogelsberger et al., 2014; Schaye et al., 2015; Hopkins et al., 2018; Marinacci et al., 2019; Davé et al., 2019). This is commonplace in many fields, and is essential in galaxy formation to reproduce many of the observed properties of galaxies. One key piece of sub-grid physics is star formation, which occurs on mass scales smaller than a solar mass. Cold, dense, gas is required to enable stars to form; to reach these temperatures and densities radiative cooling (which occurs on atomic scales) must be included in a sub-grid fashion. Finally, when these stars have reached the end of their life some will produce supernovae explosions, which are modelled using sub-grid 'feedback' schemes (such a sub-grid scheme is chosen for many reasons, including but not limited to limited resolution and the 'overcooling problem'; see Navarro & White, 1993; Dalla Vecchia & Schaye, 2012, and references for more information).

Modern galaxy formation simulations typically include radiative cooling, chemical enrichment, star formation, stellar feedback, and AGN feedback. Using these models it is now possible to reproduce many of the key observed properties of galaxies at a range of cosmic epochs (e.g. Revaz & Jablonka, 2012; Vogelsberger et al., 2014; Schaye et al., 2015; Hopkins et al., 2018).

1.7 This Thesis

This thesis focuses on the understanding (Chapter 2) and development (Chapters 3 and 4) of numerical methods for cosmological galaxy formation simulations. In Chapter 2 we investigate the Pressure-Smoothed Particle Hydrodynamics (P-SPH) technique for numerical, finite mass, hydrodynamics and its coupling to subgrid models for galaxy formation. In Chapter 3 we develop a novel numerical hydrodynamics scheme based on Density-Smoothed Particle Hydrodynamics (also called

Traditional Smoothed Particle Hydrodynamics, or T-SPH) that couples well to subgrid models for use in galaxy formation simulations, and test its performance on a number of idealised scenarios. In Chapter 4 we develop two novel numerical method for investigating the transport of matter throughout the SIMBA cosmological galaxy formation simulation suite. These methods track the histories of particles heuristically and allow us to gain insight into the origin of matter in various types of galaxies. Finally, in Chapter 5, we conclude and discuss the future of numerical methods in galaxy formation and specifically the future for our understanding of numerics in galaxy formation suites.

Chapter 2

Inconsistencies arising from the coupling of galaxy formation sub-grid models to Pressure-Smoothed Particle Hydrodynamics

2.1 Introduction

As a consequence of the non-diffusive nature of the SPH equations, dissipative shock-capturing terms must be added, similar to other shock capturing schemes required in all numerical methods. In SPH this is resolved by the addition of a diffusive 'artificial viscosity' term (Monaghan & Gingold, 1983). This added diffusivity is only required in shocks, and so many schemes include particle-carried switches for the viscosity (Morris & Monaghan, 1997; Cullen & Dehnen, 2010) to prevent unnecessary conversion between kinetic and thermal energy in e.g. shearing flows and preserve greater than first order accuracy in smooth parts of the flow. Another consequence of the non-diffusive equations is the artificial surface tension appearing in contact discontinuities (e.g. Agertz et al., 2007), which has led to the development of several mitigation procedures. One possible solution is artificial conductivity (also known as energy diffusion) to smooth out the discontinuity (e.g. Price, 2008; Read & Hayfield, 2012; Rosswog, 2020a); this method applies an extra diffusion term in the energy equation to transfer energy between particles. The alternative solution, generally favoured in the cosmology community, is to reconstruct a smooth pressure field (Ritchie & Thomas, 2001; Saitoh & Makino, 2013; Hopkins, 2013). This smooth pressure field allows for a gradual transition pressure between

hot and cold fluids, suppressing any variation in the thermodynamic variable at scales smaller than the resolution limit. This can be beneficial in fluids where there is a high degree of mixing between phases, such as in gas flowing into galactic haloes (e.g. Tumlinson et al., 2017; Stern et al., 2019).

As noted in the introduction, the inclusion of sub-grid physics is now commonplace. Each of these processes has an impact on the hydrodynamics solver which must be carefully examined. Here we employ a simple galaxy formation model including implicit cooling and energetic feedback, based on the EAGLE galaxy formation model (Schaye et al., 2015), to understand how the inclusion of such a model may affect simulations employing Density- or Pressure-based SPH differently. We note, however, that the results obtained in the following sections are applicable to all kinds of galaxy formation models, including those that instead use instantaneous or 'operator-split' cooling.

The rest of this chapter is organised as follows: In §2.2 the SPH method is described, along with the Density- and Pressure-based schemes; in §2.3 the basics of a galaxy formation model are discussed in more detail; in §2.4 issues relating to injection of energy into Pressure-based schemes are explored; in §2.5 the SPH equations of motion are discussed; in §2.6 the time-integration schemes used in cosmological simulations are presented and issues with sub-grid cooling are explored, and in §4.6 it is concluded that while Pressure-SPH schemes can introduce significant errors it is possible in some cases to use measures (albeit computationally expensive ones) to remedy them. Because of this added expense it is suggested that a Density-based scheme is preferred, with an energy diffusion term used to mediate contact discontinuities.

2.2 Smoothed Particle Hydrodynamics

SPH is a Lagrangian method that uses particles to discretise the fluid. To find the equation of motion for the system, and hence integrate a fluid in time, the forces acting on each particle are required. In a fluid, these forces are determined by the local pressure field acting on the particles. The ultimate goal of the SPH method, then, is to find the pressure gradient associated with a set of discretised particles; once this is obtained finding the equations of motion is a relatively simple task. The reader is referred to the first few pages of the review by Price (2012) for more information on the fundamentals of the SPH method.

Before continuing, it is important to separate the two types of quantities present in SPH. The first, *particle carried properties* (denoted as symbols with an index corresponding to their particle, e.g. m_i is the mass of particle i), are valid only at the positions of particles in the system and include variables such as mass. The second, *field properties* (denoted as symbols with a hat, and with a corresponding

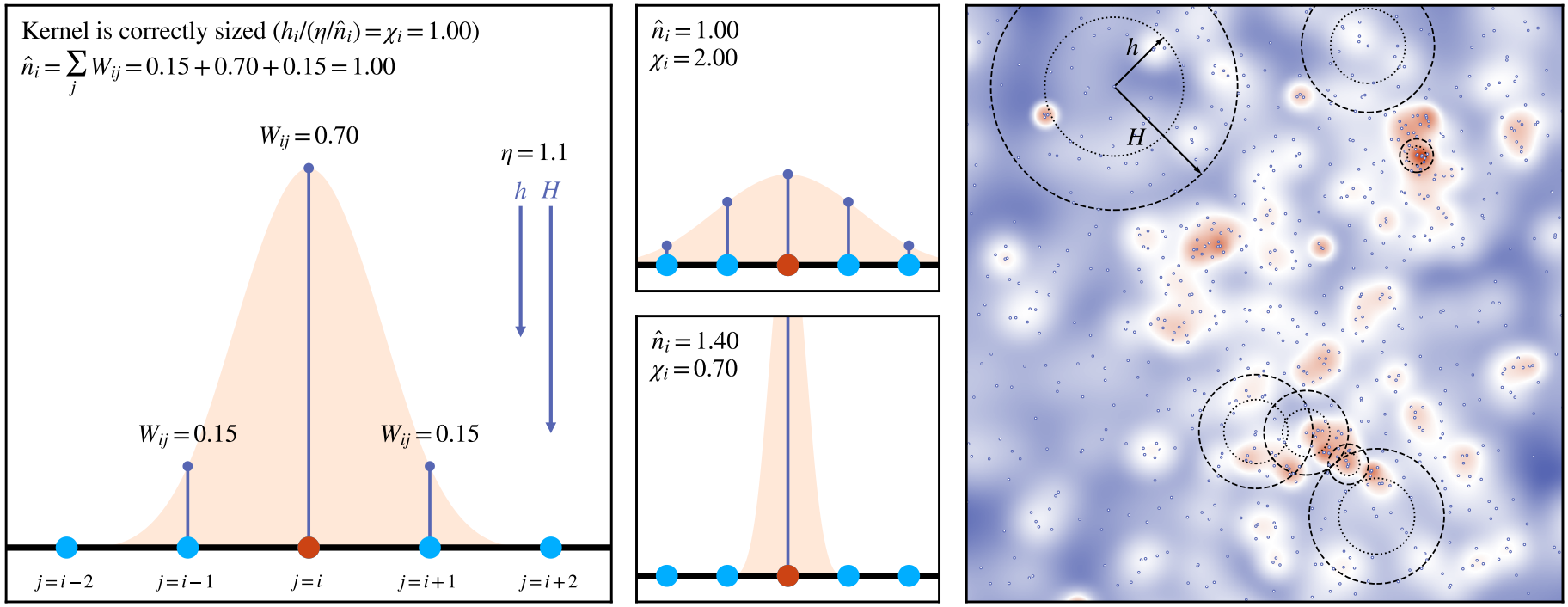


Figure 2.1: The three leftmost panels show the consequences of choosing a correct (large, left), too large (top right), and too small (bottom right) smoothing length (for $\eta = 1.1$) in 1D on a set of particles with an expected density $\hat{n} = 1$. This is quantified through both the density, $\hat{n}$, for the central particle i, and the ratio between the chosen smoothing length h_i and the expected smoothing length given by $\eta/\hat{n}_i$, parametrised as χ_i. χ_i is a well behaved function of the smoothing length, and finding the root of $\chi_i - 1$ is a reliable way to choose the value of h_i that corresponds to a given choice of η. Note how the density is only erroneous in the case with a smoothing length that is too small (bottom panel); the larger smoothing length (top panel) produces the correct density but would be less computationally efficient and inconsistent with the chosen value of η. The rightmost panel shows a 2D case with a random particle distribution, with the background colour map showing the low (blue) to high (white and then red) density regions and the associated variation in smoothing length. Here, for selected particles, the smoothing length h and kernel cut-off radius H are shown with dotted and dashed lines respectively. In particular, note how the higher density regions show smaller smoothing lengths such that Equation 2.2.1 is respected.

index if they are evaluated at particle positions, such as $\hat{\rho}_i$, the density at the position of particle i), are valid at all points in the computational domain, and generally are volumetric quantities. These field properties are built out of particle-carried properties by convolving them with the smoothing kernel.

The smoothing kernel is a weighting function of two parameters, inter-particle separation ($|\mathbf{r}_i - \mathbf{r}_j| = r_{ij}$) and smoothing length h_i, with a shape similar to a Gaussian with a full-width half maximum of $\sqrt{2\ln 2}h_i$. Smoothing kernels are also generally chosen to have 'compact support', with the function cut off at some distance $H_i = \gamma_k h_i$, where γ_k is a kernel-dependent quantity larger than one. The smoothing lengths of particles are chosen such that, for each particle, the following equation is satisfied:

$$\hat{n}_i = \sum_{\text{All particles } j} W(r_{ij}, h_i) = \left(\frac{\eta}{h_i}\right)^{n_D}, \tag{2.2.1}$$

where $\hat{n}_i$ is the local number density, n_D the number of spatial dimensions, and the kernel $W(r_{ij}, h_i)$ (henceforth written as W_{ij}) has the same dimensions as number density, typically being composed of a dimensionless weighting function $w_{ij} = w(r_{ij}/h_i)$ such that $W_{ij} \propto w_{ij} h_i^{-n_D}$. η is a dimensionless parameter that determines how smooth the field reconstruction should be (effectively setting the spatial resolution), with larger values leading to kernels that encompass more particles and typically takes values around $\eta \approx 1.2$[1]. An important distinction is the difference between the smoothing length, h_i, related to the full-width half-maximum (FWHM) of the Gaussian that the kernel approximates, and the kernel cut-off radius H_i. This cut-off radius is parametrised as $H_i = \gamma_K h_i$, with γ_K a kernel-dependent quantity taking values around $1.5 - 2.5$, such that H_i gives the maximum value of r_{ij} at which the kernel will be non-zero[2].

An example kernel (the cubic spline kernel, see Dehnen & Aly, 2012, for significantly more information on kernels) is shown in Fig. 2.1, with three choices for the smoothing length that satisfy Equation 2.2.1: one that is too large; one that is 'just right' for the given choice of η, and one that is too small. The choice to satisfy both equations is not strictly equivalent to ensuring that the kernel encompasses a fixed number of neighbouring particles; note how the edges of the kernel in the left panel do not coincide with a particle, even despite their uniform spacing.

To evaluate the mass density of the system, at the particle positions, the kernel is again used to re-evaluate the above equation now including the particle masses such that the density

$$\hat{\rho}_i = \sum_j m_j W_{ij} \tag{2.2.2}$$

[1] This corresponds to the popular choice of around 48 neighbours for a cubic spline kernel.

[2] The choice of which variable to store, h or H, is tricky; h is more easily motivated (Dehnen & Aly, 2012) and independent of the choice of kernel, but H is much more practical in the code as outside this radius interactions do not need to be considered.

is the sum over the kernel contributions and neighbouring masses m_j that may differ between particles. Note that this summation includes the self-contribution from the particle i, $m_i W(0, h_i)$.

Typically in SPH, the particle-carried property of either internal energy u_i, or entropy A_i (per unit mass)[3] is chosen to encode the thermal properties of the particle. These are related to each other, and the particle-carried pressure, through the ideal gas equation of state

$$P_i = (\gamma - 1) u_i \hat{\rho}_i = A_i \hat{\rho}^\gamma, \tag{2.2.3}$$

with the ratio of specific heats $\gamma = C_P/C_V = 5/3$ for the fluids usually considered in cosmological hydrodynamics models.

Alternatively, it is possible to construct a smooth pressure field that is evaluated at the particle positions such that

$$\hat{P}_i = \sum_j (\gamma - 1) m_j u_j W_{ij} = \left(\sum_j m_j A_j^{1/\gamma} W_{ij} \right)^\gamma, \tag{2.2.4}$$

directly includes the particle-carried thermal quantities of the neighbours into the definition of the pressure.

The differences between SPH models that use the particle pressures evaluated through the equation of state and smoothed density (i.e. those that use Equations 2.2.2 and 2.2.3), known as Density SPH, and those that use the smooth pressures (i.e. those that use Equation 2.2.4), known as Pressure SPH, is the central topic of this chapter. The SPH scheme may be referred to by its choice of thermodynamic variable, internal energy or entropy, as Density-Energy (Density-Entropy) or Pressure-Energy (Pressure-Entropy).

SPH schemes are usually implemented as a fixed number of 'loops over neighbours' (often just called loops). For a basic scheme like the ones presented above, two loops are usually used. The first loop, frequently called the 'density' loop, goes over all neighbours j of all particles i to calculate their SPH density (Equation 2.2.2) or smooth pressure (Equation 2.2.4). The second loop, often called the 'force' loop, evaluates the equation of motion for each particle i through the use of the pre-calculated smoothed quantities of all neighbours j. Each loop is computationally expensive, and so schemes that require extra loops are generally unfavourable unless they provide a significant benefit. State-of-the-art schemes typically use three loops, inserting a 'gradient' loop between the 'density' and 'force' loops to calculate either improved gradient estimators (Rosswog, 2020a) or coefficients for artificial viscosity and diffusion schemes (Price, 2008; Cullen & Dehnen, 2010).

[3] Note that this quantity is not really the 'entropy', but rather the adiabat that corresponds to this choice of entropy, hence the choice of symbol A.

2.3 A simple galaxy formation model

The discussion that follows requires an understanding of two pieces of a galaxy formation model: energy injection into the fluid and energy removal from the fluid. These are used to model the processes of supernovae and AGN feedback, and radiative cooling respectively. The results presented here are not necessarily tied to the model used, and are applicable to a wide range of current galaxy formation models that use Pressure-based SPH schemes. Here we use a simplified version of the EAGLE galaxy formation model as an instructive example, as this used Pressure-Entropy SPH for its hydrodynamics model in Schaye et al. (2015) and associated works (of particular note is Schaller et al., 2015, that discusses the effects of the choice of numerical SPH scheme on galaxy properties).

2.3.1 Cooling

The following equation is solved implicitly for each particle separately:

$$u(t+\Delta t) = u(t) + \frac{\mathrm{d}u}{\mathrm{d}t}(t, \Delta t)\Delta t, \tag{2.3.1}$$

where $\mathrm{d}u/\mathrm{d}t$ being the integrated 'cooling rate' calculated from the underlying atomic processes (assuming the gas density does not change).

The resulting final internal energy is transformed into an average rate of change of internal energy as a function of time over the step,

$$\frac{\bar{\mathrm{d}u}}{\mathrm{d}t} = \frac{u(t+\Delta t) - u(t_i)}{\Delta t}. \tag{2.3.2}$$

This secondary process occurs to ensure that in cases where particles with multiple time-steps interact, an accurate intermediate value for the internal energy can be calculated. After the cooling rate is calculated, it is limited in some circumstances (see Schaye et al., 2015, for more detail) that are not relevant to the discussion here. This average 'cooling rate' is then applied as either an addition to the $\mathrm{d}u/\mathrm{d}t$ or $\mathrm{d}A/\mathrm{d}t$ from the hydrodynamics scheme for each particle depending on the variable that the scheme tracks.

The resulting cooling rate may be large enough that it leads to orders of magnitude change in the internal energies of particles, with the cooling curve not explicitly resolved when using only the CFL condition (see §2.5.3). This is the desired behaviour, as running a simulation where the cooling curves of all particles are resolved would not be computationally feasible.

2.3.2 Energy Injection Feedback

A common, simple, feedback model is implemented as instantaneously heating particles *by* a constant temperature jump. It is possible to implement different types of feedback with this method, all being represented with a separate change in temperature ΔT. For supernovae feedback, $\Delta T_{\mathrm{SNII}} = 10^{7.5}$ K, and for AGN $\Delta T_{\mathrm{AGN}} = 10^{8.5}$ K (in EAGLE). The change in temperature does not actually ensure that the particle has this temperature once the feedback has taken place, however; the amount of energy corresponding to heating a particle from 0 K to this temperature is added to the particle. This ensures that even in cases where the particle is hotter than the heating temperature energy is still injected.

To apply feedback to a given particle, this change in temperature must be converted to a change in internal energy. This is performed by using a linear relationship between temperature and internal energy to find the internal energy that corresponds to a temperature of ΔT, and adding this additional energy onto the internal energy of the particle.

2.4 Energy injection in Pressure-Entropy

In cosmology codes it is typical to use the particle-carried entropy as the thermodynamic variable rather than the internal energy. This custom originated because in many codes (of particular note here is GADGET; Springel, 2005) the choice of co-ordinates in a space co-moving with expansion due to dark energy is such that the entropy variable is cosmology-less, i.e. it is the same in physical and co-moving space. Entropy is also conserved under adiabatic expansion, meaning that fewer equations of motion are required. This makes it convenient from an implementation point of view to track entropy rather than internal energy. However, at the level of the equation of motion, it makes no difference, as this is essentially just a choice of co-ordinate system.

This naturally leads the Pressure-Entropy variant (i.e. as opposed to Pressure-Energy) of the Pressure-based schemes to be frequently chosen; here the main smoothed quantity is pressure, with entropy being the thermodynamic variable.

The Pressure-Entropy and Pressure-Energy scheme perform equally well on hydrodynamics tests (see Hopkins (2013) for a collection), but when coupling to sub-grid physics there are some key differences.

For an entropy-based scheme, energy injection naturally leads to a conversion between the requested energy input and an increase in entropy for the relevant particle. Considering a Density-Entropy

scheme to begin with (e.g. Springel & Hernquist, 2002), with only a smooth density $\hat{\rho}$,

$$P_i = (\gamma - 1) u_i \hat{\rho}_i, \tag{2.4.1}$$

with P the pressure from the equation of state, γ the ratio of specific heats, and u_i the particle energy per unit mass. In addition, the expression for the pressure as a function of the entropy A_i,

$$P_i = A_i \hat{\rho}^{\gamma}. \tag{2.4.2}$$

Given that these should give the same thermodynamic pressure, the *derived* pressure variable can be eliminated to give

$$u_i = \frac{A_i \hat{\rho}^{\gamma-1}}{\gamma - 1} \tag{2.4.3}$$

and as these variables are independent for a change in energy Δu the change in entropy can be written

$$\Delta A_i = (\gamma - 1) \frac{\Delta u_i}{\hat{\rho}^{\gamma-1}}. \tag{2.4.4}$$

For any energy based scheme (either Density-Energy or Pressure-Energy), it is possible to directly modify the internal energy per unit mass u of a particle, and this directly corresponds to the same change in total energy of the field. This is clearly also true here too for the Density-Entropy scheme. Then, the sum of all energies (converted from entropies in the Density-Entropy case) in the box will be the original value plus the injected energy, without the requirement for an extra loop over neighbours[4].

Now considering Pressure-Entropy, the smoothed pressure shown in Equation 2.2.4 at a particle depends on a smoothed entropy over all of its neighbours. To connect the internal energy and entropy of a particle again the equation of state can be used by introducing a new *derived* variable, the weighted density $\bar{\rho}$[5],

$$\hat{P}_i = (\gamma - 1) u_i \bar{\rho}_i = A_i \bar{\rho}_i^{\gamma} \tag{2.4.5}$$

These two equations can be rearranged to eliminate the *derived* weighted density $\bar{\rho}$ such that

$$A_i(u_i) = \hat{P}_i^{1-\gamma} (\gamma - 1)^{\gamma} u_i^{\gamma}, \tag{2.4.6}$$

[4]This is only true given that the values entering the smooth quantities, here the density, are not changed at the same time. In practice, the mass of particles in cosmological simulations either does not change or changes very slowly with time (due to sub-grid stellar enrichment models for instance).

[5] Another conceptual issue with pressure-based schemes is the decision over which density to use within sub-grid routines (e.g. for cooling rates that depend on density). See Oppenheimer et al. (2018) Appendix D for more information.

$$u_i(A_i) = \frac{A_i^{1/\gamma} \hat{P}_i^{1-1/\gamma}}{\gamma - 1}. \tag{2.4.7}$$

To inject energy into the *field* by explicitly heating a single particle i in any entropy-based scheme the key is to find ΔA_i for a given Δu_i. In a pressure-based scheme this is problematic, as (converting Equation 2.4.6 to a set of differences),

$$\Delta A = \hat{P}_i(A_i, A_j)^{1-\gamma}(\gamma - 1)^{\gamma}(u_i + \Delta u)^{\gamma} - A_i, \tag{2.4.8}$$

to find this difference requires conversion via the smoothed pressure which directly (and non-linearly) depends on the value of A_i. This also occurs for the particles that neighbour i, meaning that there will be a non-zero change in the energy u_j that they report. Hence, this means that simply solving a linear equation for $\Delta A(\Delta u)$ is not enough; whilst this may at first appear trivial for a single particle, the true change in energy of the whole field will not be Δu (as it was in Density-Entropy) because of the changing pressures of the neighbours, and hence the local energy density of the field that they report.

The problem of injecting the required quantity of energy instantaneously can be reduced to producing a valid unique solution for ΔA as a function of the requested Δu for the entire field, which in principle can be performed by solving a system of $N_{\rm neigh}$ (the number of neighbours of particle i) non-linear equations (see §2.4.1). Any errors in this injection procedure effectively enter as errors in the initial conditions of the problem, and hence are carried through to any solution point in the future. For clarity, we begin with a more practical iterative solution, before moving onto the computationally and conceptually complex full solution.

Given the conservative nature of the SPH equations of motion, any error in the injection of energy will be carried through to the end of the simulation and impact the physical interpretation of the results. For instance, injecting additional energy in numerical experiments like the tests presented in Balsara et al. (2004) for supernovae, Booth & Schaye (2009) for AGN, or even a Sedov (1959) blast wave problem would correspond to errors from the initial time-step of the simulation to the last. All problems where energy must be accurately injected will be negatively affected by the use of a scheme where this is not practically possible. The thought experiment that follows corresponds to the initial injection phase of such a blast wave event, where an error in the injected energy will lead to a change in the speed and pressure of the shock front following the solution presented by Taylor (1950) and Sedov (1959).

A simple algorithm for injecting energy Δu in this case would be as follows:

1. Calculate the total energy of all particles that neighbour the one that will have energy injected,

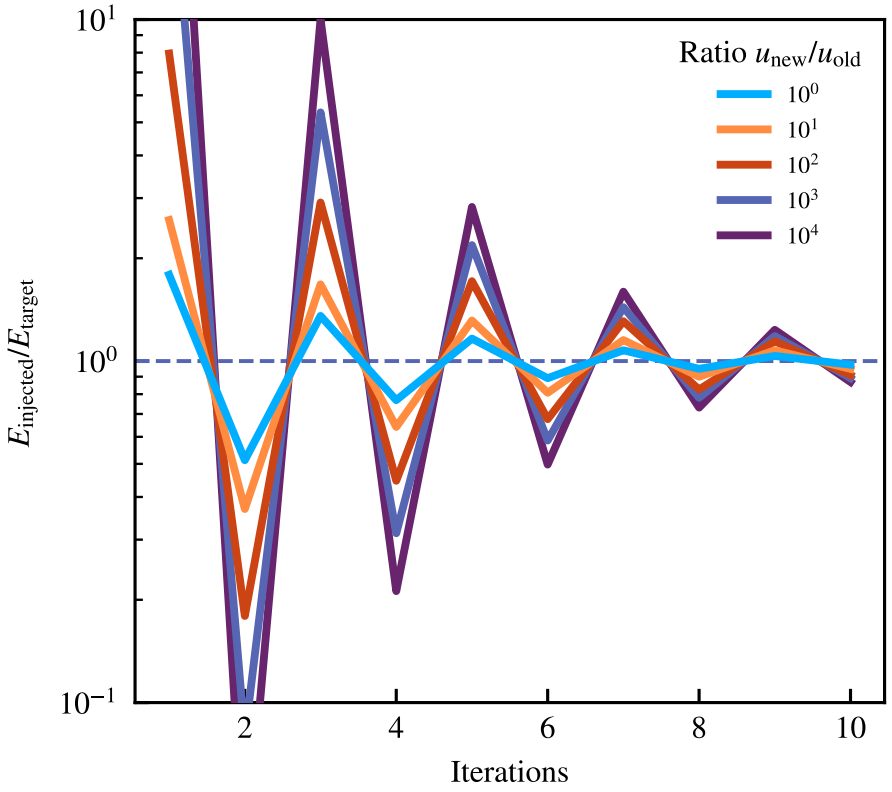


Figure 2.2: Energy injection as a function of iterations of the neighbour loop-based algorithm in Pressure-Entropy. Different coloured lines show ratios of injected energy to the original energy of the chosen particle, increasing in steps of 10. This algorithm allows for the correct energy to be injected into each particle after around 10 iterations, however more complex convergence criteria could be incorporated. A better estimate of the change in the smoothed pressure $\hat{P}$ could also significantly improve convergence.

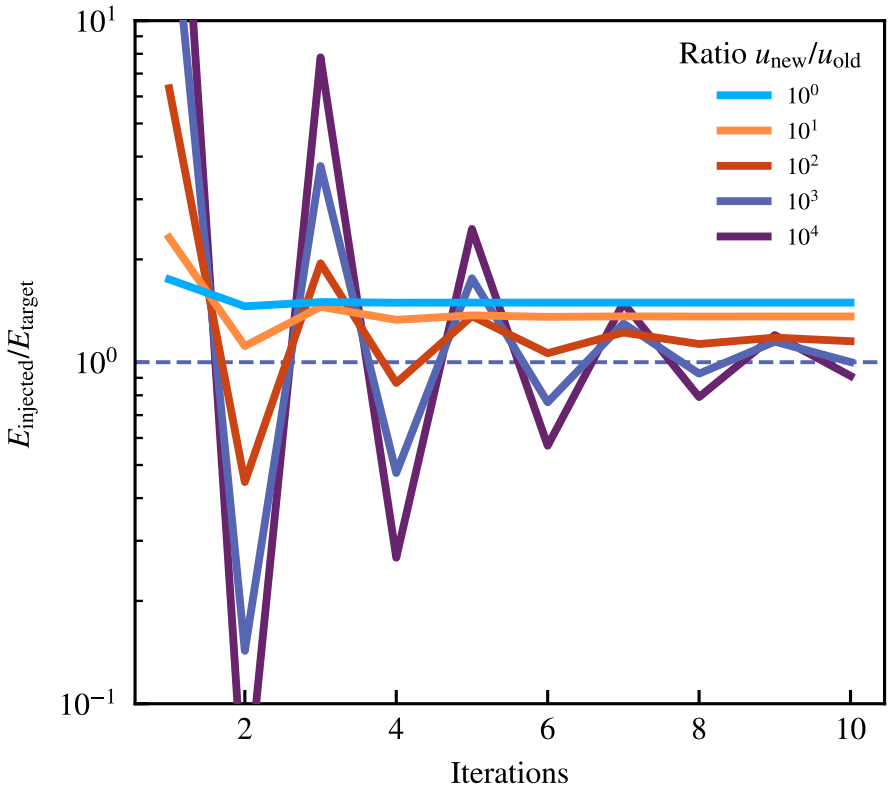


Figure 2.3: The same as Fig. 2.2, however this time using an approximate algorithm that only updates the self-contribution of the heated particle. This version of the algorithm shows non-convergent behaviour at low energy injection values, but is significantly computationally cheaper than solutions that require neighbour loops during the iteration procedure.

$u_{\mathrm{field},i} = \sum_j u(A_j, \hat{P}_j)$[6].

2. Find a target energy for the field, $u_{\mathrm{field},t} = u_{\mathrm{field},i} + \Delta u$.

3. While the energy of the field $u_{\mathrm{field}} = \sum_j u(A_j, \hat{P}_j)$ is outside of the bounds of the target energy:

   (a) Calculate $A_{\mathrm{inject}} = A(u_{\mathrm{field},t} - u_{\mathrm{field}}, \hat{P})$ for the particle that will have energy injected (i.e. apply Equation 2.4.8 assuming that $\hat{P}_i$ does not change).

   (b) Add on A_{inject} to the entropy of the chosen particle.

   (c) Re-calculate the smoothed pressures for all neighbouring particles.

   (d) Re-calculate the energy of the field u_{field} (i.e. go to item *iii* above).

The results of this process, for various injection energies, are shown in Fig. 2.2. After around 10 iterations, the requested injection of energy is reached. This process is valid only for working on a single particle at a time, however, and as such would be non-trivial to parallelise without the use of locks on particles that were currently being modified. Suddenly changing the energy of a neighbouring particle while this process was being performed would destroy the convergent behaviour that is demonstrated in Fig. 2.2.

Even without locks, this algorithm is computationally expensive, with many thousands of operations required to change a single variable. Re-calculating the smoothed pressure (step *c*) for every particle multiple times per step is generally infeasible as it would require many thousands of operations per particle per step. An ideal algorithm would not require neighbour loops; only updating the self contribution for the heated particle[7]:

1. Calculate the total energy of the particle that will have the energy injected, $u_{i,\mathrm{initial}} = u(A_i, \hat{P}_i)$.

2. Find a target energy for the particle, $u_{i,\mathrm{target}} = u_{i,\mathrm{initial}} + \Delta u$.

3. While the energy of the particle $u_i = u(A_i, \hat{P}_i)$ is outside of the bounds of the target energy (tolerance here is 10^{-6}, and is rarely reached) and the number of iterations is below the maximum (10):

[6] More specifically we actually require all particles j that see particle i as a neighbour (rather than all particles j that i sees as a neighbour), which may be different in regions where the smoothing length varies significantly over a kernel, but this detail is omitted from the main discussion for clarity.

[7] This algorithm was implemented in the original EAGLE code using the weighted density, $\bar{\rho}$ as the smoothed quantity, however this algorithm has been re-written to act on the smoothed pressure for simplicity. See Appendix A1.1 of Schaye et al. (2015) for more details.

(a) Calculate $A_{\rm inject} = A(u_{i,t} - u_i, \hat{P})$ for the particle that will have the energy injected.

(b) Add on $A_{\rm inject}$ to the entropy of that particle.

(c) Update the self contribution to the smoothed pressure for the injection particle by $\hat{P}_{i,\rm new} = \left[\hat{P}^{1/\gamma}_{i,\rm old} + (A^{1/\gamma}_{\rm new} - A^{1/\gamma}_{\rm old})W_0\right]^{\gamma}$ with $W_0 = W(0, h_i)$ the kernel self-contribution term.

(d) Re-calculate the energy of the particle $u_i = u(A_i, \hat{P}_i)$ using the new entropy and energy of that particle (i.e. go to *iii* above).

The implementation of the faster procedure is shown in Fig. 2.3. This simple algorithm leads to significantly higher than expected energy injection for low (relative) energy injection events. For the case of the requested energy injection being the same as the initial particle energy, over 50% too much energy is injected into the field. For events that inject more entropy into particle i, the value $A_i^{1/\gamma} W_{ij}$ for all neighbouring kernels becomes the leading component of the smoothed pressure field. This allows the pressure field to be dominated by this one particle, meaning that changes in $A_i^{1/\gamma}$ represent linear changes in the pressures of neighbouring particles, and hence allowing the simple methodology to correctly predict the changes in the global internal energy field.

The error in the computationally cheaper injection method is directly compared against the neighbour loop procedure from Fig. 2.2 in Fig. 2.4. The extra energy injected per event is clear here; the method using a full neighbour loop each iteration manages to reduce the error each iteration, with the non neighbour loop method showing a fixed offset after a few iterations. This also shows that the energy injection error grows as the amount injected grows, despite this becoming a lower relative fraction of the requested energy.

It is unclear exactly how much these errors impact the results of a full cosmological run. For the case of supernovae following Dalla Vecchia & Schaye (2012), which has a factor of $u_{\rm new}/u_{\rm old} \approx 10^4$ this should not represent a significant overinjection (the energy converges within 10 iterations to around a percent or so). For feedback pathways that inject a relatively smaller amount of energy (for instance SNIa, AGN events on particles that have been recently heated, events on particles in haloes with a high virial temperature, or schemes that inject using smaller steps of energy or into multiple particles simultaneously) there will be a significantly larger amount of energy injected than initially expected.

This uncontrolled energy injection is clearly undesirable, however as energy injection models are usually calibrated against an observational dataset, such errors may well be built into the eventual parameters of the model.

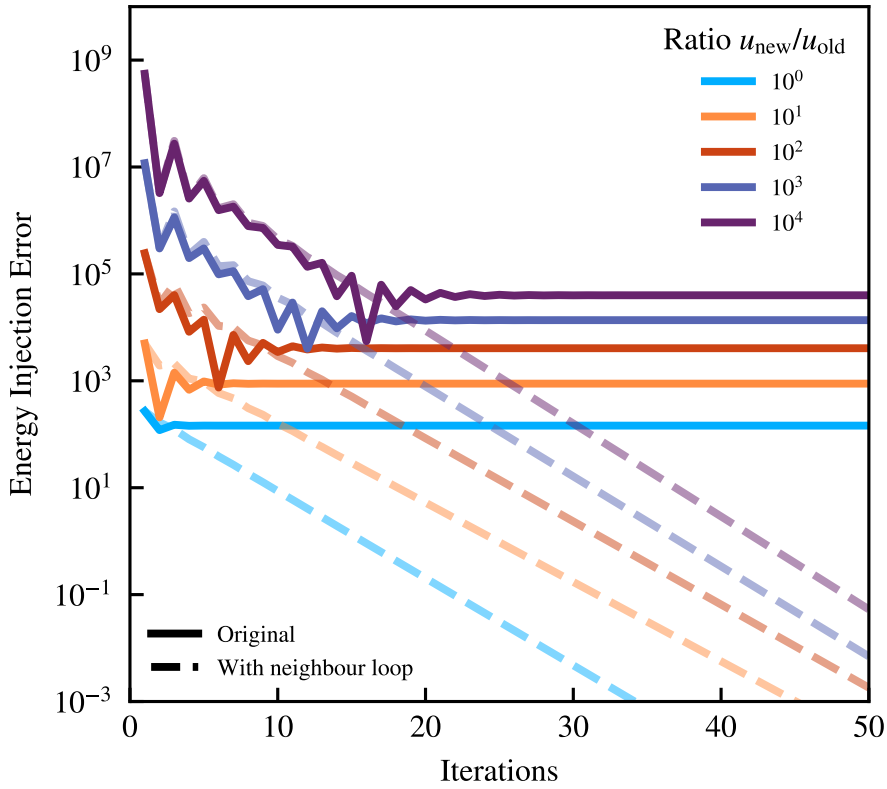


Figure 2.4: Comparison between the simple energy injection procedure (Fig. 2.3, solid lines) against the method including a neighbour loop each iteration (Fig. 2.2, dashed lines) for various energy injection values. The vertical axis here shows the energy offset from the true requested energy (in absolute arbitrary code units). The neighbour loop approach allows for the injected energy error to decrease with each iteration, where the simple procedure has a fixed (injection dependent) energy error that is reached rapidly at low values of energy injection where the entropies of neighbouring particles remain dominant.

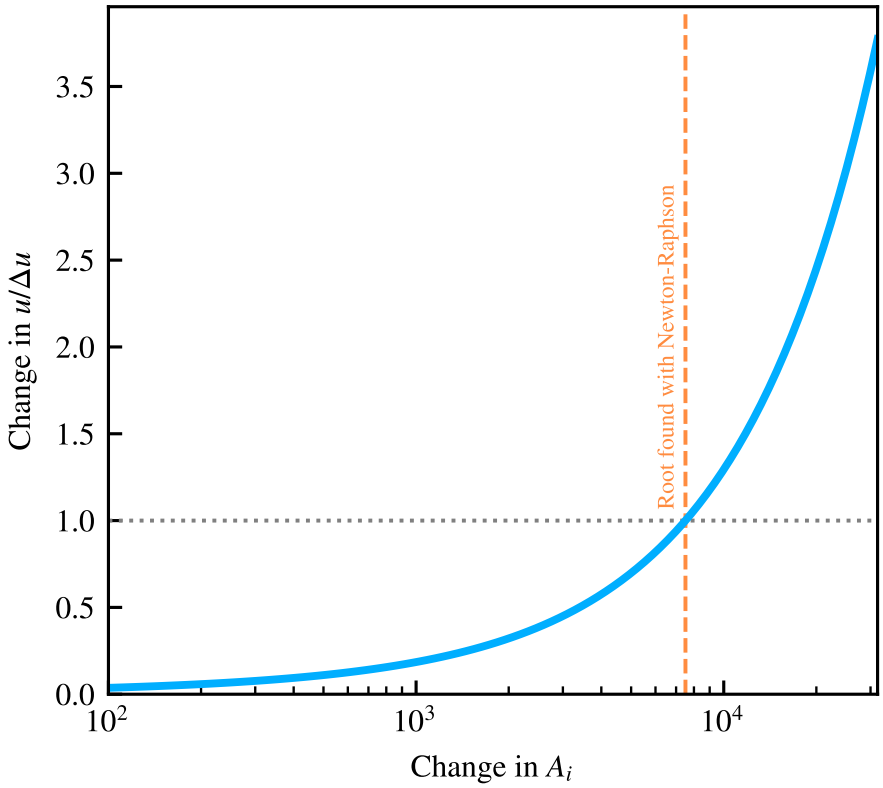


Figure 2.5: The blue line shows the dependence of change in field energy u as a function of the change in the entropy A_i of a single particle, for a requested change in energy Δu. This change in energy Δu corresponds to a heating event from $10^{3.5}$ K to $10^{7.5}$ K (a factor of 10^4 in u), which corresponds to a typical energetic supernovae feedback event. The orange dashed line shows the predicted change in A_i for this change Δu from the iterative solution (using the Newton-Raphson method) of Equation 2.4.12.

2.4.1 A Different Injection Procedure

Pressure-Entropy based schemes have been shown to be unable to inject the correct amount of energy using a simple algorithm based on updating only a single particle (i.e. without neighbour loops), however it is possible to perform this task exactly within a single step by using an iterative solver to find the change in entropy ΔA.

To inject a set amount of energy Δu the total energy of the field $u_{\rm tot}$ must be modified by changing the properties of particle i (with neighbouring particles j), with

$$u_{\rm tot} = \frac{1}{\gamma - 1}\left(\sum_j m_j A_j^{1/\gamma} W_{ij}\right)^{\gamma-1}. \tag{2.4.9}$$

This can be re-arranged to extract components specifically dependent on the injection particle i,

$$\begin{aligned} u_{\rm tot} = \frac{1}{1-\gamma}\sum_{j\neq i} A_j^{1/\gamma}\left(p_{j,i} + m_i A_i^{1/\gamma} W_{ij}\right)^{\gamma-1} \\ + A_i^{1/\gamma}\left(p_{i,i} + m_i A_i^{1/\gamma} W_{ii}\right)^{\gamma-1}, \end{aligned} \tag{2.4.10}$$

with

$$p_{a,b} = \hat{P}_b - m_b A_b^{1/\gamma} W_{ab}. \tag{2.4.11}$$

Finally, now considering a change in energy Δu as a function of the change in entropy for particle i, ΔA,

$$\begin{aligned} \Delta u = \frac{1}{1-\gamma}\sum_{j\neq i} A_j^{1/\gamma}\left(p_{j,i} + m_i (A_i + \Delta A)^{1/\gamma} W_{ij}\right)^{\gamma-1} \\ + (A_i + \Delta A)^{1/\gamma}\left(p_{i,i} + m_i (A_i + \Delta A)^{1/\gamma} W_{ii}\right)^{\gamma-1} \\ - u_{\rm tot}, \end{aligned} \tag{2.4.12}$$

which can be solved iteratively using, for example, the Newton-Raphson method. This method converges very well in just a few steps to calculate the change in entropy ΔA as demonstrated in Fig. 2.5. In practice, this method would require two loops over the neighbours of particle i per injection event. In the first loop, the values of $p_{j,i}$ and W_{ij} would be calculated and stored, with the iterative solver then used to find the appropriate value of ΔA. These changes would then need to be back-propagated to the neighbouring particles, as their smoothed pressures $\hat{P}_j$ will have changed significantly, reversing the procedure in Equation 2.4.11.

Such a scheme could potentially make a Pressure-Entropy based SPH method viable for a model that uses energy injection. This procedure requires tens of thousands of operations per thermal injection event, however, and as such would be impractical to implement efficiently.

This also highlights a possible issue with Pressure-Energy based SPH schemes, as even in this case, where it is much simpler to make changes to the global energy field, changes to the internal energy of a particle must be back-propagated to neighbours to ensure that the pressure and internal energy fields remain consistent. These errors also compound, should more than one particle in a kernel be heated without the back-propagation of changes.

2.5 Equations of Motion

So far only static fields have been under consideration; before moving on to discussing the effects of sub-grid cooling on pressure-based schemes, the *dynamics* part of SPH must be considered. Below only two equations of motion are described, the one corresponding to Density-Energy, and the equation of motion for Pressure-Energy SPH. For a more expanded derivation of the following from a Lagrangian and the first law of Thermodynamics see Hopkins (2013), or the Swift simulation code theory documentation[8].

2.5.1 Density-Energy

For Density-Energy the smoothed quantity of interest is the smoothed mass density (Equation 2.2.2). This leads to a corresponding equation of motion for velocity of

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_j m_j \left[f_i \frac{P_i}{\hat{\rho}_i^2} \nabla W(r_{ij}, h_i) + f_j \frac{P_j}{\hat{\rho}_j^2} \nabla W(r_{ji}, h_j) \right], \tag{2.5.1}$$

with the f_i here representing correction factors for interactions between particles with different smoothing lengths

$$f_i = \left(1 + \frac{h_i}{n_d \hat{\rho}_i} \frac{\partial \hat{\rho}_i}{\partial h_i} \right)^{-1}. \tag{2.5.2}$$

This factor also enters into the equation of motion for the internal energy

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_j m_j f_i \frac{P_i}{\hat{\rho}_i^2} \mathbf{v}_{ij} \cdot \nabla W(r_{ij}, h_i). \tag{2.5.3}$$

2.5.2 Pressure-Energy

For Pressure-Energy SPH, the thermodynamic quantity u remains the same as for Density-Energy, but the smoothed pressure field $\hat{P}$ is introduced (see Equation 2.2.4). This is then used in the equation

[8] `http://www.swiftsim.com`

of motion for the particle velocities

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_j (\gamma-1)^2 m_j u_j u_i \left[\frac{f_{ij}}{\hat{P}_i}\nabla W(r_{ij}, h_i) + \frac{f_{ji}}{\hat{P}_j}\nabla W(r_{ji}, h_j)\right]. \tag{2.5.4}$$

with the f_{ij} now depending on both particle i and j

$$f_{ij} = 1 - \left[\frac{h_i}{n_d(\gamma-1)\hat{n}_i m_j u_j}\frac{\partial \hat{P}_i}{\partial h_i}\right]\left(1 + \frac{h_i}{n_d \hat{n}_i}\frac{\partial \hat{n}_i}{\partial h_i}\right)^{-1}, \tag{2.5.5}$$

with $\hat{n}$ the local particle number density (Equation 2.2.1). Again, this factor enters into the equation of motion for the internal energy

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = (\gamma-1)^2 \sum_j m_j u_i u_j \frac{f_{ij}}{\hat{P}_i}\mathbf{v}_{ij}\cdot\nabla W_{ij}. \tag{2.5.6}$$

2.5.3 Choosing an Appropriate Time-Step

To integrate these forward in time, an appropriate time-step between the evaluation of these smoothed equations of motion must be chosen. SPH schemes typically use a modified version of the Courant–Friedrichs–L (CFL, Courant et al., 1928) condition to determine this step length. The CFL condition takes the form of

$$\Delta t = C_{\mathrm{CFL}}\frac{H_i}{c_\mathrm{s}}, \tag{2.5.7}$$

with c_s the local sound-speed, $H_i = \gamma_k h_i$ the kernel compact support radius, and C_{CFL} a constant that should be strictly less than 1.0, typically taking a value of 0.1-0.3[9]. Computing this sound-speed is a simple affair in density-based SPH, with it being a particle-carried property that is a function solely of other particle carried properties,

$$c_\mathrm{s} = \sqrt{\gamma\frac{P}{\hat{\rho}}} = \sqrt{\gamma(\gamma-1)u}. \tag{2.5.8}$$

For pressure-based schemes this requires a little more thought. The same sound-speed can be used, but this is not representative of the variables that actually enter the equation of motion. To clarify this, first consider the equation of motion for Density-Energy (Equation 2.5.1) and re-write it in terms of the sound-speed,

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} \sim \frac{c_{\mathrm{s},i}^2}{\hat{\rho}_i}\nabla_i W_{ij},$$

[9]In practice this c_s is usually replaced with a signal velocity v_{sig} that depends on the artificial viscosity parameters. As the implementation of an artificial viscosity is not discussed here, this detail is omitted for simplicity. The CFL condition given here is only hence valid for subsonic flows.

and for Pressure-Energy (Equation 2.5.4)

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} \sim (\gamma - 1)^2 \frac{u_i u_j}{\hat{P}_i} \nabla_i W_{ij}.$$

From this it is reasonable to assume that the sound-speed, i.e. the speed at which information propagates in the system through pressure waves, is given by the expression

$$c_{\mathrm{s}} = (\gamma - 1) u_i \sqrt{\gamma \frac{\hat{\rho}_i}{\hat{P}_i}}. \tag{2.5.9}$$

This expression is dimensionally consistent with a sound-speed, and includes the gas density information (through $\hat{\rho}$), traditionally used for sound-speeds, as well as including the extra information from the smoothed pressure $\hat{P}$. However, such a sound-speed leads to a considerably *higher* time-step in front of a shock wave (where the smoothed pressure is higher, but the smooth density is relatively constant), leading to time integration problems. Using

$$c_{\mathrm{s}} = \sqrt{\gamma \frac{\hat{P}_i}{\hat{\rho}_i}} \tag{2.5.10}$$

instead of Equation 2.5.9 leads to a sound-speed that does not represent the equation of motion as directly but does not lead to time-integration problems, and effectively represents a smoothed internal energy field. It is also possible to use the same sound-speed using the particle-carried internal energy directly above.

2.6 Time Integration

A typical astrophysics SPH code will use Leapfrog integration or a velocity-verlet scheme to integrate particles through time (see e.g. Hernquist & Katz, 1989; Springel, 2005; Borrow et al., 2018). This approach takes the accelerations, $\mathbf{a}_i = \mathrm{d}\mathbf{v}_i/\mathrm{d}t$, and the velocities, $\mathbf{v}_i = \mathrm{d}\mathbf{r}_i/\mathrm{d}t$ and solves the system for the positions $r_i(t)$ as a function of time. It is convenient to write the equations as follows (for each particle):

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}(t) + \frac{\Delta t}{2}\mathbf{a}(t), \tag{2.6.1}$$

$$\mathbf{r}\left(t + \Delta t\right) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t, \tag{2.6.2}$$

$$\mathbf{v}\left(t + \Delta t\right) = \mathbf{v}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2}\mathbf{a}(t + \Delta t), \tag{2.6.3}$$

commonly referred to (in order) as a Kick-Drift-Kick scheme. Importantly, these equations must be solved for all variables of interest.

This leapfrog time-integration is prized for its second order accuracy (in Δt) despite only including first order operators, due to cancelling second order terms as well as its manifest conservation of energy (Hernquist & Katz, 1989).

2.6.1 Multiple Time-Stepping

As noted above, it is possible to find a reasonable time-step to evolve a given hydrodynamical system with using the CFL condition (Equation 2.5.7). This condition applies on a particle-by-particle basis, meaning that to evolve the whole *system* a method for combining these individual time-steps into a global mechanism must be devised. In less adaptive problems than those considered here (e.g. those with little dynamic range in smoothing length), it is reasonable to find the minimal time-step over all particles, and evolve the whole system with this time-step. This scenario is frequently referred to as 'single-dt'.

For a cosmological simulation, however, the huge dynamic range in smoothing length (and hence time-step) amongst particles means that evolving the whole system with a single time-step would render most simulations infeasible (Borrow et al., 2018). Instead, each particle is evolved according to its own time-step (referred to as a multi-dt simulation) using a so-called 'time-step hierarchy' as originally described in Hernquist & Katz (1989). This choice is common-place in astrophysics codes (Teyssier, 2002; Springel, 2005).

In some steps in a multi-dt simulation only the particles on the very shortest time-steps are updated in a loop over their neighbours to re-calculate, for example, $\hat{\rho}$ (with these particles being called 'active'). The rest of the particles are referred to as being 'inactive'. As the inactive particles may interact with the active ones, their properties must be interpolated, or drifted, to the current time.

For particle-carried quantities, such as the internal energy u, a simple first-order equation is used,

$$u(t + \Delta t) = u + \frac{\mathrm{d}u}{\mathrm{d}t}\Delta t. \tag{2.6.4}$$

2.6.2 Drifting Smoothed Quantities

As a particle may experience many more drift steps than loops over neighbours (that are only performed for active particles), it is important to have drift operators ($\mathrm{d}\hat{x}/\mathrm{d}t$) for smoothed quantities $\hat{x}$ to interpolate their values between full time-steps. This is achieved through taking the time differential of smoothed quantities. Starting with the simplest, the smoothed number density,

$$\frac{\mathrm{d}\hat{n}_i}{\mathrm{d}t} = \sum_j \frac{\mathrm{d}W(r_{ij}, h_i)}{\mathrm{d}t},$$

$$= \sum_j \mathbf{v}_{ij} \cdot \nabla_j W(r_{ij}, h_i). \tag{2.6.5}$$

Following this process through for the smoothed quantities of interest yields

$$\frac{\mathrm{d}\hat{\rho}_i}{\mathrm{d}t} = \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_j W(r_{ij}, h_i), \tag{2.6.6}$$

$$\frac{\mathrm{d}\hat{P}_i}{\mathrm{d}t} = (\gamma - 1) \sum_j m_j \left(W_{ij} \frac{\mathrm{d}u_j}{\mathrm{d}t} + u_j \mathbf{v}_{ij} \cdot \nabla_j W_{ij} \right), \tag{2.6.7}$$

for the smoothed density and pressure respectively, with $W_{ij} = W(r_{ij}, h_i)$. In the smoothed density case, the pressure is re-calculated at each drift step from the now drifted internal energy and density using the equation of state[10].

The latter drift equation, due to its inclusion of $\mathrm{d}u_j/\mathrm{d}t$ (i.e. the rate of change of internal energy of all neighbours of particle i), presents several issues. This sum is difficult to compute in practice; it requires that all of the $\mathrm{d}u_j/\mathrm{d}t$ are set before a neighbour loop takes place. This would require an extra loop over neighbours after the 'force' loop, which has generally been considered computationally infeasible for a scheme that purports to be so cheap. In practice, the following is used to drift the smoothed pressure:

$$\frac{\mathrm{d}\hat{P}_i}{\mathrm{d}t} = \frac{\mathrm{d}\hat{\rho}_i}{\mathrm{d}t} u_i + \hat{\rho}_i \frac{\mathrm{d}u_i}{\mathrm{d}t} \,, \tag{2.6.8}$$

which clearly does not fully capture the expected behaviour of Equation 2.6.7 as it only includes the rate of change of the internal energy for particle i, discarding the contribution from neighbours.

Such behaviour becomes particularly problematic in cases where sub-grid cooling is used, where particles within a kernel may have both very large $\mathrm{d}u_j/\mathrm{d}t$ (where $(\mathrm{d}u_j/\mathrm{d}t)\Delta t$ is comparable to u_j), and $\mathrm{d}u_j/\mathrm{d}t$ that vary rapidly with time. Consider the case where an active particle cools rapidly from some temperature to the equilibrium temperature in one step (which occurs frequently in a typical cosmological simulation where no criterion on the time-step for $\mathrm{d}u/\mathrm{d}t$ is included to ensure the number of steps required to complete the calculation remains reasonable whilst employing implicit cooling). If this particle has a neighbour at the equilibrium temperature that is inactive, the pressure for the neighbouring particle will remain significantly (potentially orders of magnitude) higher than what is mandated by the local internal energy field, leading to force errors of a similar level.

To apply these drift operators to smoothed quantities, instead of using a linear drift as in Equation 2.6.4, the analytic solution to these first order differential equations is used. For a smooth quantity $\hat{x}$

[10] Note that the first equation for the smoothed density corresponds to the SPH discretisation of the continuity equation (Monaghan, 1992), but the second equation makes little physical sense.

it is drifted forwards in time using

$$\hat{x}(t + \Delta t) = \hat{x}(t) \cdot \exp\left(\frac{\Delta t}{\hat{x}} \frac{\mathrm{d}\hat{x}}{\mathrm{d}t}\right). \tag{2.6.9}$$

This also has the added benefit of preventing the smoothed quantities from becoming negative. For this to be accurate, it requires an accurate $\mathrm{d}\hat{x}/\mathrm{d}t$ term.

2.6.3 Impact of Drift Operators in multi-dt

Whilst the true drift operator for $\hat{P}$ appears to be impractical from a computational perspective due to the requirement of another loop over neighbours, at first glance it appears that the use of this correct drift operator would remedy the issues with cooling. Unfortunately, in a multi-dt simulation where active and in-active particles are mixed, this 'correct' operator can still lead to negative pressures when applied.

In Fig. 2.6 the different ways of drifting smooth pressure in a multi-dt simulation are explored. In this highly idealised test, a cubic volume of uniform 'cold' fluid is considered. A single particle at the center is set to have a 'hot' temperature of 100 times higher than the background fluid, and is set to have a cooling rate that ensures that it cools to the 'cold' temperature within its first time-step. This scenario is similar to a hot 10^6 K particle in the CGM cooling to join particles in the ISM at the 10^4 K equilibrium temperature. The difference between the time-step of the hot and cold particles, implied by Equation 2.5.7, is a factor of 10 (when using the original definition of sound-speed, see Equation 2.5.8). Here the cold particle is drifted ten times to interact with its hot neighbour over a single time-step of its own. In practice, this scenario would evolve slightly differently, with the previously hot particle having its time-step re-set to $\mathrm{d}t_{\mathrm{cool}}$ after it has cooled to the equilibrium temperature, but the nuances of the time-step hierarchy are ignored here for simplicity.

The three drifting scenarios proceed very differently. In Fig. 2.7 the fractional errors relative to the single-dt case are shown.

In the case of the drift using Equation 2.6.7, the pressure rapidly drops to zero. This is prevented from becoming negative thanks to the integration strategy that is employed (Equation 2.6.9); the rate of $\mathrm{d}\hat{P}/\mathrm{d}t$ is high enough to lead to negative pressures within a few drift steps should a simple linear integration strategy like that employed for the internal energy (Equation 2.6.4) be used. Because there is only a linear time integration (with a poorly chosen time-step for the equation to be evolved) method for a now non-linear problem (as there is a significant $\mathrm{d}^2u/\mathrm{d}t^2$ from changes in cooling rate) errors naturally manifest.

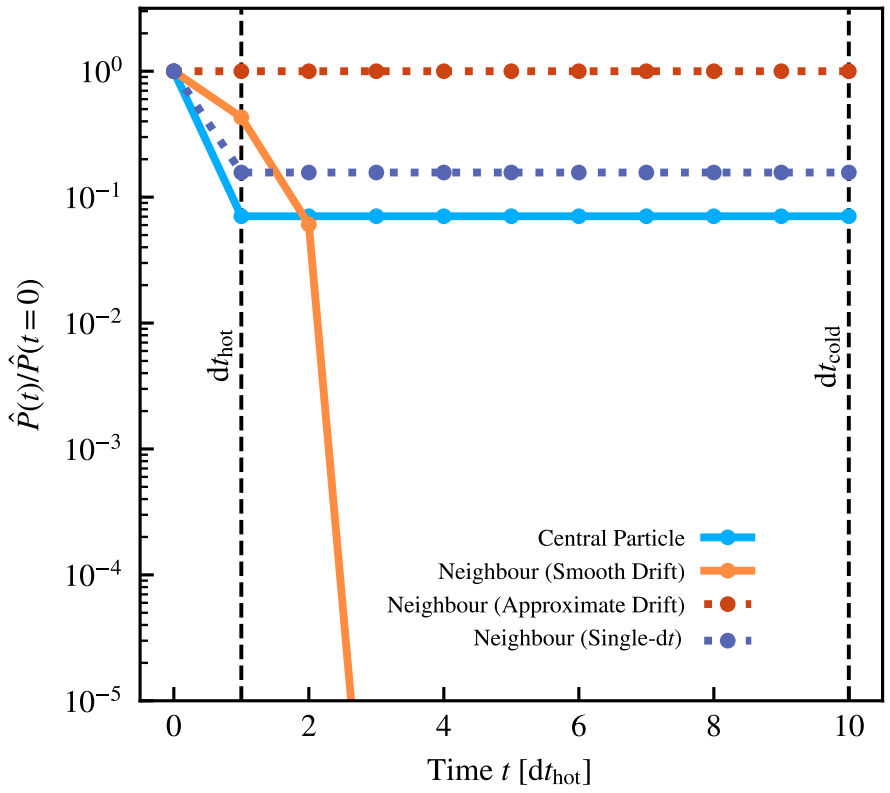


Figure 2.6: Smooth pressure as a function of time for different strategies in a uniform fluid of 'cold' particles, with one initially 'hot' particle with a temperature 100 times higher than the cold particles that cools to the 'cold' temperature in one time-step. The solid blue line shows the pressure of the central particle as a function of time (relative to its initial pressure). The dashed blue line shows the pressure of the closest 'hot' neighbour in a single-dt scenario, i.e. the whole system is evolved with time-step $\mathrm{d}t_{\mathrm{hot}}$. This shows the true answer for the pressure of the neighbour particle. The dotted red line shows the result of drifting the cold particle with Equation 2.6.8. As this particle has no cooling rate, and the fluid is stationary, the pressure does not change. The solid orange line shows the result of drifting using Equation 2.6.7. This rapidly leads to the particle having a pressure of zero, a highly undesirable result. Note that the orange line does not follow the dashed blue line in the first few steps due to different drifting schemes for smoothed and particle-carried quantities (Equation 2.6.4 and 2.6.9).

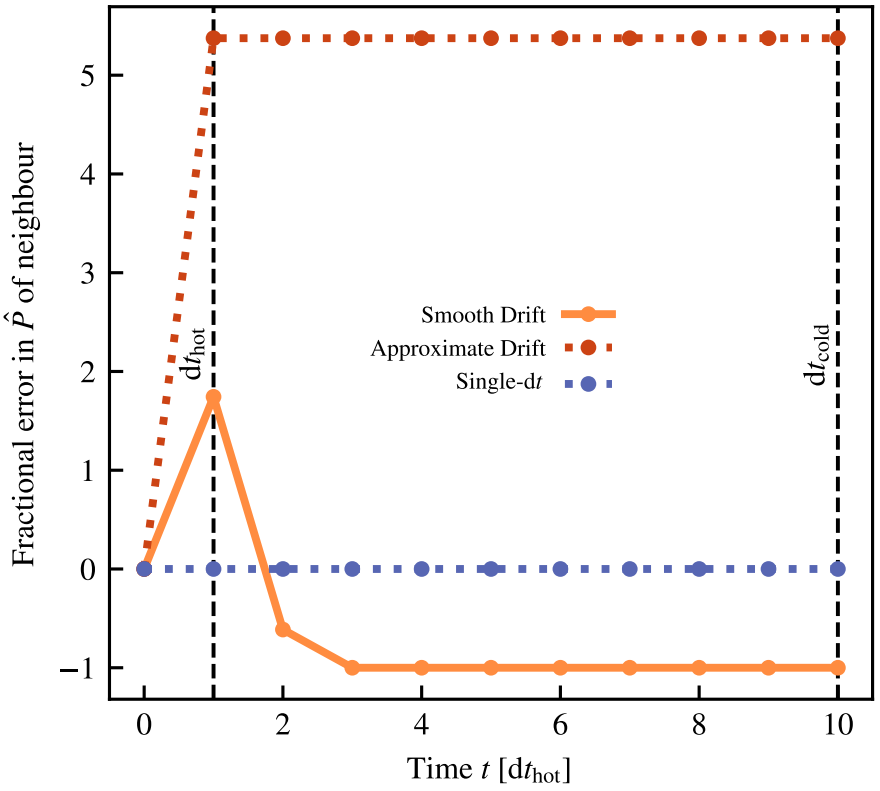


Figure 2.7: The same lines as Fig. 2.6, except now showing the 'error' as a function of time relative to the single-dt case (blue dashed line) of the pressure $\hat{P}$ of the nearest neighbour to the 'hot' particle. Here the fractional error is defined as $\hat{P}(t) - \hat{P}_{\rm single-dt}/\hat{P}_{\rm single-dt}$. The orange line showing the drifting using Equation 2.6.7 shows that the pressure rapidly drops to zero after around four steps. The red dotted line (Equation 2.6.8) shows the offset in pressure that is maintained even after the central 'hot' particle cools.

The drift operator using a combination of the local cooling rate and density time differential (Equation 2.6.8) is the safest, leading to pressures that are higher than expected; this does however come at the cost of larger relative errors in the pressure (500% increase v.s. 100% decrease; both of these are highly undesirable).

Limiting time-steps

One way to address the issues presented in Fig. 2.6 is to limit the time-steps between neighbouring particles. Such a 'time-step limiter' is common-place in galaxy formation simulations, as they are key to capturing the energy injected during feedback events (see e.g. Durier & Dalla Vecchia, 2012). In addition, the use of the 'smoothed' sound-speed (from Equation 2.5.10) ensures that the neighbouring particle has a time-step that is much closer to the time-step of the 'hot' particle than the sound-speed based solely on the internal energy of each particle alone. For simplicity, here we ignore the multi-dt nature of the simulation. However, as Fig. 2.7 shows, even only after one intervening time-step (i.e. after d$t_{\rm hot}$), there is a 50% to 500% error in the pressure of the neighbouring particle.

This error in the pressure of the neighbouring particle represents a poorly tracked non-conservation of energy. An incorrect relationship between the local internal energy and pressure field of the particles leads directly to force errors of the same magnitude. Because of the conservative and symmetric structure of the applied equations of motion, however, this does not lead to the total energy of the fluid changing over time (i.e. the sum of the kinetic and internal energy of the fluid remains constant), instead manifesting as unstable dynamics.

2.7 Conclusions

The Pressure-Energy and Pressure-Entropy schemes have been prized for their ability to capture contact discontinuities significantly better than their Density-based cousins due to their use of a directly smoothed pressure field (Hopkins, 2013). However, there are several disadvantages to using these schemes that have been presented:

- Injecting energy in a Pressure-Entropy based scheme requires the use of an iterative solver and many transformations between variables. This makes this scheme computationally expensive, and as such for this to be used in practice an efficient implementation is required. Approximate solutions do exist, but result in incorrect amounts of energy being injected into the field when particles are heated only by a (relatively, for astrophysics) small amount (typically by less than

100 times their own internal energy). This occurs even in the case where the fluid is evolved with a single, global, time-step, and is complicated even further by the inclusion of the multiple time-stepping scheme that is commonplace in cosmological simulations.

- In a Pressure-Energy based scheme, the injection of energy in a multi-dt simulation requires either 'waking up' all of the neighbours of the affected particle (and forcing them to be active in the next time-step), or a loop over these neighbours to back-port changes to their pressure due to the changes in internal energy of the heated particle. This is a computationally expensive procedure, and is generally avoided in the practical use of these schemes. As such, while no explicit energy conservation errors manifest, there is an offset between the energy field represented by the particle distribution and the associated smooth pressure field in practical implementations.

- These issues also manifest themselves in cases where energy is removed from active particles, such as an 'operator-splitting' radiative cooling scheme where energy is directly removed from particles.

- Correctly 'drifting' the smoothed pressure of particles (as is required in a multi-dt simulation) requires knowing the time differential of the smoothed pressure. To compute this, either an extra loop over neighbours is required for active particles, or an approximate solution based on the time differential of the density field and internal energy field is used. This approximate solution does not account for the changes taking place in the local internal energy field and as such does not correctly capture the evolution of the smoothed pressure.

- Even when using the 'correct' drift operator for the smoothed pressure significant pressure, and hence force, errors can occur when particles cool rapidly. This can be mitigated somewhat with time-step limiting techniques (either through the use of a time-step limiter like the one described in Durier & Dalla Vecchia (2012) or through a careful construction of a more representative sound-speed) but it is not possible to prevent errors on the same order as the relative energy difference between the cooling particle and its neighbours.

All of the above listed issues are symptomatic of one main flaw in these schemes; the SPH method assumes that the variables being smoothed over vary slowly during a single time-step. This is often true for the internal energy or particle entropy in idealised hydrodynamics tests, but in practical simulations with sub-grid radiative cooling (and energy injection) this leads to significant errors. These errors could be mitigated by using a different cooling model, where over a single time-step only small changes in the energies of particles could be made (i.e. by limiting the time-steps of particles to

significantly less than their cooling time), however this would render most cosmological simulations impractical to complete due to the huge increase in the number of time-steps to finish the simulation that this would imply.

Thankfully, due to the explicit connection between internal energy and pressure in the Density-based SPH schemes, they do not suffer the same ills. They also smooth over the mass field, which either does not vary or generally varies very slowly (on much larger timescales than the local dynamical time). As such, the only recommendation that it is possible to make is to move away from Pressure-based schemes in favour of their Density-based cousins, solving the surface tension issues at contact discontinuities with artificial conduction instead of relying on the smoothed pressure field from Pressure-based schemes. It is worth noting that most modern implementations of the Pressure-based schemes already use an artificial conduction (also known as energy diffusion) term to resolve residual errors in fluid mixing problems Hu et al. (2014); Hopkins (2015). Of particular note is the lack of phase mixing (due to the non-diffusive nature of SPH) between hot and cold fluids, even in Pressure-SPH.

Chapter 3

SPHENIX: Smoothed Particle Hydrodynamics for the next generation of galaxy formation simulations

3.1 Introduction

Because Smoothed Particle Hydrodynamics (SPH) strikes the sweet spot between computational cost, stability, and adaptivity, it has been used throughout the astronomical community for nearly five decades. The practical use of SPH in a cosmological context began with Hernquist & Katz (1989), which provided a novel solution to the large dynamic range of time-steps required to evolve a cosmological fluid, and was cemented by the Gadget-2 code (Springel, 2005) that was made public and exploited worldwide to model galaxy formation processes within this context for the first time (e.g. Dolag et al., 2004; Ettori et al., 2006; Crain et al., 2007). The base SPH model released in Gadget-2, however, was relatively simple, consisting of a fixed artificial viscosity coefficient and scheme based on Monaghan (1992). Improved models existed, such as those presented in Monaghan (1997), but the key that led to the community rallying around Gadget-2 was both its open source nature and scalability, with Gadget-2 able to run on hundreds or thousands of cores.

The popularity of Gadget-2, and similar codes like GASOLINE (Wadsley et al., 2004), along with its relatively simple hydrodynamics model, led to critical works such as Agertz et al. (2007) and Bauer & Springel (2012) that pointed out flaws in their SPH modelling, relative to mesh-based codes of the time. The SPH community as a whole, however, already had solutions to these problems (see e.g. Price, 2008) and many robust solutions were proposed and integrated into cosmological modelling

codes. In Heß & Springel (2010), the authors experimented with an extension to Gadget-2 using a Voronoi mesh to reduce errors inherrent in SPH and allow for better results on fluid mixing problems, eventually giving rise to the AREPO moving mesh scheme, allowing for significantly improved accuracy per particle but drastically increasing computational cost (Springel, 2010; Weinberger et al., 2020). In this case, the authors have steadily increased their computational cost per particle in an attempt to reduce errors inherrent in their hydrodynamics model as much as practicable.

Other authors took different directions, with the GASOLINE code (Wadsley et al., 2004, 2008, 2017) choosing to explicitly average pressures within the SPH equation of motion to alleviate the problems of artificial surface tension; the PHANTOM developers (Price, 2008, 2012; Price et al., 2018) advocating for artificial conduction of energy; and further developments on the Gadget-2 and updated Gadget-3 code by Hopkins (2013) and Hu et al. (2014) based on the work by Saitoh & Makino (2013) using an explicit smoothed pressure scheme to ensure a consistent pressure field over the contact discontinuities that artificial surface tension arises from.

Simultaneously, there was work to reduce the fundamental numerical errors present in SPH taking place by (Cullen & Dehnen, 2010; Dehnen & Aly, 2012; Read et al., 2010; Read & Hayfield, 2012) through the use of improved choices for the SPH kernel, which up until this point was assumed to have little effect on results from SPH simulations. These improved kernels typically have larger 'wings', encompassing more neighbours and providing more accurate reconstructions for smoothed quantities. These more accurate reconstructions are particularly important for the construction of accurate gradients, which enter into 'switches' that control the strength of the artificial viscosity and conduction terms.

The rise of more complex SPH models occurred alongside a significant jump in the complexity of the corresponding galaxy formation models; such an increase in complexity was required as resolutions increased over time, meaning more physics could be modelled directly. Many astrophysical processes take place on scales smaller than what can be resolved in simulations and are included in these so-called galaxy formation 'sub-grid' models. These processes include radiative cooling, which has progressed from a simple one parameter model to element and even ionisation state dependent rates (see e.g. Wiersma et al., 2009; Ploeckinger & Schaye, 2020); star formation (see e.g. Cen & Ostriker, 1992; Schaye & Dalla Vecchia, 2008, and references therein); and stellar feedback to model supernovae and element outflows (see e.g. Navarro & White, 1993; Springel & Hernquist, 2003; Dalla Vecchia & Schaye, 2008, 2012, and references therein). The coupling of these processes to hydrodynamics is complex and often overlooked; careful treatment of conservation laws and quirks of the chosen variables used to represent the fluid can frequently hide errors in plain sight (Borrow

et al., 2020).

The development of the SWIFT code (Schaller et al., 2016) led to a re-implementation of the sub-grid model used for the EAGLE simulation (Schaye et al., 2015), and a chance to re-consider the ANARCHY SPH scheme that was used in the original (Gadget-3 based) code (Schaller et al., 2015). The findings in Oppenheimer et al. (2018) (their Appendix D) and Chapter 2 meant that a switch away from the original Pressure-Entropy scheme to one based on a smoothed density field was preferred, along with the key design goals outlined below. This chapter describes the SPHENIX[1] scheme and demonstrates its performance on many hydrodynamics tests. We note here that SPHENIX does not give the best performance-per-particle (i.e. it will return higher values of the L1 norm, see §3.5.1, and a slower convergence rate) compared to other schemes. The moving mesh AREPO (Springel, 2010), finite-volume GIZMO (Hopkins, 2015), and corrected scheme presented in Rosswog (2020a) will produce lower error results. SPHENIX however lies in the very low-cost (memory and computation) per particle sweet-spot that traditional SPH schemes occupy, whilst maximising performance with some novel limiters for artificial conduction and viscosity. This makes it an excellent choice for the default SPH scheme in SWIFT. This allows SWIFT to be used to perform very large (tens to hundreds of billions of particles) simulations to study the statistical properties of coupled galaxies in detail.

The remainder of this chapter is organised as follows: in §3.2 we describe the SWIFT cosmological simulation code and the time-stepping algorithms present within it. In §3.3 we describe SPHENIX in its entirety. In §3.4 we describe the artificial conduction limiter used for energetic feedback schemes. Finally, in §3.5 we show how SPHENIX performs on various hydrodynamics problems.

3.2 The SWIFT simulation code

The SWIFT[2] simulation code (Schaller et al., 2016, 2018) is a hybrid parallel SPH and gravity code, designed to run across multiple compute nodes using MPI, but to utilise threads on each node (rather than the traditional method of using one MPI rank per core). This, along with its task-based parallelism approach, asynchronous communication scheme, and work-splitting domain decomposition system allow for excellent strong- and weak-scaling characteristics (Borrow et al., 2018).

SWIFT is also designed to be hugely modular, with hydrodynamics schemes, gravity schemes, and sub-grid models able to be easily swapped out. SWIFT can be configured to use a replica of the Gadget-2

[1] Note that, similar to the popular GIZMO schemes, SPHENIX is not an acronym.

[2] For the interested reader, the implementation of the SPHENIX scheme was developed fully in the open and is available in the SWIFT repository at `http://swiftsim.com` (Schaller et al., 2018), including all of the tests and examples shown below. We use version 0.9.0 of the SWIFT code for the tests in this work.

hydrodynamics scheme (Springel & Hernquist, 2002), a simplified version of the base PHANTOM scheme (Price et al., 2018), the MFM and MFV schemes described in Hopkins (2015), SPHENIX, or a host of other schemes. It can also be configured to use multiple different galaxy formation sub-grid models, including a very basic scheme (constant Λ cooling, no star formation), the EAGLE sub-grid model (Schaye et al., 2015), a 'Quick Lyman-α" model, the GEAR sub-grid model (Revaz & Jablonka, 2012), and some further evolutions including cooling tables from Ploeckinger & Schaye (2020). The gravity solver is interchangeable but the one used here, and throughout all SWIFT simulations, uses the Fast Multipole Method (Greengard & Rokhlin, 1987) with an adaptive opening angle, similar to Dehnen (2014).

3.2.1 Time integration

SWIFT uses a velocity-verlet scheme to integrate particles through time. This takes their acceleration ($\vec{a}$) from the equation of motion and time-step (Δt) and integrates their position forward in time through a Kick-Drift-Kick scheme as follows:

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}(t) + \frac{\Delta t}{2}\vec{a}(t), \tag{3.2.1}$$

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}\left(t + \frac{\Delta t}{2}\right)\Delta t, \tag{3.2.2}$$

$$\vec{v}(t + \Delta t) = \vec{v}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2}\vec{a}(t + \Delta t), \tag{3.2.3}$$

where the first and last equations, updating the velocity, are referred to as the 'kick', and the central equation is known as the 'drift'. The careful observer will note that the 'drift' can be split into as many pieces as required allowing for accurate interpolation of the particle position in-between kick steps. This is important in cosmological galaxy formation simulations, where the dynamic range is large. In this case, particles are evolved with their own, particle-carried time-step, given by

$$\Delta t_i = \mathrm{C_{CFL}} \frac{2\gamma_K h_i}{v_{\mathrm{sig},i}}, \tag{3.2.4}$$

dependent on the Courant–Friedrichs–Lewy ($\mathrm{C_{CFL}}$ Courant et al., 1928) constant, the kernel-dependent relationship between cut-off and smoothing length γ_K, particle-carried smoothing length h_i, and signal velocity $v_{\mathrm{sig},i}$ (see Equation 3.3.26). The discussion of the full time-stepping algorithm is out of the scope of this work, but see Hernquist & Katz (1989) and Borrow et al. (2019) for more information.

Time-step Limiter

As the time-step of the particles is particle-carried, there may be certain parts of the domain that contain interacting particles with vastly different time-steps (this is particularly promoted by particles with varied temperatures within a given kernel). Having these particles interact is problematic for a number of reasons, and as such we include the time-step limiter described in Durier & Dalla Vecchia (2012) in all problems solved below. SWIFT chooses to limit neighbouring particles to have a maximal time-step difference of a factor of 4 following the recommendations of the aforementioned study.

3.3 SPHENIX

The SPHENIX scheme was designed to replace the ANARCHY scheme used in the original EAGLE simulations for use in the SWIFT simulation code. This scheme had three major design goals:

- Be a Lagrangian SPH scheme, as this has many advantages and is compatible with the EAGLE subgrid model.
- Work well with the EAGLE subgrid physics, namely instantaneous energy injection and subgrid cooling.
- Be highly computationally and memory efficient.

The last requirement precludes the use of any Riemann solvers in so-called GIZMO-like schemes (although these do not necessarily give improved results for astrophysical problem sets, see Borrow et al., 2019); see Appendix A.1. The second requirement also means that the use of a pressure-based scheme (such as ANARCHY) is not optimal, see Chapter 2 for more details.

The SPHENIX scheme is based on so-called 'Traditional' Density-Energy SPH. This means that it uses the smoothed mass density,

$$\hat{\rho}(\vec{x}) = \sum_j m_j W(|\vec{x} - \vec{x}_j|, h(\vec{x})) \tag{3.3.1}$$

where here j are indices describing particles in the system, $h(\vec{x})$ is the smoothing length evaluated at position $\vec{x}$, and $W(r, h)$ is the kernel function.

In the examples below, the Quartic Spline (M5) kernel,

$$w(q) = \begin{cases} \left(\frac{5}{2}-q\right)^4 - 5\left(\frac{3}{2}-q\right)^4 + 10\left(\frac{1}{2}-q\right)^4 & q < \frac{1}{2} \\ \left(\frac{5}{2}-q\right)^4 - 5\left(\frac{3}{2}-q\right)^4 & \frac{1}{2} \leq q < \frac{3}{2} \\ \left(\frac{5}{2}-q\right)^4 & \frac{3}{2} \leq q < \frac{5}{2} \\ 0 & q \geq \frac{5}{2} \end{cases} \tag{3.3.2}$$

with $W(r,h) = \kappa_{n_D} w(r/h)/h^{n_D}$, n_D the number of dimensions, and $\kappa_3 = 1/20\pi$ for three dimensions, is used. The SPHENIX scheme has also been tested with other kernels, notably the Cubic and Quintic Spline (M4, M6) and the Wendland (C2, C4, C6) kernels (Wendland, 1995). The choice of kernel does not qualitatively affect the results in any of the tests in this work (see Dehnen & Aly, 2012, for significantly more information on kernels). Higher order kernels do allow for lower errors on tests that rely on extremely accurate reconstructions to cancel forces (for instance the Gresho-Chan vortex, §3.5.3), but we find that the Quintic Spline provides an excellent trade-off between computational cost and accuracy in practical situations. Additionally, the Wendland kernels do have the benefit that they are not susceptible to the pairing instability, but they must have an ad-hoc correction applied in practical use (Dehnen & Aly, 2012, Section 2.5). We find no occurrences of the pairing instability in both the tests and our realistic simulations. The SPHENIX scheme is kernel-invariant, and as such can be used with any reasonable SPH kernel.

The smoothing length h is determined by satisfying

$$\hat{n}(\vec{x}) = \sum_j W\left(|\vec{x} - \vec{x}_j|, h(\vec{x})\right) = \left(\frac{\eta}{h(\vec{x})}\right)^{n_D}, \tag{3.3.3}$$

with η setting the resolution scale. The precise choice for η generally does not qualitatively change results; here we choose $\eta = 1.2$ due to this value allowing for a very low E_0 error (see Read et al., 2010; Dehnen & Aly, 2012)[3], which is a force error originating from particle disorder within a single kernel. In SWIFT, the constraint in Equation 3.3.3 is solved numerically to a relative accuracy of 10^{-4}.

The smoothed mass density, along with a particle-carried internal energy per unit mass u, is used to determine the pressure at a particle position through the equation of state

$$P(\vec{x}_i) = P_i = (\gamma - 1) u_i \hat{\rho}_i, \tag{3.3.4}$$

[3]This corresponds to ~58 weighted neighbours for our Quartic Spline in a scenario where all neighbours have uniform smoothing lengths. In practical simulations the 'number of neighbours' that a given particle interacts with can vary by even orders of magnitude but Equation 3.3.3 must be satisfied for all particles ensuring an accurate reconstruction of the field. More discussion on this choice of smoothing length can be found in (Springel & Hernquist, 2002; Monaghan, 2002; Price, 2007, 2012; Borrow et al., 2020). We chose $\eta = 1.2$ based on Figure 3 in Dehnen & Aly (2012), where this corresponds to a very low reconstruction error in the density.

with γ the ratio of specific heats, taken to be $5/3$ throughout unless specified. This pressure enters the first law of thermodynamics,

$$\left.\frac{\partial u_i}{\partial \vec{q}_i}\right|_{A_i} = -\frac{P_i}{m_i}\frac{\partial V_i}{\partial \vec{q}_i}, \tag{3.3.5}$$

with $\vec{q}_i$ a state vector containing both $\vec{x}_i$ and h_i as *independent* variables, A_i the entropy of particle i (i.e. this equation only applies to dissipationless dynamics), and $V_i = m_i/\hat{\rho}_i$ describing the volume represented by particle i. This constraint, along with the one on the smoothing length, allows for an equation of motion to be extracted from a Lagrangian (see e.g. the derivations in Springel & Hernquist, 2002; Hopkins, 2013),

$$\frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = -\sum_j m_j \left[\frac{f_{ij}P_i}{\hat{\rho}_i^2}\nabla_i W_{ij} + \frac{f_{ji}P_j}{\hat{\rho}_j^2}\nabla_j W_{ji}\right], \tag{3.3.6}$$

where $W_{ab} = W(|\vec{x}_b - \vec{x}_a|, h(\vec{x}_a))$, $\nabla_a = \partial/\partial\vec{x}_a$, and f_{ab} a dimensionless factor encapsulating the non-uniformity of the smoothing length field

$$f_{ab} = 1 - \frac{1}{m_b}\left(\frac{h_a}{n_D \hat{n}_a}\frac{\partial\hat{\rho}_i}{\partial h_i}\right)\left(1 + \frac{h_i}{n_D\hat{n}_i}\frac{\partial\hat{n}_i}{\partial h_i}\right)^{-1} \tag{3.3.7}$$

and is generally of order unity[4]. There is also an associated equation of motion for internal energy,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -\sum_j m_j f_{ij}\frac{P_i}{\hat{\rho}_i^2}\vec{v}_{ij}\cdot\nabla_i W_{ij}, \tag{3.3.8}$$

with $\vec{v}_{ij} = \vec{v}_j - \vec{v}_i$. Note that other differences between vector quantities are defined in a similar way, including for the separation of two particles $\vec{x}_{ij} = \vec{x}_j - \vec{x}_i$.

3.3.1 Artificial Viscosity

These equations, due to the constraint of constant entropy introduced in the beginning, lead to naturally dissipationless solutions; they cannot capture shocks. Shock capturing in SPH is generally performed using 'artificial viscosity'.

The artificial viscosity implemented in SPHENIX is a simplified and modified extension to the Cullen & Dehnen (2010) 'inviscid SPH' scheme. This adds the following terms to the equation of motion (see Monaghan, 1992, and references within),

$$\frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = -\sum_j m_j \zeta_{ij}\left[f_{ij}\nabla_i W_{ij} + f_{ji}\nabla_j W_{ji}\right], \tag{3.3.9}$$

[4]Here the derivatives of the smoothed quantities with respect to h are simply calculated as an explicit sum over neighbours as would be expected from differentiating Equation 2.2.2.

and to the associated equation of motion for the internal energy,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -\frac{1}{2}\sum_j m_j \zeta_{ij} \vec{v}_{ij} \cdot \left[f_{ij} \nabla_i W_{ij} + f_{ji} \nabla_j W_{ji} \right], \tag{3.3.10}$$

where ζ_{ij} controls the strength of the viscous interaction. Note here that the internal energy equation of motion is explicitly symmetrised, which was not the case for the SPH equation of motion for internal energy (Eqn. 3.3.8). In this case, that means that there are terms from both the ij and ji interactions in Equation 3.3.10, whereas in Equation 3.3.8 there is only a term from the ij interaction. This choice was due to the symmetric version of this equation performing significantly better in the test examples below, likely due to multiple time-stepping errors within regions where the viscous interaction is the strongest[5].

There are many choices available for ζ_{ij}, with the case used here being

$$\zeta_{ij} = -\alpha_V \mu_{ij} \frac{v_{\mathrm{sig},ij}}{\hat{\rho}_i + \hat{\rho}_j}, \tag{3.3.11}$$

where

$$\mu_{ij} = \begin{cases} \frac{\vec{v}_{ij} \cdot \vec{x}_{ij}}{|\vec{x}_{ij}|} & \vec{v}_{ij} \cdot \vec{x}_{ij} < 0 \\ 0 & \vec{v}_{ij} \cdot \vec{x}_{ij} \geq 0 \end{cases} \tag{3.3.12}$$

is a basic particle-by-particle converging flow limiter (meaning that the viscosity term vanishes when $\nabla \cdot \vec{v} \geq 0$), and

$$v_{\mathrm{sig},ij} = c_i + c_j - \beta_V \mu_{ij}, \tag{3.3.13}$$

is the signal velocity between particles i and j, with $\beta_V = 3$ a dimensionless constant, and with c_i the soundspeed of particle i defined through the equation of state as

$$c_i = \sqrt{\gamma \frac{P_i}{\hat{\rho}_i}} = \sqrt{(\gamma - 1)\gamma u_i}. \tag{3.3.14}$$

Finally, the dimensionless viscosity coefficient α_V (Monaghan & Gingold, 1983) is frequently taken to be a constant of order unity. In SPHENIX, this becomes an interaction-dependent constant (see Morris & Monaghan, 1997; Cullen & Dehnen, 2010, for similar schemes), with $\alpha_V = \alpha_{V,ij}$, dependent on two particle-carried α values as follows:

$$\alpha_{V,ij} = \frac{1}{4}(\alpha_{V,i} + \alpha_{V,j})(B_i + B_j), \tag{3.3.15}$$

where

$$B_i = \frac{|\nabla \cdot \vec{v}_i|}{|\nabla \cdot \vec{v}_i| + |\nabla \times \vec{v}_i| + 10^{-4} c_i / h_i} \tag{3.3.16}$$

[5] We note additionally that the authors are unfamiliar with any other works that choose not to symmetrise these equations.

is the Balsara (1989) switch for particle i, which allows for the deactivation of viscosity in shear flows, where there is a high value of $\nabla \cdot \vec{v}$ relative to $\nabla \cdot \vec{v}$, but the associated shear viscosity is unnecessary. This, in particular, affects rotating shear flows such as galaxy disks, where the scheme used to determine $\alpha_{V,i}$ described below will return a value close to the maximum.

The equation for $\alpha_{V,i}$ is solved independently for each particle over the course of the simulation. Note that $\alpha_{V,i}$ is never drifted, and is only ever updated at the 'kick' steps. The source term in the equation for $\alpha_{V,i}$, designed to activate the artificial viscosity within shocking regions, is the shock indicator

$$S_i = \begin{cases} -h_i^2 \max\left(\dot{\nabla} \cdot \vec{v}_i, 0\right) & \nabla \cdot \vec{v}_i \leq 0 \\ 0 & \nabla \cdot \vec{v}_i > 0 \end{cases} \tag{3.3.17}$$

where here the time differential of the local velocity divergence field

$$\dot{\nabla} \cdot \vec{v}_i(t + \Delta t) = \frac{\nabla \cdot \vec{v}_i(t + \Delta t) - \nabla \cdot \vec{v}_i(t)}{\Delta t} \tag{3.3.18}$$

with $\nabla \cdot \vec{v}_i$ the local velocity divergence field and Δt the time-step associated with particle i. The primary variable in the shock indicator S_i of $\dot{\nabla} \cdot \vec{v}$ is high in pre-shock regions, with the secondary condition for the flow being converging ($\nabla \cdot \vec{v} \leq 0$) helps to avoid false detections. As the Balsara (1989) switch is used independently from the equation that evolves $\alpha_{V,i}$, a choice that is notably different from most other schemes that use B_i directly in the shock indicator S_i, it can neutralize viscosity instantaneously. This choice allows for improved shock capturing in shearing flows (e.g. feedback events occurring within a galaxy disk). In these cases, the Balsara (1989) switch (which is instantaneously evaluated) rapidly becomes close to 1.0, and the already high value of $\alpha_{V,i}$ allows for a strong viscous reaction from the fluid. The shock indicator is then transformed into an optimal value for the viscosity coefficient as

$$\alpha_{V,\mathrm{loc},i} = \alpha_{V,\mathrm{max}} \frac{S_i}{c_i^2 + S_i}, \tag{3.3.19}$$

with a maximum value of $\alpha_{V,\mathrm{loc}} = \alpha_{V,\mathrm{max}} = 2.0$ The value of $\alpha_{V,i}$ is then updated as follows:

$$\alpha_{V,i} = \begin{cases} \alpha_{V,\mathrm{loc},i} & \alpha_{V,i} < \alpha_{V,\mathrm{loc},i} \\ \frac{\alpha_{V,i} + \alpha_{V,\mathrm{loc},i} \frac{\Delta t}{\tau_{V,i}}}{1 + \frac{\Delta t}{\tau_{V,i}}} & \alpha_{V,i} > \alpha_{V,\mathrm{loc},i} \end{cases} \tag{3.3.20}$$

where $\tau_{V,i} = \gamma_K \ell_V h_i / c_i$ with γ_K the 'kernel gamma' a kernel dependent quantity relating the smoothing length and compact support ($\gamma_K = 2.018932$ for the quartic spline in 3D, Dehnen & Aly, 2012) and ℓ_V a constant taking a value of 0.05. The final value of $\alpha_{V,i}$ is checked against a minimum, however the default value of this minimum is zero and the evolution strategy used above guarantees that $\alpha_{V,i}$ is

strictly positive and that the decay is stable regardless of time-step.

3.3.2 Artificial Conduction

Attempting to resolve sharp discontinuities in non-smoothed variables in SPH leads to errors. This can be seen above, with strong velocity discontinuities (shocks) not being correctly handled and requiring an extra term in the equation of motion (artificial viscosity) to be captured. A similar issue arises when attempting to resolve strong discontinuities in internal energy (temperature). To resolve this, we introduce an artificial energy conduction scheme similar to the one presented by Price (2008). This adds an extra term to the equation of motion for internal energy,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_j m_j v_{D,ij}(u_i - u_j)\hat{r}_{ij} \cdot \left(f_{ij}\frac{\nabla_i W_{ij}}{\hat{\rho}_i} + f_{ji}\frac{\nabla_j W_{ji}}{\hat{\rho}_j} \right) \tag{3.3.21}$$

with $\hat{r}_{ij}$ the unit vector between particles i and j, and where

$$v_{D,ij} = \frac{\alpha_{D,ij}}{2}\left(\frac{|\vec{v}_{ij} \cdot \vec{x}_{ij}|}{|\vec{x}_{ij}|} + \sqrt{2\frac{|P_i - P_j|}{\hat{\rho}_j + \hat{\rho}_j}} \right). \tag{3.3.22}$$

This conductivity speed is the average of two commonly used speeds, with the former velocity-dependent term taken from Price et al. (2018) (modified from Wadsley et al., 2008), and the latter pressure-dependent term taken from Price (2008). These are usually used separately for cases that aim to reduce entropy generation in disordered fields and contact discontinuities respectively (where initially there is a strong discontinuity in pressure that is removed by the artificial conduction scheme), but we combine them here as both cases are relevant in galaxy formation simulations and use this same velocity throughout our testing, a notable difference with other works using conduction schemes (e.g. Price et al., 2018). This is additionally somewhat similar to the conduction speed used in ANARCHY and Hu et al. (2014), which is a modified version of the signal velocity (Eqn. 3.3.13) with our speed replacing the sum of sound speeds with a differenced term. Appendix A.2 contains an investigation of the individual terms in the conduction velocity. The interaction-dependent conductivity coefficient,

$$\alpha_{D,ij} = \frac{P_i \alpha_{D,i} + P_j \alpha_{D,j}}{P_i + P_j}, \tag{3.3.23}$$

is pressure-weighted to enable the higher pressure particle to lead the conduction interaction, a notable departure from other thermal conduction schemes in use today. This is critical when it comes to correctly applying the conduction limiter during feedback events, described below. The individual particle-carried $\alpha_{D,i}$ are ensured to only be active in cases where there is a strong discontinuity in

internal energy. This is determined by using the following discontinuity indicator,

$$K_i = \beta_D \gamma_K h_i \frac{\nabla^2 u_i}{\sqrt{u_i}}, \tag{3.3.24}$$

where β_D is a fixed dimensionless constant taking a value of 1, and this form of the source term was taken to be consistent with the ANARCHY scheme in (Schaye et al., 2015). The discontinuity indicator enters the time differential for the individual conduction coefficients as a source term,

$$\frac{\mathrm{d}\alpha_{D,i}}{\mathrm{d}t} = K_i + \frac{\alpha_{D,\mathrm{min}} - \alpha_{D,i}}{\tau_{D,i}}, \tag{3.3.25}$$

with $\tau_{D,i} = \gamma_K h_i / v_{\mathrm{sig},i}$, $\alpha_{D,\mathrm{min}} = 0$ the minimal allowed value of the artificial conduction coefficient, and with the individual particle signal velocity,

$$v_{\mathrm{sig},i} = \max_j (v_{\mathrm{sig},ij}), \tag{3.3.26}$$

controlling the decay rate. $\nabla^2 u$ is used as the indicator for a discontinuity, as opposed to ∇u, as it allows for (physical, well represented within SPH) linear gradients in internal energy to be maintained without activating artificial conduction. This is then integrated during 'kick' steps using

$$\alpha_{D,i}(t + \Delta t) = \alpha_{D,i}(t) + \frac{\mathrm{d}\alpha_{D,i}}{\mathrm{d}t} \Delta t. \tag{3.3.27}$$

The final stage of evolution for the individual conduction coefficients is to limit them based on the local viscosity of the fluid. This is necessary because thermal feedback events explicitly create extreme discontinuities within the internal energy field that lead to shocks (see §3.4 for the motivation leading to this choice). The limit is applied using the maximal value of viscous alpha among the neighbours of a given particle,

$$\alpha_{V,\mathrm{max},i} = \max_j (\alpha_{V,j}), \tag{3.3.28}$$

with the limiter being applied using the maximally allowed value of the conduction coefficient,

$$\alpha_{D,\mathrm{max},i} = \alpha_{D,\mathrm{max}} \left(1 - \frac{\alpha_{V,\mathrm{max},i}}{\alpha_{V,\mathrm{max}}} \right), \tag{3.3.29}$$

with $\alpha_{D,\mathrm{max}} = 1$ a constant, and

$$\alpha_{D,i} = \begin{cases} \alpha_{D,i} & \alpha_{D,i} < \alpha_{D,\mathrm{max}} \\ \alpha_{D,\mathrm{max}} & \alpha_{D,i} > \alpha_{D,\mathrm{max}}. \end{cases} \tag{3.3.30}$$

This limiter allows for a more rapid increase in conduction coefficient, and a higher maximum, than would usually be viable in simulations with strong thermal feedback implementations. In ANARCHY, another scheme employing artificial conduction, the rate at which the artificial conduction could grow

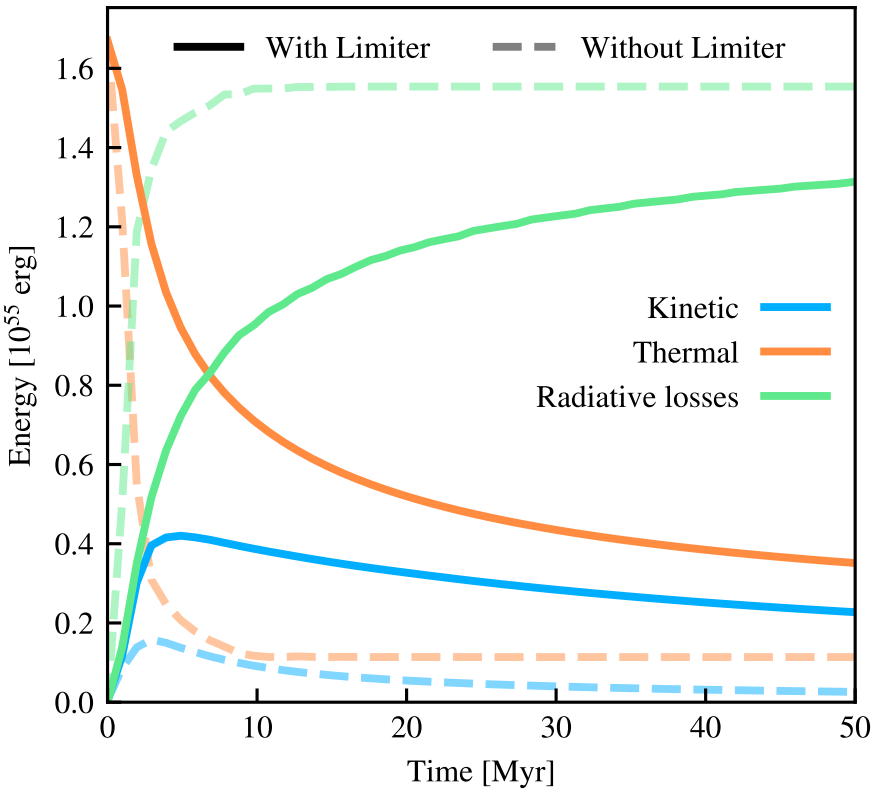


Figure 3.1: Energy in various components as a function of time for a simulated supernova blast (see text for details of the set-up). Blue shows energy in the kinetic phase, orange shows energy in the thermal phase (neglecting the thermal energy of the background) and green shows energy lost to radiation. The solid lines show the simulation performed with the artificial conduction limiter applied, and the dashed lines show the simulation without any such limiter. Simulations performed without the limiter show huge, rapid, cooling losses.

was chosen to be significantly smaller. In Anarchy, $\beta_D = 0.01$, which is 100 times smaller than the value chosen here (Schaye et al., 2015, Appendix A3). This additional conduction is required to accurately capture contact discontinuities with a Density-Energy SPH equation of motion.

3.4 Motivation for the Conduction Limiter

The conduction limiter first described in §3.3 is formed of two components; a maximal value for the conduction coefficient in viscous flows (Eqn. 3.3.30), and one that ensures that a particle with a higher pressure takes preference within the conduction interaction (Eqn. 3.3.23).

This limiter is necessary due to interactions of the artificial conduction scheme with the sub-grid physics model. Here the EAGLE sub-grid model is shown as this is what Sphenix was designed for use with, however all schemes employing energetic feedback and unresolved cooling times will suffer from the same problems when using un-limited artificial conduction. In short, when an energetic

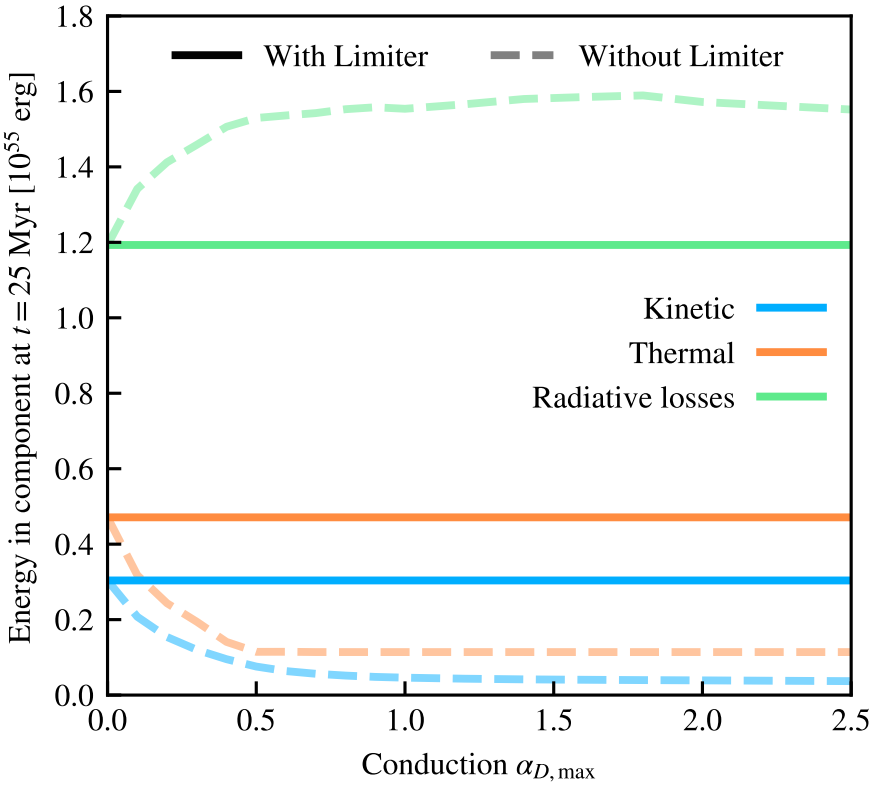


Figure 3.2: The set-up from Fig. 3.1 performed for different values for the maximum artificial conduction coefficient $\alpha_{D,\mathrm{max}}$ (i.e. a different horizontal axis as Fig. 3.1, with the same vertical axis), now showing the components of energy in each phase at a fixed time of $t = 25$ Myr. Colours and line styles are the same as in Fig. 3.1. As well as demonstrating the issue with un-limited conduction, this figure shows that the conduction limiter prevents the loss of additional energy energy relative to a simulation performed without any artificial conduction.

feedback event takes place, the artificial conduction switch is activated (as this is performed by injecting lots of energy into one particle, leading to an extreme value of $\nabla^2 u$). This then leads to energy leaking out of the particle ahead of the shock front, which is then radiated away as the neighbouring particles can rapidly cool due to their temperature being lower leading to smaller cooling times.

To show the effect of this problem on a real system, we set up a uniform volume containing 32^3 gas particles at approximately solar metallicity ($Z = 0.014$) and equilibrium temperature (around 10^4 K), at a density of $n_{\rm H} = 0.1$ cm^{-3}. The central particle in the volume has approximately the same amount of energy injected into it as in a single EAGLE-like stellar feedback event (heating it to $\sim 10^{7.5}$ K) at the start of the simulation with full sub-grid cooling (using the tables from Wiersma et al., 2009) enabled. The initial values for the artificial viscosity and conduction coefficients are set to be zero (whereas in practice they are set to be their maximum and minimum in 'real' feedback events; this has little effect on the results as the coefficients rapidly stabilise).

Fig. 3.1 shows the energy in the system (with the thermal energy of the 'background' particles removed to ensure a large dynamic range in thermal energy is visible on this plot) in various components. We see that, at least for the case with the limiter applied, at $t = 0$ there is the expected large injection of thermal energy that is rapidly partially transformed into kinetic energy as in a classic blastwave problem (like the one shown in Fig. 3.5; in our idealised, non-radiative, Sedov blasts only 28% of the injected thermal energy is converted to kinetic energy). A significant fraction, around two thirds, of the energy is lost to radiation, but the key here is that there is a transformation of the initial thermal injection to a kinetic wave.

In the same simulation, now with the conduction limiter removed (dashed lines), almost all of the injected energy is immediately lost to radiation (i.e. the feedback is unexpectedly inefficient). The internal energy in the affected particle is rapidly conducted to its neighbours (that are then above, but closer to, the equilibrium temperature) which have a short cooling time and hence the energy is quickly lost.

The direct effect of the conduction limiter is shown in Fig. 3.2, where the same problem as above is repeated ten times with maximal artificial conduction coefficients $\alpha_{D,\rm max}$ of 0 to 2.5 in steps of 0.1 (note that the value of $\alpha_{D,\rm max}$ used in production simulations is 1). We choose to show these extreme values to demonstrate the efficacy of the limiter even in extreme scenarios. The simulations with and without the limiter show the same result at $\alpha_{D,\rm max} = 0$ (i.e. with conduction disabled) but those without the limiter show a rapidly increasing fraction of the energy lost to cooling as the maximal conduction coefficient increases. The simulations with the limiter show a stable fraction of energy (at this fixed time of $t = 25$ Myr) in each component, showing that the limiter is working as expected

and is curtailing these numerical radiative losses. This result is qualitatively unchanged for a factor of 100 higher, or lower, density background gas (i.e. gas between $n_{\rm H} = 0.001$ cm^{-3} and $n_{\rm H} = 10.0$ cm^{-3}). In both of these cases, the conduction can rapidly cause numerical radiative losses, but with the limiter enabled this is remedied entirely. We also note that the limiter remains effective even for extreme values of the conduction parameter (e.g. with $\alpha_{D,\rm max} = 100$), returning the same result as the case without artificial conduction for this test.

3.5 Hydrodynamics Tests

In this section the performance of Sphenix is shown on hydrodynamics tests, including the Sod (1978) shock tube, Sedov (1959) blastwave, and the Gresho & Sani (1990) vortex, along with many other problems relevant to galaxy formation. All problems are performed in hydrodynamics-only mode, with no radiative cooling or any other additional physics, and all use a $\gamma = 5/3$ equation of state ($P = (2/3)u_i\hat{\rho}$).

Crucially, all tests were performed with the same scheme parameters and settings, meaning that all of the switches are consistent (even between self-gravitating and pure hydrodynamical tests) unless otherwise stated. This departs from most studies where parameters are set for each problem independently, in an attempt to demonstrate the maximal performance of the scheme for a given test. The parameters used are as follows:

- The quartic spline kernel.
- CFL condition $C_{\rm CFL} = 0.2$, with multiple time-stepping enabled (see e.g. Lattanzio et al., 1986).
- Viscosity alpha $0.0 \leq \alpha_V \leq 2.0$ with the initial value being $\alpha_V = 0.1$ (similar to Cullen & Dehnen, 2010).
- Viscosity beta $\beta_V = 3.0$ and length $\ell_V = 0.05$ (similarly to Cullen & Dehnen, 2010).
- Conduction alpha $0.0 \leq \alpha_D \leq 1.0$ (a choice consistent with Price, 2008) with the viscosity-based conduction limiter enabled and the same functional form for the conduction speed (Eqn. 3.3.22) used in all simulations.
- Conduction beta $\beta_D = 1.0$ with the initial value of $\alpha_D = 0.0$.

These choices were all 'calibrated' to achieve an acceptable result on the Sod shock tube, and then left fixed with the results from the rest of the tests left unseen. We choose to present the tests in this

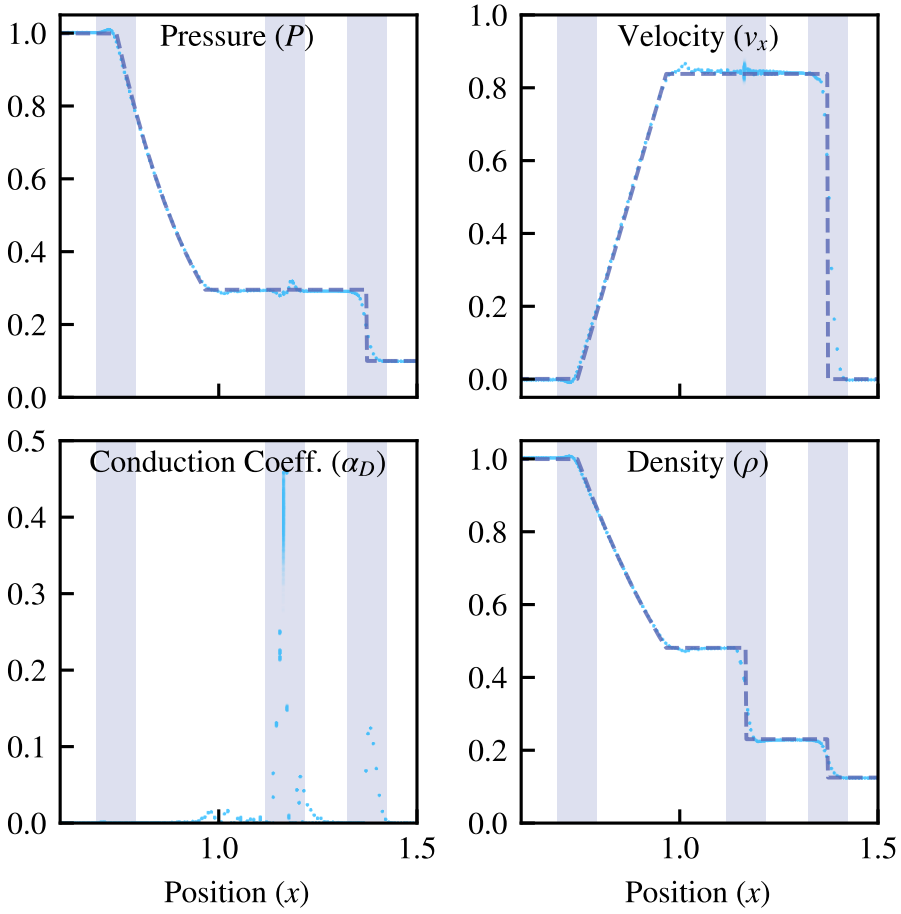


Figure 3.3: Individual quantities plotted against the analytic solution (purple dashed line) for the Sod shock tube in 3D. The horizontal axis shows the x position of the particles. All particles are shown in blue, with the purple shading in the background showing the regions considered for the convergence (Fig. 3.4) with the rarefaction wave, contact discontinuity, and shock, shown from left to right. All panels are shown at the same time $t = 0.2$, and for the same resolution level, using the 64^3 and 128^3 initial conditions for $x < 1$ and $x > 1$ respectively.

manner in an effort to show a representative overview of the performance of Sphenix in real-world conditions as it is primarily designed for practical use within future galaxy formation simulations.

For many tests, we show convergence tests, where the particle resolution is increased for a given test and the relative error compared. In these tests we choose to show the mean smoothing length rather than the equivalent one dimensional particle number as this is more natural in tests that are not in a dimensionless space (for instance the Evrard Collapse).

The source code required to produce the initial conditions (or a link to download the initial conditions themselves if this is impractical) are available open source from the Swift repository.

3.5.1 Sod shock tube

The Sod (1978) shock tube is a classic Riemann problem often used to test hydrodynamics codes. The tube is made up of three main sections in the final solution : the rarefaction wave (between $0.7 < x < 1.0$), contact discontinuity (at position $x \approx 1.2$), and a weak shock (at position $x \approx 1.4$) at the time that we show it in Figure 3.3.

Initial Conditions

The initial conditions for the Sod shock tube uses body centred cubic lattices to ensure maximally symmetric lateral forces in the initial state. Two lattices with equal particle masses, one at a higher density by a factor of 8 (e.g. one with 32^3 particles and one with 64^3 particles) are attached at $x = 1.0$[6]. This forms a discontinuity, with the higher density lattice being placed on the left with $\rho_L = 1$ and the lower density lattice on the right with $\rho_R = 1/8$. The velocities are initially set to zero for all particles and pressures set to be $P_L = 1$ and $P_R = 0.1$.

Results

Fig. 3.3 shows the shock tube at $t = 1$, plotted against the analytic solution. This figure shows the result from the 64^3 and 128^3 initial condition. In general the simulation data (blue points) shows very close agreement with the analytic solution (purple dashed line).

The three purple bands correspond to three distinct regions within the shock tube. The furthest left is the rarefaction wave, which is an adiabatically expanding fluid. The band covers the turnover point of the wave, as this is where the largest deviation from the analytic solution is present. There is a slight overestimation of the density at this turnover point, primarily due to the symmetric nature of the SPH kernel.

The next band shows the contact discontinuity. No effort is made to suppress this discontinuity in the initial conditions (i.e. they are not relaxed). There is a small pressure blip, of a similar size to that seen with schemes employing Riemann solvers such as GIZMO (Hopkins, 2015). There is no large velocity discontinuity associated with this pressure blip as is seen with SPH schemes that do not explicitly treat the contact discontinuity (note that every particle present in the simulation is shown

[6]This simplistic particle arrangement does cause a slight problem at the interface at higher (i.e. greater than one) dimensions. In 3D, some particles may have spurious velocities in the y and z directions at the interface, due to asymmetries in the neighbours found on the left and right side of the boundary. To offset this, the lattices are placed so that the particles are aligned along the x-axis wherever possible over the interface, however some spurious forces still result.

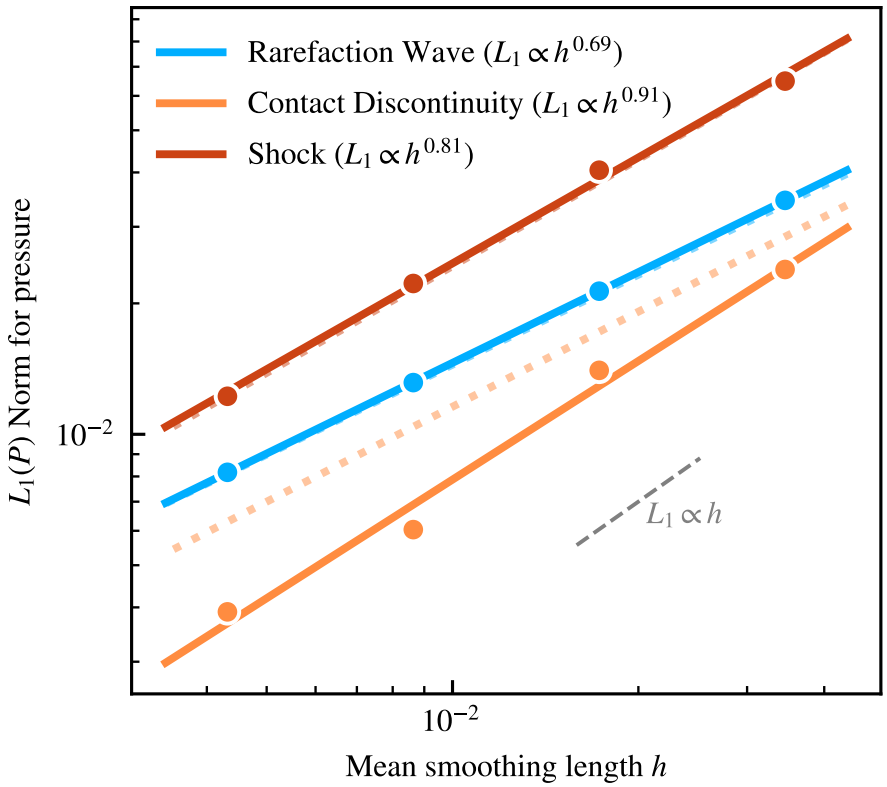


Figure 3.4: Pressure convergence for the three regions in Fig 3.3. The solid lines show fits to the data at various resolution levels (points) for each region, with the dotted lines showing convergence speed when the artificial conduction term is removed. The dashed grey line shows the expected speed of convergence for shocks in SPH simulations, to guide the eye, with a dependence of $L_1 \propto h$.

here) with some form of conduction, a smoothed pressure field, or other method. Due to the strong discontinuity in internal energy present in this region, the artificial conduction coefficient α_D peaks, allowing for the pressure 'blip' to be reduced to one with a linear gradient.

The final section of the system, the rightmost region, is the shock. This shock is well captured by the scheme. There is a small activation of the conduction coefficient in this region, which is beneficial as it aids in stabilising the shock front (Hu et al., 2014). This shows that the conduction limiter (§3.4) does not eliminate this beneficial property of artificial conduction within these frequently present weak (leading to $\alpha_V \lesssim 1.0$) shocks.

In an ideal case, the scheme would be able to converge at second order $L_1 \propto h^2$ away from shocks, and at first order $L_1 \propto h$ within shocks (Price et al., 2018). Here the L_1 norm of a band is defined as

$$L_1(K) = \frac{1}{n} \sum_n |K_{\text{sim}}(\vec{x}) - K_{\text{ref}}(\vec{x})| \tag{3.5.1}$$

with K some property of the system such as pressure, the subscripts sim and ref referring to the simulation data and reference solution respectively, and n the number of particles in the system.

Fig. 3.4 shows the convergence properties of the Sphenix scheme on this problem, using the pressure field in this case as the convergence variable. Compared to a scheme without artificial conduction (dotted lines), the Sphenix scheme shows significantly improved convergence and a lower norm in the contact discontinuity, without sacrificing accuracy in other regions.

3.5.2 Sedov-Taylor Blastwave

The Sedov-Taylor blastwave (Sedov blast; Taylor, 1950; Sedov, 1959) follows the evolution of a strong shock front through an initially isotropic medium. This is a highly relevant test for cosmological simulations, as this is similar to the implementations used for sub-grid (below the resolution scale) feedback from stars and black holes. In SPH schemes this effectively tests the artificial viscosity scheme for energy conservation; if the scheme does not conserve energy the shock front will be misplaced.

Initial Conditions

Here, we use a glass file generated by allowing a uniform grid of particles to settle to a state where the kinetic energy has stabilised. The particle properties are then initially set such that they represent a gas with adiabatic index $\gamma = 5/3$, a uniform pressure of $P_0 = 10^{-6}$, density $\rho_0 = 1$, all in a 3D box of side-length 1. Then, the $n = 15$ particles closest to the centre of the box have energy $E_0 = 1/n$ injected

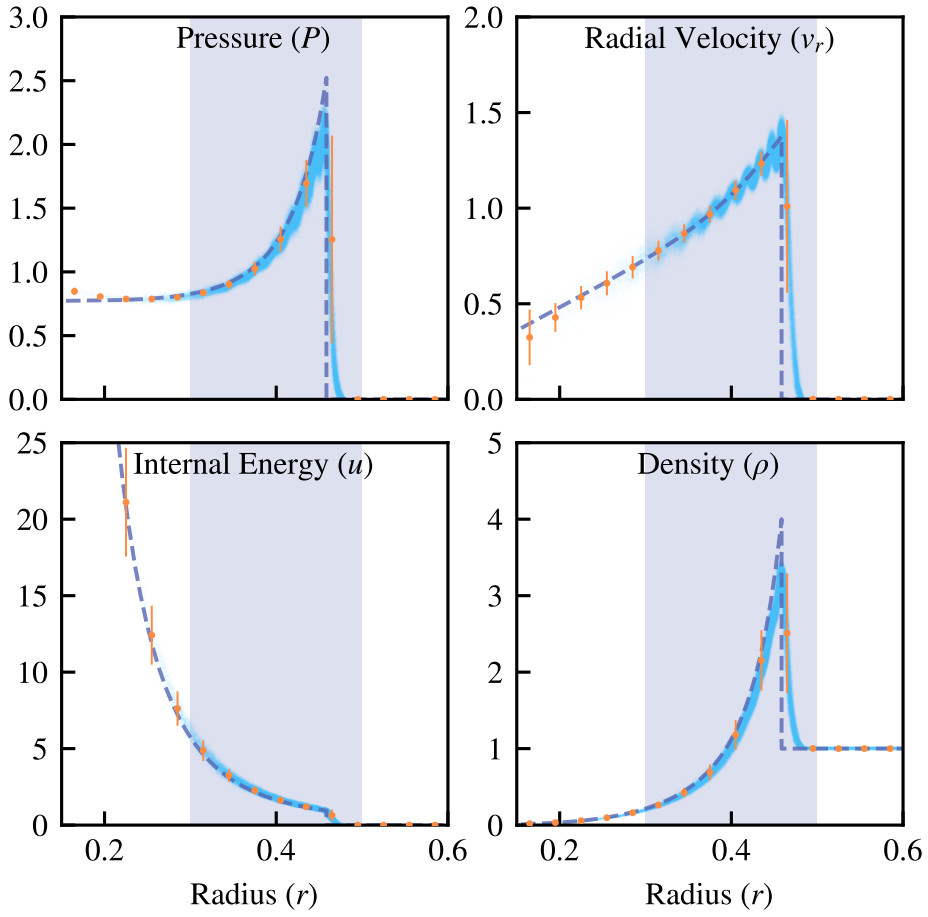


Figure 3.5: Particle properties at $t = 0.1$ shown against the analytic solution (purple dashed line) for the Sedov-Taylor blastwave. A random sub-set of 1/5th of the particles are shown in blue, with the orange points showing the mean value within equally spaced horizontal bins with one standard deviation of scatter. The background purple band shows the region considered for measuring convergence in Fig. 3.6. This figure shows the results for a 128^3 particle glass file.

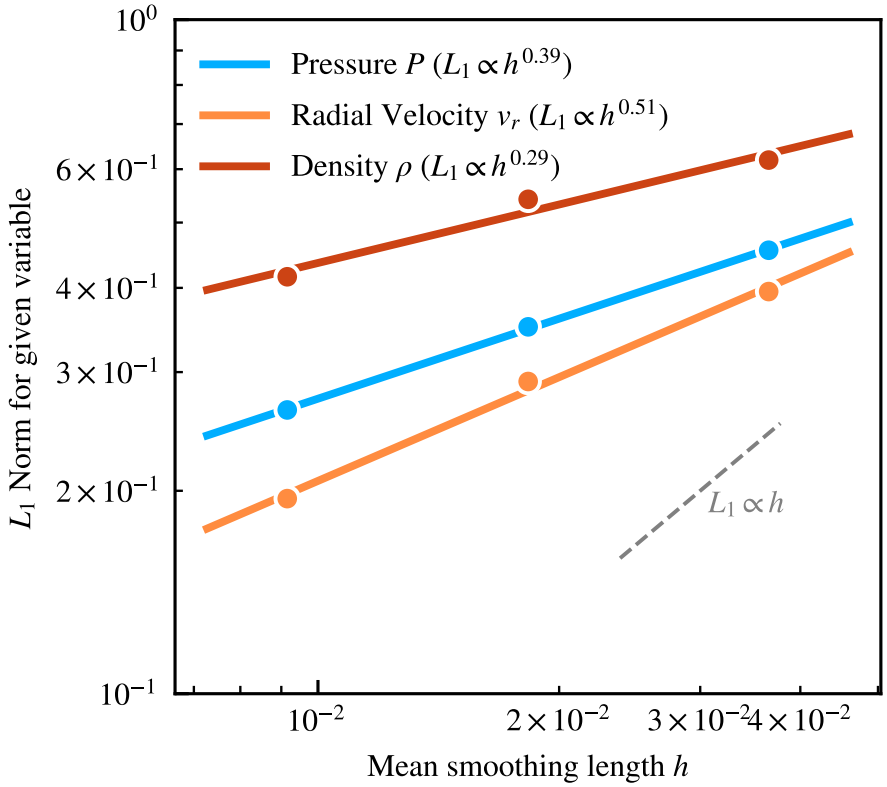


Figure 3.6: L_1 Convergence with mean smoothing length for various particle fields in the Sedov-Taylor blastwave test, measured at $t = 0.1$ against the analytic solution within the purple band of Fig. 3.5. Each set of points shows a measured value from an individual simulation, with the lines showing a linear fit to the data in logarithmic space. Dotted lines for the simulation without conduction are not shown as they lie exactly on top of the lines shown here.

into them. This corresponds, roughly, to a temperature jump of a factor of $\sim 10^5$ over the background medium.

Results

Fig. 3.5 shows the particle properties of the highest resolution initial condition (128^3) at $t = 0.1$ against the analytic solution. The Sphenix scheme closely matches the analytic solution in all particle fields, with the only deviation (aside from the smoothed shock front, an unavoidable consequence of using an SPH scheme) being a slight upturn in pressure in the central region (due to a small amount of conduction in this region). Of particular note is the position of the shock front matching exactly with the analytic solution, showing that the scheme conserves energy in this highly challenging situation thanks to the explicitly symmetric artificial viscosity equation of motion. The Sphenix scheme shows qualitatively similar results to the PHANTOM scheme on this problem (Price et al., 2018).

SPH schemes in general struggle to show good convergence on shock problems due to their inherent

discontinuous nature. Ideal convergence for shocks with the artificial viscosity set-up used in Sphenix is only first order (i.e. $L_1 \propto h$).

Fig. 3.6 shows the L_1 convergence for various fields in the Sedov-Taylor blastwave as a function of mean smoothing length. Convergence here has a best-case of $L_1(v) \propto h^{1/2}$ in real terms, much slower than the expected $L_1 \propto h^{-1}$. This is primarily due to the way that the convergence is measured; the shock front is not resolved instantaneously (i.e. there is a rise in density and velocity over some small distance, reaching the maximum value at the true position) at the same position as in the analytic solution. However, all resolution levels show an accurately placed shock front and a shock width that scales linearly with resolution (see Appendix A.4 for more information and an additional visualisation to demonstrate the convergence of the shock width).

3.5.3 Gresho-Chan Vortex

The Gresho-Chan vortex (Gresho & Chan, 1990) is typically used to test for the conservation of vorticity and angular momentum, and is performed here in two dimensions.

Initial Conditions

The initial conditions use a two dimensional glass file, and treat the gas with an adiabatic index $\gamma = 5/3$, constant density $\rho_0 = 1$, in a square of side-length 1. The particles are given azimuthal velocity

$$v_\phi = \begin{cases} 5r & r < 0.2 \\ 2 - 5r & 0.2 \leq r < 0.4 \\ 0 & r \geq 0.4 \end{cases} \tag{3.5.2}$$

with the pressure set so that the system is in equilibrium as

$$P_0 = \begin{cases} 5 + 12.5r^2 & r < 0.2 \\ 9 + 12.5r^2 - 20r + 4\log(5r) & 0.2 \leq r < 0.4 \\ 3 + 4\log(2) & r \geq 0.4 \end{cases} \tag{3.5.3}$$

where $r = \sqrt{x^2 + y^2}$ is the distance from the box centre.

Results

Fig. 3.7 shows the state of a high resolution (using a glass containing 512^2 particles) result after one full rotation at the peak of the vortex ($r = 0.2$, $t = 1.3$). The vortex is well supported, albeit with

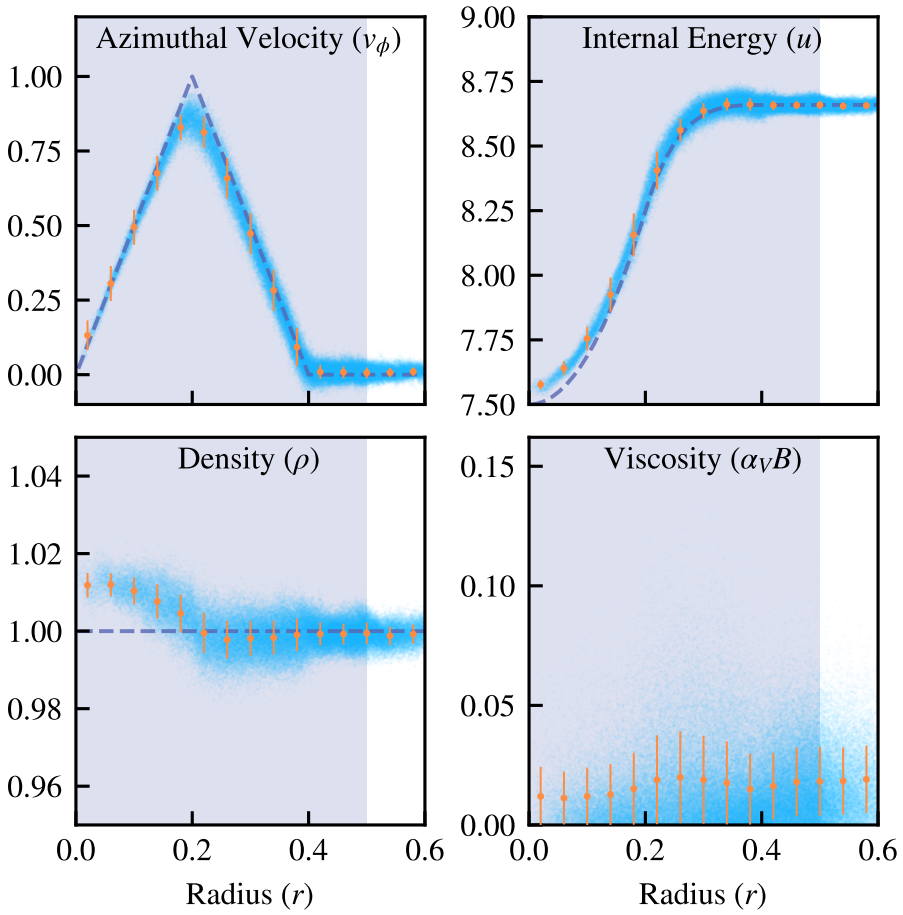


Figure 3.7: Gresho vortex at $t = 1.3$ after one rotation of the vortex peak with the Sphenix scheme using a background resolution of 512^2 and with a mach number of $\mathcal{M} = 0.33$. Here the blue points show all particles in the volume, the purple band the region used for convergence testing in Fig. 3.8, and the purple dashed line shows the analytic solution. The viscosity switch panel shows a very low maximal value (0.15) relative to the true maximum allowed by the code ($\alpha_V B = 2.0$), with the mean value (orange points with error bars indicating one standard deviation of scatter) of around 0.02 showing an excellent activation of the viscosity reducing switches throughout the Sphenix scheme.

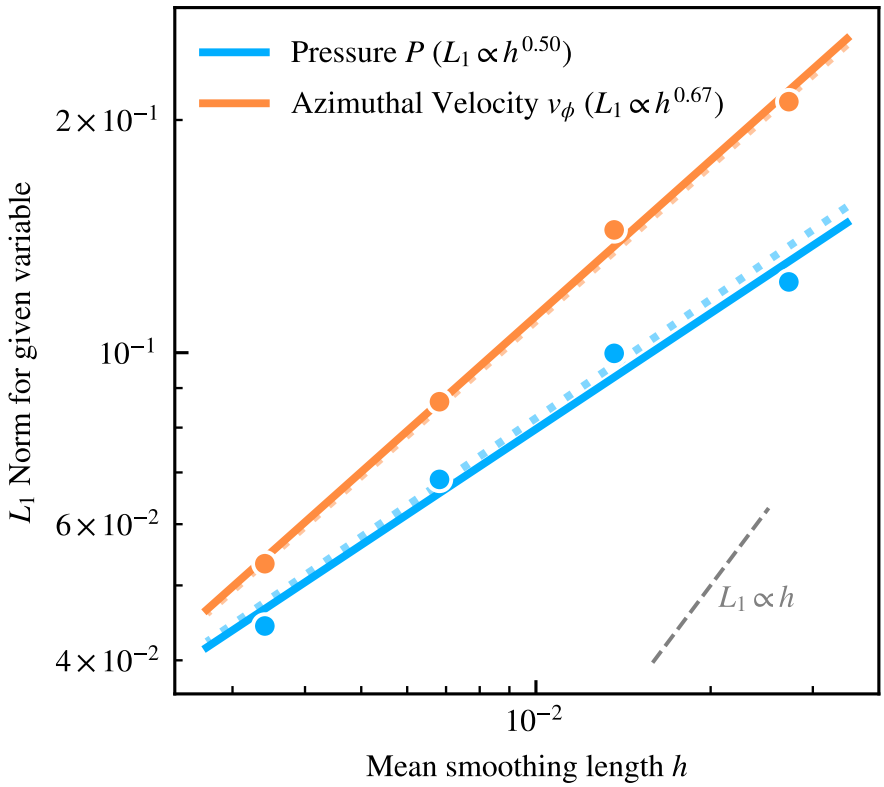


Figure 3.8: L_1 Convergence with mean smoothing length for various particle fields in the Gresho vortex test, measured against the analytic solution within the shaded region of Fig. 3.7. Each set of points shows a measured value from an individual simulation, with the lines showing a linear fit to the data in logarithmic space. The solid lines show results obtained with the full Sphenix scheme, with dotted lines showing the results with the artificial conduction scheme disabled.

some scatter, and the peak of the vortex is preserved. There has been some transfer of energy to the centre with a higher density and internal energy than the analytic solution due to the viscosity switch (shown on the bottom right) having a very small, but nonzero, value. This then allows for some of the kinetic energy to be transformed to thermal, which is slowly transported towards the centre as this is the region with the lowest thermal pressure.

Fig. 3.8 shows the convergence properties for the vortex, with the Sphenix scheme providing convergence as good as $L_1 \propto h^{0.7}$ for the azimuthal velocity. As there are no non-linear gradients in internal energy present in the simulation there is very little difference between the simulations performed with and without conduction at each resolution level due to the non-activation of Eqn. 3.3.27. This level of convergence is better than the rate seen in Dehnen & Aly (2012) implying that the Sphenix scheme, even with its less complex viscosity limiter, manages to recover some of the benefits of the more complex *Inviscid* scheme thanks to the novel combination of switches employed.

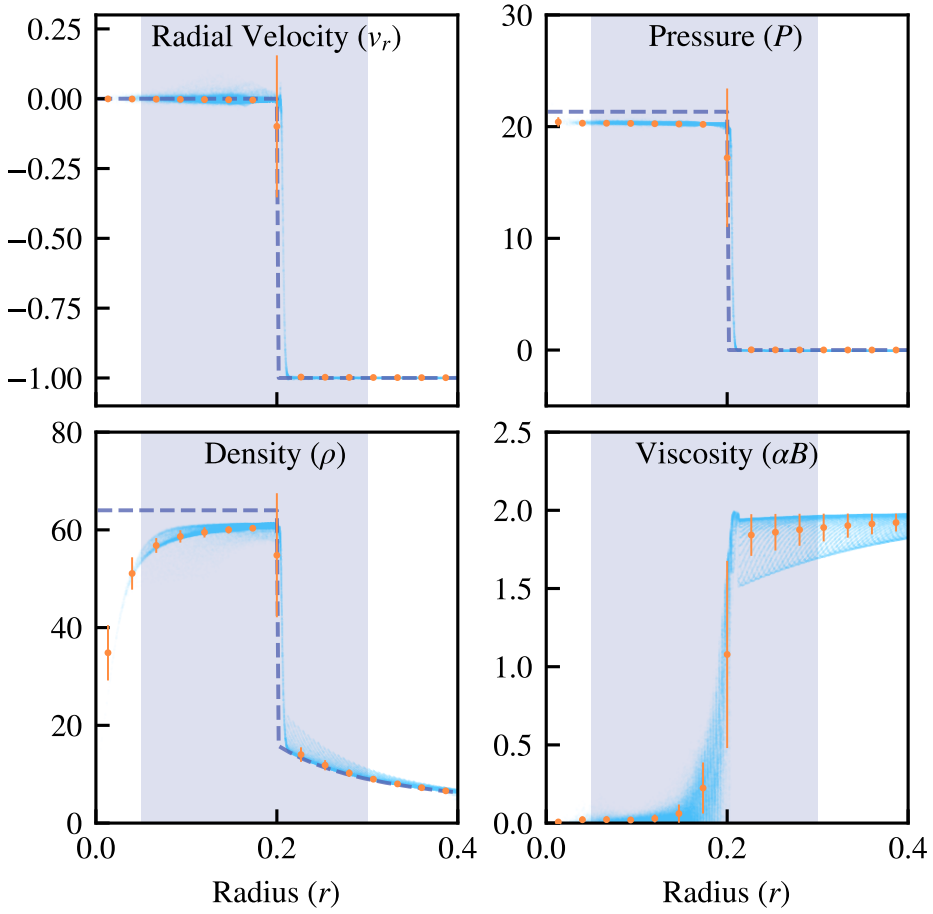


Figure 3.9: Noh problem simulation state at $t = 0.6$, showing a random sub-set of 1/100th of all of the particles plotted as blue points, the analytical solution as a dashed purple line, and binned quantities as orange points with error bars showing one standard deviation of scatter in that bin. The background shaded band shows the region considered for convergence in Fig. 3.11, with this figure showing the highest resolution simulation performed, using 512^3 particles. This simulation state is also visualised in Fig. 3.10.

3.5.4 Noh Problem

The Noh (1987) problem is known to be extremely challenging, particularly for particle-based codes, and generally requires a high particle number to correctly capture due to an unresolved convergence point. It tests a converging flow that results in a strong radial shock. This is an extreme, idealised, version of an accretion shock commonly present within galaxy formation simulations.

Initial Conditions

There are many ways to generate initial conditions, from very simple schemes to schemes that attempt to highly optimise the particle distribution (see e.g. Rosswog, 2020a). Here, we use a simple initial condition, employing a body-centred cubic lattice distribution of particles in a periodic box. The

velocity of the particles is then set such that there is a convergent flow towards the centre of the box,

$$\vec{v} = -\frac{\vec{C} - \vec{x}}{|\vec{C} - \vec{x}|} \tag{3.5.4}$$

with $\vec{C} = 0.5L(1,1,1)$, where L is the box side-length, the coordinate at the centre of the volume. This gives every particle a speed of unity, meaning those in the centre will have extremely high relative velocities. We cap the minimal value of $|\vec{C} - \vec{x}|$ to be $10^{-10}L$ to prevent a singularity at small radii.

The simulation is performed in a dimensionless co-ordinate system, with a box-size of $L = 1$.

Results

The state of the simulation is shown at time $t = 0.6$ in Fig. 3.9 and visualised in Fig. 3.10, which shows the radial velocity, which should be zero inside of the shocked region (high density in Fig. 3.10), and the same as the initial conditions (i.e. -1 everywhere) elsewhere. This behaviour is captured well, with a small amount of scatter, corresponding to the small radial variations in density shown in the image.

The profile of density as a function of radius is however less well captured, with some small waves created by oscillations in the artificial viscosity parameter (see e.g. Rosswog, 2020b, for a scheme that corrects for these errors). This can also be seen in the density slice, and is a small effect that also is possibly exacerbated by our non-perfect choice of initial conditions, but is also present in the implementation shown in Rosswog (2020a). The larger, more significant, density error is shown inside the central part of the shocked, high-density, region. This error is ever-present in SPH schemes, and is likely due to both a lack of artificial conduction in this central region (as indicated by Noh, 1987, note the excess pressure in the centre caused by 'wall heating') and the unresolved point of flow convergence.

The Noh problem converges well using Sphenix, with better than linear convergence for the radial velocity (Fig. 3.11; recall that for shocks SPH is expected to converge with $L_1 \propto h$).

This problem does not activate the artificial conduction in the Sphenix implementation because of the presence of Equation 3.3.30 reducing conductivity in highly viscous flows, as well as our somewhat conservative choice for artificial conduction coefficients (see Appendix A.5 for more details on this topic). However, as these are necessary for the practical functioning of the Sphenix scheme in galaxy formation simulations, and due to this test being highly artificial, this outcome presents little concern.

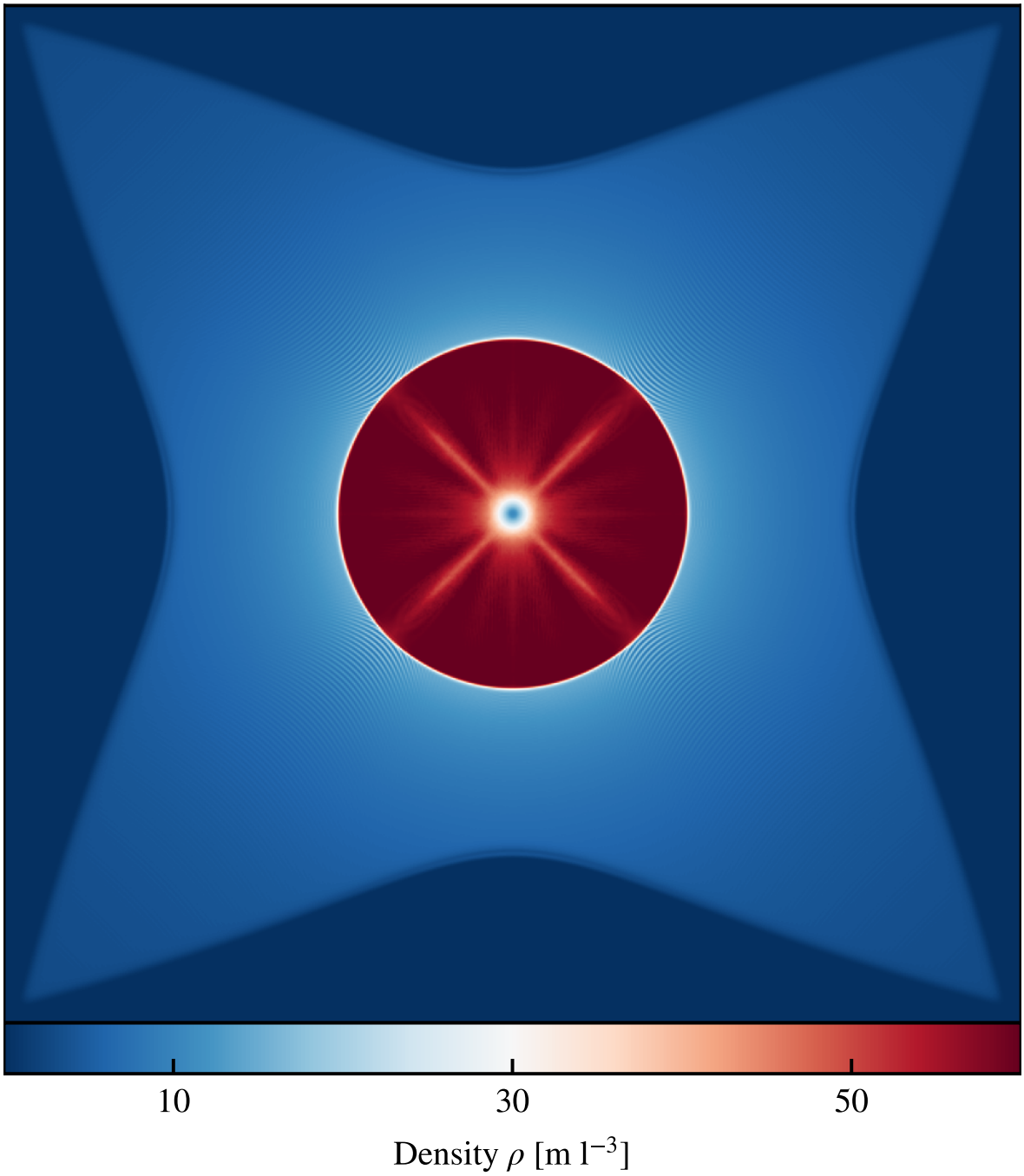


Figure 3.10: A density slice through the centre of the Noh problem at $t = 0.6$ corresponding to the particle distribution shown in Fig. 3.9. The SPHENIX scheme yields almost perfect spherical symmetry for the shock, but does not capture the expected high density in the central region, likely due to lower than required artificial conductivity (see Appendix A.5 for more information).

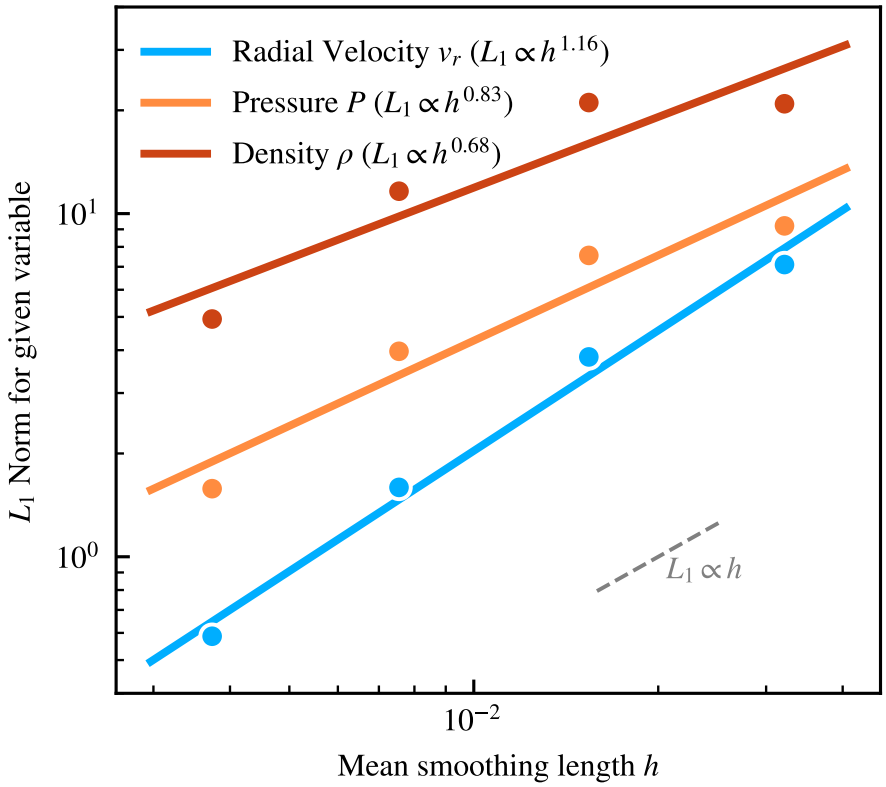


Figure 3.11: L_1 convergence test for various particle properties at $t = 0.6$ for the Noh problem, corresponding to the particle distribution shown in Fig. 3.9. The lines without conduction are not shown here as there is little difference between the with and without conduction case, due to the extremely strong shock present in this test (leading to low values of the viscosity alpha, Equation 3.3.30).

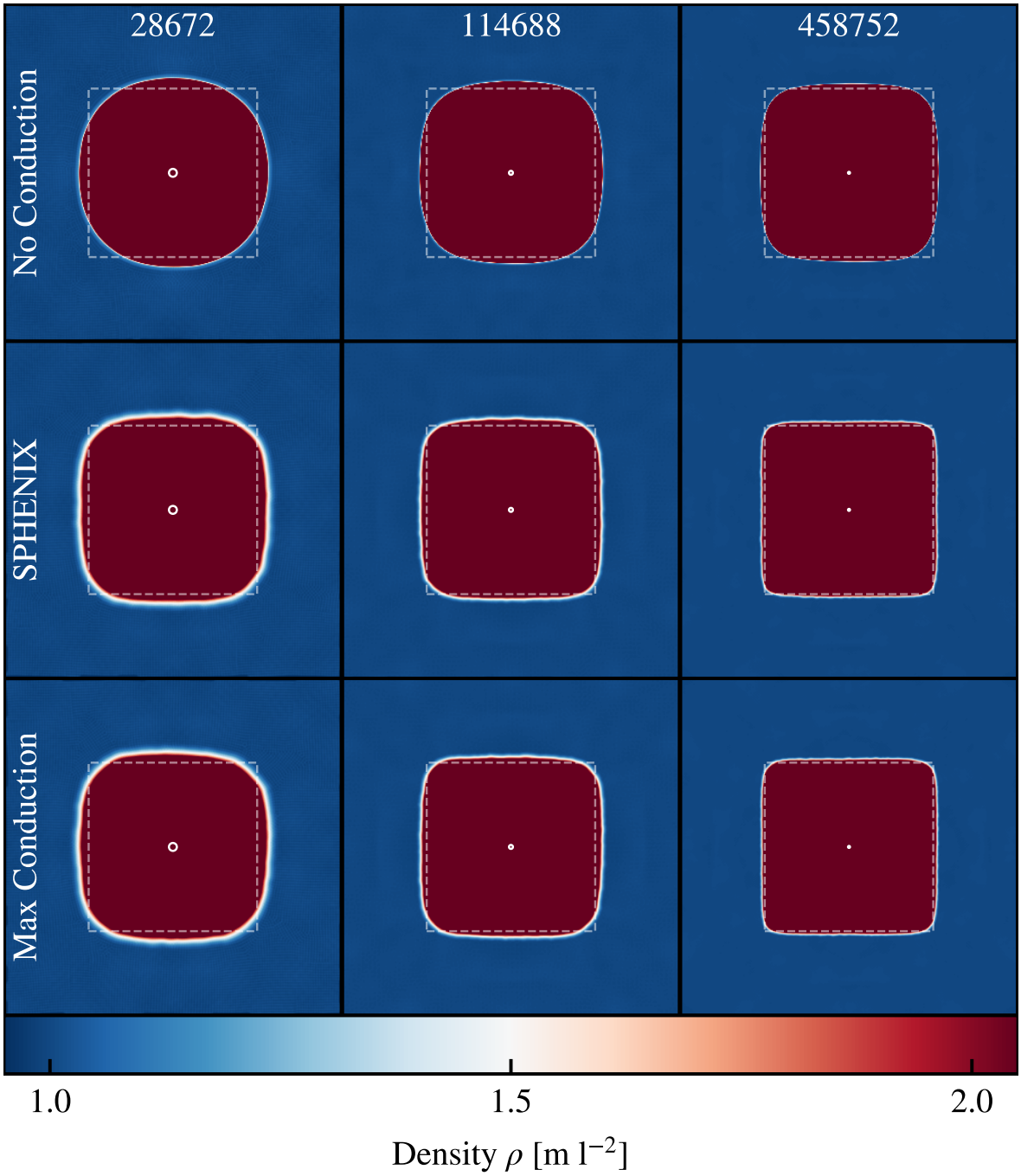


Figure 3.12: The density field for the square test at $t = 4$, shown at various resolution levels (different columns, numbers at the top denote the number of particles in the system) and with various modifications to the underlying SPH scheme (different rows). The dashed line shows the initial boundary of the square that would be maintained with a perfect scheme due to the uniform pressure throughout. The white circle at the centre of the square shows a typical smoothing length for this resolution level. Vertically, the scheme with no conduction is shown at the top, with the Sphenix scheme in the middle and a scheme with the conduction coefficient set to the maximum level throughout at the bottom. The schemes with conduction maintain the square shape significantly better than the one without conduction, and the Sphenix limiters manage to provide the appropriate amount of conduction to return to the same result as the maximum conduction case.

3.5.5 Square Test

The square test, first presented in Saitoh & Makino (2013), is a particularly challenging test for schemes like Sphenix that do not use a smoothed pressure in their equation of motion, as they typically lead to an artificial surface tension at contact discontinuities (the same ones that lead to the pressure blip in §3.5.1). This test is a more challenging variant of the ellipsoid test presented in Heß & Springel (2010), as the square includes sharp corners which are more challenging for conduction schemes to capture.

Initial conditions

The initial conditions are generated using equal mass particles. We set up a grid in 2D space with $n \times n$ particles, in a box of size $L = 1$. The central 0.5×0.5 square is set to have a density of $\rho_C = 4.0$, and so is replaced with a grid with $2n \times 2n$ particles, with the outer region having $\rho_O = 1.0$. The pressures are set to be equal with $P_C = P_O = 1.0$, with this enforced by setting the internal energies of the particles to their appropriate values. All particles are set to be completely stationary in the initial conditions with $\vec{v} = 0$. The initial conditions are not allowed to relax in any way.

Results

Fig. 3.12 shows the square test at $t = 4$ for four different resolution levels and three different variations on the Sphenix scheme. By this time the solutions are generally very stable. The top row shows the Sphenix scheme without any artificial conduction enabled (this is achieved by setting $\alpha_{D,\max}$ to zero) and highlights the typical end state for a Density-Energy SPH scheme on this problem. Artificial surface tension leads to the square deforming and rounding to become more circular.

The bottom row shows the Sphenix scheme with the artificial conduction switch removed; here $\alpha_{D,\min}$ is set to the same value as $\alpha_{D,\max} = 1$. The artificial conduction scheme significantly reduces the rounding of the edges, with a rapid improvement as resolution increases. The rounding present here only occurs in the first few steps as the energy outside the square is transferred to the boundary region to produce a stable linear gradient in internal energy.

Finally, the central row shows the Sphenix scheme, which gives a result indistinguishable from the maximum conduction scenario. This is despite the initial value for the conduction coefficient $\alpha_D = 0$, meaning it must ramp up rapidly to achieve such a similar result. The Sphenix result here shows that the choices for the conduction coefficients determined from the Sod tube (§3.5.1) are not only appropriate for that test, but apply more generally to problems that aim to capture contact discontinuities.

3.5.6 2D Kelvin-Helmholtz Instability

The two dimensional Kelvin-Helmholtz instability is presented below. This test is a notable variant on the usual Kelvin-Helmholtz test as it includes a density jump at constant pressure (i.e. yet another contact discontinuity). This version of the Kelvin-Helmholtz instability is performed in two dimensions. It is possible to design a three dimensional version of this specific test, but we choose to show it in two dimensions to more clearly show the vortex structure (as in three dimensions other layers of particles will wash out small-scale structure in projection). A recent, significantly more detailed, study of Kelvin-Helmholtz instabilities within SPH is available in Tricco (2019). In this section we focus on qualitative comparisons and how the behaviour of the instability changes with resolution within SPHENIX.

Initial conditions

The initial conditions presented here are similar to those in Price (2008), where they discuss the impacts more generally of the inclusion of artificial conduction on fluid mixing instabilities. This is set up in a periodic box of length $L = 1$, with the central band between $0.25 < y < 0.75$ set to $\rho_C = 2$ and $v_{C,x} = 0.5$, with the outer region having $\rho_O = 1$ and $v_{O,x} = -0.5$ to set up a shear flow. The pressure $P_C = P_O = 2.5$ is enforced by setting the internal energies of the equal mass particles. Particles are initially placed on a grid with equal separations. This is the most challenging version of this test for SPH schemes to capture as it includes a perfectly sharp contact discontinuity; see Agertz et al. (2007) for more information.

We then excite a specific mode of the instability, as in typical SPH simulations un-seeded instabilities are dominated by noise and are both unpredictable and unphysical, preventing comparison between schemes.

Results

Fig. 3.13 shows the simulation after various multiples of the Kelvin-Helmholtz timescale for the excited instability, with $\tau_{\rm KH}$ given by

$$\tau_{\rm KH} = \frac{(1+\chi)\lambda}{\bar{v}\sqrt{\chi}} \tag{3.5.5}$$

where $\chi = \rho_C/\rho_O = 2$ is the density contrast, $\bar{v} = v_{I,x} - v_{O,x} = 1$ the shear velocity, and $\lambda = 0.5$ the wavelength of the seed perturbation along the horizontal axis (e.g Hu et al., 2014). The figure shows three initial resolution levels, increasing from left to right. Despite this being the most challenging version of the Kelvin-Helmholtz test (at this density contrast) for a Density-Energy based SPH scheme, the

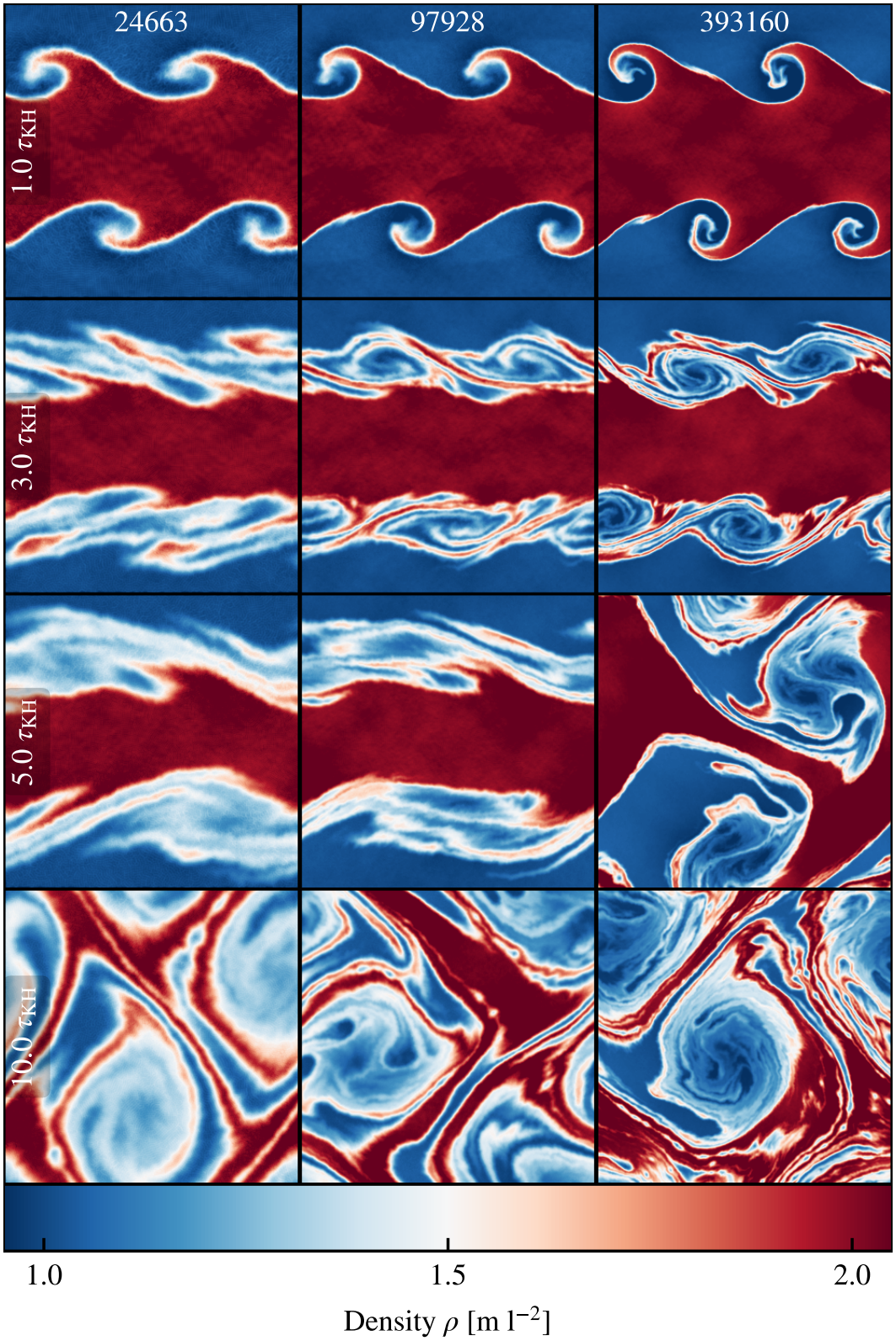


Figure 3.13: Density map of the standard Kelvin-Helmholtz 2D test at various resolutions (different columns, with the number of particles in the volume at the top) and at various times (different rows showing times from $t = \tau_{KH}$ to $t = 10\tau_{KH}$). Despite this being a challenging test for SPH, the instability is captured well at all resolutions, with higher resolution levels capturing finer details.

instability is captured well at all resolutions, with higher resolutions allowing for more rolls of the 'swirl' to be captured. In particular, the late-time highly mixed state shows that with the conduction removed after a linear gradient in internal energy has been established, the Sphenix scheme manages to preserve the initial contact discontinuity well.

The non-linear growth rate of the swirls is resolution dependent within this test, with higher-resolution simulations showing faster growth of the largest-scale modes. This is due to better capturing of the energy initially injected to perturb the volume to produce the main instability, with higher resolutions showing lower viscous losses.

Fig. 3.14 shows a different initial condition where the density contrast is $\chi = 8$, four times higher than the one initially presented. Because SPH is fundamentally a finite mass method, and we use equal-mass particles throughout, this is a particularly challenging test as the low-density region is resolved by so few particles. Here we also excite an instability with a wavelength $\lambda = 0.125$, four times smaller than the one used for the $\chi = 2$ test. This value is chosen for two reasons; it is customary to lower the wavelength of the seeded instability as the density contrast is increased when grid codes perform this test as it allows each instability to be captured with the same number of cells at a given resolution level; and also to ensure that this test is as challenging as is practical for the scheme.

Sphenix struggles to capture the instability at very low resolutions primarily due to the lack of particles in the low-density flow (an issue also encountered by Price, 2008). In the boundary region the artificial conduction erases the small-scale instabilities on a timescale shorter than their formation timescale (as the boundary region is so large) and as such they cannot grow efficiently. As the resolution increases, however, Sphenix is able to better capture the linear evolution of the instability, even capturing turn-overs and the beginning of nonlinear evolution for the highest resolution.

3.5.7 Blob Test

The Blob test is a challenging test for SPH schemes (see Klein et al., 1994; Springel, 2005) and aims to replicate a scenario where a cold blob of gas falls through the hot IGM/CGM surrounding a galaxy. In this test, a dense sphere of cold gas is placed in a hot, low density, and supersonic wind. Ideally, the blob should break up and dissolve into the wind, but Agertz et al. (2007) showed that the inability of traditional SPH schemes to exchange entropy between particles prevents this from occurring. The correct, specific, rate at which the blob should mix with its surroundings, as well as the structure of the blob whilst it is breaking up, are unknown.

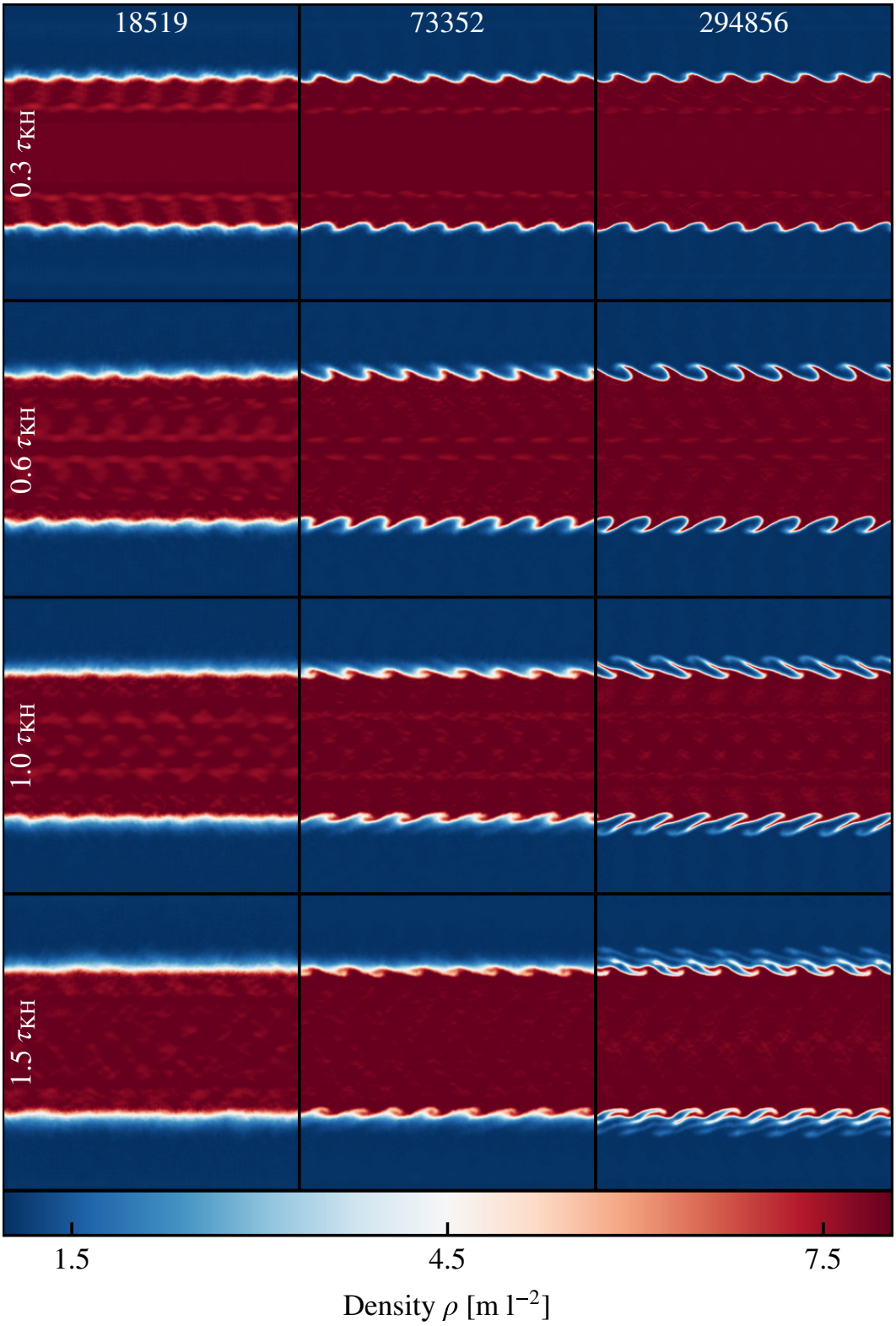


Figure 3.14: The same as Fig. 3.13, but this time using initial conditions with a significantly higher (1:8 instead of 1:2) density contrast. The initial instabilities are captured well at all resolution levels, but at the lowest level they are rapidly mixed by the artificial conduction scheme due to the lack of resolution elements in the low-density region.

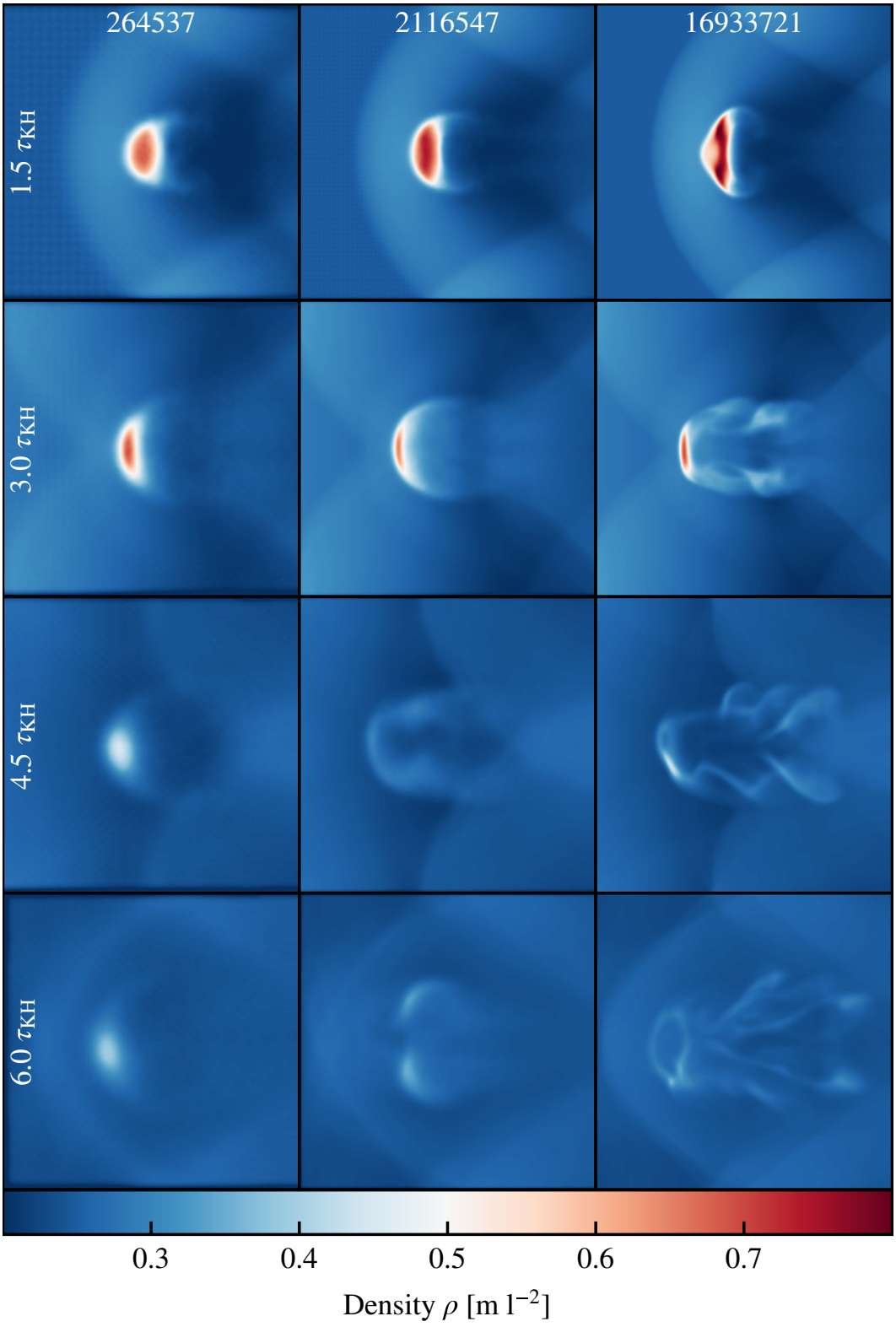


Figure 3.15: Time-evolution of the blob within the supersonic wind at various resolution levels (different columns; the number of particles in the whole volume is noted at the top) and at various times (expressed as a function of the Kelvin-Helmholtz time for the whole blob τ_{KH}; different rows). The projected density is shown here to enable all layers of the three dimensional structure to be seen. At all resolution levels the blob mixes with the surrounding medium (and importantly mixes phases with the surrounding medium), with higher resolution simulations displaying more thermal instabilities that promote the breaking up of the blob.

Initial Conditions

There are many methods to set up the initial conditions for the Blob test, including some that excite an instability to ensure that the blob breaks up reliably (such as those used in Hu et al., 2014). Here we excite no such instabilities and simply allow the simulation to proceed from a very basic particle set-up with a perfectly sharp contact discontinuity. The initial conditions are dimensionless in nature, as the problem is only specified in terms of the Mach number of the background medium and the blob density contrast.

To set up the initial particle distribution, we use two body centred cubic lattices, one packed at a high-density (for the blob, $\rho_{\rm blob} = 10$) and one at low density (for the background medium, $\rho_{\rm bg} = 1$). The low-density lattice is tiled four times in the x direction to make a box of size $4 \times 1 \times 1$, and at $[0.5, 0.5, 0.5]$ a sphere of radius 0.1 is removed and filled in with particles from the high-density lattice. The particles in the background region are given a velocity of $v_{\rm bg} = 2.7$ (with the blob being stationary), and the internal energy of the gas everywhere is scaled such that the background medium has a mach number of $\mathcal{M} = 2.7$ and the system is in pressure equilibrium everywhere.

Results

The blob is shown at a number of resolution levels at various times in Fig. 3.15. At all resolution levels the blob mixes well with the background medium after a few Kelvin-Helmholtz timescales (see Eqn. 3.5.5 for how this is calculated; here we assume that the wavelength of the perturbation is the radius of the blob)[7]. The rate of mixing is consistent amongst all resolution levels, implying that the artificial conduction scheme is accurately capturing unresolved mixing at lower resolutions.

The rate of mixing of the blob is broadly consistent with that of modern SPH schemes and grid codes, however our set of initial conditions appear to mix slightly slower (taking around $\sim 4 - 6\tau_{\rm KH}$ to fully mix) than those used by other contemporary works (Agertz et al., 2007; Read & Hayfield, 2012; Hu et al., 2014), possibly due to the lack of perturbation seeding (see Read et al., 2010, Appendix B for more details). When testing these initial conditions with a scheme that involves a Riemann solver or a Pressure-based scheme (see Appendix A.6) the rate of mixing is qualitatively similar to the one presented here. Sphenix is unable to fully capture the crushing of the blob from the centre outwards seen in grid codes and other SPH formulations using different force expressions (Wadsley et al., 2017), rather preferring to retain a 'plate' of dense gas at the initial point of the blob that takes longer to break up.

[7] Note that here the Kelvin-Helmholtz timescale is 1.1 times the cloud crushing timescale (Agertz et al., 2007).

At low resolutions it is extremely challenging for the method to capture the break-up of the blob as there are very few particles in the background medium to interact with the blob due to the factor of 10 density contrast.

3.5.8 Evrard Collapse

The Evrard collapse (Evrard, 1988) test takes a large sphere of self-gravitating gas, at low energy and density, that collapses in on itself, causing an outward moving accretion shock. This test is of particular interest for cosmological and astrophysical applications as it allows for the inspection of the coupling between the gravity and hydrodynamics solver.

Initial Conditions

Gas particles are set up in a sphere with an adiabatic index of $\gamma = 5/3$, sphere mass $M = 1$, sphere radius $R = 1$, initial density profile $\rho(r) = 1/2\pi r$, and in a very cold state with $u = 0.05$, with the gravitational constant $G = 1$. These initial conditions are created in a box of size 100, ensuring that effects from the periodic boundary are negligible. Unfortunately, due to the non-uniform density profile, it is considerably more challenging to provide relaxed initial conditions (or use a glass file). Here, positions are simply drawn randomly to produce the required density profile.

The Evrard collapse was performed at four resolution levels, with total particle numbers in the sphere being 10^4, 10^5, 10^6, and 10^7. The gravitational softening was fixed at 0.001 for the 10^6 resolution level, and this was scaled with $m^{-1/3}$ with m the particle mass for the other resolution levels. The simulations were performed once with artificial conduction enabled (the full Sphenix scheme), and once with it disabled.

Results

The highest resolution result (10^7 particles) with the full Sphenix scheme is shown in Fig. 3.16. This is compared against a high resolution grid code[8] simulation performed in 1D, and here Sphenix shows an excellent match to the reference solution. The shock at around $r = 10^{-1}$ is sharply resolved in all variables, and the density and velocity profiles show excellent agreement. In the centre of the sphere, there is a slight deviation from the reference solution for the internal energy and density (balanced to accurately capture the pressure in this region) that remains even in the simulation performed without

[8] `HydroCode1D`, see `https://github.com/bwvdnbro/HydroCode1D` and the Swift repository for more details.

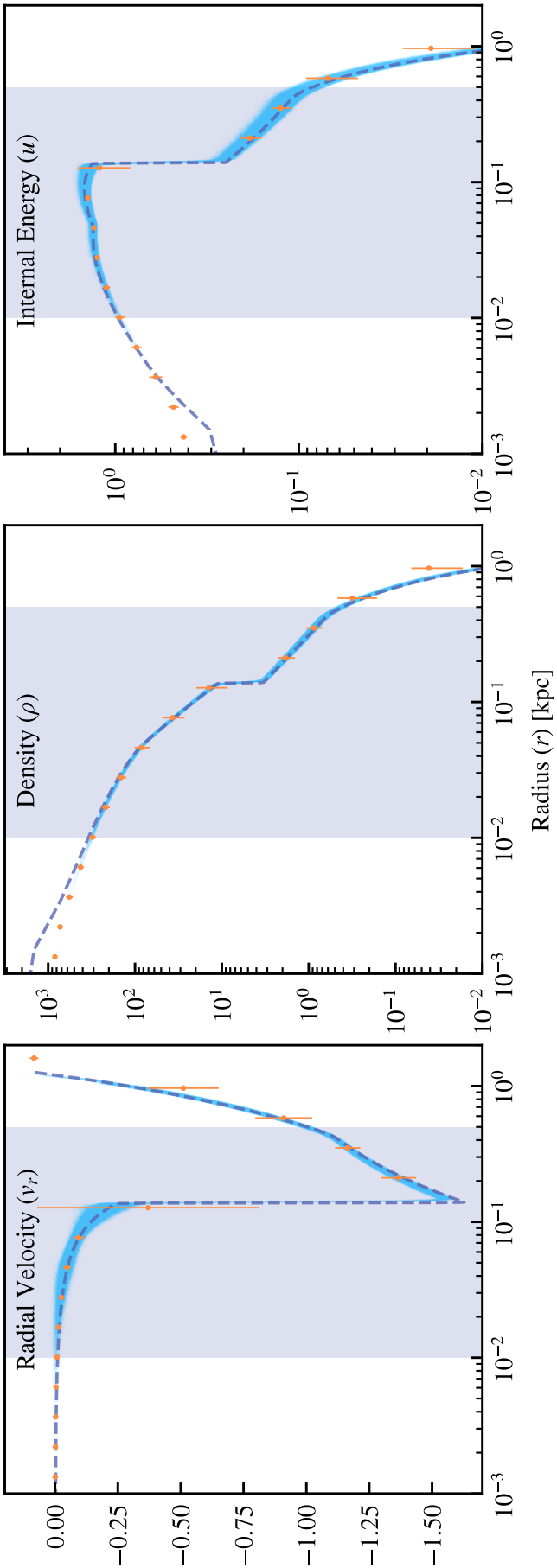


Figure 3.16: State of the Evrard sphere at $t = 0.8$ for a resolution of 10^7 particles. A random sub-set of 1/10th of the particles is shown in blue, with the solution from a high resolution 1D grid code shown as a purple dashed line. The orange points with error bars show the median within a radial equally log-spaced bin with the bar showing one standard deviation of scatter. The shaded band in the background shows the region considered for the convergence test in Fig. 3.17.

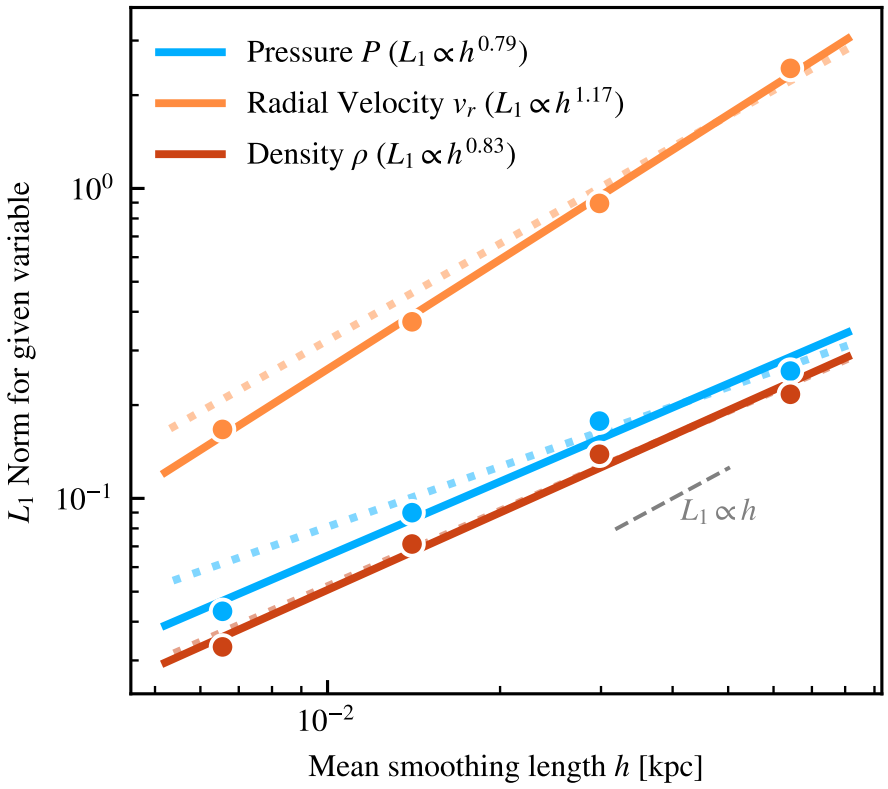


Figure 3.17: L_1 convergence for various gas properties for the Evrard collapse sphere at $t = 0.8$. The region considered for convergence here is the purple band shown in Fig. 3.16. The SPHENIX scheme is shown with the points and linear fits in solid, and the same scheme is shown with artificial conduction turned off as dotted lines. Artificial conduction significantly improves convergence here as it helps stabilise the thermal properties of the initially randomly placed particles.

artificial conduction (omitted for brevity, as the simulation without conduction shows similar results to the simulation with conduction, with the exception of the conduction reducing scatter in the internal energy profile). This is believed to be an artefact of the initial conditions, however it was not remedied by performing simulations at higher resolutions.

The convergence properties of the Evrard sphere are shown in Fig. 3.17. The velocity profile shows a particularly excellent result, with greater than linear convergence demonstrated. The thermodynamic properties show roughly linear convergence. Of particular note is the difference between the convergence properties of the simulations with and without artificial conduction; those with this feature of Sphenix enabled converge at a more rapid rate. This is primarily due to the stabilising effect of the conduction on the internal energy profile. As the particles are initially placed randomly, there is some scatter in the local density field at all radii. This is quickly removed by adiabatic expansion in favour of scatter in the internal energy profile, which can be stabilised by the artificial conduction.

3.5.9 nIFTy Cluster

The nIFTy cluster comparison project, Sembolini et al. (2016), uses a (non-radiative, cosmological) cluster-zoom simulation to evaluate the efficacy of various hydrodynamics and gravity solvers. The original paper compared various types of schemes, from traditional SPH (Gadget, Springel, 2005) to a finite volume adaptive mesh refinement scheme (RAMSES, Teyssier, 2002). The true answer for this simulation is unknown, but it is a useful case to study the different characteristics of various hydrodynamics solvers.

In Fig. 3.18 the Sphenix scheme is shown with and without artificial conduction against three reference schemes from Sembolini et al. (2016). Here, the centre the cluster was found using the VELOCIraptor (Elahi et al., 2019) friends-of-friends halo finder, and the particle with the minimum gravitational potential was used as the reference point.

The gas density profile was created using 25 equally log-spaced radial bins, with the density calculated as the sum of the mass within a shell divided by the shell volume. The Sphenix scheme shows a similar low-density core as AREPO, with the no conduction scheme resulting in a cored density profile similar to the traditional SPH scheme from Sembolini et al. (2016).

The central panel of Fig. 3.18 shows the 'entropy' profile of the cluster; this is calculated as $T n_{\rm e}^{-2/3}$ with $n_{\rm e}$ the electron density (assuming primordial gas, this is $n_{\rm e} = 0.875\rho/m_{\rm H}$ with $m_{\rm H}$ the mass of a hydrogen atom) and T the gas temperature. Each was calculated individually in the same equally log-spaced bins as the density profile, with the temperature calculated as the mass-weighted temperature

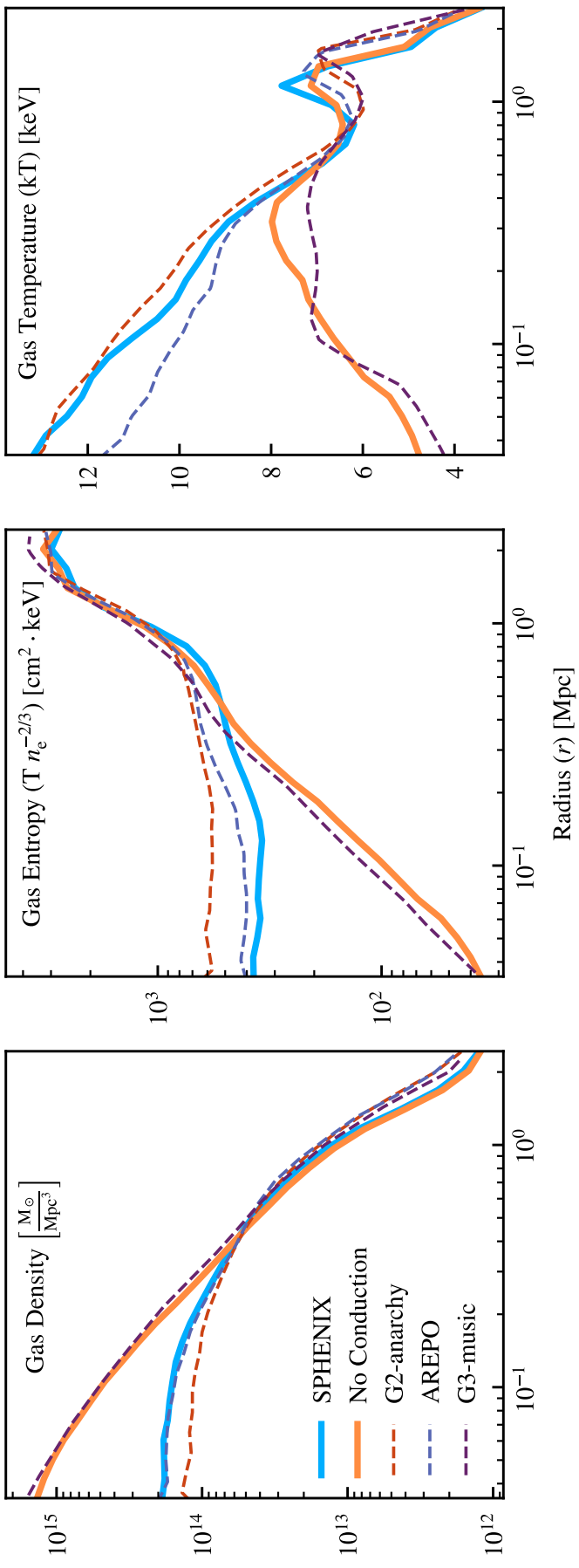


Figure 3.18: Thermodynamics profiles for the nIFTy cluster at $z = 0$ with five codes and schemes. The solid lines show those simulated with Swift, with the blue line showing the full Sphenix scheme, and the orange line showing Sphenix without artificial conduction . The dashed lines were extracted directly from Sembolini et al. (2016) and show a modern Pressure-Entropy scheme (G2-anarchy; Schaye et al., 2015, appendix A), a moving mesh finite volume scheme (AREPO; Springel, 2010), and a traditional SPH scheme (G3-music; Springel, 2005).

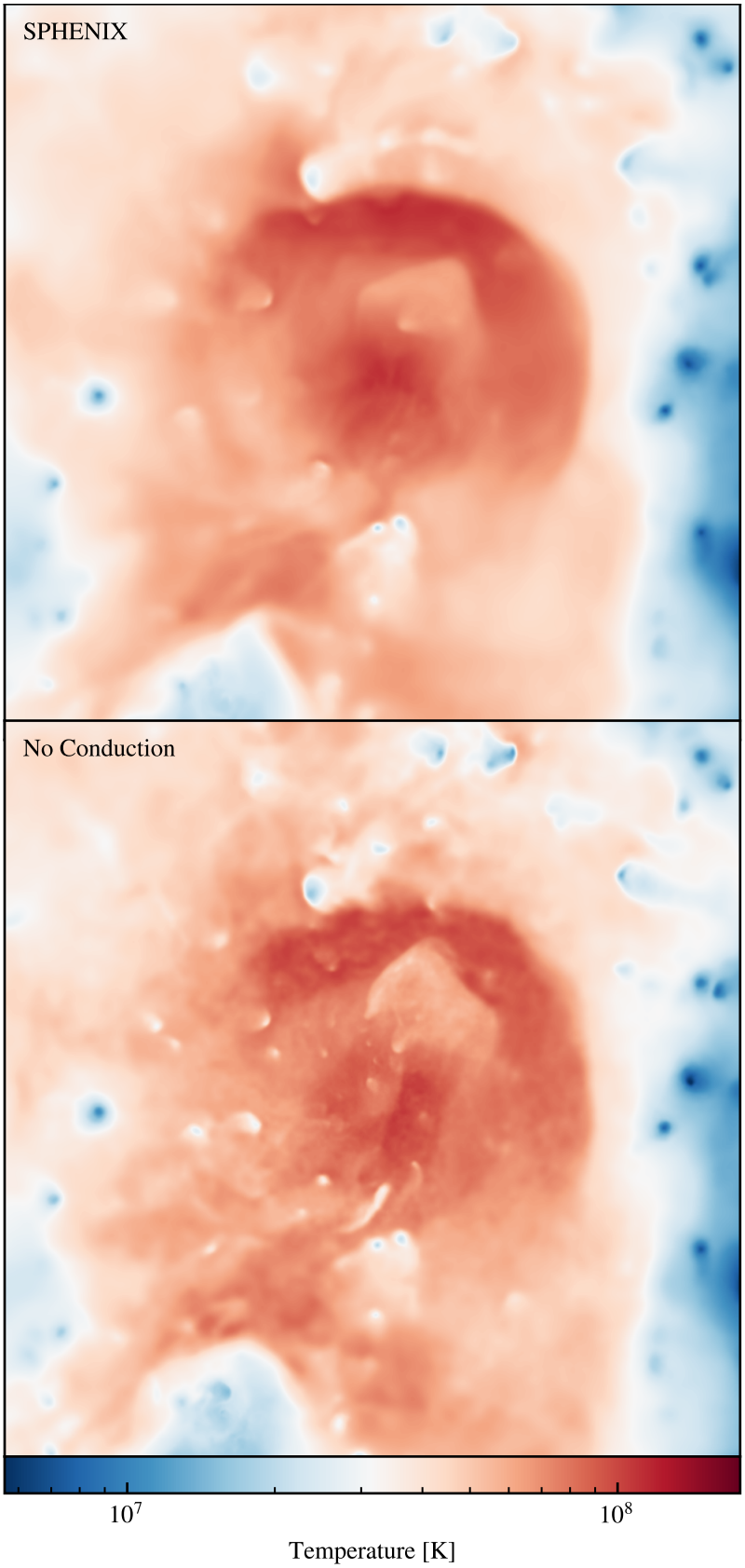


Figure 3.19: Image of the nIFTY cluster, as a projected mass-weighted temperature map, shown for the Sphenix scheme with (top) and without artificial conduction enabled (bottom). The image shows a 5 Mpc wide view, centred on the most bound particle in the halo.

within that shell. The rightmost panel shows this mass-weighted temperature profile, with Sphenix showing slightly higher temperatures in the central region than AREPO, matching G2-anarchy instead. This high-temperature central region, along with a low-density centre, lead to the 'cored' (i.e. flat, with high values of entropy, at small radii) entropy profile for Sphenix.

The cored central entropy profile with Sphenix is attained primarily due to the artificial conduction scheme and is not due to the other improvements over the traditional SPH base scheme (including for example the artificial viscosity implementation). We note again that there was no attempt to calibrate the artificial conduction scheme to attain this result on the nIFTy cluster, and any and all parameter choices were made solely based on the Sod shock tube in §3.5.1.

In Fig. 3.19, a projected mass-weighted temperature image of the cluster is shown. The image demonstrates how the artificial conduction present in the Sphenix scheme promotes phase mixing, resulting in the cored entropy profile demonstrated in Fig. 3.18.

The temperature distribution in the SPH simulation without conduction appears noisier, due to particles with drastically different phases being present within the same kernel. This shows how artificial conduction can lead to sharper shock capture as the particle distribution is less susceptible to this noise, enabling a cleaner energy transition between the pre- and post-shock region.

3.6 Conclusions

We have presented the Sphenix SPH scheme and its performance on seven hydrodynamics tests. The scheme has been demonstrated to show convergent (with resolution) behaviour on all these tests. In summary:

- Sphenix is an SPH scheme that uses Density-Energy SPH as a base, with added artificial viscosity for shock capturing and artificial conduction to reduce errors at contact discontinuities and to promote phase mixing.

- A novel artificial conduction limiter allows Sphenix to be used with energy injection feedback schemes (such as those used in EAGLE) by reducing conduction across shocks and other regions where the artificial viscosity is activated.

- The artificial viscosity and conduction scheme coefficients were determined by ensuring good performance on the Sod Shock tube test, and remain fixed for all other tests.

- The modified Inviscid SPH (Cullen & Dehnen, 2010) scheme captures strong shocks well, ensuring energy conservation, as shown by the Sedov-Taylor blastwave test, but the smooth nature of SPH prevents rapid convergence with resolution.

- The use of the Balsara (1989) switch in SPHENIX was shown to be adequate to ensure that the Gresho-Chan vortex is stable. Convergence on this test was shown to be faster than in Cullen & Dehnen (2010).

- The artificial conduction scheme was shown to work adequately to capture thermal instabilities in both the Kelvin-Helmholtz and Blob tests, with contact discontinuities well preserved when required.

- SPHENIX performed well on both the Evrard collapse and nIFTY cluster problems, showing that it can couple to the FMM gravity solver in SWIFT and that the artificial conduction scheme can allow for entropy cores in clusters.

- SPHENIX is implemented in the SWIFT code and is available fully open source to the community.

SPHENIX hence achieves its design goals; the Lagrangian nature of the scheme allows for excellent coupling with gravity; the artificial conduction limiter allows the injection of energy as in the EAGLE sub-grid physics model; and the low cost-per-particle and lack of matrices carried on a particle-by-particle basis provide for a very limited computational cost (see Borrow et al., 2019, for a comparison of computational costs between a scheme like SPHENIX and the GIZMO-like schemes also present in SWIFT).

Chapter 4

Baryon Redistribution & Inter-Lagrangian Transfer

4.1 Introduction

It has been recognized that feedback processes from the formation of stars and black holes have an important effect on the resulting observable baryonic component, though they have a small effect on the collisionless dark matter. Such feedback often takes the form of large-scale winds that eject substantial amounts of gas from galaxies due to energetic input from young stars, supernovae, and active galactic nuclei (AGN). This gas can then be deposited far out in the intergalactic medium (IGM), remain as halo gas in the Circumgalactic Medium (CGM), or be re-accreted in 'wind recycling' (Oppenheimer et al., 2010; Christensen et al., 2016; Anglés-Alcázar et al., 2017b; Hafen et al., 2019b; Christensen et al., 2018). This cycling of baryons is an integral part of modern galaxy formation theory, and is believed to be a key factor in establishing the observed properties of both galaxies and intergalactic gas (Somerville et al., 2015). Despite playing a critical role in regulating galaxy growth (Naab & Ostriker, 2017), feedback remains poorly understood. These models must prevent too much star formation, as well as the 'overcooling problem', suffered by the earliest hydrodynamical simulations (Davé et al., 2001; Balogh et al., 2001).

Feedback processes transport baryons far from their originating dark matter haloes. Early observational evidence for this was that the diffuse intergalactic medium at high redshift is enriched with metals produced by supernovae, requiring winds with speeds of hundreds of $\mathrm{km\,s^{-1}}$ to be ejected ubiquitously (e.g. Aguirre et al., 2001; Springel & Hernquist, 2003; Oppenheimer & Davé, 2006). More recently, feedback from AGN is seen to eject ionised and molecular gas outflows with velocities exceeding 1000 $\mathrm{km\,s^{-1}}$ (e.g. Sturm, 2001; Greene et al., 2012; Maiolino et al., 2012; Zakamska

et al., 2016). It has long been known that some AGN also power jets, carrying material out at relativistic velocities (Fabian, 2012). These processes decouple the baryonic matter from the dark matter on cosmological scales, which could potentially complicate approaches to populating dark matter simulations with baryons. Hence it is important to quantify the amount of baryons that are participating in such large-scale motions, within the context of modern galaxy formation models that broadly reproduce the observed galaxy population.

This chapter thus examines the large-scale redistribution of baryons relative to the dark matter, using the Simba cosmological simulations that include kinetic feedback processes which plausibly reproduces the observed galaxy population (Davé et al., 2019). To do this, we pioneer a suite of tools to compare the initial and final location of baryons relative to their initial 'Lagrangian region', defined as the region in the initial conditions that collapses into a given dark matter halo. In classical galaxy formation theory, the baryons follow the dark matter into the halo, and only then significantly decouple thanks to radiative processes; this would result in the baryons lying mostly within the Lagrangian region of the halo. However, outflows can disrupt this process, and result in the transfer of baryons outside the Lagrangian region or even transfer *between* Lagrangian regions. It is these effects we seek to quantify in this chapter.

The importance of ejecting baryons and the resulting transfer of material to other galaxies was highlighted using recent cosmological 'zoom-in' simulations from the FIRE project (Hopkins et al., 2014, 2018). Tracking individual gas resolution elements in the simulations, Anglés-Alcázar et al. (2017b) showed that gas ejected in winds from one galaxy (often a satellite) can accrete onto another galaxy (often the central) and fuel in-situ star formation. This mechanism, dubbed 'intergalactic transfer', was found to be a significant contributor to galaxy growth. The galaxies that provided intergalactic transfer material often ended up merging with the central galaxy by $z = 0$, with the mass of gaseous material provided by galactic winds greatly exceeding the mass of interstellar medium (ISM) gas delivered via merger events. However, this work did not examine the extent to which galactic winds can push gas to larger scales and connect individual haloes at $z = 0$, since it is not feasible to examine this in zoom-in simulations that by construction focus on modelling a single halo.

In this chapter, we consider matter flows in a large cosmological volume ($50h^{-1}$Mpc) using the Simba simulations (Davé et al., 2019), whose star formation feedback employs scalings from FIRE, and whose black hole model includes various forms of AGN feedback including high-velocity jets. More generally, we present a framework for analysing the relative motion of dark matter and baryons on large scales due to hydrodynamic and feedback processes. With this, we quantify the large scale gas flows out of Lagrangian regions into the surrounding IGM and the importance of 'inter-Lagrangian

transfer' in galaxy evolution.

The remainder of this chapter is organised as follows: in §4.2, we discuss the Simba simulation suite that is used for analysis; in §4.3, we discuss a distance-based metric for the investigation of feedback strength; in §4.4, we discuss halo-level metrics based on Lagrangian regions to study inter-Lagrangian transfer; in §4.5 we discuss the convergence of the method; and in §4.6 we conclude and summarise the results.

4.2 The Simba Simulation Suite

4.2.1 Code and sub-grid model

This work uses the Simba simulation suite (Davé et al., 2019), which inherits a large amount of physics from Mufasa (Davé et al., 2016). Simba uses a variant of the GIZMO code (Hopkins, 2015), with the Meshless-Finite-Mass (MFM) hydrodynamics solver using a cubic spline kernel with 64 neighbours. The gravitational forces are solved using the Tree-PM method as described in Springel et al. (2005a) for Gadget-2, of which GIZMO is a descendent. In the $50h^{-1}$Mpc, 512^3 particle box used here, the mass resolution for the gas elements is 1.8×10^7 $M_\odot$, and for the dark matter is 9.6×10^7 $M_\odot$. The cosmology used in Simba is consistent with results from Planck Collaboration et al. (2016), with $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{m} = 0.3$, $\Omega_\mathrm{b} = 0.048$, $H_0 = 68$ km s^{-1}, $\sigma_8 = 0.82$, and $n_s = 0.97$.

On top of this base code, the Simba sub-grid model is implemented. This model is fully described in Davé et al. (2019), but it is summarised here. Radiative cooling and photoionisation are included from Grackle-3.1 (Smith et al., 2016). Stellar feedback is modelled using decoupled two-phase winds that have 30% of their ejected particles set at a temperature given by the supernova energy minus the kinetic energy of the wind. The mass loading factor of these winds scales with stellar mass using scalings from Anglés-Alcázar et al. (2017b), obtained from particle tracking in the FIRE zoom-in simulations.

Black hole growth is included in Simba using the torque-limited accretion model from Anglés-Alcázar et al. (2017a) for cold gas and Bondi (1952) accretion for the hot gas. The AGN feedback model includes both kinetic winds and X-ray feedback. At high Eddington ratios ($f_\mathrm{Edd} > 0.02$) or low black holes mass ($M_\mathrm{BH} < 10^{7.5}$ $M_\odot$), the radiative-mode winds are high mass-loaded and ejected at ISM temperature with velocities $\lesssim 10^3$ km s^{-1}. At low Eddington ratios and high black hole mass, the jet-mode winds are ejected at velocities approaching $\sim 10^4$ km s^{-1}. We refer the interested reader to the full description of this feedback model in Davé et al. (2019).

In addition to the fiducial model, we also use two comparison models. The first, described as `NoJet`, includes all of the SIMBA physics but has the high-energy black hole jet-mode winds disabled. All other star formation and AGN feedback is included. The second, described as non-radiative, uses the same initial conditions as the fiducial model but only includes gravitational dynamics and hydrodynamics, i.e. without sub-grid models. This latter simulation was performed with the SWIFT simulation code (Schaller et al., 2016) using a Density-Entropy Smoothed Particle Hydrodynamics (SPH) solver as it performs orders of magnitude faster than the original GIZMO code (Borrow et al., 2018). The use of this hydrodynamics model, over the MFM solver, will have a negligible effect on the quantities of interest in this chapter, as it has been shown that such a solver produces haloes of the same baryonic mass when ran in non-radiative mode (see e.g. Sembolini et al., 2016).

4.2.2 Defining haloes

Haloes are defined using a modified version of the Amiga Halo Finder (AHF, Gill et al., 2004; Knollmann & Knebe, 2009) presented in Muratov et al. (2015). This spherical overdensity finder determines the halo centers by using a nested grid, and then fits parameters based on the Navarro-Frenk-White (NFW, Navarro et al., 1996) profile. Here we define the virial radius, $R_{\rm vir}$, as the spherical overdensity radius retrieved from AHF, consistent with the definition in Bryan & Norman (1998). Substructure search was turned off, such that the code only returned main haloes.

4.2.3 Defining Lagrangian regions

The Lagrangian region (LR) associated with a halo is the volume in the initial conditions that contains the dark matter that will eventually collapse to form that halo.

Many methods exist for defining Lagrangian regions (see e.g. Onorbe et al., 2014, for a collection of methods). In this work the Lagrangian regions are defined in the following way:

1. Find all haloes at redshift $z = 0$, and assign them a unique halo ID.

2. For each halo, match the dark matter particles contained within it with those in the initial conditions. These particles are then assigned a Lagrangian region ID that is the same as this halo ID, with particles outside of haloes (and hence Lagrangian regions) assigned an ID of -1. This defines the initial Lagrangian regions based on the dark matter.

3. In some cases, discussed below, fill in the holes in this Lagrangian region by using a nearest-neighbour search. In the fiducial case, skip this step (see §4.5).

4. For every gas particle in the initial conditions, find the nearest dark matter neighbour. This gas particle is assigned to the same Lagrangian region as that dark matter particle.

In this way, Lagrangian regions contain all dark matter particles that end up within $R_{\rm vir}$ of each halo at $z = 0$, by definition, as well as the baryons that should also in principle collapse into the corresponding halo. In §4.5, we explore alternative definitions of LRs and their impact in our results.

4.3 Quantifying Baryon Redistribution

Feedback is a complex process that impacts a wide range of baryonic observables, from the galaxy stellar mass function, to galaxy sizes, to the density profiles of galaxies (e.g. Anglés-Alcázar et al., 2014; Nelson et al., 2015; Hellwing et al., 2016; Benítez-Llambay et al., 2018). It is interesting, therefore, to develop tools to study the global effects of feedback directly, as a complement to the many indirect constraints obtainable from comparing to astrophysical observables. Here we describe the *spread metric* as a general tool to examine the redistribution of baryons via feedback relative to the underlying dark matter distribution.

4.3.1 The Spread Metric

Our approach to quantifying the large-scale impact of feedback is to develop a simple and robust metric that directly captures the displacement of gas due to feedback. This *spread metric*, illustrated in Fig. 4.1, works as follows:

1. For every gas particle i in the initial conditions, find the nearest n dark matter neighbours j (with $n = 3$ for our fiducial results).
2. Then, in the final conditions at $z = 0$:
    - Match all remaining baryonic particles with their initial conditions progenitor (in this case, stars are matched with their gas particle progenitor).
    - Find the distance r_{ij} between particles i and j in the final conditions.
    - The spread metric for particle i, denoted S_i, is given by the *median* of the n distances at $z = 0$ to the original dark matter neighbours, r_{ij}.

The spread metric is introduced to measure the net displacement of baryons over cosmic time. This is somewhat difficult to do in practice, as to measure the net movement of particles we require a

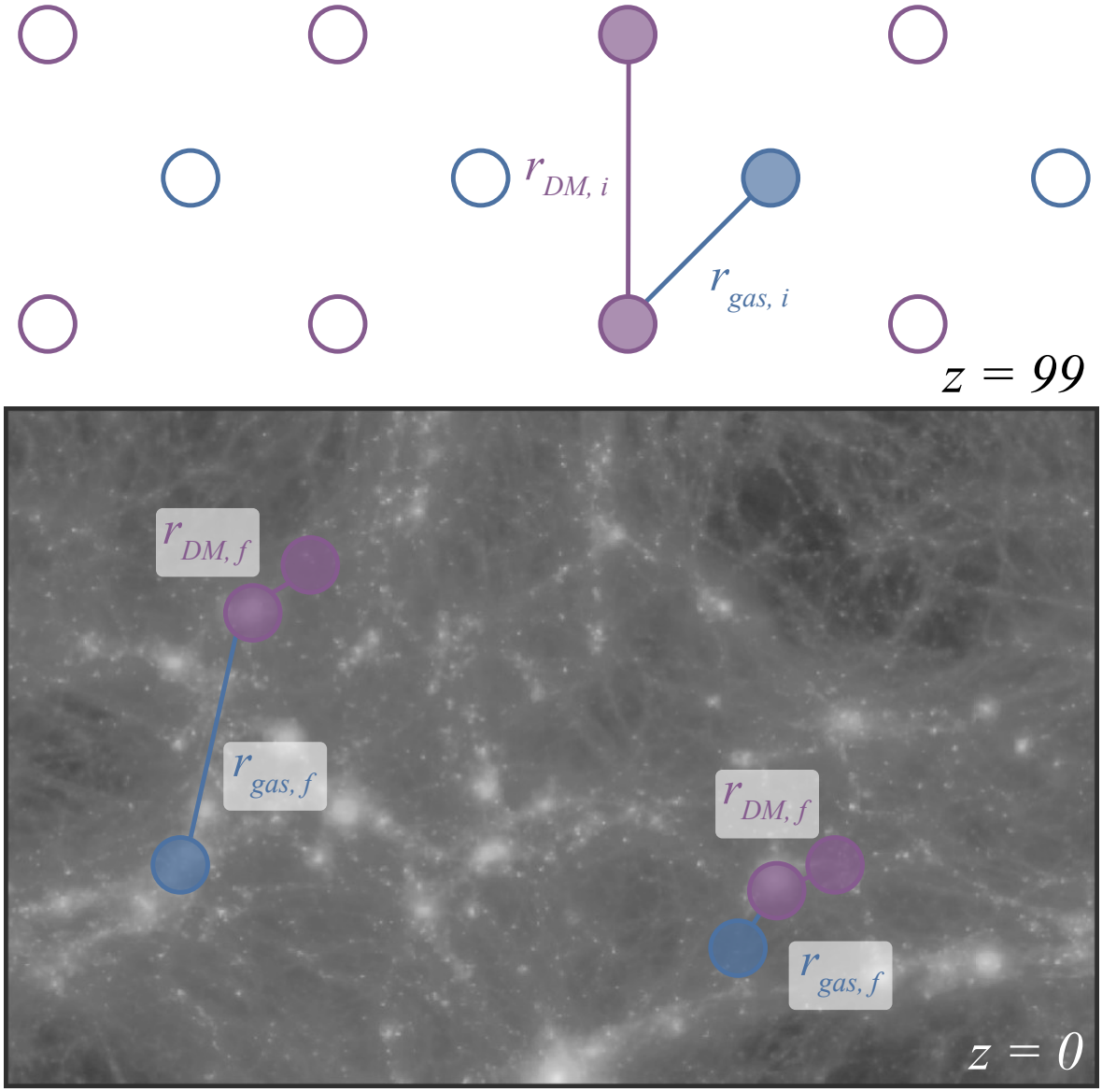


Figure 4.1: Illustration of the matching procedure between initial and final conditions to define the spread metric. Gas particles are shown in blue, with dark matter particles shown in purple. The top panel shows the $z = 99$ initial conditions, where every particle finds its nearest dark matter neighbour. The bottom panel shows the distances between two sets of example particles at $z = 0$, demonstrating that the distances to the gas and dark matter particle may or may not be similar. For our fiducial results, each particle is matched to the three nearest neighbours at $z = 99$ and the spread metric is computed as the median of the corresponding distances at $z = 0$ (see text for details).

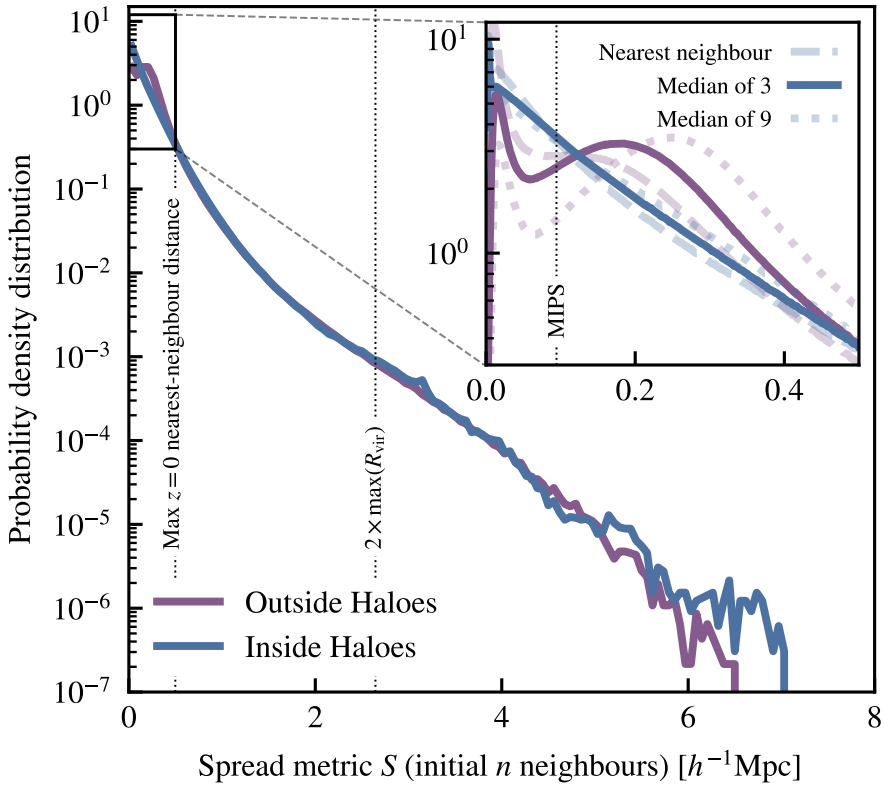


Figure 4.2: The redshift $z = 0$ spread metric distribution for the dark matter component in the full Simba model. The distribution is split between particles that lie within haloes (blue) and outside haloes (purple), with this being an approximately even split at $z = 0$. Vertical dotted lines indicate the maximal distance between any two nearest dark matter particles at $z = 0$ ($\sim 0.5h^{-1}$Mpc) and twice the maximal virial radius of any halo in the box ($\max(R_{\rm vir}) \sim 1.3h^{-1}$Mpc). The inset figure shows the inner $0.5h^{-1}$Mpc of the distribution, with the mean inter-particle separation in the initial conditions (MIPS $\sim 0.1h^{-1}$Mpc) indicated by the vertical dotted line. The fainter lines show how the spread metric changes when taking the median over a different number of initial nearest neighbours. This figure shows that initially neighbouring dark matter particles can be spread out to $7h^{-1}$Mpc due to gravitational dynamics alone.

reference point. We take that reference point to be the initially neighbouring dark matter particle as to respect the Lagrangian nature of the simulation. This is different to taking the relative motion of the particle compared to its initial point in co-moving space as it ensures that there is zero 'spread' in bulk motions.

The spread metric is presented first for dark matter in Fig. 4.2, showing the probability density distribution of the spread S for dark matter particles either inside (blue) or outside (purple) of virialized haloes at $z = 0$. This quantifies the redistribution of the dark matter due to any gravitational effects. We see here that the largest spread distances are significantly larger than any of the characteristic distances shown in this figure; this is even compared to the largest separation for any two particles at $z = 0$, implying that these distances are much further than can be achieved from Hubble expansion in voids alone. The overall distribution follows an exponential decay, with exponentially fewer particles (once outside the inner $\sim 0.5h^{-1}$Mpc) being found at larger distances. There are many possible explanations for these results, from tidal stripping of objects that end up never merging, accretion of dark matter from satellites (see e.g. the effects in van den Bosch & Ogiya, 2018), or even particles on randomised orbits from recently accreted material that end up on opposite sides of the 'splashback' region (Diemer & Kravtsov, 2014; Adhikari et al., 2014). This splashback region is sometimes larger than the virial radius of the halo, meaning that two particles may be separated by up to $4R_{\rm vir}$ through this process (Diemer et al., 2017). Finally, we may expect three-body interactions between substructures, leading to some being ejected to very large distances (up to $6R_{\rm vir}$; see Ludlow et al., 2009). This is the only plausible explanation that we have for such large spread distances in the dark matter. In practice, we expect the final spread distribution to reflect the effects of multiple dynamical mechanisms.

In Fig. 4.2 we also show the consequences of choosing to average over different numbers of initial neighbours. The simplest metric would use a single nearest neighbour in the initial conditions. However, the distance between any two nearest neighbours would be 'double counted' and not representative of motion relative to the surrounding matter distribution in the case of a single neighbour travelling a long distance. The choice of $n = 3$ is the lowest that ensures that the metric S always represents the distance between two real pairs of particles, whilst simultaneously solving this conceptual problem. In practice, the overall distribution of the spread metric does not depend much on the number of neighbours considered, but we find that larger choices of n yield a more direct connection between spread distance and hierarchical structure (with low-spread particles dominating substructures and high-spread particles corresponding to more diffuse components, as shown in Fig. 4.4). The only minor difference when considering changes in the choice of neighbours is the placement of the 'bump' in gas outside haloes at around $0.2h^{-1}$Mpc. In the case with no averaging, this bump corre-

sponds to approximately the mean inter-particle separation, with the distance increasing as a function of $n^{1/3}$.

The distributions of spread metrics at large distances for particles inside and outside haloes are very similar, even when using the averaging scheme. With such large separations, the particles are essentially entirely decoupled; as such, the probability that they either reside in a halo or not is simply just the probability of a particle to lie within a halo (which, here, is approximately 50%).

4.3.2 Baryon Spreading in Simba

Fig. 4.3 shows how the distribution of spread distances for the gas particles is significantly different to that for the dark matter. Gas particles are able to spread to much larger distances, up to $12h^{-1}$Mpc (approximately 10 times the virial radius of the largest halo in the box!), compared to the $7h^{-1}$Mpc that dark matter can reach. We also see that even gas inside of haloes at $z = 0$ has spread significantly more than the dark matter when explicitly selecting for this component. This suggests a different origin for the gas and dark matter content of haloes.

Another interesting component is the gas that originated in Lagrangian regions (i.e. next to the dark matter that will reside in haloes at $z = 0$), indicated by the blue dashed line. With the baryon fraction of haloes being typically less than 50% of the cosmic mean, we should expect that a significant amount of Lagrangian gas is lost over time, possibly spreading to large distances out of haloes due to high energy feedback events, either through galactic winds or AGN feedback. In Simba, we see that gas from Lagrangian regions indeed spreads systematically further, with a factor of ~ 2 more particles at distances larger than $\sim 4h^{-1}$Mpc than an unbiased selection would suggest.

A visualisation of the projected surface densities corresponding to the low- and high-spread particles is shown in Fig. 4.4 for both dark matter and gas, for the fiducial Simba model. We define 'low-spread' particles as those in the lower tertile (33%) of the distribution, and 'high-spread' particles as those in the upper tertile. By making these cuts in the distance distribution, we are able to show that the low-spread particles correspond to substructure, with the high-spread particles contribution being the larger-scale, more diffuse, CGM and intergalactic medium (IGM).

Considering first the dark matter in the largest halo (top row), we see that the very small-scale substructure of the halo is preferentially picked up by the low-spread particles, including the central density peak itself and the centers of subhaloes. In contrast, the more diffuse dark matter component that fills the space between these individual density peaks is significantly more prominent in the high-spread particles, with only a small amount of residual sub-structure remaining. These trends

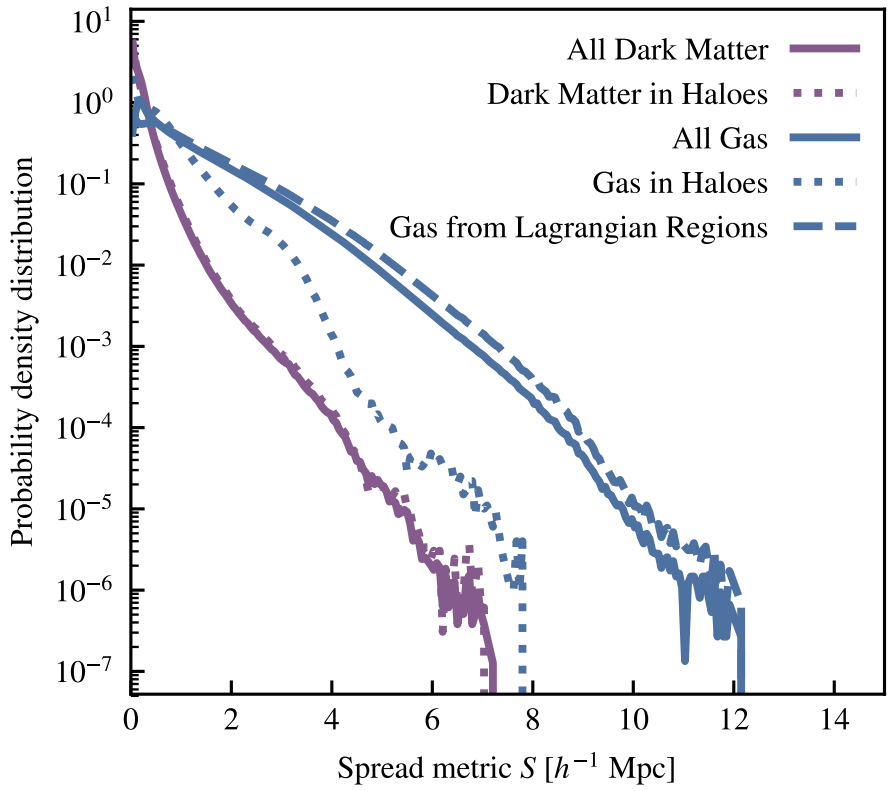


Figure 4.3: Spread distance distribution for gas at $z = 0$ (blue) compared to that of the dark matter component (purple). Solid lines indicate the full distribution, dotted lines correspond to matter inside $z = 0$ haloes, and the blue dashed line shows the distribution for gas that was inside of Lagrangian regions at $z = 99$. The distributions for gas inside haloes and outside haloes are significantly different, with gas that resides outside haloes being preferentially spread to larger distances than gas on average. Note that only 10% of the gas in the entire simulation is in haloes at $z = 0$. Gas that originated in Lagrangian regions is preferentially spread the most, with a factor of 2 offset over the unbiased selection at large spread distances.

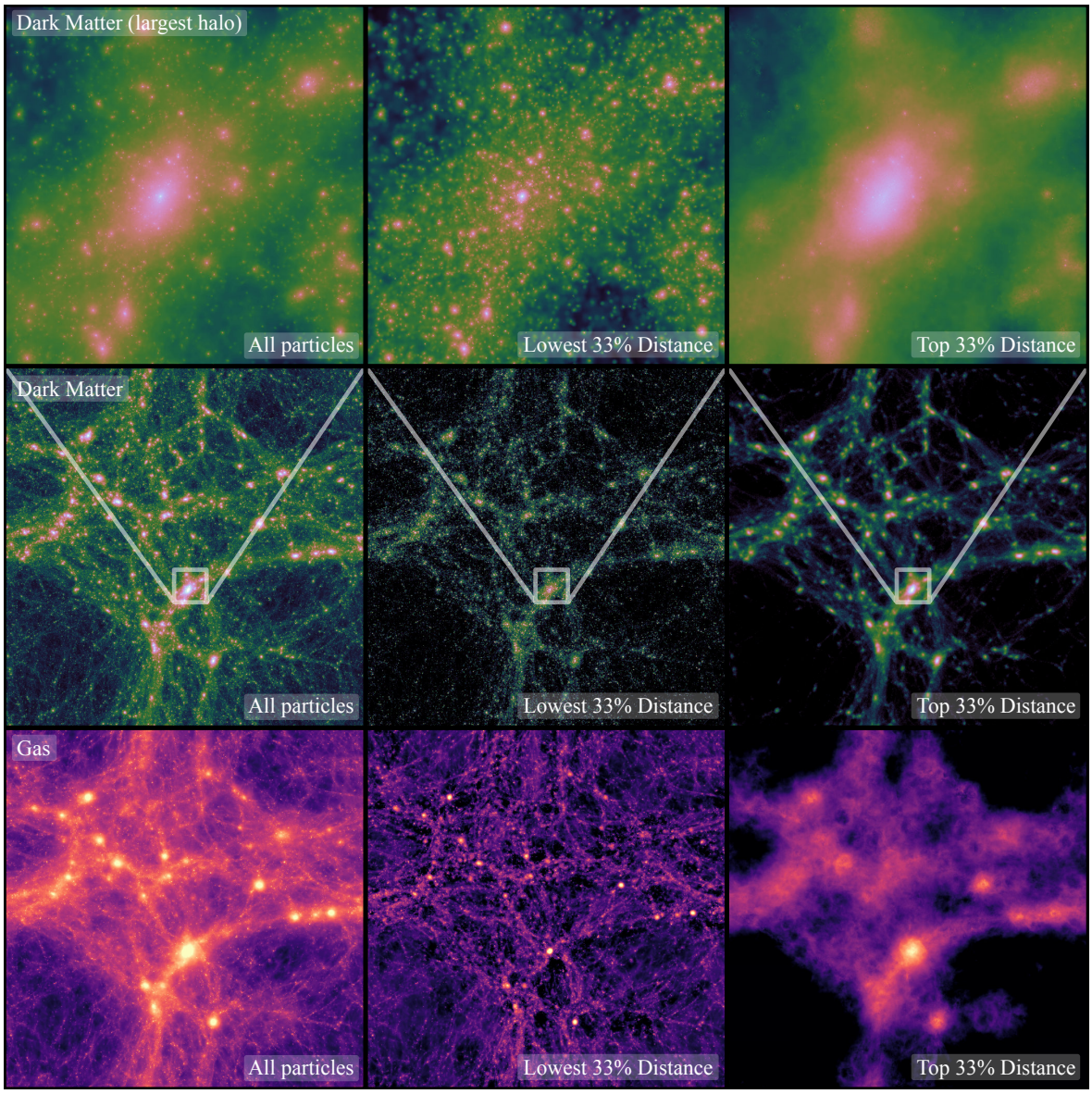


Figure 4.4: Projected mass surface density distributions for different particle selections at $z = 0$. The three rows show, from top to bottom, the dark matter in a $4.5h^{-1}$Mpc cubic volume centred around the largest halo ($R_{\rm vir} \sim 1.3h^{-1}$Mpc), the dark matter distribution in the whole $50h^{-1}$Mpc box, and gas distribution again in the whole volume. Columns show, from left to right, all particles inside of the corresponding volume, the 33% of the particles with the lowest spread distance, and the 33% of the particles that have spread the most. For the dark matter, these cuts correspond to particles that have travelled less than $0.1h^{-1}$Mpc and more than $0.25h^{-1}$Mpc, respectively. For the gas, these numbers increase to $0.45h^{-1}$Mpc and $1.25h^{-1}$Mpc, respectively, due to the larger spread that gas particles experience. Each density projection is generated using smoothing lengths defined to encompass the 64 nearest neighbours and smoothing lengths are kept consistent across columns (i.e. they are not recomputed for different particle distributions). All density projections in a given row also use the exact same (logarithmic) normalisation and colour map to enable direct comparisons. Note the significant difference between the spatial distribution of material with different spread metric, with sub-structure preferentially picked out by the low spread distance selection while the large spreads trace large scale structure.

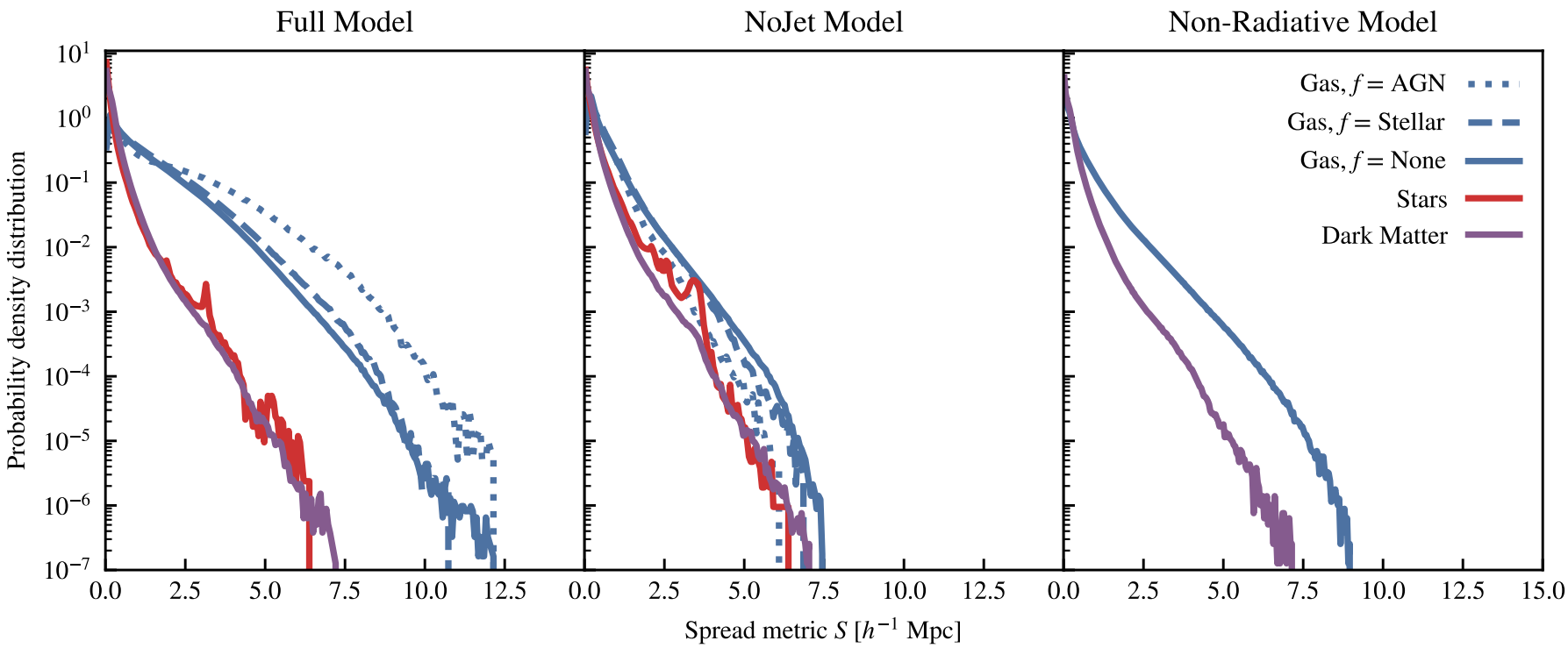


Figure 4.5: Distribution of spread distances split by particle type for gas (blue), stars (red), and dark matter (purple). This is shown for the $z = 0$ particle distribution in the reference model (left), the `NoJet` model (center), and the non-radiative simulation (right). The left and middle panels separate gas particles that have not been involved in any feedback event (solid) from those that have participated directly in either stellar (dashed) or AGN (dotted) feedback events. Jets are primarily responsible for spreading baryons to the largest distances in Simba, with significant entrainment of gas that did not directly participate in feedback events. The stellar distribution is significantly more noisy than the others due to the smaller number of star particles (compared to gas or dark matter) in the simulation.

are also clear at larger scales, as shown by the view of the $50h^{-1}$Mpc box in the second row, with large-scale dark matter filaments primarily traced by high-spread particles. It is interesting to note that a large amount of structure in voids is not present in either of these panels, with it being captured by the medium-spread particles with values $0.1h^{-1}\text{Mpc} < S < 0.25h^{-1}\text{Mpc}$. The spread metric is thus a very useful tool to connect hierarchical structure and dynamical evolution in cosmological N-body simulations.

The bottom row in Fig. 4.4 shows the large scale gas distributions separated with the same proportions, with a third of the total gas mass contained in each of the middle and right panels (this corresponds to different absolute values of the spread metric compared to the dark matter panels). The low-spread particles trace the densest gas in haloes along with lower density gas in the central parts of large scale filaments. Of particular interest is the high-spread gas, which traces the large bubbles around the most massive haloes that strong AGN jets produce in the Simba model (see §4.3.3). As expected from Fig. 4.3, the top third of the gas distribution has been pushed out to significantly larger distances compared to the third of the dark matter that moved the most due to gravitational dynamics only. The spread metric hence captures the impact of feedback in a global sense.

4.3.3 Connecting feedback and the spread of baryons

The kinetic feedback scheme used in Simba for both star formation and AGN feedback makes it straightforward to identify the gas elements that have been directly impacted by feedback. However, these gas elements will then go on to entrain and deposit energy into other gas elements as they travel. This makes it challenging to fully capture the impact of feedback solely from particle tagging. Here, we use the additional `NoJet` and non-radiative simulations in order to explore how baryon redistribution is sensitive to different physics modules in Simba, although we caution that these are not fully independent sub-grid models with their own calibration process.

The left panel of Fig. 4.5 shows the spread distribution for the full Simba model, splitting the gas component into particles that have been affected by different types of feedback. Here, AGN feedback takes precedence over stellar feedback, such that if a particle has been affected by both it is only classified as being part of the f = AGN group. We see that the particles that have directly interacted with the AGN are spread to significantly larger distances, with a vertical offset of 0.5-1 dex compared to no-feedback particles for $S \gtrsim 5h^{-1}\text{Mpc}$. Particles that have been directly kicked by stellar feedback also have systematically higher spread metric values, albeit with a smaller offset. This implies that particles are indeed being spread to these large distances by feedback events.

The left panel also now includes the stellar component, which shows a very similar distribution to

that of dark matter. This could be seen as surprising given that stars form out of the most bound gas at the center of haloes. Despite this, it would be unlikely for a star particle to form from a gas particle with a high spread value, as these must have been separated dynamically from their closest dark matter neighbour requiring some form of strong energy injection. This would eject and heat the particle making it less likely to cool down, accrete back onto the galaxy, and condense to high enough density to form a star by redshift $z = 0$. This suggests that the stellar spread distribution is produced by dynamical effects after the star has formed, affected by the same physics that shapes the spread distribution for the dark matter, including tidal disruption and stripping of satellites, merger events, and orbital divergence through N-body dynamics.

The middle panel of Fig. 4.5 shows the spread distribution for the `NoJet` simulation, where we still include AGN feedback in the form of radiative winds and X-ray heating but the high velocity jet feedback mode is disabled. With this change, the spread metric is significantly affected, with much less difference between the distributions of the dark matter, gas, and stellar components. While galactic winds and AGN feedback in radiative mode can still decouple the dark matter and gas components, high-velocity jets are clearly the dominant mechanism responsible for spreading baryons to the largest distances in Simba. Surprisingly, gas particles directly kicked by feedback in this case show a lower spread distribution compared to gas not directly impacted by feedback, in contrast to the trend seen for the fiducial Simba model. This suggests that feedback in the `NoJet` simulation is not strong enough to compensate for the fact that feedback events occur in the densest regions (inside galaxies). It is intrinsically more difficult to escape these deep potential wells, especially now that a crucial energy injection mechanism from the AGN jets is missing.

This result is surprising given that less than 0.4% of gas particles in the simulation have ever interacted directly with the AGN jets; this has been enough to significantly decouple the gas from the dark matter dynamically. Such a high degree of separation points to substantial amounts of gas being entrained by these powerful jets. It is not simply the case that higher mass ($M_H > 10^{11}$ $M_\odot$) haloes are quenched internally reducing their star formation rate; the energetics and dynamics of the CGM and IGM are significantly altered, as is already seen by the more complex interaction between the turn-off of the galaxy stellar mass function (GSMF) and the power of the AGN jets in many studies (Weinberger et al., 2018; Davé et al., 2019).

The final contrast to highlight is the difference between the `NoJet` and non-radiative model. The non-radiative model shows increased distance between gas particles and their associated dark matter neighbour compared to the `NoJet` run; this is due to the lack of cooling preventing particles that lie in small haloes from remaining as tightly bound. It also highlights how difficult it is to drive gas

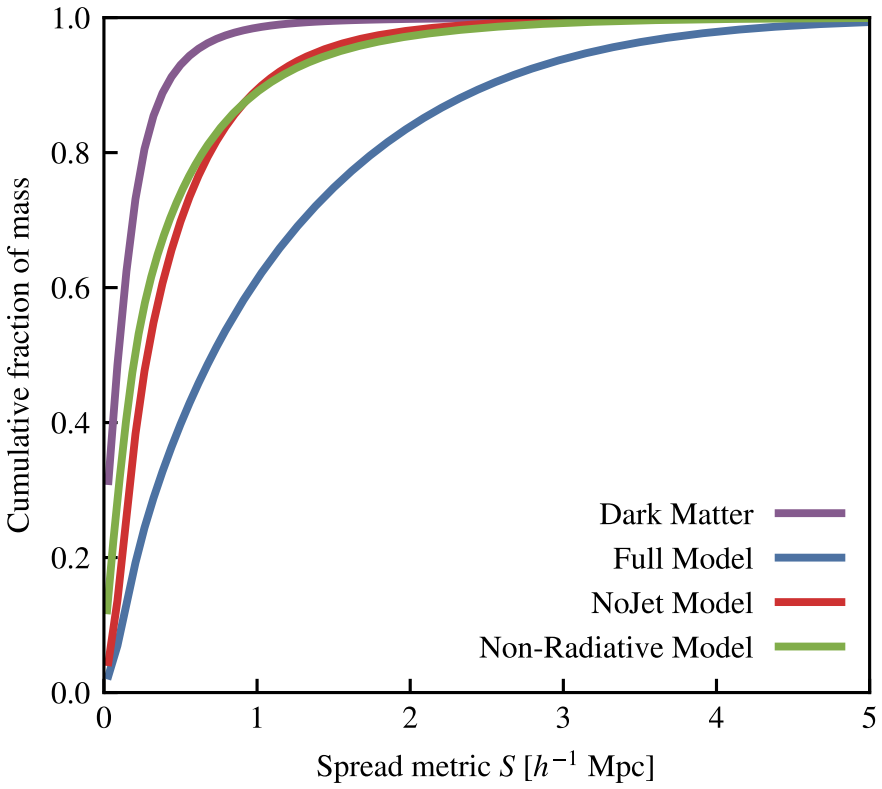


Figure 4.6: Cumulative version of Fig. 4.5 for the spread of gas in the three different models alongside the dark matter from the full model. This shows that 10% of the gaseous matter has spread at least $3h^{-1}$Mpc, while 90% of the dark matter resides within $0.5h^{-1}$Mpc.

into the centers of structures without cooling. The collisionless dark matter can continue to fall in to bound structures, with the gas being prevented due to strong accretion shocks. This allows for a very different kind of separation than what we have shown above for the full physics model including cooling and feedback.

In Fig. 4.6 we show the cumulative version of Fig. 4.5 to better show the amounts of mass that are spread to large distances, showing that 40% (10%) of cosmological baryons have moved $> 1h^{-1}$Mpc ($3h^{-1}$Mpc) by $z = 0$, with a slow tail off ending with nearly all of the mass being constrained to be spread less than $5h^{-1}$Mpc.

4.3.4 Redshift evolution of the Spread Metric

From Fig. 4.5 it is clear that the AGN jets have a significant impact on the spread metric, causing the maximal spread distance in the gas to almost double. In Fig. 4.7 we explore how this deviation between gas and dark matter depends on redshift. The dashed lines show the spread metric distribution at $z = 2$, and from this we see that in the full model gas has spread to over $5h^{-1}$Mpc (more than twice that of the largest dark matter spread) even by this early epoch. The `NoJet` model shows no such behaviour, showing a very close convergence between the spread metrics of all three particle types. This long-distance baryon spreading is then not a late-time effect; it occurs at all times that the jets are active, gradually filling in the final spread metric distribution.

4.4 Lagrangian baryon transfer

We have explored the relative motion of dark matter and baryons using a particle-level metric, showing that AGN jets in the Simba cosmological simulations can spread baryons up to $12h^{-1}$Mpc relative to the neighbouring dark matter. In this section, we consider the movement of baryons relative to dark matter haloes and their corresponding Lagrangian regions. The definitions of haloes and Lagrangian regions used here are described in §4.2.

This topic has been considered recently by Liao et al. (2017), where they used a $10h^{-1}$Mpc non-radiative simulation to show that the gas in haloes may originate from different places than the dark matter in those same haloes in the initial conditions.

4.4.1 The different origins of baryons and dark matter in haloes

Fig. 4.8 illustrates the mixed origins of the gas and dark matter components in bound structures at $z = 0$ by showing simultaneously the initial and final states of the simulation. A common trend for

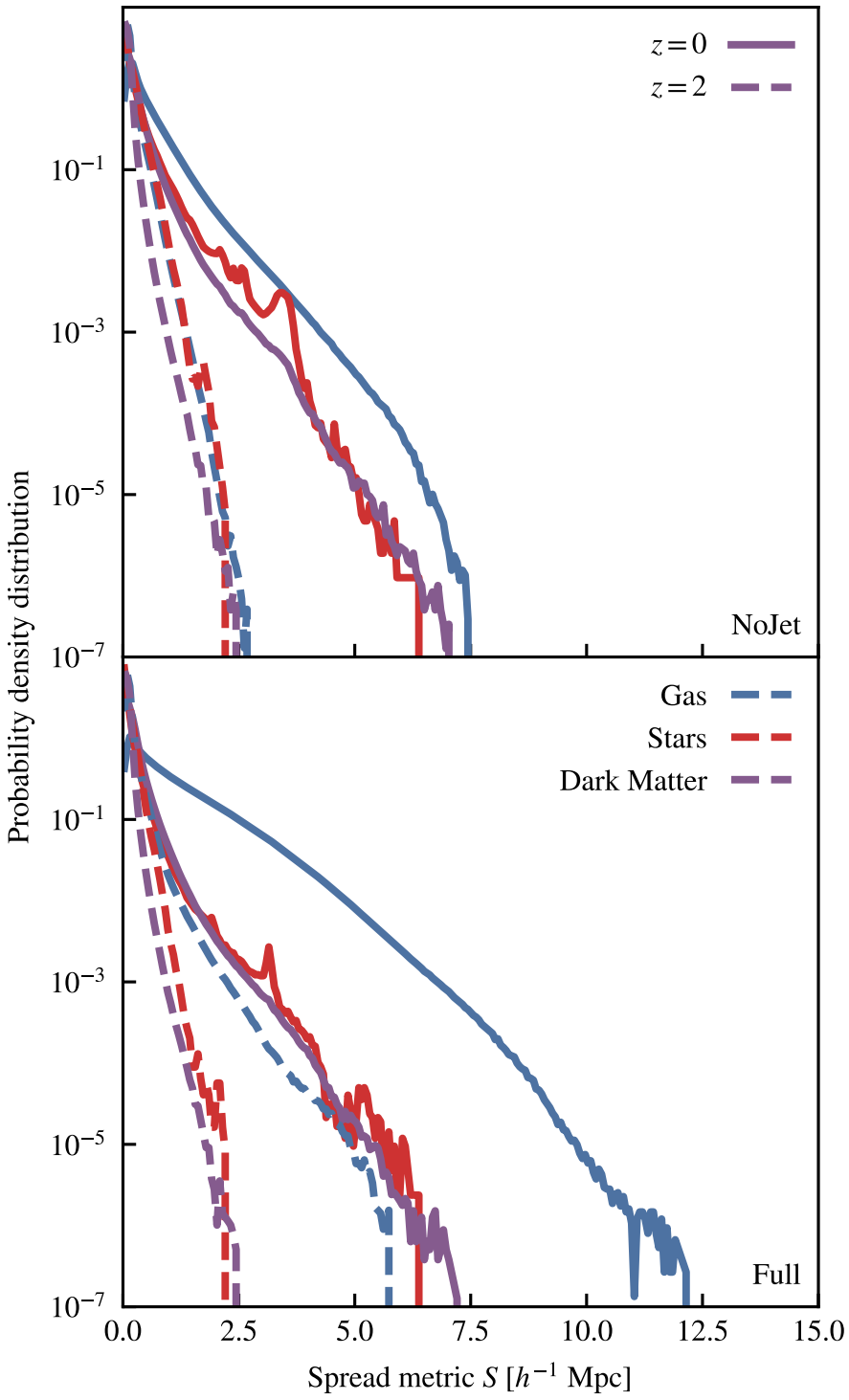


Figure 4.7: Spread metric distributions shown again for the `NoJet` (top) and full Simba model (bottom) simulations, now including the redshift $z = 0$ (solid) and $z = 2$ (dashed) results. We see that at all redshifts the `NoJet` model produces spread metric distributions that are highly similar for all three particle types, with the full Simba model showing divergence between the dark matter and gas even at redshift $z = 2$. The AGN jets cause a significant difference between these gas distributions, and are able to power winds out to a spread of $5h^{-1}$Mpc even by $z = 2$.

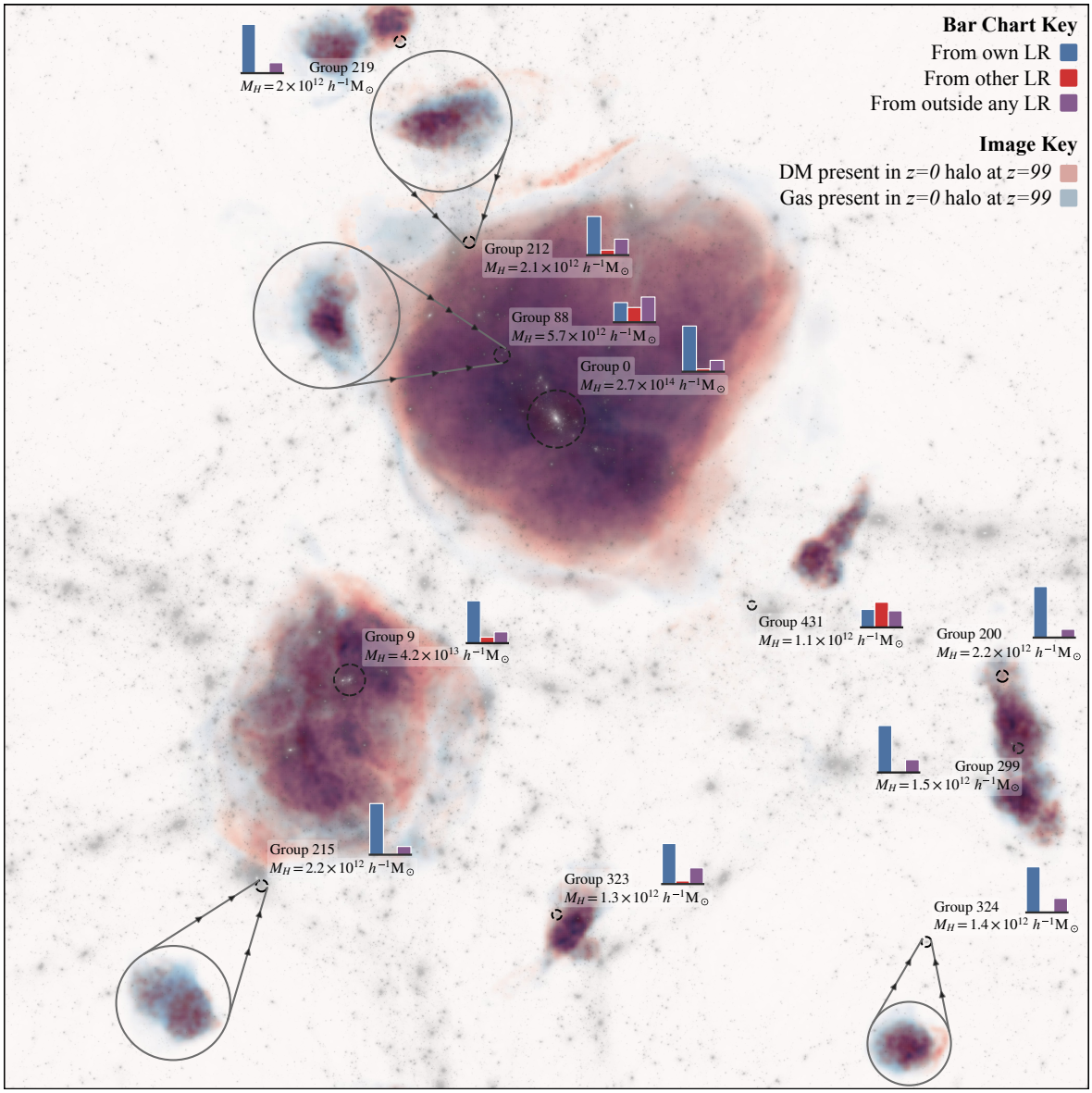


Figure 4.8: This visualisation shows two epochs at once, simultaneously showing the initial conditions (in blue and red) and the final simulation volume at redshift $z = 0$ in white/grey. The blue and red show the positions of the gas and dark matter (respectively) *in the initial conditions* for particles that reside in selected haloes at redshift $z = 0$. The overlaid white/grey map shows the dark matter at redshift $z = 0$ to enable comparisons between the initial and final comoving positions for various bound structures. For each selected halo, the dashed black circles show their virial radii as defined in §4.2. For some haloes in crowded regions, we have overlaid a circle and arrows showing which blob of dark matter and gas in the initial conditions collapses to form this halo. Finally, for each halo we show a small bar chart showing how their gas is composed from Lagrangian components, as described later in the text. The blue bar shows the fraction of gas in each halo that originated from that haloes own Lagrangian region, the red bar shows the gas from another haloes Lagrangian region, and the purple bar shows the fraction of gas that originated outside any Lagrangian region. This figure illustrates the significant differences in origin between the gas (blue) and dark matter (red) for these selected haloes of various masses. We also see how the environment of each halo changes its Lagrangian make-up. In particular, group 431 shows a large baryonic component originating from the Lagrangian region of another halo, with this halo entering a small cluster environment near the end of the simulation. Note that individual regions are colour-mapped separately, i.e. the intensity of colour for a single halo is unique to that halo only, as to enable all Lagrangian regions to be seen. Without this choice, the structure for the lower mass haloes would be completely washed out.

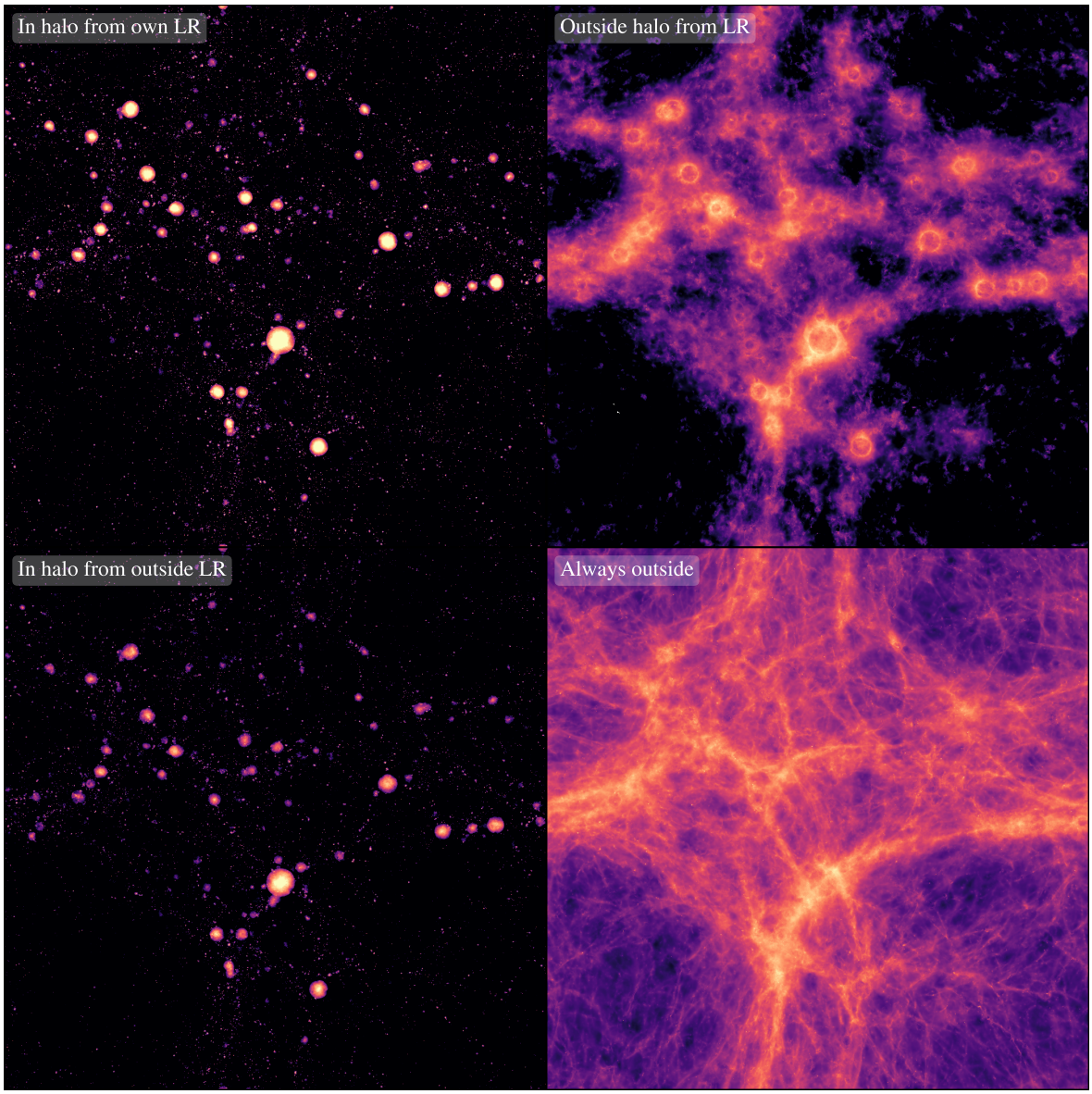


Figure 4.9: Gas distribution in the fiducial Simba model for the full $50h^{-1}$Mpc volume, split by the following Lagrangian components (clockwise, starting from top left): particles that began in Lagrangian regions at $z = 99$ and have remained in the associated haloes at $z = 0$; particles that began in Lagrangian regions and ended up outside of the destination halo; particles that began outside any Lagrangian region and ended up outside any halo; and particles that ended up in a halo but originated outside any Lagrangian region. All images are shown with the same (logarithmic) colour-map and normalisation and taking their linear sum would reproduce the full gas distribution at $z = 0$. Gas particles that began in Lagrangian regions but ended up outside of haloes (top right) show a striking similarity to the distribution of gas with the 33% highest spread distance shown in Fig. 4.4. As expected, particles that began outside of Lagrangian regions and remained outside of haloes (bottom right) trace the filaments and voids.

all A common trend for all lagrangian regions is the gas extending to larger radii than the main dark matter component in the initial conditions, showing that gas in general is able to collapse further (due to cooling and other dissapational processes like shock heating) than the dark matter, which is unable to lose angular momentum as efficiently. This is consistent with the larger values of the spread metric for gas in haloes relative to the dark matter in haloes, as shown in Fig. 4.3. Dark matter haloes in mergers can also pass through each other, whereas the gas will shock heat and combine into one larger halo. Through this process the two dark matter components will remain distinct (and hence compact in the initial conditions), but the gas regions will combine into a larger, more amorphous, collection.

The origin of the dark matter in the initial conditions corresponds exactly to our definition of Lagrangian region for that component in §4.2. These Lagrangian regions have very complex shapes, with larger haloes tending to have more spherical Lagrangian regions, as can be seen with the largest halo in the box (Group 0) in Fig. 4.8. These complex non-spherical shapes are why we chose to identify our Lagrangian regions for gas through neighbour searching, as other methods (e.g. constructing a convex hull enclosing all dark matter particles that end up in a given halo) would not allow us to capture the surprisingly intricate structure that is at play here.

There are many possible reasons for the complex shapes that we see here. Consider a simple case where we have one 'main' halo, and a satellite that is being accreted. The gas and dark matter in the satellite galaxy have several potential fates. For instance, when accreting onto the main halo, the gas in the satellite may be shock heated, and stalled in the CGM, with the dark matter being able to continue to move towards the center of the main halo. This process dynamically separates the dark matter and gas, and now the gas may have several fates; it could be pushed out in a feedback event, rise out of the halo due to buoyancy, or fall to the centre of the halo after cooling and re-join the dark matter. Once the gas has been removed from the CGM into the IGM, it is free to be picked up by other passing galaxies.

The other possibility for the fate of this substructure is the dark matter failing to accrete onto the central. In this case, the dark matter continues moving out into the IGM, with the gas being shocked and captured by the main halo. It is this complex difference in assembly between dark matter and baryons, due to the latter behaving as a collisional fluid, that we aim to capture here.

4.4.2 Computing transfer between Lagrangian regions

Given the definitions of haloes and Lagrangian regions in §4.2, it is possible to classify every particle in the simulation according to their Lagrangian ID and halo ID (if any) in the initial and final conditions. The algorithm is as follows:

1. ID match all particles between the initial and final conditions, including star particles (these are matched to their gas progenitor). Black holes are ignored in this analysis since globally they represent a minimal amount of mass.

2. Every particle at $z = 0$ has several possible final states and origins, based on its halo ID (i) and Lagrangian region ID (j):

   - Particle resides in halo ($i \neq -1$)
     - Particle originated in the same Lagrangian region, $j = i$
     - Particle originated outside any Lagrangian region, $j \equiv -1$
     - Particle originated in some other Lagrangian region, $j \neq i$
   - Particle resides outside of any halo ($i \equiv -1$)
     - Particle originated outside any Lagrangian region, $j = i$
     - Particle originated in some Lagrangian region, $j \neq i$

3. For every halo and Lagrangian region, the mass originating from each of the above components is computed and stored.

A visualisation of this particle classification scheme is shown in Fig. 4.9, where we split the gas distribution in the Simba $50h^{-1}$Mpc box into the four main Lagrangian components that we consider in the remainder of this chapter. Considering each panel clockwise from the top left, we select first the gas that is in the same halo at redshift $z = 0$ as the Lagrangian region that it originated in. As expected, we see a population of spherical shapes corresponding to every halo in the box, with their sizes corresponding to $R_{\rm vir}$ as defined by AHF. The centers of haloes, where the gas is densest, are the brightest.

In the top right panel we have the gas that is outside any halo at $z = 0$, but is assigned to a Lagrangian region at $z = 99$; this is the gas that should have ended up in haloes by the end of the simulation if the baryonic matter was also collisionless. We see that this component traces gas primarily around massive haloes, resembling the large-scale bubbles that the AGN jets power in Simba (Davé et al., 2019). Note that some of this gas piles up just outside of haloes due to the somewhat arbitrary boundary defined by the virial radius of haloes. This gas resides primarily in filaments, with some reaching out into the voids.

In the bottom right panel, we visualise the gas that begun outside any Lagrangian region and resides outside any halo at redshift $z = 0$. This gas traces the majority of the filamentary structure, and shows all of the structure in the voids.

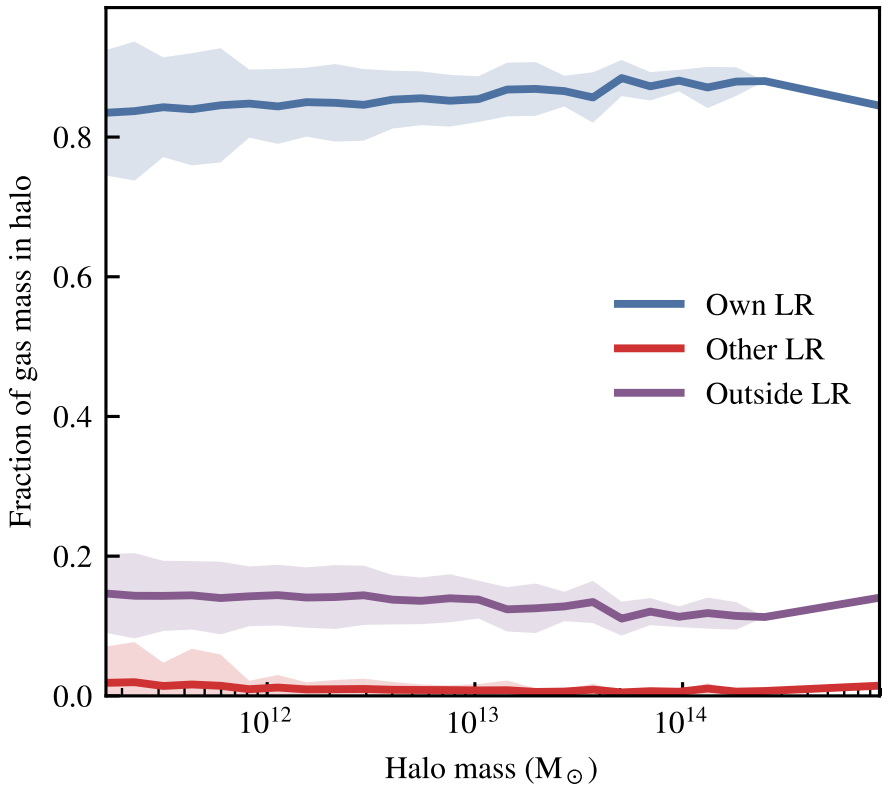


Figure 4.10: The fraction of baryonic mass originating from each Lagrangian component in the non-radiative model (i.e. without sub-grid physics) is shown as a function of redshift $z = 0$ halo mass. The gas particles are binned by their origin, with the baryons originating from their own Lagrangian region shown in blue, the Lagrangian region of other haloes (red), and outside of any Lagrangian region (purple). Shaded regions show the 1σ scatter in a given bin, which is given by one standard deviation of variation. The lines represent the mean value within each bin. Approximately 85% of the baryonic mass of a given halo originates from its own Lagrangian region, showing very little transfer of baryons from either outside or from another Lagrangian region. This is provided for comparison to the full model result in Fig. 4.11.

Finally, in the bottom left panel, we have the gas that is in haloes at $z = 0$ but originated from outside any Lagrangian region. As expected, this shows a very similar structure (albeit less bright) to the gas that resides in its own halo (top left), but this component originates from regions where the dark matter now resides outside of haloes. This gas is likely dragged into these bound structures by cooling flows, while the dark matter is not able to lose angular momentum quickly enough to assemble by $z = 0$.

4.4.3 Transfer in a non-radiative Model

Before considering the numerical results of the full model, we first present the non-radiative simulation as a null model to investigate the effects of hydrodynamics alone. In this case, we run the

simulation without cooling, star formation, or feedback, only including hydrodynamics, cosmology, and gravity. In Fig. 4.10 we present the fraction of baryonic mass for each halo contributed from each Lagrangian component, as a function of halo mass. The blue line shows the fraction of mass in each halo from its own Lagrangian region (top left in Fig. 4.9), the red shows transfer into a halo from another Lagrangian region, and the purple line shows the fraction of baryonic mass from outside any Lagrangian region (bottom left in Fig. 4.9). There is no dependence on halo mass (as the simulation is effectively scale-free above some resolution limit), and apart from some small level of transfer from outside any Lagrangian region (of around $10-15\%$), the baryonic mass in each halo consists of that which originated in its own Lagrangian region.

The difference in origins of the baryons in the final haloes, from hydrodynamical effects alone, is then around the $10-15\%$ level. This is close to the 25% level of segregation between gas and dark matter reported by Liao et al. (2017) (who also used a non-radiative simulation), with the difference likely rooted in the definitions that we use. We consider the fraction of gas particles in the final redshift $z = 0$ halo whose initially pairing dark matter is also resident in that halo; hence what we are really counting is the 'contamination' of the halo by gas particles from outside of its Lagrangian region. Liao et al. (2017) count all particles in the final halo, treating gas and dark matter equally, then finding all particles that were gas-dark matter pairs in the initial conditions. Their higher level of segregation is expected due to contributions from dark matter particles that are resident in a halo but whose initial gas pair is not. Fundamentally this represents the difference in our approaches; here we are interested in treating the dark matter as a ground source of truth, and asking if the gas nearest to that dark matter follows it into the same haloes. Liao et al. (2017), on the other hand, were interested in treating *all* occupants of the final halo as the ground source of truth, and asking what differences there were in their origin.

The causes for our contamination here are less clear than in the case of Liao et al. (2017); we would report a halo that has had gas only *removed* as being completely uncontaminated, and hence stripping of gas is an unsatisfactory explanation of these differences. The likeliest explanation for the contamination in this case is that the baryons and dark matter go through a phase of mixing as they enter the cosmic web, before going on to fully collapse into bound structures.

4.4.4 Transfer *into* haloes

Moving on to the full Simba model, we consider again the fractions of baryonic mass as a function of halo mass, split by Lagrangian component. Fig. 4.11 shows three panels: the left panel shows all baryons, the centre shows only gas, and the right panel shows the contribution from only the stars.

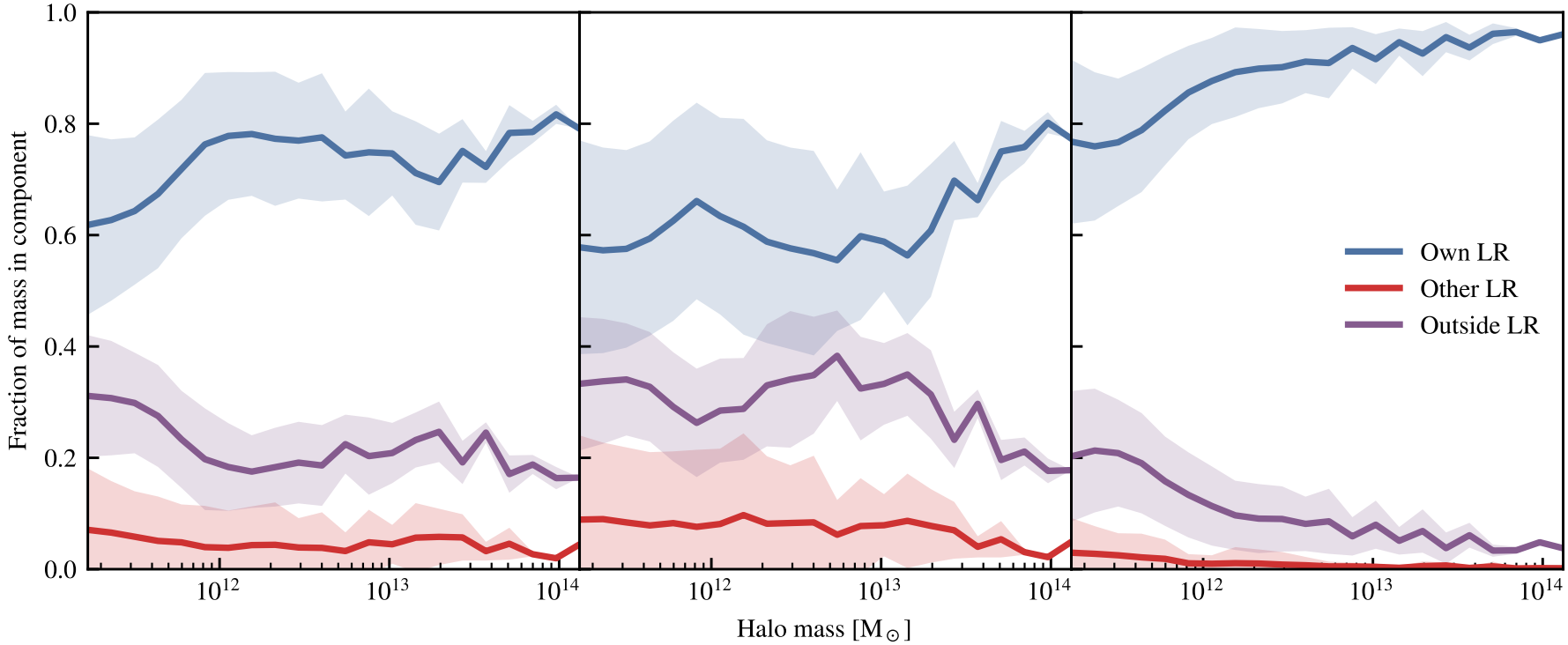


Figure 4.11: The fraction of baryonic mass in haloes at $z = 0$ originating from their own Lagrangian region (blue), the Lagrangian region of other haloes (red), and outside of any Lagrangian region (purple), shown as a function of $z = 0$ halo mass for the fiducial Simba model. We consider all baryons in haloes (left) as well as their gas (center) and stellar (right) components separately.

The lines are coloured the same as the non-radiative model shown in Fig. 4.10. Now that we have introduced scale into the simulation through density-dependent energy injection mechanisms, these components scale with halo mass. The general trend is that for an increasing halo mass, a Lagrangian region is able to hold on to more of the original baryonic mass, with this flattening off around $M_H = 10^{12}$ $M_\odot$. For a given halo, significantly more of the gaseous mass originates outside the original Lagrangian region as compared to the stellar mass ($\sim 40\%$ versus $\sim 10\%$). The transfer between haloes is at around the $\sim 10\%$ baryonic mass level, with this transfer predominantly originating from the gaseous component, as compared to the stellar component. This combines nicely with the distance metrics shown in §4.3, which showed that the dark matter and stars have very similar dynamics and hence should be similarly well bound.

This transfer into, and between, Lagrangian regions can have several physical origins. The first, as shown in the non-radiative run, is caused by the collisional dynamics of the gas preventing gas from following the dark matter in all cases. We found that this can account for up to 15% of the baryonic mass of a bound structure at redshift $z = 0$ originating from a different region than the dark matter (see Fig. 4.10), but this could not account for any *inter-Lagrangian* region transfer.

The galaxy formation sub-grid model clearly has a significant effect on the baryonic make-up of haloes at redshift $z = 0$. The fraction of mass from outside any Lagrangian region has increased to 20-40%. This increase is explained by the inclusion of sub-grid cooling and feedback processes, with the baryons now able to cool before accreting and lose angular momentum at a much higher rate than the dark matter component is able to.

Around 10% of the baryonic mass of haloes is now made up of gas that has experienced inter-Lagrangian transfer. It is important to recall that this is transfer between bound structures at redshift $z = 0$, and that it only takes into account the initial and final conditions of the simulation; this analysis does not consider the complete history of these particles.

The transfer between haloes has several possible sources: stripped gas from nearby galaxies that are still classified as their own bound structures at redshift $z = 0$, gas that has been expelled from galaxies through stellar winds or AGN feedback and re-captured by a halo, and transfer due to boundary effects caused by the complex shapes of Lagrangian regions according to the definition adopted. With the non-radiative simulation showing zero transfer between haloes, and there being little transfer before $z = 2$ in the fiducial model (see below in Fig. 4.12), we believe that the contribution from pure dynamics alone to inter-Lagrangian transfer is likely very small. When repeating this analysis with the `NoJet` run, the inter-Lagrangian transfer is reduced, but still remains at the 10% level. The feedback events that power this transfer must be dominated by the expulsion (or alternatively preventative

pathways) from stellar winds and the residual thermal AGN feedback.

A given mass bin contains haloes that entertain a range of 10x in transfer, which is likely dependent on environment. Future work should investigate in more detail the physical mechanisms driving the scatter in these relations.

The level of transfer above a halo mass of 10^{13} $M_\odot$ must be interpreted carefully, as there are very few haloes above this mass present in the box (less than 50), with the small scatter being misleading. It is also important to note that the shaded regions in Fig. 4.11 represent the 1σ scatter in a given bin and explicitly do *not* include any dispersion that would occur from a finite sampling of haloes or halo assembly bias.

4.4.5 Redshift evolution of transfer into haloes

To further investigate the origin of the inter-Lagrangian transfer, in Fig. 4.12, we consider the `NoJet` model and show how the gas in haloes at redshift $z = 2$ is composed in this and the full Simba model.

We see that both the `NoJet` and Simba models broadly reproduce the same fractions of gas in each Lagrangian component, with some interesting differences. In the full model, a higher fraction of the halo gas originates from inter-Lagrangian transfer than the `NoJet` model at all masses, with no change in the shape of this function observed. The level of inter-Lagrangian transfer is increased by around $25-50\%$ such that it represents approximately 15% of the gaseous mass in the halo, with the `NoJet` results showing an inter-Lagrangian fraction of $\sim 10\%$. The fraction of gas originating outside of any Lagrangian regions shows a dip at around 10^{12} $M_\odot$ being removed in the `NoJet` model, however this is well within the scatter that we observe in the full model results.

All of this is despite both models producing very different $z = 0$ halo baryon fractions (see Fig. 4.14 for the full model; the `NoJet` model produces baryon fractions at approximately the cosmic mean for all halo masses above $\sim 10^{11}$ $M_\odot$). For a further investigation, halo matching should be performed between the two models and individual cases compared, but this is out of the scope of the current work.

The fraction of gas in haloes originating from the different Lagrangian components shows a closer match at $z = 2$, with the shape and normalisation of all components being well within the reported scatter. The higher-mass end of these results ($M_H > 10^{13}$ $M_\odot$) also lacks objects here, with there being even fewer in this mass range than at $z = 0$.

We see that between redshift $z = 2$ and $z = 0$ a change in the slope of these functions takes place, and that the level of inter-Lagrangian transfer increases significantly. The fraction of gas originating from

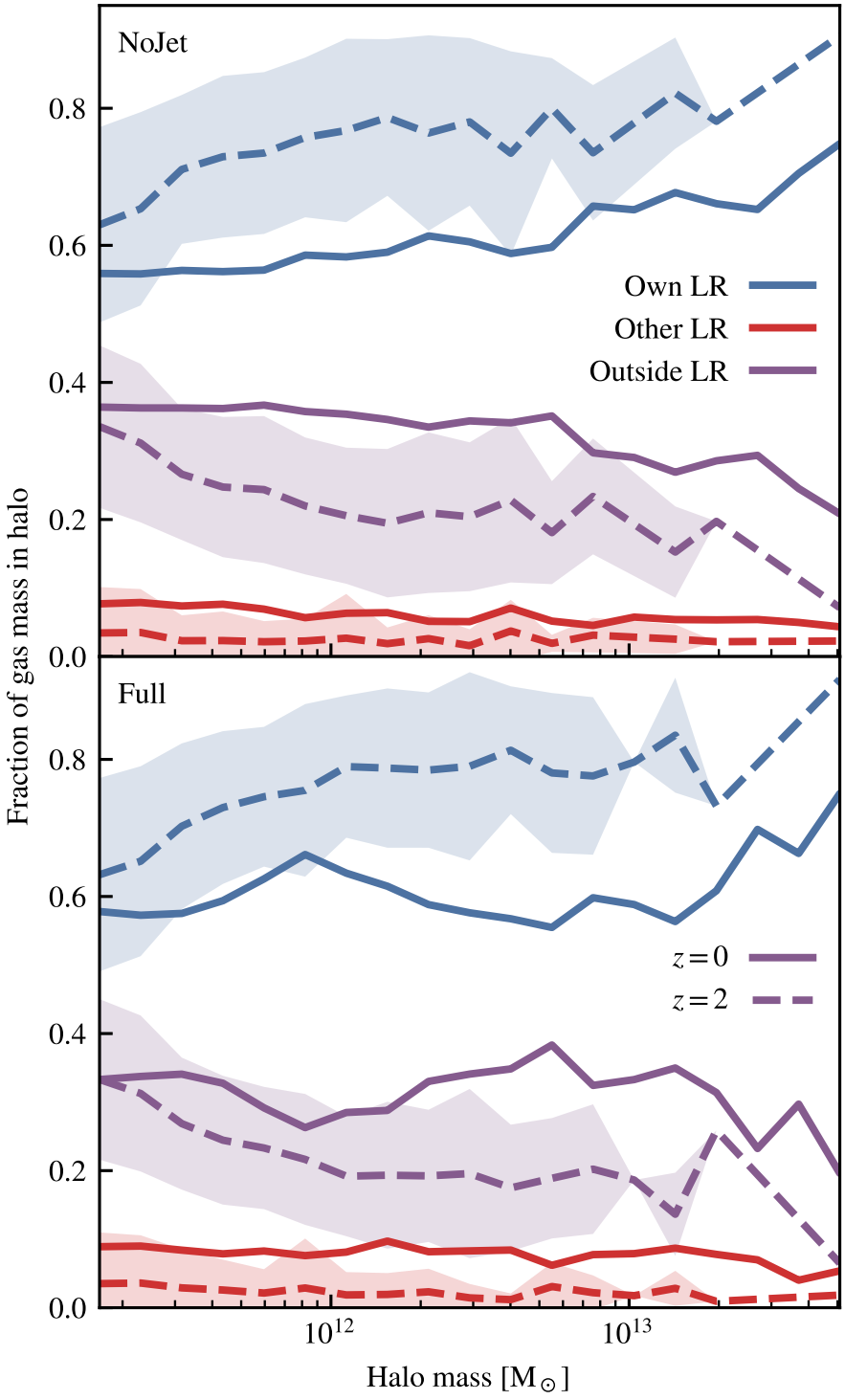


Figure 4.12: The fraction of gas mass in haloes at redshift $z = 0$ (solid) and redshift $z = 2$ (dashed) as a function of halo mass at that redshift, split by Lagrangian component. Scatter is shown only for the $z = 2$ results. The top panel shows the results from the `NoJet` simulation, with the bottom showing the full Simba model.

the Lagrangian regions of other haloes increases by a factor of two (or more) at all halo masses, with the fraction of transfer from outside Lagrangian regions remaining constant or again increasing by a factor of two dependent on the resident halo mass.

All of this must be explained within the context of very different baryon fractions for all haloes at $z = 0$. One possibility is that the majority of gas gained from outside of a haloes own Lagrangian region remains in the CGM, with very little of it making it into the disk (this is supported by the very low fraction of halo stars that originate from transfer, see Fig. 4.11). This gas can then be swept out of the halo either by stellar winds or (ejective) AGN feedback. Alternatively, if the main pathway for feedback is preventative, and the gas outside of haloes is well mixed, then this assembly of baryons would be curtailed equally for all Lagrangian components. A further investigation of these transfer properties (considering differences between the galaxy disks and the CGM) would be well suited for follow-up work using higher resolution simulations.

4.4.6 Transfer *out* of Lagrangian Regions

Let us now consider the fates of baryons that begin their lives in Lagrangian regions. This material has three possible fates, as shown in Fig. 4.13: it can end up in the same halo as the dark matter from that Lagrangian region (blue line), in another halo (red line), or outside of any halo in the IGM (purple line). Here, we we plot the fraction of LR mass at $z = 0$ from each component as a function of their Lagrangian region mass (this is the sum of the baryons and dark matter contained within that Lagrangian region). The Lagrangian region mass is somewhat higher than the eventual halo mass due to the baryon fractions of redshift $z = 0$ haloes being below the cosmic mean. We see that, below a halo mass of $10^{13.5}$ $M_\odot$, only around 20-30% of the baryons initially present in the Lagrangian region make it in to the halo by $z = 0$. Only above a halo mass of $10^{13.5}$ $M_\odot$ do haloes become strong enough attractors to retain the majority of their baryons. Despite the clear trend, this result is somewhat uncertain due to the very small number of these very large haloes present in our $50h^{-1}$Mpc box. On top of this initial structure, we see that there is a dip in the retained fraction of baryons between 10^{12} and 10^{13} $M_\odot$. We speculate that this is due to the increased efficiency of AGN feedback in haloes in this mass range, allowing for more gas in central objects to be expelled, however making a direct connection would require significant investigation. It is worth noting that without the AGN jets (i.e. in the `NoJet` run), the baryon fraction of haloes in this mass range is approximately $f_b/f_{b,c} = 1$.

Finally, we find that up to 10% of the Lagrangian region gas of low-mass haloes ($< 10^{12}$ $M_\odot$) can be transferred to other haloes, decreasing at higher masses. A larger cosmological volume with more objects is required for a full study of objects at masses higher than $M_H > 10^{13}$ $M_\odot$, but these trends

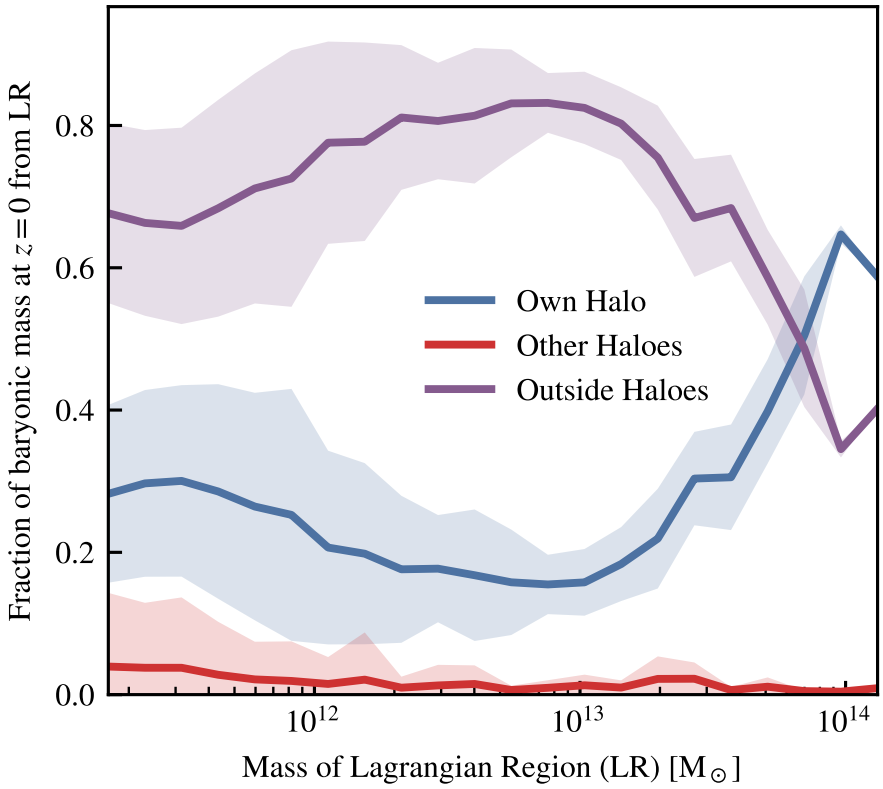


Figure 4.13: The fate of gas that begins in Lagrangian regions, as a function of initial Lagrangian region mass. The blue line shows the fraction of baryons that reside in the halo that defines the Lagrangian region at redshift $z = 0$, the red line shows the the fraction of baryons that lie in a different halo, and the purple line shows the baryons that lie outside of any halo at redshift $z = 0$. All but the most massive objects in the box struggle to retain more than 30% of their baryons due to various factors, see the text for details. The fraction of mass retained in the corresponding halo (blue) is the lowest in the mass range $10^{12} - 10^{13}$ $M_\odot$.

point towards inter-Lagrangian transfer being fuelled by accretion of gas that is either expelled or stripped from lower mass haloes by higher mass objects. A plausible physical scenario is that early feedback leading up to redshift $z = 2$, where star formation (and hence stellar feedback) peaks, expels significant quantities of gas from lower mass haloes that can then be swept up at later times from the IGM by all haloes. Higher mass haloes at this redshift may have a strong enough gravitational potential to enable their stellar winds to be more efficiently recycled, preventing them from being sources of inter-Lagrangian transfer.

The combination of the baryons that are retained by haloes (Fig. 4.13) and the baryons that they manage to accrete from sources outside their Lagrangian region (Fig. 4.11) is seen in the baryon fraction of haloes, shown in Fig. 4.14 split by Lagrangian component. Here, we split the overall baryon fraction (relative to the cosmic mean) into three Lagrangian components, coloured by the baryons from the haloes own Lagrangian region (blue), other Lagrangian regions (red), and from outside any Lagrangian region (purple). In general, we see that there is a trough in the baryon fractions of haloes with a mass between 10^{12} $M_\odot$ and 10^{13} $M_\odot$, with the baryon fraction reaching the cosmic mean for the largest objects in the box (with a halo mass of 10^{14} $M_\odot$). The baryon fraction returning to $f_b = 1$ for these very large haloes is not due to these haloes retaining all of their Lagrangian gas, however; it is a complex interplay between their accretion from outside, from other Lagrangian regions, and from the significant component that originates outside of any Lagrangian region. These objects are clearly able to mix outside of their halo boundaries, swapping gas with the IGM, as has been shown in several studies through 'splashback' (Mansfield et al., 2017; Diemer et al., 2017).

The dip in baryon fraction between 10^{12} $M_\odot$ and 10^{13} $M_\odot$ in halo mass corresponds to the dip in retained baryons in a similar mass range in Fig. 4.13. However, within this mass range, it appears that the fraction of baryons originating from outside the Lagrangian region is more significantly affected than the fraction of baryons from the haloes own Lagrangian region (reduced by 50% as opposed to 20%). This points to a more complex accretion history for these objects, with a mixture of ejective feedback (in general reducing the amount of retained baryons) and preventative feedback (in general reducing the amount of baryons from outside of the corresponding Lagrangian region) shaping their baryonic content.

The halo mass range where this dip occurs, 10^{12} $M_\odot$ to 10^{13} $M_\odot$, corresponds to the range in stellar mass (10^{10} $M_\odot$ to 10^{11} $M_\odot$, Moster et al., 2013) where black holes can begin to efficiently quench galaxies, as seen in observations (Kauffmann et al., 2003) and simulations (Bower et al., 2017; Taylor et al., 2017). The baryon fraction of haloes is shown as a function of the host stellar mass in Fig. 4.15, with a significant dip in the baryon fraction, primarily caused by the reduction in material from outside

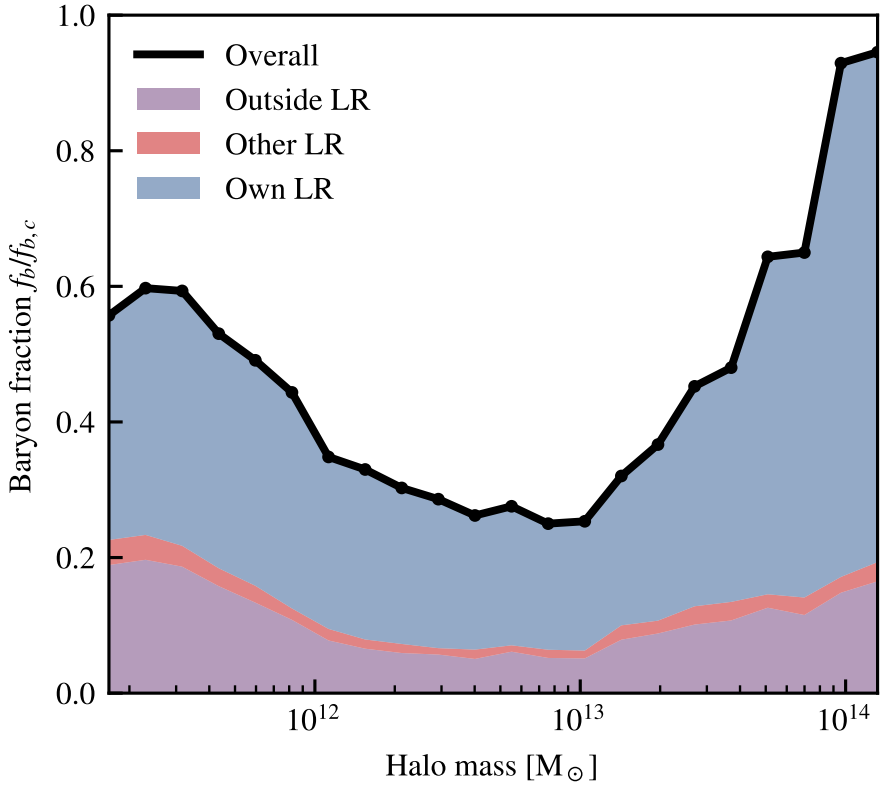


Figure 4.14: The baryon fraction f_b relative to the cosmic baryon fraction $f_{b,c}$ shown as a function of halo mass. The coloured bands show the contributions to the baryon fraction from various Lagrangian components. Note that the halo mass is again the mass of particles inside the spherical overdensity radius consistent with Bryan & Norman (1998).

Lagrangian regions, around a stellar mass of $M_{*,\mathrm{crit}} = 3 \times 10^{10}$ $\mathrm{M}_\odot$ (Kauffmann et al., 2003). At this stellar mass, black holes begin to accrete efficiently (Anglés-Alcázar et al., 2017a) and can effect non-linear reactions on the galaxy to quench it, either by expelling gas by making it buoyant (Bower et al., 2017), or in the case of Simba preventing gas from assembling into the halo by heating it (Davé et al., 2019). In Simba, around $M_{*,\mathrm{crit}}$, preventative feedback specifically helps restrict the baryonic content of the galaxy by preventing the influx of material from outside the Lagrangian region, suggesting that this material is more prone to preventative effects.

4.5 Variations on numerical parameters

The above halo-based metrics will have a certain level of dependence on the choice of halo finder used. In an attempt to ensure independence of the results from such factors, the above analysis was repeated with the 3D friends-of-friends (FoF) halo finder included in the `yt` package (Turk et al., 2011). We also repeated the analysis with the VELOCIraptor 6D FoF finder (Elahi et al., 2019). The

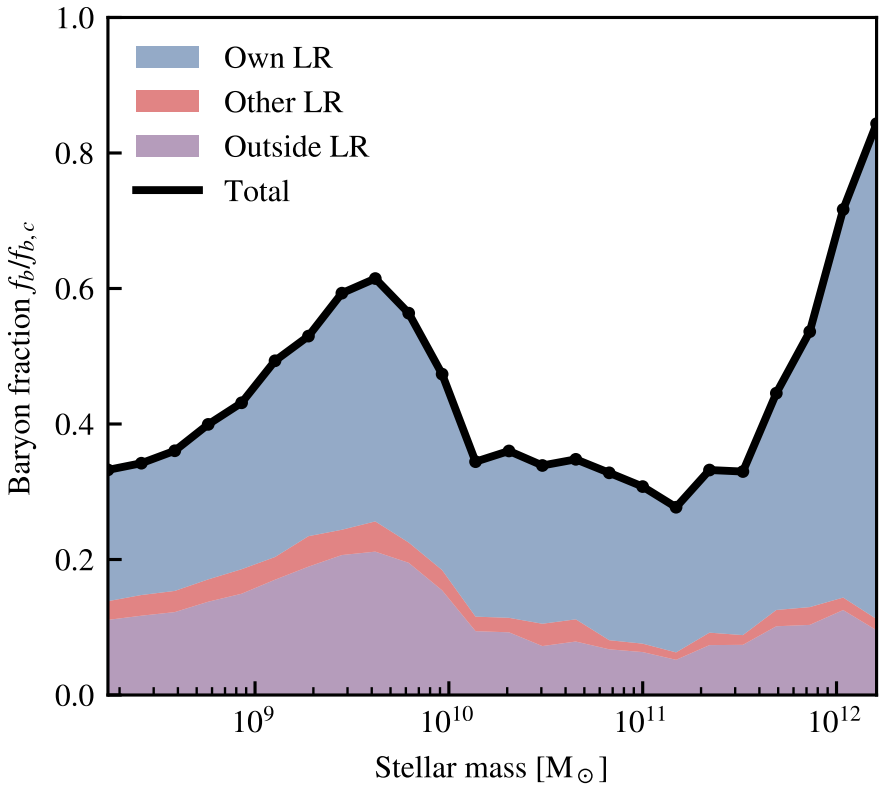


Figure 4.15: The baryon fraction f_b of haloes as a function of *halo* stellar mass. This is the same as Fig. 4.14, but more clearly shows the restricting effect of AGN feedback around a stellar mass of 10^{10} $M_\odot$.

latter will disentangle active mergers, but as active mergers make up a small fraction of the galaxy population, the above results are qualitatively unaffected and only change quantitatively to the 5% level. The use of a FoF finder, rather than the spherical overdensity finder found in AHF, did not qualitatively change the results.

In this section, we explore the implications of extending the Lagrangian region of haloes while retaining the ability to capture non-uniform shapes. We find that, in general, including more particles in the definition of the Lagrangian region (than are present in the halo) leads to a fractionally higher level of inter-Lagrangian transfer and more self-contribution to the final halo mass at the expense of transfer from outside any Lagrangian region. This is expected, as now many more particles are classified as being present in the Lagrangian region.

4.5.1 Filling in Holes in Lagrangian Regions

Our method for producing Lagrangian regions simply uses the dark matter particles from a given halo; this naturally leads to a very diffuse Lagrangian region. To see how the diffuse nature of these regions affects our results, we smooth out the Lagrangian regions, by extending the procedure that was used to extend the regions from the dark matter to the gas. This works as follows:

1. For every dark matter particle not in a Lagrangian region in the initial conditions, find the nearest n neighbours.

2. Find among the neighbours the maximal Lagrangian region ID, ensuring the particles are always assigned to the Lagrangian region corresponding to the lowest mass $z = 0$ halo.

3. Assign the particle the same Lagrangian region ID.

The choice to assign the particles to the lowest mass halo, rather than the higher mass halo, was made to ensure that spurious transfer into the lower mass halo was avoided wherever possible. This means that the expectation is that with this metric the level of inter-Lagrangian transfer will increase with respect to the fiducial Lagrangian region identification method. This results with the particles given to the haloes of a higher mass showing negligible deviation from the fiducial result (see Fig 4.16).

Note how smoothing the Lagrangian regions does have the expected effect of inducing more inter-Lagrangian transfer, and does increase the proportion of baryons that are classified as retained as the Lagrangian regions are filled out. Despite this, the overall trends with respect to halo mass remain, with a significant (>20%) contribution from gas from outside Lagrangian regions in haloes.

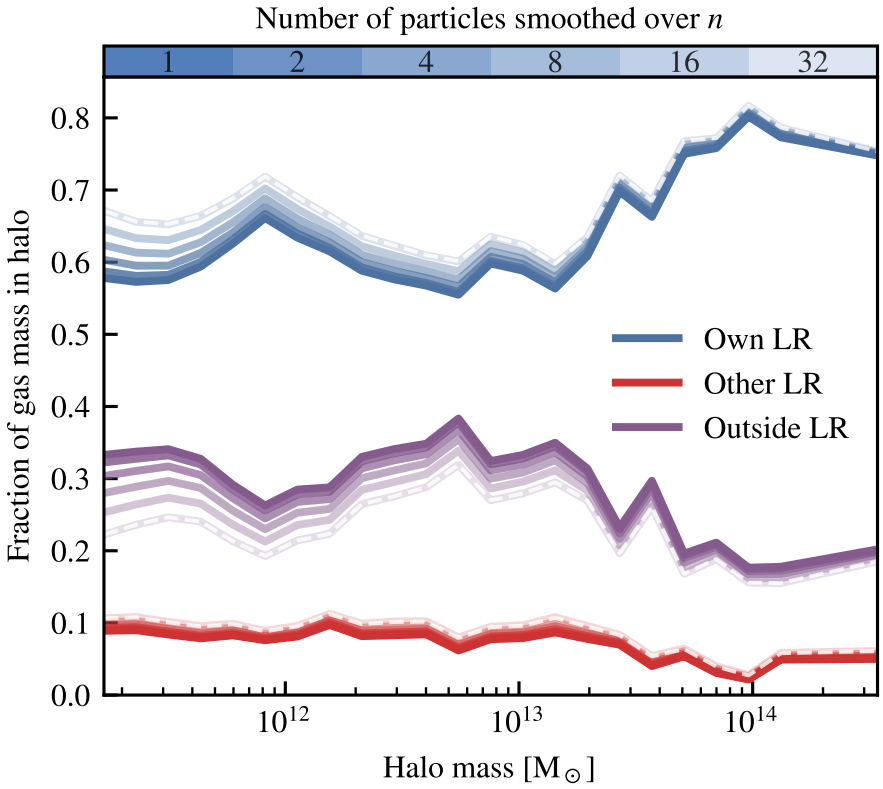


Figure 4.16: The same as Fig. 4.11, but including Lagrangian region smoothing. Each line, coded by transparency, shows the fraction of gas mass in a halo from each component when the Lagrangian regions have been smoothed by 1 (i.e. the fiducial result), 2, 4, 8, 16, or 32 particles (from darkest to lightest respectively). The white dashed line shows the result for the 32-smoothing case where the particles are given to the highest, rather than lowest, mass haloes; no difference is seen here suggesting that there is little overlap between the Lagrangian regions on these scales. See the text for the details of how this smoothing is constructed.

4.5.2 The sizes of Lagrangian regions

In Fig. 4.8 we saw that there was a large amount of gaseous matter inside haloes from outside any Lagrangian region. It may be reasonable to assume that this gas corresponds to dark matter that is simply sitting just outside of the halo edge, perhaps within the so-called 'splashback radius'. The estimates for this radius range between 0.8 and $1.5R_{\rm vir}$ (More et al., 2015; Diemer et al., 2017), and hence below we consider the situation where we extend the region around the halo that contributes to the Lagrangian region. This is done in the following way:

1. For every halo, find its current virial radius $R_{\rm vir}$. This contains all particles at redshift $z = 0$ that we consider to be within the halo.

2. Now consider a new radius, $R_{\rm vir} \leq R_{\rm LR} \leq 1.5R_{\rm vir}$, and find all dark matter particles within this region from the halo centre. These dark matter particles are now defined to lie within the Lagrangian region of that halo.

3. ID match these particles in the initial conditions to define the new Lagrangian region, extending to the gas in the usual way.

The effects of this process on the gas component of Fig. 4.11 (where it is most significant) are shown in Fig. 4.17.Here we see that there is a significant change in the fraction of mass in the halo at redshift $z = 0$ from outside any Lagrangian region, especially when going to $R_{\rm LR} = 1.5R_{\rm vir}$. This large change is expected, though, as we now have included a volume that is three times larger than the initial halo in the Lagrangian region classification; taking this extreme value for all haloes really is a 'worst-case' scenario. The inter-Lagrangian transfer remains at a similar level despite the increase in radius. Note that there will be no extra mass included in the haloes here, with particles simply changing their Lagrangian allegiances.

We chose this specific process, increasing the radius of our Lagrangian region rather than the whole halo, to prevent us from simply re-defining our halo size and including more gas as well (as in this case, the transfer across the halo boundary would simply be moved to a larger radius). There is the additional complication that haloes are known to not be spherical, and there may be particles frequently moving in and out of haloes at the boundary region. However, the main aim of this chapter is to use a very simplistic approach to gather insight into the physical processes and composition of haloes.

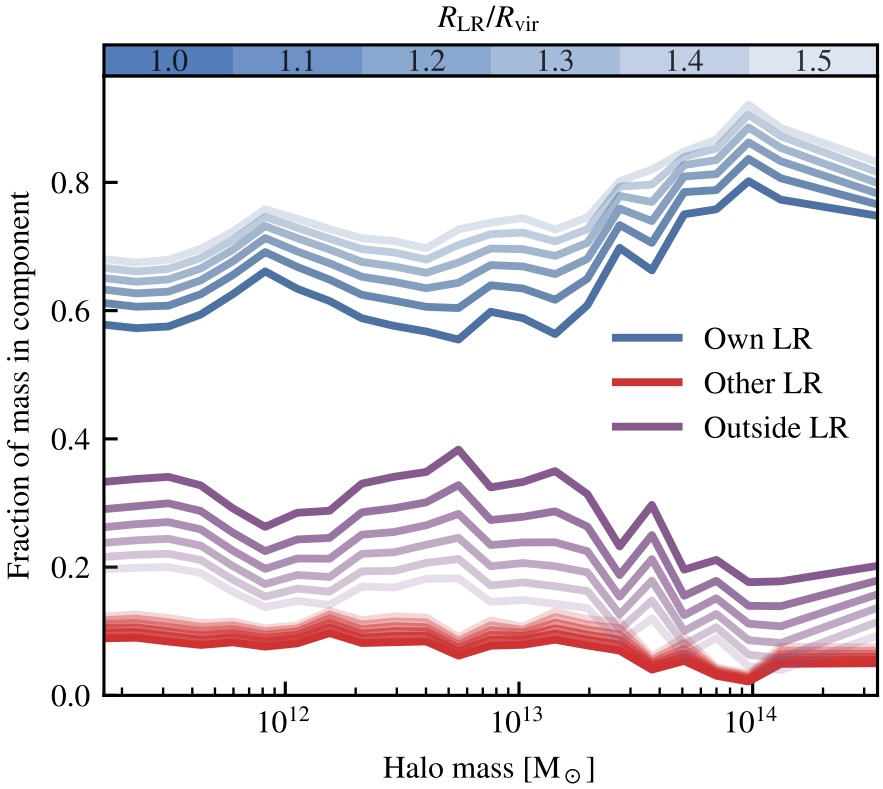


Figure 4.17: The same as Fig. 4.11, but now showing how the Lagrangian make-up of haloes is changed with an increasing radius for the definition of the Lagrangian region. Lighter colours correspond to larger radii, going in steps of $0.1R_{\rm vir}$ from 1.0 to 1.5.

4.6 Discussion and Conclusions

We have developed two novel metrics that describe the movement of baryons throughout a cosmological simulation with respect to the dark matter, and employed them to investigate the Simba simulations and sub-grid model. The first of these metrics, the *spread metric*, shows that:

- Dark matter can be spread up to $7.5h^{-1}$Mpc away from their inital mass distribution throughout the course of a cosmological simulation. This has been validated with two simulation codes, Gizmo and Swift.
- Gas can be spread to even larger distances, with the distance dependent on the physics included in the sub-grid model. For the Simba galaxy formation model with AGN jets, we find that gas can be spread to up to $12h^{-1}$Mpc (notably this is increased fom $7h^{-1}$Mpc in the model without AGN jets) throughout the course of the simulation in a box that is only $50h^{-1}$Mpc in size, with 40% (10%) of baryons having moved $> 1h^{-1}$Mpc ($3h^{-1}$Mpc). This is despite this powerful form of feedback only directly interacting with 0.4% of particles, and points towards significant quantities of gas being entrained by these jets. It remains to be seen if this will increase further with higher mass objects in larger boxes.
- Stars in the simulation show a very similar level of spread to the dark matter, suggesting that the gas particles that stars form out of remain tightly coupled to the dark matter. This implies that the spreading of stars by gravitational dynamics dominates over the spreading of their gas particle progenitors by feedback.
- Using the spread metric to select particles, we have shown that dark matter that is spread to large distances forms the diffuse structure within and around haloes, with lower spread dark matter forming substructure within haloes. When extending this to the gas, we find that the baryons that are spread the most are those that reside in the diffuse structure around haloes, with this structure being created by the energetic feedback present in the Simba model. We suggest that this spread metric may be a useful, highly computationally efficient, way of selecting particles that have been entrained by feedback processes that are not tagged during the injection of energy.

The second of these metrics, which considers the baryonic make-up of haloes at $z = 0$ split by the Lagrangian origin of the particles, shows that:

- Approximately 40% of the gas in an average $z = 0$ halo did not originate in the Lagrangian region of that halo, with around 30% originating outside any Lagrangian region, and 10% orig-

inating in the Lagrangian region of another halo. This suggests that *inter-Lagrangian transfer* is prevalent throughout the simulation, with haloes interchanging particles between $z = 2$ and $z = 0$ thanks to energetic feedback pathways. There will be some transfer of dark matter particles between haloes, but as the adiabatic tests shows this is a relatively limited effect.

- The majority of the stellar component of haloes (90% above a halo mass 10^{12} $M_\odot$) originates from the Lagrangian region of the same halo, as expected given the similar large-scale spreads of the stellar and dark matter.

- Below a halo mass of 10^{13} $M_\odot$, haloes can only retain approximately 20-30% of the baryons from their Lagrangian region, with the majority of these baryons being lost to the IGM. Above this mass, haloes become strong enough gravitational wells to retain the majority of their baryons (up to 60%) by around 10^{14} $M_\odot$ halo mass, although this result is somewhat uncertain due to the lack of objects in this mass range in the $50h^{-1}$Mpc simulation box used here.

- Haloes with mass $M_H > 10^{13.5}$ $M_\odot$, despite having a baryon fraction comparable to the cosmic mean, still show significant levels of transfer from other haloes and from outside any Lagrangian region. This suggests a complex cycling of baryons with approximately 20% of their baryonic mass being 'swapped' with the IGM by $z = 0$.

- Different Lagrangian components, as they make up the baryon fraction of haloes, are affected differently by feedback mechanisms at different halo masses. In the halo mass range 10^{12}–10^{13} $M_\odot$, the component of baryonic mass from outside of the Lagrangian region is halved, whereas the component from the haloes own Lagrangian region is only reduced within 20%; this highlights the importance of preventive feedback for the baryon fraction of haloes.

Our results add a new perspective to the connection between baryon cycling and galaxy evolution. Using large volume simulations including momentum-driven winds, Oppenheimer et al. (2010) showed that most stars likely form out of gas that has previously been ejected in winds, and more recent zoom-in simulations agree with the prevalence of wind recycling (Christensen et al., 2016; Anglés-Alcázar et al., 2017a; Tollet et al., 2019). Using the FIRE simulations, Anglés-Alcázar et al. (2017b) further showed that the intergalactic transfer of gas between galaxies via winds can provide up to a third of the stellar mass of Milky Way-mass galaxies. Here we have introduced the concept of inter-Lagrangian transfer, which represents the extreme case of transfer of baryons between individual central haloes. For the Simba simulations, we find that only a small fraction (<5%) of the stellar mass of haloes can be made up from inter-Lagrangian transfer gas, suggesting that most intergalactic transfer originates

from satellite galaxies and is thus confined within Lagrangian regions. It is nonetheless quite significant that gas exchanged between Lagrangian regions can fuel star formation in a different halo at all. In addition, we do find a significant contribution (<20%) of inter-Lagrangian transfer to the gas content of haloes at z = 0. Recently, Hafen et al. (2019b,a) has highlighted the contribution of satellite winds to the gas and metal content of the CGM in the FIRE simulations. Our results suggest that the origin of the CGM of galaxies is linked to larger scales than previously considered.

These results provide two possible main implications for current works. The first is the implications for semi-analytic models of galaxy formation. These models, by construction, tie the baryonic matter to dark matter haloes; they contain no prescription for gas that explicitly originates from regions where the dark matter does not end the simulation in a bound object. Also, whilst there has been some effort by Henriques et al. (2015); White et al. (2015) and others to include wind recycling into these models, there is currently no semi-analytic model that includes any concept of baryon transfer between un-merged haloes or baryonic accretion rates significantly different to that expected from the dark matter component.

The second implication is for zoom-in simulation suites. These suites typically construct their initial conditions by considering the cubic volume, ellipsoid, or convex hull in the initial conditions containing the dark matter particles that are located within a given distance (typically $2-3R_{\rm vir}$) of the selected halo at $z = 0$ (see e.g. Onorbe et al., 2014). However, our results highlight that the shapes of the causally connected regions in gas and dark matter may be significantly different. For example, the Latte (Wetzel et al., 2016) suite uses an exclusion region for high resolution particles of around $1.5h^{-1}$Mpc while we find that 10% of cosmological baryons can move >3 Mpc away relative to the original neighbouring dark matter distribution. While zoom-in simulations are constructed to avoid contamination of low-resolution particles into the high-resolution region, our results suggest that they may miss a flux of external baryons into the high resolution region. In practice, contamination from external sources will be somewhat mitigated by the usual choice of isolated haloes, but future work should consider these effects for zoom-in suites that have a full hydrodynamical simulation for their parent.

The results presented here are based on the Simba model, which is in good agreement with a wide range of galaxy (Davé et al., 2019) and black hole (Thomas et al., 2019) observables, but are clearly dependent on the feedback implementation. Other galaxy formation models may yield different results, especially those with drastically different implementations for AGN feedback, such as the purely thermal feedback in the EAGLE model (Schaye et al., 2015). The spread metric represents a unique tool to characterize the global effects of feedback and will enable novel comparisons between existing

cosmological simulations. Future work should also address the connection between baryon spreading and galaxy/CGM observables, as well as investigate baryonic effects on cosmological observables (Schneider & Robertson, 2015; Chisari et al., 2018) in the context of the spread metric.

Chapter 5

Conclusions & Future Work

In this thesis we have explored the design and implementation of a selection of numerical methods and their application to astrophysical cosmological simulations.

5.1 Summary

In Chapter 2 we investigated a commonly used numerical scheme for solving the fluid equations called Pressure-SPH. This scheme was shown to lead to significant errors when injecting (common in cosmological astrophysical simulations for modelling stellar and AGN feedback) or removing energy (again common, usually for modelling radiative cooling). In these scenarios the evolution of the energetic variable (either internal energy u or entropy A) is corrupted and can lead to significant force errors.

The P-SPH method has been used in many large simulation suites, such as Fire (Wheeler et al., 2015), Mufasa (Davé et al., 2016), and EAGLE (Schaye et al., 2015), with these simulation suites and their associated publications either not considering these issues, or providing solutions that we have now shown to be ineffective. This chapter really highlights how important it is that we fully understand the numerical methods we employ in these large scale simulations.

In Chapter 3, we developed a new SPH model that does not suffer from the same errors as the Pressure-SPH models in Chapter 2, but is just as computationally efficient. Many alternative hydrodynamics models are available, but one of the largest challenges in large-scale simulations of the universe is minimising memory footprint, to run a simulation as large and with as many particles as possible. This then makes it imperative to increase the accuracy to cost ratio, rather than simply increasing accuracy.

The Sphenix model was shown to perform well on a large suite of hydrodynamics test problems (and converge at the expected rate) whilst remaining computationally cheap and being able to interact with the sub-grid models ubiquitous throughout cosmological simulations.

In Chapter 4 we developed two new techniques to analyse numerical simulations: the spread metric, and inter-Lagrangian transfer. These two techniques were applied to the Simba simulation suite (Davé et al., 2019). These novel techniques allowed for the investigation of the distribution of baryons throughout the universe in a model-independent way, with very few base assumptions.

Through the spread metric, we were able to demonstrate that AGN feedback has a significant impact on the redistribution of baryons throughout the universe, with the powerful jets in Simba spreading particles 15 Mpc away from their expected position at $z = 0$.

By analysing the redistribution of baryons from Lagrangian regions to haloes at $z = 0$, we were able to show that matter, in particular halo gas, is transferred between bound objects. We further discussed how this may lead to problems for zoom-in simulations, where a single $z = 0$ halo is simulated in an under-refined background region.

Although the Lagrangian transfer and spread metric are not observationally viable (due to their reliance on discrete particles and a knowledge of the history of the particles), they do allow us to confirm our existing intuition with few underlying assumptions. Other methods for studying the baryon cycle (see e.g. Anglés-Alcázar et al., 2017b; Mitchell et al., 2020b) require complex particle tracking techniques, relying on inflows and outflows at specific radii. Simple numerical methods can capture the same concepts without the need to characterise all of these potential systematics, and can allow us to build our intuition.

The work in this thesis demonstrates two major points: the importance of having a deep understanding of the numerical methods employed in simulations, and that purely numerical techniques can allow us to develop intuition for the important physical processes at play.

5.2 Future Work

The past ten years have been a massive turning point in our understanding of galaxy formation, thanks to the production of three key simulations: Horizon-AGN, EAGLE, and Illustris-TNG (Dubois et al., 2014; Schaye et al., 2015; Pillepich et al., 2018). These three simulations allowed, for the first time, a statistical sample of inter-connected galaxies, simulated with full hydrodynamics, to be studied by the community.

In this thesis we have discussed much about the sub-grid models in galaxy formation simulations, in particular their coupling to the hydrodynamics scheme chosen. These sub-grid models are the leading order component in determining the output galaxy properties, with things like the accuracy of the hydrodynamics or galaxy calculation very much taking a secondary role (Hopkins et al., 2018). The complexity of sub-grid physics is growing rapidly with time, with a number of groups now explicitly including a cold gas phase (e.g. Marinacci et al., 2019; Ploeckinger & Schaye, 2020). Simultaneously, confidence in simulations is increasing, and as such various studies are studying the formation history of individual objects in detail (e.g. Evans et al., 2020; Mitchell et al., 2020a).

Despite this increased confidence and rapid pace of development, we have still not yet fully understood the models currently in use. These models typically contain several free parameters, with the main simulation simply choosing some 'best' selection of parameters based on comparison to various observational datasets (usually by eye). Sub-grid models are typically stochastic, and yet there is little literature available discussing the impact of this stochasticity on the resultant properties of galaxies.

With these points in mind, I outline some avenues for future work.

5.2.1 Understanding Random Variation

Cosmological galaxy formation simulations typically have constrained spatial resolution and as such require sub-grid models to emulate physics that occurs below the resolved length scale. They are, additionally, temporally constrained; the simulation does not proceed in real-time[1], which must additionally be respected by the sub-grid scheme and a very coarse temporal brush must be used to ensure the completion of simulations.

Let us for instance consider star formation. A typical simulation will compute the star formation rate density ($\dot{\rho}_*$) to depend on the local gas density (ρ_g) following a (Schmidt, 1959) law,

$$\dot{\rho}_* = \frac{\rho_g}{t_{\rm ff}}, \tag{5.2.1}$$

with $t_{\rm ff}$ the free-fall time of the gas. Discretising this to a finite mass simulation (converting densities to masses m), and folding in the time-step of the gas Δt,

$$m_* = m_g \frac{\Delta t}{t_{\rm ff}}, \tag{5.2.2}$$

leaves us in a predicament; if the time-step is much less than the free-fall time (which would be required for a stable evolution), we will need to form many star particles with low masses, which will

[1]Galaxy formation is one of the very few fields where simulations have a much smaller wall-clock time than their simulated evolution.

lead to both memory issues and errors in the gravity calculation (Ludlow et al., 2020). Instead, this form of the equation is transformed into a stochastic model, with the probability that the gas particle is transformed into a star particle

$$P_*(\Delta t) = \frac{\Delta t}{t_{\rm ff}}. \tag{5.2.3}$$

This solves our discretisation issue, but brings in a new problem; our model is no longer deterministic. The timing of individual star formation events (and more importantly their associated stellar feedback events) can have a significant impact on the evolution of individual galaxies (Davies et al., 2020). It is additionally not clear that this process is well converged relative to the number of events, so convergence with resolution is unlikely for such a scheme without changes to other parameters.

The inherent stochasticity of the current range of sub-grid models then means that correlating the formation history of haloes and the galaxies that reside in them (the so-called galaxy-halo connection, see Wechsler & Tinker, 2018, for a review) is infeasible in anything other than a statistical sense with the current range of large volume simulations (Keller et al., 2019; Genel et al., 2019).

The first step to understanding the impact of the stochastic implementation of sub-grid models is to characterise it. This, so far, has been missing from the literature - only Keller et al. (2019) and Genel et al. (2019) have attempted such a study.

To characterise the level of random variation in galaxy properties, the same simulation must be performed multiple times with some change to the random numbers. This can be a change to the random seeds if a pseudorandom number generator is used, or in the case of Swift such divergence occurs naturally due to the variation in roundoff errors originating from the task-based parallelism implementation.

In Fig. 5.1 we show the level of variation in the most basic of output galaxy properties, the stellar mass, in a simulation performed with the Swift code (Schaller et al., 2018) and the EAGLE model (Schaye et al., 2015). The halo catalogues were created with VELOCIraptor (Elahi et al., 2019)[2]. In this figure, $16 \times (25\mathrm{Mpc})^3$ volumes are shown (at $z = 0$), with each point showing the variation in the stellar mass of the same galaxy, realised 16 times independently with different random numbers.

Fig. 5.1 shows how the introduction of new stochastic physics, in this case AGN feedback (which occurs in galaxies of $M_* \gtrapprox 10^{10}\ \mathrm{M}_\odot$), can lead to an increase in the variability between random realisations. Below this threshold, the scatter in stellar mass decreases as would be expected for a Poisson sampling of the field (as shown in Keller et al., 2019).

[2] Haloes are matched by ensuring that they have at least 80% of the same dark matter particles as a similar halo in a dark-matter only version of the simulation, and we only match central haloes.

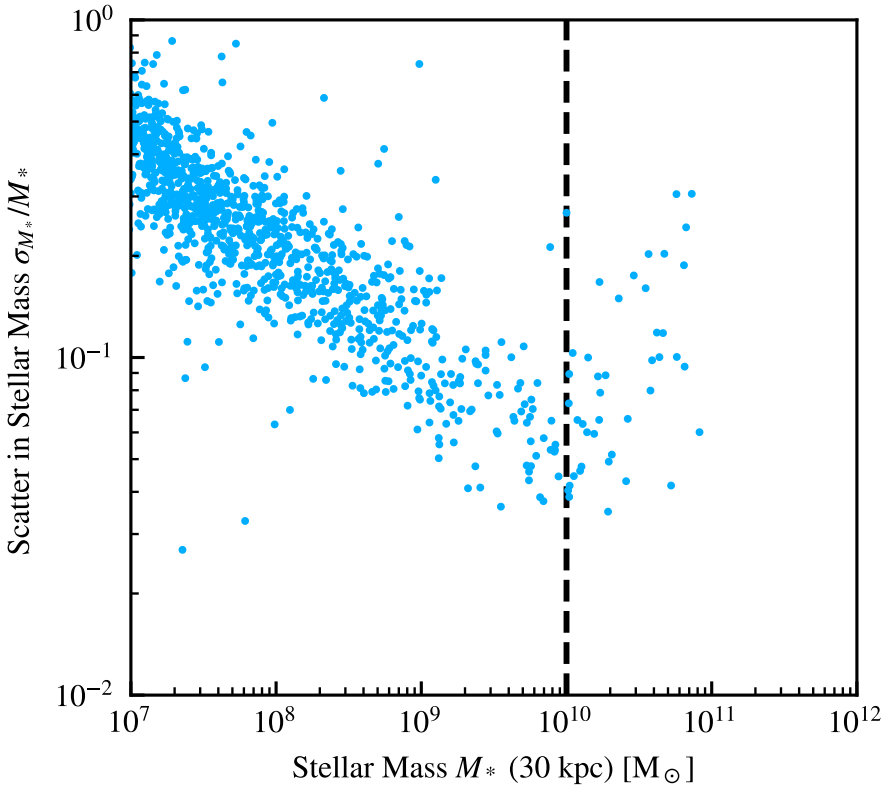


Figure 5.1: The scatter between 16 random realisations of galaxies in a $(25\ \mathrm{Mpc})^3$ volume with the EAGLE model. The scatter is calculated as the standard deviation of the stellar mass in the matched haloes. Below $M_* = 10^{10}\ \mathrm{M}_\odot$ the scatter reduces as roughly $\sigma_* \propto 1/\sqrt{M_*}$ as would be expected for Poisson noise. However, the introduction of new stochastic physics (AGN feedback) at roughly $M_* = 10^{10}\ \mathrm{M}_\odot$ (black dashed line) leads to an increase in scatter with mass as AGN events become more frequent.

There are many open questions that naturally arise that should be addressed in future work:

- What level of variability is there in various galaxy properties that are key for comparisons to observations (e.g. star formation rates, stellar masses, galaxy sizes)?
- Is the level of variation in various galaxy properties correlated with other galaxy or halo properties (we have already seen that the scatter in stellar mass is correlated with stellar mass, for instance)?
- Is the level of variation in one galaxy property correlated with variation in another galaxy property (e.g. if the stellar mass is high for one realisation is the black hole mass of that galaxy also high)?
- Is the level of stochasticity consistent between different galaxy formation models (e.g. EAGLE, Simba, Illustris-TNG)?
- Is the level of stochasticity in a population for a given model converged with mass and time resolution?
- Is the random scatter in galaxy properties responsible for the majority of the scatter in simulated scaling relations, and if so what are the implications for comparisons to observations?

5.2.2 Understanding the Impact of Free Parameters

Galaxy formation sub-grid models contain a number of free parameters that are specific to each suite. Some simulation suites attempt to model the Universe with as few parameters as possible (e.g. EAGLE, Crain et al., 2015), whilst others attempt to match the observed relations as closely as possible, usually necessitating more free parameters (e.g. Illustris-TNG, Pillepich et al., 2018).

The impact that the individual parameters have is generally studied by using physical reasoning, with several simulations performed with individual changes in a single parameter at a time (e.g. `WeakFB` and `StrongFB` in Crain et al., 2015, studying the impact of stronger or weaker stellar feedback). The problem with this approach is that frequently the impact of individual parameters is degenerate with others.

Techniques such as Sensitivity Analysis (see Oleskiewicz & Baugh, 2020, for an application to a SAM) allow for an investigation of the underlying impact of a single parameter in a model, but this usually requires a large number of model outputs (100-1000s per parameter) to yield reliable results,

which is usually infeasible for galaxy formation simulations due to computational constraints. Thankfully, techniques such as Gaussian Process Regression (see Bower et al., 2010, for an application to a SAM) allow for the generation of a large number of model outputs based on a small training set ($\sim$ a few per parameter).

Such a procedure allows for the *model* to be studied, rather than a specific pre-calibrated *instance* of the model, and hence allows for the discovery of truly missing physics. For instance, the original EAGLE simulation suite (using a single set of parameters) could not capture the sharp peak in the galaxy stellar mass-halo mass relation (Crain et al., 2015; Behroozi et al., 2019), showing a deficit in stellar mass in galaxies with halo mass $M_H \approx 10^{12}$ M$_\odot$. As this was not explicitly calibrated against (EAGLE was calibrated to the galaxy stellar mass function, stellar mass-size relation, and stellar mass-black hole mass relation) this is perhaps unsurprising, but it begs the question: was it the specific choice of parameters that led to this mis-match, or the underlying model and hence some missing piece of physics (e.g. magnetic fields, stellar radiation pressure, alternative dark matter models)? Only through a systematic exploration of the available parameter space for the whole model can we attempt to answer such questions.

5.2.3 Calibrating Sub-Grid Models

In the above discussion on understanding free parameters, there was a distinction made between the galaxy formation model and the model simulated with a specific choice of 'calibrated' parameters. The systematic exploration of the parameter space provides a significant improvement for calibrating the model than the current status quo. Currently, sub-grid models are calibrated by making changes to individual parameters based on physical reasoning (such as needing a lower stellar mass in low mass galaxies and increasing the stellar feedback strength) until a desired accuracy (by eye) to a set of galaxy formation scaling relations is attained. As such, it is not clear that the final choice of parameters represents a global best fit, rather just that it likely lies in a local error minimum. Additionally, it is not clear at the end of the process where the tensions between the calibration datasets lie (with the bias between the datasets purely determined by the human carrying out the experiment, rather than a pre-determined error structure), or how to repeat the calibration procedure in the future.

With the exploration of parameter space described above, we will be able to answer the following questions:

- Given a specific choice for the galaxy stellar mass function and galaxy mass-size relation, what is the global best fit simulation choice of parameters to these relations?

- What other scaling relations are naturally fit by this choice of parameters, and is there a unique tension in the parameter space between fitting a given relation and fitting the calibration datasets?
- What are the unique limitations of a given galaxy formation model and in what areas must the model be improved to better match the observed galaxy population?
- If a new external piece of physics (for instance, magnetic fields) is added, does this help address any *tensions* in parameter space between scaling relations (suggesting the model is closer to reality)?

5.3 Final Thoughts

The past decade has seen an explosion in the availability and use of cosmological galaxy formation simulations, and the future for them looks bright. By addressing issues relevant to the models themselves, such as further developing the underlying numerical methods and better calibration procedures, we can ensure that the predictions from galaxy formation simulations are reliable and repeatable.

Appendix A

Additional Details on SPHENIX Tests

A.1 Particle Costs

Different SPH models require different information stored per particle. Compared to a basic, 'Traditional' SPH model, SPHENIX requires a small amount of extra data to store things like the particle-carried artificial conduction and viscosity coefficients. The amount of data required increases for more complex models, such as those making use of the full shear tensor, like Wadsley et al. (2017), or additional corrections, like Rosswog (2020b). SPH models using an ALE (Arbitrary Lagrangian-Eulerian) framework (see Vila, 1999) require even more information as the particles carry flux information for use in the Riemann solver.

The amount of data required to store a single element in memory is of upmost importance when considering the speed at which a simulation will run. SPH codes, and SWIFT in particular, are bound by the memory bandwidth available, rather than the costs associated with direct computation. This means any increase in particle cost corresponds to a linear increase in the required computing time for simulation; this is why keeping the particles lean is a key requirement of the SPHENIX model. Additionally, in large simulations performed over many nodes, the bandwidth of the interconnect can further become a limitation and hence keeping the memory cost of particles low is again beneficial.

In Fig. A.1 we show the memory cost of four models: Traditional SPH (similar to to the one implemented in Gadget-2; Springel, 2005), SPHENIX, a model with the full shear matrix, and a SPH-ALE model similar to GIZMO-MFM (Hopkins, 2015), all implemented in the SWIFT framework. We see that SPHENIX only represents a 25% increase in memory cost per particle for significant improvement over the traditional model.

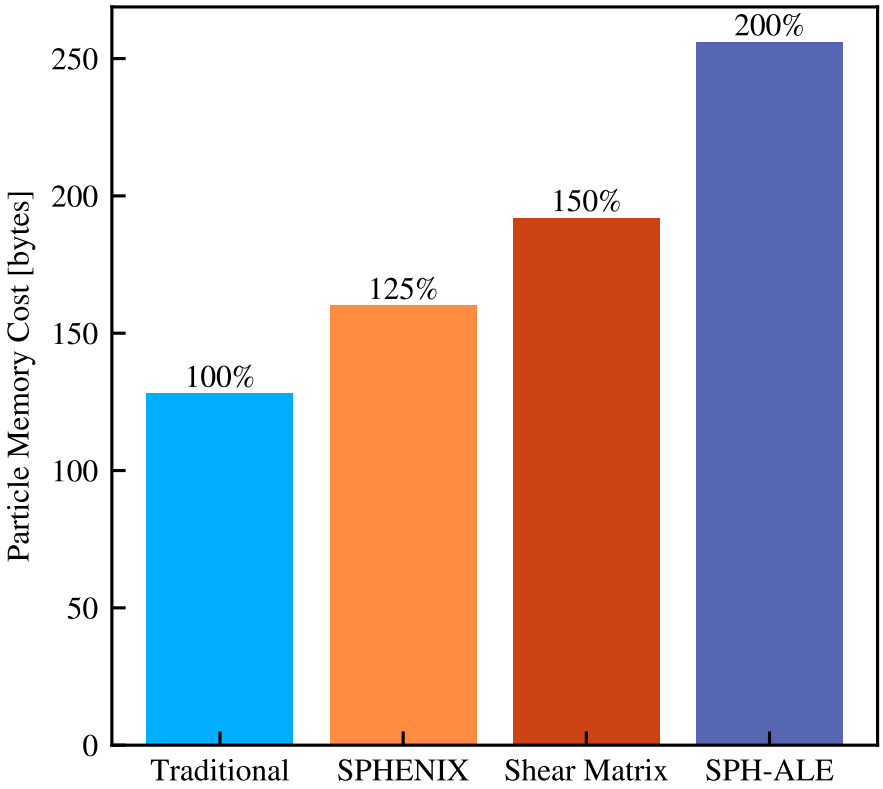


Figure A.1: Cost per particle (in bytes) for four different hydrodynamics models (see text for details). Percentages are give relative to the Traditional (similar to Gadget-2, with no artificial conduction and fixed artificial viscosity coefficients) model.

A.2 Conduction Speed

The choice for the conduction speed (Eqn. 3.3.22) in SPHENIX was primarily selected for numerical reasons. In a density based scheme, it is common to see significant errors around contact discontinuities where there are large changes in density and internal energy simultaneously to produce a uniform pressure field. In Fig. 3.3 we demonstrated the performance of the SPHENIX scheme on one of these discontinuities, present in the Sod Shock.

In Fig. A.2 we zoom in on the contact discontinuity, this time using glass files for the base initial conditions (of resolution 32^3 and 64^3), allowing for a more even distribution of particles along the horizontal axis. We use five different models,

- SPHENIX, the full SPHENIX model using the conduction speed from Eqn. 3.3.22.
- Pressure Term, which uses only the pressure-based term from Eqn. 3.3.22.
- Velocity Term, which only uses the velocity-based term from Eqn. 3.3.22.
- No Conduction, which sets the conduction speed to zero.
- Max Conduction, which sets $\alpha_{\rm D}$ to unity everywhere, and uses the conduction speed from Eqn. 3.3.22.

The first thing to note here is that the pressure term provides the vast majority of the conduction speed, highlighting the importance of this form of bulk conduction in SPHENIX and other models employing a density based equation of motion. Importantly, the conduction allows for the 'pressure blip' to be reduced to a level where there is no longer a discontinuity in pressure (i.e. there is a smooth gradient with x). Although the velocity term is able to marginally reduce the size of the blip relative to the case without conduction, it is unable to fully stabilise the solution alone. Pressure blips can lead to large pressure differences between individual particles, then leading to the generation of a divergent flow around the point where the contact discontinuity resides. This is the primary motivation for the inclusion of the velocity divergence-based term in the conduction speed. Along with the conduction limiter (see Eqn. 3.3.24 for the source term), if there is a large discontinuity in internal energy that is generating a divergent flow (and not one that is expected to do so, such as a shock), the velocity-dependent term can correct for this and smooth out the internal energy until the source of divergence disappears.

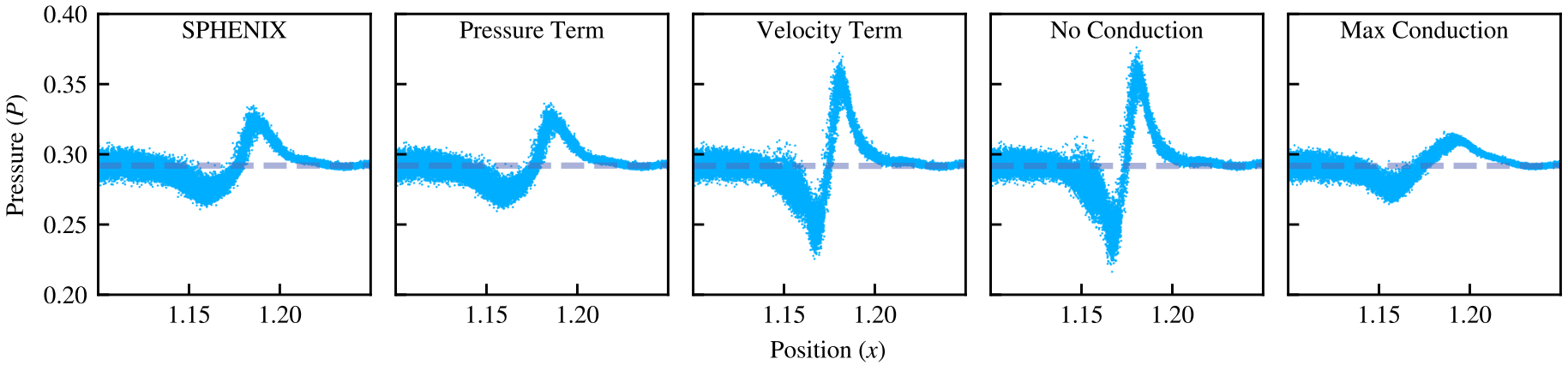


Figure A.2: The pressure contact discontinuity in the Sod Shock (Fig. 3.3) at a resolution of 32^3 and 64^3, but using glass files instead of the BCC lattices (this leads to significantly increased particle disorder, but more evenly distributes particles in the x direction enabling this figure to be clearer). Here we show a zoomed-in representation of all particles (blue points) against the analytical solution (purple dashed line). Each sub-figure shows the simulation at the same time $t = 0.2$, but with different forms for the conduction velocity (see text for details).

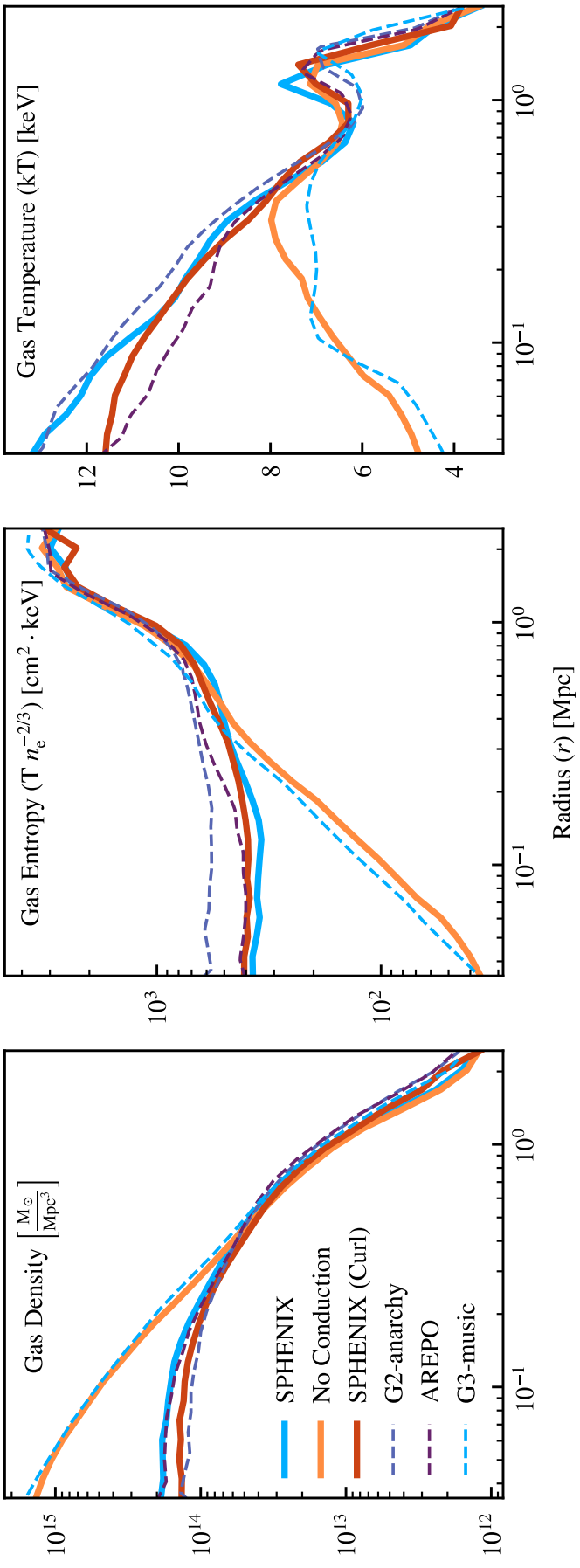


Figure A.3: Reproduction of Fig. 3.18 but including the line (red) for the version of SPHENIX performed with an explicit shear component in the conduction speed. We see no qualitative differences between the two models, with them both providing an entropy core at a similar level.

A.2.1 Alternative Conduction Speeds

The Sphenix conduction speed (Eqn. 3.3.22) contains two components: one based on pressure differences and one based on a velocity component. In Sphenix, as in a number of other models, this velocity component really encodes compression or expansion along the axis between particles.

The motivation for some alternative schemes (e.g. those presented in Wadsley et al., 2008, 2017) is shear between particles. To test if we see significant differences in our tests, we formulate a new conduction speed,

$$v_{D,ij} = \frac{\alpha_{D,ij}}{2}\left(\frac{|\vec{v}_{ij}\times\vec{x}_{ij}|}{|\vec{x}_{ij}|} + \sqrt{2\frac{|P_i - P_j|}{\hat{\rho}_j + \hat{\rho}_j}}\right). \tag{A.2.1}$$

that focuses on capturing the shear component of the velocity between two particles.

We again test this new formulation on some of our example problems. First, the nIFTy cluster, presented in Fig. A.3, shows little difference between the two formulations, with both providing a solution similar to other modern SPH schemes and grid codes.

The Kelvin-Helmholtz test again shows little difference (Fig. A.4), although there is a slightly increased growth rate of the perturbation at late times for the shear formulation.

We find no discernible difference between the two formulations on the blob test, as this is mainly limited by the choice of Density-SPH as the base equation of motion to correctly capture the initial break up of the blob from the centre outwards.

A.3 Maintenance of Hydrostatic Balance

The form of the conduction speed used in Sphenix based on pressure differences (Eqn. 3.3.22) has been conjectured to not allow for the maintenance of a pressure gradient against some external body force (for example a halo in hydrostatic equilibrium). The main concern here is that the pressure difference form of the conduction speed may allow thermal energy to travel down into the gravitational potential, heating the central regions of the halo. As Sphenix uses an additional limiter (Eqn. 3.3.24 for the source term) that only activates conduction in regions where the internal energy gradient cannot be represented by SPH anyway, this may be less of a concern. Additionally, there will be no conduction across accretion shocks due to the limiter in Eqn. 3.3.29.

In Fig. A.5 we show an idealised simulation of an adiabatic halo with an NFW (Navarro et al., 1996) dark matter density profile, and gas in hydrostatic equilibrium. The halo uses a fixed NFW potential in the background, with a mass of 10^{13} $M_\odot$, concentration 7.2, and a stellar bulge fraction of 1%. The

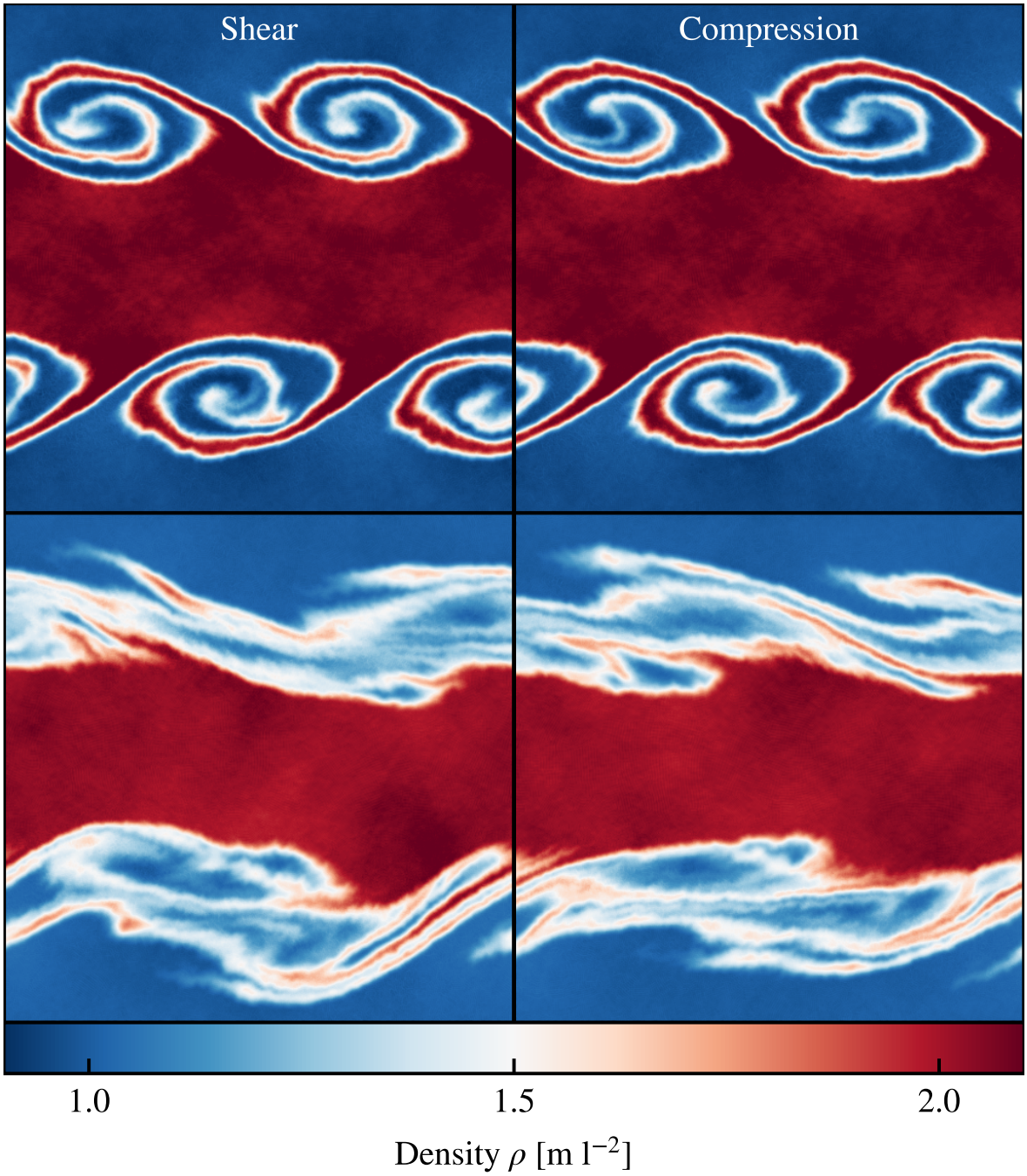


Figure A.4: Kelvin-Helmholtz test with a density contrast of $\rho_C = 2$ as in Fig. 3.13, shown at $t = 2\tau_{KH}$ (top) and $t = 4\tau_{KH}$ (bottom). We show on the left the simulation with the shear-based conduction speed, and again the compression-based speed on the right. No significant qualitative differences are seen between the two models.

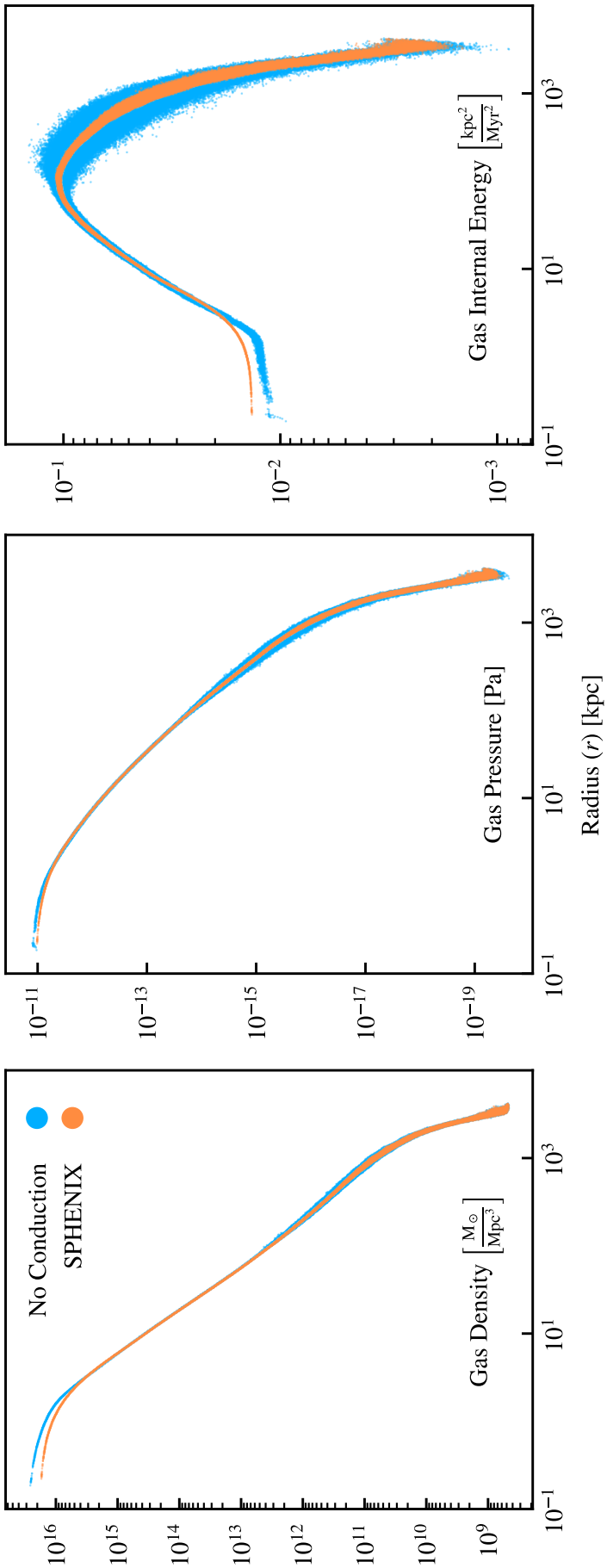


Figure A.5: Profiles of the idealised NFW halo (of mass $\approx 10^{13}$ $M_\odot$, at a gas particle mass resolution of 10^5 $M_\odot$) at $t = 5$ Gyr after the initial state. Blue points show every particle in the simulation without artificial conduction enabled, with orange showing the simulation with conduction enabled. Here the conduction can allow for a reduction in the scatter in internal energy without leading to significant conduction into the centre. The offset seen in the centre of about a factor of 1.5x originates from the smoothing of the kink around ≈ 0.7 kpc during the initial settling of the halo, and remained stable from that point at around $\approx t = 0.5$ Gyr until the end of the simulation.

halo has a gas mass of 1.7×10^{12} $M_\odot$, resolved by 1067689 particles with varying mass from 10^5 to 1.7×10^{12} $M_\odot$ with the highest resolution in the centre.

The gas in the halo is set up to be isothermal, following (Stern et al., 2019),

$$\frac{\mathrm{d} \ln P}{\mathrm{d} \ln r} = -\gamma \frac{v_c^2}{c_s^2} \tag{A.3.1}$$

where v_c is the circular velocity of the halo. The condition used to set the initial temperatures is $v_c = c_s$, and to get the correct normalisation for pressure and density the gas fraction at $R_{500,\mathrm{crit}}$ is used following Debackere et al. (2020).

Fig. A.5 shows that there is little difference between the result with conduction, and without. There is a small offset in the centre where the simulation with conduction has a slightly higher energy and slightly lower density, giving a very small overall offset in pressure. This figure is shown at $t = 5$ Gyr, much longer than any realistic cluster of a similar mass would go without accretion or some other external force perturbing the pressure profile anyway. Finally, the conduction allows the noisy internal energy distribution (and additionally density distribution) to be normalised over time thanks to the inclusion of the pressure differencing term.

A.4 Sedov Blast

In Fig. 3.6 we presented the convergence properties of the Sedov blast with the Sphenix scheme. The scheme only demonstrated convergence as $L_1(v) \propto h^{\sim 0.5}$, which is much slower than the expected convergence rate of $L_1 \propto h^1$ for shock fronts in SPH (that is demonstrated and exceeded in the Noh problem in Fig. 3.11). This is, however, simply an artefact of the way that the convergence is measured.

In Fig. A.6 we show the actual density profiles of the shock front, by resolution (increasing as the subfigures go to the right). Note here that the width of the shock front (from the particle distribution to the right of the vertical line to the vertical line in the analytical solution) does converge at the expected rate of $L_1 \propto 1/n^{1/3} \propto h$ with n the number of particles in the volume (in 3D).

The Sedov blast, unlike the Noh problem and Sod tubes, does not aim to reproduce a simple step function in density and velocity, but also a complex, expanding, post-shock region. The L_1 convergence is measured 'vertically' in this figure, but it is clear here that the vertical deviation from the analytical solution is not representative of the 'error' in the properties of a given particle, or in the width of the shock front. Small deviations in the position of the given particle could result in changes of orders of magnitude in the value of the L_1 norm measured for it.

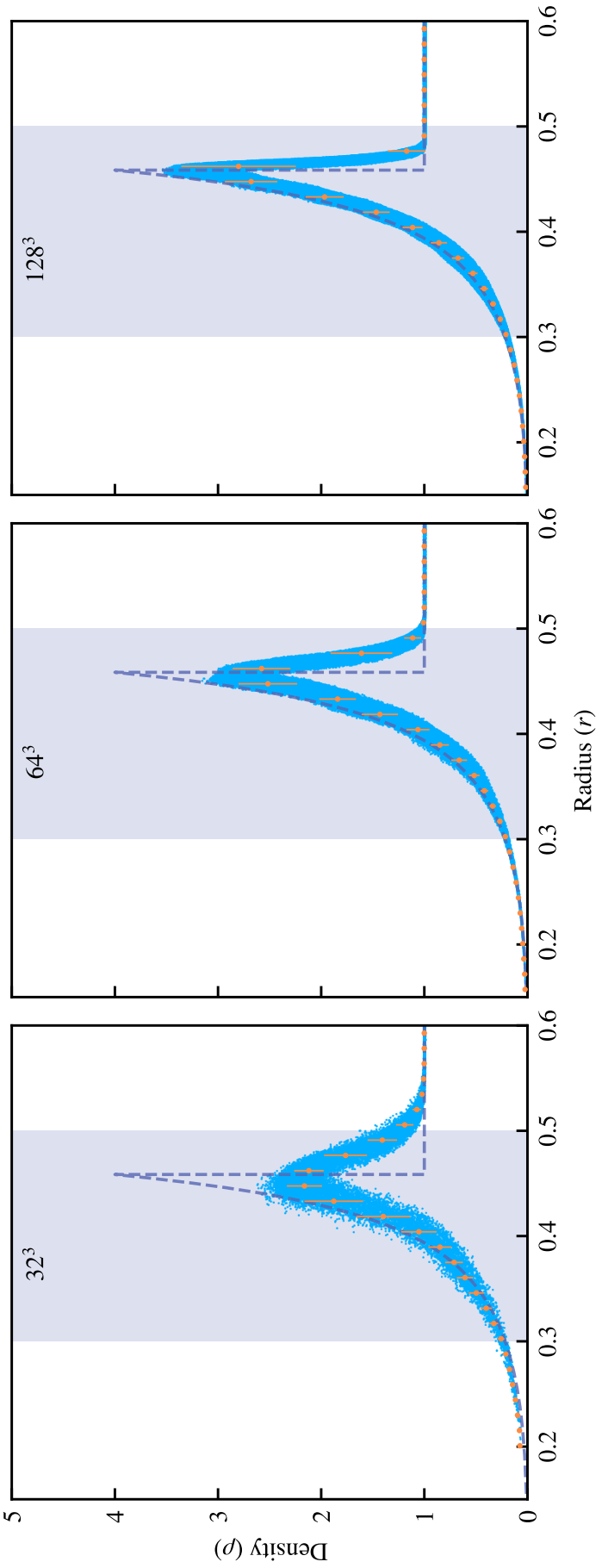


Figure A.6: The density profile of the Sedov blasts initially presented in Fig. 3.6. The blue points show the positions of every particle in the volume, the purple dashed line the analytical prediction, and the orange points binned means with error bars showing one standard deviation. The shaded band is the region over which the convergence properties were measured. The text at the top notes the total number of particles in each volume.

Because of this, and because we have demonstrated in other sections that Sphenix is able to converge on shock problems at faster than first order, we believe the slow convergence on the Sedov problem to be of little importance in practical applications of the scheme.

A.5 Conduction in the Noh Problem

In §3.5.4 we presented the Noh problem, and showed that the Sphenix scheme (like other SPH schemes in general) struggles to capture the high density in the central region due to so-called 'wall heating'.

The Sphenix scheme includes a switch to reduce artificial conduction in viscous flows. This is, as presented in §3.4, to allow for the capturing of energetic feedback events. It does, however, lead to a minor downside; the stabilising effect of the conduction in these shocks is almost completely removed. Usually, the artificial conduction lowers the dispersion in local internal energy values, and hence pressures, allowing for a more regular particle distribution.

In Fig. A.7 we show three re-simulations of the Noh problem (at 256^3 resolution) with three separate schemes. The first, the full Sphenix scheme, is simply a lower resolution version of Fig. 3.10. The second, 'No Conduction Limiter', is the Sphenix scheme, but with Equation 3.3.30 removed; i.e. the particle-carried artificial conduction coefficient depends solely on the local internal energy field (through $\nabla^2 u$ and Eqn. 3.3.24), instead of also being mediated by the velocity divergence field. The final case, 'Fixed $\alpha_D = 1.0$', shows the case where we remove all conduction switches and use a fixed value for the conduction α_D of 1.0. The former two look broadly similar, suggesting that the post-shock region is not significantly affected by the additional Sphenix conduction limiter.

The final panel, however, shows the benefits available to a hypothetical scheme that can remove the artificial conduction switch; the central region is able to hold a significantly higher density thanks to energy being conducted out of this region, allowing the pressure to regularise. In addition to the above, this case shows significantly weaker spurious density features (recall that the post-shock, high-density, region should have a uniform density) because these have been regularised by the conduction scheme.

We present this both to show the drawbacks of the Sphenix artificial conduction scheme, and to show the importance of demonstrating test problems with the same switches that would be used in a production simulation.

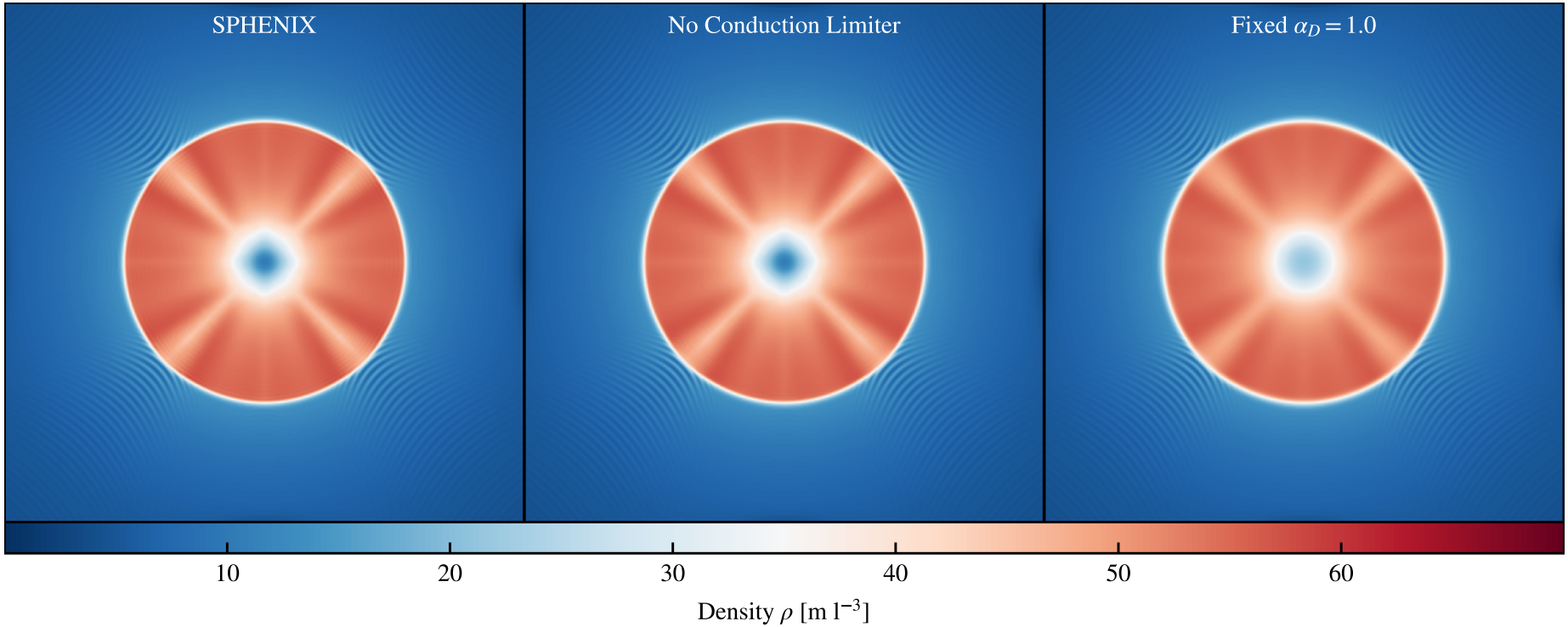


Figure A.7: Density slice through the centre of the Noh problem (analogue of Fig. 3.10) shown for three different artificial conduction schemes (see text). The colour bar is shared between all, and they all use the same, 256^3, initial condition, and are also all shown at $t = 0.6$. The case with the fixed, high, conduction coefficient (right) shows the smallest deviation in density in the centre, as the conduction can treat the wall heating present in this test.

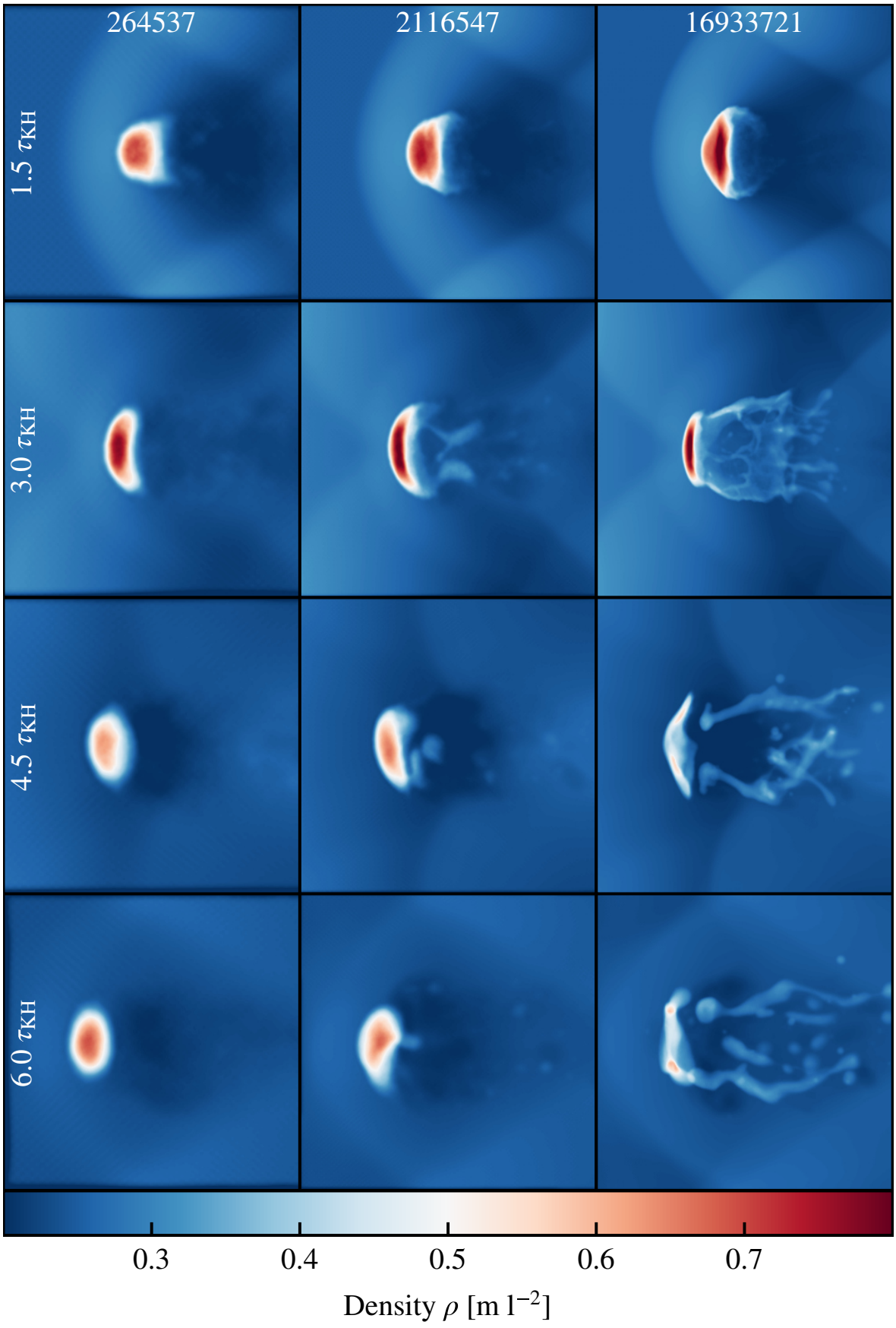


Figure A.8: A repeat of Fig. 3.15 but using a 'traditional' SPH scheme without diffusive switches.

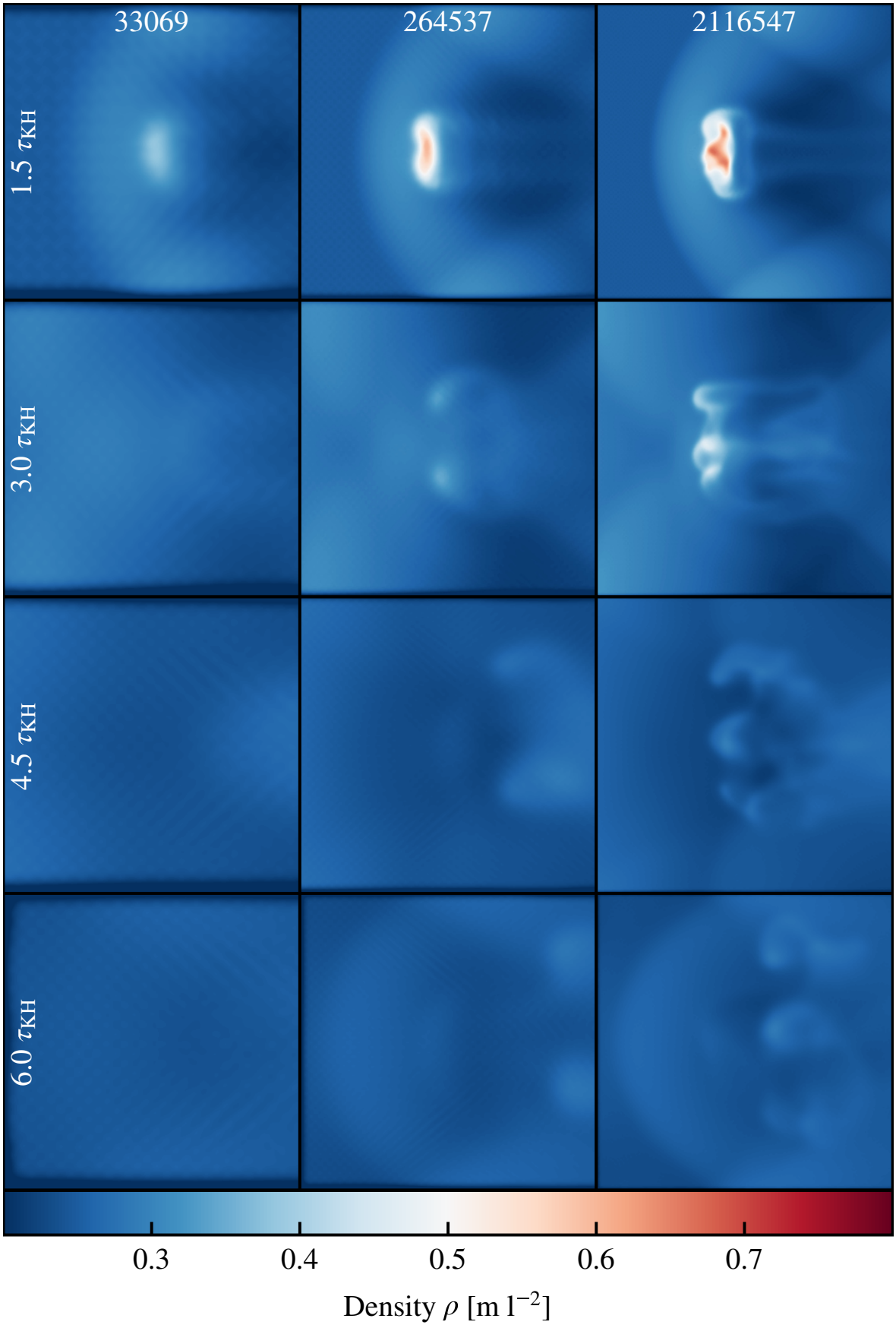


Figure A.9: A repeat of Fig. 3.15 but using an SPH-ALE scheme with a diffusive slope limiter. Note however that this is one step lower in resolution, due to the additional computational cost required to perform a simulation including a Riemann solver.

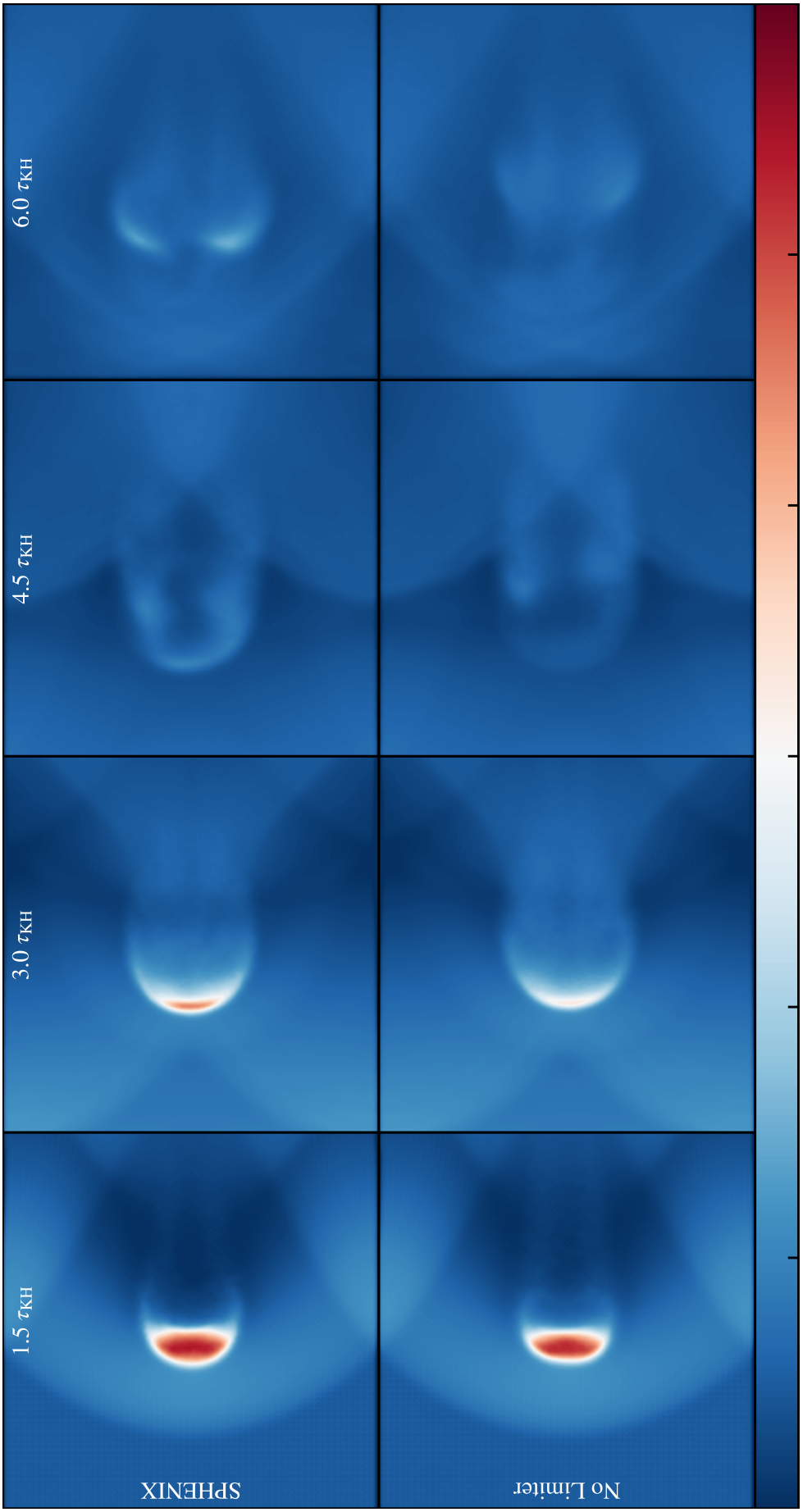


Figure A.10: The evolution of a single blob (using the medium resolution, 2116547 particle, initial conditions from Fig. 3.15), to illustrate the effect of turning off the conduction limiter (Eqn. 3.3.30; bottom row) in comparison to the full SPHENIX scheme (top row). The limiter suppresses some of the initial mixing during the cloud crushing, but does not cause significant qualitative changes in the mixing of the cloud.

A.6 Blob Test

In Fig. 3.15 we demonstrated the performance of the Sphenix scheme on an example 'blob' test. Here, we show how the same initial conditions are evolved using two schemes: a 'traditional SPH' scheme with fixed artificial viscosity ($\alpha_V = 0.8$) and no artificial conduction (e.g. Monaghan, 1992)[1], and a SPH-ALE (Vila, 1999) scheme similar to GIZMO-MFM[2] (Hopkins, 2015) with a diffusive slope limiter. This is in an effort to demonstrate how the initial conditions are evolved with a minimally viable non-diffusive scheme, through to what could be considered the most diffusive viable scheme.

Fig. A.8 shows the result of the blob test with the traditional SPH scheme. Here, as expected, there is a severe lack of mixing, with the artificial surface tension holding the blob together even at the highest resolutions. The lack of phase mixing also contributes to a lack of overall mixing, with the stripped trails (shown most clearly at $t = 3\tau_{\rm KH}$) adiabatically expanding but crucially remaining distinct from the hot background medium.

Fig. A.9 shows the result of the blob test with the SPH-ALE (GIZMO-MFM) scheme. This scheme is known to be highly diffusive (due to the less conservative slope limiter employed in the Swift implementation). This follows closely the results seen in e.g. Agertz et al. (2007) for diffusive grid-based codes. Here, the blob is rapidly shattered, and then dissolves quickly into the surrounding media, especially at the lowest resolutions.

The Sphenix results in Fig. 3.15 showed that the blob mixed with the surrounding media, but at a less rapid rate than in the SPH-ALE case. This is somewhat expected, given the trade-off required in the artificial conduction switches (Eqn. 3.3.30). We do note, however, that no analytical solution exists for the blob test, and as such all of these comparisons may only be made qualitatively.

In Fig. A.10 we examine the effect of removing the conduction limiter from the Sphenix implementation (i.e. Eqn. 3.3.30 is removed, allowing α_D to vary irrespective of the values of α_V). We see that the inclusion of the limiter does slightly reduce the rate of initial mixing within the blob, but that the effect of the limiter is not particularly strong within this case.

[1] The `minimal` scheme in Swift.
[2] The `gizmo-mfm` scheme in Swift with a HLLC Riemann solver.

Bibliography

Adhikari, S., Dalal, N., Chamberlain, R.T. *Splashback in Accreting Dark Matter Halos*. Journal of Cosmology and Astroparticle Physics, **2014**(11) (2014), 019. ISSN 1475-7516.

Agertz, O., Moore, B., Stadel, J., et al. *Fundamental Differences between SPH and Grid Methods: Simulating Fluids Using SPH and Grid Techniques*. Monthly Notices of the Royal Astronomical Society, **380**(3) (2007), 963. ISSN 00358711.

Aguirre, A., Hernquist, L., Schaye, J., et al. *Metal Enrichment of the Intergalactic Medium in Cosmological Simulations*. **561** (2001), 521. astro-ph/0105065.

Anglés-Alcázar, D., Davé, R., Faucher- Giguère, C.A., et al. *Gravitational Torque-Driven Black Hole Growth and Feedback in Cosmological Simulations*. **464** (2017a), 2840. 1603.08007.

Anglés-Alcázar, D., Davé, R., Özel, F., et al. *COSMOLOGICAL ZOOM SIMULATIONS OF* z = *2 GALAXIES: THE IMPACT OF GALACTIC OUTFLOWS*. The Astrophysical Journal, **782**(2) (2014), 84. ISSN 0004-637X, 1538-4357.

Anglés-Alcázar, D., Faucher-Giguère, C.A., Kereš, D., et al. *The Cosmic Baryon Cycle and Galaxy Mass Assembly in the FIRE Simulations*. Monthly Notices of the Royal Astronomical Society, **470**(4) (2017b), 4698. ISSN 0035-8711, 1365-2966.

Balogh, M.L., Pearce, F.R., Bower, R.G., et al. *Revisiting the Cosmic Cooling Crisis*. **326** (2001), 1228. astro-ph/0104041.

Balsara, D.S. Ph.D. thesis (1989).

Balsara, D.S., Kim, J., Mac Low, M.M., et al. *Amplification of Interstellar Magnetic Fields by Supernova-driven Turbulence*. The Astrophysical Journal, **617**(1) (2004), 339. ISSN 0004-637X, 1538-4357.

Barnes, J., Hut, P. *A Hierarchical O(N Log N) Force-Calculation Algorithm*. Nature, **324**(6096) (1986), 446. ISSN 0028-0836, 1476-4687.

Bauer, A., Springel, V. *Subsonic Turbulence in Smoothed Particle Hydrodynamics and Moving-Mesh Simulations:* Turbulence in SPH and on a Moving-Mesh. Monthly Notices of the Royal Astronomical Society, **423**(3) (2012), 2558. ISSN 00358711.

Behroozi, P., Wechsler, R.H., Hearin, A.P., et al. *UNIVERSEMACHINE: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from z = 0-10*. **488**(3) (2019), 3143. 1806.07893.

Benitez-Llambay, A. *Py-Sphviewer: Py-SPHViewer v1.0.0* (2015).

Benítez-Llambay, A., Navarro, J.F., Frenk, C.S., et al. *The Vertical Structure of Gaseous Galaxy Discs in Cold Dark Matter Haloes*. **473** (2018), 1019. 1707.08046.

Benz, W. *Applications of Smooth Particle Hydrodynamics (SPH) to Astrophysical Problems.* Computer Physics Communications, **48** (1988), 97. ISSN 0010-4655.

Bondi, H. *On Spherically Symmetrical Accretion*. **112** (1952), 195.

Booth, C.M., Schaye, J. *Cosmological Simulations of the Growth of Supermassive Black Holes and Feedback from Active Galactic Nuclei: Method and Tests*. Monthly Notices of the Royal Astronomical Society, **398**(1) (2009), 53. ISSN 00358711, 13652966.

Borrow, J., Borrisov, A. *Swiftsimio: A Python Library for Reading SWIFT Data*. Journal of Open Source Software, **5**(52) (2020), 2430. ISSN 2475-9066.

Borrow, J., Bower, R.G., Draper, P.W., et al. *SWIFT: Maintaining Weak-Scalability with a Dynamic Range of $ 10â$ in Time-Step Size to Harness Extreme Adaptivity*. Proceedings of the 13th SPHERIC International Workshop, Galway, Ireland, June 26-28 2018 (2018), 44.

Borrow, J., Kelly, A.J. *Projecting SPH Particles in Adaptive Environments*. arXiv e-prints (2021), arXiv:2106.05281. 2106.05281.

Borrow, J., Schaller, M., Bower, R.G. *Inconsistencies Arising from the Coupling of Galaxy Formation Sub-Grid Models to Pressure-Smoothed Particle Hydrodynamics*. arXiv e-prints (2020), arXiv:2011.11641.

Borrow, J., Vandenbroucke, B., Schaller, M. *Choosing the Best Flavour of SPH for Astrophysics Problems: More Complex May Not Be Better*. In *Proceedings of the 14th SPHERIC International Workshop, Exeter, United Kingdom, June 25-27 2019*. Exeter, United Kingdom (2019).

Bower, R.G., Schaye, J., Frenk, C.S., et al. *The Dark Nemesis of Galaxy Formation: Why Hot Haloes Trigger Black Hole Growth and Bring Star Formation to an End.* Monthly Notices of the Royal Astronomical Society, **465**(1) (2017), 32. ISSN 0035-8711, 1365-2966.

Bower, R.G., Vernon, I., Goldstein, M., et al. *The Parameter Space of Galaxy Formation*. **407**(4) (2010), 2017. 1004.0711.

Bryan, G.L., Norman, M.L. *Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons*. The Astrophysical Journal, **495**(1) (1998), 80. ISSN 0004-637X, 1538-4357.

Caswell, J., Wallis, J. *A Treatise of Algebra, Both Historical and Practical ...: With Some Additional Treatises I. of the Cono-Cuneus; Being a Body Representing in Part a Conus, an Part a Cuneus ; II. of Angular Sections; and Other Things Relating There unto, and to Trigonometry ; III. of the Angle of Contact; with Other Things Appertaining to the Composition of Magnitudes, the Inceptive of Magnitudes, and the Composition of Motions, with the Results Thereof ; IV. of Combination, Alternations, and Aliquot Parts*. printed by John Playford, London (1685).

Cen, R. *A Hydrodynamic Approach to Cosmology: Methodology*. The Astrophysical Journal Supplement Series, **78** (1992), 341. ISSN 0067-0049.

Cen, R., Ostriker, J.P. *Galaxy Formation and Physical Bias*. The Astrophysical Journal Letters, **399** (1992), L113.

Chisari, N.E., Richardson, M.L.A., Devriendt, J., et al. *The Impact of Baryons on the Matter Power Spectrum from the Horizon-AGN Cosmological Hydrodynamical Simulation*. Monthly Notices of the Royal Astronomical Society, **480**(3) (2018), 3962. ISSN 0035-8711, 1365-2966.

Christensen, C.R., Davé, R., Brooks, A., et al. *Tracing Outflowing Metals in Simulations of Dwarf and Spiral Galaxies*. The Astrophysical Journal, **867**(2) (2018), 142. ISSN 1538-4357.

Christensen, C.R., Davé, R., Governato, F., et al. *IN-N-OUT: THE GAS CYCLE FROM DWARFS TO SPIRAL GALAXIES*. The Astrophysical Journal, **824**(1) (2016), 57. ISSN 1538-4357.

Clagett, M. *Ancient Egyptian Science : A Source Book*. American Philosophical Society, Philadelphia (1989). ISBN 978-0-87169-232-0.

Courant, R., Friedrichs, K., Lewy, H. *{\"U}ber Die Partiellen Differenzengleichungen Der Mathematischen Physik*. Mathematische Annalen, **100** (1928), 32.

Crain, R.A., Eke, V.R., Frenk, C.S., et al. *The Baryon Fraction of CDM Haloes*. Monthly Notices of the Royal Astronomical Society, **377**(1) (2007), 41. ISSN 0035-8711, 1365-2966.

Crain, R.A., Schaye, J., Bower, R.G., et al. *The EAGLE Simulations of Galaxy Formation: Calibration of Subgrid Physics and Model Variations*. Monthly Notices of the Royal Astronomical Society, **450**(2) (2015), 1937. ISSN 1365-2966, 0035-8711.

Croton, D.J., Stevens, A.R.H., Tonini, C., et al. *Semi-Analytic Galaxy Evolution (SAGE): Model Calibration and Basic Results*. **222**(2) (2016), 22. 1601.04709.

Cui, W., Knebe, A., Libeskind, N.I., et al. *The Large-Scale Environment from Cosmological Simulations II: The Redshift Evolution and Distributions of Baryons*. Monthly Notices of the Royal Astronomical Society, **485** (2019), 2367. ISSN 0035-8711.

Cullen, L., Dehnen, W. *Inviscid SPH*. Monthly Notices of the Royal Astronomical Society, **408**(2) (2010), 669. ISSN 00358711. 1006.1524.

Dalla Vecchia, C., Schaye, J. *Simulating Galactic Outflows with Kinetic Supernova Feedback*. Monthly Notices of the Royal Astronomical Society, **387**(4) (2008), 1431. ISSN 00358711, 13652966.

Dalla Vecchia, C., Schaye, J. *Simulating Galactic Outflows with Thermal Supernova Feedback: Galactic Outflows with Thermal SN Feedback*. Monthly Notices of the Royal Astronomical Society, **426**(1) (2012), 140. ISSN 00358711.

Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. *Simba: Cosmological Simulations with Black Hole Growth and Feedback*. Monthly Notices of the Royal Astronomical Society, **486**(2) (2019), 2827. ISSN 0035-8711, 1365-2966.

Davé, R., Cen, R., Ostriker, J.P., et al. *Baryons in the Warm-Hot Intergalactic Medium*. **552** (2001), 473. astro-ph/0007217.

Davé, R., Thompson, R., Hopkins, P.F. *Mufasa: Galaxy Formation Simulations with Meshless Hydrodynamics*. Monthly Notices of the Royal Astronomical Society, **462**(3) (2016), 3265. ISSN 0035-8711, 1365-2966.

Davies, J.J., Crain, R.A., Pontzen, A. *Quenching and Morphological Evolution Due to Circumgalactic Gas Expulsion in a Simulated Galaxy with a Controlled Assembly History*. arXiv:2006.13221 [astro-ph] (2020). 2006.13221.

Debackere, S.N.B., Schaye, J., Hoekstra, H. *The Impact of the Observed Baryon Distribution in Haloes on the Total Matter Power Spectrum*. Monthly Notices of the Royal Astronomical Society, **492**(2) (2020), 2285. ISSN 0035-8711, 1365-2966.

Dehnen, W. *A Fast Multipole Method for Stellar Dynamics*. Computational Astrophysics and Cosmology, **1**(1) (2014), 1. ISSN 2197-7909.

Dehnen, W., Aly, H. *Improving Convergence in Smoothed Particle Hydrodynamics Simulations without Pairing Instability*. Monthly Notices of the Royal Astronomical Society, **425**(2) (2012), 1068. ISSN 00358711. 1204.2471.

Diemer, B., Kravtsov, A.V. *DEPENDENCE OF THE OUTER DENSITY PROFILES OF HALOS ON THEIR MASS ACCRETION RATE*. The Astrophysical Journal, **789**(1) (2014), 1. ISSN 0004-637X, 1538-4357.

Diemer, B., Mansfield, P., Kravtsov, A.V., et al. *The Splashback Radius of Halos from Particle Dynamics. II. Dependence on Mass, Accretion Rate, Redshift, and Cosmology*. The Astrophysical Journal, **843**(2) (2017), 140. ISSN 1538-4357.

Dolag, K., Borgani, S., Murante, G., et al. *Substructures in Hydrodynamical Cluster Simulations*. Monthly Notices of the Royal Astronomical Society, **399**(2) (2009), 497. ISSN 00358711, 13652966.

Dolag, K., Jubelgas, M., Springel, V., et al. *Thermal Conduction in Simulated Galaxy Clusters*. The Astrophysical Journal, **606**(2) (2004), L97. ISSN 0004-637X, 1538-4357.

Dubois, Y., Pichon, C., Welker, C., et al. *Dancing in the Dark: Galactic Properties Trace Spin Swings along the Cosmic Web*. **444**(2) (2014), 1453. 1402.1165.

Durier, F., Dalla Vecchia, C. *Implementation of Feedback in Smoothed Particle Hydrodynamics: Towards Concordance of Methods: Towards Concordance of Feedback Methods in SPH*. Monthly Notices of the Royal Astronomical Society, **419**(1) (2012), 465. ISSN 00358711.

Elahi, P.J., Cañas, R., Poulton, R.J.J., et al. *Hunting for Galaxies and Halos in Simulations with VELOCIraptor*. Publications of the Astronomical Society of Australia, **36** (2019), e021. ISSN 1323-3580, 1448-6083. 1902.01010.

Ettori, S., Dolag, K., Borgani, S., et al. *The Baryon Fraction in Hydrodynamical Simulations of Galaxy Clusters*. Monthly Notices of the Royal Astronomical Society, **365**(3) (2006), 1021. ISSN 0035-8711, 1365-2966.

Evans, T.A., Fattahi, A., Deason, A.J., et al. *How Unusual Is the Milky Way's Assembly History?* Monthly Notices of the Royal Astronomical Society, **497**(4) (2020), 4311. ISSN 0035-8711, 1365-2966.

Evrard, A.E. *Beyond N-Body: 3D Cosmological Gas Dynamics.* Monthly Notices of the Royal Astronomical Society, **235** (1988), 911. ISSN 0035-8711.

Evrard, A.E., Summers, F.J., Davis, M. *Two-Fluid Simulations of Galaxy Formation*. The Astrophysical Journal, **422** (1994), 11. ISSN 0004-637X, 1538-4357.

Fabian, A. *Observational Evidence of Active Galactic Nuclei Feedback*. Annual Review of Astronomy and Astrophysics, **50**(1) (2012), 455. ISSN 0066-4146, 1545-4282.

Fowler, D., Robson, E. *Square Root Approximations in Old Babylonian Mathematics: YBC 7289 in Context*. Historia Mathematica, **25**(4) (1998), 366. ISSN 03150860.

Frenk, C.S., White, S.D.M., Davis, M., et al. *The Formation of Dark Halos in a Universe Dominated by Cold Dark Matter*. The Astrophysical Journal, **327** (1988), 507. ISSN 0004-637X, 1538-4357.

Frenk, C.S., White, S.D.M., Efstathiou, G., et al. *Galaxy Clusters and the Amplitude of Primordial Fluctuations*. **351** (1990), 10.

Genel, S., Bryan, G.L., Springel, V., et al. *A Quantification of the Butterfly Effect in Cosmological Simulations and Implications for Galaxy Scaling Relations*. The Astrophysical Journal, **871**(1) (2019), 21. ISSN 1538-4357.

Gill, S.P.D., Knebe, A., Gibson, B.K. *The Evolution of Substructure — I. A New Identification Method*. Monthly Notices of the Royal Astronomical Society, **351**(2) (2004), 399. ISSN 0035-8711, 1365-2966.

Gingold, R.A., Monaghan, J.J. *Smoothed Particle Hydrodynamics: Theory and Application to Non-Spherical Stars.* Monthly Notices of the Royal Astronomical Society, **181** (1977), 375. ISSN 0035-8711.

Greene, J.E., Zakamska, N.L., Smith, P.S. *A SPECTACULAR OUTFLOW IN AN OBSCURED QUASAR*. The Astrophysical Journal, **746**(1) (2012), 86. ISSN 0004-637X, 1538-4357.

Greengard, L., Rokhlin, V. *A Fast Algorithm for Particle Simulations*. Journal of Computational Physics, **73** (1987), 325. ISSN 0021-9991.

Gresho, P.M., Chan, S.T. *On the Theory of Semi-Implicit Projection Methods for Viscous Incompressible Flow and Its Implementation via a Finite Element Method That Also Introduces a Nearly Consistent Mass Matrix. II - Implementation*. International Journal for Numerical Methods in Fluids, **11** (1990), 621. ISSN 0271-2091.

Gresho, P.M., Sani, R.L. *Introducing Four Benchmark Solutions*. International Journal for Numerical Methods in Fluids, **11**(7) (1990), 951. ISSN 0271-2091, 1097-0363.

Guillet, T., Pakmor, R., Springel, V., et al. *High-Order Magnetohydrodynamics for Astrophysics with an Adaptive Mesh Refinement Discontinuous Galerkin Scheme*. Monthly Notices of the Royal Astronomical Society, **485**(3) (2019), 4209. ISSN 0035-8711, 1365-2966.

Hafen, Z., Faucher-Giguere, C.A., Angles-Alcazar, D., et al. *The Fates of the Circumgalactic Medium in the FIRE Simulations*. arXiv e-prints (2019a), arXiv:1910.01123. 1910.01123.

Hafen, Z., Faucher-Giguère, C.A., Anglés-Alcázar, D., et al. *The Origins of the Circumgalactic Medium in the FIRE Simulations*. Monthly Notices of the Royal Astronomical Society, **488**(1) (2019b), 1248. ISSN 0035-8711, 1365-2966.

Harris, C.R., Jarrod Millman, K., van der Walt, S.J., et al. *Array Programming with NumPy*. arXiv e-prints (2020), arXiv:2006.10256.

Hellwing, W.A., Schaller, M., Frenk, C.S., et al. *The Effect of Baryons on Redshift Space Distortions and Cosmic Density and Velocity Fields in the EAGLE Simulation*. Monthly Notices of the Royal Astronomical Society: Letters, **461**(1) (2016), L11. ISSN 1745-3925, 1745-3933.

Henriques, B.M.B., White, S.D.M., Thomas, P.A., et al. *Galaxy Formation in the Planck Cosmology - I. Matching the Observed Evolution of Star Formation Rates, Colours and Stellar Masses*. **451**(3) (2015), 2663. 1410.0365.

Hernquist, L., Katz, N. *TREESPH - A Unification of SPH with the Hierarchical Tree Method*. The Astrophysical Journal Supplement Series, **70** (1989), 419. ISSN 0067-0049, 1538-4365.

Heß, S., Springel, V. *Particle Hydrodynamics with Tessellation Techniques: Tessellation Based Particle Hydrodynamics*. Monthly Notices of the Royal Astronomical Society, **406**(4) (2010), 2289. ISSN 00358711.

Hobson, E.W. *John Napier and the Invention of Logarithms, 1614; a Lecture*. Cambridge University Press (1914).

Holmberg, E. *On the Clustering Tendencies among the Nebulae. II. a Study of Encounters between Laboratory Models of Stellar Systems by a New Integration Procedure.* **94** (1941), 385.

Hopkins, P.F. *A General Class of Lagrangian Smoothed Particle Hydrodynamics Methods and Implications for Fluid Mixing Problems*. Monthly Notices of the Royal Astronomical Society, **428**(4) (2013), 2840. ISSN 0035-8711, 1365-2966.

Hopkins, P.F. *GIZMO: A New Class of Accurate, Mesh-Free Hydrodynamic Simulation Methods*. Monthly Notices of the Royal Astronomical Society, **450**(1) (2015), 53. ISSN 0035-8711, 1365-2966. 1409.7395.

Hopkins, P.F. *A New Public Release of the GIZMO Code*. arXiv e-prints (2017), arXiv:1712.01294. 1712.01294.

Hopkins, P.F., Kereš, D., Oñorbe, J., et al. *Galaxies on FIRE (Feedback In Realistic Environments): Stellar Feedback Explains Cosmologically Inefficient Star Formation*. Monthly Notices of the Royal Astronomical Society, **445**(1) (2014), 581. ISSN 0035-8711, 1365-2966.

Hopkins, P.F., Wetzel, A., Keres, D., et al. *FIRE-2 Simulations: Physics versus Numerics in Galaxy Formation*. Monthly Notices of the Royal Astronomical Society, **480**(1) (2018), 800. ISSN 0035-8711, 1365-2966. 1702.06148.

Hu, C.Y., Naab, T., Walch, S., et al. *SPHGal: Smoothed Particle Hydrodynamics with Improved Accuracy for Galaxy Simulations*. Monthly Notices of the Royal Astronomical Society, **443**(2) (2014), 1173. ISSN 1365-2966, 0035-8711.

Hubble, E. *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*. Proceedings of the National Academy of Science, **15**(3) (1929), 168.

Hunter, J.D. *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, **9**(3) (2007), 90. ISSN 1521-9615.

Kauffmann, G. *The Age of Elliptical Galaxies and Bulges in a Merger Model*. **281**(2) (1996), 487. astro-ph/9502096.

Kauffmann, G., Heckman, T.M., White, S.D.M., et al. *The Dependence of Star Formation History and Internal Structure on Stellar Mass for 10^5 Low-Redshift Galaxies*. **341**(1) (2003), 54. astro-ph/0205070.

Keller, B.W., Wadsley, J.W., Wang, L., et al. *Chaos and Variance in Galaxy Formation*. Monthly Notices of the Royal Astronomical Society, **482**(2) (2019), 2244. ISSN 0035-8711, 1365-2966.

Klein, R.I., McKee, C.F., Colella, P. *On the Hydrodynamic Interaction of Shock Waves with Interstellar Clouds. I. Nonradiative Shocks in Small Clouds*. The Astrophysical Journal, **420** (1994), 213. ISSN 0004-637X.

Knabenhans, M., Stadel, J., Potter, D., et al. *Euclid Preparation: IX. EuclidEmulator2 - Power Spectrum Emulation with Massive Neutrinos and Self-Consistent Dark Energy Perturbations* (2021).

Knollmann, S.R., Knebe, A. *AHF: AMIGA'S HALO FINDER*. The Astrophysical Journal Supplement Series, **182**(2) (2009), 608. ISSN 0067-0049, 1538-4365.

Lacey, C.G., Baugh, C.M., Frenk, C.S., et al. *A Unified Multiwavelength Model of Galaxy Formation*. **462**(4) (2016), 3854. 1509.08473.

Lam, S.K., Pitrou, A., Seibert, S. *Numba: A LLVM-Based Python JIT Compiler*. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15. Association for Computing Machinery, New York, NY, USA (2015). ISBN 978-1-4503-4005-2.

Lattanzio, J.C., Monaghan, J.J., Pongracic, H., et al. *Controlling Penetration*. SIAM Journal on Scientific and Statistical Computing, **7**(2) (1986), 591.

Liao, S., Gao, L., Frenk, C.S., et al. *The Segregation of Baryons and Dark Matter during Halo Assembly*. Monthly Notices of the Royal Astronomical Society, **470**(2) (2017), 2262. ISSN 0035-8711, 1365-2966.

Lucy, L.B. *A Numerical Approach to the Testing of the Fission Hypothesis*. The Astronomical Journal, **82** (1977), 1013. ISSN 00046256.

Ludlow, A.D., Navarro, J.F., Springel, V., et al. *THE UNORTHODOX ORBITS OF SUBSTRUCTURE HALOS*. The Astrophysical Journal, **692**(1) (2009), 931. ISSN 0004-637X, 1538-4357.

Ludlow, A.D., Schaye, J., Schaller, M., et al. *Numerical Convergence of Hydrodynamical Simulations of Galaxy Formation: The Abundance and Internal Structure of Galaxies and Their Cold Dark Matter Haloes*. **493**(2) (2020), 2926. 1908.05019.

Maiolino, R., Gallerani, S., Neri, R., et al. *Evidence of Strong Quasar Feedback in the Early Universe*. **425** (2012), L66. 1204.2904.

Mansfield, P., Kravtsov, A.V., Diemer, B. *Splashback Shells of Cold Dark Matter Halos*. The Astrophysical Journal, **841**(1) (2017), 34. ISSN 1538-4357.

Marinacci, F., Sales, L.V., Vogelsberger, M., et al. *Simulating the Interstellar Medium and Stellar Feedback on a Moving Mesh: Implementation and Isolated Galaxies*. arXiv:1905.08806 [astro-ph] (2019). 1905.08806.

McCarthy, I.G., Schaye, J., Bird, S., et al. *The Bahamas Project: Calibrated Hydrodynamical Simulations for Large-Scale Structure Cosmology*. Monthly Notices of the Royal Astronomical Society, **465**(3) (2017), 2936. ISSN 0035-8711, 1365-2966.

Metropolis, N. *The Beginning of the Monte Carlo Method*. Los Alamos Science, (Special Issue dedicated to Stanislaw Ulam) (1987), 125.

Mitchell, P., Blaizot, J., Cadiou, C., et al. *Tracing the Simulated High-Redshift Circum-Galactic Medium with Lyman Alpha Emission*. arXiv:2008.12790 [astro-ph] (2020a). 2008.12790.

Mitchell, P.D., Schaye, J., Bower, R.G. *Galactic Inflow and Wind Recycling Rates in the Eagle Simulations*. Monthly Notices of the Royal Astronomical Society, **497**(4) (2020b), 4495. ISSN 0035-8711, 1365-2966.

Monaghan, J. *SPH and Riemann Solvers*. Journal of Computational Physics, **136**(2) (1997), 298. ISSN 00219991.

Monaghan, J., Gingold, R. *Shock Simulation by the Particle Method SPH*. Journal of Computational Physics, **52**(2) (1983), 374. ISSN 00219991.

Monaghan, J.J. *Smoothed Particle Hydrodynamics.* Annual Review of Astronomy and Astrophysics, **30** (1992), 543.

Monaghan, J.J. *SPH Compressible Turbulence*. Monthly Notices of the Royal Astronomical Society, **335**(3) (2002), 843. ISSN 00358711, 13652966.

More, S., Diemer, B., Kravtsov, A.V. *The Splashback Radius as a Physical Halo Boundary and the Growth of Halo Mass*. **810**(1) (2015), 36. 1504.05591.

Morris, J., Monaghan, J. *A Switch to Reduce SPH Viscosity*. Journal of Computational Physics, **136**(1) (1997), 41. ISSN 00219991.

Moster, B.P., Naab, T., White, S.D.M. *Galactic Star Formation and Accretion Histories from Matching Galaxies to Dark Matter Haloes*. **428**(4) (2013), 3121. 1205.5807.

Muratov, A.L., Kereš, D., Faucher-Giguère, C.A., et al. *Gusty, Gaseous Flows of FIRE: Galactic Winds in Cosmological Simulations with Explicit Stellar Feedback*. Monthly Notices of the Royal Astronomical Society, **454**(3) (2015), 2691. ISSN 0035-8711, 1365-2966.

Naab, T., Ostriker, J.P. *Theoretical Challenges in Galaxy Formation*. Annual Review of Astronomy and Astrophysics, **55** (2017), 59. 1612.06891.

Navarro, J.F., Frenk, C.S., White, S.D.M. *The Structure of Cold Dark Matter Halos*. The Astrophysical Journal, **462** (1996), 563. astro-ph/9508025.

Navarro, J.F., White, S.D.M. *Simulations of Dissipative Galaxy Formation in Hierarchically Clustering Universes - Part One - Tests of the Code*. Monthly Notices of the Royal Astronomical Society, **265** (1993), 271.

Nelson, D., Genel, S., Vogelsberger, M., et al. *The Impact of Feedback on Cosmological Gas Accretion*. Monthly Notices of the Royal Astronomical Society, **448**(1) (2015), 59. ISSN 1365-2966, 0035-8711.

Noh, W.F. *Errors for Calculations of Strong Shocks Using an Artificial Viscosity and an Artificial Heat Flux*. Journal of Computational Physics, **72** (1987), 78. ISSN 0021-9991.

Oleskiewicz, P., Baugh, C.M. *Sensitivity Analysis of a Galaxy Formation Model* (2020), 15.

Onorbe, J., Garrison-Kimmel, S., Maller, A.H., et al. *How to Zoom: Bias, Contamination and Lagrange Volumes in Multimass Cosmological Simulations*. Monthly Notices of the Royal Astronomical Society, **437**(2) (2014), 1894. ISSN 0035-8711, 1365-2966.

Oppenheimer, B.D., Davé, R. *Cosmological Simulations of Intergalactic Medium Enrichment from Galactic Outflows*. **373** (2006), 1265. astro-ph/0605651.

Oppenheimer, B.D., Davé, R., Kereš, D., et al. *Feedback and Recycled Wind Accretion: Assembling the z = 0 Galaxy Mass Function*. **406** (2010), 2325. 0912.0519.

Oppenheimer, B.D., Schaye, J., Crain, R.A., et al. *The Multiphase Circumgalactic Medium Traced by Low Metal Ions in EAGLE Zoom Simulations*. Monthly Notices of the Royal Astronomical Society, **481**(1) (2018), 835. ISSN 0035-8711, 1365-2966.

Ossendrijver, M. *Ancient Babylonian Astronomers Calculated Jupiter's Position from the Area under a Time-Velocity Graph*. Science, **351**(6272) (2016), 482.

Pillepich, A., Springel, V., Nelson, D., et al. *Simulating Galaxy Formation with the IllustrisTNG Model*. Monthly Notices of the Royal Astronomical Society, **473**(3) (2018), 4077. ISSN 0035-8711, 1365-2966.

Planck Collaboration, Adam, R., Aghanim, N., et al. Planck *Intermediate Results: XLVII. Planck Constraints on Reionization History*. Astronomy & Astrophysics, **596** (2016), A108. ISSN 0004-6361, 1432-0746.

Ploeckinger, S., Schaye, J. *Radiative Cooling Rates, Ion Fractions, Molecule Abundances, and Line Emissivities Including Self-Shielding and Both Local and Metagalactic Radiation Fields*. Monthly Notices of the Royal Astronomical Society, **497** (2020), 4857. ISSN 0035-8711.

Porter, L.A., Somerville, R.S., Primack, J.R., et al. *Understanding the Structural Scaling Relations of Early-Type Galaxies*. **444**(1) (2014), 942. 1407.0594.

Price, D.J. SPLASH *: An Interactive Visualisation Tool for Smoothed Particle Hydrodynamics Simulations*. Publications of the Astronomical Society of Australia, **24**(3) (2007), 159. ISSN 1323-3580, 1448-6083.

Price, D.J. *Modelling Discontinuities and Kelvin-Helmholtz Instabilities in SPH*. Journal of Computational Physics, **227**(24) (2008), 10040. ISSN 00219991. 0709.2772.

Price, D.J. *Smoothed Particle Hydrodynamics and Magnetohydrodynamics*. Journal of Computational Physics, **231**(3) (2012), 759. ISSN 00219991. 1012.1885.

Price, D.J., Wurster, J., Tricco, T.S., et al. *Phantom: A Smoothed Particle Hydrodynamics and Magnetohydrodynamics Code for Astrophysics*. Publications of the Astronomical Society of Australia, **35** (2018), e031. ISSN 1323-3580, 1448-6083. 1702.03930.

Read, J.I., Hayfield, T. *SPHS: Smoothed Particle Hydrodynamics with a Higher Order Dissipation Switch: SPH with a Higher Order Dissipation Switch*. Monthly Notices of the Royal Astronomical Society, **422**(4) (2012), 3037. ISSN 00358711.

Read, J.I., Hayfield, T., Agertz, O. *Resolving Mixing in Smoothed Particle Hydrodynamics: Resolving Mixing in SPH*. Monthly Notices of the Royal Astronomical Society (2010), no. ISSN 00358711, 13652966.

Revaz, Y., Jablonka, P. *The Dynamical and Chemical Evolution of Dwarf Spheroidal Galaxies with GEAR*. Astronomy & Astrophysics, **538** (2012), A82. ISSN 0004-6361, 1432-0746.

Richardson, L.F. *Weather Prediction by Numerical Process*. Cambridge university press (2007).

Ritchie, B.W., Thomas, P.A. *Multiphase Smoothed-Particle Hydrodynamics*. Monthly Notices of the Royal Astronomical Society, **323**(3) (2001), 743. ISSN 0035-8711, 1365-2966.

Rosswog, S. *The Lagrangian Hydrodynamics Code MAGMA2*. Monthly Notices of the Royal Astronomical Society, **498** (2020a), 4230. ISSN 0035-8711. 1911.13093.

Rosswog, S. *A Simple, Entropy-Based Dissipation Trigger for SPH*. The Astrophysical Journal, **898**(1) (2020b), 60. ISSN 1538-4357.

Saitoh, T.R., Makino, J. *A DENSITY-INDEPENDENT FORMULATION OF SMOOTHED PARTICLE HYDRODYNAMICS*. The Astrophysical Journal, **768**(1) (2013), 44. ISSN 0004-637X, 1538-4357.

Schaller, M., Dalla Vecchia, C., Schaye, J., et al. *The Eagle Simulations of Galaxy Formation: The Importance of the Hydrodynamics Scheme*. Monthly Notices of the Royal Astronomical Society, **454**(3) (2015), 2277. ISSN 0035-8711, 1365-2966.

Schaller, M., Gonnet, P., Chalk, A.B.G., et al. *SWIFT: Using Task-Based Parallelism, Fully Asynchronous Communication, and Graph Partition-Based Domain Decomposition for Strong Scaling on More than 100,000 Cores*. Proceedings of the Platform for Advanced Scientific Computing Conference on - PASC '16 (2016), 1. 1606.02738.

Schaller, M., Gonnet, P., Draper, P.W., et al. *SWIFT: SPH With Inter-Dependent Fine-Grained Tasking*. Astrophysics Source Code Library (2018), ascl:1805.020.

Schaye, J., Crain, R.A., Bower, R.G., et al. *The EAGLE Project: Simulating the Evolution and Assembly of Galaxies and Their Environments*. Monthly Notices of the Royal Astronomical Society, **446**(1) (2015), 521. ISSN 1365-2966, 0035-8711.

Schaye, J., Dalla Vecchia, C. *On the Relation between the Schmidt and Kennicutt-Schmidt Star Formation Laws and Its Implications for Numerical Simulations*. Monthly Notices of the Royal Astronomical Society, **383**(3) (2008), 1210.

Schaye, J., Dalla Vecchia, C., Booth, C.M., et al. *The Physics Driving the Cosmic Star Formation History*. **402**(3) (2010), 1536. 0909.5196.

Schmidt, M. *The Rate of Star Formation.* **129** (1959), 243.

Schneider, E.E., Robertson, B.E. *Cholla : A New Massively-Parallel Hydrodynamics Code For Astrophysical Simulation*. The Astrophysical Journal Supplement Series, **217**(2) (2015), 24. ISSN 1538-4365. 1410.4194.

SciPy 1.0 Contributors, Virtanen, P., Gommers, R., et al. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, **17**(3) (2020), 261. ISSN 1548-7091, 1548-7105.

Sedov. *Similarity and Dimensional Methods in Mechanics - NASA/ADS* (1959).

Sembolini, F., Yepes, G., Pearce, F.R., et al. *nIFTy Galaxy Cluster Simulations – I. Dark Matter and Non-Radiative Models*. Monthly Notices of the Royal Astronomical Society, **457**(4) (2016), 4063. ISSN 0035-8711, 1365-2966.

Smith, B.D., Bryan, G.L., Glover, S.C.O., et al. *Grackle: Chemistry and Radiative Cooling Library for Astrophysical Simulations* (2016). 1612.020.

Sod, G.A. *A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws*. Journal of Computational Physics, **27**(1) (1978), 1. ISSN 00219991.

Somerville, R.S., Popping, G., Trager, S.C. *Star Formation in Semi-Analytic Galaxy Formation Models with Multiphase Gas*. Monthly Notices of the Royal Astronomical Society, **453**(4) (2015), 4338. ISSN 0035-8711, 1365-2966.

Somerville, R.S., Primack, J.R. *The Star Formation History in a Hierarchical Universe*. arXiv e-prints (1998), astro-ph/9811001. astro-ph/9811001.

Springel, V. *The Cosmological Simulation Code Gadget-2*. Monthly Notices of the Royal Astronomical Society, **364**(4) (2005), 1105. ISSN 0035-8711, 1365-2966.

Springel, V. E Pur Si Muove: *Galilean-Invariant Cosmological Hydrodynamical Simulations on a Moving Mesh*. Monthly Notices of the Royal Astronomical Society, **401**(2) (2010), 791. ISSN 00358711, 13652966.

Springel, V., Di Matteo, T., Hernquist, L. *Modelling Feedback from Stars and Black Holes in Galaxy Mergers*. **361**(3) (2005a), 776. astro-ph/0411108.

Springel, V., Hernquist, L. *Cosmological Smoothed Particle Hydrodynamics Simulations: The Entropy Equation*. Monthly Notices of the Royal Astronomical Society, **333**(3) (2002), 649. ISSN 0035-8711, 1365-2966.

Springel, V., Hernquist, L. *Cosmological Smoothed Particle Hydrodynamics Simulations: A Hybrid Multiphase Model for Star Formation*. Monthly Notices of the Royal Astronomical Society, **339**(2) (2003), 289. ISSN 0035-8711, 1365-2966.

Springel, V., White, S.D.M., Jenkins, A., et al. *Simulations of the Formation, Evolution and Clustering of Galaxies and Quasars*. **435**(7042) (2005b), 629. astro-ph/0504097.

Steinwandel, U.P., Moster, B.P., Naab, T., et al. *Hot Phase Generation by Supernovae in ISM Simulations: Resolution, Chemistry, and Thermal Conduction*. Monthly Notices of the Royal Astronomical Society, **495**(1) (2020), 1035. ISSN 0035-8711, 1365-2966.

Stern, J., Fielding, D., Faucher-Giguère, C.A., et al. *Cooling Flow Solutions for the Circumgalactic Medium*. arXiv:1906.07737 [astro-ph] (2019). 1906.07737.

Sturm, E. *The Nature of Ultraluminous Infrared Galaxies*. In M. Harwit, M.G. Hauser, editors, *The Extragalactic Infrared Background and Its Cosmological Implications*, volume 204 of *IAU Symposium* (2001), page 179.

Taylor, G. *The Formation of a Blast Wave by a Very Intense Explosion. I. Theoretical Discussion*. Proceedings of the Royal Society of London Series A, **201**(1065) (1950), 159.

Taylor, P., Federrath, C., Kobayashi, C. *Star Formation in Simulated Galaxies: Understanding the Transition to Quiescence at 3 × 1010 M⊙*. Monthly Notices of the Royal Astronomical Society, **469**(4) (2017), 4249. ISSN 0035-8711, 1365-2966.

Teklu, A.F., Remus, R.S., Dolag, K., et al. *CONNECTING ANGULAR MOMENTUM AND GALACTIC DYNAMICS: THE COMPLEX INTERPLAY BETWEEN SPIN, MASS, AND MORPHOLOGY*. The Astrophysical Journal, **812**(1) (2015), 29. ISSN 1538-4357.

Teyssier, R. *Cosmological Hydrodynamics with Adaptive Mesh Refinement: A New High Resolution Code Called RAMSES*. Astronomy & Astrophysics, **385**(1) (2002), 337. ISSN 0004-6361, 1432-0746.

Thomas, N., Davé, R., Anglés-Alcázar, D., et al. *Black Hole – Galaxy Correlations in Simba*. Monthly Notices of the Royal Astronomical Society, **487**(4) (2019), 5764. ISSN 0035-8711, 1365-2966.

Tollet, É., Cattaneo, A., Macciò, A.V., et al. *NIHAO XIX: How Supernova Feedback Shapes the Galaxy Baryon Cycle*. Monthly Notices of the Royal Astronomical Society, **485**(2) (2019), 2511. ISSN 0035-8711, 1365-2966.

Tremmel, M., Karcher, M., Governato, F., et al. *The Romulus Cosmological Simulations: A Physical Approach to the Formation, Dynamics and Accretion Models of SMBHs*. Monthly Notices of the Royal Astronomical Society, **470**(1) (2017), 1121. ISSN 0035-8711, 1365-2966.

Tricco, T.S. *The Kelvin-Helmholtz Instability and Smoothed Particle Hydrodynamics*. Monthly Notices of the Royal Astronomical Society, **488**(4) (2019), 5210. 1907.03935.

Tumlinson, J., Peeples, M.S., Werk, J.K. *The Circumgalactic Medium*. Annual Review of Astronomy and Astrophysics, **55**(1) (2017), 389. ISSN 0066-4146, 1545-4282.

Turk, M.J., Smith, B.D., Oishi, J.S., et al. *Yt: A MULTI-CODE ANALYSIS TOOLKIT FOR ASTROPHYSICAL SIMULATION DATA*. The Astrophysical Journal Supplement Series, **192**(1) (2011), 9. ISSN 0067-0049, 1538-4365.

van den Bosch, F.C., Ogiya, G. *Dark Matter Substructure in Numerical Simulations: A Tale of Discreteness Noise, Runaway Instabilities, and Artificial Disruption*. Monthly Notices of the Royal Astronomical Society, **475**(3) (2018), 4066. ISSN 0035-8711, 1365-2966.

van Rossum, G., Drake Jr, F.L. *Python Tutorial*. In *Python Tutorial*, volume 620. Centrum voor Wiskunde en Informatica, Amsterdam (1995).

Vila, J.P. *On Particle Weighted Methods and Smooth Particle Hydrodynamics*. Mathematical models and methods in applied sciences, **9**(02) (1999), 161.

Vogelsberger, M., Genel, S., Springel, V., et al. *Introducing the Illustris Project: Simulating the Co-evolution of Dark and Visible Matter in the Universe*. Monthly Notices of the Royal Astronomical Society, **444**(2) (2014), 1518. ISSN 0035-8711, 1365-2966.

Vogelsberger, M., Marinacci, F., Torrey, P., et al. *Cosmological Simulations of Galaxy Formation*. Nature Reviews Physics, **2**(1) (2020), 42. ISSN 2522-5820.

Wadsley, J., Stadel, J., Quinn, T. *Gasoline: A Flexible, Parallel Implementation of TreeSPH*. New Astronomy, **9**(2) (2004), 137. ISSN 13841076.

Wadsley, J.W., Keller, B.W., Quinn, T.R. *Gasoline2: A Modern Smoothed Particle Hydrodynamics Code*. Monthly Notices of the Royal Astronomical Society, **471**(2) (2017), 2357. ISSN 0035-8711, 1365-2966.

Wadsley, J.W., Veeravalli, G., Couchman, H.M.P. *On the Treatment of Entropy Mixing in Numerical Cosmology*. Monthly Notices of the Royal Astronomical Society, **387**(1) (2008), 427. ISSN 0035-8711, 1365-2966.

Wechsler, R.H., Tinker, J.L. *The Connection between Galaxies and Their Dark Matter Halos*. **56** (2018), 435. 1804.03097.

Weinberger, R., Springel, V., Pakmor, R. *The AREPO Public Code Release*. The Astrophysical Journal Supplement Series, **248** (2020), 32. ISSN 0067-0049. 1909.04667.

Weinberger, R., Springel, V., Pakmor, R., et al. *Supermassive Black Holes and Their Feedback Effects in the IllustrisTNG Simulation*. **479**(3) (2018), 4056. 1710.04659.

Wendland, H. *Piecewise Polynomial, Positive Definite and Compactly Supported Radial Functions of Minimal Degree*. Advances in Computational Mathematics, **4**(1) (1995), 389. ISSN 1572-9044.

Wetzel, A.R., Hopkins, P.F., Kim, J.h., et al. *Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-Mass Galaxy*. **827** (2016), L23. 1602.05957.

Wheeler, C., Oñorbe, J., Bullock, J.S., et al. *Sweating the Small Stuff: Simulating Dwarf Galaxies, Ultra-Faint Dwarf Galaxies, and Their Own Tiny Satellites*. Monthly Notices of the Royal Astronomical Society, **453**(2) (2015), 1305. ISSN 0035-8711, 1365-2966.

White, C.E., Somerville, R.S., Ferguson, H.C. *A PARAMETRIC STUDY OF POSSIBLE SOLUTIONS TO THE HIGH-REDSHIFT OVERPRODUCTION OF STARS IN MODELED DWARF GALAXIES*. The Astrophysical Journal, **799**(2) (2015), 201. ISSN 1538-4357.

Wiersma, R.P.C., Schaye, J., Theuns, T., et al. *Chemical Enrichment in Cosmological, Smoothed Particle Hydrodynamics Simulations*. Monthly Notices of the Royal Astronomical Society, **399**(2) (2009), 574. ISSN 00358711, 13652966.

Zakamska, N.L., Hamann, F., Pâris, I., et al. *Discovery of Extreme [O Iii]* Λ*5007 Å Outflows in High-Redshift Red Quasars*. Monthly Notices of the Royal Astronomical Society, **459**(3) (2016), 3144. ISSN 0035-8711, 1365-2966.