Simulating the Universe: The evolution of the most massive galaxies

by

Douglas Rennehan
B.Sc., University of Victoria, 2015

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

Simulating the Universe: The evolution of the most massive galaxies

by

Douglas Rennehan
B.Sc., University of Victoria, 2015

Supervisory Committee

Dr. A. Babul, Supervisor
(Department of Physics and Astronomy)

Dr. J. Willis, Departmental Member
(Department of Physics and Astronomy)

Dr. A. Brolo, Outside Member
(Department of Chemistry)

ABSTRACT

The cores of galaxy clusters contain the most massive galaxies in the Universe, the brightest cluster galaxies. These galaxies are unique compared to their counterpart galaxies outside of clusters as they have much brighter cores, and vast spatially-extended stellar envelopes. The theoretical picture of how they reached their huge masses relied on the idea of gradual stellar mass growth during the second half of the history of the Universe. However, recent observational evidence of highly-overdense protoclusters, the progenitors of these galaxies, demonstrates that some brightest cluster galaxies may have assembled within the first few billion years after the Big Bang – seemingly contradicting our theoretical predictions. I include my theoretical work that shows the short timescales over which these observed protoclusters transform into the brightest cluster galaxies and discuss the likelihood of finding these rare protoclusters in the early Universe.

To push our understanding of the rapid evolution of these galaxies even further forward demands the use of numerical simulations due to the highly coupled, non-linear astrophysical processes that occur during the process. In this dissertation, I include improvements to our numerical models of hydrodynamical turbulence and supermassive black holes that I incorporated into a state-of-the-art hydrodynamical+gravity simulation code, in effort to provide the groundwork to improving our understanding of the build-up of the brightest cluster galaxies in the early Universe, and galaxy evolution in general.

Contents

List of Tables

List of Figures

List of Abbreviations

ADAF Advection dominated accretion flow.

AGB Asymptotic giant branch.

AGN Active galactic nuclei.

BCG Brightest cluster galaxy.

CCD Charged-coupled device.

CDM Cold dark matter.

CGM Circumgalactic medium.

CLASH Cluster Lensing And Supernova survey with Hubble.

DIFF The cold, diffuse gas phase.

DN No sub-grid dynamical friction.

DY Sub-grid dynamical friction.

Dyn Dynamic model.

FIRE Feedback in realistic environments.

Grad Gradient model.

GSMF Galaxy stellar mass function.

HHG Hot halo gas.

ICM Intracluster medium.

IMF Initial mass function.

ISM Interstellar medium.

JWST James Webb Space Telescope.

KH Kelvin-Helmholtz.

LOS Line of sight.

MDF Metal distribution function.

MFM Mesh-free finite mass.

MM Moving mesh.

MNRAS Monthly Notices of the Royal Astronomical Society.

NFW Navarro-Frenk-White.

NIRCam Near Infrared Camera.

pBB post-Big Bang.

RMSE Root mean squared error.

Smag Smagorinsky model.

SMBH Supermassive black hole.

SNIa Supernova type Ia.

SNII Supernova type II.

SpARCS Spitzer Adaptation of the Red-Sequence Cluster Survey.

SPH Smoothed particle hydrodynamics.

SPT South Pole Telescope.

WHIM Warm-hot intergalactic medium.

XCS XMM Cluster Survey.

XMM X-ray Multi-mirror Mission.

Acknowledgements

I would like to thank:

Arif Babul, for mentoring and unwavering encouragement. Especially, thank you for teaching me how to make an clear, concise argument in writing and in person.

Jon Willis and Alex Brolo, for asking insightful questions that challenged me during the PhD process.

Belaid Moa, for consistently bringing joy and excitement to both physics and computer science.

Fabrice Durier, for introducing me to the field of computational galaxy evolution.

Austin Davis, for discussing turbulence with me in the forest.

Collin Kielty, for giving me the key to unlock my mind.

Logan Francis, for the Nectar of the Gods.

Adrian and Kali, for existing.

Fiona Mortimer, for unwavering friendship.

Chervin, Isabel, Salvatore, Guillaume, Vida, and Nabi for introducing me to new cultures and ideas. I hope these friendships last a life-time.

My parents, for instilling in me a love of reading and answering all of my questions as a child.

Ondrea Clarkson, for making everything worth it.

If something is boring after two minutes, try it for four. If still boring, try it for eight, sixteen, thirty-two, and so on. Eventually one discovers that it's not boring at all but very interesting.

John Cage

Chapter 1

Introduction

Galaxy clusters are the most massive gravitationally bound structures that exist in the cosmos and contain thousands of individual galaxies. Clusters not only provide constraints on our theories of the evolution of the universe as a whole, but their dense environments also provide a laboratory for studying the impact of a plethora of physical processes on galaxy evolution. The dense cores of galaxy clusters house the most massive galaxies in the universe, the brightest cluster galaxies (BCGs) (Peebles, 1968; Sandage, 1976; Tremaine & Richstone, 1977). These galaxies are morphologically spheroidal, have a large spatial[1] extent ($\sim 100\,\mathrm{kpc}$), and are usually a factor of ~ 10 more luminous than a typical galaxy (Lin & Mohr, 2004; Pipino et al., 2011; Loubser et al., 2018). Importantly, many of the observed properties of the BCGs scale with the properties of their host cluster (Brough et al., 2008; Lidman et al., 2012; Kravtsov & Borgani, 2012; Lavoie et al., 2016) and they are usually found close to the gravitational potential center of their hosts (Bildfell et al., 2008; Lidman et al., 2013; Lauer et al., 2014). These observations suggest that the evolution of the BCGs and their host cluster are intimately linked, and that studying the former will provide clues about the formation and evolution of the latter. However, understanding evolutionary history of the BCGs has proved challenging and a complete picture remains elusive. The goal of my research is twofold: (a) to advance our understanding of the evolution the BCGs – the cores of galaxy clusters and (b) improve our numerical, astrophysical models to help achieve goal (a).

Our current understanding of how all galaxies, including the BCGs, form and evolve relies on the Λ Cold Dark Matter (ΛCDM) paradigm where an initial random field of mass-density fluctuations in the early universe lead to the formation of all

[1] Astronomical distances are based on the parsec, where $1\,\mathrm{pc} = 3.086 \times 10^{13}$ km. In these units $1\,\mathrm{Mpc} = 10^6\,\mathrm{pc}$ and $1\,\mathrm{Gpc} = 10^9\,\mathrm{pc}$.

of the structure we observe today, driven by gravity. The initial fluctuations, or over-densities, were comprised of modes of various scales – from the sub-galactic to the galaxy cluster scale, and beyond. Over time the universe expands and the matter in those over-densities that are above the threshold for gravitational collapse form gravitationally-bound dark matter structures which are referred to as *halos*. In this paradigm, the smallest structures collapse first and then gradually merge to build more and more massive structures (i.e. hierarchical assembly) – with galaxy clusters at the peak of the mass-assembly hierarchy at the present time. The initial assemblages of structures that eventually go on to form clusters are *protoclusters*, and they exist at early epochs in the universe (Overzier, 2016).

While dark matter is the prevalent form of matter in the universe ($\sim 25\%$ of the total mass-energy density) (Ade et al., 2016), it is the baryons[2] ($\sim 5\%$ of the total mass-energy density) that go on to form the structures that we observe. So, how exactly does a cluster form?

Consider the formation of a single dark matter halo that will eventually host a galaxy. The matter in the region is overdense compared to the mean background density and, therefore, gravity forces the dark matter and gas to collapse toward the overdensity peak. Our current understanding is that dark matter is collisionless[3] and, therefore, the dark matter falls through the central region without immediately settling in equilibrium. Simultaneously, the gas follows the dark matter. However, gas *is* collisional and self-interacts when it collide. As the infalling gas collides with gas that fell in earlier, it shocks and the kinetic energy of the infalling gas converts to thermal energy. Simultaneously, the entire system begins to approach a dynamical equilibrium state known as *virial equilibrium*[4]. The hot gas additionally reaches a state of hydrostatic equilibrium where the pressure balances gravity. However, in order for a galaxy to form at the center of this halo, the gas must lose thermal energy.

Two-body processes are the most important processes through which gas loses its internal energy. This includes two-body processes within a gas where the gas loses its energy through radiation, or *radiative cooling*. At gas temperatures $10^4\,\mathrm{K} \lesssim T \lesssim 10^6\mathrm{K}$, there are three dominant two-body mechanisms that cause the gas to lose energy: collisional ionization, recombination, and collisional excitation. Collisional

[2]When I say *baryon*, I am referring to all matter that is not dark matter and should not be interpreted in the particle physics sense.

[3]It is free to pass through itself without interaction, besides gravity.

[4]We say the halo is *virialized*. Virialization occurs when the kinetic energy is half of the potential energy in the system, $KE = -\frac{1}{2}PE$.

ionization occurs when atoms and electrons collide in a gas causing atoms to lose energy equal to the ionization threshold as they lose an electron. Recombination occurs in an ionized gas where there are free electronic states. Any free electrons may be able to recombine into the atoms, leading to photon emission. Finally, collisional excitation occurs when electron-atom collisions cause the electron to move up energy levels, and then fall back down – emitting a photon. All three processes lead to kinetic energy in the gas leaving the system.

Below $T \lesssim 10^4$ K, the gas is mostly neutral and two-body processes become negligible. Conversely, above $T \sim 3{\times}10^7$ K, pristine and enriched gas is mostly ionized. In that regime, the dominant radiative cooling mechanism is *bremsstrahlung*[5]. The free electrons that move through the gas accelerate as they pass free ions and emit photons. Those photons represent energy loss from the system, and the gas cools.

Very importantly, these radiative processes are highly dependent on the *metallicity*[6] Z of the gas, the temperature T (as I mentioned), as well as the density of the gas, ρ. Gas that is enriched to the same metallicity as the Sun has factors of up to ~ 100 times faster cooling rates for enriched gas compared to pristine (i.e., $Z = 0$) gas. As a consequence of two-body processes, the overall cooling rate is proportional to ρ^2, i.e. $\mathrm{d}E/\mathrm{d}t = \rho^2\Lambda(Z,T)$. That leads much more rapid cooling in the centers of the overdense regions compared to the outskirts.

As the gas in the center of the dark matter halo cools and leaves equilibrium, the process of creating stars begins. Gas that is sufficiently cold ($T \lesssim 10^4$ K) may form molecular[7] cloud structures that will eventually become the birthplaces of stars. Some of those stars, when they reach the end of their lives, end up as supernova explosions that inject enormous amounts of energy back into the cooling gas – potentially preventing further cooling – creating a feedback loop (see reviews in Naab & Ostriker 2017, for example).

The feedback loop is of prime importance to the field of galaxy evolution. In the absence of stellar feedback, gas within virialized halos would simply continue to cool until it was all converted into stars (e.g., White & Rees 1978; Katz 1992; Benson 2010) – producing galaxies factors of $\sim 10{-}100$ more massive than observed. Assuming each dark matter halo contains a fraction f_b of their total mass in baryons, we would expect a stellar mass of $M_* = f_\mathrm{b} M_\mathrm{halo}$ after the conversion process. However, our observations

[5]German for *braking*.

[6]Metallicity is the combined mass fraction of all elements more massive than helium in a gas. We define $X + Y + Z = 1$, where X is the mass fraction of hydrogen, Y helium, and Z all other elements.

[7]Composed predominantly of H_2.

of the number density of galaxies as a function of their stellar mass show that the stellar mass underestimates this prediction (Tomczak et al., 2014; Somerville & Davé, 2015). Stellar feedback is the leading theory as to why this happens in galaxies with stellar masses $M_* \lesssim M_{\rm MW}$, where $M_{\rm MW}$ is the mass of our Milky way galaxy. In those systems, the energy from stellar feedback is sufficient to expel gas from the galaxy, preventing it from cooling and forming stars. However, at masses $M_* \gtrsim M_{\rm MW}$, stellar feedback is too weak to expel gas from the system (see, for example, Davé et al. 2016) and we require additional explanation.

One solution to the apparent lack of stellar mass in the most massive systems is active galactic nuclei (AGN) feedback (Babul et al., 2002; McNamara & Nulsen, 2007). AGN are found in a large fraction of galaxies (Perna et al., 2017; Fiore et al., 2017), and are associated with the bright core regions of galaxies that present non-thermal emission. Their high luminosities have been associated with galactic scale energetic feedback that could prevent star formation (Springel, 2005). There are also spectacular examples of AGN that operate on much larger scales than galactic cores. These are found in the BCGs and produce powerful relativistic jets that inflate bubbles in the intracluster medium (i.e. the gas within clusters; the ICM), possibly offsetting gas cooling (Reynolds et al., 2005; McCarthy et al., 2008; Zhuravleva et al., 2014). Our current understanding of AGN is that they are powered by the accretion of gas onto a supermassive black hole (SMBH). SMBHs are observed to typically have masses in the range $M \sim 10^6 - 10^{10}$ $\mathrm{M}_\odot$[8] (Mcconnell et al., 2011; Kormendy & Ho, 2013) and are typically found near the potential centers of their host galaxies, with the most massive in the Universe found in the BCGs.

The outflows driven by stellar and AGN feedback processes interact strongly with the gas that permeates and surrounds galaxies. Specifically, they inject kinetic energy that drives turbulence in these media. If the impacted gas cannot convert the injected kinetic energy rapidly into thermal energy, a turbulent cascade of motion will form with eddies on the largest scales breaking up on smaller and smaller scales, until they dissipate at the smallest scales (Pope, 2000). These eddies transport momentum, thermal energy, and metals throughout the gas. In fact, turbulence is one mechanism that may explain the redistribution of AGN jet energy in the ICM, allowing the suppression of gas cooling onto the BCG (Reynolds et al., 2005). However, perhaps counterintuitively, turbulence driven by AGN feedback may enhance gas cooling in the intracluster medium since turbulence promotes the production of density fluctuations

[8]Where $\mathrm{M}_\odot = 1.989 \times 10^{33}$ g is the mass of the Sun.

that subsequently cool and fall into the BCG (Prasad et al., 2015; Gaspari et al., 2017; Prasad et al., 2018).

In the case of BCGs, the aforementioned physical processes work in concert over cosmic time to produce extremely massive systems. With these processes as a backdrop, there are three main broad scenarios that could explain the growth of BCGs across cosmic time: (i) extended star formation over the history of the universe, (ii) monolithic collapse, and (iii) early star formation in separate galaxies and late merging.

In scenario (i), efficient gas cooling in the ICM is invoked to explain the growth of stellar mass in the BCGs. In this picture, gas in the cluster core cools and coalesces toward the center of the cluster potential where it should then go on to supply fuel for star formation. Indeed, the cooling rate in the cores of some galaxy clusters is quite short[9] ($\tau_{\rm cool} \sim 0.1$ **Gyr**) and, therefore, rapid star formation at rates ~ 1000 $\mathbf{M_\odot\,yr^{-1}}$ are expected at the centers of these systems (Cowie & Binney, 1977; Fabian & Nulsen, 1977). However, the majority of BCGs are quiescent[10] and have star formation rates $\lesssim 100$ $\mathbf{M_\odot\,yr^{-1}}$. This is the *cooling flow* problem and usually AGN feedback on large-scales in the ICM is invoked to explain the discrepancy. However, it is not clear that AGN feedback is the main agent responsible for the lack of star formation. Additionally, observations indicate that the majority of BCG stars are old, implying that continual star formation could not be the main driver in BCG stellar mass growth.

Scenario (ii) suggests that all massive elliptical galaxies, including BCGs, form from a single mass over-density collapse. This is the *monolithic collapse* scenario (Eggen et al., 1962) and does not rely on galaxies merging across all of cosmic time such as in **ΛCDM**. If the galaxies formed under this scenario, then their sizes should remain relatively constant over time since their size would be set by the initial mass overdensity. However, there is evidence of size growth of the BCG population when comparing across various epochs, indicating that mergers must play a role in the growth of the BCGs.

Finally, scenario (iii) explains the growth of BCGs by invoking the **ΛCDM** picture and is the commonly accepted scenario. First, in protocluster regions in the early-universe (~ 1.5 **Gyr** post-Big Bang; henceforth, **pBB**) the galaxies that will eventually become the BCG have already formed the bulk of their stars (De Lucia et al., 2006).

[9]Compare with a typical dynamical timescale $\tau_{\rm dyn} \sim 1$ Gyr.

[10]Quiescent means low star formation rates per unit stellar mass.

Later, these galaxies gradually and continually merge into the main progenitor of the BCG and constantly build up more and more stellar mass. In this picture, the galaxy mergers lack sufficient gas to fuel significant star formation and only bring their old stars into the BCG. Theoretical studies estimate that $50-60\%$ of the BCGs' present-day stellar mass is incorporated at late times, over the last ~ 10 Gyr (Dubinski, 1998; Conroy et al., 2007; De Lucia & Blaizot, 2007; Ruszkowski & Springel, 2009; Laporte et al., 2013; Laporte & White, 2015).

Recent observations, however, suggest that the accepted picture may not be fully complete and that early assembly may play a role in the evolution of the BCGs. First, several BCGs have been discovered $\sim 4-6$ Gyr pBB (Collins et al., 2009) that have stellar masses comparable to present-day[11] BCGs ($\sim 10^{12}$ $M_\odot$). If those BCGs grew at the rates predicted by theoretical models in the following $\sim 8-10$ Gyr until present times, they would be more massive than any known galaxy. Additionally, a new paradigm wherein BCGs are no longer thought to be completely quiescent is emerging due to increasing larger observational surveys of more of the BCG population (Webb et al., 2015; McDonald et al., 2016; Bonaventura et al., 2017). As the surveys look further in the past it appears that the star formation rates in BCGs increase to levels of ~ 1000 $M_\odot$ yr^{-1} similar to those predicted by the cooling flow scenario in the cluster cores. This then begs the question: why are the star formations rates so high if AGN feedback offsets the star formation? Why would AGN not operate effectively at early times in the Universe, but very effectively at late times? Clearly the picture is not complete.

Furthermore, the story is muddled by the unprecedented existence of highly-overdense (in terms of galaxy density) protoclusters. A fraction of the galaxies that comprise protoclusters are the progenitors of BCGs and, therefore, they provide clues to their formation and evolution. While there have been example of protoclusters (see Overzier 2016 for a review) and recently, fully-developed clusters (Willis et al., 2020; Tilvi et al., 2020) at early times ($\sim 0.5-3$ Gyr pBB), even more spectacular systems have emerged that are pushing the boundaries of our understanding of the Universe. Recent observations show an increasing number of protoclusters with multiple ($\gtrsim 6$) massive galaxies on the verge of collapse in compact regions on the sky. These galaxies will become the BCGs of the resulting galaxy clusters and, given their overdensities, must merge more quickly than in the standard picture. One of the most stunning examples of an over-dense protocluster at early times is SPT2349-56 (Miller et al.,

[11]The age of the Universe is ~ 13.8 Gyr.

2018) – an assemblage of 14 galaxies with stellar masses ranging from 5×10^9 $M_\odot$ to 5×10^{10} $M_\odot$, all within a 130 kpc region on the sky, observed at $\sim$ 1.4 Gyr pBB, and has an estimated total star formation rate of $\sim$ 6500 $M_\odot\,yr^{-1}$. For comparison, the Milky Way has a stellar mass of $\sim 5 \times 10^{10}$ $M_\odot$ and a characteristic diameter of $\sim$ 30 kpc, with a star formation rate of $\sim$ 1 $M_\odot\,yr^{-1}$. The observational evidence for early, highly star forming protoclusters is growing (Ishigaki et al., 2016; Jiang et al., 2018; Higuchi et al., 2019), suggesting that at least a subset of the BCG must form very quickly in the early universe (Ito et al., 2019; Rennehan et al., 2020).

The astrophysical community is entering a new era in the understanding of the evolution of BCGs as our observations reach further in distance and time. Clearly, in light of the observations outlined above, the once solidified picture of gradual hierarchical assembly of the most massive galaxies in the Universe is faltering and new ideas must be injected. My thesis will add pieces to a new picture for how the BCGs evolve through addressing the follow questions: How quickly do the observed early protoclusters assemble a BCG? What fraction of clusters host these highly over-dense protoclusters in their history? However, to address these questions, I require numerical cosmological simulations that have physically-motivated and accurate models for the physical processes that I outlined above.

1.1 Numerical simulations

The complexity, interdependence, and inherent non-linearity of the physical processes outlined above demands the use of numerical simulations. The majority of open theoretical problems in the field of galaxy evolution may no longer be approached with simple arguments and analytic methods. There are two classes of simulations that have been successful at advancing our knowledge of galaxy evolution in general: semi-analytic modelling and numerical integration. In the former, simplified analytic models are employed that treat galaxies as abstract objects whose properties change via updates to analytic approximations to known physical processes. These have been successful at advancing the picture of late brightest cluster galaxy formation (for example, see De Lucia & Blaizot 2007). In the latter case, the dynamical, hydrodynamical, and thermodynamical equations are integrated in an expanding universe through time (i.e. ΛCDM) following the interactions of dark matter, gas, stars, and supermassive black holes. I will focus on the latter and the treatment of models necessary to accurately track the complex physics involved.

I want to stress the importance of numerical simulations as a tool for understanding our universe. I consider computational astrophysics to be the fourth branch of astronomy, the other three being observational astronomy, instrumentation, and pure theoretical astrophysics. Computational astrophysics is unique in that it inherently combines all of the branches into a single unified platform. As computational astrophysicists, we must *observe* our simulations, *improve* the tools, and *propose explanations* of what we see. I must clarify the latter point: in some cases new developments arise from simulation and we must explain those, but a great power lies in using simulations to test theories proposed *outside* of simulation. That power comes from the ability to test the parameter space and qualitatively show that a theory is plausible. However, my problem with the state of computational astrophysics is that many of the underlying tools, or models, are not well tested and motivated (see Chapter 3). There are also models that have been tried and tested from fields outside of astrophysics but not incorporated in contemporary simulations. I focus heavily on that topic in Chapter 2. My main motivator for studying how simulations work is a question I ask myself regularly: how can we trust the qualitative results of our simulations if we do not have a strong numerical base? For that reason, I will spend the next half of the introduction discussing numerical simulation and how it impacts galaxy evolution.

I spoke of numerical integration, but what does that mean? How does a numerical simulation actually work? At the fundamental level, a numerical simulation is really calculation that one could do by hand given sufficient time. Take the simplest example of uniform motion of a particle that starts at some position x_0 with initial speed v_0 in some coordinate system. The distance from the origin evolves following the differential equation $v = \mathrm{d}x/\mathrm{d}t$ or $\mathrm{d}x = v \cdot \mathrm{d}t$. We find the position as a function of time by integrating over time, $x(t) = \int_0^t v_0 \cdot dt' = \sum_i v_0 \Delta t_\mathrm{i} + x_0$. Here I replace the integral with a discrete sum, or I discretize time. That equation is trivial to solve by hand – you need to know the initial velocity v_0 and position x_0, and sum up all of the contributions over small times Δt. These are the initial conditions and timestep that one would use to *simulate* the particle's motion, if one implemented an algorithm into a computer to solve the equation.

Indeed, numerical simulation is simply a collection of algorithms that force a computer to solve (partial-)differential equations (usually coupled) that describe some physical processes evolving in time given some initial conditions in space and time. We then make a fundamental *assumption* that the numerical simulation solution maps to reality. It is important to note that it is a misnomer to say that the computer solves

anything; at best, the computer approximates the solutions to the differential equations. In a purely mathematical approach, we would be able to work in continuous space but computers cannot provide continuous solutions. The computer requires initial conditions that are discretized in space, i.e. at sample points across the spatial domain of interest. The word *integration* appears because we integrate in time and occasionally in space in order to approximate the solutions to the equations. Therefore, in addition to not providing a true solution (if one even exists), the computer also provides results that are averaged over some spatio-temporal scale. Or, to say in another way, they are limited in spatio-temporal *resolution*.

The extremely simplified picture above applies to the equations that govern the evolution of galaxies. I raise two important questions: (a) What are the fundamental equations that govern galaxy evolution? And, (b) how do we solve them?

1.1.1 What are the fundamental equations of galaxy evolution?

There are numerous components that go into calculating how a galaxy forms and evolves in time but the first two-thirds of this dissertation focuses on the gaseous component of the Universe.

We treat gas on cosmological scales as an ideal, perfect gas that behaves following the hydrodynamical equations. The equations used in galaxy evolution simulations usually are the Euler equations, a set of three conservation equations for mass, momentum, and energy:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \Big(\rho \boldsymbol{u} \Big) = 0, \tag{1.1}$$

$$\frac{\partial (\rho \boldsymbol{u})}{\partial t} + \boldsymbol{\nabla} \cdot \Big(\rho \boldsymbol{u} \otimes \boldsymbol{u} + P \cdot \boldsymbol{I} \Big) = 0, \tag{1.2}$$

$$\frac{\partial (\rho e)}{\partial t} + \boldsymbol{\nabla} \cdot \Big(\rho \boldsymbol{u} e + \boldsymbol{u} P \Big) = 0, \tag{1.3}$$

where ρ is the gas density, $\boldsymbol{u}$ is the gas velocity, P is the gas pressure, e is the gas specific total energy, and $\boldsymbol{I}$ is the identity matrix. Equations 1.1, 1.2, & 1.3 represent mass, momentum, and energy conservation, respectively. The simplest way to understand the Euler equations is through:

$$\frac{\partial}{\partial t} \Big(\text{property that changes} \Big) = -\boldsymbol{\nabla} \cdot \Big(\text{properties that drive changes} \Big). \tag{1.4}$$

The spatial differences of properties on the right hand side of equation 1.4 determine the change in time of the quantities on the left hand side. The hydrodynamical state of the gas, $(\rho, \boldsymbol{u}, e)$ at any point $(\boldsymbol{x}, t)$ requires knowledge of the gas pressure P. To close the equations, we assume the gas is an ideal, polytropic gas with $P = \rho e(\gamma - 1)$. γ is related to the degrees of freedom in the gas and we assume that the gas is monatomic, so $\gamma = 5/3$.

In order for gas to sink to the centres of dark matter halos to form the stars that form galaxies, we must consider additional processes. Specifically, gas may lose energy through radiative cooling, or gain energy from supernova or supermassive black hole feedback. These two contributions appear as sink and source terms on the right side of equation 1.3, commonly labelled as Λ and Γ, respectively.

The Euler equations give solutions for gas density, not for dark matter or stars. We consider dark matter and stars to be collisionless, i.e. they do not self-interact – therefore, they require different[12] treatment (Binney & Tremaine, 2011). Stars and dark matter are not the focus of this dissertation and, therefore, I intentionally leave discussion of their conservation equations[13] out of this work.

The most fundamental interaction for all matter is gravity. Dark matter, gas, and stars all interact gravitationally and we account for their contribution to motion through their contributions to the gravitational potential, $\phi = \phi(x, y, z, t)$. Specifically, Poisson's equation allows us to find the gravitational potential given the densities of each component of matter,

$$\nabla^2 \phi = 4\pi G(\rho_{\rm DM} + \rho_{\rm gas} + \rho_{\rm stars}), \tag{1.5}$$

where G is Newton's gravitational constant and $\rho_{\rm DM}$, $\rho_{\rm gas}$, and $\rho_{\rm stars}$ are the densities of dark matter, gas, and stars, respectively. The gravitational force follows directly via $\boldsymbol{F} = -\boldsymbol{\nabla}\phi$.

1.1.2 How do we solve the equations?

There are two main approaches to solving the equations of motion I presented above: the Eulerian approach and the Lagrangian approach.

The Eulerian approach involves using a grid of sample points that are fixed in space

[12] Stars are collisionless on our scales of interest. Only in the densest regions of stellar clusters within individual galaxies would additional interactions be necessary. Those scales are much below the simulation scales I discuss in this dissertation.

[13] They are derived from the collisionless Boltzmann equation.

in order to calculate the time rate of change of the dynamical variables from conservation equations (equations 1.1, 1.2, & 1.3). In a simpler sense, the Eulerian approach treats the frame of reference as a stationary observer transporting mass, energy, and momentum between sample points on the grid. There are numerous improvements and technicalities that I ignore here as my work focuses on the Lagrangian method but I point the interested reader to Bryan et al. (2014) for the current state-of-the-art for Eulerian methods.

Lagrangian hydrodynamical simulations decompose the matter fields into particles, usually of finite (constant) mass. In my work, I only consider finite-mass methods since these are the most relevant in the astrophysical community (see the introductions of Springel 2010 and Hopkins 2015 for detailed comparisons). The particles follow the motion of the fluid as it moves through the simulated volume rather than acting as a fixed stationary observer like in the Eulerian case. However, unlike the Eulerian method, the Lagrangian method cannot move mass between its elements as an inherent feature.

There are a several important reasons why I choose Lagrangian methods over Eulerian methods: their (a) conservation properties, (b) qualitative interpretation, and (c) simple treatment of gravity.

For (a), Lagrangian methods naturally conserve angular momentum whereas Eulerian methods do not. I consider this to be the most important point for astrophysical simulations since rotation is ubiquitous, i.e. almost all galaxies are rotating. The difference between the methods arises due to the differences in errors within the numerical approximations. In the Eulerian method, errors propagate in the direction of the grid since that is the direction over which the fluid properties move. Imagine simulating a galaxy that is tilted with respect to the grid. The errors will cause the galaxy to torque toward alignment with the grid, causing artificial gas mixing in that direction (for example, see Hahn et al. 2010 or Hopkins 2015). That leads to galaxies that may have different physical properties in the simulation than reality, changing our understanding of galaxy evolution just by the numerical method alone!

For (b), having particles allows for very simple tracking of mass movement throughout a simulation volume. For example, stellar feedback produced in the interstellar medium of galaxy produces a *galactic-scale wind* that creates, and adds to, the surrounding medium – the circumgalactic medium. As the star formation rate drops, and the stellar feedback diminishes, the gas in the circumgalactic medium may cool back into the interstellar medium. Tracking the movement of that gas in a simulation

is simple in a Lagrangian simulation since the particles track the mass movement directly. There are a series of important works that explore this exact physical picture in Oppenheimer & Davé (2006), Oppenheimer & Davé (2008), and Oppenheimer et al. (2010) using Lagrangian simulations. If these studies had used the Eulerian approach, additional equations (e.g. tracer particles) would be necessary to trace the mass as it moves in and out of the circumgalactic and interstellar media (see, for example, Cadiou et al. 2019). Additional equations leads to additional error that would have to be dealt with.

For (c), using finite mass particles allows us to treat gravity in a simple manner between dark matter, stars, gas, and supermassive black holes. All particles that interact gravitationally follow Newton's law of gravity,

$$\boldsymbol{F} = \frac{-Gm_1m_2}{r^2}\hat{\boldsymbol{r}} \tag{1.6}$$

where r is the distance between the particles, $\hat{\boldsymbol{r}}$ is the unit vector pointing in the direction between the particles, and m_1 and m_2 are the masses of the particles. Interestingly, in some Eulerian simulations dark matter and stars are often treated as finite mass particles whereas gas is treated on a grid since this method of treating gravity is so simple (O'shea et al., 2004; Bryan et al., 2014).

Historically, the most important Lagrangian hydrodynamics method for studying galaxy evolution is smoothed particle hydrodynamics (SPH) (Gingold & Monaghan, 1977; Lucy, 1977; Gingold & Monaghan, 1982; Hernquist & Katz, 1989; Katz, 1992; Springel & Hernquist, 2002). The method computes the gas density at some point (x, y, z) of a gas by averaging (i.e., weighted integration) the mass over N neighbouring interpolation points (i.e., particles) within a small spherical radius h (Monaghan, 2005). The inherent *smoothing* of properties leads to a loss of small-scale structure in the gas (Monaghan, 2002), and difficulty in simulating turbulence (Bauer & Springel, 2012). Regardless of the difficulties with the method, the public release of the SPH-based simulation code `GADGET-2` (Springel, 2005) initiated a revolution in our ability to simulate the Universe and in our understanding of dark matter (Springel et al., 2005b; Boylan-Kolchin et al., 2009) and galaxy evolution (Kereš et al. 2005; Oser et al. 2010; Oppenheimer et al. 2010; Behroozi et al. 2013b; Hopkins et al. 2014; Schaye et al. 2015; see Somerville & Davé 2015 for an overview).

In this dissertation, I use the mesh-free finite mass (MFM) method (Lanson &

Vila, 2008a,b; Gaburov & Nitadori, 2011). The method has origins in SPH[14], but is fundamentally different in how it approaches approximating the conservation equations. The method does not smooth the gas properties over a neighbouring particles within a distance h, but rather treats the interactions between the N neighbours as N individual 1D hydrodynamical Riemann problems (Hopkins, 2015). In short, the Riemann problem for the Euler conservation equations involves a 1D gas that has a left and right state of mass, momentum, and energy: $(\rho_{\rm L}, \rho u_{\rm L}, \rho e_{\rm L})$ and $(\rho_{\rm R}, \rho u_{\rm R}, \rho e_{\rm R})$, respectively. These states are separated by a discontinuity in their properties (Toro, 2009). The solution to the Riemann problem gives the *flux* of a quantity from one state to the other or, in our case, the flux of the gas property from one particle to another. It is possible to solve the Riemann problem exactly (Roe, 1981), but usually approximations are made since the exact solution is computationally expensive (Springel, 2010). The contributions from the N approximations to the Riemann solution are then weighted and summed into a final value for the time derivatives in the conservation equations (Hopkins, 2015). The method boasts the benefits of a Lagrangian method, while giving much more accuracy than the original SPH method.

The equations I discussed are fundamental in the sense that they describe the hydrodynamical and gravitational basis for all cosmological simulations. There are many different approaches to solving these equations across many implementations in public and private code-bases, each with their own intricacies and problems (see, for example, Springel 2005; O'shea et al. 2004; Springel 2010; Hopkins 2013, 2015; Wadsley et al. 2017; also Somerville & Davé 2015). I choose to use the `GIZMO` code (Hopkins, 2015) that has the MFM method built-in.

1.2 Cosmological simulations and sub-grid models

Numerical simulations of galaxy evolution usually begin at redshifts[15] $z \gtrsim 100$ when the distribution of matter is relatively uniform save for a field of random over-density fluctuations. Upon integrating the Newtonian gravitational interactions the over-densities lead to structure formation following the **ΛCDM** picture presented in the above introduction. These simulations need to treat large volumes of the Universe to

[14]Outside of the astrophysics community, the MFM method is a subset of SPH-ALE methods (**A**rbitrary **L**agrangian-**E**ulerian). See, for example, Frontiere et al. 2017.

[15]In an expanding universe, redshift corresponds to the relative size of the Universe compared to today. The scale factor $a(t) = 1/(1+z)$ is $a(0) = 1$ today and, for example, at $z = 1$ the Universe was *half* of its current size (i.e. $a = 1/2$). For reference, $z = 1$ was about 7 Gyr ago and $z = 2$ was about 10 Gyr ago (approximately 4 Gyr after the Big Bang).

gain insight into galaxy populations in addition to reducing statistical fluctuations. Usually, the cubical volumes cover side-lengths of $L \sim 150$ Mpc which should contain sufficient galaxies (hundreds of thousands) to provide a representative sample of the entire universe (Vogelsberger et al., 2014; Schaye et al., 2015) while simultaneously capturing the large-scale gravitational tidal fields that govern matter flow. However, contemporary computational limitations restrict the smallest scales that can be simulated simultaneously with the largest scales and require additional models to account for small-scale astrophysical processes.

The smallest scales that modern cosmological simulations reach is ~ 200 pc (on average typically higher at ~ 1 kpc) whereas star formation, stellar feedback, active galactic nuclei feedback, and turbulence operate on the sub-pc scale. Therefore, these physical processes are broadly modelled using *sub-grid* models, aptly named since they treat the processes occurring below the numerical grid scale (i.e. the resolution scale). Sub-grid models may be phenomenological, physically derived approximations, or even completely ad hoc and the applicability and validity of any approach is subject to debate (see Winsberg 2010 for an excellent philosophical overview; also see the introduction of Schaye et al. 2015).

A large portion of this dissertation deals with the applicability and impact of the sub-grid physics involved in understanding massive brightest cluster galaxy evolution. Specifically, I consider the star formation and stellar feedback models as accurate and well-calibrated, and work to improve the turbulence and active galactic nuclei aspects. Therefore, I require a set of sub-grid models on which to base my work. My work on improving treatments of sub-grid models involves extending the `MUFASA` (Davé et al., 2016) and `SIMBA` (Davé et al., 2019) galaxy evolution sub-grid models, which include radiative cooling and models for star formation, stellar feedback, and active galactic nuclei feedback. These simulations are calibrated to reproduce observed galaxy properties across cosmic time and act as a good starting point for improving turbulence and active galactic nuclei models. I choose these models since they are the only set available for `GIZMO` that uses the state-of-the-art MFM method.

1.3 Dissertation overview

This dissertation contains a compilation of projects on sub-grid models and investigation of the evolution of brightest cluster galaxies (BCGs). Chapter 2 contains two projects that improve the treatment of turbulence in cosmological simulations

that have impact across all fields of astrophysics, although my focus is on galaxy evolution and the evolution of metal distribution functions. Specifically, Section 2.1 and Section 2.2 are two papers on the topic that were accepted into the *Monthly Notices of the Royal Astronomical Society* (MNRAS) journal. Chapter 3 contains my unpublished work on supermassive black hole modelling in cosmological simulations. That work is a viability study of combining the most physical models of supermassive black holes into a single code-base for eventual use to study the early evolution of BCGs. Chapter 4 contains work published in MNRAS on the early evolution of BCGs from highly-overdense protoclusters, and the probability of finding these objects in the Universe. Chapter 5 contains a summary of my overall conclusions on each topic.

Chapter 2

Numerical models of turbulence

2.1 Dynamic localised turbulent diffusion and its impact on the galactic ecosystem

Originally published in Monthy Notices of the Royal Astronomical Society: 2019, vol. 483, issue 3, pp 3810-3831. Reproduced by permission of Oxford University Press. https://doi.org/10.1093/mnras/sty3376

Authours: Douglas Rennehan[1], Arif Babul[1], Philip F. Hopkins[2], Romeel Davé[3], and Belaid Moa[4]

Personal contribution clarification: I assembled the methodology necessary to incorporate the sub-grid turbulence model into a Lagrangian hydrodynamical simulation, including the novel derivations. The additions to the simulation code `GIZMO` was written by myself in consultation with Belaid Moa. I wrote the full text and Arif Babul, Romeel Davé, and Belaid Moa provided important comments and suggestions for improvement. Philip F. Hopkins provided insight into the energy conservation in the `GIZMO` code, as he is the authour of that simulation package. I designed, simulated, and reduced all of the data in this paper. I subsequently produced all of the figures.

Abstract: Modelling the turbulent diffusion of thermal energy, momentum, and metals is required in all galaxy evolution simulations due to the ubiquity of turbu-

[1]Department of Physics & Astronomy, University of Victoria, BC V8X 4M6, Canada
[2]TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA
[3]Institute for Astronomy, Royal Observatory, Edinburgh EH9 3HJ, UK
[4]Compute Canada/WestGrid/University Systems, University of Victoria, BC V8P 5C2, Canada

lence in galactic environments. The most commonly employed diffusion model, the Smagorinsky model, is known to be over-diffusive due to its strong dependence on the fluid velocity shear. We present a method for dynamically calculating a more accurate, locally appropriate, turbulent diffusivity: the dynamic localised Smagorinsky model. We investigate a set of standard astrophysically-relevant hydrodynamical tests, and demonstrate that the dynamic model curbs over-diffusion in non-turbulent shear flows and improves the density contrast in our driven turbulence experiments. In galactic discs, we find that the dynamic model maintains the stability of the disc by preventing excessive angular momentum transport, and increases the metal-mixing timescale in the interstellar medium. In both our isolated Milky Way-like galaxies and cosmological simulations, we find that the interstellar and circumgalactic media are particularly sensitive to the treatment of turbulent diffusion. We also examined the global gas enrichment fractions in our cosmological simulations, to gauge the potential effect on the formation sites and population statistics of Population III stars and supermassive black holes, since they are theorised to be sensitive to the metallicity of the gas out of which they form. The dynamic model is, however, not for galaxy evolution studies only. It can be applied to all astrophysical hydrodynamics simulations, including those modelling stellar interiors, planetary formation, and star formation.

2.1.1 Introduction

Galaxies form at the confluence of gas streams and cooling flows at the centres of virialized halos and evolve via a constant exchange of baryons with their environments. Despite significant recent progress on understanding the details of this picture, developing predictive models for the evolution and observed properties of galaxies has proven to be an immense challenge (see, for example, Guedes et al. 2011; Hopkins et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Genel 2016; Davé et al. 2016; – see also Somerville & Davé 2015 and Naab & Ostriker 2017 for a recent review and additional references). The problem lies in the large number of complex interconnected processes involved, and the huge dynamic range in spatio-temporal scales over which they operate.

One important interplay involves the interstellar medium (ISM), the gas that permeates a galaxy and provides fuel for star formation, and circumgalactic medium (CGM), the gas that cocoons the galaxy. The amount of gas in the CGM and the

efficiency with which it can cool, fall into the galaxy and replenish the ISM, are important variables in setting the duration and the rate of star formation (Somerville & Davé, 2015). Stellar winds and supernova explosions (SNe) – processes directly related to star formation – provide competition for gas cooling (Springel & Hernquist, 2003b; Oppenheimer & Davé, 2008). These deposit energy and momentum into the ISM, engendering outflows of gas. If these outflows are sufficiently powerful, they not only heat the CGM, but also expel most of the CGM from the galaxy's halo. This ongoing competition between gas into and out of galaxies provides a basic framework for understanding a number of observed properties (Shen et al., 2012; Crain et al., 2013; Christensen et al., 2016; Oppenheimer et al., 2016; Sokołowska et al., 2016).

The cooling efficiency and ionisation state of the gas in the CGM depend sensitively not only on the spatial injection and redistribution of thermal energy and momentum (Suresh et al., 2017), but also on the metals that are transported from the galaxy via the galactic outflows (Davé et al., 2006; Oppenheimer & Davé, 2006; Finlator & Davé, 2008; Hani et al., 2018). Metals, although negligible in terms of mass fraction, play an out-sized role in galaxy evolution because they can dramatically alter the CGM's cooling profile (van de Voort et al., 2012; Oppenheimer & Schaye, 2013; Sokołowska et al., 2018) and, hence, the delicate balance between gas in- and out- flow.

In addition to cooling, the redistribution of metals can impact other processes, such as the sites and formation history of the putative Population III (Pop III) stars and supermassive black holes (SMBHs). Pop III stars are associated with star formation involving near-pristine gas with an upper limit on metallicities somewhere in the range $[Z] \sim -6$ to $[Z] \sim -3$ (Sarmento et al., 2016). As for SMBHs, a number of authors (e.g. Volonteri 2010 and for use in recent simulations Tremmel et al. 2017) postulate that they form via direct collapse of gas clouds with metallicity $[Z] \lesssim -4.0$, with the idea that very low metallicity would prevent the gas from cooling rapidly and fragmenting into Pop III stars during collapse. It is therefore crucial to identify which physical processes redistribute thermal energy, momentum, and metals in galactic environments spatially, and to include these accurately in the numerical galaxy evolution experiments.

One critical, often overlooked, redistribution mechanism is gas turbulence. Turbulence occurs when inertial forces dominate viscous forces in a gaseous environment, and kinetic energy injected on large scales cannot immediately dissipate as heat. This leads to the formation of a kinetic energy cascade, as coherent turbulent eddies on

large scales spawn eddies on successively smaller scales, until the energy thermalises. Galactic environments, for example, are expected to be highly turbulent (Evoli & Ferrara, 2011; Iapichino et al., 2013). In the case of the CGM, this is strongly suggested by the kinematic complexity revealed by absorption and emission line measurements (Tumlinson et al., 2017) and it has long been recognized that the cold ISM is also highly turbulent. The susceptibility of a medium to turbulence is quantified by its dimensionless Reynolds number[5], Re. The Reynolds number for the cold ISM has been estimated to be as high as $Re \sim 10^7$ (Elmegreen & Scalo, 2004), whereas the onset of turbulence usually occurs at $Re \sim 10^3$. The Reynolds number is also a measure of the separation of scales in the energy cascade and, in the case of incompressible turbulence, $L/\eta \sim Re^{3/4}$, where L is the kinetic energy injection scale, and η is the dissipation scale[6]. Therefore the degrees of freedom in a 3D simulation scales as $Re^{9/4}$, and simulating a $Re \sim 10^7$ flow would require 10^{15} fluid elements! Contemporary cosmological simulations have reached $\sim 10^{12}$ fluid elements and a dynamic range typically of the order $\sim 10^6$ (Somerville & Davé, 2015), with the smallest scale being the resolution limit h. Therefore all cosmological simulations that involve turbulence – independent of hydrodynamical method – have a natural cut-off scale h, in the range $\eta \ll h \ll L$, where discretisation truncates the turbulent cascade.

Physically, small-scale turbulent fluctuations, by promoting mixing, provide a transport mechanism for the fluid properties such as momentum, thermal energy, and metals. In numerical simulations, this implies that turbulence on scales smaller than h can potentially impact the resolved properties of the flow and consequently must be accounted for (Germano et al., 1991). The crux of the issue is that, as stated above, the kinetic energy flux down the turbulent cascade is truncated at h in the simulations whereas, in reality, the kinetic energy cascade should continue to smaller scales until it is dissipated. Numerical simulations break the physical coupling between the scales that are resolved ($> h$) and unresolved ($< h$), and, therefore, they require models of (i) the kinetic energy flux from the resolved to the unresolved scales, (ii) the effect of unresolved eddies on the resolved scales, and (iii) the transport properties of kinetic energy on unresolved scales. Typically, this is done using the *turbulent eddy-diffusion models* that treat sub-grid turbulent eddy motion as a diffusive process. Several implementations of this approach have been proposed (cf. Schmidt & Federrath, 2011; Wadsley et al., 2017; Di Mascio et al., 2017; – see Sagaut, 2006

[5]Defined as the ratio of inertial forces to viscous (dissipative) forces.

[6]In compressible turbulence, the scaling is much steeper (see Kritsuk et al. 2007; Federrath 2013).

and Garnier et al., 2009 for extensive lists); nonetheless, many cosmological hydrodynamical simulation studies assume that numerical diffusion adequately accounts for sub-grid turbulent transport (Schmidt, 2015). Numerical diffusion, however, can lead to diffusive behaviour that poorly represents turbulent flow statistics (Sagaut, 2006).

Extensive effort has been devoted toward developing models for treating turbulent diffusion in Eulerian cosmological simulations, which employ grids to discretise the fluid. Here we refer the reader to Scannapieco & Brüggen (2008); Pan et al. (2013); Federrath (2013); Schmidt et al. (2014); Schmidt (2015); Semenov et al. (2016a); Sarmento et al. (2016). In this paper we focus on the Lagrangian hydrodynamical approach.

Within the Lagrangian framework, the fluid equations of motion are approximated by tracking individual fluid elements as they move with the flow. Commonly used Lagrangian methods in computational cosmology include smoothed particle hydrodynamics (SPH) (Lucy, 1977; Gingold & Monaghan, 1977; Hernquist & Katz, 1989), moving-meshes (MM) (Springel, 2010), and higher-order mesh-free methods (MF) (Lanson & Vila, 2008a,b; Gaburov & Nitadori, 2011; Hopkins, 2015) – see Springel (2010) and Hopkins (2015) for extensive discussion.

SPH has no inherent diffusion and therefore, by construction, explicitly requires additional transport terms, one benefit of which is full control over the strength of mixing. In MM and MF methods, numerical diffusion arises as a by-product of the numerical scheme used to solve the Riemann problem between fluid elements. However, as noted above, inherent numerical diffusion does not, in general, reproduce the correct turbulent properties of a simulated flow. The key problem here is: how can we address this? Or, more precisely, how do we model the interaction between the unresolved and the resolved scales?

A common approach is to assume that the interaction between the resolved and the unresolved scales reduces to a local transfer of kinetic energy from the large to the small scales. If we additionally assume that the kinetic energy transfer mechanism is analogous to a diffusive process, where kinetic energy is redistributed on progressively smaller scales via momentum diffusion, then we can treat the action of unresolved eddies in a similar fashion to molecular viscosity. Under this equivalency, on scales close to the resolution scale h, the unresolved eddies extract kinetic energy from the resolved flow via momentum diffusion, and simultaneously allow for the dissipation of kinetic energy. This is the *eddy-viscosity* hypothesis, and is fundamentally a model assumption that describes the simplest view of the interactions between the scales.

The action of the unresolved eddies can then be modelled by a viscous term in the fluid equations of motion – diffusing momentum and dissipating kinetic energy (Pope, 2000). The above assumptions lead to a simple model but it does not provide guidance about the appropriate choice for the effective diffusivity/viscosity.

In a turbulent cascade, the diffusive action of the eddies depends on their velocity and length scale. On the sub-grid level, this involves estimating the velocity scale that transports fluid properties over the resolution scale h (Wadsley et al., 2008; Greif et al., 2009). In the simplest model, the Smagorinsky model (Smagorinsky, 1963), this velocity is assumed to be proportional to the gradients of the velocity field. In this model, the sub-grid eddies both diffuse (characterised by diffusivity D) and dissipate via eddy viscosity, $\nu_{\rm sgs}$, given by $\nu_{\rm sgs} = D = (C_{\rm s}h)^2\,|S^*|$, where $|S^*|$ is the norm of the trace-free shear tensor, and $C_{\rm s}$ is the Smagorinsky model constant. Contemporary fluid mechanics literature notes that in general $C_{\rm s}$ needs to be tuned to a value between 0.1 and 0.2 for optimal results under different flow conditions (Garnier et al., 2009).

The main advantage of the Smagorinsky model is its simplicity and as a result, many researchers have started to incorporate this model into their cosmological simulation codes – specifically as a model for the diffusivity when treating thermal energy and metal mixing (Shen et al., 2010; Brook et al., 2014; Williamson et al., 2016; Tremmel et al., 2017; Sokołowska et al., 2018; Escala et al., 2018). The model, however, has some drawbacks: (i) A single valued C_s is incapable of correctly describing different types of turbulent flows. Studies show that the Smagorinsky model introduces too much diffusion into the flow in almost all cases except for homogeneous, isotropic turbulence (Garnier et al., 2009). And, (ii) the sub-grid eddy viscosity does not vanish for laminar shear flows where there ought to be no diffusion due to turbulence. Over-diffusion is especially worrisome given the push to resolve the multiphase structure in the ISM and CGM at greater levels, and the recent results that differential, localised metal mixing could change our understanding of galactic chemical evolution (Emerick et al., 2018).

As noted above, most implementations of the Smagorinsky model only consider thermal energy and metal diffusion, but not momentum diffusion because the latter drawback above is a concern for differentially rotating structures, such as galactic discs. Specifically, it results in undesired angular momentum transport and, consequently, unphysical flows in the discs. Most galaxy evolution studies ignore momentum diffusion in order to avoid this viscous instability. The crux of the problem is

that a constant C_s cannot automatically adjust to either the local conditions or the changing character of the flow with time. This led Germano et al. (1991) to propose the dynamic Smagorinsky model where C_s is a function of space and time, i.e. $C_s = C_s(\boldsymbol{x}, t)$.

The dynamic Smagorinsky model has, to our knowledge, not yet been implemented and investigated in cosmological simulations. The model has, however, been validated extensively in the fluid simulation community by comparing to the results of standard numerical tests and experiment data, and has been shown to improve upon the constant-coefficient model (Kleissl et al., 2006; Kirkpatrick et al., 2006; Benhamadouche et al., 2017; Lee & Cant, 2017; Kara & ÇaÄ§lar, 2018; Taghinia et al., 2018). In the dynamic model, the sub-grid properties of a turbulent fluid are computed under two assumptions[7]: (i) the behaviour of the largest *unresolved* eddies is entirely determined by their interactions with the eddies on the smallest *resolved* scales, and (ii) these interactions are analogous to those between the fluid motions on the smallest resolved scale and the motions on larger scales. In practice, determining the characteristics of these interactions involves filtering (or smoothing) the resolved velocity field on two different scales. When the sub-grid turbulent properties are calculated based on the local fluid properties, $C_s(\boldsymbol{x}, t)$ reduces to zero (i.e. the eddy viscosity/diffusivity vanishes) in non-turbulent (or laminar) shear flows (Piomelli & Liu, 1995). Consequently, this allows for the self-consistent treatment of momentum diffusion, along with thermal energy and metal diffusion, in numerical studies of cosmic baryons and galaxy evolution.

In this study, we introduce an implementation of the dynamic Smagorinsky model for the first time in Lagrangian astrophysical simulations, focusing on the higher order MF approach. We discuss the implementation details in Section 2.1.2. In order to test the impact of dynamic localised turbulent mixing on the galactic ecosystem we run a series of hydrodynamical and physical experiments relevant to galaxy evolution. In Section 2.1.3 we show the results for a set of standard hydrodynamic tests and explicitly check on the extent of sub-grid diffusion in laminar shear flows. We then go on to examine the effects of dynamic mixing in an isolated disc galaxy in Section 2.1.4, followed by a set of cosmological simulations in Section 2.1.5. We investigated SPH and performed all of the experiments presented below, but we do not include them in this paper because the results are qualitatively similar to the MF results, as

[7]The two assumptions are often combined together and referred to as the scale-similarity hypothesis.

we mention in Section 2.1.6.

2.1.2 Methods

Hydrodynamics

In order to test the impact and robustness of the localised turbulent diffusion model, we employ a modified version of the GIZMO gravity plus hydrodynamics solver code (Hopkins, 2015). GIZMO builds on the GADGET-3 code base (Springel, 2005), with improvements in numerical accuracy and includes an implementation of the novel mesh-free finite mass (MFM) method, in addition to various implementations of smoothed particle hydrodynamics (SPH) methods.

The MFM method evolves the fluid equations of motion in a Lagrangian manner similar to SPH. However, while the fluid mass elements in SPH are discretised into particles and their motions are determined by fluid properties smoothed over neighbouring particles, the conservation laws in MFM are evolved by calculating the flux of basic variables between neighbouring particles[8]. These fluxes depend on the effective face area between the two particles and are determined by solving the Riemann problem along the line connecting them. This removes the need for additional terms, such as artificial viscosity and conductivity as is necessary in SPH to ensure proper treatment of shocks. The numerical Riemann solvers have inherent numerical dissipation[9], hence improved shock capturing capabilities. We, therefore, focus on the use of the MFM method in our investigation of the dynamic diffusion model. For a thorough exposition of GIZMO and an extensive comparison between MFM and SPH, see Hopkins (2015).

Sub-grid turbulent diffusion terms

GIZMO solves the conservation equations for momentum, energy, and mass using the MFM method, and like other hydrodynamic methods, it is limited in resolution down to a scale h. The minimum resolution limits the ability to resolve high Reynolds number flows down to the viscous dissipation scale, impacting its ability to resolve the turbulent cascade. As mentioned in Section 2.1.1, the interaction of the resolved

[8]We refer to any fluid element as a particle, for simplicity. Fluid elements in the MFM method are not particles in an SPH sense, and are defined by the effective geometrical faces moving along lines connecting cells enclosing a finite mass.

[9]Riemann solver dissipation arises from the high, even-order, truncated terms in the Taylor series expansion of the basic variables in the conservation equations.

and unresolved scales must be modelled. These models apply to the additional sub-grid scale terms that appear in the equations of motion when treating discretisation as a filtering process (Sagaut, 2006; Garnier et al., 2009; Schmidt, 2015). In this section, we detail the origin of the sub-grid scale terms and discuss which terms we include in the `GIZMO` code.

Discretisation of the conservation equations is equivalent to applying a low-pass filter, damping out high frequency turbulent fluctuations. When we discuss filtering, we refer to the definition of a general filtered scalar field $f(\boldsymbol{x})$,

$$\overline{f}(\boldsymbol{x}) \equiv \int_D f(\boldsymbol{x}')G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h})\boldsymbol{dx}', \tag{2.1}$$

where $G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h})$ is a filter function. $\overline{h}$ is the characteristic size of the filtering operation below which fluctuations are damped (essentially the resolution scale, in the present context), $\boldsymbol{dx}'$ is a volume element, and the integral is evaluated over the entire domain. We discuss our filtering implementation in more detail, in Section 2.1.2. We now apply this equation to the conservation equations in order to see that additional terms in the hydrodynamical equations are required.

The momentum conservation equation for a compressible fluid follows[10] (Landau & Lifshitz, 1987),

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}[\rho u_i u_j + p\delta_{ij}] = 0, \tag{2.2}$$

where $u_i = u_i(\boldsymbol{x}, t)$ is the fluid velocity vector in the $i = \{x, y, z\}$ direction, $\rho = \rho(\boldsymbol{x}, t)$ is the fluid density, and $p = p(\boldsymbol{x}, t)$ is the pressure. When we filter the momentum equation, assuming the filtering operation and derivatives commute, we end up with an extra term τ_{ij},

$$\frac{\partial}{\partial t}(\overline{\rho}\tilde{u}_i) + \frac{\partial}{\partial x_j}[\tau_{ij} + \overline{p}\delta_{ij} + \overline{\rho}\tilde{u}_i\tilde{u}_j] = 0, \tag{2.3}$$

where we have also switched to density weighted variables such that $\tilde{u}_i = \overline{\rho u_i}/\overline{\rho}$. τ_{ij} is the *subgrid-scale turbulent stress tensor* or *residual stress tensor* and is defined as,

$$\tau_{ij} \equiv \overline{\rho}(\widetilde{u_i u_j} - \tilde{u}_i\tilde{u}_j). \tag{2.4}$$

This term must be modelled because $\widetilde{u_i u_j}$ is unknown at the time of simulation, i.e.,

[10] We follow this Einstein notation throughout this paper.

the system of equations is not closed. A common model, or closure, involves the eddy-viscosity assumption where the sub-grid scales impart a momentum flux on the resolved scales that is linearly dependent on the rate of strain of the resolved scale,

$$\tau_{ij} = -2\overline{\rho}\nu_{\text{sgs}}\widetilde{S^*}_{ij}, \tag{2.5}$$

where ν_{sgs} is the sub-grid eddy viscosity, and $\widetilde{S^*}_{ij}$ is the trace-free resolved rate of strain tensor. The same logic can be applied to any of the conservation equations and any filtered multiplicative terms require modelling. In the derivations below, we model (as in equation 2.5 above) the unknown terms under the assumption that they act as diffusive processes.

There is a similar term when filtering the total energy equation (e is the specific total energy),

$$\frac{\partial}{\partial t}(\rho e) + \frac{\partial}{\partial x_j}[\rho u_j e + u_j p] = 0. \tag{2.6}$$

leads to additional terms,

$$\frac{\partial}{\partial t}(\overline{\rho}\tilde{e}) + \frac{\partial}{\partial x_j}[Q_j + P_j + \overline{\rho}\tilde{u}_j\tilde{e} + \overline{u}_j\overline{p}] = 0, \tag{2.7}$$

where $e = \theta + \frac{1}{2}|\boldsymbol{u}|^2$ (θ is the specific internal energy). We also have defined:

$$Q_j \equiv \overline{\rho}(\widetilde{u_j e} - \tilde{u}_j\tilde{e}), P_j \equiv \overline{u_j p} - \overline{u}_j\overline{p}. \tag{2.8}$$

In this study, we ignore the term associated with pressure, P_j, and the term in Q_j that arises from $\frac{1}{2}|\boldsymbol{u}|^2$ in e, and focus on the application of the dynamic model to the sub-grid momentum term working in concert with the widely employed thermal energy term (Shen et al., 2013; Brook et al., 2014; Tremmel et al., 2017; Wadsley et al., 2017),

$$q_j \equiv \overline{\rho}(\widetilde{u_j\theta} - \tilde{u}_j\tilde{\theta}) = -\overline{\rho}\nu_{\text{sgs}}\frac{\partial\tilde{\theta}}{\partial x_j}. \tag{2.9}$$

Not only are there additional terms for momentum and thermal energy, but any scalar quantities, such as the concentration of different metal species, in the gas should be transported in a turbulent flow. In order to model this, we treat metal concentrations, ϕ_z, as passive scalars that obey a diffusion equation (Pope, 2000; Shen et al., 2010),

$$\frac{\partial \phi_z}{\partial t} = \frac{\partial}{\partial x_j}\left(\overline{\rho}\nu_{\rm sgs}\frac{\partial \phi_z}{\partial x_j}\right). \tag{2.10}$$

For a detailed description of the incorporation of these fluxes into `GIZMO`, see Hopkins (2017).

Throughout this paper, we refer to the action of the terms in equations (2.5), (2.9), and (2.10) as *turbulent diffusion* because they contribute to the conservation equations as $\nabla^2 f$, where f is the flux quantity. Additionally, when we mention energy diffusion, we are referring to the term $\boldsymbol{\nabla}\cdot\boldsymbol{q}$ (thermal energy diffusion), and, similarly, when we mention momentum diffusion we are referring to the action of the stress tensor through $\boldsymbol{\nabla}\cdot\boldsymbol{\tau}$, along with the corresponding kinetic energy dissipation.

Diffusivity

Physically, turbulent mixing can be modelled as a diffusive process with diffusivity D and, in the simplest model, the fluid properties are assumed to mix over the resolution scale h with a velocity $h|S^*|$, where $|S^*|$ is the norm of the trace-free shear tensor. This is, as we mentioned in the introduction, the Smagorinsky model and the corresponding diffusivity is parametrised as $D = \nu_{\rm sgs} = (C_{\rm s}h)^2|S^*|$.

The Smagorinsky model inherently assumes that the kinetic energy transfer rate down the turbulent cascade is equal on all scales, and is equal to the physical dissipation rate (i.e. the flow is in local equilibrium). In simulations, the resolution scale h inhibits kinetic energy from moving to progressively smaller scales, and results in a build-up of kinetic energy at the resolution scale – so long as numerical dissipation cannot extract kinetic energy sufficiently rapidly. The Smagorinsky model combined with the local equilibrium assumption only provides a model for the turbulent stress, τ_{ij}, and dissipation in the flow, $\Sigma = \tau_{ij}\tilde{S}^*_{ij}$, and ignores the additional terms we discuss in Section 2.1.2. In order to consistently model all of the energetic terms, one must relax the local equilibrium assumption and follow the sub-grid kinetic energy, K, directly. It is possible to derive a one-equation model for K that includes a third-order term for the transport of K on sub-grid scales (Schmidt, 2015), and self-consistently follows all of the sub-grid quantities. Each sub-grid term can then be modelled using diffusive terms similar to the models in Section 2.1.2, in order to close the system of equations.

If the Smagorinsky model only considers the turbulent stress, is it then valid to apply this model (as we have done following Shen et al. 2010) to the thermal energy

and metal sub-grid terms? In order for this to be possible, the local equilibrium condition must be approximately true in the regime of interest. We are specifically interested in cosmological-scale gas, and Schmidt et al. (2016) show that the local equilibrium condition holds – on average – in a cosmological-scale volume. Introducing the dynamic model on top of these approximations further supports our model assumption, because the dynamic model inherently accounts for the deviations from local equilibrium.

The Smagorinsky model diffusivity is parametrised in `GIZMO` for a particle a as,

$$D_a = \rho_a (C_s h_a)^2 |S^*|_a, \tag{2.11}$$

where C_s is the Smagorinsky constant, h_a is the mean inter-particle spacing in the kernel, and $|S^*|_a$ is the magnitude of the trace-free symmetric shear tensor. Note that we absorb the densities $\overline{\rho}$ from Section 2.1.2 into ν_{sgs} via ρ_a. D_a is used in the diffusion equations for thermal energy, momentum, and metal mass fractions as described in Hopkins (2017). There are a myriad of values quoted for C_s in the literature (see Section 5 of Sagaut (2006) for an extensive list), but we choose the value calibrated for fully-developed isotropic turbulence, $C_s = 0.2$ (Clark et al., 1979), because the Smagorinsky model was developed for this specific regime.

In order to compare our C_s with other values in the literature, it is important to consider the definitions of the quantities in equation (2.11). In many SPH studies, the length scale in the diffusion coefficient definition is taken as the kernel-support radius, h_{SPH}, which is the maximum extent from a particle that gives a non-zero weight. In contrast, as we mentioned above, we employ the mean inter-particle spacing within the kernel or $h_a \approx 0.5 h_{SPH}$. Additionally, some studies use $C \equiv \sqrt{2} {C_s}^2$ whereas for this study $\sqrt{2}$ is absorbed into our definition of the norm of the shear tensor (see equation 2.13 below), closely following the fluid simulation literature (Piomelli & Liu, 1995). Using these definitions, our adopted value of C_s is lower than those quoted in the astrophysics literature. For example, the value $C = 0.05$ (see Shen et al. 2010, 2013; Brook et al. 2014), corresponds to $C_s = 0.37$ whereas $C = 0.03$ corresponds to $C_s = 0.29$ (see Wadsley et al. 2017). However, our C_s is higher than the value recently calibrating from studying metal mixing in dwarf galaxies, where Escala et al. (2018) found $C_s = 0.046$ reproduced more realistic stellar metal distribution functions via supersonic mixing in the interstellar medium.

We compute the trace-free symmetric shear tensor via the high-order accurate

gradient estimators in GIZMO,

$$S^*_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) - \frac{1}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}, \tag{2.12}$$

where u_i is the fluid velocity vector and x_i the spatial coordinate, and $i, j = \{x, y, z\}$. The magnitude of equation (2.12) is implemented using the Frobenius norm (Piomelli & Liu, 1995),

$$|S^*| = \sqrt{2S^*_{ij}S^*_{ij}}. \tag{2.13}$$

We note that in GIZMO, Escala et al. (2018) impose an ad hoc cap on the diffusivity based on the expected maximum mass flux between resolution elements. The cap does not, however, mitigate the fact that the constant Smagorinsky model induces diffusion whenever there is shear, regardless of whether the fluid is laminar or turbulent. It is only intended to prevent unphysical mass/energy transport that can potentially arise due to the excessive diffusivity of the Smagorinsky model. We adopt the same limiter in this study, but note that the diffusivity rarely reaches the maximum limit.

Dynamic Model

The Smagorinsky model provides an approximate model of sub-grid mixing for fully developed, homogeneous turbulence but it is far too simple for complex flows. In fact, in laminar shear flows, the constant-coefficient Smagorinsky model predicts a non-zero diffusivity through its dependence on the shear strength (equation 2.11). However, in this situation, the diffusivity should be zero since the fluid is not turbulent. In more complex flows, such as those in astrophysical contexts, the value of the constant ought to depend on the spatio-temporal coordinates $C_s = C_s(\boldsymbol{x}, t)$ (Germano et al., 1991). Piomelli & Liu (1995) showed that by assuming scale-similarity (cf. Section 2.1.1), the local Smagorinsky constant in a neighbourhood can be calculated at each point (at a fixed simulation timestep) as follows:

$$C_{\rm dyn}(\boldsymbol{x}) = C_s^2 = -\frac{1}{2}\frac{(L_{ij} - 2\widehat{C^p_{\rm dyn}\beta_{ij}})\alpha_{ij}}{\alpha_{mn}\alpha_{mn}}. \tag{2.14}$$

Here L_{ij} is the Leonard tensor,

$$L_{ij} = \widehat{\overline{u}_i\overline{u}_j} - \widehat{\overline{u}}_i\widehat{\overline{u}}_j, \tag{2.15}$$

C^p_{dyn} is the value of C_{dyn} at the previous timestep, and α_{ij} and β_{ij} are defined as,

$$\begin{aligned} \alpha_{ij} &\equiv \widehat{h}^2 |\widehat{\overline{S^*}}| \widehat{\overline{S^*}}_{ij}, \\ \beta_{ij} &\equiv \overline{h}^2 |\overline{S^*}| \overline{S^*}_{ij}. \end{aligned} \tag{2.16}$$

Here $\overline{f}$ represents a filtering (or smoothing) operation on f over a length-scale $\overline{h}$, and $\widehat{f}$ represents smoothing on a scale $\widehat{h}$. $\overline{h}$ usually is equated with the lowest resolvable scale, and extensive work has been done in the fluid simulation community to show that the optimal value for $\widehat{h}$ is $\widehat{h} = 2\overline{h}$ (Germano et al., 1991; Piomelli & Liu, 1995; Spyropoulos & Blaisdell, 1996; Schmidt et al., 2006; Grete et al., 2018). For this work, we choose $\overline{h}$ to be the compact support radius of the kernel, and $\widehat{h} = 2\overline{h}$.

More precisely, we smooth a scalar field $f(\boldsymbol{x})$ by convolving it with the filter function $G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h})$ over the domain[11],

$$\overline{f}(\boldsymbol{x}) = \int_D f(\boldsymbol{x}') G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h}) \boldsymbol{dx}'. \tag{2.17}$$

This is similar to the SPH method of interpolating a scalar function, with $G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h})$ sharing the same properties as the smoothing kernel $W(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h})$,

$$\begin{aligned} \int G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h}) \boldsymbol{dx}' &= 1, \\ \lim_{\overline{h} \to 0} G(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h}) &= \delta(|\boldsymbol{x}' - \boldsymbol{x}|). \end{aligned} \tag{2.18}$$

The MFM method employs a similar technique for evaluating integrals and in order to be consistent, we choose $G = W$. The integral in equation (2.17) is expensive, but we simplify the computation using XSPH smoothing (Monaghan, 1989, 2005, 2011),

$$\overline{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \epsilon \int_D (f(\boldsymbol{x}') - f(\boldsymbol{x})) W(|\boldsymbol{x}' - \boldsymbol{x}|, \overline{h}) \boldsymbol{dx}'. \tag{2.19}$$

Following Monaghan (2011), the Fourier coefficients a_{k} of the velocity satisfy $\overline{a}_{\text{k}} = a_{\text{k}}[1 + \epsilon(\widetilde{G}(k) - 1)]$ where $\overline{a}_{\text{k}}$ are the coefficients of the smoothed field, $\widetilde{G}(k)$ is the Fourier transform of the filter function, and k is the spatial frequency $k = 2\pi/x$. In the limit $k \to \infty$, the coefficients satisfy $\overline{a}_{\text{k}} \to (1 - \epsilon) a_{\text{k}}$. The value of ϵ controls the

[11] Our filtering implementation naturally density-weights quantities because we follow the hydrodynamical weighting scheme.

magnitude of the smoothing on a scale of $\leq \overline{h}$, and is constrained to $0 < \epsilon \leq 1$. We choose $\epsilon = 0.8$ to be consistent with the tests in Monaghan (2011).

We discretise equation (2.19) as,

$$\overline{f}_a = f_a + \epsilon \sum_b \frac{m_b}{\langle \rho_{ab} \rangle_{\overline{h}}} (f_b - f_a) W(|\boldsymbol{x_a} - \boldsymbol{x_b}|, \overline{h}_{ab}), \tag{2.20}$$

where f_a represents the quantity at particle a, $\overline{h}_{ab}$ is the arithmetic mean of $\overline{h}_a$ and $\overline{h}_b$[12], $\langle \rho_{ab} \rangle_{\overline{h}}$ is the harmonic mean of the densities $\overline{\rho}_a$ and $\overline{\rho}_b$[13], and the sum is taken over b nearest neighbours. We also require doubly-filtered quantities, which involves another application of equation (2.20) to the singly-filtered quantities,

$$\widehat{\overline{f}}_a = \overline{f}_a + \epsilon \sum_b \frac{m_b}{\langle \rho_{ab} \rangle_{\widehat{h}}} (\overline{f}_b - \overline{f}_a) W(|\boldsymbol{x_a} - \boldsymbol{x_b}|, \widehat{h}_{ab}). \tag{2.21}$$

In order to calculate the average densities, or weights, in equations (2.20) and (2.21), we require the density at each particle for a given scale,

$$\begin{aligned} \overline{\rho}_a &= \sum_b m_b W(|\boldsymbol{x}_a - \boldsymbol{x}_b|, \overline{h}_a), \\ \widehat{\rho}_a &= \sum_b m_b W(|\boldsymbol{x}_a - \boldsymbol{x}_b|, \widehat{h}_a). \end{aligned} \tag{2.22}$$

It is important to note that the values of $\overline{S^*}_{ij}$ and $\widehat{\overline{S^*}}_{ij}$ are built from the smoothed velocity field and are not smoothed versions of the trace-free symmetric shear tensor in equation (2.12) (Schmidt, 2015). They have the following corresponding equations,

$$\begin{aligned} \overline{S^*}_{ij} &= \frac{1}{2}\left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i}\right) - \frac{1}{3}\delta_{ij}\frac{\partial \overline{u}_k}{\partial x_k}, \\ \widehat{\overline{S^*}}_{ij} &= \frac{1}{2}\left(\frac{\partial \widehat{\overline{u}}_i}{\partial x_j} + \frac{\partial \widehat{\overline{u}}_j}{\partial x_i}\right) - \frac{1}{3}\delta_{ij}\frac{\partial \widehat{\overline{u}}_k}{\partial x_k}. \end{aligned} \tag{2.23}$$

The magnitudes $|\overline{S^*}|$ and $|\widehat{\overline{S^*}}|$ are given via equation 2.13.

The dynamic method relies on local scale-similarity in the neighbourhood of a

[12]Taken as $\overline{h}_{ab} = \frac{1}{2}(\overline{h}_a + \overline{h}_b)$ in order to equally weight each smoothing scale.

[13]The harmonic mean is of the form $\langle \rho_{ab} \rangle_{\overline{h}} = 2\overline{\rho}_a\overline{\rho}_b/(\overline{\rho}_a + \overline{\rho}_b)$ and weights toward the lowest value. This allows high density particles to have a fair contribution to the differences within the kernel.

Name	Dynamic	Thermal Energy	Velocity	Metals
None	N/A	×	×	×
S-uz	×	✓	×	✓
D-uz	✓	✓	×	✓
S-uvz	×	✓	✓	✓
D-uvz	✓	✓	✓	✓

Table 2.1: We compare five mixing models involving combinations of the standard implementation, dynamic implementation, as well as the mixing of energy, velocity, and metals. We prefix models using the standard implementation by S-, and those involving the dynamic model by D-. Models which mix thermal energy, velocity (momentum), or metals have combinations of the suffixes u, v, or z, respectively.

point $\boldsymbol{x}$, which in turn implies that the Smagorinsky model is an accurate description of the flow, albeit with a variable constant. This assumption breaks down in highly complex flows and in some cases, the dynamic model predicts negative values (Piomelli & Liu, 1995; Urzay et al., 2013). Negative values of $C_{\rm dyn}$ are usually explained as *backscatter* in the cascade (Piomelli et al., 1991; Piomelli & Liu, 1995; Meneveau & Katz, 2000; Vreman, 2004; Urzay et al., 2013) where in some circumstances a fraction of the energy cascading to small scales can return to large scales as smaller eddies unite to form larger eddies. An alternate explanation for $C_{\rm s} < 0$ is that the Smagorinsky model fails, and a more appropriate model should be employed. For now, we adopt the latter view and follow the usual approach in restricting $C_{\rm dyn}$ to the range $C_{\rm dyn} \in [0, C_s]$ with values $C_{\rm dyn} < 0$ set to $C_{\rm dyn} = 0$ (Garnier et al., 2009; Schmidt, 2015). The upper limit is imposed since large values are thought to be due to numerical instability. For the remainder of this paper, we identify and discuss $C_{\rm s} = \sqrt{C_{\rm dyn}}$. The distinction between the simple and dynamic model is distinguishable based on context.

2.1.3 Hydrodynamical tests

As indicated previously, a number of studies have carried out robust validation of the dynamic Smagorinsky model against, for example, experimental results within the fluid mechanics community (Kleissl et al., 2006; Benhamadouche et al., 2017; Lee & Cant, 2017; Kara & ÇaÄ§lar, 2018; Taghinia et al., 2018) and the model has also been adopted by other users, including researchers studying atmospheric phenomena (e.g. Kirkpatrick et al. 2006). To motivate its use in cosmological and astrophysical simulations, we start by discussing the model within the context of three

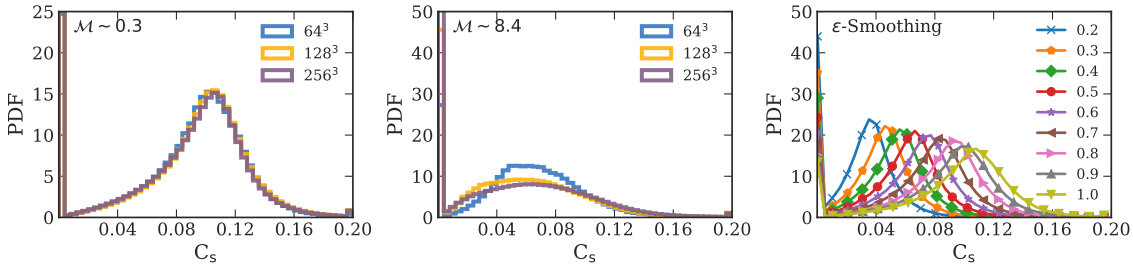

Figure 2.1: (*left*) Probability density function of C_s, calculated with the dynamic model, in homogeneous subsonic turbulence at three resolutions – 64^3, 128^3, and 256^3 particles. The MFM method with a cubic spline kernel is employed, with $N_{ngb} = 32$. Median values of the predicted Smagorinsky constant are $C_{s,64} = 0.1005$, $C_{s,128} = 0.1017$, $C_{s,256} = 0.1009$ and are well within 2% at maximum difference. (*middle*) Supersonic, homogeneous turbulence at three resolutions, with $C_{s,64} = 0.0668$, $C_{s,128} = 0.0632$, $C_{s,256} = 0.0679$. (*right*) Probability density function of C_s as we vary the smoothing parameter, ϵ, in homogeneous, subsonic turbulence using the MFM method, with a quintic spline kernel and $N_{ngb} = 64$.

hydrodynamical tests.

First, we investigate the distribution of predicted C_s values in homogeneous driven turbulence, and the sensitivity of the distributions to variations in the smoothing parameter, ϵ. Next, we examine the evolution of a Keplerian disc where turbulence is not expected to develop *a priori* yet numerical instabilities and noise lead to disorder in the velocity fields, and subsequent artificially enhanced diffusivities via the trace-free shear strength, $|S^*|$. Last, we consider the linear regime of the Kelvin-Helmholtz instability which suffers from similar challenges as the Keplerian disc. In both cases, numerical instability causes the constant-coefficient Smagorinsky model to fail, and we investigate whether the dynamic model can mitigate spurious sub-grid turbulent mixing.

Homogeneous turbulence

We investigate homogeneous, isotropic, driven turbulence to determine to what degree the dynamic model predicts different diffusivities on a per particle basis.

We initialise periodic boxes of side length $L = 1$ with 64^3, 128^3, and 256^3 equal mass particles of an ideal isothermal gas, initial density $\rho = 1$, and energy per unit mass $u = 1000$, placed on uniform Cartesian grids[14]. Following the methods of Bauer

[14]Units are arbitrary code units.

& Springel (2012) and Hopkins (2015), as the system is driven, the thermal energy of the gas particles are reset to the initial value in order to simulate isothermal turbulence. We investigate the five combinations of mixing models described in Table 2.1, and examine subsonic ($\mathcal{M} \approx 0.3$) and supersonic ($\mathcal{M} \approx 8.4$) test cases.

In order to mix the fluid over time, we use an identical forcing routine as in Bauer & Springel (2012). The accelerations are calculated in Fourier space and only contain power over a small range of modes corresponding to a spatial range $\ell \in [L/2, L]$ (i.e., the largest scales), and the Fourier mode phases are drawn from an Ornstein-Uhlenbeck process. In the subsonic case, the forcing is purely solenoidal (or incompressible) since the compressive part of the acceleration is removed via a Helmholtz decomposition in Fourier space. It is important to note that Grete et al. (2018) showed that this is not completely correct, and that compressive modes still exist even with purely solenoidal forcing. However, we are comparing the effects of the dynamic model using the same forcing methodology across all of our test cases, and additionally we construct the shear and sub-grid scale stress tensor to be trace-free, removing any contributions from compression of the fluid. We use the exact parameters in Table 1 of Bauer & Springel (2012), and point the interested reader to their Section 2.2 for the precise details of the driving routine. The systems enter an approximate steady state after $t \gtrsim 5$. We measure the probability density functions (PDFs) of C_s in each test in order to determine its sensitivity to the smoothing parameter ϵ in equation (2.20). In addition, we measure the PDF of the density field in order to gauge the ability of each model to resolve different density ranges in the turbulent flow.

The left panel in Fig. 2.1 shows the distribution of C_s in the subsonic case as predicted with the dynamic model for three separate resolutions: 64^3, 128^3, and 256^3. The median value and the shape of the distributions do not change much with resolution, indicating excellent convergence. At 64^3 resolution we find a median value $C_s = 0.1$ and approximately 9.61% of the particles have $C_s = 0$. In the fluid mechanics literature, as many as 50% of the fluid elements have been reported to have $C_s = 0$ (Piomelli et al., 1991; Urzay et al., 2013). We also test for convergence in supersonic turbulence (see the middle plot in Fig. 2.1). Compared to the subsonic case, the dynamic model predicts more particles at $C_s = 0$, with a total fraction below 50%. The median value is much lower than in the subsonic case, $C_s = 0.066$. The lower median agrees with calibration results from Colbrook et al. (2017) who found that $C_s \approx 0.05$ reproduces the turbulent scaling relationships in supersonic turbulence.

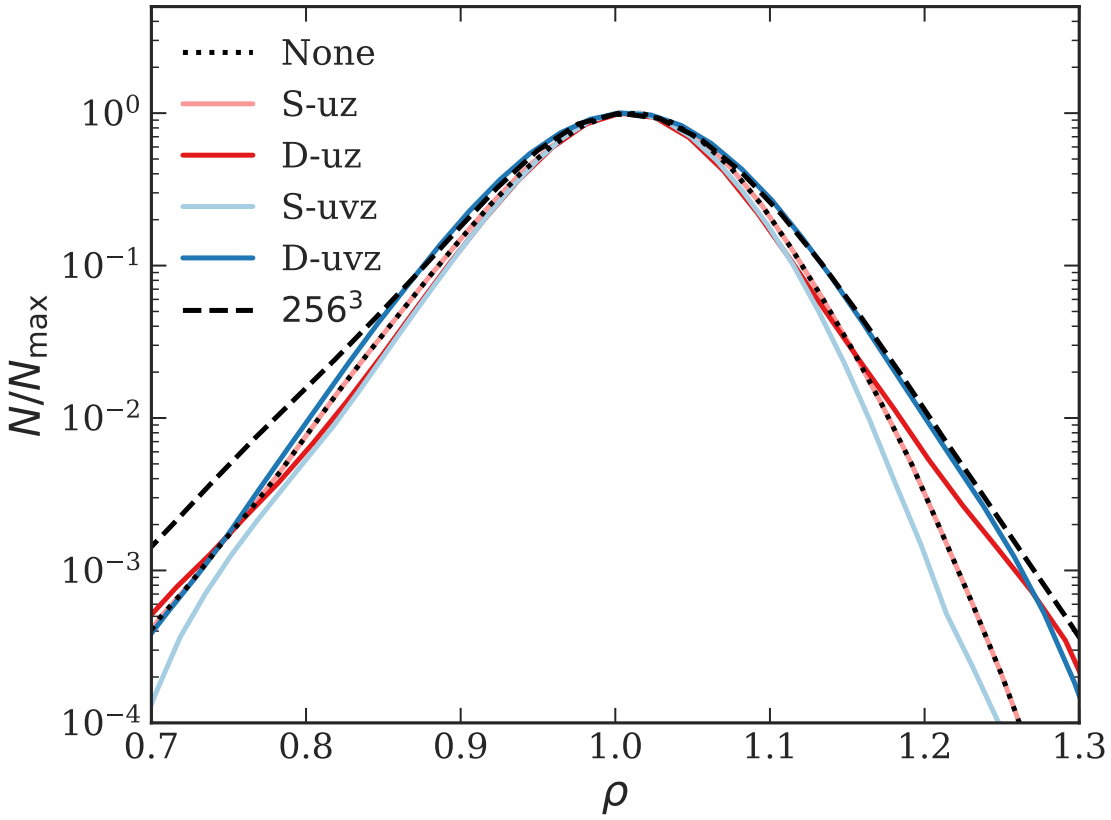

Figure 2.2: Histograms of the density field in homogeneous subsonic turbulence (64^3 case), normalised to the bin with the maximum particle count. Density measurements taken from 150 snapshots between $t = 10$ and $t = 25$. Here, the `None` and `S-uz` cases coincide. The particle density contrast in the `S-uvz` case is much tighter than the other cases with particles closer to the mean density ($\rho = 1$). The `D-uvz` case is able to represent the widest range of densities, including much higher density regions, with the same number of fluid elements.

We also investigate the sensitivity of the C_s to the smoothing parameter, ϵ, from equation (2.20). The right plot in Fig. 2.1 shows the PDFs of C_s in homogeneous, subsonic turbulence using the MFM method. We vary ϵ between 0.2 and 1.0 since ϵ cannot be greater than 1.0, as it is derived from a series expansion, and should be $\geqslant 0$ in order to have positive kinetic energies in the smoothed fields (Monaghan, 2011). Values in the range $0.7 \leqslant \epsilon < 1.0$ produce comparable distributions with medians $C_s \approx 0.1$. Monaghan (2011) found, using a version of SPH with smoothed velocities, that $\epsilon = 0.8$ reproduced turbulent flow trends in decaying wall-bounded turbulence. For this reason we employ $\epsilon = 0.8$ in all of our tests.

Turning to the gas properties, Fig. 2.2 shows histograms of the gas densities in the 64^3 homogeneous subsonic turbulence simulations, averaged at 150 equally spaced times from $t = 10$ to $t = 25$ (inclusive), normalised such that each maximum is $N_{\max} = 1$. After $t = 10$, each turbulent field across all models is in an approximate statistically stationary state with $\overline{\rho} \approx 1$. Although the model labels include the `-z` flag, we do not include metal mixing in our driven turbulence simulations. We retain the suffix for cross-comparison across the different cases in paper, and point the reader to Section 2.1.3 for a discussion of turbulent metal mixing in an idealised experiment. We note that the `None` and `S-uz` cases coincide and the lines in the figure overlap, indicating that internal energy diffusion with a global C_s value has no effect on the density distribution.

First, using the `None` case as a reference curve, the `D-uz` case shows a narrower distribution between $0.75 < \rho < 1.15$ with a prominent extended tail toward higher densities. This indicates that the `D-uz` case can represent a wider range of densities in the turbulent cascade. Since the `S-uz` case follows the `None` case exactly, this suggests that the localisation of the diffusivity D impacts the density resolution much more than the dependence on velocity shear.

Introducing momentum diffusion alters the density distributions significantly compared to the `None` case. In the `S-uvz` case, the density distribution is tighter and exhibits no apparent wings, with the majority of densities falling in the range $0.75 < \rho < 1.2$. Here the increased diffusivity destroys any small scale structure by causing densities to remain closer to the mean. However, when we employ the `D-uvz` model, we find the opposite effect – localising momentum diffusion leads to a wider range of densities in the turbulent gas compared to the all other cases. The effect is strongest at higher densities and therefore we can conclude that the dynamic model is able to resolve higher densities in a turbulent flow at the same mass resolution.

Keplerian disc

In numerical studies of galaxy formation, inherent or artificial dissipation in the hydrodynamical method can cause gas to lose angular momentum and flow radially inward (Hosono et al., 2016), i.e. the viscous instability. Numerical simulations require sub-grid diffusive terms as they cannot resolve the viscous scale, however, additional momentum diffusion enhances the viscous instability. This is an important consideration for the constant-coefficient Smagorinsky model: in a simulation of a gaseous

disc where the rotational velocity curve depends on the radius, including momentum diffusion will trigger the viscous instability even if there is no turbulence as the radial velocity gradient contributes to $|S^*|$. A quick analytic calculation demonstrates this. Let us consider a 2D idealised rotating gaseous disc with constant surface density that follows a Keplerian velocity profile $v_\phi \propto r^{-1/2}$. For this disc, $|S^*| \propto r^{-3/2}$ and inserting this into equation (2.11) with $h = \text{const.}$, we find $D \propto r^{-3/2}$. Generally, for any non-constant velocity profile $v_\phi = v_\phi(r)$, $D \propto \partial_r v_\phi$ in the constant density case.

In a Keplerian disc, particles near the inner radii of a rotating disc will diffuse the strongest in the standard Smagorinsky model as the difference in velocity between each concentric ring is much higher in this region; leading to the rapid break-up of the disc. One could mitigate the over-diffusion by using a smaller value of C_s in equation (2.11) but then the model would lose its advantages in turbulent flows. The dynamic model provides a solution to this problem.

We use the 2D idealized Keplerian disc as a representative case of an astrophysical laminar shearing flow to illustrate the aforementioned over-diffusion and compare to the results of the dynamic model. We simulate a gas annulus of constant surface density using the MFM method, with particles initialised on circular orbits about the centre. We smooth the inner and outer edges of the annulus in order to suppress numerical instabilities that occur at sharp boundaries. The particles are subject to an external softened gravitational acceleration ($\boldsymbol{a} = -\boldsymbol{r}(r^2 + \epsilon^2)^{-3/2}$) directed towards the centre of the annulus, and follow a corresponding Keplerian velocity profile. This initial condition is identical to that in Section 4.2.4 of Hopkins (2015), with surface density as a function of radial distance, r,

$$\Sigma(r) = \begin{cases} (2r)^3 & r < 0.5, \\ 1 & 0.5 \leqslant r \leqslant 2, \\ (1 + 10(r-2))^{-3} & r > 2. \end{cases} \tag{2.24}$$

In the ideal case, the disc should remain intact at any time $t > 0$. We study the surface densities of the test cases involving thermal energy and momentum diffusion, described in Table 2.1. As with the driven turbulence experiments, we do not include metals despite the model suffix `-z`. We include the suffix to allow for cross-comparison across the various physical tests. The leftmost plot in Fig. 2.3 shows the surface density of the disc in the `None` model at $t \approx 2t_{\text{orb,r=1}}$. We focus on relatively early times to decouple the effects of inherent numerical diffusion in the MFM method

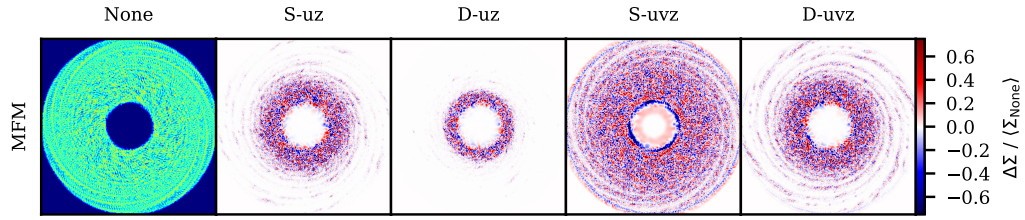

Figure 2.3: Normalised surface density differences between each mixing model. The leftmost plot shows the surface density profile of the `None` case for comparison. The `S-` models lead to a more rapid break-up of the disc, especially in the case including momentum diffusion (`S-uvz`). The `D-uz` model minimises the difference to the `None` case. The dynamic diffusion of thermal energy and momentum (`D-uvz`) leads to an equivalent amount of differences to the `S-uz` case but the break-up of the disc, due to the over-diffusive `S-uvz` case, is avoided.

with those of the turbulent mixing models. In the `None` case, we see that the inner half of the disc is noisy and in the outer region, there are density waves propagating outward, similar to the results in Hopkins (2015). The noise in the inner region, where the orbital time is short, is due to numerical diffusion randomising the particle motions; short of altering the hydrodynamical solver, this effect is unavoidable and is present in all of the tests we investigate here, with or without mixing. We therefore use the `None` model as a baseline experiment to compare the four mixing models.

In the rest of the four panels in Fig. 2.3, we show the point-wise difference in surface density between the model in question, $\Sigma_{\rm i}(r)$, and the `None` case, $\Sigma_{\rm None}(r)$, normalised to the mean surface density in the `None` case; i.e. $\Delta\Sigma(r)/\langle\Sigma_{\rm None}\rangle = (\Sigma_{\rm i}(r) - \Sigma_{\rm None}(r))/\langle\Sigma_{\rm None}\rangle$. This allows us to compare the diffusion of energy and momentum spatially, by observing the differences directly on the surface of the disc.

In the mixing tests without momentum diffusion, `S-uz` (second panel in Fig. 2.3) and `D-uz` (third panel, Fig. 2.3), the inner region ($0.5 < r < 1.0$) of each annulus shows differences compared to the `None` case. These are due to the false identification of turbulence caused by two effects: (i) both models identify random particle motions, like those in the central region, with turbulence and (ii) the diffusivity scales as $D \propto r^{-3/2}$ in the `S-uz` case. The advantage of the dynamic model is that the radial extent of the affected region is significantly smaller compared to the `S-uz` case. `D-uz` predicts much smaller values of $C_{\rm s}$ (median of $C_{\rm s} \approx 0.026$ in the `D-uz` case compared to $C_{\rm s} = 0.2$ in `S-uz`) and, therefore, the differences with the `None` case are mostly limited to the relatively noisy central annulus.

From Fig. 2.3 we see that the addition of momentum diffusion (the `S-uvz` and `D-uvz` cases, two rightmost panels) causes increased noise throughout the disc. Specifically for the `S-uvz` case, the noise in the disc extends radially to the outer boundary, and the material spreads into the central region, compared to the `None` case (notice the faint pink annulus). There is also a corresponding deficit of gas at $R \approx 0.5$ (blue ring) showing that gas has collapsed due to the viscous instability. Comparing `D-uvz` to the `None` case, we see that the differences do not extend to the boundary, but rather follow the density waves in the disc caused by natural dissipation in MFM. The central noisy region has the same extent as in the `S-uz` case but with no apparent in-fall of material into the central region. These results indicate that the dynamic model allows momentum diffusion in laminar shear flows without instigating the viscous instability.

In Fig. 2.4, we show the azimuthally averaged difference between the surface

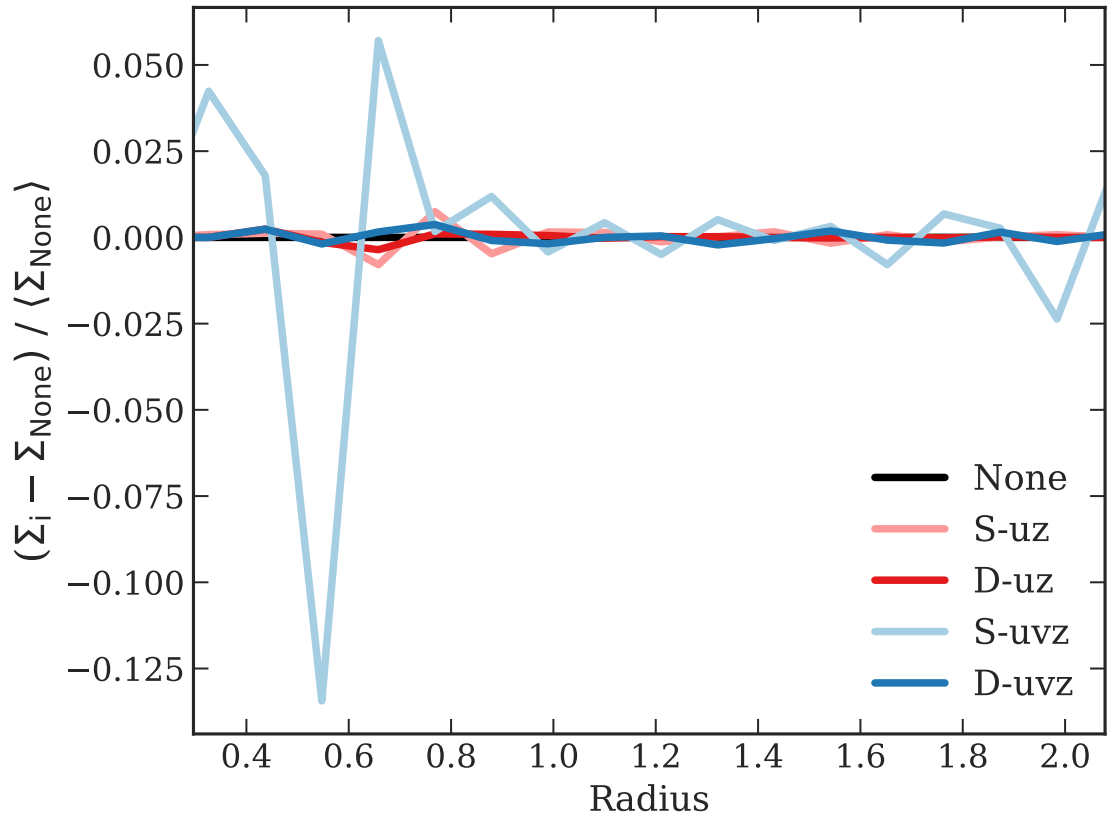

Figure 2.4: The azimuthally averaged differences in surface density for the Keplerian disc experiments, between each mixing model i and the `None` case at $t \approx 2t_{\rm orb}$. The `D-uz`, `D-uvz`, and `S-uz` cases show small fluctuations around the `None` case whereas the `S-uvz` model causes large differences due to over-diffusing momentum.

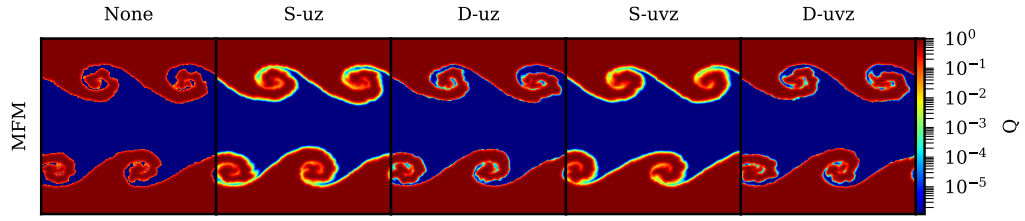

Figure 2.5: A comparison of tracer concentrations Q, in the 2D Kelvin-Helmholtz instability test at $t = 0.28\tau_{\rm KH}$, simulated with the MFM method. The columns represent the five turbulent mixing models. In the `None` case, gas particles cannot exchange the tracer concentrations and the interface of the instability remain unsmoothed. In this case, inter-fluid mixing only occurs when particles move across the boundary. The `S-` cases diffuse rapidly due to the presence of strong shear at the boundary, and the tracer engulfs the whorls during the early evolution of the instability. The `D-` cases provide a compromise between the two situations – they limit diffusion strictly to the interface between the two fluids, and the internal structure of the whorls are distinguishable.

density in each mixing model, $\Sigma_{\rm i}$, and the `None` case, $\Sigma_{\rm None}$, in order to get a more quantitative estimate of the extent of over-diffusion. Both `S-uz` and `D-uz` are nearly identical to the `None` case, except in the innermost region ($R \lesssim 1$ for `S-uz` and $R \lesssim 0.7$ for `D-uz`) where there are slight fluctuations about the `None` value. The differences compared to the `None` case are reduced because of the azimuthal averaging.

Introducing momentum diffusion results in greater quantitative differences in the disc. The `S-uvz` model results in significant flows (inward and outward) in the annulus relative to the `None` case. For example, angular momentum transport causes some of the material at $R \approx 0.5$–0.6 to flow inward due to loss of angular momentum, and some to flow outward. This occurs throughout the disc, resulting in fluctuations extending to the outer edge of the disc. These flows lead to the `S-uvz` case showing regions of higher densities between $0.6 \lesssim R \lesssim 1.8$. Comparatively, the small fluctuations in the `D-uvz` case are similar to the models without momentum diffusion. The dynamic model clearly reduces the impact of the viscous instability, but a question then arises: why does the dynamic model not eliminate the viscous instability completely?

The dynamic model formally predicts $C_{\rm s} = 0$ for all of the gas particles. Prior to $t \approx 0.4t_{\rm orb}$, virtually all of the particles have a near zero value of $C_{\rm s}$. However, by $t \approx 2t_{\rm orb}$, turbulent momentum diffusion and the inherent noise in the inner regions lead to a small non-zero distribution centred at approximately $C_{\rm s} \approx 0.04$ (ignoring particles with $C_{\rm s} = 0$). Overall, the particles have a median value of $C_{\rm s} \approx 0.026$ when

including the 21.4% of particles at $C_s = 0$. Although we cannot avoid the inherent numerical noise, due to MFM's Riemann solver, the dynamic method does minimise the damage: in standard implementations, C_s is in the range $\approx 0.1 - 0.2$ (Garnier et al., 2009) while in the dynamic model, only a small fraction ($\approx 7\%$) of the particles attain such values.

These tests show that in the case of a rotating, laminar shear flow, the dynamic model (`D-uvz`) can indeed minimise turbulent mixing of thermal energy and momentum – preventing unphysical viscous flows within the disc. Therefore, we recommend incorporating the dynamic Smagorinsky model in all numerical simulations involving rotating galactic discs which simultaneously include turbulent mixing models.

The Kelvin-Helmholtz instability

In a fluid with high velocity shear, or at the shearing interface between two fluids, rapidly growing perturbations cause mixing within the fluid. In the case of a shear interface, the perturbations cause the two fluids to encroach the boundary, transporting and mixing fluid properties such as thermal energy, momentum, and metals. This is the Kelvin-Helmholtz (KH) instability, and the timescale characterising the growth of perturbations is given by,

$$\tau_{\rm KH} = \frac{(\rho_1 + \rho_2)}{\sqrt{\rho_1 \rho_2}} \frac{\lambda}{\Delta v}, \tag{2.25}$$

where ρ_1, ρ_2 are the densities of the two fluids, Δv is the velocity difference, and λ is the wavelength of the perturbation (Chandrasekhar, 1961). In the $t < \tau_{\rm KH}$ regime (the linear regime), the flow has not completely transitioned to turbulence and sub-grid turbulent mixing does not dominate the resolved mixing.

Shear flows, and hence KH instabilities, are ubiquitous in galactic environments: ram pressure stripping of galaxies falling into groups and clusters, galaxy mergers, galactic winds streaming into the circumgalactic medium, and a myriad of other processes involve the KH instability. The constant-coefficient Smagorinsky model over-diffuses in such situations because it identifies shearing motion with turbulence, as discussed in Section 2.1.3, and the presence of high shear increases the diffusivity to the maximum at the interface. Ideally, sub-grid turbulent mixing models that better reflect physical reality are preferable. More precisely, models that capture unresolved mixing in turbulent situations and avoid diffusion in laminar shear flows.

We investigate a simple 2D KH test in order to demonstrate the over-diffusive

nature of the standard Smagorinsky model and determine if the dynamic model mitigates the problem. We set up a 2D configuration of 256^2 ideal gas particles in a square ($L = 1$) domain, with constant pressure, and with an initial density profile,

$$\rho(y) = \begin{cases} \rho_2 - (\Delta\rho/2)\exp\left[(y - 1/4)/\Delta y\right], & 0 \leqslant y < 1/4 \\ \rho_1 + (\Delta\rho/2)\exp\left[(1/4 - y)/\Delta y\right], & 1/4 \leqslant y < 1/2 \\ \rho_1 + (\Delta\rho/2)\exp\left[(y - 3/4)/\Delta y\right], & 1/2 \leqslant y < 3/4 \\ \rho_2 - (\Delta\rho/2)\exp\left[(3/4 - y)/\Delta y\right], & 3/4 \leqslant y < 1 \end{cases} \tag{2.26}$$

and initial velocity profile,

$$v_x(y) = \begin{cases} -1/2 + (1/2)\exp\left[(y - 1/4)/\Delta y\right], & 0 \leqslant y < 1/4 \\ 1/2 - (1/2)\exp\left[(1/4 - y)/\Delta y\right], & 1/4 \leqslant y < 1/2 \\ 1/2 - (1/2)\exp\left[(y - 3/4)/\Delta y\right], & 1/2 \leqslant y < 3/4 \\ -1/2 + (1/2)\exp\left[(3/4 - y)/\Delta y\right], & 3/4 \leqslant y < 1. \end{cases} \tag{2.27}$$

We choose $\rho_2 = 2$, $\rho_1 = 1$, and $\Delta y = 0.025$, and introduce a sine wave velocity perturbation, with period $T = 2$ and amplitude $A = 0.01$, at $t = 0$. This gives a perturbation wavelength $\lambda = 1/2$, and therefore $\tau_{\mathrm{KH}} \approx 0.71$. In addition, we add a uniform passive scalar tracer of concentration $Q = 1$ to all gas particles in the range $0 \leqslant y < 1/4$ and $3/4 \leqslant y < 1$, which is the higher density gas. We focus our analysis on the evolution of the tracer concentration Q.

Fig. 2.5 shows the tracer concentration at $t = 0.28\tau_{\mathrm{KH}}$ (i.e. in the linear regime). The five panels show the five mixing models described in Table 2.1. We first consider the `None` case. The tracer concentration Q follows the high density regions of the experiment and we see individual particles advecting across the shear interface. Although there appears to be less tracer on this interface (orange-red line), it is impossible for particles to exchange tracer in the `None` case. The interpolation method we employ causes this effect as it smooths the particle properties over the resolution scale h, leading to a minuscule amount of artificial mixing.

When we allow for the turbulent mixing of the tracer in the `S-` and `D-` cases, we see diffusion occurring along the shear interface. In both `S-` models, $Q \sim 10^{-2}$ at the interface and the tracer engulfs the initially pristine, lower density gas in the whorls. This is in contrast to the two `D-` cases where the majority of diffusion occurs in the whorls themselves with comparatively little along the rest of the interface.

The constant-coefficient Smagorinsky model diffuses the most because of the false identification of strong turbulence through the norm of the shear tensor $|S^*|$. The velocity profile in equation (2.27) shows that although the flow is laminar, there is a gradient, $|\partial v_x/\partial y| > 0$, across the entire domain. Based on our arguments for the Keplerian disc in Section 2.1.3, the fact that $D \propto |\partial v_x/\partial y|$ directly leads to the over-diffusion of the tracer. The diffusion coefficient D also depends on $|S^*|$ in the `D-` models but in the dynamic model most of the values of C_s are near zero. Only in the whorls do we find values of C_s as large as $C_s = 0.2$ but these are limited to this region – where the transition to turbulence begins.

Over-diffusion in non-turbulent shear flows can have important consequences for the gas in numerical studies of galaxy formation. If the KH timescale is longer than the gas cooling time, then the gas is susceptible to over-cooling in the whorls due to metals transferred to the region. This does not accurately capture what physically occurs at the sub-grid level; in the linear regime, the fluids do not mix completely. The dynamic model solves this issue by minimising mixing along the interfaces while allowing it to proceed in the whorls where, in principle, the KH instability continues down to unresolved scales.

2.1.4 Isolated disc galaxy

Here we investigate an isolated galaxy in order to test the effects of localised diffusion in a more realistic, physical environment.

Initial conditions

We follow the method outlined in Springel et al. (2005a) to set up the initial conditions, using the `GALSTEP` package[15] (Ruggiero & Lima Neto, 2017) in combination with `DICE`[16] (Perret et al., 2014). The galaxy is a Milky Way-like system (Sokołowska et al., 2016) consisting of a dark matter halo of mass $M_h = 10^{12}$ $M_\odot$, a gaseous halo of mass $M_{gh} = 3 \times 10^{10}$ $M_\odot$, a stellar bulge of mass $M_b = 10^{10}$ $M_\odot$, and gas and stellar discs of masses $M_g = 10^{10}$ $M_\odot$ and $M_s = 5 \times 10^{10}$ $M_\odot$, respectively. The dark matter and bulge components follow a Hernquist density profile with scale factors $a = 47$ kpc and $a = 1.5$ kpc, respectively. The stellar and gaseous discs follow an exponential density profile with a radial scale $R_d = 3.5$ kpc, and the scale-heights for these com-

[15] https://github.com/ruggiero/galstep
[16] https://bitbucket.org/vperret/dice

Table 2.2: Parameters for the isolated galaxy.

Component	Particle Mass ($M_\odot$)	Min. Softening (pc)	N_{part}
Gas	5.0×10^3	1.4	8×10^6
Halo	5.0×10^5	12.0	2×10^6
Disc	5.0×10^5	3.2	1×10^5
Bulge	2.5×10^5	1.4	4×10^4

ponents are $z_0 = 0.7$ kpc and $z_0 = 0.0175$ kpc, respectively. We initialize the gaseous halo metallicity at $Z_{gh} = 10^{-3} Z_\odot$, and the gaseous disc metallicity at $Z_g = Z_\odot/3$, with $Z_\odot = 0.02$ (Anders & Grevesse, 1989). Our fiducial run is carried out at a gas mass resolution of $M_{g,res} = 5 \times 10^3$ $M_\odot$, along with the softening values specified in Table 2.2. We evolve the disc for 2 Gyr in an isolated (non-cosmological) setting.

Galactic physics

Star formation & cooling

We employ the sub-grid multiphase ISM model of Springel & Hernquist (2003a), which places the ISM gas ($n_H > n_{*,crit}$, where $n_{*,crit}$ is the star formation density threshold) on an effective equation of state (EoS). This model has been used extensively in numerical galaxy formation studies (for recent examples, see Genel et al., 2014; Schaye et al., 2015; Grand et al., 2017), and provides well-converged results (Springel et al., 2005a). For primordial and metal-line cooling, we use the `GRACKLE-2.1` cooling library (Smith et al., 2017) in combination with the UV background from Faucher-Giguère et al. (2009).

In this sub-grid model, stars form stochastically, on a star formation timescale t_{sfr}, from gas that reaches a density $n_{*,crit}$. Here we take $n_{*,crit} = 0.1$ cm^{-3} and $t_{sfr} = 2.1$ Gyr, which give a good fit to the Kennicutt law (Kennicutt, 1998; Springel & Hernquist, 2003a). We differ from the original model in the choice of the initial mass function (IMF). We assume the Chabrier IMF (Chabrier, 2003) instead of the Salpeter IMF (Salpeter, 1955).

Feedback

Due to lack of resolution, it is necessary to include a sub-grid model for feedback from stars, including the effects of supernovae, stellar radiation, and stellar winds.

For massive stars, we adopt the scheme used in the MUFASA simulations (Davé et al., 2016, 2017) and refer the reader to these references for a detailed description. In brief, stellar feedback is expected to drive galactic outflows and, in the MUFASA

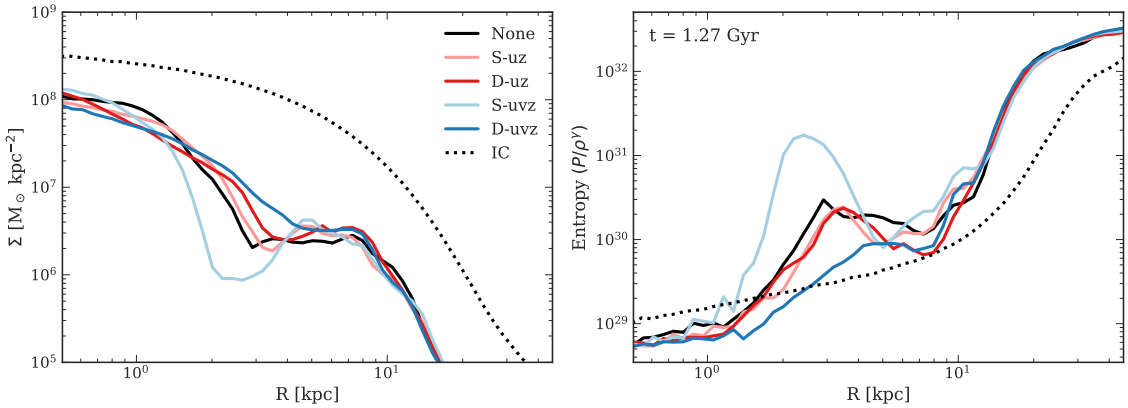

Figure 2.6: Isolated galaxy radial disc profiles at $t = 1.27$ Gyr, averaged azimuthally and vertically between ±0.5 kpc from the plane of the disc. (*left*) The gas surface density. (*right*) A monotonic measure of the gas entropy. The dotted line in both plots represents the corresponding profile in the initial condition, which is identical for all models. Over time, the normalisation of the surface density decreases due to gas consumption and stellar feedback. The dynamic model with thermal energy and momentum diffusion (`D-uvz`) produces a more stable disc whereas the constant-coefficient model (`S-uvz`) shows a more concentrated central region.

approach, stellar feedback directly launches a kinetic wind via a two-parameter model that characterises the net effect of stellar feedback into a mass loading factor η, and the wind speed v_w. These parameters are calibrated to the FIRE wind scalings (Muratov et al., 2015), where η scales with the stellar mass of the host galaxy and v_w scales with the galaxy's circular velocity. We fix η and v_w to the values for our isolated system based on equations (6) and (7) in Davé et al. (2016). At launch, the outflow hydro-dynamically decouples, and only recouples if the wind speed drops to 50% of the local sound speed, the density of the ambient medium is 1% of the interstellar medium density, or the outflow has travelled for 2% of the Hubble time at launch (Davé et al., 2016). For a more detailed description of the decoupled outflow model, see Springel et al. (2005a), Oppenheimer & Davé (2008), Liang et al. (2016), and Davé et al. (2016).

The contribution from supernovae type Ia (SNIa) are modelled following Scannapieco & Bildsten (2005) as a prompt and delayed component, where the prompt component occurs simultaneously with SNII, and the delayed component begins 0.7 Gyr after the star formation time. The prompt component is assumed to release 10^{51} erg of thermal energy to the star-forming gas, whereas the delayed component is added in a kernel-weighted manner to the nearest 16 gas particles.

Chemical enrichment

The chemical enrichment of gas is paramount to the study of turbulent mixing as the metallicity follows the mixing of energy and momentum, and, therefore, provides a tracer for the diffusion equation. We track 11 chemical elements in our isolated and cosmological simulations: H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe. These elements are produced from three sources in the simulations: SNIa, SNII, and the winds from AGB stars.

For SNIa, the prompt component returns mass to the ISM and enriches the star-forming gas instantaneously. Each SNIa is assumed to release $1.4 M_\odot$ of metals, with yields from Iwamoto et al. (1999). For the delayed component, stars deposit metals over their nearest 16 neighbouring gas particles in a kernel-weighted fashion.

SNII return mass and enrich the gas via the instantaneous recycling approximation (Springel & Hernquist, 2003a; Oppenheimer & Davé, 2008; Davé et al., 2016) following,

$$\Delta Z_i = (1 - f_{SN}) \cdot y_i(Z) \cdot \frac{\Delta t}{t_{sfr}} \tag{2.28}$$

where f_{SN} is the fraction of stars in the Chabrier IMF expected to go supernova, $y_i(Z)$ is the metallicity dependent yield of species i, Δt is the timestep, and t_{sfr} is the aforementioned star-formation time-scale. The SNII yields follow Nomoto et al. (2006) and are a function of the metallicity of the gas receiving the metals. Following Davé et al. (2016), the SNII yields are reduced by a factor of 0.5 in order to match the mass-metallicity relationship. SNII also return mass into the gas via the instantaneous recycling approximation.

For AGB stars, chemical enrichment is done in a kernel-weighted fashion over the nearest 16 neighbours. AGB yields are obtained from a lookup table as a function of age and metallicity based on the study in Oppenheimer & Davé (2008). The mass-loss rates of the AGB stars are calculated from a lookup table based on Bruzual & Charlot (2003) stellar models.

Results: Disc stability

Fig. 2.6 shows radial disc profiles of the surface density (left panel) and entropy[17] (right panel) in our isolated galaxies after 1.27 Gyr of evolution. The dotted line represents the initial condition (IC) for all models. The differences between models only appear

[17]The entropy scales as $S \propto \ln(P/\rho^\gamma)$, and therefore our measure is off by a multiplicative constant.

after 1 Gyr (≈ 4 rotations in the mid-disc) and continue until star formation consumes the bulk of the gas after 2 Gyr.

First, we consider the `None` case. The radial surface density (left panel, Fig. 2.6) gives a measure of the stability of the disc. We see that by $t = 1.27$ Gyr, compared to the shape of the IC density profile, the gas has moved inward toward the centre, especially within $R < 3$ kpc. A combination of the bar instability and inherent numerical dissipation causes the gas to concentrate inside $R \approx 2$ kpc, whereas gas consumption and galactic winds due to supernova feedback cause the difference in normalisation compared to the IC. Correspondingly in the right panel, there is an order-of-magnitude increase in entropy from $R = 1$ kpc to $R = 3$ kpc.

In the `S-uz` and `D-uz` models, we see only minor difference in the surface density and entropy compared to the `None` case. Evidently, thermal energy diffusion, combined with metal diffusion, has negligible impact on the structure of the disc.

When we introduce momentum diffusion in the `S-uvz` and `D-uvz` cases, the differences compared to the `None` case at $t = 1.27$ Gyr are more significant. Fig. 2.6 shows that in the `S-uvz` case there is an order-of-magnitude deficit of gas surface density (left plot) between $R = 1$ kpc and $R = 3$ kpc with a corresponding jump in entropy (right plot). Over-diffusion of momentum due to the diffusivity scaling strongly with the shear causes the inward flowing gas to be more centrally concentrated compared to the `None` case, and also engenders an outward flow leading to slightly higher density (again, compared to `None`) at $R \approx 4 - 7$ kpc. Once the instability occurs, the effect accelerates and the trend remains throughout the evolution of the disc. However, with the dynamic model (`D-uvz`), less gas gets redistributed. From the surface density, we see that the disc stabilises when momentum is diffused locally, based on the turbulent character of the flow.

Results: Metal distribution functions

Supernovae and stellar winds inject energy and metals into the interstellar medium (ISM), engendering turbulent motion that mixes and spreads thermal energy, momentum, and metals throughout the medium. These processes also drive a galactic-scale wind that deposits metals and drives turbulence in circumgalactic medium (CGM) (Evoli & Ferrara, 2011).

In order to obtain a measure of energy and metal mixing in the five models under consideration (see Table 2.1), we examine metal distribution functions (MDFs) in

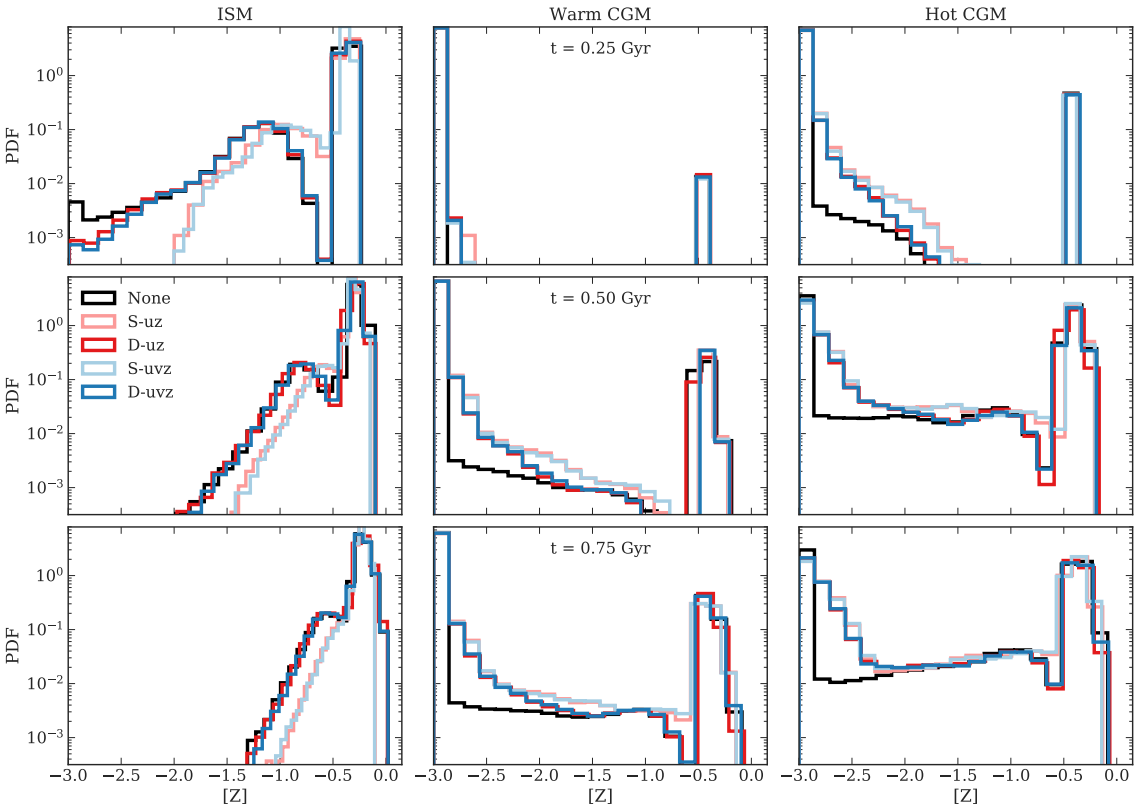

Figure 2.7: The metal distribution functions (MDFs) for the isolated disc galaxy in solar units ($Z_\odot = 0.02$). Each column represents a specific region of $T - \rho$ phase space. The leftmost column represents the interstellar medium, whereas the middle and rightmost columns show the MDFs for the warm (10^5 K $< T < 10^6$ K) and hot ($T > 10^6$ K) circumgalactic medium, respectively. Each row represents a different time in Gyr: $t = 0.25$ Gyr, 0.50 Gyr, and 0.75 Gyr from top to bottom, respectively. Dynamic diffusion slows metal mixing in the interstellar medium and causes increased mixing in the circumgalactic medium, at approximately the same levels as the constant-coefficient Smagorinsky model.

three phases of gas: (1) gas with density above the star formation critical density ($n_{\rm H} > 0.1$ cm^{-3}) – i.e, the ISM gas, (2) warm CGM gas in the halo in the range $10^5{\rm K} < T < 10^6{\rm K}$, and (3) hot CGM gas with $T > 10^6{\rm K}$. The gas density in the latter two phases is $n_{\rm H} \leqslant 0.1$ cm^{-3}. In our isolated system, we find that the cool, non-star-forming gas ($T < 10^4$ K) is mostly at the outskirts of the halo and that there are no significant variation in its MDFs between models. Therefore, we do not discuss this phase further in the isolated case. Fig. 2.7 shows the MDFs for the the ISM, the warm CGM, and the hot CGM, at three separate evolutionary times: $t = 0.25$ Gyr (top row), 0.50 Gyr (middle row), and 0.75 Gyr (bottom row) after the initial conditions for the MFM method. Henceforth, we use $[Z] \equiv \log(Z/Z_\odot)$ as a proxy for metallicity.

First, we focus on the ISM shown in the left column of Fig. 2.7. At $t = 0.25$ Gyr, the MDF of the `None` case is bimodal with a narrow peak at $[Z] \approx -0.3$, a broad distribution at $[Z] < -0.5$ with a peak at $[Z] = -1.2$, and a dearth at $[Z] = -0.5$. Gas in the ISM with metallicity lower than the initial $[Z] \approx -0.5$ comes from the lower metallicity (initial $[Z] = -3.0$) halo gas that cools onto the disc and is steadily enriched. Recall that in the `None` case, gas cannot exchange metals between particles and therefore it gives an upper bound on the mixing timescale for a given metal injection rate from stars. The `D-` cases follow the `None` case with minimal differences, therefore we conclude that the `D-` cases provide minimal turbulent mixing in the ISM. The `S-` cases share the basic shape as the other cases but the lower metallicity component is narrower (there is very little gas with $[Z] < -2.0$), and is shifted to the right, with peak $[Z] \approx -0.9$. The over-diffusivity of the `S-` cases is apparent and is caused by the high levels of shearing motion in the supersonic ISM, leading to high values of the diffusivity D, similar to the Keplerian disc in Section 2.1.3.

At $t = 0.5$ Gyr in the ISM (left column, middle row), the low metallicity component of the MDF for all cases has narrowed and shifted toward higher metallicity. In the `None` case, the tightening is due to stars depositing metals into the medium and driving all of the gas toward highly metallicities. Here, the `D-` cases follow the `None` case closely – as at $t = 0.25$ Gyr – while the `S-` cases have a smaller spread. We attribute the differences between the `D-` and `S-` cases to the dynamic model predicting median diffusivities orders-of-magnitude lower than the constant-coefficient Smagorinsky model in the ISM at $t > 0.25$ Gyr.

By $t = 0.75$ Gyr, all cases have tight distributions with very few particles having $[Z] < -1.5$. `None` and `D-` cases continue to exhibit very similar distributions with a

larger spread compared to the S- cases.

Now we consider the MDFs of the warm CGM, in the middle column of Fig. 2.7. In the None case at $t = 0.25$ Gyr, the distribution is bimodal with peaks at $[Z] \approx -0.5$ and $[Z] = -3.0$. The peak at $[Z] = -3.0$ is the initial halo metallicity whereas the peak at $[Z] \approx -0.5$ is due to enriched gas either pushed out of the ISM by stellar winds or at the interface between the CGM and ISM. The D- and S- cases closely follow the None case, but with a slight spread toward $[Z] > -3.0$ for the low metallicity gas.

Next we consider the $t = 0.5$ Gyr and $t = 0.75$ Gyr cases. In all of the experiments, the flow of the enriched ISM gas into the CGM results in an increase in the fraction of particles near $[Z] \approx -0.5$. The differences near $[Z] = -3.0$ between the None and S-/D- cases are due to the inability of the gas in former case to mix metals. For the None case, a spread only occurs if the gas enters the CGM at $[Z] > -3.0$ from the ISM, or the CGM gas at the ISM-CGM boundary becomes enriched via the kernel-weighting approach associated with delayed SNIa and AGB wind feedback. Recall that in this implementation, metals are deposited in a kernel-weighted fashion over the nearest 16 neighbouring particles. Therefore, gas classified as CGM, that is spatially adjacent to the star-forming gas in the ISM, can be enriched at low levels. The gas in the turbulent mixing cases consistently exchange metals if the value of C_s is non-zero, and the gas-phase metals in the halo further mix – leading to the greater spread in the lower peak at $[Z] = -3.0$, corresponding to the bulk of the initial gas. The D- cases show the aforementioned spread in the metallicity in the range $-3.0 < [Z] < -1.75$, indicating increased mixing due to sub-grid turbulence, but then coincides with the None case for metallicities $[Z] > -2.0$. The S- and D- cases share a similar distribution below $[Z] < -2.25$ but the S- cases show more particles with metallicities in the range $-2.25 < [Z] < -0.75$, indicating even greater mixing than the D- cases.

The MDFs in the hot CGM, the rightmost column of Fig. 2.7, follow similar trends to the those in the warm CGM. At $t = 0.25$ Gyr for the None case, the peak due to enriched ($[Z] \approx -0.3$) ISM gas entering the halo is prominent. There is also a small spread at $[Z] = -3.0$ for reasons already noted.

At $t \geqslant 0.5$ Gyr, the None case undergoes slight evolution and by $t = 0.75$ Gyr, a mild positive slope develops in the range $-2.5 \leqslant [Z] \leqslant -0.75$. The D- and S- cases follow near identical evolution at these later stages of the simulation, each with a spread in the distribution at $[Z] = -3.0$. Here the differences between the S- and D- cases are minimal since the hot gaseous halo is turbulent and dominated by random motions, driven by the galactic outflows.

Our isolated experiment demonstrates that by endowing particles with diffusivities based on the local fluid properties, we obtain significant differences in the ISM. The constant-coefficient Smagorinsky model causes the MDFs in the ISM to rapidly tighten toward the mean value, whereas the dynamic model predicts diffusivities orders-of-magnitude smaller and the corresponding MDFs closely follow the `None` case. Simultaneously, in the hot turbulent halo, the constant-coefficient and dynamic Smagorinsky models produce similar distributions by $t = 0.5$ Gyr due to the latter having higher values of C_s. Overall, for the constant-coefficient model, non-negligible shear in all gas phases causes the rapid diffusion of fluid properties, whereas the dynamic model allows different regions of phase space to undergo unique evolution in terms of the MDFs. We stress the importance of the unique evolution of both phases of gas: with a constant C_s, it is impossible to capture the decreased mixing in the ISM while simultaneously capturing the high level of mixing in the hot turbulent halo. The dynamic model provides an interesting avenue for follow-up study with zoom-in simulations, in order to gauge the impact on the ISM and CGM in a cosmological context.

2.1.5 Cosmological volumes

As we demonstrate in the previous section, turbulent mixing can alter the distribution of metals in various gas phases of an isolated disc experiment depending on the localisation of C_s. In a realistic cosmological environment, the evolution of galaxies is much more complex due to interactions between the galaxies and their environments. These interactions include galaxy mergers, gas inflows and outflows, tidal interactions, ram pressure stripping, etc. – all of which contribute to the production of turbulence in the galactic environments (Iapichino et al., 2013; Schmidt et al., 2016). The resulting turbulence redistributes thermal energy, momentum, and metals, and must be included in numerical studies of galaxy evolution in order to have a self-consistent treatment of the physical models.

In contemporary Lagrangian-based numerical cosmological experiments, smoothed particle hydrodynamics (SPH) has been employed in simulation programs such as EAGLE (Schaye et al., 2015), OWLS (Schaye et al., 2010), and Romulus (Tremmel et al., 2017), whereas recently mesh-free finite mass (MFM) method has been employed in the MUFASA (Davé et al., 2016, 2017) simulations. These experiments have produced a wealth of results for understanding galaxy evolution and gas properties (see

Somerville & Davé 2015; Naab & Ostriker 2017 for a summary), despite the fact that it is not possible for contemporary models to include all of the relevant physics[18]. Including sub-grid turbulent mixing could alter the results of such large-scale simulations. Indeed, Tremmel et al. (2019) argue that turbulent mixing is critical for efficient redistribution of thermal energy released during active galactic nuclei episodes, and in previous sections, we showed that metal redistribution is also affected by turbulent mixing. In models where the over-diffusive Smagorinsky model is used, the dynamic model can lead to differences in, for example, gas-phase metal abundances and hence, stellar abundances. In principle, the differences in the manner and the rate at which metals are distributed could also affect the formation sites and population statistics of Population III stars and direct collapse seed supermassive black holes (see Section 2.1.1).

In this section, we examine a set of cosmological simulations in order to test the effects of the dynamic model on the global gas enrichment levels and distributions. In what follows, we adopt the MUFASA model (Davé et al., 2016) in combination with the diffusion models we describe in Table 2.1. We choose to use the MUFASA model partly because it is the only cosmological model that has been implemented using the MFM method at the present.

The initial conditions were created using a modified version of `GRAFIC-2`[19] (Bertschinger, 2011) and the parameters describing our simulations are listed in Table 2.3. In the following subsection, we briefly describe the MUFASA models, and point interested readers to Davé et al. (2016) and Davé et al. (2017) for a more detailed explanation of the physical models. In Sections 2.1.5 & 2.1.5 we examine the global gas-phase metallicity fractions and the metal distribution functions (MDFs), respectively, from the simulation suite.

MUFASA

The MUFASA simulations include the sub-grid models we described in Section 2.1.4 and sub-sections therein, with some modifications to the star formation recipe and feedback described in the following sub-sections.

Star formation

Star formation is based on the molecular gas model of Krumholz et al. (2009), and the implementation dynamically calculates the fraction of molecular hydrogen in gas

[18] For several examples see Naab & Ostriker (2017).

[19] `http://web.mit.edu/edbert/`

Table 2.3: Parameters for the cosmological simulations. We use the Planck Collaboration XIII (2015) cosmological model

Simulation Parameters	
L	25 Mpc h^{-1}
N	2×256^3
m_g	1.26×10^7 $M_\odot$ h^{-1}
m_{dm}	6.88×10^7 $M_\odot$ h^{-1}
$\epsilon_{soft,min}$	0.5 kpc h^{-1}
z_{init}	70
T_{init}	59 K
Cosmological Model	
$\Omega_{m,0}$	0.308
$\Omega_{\Lambda,0}$	0.692
$\Omega_{b,0}$	0.048
h	0.678
σ_8	0.815
n_s	0.968

particles, f_{H_2}, based on the gas surface density and the metallicity. For more precise details, see Davé et al. (2016) and Krumholz et al. (2009). The star formation rate follows,

$$\frac{\mathrm{d}\rho_*}{\mathrm{d}t} = \epsilon_* \frac{\rho f_{H_2}}{t_{\rm dyn}}, \tag{2.29}$$

where $t_{\rm dyn} = (G\rho)^{-1/2}$ is the local dynamical time, ρ is the density of the gas, and $\epsilon_* = 0.02$ is the efficiency of star formation (Kennicutt, 1998). The critical density of star formation is taken at $n_{*,\rm crit} = 0.2$ cm^{-3} following Davé et al. (2016).

For our cosmological simulations, we do not employ the sub-grid interstellar medium (ISM) model of Springel & Hernquist (2003a), however we still require that the Jeans mass is resolved. Therefore, in order to prevent numerical fragmentation at high densities, an artificial pressure is applied above a density

$$n_{\rm th} = \frac{3}{4\pi\mu m_{\rm p}} \left(\frac{5 k_{\rm b} T_0}{G\mu m_{\rm p}}\right)^3 \left(\frac{1}{N_{\rm ngb} m_{\rm g}}\right)^2, \tag{2.30}$$

where $m_{\rm g}$ is the gas particle mass, $\mu = 1.22$, $T_0 = 10^4$ K, and $N_{\rm ngb} = 64$ is the number of neighbours. The pressure is applied in the form of a minimum temperature (Teyssier et al., 2011; Davé et al., 2016),

$$T_{\rm JMT} = T_0 \left(\frac{n}{n_{\rm th}}\right)^{1/3}. \tag{2.31}$$

Feedback

In Section 2.1.4, we described the decoupled-wind model for massive star feedback in addition to feedback from supernova type-Ia (SNIa) and asymptotic giant branch (AGB) stars. We use the same in the following experiments.

The simulations we present here do not include explicit active galactic nuclei (AGN) feedback. AGN feedback is thought to be necessary to prevent excessive cooling and quench star formation in massive systems (see King & Pounds, 2015 for a recent review), but our smaller simulation volumes do not include many such systems. We do, however, include an effective AGN feedback model from the MUFASA simulations that mimics the action of AGN feedback, and suppresses cooling of the diffuse halo gas. Specifically, gas that is not self-shielded in halos with $M_{\rm halo} > M_{\rm q}$ (where $M_{\rm q} = (0.96 + 0.48z) \times 10^{12}$ M$_\odot$) is heated to 20% above the virial temperature of the halo (Mitra et al., 2015). The virial temperature follows (Balogh et al., 1999;

Table 2.4: We separate gas in our cosmological simulations into five phases: the interstellar medium (ISM), cool circumgalactic medium (CGM), hot halo gas (HHG), warm-hot intergalactic medium (WHIM), and cool diffuse gas (DIFF). ρ_* is the star formation threshold, $n_{*,\mathrm{crit}} = 0.2$ cm^{-3} and we give ρ_{bound} in equation (2.33).

Name	Density Range	Temperature Range
ISM	$\rho > \rho_*$	Any
CGM	$\rho_* > \rho > \rho_{\mathrm{bound}}$	Below equation (2.36)
HHG	$\rho > \rho_{\mathrm{bound}}$	Above equation (2.36)
WHIM	$\rho < \rho_{\mathrm{bound}}$	$T > 10^5$ K
DIFF	$\rho < \rho_{\mathrm{bound}}$	$T < 10^5$ K

Voit, 2005),

$$T_{\mathrm{vir}} = 9.52 \times 10^7 \left(\frac{M_{\mathrm{halo}}}{10^{15} \mathrm{M}_\odot \mathrm{h}^{-1}} \right)^{2/3} \mathrm{K}. \tag{2.32}$$

Gas phases

We define five separate gas phases for the following subsections, and examine their properties in a global sense over the entire simulation volume. We give a summary in Table 2.4. The definitions are largely from Davé et al. (2010), except for the definition of the gas associated with the circumgalactic medium of galaxies. For density, we cut the gas phase space using two thresholds ρ_{bound} and ρ_*, with $\rho_{\mathrm{bound}} = \rho_{\mathrm{bound}}(z)$ following,

$$\frac{\rho_{\mathrm{bound}}(z)}{\Omega_{\mathrm{b}}(z)\rho_{\mathrm{c}}(z)} = 6\pi^2 \left(1 + 0.4093 \left(\frac{1}{\Omega_{\mathrm{m}}(z)} - 1 \right)^{0.9052} \right) - 1, \tag{2.33}$$

where

$$\Omega_{\mathrm{m}}(z) = \frac{\Omega_{\mathrm{m},0}(1+z)^3}{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\Lambda,0}}, \tag{2.34}$$

$$\Omega_{\mathrm{b}}(z) = \frac{\Omega_{\mathrm{b},0}(1+z)^3}{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\Lambda,0}}, \tag{2.35}$$

$\rho_c(z) = 3(H(z))^2/(8\pi G)$, and $H(z) = H_0\sqrt{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\Lambda,0}}$. For the second density cut, we adopt $\rho_* = 4.4 \times 10^{-25}$ g cm^{-3}, the star formation density threshold. We have also applied a single temperature cut at $T_5 = 10^5$K to separate the warm-hot intergalactic medium (WHIM) from the cold diffuse gas (DIFF). Following Torrey

et al. (2019), we apply a cut to distinguish the hot halo gas (HHG) from the cool circumgalactic medium (CGM),

$$\log\left(\frac{T}{10^6\ \mathrm{K}}\right) = 0.25\log\left(\frac{n}{405\ \mathrm{cm}^{-3}}\right), \tag{2.36}$$

where T is the gas temperature, and n is the gas density. We define the HHG to be above the temperature threshold in equation (2.36), and the CGM to be below.

It is important to note that these density and temperature cuts do not distinguish gas domains that precisely correspond to their associated galactic or inter-galactic regions. At high redshift ($z \gtrsim 5$), most of the gas classified as the WHIM phase is spatially located in the region that mostly corresponds to the CGM/HHG, and corresponds to outflowing galactic winds from early star formation, which gives rise to low-density, high-temperature gas. Similarly, a fraction of the gas classified as the CGM at $z \gtrsim 1$ is in the cores of the cosmic filaments. However, by $z = 0$, the majority of what we consider the CGM is indeed inside the halos, and the cores of the filaments eventually end up in the WHIM phase. This introduces a transition from CGM to WHIM gas that might not be obvious at first glance. We have not tried to address these trends or optimise the phase cuts because the present study does not pertain to the evolution of the phases *per se*, but rather the effect of differences in mixing strength on approximately physical phase-space cuts.

Global gas evolution

Fig. 2.8 shows the enriched fraction as a function of redshift, i.e. the ratio of the enriched gas mass to the total gas mass in each phase for two metallicity cuts: $[Z] > -5.0$ (left) and $[Z] > -3.0$ (right). We use $[Z] \equiv \log Z/Z_\odot$ as a proxy for metallicity, where Z is the mass fraction of metals in a gas particle, and $Z_\odot = 0.02$ is the solar metal mass fraction (Anders & Grevesse, 1989). The rows represent the five phases defined in Table 2.4.

We first focus on the $[Z] > -5.0$ cut and start by examining the ISM results in the top row of Fig. 2.8. In the `None` case, the fraction of gas enriched to $[Z] > -5.0$ exceeds 90% at $z = 4$, reaches a peak at $z = 0.5$, and then very slightly downturns by $z = 0$ due to accretion of low-metallicity gas. Note that, in the following discussion, the differences between models are more important than the absolute values. In the `S-` models, gas is enriched much earlier and we see 95% of gas above $[Z] > -5.0$ by $z = 9$. We do not observe the same slight downturn as in the `None` case. This is not

surprising. The over-diffusive nature of the S- model leads to a reduced fraction of low-metallicity gas. By $z = 0$, 100% of the gas is above $[Z] > -5.0$ in the S- models. Until $z \sim 1$, the D- models show enrichment levels intermediate between the None and S- cases. At $z < 1$, the D-uz case continues this trend while the D-uvz model exhibits the biggest downturn.

In the CGM, the second row in Fig. 2.8, we notice similar trends to the ISM, at enrichment levels $[Z] > -5.0$. For the None case, the gas is 30% enriched at $z = 3$, 85% (the peak) at $z = 0.5$, followed by a slight downturn. The S- models follow the same qualitative trend but are at a constant 10% above the None case, while the D- models remain in between the S- and None curves at all times.

We do not discuss the details of the trends in the following three rows: the HHG, DIFF, and WHIM phases, respectively, but include them for completeness. In these gas phases, the differences between mixing models are qualitatively the same as the CGM, whereas the S- and D- cases show significant increased enrichment above $[Z] > -5.0$.

Now we focus on the higher metallicity cut, $[Z] > -3.0$, in the ISM (top, right panel in Fig. 2.8). In the None case, gas is enriched over 90% starting at redshift $z = 2$, reaches a peak at $z \approx 0.5$, and turns down by $z = 0$. The trend for the $[Z] > -3.0$ gas is similar to that for the $[Z] > -5.0$ gas. The normalisation of the curve is lower than at the $[Z] > -5.0$ threshold, as expected, since gas is enriched at higher metallicities later in cosmic evolution. In the S- cases, gas is enriched above 90% (for the $[Z] > -3.0$ cut) earlier compared to the None case, starting at $z = 3$, and remains above the None case at all times. The S- cases show a similar downturn to the None case near $z \approx 0.5$. Enrichment levels in the D-uz case remain slightly above the None case, but in D-uvz, there is a sharp downturn at $z \approx 0.75$ and the final enrichment level is below the None case. The enrichment downturn in all of the mixing models is due a fresh supply of lower metallicity gas entering the medium, and the D-uvz model appears to amplify this effect.

We now discuss the enrichment levels above $[Z] > -3.0$ for the CGM in the right column of Fig. 2.8, in the second row from the top. In the None case, the gas is enriched above 30% by $z \approx 3$, rises to a maximum of 80% at $z \approx 0.4$, and slightly decreases to 75% at $z = 0$. Compared to the None case at the $[Z] > -5.0$ metallicity cut, the qualitative trend remains the same while the normalisation has decreased, for the same reason we describe above. In the S- cases, the gas reaches 30% enrichment levels somewhat earlier, by $z \approx 3.5$, and reaches a maximum enrichment level of 90%

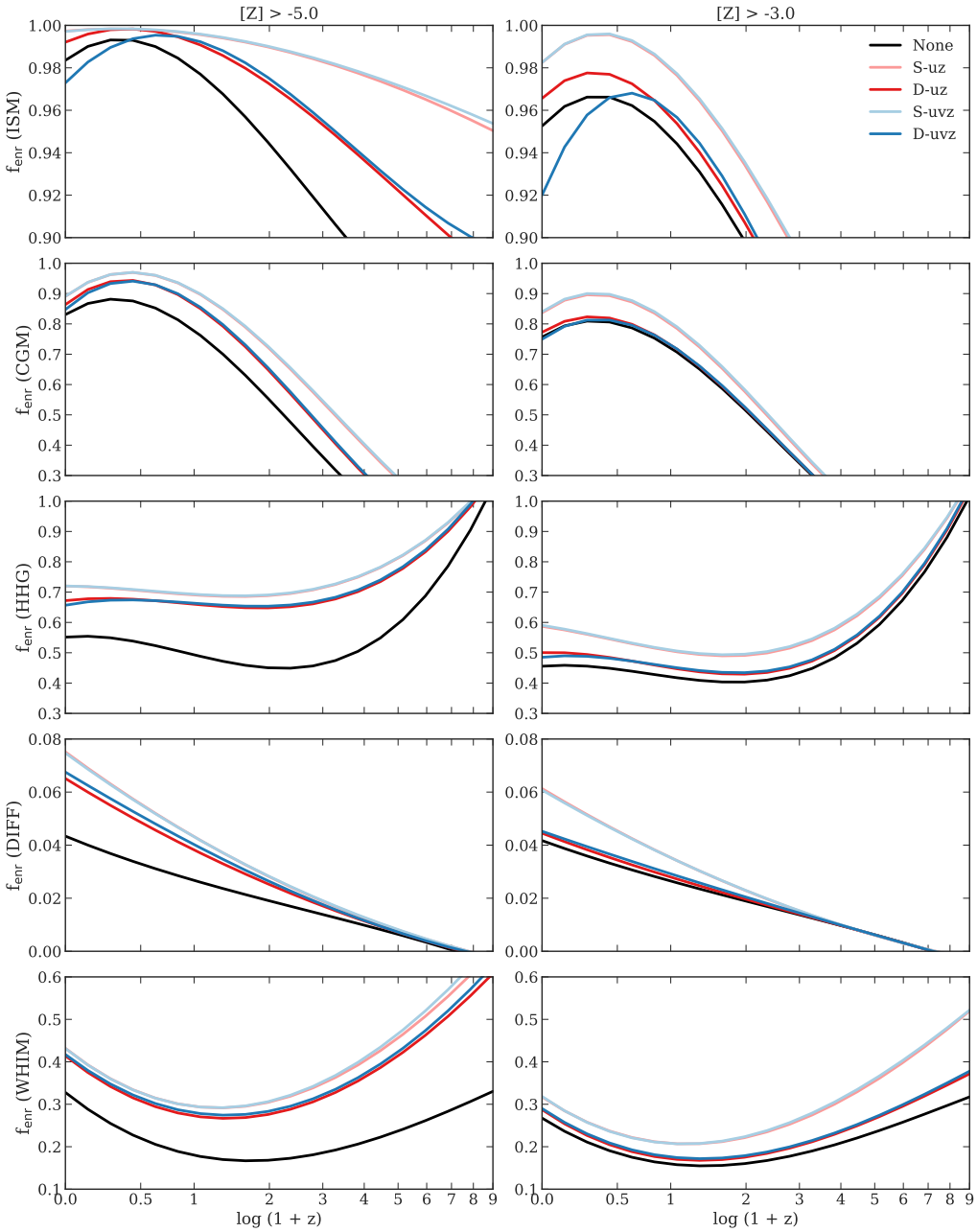

Figure 2.8: The enriched fraction as a function of redshift, defined as the mass fraction of gas with $Z > 10^{-5}Z_{\odot}$ (left) and $Z > 10^{-3}Z_{\odot}$ (right) in each phase. The phases are, from top to bottom: interstellar medium (ISM), cool circumgalactic medium (CGM), hot halo gas (HHG), cold diffuse gas (DIFF), and the warm-hot intergalactic medium (WHIM). Dynamic diffusion results in higher enriched fraction at the $[Z] > -5$ level, while maintaining a similar fraction to the no-mixing case above $[Z] > -3$. The constant-coefficient Smagorinsky model increases the enriched fractions at all metallicity cuts.

at $z \approx 0.4$. The maximum enrichment is 10% higher in the `S-` cases compared to the `None` case showing that the constant-coefficient Smagorinsky model affects higher and lower metallicities equally. The enrichment levels of the gas in the `D-` cases closely follows the `None` case for metallicities $[Z] > -3.0$, with only a slight divergence at $z = 0$.

The HHG, DIFF, and WHIM phases follow the same qualitative trends as the CGM and, as noted above, we do not examine them in detail here. However, these phases show that the dynamic model increases the gas enrichment levels above $[Z] > -3.0$ more so than in the CGM, but only slightly.

As in the isolated galaxy experiment (see Section 2.1.4), the diffusivity plays a significant role in the enrichment levels of cosmic gas on a global scale. The over-diffusive character of the constant-coefficient Smagorinsky model consistently results in the highest gas enrichment levels throughout the evolution of the simulations. Interestingly, these same results also indicate that the dynamic model has the most effect on the metallicities in the range $-5.0 < [Z] < -3.0$ while leaving those above $[Z] > -3.0$ near the no-mixing level. Early star formation and supermassive black hole formation are very sensitive to metallicities in this range (see Section 2.1.1), and we have demonstrated that the dynamic model maximally affects those metallicities. Although we do not resolve early star formation nor include supermassive black holes in our simulations, our results show that the dynamic model ought to be investigated further in cosmological simulations.

Global gas-phase metallicity

Metal distribution functions (MDFs) provide additional insight into the spatial redistribution of metals when compared with global enriched fractions, and we therefore investigate the MDFs of galactic gas at three separate evolutionary times: $z = 2, 1, 0$. Our MDFs are probability density functions ($\mathrm{d}n/\mathrm{d}[Z]$, where n is the number of particles) and are normalised such that the area under each curve is unity. To facilitate plotting, we set a metallicity minimum of $[Z] = -10$ for all gas particles. Fig. 2.9 shows the MDFs in the ISM, CGM, and HHG in the columns, left to right, and at $z = 2$, 1, and 0, from top to bottom. The narrow, leftmost spike in the panels corresponds to the metallicity minimum.

First, as we mention at the end of Section 2.1.5, it is important to note that the differences in the MDFs are not solely due to turbulent mixing. There are transitions

that occur across the sharp boundaries of the phase cuts we use. In all of our mixing models, the cuts mostly affect the gas that falls within the CGM region of phase space and, in examining its spatial distribution at $z \gtrsim 1$, we find that the majority of the low metallicity ($[Z] < -3$) gas in the CGM is in the cosmic filaments. This gas is not spatially associated with dark matter halos, but has the correct density and temperature to belong to the CGM phase. Conversely, the gas in the ISM and HHG phase-space regions do correspond to what we consider their spatial counterparts, and the differences in their MDFs across the evolution of the simulation are dominated by turbulent mixing.

Turning to the distributions in Fig. 2.9, we first investigate the ISM. The distributions across all models appear qualitatively similar, with slight differences at lower metallicities. At $z = 2$, there is a peak at $[Z] \approx -1$ in all models, and in the `None` case, there is also near-pristine gas in the ISM. The `D-` cases show a slightly extended tail covering the range $-6 \lesssim [Z] \lesssim -4$, compared to the `S-` and `None` cases, corresponding to the slight enrichment of the low-metallicity gas in the `None` case. In the `S-` cases, the enrichment process is more efficient. By $z = 1$, the tail tightens and all the distributions have negligible differences. At $z = 0$, however, the `None` and `D-` cases share the same distribution whereas the `S-` models show a tighter distribution. Additionally, there is much more gas at the minimum metallicity in the `None` case compared to both the `D-` and `S-` cases because metal enrichment of pristine particles in the `None` case only occurs when they are spatially adjacent to the stellar feedback sources. In the `D-` and `S-` cases, any particle that is enriched acts as a local source of metals for neighbouring pristine particles, driving down the number of particles at the metallicity minimum. The trends here are similar to, and caused by, the same effect we see in the ISM of the isolated galaxy in Section 2.1.4, where the spread in the ISM MDF strongly depends on the diffusivity.

Next, we examine the MDFs of the CGM. In all models at $z = 2$, the gas in the CGM is a combination of the dense cores of the cosmic filaments and the cool, dense gas within dark matter halos. In the `None` case, the MDF resembles the ISM distribution albeit with a more extended tail toward lower metallicities and more gas at the minimum metallicity. Nonetheless, most of the enriched gas is above $[Z] \geqslant -6.0$. The CGM distribution in the `None` case builds from the galactic winds transporting metals into the medium, sampling the underlying ISM distribution. Additionally, SNIa and AGB stars contribute to varying metal distributions via the kernel-weighting procedure as gas spatially adjacent to the ISM is enriched at low levels. In the `None`

case, the gas that is spatially in the cosmic filaments is at the metallicity minimum. The $z = 2$ D- models show a bi-modality with peaks at $[Z] \approx -1.0$ and $[Z] \approx -5.5$, and in the S- models we also see a bi-modality but with peaks at $[Z] \approx -1.0$ and $[Z] \approx -3.0$. The higher metallicity peaks in the D- and S- cases correspond to the peak metallicity in the ISM whereas the lower peak in each case is due to the gas that is classified as the CGM, yet is spatially in the cosmic filaments. The spatial location of the gas does not change the effect of varying turbulent mixing strength; at redshift $z = 2$, the increased diffusivity in the S- models leads to a 2 - 3 order of magnitude shift compared to the D- models in the secondary low metallicity peak, indicating that the metals are much more dispersed in the S- cases. Comparing to the None case, the question arises as to why the D- and S- cases have a broad lower metallicity peak. At high redshift ($z \gtrsim 2$), in the D- and S- cases, the metal-enriched galactic winds escape the galactic halos and under the action of turbulent mixing, contaminate the gas in the filaments. This is evidenced by the lower amplitude spike at $[Z] = -10$ in these models. This does not occur in the None case, because the particles are unable to exchange metals directly.

By $z = 1$ the fraction of CGM gas in the filaments has dropped as the dense filament cores are heated and enter the WHIM. However, there is still a small fraction of gas associated with the filaments. The None case shows a qualitatively similar distribution compared to $z = 2$ because the gas leaving the cosmic filaments is at the metallicity minimum. In the D- cases, the previous peak at $[Z] \approx -5.5$ becomes a broad shelf between $-5.0 < [Z] < -2.0$. In the S- cases there is also a shelf of gas at $[Z] \approx -2.0$ but the main peak dominates. The strong bi-modality from $z = 2$ has disappeared by $z = 1$ in the D- and S- models partly because the filamentary structure is increasingly classified as the WHIM and partly because the metallicity of the gas particles continues increasing due to mixing.

We turn now to the MDF of the CGM at $z = 0$. At this redshift, most of the gas in the CGM region of phase space is associated with dark matter halos. The None case has a similar distribution to $z = 1$ and $z = 2$. The D- MDFs have narrowed further, with the low extended metallicity shelf at $z = 1$ transforming into a tail that extends from $[Z] \approx -7.5$ to $[Z] \approx -3.0$, and a small shelf at $[Z] \approx -3.5$. Above $[Z] > -2.0$, the D- MDFs coincide with the None case. In the S- cases, the distribution also tightens and the tail is approximately a power law from $[Z] \approx -6.0$ to $[Z] \approx -1.0$, where the latter value is the peak of the distribution. The variations in the MDFs between the None case and both the D- and S- cases represent enriched fractions that have

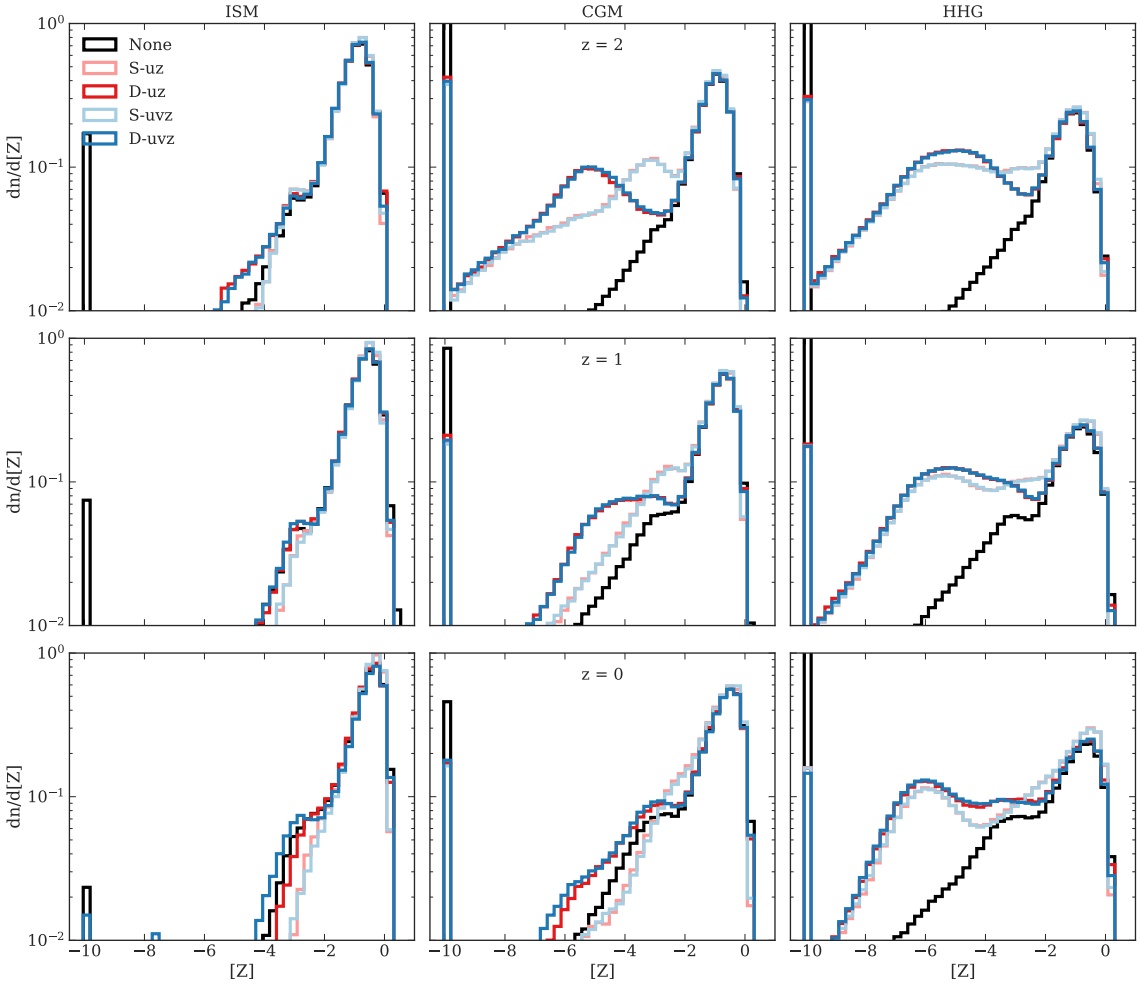

Figure 2.9: The metal distribution functions (MDFs) across three global gas phases (columns) in the simulation volume, at three separate evolutionary times (rows from top to bottom): $z = 2$, 1, and 0. We set a minimum metallicity of $[Z] = -10$ in our simulations, in order to show the abundance of unenriched gas. The overall trends in each phase do not change significantly over time, yet comparing between mixing models reveal slight differences in the distributions of gas-phase metals. Without metal mixing (`None`), galactic winds build MDFs in the non-star-forming gas that are similar to the distribution of the interstellar medium. Turbulent mixing allows for further redistribution of metals once they arrive in the exterior media, altering the original distribution. The diffusivity has a larger role in this method rather than the quantities diffused.

been altered via turbulent mixing rather than the aforementioned transitions between phases.

Shifting focus to the HHG at $z = 2$, we find that the gas in this phase is spatially associated with the dark matter halos. We emphasise again that in the `None` case it is not possible for gas particles to exchange metals directly and consequently, the metal distributions only change via direct enrichment, or if the gas from the ISM reaches the phase under consideration via winds or via gas cooling from the intergalactic medium. We see that the `None` case shows a similar distribution to both the ISM and the CGM, albeit with a slightly broader low-metallicity tail. Like the $z = 2$ CGM distributions, the `D-` MDFs for the HHG show a bi-modality with peaks at $[Z] \approx -5.0$ and $[Z] \approx -1.0$ and a valley between $-4.0 < [Z] < -2.0$. Unlike the CGM MDF, the HHG low-metallicity peak is broader and it is not due to gas transitioning from the HHG phase, but rather it is due to inflowing, low-metallicity gas mixing with the already enriched gas in the halos. The `S-` cases share the peak at $[Z] \approx -1.0$, but the distribution is flat between $-6.0 < [Z] < -2.0$, before dropping-off toward low metallicities, in lockstep with the `D-` results. The gap that is apparent in the `D-` cases has disappeared, indicating that in the `S-` cases there is much more gas with metallicities in the range $-4.0 < [Z] < -2.0$ than in the `D-` cases, which is not surprising given the over-diffusive nature of the constant-coefficient Smagorinsky model.

At $z = 1$ in the HHG, the `None` case remains unchanged except for a slight increase in the amount of gas near $[Z] \approx -3.0$, and the tail has extended slightly toward lower metallicity. The latter is due to less enriched gas from the WHIM and DIFF phases accreting onto halos, diluting the metal distribution, and this dilution continues through to $z = 0$. In the `D-` and `S-` cases, the low metallicity feature (peak/shelf and extended tail) has shifted slightly toward higher metallicity as turbulent mixing redistributes the metals from the highly enriched particles.

Next, we examine the HHG at $z = 0$. The enrichment level, in the `None` case, is decreasing as evidenced by the tail of the MDF due to less enriched gas from the WHIM and DIFF phases accreting onto halos, diluting the metal distribution. In the `D-` and `S-` cases, there are coincident peaks in the MDFs at $[Z] \approx -6.0$, although the `D-` cases show more gas at lower metallicities.

Comparing across phases, the MDFs of the CGM gas are more sensitive to non-zero turbulent mixing strength than the ISM or the HHG. For instance, even though the bi-modalities at $z = 2$ in the CGM MDFs, in the `S-` and `D-` cases, disappear by

$z = 0$ due to the CGM to WHIM transition, the dynamic model results in a residual extended tail in the CGM MDF as metals mix throughout the medium, whereas the constant-coefficient Smagorinsky model tightens the distribution in this phase and the gas metallicities rapidly approach the mean. The peak in the CGM moves toward higher metallicities as gas flows between phases (CGM to WHIM).

The consequences of not including sub-grid turbulent mixing, regardless of the diffusivity, are clear – complex structure in the MDFs is highly dependent on the mixing strength. This is in contrast to Su et al. (2017) and Escala et al. (2018), who found that turbulent metal mixing strength had low-level effects in their simulated ISM. We posit that the low level effects were due the authours use of a constant value of C_s, rather than localising the mixing strength to the appropriate regions. We do not see the aforementioned trends in the `None` case because it is not possible for gas particles to exchange metals in our simulations. Our stellar feedback model drives a decoupled wind from the ISM and that wind samples the MDF in the ISM, building up a similar metal distribution in the other phases that cannot change over time, unless material flows between the phases or due to delayed SNIa and AGB stars, via the kernel-weighted enrichment procedure.

One important difference between the model we present and realistic environments is that our winds do not mix as they free-stream out of the galaxies. While the coupling and mixing strength of galactic winds is uncertain, the wind could redistribute thermal energy, momentum, and metals internally as a cohesive unit, even if they do not couple strongly to the surrounding medium (Huang & Katz, private communication). In our model, once the winds reach the criteria for recoupling in our simulations, they are free to mix their fluid properties and this, subsequently, allows for diverse MDFs in the gas phases exterior to the ISM.

We do not investigate the details of individual galaxies here but note that the differences outlined above will impact the gas-phase and stellar metallicities of those systems. Therefore, including the dynamic model with a more accurate estimation of C_s is necessary, moving forward, in order to capture the physical redistribution of metals in galaxy evolution.

2.1.6 Conclusions

All hydrodynamical methods that are used to investigate galaxy evolution, whether Lagrangian or Eulerian, require additional sub-grid thermal energy and momentum

diffusion terms in order to account for sub-grid turbulence. In Lagrangian methods, such as mesh-free finite mass (MFM) and smoothed particle hydrodynamics (SPH), metal diffusion is also required due to the inability of fluid elements to exchange metals by construction. Most implementations use the constant-coefficient Smagorinsky model – one that has been shown to be over-diffusive in almost all cases, especially laminar shear flows.

We implemented and investigated the impact of the localised dynamic Smagorinsky model on global gas-phase properties in a series of numerical experiments using the `GIZMO` code. In the dynamic case, the model coefficient depends on the local turbulent flow conditions, hence on the spatio-temporal coordinates. This is in contrast to the constant-coefficient Smagorinsky model where diffusivities depend directly on the magnitude of the velocity shear in the fluid. Compared to the constant-coefficient Smagorinsky model, the dynamic model has been shown to produce more accurate representations of fluid mechanical experiments (Kleissl et al., 2006; Kirkpatrick et al., 2006; Khani & Waite, 2015; Benhamadouche et al., 2017; Lee & Cant, 2017; Kara & ÇaÄ§lar, 2018; Taghinia et al., 2018). While we focused on cosmological experiments, the dynamic model has applications to any numerical experiment involving turbulent astrophysical flows, including stellar interiors, planetary formation, and star formation. Moreover, the method we describe in this paper, following Germano et al. (1991) and Piomelli & Liu (1995), is general and not only limited to Lagrangian hydrodynamics, but also applicable to the Eulerian cases (see Schmidt 2015).

For the MFM method, we showed that the dynamic model improves the density contrast in subsonic turbulence, allowing higher and lower density regions at fixed mass resolution. In an idealised Keplerian disc, an example of a laminar shear flow where the Smagorinsky model is known to be over-diffusive from basic analytic arguments, the dynamic model produced near-zero values of turbulent diffusivity. When we included thermal energy and momentum diffusion, the lower diffusivities prevented the rapid break-up of the disc due to excessive angular momentum transport. We observed similar minimised diffusivities in a Kelvin-Helmholtz instability experiment, where the constant-coefficient Smagorinsky model smoothed, and rapidly diffused, our metal tracer, whereas the dynamic model captured the fine level of mixing at the interface of the two fluids.

We also investigated an isolated, Milky Way-like galaxy in order to test the dynamic model in a more complex, but still controlled, environment. The dynamic model in combination with momentum diffusion improved the stability of the gaseous

disc compared to the constant-coefficient Smagorinsky model, and affected the spatial metal distributions, as indicated by the metal distribution functions (MDFs), as shown in Fig. 2.7. Rapid star formation early in the evolution of the disc leads to a higher diffusivity of thermal energy, momentum, and metals, and the subsequent exponential decay of star formation lowers the diffusivity in the dynamic model significantly. This results in a broader MDF in the ISM in the dynamic case, pointing toward less mixing in the ISM. When using the constant-coefficient Smagorinsky model, the diffusivity remained high throughout the evolution of the disc because of its strong dependence on the fluid velocity shear. We found similar variations in the circumgalactic medium (CGM) for both the dynamic and constant-coefficient Smagorinsky models that we attribute to the turbulence generated from stellar feedback increasing the diffusivity in both cases.

We also examined the global gas enrichment fractions in a set of cosmological simulations. Global gas enrichment fractions are important for the formation of Population III stars and supermassive black holes because they are theorised to be sensitive to the metal content in the gas out of which they form (see Section 2.1.1 for more details). We found that the dynamic model lowers overall enrichment compared to the standard Smagorinsky model, and that it maximally impacts metallicities in the range $-5.0 < [Z] < -3.0$. This is precisely the metallicity regime that constrains the formation sites of supermassive black holes and Population III stars (Volonteri, 2010; Sarmento et al., 2016). Specifically, the dynamic model increases the amount of gas above $[Z] > -5.0$ while maintaining the same enriched fraction of gas above $[Z] > -3.0$, compared to the no-mixing case. The standard Smagorinsky model increased the enriched fraction at all metal thresholds and in all gas phases.

In our cosmological simulations without turbulent mixing, we found that each gas phase external to the ISM has a qualitatively similar MDF to the ISM itself. Turbulent mixing allows for regions to mix their metals, and introduces additional structure in MDFs of each phase. We found that the diffusivity had a significant impact on the MDFs of the ISM and CGM – the dynamic model shows broader MDFs in both phases at $z = 0$. In these regions, we found a bi-modality in the CGM at $z = 2$ which disappeared by $z = 0$ in both cases, yet more lower metallicity gas remained in the dynamic case. Our broad density and temperature phase-space criteria led to the bi-modality, as we found low-temperature and dense gas in the cores of the cosmic filaments at $z \sim 2$. These spatial regions were of lower metallicity, and eventually return to the WHIM phase by $z = 0$. The peaks of the bi-modality, however, depend

on the diffusivity: the dynamic model produced more metal poor (by several orders of magnitude) gas than the constant-coefficient Smagorinsky model.

Finally, we briefly touch on our conclusions for SPH. Most authors apply the constant-coefficient Smagorinsky model to SPH (Wadsley et al., 2008; Shen et al., 2010, 2013; Williamson et al., 2016; Tremmel et al., 2017; Wadsley et al., 2017) and only include thermal energy and metal mixing. In reality, there are additional turbulent transport terms that are unique to SPH (see Di Mascio et al. 2017 for an introduction and derivation) that must be included[20]. Introducing momentum diffusion (e.g. via turbulent mixing as in `D-uvz` and `S-uvz`) in SPH is problematic because of unknown interactions with artificial viscosity. Additionally, by construction, the smoothing kernel in SPH acts to produce coherent flows rather than fine structure observed in mesh-free or grid methods. When we introduced momentum diffusion into SPH, the results from all of the experiments in this study were amplified when compared to the MFM method, but the qualitative trends remained. Specifically, in our cosmological experiments, we found that momentum diffusion with the constant-coefficient model causes a delay in the formation of the interstellar medium (ISM) by ~ 1 Gyr, compared to the `None` case. The dynamic model reduces the delay, but not to the no-mixing case. We attribute this to momentum diffusion and dissipation causing the Jeans mass to increase resulting in the damping of mass fluctuations.

While we note that the dynamic turbulent mixing model introduced here is a step forward in understanding the redistribution of fluid properties in Lagrangian codes, there are caveats that must be explored. The dynamic model predicts the correct behaviour in supersonic flows, but we did not include compressive mixing terms into our equations of motion. It may be that compressive sub-grid mixing models further improve super-sonic turbulence in the MFM method. Also, we justified using the Smagorinsky model by assuming that the local equilibrium condition holds, where the kinetic energy transfer rate down the turbulent cascade is equal on all scales. While the assumption is approximately true on average in the regimes we investigated (Schmidt et al., 2016), a fully-consistent turbulence model involves tracking the sub-grid kinetic energy via an additional transport equation that includes all of the necessary, higher-order, sub-grid scale terms (Schmidt, 2015). However, the dynamic model mitigates the issue by inherently calculating the deviations from local equilibrium. Furthermore, the approximations for filtering the fluid fields require care

[20]It is possible to apply the dynamic model to the transport terms in Di Mascio et al. (2017), further improving upon their work.

and attention. While Monaghan's filtering approximation (equation 2.19) holds on the singly-filtered quantities for variations on scales larger than the resolution scale h, doubly-filtered quantities may be over- or under-smoothed. A more robust, efficient, filtering procedure will need to be derived specifically for Lagrangian methods in the highly-compressible case, for filtering scales larger than h.

In summary, the dynamic Smagorinsky model localises the strength of turbulent mixing to only turbulent regions of the flow. This provides a turbulent mixing model that does not rely on pre-calibrated parameters – therefore simultaneously allowing near-zero diffusion in laminar shear flows and the expected diffusion in turbulent flows. The physical experiments to which we subjected the model show that the dynamic model significantly alters the MDFs of the ISM and CGM in a global sense. In future work we will examine the extent of small-scale differences associated with dynamic diffusion, and its impact on galaxy properties.

2.1.7 Acknowledgements

This research was enabled in part by support provided by WestGrid and Compute/Calcul Canada. DR and AB acknowledge support from NSERC (Canada) through the Discovery Grant program. DR thanks the organisers of the *Computing the Universe: At the Intersection of Computer Science and Cosmology* conference in Oaxaca, Mexico for an invited talk, and also James Wadsley and Andrey Kravtsov for their recommendations at the conference that led to the further refinement of this research project. DR also thanks Valentin Perret for the `DICE` code, and Fabrice Durier, Ondrea Clarkson, Austin Davis, and Maan Hani for many useful discussions during the course of this research. Support for PFH was provided by an Alfred P. Sloan Research Fellowship, NSF Collaborative Research Grant #1715847 and CAREER grant #1455342, and NASA grants NNX15AT06G, JPL 1589742, 17-ATP17-0214. We would also especially like to thank our referee, Wolfram Schmidt, for his contributions in improving the final version of this study.

2.2 Mixing matters

Originally published in Monthy Notices of the Royal Astronomical Society: 2021, vol. 506, issue 2, pp 2836-2852. Reproduced by permission of Oxford University Press. https://doi.org/10.1093/mnras/stab1813

Authour: Douglas Rennehan[21]

Personal contribution clarification: I am the sole contributor to this work.

Abstract: All hydrodynamical simulations of turbulent astrophysical phenomena require sub-grid scale models to properly treat energy dissipation and metal mixing. We present the first implementation and application of an anisotropic eddy viscosity and metal mixing model in Lagrangian astrophysical simulations, including a dynamic procedure for the model parameter. We compare these two models directly to the common Smagorinsky and dynamic variant. Using the mesh-free finite mass method as an example, we show that the anisotropic model is best able to reproduce the proper Kolmogorov inertial range scaling in homogeneous, isotropic turbulence. Additionally, we provide a method to calibrate the metal mixing rate that ensures numerical convergence. In our first application to cosmological simulations, we find that all models strongly impact the early evolution of galaxies leading to differences in enrichment and thermodynamic histories. The anisotropic model has the strongest impact, with little difference between the dynamic and the constant-coefficient variant. We also find that the metal distribution functions in the circumgalactic gas are significantly tighter at all redshifts, with the anisotropic model providing the tightest distributions. This is contrary to a recent study that found metal mixing to be relatively unimportant on cosmological scales. In all of our experiments the constant-coefficient Smagorinsky and anisotropic models rivaled their dynamic counterparts, suggesting that the computationally inexpensive constant-coefficient models are viable alternatives in cosmological contexts.

2.2.1 Introduction

Galaxies form and evolve in tempestuous gaseous environments where hydrodynamics, radiative cooling, and gravity synergize to produce rich emergent phenomena on a myriad of spatial scales. The immense dynamic range of scales involved and their interconnectedness prove to be limiting factors in advancing our understanding of the complete picture of galaxy evolution (see Naab & Ostriker 2017 for an excellent review). At the forefront of the issue is hydrodynamical turbulence as it is a multi-scale, non-linear phenomenon that occurs in almost all galactic environments

[21]Department of Physics & Astronomy, University of Victoria, BC V8P 5C2, Canada

— directly impacting our theoretical understanding of galaxy evolution.

While the importance of turbulence in the interstellar medium of galaxies has been long recognized (see Elmegreen & Scalo 2004 and Scalo & Elmegreen 2004 for reviews), only recently has the role of turbulence in halo gas come under careful consideration. Indeed, both the circumgalactic medium (CGM) of L^* galaxies and the intracluster medium (ICM) of groups and clusters of galaxies show signs of turbulence playing an important role in their evolution (see, for example, Prasad et al. 2018 and Wang et al. 2020). Observationally, there is evidence of complex kinematic structure in the CGM of L^*-galaxies that emerged through the revolutionary *Cosmic Origins Spectrograph* halo survey on the Hubble space telescope (*COS*-halos; Tumlinson et al. 2013, 2017). For instance, Werk et al. (2016) found that turbulent velocities of 50-75 $\mathrm{km\,s^{-1}}$ explain the broadening of absorption lines in the CGM that is not otherwise explainable that has subsequently been confirmed using numerical studies (Buie et al., 2020). Moving up in mass scale, the ICM also shows evidence of turbulence through indirect observational methods such as X-ray surface brightness and Sunyaev-Zeldovich fluctuations (Zhuravleva et al., 2014; Pinto et al., 2015; Zhuravleva et al., 2015; Khatri & Gaspari, 2016; Zhuravleva et al., 2018).

There are two main drivers of turbulence in galactic environments: (a) global outflows that emerge from star formation processes and supermassive black holes (SMBHs) within galaxies (Prasad et al., 2015, 2018; Karen Yang & Reynolds, 2016; Bourne & Sijacki, 2017; Fielding et al., 2017, 2018; Sokołowska et al., 2018; Li et al., 2020) and (b) shearing motion driven by gas in-fall during structure formation (Dekel et al., 2009; Vazza et al., 2010, 2012, 2017; Wittor et al., 2017; Bennett & Sijacki, 2020), mergers (ZuHone et al., 2013), and ram-pressure stripping of galaxies moving through the ICM (Ruggiero & Lima Neto, 2017; Simons et al., 2020). In both the CGM and ICM, turbulence could provide additional pressure support (Poole et al., 2006; Vazza et al., 2018; Lochhaas et al., 2020) that prevents the gas from collapsing and rapidly converting into stars as well as a physical mechanism to transport energy and metals throughout gas directly – impacting the cooling profile, star formation cycle, and metal distribution functions (Shen et al., 2010, 2012, 2013; Brook et al., 2014; Sokołowska et al., 2018; Escala et al., 2018; Tremmel et al., 2019; Rennehan et al., 2019; Hafen et al., 2019, 2020). Therefore, understanding the nature of turbulence is imperative to understand the complete picture of galaxy evolution.

While there are many successful cosmological simulations that use a variety of sub-grid assumptions to broadly reproduce galaxy populations (e.g. Guedes et al.

2011; Hopkins et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Genel et al. 2014; Davé et al. 2017, 2019; Tremmel et al. 2019; Pillepich et al. 2018; Huang et al. 2020), one aspect that is often overlooked is the numerical modelling of sub-grid turbulence. The crux of the problem is in the fact that in hydrodynamical simulations, the physical dissipation scale η is almost always much smaller than the resolution scale, $\eta \ll h$ (Pope, 2000). For that reason, the kinetic energy that is flowing in the turbulent cascade reaches some scale $H \gtrsim h$ where it may no longer progress. If the numerical viscosity of the hydrodynamical method is not sufficiently strong to thermalise the kinetic energy, there will be a build-up of kinetic energy at that scale $H \gtrsim h \gg \eta$ (Garnier et al., 2009). The kinetic energy build-up is a completely unphysical representation of turbulence and not only impacts the energetics, but the large scale flow properties such as the redistribution of metals.

Although not explicitly stated, many cosmological hydrodynamical simulation studies implicitly assume that *numerical dissipation* is sufficient to mimic sub-grid turbulence, rather than modelling sub-grid turbulence with additional terms in the hydrodynamical equations of motion. However, numerical dissipation is not sufficient to reduce the kinetic energy build-up and may not represent turbulent flow statistics in all cases (Sagaut, 2006). Indeed, Lecoanet et al. (2016) showed that numerical noise at the resolution scale seeds instability in the Kelvin-Helmholtz experiment that causes the long-term evolution to be unconverged as the small-scale instabilities grow. When they introduced explicit sub-grid diffusion to their simulations, the results of the Kelvin-Helmholtz experiment converged – showing that more small-scale structure (i.e. resolution-scale noise) is not necessarily better, and that explicit sub-grid diffusion can cause *less* large-scale mixing.

In simulations that use Eulerian (i.e. grid-based) hydrodynamics, there has been extensive effort to developing models for sub-grid turbulence (see Sagaut 2006; Garnier et al. 2009 and Schmidt 2015 for extensive reference lists). For astrophysically-relevant work, we refer the reader to Scannapieco & Brüggen (2008), Pan et al. (2013), Federrath (2013), Schmidt et al. (2014), and Semenov et al. (2016b) (for a review, see Schmidt 2015). In this paper, we focus on the Lagrangian hydrodynamical methods.

In Lagrangian hydrodynamics, the fluid equations are approximated via fluid elements that move with the flow. There are three main approaches that are most common in cosmological simulations: smoothed particle hydrodynamics (SPH) (Gingold & Monaghan, 1977; Lucy, 1977; Hernquist & Katz, 1989; Hopkins, 2013), the moving-mesh method (MM) (Springel, 2010), and mesh-free methods (MF) (Lanson

& Vila, 2008a,b; Gaburov & Nitadori, 2011; Hopkins, 2015). Each of these methods track fluid elements using different discretisation techniques that lead to different levels of numerical dissipation. In SPH, there is no inherent numerical dissipation and it, counter-intuitively, produces a deficit of kinetic energy near the resolution scale rather than a build-up (Bauer & Springel, 2012; Price, 2012b). The MM and MF methods use Riemann solvers to approximate the fluid equations of motion between neighbouring fluid elements. Riemann solvers are generally diffusive due to their approximate nature, and the build-up of kinetic energy is present in both methods on scales up to ~ 10 times the resolution scale (Bauer & Springel, 2012; Hopkins, 2015).

A solution to kinetic energy build-up is to model the action of turbulent eddies as a viscous process that diffuses momentum (and metals) and dissipates kinetic energy via a diffusion equation and source term in the energy equation, respectively. Usually the assumption is that the viscosity, or diffusivity, depends on velocity fluctuations $v_{\rm eddy}$ near the resolution scale h. The resulting diffusivity is $D \propto h v_{\rm eddy}$, where $v_{\rm eddy}$ may be a characteristic velocity, or velocity difference, within the neighbourhood of a fluid element (Wadsley et al., 2008; Greif et al., 2009).

One particularly important choice of $v_{\rm eddy}$ is the Smagorinsky model (Smagorinsky, 1963), which assumes that velocity shear fluctuations drive dissipation and mixing through $v_{\rm eddy} \sim h|S^*|$, where $|S^*|$ is the magnitude of the trace-free shear tensor. The Smagorinsky model has been successfully used to treat metal and thermal energy mixing in SPH (Wadsley et al., 2008; Shen et al., 2010, 2012; Brook et al., 2014; Williamson et al., 2016; Tremmel et al., 2017; Wadsley et al., 2017; Su et al., 2017) and has been extended to other Lagrangian hydrodynamical methods, such as the MFM method (Colbrook et al., 2017; Escala et al., 2018; Rennehan et al., 2019; Hafen et al., 2019, 2020). While the Smagorinsky model improves mixing in Lagrangian simulations, it is important to consider two assumptions in the model: (a) that shear fluctuations (i.e., changes in $|S^*|$) always represent turbulence and (b) that the diffusive process acts isotropically through the magnitude of the trace-free (i.e. ignoring compression) shear tensor over the scale h. Are either of these assumptions reasonable?

To address point (a), Piomelli & Liu (1995) proposed a method of dynamically calculating the model coefficient at simulation time. That method was employed in Rennehan et al. (2019) for the first time in Lagrangian hydrodynamics and they found that the model significantly reduced over-diffusion in non-turbulent shear flows, such as in rotating galactic disks and the Kelvin-Helmholtz instability.

The second point (b) concerns the isotropy in $\boldsymbol{D}$ and the discounting of compression in $|S^*|$. Recently Hu & Chiang (2020) showed that a better representation of sub-grid scale turbulence is obtained by treating the full velocity tensor $\boldsymbol{\nabla} \otimes \boldsymbol{u}$ for the diffusivity $\boldsymbol{D}$, now a tensor. The model is the *gradient* model (Clark et al., 1979) and is a major change since diffusion now depends on the directionality encoded in $\boldsymbol{\nabla} \otimes \boldsymbol{u}$ rather than equally in every spatial direction. The trace of $\boldsymbol{\nabla} \otimes \boldsymbol{u}$ is automatically included and, therefore, compression is automatically handled[22] – an important point for highly-compressible turbulence in cosmological flows. However, Hu & Chiang (2020) post-processed their driven turbulence simulations to check if the model *would have* improved the results at simulation time. Our goal is to implement the model for the mesh-free finite mass method, and determine its feasibility *at simulation time* in combination with the dynamic procedure from Rennehan et al. (2019).

We introduce an implementation of the gradient model for Lagrangian astrophysical simulations and additionally provide methods for computing the model coefficient at simulation time. In Section 2.2.2 we provide a derivation of the model, as well as a derivation of the dynamic procedure in Section 2.2.2 that allows calculation of the model coefficient at simulation time. Section 2.2.3 describes driven turbulence validation tests of the gradient model, included at run time, for both eddy viscosity and metal mixing. As a first application, we describe the qualitative impact of the eddy viscosity and metal mixing model on cosmological gas phases in Section 2.2.4. We present our conclusions and recommendations in Section 2.2.5.

2.2.2 The gradient model

In finite-mass Lagrangian hydrodynamics, such as the mesh-free finite mass (MFM) method (Lanson & Vila, 2008a,b; Gaburov & Nitadori, 2011; Hopkins, 2015), the build-up of kinetic energy at the resolution scale demands an additional dissipation mechanism. Additionally, metals follow the fluid mass elements throughout the simulation volume and, therefore, the exchange of metals between fluid elements due to sub-grid scale turbulent motion does not occur. The crux of the issue is that discretisation of the fluid field leads to damping out of the high-frequency turbulent fluctuations that should continue down to the *physical* dissipation scale. It is useful to think of the damping action as a high-pass filter acting on the fluid equations of motion. By applying a general filter to the momentum conservation equation, it is

[22] The trace of $\boldsymbol{\nabla} \otimes \boldsymbol{u}$ is $\boldsymbol{\nabla} \cdot \boldsymbol{u}$, the divergence of the flow.

possible to derive the correct level of mixing that should occur between fluid elements due to unresolved turbulence.

In general, the filtering action over a scalar field $f_i(\boldsymbol{r})$ can be represented as,

$$\overline{f}_i(\boldsymbol{r}) = \int_D f_i(\boldsymbol{r}')G(|\boldsymbol{r} - \boldsymbol{r}'|, h)d\boldsymbol{r}', \tag{2.37}$$

where h is the smoothing scale over the domain $\boldsymbol{D}$. We apply this to the momentum equation to determine the correction terms due to unresolved turbulence,

$$\frac{\partial(\rho \boldsymbol{u})}{\partial t} + \nabla \cdot (\rho \boldsymbol{u} \otimes \boldsymbol{u} + P\boldsymbol{I}) = 0, \tag{2.38}$$

where ρ is the gas density, $\boldsymbol{P}$ is the pressure, and $\boldsymbol{u}$ is the velocity vector. If we apply equation (2.37) to equation (2.38) it follows that, assuming the filtering operation and derivatives commute,

$$\frac{\partial(\overline{\rho \boldsymbol{u}})}{\partial t} + \nabla \cdot (\overline{\rho \boldsymbol{u} \otimes \boldsymbol{u}} + \overline{P}\boldsymbol{I}) = 0. \tag{2.39}$$

For simplicity, we switch to density-weighted variables such that $\widetilde{\boldsymbol{u}} \equiv \overline{\rho \boldsymbol{u}}/\overline{\rho}$ and

$$\frac{\partial(\overline{\rho}\tilde{\boldsymbol{u}})}{\partial t} + \nabla \cdot (\overline{\rho}\widetilde{\boldsymbol{u} \otimes \boldsymbol{u}} + \overline{P}\boldsymbol{I}) = 0. \tag{2.40}$$

The term $\widetilde{\boldsymbol{u} \otimes \boldsymbol{u}}$ is unknown at simulation time because it relies on information below the resolution scale. To put the equation in a more manageable form, we add $\nabla \cdot (\overline{\rho}[\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}} - \widetilde{\boldsymbol{u} \otimes \boldsymbol{u}}])$ and rearrange,

$$\frac{\partial(\overline{\rho}\widetilde{\boldsymbol{u}})}{\partial t} + \nabla \cdot (\overline{\rho}\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}} + \overline{P}\boldsymbol{I}) = \nabla \cdot (\overline{\rho}[\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}} - \widetilde{\boldsymbol{u} \otimes \boldsymbol{u}}]). \tag{2.41}$$

Therefore, we define the sub-grid scalar flux $\boldsymbol{F}$,

$$\boldsymbol{F} \equiv \overline{\rho}(\widetilde{\boldsymbol{u} \otimes \boldsymbol{u}} - \widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}}), \tag{2.42}$$

and retrieve a new equation,

$$\frac{\partial(\overline{\rho}\widetilde{\boldsymbol{u}})}{\partial t} + \nabla \cdot (\overline{\rho}\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}} + \overline{P}\boldsymbol{I}) = -\nabla \cdot \boldsymbol{F}. \tag{2.43}$$

The sub-grid scale momentum flux $\boldsymbol{F}$ is unknown at simulation time and must be modelled yet is widely ignored in cosmological simulation studies which usually focus only on the thermal energy and metal fluxes via the Smagorinsky model.

There are a myriad of models in the literature for $\boldsymbol{F}$ (see Garnier et al. 2009 for extensive lists) but here we take the direct approach of using a Taylor series approximation following Clark et al. (1979) and Hu & Chiang (2020). We expand[23] $\widetilde{\boldsymbol{u}\otimes\boldsymbol{u}}$ via a Taylor expansion as

$$\widetilde{\boldsymbol{u}\otimes\boldsymbol{u}} \approx \widetilde{\boldsymbol{u}}\otimes\widetilde{\boldsymbol{u}} + \epsilon\nabla^2(\widetilde{\boldsymbol{u}}\otimes\widetilde{\boldsymbol{u}}) \tag{2.44}$$

and, therefore, the flux becomes

$$\boldsymbol{F} = \overline{\rho}(\widetilde{\boldsymbol{u}\otimes\boldsymbol{u}} - \widetilde{\boldsymbol{u}}\otimes\widetilde{\boldsymbol{u}}) \approx \overline{\rho}\epsilon\nabla^2(\widetilde{\boldsymbol{u}}\otimes\widetilde{\boldsymbol{u}}), \tag{2.45}$$

where ϵ is a constant that depends on the kernel scale h as $\epsilon \propto h^2$ (Monaghan, 1989, 2002, 2011). Expanding the term on the right hand side of equation (2.45) and keeping only the first derivative terms we find

$$\boldsymbol{F} = 2\overline{\rho}Ch^2(\nabla\otimes\widetilde{\boldsymbol{u}})(\nabla\otimes\widetilde{\boldsymbol{u}})^{\mathrm{T}}, \tag{2.46}$$

where C is our model parameter.

A similar result emerges when considering the mass flux of metals in a fluid,

$$\frac{\partial(\rho Z)}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}Z) = 0. \tag{2.47}$$

Applying the same approach as before gives an equation for the sub-grid flux of metals,

$$\boldsymbol{F} = 2\overline{\rho}C_Z h^2(\nabla\otimes\widetilde{\boldsymbol{u}})\cdot\nabla Z \tag{2.48}$$

The method for solving equation (2.43) is detailed in Hopkins (2017), and we point the reader to that work for further information. From their equation (2),

$$\boldsymbol{F} = \boldsymbol{K}\cdot(\nabla\otimes\boldsymbol{q}), \tag{2.49}$$

where $\boldsymbol{K}$ is the tensor describing the diffusive strength, and $\boldsymbol{q}$ is the fluid field property. Therefore, we identify,

$$\boldsymbol{K} \equiv 2\overline{\rho}Ch^2(\nabla\otimes\widetilde{\boldsymbol{u}}), \tag{2.50}$$

$$\boldsymbol{q} \equiv \widetilde{\boldsymbol{u}}. \tag{2.51}$$

[23] See Appendix 2.2.8 for a full derivation.

It is important to note that following Section 3.0.6 of Hopkins (2017), we also include the dissipation term corresponding to $\boldsymbol{\nabla} \cdot \boldsymbol{F}$ in the energy flux to ensure energy conservation.

The model in equation (2.46) is known to lead to numerical instability due to particles attracting rather than repelling (Nomura & Post, 1998; Balarac et al., 2013), similar to the well-studied tensile instability in smoothed particle magnetohydrodynamics (Phillips & Monaghan, 1985; Morris, 1996; Monaghan, 2000; Price, 2012a). Balarac et al. (2013) specifically showed for the anisotropic eddy viscosity model that ignoring the positive eigenvalues of $\boldsymbol{S}$ ensures the model is always well behaved and that the action of the model is to repel particles rather than attract. Therefore, we follow Balarac et al. (2013) and only keep the contribution due to the negative eigenvalues of the shear tensor. However, we must first decompose $\nabla \otimes \widetilde{\boldsymbol{u}}$ into the symmetric and anti-symmetric parts,

$$\nabla \otimes \widetilde{\boldsymbol{u}} = \frac{1}{2}(\nabla \otimes \widetilde{\boldsymbol{u}} + [\nabla \otimes \widetilde{\boldsymbol{u}}]^{\mathrm{T}}) + \frac{1}{2}(\nabla \otimes \widetilde{\boldsymbol{u}} - [\nabla \otimes \widetilde{\boldsymbol{u}}]^{\mathrm{T}}) \equiv \boldsymbol{S} + \boldsymbol{\Omega}. \tag{2.52}$$

We further decompose $\boldsymbol{S}$ into the contribution from the positive eigenvalues and negative eigenvalues as $\boldsymbol{S} \equiv \boldsymbol{S}_{\oplus} + \boldsymbol{S}_{\ominus}$. It then follows that

$$\boldsymbol{S}_{\ominus} \equiv \sum_{k=1}^{3} \min(0, \lambda^{(k)}) \boldsymbol{e}^{(k)} \otimes \boldsymbol{e}^{(k)}, \tag{2.53}$$

where $\lambda^{(k)}$ is the kth eigenvalue and $\boldsymbol{e}^{(k)}$ is the corresponding eigenvector. Therefore, the new diffusion coefficient tensor is

$$\boldsymbol{K} = 2\overline{\rho} C h^2 \boldsymbol{S}_{\ominus}, \tag{2.54}$$

where we have reduced the shear contribution into $\boldsymbol{S}_{\ominus}$. Removing the negative eigenvalues appears to not be necessary for the metal mixing case as we found no cases of numerical instability in all of our hydrodynamical tests and cosmological simulations. However, we find that it is absolutely necessary in cosmological simulations for the momentum flux.

It is important to note that our choice of h differs from that in Hopkins et al. (2018). We choose h as the kernel radius of compact support as this is the maximum interaction distance for the flux in the MFM method. Physically, it is the maximum distance over which sub-grid eddies transport their fluid properties. We find much better results in our tests in Section 2.2.3 using the maximum interaction distance.

However, we do note that the radius of compact support is different for each kernel and may not be the most accurate length-scale. Given that it is normally twice the smoothing scale for the kernel (Dehnen & Aly, 2012), it provides the best compromise. Any kernel that is used having a compact support radius larger than twice the smoothing scale should be further investigated before using our definition of h. Our value is approximately twice that found in the `FIRE` studies (Su et al., 2017; Escala et al., 2018; Hafen et al., 2019, 2020), but produces 4 times as much dissipation and metal mixing since the dependence is squared in equation (2.54). Our choice of h is the same as Wadsley et al. (2017) who use the radius of compact support for turbulent mixing in the `GASOLINE-2` code.

The dynamic gradient model

We apply the same procedure in Balarac et al. (2013) combined with the density-weighted filtering procedure in Rennehan et al. (2019). Following their notation, we replace $\tilde{f}$ with $\overline{f}$ since fluid properties are inherently density-weighted in the mesh-free finite mass method. Although we focus on the velocity fluctuations in the following procedure, we note that it applies equally to the metal field.

The resolved fluctuations in the flow are,

$$\boldsymbol{\mathcal{L}} = \widehat{\overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}} - \widehat{\overline{\boldsymbol{u}}} \otimes \widehat{\overline{\boldsymbol{u}}}, \tag{2.55}$$

where $\widehat{\overline{\boldsymbol{u}}}$ represents the velocity vector $\boldsymbol{u}$ filtered once on the resolution scale (to produce $\overline{\boldsymbol{u}}$), then filtered again on twice the resolution scale ($\widehat{h} \equiv \widehat{\overline{h}} \sim 2\overline{h}$, see Section 2.4 of Rennehan et al. 2019). Explicitly, we represent the filtering operation on any scalar quantity f from equation (2.37) as a sum (Monaghan, 1989, 2005, 2011; Rennehan et al., 2019),

$$\overline{f}_a = f_a + \epsilon \sum_b \frac{m_b}{\langle \rho_{ab} \rangle_{\overline{h}}} (f_b - f_a) W(|\boldsymbol{x_a} - \boldsymbol{x_b}|, \overline{h}_{ab}), \tag{2.56}$$

where f_a is any scalar quantity at particle a, $\overline{h}_{ab}$ is the mean smoothing lengths between the two particles, $\langle \rho_{ab} \rangle_{\overline{h}}$ is the mean of the densities of the two particles, and the sum is taken over b nearest neighbours. We take the smoothing factor $\epsilon = 0.8$ following Rennehan et al. (2019). To obtain double-filtered quantities, we apply equation (2.56) to the singly-filtered quantities,

$$\widehat{\overline{f}}_a = \overline{f}_a + \epsilon \sum_b \frac{m_b}{\langle \rho_{ab} \rangle_{\widehat{h}}} (\overline{f}_b - \overline{f}_a) W(|\boldsymbol{x_a} - \boldsymbol{x_b}|, \widehat{h}_{ab}). \tag{2.57}$$

If we use only resolved quantities (i.e. doubly-filtered) then the gradient model in equation (2.46) should reproduce $\boldsymbol{\mathcal{L}}$,

$$\boldsymbol{\mathcal{L}} = \widehat{\overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}} - \widehat{\overline{\boldsymbol{u}}} \otimes \widehat{\overline{\boldsymbol{u}}} = 2C\widehat{\overline{h}}^2 \widehat{\overline{\boldsymbol{S}}}_{\ominus} (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}})^{\mathrm{T}}. \tag{2.58}$$

The above equation results in a solution for the one unknown parameter C,

$$C = \frac{\boldsymbol{\mathcal{L}} \cdot \left[2\widehat{\overline{h}}^2 \widehat{\overline{\boldsymbol{S}}}_{\ominus} (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}})^{\mathrm{T}} \right]}{||2\widehat{\overline{h}}^2 \widehat{\overline{\boldsymbol{S}}}_{\ominus} (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}})^{\mathrm{T}}||^2} = \frac{1}{2} \frac{\boldsymbol{\mathcal{L}} \cdot \boldsymbol{\alpha}}{||\boldsymbol{\alpha}||^2}, \tag{2.59}$$

where we have defined $\boldsymbol{\alpha} \equiv \widehat{\overline{h}}^2 \widehat{\overline{\boldsymbol{S}}}_{\ominus} (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}})$. Using this determined value of C, we use equation (2.54) as the diffusivity tensor in the additional flux term from equation (2.46). Note that $\overline{h}$ is the *resolution scale* and not the radius of compact support of the kernel. The resolution scale is approximately half of the radius of compact support, or the mean interparticle spacing.

Applying the same procedure to the metal field yields a separate equation for C_{Z}. The resolved fluctuations take a similar form to equation (2.55),

$$\boldsymbol{\mathcal{L}}_{\mathrm{Z}} = \widehat{\overline{\boldsymbol{u}} \, \overline{Z}} - \widehat{\overline{\boldsymbol{u}}} \, \widehat{\overline{Z}}. \tag{2.60}$$

Note that $\boldsymbol{\mathcal{L}}$ is now a vector rather than a rank-2 tensor. Following the procedure we outline above results in an equation for C_{Z}

$$C_{\mathrm{Z}} = \frac{\boldsymbol{\mathcal{L}}_{\mathrm{Z}} \cdot \left[2\widehat{\overline{h}}^2 (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}}) \cdot \boldsymbol{\nabla} \widehat{\overline{Z}} \right]}{||2\widehat{\overline{h}}^2 (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}}) \cdot \boldsymbol{\nabla} \widehat{\overline{Z}}||^2} = \frac{1}{2} \frac{\boldsymbol{\mathcal{L}}_{\mathrm{Z}} \cdot \boldsymbol{\beta}}{||\boldsymbol{\beta}||^2}, \tag{2.61}$$

where we have defined $\boldsymbol{\beta} \equiv \widehat{\overline{h}}^2 (\boldsymbol{\nabla} \otimes \widehat{\overline{\boldsymbol{u}}}) \cdot \boldsymbol{\nabla} \widehat{\overline{Z}}$.

Comparison to the Smagorinsky model

The gradient model differs from the widely used Smagorinsky model in a subtle yet important way. Returning to the definition of the sub-grid flux in equation (2.46), the Smagorinsky model represents $\boldsymbol{F}$ as,

$$\boldsymbol{F} = 2\overline{\rho}(C_{\rm s}h)^2||\boldsymbol{S}^*||\boldsymbol{S}^*, \tag{2.62}$$

where $\boldsymbol{S}^*$ is the trace-free symmetric shear tensor[24]. The diffusivity tensor is isotropic,

$$\boldsymbol{K}_{\rm Smag} \equiv 2\overline{\rho}(C_{\rm s}h)^2||\boldsymbol{S}^*||\,\boldsymbol{I}. \tag{2.63}$$

and $||\boldsymbol{K}_{\rm Smag}|| = 2\overline{\rho}(C_{\rm s}h)^2||\boldsymbol{S}^*||$ with $C_{\rm s} \sim 0.15$. The constant nature of $C_{\rm s}$ implies that the diffusivity scales with the magnitude of the symmetric shear or, more directly, that the strength of turbulent fluctuations is only determined by fluctuations in the shear strength. That is a good assumption in purely turbulent flows but fails dramatically in laminar shear flows, where the shear is not a good indicator of the presence of turbulence.

One solution to over-diffusion in the Smagorinsky model is to dynamically calculate the coefficient $C_{\rm s}$ at simulation time based on the local fluid properties. We showed in Rennehan et al. (2019) that the dynamic procedure predicts much lower values of $C_{\rm s}$ in the majority of our simple hydrodynamical tests. However, we did not consider the impact of altering the isotropic nature of the diffusivity.

The gradient model has $\boldsymbol{K}_{\rm Grad} \propto \boldsymbol{S}_{\ominus}$ with $||\boldsymbol{K}_{\rm Grad}|| = 2\overline{\rho}Ch^2||\boldsymbol{S}_{\ominus}||$. The constant C is yet to be determined but the direction differs from the dynamic and non-dynamic Smagorinsky models. The diffusivity itself no longer acts in each direction equally but acts in the direction of the eigenvectors of $\boldsymbol{S}_{\ominus}$. However, in a simple incompressible, low Mach number turbulent flows we expect that $||\boldsymbol{K}_{\rm Smag}|| \sim ||\boldsymbol{K}_{\rm Grad}||$ given that the velocity derivative tensor $\boldsymbol{\nabla} \otimes \tilde{\boldsymbol{u}}$ is approximately isotropic in that regime.

Our application of the dynamic method (Piomelli & Liu, 1995) to the gradient model simultaneously allows $C = C(\boldsymbol{x}, t)$ and the diffusivity to be anisotropic, $\boldsymbol{K}_{\rm Grad} \propto \boldsymbol{S}_{\ominus}$. This should be important for any complicated astrophysical flows such as those we investigate in the following sections.

Models

Table 2.5 contains a compact description of our model set. In all cases where there is a sub-grid turbulence model, we treat both metals and viscosity simultaneously. There are three categories of models: no sub-grid model (`None`), the Smagorinksy model, and the gradient model. The dynamic procedure allows us to extend the Smagorinsky and

[24] $\boldsymbol{S}^* \equiv \boldsymbol{S} - \frac{1}{3}\mathrm{tr}(\boldsymbol{S}) \cdot \boldsymbol{I}$

Table 2.5: Turbulence models and parameters. The dynamic model calculates the model parameters at simulation and, therefore, the values listed below are the forced upper limit.

Label	Dynamic	(Max.) Parameter Value	Anisotropic
None	N/A	N/A	N/A
Smag.	×	0.15	×
Dyn. Smag.	✓	0.20	×
FIRE	×	0.05	×
Grad.	×	0.22	✓
Dyn. Grad.	✓	1.00	✓

gradient models with a model parameter that depends on spatio-temporal coordinates. Additionally, we test the only other calibration of the Smagorinsky model in the mesh-free finite mass method from the FIRE collaboration (Escala et al., 2018).

For the Smagorinsky model, `Smag.`, we use the theoretical value of $C_s \sim 0.15$ and limit C_s to 0.20 for the dynamic Smagorinsky model (`Dyn. Smag.`) to avoid numerical instability. The FIRE calibration of the Smagorinsky model (labelled `FIRE`) is $C_s \approx 0.046$ and we adopt $C_s = 0.05$ for simplicity. For these three versions of the Smagorinsky model the value of C_s is the same for both metals and eddy-viscosity.

In the new gradient model we use fixed values of $C = 0.22$ and $C_Z = 0.22$ for the baseline comparison and label these as `Grad`. In our other tests, we use the dynamic procedure outlined in Section 2.2.2 and label these tests as `Dyn. Grad`. We derive our fixed values of C and C_Z from the approximate median value predicted by the dynamic procedure in the driven turbulence tests in Section 2.2.3.

2.2.3 Homogeneous Turbulence

Turbulence is ubiquitous in astrophysical flows on a myriad of scales and Mach numbers. Therefore, in this section, we investigate the impact of the gradient model on the velocity statistics and metal distributions in homogeneous, isotropic, driven turbulence at Mach numbers $\mathcal{M} \in \{0.3, 0.7, 2.1\}$.

For each $\mathcal{M}$, our control (i.e., no sub-grid turbulence model) simulation set comprises 5 simulations with particle counts $N \in \{64^3, 128^3, 256^3, 512^3, 768^3\}$ within a box of side length $L = 1$ with initial pressure, density, and specific internal energy of $P = 1$, $\rho = 1$, $u = 1000$, respectively[25]. Initially, we place equal mass particles on a uniform grid and then subsequently mix the gas via the prescription of Schmidt et al. (2006), ported to particle-based simulations in Price & Federrath (2010) and subsequently later implemented in `GADGET` and `GIZMO` (Bauer & Springel, 2012; Hopkins, 2015). We list our turbulent driving parameters for `GIZMO` in Table 2.6 (cf. Table 1 in Bauer & Springel 2012). For more details of our approach, see Section 3.1 of Rennehan et al. (2019).

Our interest lies in measuring the impact of (1) the eddy-viscosity model on the velocity power spectra of these driven turbulence volumes and (2) the convergence of metal distribution functions in these volumes. However, these properties rely on driven turbulence volumes that are in statistical equilibrium. To gauge whether our

[25]We use arbitrary code units in all of our hydrodynamical tests.

Table 2.6: Parameters for our driven turbulence experiments. Units are arbitrary code units.

$\sim \mathcal{M}$	σ	Δt	t_s	$k_{\rm min}$	$k_{\rm max}$	$\tau_{\rm mix}$
0.3	0.014	0.005	1	6.27	12.57	3.33
0.7	0.045	0.005	1	6.27	12.57	1.67
2.1	0.21	0.005	0.5	6.27	12.57	0.50

simulations are in equilibrium, we define a mixing timescale $\tau_{\rm mix} \equiv L/\langle v \rangle$, where $\langle v \rangle$ is the expected average velocity of the particles in each volume. However, $\langle v \rangle = v_{\rm s}\mathcal{M}$, where $v_s = 1$ is the isothermal sound speed of the gas and, additionally, $L = 1$. Therefore, $\tau_{\rm mix} = 1/\mathcal{M}$. We evolved each control simulation for several mixing timescales ($\sim 4\tau_{\rm mix}$) to ensure that the gas is in steady-state statistical equilibrium and we confirmed the stability of the Mach number. The mixing timescale and steady-state Mach numbers are listed in Table 2.6. We measure the velocity power spectra following the same method as Bauer & Springel (2012), which is available in the public version of `GIZMO`.

We note that there are debates in the literature over the most accurate method to compute the velocity power spectrum in Lagrangian hydrodynamical methods, particularly for classic smoothed particle hydrodynamics (Bauer & Springel, 2012; Price, 2012b). The biggest issue is reproducing the correct power on the smallest scales, near the maximum resolution. Shi et al. (2013) compared several methods and showed that a second-order moving least squares method produces the least error in reproducing the correct velocity power spectra on the smallest scales. However, it is not clear how this will generalise to the MFM method where the power on smaller scales, as we discuss below, more closely resembles grid-based hydrodynamical methods. We choose to use the module available in the `GIZMO` for easy comparison to the turbulence results Bauer & Springel (2012) and Hopkins (2015), and leave comparison of the different power spectra calculation methods to future work.

It may seem out of place to study low resolutions, such as 64^3 and 128^3, in this study when it is clearly possible to study resolutions up to 768^3. The ultimate goal of our work is to apply the model to cosmological simulations which have a huge dynamic range and, therefore, low resolutions in individual galaxies. For example, IllustrisTNG (Pillepich et al., 2018) and the RomulusC (Tremmel et al., 2019) simulations both have particle mass resolutions of $\sim 10^5$ $\mathrm{M}_\odot$ for each gas particle, at their best. Consider that in an L^*-galaxy, we expect perhaps $\lesssim 10^{10}$ $\mathrm{M}_\odot$ of hot gas in the circumgalactic medium (Anderson & Bregman, 2010). At the best resolutions we have today, that gives $\sim 10^5$ particles per L^*-halo or $\sim 50^3$ particles. Evidently, contemporary cosmological simulations that capture both hundreds of Mpc on the large-scale as well as individual galaxies are far off from $\gtrsim 256^3$ per galactic halo.

Velocity power spectra

A standard measure to determine if eddy viscosity models improve the accuracy of hydrodynamical simulations is whether the velocity power spectra reproduces the theoretically predicted Kolmogorov scaling, $E(k) \sim k^{-5/3}$. That scaling holds for incompressible, low Mach number turbulence ($\mathcal{M} \lesssim 1$) but is shallower than the apparent scaling in supersonic turbulence, $E(k) \sim k^{-2}$ (Federrath, 2013). In physical turbulence, the dissipation scale is demarcated by a sharp decline from the Kolmogorov slope on the smallest scales. In simulations, the resolution scale forces dissipation to occur on a much larger scale than what would occur in nature as the physical dissipation scale is unresolved. If the numerical viscosity of a hydrodynamical method cannot rapidly dissipate that energy, there will be a build-up of kinetic energy near the resolution scale that causes an unphysical representation of turbulence.

Eddy-viscosity models introduce additional dissipation in the gas by accounting for the unrepresented scales in the flow, or equivalently by minimizing the error from the missing terms in the equations of motion. The build-up of kinetic energy is usually observed as a "bump" in the velocity power spectra where there is artificial correlation in the velocities on small scales. However, before we discuss the impact of eddy viscosity models on the power spectrum we must first test the convergence of the mesh-free finite mass (MFM) method in simulations without eddy viscosity.

Fig. 2.10 shows the velocity power spectra for our set of simulations with particle counts 64^3, 128^3, 256^3, 512^3, and 768^3 coloured by lines from lightest to darkest, respectively. The power spectra are compensated by k for easy comparison to Bauer & Springel (2012). The panels are ordered from lowest to highest Mach number from top to bottom — $\mathcal{M} \sim 0.3$, 0.7 and 2.1, respectively. In each panel, the dotted line represents the predicted scaling but at an arbitrary normalisation.

From the top panel of Fig. 2.10, it is apparent that 64^3 and 128^3 do not faithfully represent a turbulent gas as they are dominated by the bump. More precisely, the $E(k)$ scaling is much too shallow compared to Kolmogorov turbulence for a wide range of k. Our 256^3 simulation shows an inkling of the inertial range scaling but is slightly too steep below $k \lesssim 20$ and dominated by the bump at $k \gtrsim 30$. As we move up in resolution the inertial range only grows slightly. At our highest resolution, the inertial range spans $k \sim 40$ to $k \sim 60$ and the bump dominates the small scales. We skip discussion of the middle panel as the results are qualitatively equivalent between $\mathcal{M} \sim 0.3$ and 0.7.

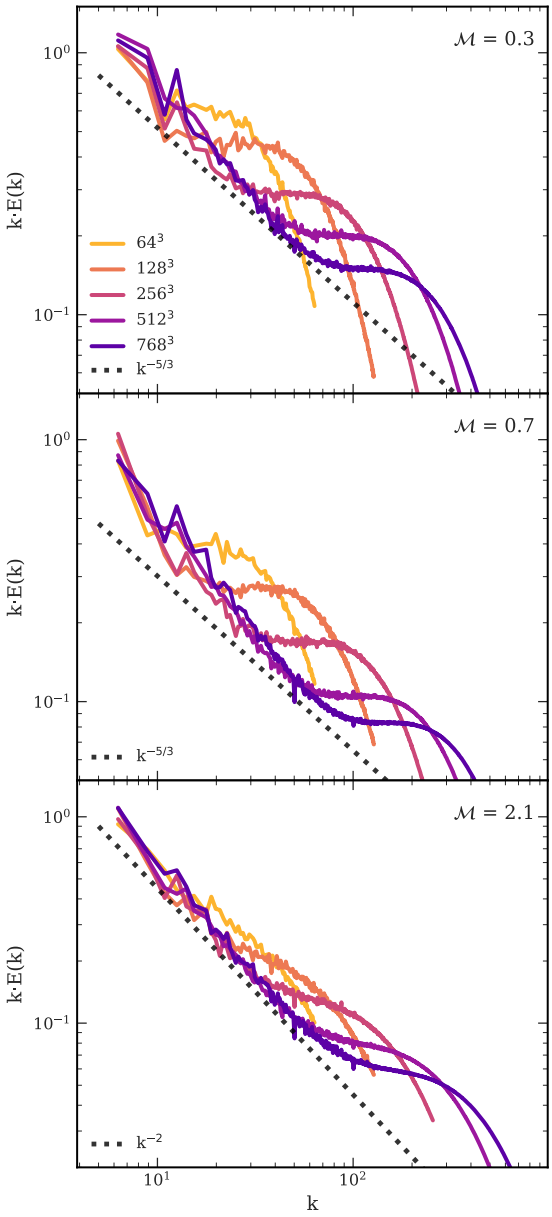

Figure 2.10: The velocity power spectra of our simulations with no sub-grid eddy viscosity model at three Mach numbers from top to bottom, respectively: $\mathcal{M} = 0.3$, 0.7, and 2.1. We compensate the power spectra by k for easy comparison with Bauer & Springel 2012. The coloured lines show resolutions 64^3, 128^3, 256^3, 512^3, and 768^3 from lightest to darkest, respectively. In each panel, we show the predicted scaling with a dotted line of arbitrary normalisation. Only in the highest Mach number case do we see many of the scales in the inertial range represented and clear convergence of the hydrodynamical method.

In the bottom panel of Fig. 2.10, we show the compensated power spectra for supersonic turbulence at $\mathcal{M} \sim 2.1$. It is immediately evident that the simulations converge much more rapidly to the proper scaling than in subsonic turbulence. Still, at 64^3 resolution the power spectra is dominated by the build-up of kinetic energy near the resolution scale, with the inertial range only beginning to appear at 128^3 resolution. At our highest resolution, 768^3, there is arguably an entire order-of-magnitude resolved in the inertial range before the build-up of kinetic energy dominates at the smallest scales.

Evidently, sub-grid eddy viscosity models are required for the mesh-free finite mass (MFM) method at any resolutions that may be used in astrophysical environments. We show the impact of our eddy viscosity models in Fig. 2.11. The left column shows the compensated power spectra, $E(k)$, as a function of wavenumber k. The rows represent the same three Mach numbers $\mathcal{M} = 0.3$, 0.7, and 2.1 from top to bottom, respectively. The dotted black line shows the Kolmogorov scaling at each Mach number. All of the simulations were run at 256^3 resolution since that is the resolution where, with no eddy viscosity model, we begin to see an extended inertial range and distinguish the kinetic energy "bump". The coloured curves show the eddy viscosity models, shaded from lightest to darkest: `Dyn. Smag` (solid salmon), `Smag.` (solid magenta), `Dyn. Grad.` (dotted purple), and `Grad.` (dashed purple). To explain the right column, we must first explain the results in the left column.

The left column of Fig. 2.11 shows the velocity power spectra of the turbulent gas in our simulated volumes with the same models presented in Section 2.2.2. The black curve shows the control experiment at 256^3 resolution, i.e. the simulation with no eddy viscosity model and only numerical dissipation as in Fig. 2.10. All Mach numbers show the same trend: the eddy viscosity models have little impact on reducing the build-up of kinetic energy at small scales. Especially important is that the sub-sonic ($\mathcal{M} \lesssim 1$) simulations are much less improved than the supersonic case. However, the new gradient model variants, `Grad.` and `Dyn. Grad.`, dissipate slightly more rapidly and allow for a steeper slope closer to the Kolmogorov scaling.

We did not expect that all of the eddy viscosity implementations would fail to reduce the kinetic energy build-up *a priori*. The fact that there is not enough dissipation suggests that some physical property in the diffusion tensor was assigned incorrectly. As we outlined in Section 2.2.2, the diffusion strength for the Smagorinsky and the gradient model classes should effectively scale with each other ($||\boldsymbol{K}_{\rm Smag}|| \sim ||\boldsymbol{K}_{\rm Grad}||$) in isotropic, homogeneous turbulence. Therefore, in both classes of models, there are

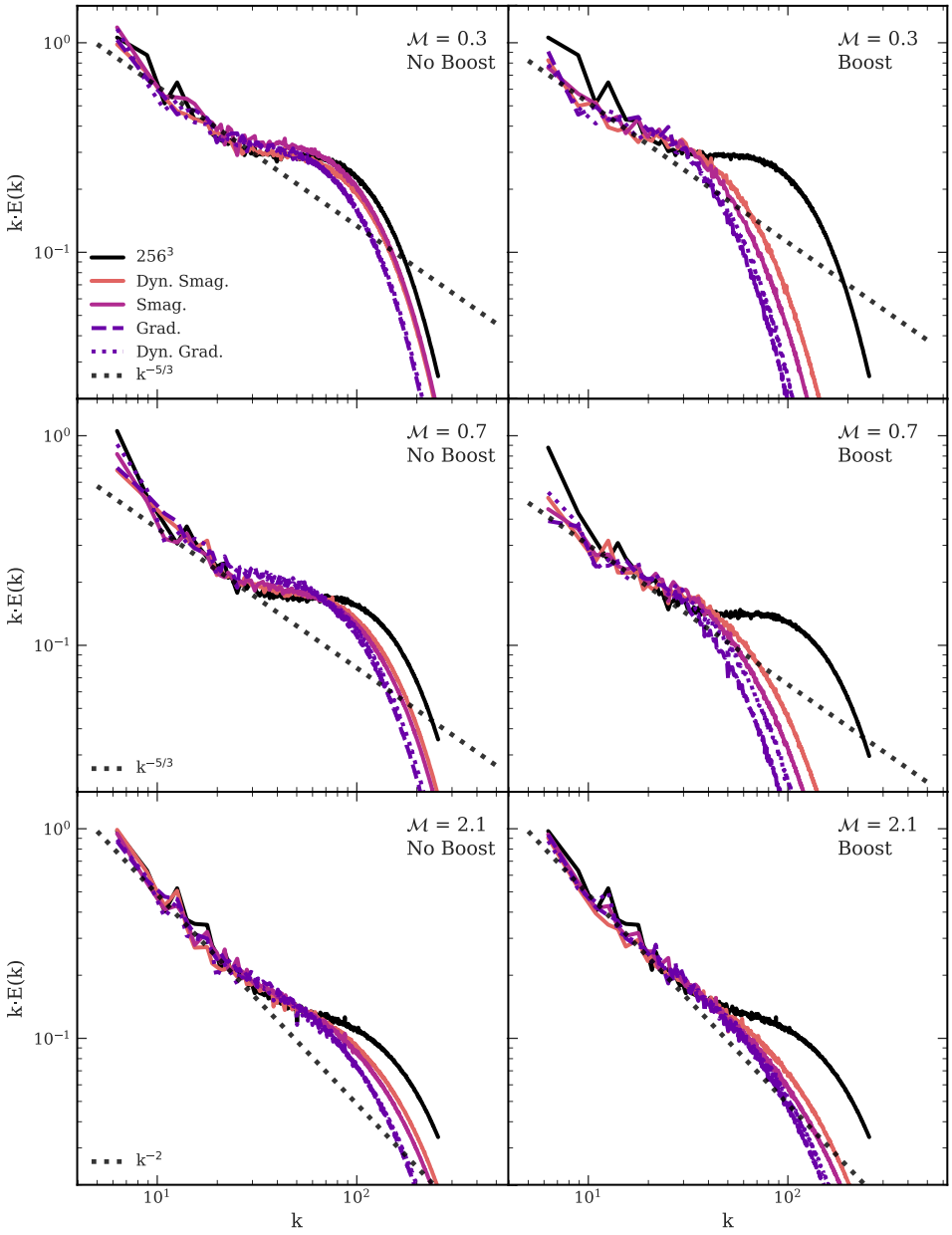

Figure 2.11: The velocity power spectrum of our turbulence volumes with various eddy viscosity models. Rows are Mach numbers 0.3, 0.7, and 2.1 from top to bottom, respectively. The left column has no additional increase in dissipation whereas the right column has an order-of-magnitude boost in dissipation on particles with $\mathcal{M} < 1$. The black curve shows the 256^3 simulation with no eddy viscosity model. The other solid lines show the `Smag.` and `Dyn. Smag.` models, and the dashed and dotted lines show the `Grad.` and `Dyn. Grad.` models, respectively. While all of the models improve the inertial scaling, an additional boost factor ($\gamma \sim 10$) for subsonic particles is required to reproduce the proper scaling at all Mach numbers.

only two physically-motivated quantities that control the strength of diffusion: the length-scale h and the velocity tensor $\boldsymbol{\nabla} \otimes \boldsymbol{u}$. The velocity tensor should not be the issue since it has been verified through the hydrodynamical tests in Hopkins (2015) and would cause the MFM method to fail drastically if the velocity gradients were incorrect. That leaves h as the issue, suggesting that our estimate of the scale over which the eddy viscosity interactions propagate is underestimated[26]. Therefore, we introduce a boost factor γ to the diffusion tensor $||\boldsymbol{K}||$ (i.e., $||\boldsymbol{K}|| \rightarrow \gamma||\boldsymbol{K}||$) in order to get a more reasonable scaling in the inertial range.

We determined the boost factor by running a series of driven turbulence tests with discrete γ from $\gamma = 1$ to $\gamma = 100$. We did not perform a quantitative fit to the Kolmogorov scaling as our interest is in the approximate offset required to improve the inertial scaling and the exact γ is unimportant for the statistics of the flow. We additionally found that we only need to correct the diffusion strength in particles that are subsonic, $\mathcal{M} < 1$, with the Mach number for each particle derived from the current velocity of the particle divided by its thermal sound speed.

The right column of Fig. 2.11 shows the velocity power spectra of our simulations with a dissipation boost factor of $\gamma \sim 10$ on only the subsonic particles in each simulation. At all Mach numbers the additional dissipation causes the kinetic energy to convert into thermal energy much more readily, causing the disappearance of the additional power at small scales, $k \gtrsim 40$. The gradient models `Grad.` and `Dyn. Grad.` perform better than the Smagorinksy variants, but only slightly. That is expected in isotropic homogeneous turbulence since the dissipation strength is effectively the same, and confirms that our implementation of the gradient model is a successful eddy viscosity model.

It is very important that we emphasise the results we observe in Fig. 2.10 are very similar to those found in Bauer & Springel (2012) and Hopkins (2015) for the moving-mesh method (MM; as implemented in `AREPO`) and the mesh-free finite mass method (MFM; as implemented in `GIZMO`), respectively. The only difference is that we show results for much higher resolutions where we begin to see an extended inertial range. Both the MFM and MM hydrodynamical methods produce the same kinetic energy bump that exists in grid-based methods and, therefore, we would expect an eddy viscosity model to also solve the problem in the MM method, although we do not test that in this work. To reiterate, it is necessary that *all* hydrodynamical simulations

[26]Recall that our definition of h is already twice as large as Hopkins (2017), leading to 4 times as much dissipation and mixing.

resolve the inertial range combined with an immediate sharp drop-off in power, or they are not reproducing what we physically observe as turbulence below a certain scale, where the build-up of kinetic energy begins to dominate.

Metal mixing

After each simulation reached $\sim 4\tau_{\rm mix}$, we treated each steady-state volume as new initial conditions for our metal mixing study. In each volume, we gave the densest 50% of particles a metal mass fraction of $Z = 1$ while keeping the rest $Z = 0$. The metals in our simulations act as passive scalars and have no impact on the flow properties. We will test the model with more realistic metal distributions in Section 2.2.4.

First, we must determine if our simulations converge toward a solution for the metal distribution as we increase resolution. We ran each of the metal enriched volumes for an additional $4\tau_{\rm mix}$ to sample a wide variety of metal distribution states. We expect *a priori* that by $\sim 2\tau_{\rm mix}$ the metal-enriched particles should be scattered approximately homogeneously since a particle with the typical velocity $\langle v \rangle$ should have crossed the volume twice in that time. Although that is true for all resolutions, how can we compare each resolution on equal footing after it has reached equilibrium?

The appropriate comparison involves smoothing the spatial distribution of metals on the same scale in all of our simulations. The main assumption is that our simulations with particle counts $\geqslant 128^3$ contain more accurate information on the scales equivalent to our 64^3 simulations. Equivalently, if we *degrade* the resolution of the highest resolution simulations to the lowest resolution, we should hope to obtain a result similar to the lowest resolution simulation. To degrade the resolution for each simulation, we first kernel-weight the particle data onto a grid with resolution twice as fine as minimum smoothing length in the simulation, $\Delta x_{\rm sim,i}$. Next, we smooth the grid data on a physical scale equivalent to our 64^3 simulation using a uniform top-hat filter[27] with width $w = \Delta x_{\rm low}/\Delta x_{\rm sim,i}$, where $\Delta x_{\rm low}$ is always $\Delta x_{\rm low} \equiv 1/64$ since our box has length $L = 1$.

Fig. 2.12 shows the normalised histograms of the filtered metal field. The panels are Mach numbers $\mathcal{M} = 0.3$, 0.7, and 2.1 from top to bottom, respectively. The black curves with markers show the convergence of the filtered metal field for resolutions 64^3, 128^3, 256^3, 512^3, and 768^3. The coloured lines show, from lightest to darkest: `Dyn. Smag` (solid salmon), `Smag.` (solid magenta), `FIRE` (dotted magenta), `Dyn.`

[27]Specifically, we use the `uniform_filter` function from the `scipy` package in Python, with periodicity enabled.

`Grad.` (dashed purple), and `Grad.` (solid purple) at 64^3 resolution. We obtained all of the information for this figure after approximately two mixing timescales, $t \sim 2\,\tau_{\rm mix}$, where $\tau_{\rm mix} = 1/\mathcal{M}$. All of the sub-grid models except the `FIRE` calibration predict a more accurate large-scale metal distribution, at lower resolution.

Fig. 2.13 shows the standard deviation ($\sigma_{\rm Z}$) of the smoothed metal distribution as a function of $N_{\rm x} = \{64, 128, 256, 512, 768\}$ in our simulations at $t \sim 2\tau_{\rm mix}$, for Mach numbers 0.3, 0.7, and 2.1 from top to bottom, respectively. The stars correspond to the simulations without a sub-grid metal mixing model at the resolution given by their labels. The remaining symbols in the legend correspond to the simulations at 64^3 with a sub-grid metal mixing model (see Table 2.5 for a description). The dotted line shows an exponential decay fit to test convergence in the simulations without metal mixing. We expect decreasing $\sigma_{\rm Z}$ with increasing resolution since hydrodynamical mixing is more resolved and the metal value in each grid cell approaches the mean.

At $\mathcal{M} \sim 0.3$ in the top panel of Fig. 2.13, $\sigma_{\rm Z}$ follows an exponentially decreasing trend with resolution in the simulations without metal mixing, as expected. However, there is not evidence for strong convergence at our highest resolution — although it appears the curve is beginning to flatten. The inverted triangle shows the results for the `Dyn. Smag.` model and the left-pointing triangle shows the result for the `Smag.` model. Both the dynamic and standard Smagorinsky models predict a more reasonable $\sigma_{\rm Z}$, reproducing a metal distribution closer to a resolution of 512^3. The `FIRE` calibration of the Smagorinsky model, marked as $\times$, shows little improvement in $\sigma_{\rm Z}$; the result is effectively equivalent to having no model at all. The `Dyn. Grad.` and `Grad.` models are marked by a plus sign and diamond, respectively. The gradient model obviously improves $\sigma_{\rm Z}$ and can, at 64^3 resolution, also reproduce a metal distribution equivalent to a resolution 512^3. The trends are equivalent for $\mathcal{M} \sim 0.7$ turbulence, so we continue to the next panel.

The bottom panel of Fig. 2.13 shows how the sub-grid metal mixing models impact supersonic turbulence at $\mathcal{M} \sim 2.1$. There is much better convergence of the metal distribution at $t \sim 2\tau_{\rm mix}$ than in subsonic turbulence at this time. The qualitative trend remains the same for the 64^3 simulations with sub-grid metal mixing: both the gradient models and Smagorinsky models perform equally as well. However, the width $\sigma_{\rm Z}$ of the distributions are much wider in $\mathcal{M} \sim 2.1$ turbulence. With the exception of the `FIRE` calibration, the other metal mixing models at 64^3 resolution predict metal distributions similar to 256^3.

Our turbulence tests with metals demonstrate that sub-grid metal mixing models

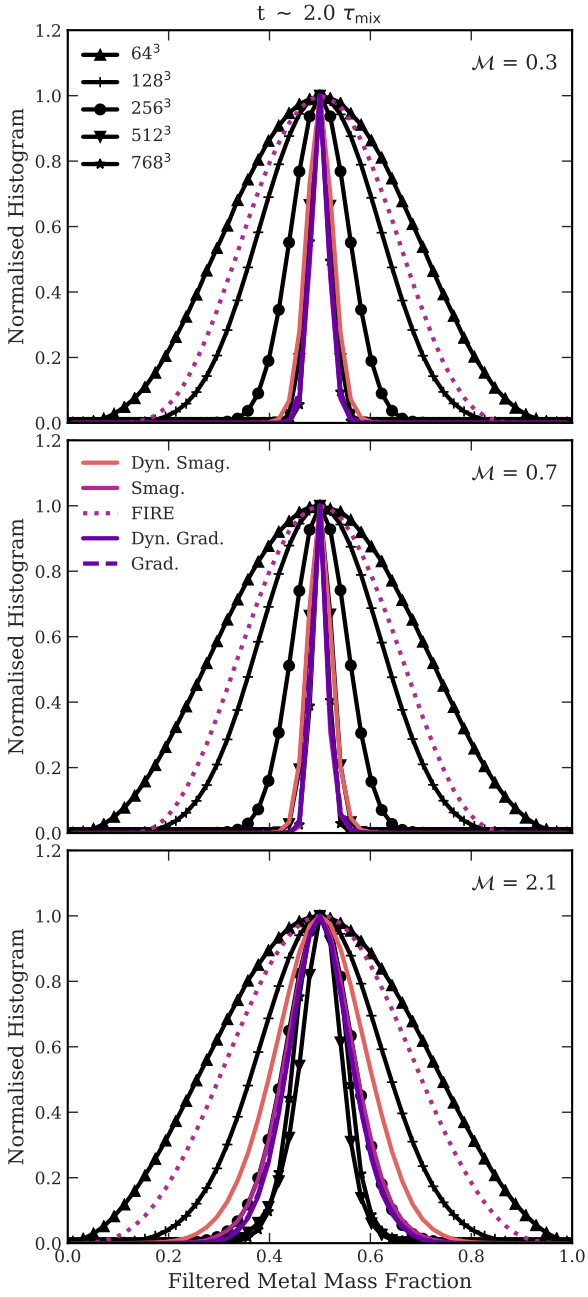

Figure 2.12: The normalised histograms of the filtered metal mass fraction field from the simulations in Section 2.2.3. The panels show Mach numbers $\mathcal{M} = 0.3$, 0.7, and 2.1 from top to bottom, respectively after two mixing timescales. The black curves show the simulations with no sub-grid metal mixing. The coloured curves show the simulations at 64^3 resolution with a sub-grid metal mixing model given by the label. All of the sub-grid metal mixing models show improvement except for `FIRE` which lags due to the lower calibration coefficient.

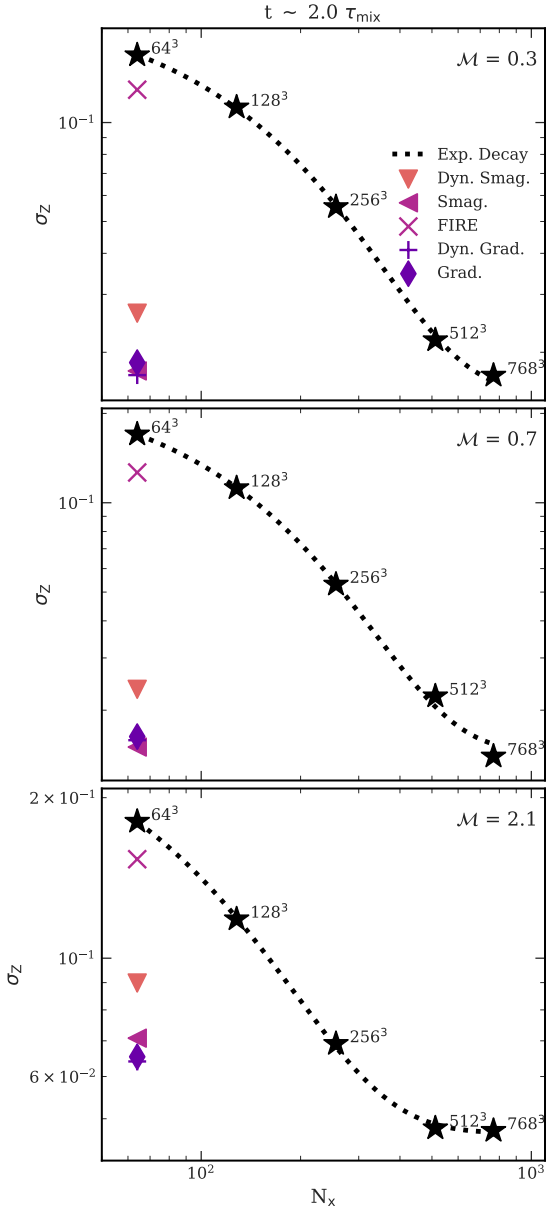

Figure 2.13: The standard deviation of the Gaussian fit to the metal distributions in our driven turbulence experiments as a function of resolution, N_x. Mach numbers are $\mathcal{M}$ = 0.3, 0.7, and 2.1 from top to bottom, respectively. The stars show the simulations with no sub-grid metal mixing model at resolutions given by the number associated with each point. The simulations at 64^3 resolution with sub-grid metal mixing models are given by the figure labels. The dotted line shows an exponential fit to the no metal mixing model simulations. Clearly, the gradient model allows us, at 64^3 resolution, to reproduce the spatial redistribution of metals due to turbulence at a level better than $4-8$ times the resolution.

are necessary in the mesh-free finite mass (MFM) method if one desires more accurate metal distributions. The method we provided for calibrating the metal mixing models is important and, we argue, must be investigated whenever one implements a new metal mixing model into *any* hydrodynamics solver. Particularly, one must calibrate the model parameter at the resolution they desire for their mixing model if a dynamic procedure is not applied. That is an important point: the dynamic procedure (`Dyn. Smag.` and `Dyn. Grad.`) allows us to approximate the calibrated model parameter for the corresponding model (`Smag.` and `Grad.`) without carrying out the calibration. However, the true power of the dynamic procedure is in simulations with mixtures of non-turbulent and turbulent gas at various Mach numbers. In those complex environments the dynamic procedure automatically adjusts the model parameter and, as we showed in Rennehan et al. (2019) for the `Dyn. Smag.` model, drastically alters the resulting metal distributions.

As expected, in pure homogeneous, isotropic turbulence all of the sub-grid metal mixing models improved the accuracy of the metal distributions when using a proper calibration. That is expected in the homogeneous, isotropic case since all of the models effectively act in the same way, on average. The power of the dynamic procedure, and the new anisotropic model is in cosmological simulations where many complex flows interact.

Applicability to other hydrodynamical methods

While our interest lies in the mesh-free finite mass (MFM) method, the models in Section 2.2.2 and experiments in Section 2.2.3 are applicable to other hydrodynamical solvers. Specifically, our results extrapolate with minor modification to grid-based hydrodynamics. Only slight changes to the filtering method must be implemented, as outlined in Schmidt (2015). More care must be taken when applying the model to smoothed particle hydrodynamics (SPH). However, as we outline below, there is much broader applicability to the moving-mesh (MM) method.

First we consider eddy viscosity. In SPH, it is well-established that there is a *deficit* in power near the resolution scale rather than a build-up of kinetic energy (Bauer & Springel, 2012; Price, 2012b; Hopkins, 2013, 2015). That fact suggests that SPH reproduces turbulence better than the MFM or MM methods, but produces results at a much lower effective resolution. The lack of power, rather than the overabundance of power, implies that an eddy viscosity model would only further

degrade the resolution of SPH results and not improve the inertial scaling. Therefore, we do not recommend eddy viscosity models for SPH but rather the work of Di Mascio et al. (2017), who recently provided an SPH equivalent.

For the MM method, the build-up of kinetic energy at the resolution scale is present (Bauer & Springel, 2012) and of equivalent magnitude to our results in the MFM method. Therefore, we recommend investigation into eddy viscosity models for the MM method as they could improve the inertial scaling. In terms of implementation, all of the derivations in Section 2.2.2 apply to the MM method.

Metal mixing using the Smagorinsky model has been studied in cosmological simulations involving SPH but widely ignored in the MM method. However, no calibration technique has been provided by the community for SPH and the calibrations usually follow the theoretical value for the Smagorinsky model (e.g., Shen et al. 2010; Williamson et al. 2016) or calibrations that depend on sub-grid astrophysics models (e.g., Wadsley et al. 2017; Escala et al. 2018). The metal mixing calibration technique in Section 2.2.3 is completely applicable to SPH since metals are treated equivalently to the MFM method, and they are both constant mass methods. Additionally, our calibration technique does not depend on the uncertainties within astrophysical sub-grid models — only pure hydrodynamics.

For the MM method, all of Section 2.2.2 is applicable for metal mixing since the MM method relies on transport equations such as equation (2.38) for advecting metals throughout the fluid (Springel, 2010). In fact, Balarac et al. (2013) find that the gradient model improves the inertial scaling in the power in the metal field through identical transport equations. However, convergence tests such as those in Fig. 2.13 are necessary in order to determine whether they are truly required, or if numerical dissipation is adequate.

2.2.4 Cosmological Simulations

Understanding the evolution of galaxies is a complex enterprise involving highly non-linear coupled physical processes. Not only do stellar feedback and active galactic nuclei produce powerful outflows that drive turbulence locally in the interstellar media of galaxies, but also in the gas reservoirs surrounding galaxies. Turbulence also appears through the Kelvin-Helmholtz instability in ram pressure stripping of galaxies moving through a hot medium, and through the stellar winds from stars making their way out from the galaxy into the circumgalactic medium.

The physical processes above occur on spatial scales much smaller than currently possible to resolve in the average contemporary cosmological simulation. For that reason, the majority of astrophysics in cosmological simulations are encoded into parametrised sub-grid models that use the information on the largest scales to predict what occurs below the resolution of the simulation. The resulting calculations usually indicate how much mass and energy should be injected (or removed) from the large-scale gas and stellar components. However, there is no one correct way to approximate the astrophysics on the sub-grid scale since it highly depends on the maximum possible resolution, hydrodynamical method, as well as other complex numerical effects. All of the issues with numerics and missing physics usually ends up in one or more tunable free parameters in the model.

Assuming such a sub-grid astrophysical model is developed, how do we know that it is correct? Or, at least approximating reality? Normally, one or more trusted astronomical observation is used to test the validity of all of the sub-grid astrophysics that may exist. A common example would be the galaxy stellar mass function, or the $\mathrm{M_{BH}}$ - σ_* relationship that links supermassive black hole masses to the stellar masses of their host galaxies. However, two different hydrodynamical methods may provide different parameter values for the same sub-grid astrophysical models. Additionally, there may be two completely different approaches to modelling the same physical phenomenon with no clear mapping between free parameters! Calibrating sub-grid astrophysical models is obviously a complicated endeavour and must be built on a strong hydrodynamics base. How can we begin to trust that our understanding of the astrophysics of small scale approximate reality if the hydrodynamics, as we showed in Section 2.2.3, does not reproduce reality?

Our goal is in determining whether the converged and separately calibrated sub-grid turbulence models we presented in Section 2.2.2 have any significant impact over the sub-grid astrophysical models that we use in large-volume cosmological simulations. As a first step, we only investigate the broad, qualitative impact on the gas properties in gaseous halos in a single set of sub-grid astrophysical models. We stress that we do not intend to reproduce the full galaxy population in a calibrated and predictive sense. Additionally, we note that more testing is required across the all of sub-grid astrophysical models that exist in the literature, as the turbulent mixing models may interact in unexpected ways due to the non-linearity of the problem.

Table 2.7: Cosmological parameters and simulation information for our simulation set. Our parameters follow Davé et al. 2019 except begin at a higher redshift.

Cosmological Parameters	
$\Omega_{\rm m,0}$	0.3
$\Omega_{\Lambda,0}$	0.7
$\Omega_{\rm b,0}$	0.048
h	0.68
σ_8	0.82
$n_{\rm spec}$	0.97
Simulation Information	
$z_{\rm begin}$	249
$N_{\rm part}$	2×256^3
$L_{\rm side}$	25 cMpc h^{-1}
$m_{\rm part,gas}$	1.26×10^7 M$_\odot h^{-1}$
$m_{\rm part,dark}$	6.88×10^7 M$_\odot h^{-1}$
$\epsilon_{\rm soft,min}$	0.5 kpc h^{-1}

The simulations

For our comparison, we choose to use the SIMBA galaxy formation model. SIMBA includes robust sub-grid models of star formation, cooling, stellar feedback, chemical enrichment, active galactic nuclei feedback, and dust evolution — all evolved in concert with the mesh-free mass method (MFM) (Davé et al., 2016, 2019). For this study, we implemented the SIMBA models into the public version of GIZMO as described in Davé et al. (2019) and we point the interested reader to that study for the details of the sub-grid models. We follow the approach of Schaye et al. (2015) and calibrate our implementation of the SIMBA model only to the galaxy stellar mass function and the $M_{\rm BH}$-M_* relationship at $z = 0$ for the purposes of this study.

We run 6 cosmological-scale volumes of side-length $L = 25$ cMpc h^{-1} (~ 37 cMpc) in order to compare our various mixing models. The simulations begin from initial conditions generated with MUSIC (Hahn & Abel, 2011) at a redshift of $z = 249$, with a standard Λ Cold Dark Matter cosmology (see Table 2.7 for values). The mass resolution in gas and dark matter follows the SIMBA simulations with $m_{\rm part,gas} = 1.26 \times 10^7$ M$_\odot h^{-1}$ and $m_{\rm part,dark} = 6.88 \times 10^7$ M$_\odot h^{-1}$, respectively. We use adaptive gravitational softening (Hopkins, 2015; Hopkins et al., 2018) to compute the softening lengths of all of our particles, and enforce a minimum softening length of $\epsilon_{\rm soft,min} = 0.5$ kpc h^{-1}.

Global metal mixing

While sub-grid scale turbulence models maximally impact the smallest scales in a cosmological simulation, their integrated effect impact the properties of the largest scales, such as global metal distribution functions (Shen et al., 2010; Escala et al., 2018; Rennehan et al., 2019). Therefore, in this Section, we examine the impact of the gradient model on the metal distribution functions (MDFs) in the circumgalactic medium (CGM) and warm-hot intergalactic medium (WHIM) — both known to be turbulent environments (Iapichino et al., 2011, 2013; Tumlinson et al., 2017).

Our definition of CGM and WHIM depends on separating gas that is bound to halos from that which is unbound, at a given epoch. A good estimation comes from Davé et al. (2010),

$$\frac{\rho_{\rm bound}(z)}{\Omega_{\rm b}(z)\rho_{\rm c}(z)} = 6\pi^2\left(1 + 0.4093\left(\frac{1}{\Omega_{\rm m}(z)} - 1\right)^{0.9052}\right) - 1, \tag{2.64}$$

where $\Omega_{\rm b}(z)$ is the baryon fraction as a function of redshift, $\Omega_{\rm m}(z)$ the matter fraction, $\rho_c(z) = 3(H(z))^2/(8\pi G)$, and $H(z)$ the redshift-dependent Hubble function. All gas above $\rho_{\rm bound}(z)$ we consider bound to halos, and have confirmed that the approximation holds well.

We define the CGM to be all gas in the volume that is above $\rho_{\rm bound}(z)$ in equation (2.64) and below the star formation density threshold, $\rho_{*,\rm crit} = 4.4\times10^{-25}$ g cm^{-3}, at any temperature. That includes gas in the intragroup medium of our most massive halos in the $(25\,{\rm cMpc}\,h^{-1})^3$ volumes. The WHIM is all gas that is below $\rho_{\rm bound}(z)$ and above a temperature of $T = 10^5$ K.

Fig. 2.14 shows the metal distribution functions (MDFs) for our two gas phases in columns: WHIM (left) and CGM (right), and at $z = 6$, 4, 2, and 0 in rows from top to bottom, respectively. These are probability density functions, and were constructed by binning the particle metallicities in the range $10^{-6} < \log(Z/Z_\odot) < 10^1$, where $Z_\odot = 0.0134$ (Asplund et al., 2009). The black curves show the control simulation, `None`, with no sub-grid metal mixing. The coloured curves show the simulations with sub-grid metal mixing and are, from lightest to darkest: `Dyn. Smag` (solid salmon), `Smag.` (solid magenta), `FIRE` (dotted magenta), `Dyn. Grad.` (dotted purple), and `Grad.` (dashed purple). See Table 2.5 for more details.

First we focus on the WHIM. At $z = 6$, there are two distinct components across all of our model variants. The peak at $Z \sim 10^{-1}$ Z$_\odot$ is the highly enriched interstellar medium (ISM) gas that recently joined the WHIM via stellar winds from the

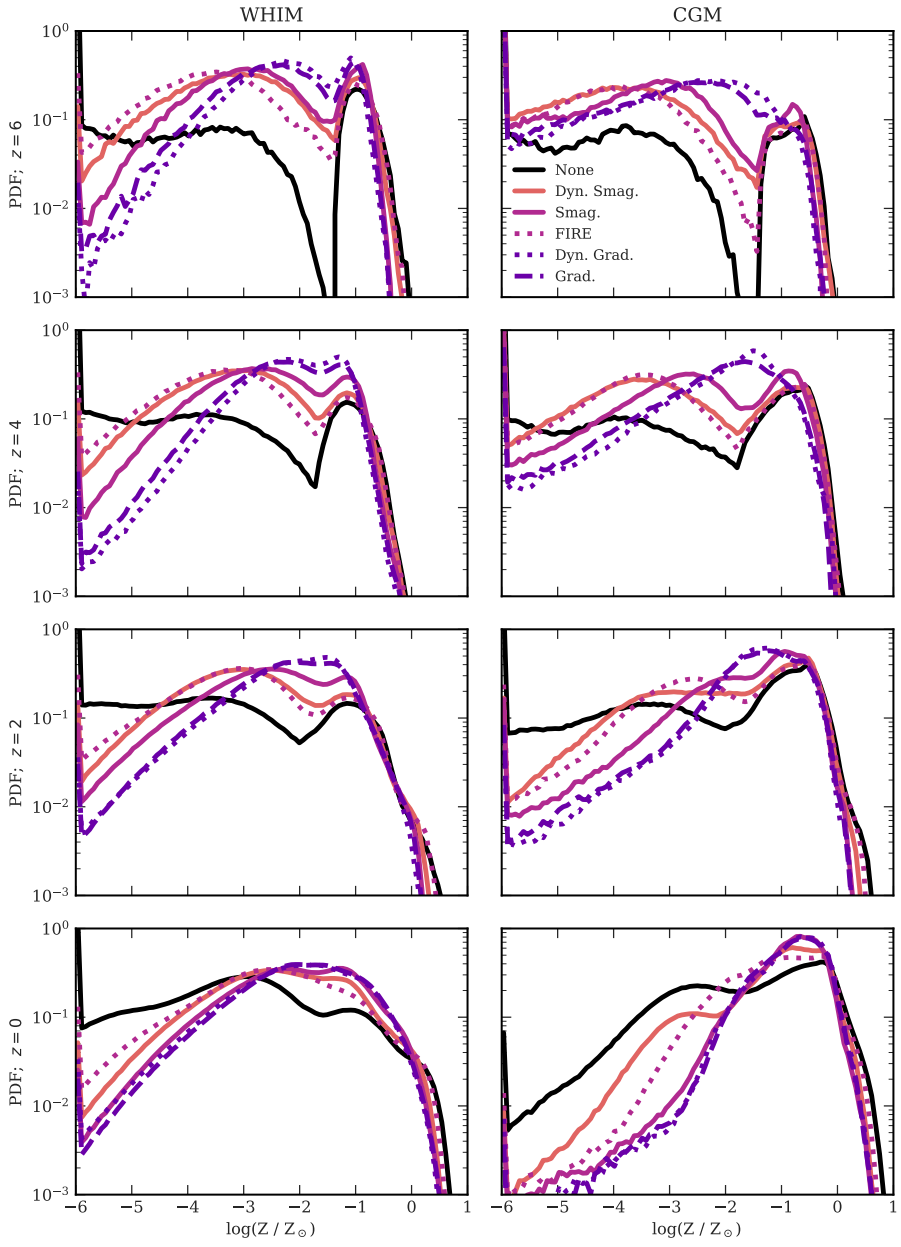

Figure 2.14: The metal distribution function across global phases: the warm-hot intergalactic medium (WHIM; left column) and the circumgalactic medium (CGM; right column). Redshifts are given by the rows from top to bottom: $z = 6$, 4, 2, and 0, respectively. The gradient model variants mix metals much more rapidly in the early stages of the simulation, suggesting that the choice of model will impact studies that focus on enrichment timing of halo gas.

integrated star formation in the early universe. The lower distribution is gas that has mixed into the WHIM but did not recycle through the ISM, missing the opportunity for further enrichment via supernova feedback. It is important to note that since there is no mixing in the `None` case, when a particle leaves the ISM it *cannot* change its metallicity. The models that include sub-grid mixing show varying spread in the MDFs, with the `FIRE` and `Dyn. Smag.` showing the widest spread. The `Dyn. Grad.` and `Grad.` models show the tightest distributions, with the two peaks in the distributions seemingly merging at $Z \sim 10^{-1.5}$ $Z_\odot$. The `Smag.` model matches the `Dyn. Smag.` model at $Z \gtrsim 10^{-3}$ $Z_\odot$, but is biased toward higher metallicities below that threshold.

The next 3 panels in the left column of Fig. 2.14 show the MDFs in the WHIM for $z = 4$, 2, and 0, from top to bottom, respectively. The trend for all models is to approach a singly-peaked distribution as the simulation evolves through cosmic time. Most of the evolution in the MDFs occurs from $z = 6$ to $z = 2$ after which the distributions are mostly stationary. The transition from $z = 6$ to $z = 4$ demonstrates how rapidly the WHIM evolves at high redshift, and how each sub-grid metal mixing model impacts the MDFs with different mixing rates. Specifically, the `Dyn. Grad.` and `Grad.` models predict similar distributions at $z = 4$, and produce the tightest MDFs compared to all of the other models. In fact, there is the same trend at $z = 4$ as at $z = 6$ — the gradient model variants (`Dyn. Grad.` and `Grad.`) predict tighter MDFs, followed by wider distributions in the Smagorinsky models (`Smag.`, `Dyn. Smag.`, and `FIRE`, respectively).

There are similar trends in the CGM, as we see in the right column of Fig. 2.14. To reiterate, the panels show redshifts $z = 6$, 4, 2, and 0 from top to bottom, respectively.

At $z = 6$, there is a clear distinction between the distribution at $Z \lesssim 10^{-2}$ $Z_\odot$ and $Z \gtrsim 10^{-2}$ $Z_\odot$ in the `None` case, and the Smagorinsky variants (`Smag.`, `Dyn. Smag.`, and `FIRE`). Stellar feedback drives the peak at $Z \sim 10^{-1}$ $Z_\odot$ similarly to the WHIM at this redshift, whereas the distribution at $Z \lesssim 10^{-2}$ $Z_\odot$ is from the very first generations of stars. By this time the `Dyn. Grad.` and `Grad.` models have mixed the most rapidly to create a single broad distribution in their MDFs. All of the models with sub-grid metal mixing have much more gas mass enriched above $Z > 10^{-6}$ $Z_\odot$ than the `None` case, especially compared to the deficit at $Z \sim 10^{-1.5}$ $Z_\odot$ in the `None` case. The Smagorinsky models vary in mixing rate as `Smag.`, `Dyn. Smag.`, and `FIRE`, from fastest to slowest, respectively. The MDFs in the CGM at redshifts $z = 4$ to $z = 0$ demonstrate the same trends as in the WHIM phase at the

same redshifts: the gradient models mix much more rapidly at early stages than the Smagorinsky models. At $z = 0$ the `Dyn. Grad.`, `Grad.`, and `Smag.` models predict the same distribution in the global CGM phase, whereas the `Dyn. Smag.` and `FIRE` models predict slightly less enriched gas.

The MDFs in the turbulent WHIM and CGM show the importance of sub-grid metal mixing models in cosmological simulations as well as the importance of model choice. In all cases we include metal mixing, the MDFs are significantly tighter at all redshifts we measure and significantly tighter for the CGM at $z = 0$. This is contrary to the study in Su et al. (2017) that found metal mixing to be relatively unimportant on cosmological scales. However, we find that the `FIRE` calibration is much too low to reproduce the correct converged hydrodynamical mixing of metals (see Section 2.2.3). With our new calibrations of the Smagorinsky model, `Smag.`, and the new gradient models, `Dyn. Grad.` and `Grad.`, we see significant differences at all redshifts.

A common theme in theoretical galaxy evolution is that equivalent results between different models at $z = 0$ does not necessarily imply a similar integrated history. The evolutionary paths for each sub-grid metal mixing model all evolve at slightly different rates as we would expect based on their diffusivities from Section 2.2.2. The lesson from our results is that the metal mixing model choice impacts the early development phases of galaxies rather than the long term equilibrium stages. At higher redshifts, $z \gtrsim 2$, gas is collapsing to form galaxies, while stellar feedback and supermassive black holes are driving outflows out of the potential wells and forcing turbulence. Our inclusion of the full diffusion tensor in the `Dyn. Grad.` and `Grad.` models allows the gas that is compressing from feedback and infall to further mix its metal mass with nearby neighbours, tightening the MDFs.

The Smagorinsky variants also improve the results and allow metal mixing between gas particles but produce broader distributions, notwithstanding the `Dyn. Smag.` and `Smag.` models showing good matches in the convergence tests of Section 2.2.3. This is an important point: the simple turbulence tests in Section 2.2.3 showed agreement between the gradient and Smagorinsky models (except for the lower `FIRE` calibration) but now we see disagreement in complex cosmological environments. Evidently, ignoring the trace of the velocity tensor is not the correct approach for cosmological contexts and we recommend either the `Grad.` or `Dyn. Grad.` models as we see no difference with the dynamic procedure applied to the constant coefficient gradient case.

The impact of eddy viscosity

While the velocity power spectra results in Section 2.2.3 show that eddy viscosity is required in Lagrangian finite mass methods, we find no significant impact of eddy viscosity on the *average* gas and galaxy properties in our cosmological simulations. However, we should expect these results considering that the eddy viscosity models we tested in Section 2.2.3 had no impact on the largest scales of the simulation (as they should not). In terms of global gas properties, we investigated the vorticity, temperature, and density distributions of the warm-hot intergalactic medium (WHIM) and circumgalactic medium (CGM) phases[28] and found only minor differences in data binned by galaxy stellar mass. Additionally, there were little differences between stellar mass distributions in our small volume, low resolution tests. To compare the impact of eddy viscosity between the models we need to investigate the small scales of the simulations in a controlled manner. For that reason, we will use the most massive galaxy in our cosmological simulations as a qualitative case study of the impact of the eddy viscosity model on the halo gas since this galaxy represents the same system in each simulation we ran.

First, we investigate the temperature projections of the most massive (in stellar mass) galaxy at redshifts $z = 2$, 1, and 0. We confirmed that the most massive galaxy at $z = 2$ ends up as part of the most massive galaxy at $z = 1$ and $z = 0$. We restrict our analysis to a radius of 500 kpc (physical) for the purposes of this introductory study. Additionally, given that we see the `Smag.`, `Grad.`, and `Dyn. Grad.` models as the best choices from the results in Section 2.2.4, we restrict our analysis to the `None`, `Smag.`, and `Grad.` models. The `Grad.` model is much less computationally expensive than the `Dyn. Grad.` model and is, therefore, the better choice for cosmological simulations[29].

Fig. 2.15 shows the density-weighted temperature projections of the most massive galaxy at $z = 2$, 1, and 0 in rows from top to bottom, respectively. The columns show the `None`, `Smag.`, and `Grad.` models from left to right, respectively.

First, we compare the results at $z = 2$ across mixing model variants. The dark clumps in the `None` case are individual cold galaxies that are being fed into the main structure via filaments. Ongoing stellar and AGN feedback leads to the temperature increase in the central region, while the extended $T \sim 10^6$ K halo is a mixture of

[28] Definitions in Section 2.2.4.

[29] Further study is required for tests that rely on getting the small scale structure as correct as possible, such as cold clouds interacting with a hot medium.

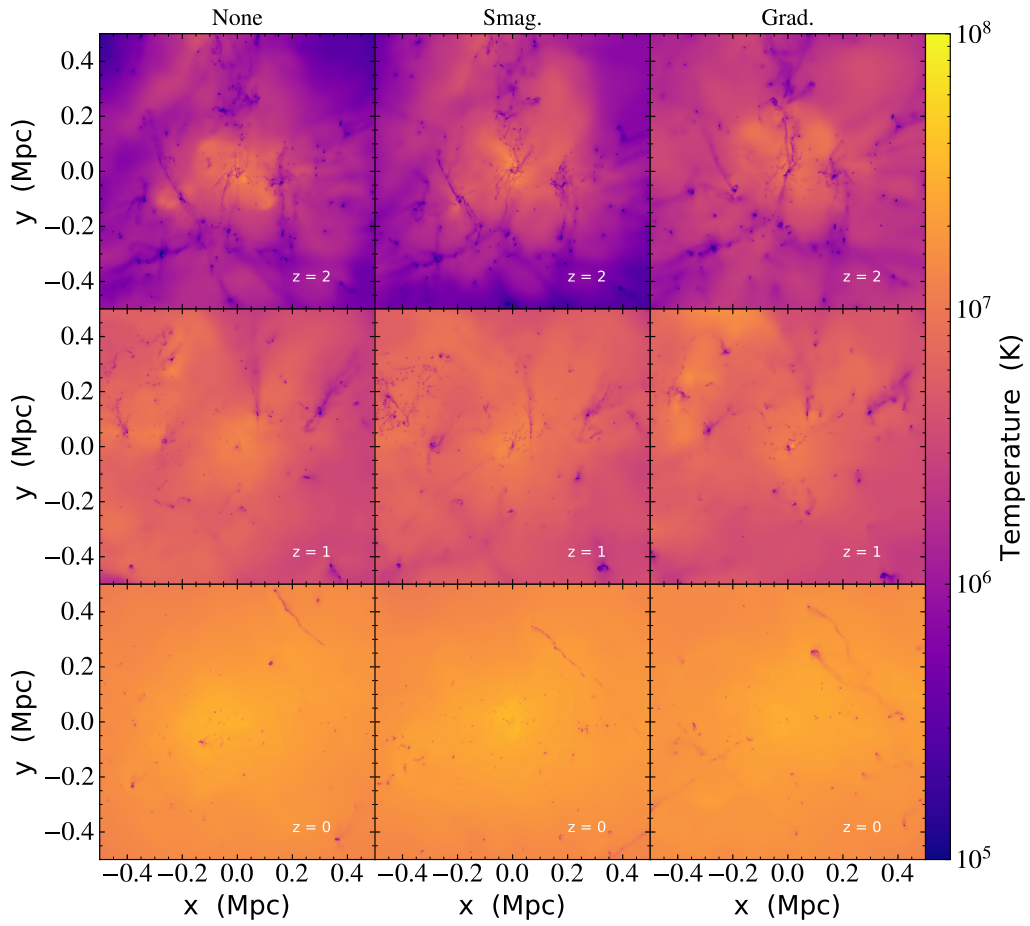

Figure 2.15: Temperature projections of the most massive halo in three of our cosmological simulations at redshifts $z = 2$, 1, and 0 in rows from top to bottom, respectively. The columns show our `None`, `Smag.`, and `Grad.` models from left to right, respectively. Each of the panels represents a 1 Mpc by 1 Mpc (physical) region centred on the most massive galaxy at each redshift. The `Smag.` simulation shows the most small-scale structure at all redshifts, and a smoother distribution of temperature at high-redshift compared to the `None` case. The `Grad.` model produces less small-scale structure than any other model, and much more hot gas at high redshift. Surprisingly, the `Grad.` model also produces more extended tails from sub-structure moving through the hot halo at lower redshift, suggesting it may impact future studies of jellyfish galaxies.

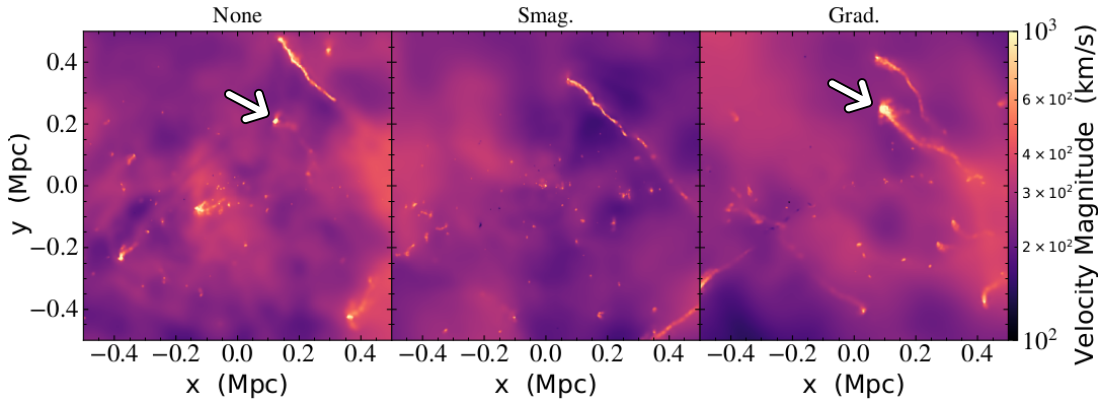

Figure 2.16: Velocity magnitude projections of the most massive halo at $z = 0$ in our cosmological simulations. From left to right we show the `None`, `Smag.`, and `Grad.` simulations, respectively. The images are centred on the most massive galaxy and each show a 1 Mpc by 1 Mpc (physical) region. The `Smag.` model shows more substructure than the `None` and `Grad.` models. The `Grad.` model shows many extended tails from substructure moving through the hot halo compared to the control and `Smag.` models, in addition to an overall smoother distribution in velocity space.

gravitational heated gas and gas heated by previous generations of stellar feedback and AGN. There are many filamentary structures visible feeding the main galaxy (centred), and there is some small scale structure visible surrounding the galaxies. The `Smag.` case resembles the `None` case, except there less cold gas in the upper infalling structure. Additionally, the satellite galaxies have more small-scale structure in the cold gas surrounding them. Importantly, the distribution of hotter gas appears mores smoothly distributed throughout the volume due to an increase in conversion of kinetic energy in thermal energy via the eddy viscosity.

The temperature projection of the `Grad.` model simulation is strikingly different at $z = 2$ than both the `Smag.` and `None` cases. While in the `None` and `Smag.` cases there appears to be large-scale gas at $T \lesssim 10^5$ K at the boundary of the 500 kpc radius, there is no such gas in the `Grad.` case. There is much less small-scale structure in the `Grad.` case, and the filaments are much smoothly distributed in space. We conclude that the stellar feedback, gravitational heating, and AGN feedback are much more effective at heating the gas on small scales in the `Grad.` simulation, since the eddy viscosity is the strongest.

At $z = 1$, in the middle row of Fig. 2.15, a similar picture emerges. The `Smag.` case has much more small scale structure than the `None` case as is evident in the central region and to the leftmost satellite galaxy structure. There is still a much smoother distribution of cold gas that extends outwards from the satellite galaxies. There are only a few centrally-located cold gas clumps in the `Grad.` case compared with the `Smag.` case and there is effectively no fragmentation identifiable in the satellite galaxy structure to the left of panel.

The last row of Fig. 2.15 shows $z = 0$ across the three model comparisons. By this redshift, there is little structure remaining in the group-sized halo and the temperature distribution of the intragroup gas is very smooth. Similarly to the other redshifts, the `Smag.` shows the most fragmentation of cold gas, followed by the `None` case, and then the `Grad.` case. All three show a satellite being stripped of gas in the upper right of the panels but the `Smag.` and `Grad.` models each differ in an important unique way compared to the `None` case.

It is much easier to see the differences in ram pressure stripping in velocity space rather than temperature space. Fig. 2.16 shows the density-weighted velocity magnitude projections of the most massive galaxy at $z = 0$ for the three eddy viscosity models `None`, `Smag.`, and `Grad.` from left to right, respectively. It is apparent across all three cases that the substructure is moving at least a factor of ≈ 5 faster than the

background gas yet only the `Grad.` case shows the most clearly defined long stripping tails from the cold gas in the galaxies. The `Smag.` case does not have the cold gas in the satellite galaxy marked by the arrows in the `None` and `Grad.` cases, as the gas has been completely stripped away. The `Grad.` case produces the cold gas in that satellite galaxy, and the tail is much more extended than in the `None` case. In fact, it is evident upon close inspection of the cold gas structures in the halo that the `Grad.` case produces tails from the cold gas in the galaxies more readily than the `None` or `Smag.` cases. That has implications for the study of ram pressure stripping in general, and should be further investigated in the future.

2.2.5 Conclusions

Turbulence is a key physical process in the study galaxy evolution and one of many highly complex non-linear interactions that must be understood to advance our knowledge of the Universe. The complexity demands the use of simulations that combine astrophysical sub-grid models with hydrodynamics and gravity in an expanding universe. All hydrodynamical simulations are known to require additional sub-grid models to accurately treat the impact sub-grid turbulence, yet these models have been widely ignored in the astrophysical community. We have, for the first time in Lagrangian hydrodynamics, implemented and studied the gradient model (Clark et al., 1979) – an anisotropic sub-grid turbulence model for viscosity and metal mixing (Hu & Chiang, 2020). The model is based on directly modelling the error terms that arise from discretisation of the fluid field via a Taylor series expansion, including the compression terms that are missing from the standard Smagorinsky model. We additionally implemented a dynamic procedure that computes the model parameter on-the-fly for the gradient model following the approach of Rennehan et al. (2019). We used the mesh-free finite mass method in the `GIZMO` code (Hopkins, 2015) as our numerical hydrodynamics solver for all of our experiments.

We ran driven turbulence simulations at Mach numbers $\mathcal{M}$ = 0.3, 0.7, and 2.1 to validate the gradient model and compare with the popular Smagorinsky model. Hu & Chiang (2020) recently showed, by post-processing driven their turbulence simulations, that the gradient model should be able to reduce the build-up of kinetic energy near the resolution scale in isotropic, homogeneous turbulence and better reproduce the sub-grid metal flux. We confirmed these results in Section 2.2.3 by using the gradient model at simulation time.

Our analysis of the velocity power spectra in driven turbulence produced unexpected results. We found that the gradient and Smagorinsky models, and their dynamic variants, predicted insufficient dissipation to reduce the artificial build-up of kinetic energy near the grid-scale in our 256^3 simulations. For that reason we introduced a boost factor γ for the dissipation strength (i.e. $||\boldsymbol{K}|| \rightarrow \gamma||\boldsymbol{K}||$) and experimented with various values in the range $\gamma \in [1, 100]$. We found that a factor of $\gamma \sim 10$ is sufficient to reduce the build-up of kinetic energy and is required for all of the Smagorinsky and gradient model variants. Additionally, we found that the boost factor only needs to be applied to the subsonic ($\mathcal{M} < 1$) particles in our simulations to produce the correct statistics in supersonic turbulence (our $\mathcal{M} \sim 2.1$ test). The true boost factor is higher since we used the maximum interaction distance between neighbouring gas particles to calculate the diffusion tensor, leading to a ~ 4 times *additional* boost (total ~ 40) over other the default implementation in `GIZMO` (Hopkins et al., 2018).

Our converged metal mixing simulations in Section 2.2.3 show that when we use the gradient and Smagorinsky models at lower resolution (64^3 particles) we are able to produce MDFs that are equivalent to 4 to 12 times the resolution. That is true for both the constant-coefficient and dynamic variants of the gradient and Smagorinsky models with standard parameter values, with the additional factor of ~ 4 boost from using the maximum interaction distance in the kernel. However, lower calibrations such as those from the `FIRE` simulations do not produce the correct rate of mixing as they are at least a factor of ~ 20 too low. We posit that this is because of the common calibration approach in cosmological simulations: calibration in tandem with the full suite of astrophysical sub-grid models.

We argue that calibration of fundamental hydrodynamics models, such as the metal mixing model here, *must be done in the absence of sub-grid astrophysical models.* There must be a strong hydrodynamics base before the complexity of astrophysics is built on top. Note that our dynamic Smagorinsky and dynamic gradient models *do not require calibration* and produce accurate predictions for the rate of metal mixing in isotropic, homogeneous turbulence.

As a first application of the new gradient model, we investigated a set of cosmological simulations to determine if there is any dominant impact on the galaxy and gas properties. We investigated the metal mixing and eddy viscosity separately in Sections 2.2.4 & 2.2.4, respectively.

We found that the choice of sub-grid metal mixing model strongly impacts the

MDF evolution in the warm-hot intergalactic medium (WHIM) and circumgalactic medium (CGM). We found that the gradient and dynamic gradient models mix metals much more rapidly than the Smagorinsky variants and produce tighter MDFs up until $z \sim 1$ when they approach a similar distribution down to $z = 0$. In our simulation without sub-grid metal mixing, the MDFs in the WHIM and CGM are significantly broader than any of the simulations with sub-grid metal mixing demonstrating that, at the very minimum, including any sub-grid metal mixing model is an improvement. The most important result we discovered is that the metal mixing models are most impactful during the tempestuous early stages of galaxy evolution. On very long timescales, the equilibrium distributions match quite closely across the models.

Including eddy viscosity in our cosmological simulations did not significantly impact the galaxy properties we investigated when averaging in bins of stellar mass after $z = 2$. The galaxy stellar mass function was relatively unchanged, along with only slight variations in the stellar mass to halo mass function. We also found that including eddy viscosity did not significantly impact the averaged gas distributions of vorticity, temperature, and density across galaxies of similar stellar mass. We expected *a priori* that the large-scale properties of galaxies would be unaffected as the sub-grid eddy viscosity mainly impacts the small-scale. For that reason, we investigated a single system that could be linked across all of our cosmological simulations to gain a qualitative view of the impact.

In Section 2.2.4 we showed the temperature projections of the most massive halo traced from $z = 2$, 1 and 0 in our cosmological simulations for three eddy viscosity models: no model, the standard Smagorinsky model, and the new constant-coefficient gradient model. We found that the Smagorinsky model produced much more fragmentation in the halo gas of the most massive galaxy on the small scale compared to having no eddy viscosity, at all redshifts. We also found that the spatial temperature distribution was much smoother at $z = 2$ when stellar and active galactic nuclei feedback was much stronger, showing that the small-scale kinetic energy was being efficiently converted into thermal energy. Although the constant-coefficient gradient model seemingly dissipates faster based on our results in Section 2.2.3, we observed that its inherent anisotropy does not lead to the same fragmentation we saw in the Smagorinsky model. At high redshift, $z = 2$ the gradient model produced a much more widespread hot gas. The filamentary structure at all redshifts was much smoother and, after $z = 1$, the satellite galaxies in the halo had many more clearly defined tails due to an improved treatment of ram-pressure stripping.

Sub-grid metal mixing and eddy viscosity models have a strong impact on galaxy evolution simulations. In this work, we showed in the simplest case of isotropic, homogeneous turbulence that the all of the models tested here improved the accuracy of metal mixing and turbulent kinetic energy dissipation in the mesh-free finite mass method. The most significant differences between model choice appeared at high redshift in the early stages of galaxy evolution, before any equilibrium is reached. Given that contemporary cosmological simulations have resolutions of less than $\sim 50^3$ particles in a typical galaxy, we recommend that future studies must at least use the constant-coefficient gradient model as it is (a) computationally inexpensive compared to the dynamic version, while producing similar results and (b) includes the full velocity tensor in the diffusion tensor to give the most accurate solution for sub-grid turbulence. Our recommendation is especially pertinent given the recent push to study higher redshift systems driven by the upcoming launch of the *James Webb Space Telescope*, as our theoretical understanding of enrichment and thermodynamic histories will depend directly on sub-grid turbulence model choice.

2.2.6 Acknowledgements

This research was enabled in part by support provided by WestGrid and Compute/Calcul Canada. The simulations in this research were made possible by SciNet and the Niagara supercomputing cluster. DR acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 534263] and through the Discovery Grant program. DR thanks Arif Babul, Belaid Moa, Ondrea Clarkson, Austin Davis, Drummond Fielding, Phil Hopkins, and Wolfram Schmidt for useful advice during the course of this research. DR gives special acknowledgement to Arif Babul and Belaid Moa for providing invaluable support to this research, without which it would not have been possible. DR also thanks Daniel Price for his helpful comments during the reviewing stage that improved this work.

Our analysis was performed using the Python programming language (Python Software Foundation, `https://www.python.org`). The following packages were used throughout the analysis: `h5py` (Collette, 2013), `numpy` (Harris et al., 2020), `scipy` (Virtanen et al., 2020), `yt` (Turk et al., 2011), and `matplotlib` (Hunter, 2007). Prototyping of the analysis scripts was performed in the IPython environment (Perez & Granger, 2007).

2.2.7 Data availability

The data underlying this article will be shared on reasonable request to the authour.

2.2.8 FILTERING APPROXIMATION

Consider a scalar field $f_i(\boldsymbol{r})$ where $\boldsymbol{r} = (x, y, z)$ in a general coordinate system. We expand $f_i(\boldsymbol{r'})$ about $\boldsymbol{r}$ in equation (2.37) via a Taylor expansion,

$$f_i(\boldsymbol{r'}) = f_i(\boldsymbol{r}) + [(\Delta\boldsymbol{R}) \cdot \boldsymbol{\nabla} f(\boldsymbol{r})] + \frac{1}{2}[(\Delta\boldsymbol{R}) \otimes (H(\boldsymbol{r}) \cdot (\Delta\boldsymbol{R}))] + O(|\Delta\boldsymbol{R}|^3), \tag{2.65}$$

where $\Delta\boldsymbol{R} = \boldsymbol{r'} - \boldsymbol{r}$ and $H(\boldsymbol{r})$ is the Hessian matrix. Putting this into equation 2.37 gives,

$$\overline{f}_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \frac{1}{2}H(\boldsymbol{r}) \int_{\text{kern}} ((\Delta\boldsymbol{R}) \otimes (\Delta\boldsymbol{R}))G(|\Delta\boldsymbol{R}|, h)d\boldsymbol{r'} + O(|\Delta\boldsymbol{R}|^4). \tag{2.66}$$

using the fact that the kernel function is normalized and the integral of an odd function over the domain is zero. The integral in equation 2.66 can be tabulated for a specific function $G = G(|\Delta\boldsymbol{R}|, h)$, so let's define,

$$\epsilon(\boldsymbol{r}) \equiv \frac{1}{2} \int_{\text{kern}} ((\Delta\boldsymbol{R}) \otimes (\Delta\boldsymbol{R}))G(|\Delta\boldsymbol{R}|, h)d\boldsymbol{r'}. \tag{2.67}$$

We then have,

$$\overline{f}_i(\boldsymbol{r}) \approx f_i(\boldsymbol{r}) + \epsilon(\boldsymbol{r}) \cdot H(\boldsymbol{r}). \tag{2.68}$$

However, $\epsilon(\boldsymbol{r})$ is isotropic $\epsilon(\boldsymbol{r}) = \epsilon\, \boldsymbol{I}$, so we can write

$$\overline{f}_i(\boldsymbol{r}) \approx f_i(\boldsymbol{r}) + \epsilon \boldsymbol{\nabla}^2 f_i(\boldsymbol{r}). \tag{2.69}$$

Chapter 3

Numerical models of supermassive black holes

Stellar thermal emission dominates the light that we receive from the majority of galaxies, yet there is a small subset of galaxies for which another non-thermal source dominates – we call these active galaxies. These sources typically emit over the entire range from radio to X-ray wavelengths, and the emission usually emanates from the central few parsecs of the galaxies (Fabian, 2012; EHT-Collaboration, 2019; Pasini et al., 2020). For this reason, we refer to them as active galactic nuclei (AGN) and, in some cases, the power output of these sources outshine the entire host galaxy itself (Fiore et al., 2017). This begs two questions: (a) what causes AGN and (b) does this massive power output affect galaxy evolution?

The contemporary explanation for AGN relies on the supermassive black hole (SMBH) paradigm, where matter accreting onto a SMBH drives the powerful emission from an active galaxy's nucleus (Lynden-Bell, 1969; Netzer, 2015). Two arguments support this hypothesis: dynamical evidence indicates that most nearby galaxies host SMBHs with masses in the range $10^6 - 10^9$ $M_\odot$ (McConnell & Ma, 2013), and SMBHs are the only astrophysical objects spatially small and massive enough to explain the observed large concentration of energy output.

AGN are of paramount importance in the evolution of galaxies across cosmic time. At the center of each AGN is a supermassive black hole (SMBH) that accretes gas, efficiently converting gravitational potential energy into energetic feedback. Their intense energetics are usually invoked to explain the suppression of cooling flows in the intracluster medium (see Chapter 1) and to explain the rapid quenching of star formation in high-redshift galaxies (Bower et al., 2006). However, it is difficult to prove observationally that their energetic feedback is truly the cause of quenching in these systems, or that the energy couples strongly to their environments. For that reason it is important that theoretical astrophysicists have physically motivated

models of AGN (i.e. SMBHs) that will provide insight from cosmological simulations as to their true impact on galaxy evolution.

Energetic feedback is the main route through which SMBHs impact galaxy evolution, and there are two classes of feedback that have been identified observationally: radio mode and quasar mode (King & Pounds, 2015; Morganti, 2017). These modes are usually divided based on the luminosity of the AGN compared to the Eddington luminosity,

$$L_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p} c}{\sigma_{\rm T}}, \tag{3.1}$$

where $M_{\rm BH}$ is the mass of the SMBH, $m_{\rm p}$ is the mass of the proton, c is the speed of light, and $\sigma_{\rm T}$ is the Thomson cross section of the electron. The Eddington luminosity represents the maximum luminosity after which gas will no longer be able to accrete onto a SMBH due to radiation pressure on the in-falling gas. Powerful large-scale jets are observed when the AGN has luminosities $L_{\rm AGN}$ that are a few percent of $L_{\rm Edd}$ (i.e. low accretion efficiency) (McNamara & Nulsen, 2007; Nemmen et al., 2007). Conversely, the quasar mode is associated with highly efficient accretion, i.e. high $L_{\rm BH}/L_{\rm Edd}$ (Nemmen et al., 2007), and are more ubiquitous in a variety of galaxies than the radio mode (King & Pounds, 2015).

Radio mode feedback is observed in elliptical galaxies, galaxy groups, and clusters where there is abundance hot gas, or an intracluster medium (ICM). The hot medium is heated by narrow supersonic ($\sim 10^4$ km s^{-1}) jets originating from the brightest cluster galaxies (Fabian, 2012). The jets inflate enormous radio bubbles – cavities in the ICM that emit strongly in the radio bands – that shock the ICM and deposit heat as they inflate on $\sim$ 10kpc scales (Karen Yang & Reynolds, 2016; Cielo et al., 2018). AGN jets are not always powered-on but exhibit variability and experience periods of inactivity. During the inactive periods, the inflated bubbles rise buoyantly in the ICM and drive turbulence at the boundaries from shearing instabilities (Prasad et al., 2015).

The quasar mode is observed as a very luminous source but also associated with highly ionized outflows with lower velocities ($\sim 100 - 1000$ km s^{-1}) on much smaller scales ($\sim$ kpc) than the radio mode jets (Perna et al., 2017). The SMBH still drives jets on smaller scales, but the energy becomes more isotropic on the larger scales, possibly impacting the galaxy's interstellar medium directly.

One of the primary suggestions that AGN have an effect on their host galaxies

is the observed relationship between SMBH mass and the stellar properties of their galactic hosts. That relationship is the $M_{\rm BH}$ - σ_* relationship that relates the SMBH mass, $M_{\rm BH}$, to the velocity dispersion, σ_*, of the stellar bulge[1] of its host (Magorrian et al., 1998; Gebhardt et al., 2000; Tremaine et al., 2002; Gültekin et al., 2009). That relationship is usually further cast into a relationship between the SMBH mass and the stellar mass of the galaxy's bulge, $M_{\rm BH}$ - M_* (Kormendy & Ho, 2013). Typical values are $M_{\rm BH} \sim 10^{-3} M_*$, with a logarithmic slope of $\sim 1-2$ (Bentz & Manne-Nicholas, 2018).

AGN modelling is usually divided into four aspects involving the physics of the SMBHs: seeding, accretion, feedback, and dynamics. Seeding refers to how the SMBHs came to exist in the Universe, accretion is how the SMBHs take gas from their environments and grow over time, feedback is how the SMBHs impact their environments, and dynamics is how the SMBHs interact gravitationally. Each aspect has modelling challenges, both numerical and physical. In this Chapter, I put all of my work into a single project that involves improving the sub-grid models for each one of the aforementioned aspects of SMBH physics.

3.1 The Four Aspects

To better organize the physics and implementation of the AGN models, it is easier to divide the topic into 4 separate categories (or aspects): (a) formation, (b) accretion and growth, (c) feedback, and (b) dynamics. There are a wide range of physical models for each aspect, and for each model there are a variety of implementations for testing their applicability in specific simulation codes.

3.1.1 Seeding

There are three main avenues through which SMBHs in the early Universe could have formed (Volonteri, 2010):

- The very first stars in the early Universe should form in dark matter mini-halos of mass $M \approx 10^6$ $M_\odot$ at $z \sim 20-50$, and in some situations end their lives by collapsing into black holes. These are the Population III (henceforth Pop-III)

[1]The bulge is the spheroidal component of a galaxy. Some galaxies might be bulge-dominated while others may have no bulge. Our Milky Way has a bulge and stellar disc.

stars[2] and studies suggest that they were very massive, with masses estimated greater than 100 $M_\odot$ (Madau & Rees, 2001). The most massive of the Pop-III stars ($M_* \gtrsim 260\ M_\odot$) should directly collapse into black holes of sufficient mass to explain the existence of SMBHs in the early Universe (Fryer et al., 2001). While this is a plausible avenue to explore, the uncertainties in Pop-III stellar models are large (Clarkson & Herwig, 2020), and the initial mass distribution of Pop-III stars is unknown.

- Early, unenriched proto-galaxies could provide an adequate environment for the direct collapse of a gas cloud into a black hole. The gas must be unenriched as enriched gas cools quickly – favouring fragmentation and star formation rather than direct collapse. Spaans & Silk (2006) suggest that this occurs for $T \gtrsim 10^4$ K and $Z \lesssim 10^{-4}\ Z_\odot$, under the assumption of an isothermal collapse.

- If the gas in the early proto-galaxies favoured fragmentation as opposed to direct-collapse (these are mutually exclusive), then it is possible that the first episode of star formation caused the formation of very compact nuclear star clusters (Biernacki et al., 2017). The collisions between these stars could cause the formation of a very massive star, possibly leaving a massive black hole as a remnant, in the range $M_{\rm BH} \sim 10^2 - 10^4\ M_\odot$.

The majority of current simulation studies do not attempt to physically model the formation of SMBHs in the early Universe, and instead rely on the observed $M_{\rm BH} - \sigma_*$ relationship (Schaye et al., 2015; Anglés-Alcázar et al., 2015; Sijacki et al., 2015; Choi et al., 2017; Weinberger et al., 2017; Davé et al., 2019). These studies usually artificially place a SMBH at the center of a dark matter halo or galaxy that has reached a certain mass threshold, since σ_* usually correlates with mass of the host system. This presents a major issue as simulations should predict this relationship, not start with the assumption of its existence. Even more troubling is that the assumption relies on having the same relationship at all times in cosmic history. That may not be the case as it is perfectly plausible that the relationship does not hold in the early Universe. Only two groups physically motivate the formation of the first SMBHs in their simulations: Biernacki et al. (2017) and Tremmel et al. (2017).

Biernacki et al. (2017) choose to place SMBHs of mass $10^6\ M_\odot$ in the centers of gas clumps with masses $10^8\ M_\odot$, and follow the co-evolution of the SMBH with

[2]They are not the third generation of stars, but rather the first. Population II stars are the next generation, followed by Population I stars.

a nuclear star cluster (NSC). They model the NSC using a sub-grid model as they cannot resolve the individual stars and SMBH system at simulation time.

Tremmel et al. (2017) choose to assume the gas cloud direct collapse model for SMBH seed formation in their simulations. At simulation time, if a gas particle is transforming into a star particle they apply additional criteria to decide whether it should instead become a SMBH. These criteria are: a low-mass fraction of metals ($Z < 3{\times}10^{-4}$), a gas density 15 times the star formation threshold, and a temperature between 9.5×10^3 K and 10^4 K. They selected a slightly higher metallicity threshold than predicted by theory in order to ensure that SMBHs do not form far away from the densest regions in the simulation. If the gas converting into a star particle meets the criteria, the gas particle is instead converted into a SMBH of mass 10^6 $\mathrm{M}_\odot$. To conserve mass, they force the new SMBH to consume gas from its immediate neighbourhood. The seed mass of 10^6 $\mathrm{M}_\odot$ is arbitrary and large, but is necessary for their dynamical model to work, as I discuss in Section 3.5.3.

3.1.2 Accretion and growth

Observations show that SMBHs exist with masses in the range 10^6 $\mathrm{M}_\odot$ to monstrous SMBHs of 10^{10} $\mathrm{M}_\odot$ (Kormendy & Ho, 2013; McConnell & Ma, 2013; EHT-Collaboration, 2019). It is not plausible that the physical scenarios I outlined in Section 3.1.1 could produce SMBHs of these masses. Therefore, SMBHs must grow significantly over the course of their lives. There are two pathways through which this occurs: (a) accretion of material and (b) mergers with other SMBHs. Our current theoretical understanding of how SMBHs grow through route (a) relies on diffuse gas accretion, tidal torques that push gas toward the SMBH, or through cold precipitation in certain environments (Gaspari et al., 2017). My primary focus is on gas accretion through diffuse gas and gravitational torques in this Section. I also leave SMBH growth via mergers for Section 3.1.4.

The simplest model for SMBH accretion is the Bondi accretion model (Bondi & Hoyle, 1944). Consider a point mass $M_{\rm BH}$ moving with speed $v_{\rm BH}$ through an infinite gaseous medium of density ρ_∞ and sound speed $c_{\rm s,\infty}$. Ignoring the self-gravity of the gas, the accretion of material onto the point mass follows (Bondi, 1952):

$$\dot{M}_{\rm BH} \equiv \frac{{\rm d}M_{\rm BH}}{{\rm d}t} = \frac{4\pi G^2 M_{\rm BH}^2 \rho_\infty}{(v_{\rm BH}^2 + c_{\rm s,\infty}^2)^{3/2}}. \tag{3.2}$$

There are several problems if we apply the Bondi model to gas within a cosmological

simulation. It assumes that the angular momentum of the infalling gas is negligible, and that the infalling gas undergoes isothermal collapse. For the former point: it is impossible for the infalling gas to have zero angular momentum, and any additional angular momentum would lead to a decrease in accretion rate. Realistically, the gas settles into a massive accretion disc before infalling and needs a dissipation mechanism to lose angular momentum before feeding the SMBH. In addition, the luminosity from the accretion disc leads to a radiation pressure that halts the infalling gas, and limits the accretion rate to a value of

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p}}{\epsilon_{\rm r} \sigma_{\rm T} c}, \tag{3.3}$$

where $m_{\rm p}$ is the proton mass, $\sigma_{\rm T}$ is the Thomson electron cross-section, and $\epsilon_{\rm r}$ is the radiative efficiency of converting accreting matter into energy.

Despite the drawbacks of the Bondi accretion model, it is the most common model implemented in contemporary cosmological simulations (Dubois et al., 2012; Choi et al., 2012; Schaye et al., 2015; Sijacki et al., 2015; Weinberger et al., 2017; Biernacki et al., 2017; Tremmel et al., 2017; Trebitsch et al., 2021). All implementations follow the same basic principle: the SMBH accretes gas stochasically to obey equation 3.2 and is limited to the rate in equation 3.3. However, there are variations that attempt to take into account further sub-grid physics and numerical effects. For example, some groups spatially average the accretion rate (Choi et al., 2012, 2017), limit the accretion based on sub-grid accretion disc models (Schaye et al., 2015), or in many cases boost the accretion rate by an arbitrary factor (Springel, 2005; Sijacki et al., 2007, 2015; Tremmel et al., 2017). The major issue with all the implementations is that equation 3.2 only applies to diffuse, homogeneous gas unlike a lot of cases in nature.

An alternative mechanism involves strong perturbations in the discs of galaxies that drive gas to the center of the gravitational potential. In this situation, merging galaxies or other disc instabilities perturb the stellar disc and, subsequently, the gaseous disc. This causes the gas to shock and dissipate its angular momentum, forcing it toward the center of the galaxy and into the sphere of influence of the SMBH (Hopkins & Quataert, 2011). The resolution of contemporary cosmological simulation is not high enough to resolve such instabilities, and requires a sub-grid model. There is an analytic approximation to the process based on relationships derived from very high resolution simulations (Hopkins & Quataert, 2010, 2011; Anglés-Alcázar et al.,

2013, 2015, 2017b),

$$\frac{\dot{M}_{\rm Torque}}{\rm M_\odot\,yr^{-1}} = \epsilon_{\rm T} f_{\rm d}^{5/2} \left(\frac{M_{\rm BH}}{10^8\,\rm M_\odot}\right)^{1/6} \left(\frac{M_{\rm d}(R_0)}{10^9\,\rm M_\odot}\right) \left(\frac{R_0}{100\,\rm pc}\right)^{-3/2} \left(1+\frac{f_0}{f_{\rm gas}}\right)^{-1}. \tag{3.4}$$

where $\epsilon_{\rm T}$ is a normalization factor taking into account unresolved transport, $f_{\rm d}$ is the gas+stellar disc fraction in the environment, $M_{\rm BH}$ is the SMBH mass, $M_{\rm d}$ is the total disc mass within R_0, $f_0 \approx 0.31 f_{\rm d}^2 (M_{\rm d}(R_0)/10^9\,\rm M_\odot)^{-1/3}$, and $f_{\rm gas}$ is the gas mass fraction of the disc.

One difficulty with this model is that it requires determining how much of a simulated galaxy is in a disc (through $f_{\rm d}$) and how much is more spherically distributed. That could lead to an incorrect estimation of the accretion rate, although it has been shown to converge (Anglés-Alcázar et al., 2015).

A key difference between this model and the Bondi accretion model is that the accretion rate scales as $\dot{M}_{\rm BH} \propto M_{\rm BH}^{1/6}$ rather than $\dot{M}_{\rm BH} \propto M_{\rm BH}^2$. The weak dependence on black hole mass in the torque accretion model ensures that the $M_{\rm BH}$–σ_* relationship holds (Anglés-Alcázar et al., 2017b), whereas Bondi accretion causes runaway growth at high $M_{\rm BH}$. As the SMBHs accrete material, the corresponding feedback must always drive gas out of the system such that the mass of the black hole scales with the host system. However, increasing the accretion strength requires increasing feedback to stop gas from moving into the SMBH, and keep it on the $M_{\rm BH} - \sigma_*$ relationship. This implies there is a strong degeneracy between SMBH accretion strength and AGN feedback in cosmological simulations that include Bondi accretion. The torque-driven accretion model removes the degeneracy, and ensures the observed relationship holds[3].

3.1.3 Feedback

The SMBH engine that powers bright AGN releases tremendous amounts of energy. However, SMBHs are only approximately $1/1000^{\rm th}$ the mass of the spheroidal component of their host galaxies. How could such an under-massive object affect the properties of a galaxy?

Consider a galactic bulge with a velocity dispersion σ^2. The binding energy of the bulge is $E_{\rm bulge} \sim M_{\rm bulge}\sigma^2$ and the energy released from a SMBH is $E_{\rm BH} \sim \eta M_{\rm BH} c^2$ with some radiative efficiency η. If we assume a radiative efficiency of $\approx 10\%$ and $\sigma \sim 10^2$

[3] I would debate that *ensuring* the relationship holds is the wrong approach. The $M_{\rm BH} - \sigma_*$ relationship might be different at high redshifts, and if we trust the torque accretion model for reproducing it at $z = 0$, we may lead ourselves to trouble.

$\mathrm{km\,s^{-1}}$ for a typical galaxy, then $E_{\mathrm{BH}}/E_{\mathrm{bulge}} \sim 10^3$ (Fabian, 2012). This simple, order-of-magnitude argument shows that a SMBH may release a huge amount of energy compared to the binding energy of the bulge – more than enough to impact the host system. However, the true extent and impact remains to be determined because it depends on how efficiently the energy couples to the surrounding medium. There are two main modes through which this energetic feedback occurs: the radiative mode and the kinetic mode.

First, the radiative mode (or *quasar* mode) involves SMBHs that accrete near the Eddington limit (equation 3.3). As gas makes its way to the vicinity of the SMBH, it flattens into an accretion disc as it is difficult for the gas to lose angular momentum. Internal friction within the accretion disc gas causes heating, and the system resembles a black body. The gas heats to temperatures high enough to produce ultraviolet radiation that is able to drive a wind of hot gas away from the disc. Eventually, the hot gas may end up pushing out cold gas from the system as it is swept up in a mechanical outflow. Radiative mode feedback usually involves moving cold gas about, and preventing additional gas from reaching the SMBH and has been suggested to be one of the primary drivers of the $M_{\mathrm{BH}} - \sigma_*$ relationship (Murray et al., 2005).

Second, the kinetic mode (or *radio mode*) involves SMBHs that do not accrete efficiently. They impact their surrounding environments through mechanical means. The primary source of energy here is powerful relativistic jets that begin close to the SMBH and eventually extend far outside of the host galaxies (~ 10 kpc). The jets inflate bubbles into the hot gas within galaxy clusters, possibly preventing the gas from cooling and accreting into the massive central brightest cluster galaxy.

In cosmological simulations, the quasar mode was the first AGN feedback model introduced (Springel et al., 2005a) and has remained the primary model in most simulations. The most common implementation involves taking the calculated accretion rate $\dot{M}_{\mathrm{BH}}$ (usually through Bondi accretion) and calculating the energy that the SMBH would release in a time step Δt, given a radiative efficiency (usually $\eta \sim 0.1$). The SMBH would then release energy

$$\Delta E_{\mathrm{BH}} = \epsilon_{\mathrm{f}} \eta \dot{M}_{\mathrm{BH}} c^2 \Delta t, \tag{3.5}$$

where ϵ_{f} is the coupling fraction (usually $\epsilon_{\mathrm{f}} \sim 0.05 - 0.15$) to the surrounding medium. The original study in Springel et al. (2005a) showed that combining Bondi accretion (equation 3.2) with the feedback in equation 3.5 allowed SMBHs to remain on the

$M_{\rm BH} - \sigma_*$ relationship.

The kinetic mode has received less attention, especially the treatment of a jet in a cosmological simulation. Dubois et al. (2012) and Davé et al. (2019) add kinetic jets to their full cosmological simulations. Other groups, such as the Illustris collaboration, add kinetic energy surrounding their SMBHs at simulation time in order to mimic the physical injection of kinetic energy (Weinberger et al., 2017). Despite the intense development of AGN feedback models, none of the groups have successfully captured all of the observed impact of AGN feedback on their host systems, especially the more massive systems (Oppenheimer et al., 2021).

3.1.4 Dynamics

The dynamical evolution of SMBHs remains a challenge for cosmological simulations. There are two aspects to consider: How do we capture the merger of two SMBHs? And, how do SMBHs move through their environments?

First, almost all studies treat SMBH mergers in the same fashion, where two SMBHs automatically merge when they are within their respective gravitational resolution scales and their relative velocities are lower than some characteristic velocity. Implementations include, for example: the relative velocity of the SMBHs being less than their mutual escape velocity (Schaye et al., 2015; Tremmel et al., 2017; Anglés-Alcázar et al., 2017a), and the relative velocity being less than the local soundspeed of the gas (Choi et al., 2017). Sijacki et al. (2015) argued that the velocity is meaningless in artificial repositioning schemes and, therefore, they do not take into account the relative velocities – promptly merging the SMBHs (also see Weinberger et al. 2017 for IllustrisTNG). These two conditions are adequate for capturing whether or not two SMBHs merge, but they do not capture the timescale of the process[4].

Second, a supermassive blackhole (SMBH) moving through a gravitational field experiences a friction force due to its own gravitational wake. This dynamical friction causes SMBHs to lose energy in their orbits and slowly sink to the center of their host galaxies. In cosmological simulations, the resolution is often at the $\sim$ kpc scale and, therefore, they are unable to capture the small-scale dynamical friction forces. The inability of simulations to resolve the small-scale forces has led most researchers to simply force the SMBHs in their simulations to the gravitational potential center of their respective dark matter halos. This method guarantees that each galaxy has its

[4]The timescale of two merging SMBHs is actually unknown because of the *final parsec problem (Milosavljevic & Merritt, 2003)*.

own SMBH and that they will merge when the galaxies merge. Although this is the dominant method, there are alternatives.

The best alternative is to model the dynamical friction on the sub-grid scale. That idea was introduced by Tremmel et al. (2015) and involves calculating the acceleration $\boldsymbol{a}_{\rm DF}$ due to dynamical friction formula from Chandrasekhar (1943),

$$\boldsymbol{a}_{\rm DF} = -4\pi G M_{\rm BH} \rho(< v_{\rm BH}) \ln(\Lambda) \frac{\boldsymbol{v}_{\rm BH}}{v_{\rm BH}^3}, \tag{3.6}$$

where $\rho(< v_{\rm BH})$ is the density of matter moving slower than the SMBH, $\ln(\Lambda)$ is the Coulomb logarithm, $v_{\rm BH}$ is the velocity of the SMBH, and $v_{\rm a}$ is the velocity of a nearby particle. The Coulomb logarithm depends on the maximum impact parameter $b_{\rm max}$ and the minimum $b_{\rm min}$ as $\ln(\Lambda) \sim \ln(b_{\rm max}/b_{\rm min})$. To ensure only the sub-grid contribution counts, Tremmel et al. (2015) chooses the values of $b_{\rm max}$ equal to the resolution scale and $b_{\rm min}$ to be the Schwarzchild radius of the SMBH[5].

3.2 My approach

The main idea of my work on supermassive black hole (SMBH) modelling is to combine the *best* models available into a single simulation so that we may better trust our understanding of brightest cluster galaxy evolution. It is important to note that *best* is subjective without a reasonable metric, something like a convergence test may be appropriate. However, the SMBH sub-grid approximations are dependent on numerical resolution and, therefore, the parameter values should change to reflect the new way the SMBH couples to the medium. In my tests, I use the galaxy stellar mass function and $M_{\rm BH} - M_*$ relationship to calibrate my models and determine their trustworthiness as I show in Section 3.5. First, I discuss the models I determined to be the most physically motivated.

I choose to model the movement of gas into the environment of the SMBHs via a Bondi accretion estimator (equation 3.2) combined with an estimator based on gravitational torques (equation 3.4; henceforth, *torque accretion*) in surrounding gas discs,

$$\dot{M}_{\rm inflow} = \dot{M}_{\rm Bondi} + \dot{M}_{\rm Torque}. \tag{3.7}$$

However, I only compute the Bondi accretion rate based on the hot gas ($T > 10^5$ K)

[5]The *point of no return* for all material moving towards a black hole in a flat space-time.

surrounding the SMBH since the Bondi estimator is derived for hot, diffuse gas. I use the torque estimator on the remaining cold gas in the region. This method was contemporaneously and independently implemented in Davé et al. (2019) as I was working on the model. I use the code from that work for my research.

For the seeding procedure, I follow the prescription of Tremmel et al. (2017) who use a density, temperature, and metallicity criteria for determining when a SMBH should form out of a gas particle. However, the nature of my simulations causes a problem with the temperature. I am using the star formation and stellar feedback model from Davé et al. (2017) and Davé et al. (2019) that relies on having an artificial equation of state in dense gas. In brief, at the highest gas densities, the resolution of cosmological simulations is insufficient to capture the multiphase gas within the interstellar medium. Therefore, a common solution is to force gas above the star formation density threshold to rest on an equation of state $T \propto \rho^{4/3}$ to prevent artificial fragmentation. Unfortunately, this makes it impossible to use the temperature condition for seeding SMBHs because the temperature is set by hand to always obey that scaling relationship. I treat the density threshold, $\rho_{\rm seed}$, as a free parameter and fix the metallicity threshold to the value in Tremmel et al. (2017), $Z_{\rm seed} < 3 \times 10^{-4}$.

For the dynamical evolution of SMBHs, I follow the work of Tremmel et al. (2015). I implemented equation 3.6 into the `GIZMO` code. I show the calibration of the model in Section 3.5.3.

3.3 New feedback model

The most significant model development I am contributing to the community is a new finite state machine SMBH model for cosmological simulations. There are three physical regimes which we must consider:

1. $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd} > R_{\rm upper}$,
2. $R_{\rm lower} < \dot{M}_{\rm BH}/\dot{M}_{\rm Edd} \leqslant R_{\rm upper}$, and
3. $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd} \leqslant R_{\rm lower}$,

where $R_{\rm upper}$ is the boundary between the slim disk mode and the traditional quasar mode and $R_{\rm lower}$ is the boundary between the advection-dominated accretion flow (ADAF) mode and quasar mode. The upper boundary is $R_{\rm upper} \sim 0.3$ (Madau et al., 2014) and the lower boundary is $R_{\rm lower} \sim 0.03$ (King & Pounds, 2015). I treat each

regime as a unique state in which the SMBH exists until the conditions are met to transfer to a new state. The transfer condition is dependent on the current state of the SMBH through the *true accretion rate* onto the SMBH ($\dot{M}_{\rm BH}$) not the *calculated mass inflow rate* ($\dot{M}_{\rm inflow}$).

3.3.1 Slim disk

For the high accretion rate regime, I follow Lupi et al. (2016) and use the radiative efficiency from Sądowski et al. (2014) and Madau et al. (2014).

$$\eta_{\rm high}(j,r) = \frac{r}{16}A(j)\left[\frac{0.985}{r+B(j)} + \frac{0.015}{r+C(j)}\right], \tag{3.8}$$

where j is the black hole spin parameter and $r \equiv \dot{M}^{*}_{\rm Edd}/\dot{M}_{\rm BH}$. Note that $\dot{M}^{*}_{\rm Edd} \neq \dot{M}_{\rm Edd}$, it is defined as $\dot{M}^{*}_{\rm Edd} \equiv 16\,L_{\rm Edd}/c^2$ or $\dot{M}^{*}_{\rm Edd} = 1.6\,\dot{M}_{\rm Edd}$. Additionally,

$$A(j) \equiv (0.9663 - 0.9292j)^{-0.5639}, \tag{3.9}$$

$$B(j) \equiv (4.627 - 4.445j)^{-0.5524}, \tag{3.10}$$

and

$$C(j) \equiv (827.3 - 718.1j)^{-0.7060}. \tag{3.11}$$

The true accretion rate onto the supermassive black hole $\dot{M}_{\rm BH}$ is the difference between the large scale mass flow rate and the out-flowing wind (Choi et al., 2012)

$$\dot{M}_{\rm BH} = \dot{M}_{\rm inflow} - \dot{M}_{\rm wind} = \dot{M}_{\rm inflow} - \psi_{\rm slim}\dot{M}_{\rm BH}, \tag{3.12}$$

where I assume a mass loading factor of $\psi_{\rm slim}$ between $\dot{M}_{\rm BH}$ and the wind mass outflow rate, $\dot{M}_{\rm wind}$. Explicitly, I assume the SMBH powers a kinetic wind

$$\frac{1}{2}\dot{M}_{\rm wind}v^2_{\rm wind} = \epsilon_{\rm f,slim}\eta(j,\dot{M}_{\rm BH})\dot{M}_{\rm BH}c^2, \tag{3.13}$$

where $\dot{M}_{\rm wind}$ is the mass outflow rate, $v_{\rm wind}$ is the wind velocity, $\epsilon_{\rm f,slim}$ is a coupling factor, and $\eta(j,\dot{M}_{\rm BH})$ is the radiative efficiency. Therefore,

$$\psi_{\rm slim} = \frac{\dot{M}_{\rm wind}}{\dot{M}_{\rm BH}} = \left(\frac{2\epsilon_{\rm f,slim}}{v^2_{\rm wind}}\right)\eta(j,\dot{M}_{\rm BH})c^2 \equiv \phi\eta(j,\dot{M}_{\rm BH}), \tag{3.14}$$

where I have further defined ϕ in order to show the explicit dependence on η. I treat

$\epsilon_{\rm f,slim}$ and $v_{\rm wind}$ as free parameters and I discuss their calibration in Sections 3.5 & 3.6. Recall that I fix $\eta = \eta(j, \dot{M}_{\rm BH})$ in equation 3.8.

I further define the following accretion rate ratios for $\dot{M}_{\rm BH}$ and $\dot{M}_{\rm inflow}$:

$$\mathcal{R} \equiv \frac{\dot{M}_{\rm BH}}{\dot{M}_{\rm Edd}}, \ \mathcal{R}' \equiv \frac{\dot{M}_{\rm inflow}}{\dot{M}_{\rm Edd}}, \tag{3.15}$$

where $\dot{M}_{\rm Edd}$ is the Eddington accretion rate.

Using equation 3.8 for η, I rewrite equation 3.12 as,

$$\mathcal{R} + \frac{\phi}{16} A(j) \left[\frac{0.985\mathcal{R}}{1 + (5/8)B(j)\mathcal{R}} + \frac{0.015\mathcal{R}}{1 + (5/8)C(j)\mathcal{R}} \right] = \mathcal{R}'. \tag{3.16}$$

I further simplify this equation by finding a common denominator for the terms in the brackets and multiplying through the entire equation. Explicitly, after finding the common denominator, I multiply the entire equation by $[1 + (5/8)B(j)\mathcal{R}][1 + (5/8)C(j)\mathcal{R}] = 1 + (5/8)(B(j) + C(j))\mathcal{R} + (5/8)^2 B(j)C(j)\mathcal{R}^2$ and obtain,

$$\begin{aligned} &\mathcal{R} + \frac{5}{8}(B(j) + C(j))\mathcal{R}^2 + \left(\frac{5}{8}\right)^2 B(j)C(j)\mathcal{R}^3 + \\ &\frac{\phi}{16} A(j) \left(\mathcal{R} + \frac{5}{8}[0.015B(j) + 0.985C(j)]\mathcal{R}^2 \right) \\ &= \mathcal{R}' + \frac{5}{8}(B(j) + C(j))\mathcal{R}\mathcal{R}' + \left(\frac{5}{8}\right)^2 B(j)C(j)\mathcal{R}^2\mathcal{R}'. \end{aligned} \tag{3.17}$$

This is a cubic equation in $\mathcal{R}$ (i.e. $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd}$) since I know $A(j)$, $B(j)$, $C(j)$, $\dot{M}_{\rm Edd}$, and $\mathcal{R}'$ (i.e. $\dot{M}_{\rm inflow}/\dot{M}_{\rm Edd}$) at simulation time. I reduce the equation to

$$\alpha_3\mathcal{R}^3 + \alpha_2\mathcal{R}^2 + \alpha_1\mathcal{R} + \alpha_0 = 0, \tag{3.18}$$

where

$$\begin{aligned}
\alpha_3 &\equiv \left(\frac{5}{8}\right)^2 B(j)C(j) \\
\alpha_2 &\equiv \frac{5}{8}\Bigg[B(j) + C(j) + \frac{\phi}{16}A(j)(0.015B(j) \\
&\qquad +0.985C(j)) - \frac{5}{8}B(j)C(j)\mathcal{R}'\Bigg] \\
\alpha_1 &\equiv 1 + \frac{\phi}{16}A(j) - \frac{5}{8}(B(j) + C(j))\mathcal{R}' \\
\alpha_0 &\equiv -\mathcal{R}'
\end{aligned} \tag{3.19}$$

I solve this equation at simulation time to find $\mathcal{R}$, giving the true accretion rate $\dot{M}_{\rm BH}$.

3.3.2 Quasar mode

I treat the energetics in the quasar mode identically as the slim disk mode, except with a different radiative efficiency η. In the quasar mode, the radiative efficiency is independent of $\dot{M}_{\rm BH}$, or $\eta = \eta(j)$. To reiterate from the above Section, the true accretion rate onto the SMBH is the difference between the large-scale inflow rate and the mass outflow rate,

$$\dot{M}_{\rm BH} = \dot{M}_{\rm inflow} - \dot{M}_{\rm wind} = \dot{M}_{\rm inflow} - \psi_{\rm quasar}\dot{M}_{\rm BH} \tag{3.20}$$

Specifically, the SMBH powers a fully kinetic wind

$$\frac{1}{2}\dot{M}_{\rm wind}v_{\rm wind}^2 = \epsilon_{\rm f,quasar}\eta(j)\dot{M}_{\rm BH}c^2, \tag{3.21}$$

where $\eta(j)$ is the radiative efficiency and $\epsilon_{\rm f,quasar}$ is the fraction of the energy that couples to the surrounding medium. I take $\epsilon_{\rm f,quasar} \equiv \epsilon_{\rm f,slim}$.

3.3.3 Advection dominated accretion flow mode

At the lowest accretion rates, I model an isotropic wind and a jet component emanating from the SMBH (cf. Benson & Babul 2009). Each component has a corresponding independent radiative efficiency $\eta_{\rm wind}$ and $\eta_{\rm jet}$. Additionally, I only use a small fraction of the estimated accretion rate based on observations,

$$\dot{M}_{\rm acc} = f\,\dot{M}_{\rm inflow}, \tag{3.22}$$

where $f \sim 0.01 - 0.1$ (Russell et al., 2015) and $\dot{M}_{\rm acc}$ is the mass inflow rate close to the SMBH. I choose $f = 0.05$. The remaining $(1-f)\,\dot{M}_{\rm inflow}$ leaves the system as an isotropic wind.

Not all of $\dot{M}_{\rm acc}$ actually reaches the SMBH as some of that material is outflowing as a jet or converted into energy. That is represented by the mass balance

$$\dot{M}_{\rm BH} = \dot{M}_{\rm acc} - \dot{M}_{\rm jet} - \frac{L_{\rm rad}}{c^2} - \frac{P_{\rm jet,flow}}{c^2}, \tag{3.23}$$

where $L_{\rm rad} = \eta(j, \dot{M}_{\rm BH})\dot{M}_{\rm BH}c^2$, $P_{\rm jet,flow} = \eta_{\rm jet}(j)\dot{M}_{\rm BH}c^2$, and c is the speed of light. Using these definitions I obtain a result for the true accretion rate,

$$\dot{M}_{\rm BH}(1 + \eta(j, \dot{M}_{\rm BH}) + \eta_{\rm jet}(j)) = \dot{M}_{\rm acc} - \dot{M}_{\rm jet}. \tag{3.24}$$

In the low accretion rate regime, $\eta(j, \dot{M}_{\rm BH}) \propto \dot{M}_{\rm BH}/\dot{M}_{\rm Edd}$. For $\eta_{\rm jet}(j)$ I fit the results of Benson & Babul (2009). I also assume a mass-loading factor of $\psi_{\rm ADAF}$ for the jet such that $\dot{M}_{\rm jet} = \psi_{\rm ADAF}\dot{M}_{\rm BH}$. To obtain our final equation, I first define a new variable to ensure continuity of the radiative efficiency as a function of $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd}$,

$$C \equiv \frac{100}{3}\frac{\eta(j, R_{\rm upper}\,\dot{M}_{\rm Edd})}{\dot{M}_{\rm Edd}}, \tag{3.25}$$

leading to

$$\dot{M}_{\rm BH}^2 + \frac{(1 + \eta_{\rm jet}(j) + \psi_{\rm ADAF})}{C}\dot{M}_{\rm BH} - \frac{\dot{M}_{\rm acc}}{C} = 0. \tag{3.26}$$

I take the positive solution to the quadratic,

$$\dot{M}_{\rm BH} = \frac{1}{2}\left(\left[\left\{\frac{1 + \eta_{\rm jet}(j) + \psi_{\rm ADAF}}{C}\right\}^2 + \frac{4\dot{M}_{\rm acc}}{C}\right]^{1/2} - \left[\frac{1 + \eta_{\rm jet}(j) + \psi_{\rm ADAF}}{C}\right]\right). \tag{3.27}$$

3.3.4 Summary

Fig. 3.1 shows the expected AGN luminosities and radiative efficiencies of my model as a function of true accretion rate onto the SMBH. To compute these values, I use the high accretion rate regime radiative efficiency as a starting point. Then I ensure continuity at each of the accretion rate boundaries, assuming that in the low accretion

rate regime the radiative efficiency scales as $\eta_{\rm low} \propto \dot{M}_{\rm BH}/\dot{M}_{\rm Edd}$. The luminosity comes directly from the radiative efficiency combined with a range of input $\dot{M}_{\rm BH}$. I show the jet efficiencies with a dot-dash line which come from a fit to the results in Benson & Babul (2009). In this work, I fix the jet spin parameter to $j = 0.9$ in order to reduce the final parameter space search.

I have also provided a comparison of sub-grid AGN models across the literature in Table 3.1. The model that I propose is unique in that it combines a physical seeding model, kinetic jet in the radio mode, and has the dynamical friction model.

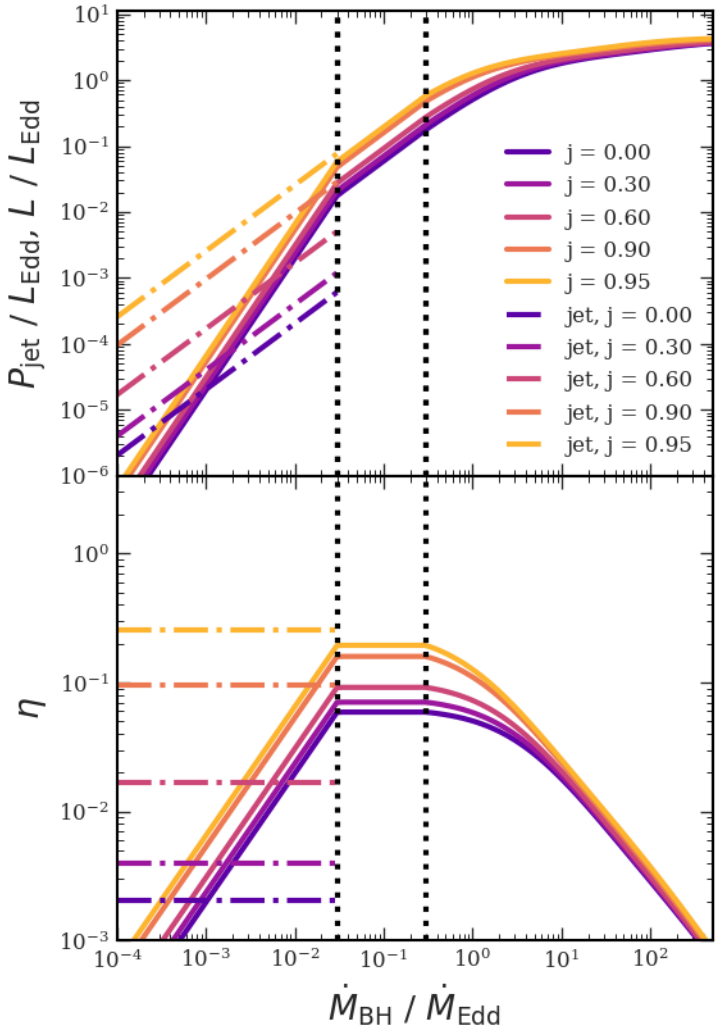

Figure 3.1: (*top*) Power of the active galactic nuclei as a function of its accretion rate. The colors show the supermassive black hole jet spin parameter from $j = 0$ to $j = 0.95$, darkest-to-light shade, respectively. (*bottom*) The radiative efficiency of the active galactic nuclei as a function of its accretion rate.

Table 3.1: SMBH model comparisons. Quasar mode = QM, Radio mode = RM, Eddington-Limited = EL, Density dependent boost factor (Booth & Schaye, 2009) = BS09. Simba: Davé et al. (2019), Romulus-C: Tremmel et al. (2019), EAGLE: Barnes et al. (2017), IllustrisTNG: Weinberger et al. (2017), BAHAMAS: McCarthy et al. (2017), Horizon-AGN: Dubois et al. (2012), Rhapsody-G: Martizzi et al. (2016), Magneticum: Hirschmann et al. (2014), FIRE-2: Anglés-Alcázar et al. (2017a,b).

Suite	Seeding	Accretion	Feedback	Dynamics
This model	Gas properties (ρ, Z)	Bondi, Torque	QM (kinetic) RM (isotropic wind, kinetic jet)	Dynamical friction
Simba	Artificial	Bondi, Torque	QM (kinetic), RM (kinetic jet)	Artificial
Romulus-C	Gas properties (ρ, Z, T)	Bondi (EL, modified)	QM (thermal)	Dynamical friction
EAGLE	Artificial	Bondi (visc. limited)	QM (thermal)	Artificial
IllustrisTNG	Artificial	Bondi (EL)	QM (thermal), RM (no jet)	Artificial
BAHAMAS	Artificial	Bondi (EL, BS09)	QM (thermal)	Artificial
Horizon-AGN	Gas properties (ρ)	Bondi (EL, BS09)	QM (thermal), RM (kinetic jet)	None
Rhapsody-G	Gas properties (ρ)	Bondi (EL, BS09)	QM (thermal)	None
Magneticum	Artificial	Bondi (EL, boosted)	QM (thermal), RM (no jet)	Dynamical friction
FIRE-2	Artificial	Gravitational torques	QM (thermal, kinetic)	Artificial
Barai et al.	Artificial	Bondi (EL, boosted)	QM (thermal, kinetic jet)	None
Choi et al.	Artificial	Bondi (EL)	QM (thermal, kinetic, X-ray)	None

3.4 Particle implementation

In cosmological simulations, we deal with discrete massive particles and discrete time, whereas the preceding formulae are continuous mass estimators. To discretize the process, I follow Anglés-Alcázar et al. (2017b) and Davé et al. (2019) and use a smooth accretion model that accretes fractions of gas particles $f_{\rm acc} M_{\rm gas,part}$ and ejects the remaining $(1 - f_{\rm acc}) M_{\rm gas,part}$ as a kinetic wind. $f_{\rm acc}$ comes directly from the true accretion rate onto the supermassive black hole (SMBH), $\dot{M}_{\rm BH}$, as well as the total inflow rate, $\dot{M}_{\rm inflow}$, as,

$$f_{\rm acc} \equiv \frac{\dot{M}_{\rm BH}}{\dot{M}_{\rm inflow}}. \tag{3.28}$$

For the advection dominated accretion flow (ADAF) mode, I fix $f_{\rm acc,ADAF} = 0.05$ and use $\dot{M}_{\rm acc}$ above instead of $\dot{M}_{\rm BH}$. For the slim disk mode, I use the solution of equation 3.18 as $\dot{M}_{\rm BH}$ which depends on the mass loading factor $\psi_{\rm slim}$ from equation 3.14, giving $f_{\rm acc,slim}$. The mass loading factor in the quasar mode is similar to the slim disk mode,

$$\psi_{\rm quasar} = \frac{\dot{M}_{\rm wind}}{\dot{M}_{\rm BH}} = \frac{2\epsilon_{\rm f,quasar}\eta(j)c^2}{v^2_{\rm wind,quasar}}, \tag{3.29}$$

and, since all of the parameters are constant[6], in the quasar model I then explicitly write

$$f_{\rm acc,quasar} = \frac{1}{1+\psi_{\rm quasar}} = \frac{1}{1+\frac{2\epsilon_{\rm f,quasar}\eta(j)c^2}{v^2_{\rm wind,quasar}}}. \tag{3.30}$$

To further complicate matters, there are two masses for the SMBHs at simulation time: (a) the physical sub-grid mass, $M_{\rm BH,sub}$, and (b) the particle mass, $M_{\rm BH,part}$. The sub-grid mass is initially $M_{\rm BH,sub} < M_{\rm BH,part}$ since our particle resolution cannot reach the smaller physical initial seed mass of SMBHs in the Universe. The initial particle mass of seed SMBHs that originate from gas particles have masses equal to the gas particles themselves, $M_{\rm BH,part,init} = M_{\rm gas,part}$. That presents an issue for the dynamical friction model since the SMBH particle mass must be *at least* as massive as the dark matter particles, as I demonstrate in Section 3.5.3. I heavily modified the `GIZMO` code so that the initial particle masses of the SMBHs would be equal to the dark matter

[6]Recall η is independent of $\dot{M}_{\rm BH}$ in the quasar mode, i.e. $\eta = \eta(j)$.

particle masses. In the very first timestep of the SMBH, I ensure mass conservation by accreting the nearest gas particles to make up for the difference between the gas particle mass and dark matter mass. The SMBH consumes all of that gas in a single timestep and does not feedback into the environment. After that initial step, the simulation continues as described below.

Hereafter, everything I discuss happens in a single timestep Δt of a single SMBH particle. The algorithm is as follows[7]:

1. Search the neighbourhood of the SMBH particle in order to compute the properties needed for the accretion and dynamical friction model. For example, I need the amount of hot ($T > 10^5$ K) and cold ($T < 10^5$ K) gas surrounding the SMBH, within its gravitational resolution limit.

2. Compute $\dot{M}_{\rm inflow}$ via equation 3.7.

3. Assume the SMBH is in its previous state, $S_{\rm i-1}$, to compute $\dot{M}_{\rm BH} = f_{\rm acc,i-1}\dot{M}_{\rm inflow}$. $f_{\rm acc,i-1}$ is the accretion factor for the previous state. Compute the new state, $S_{\rm i}$:

   (a) If $\dot{M}_{\rm BH} \leq R_{\rm lower}$, then $S_{\rm i} \rightarrow$ ADAF.

   (b) If $R_{\rm lower} < \dot{M}_{\rm BH} \leq R_{\rm upper}$, $S_{\rm i} \rightarrow$ Quasar.

   (c) If $R_{\rm upper} < \dot{M}_{\rm BH}$, then $S_{\rm i} \rightarrow$ Slim disk.

4. Based on the new state $S_{\rm i}$, compute the new accretion factor $f_{\rm acc,i}$.

   (a) Assign $\dot{M}_{\rm BH} = f_{\rm acc,i}\dot{M}_{\rm inflow}$.

   (b) Update the sub-grid mass $M_{\rm BH,sub}$ by adding $\dot{M}_{\rm BH}\Delta t$.

5. Search the local neighbourhood for other SMBHs that should merge.

   (a) If a SMBH neighbour is found, check the relative velocity, $v_{\rm rel}$. If $v_{\rm rel} < v_{\rm bound}$, then merge the two SMBHs.

6. Search the local neighbourhood of the SMBH for gas particles to accrete. Each have a mass $\approx M_{\rm gas,part}$.

   (a) Loop over each gas particle neighbour, say k, individually. Keep track of total accreted mass by starting with, $M_{\rm accrete,cumulative} = 0$.

[7] A large part of steps 5-10 of this algorithm was already implemented in `GIZMO`. However, I had to modify the steps to fit with my particular model.

(b) Determine the probability for accretion,
$p_{\rm k} = (M_{\rm BH,sub} - (M_{\rm BH,part} + M_{\rm accrete,cumulative})\frac{W_{\rm k}}{f_{\rm acc,i}\sum W_{\rm k}}$.
$W_{\rm k}$ is the kernel weight of particle k.

(c) Choose a random number $w_{\rm k}$, and if $w_{\rm k} < p_{\rm k}$, add $f_{\rm acc,i}M_{\rm gas,part,k}$ to $M_{\rm accrete,cumulative}$.

(d) Flag the particle as partially accreted.

(e) **Note**: $p_{\rm k} \rightarrow 0$ when $M_{\rm BH,part} + M_{\rm accrete,cumulative} \rightarrow M_{\rm BH,sub}$.

7. By the end of the previous loop, the accreted particles are known. The remaining $(1 - f_{\rm acc,i})M_{\rm gas,part,k}$ of each particle automatically satisfies $\dot{M}_{\rm wind}$ since $\dot{M}_{\rm wind} = (1 - f_{\rm acc})\dot{M}_{\rm inflow}$ from equation 3.7.

   (a) Calculate the wind velocity, $v_{\rm wind,k}$, and update the particle velocity in the direction of the particle's angular momentum with respect to the SMBH, $\pm\hat{\mathbf{j}}$. Note: each particle has a 50% chance of $+\hat{\mathbf{j}}$ and $-\hat{\mathbf{j}}$.

   $$\boldsymbol{v}_{\rm new,k} = \boldsymbol{v}_{\rm old,k} + (v_{\rm wind,k} \cdot \pm\hat{\mathbf{j}})$$

   (b) If we are in state S_i = **ADAF**, then we have to take the jet into consideration.

      i. Calculate the probability of having a jet particle in the ADAF mode via $p_{\rm jet} = \psi_{\rm ADAF}(\dot{M}_{\rm BH}/[\dot{M}_{\rm inflow} - \dot{M}_{\rm BH}])$.

      ii. If some uniformly random number is generated and is less than $p_{\rm jet}$, then the particle is identified as a jet event and its properties are set accordingly.

      iii. If the jet particle is not identified, continue to 7a for the isotropic ADAF wind.

8. Subtract the equivalent mass that was radiated to power the wind:
$M_{\rm BH,sub,new} = M_{\rm BH,sub} - \eta(j, \dot{M}_{\rm BH})\dot{M}_{\rm BH}c^2\Delta t$.

9. Ensure momentum is conserved by transferring momentum from accreted gas into the SMBH.

10. Apply the dynamical friction acceleration to the SMBH from equation 3.6:
$\boldsymbol{a}_{\rm BH,new} = \boldsymbol{a}_{\rm BH} + \boldsymbol{a}_{\rm DF}$.

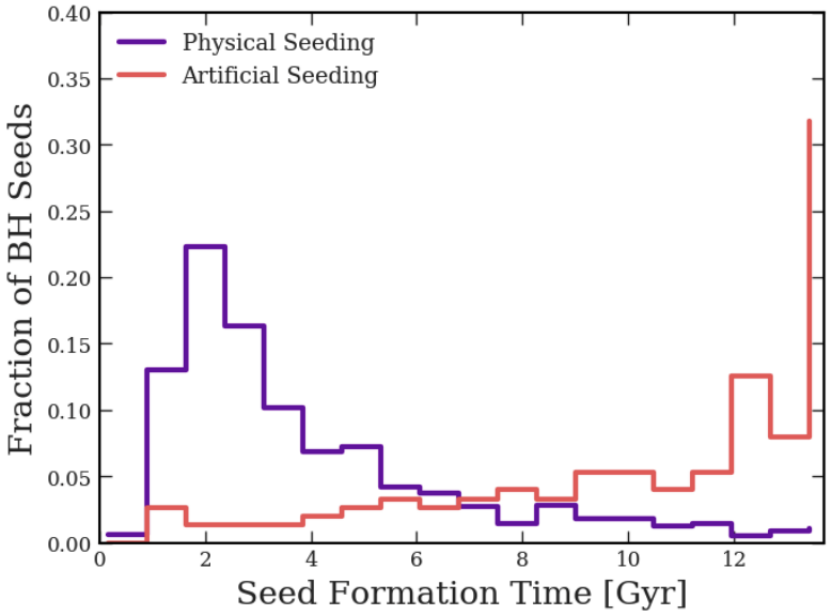

Figure 3.2: A histogram of formation times of supermassive black hole seeds in a cosmological simulation. The darker curve shows the formation times of the physical seeding prescription based on the gas properties, whereas the lighter curve shows an example of artificial seeding. The physical seeding prescription produces many more supermassive black holes early in the history of the Universe, aligning more with our theoretical understanding.

3.5 Verification

The goal of combining these models is to have a functioning active galactic nuclei (AGN) model for cosmological-scale simulations, in order to study brightest cluster galaxy evolution. However, before a full calibration of all of the models, it is necessary to check that each of the components work individually. In some cases, there is no *known* solution and we must accept qualitative verification of the models.

3.5.1 Seeding

To illustrate the impact of the more physical seeding model, Fig. 3.2 shows a histogram of the supermassive black hole (SMBH) formation times in two cosmological simulations. Each simulation has identical initial conditions, and I ran them with the `Simba` model (Davé et al., 2019) (see Chapter 2, Section 2.2.4 for more details) – only altering the creation method for SMBHs in the simulation. I label the model that places SMBHs in resolved galaxies as *Artificial Seeding* (from Davé et al. 2019,

as one example) and the model based on the gas properties as *Physical Seeding*. The differences are stark early in the simulation. Most of the SMBHs in the physical seeding scenario form before ~ 5 Gyr whereas the distribution reverses in the artificial seeding case. The *Physical Seeding* results are more accurate given our knowledge of how SMBHs might form through the direct collapse of a gas cloud into a single massive object (Volonteri, 2010; Ardaneh et al., 2018).

3.5.2 Accretion and feedback

The impact of SMBH accretion and feedback on verification measures are both highly intertwined. For example, if a SMBH does not accrete it naturally does not feedback simply because the power is $L_{\rm BH} = {\rm d}E/{\rm d}t \propto \dot{M}_{\rm BH}$. The best method for verifying that the accretion and feedback models function properly in a cosmological simulation is to check the $M_{\rm BH}-M_*$ relationship and the galaxy stellar mass function (Schaye et al., 2015). The $M_{\rm BH}-M_*$ measure immediately reveals if there is some issue with accretion since the SMBHs will grow either too massive, or grow insufficiently and move off of the observed relationship (for example, see Anglés-Alcázar et al. 2017b). However, it requires that I assume that the $M_{\rm BH}-M_*$ relationship holds at all redshifts since there is no other measure to test at high-redshift. The galaxy stellar mass function shows the impact of feedback as simulated galaxies may be destroyed by too much energy being deposited by the SMBHs, or grow too much if there is insufficient feedback. Luckily, there are observations of the galaxy stellar mass function out to high redshift ($z \sim 8$; Song et al. 2016).

Testing both of the above relationships requires a parameter study in cosmological-scale simulations for sufficient galaxy statistics. In my initial tests, I chose to first calibrate the wind velocity of the quasar/slim disk mode. *I will only discuss this calibration, as it has important impact on the model as a whole.* For the other sub-grid models, I used the `Simba` model that has models for cooling, star formation, and stellar feedback (Davé et al., 2019). To isolate the impact of my new accretion and feedback model, I also used the seeding and dynamics prescriptions that are built into `Simba` since they are well-tested. Table 3.2 summarizes the parameters for the feedback model, cosmology, and the simulation details. My simulations use the same resolution as `Simba` with 256^3 dark matter and gas particles ($M_{\rm gas} = 2\times10^7$ M$_\odot$, $M_{\rm DM} = 10^7$ M$_\odot$) and volume $\approx (37{\rm Mpc})^3$.

I tested values for the quasar/slim disk mode wind velocity in the range $500\,{\rm km\,s^{-1}}$

Table 3.2: I use the cosmological parameters from `Simba` (Davé et al., 2019), that are consistent with the values in Planck Collaboration XIII (2015). The second set of parameters below are the simulation parameters.

Parameter	Value
$\Omega_{\rm m}$	0.3
Ω_Λ	0.7
$\Omega_{\rm b}$	0.048
h	0.68
σ_8	0.82
$n_{\rm s}$	0.97
$z_{\rm start}$	127
$N_{\rm gas}$	256^3
$N_{\rm DM}$	256^3
Volume	$\approx (37\,{\rm Mpc})^3$
$\epsilon_{\rm f,quasar}$	0.05
$\epsilon_{\rm f,slim}$	0.05
$f_{\rm acc,ADAF}$	0.05

to 10^4 km s^{-1}. The high upper limit in wind velocity is motivated by observations of wind velocities measured at small distances from the SMBH (Fiore et al., 2017). After searching parameter space, I found that the best solution is $v_{\rm wind} = 10^4$ km s^{-1} (cf. Choi et al. 2012).

Fig. 3.3 shows the $M_{\rm BH} - M_*$ relationship at $z = 0$ for the best calibration. The star-points are individual galaxies from the simulation and are coloured by their host dark matter halo virial mass. The dotted line shows a fit to the observed relationship taken from Kormendy & Ho (2013). The results from this single calibration shows that the accretion and feedback combination is able to keep the SMBHs following the observed trend, although going higher in stellar mass the SMBHs are moving above the relation.

Fig. 3.4 shows the galaxy stellar mass function at $z = 0$ for the best calibration. I show two observational results from Tomczak et al. (2014) and Baldry et al. (2012) which have slightly different values around $M_* \sim 10^{11}$ M$_\odot$. The vertical dotted line shows the galaxy resolution cut-off of 64 star particles. I show the best calibration as a purple curve with star points identifying each bin. The wind velocity $v_{\rm wind} = 10^4$ km s^{-1} provides an excellent fit to the observations at $z = 0$. There is some slight noise around $M_* \sim 10^{11}$ M$_\odot$, however that is most likely due to the volume being small and there only being $\lesssim 10$ objects in the upper stellar mass bins.

Evidently, the highest wind velocity $v_{\rm wind} = 10^4$ km s^{-1} provided the best fit to the observed data. In fact, that wind velocity is also used in a closely related model from Choi et al. (2012) who found that the high wind velocities produce realistic X-ray luminosities in their simulated galaxies. However, there are important points that I must address that lead to a fundamental restructuring of the model I outlined in Section 3.3.

Recall that my implementation of the quasar and slim disk mode involves using the SMBH luminosity to drive a completely kinetic wind, or the SMBH power $\epsilon_{\rm f}\eta\dot{M}_{\rm BH}c^2$ goes directly into the wind's kinetic power $(\dot{M}_{\rm wind}v^2_{\rm wind})/2$. Consider the quasar mode whose accretion fraction is $f_{\rm acc,quasar} = 1/(1 + \psi_{\rm quasar})$. For a value of $v_{\rm wind} = 10^4$ km s^{-1}, the parameters in Table 3.2, and the radiative efficiency $\eta(j) \approx 0.159$, I get $f_{\rm acc,quasar} \approx 0.065$. That implies that 6.5% of each particle labelled for accretion will be accreted, and the other 93.5% ejected in a kinetic wind. All of the gas particles in the simulation have approximately the same mass, $M_{\rm gas} \approx 2 \times 10^7$ M$_\odot$. That means that any time the simulation selects a particle to be ejected, it will always have an approximate mass of $M_{\rm wind} \approx 1.87 \times 10^7$ M$_\odot$. A typical timestep in these simulations

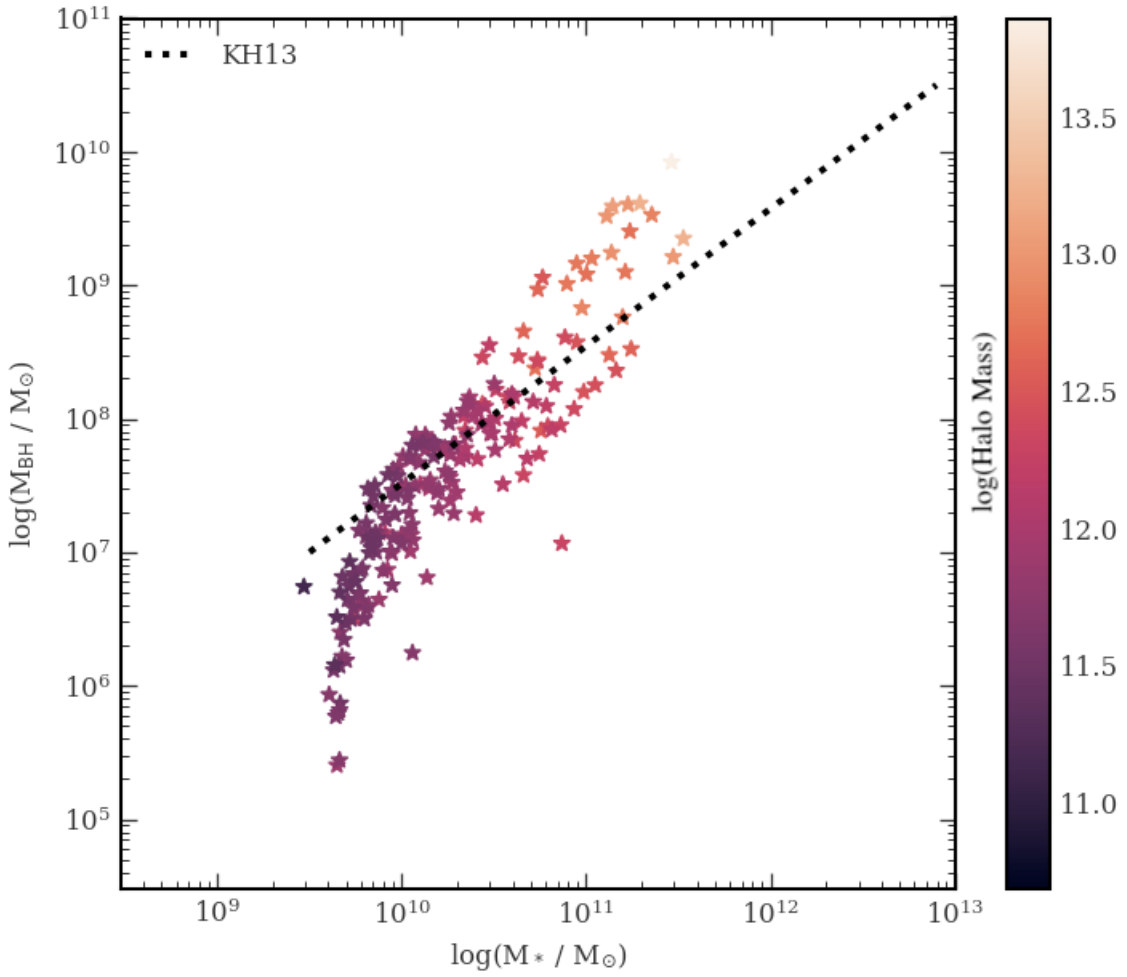

Figure 3.3: The relationship between supermassive black hole mass and galaxy stellar mass in one of my calibration simulations. The dotted line shows a fit to the observed relationship from Kormendy & Ho 2013. The colouring shows the host halo mass (virial mass) of each galaxy. The accretion model provides a good fit to the observations, although there is some divergence at high stellar masses.

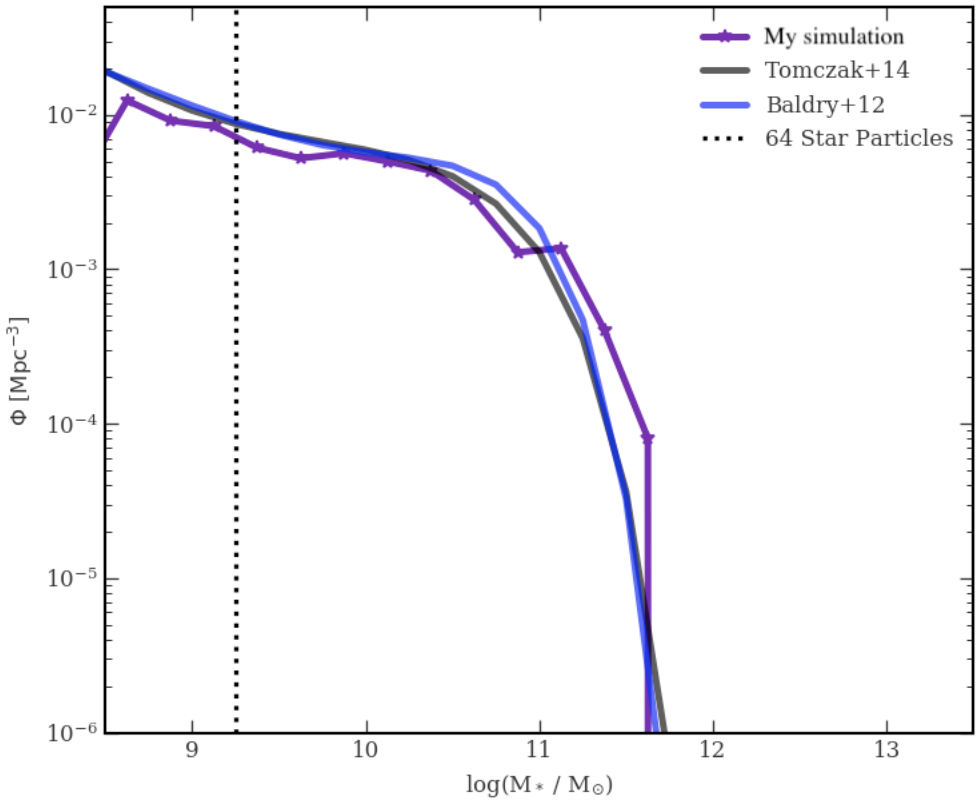

Figure 3.4: The galaxy stellar mass function at $z = 0$ for a single calibration test with the best fit. I show two observed relationships from Tomczak et al. (2014) and Baldry et al. (2012). The purple-star curve shows the results from my calibration. The dotted vertical line shows the mass of 64 star particles, a measure of resolution of the galaxies. The calibration gives a good fit to the observed relationships.

is $t \sim 1$ Myr. That implies that the luminosity, based on the kinetic energy, of that single particle is $L_{\rm particle} \sim 6 \times 10^{44}$ erg s^{-1}. Now compare with the fact that a SMBH with mass $M_{\rm BH} = 10^6$ M$_\odot$ has an Eddington limit of $\dot{M}_{\rm Edd} \approx 0.022$ M$_\odot$ yr^{-1}, or an Eddington luminosity of $L_{\rm Edd} \approx 2 \times 10^{44}$ erg s^{-1}. Therefore, every single kinetic wind event, regardless of accretion rate, is ~ 3 times more luminous than one measure of the maximum luminosity, $L_{\rm Edd}$. It is possible that there are multiple particles selected for each accretion event, amplifying the problem.

I discovered this issue while examining the entropy of the diffuse gas in dark matter halos from a subsequent test simulation. Specifically, I examined the entropy profiles of the most massive halos in a simulation with the same resolution and sub-grid models, but in a much larger volume of $\approx (74\,{\rm Mpc})^3$. The entropy of gas, K, within massive dark matter halos is predicted to be a power law as a function of radius, $K(r) \propto r^{1.1}$ (Kravtsov & Borgani, 2012). The lowest entropies are in the central regions of the halo where the densities are the highest and the cooling times are the shortest. Energetic processes such as stellar and AGN feedback increase the entropy in the core regions, flattening the profile away from the theoretically predicted slope. While there are indeed observed galaxy clusters with flatter cores, there are very few groups ($M \lesssim 10^{14}$ M$_\odot$) that deviate from a power law profile (O'Sullivan et al., 2017; Oppenheimer et al., 2021). The goal is to produce a population of galaxy groups and clusters that can reproduce both observed populations.

Fig. 3.5 shows an example of entropy profiles of the gas in the top ten most massive galaxy groups in the high volume simulation. The colored lines are the profiles of the halos and I leave them unranked by mass as it is irrelevant. The solid black curve shows the expected entropy profile. I scaled the entropy K to its value at $R500$, where $R500$ is the radius at which the enclosed mass is 500 times the critical density of the Universe. In a recent review, Oppenheimer et al. (2021) found that most simulations are unable to reproduce the predicted power law entropy profile exactly within this radius. However, none of the simulations that they studied was completely flat out to $R500$. For these dark matter halos, $R500$ is hundreds of kpc and the flat entropy profile implies that the gas is completely mixed within that region, a truly unphysical result.

After significant testing, I determined that the source of the problem is the wind velocity and the kinetic energy constrained model itself. The momentum in each wind particle is extremely high due to the high mass (i.e. low resolution) of the gas particles, causing the destruction of the halo when the wind particle recouples to

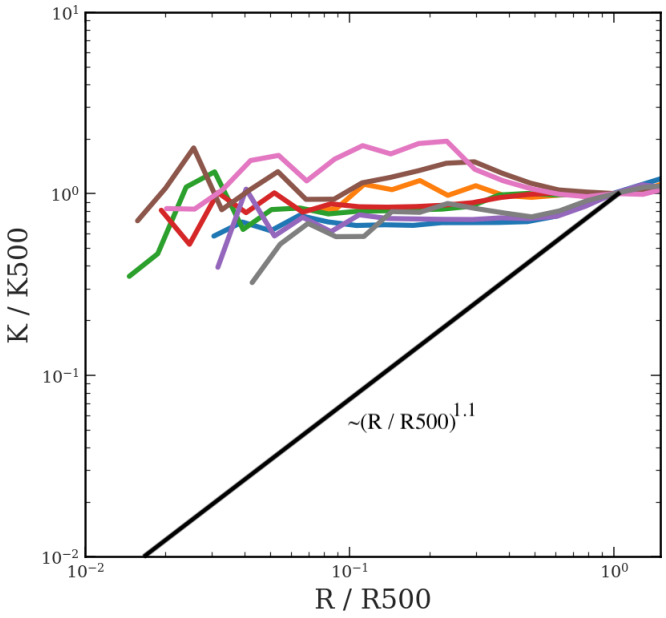

Figure 3.5: The entropy profiles (colored lines) of the gas in the ten most massive halos in a test calibration cosmological simulation with volume $\approx (74\,\mathrm{Mpc})^3$. I scale the entropy to its value at $R500$ for comparison. The straight power-law relationship shows the theoretical expectation, $K/K500 \sim (R/R500)^{1.1}$. All of the halos show unphysically flat entropy profiles in the gas.

the medium. The high-velocity particle will cause the gas to shock and heat, and push gas out of the system. For comparison, the `Simba` model has a better match to the entropy profiles in similar systems, and has quasar wind velocities of order ~ 500 km s^{-1} for a 10^6 M$_\odot$ SMBH. For a similar mass ejection event, a wind particle in `Simba` has effectively 20 times less momentum and 400 times less kinetic energy. To match a similar energy injection rate as the `Simba` simulations, I would have to increase the mass resolution of the simulation to $M_{\rm gas} = 5 \times 10^4$ M$_\odot$, which is higher than any cosmological-scale simulation that has ever existed. Simultaneously, decreasing the mass of the particles by that factor would reduce the momentum injection rate of my model to $1/20^{\rm th}$ that of `Simba`.

The most interesting result is that the calibration provided a good match to the $M_{\rm BH} - M_*$ relationship and the galaxy stellar mass function (see Figs 3.3 & 3.4). My search of parameters revealed that the best solution for the model is to *blow up all of the massive gaseous halos* so that stars never form, fitting the high-mass end of the stellar mass function quite nicely. That shows the huge degeneracy in the parameter space of sub-grid models that can give a seemingly correct solution for a subset of relationships, while simultaneously destroying other relationships[8]. However, increasing the number of relationships to fit does not necessarily help as one could over fit the model and destroy any predictability.

The biggest problem with the kinetic energy constrained model is that it is not resolution independent. What is necessary is a model that is consistent in momentum flux across resolutions. Luckily, the `Simba` model uses a model that should be resolution independent by fixing the momentum flux of the quasar directly, rather than the kinetic energy (Davé et al., 2019). Quasars usually have a momentum flux that is ~ 20 times their luminosities, $\dot{p}_{\rm quasar} \sim 20 L_{\rm quasar}/c$ (Faucher-Giguère & Quataert, 2012). We know that $L_{\rm quasar} = \epsilon_{\rm f,quasar}\eta(j)\dot{M}_{\rm BH}c^2$ and $\dot{p} = \dot{M}_{\rm wind}v_{\rm wind}$ so

$$\dot{M}_{\rm wind}v_{\rm wind} = 20\epsilon_{\rm f,quasar}\eta(j)\dot{M}_{\rm BH}c. \tag{3.31}$$

This equation implies that the mass loading factor will now much higher for fixed wind velocity compared to the kinetic energy constrained model,

$$\psi_{\rm quasar,new} \equiv \frac{\dot{M}_{\rm wind}}{\dot{M}_{\rm BH}} = \frac{20\epsilon_{\rm f,quasar}\eta(j)\dot{M}_{\rm BH}c}{v_{\rm wind}}. \tag{3.32}$$

[8]Parameter whack-a-mole, as I call it.

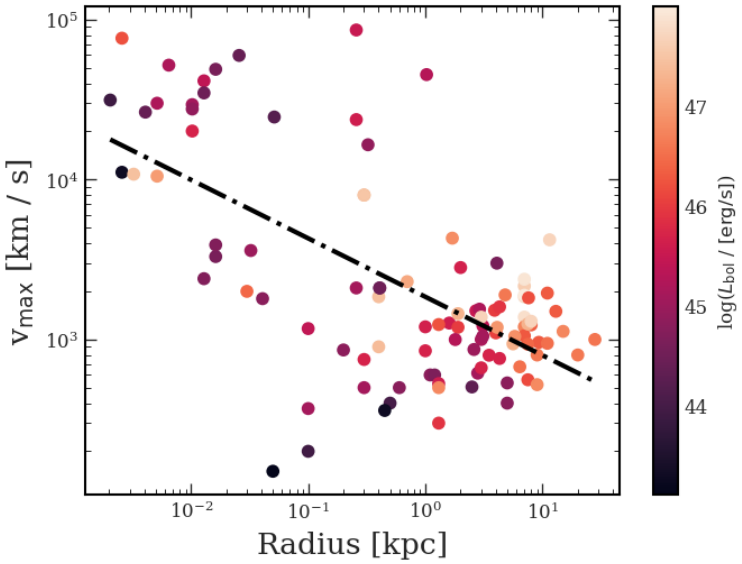

Figure 3.6: Observed active galactic nuclei wind velocities as a function of radius that I compiled from Fiore et al. 2017. I fit a power law and found a relationship $\log(v_{max}) = -0.366 \log(R) + 3.267$.

This gives $\psi_{quasar,new}/\psi_{quasar,old} \approx 30$ for $v_{wind} = 10^4$ km s^{-1}.

The wind velocity, v_{wind}, is itself an issue. I originally followed Choi et al. (2012) for the kinetic energy constrained model, and they use a large value since 10^4 km s^{-1} is observed close to the SMBH in the AGN. However, their resolution is much higher than mine since they were studying a galaxy in isolation. In fact, their low resolution test cases are 100 times better than the calibration tests I presented here! I revisited the review in Fiore et al. (2017) to determine what the best choice of wind velocity would be for the spatial resolution of my simulations ($\approx$ kpc). I compiled the tabular results in Fiore et al. (2017) in order to determine the relationship between the wind velocity measured, v_{max}, and the radius at which that velocity was measured, R. Fig. 3.6 shows the data points from their study coloured by the bolometric[9] luminosity of the AGN. I fit a power law to the scattered points and found,

$$\log(v_{max}) = -0.366 \log(R) + 3.267. \tag{3.33}$$

Using my simulation spatial resolution of $\sim$ kpc gives $v_{max} \approx 1000$ km s^{-1}. On average, it is unlikely to have a 10^4 km s^{-1} wind at that scale, although it is observed in at least one case. The new mass loading in the winds, $\psi_{quasar,new}$, also leads to a new

[9]Total across all wavelengths.

accretion fraction, $f_{\rm acc,quasar,new}$. Using the relationship in equation 3.30 with $\psi_{\rm quasar,new}$ and the new $v_{\rm wind} = 1000$ km s^{-1} gives $f_{\rm acc,quasar,new} = 0.02$.

The results I presented here are very important as they completely change the models that I presented in Section 3.3. The model more closely resembles that of `Simba` now, only I still have the two additional states in place for the low and high accretion rate regimes. In both of those regimes, I now fix the momentum as I just described. Moving forward, my main recommendation for sub-grid AGN models is that they use momentum constrained models as they should be independent of resolution, and work up until our currently achievable resolutions. Once computational power allows for resolutions above 5×10^4 M$_\odot$ in a large-volume simulation, then the kinetic energy constrained model may be worth revisiting.

3.5.3 Dynamics

I follow Tremmel et al. (2015) and use an isolated dark matter halo collapsing with a SMBH orbiting the center to validate the dynamical friction model[10]. Specifically, I place the a SMBH of mass $M_{\rm BH} = 10^6$ M$_\odot$ on an eccentric orbit ($v = 0.1 v_{\rm circ}$) with an apocenter of 2 kpc around the center of a dark matter halo with total mass $M = 3{\times}10^{11}$ M$_\odot$. I evolve the system for 7.5 Gyr and measure the distance of the SMBH from the center of the halo. I find the center of the halo using the shrinking spheres method (Power et al., 2003). Taffoni et al. (2003) provides an analytic estimate for when such an object should fall into the very center of the dark matter halo.

Fig. 3.7 shows the distance of the simulated SMBH from the center of the dark matter halo as a function of time. The horizontal dotted line shows the resolution length (i.e. gravitational softening length) of the SMBH, 1 kpc, as measured from the center. I show that curve as the SMBH must infall within that distance in order to merge with any SMBH in the center, according to the criteria in Section 3.1.4. However, there is no SMBH to merge with in the center and I consider the model successful if it stays below that horizontal line for the remainder of the simulation. The vertical dash-dot line shows the analytic estimate for when the SMBH should get to the center of the halo. The solid curves show various resolutions `R14`, `R17`, `R20`, and `R23` that are the lowest resolution to highest resolution, respectively. `DY` indicates if the simulation has the sub-grid dynamical friction model. The faint line shows the simulation at the highest resolution with no sub-grid dynamical friction

[10]Michael Tremmel graciously provided the code necessary to generate the initial conditions for this test.

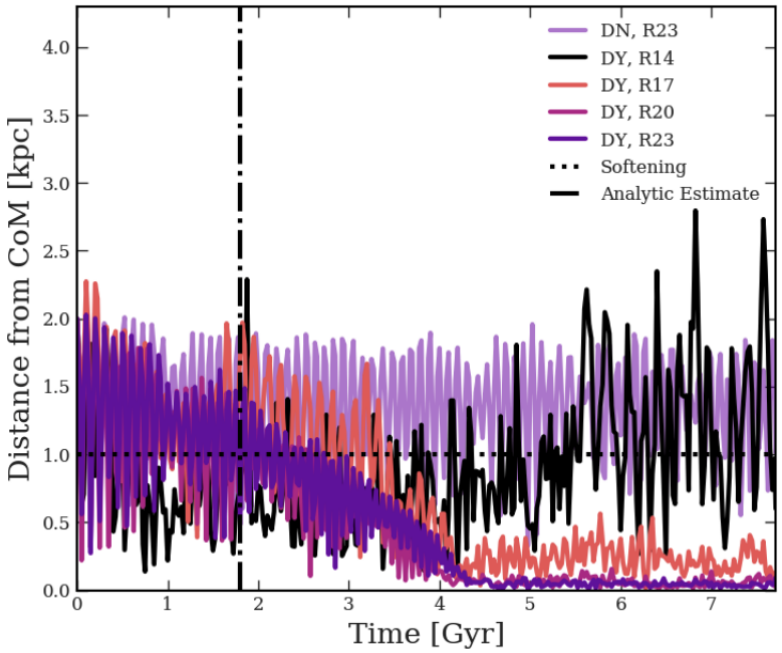

Figure 3.7: The distance (in kpc) of a simulated supermassive black hole from the center of a dark matter halo across time in Gyr. DY: Sub-grid dynamical friction is on. DN: No sub-grid dynamical friction. R14-23: Various steps in resolution, with R23 being the highest, and R14 being the lowest. The horizontal dotted line shows a single resolution scale from the center of mass of the halo. The vertical dash-dot line shows the estimated infall time for the supermassive black hole. The sub-grid model dramatically improves the result, allowing the supermassive black hole to fall in on the expected timescale.

model, labelled DN.

At the lowest resolution R14 (black curve), the dynamical friction model does not provide sufficient acceleration to sink the SMBH to the center of the dark matter halo. The resolution of the dark matter in the R14 simulations is $M_{\rm DM,part} = 10^7$ M$_\odot$, a factor of 10 larger than the SMBH mass. The disparity between the masses causes the SMBH to jiggle about from the tug of the massive particles. The other resolutions R17, R20, and R23 increase in mass resolution by factors of 10 and do not experience this effect. Importantly, R17 is the simulation where $M_{\rm BH} = M_{\rm DM,part}$ and, therefore, I conclude that the SMBH must be at least as massive as the dark matter particles in order for the dynamical friction model to accurately capture the infall time.

The difference between the DN, R23 and DY, R23 shows the true necessity of the dynamical friction model. Even at the highest resolution, where the dark matter particles do not push and pull the SMBH about, there are not adequately resolved dynamical friction forces. The the dynamical friction model successfully captures the forces that occur on a scale $\lesssim h$, where $h = 1$ kpc is the resolution length of the SMBH particle.

3.6 Calibration

The main goal of this Chapter is to examine the viability of combining all of the most physically motivated models in the literature. Calibration itself is a field of study (Calder et al., 2004) and we can learn important results from failures and successes of having a calibrated model. The computational expense of running a grid of parameters is enormous and, therefore, I was only able to explore limited regions of parameter-space up to $z = 2$ to find viable parameters. The parameters that I selected to calibrate were the supermassive black hole (SMBH) seed mass, $M_{\rm seed}$, the seeding threshold, $\mathcal{T}$, the quasar wind velocity, $v_{\rm wind}$, and the jet velocity $v_{\rm jet}$. I chose to use 3 values for each parameter, giving $3^4 = 81$ simulations. All of the simulations have the updated momentum constrained kinetic outflows for the quasar and slim disk modes.

For $M_{\rm seed}$, I chose values $\{10^4, 10^5, 4 \times 10^5\}$ M$_\odot$. These values span the range of expected masses from the direct collapse of a gas cloud into a supermassive black hole (SMBH) (Volonteri, 2010). The minimum value comes from Davé et al. (2019) whereas the maximum value is the resolution of the gas particles in the simulation, out of which the SMBH form.

The seeding threshold $\mathcal{T}$ determines when a star-forming gas particle in the simulation should form a SMBH rather than a star particle. Recall that there are three conditions that I am using from Tremmel et al. (2017) to determine when the seed should form: density, temperature, and metallicity. I cannot use the temperature condition since the gas above the star formation density threshold, $\rho > \rho_*$, is artificially placed on an equation of state, $T \propto \rho^{4/3}$. Therefore, the temperature is not trustworthy. For the density, the threshold $\mathcal{T}$ is the multiplicative factor that gives minimum gas density at which the simulation can form a seed, $\rho_{\rm min} = \mathcal{T}\rho_*$. While Tremmel et al. (2017) choose $\mathcal{T} = 15$ for their simulations, I choose to calibrate over the values $\mathcal{T} \in \{10, 50, 100\}$. For the metallicity threshold, I fix that to the value in Tremmel et al. (2017) at $Z_{\rm max} = 3 \times 10^{-4}$.

The observed physical wind velocities from AGN range from $\sim 100\,{\rm km\,s^{-1}}$ to $5{\times}10^4$ ${\rm km\,s^{-1}}$ depending on the observed scale (Fiore et al. 2017; see Section 3.5.2). I choose to calibrate over the values $\{500, 1000, 1500\}$ ${\rm km\,s^{-1}}$ since these are appropriate for the resolution of my cosmological-scale simulations.

My new calibration simulations are at a higher resolution compared to those in Section 3.5 in order to have physical seeding masses combined with the dynamical friction model. Recall that the dynamical friction model requires the SMBH particle mass to be at least as massive as the dark matter. However, why does it matter if I go to higher resolution since there is a sub-grid mass of the SMBH? I want the sub-grid mass (physical) and the particle mass (numerical) to be as close as possible, with the latter being as small as possible in order to more accurately track the dynamics. If I stayed at $M_{\rm DM} = 10^8$ ${\rm M_\odot}$ resolution, the SMBHs would fall in at the maximum rate given by a 10^8 ${\rm M_\odot}$ object, regardless of its physical sub-grid mass. Increasing the mass resolution to $M_{\rm DM} = 2 \times 10^6$ ${\rm M_\odot}$ provides a much more accurate in-fall time since *most* SMBHs grow very rapidly from their seed mass to $\sim 10^6$ ${\rm M_\odot}$ regardless. Therefore, my new calibration simulations have the same parameters as in Table 3.2 except that I reduced the volume to $\approx (9\,{\rm Mpc})^3$ which changed $M_{\rm gas,particle} = 4 \times 10^5$ ${\rm M_\odot}$ and $M_{\rm DM,particle} = 2 \times 10^6$ ${\rm M_\odot}$.

The first conclusion from my set of calibration simulations is that another major change must be made. Fig. 3.8 shows the accretion rate of the most massive SMBH at $z = 2$, for all timesteps, from one of the calibration simulations. The horizontal dotted line shows the lower boundary for when the quasar mode activates. The purple points show the Bondi accretion contribution and the coral points show the torque accretion contribution. The torque model should contribute the most to the overall

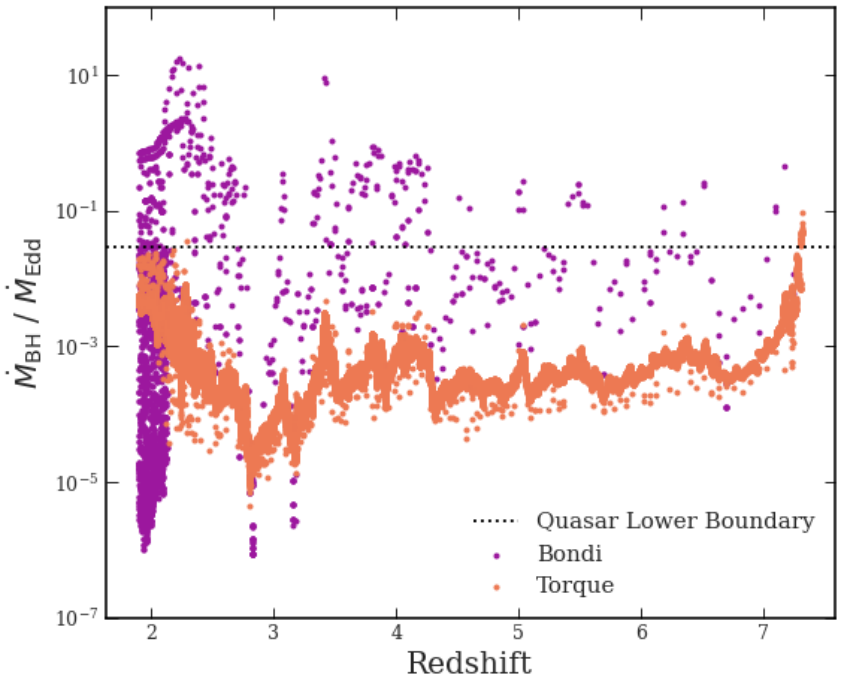

Figure 3.8: The accretion rate history of the most massive supermassive black hole at $z = 2$, broken into contributions from the Bondi accretion model and the torque accretion model. The dotted horizontal line shows the lower accretion rate boundary for entering the quasar mode. At high redshift, the torque mode should dominate, but does not in this example. The black hole seeds form off center from the galactic core and take time to fall into the center due to dynamical friction – leading to a reduced accretion estimate.

accretion rate in the early Universe when there is a lot of cold, dense gas. Later, as the gas heats up from stellar feedback, AGN, and gravitationally processes, the Bondi accretion model should dominate. I found in my calibration simulations that the SMBHs were: (a) only growing via Bondi accretion and mergers and (b) not feeding back into the environment efficiently enough to quench star formation in their host galaxies. This was a common theme I saw across all calibrations. It is important to note that this is despite the fact that I do not impose a limit on how fast the SMBHs may accrete, as is common in the literature[11].

One reason that the torque accretion mode may not contributing to the overall accretion rate is because of the change in the seeding and dynamics model. With the physical seeding model, SMBHs form away from the center of the galaxies and take time to fall into the center from dynamical friction. If they are not centrally located and formed out of the rotating gas, then they are moving with the bulk flow at their birth radius. If they are moving with the bulk flow, then the computation for the torque accretion predicts low rates of accretion since there is no rotation around the SMBH itself. Additionally, unconstrained accretion rates may lead to cold gas destruction in the regions neighbourhood of the SMBHs. Realistic growth rates in the early Universe do not seem possible with this combination of models. For that reason, I recommend to only use the Bondi accretion model moving forward with the Eddington limiter in place until an adequate solution can be found.

After changing the accretion mode, I found sufficient growth and feedback in the SMBHs. I re-ran the 81 calibrations and found that there were some regions of parameter space that gave reasonable qualitative fits to the $M_{\rm BH} - M_*$ relationship.

First, only a few calibration combinations work with $M_{\rm seed} = 10^4$ M$_\odot$, those with a seeding threshold $\mathcal{T} = 10$ and wind speed $v_{\rm wind} < 1500$ km s^{-1}. The combination of low seed mass at higher wind speeds and seeding thresholds causes the SMBHs to not grow and severely under predict the normalization of the $M_{\rm BH} - M_*$ relationship. With a seeding mass $M_{\rm seed} = 10^5$ M$_\odot$ or 4×10^5 M$_\odot$, the seeding threshold $\mathcal{T} = 10$ does not provide any good match to the $M_{\rm BH} - M_*$ relationship as too many SMBHs form and cause too many mergers in the early Universe. Additionally, the threshold $\mathcal{T} = 100$ does not provide a good match at all. In fact, I find that there are insufficient SMBHs to even begin to trace out the $M_{\rm BH} - M_*$ relationship in that region of parameter space.

[11]Recall that most simulators impose the Eddington limit, or some value near it, on accreting SMBHs to prevent overgrowth. I do not impose an upper limit and find the SMBHs self-regulate naturally.

However, the values $\mathcal{T} = 50$, $v_{\rm wind} = 500$ km s^{-1}, and any $v_{\rm jet} < 10^4$ km s^{-1} provide a good fit. There are somewhat reasonable solutions for the other thresholds and wind velocities, but they are far too scattered and would not eventually converge onto the $M_{\rm BH} - M_*$ relationship.

The galaxy stellar mass function in the small volume is more difficult to calibrate. Unfortunately, in such a small volume there could be bias toward a high or low galaxy density, and insufficient massive galaxies to calibrate the high mass end. However, there are good fits to the galaxy stellar mass function in some of parameter space.

The above discussion is qualitative and I need a more quantitative measure of what is a *good fit*. My method of determining the best fit involves finding the root mean squared error,

$$\mathrm{RMSE}_{\mathrm{i,X}} = \sqrt{\frac{\sum_j^{N_{\rm gal}} \left(X_{\mathrm{j,pred}} - X_{\mathrm{j,sim}}\right)}{N_{\rm gal}}} \tag{3.34}$$

where i is the calibration index, X is the label for the $M_{\rm BH} - M_*$ or the galaxy stellar mass function measurements[12], j indexes the galaxies in each calibration, and $N_{\rm gal}$ is the number of galaxies in each calibration simulation. After I have all of the $\mathrm{RMSE}_{\mathrm{i,BH}}$ and $\mathrm{RMSE}_{\mathrm{i,gsmf}}$, I find the best calibration C by finding the minimum Euclidean norm of the errors across all calibrations,

$$C = \mathrm{argmin}\left(\sqrt{\mathrm{RMSE}_{\mathrm{0,BH}} + \mathrm{RMSE}_{\mathrm{0,gsmf}}}, \ldots, \sqrt{\mathrm{RMSE}_{\mathrm{i,BH}} + \mathrm{RMSE}_{\mathrm{i,gsmf}}}\right) \tag{3.35}$$

The simulation that had the lowest error compared to the observation was $C = 1$. In this case, $\mathrm{RMSE}_{\mathrm{1,gsmf}} = 0.370$ and $\mathrm{RMSE}_{\mathrm{1,BH}} = 0.327$. The $C = 1$ simulation has $\mathcal{T} = 10$, $M_{\rm seed} = 10^4$ M$_\odot$, $v_{\rm wind} = 500$ km s^{-1}, and $v_{\rm jet} = 7500$ km s^{-1}.

As an example, Fig. 3.9 shows the galaxy stellar mass function (top) and $M_{\rm BH} - M_*$ (bottom). The galaxy stellar mass function is noisy due to the small volume, and seems to have an overabundance of very low-mass galaxies. That may be bias again because of the low volume. The $M_{\rm BH} - M_*$ relationship fits very well, with a single outlier at high mass. However, the scatter of the $M_{\rm BH} - M_*$ relationship does allow the possibly of that SMBH existing and my simulations are biased for only having 1 object to sample that region of the curve. Additionally, Fig. 3.10 shows a projection

[12] I use the log-values of the simulation output and observation, so that the RMSE is the log-error.

of the gas density centered on the most massive galaxy in the simulation at $z = 2$. The white x mark the positions of the top five most massive SMBHs in the region. The SMBH in the central region is the most massive in the simulation at $M_{\rm BH} = 3 \times 10^8$ $M_\odot$ and is surrounded by several $M_{\rm BH} \sim 10^5$ $M_\odot$ smaller SMBHs.

The parameters I provide are starting point for using this model in galaxy evolution studies, and I showed that it can produce reasonable results at high redshift where the brightest cluster galaxies begin to form. Interestingly, this parameter choice does not activate the jet as much as other calibrations and relies on the quasar and slim disk mode. The jet may turn on at lower redshift, but that is beyond the scope of this initial viability project.

3.7 Conclusions

Supermassive black holes (SMBHs) power highly energetic active galactic nuclei (AGN) found in the cores of galaxies throughout the cosmos. These AGN have luminosities that outshine their host systems, leading to the idea that they may be able to impact the evolution of their host galaxy (Fabian, 2012). Indeed, the evidence is convincing that AGN impact their host galaxies through the observed $M_{\rm BH} - M_*$ relationship and galaxy stellar mass function. The $M_{\rm BH} - M_*$ relationship shows that SMBH masses are highly correlated with the stellar masses of their host stellar masses, suggesting co-evolution (Kormendy & Ho, 2013). Simultaneously, the galaxy stellar mass function shows a significant drop in the density of galaxies with stellar masses $M_* \gtrsim 10^{11}$ $M_\odot$ (Tomczak et al., 2014). AGN feedback could be responsible for suppressing star formation in these massive systems. The non-linear interaction between AGN feedback and their host systems demands the use of numerical simulations in order to measure their impact on galaxy evolution.

The cosmological simulation community has provided several state-of-the-art models for the four aspects of modelling SMBHs: seeding, accretion and growth, feedback, and dynamics. My goal with this Chapter was to test the viability of combining all of the most physically motivated models for each aspect into a single simulation for use at high $z > 2$ redshift. Using the `Simba` models for cooling, star formation, and stellar feedback I combined[13] the physical seeding model from Tremmel et al. (2017), the accretion model from Davé et al. (2019), a new feedback model I described in

[13] I again note that the accretion model was already implemented in `GIZMO`, I simply ran it with the other models that I implemented.

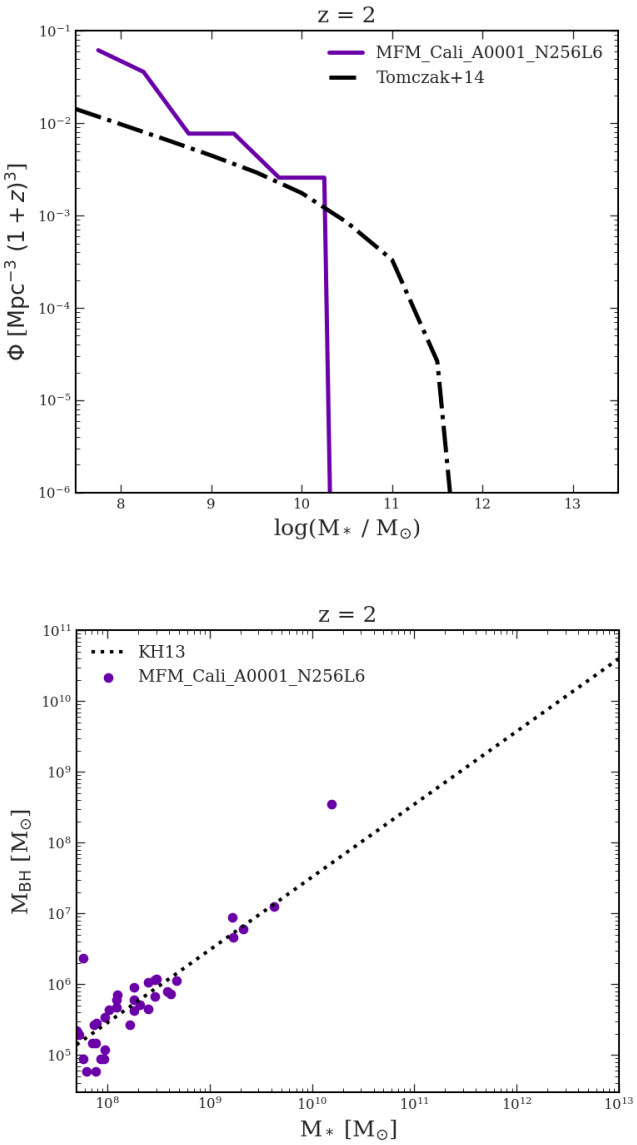

Figure 3.9: (*top*) The galaxy stellar mass function. The purple curve shows the results from the best calibration. The dotted-dashed line shows a fit to the observed results in Tomczak et al. 2014 at $z = 2$. (*bottom*) The $M_{\rm BH} - M_*$ relationship. The purple points show the supermassive black hole masses $M_{\rm BH}$ in individual galaxies of stellar mass M_* from the best calibration. The dotted line shows the fit to the observational results in Kormendy & Ho 2013 at $z = 0$.

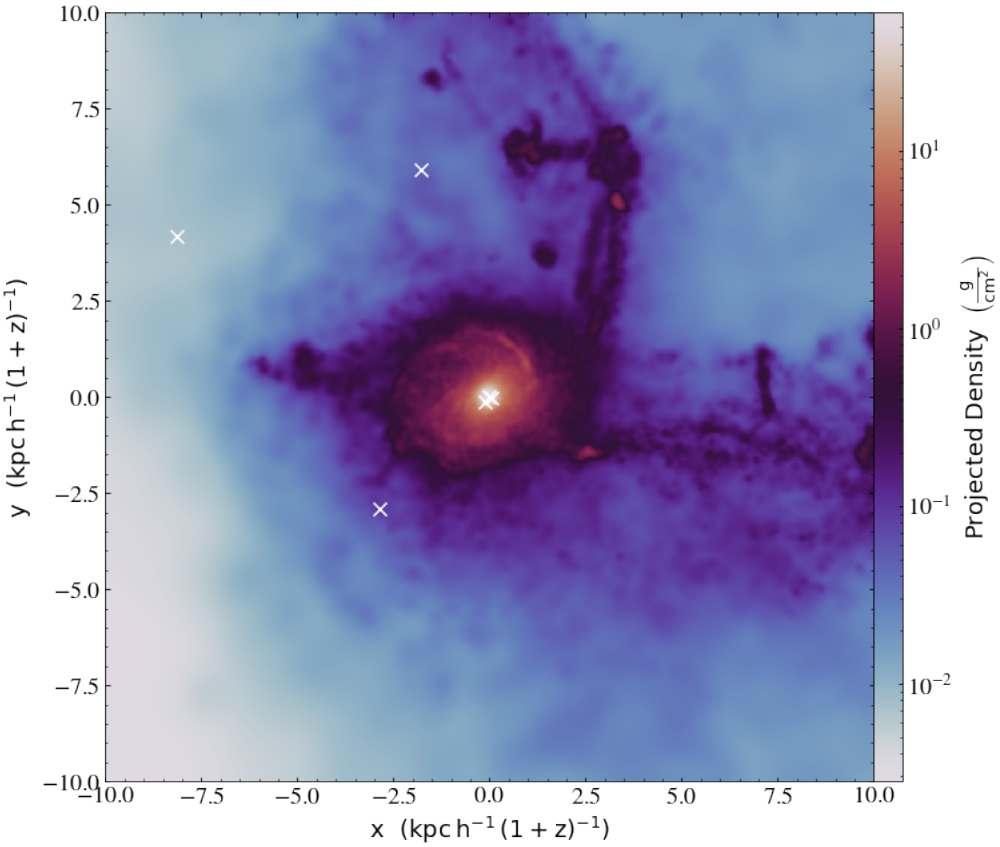

Figure 3.10: A gas density projection centered on the most massive galaxy at $z = 2$ in the best calibration. The white x mark the top five most massive supermassive black holes in the region. By this time, the most massive black hole is 3×10^8 $M_\odot$ and is very close to the center of the gas distribution. The neighbouring black holes have masses $\sim 10^5$ $M_\odot$.

Section 3.3, and the model for dynamical friction from Tremmel et al. (2015) into the GIZMO code (Hopkins, 2015).

The physical seeding model I implemented uses the gas density, ρ, and metallicity, Z, to determine if a SMBH should form out of a star-forming gas particle that is about to turn into a star particle. If the criteria $\rho > \mathcal{T}\rho_*$ (ρ_* is the star formation threshold density) and $Z < 3 \times 10^{-4}$ are met within that gas particle, it turns into a SMBH of seed mass $M_{\rm seed}$. I treated both $\mathcal{T}$ and $M_{\rm seed}$ as free parameters. The seed mass is different from the particle mass, which must start at the resolution of the dark matter particles. To conserve mass, I implemented a mechanism to instantaneously consume gas from the neighbourhood of the SMBH in its first time step – ignoring feedback. After the first step, the SMBH will accrete and feedback as I described in Section 3.4. For accretion and growth, I use the model of Davé et al. (2019) that combines both Bondi accretion for hot gas ($T > 10^5$ K) and a gravitational torque estimator for cold gas ($T < 10^5$ K). I treated the accretion model as well-calibrated and do not allow free parameters.

The new feedback model I introduced is a combination of the quasar mode from Anglés-Alcázar et al. (2017b) and Davé et al. (2019), along with two new modes: the advection dominated accretion flow (ADAF) and slim disk modes. One major change is that I limited the quasar mode to only accretion rates in the range $0.03 < \dot{M}_{\rm BH}/\dot{M}_{\rm Edd} \leq 0.3$. The ADAF mode activates in the range $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd} \leq 0.03$ with an isotropic kinetic wind and a jet acting simultaneously – a novel addition. The slim disk mode activates in the high accretion rate regime, $\dot{M}_{\rm BH}/\dot{M}_{\rm Edd} > 0.3$. The radiative efficiency drops rapidly in the high accretion rate regime (Sądowski et al., 2014; Madau et al., 2014; Lupi et al., 2016) causing most of the inflowing matter to be accreted into the SMBH. The rapid growth quickly causes the SMBHs to enter the quasar mode and self-regulate and, therefore, the model does not require an Eddington rate limiter. I chose to fix the kinetic wind coupling factors $\epsilon_{\rm f,quasar}$ and $\epsilon_{\rm f,slim}$ to the standard value 0.05, and leaving the wind velocity $v_{\rm wind}$ as a free parameter. The second free parameter for this model is the jet velocity, $v_{\rm jet}$.

For the dynamical evolution of the SMBHs, I implemented the dynamical friction model described in Tremmel et al. (2015) based on the formula of Chandrasekhar (1943). In this model, the dynamical friction forces that occur below the resolution scale are taken into account. Importantly, my tests showed that the particle mass of the SMBH must be at least as massive as the dark matter particles in the simulation to avoid gravitational noise causing spurious repositioning.

The above-described models are the ideal models with which I began my experiments. However, through the process of verifying the models several important results arose. I list below the two most important results from my viability study:

1. A kinetic energy constrained kinetic wind model is not possible at the resolution of contemporary cosmological simulations. Specifically, it is not possible in large-volume simulations given current computational restraints. The resolution of gas particles in the simulation would have to approach $M_{\rm gas,particle} = 5 \times 10^4$ $\mathrm{M}_\odot$ in order for the model to give reasonable results.

2. The gravitational torque accretion estimator does not appear to work well at high redshift with the physical seeding and dynamical friction model. The problem may arise because the SMBHs are generally not at the center of the rotational plane of the system, suppressing accretion. It may also be due to cold gas suppression due to unconstrained growth and feedback in the SMBH neighbourhood. There were hints that this might improve just before $z \sim 2$, however that is outside of the scope of this work. The torque mode should dominate in the early Universe when the gas densities are high and there is abundant cold gas.

My solution to problem (1) was to constrain the momentum flux of the SMBH kinetic wind to $20\,L_{\rm BH}/c$ (Faucher-Giguère & Quataert, 2012; Davé et al., 2019). That provides a resolution-independent model that works even at high resolution. For problem (2), I decided to eliminate the gravitational torque estimator and revert to using only Bondi accretion for all of the gas, regardless of temperature.

With these two important changes, I ran 81 calibration tests with $N = 256^3$ gas and dark matter particles in volumes approximately $(9\,\mathrm{Mpc})^3$, giving $M_{\rm gas,particle} = 4 \times 10^5$ $\mathrm{M}_\odot$ and $M_{\rm DM,particle} = 2 \times 10^6$ $\mathrm{M}_\odot$. I calibrated the parameters $\mathcal{T}$, $M_{\rm seed}$, $v_{\rm wind}$, and $v_{\rm jet}$. I found the best combination of parameters by using the root mean squared error between the simulation predictions and the (a) $M_{\rm BH} - M_*$ relationship from Kormendy & Ho (2013) and (b) the galaxy stellar mass function from Tomczak et al. (2014). The best values were: $(\mathcal{T}, M_{\rm seed}, v_{\rm wind}, v_{\rm jet}) = (10, 10^4\,\mathrm{M}_\odot, 500\,\mathrm{km\,s^{-1}}, 7500\,\mathrm{km\,s^{-1}})$.

While my original model did not function as planned, the new model is novel. To my knowledge, the combination of physical seeding, dynamical friction, an isotropic wind in an ADAF mode, and a kinetic jet have not been explored in cosmological simulations. My parameters provide a starting point and show the viability of running

simulations of the early Universe. In the long term, studying the evolution of the brightest cluster galaxies will be possible given that the model is successful at the current resolutions used in high-resolution zoom simulations of those systems.

Chapter 4

Rapid early coeval star formation and assembly of the most massive galaxies in the universe

Originally published in Monthy Notices of the Royal Astronomical Society: 2020, vol. 493, issue 4, pp 4607-4621. Reproduced by permission of Oxford University Press. https://doi.org/10.1093/mnras/staa541

Authours: Douglas Rennehan[1], Arif Babul[1], Christopher C. Hayward[2], Connor Bottrell[1], Maan H. Hani[1] and Scott C. Chapman[3]

Personal contribution clarification: I designed the experiment, initial conditions for the simulations, and ran all of the simulations for this work. I wrote the entire document except for Section 4.7, which was written by Connor Bottrell. Arif Babul and Christopher C. Hayward provided important scientific guidance and recommendations for experimental design and interpretation of the results. Arif Babul provided extensive comments on my original draft(s) of the document, leading to significant changes. Christopher C. Hayward, Connor Bottrell, Maan H. Hani, and Scott C. Chapman all provided comments on the final draft. Connor Bottrell and Maan H. Hani post-processed the simulations I ran in this work in order to produce the data for Fig. 4.3. I created all of the figures in this work.

[1]Department of Physics & Astronomy, University of Victoria, BC, V8X 4M6, Canada
[2]Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY, 10010, USA
[3]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, B3H 4R2, Canada;
NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada;
Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada

Abstract: The current consensus on the formation and evolution of the brightest cluster galaxies is that their stellar mass forms early ($z \gtrsim 4$) in separate galaxies that then eventually assemble the main structure at late times ($z \lesssim 1$). However, advances in observational techniques have led to the discovery of protoclusters out to $z \sim 7$. If these protoclusters assemble rapidly in the early universe, they should form the brightest cluster galaxies much earlier than suspected by the late-assembly picture. Using a combination of observationally constrained hydrodynamical and dark-matter-only simulations, we show that the stellar assembly time of a sub-set of brightest cluster galaxies occurs at high redshifts ($z > 3$) rather than at low redshifts ($z < 1$), as is commonly thought. We find, using isolated non-cosmological hydrodynamical simulations, that highly overdense protoclusters assemble their stellar mass into brightest cluster galaxies within ~ 1 Gyr of evolution – producing massive blue elliptical galaxies at high redshifts ($z \gtrsim 1.5$). We argue that there is a downsizing effect on the cluster scale wherein some of the brightest cluster galaxies in the cores of the most-massive clusters assemble earlier than those in lower-mass clusters. In those clusters with $z = 0$ virial mass $\geqslant 5 \times 10^{14}$ $M_\odot$, we find that 9.8% have their cores assembly early, and a higher fraction of 16.4% in those clusters above 10^{15} $M_\odot$. The James Webb Space Telescope will be able to detect and confirm our prediction in the near future, and we discuss the implications to constraining the value of σ_8.

4.1 Introduction

Galaxy clusters are at the peak of the mass-assembly hierarchy in the ΛCDM paradigm and share a common feature of hosting a distinct population of galaxies aptly referred to as the brightest cluster galaxies (BCGs) (Peebles, 1968; Sandage, 1976; Tremaine & Richstone, 1977). The BCGs are ultra-luminous (with $\sim 10 L_{K,*}$ where $L_{K,*}$ is the characteristic luminosity of the general galaxy population), morphologically spheroidal, have a large spatial extent, and exhibit core stellar velocity dispersions of order $300 - 400$ km s^{-1} (Lin & Mohr, 2004; Pipino et al., 2011; Loubser et al., 2018). The BCGs are in fact the brightest and the most-massive galaxies in the present-day Universe. Many of the observed characteristics of the BCGs seem to scale with the properties of the hosting cluster halo (Brough et al., 2008; Lidman et al., 2012; Lavoie et al., 2016; Kravtsov et al., 2018) and they are frequently found close to, and typically have relatively small velocity offsets with respect to, the potential centres of the

cluster halo (Lidman et al., 2013; Lauer et al., 2014). This is commonly interpreted as an indication that the formation and evolution of the BCGs and their host clusters are intimately linked, and that studying the former will provide clues about the formation and evolution of the latter. However, pinning down the assembly and growth histories of these gigantic galaxies is proving to be a challenge.

There are several pathways that could explain the origin of the BCGs: (i) extended in-situ star formation (e.g. via cooling flows); (ii) rapid star formation and early assembly; and (iii) early star formation in separate galaxies but relatively recent assembly via a sequence of late-time mergers. The key idea underlying the first proposal is that radiative cooling drives the hot intracluster medium to concentrate at the clusters' potential centre where it then forms stars at relatively high rates (Cowie & Binney, 1977; Fabian & Nulsen, 1977). However, observations show that not only are the bulk of the BCG stars old (Whiley et al., 2008), but heating from the central active galactic nucleus (AGN) also strongly suppresses the cooling of the intracluster medium (Tabor & Binney, 1993; Ciotti & Ostriker, 1997; Silk & Rees, 1998). The second proposal suggests that massive elliptical galaxies, including the BCGs, form via essentially monolithic collapse of a mass density peak (Eggen et al., 1962), with the galaxies' stellar mass building up rapidly in the process. One difficulty with this model is that BCGs show evidence of significant growth in their sizes over cosmic time (Daddi et al., 2005; van der Wel et al., 2008; Shankar et al., 2015). Once a leading theory, this scenario has fallen out of favour due to the emergence of the hierarchical assembly paradigm for cosmic structure formation. According to this paradigm, the third proposal, galaxies form via a series of mergers of lower-mass systems – implying that the low mass systems form first, and over time build-up the more-massive systems (Aragon-Salamanca et al., 1998; Dubinski, 1998).

Numerical studies investigating the formation and evolution of BCGs in the presently favoured hierarchical **ΛCDM** model find that the majority of stars that end up in the present-day BCGs typically form at $z \gtrsim 4$ in distinct progenitor galaxies (De Lucia et al., 2006). These galaxies then eventually merge to assemble the BCGs we observe today. As for the timing of this assembly, until recently, the theoretical consensus was that present-day BCGs are assembled through dissipationless mergers, with between $50-60\%$ of their present-day stellar mass being incorporated at relatively late times, from $z \sim 1.5$ to the present (Dubinski, 1998; Conroy et al., 2007; De Lucia & Blaizot, 2007; Ruszkowski & Springel, 2009; Laporte et al., 2013; Contini et al., 2014; Laporte & White, 2015). Recent cosmological hydrodynamical simulations suggest a slightly

modified scenario (Martizzi et al., 2016; Ragone-Figueroa et al., 2018) and highlight the importance of distinguishing between the galaxy proper and its extended stellar envelope. Focusing on the galaxy proper (i.e. the mass within 50 kpc of the centre), Ragone-Figueroa et al. (2018) find that while half of the stars that end up in the BCG have formed by $z \sim 3.7$, the assembly of the BCG occurs over an extended time-span and half of the BCGs' stellar mass only falls into place typically by $z \sim 1.5$. In detail, the stellar mass of the BCGs grows on average by a factor of 2.5 between $z = 2$ and $z = 0$, with a median stellar mass growth factor over all BCGs of 3.5. Interestingly, Ragone-Figueroa et al. (2018) also find that *in-situ* star formation is not entirely negligible; it accounts for nearly 25% of the growth in stellar mass between $z = 2$ and $z = 1$. These minor revisions aside, the central paradigm – of early formation of the stellar mass and late-time assembly of the BCG – remains intact.

Recent observations, however, suggest that the late-assembly picture may not be as concrete as once thought, and that early assembly may play a major role in the evolution of the BCGs. Several BCGs have been discovered at $z \sim 1 - 1.5$ that have stellar masses comparable to the most-massive galaxies in the universe (Collins et al., 2009). If these BCGs were to grow at the rates theoretically expected via late-time hierarchical assembly, they would greatly exceed the predicted masses of those theoretical models. Additionally, the debate on the size evolution of BCGs is far from settled, as there is also evidence that the luminosities and sizes of the BCGs, as a population, do not evolve much past $z \sim 1$ (Whiley et al., 2008; Stott et al., 2011), suggesting that little growth through the hierarchical scenario is possible. The absence of observed evolution past $z \sim 1$ implies that these massive galaxies must grow via a combination of *in-situ* star formation and early assembly. With respect to the former, Webb et al. (2015) analyse a set of BCGs in the *Spitzer* Adaptation of the Red-Sequence Cluster Survey (SpARCS; see Muzzin et al. 2009; Wilson et al. 2009) and find that a large contribution to the overall growth of the BCGs must be due to *in-situ* star formation based on the estimated star formation rates of hundreds of BCGs (in the range $0.8 < z < 1.8$), and also find an increasing star formation rate with increasing redshift. As to the latter, there is growing evidence of highly over-dense protocluster cores (e.g. Ishigaki et al. 2016; Miller et al. 2018; Jiang et al. 2018; Higuchi et al. 2019; Wang et al., in prep.; also see Ito et al. 2019, and Overzier 2016 for a broad census review) at high redshifts ($z \gtrsim 4$). As we demonstrate in Section 4.2, protocluster cores with a high density of galaxies are the birthplaces of BCGs, and highly-over-dense systems should collapse rapidly in the **ΛCDM** theory –

casting into doubt whether the theoretical consensus of late assembly is valid for the entire population of BCGs.

In this paper, we investigate the above tension between the current theoretical picture and accumulation of observational results in order to gain insight into the evolution of the BCG population. Specifically, we use a bespoke non-cosmological simulation based on the observed parameters of the SPT2349-56 protocluster (Miller et al., 2018) to track its forward evolution. Our interest lies in determining the future evolution of the protocluster core – including the fate of the observed galaxies – and the timescale of its evolution. We then use the MultiDark Planck 2 Bolshoi simulation (Riebe et al., 2013; Klypin et al., 2016) a large-volume non-baryonic simulation, to estimate the frequency of similar events in the universe. While dark-matter simulations exist that provide ample resolution and population statistics (through their large volumes) for discovering over-dense protoclusters at high redshift, simulating the equivalent volumes in tandem with the hydrodynamical equations of motion and galactic-baryonic physical processes is, at present, not feasible due to computational constraints. These constraints force us to study the forward evolution of SPT2349-56 in the bespoke simulation. In Section 4.2 we describe our setup and initial conditions for the SPT2349-56 simulation. In Section 4.3, we discuss the assembly and growth of the system. In Section 4.4, we analyse a large volume dark-matter-only simulation in order to determine how frequent such highly over-dense events may be. Lastly, we synthesise our findings and present a revised paradigm for the formation and the evolution of the BCGs in Section 4.5.

4.2 Methodology

We start by constructing a bespoke simulation of the SPT2349-56 system in order to specifically study its forward evolution. Observations indicate that the 14 galaxies that comprise the core of the protocluster are within a 130 kpc (physical) projected region on the sky at a mean redshift of $z \approx 4.3$, and we show the observed physical properties of each of the 14 galaxies in leftmost two columns of Table 4.1. The observed line-of-sight velocity distribution, $\Delta V_{\rm LOS}$, was found to approximate a Gaussian distribution with $\sigma_{\rm LOS} = 408\,{\rm km\,s^{-1}}$ (Miller et al., 2018). The cold gas masses, $M_{\rm gas,gal}$, were estimated by converting the estimated[4] CO(1-0) line luminosity to a

[4]Miller et al. (2018) measured the CO(4-3) line luminosity and then converted to a CO(1-0) luminosity by using the ratio of the line brightness temperatures, calculated from a sample of sub-millimetre galaxies with measurements of both lines.

Table 4.1: Observed and estimated physical properties of SPT2349-56 (Miller et al., 2018). We use only the observed line-of-sight offset velocities ($\Delta V_{\rm LOS}$) and cold gas masses ($M_{\rm gas,gal}$) to estimate the remaining physical parameters in this table. $M_{*,\rm gal}$ is the estimated galactic stellar mass, $C_{\rm vir}$ is the estimated NFW halo concentration, $M_{\rm vir}$ is the estimated virial mass of each halo, and $V_{\rm vir}$ is the estimated virial velocity of the galaxy's host halo.

Label	$\Delta V_{\rm LOS}$ (km s^{-1})	$M_{\rm gas,gal}$ (10^{10} M$_\odot$)	$M_{*,\rm gal}$ (10^{10} M$_\odot$)	$M_{\rm vir}$ (10^{10} M$_\odot$)	$C_{\rm vir}$	$V_{\rm vir}$ (km s^{-1})
A	-90	12.0	5.14	514	1.67	537
B	-124	11.2	4.79	479	1.69	524
C	603	6.7	2.87	287	1.81	442
D	-33	8.4	3.6	360	1.75	477
E	84	4.8	2.05	205	1.89	395
F	395	3.4	1.46	146	1.97	353
G	308	1.6	0.685	68.5	2.18	274
H	-719	4.4	1.88	188	1.91	384
I	310	2.2	0.942	94.2	2.09	305
J	481	2.2	0.942	94.2	2.09	305
K	631	3.1	1.33	133	2.00	342
L	-379	3.3	1.41	141	1.98	348
M	34	1.2	0.514	51.4	2.26	249
N	90	1.0	0.428	42.8	2.31	234

gas mass using a conservative estimate of the $\alpha_{\rm CO}$ conversion factor, $\alpha_{\rm CO} = 0.8$ M$_\odot$ / (K km s^{-1} pc^2).

In order to simulate the forward evolution of the system, we use a modified version of GIZMO[5] (Hopkins, 2015), a publicly available gravity plus hydrodynamics simulation program, that is equipped with an implementation of the mesh-free finite mass method (Lanson & Vila, 2008a,b; Gaburov & Nitadori, 2011).

We simulate the protocluster in isolation with vacuum boundary conditions. Specifically, we simulate the initial coalescence phase of the separate systems (i.e. galaxies with their own dark matter halo, gas, and stellar components) that make up the SPT2349-56 protocluster core. At this early phase, we do not expect the cluster-scale dark matter envelope to be in place. The simulation is non-cosmological, and evolves the equations of motion for 1 Gyr. Gas and star particles in the simulation have an initial mass of $M_{\rm gas} = M_* = 10^6$ M$_\odot$. Gas properties are calculated using the cubic spline kernel with 32 neighbouring particles in our simulations. The dark

[5] http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html

matter particle mass is $M_{\rm dark} = 5 \times 10^6$ M$_\odot$. Additionally, we seed each galaxy with a black hole of mass $M_{\rm BH} = 10^5$ M$_\odot$. We use adaptive gravitational softening for all gravitationally interacting particles (Hopkins et al., 2018), which requires minimum softening parameters. Baryonic particles have a minimum softening of $\epsilon_{\rm b,min} = 50$ pc and, for dark matter, we use a minimum softening of $\epsilon_{\rm dark,min} = 200$ pc.

4.2.1 Initial conditions

We assume a Planck Collaboration XVI (2014) cosmology throughout the following procedure, and generate initial conditions using the `MakeGalaxy` software (Hernquist, 1993; Springel & White, 1999; Springel, 2000; Springel et al., 2005a). Each galaxy in our synthetic SPT2349-56 system consists of a dark matter halo, gas disc and stellar disc, with no stellar bulge or surrounding gaseous circumgalactic medium. Given the high overdensity of the SPT2349-56 system, we expect that the large-scale background dark matter field of the surrounding region contributes much less to the mass budget within 130 kpc compared to the individual systems. Therefore, we do not include an extended dark-matter field as the dynamics within the 130 kpc region should not be affected. `MakeGalaxy` employs the methods of Springel et al. (2005a) to create stable spiral galaxies, which we summarise here. We model the dark matter distribution in each galaxy with a Hernquist (1990) profile where the scalelength is related to the corresponding NFW concentration $C_{\rm vir}$ of the halo (Navarro et al., 1997). We model the gas and stellar discs with exponentially declining surface densities with the scalelength H related to the angular momentum (through the spin parameter λ) of the system. We follow Robertson et al. (2006) and use $\lambda = 0.033$, which is the mode of the spin distribution from cosmological simulations (Vitvitska et al., 2002). The combined gas+stellar disc is also under the condition that the total disc mass is a fixed fraction of the total mass of the system, i.e. $M_{\rm disc} = m_{\rm d} M_{\rm vir}$. To ensure disc stability we choose $m_{\rm d} = 0.03$ for each disc as values between $0.03 \lesssim m_{\rm d} \lesssim 0.05$ lead to stable discs in the ΛCDM cosmology (Mo et al., 1998). The vertical structure of the stellar disc is that of an isothermal sheet with a radially constant scaleheight z_0 given as a free parameter proportional to the scalelength of the disc, which we assume to be $z_0 = 0.1H$. For the vertical structure in the gas disc, we set the scaleheight such that hydrostatic equilibrium is enforced. We tested each galaxy in isolation to ensure that the system is physically and numerically stable, and find an average star formation rate of ≈ 75 M$_\odot$ yr^{-1} over 1 Gyr, with a peak of ≈ 140 M$_\odot$ yr^{-1}, for the most-massive

galaxy.

In order to setup each galaxy, we require estimates of their dark matter halo, stellar, and gas masses as well as the concentrations of the dark matter halos. We base our following calculations on the observed total galactic cold gas mass, $M_{\rm gas,gal}$, shown in Table 4.1. We prepare and simulate three separate realisations of the SPT2349-56 system.

Masses

We estimate the dark matter halo mass of each individual galaxy in the protocluster core by first assuming a reasonable gas fraction $f_{\rm gas,gal}$, computing the corresponding galactic stellar mass $M_{*,\rm gal}$, and then estimating the halo virial mass $M_{\rm vir}$ from $M_{*,\rm gal}$.

We define the gas fraction as,

$$f_{\rm gas,gal} \equiv \frac{M_{\rm gas,gal}}{M_{\rm gas,gal} + M_{*,\rm gal}} \tag{4.1}$$

and, therefore, the stellar mass is

$$M_{*,\rm gal} = \left(\frac{1}{f_{\rm gas,gal}} - 1\right) M_{\rm gas,gal}. \tag{4.2}$$

We assume all of our galaxies have the same gas fraction, $f_{\rm gas,gal} = 0.7$, estimated from the results in Narayanan et al. (2012) and Tadaki et al. (2019). This assumption is reasonable for high-redshift galaxies, where gas fractions $f_{\rm gas} > 0.4$ are routinely inferred, with a large spread above this value (Carilli et al., 2010; Daddi et al., 2010; Tacconi et al., 2010, 2013). In fact, the median gas fraction of protocluster galaxies measured in Tadaki et al. (2019) is $f_{\rm gas} \approx 0.77$, comparable to our value, with a spread from $f_{\rm gas} \approx 0.4$ to $f_{\rm gas} \approx 0.9$. We discuss the impact of varying $f_{\rm gas}$ in Section 4.3.

As for relating $M_{\rm vir}$ and $M_{*,\rm gal}$, Behroozi et al. (2013b) show that the stellar-to-halo mass fraction is $M_{*,\rm gal}/M_{\rm vir} \sim 0.01$, within a factor of two for a wide range of halos at $z > 4$ that eventually become clusters of mass $\sim 10^{15}$ $M_\odot$, the predicted $z = 0$ mass of the SPT2349-56 system (see Fig. 2 of Miller et al. 2018). We expect that altering the stellar mass to halo mass ratio by a factor of ~ 2 will change our results by the same factor. We do note that the results in Behroozi et al. (2013b) apply to central galaxies; however, we assume that the 14 galaxies in the SPT2349-56 are the centrals of their own halos before commencing to merge, coalesce, and form the protocluster core region. Using these results we estimate $M_{*,\rm gal} \approx 0.428 M_{\rm gas,gal}$ and

$M_{\rm vir} \approx 42.8 M_{\rm gas,gal}$ to within a factor of two. We show the results of our calculations in Table 4.1.

Halo properties

Having an estimate of the virial mass of each system allows us to calculate the halo concentrations and virial velocities. Using the universal halo concentration model from Bullock et al. (2001), the concentration parameter is

$$C_{\rm vir}(M_{\rm vir}, z) \approx 9\left(\frac{M_{\rm vir}}{M_{\rm coll,0}}\right)^{-0.13}(1+z)^{-1}, \tag{4.3}$$

where $M_{\rm coll,0}$ is the typical collapsing halo mass at $z = 0$, with $M_{\rm coll,0} = 1.18\times10^{13}\,{\rm M}_\odot$. Diemer & Kravtsov (2015) show that, for halo masses in our regime of interest ($M_{\rm vir} \lesssim 4.42\times10^{12}\,{\rm M}_\odot$), equation (4.3) is an excellent approximation to their more general models. The virial velocity follows from the virial mass as,

$$V_{\rm vir} \approx (10{\rm G}M_{\rm vir}H(z))^{1/3}, \tag{4.4}$$

where $H(z) = H_0\sqrt{\Omega_{\rm m,0}(1+z)^3 + \Omega_{\Lambda,0}}$. We show the results of equations (4.3) and (4.4), applied to each halo, in Table 4.1.

System dynamics

For the dynamical evolution of the entire system we require the initial positions and velocities of each galaxy. We select initial positions randomly within a sphere of physical radius 65 kpc (the observed maximal separation) for each galaxy with no dependence on the true separations between the observed objects. Fig. 4.1 shows a qualitative view of the positioning of each galaxy in one of the realisations of our synthetic SPT2349-56 system. The observed velocity distribution in Table 4.1 provides the initial velocities of our simulated galaxies. We fit a Gaussian to the distribution and then sample 14 new velocity components for each spatial direction, assuming the velocity distribution is isotropic. We randomly select the orientation of the spin axes for each galaxy.

4.2.2 Galactic Physics

Our sub-grid physics models are the same as those described in Rennehan et al. (2019) (based on the model in Davé et al. 2016), except that we now include a model for

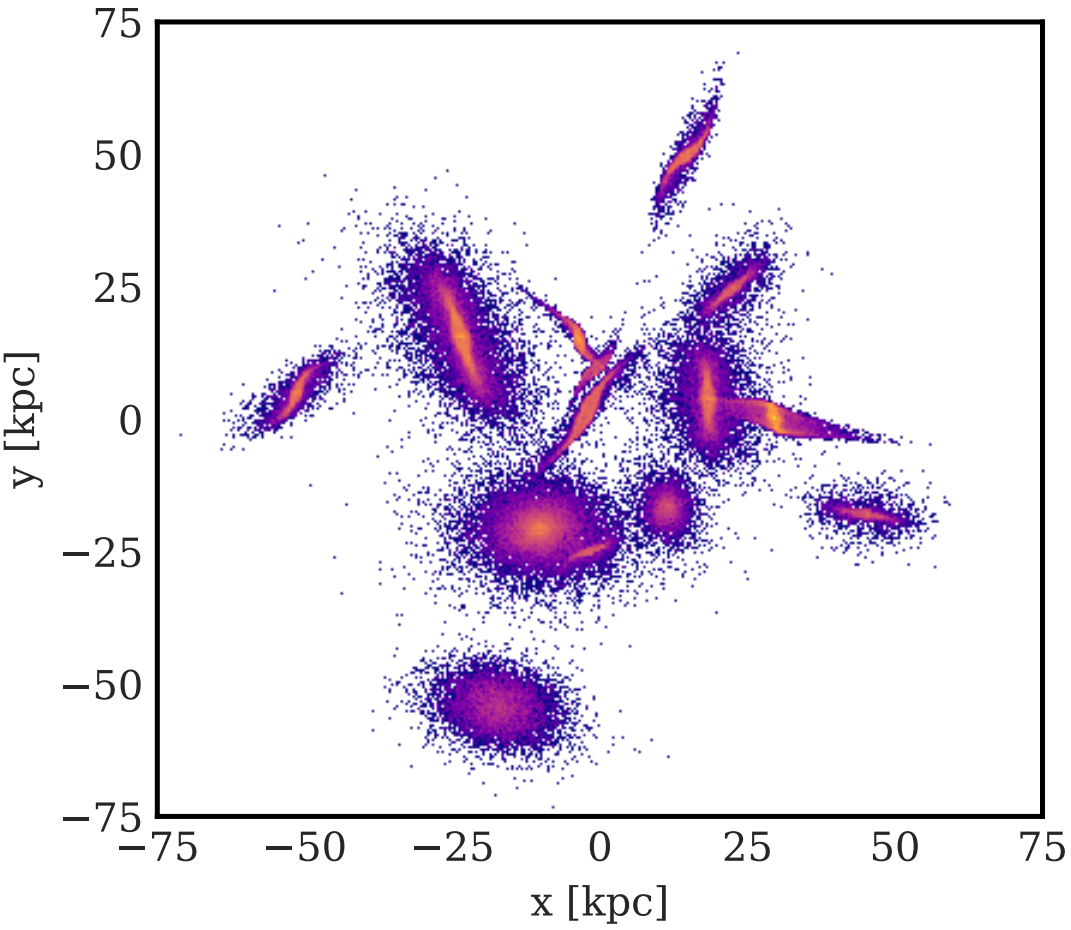

Figure 4.1: A schematic view of one of the realisations of our synthetic SPT2349-56 system. We bin the *xy*-plane positions of each stellar particle with an arbitrary logarithmic scaling at 1 **Myr** after the initial condition. We constrain the mass-weighted centres of each galaxy to lay within a sphere of radius **65 kpc** (physical) based on the observed projected separations.

supermassive black hole (SMBH) growth and feedback. We briefly describe the models below and point the reader to the aforementioned reference for more information.

Cooling and Star Formation

For radiative cooling, we calculate the cooling rates in the presence a UV background (Faucher-Giguère et al., 2009) using the `GRACKLE-3.1` cooling library[6] (Smith et al., 2017).

Our star formation implementation follows that in the MUFASA simulations (Davé et al., 2016, 2017). We determine the conversion rate of gas into stars based on the estimated fraction of molecular hydrogen ($f_{\rm H_2}$) in the gas based on the approximations in Krumholz et al. (2009). We convert gas at densities above the threshold $n_{\rm crit} = 0.2\,{\rm cm}^{-3}$ into stars at a rate ${\rm d}\rho_*/{\rm d}t = \epsilon_* f_{\rm H_2}\rho_{\rm gas}/t_{\rm dyn}$ where ρ_* is the stellar density, $\epsilon_* = 0.02$ is star formation efficiency (Kennicutt, 1998), and $t_{\rm dyn} = (G\rho_{\rm gas})^{-1/2}$ is the local dynamical time. We also force gas onto an artificial equation of state, $T_{\rm EoS} = 10^4 (n_{\rm gas}/n_{\rm crit})^{1/3}\,{\rm K}$, where $n_{\rm gas}$ is the gas hydrogen number density, above the star formation critical density ($n_{\rm gas} > n_{\rm crit}$) to suppress numerical fragmentation (Teyssier et al., 2011; Davé et al., 2016).

Stellar Feedback

We include energetic feedback from supernova types Ia and II (SNIa and SNII, respectively), stellar radiation, and stellar winds from asymptotic giant branch (AGB) stars based on the MUFASA cosmological simulation model. We also include mass injection from SNIa, SNII, and AGB stars, which is important for enriching the gas in the simulation (Davé et al., 2016; Liang et al., 2016). We account for the effects of both prompt and delayed SNIa (Scannapieco & Bildsten, 2005).

Metals are vital in determining the balance of gas cooling and heating in astrophysical gas, and therefore we include their production and account for their role in cooling. We consider metal production by SNIa, SNII as well as AGB stars (Iwamoto et al., 1999; Nomoto et al., 2006; Oppenheimer & Davé, 2008). For details, we refer the reader to Liang et al. (2016), Davé et al. (2016), and Rennehan et al. (2019).

[6] https://grackle.readthedocs.io

Active Galactic Nuclei

High luminosity galaxies often host active galactic nuclei (AGN) concurrently with intense starburst episodes in the local universe (Nardini et al., 2008), and at early epochs (Alexander et al., 2005). Therefore, we also include the effects of AGN feedback into our investigation. AGN are important in determining the correct estimate of stellar mass growth. Our model is that of Springel et al. (2005a), which we briefly describe below.

We initially place black holes of mass 10^5 $M_\odot$ in the centres of each galaxy, and allow them to grow via Eddington-limited Bondi accretion. We use the unboosted Bondi model because the mesh-free finite mass method can resolve higher densities at the same mass resolution, compared to the common smoothed particle hydrodynamics implementations (Hopkins, 2015). We also include energetic feedback, and assume that each AGN generates energy in the gas at a rate $\dot{E} = \epsilon_r \epsilon_f \dot{M}_{BH} c^2$, where $\epsilon_r = 0.1$ is the radiative efficiency, $\epsilon_f = 0.05$ is the coupling fraction to the gas, and c is the speed of light. The accretion rate $\dot{M}_{BH} = 4\pi G^2 M_{BH}^2 \rho_{gas} / (c_s^2 + v_{rel}^2)^{3/2}$ is the Bondi accretion rate onto the black hole, where c_s is the surrounding gas sound speed, v_{rel} is the relative velocity of the SMBH with respect to the gas, and ρ_{gas} is the surrounding gas density. The gas properties – density, sound speed, and relative velocity – are calculated over the nearest 128 neighbouring gas particles. The energy is deposited to the surrounding gas in a kernel-weighted manner, over the same nearest 128 neighbouring particles. To follow the dynamical evolution of the SMBHs, we use the model from Tremmel et al. (2015) in which the dynamical friction force is calculated by using the approximation from Chandrasekhar (1943).

4.3 Stellar assembly and growth

To gain a qualitative understanding of the protocluster assembly, we examine one realisation of the system visually in Fig. 4.2. We bin the positions of each star particle in the simulated xy-plane at three times: $t = 0.12$, 0.5, and 1 Gyr, from top to bottom, respectively.

In the top panel of Fig. 4.2, there are several stellar streams protruding through the system as the galaxies undergo the initial collapse after $\sim$ 120 Myr. These streams are due to tidal stripping from the companion galaxies as the initial velocity dispersion of the system, combined with the close proximity of galaxies, is unable to prevent imminent merging. At maximum distance, the tidal tails extend approximately 90

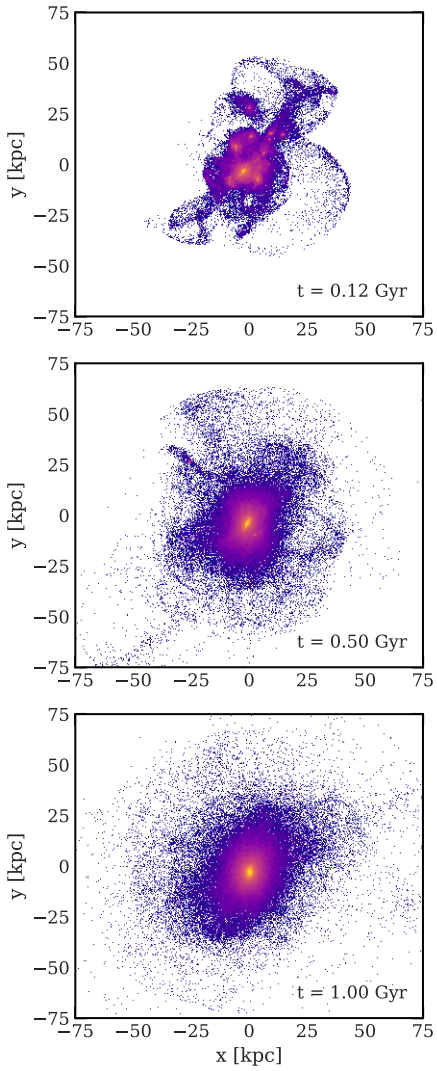

Figure 4.2: A particle view of the assembly of one of the realisations of our synthetic SPT2349-56. We bin the positions of the stellar particles in the simulation in the xy-plane (150 kpc × 150 kpc) using an arbitrary logarithmic scaling. These panels represent t = 0.12, 0.5, and 1 Gyr from the initial condition, from top to bottom, respectively. There are obvious shell-like structures and stellar streams throughout the short assembly period, until the system resembles a massive elliptical galaxy at ∼ 500 Myr.

kpc.

In the middle panel, 500 Myr after the start of the simulation, several streams are visible in addition to shell-like structures surrounding the core of the galaxy. It is difficult to distinguish any of the original structure as the stellar populations begin to mix. At this point during the simulation, many of the stellar particles are launched out to ≈ 75 kpc from the centre-of-mass of the system, building up the diffuse stellar envelope and intracluster light.

At 1 Gyr in the bottom panel of Fig. 4.2, the system resembles a massive elliptical galaxy. Many of the stars from the initial galaxies, as well as those formed *in-situ*, were kicked out and formed an extended diffuse stellar halo. There is no longer any visible structure in this halo. Using radiative transfer we are able to further investigate the evolution of the system.

Fig. 4.3 shows colour-composite mock James Webb Space Telescope (JWST) observations of one of our synthetic SPT2349-56 realisations. We generated these observations using the `SKIRT` radiative transfer code (Baes et al., 2011; Camps & Baes, 2015; Baes & Camps, 2015) and an adapted version of the observational realism suite described in Bottrell et al. (2017a,b). To produce these images, we first assigned star particles spectral templates based on the STARBURST99 (Leitherer et al., 1999) spectral energy distribution set for old stellar populations and Groves et al. (2008) templates, which include emission from HII and photodissociation regions, for young (< 10 Myr-old) star particles. We use a multi-component dust model (Zubko et al., 2004) with a constant dust-to-metal ratio of 0.3 and do not limit dust to star-forming gas particles. `SKIRT` produced rest-frame optical data cubes that we processed into noiseless, idealised photometric images in the JWST NIRCam F150W, F200W, F277W, F356W band passes[7] at redshifts corresponding to the simulation snapshots. In the right column of Fig. 4.3, we show idealised images. The idealised images are noiseless and are neither rebinned down to the NIRCam angular resolution nor convolved with the NIRCam point-spread function. At each of the redshifts we consider $z \in \{4, 3.3, 2.7\}$, corresponding to $t \in \{0.12, 0.5, 1\}$ Gyr, at least three of these filters reside in the protocluster's redshifted rest frame optical domain.

We assigned the filtered light at the smallest wavelengths to the blue values, the mid-range wavelengths to the green, and the longest to the red to construct a qualitative, visual representation of the system from $t \sim 120$ Myr to $t \sim 1$ Gyr (from

[7] https://jwst-docs.stsci.edu/near-infrared-camera/nircam-instrumentation/nircam-filters

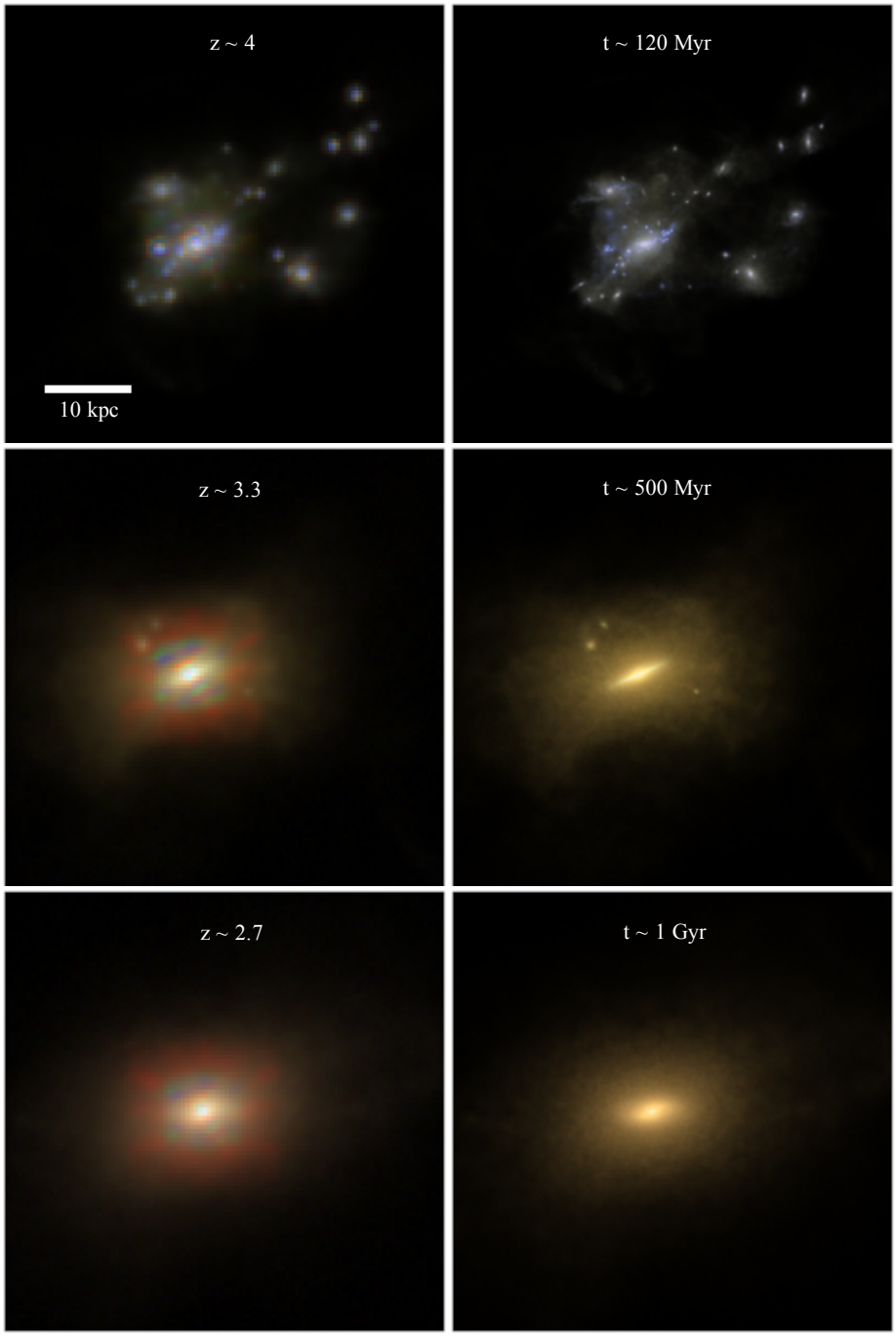

Figure 4.3: (*left*) James Webb Space Telescope mock observations of the forward evolution of SPT2349-56. Redshifts are $z \sim 4$, 3.3, and 2.7, from top to bottom, respectively. (*right*) Synthetic false colour images made from three of the James Webb Space Telescope NIRCam filters without the CCD angular scale or noise. The snapshots correspond to the redshift labels in the left column, for each row. The scale marker shows 10 kpc (physical) and each panel is 50 kpc per side.

top to bottom in Fig. 4.3, respectively)[8].

In the top-right panel of Fig. 4.3, we show $t \sim 120$ Myr after the initial condition. Already by this point in the system's evolution, we see the long stellar streams from the top panel of Fig. 4.2 as a spatially-extended low-surface brightness web surrounding the remnant cores of each initial galaxy. Several bright star-forming cores remain visible in blue and white. Initial gravitational torques cause the fragmentation we observe in the protocluster early on due to the close proximity of the 14 galaxies.

At 500 Myr in the centre row on the right, there are a few remnant cores of the original galaxies and the system assumes a more spherical shape in the diffuse stellar halo. By this point in time, the gas settles into an extended disc in the centre of the gravitational potential, and we see the system shift toward a redder appearance. We analysed the simulation data and discovered that the disc only exists from $t \sim 300$ Myr to $t \sim 600$ Myr, before the remaining cores of the original galaxies dynamically pummel the disc, leading to its demise. It is not surprising that there is a disc given the high gas fraction of the system (Hopkins et al., 2009).

At 1 Gyr, the system is elliptical and resembles a typical low-redshift cD galaxy. Although the disc vanishes, star formation remains at an absolute rate of ~ 40 $M_\odot\,yr^{-1}$, or a specific rate of 5×10^{-11} yr^{-1}, much lower than values predicted in Webb et al. (2015) for BCGs at slightly lower redshift ($z \sim 2$).

In addition to the idealised mock observations in the right column of Fig. 4.3, we show mock observations of the protocluster in the left column. We rebin the idealised images to the NIRCam CCD angular scale corresponding to the appropriate channel (0.031 arcsec pixel^{-1} for F200W and 0.063 arcsec pixel^{-1} for the others), point-spread function convolution, and added noise corresponding to the predicted NIRCam surface-brightness sensitivity for our observing strategy. Using the simulated ramp-optimization feature of the JWST `pynrc` package (Leisenring et al., "pyNRC: A NIRCam ETC and Simulation Toolset", in prep.), we determined the optimal observing strategy for the protocluster assuming 10 ks of observing time. For more details, see Appendix 4.7. We analyse these results at three selected redshifts: $z \in \{4, 3.3, 2.7\}$, from top to bottom in Fig. 4.3, respectively. These leftmost panels correspond to the same redshift as the idealised images in the same row in Fig. 4.3.

In the top-left panel, the substructure at $z \approx 4$ is clearly visible in the JWST composite image. We find that the absolute AB-magnitude in the F277W NIRCam

[8]Specifically, we use the NIRCam filters {F200W, F277W, F356W} at 120 Myr, and {F150W, F200W, F277W} at both 500 Myr and 1 Gyr.

band is $M_{\mathrm{AB,F277W}} \approx -28.7$, making our synthetic SPT2349-56 an extremely bright object. In an analysis of 430 brightest cluster galaxies in Donzelli et al. (2011), the brightest galaxy (object A0401 in the Table 2 of that study) has an absolute R-band magnitude of −27.25. Although our filters are not equivalent, we find the same approximate magnitude order in all of the JWST filters we apply and, therefore, we can conclude that SPT2349-56 would end up being one of the brightest galaxies in the observable universe. Not only is our realisation bright, it is also very blue. We calculated the rest-frame SDSS g − r colours for the system and find g − r ≈ −0.05 at $z \approx 4$ for our dust model. A typical BCG has a redder colour at g − r ≈ 0.6 (Cerulo et al., 2019).

The qualitative view of the BCG at a redshift of $z \approx 3.3$ corresponds to 500 Myr. At this time, the stars remain blue in the system (g − r ≈ 0.13), which is interesting since we do not expect fully assembled, highly star-forming BCGs at $z \approx 3.3$ with a large fraction of young stellar populations. If the predictions of substantial (i.e. factors of ~ 2 to ~ 4) growth past $z \sim 1$ in the literature (De Lucia & Blaizot, 2007; Lavoie et al., 2016; Ragone-Figueroa et al., 2018) are correct, then systems such as SPT2349-56 would go on to become the most-massive galaxies in the universe at $z = 0$, reaching stellar masses up to $\sim 3 \times 10^{12}$ $\mathrm{M}_{\odot}$. We do not include any of the substructure that may exist outside of the core of the SPT2349-56 protocluster, which could increase our estimate even further if those galaxies were to accrete between $z \sim 3$ and $z \sim 1$ via dynamical friction (i.e. galactic cannibalism; Ostriker & Tremaine 1975; White 1976). However, since they grow rapidly to a large mass, the dynamical friction timescale for lower-mass satellites to merge into the system becomes large, which could hamper late-time growth.

In the top panel in Fig. 4.4, we show the forward-evolution of the star formation rate of our synthetic SPT2349-56 system (i.e. our fiducial simulation with $f_{\mathrm{gas,gal}} = 0.7$) as a function of time, averaged over three realisations, up to 1 Gyr ($z \approx 2.7$) after the observation ($z \approx 4.3$)[9]. The star formation rate (SFR) peaks at ~ 3000 $\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$ approximately 5 Myr after the initial condition and decays exponentially to ~ 40 $\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$ at 1 Gyr. Our simulated SFR is comparable to the observed results of Miller et al. (2018) who find a total star formation rate of ≈ 6500 $\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$ for the 14 galaxies in the SPT2349-56 system. We use a non-linear least squares fitting method to fit an exponential curve to the SFR, and find that the decay time is $\tau \sim 200$

[9]In terms of redshift, our simulation began at $z \approx 4.3$ and ended 1 Gyr later at $z \approx 2.7$, assuming a Planck Collaboration XVI (2014) cosmology.

Myr. Initially, the SFR peaks due to high compression from the proximity of each individual galaxy, and the fact that they are rapidly collapsing under their mutual gravity. Additionally, strong fluctuations in SFR occur as the discs undergo tidal interactions. Stellar feedback is the main cause of the declining SFR, as the black holes provide little feedback past the initial ~ 50 Myr due to their low accretion rates. They are unable to accrete because of their high velocities relative to the medium – which suppresses accretion. On the other hand, stellar feedback is most powerful when the star formation rate is the highest, as the system coalesces, and the supernovae and stellar winds begin to excavate the gas out of the original galaxies.

The balance of stellar feedback and star formation causes a smooth decline in star formation rate, and consequently a smooth increase in the stellar mass, as we show in the bottom panel in Fig. 4.4. The total stellar mass of our fiducial simulation is given as the solid black line with initial and final (at 1 Gyr) stellar masses, $M_{\rm initial} = 2.81 \times 10^{11}$ M$_\odot$ and $M_{\rm final} = 8.14 \times 10^{11}$ M$_\odot$, respectively. The total stellar mass is the sum of all of the individual stellar particle masses in the entire simulation volume. The assembly mass, which we define as 90% of the total final mass, is $M_{\rm assembly} = 7.33 \times 10^{11}$ M$_\odot$. The time at which assembly occurs is $t_{\rm assembly} \sim 370$ Myr. Assembly occurs rapidly in the system, and our results show that the system nearly quadruples size within a giga-year.

Apart from the total stellar mass, we also examine the growth of the stellar mass of largest galaxy in the simulation, which we label as the evolving protoclusterâĂŹs BCG. Observationally, the stellar masses of BCGs are usually determined within an aperture with a specified radius since only the projected stellar light distribution is known. Therefore, to compare to observational stellar masses, it is more reasonable that we choose an aperture centred on the peak stellar brightness and sum the stellar mass along the corresponding cylinder. Kravtsov et al. (2018) advocate apertures of fixed radii (they used 30 kpc, 50 kpc, and 70 kpc) for comparing simulated BCG masses to observed masses. Burke et al. (2015), whose masses we compare to below, adopted aperture sizes for each BCG in the Cluster Lensing and Supernova survey with Hubble (CLASH) (Postman et al., 2012) sample such that it encompassed $\sim 90\%$ of the flux. Stott et al. (2010) used the same method to determine the BCG stellar masses for the systems in the XMM Cluster Survey (XCS; Romer et al. 2001; Sahlén et al. 2009), and showed that a fixed 50 kpc aperture gave similar results. A 50 kpc aperture is, however, too large for our high-redshift protocluster since the diameter of the system is only 135 kpc. A 50 kpc aperture centered on the peak in the projected

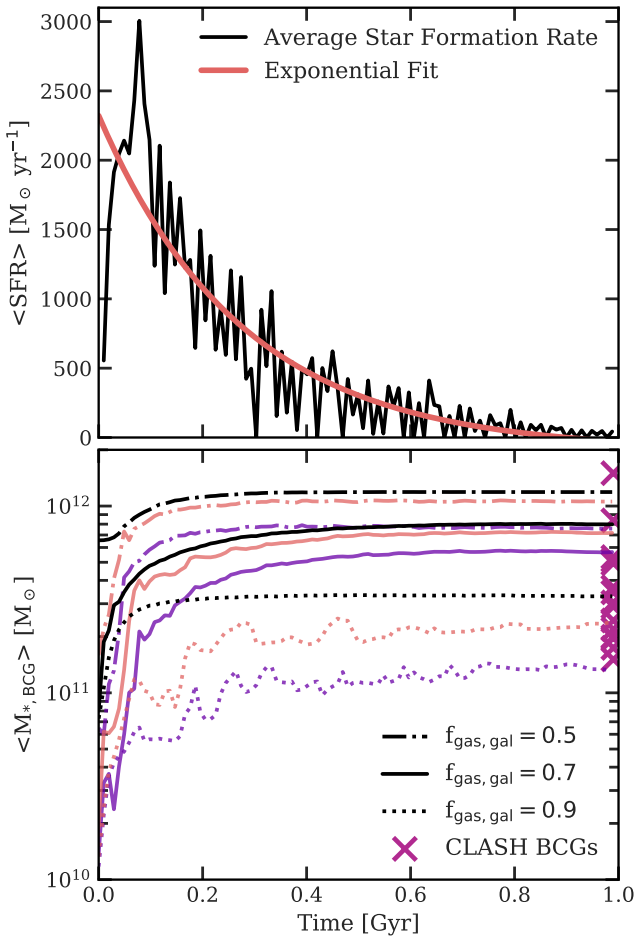

Figure 4.4: Growth and assembly history of the protocluster simulations. (*top*) Star formation rate over the first giga-year of evolution averaged over three independent realisations of our fiducial simulation with initial gas fraction $f_{\rm gas,gal} = 0.7$. We used a non-linear least squares method to fit a decaying exponential and found a decay timescale of $\tau \approx 200$ Myr. (*bottom*) The black curves show the total mass of all stellar particles in the simulation volume for each system with different initial gas fraction. The purple and coral curves show the stellar mass within 5 kpc and 15 kpc apertures, centred on the peak projected stellar mass, respectively. We show the observed masses of brightest cluster galaxies from the CLASH survey as × symbols at 1 Gyr which were derived using `MAG_AUTO` magnitudes. After 1 Gyr ($z \approx 2.7$), the total stellar masses across the three simulations with varying gas fractions is larger than most of the lower-redshift CLASH brightest cluster galaxies.

stellar mass distribution[10] – approximating the location of the peak surface brightness – encompasses 95% of the total stellar mass (black curves in Fig. 4.4) over course of the simulation. To isolate the most massive galaxy in our simulated protocluster and track the growth of its stellar mass over time, we use smaller apertures of 5 kpc and 15 kpc since the effective radius (i.e. the half-light radius) of the stellar distribution is ≈ 2.2 kpc at 1 Gyr in our fiducial simulation and, therefore, the apertures enclose a few effective radii.

The purple and coral solid curves in the bottom panel of Fig. 4.4 show the stellar mass growth within the aforementioned cylindrical apertures of radii 5 kpc and 15 kpc apertures, respectively. The smaller apertures more closely track the mass growth in the most massive galaxy across the entire simulation although the merging process is so rapid that distinguishing between the BCG and the debris from in-falling, merging, disrupting galaxies past $\sim$ 100 Myr is impractical, and the differences between the stellar masses within each aperture become less significant. The masses with the apertures rise rapidly as the galaxies coalesce over the course of the first $\sim$ 100 Myr (see the top panel of Fig. 4.2) and converge to their final values by $200-250$ Myr. At 1 Gyr, the 5 kpc and 15 kpc apertures contain $\approx 70\%$ and $\approx 90\%$ of the total stellar mass, respectively.

We also show the observed masses of brightest cluster galaxies (BCGs) in Burke et al. (2015) from the CLASH survey with $\times$ symbols at 1 Gyr. These BCGs belong to clusters with masses $\gtrsim 10^{15}$ $M_\odot$, and range in redshift from $0.187 < z < 0.890$. Although our system is at $z \sim 2.7$ after 1 Gyr, the total stellar mass and the mass within the smaller apertures are already more massive than most of the BCGs from the CLASH survey. We emphasise that we only mean to compare the observed BCG masses with our simulated BCG at 1 Gyr, after the system resembles a fully-formed BCG (see the bottom panel of Fig. 4.2).

It is pertinent that we emphasise that we designed our numerical experiment to be simple in nature to determine the star formation rate and assembly timescale of the protocluster. We did not tune the star formation rate to match the observations and, given that we are within a factor of ~ 2, the broad assumptions appear reasonable. The main sources of error are (i) our choice for the gas fraction and (ii) the abundance matching results. To test the dependence on gas fraction, we simulated two additional realisations with a lower ($f_{\rm gas,gal} = 0.5$) and higher ($f_{\rm gas,gal} = 0.9$) gas fraction. We show the total stellar mass evolution in the protocluster system for each initial gas

[10]We tested multiple viewing angles and found that our results do not change significantly.

fraction in the bottom panel of Fig. 4.4.

In the case of $f_{\rm gas,gal} = 0.5$, the total initial stellar mass and, therefore, dark matter mass are a factor of ~ 2.3 larger than in our fiducial simulation. This is because we fix our gas masses based on the observations in Miller et al. (2018). We find that the total stellar mass of this low gas fraction system at 1 Gyr is $\sim 1.2{\times}10^{12}$ $M_\odot$, which is a factor of ~ 1.5 higher than in our fiducial simulation but still much more massive than the majority of CLASH BCGs. The dashed purple and coral curves in Fig. 4.4 show that the growth within 5 kpc and 15 kpc is rapid in this system and the most massive galaxy quickly grows in size. At 1 Gyr, $\approx 99\%$ of the total stellar mass is contained within the 5 kpc aperture and $\approx 90\%$ is contained within the 15 kpc aperture. Based on our definition of the stellar assembly time, we find $t_{\rm assembly} \sim 150$ Myr which is a factor of ~ 2.5 smaller than in the fiducial simulation. Our low gas fraction system therefore reaches a slightly higher stellar mass and does so more quickly.

Next, we consider the simulation with a high gas fraction of $f_{\rm gas,gal} = 0.9$. This results in an initial stellar mass a factor of ~ 4 lower than our fiducial simulation and, therefore, the same factor lower in halo mass. The total stellar mass at 1 Gyr of the simulated system is $\sim 3.3 \times 10^{11}$ $M_\odot$ – a factor of ≈ 2.5 lower than our fiducial simulation. This BCG is much less concentrated than the other BCGs with lower gas fractions as we find $\approx 50\%$ and $\approx 85\%$ of the total stellar mass in the apertures of radii 5 kpc and 15 kpc, respectively. We find that the assembly time is a factor of ~ 3 shorter than in our fiducial simulation at $t_{\rm assembly} \sim 120$ Myr.

In summary, the above variations in the initial gas fractions result in only factors of 2–3 in the final stellar mass and assembly time. Therefore, we argue that our results are relatively robust; we expect that all scenarios involving factors of ~ 2 changes in $f_{\rm gas,gal}$ or $M_{*,\rm gal}/M_{\rm vir}$ will result in a massive BCG at high-redshift that assembles its stellar mass very quickly (within ~ 1 Gyr). We attribute the differences in assembly timescales between the simulations with different initial gas fractions as due to the varying potential well depth. In the lowest gas fraction case, the potential is deeper and the interactions are stronger, therefore the assembly process occurs on shorter timescales. In the high gas fraction case, the potential well is shallower and stellar feedback, following the initial intense star-burst, leads to the expulsion of gas – quenching the system more rapidly than in the fiducial case.

As for the choice of $M_{\rm vir} \sim 0.01 M_{*,\rm gal}$ (see Fig. 8 of Behroozi et al. 2013b), for halos at $z > 4$ which become halos of mass $\sim 10^{15}$ $M_\odot$ at $z = 0$, the stellar mass to halo mass ratio is ~ 0.01 to within a factor of two. Decreasing (increasing) the ratio

would lower (raise) the virial mass of the halo, which in turn will alter the dynamics of the system at the level of a factor $\sim 2-3$, similar to what we discussed above.

As an additional source of uncertainty, Granato et al. (2015) point out that it is difficult to reproduce high star formation rates in high-redshift protoclusters in numerical simulations using standard sub-grid models of star formation, stellar feedback, and active galactic nuclei feedback. However, the sub-grid models we used in `GIZMO` have been shown to reproduce broad galaxy population properties (Davé et al., 2017, 2019) at a wide-range of redshifts and in high-redshift ($z \sim 6$) galaxies at similar particle mass resolutions (Olsen et al., 2017).

We posit that recently discovered systems similar to SPT2349-56 (e.g. Ishigaki et al. 2016; Jiang et al. 2018; Higuchi et al. 2019) are the proto-cores of the massive galaxy clusters. The speed of assembly and growth of stellar mass in our simulated realisations of SPT2349-56 is obvious from Figs. 4.2, 4.3, & 4.4. The star formation rate declines exponentially with an e-folding time of $\sim$ **200 Myr** while simultaneously, an object that qualitatively looks like many observed BCGs forms by **500 Myr**. From these results, we predict that the observed high-redshift over-dense protoclusters (Ishigaki et al. 2016; Miller et al. 2018; Jiang et al. 2018; Wang et al., in prep.) will undergo a similar evolution and therefore form massive BCGs as early as $z \approx 4$. As we demonstrated, the JWST will be able to clearly see the massive BCGs out to redshift of $z \approx 3$, and their progenitors out to $z \approx 4.3$, opening up a new frontier for exploration, especially in collaboration with survey telescopes such as the Wide-Field Infrared Survey Telescope (WFIRST).

4.4 Implications for galaxy clusters

In the hierarchical structure formation scenario, galaxy clusters are the youngest and most-massive objects in the universe. In this theory, the more massive the cluster, the younger and rarer the system. Hence, the most-massive clusters should continue to assemble the bulk of their mass until late times. What we have demonstrated in the previous section is the possibility of a *downsizing* effect (Bower et al., 2006; Cimatti et al., 2006; Neistein et al., 2006; Fontanot et al., 2009; Oser et al., 2010) on the cluster scale, where the cores of these massive clusters could be much older than the cores of less-massive clusters, beginning to assemble at redshifts $z \gtrsim 2$ at the minimum. Our results on the rapid assembly of the brightest cluster galaxy (BCG) are then not unexpected as these high-redshift protocluster cores would be the strongest relative

overdensities in the early universe. This begs the question: why has the theoretical community not predicted high-redshift fully-assembled BCGs?

One of the issues is that the main tools of contemporary theoretical astrophysicists are numerical simulations. These are necessary because of the strong non-linearity of the structure formation process after the linear perturbation theory breaks down. However, a lack of computational power limits the spatio-temporal dynamic range of the numerical simulations. We are therefore forced into a compromise. We can model a small comoving volume of the universe and resolve the galaxies in this volume (e.g. Schaye et al. 2015; Pillepich et al. 2018; Davé et al. 2019) but such volumes generally do not contain the rare massive clusters we would expect to host an object such as SPT2349-56. Or, we can sacrifice resolution in favour of large volumes but then galaxy formation must be introduced in an ad hoc fashion (Ruszkowski & Springel, 2009), which can introduce biases. Despite these difficulties, we do expect that these protoclusters are present in the largest dark-matter-only simulations, such as the MultiDark Bolshoi (Klypin et al., 2011) and Millennium XXL (Angulo et al., 2012) simulations.

To test our theory of early, rapid assembly of the cores of the most-massive clusters with regards to the general population of galaxy clusters, we shift focus from SPT2349-56 and now investigate the assembly history of the most massive clusters (at $z = 0$) in the MultiDark Planck 2 (MDPL2) simulation (Riebe et al., 2013; Klypin et al., 2016) – a child of the MultiDark Bolshoi suite of simulations. Our goal is to determine if there is a population of BCGs that assemble early and their relationship to their host clusters. The MDPL2 simulation consists of 3840^3 particles within a simulation volume of side-length 1 cGpc h^{-1}, and has a mass resolution of $1.51{\times}10^9$ $\mathrm{M}_\odot h^{-1}$. All of the halo data is publicly available online[11] and we specifically use the `MDPL2.Rockstar` database in the following analysis. This database contains halo properties that were determined using the `ROCKSTAR` halo finder (Behroozi et al., 2013a), and includes the substructure trees for each host halo.

First, we require a set of criteria for the occurrence of highly over-dense massive collapse events at high redshift. We analysed the formation histories of all galaxy clusters that had final masses[12] of $M_{\rm vir} \geqslant 5{\times}10^{14}$ $\mathrm{M}_\odot$ at $z = 0$, and narrowed our search to those with a large number of relatively massive halos entering their progenitor's

[11] https://www.cosmosim.org

[12] Henceforth we assume a Planck Collaboration XVI (2014) cosmology for h in our quoted masses and distances.

virial radius across cosmic time. Specifically, we define an over-dense collapse event to be when $N \geqslant 5$ halos of individual mass $M_{\rm vir} \geqslant 2{\times}10^{11}$ $M_\odot$ all enter the virial radius of a more massive halo within a time Δt, where Δt is the time between two consecutive simulation snapshots. For example, at $z \approx 4.5$ this timescale corresponds to ~ 40 Myr in the `MDPL2.Rockstar` database. We also ensure that there are at least 5 lower mass halos that are no more than 20 times less massive than the more massive halo. The latter condition considers only those events in which the substructures have a chance of merging within a Hubble time at their respective redshift due to dynamical friction, estimated at the virial radius of the more massive halo (Mo et al., 2010),

$$t_{\rm df}|_{r_{\rm i}=R_{\rm vir}} \approx \frac{1.17}{\ln(M_{\rm large}/M_{\rm sub})}\left(\frac{M_{\rm large}}{M_{\rm sub}}\right)\frac{1}{10H(z)}, \tag{4.5}$$

where $M_{\rm large}$ is the virial mass of the more massive system, $M_{\rm sub}$ is the mass of the substructure, and $H(z) = H_0\sqrt{\Omega_{\rm m,0}(1+z)^3 + \Omega_{\Lambda,0}}$. Each structure is within a factor of ~ 20 of the more massive halo and, therefore, $t_{\rm df}(z) \sim 3t_{\rm H}(z)/4$ where $t_{\rm H}(z) = 1/H(z)$ is the Hubble time. We consider the dynamical friction timescale to be a conservative estimate of how quickly the systems will merge since our criteria captures all mergers with $M_{\rm large}/M_{\rm sub} \lesssim 20$ and the merging halos are not on circular orbits at the virial radius; therefore, they should merge faster (Poole et al., 2006; Boylan-Kolchin et al., 2008).

We also assume that the substructures and the host each contain a galaxy with stellar mass given by the abundance matching relations in Behroozi et al. (2013b). One important caveat to note is that the results of Behroozi et al. (2013b) apply to central galaxies and not satellites. However, since the substructures are independent halos before merging, this implies that our criteria is equivalent to a minimum of 6 central galaxies merging. For comparison, the most massive galaxy in our scenario plays a similar role to galaxy A in Table 4.1 and in the analysis of Section 4.3.

We determine the first redshift at which an over-dense event occurs for each of the galaxy clusters above our mass limit. The least massive events we find are a factor of ~ 20 less massive than our estimated mass for SPT2349-56 and ought to rapidly (if found above $z \gtrsim 3$) collapse into elliptical galaxies of mass $M_* \gtrsim 10^{11}$ $M_\odot$, if we assume the same abundance matching estimates from Section 4.2 combined with the aforementioned dynamical friction timescale constraint.

In the top panel of Fig. 4.5, we present a two-dimensional histogram that encodes the occurrence of high-density collapse regions for a cluster of a given final mass at

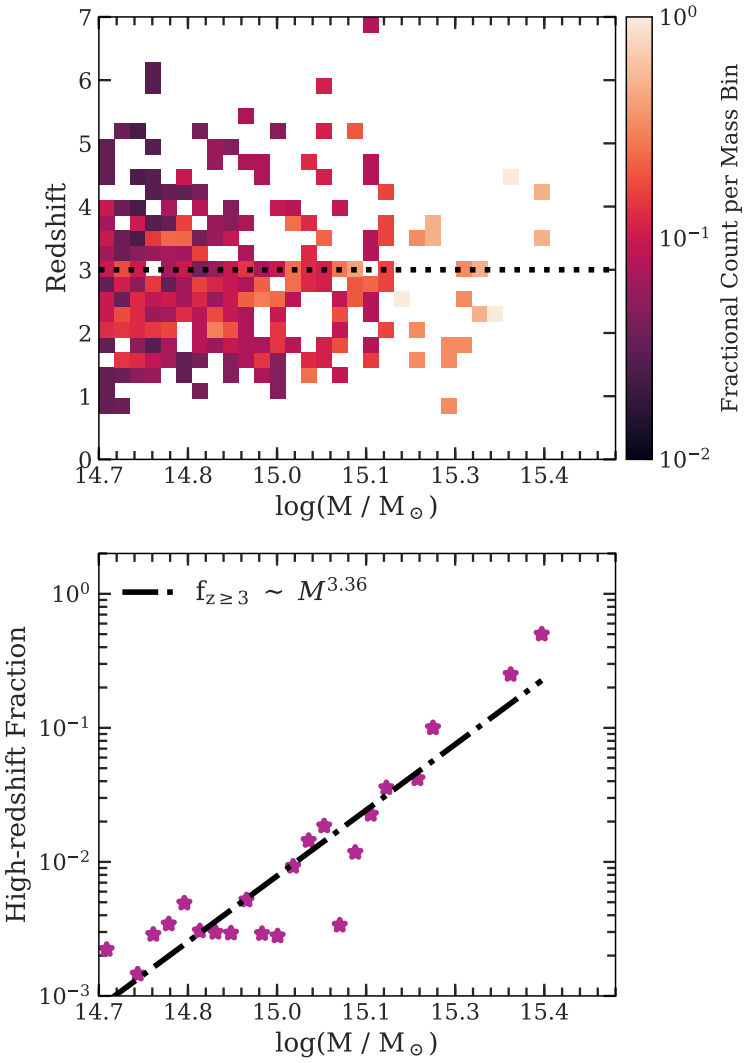

Figure 4.5: (*top*) A two-dimensional histogram showing the probability of the first occurrence of a highly over-dense region collapsing at a specific redshift, in a given bin of final cluster virial mass (i.e., the virial mass at $z = 0$) spanning from 5×10^{14} $M_\odot$ to 3.25×10^{15} $M_\odot$ (the most-massive cluster). We remove non-events from this histogram so that the fraction is truly the probability of finding an event at a specific redshift, given that such an event occurs. There is a slight trend for more instances of the events at higher final cluster masses but the scatter in the distributions shows that many events occur for all of the clusters here above $z = 3$. (*bottom*) The number of over-dense collapse events above $z = 3$, normalised to the total number of clusters in each mass bin. The bins are identical to those in the top panel. The dash-dot line shows a power-law fit with $f \sim M^{3.36}$ – high-redshift over-dense collapse events are more frequent as a function of increasing final cluster mass.

$z = 0$. Specifically, we count the number of first occurrences of an over-dense event in a grid of redshift and final cluster mass coordinates. We normalise to the total number of events in each mass bin, so that the colouring shows the probability of finding the first event for that final cluster mass at a given redshift, compared to all other events that occur. We do not include non-events in the top panel as our interest lies in the probability of an event occurring at a specific redshift, given all of the events that occur.

From the top-panel in Fig. 4.5, lower-mass clusters at $z = 0$ are less likely to have a high-density collapse at redshifts $z \gtrsim 3$ (above the dashed line in Fig. 4.5) compared to clusters with masses $M_{\rm vir} \gtrsim 10^{15}$ M$_\odot$. There are, however, clusters that did not experience an over-dense collapse at any redshift.

The trend of increasing event frequency with increasing final cluster mass is more evident in the bottom panel of Fig. 4.5, where we show the number of events above $z \geqslant 3$ in each mass bin, normalised to the total number of clusters in each mass bin. We show a fit to the fraction as a function of $z = 0$ cluster mass, and find that the fraction scales as $f \sim M^{3.36}$. There is a clear power-law trend where the fraction of events increases with increasing $z = 0$ mass of the clusters. In the highest mass bin, there is a $\approx 50\%$ chance of all clusters having a high-redshift over-dense collapse event occur at $z \geqslant 3$.

For the present purposes, our interest lies in the objects that collapse in the range $z \gtrsim 3$ since these would end up as the most-massive, blue, elliptical galaxies before $z \sim 1.5$ by our predictions. We find that 16.4% of clusters with final masses $M_{\rm vir} > 10^{15}$ M$_\odot$ have an over-dense collapse event occur above $z = 3$, and 9.8% of clusters with final masses 5×10^{14} M$_\odot \leqslant M_{\rm vir} < 10^{15}$ M$_\odot$. In total, we find 155 high-density collapse events at $z \geqslant 3$. Given the volume of the simulation, $V \approx 3.2$ cGpc^3, we expect a comoving number density of corresponding massive, blue, elliptical cluster BCGs at redshifts $z \gtrsim 1.5$ to be $n \approx 48$ cGpc^{-3} (or $n \approx 4.8 \times 10^{-8}$ cMpc^{-3}).

In Fig. 4.6, we examine the spatial, mass, and velocity distribution (from top to bottom, respectively) of an example cluster proto-core in the MDPL2 simulation[13]. This protocluster is the most-massive progenitor (MMP) at $z = 4.266$ (the approximate redshift of SPT2349-56) of the second-most-massive cluster at $z = 0$.

In the top panel of Fig. 4.6, we show the spatial distribution of the substructure entering the virial region of the MMP. The dotted line shows the observed extent of the SPT2349-56 object for comparison. Each halo is marked with a triangle that we

[13]Specifically, we examine the halo with `rockstarId`=923730455 in the `MDPL2.Rockstar` table.

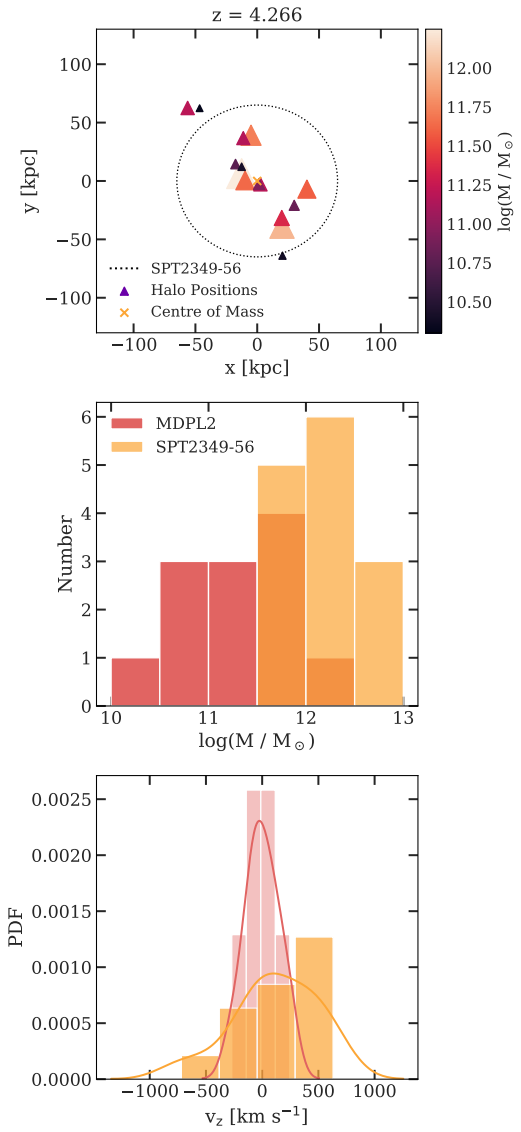

Figure 4.6: An example of one of the many high-redshift, over-dense regions we found in the Multi-Dark Planck 2 simulation. This system is the progenitor of the second-most-massive galaxy cluster at $z = 0$, and is approximately an order of magnitude less in mass than our estimation of SPT2349-56. (*top*) Spatial (physical) positions of substructure within a more-massive halo ($M_{\rm vir} \sim 10^{13}$ $\mathrm{M}_{\odot}$). The physical projection of the observed SPT2349-56 region is overlaid with a dotted line. (*middle*) Virial mass histogram of substructure compared with the estimated virial masses of the dark matter halos in SPT2349-56. We consider only those masses that lay within the dotted line. (*bottom*) Kernel density estimate of the substructure velocities along the line of sight. We show the SPT2349-56 observed values as a comparison. As above, we only consider substructure within the dotted circle in the top panel for the velocity distribution.

colour based on the logarithmic mass of the halo, and we arbitrarily scale the size of each triangle to the virial radius of each halo. Furthermore, we mark the centre of mass of the system with a × symbol. The virial radius of the MMP is $R_{\rm vir,phys} \approx 120$ kpc and we see that most of the massive substructure is within the SPT2349-56 sphere, or approximately half of the virial radius of the MMP. In this particular system, there are eight halos with masses $M_{\rm sub} > 10^{11}$ $M_\odot$, and we assume that they will merge on timescales shorter than the dynamical friction timescale taken at the virial radius of the host following equation 4.5. At $z = 4.266$ the Hubble time is $t_{\rm H} \approx 1.43$ Gyr assuming a Planck Collaboration XVI (2014) cosmology and, therefore, $t_{\rm df} \approx 1$ Gyr. Of course, this is a reasonable estimate since the substructure is within half of the virial radius of the more massive system and, therefore, will merge much more quickly. Evidently, over-dense systems such as the example in the top panel assemble rapidly in the early universe. However, will they go on to form objects as massive and bright as our synthetic SPT2349-56?

To address this question, we show the mass distribution of our cluster proto-core in the middle panel of Fig. 4.6. As we mention above, there are eight objects above $M_{\rm sub} > 10^{11}$ $M_\odot$, and we first assume that each of these halos host a galaxy. Using the same methodology as in Section 4.2.1, we estimate that each halo should host a galaxy with stellar mass that is 1% of the halo mass. Therefore, each galaxy in the system should contain, on average, a stellar mass of $M_* \sim 5 \times 10^{10}$ $M_\odot$. If we use a conservative estimate of mass doubling within 1 Gyr of evolution based on Section 4.3, we would expect this object to become an elliptical galaxy with stellar mass $M_* \sim 10^{11}$ $M_\odot$ at $z \approx 2.7$ and, like the BCG in our simulation, it would be extremely bright and blue at that redshift.

In the bottom panel of Fig. 4.6 we show a comparison of the velocity differences between SPT2349-56 and our example system. We calculate the velocity offset in the perpendicular direction to the top panel, relative to the centre of mass of the system, in order to mimic how Miller et al. (2018) determined the velocities of their object[14]. The solid lines show the kernel density estimate of the underlying distribution for both our system and SPT2349-56. The velocity dispersion of our example system is $\sigma \approx 590$ km s^{-1} and we find a tightly peaked distribution within the extent of the velocity dispersion, indicating that the system is indeed bound, with no signs of filamentary structure within the virial radius. In SPT2349-56, we see that there is

[14]The velocity distribution of the simulated halos does not change appreciably with respect to the projection plane.

a bias toward approaching velocities, indicating that the system is most likely being fed by a filament.

Our investigation into the MDPL2 simulation suggests that there is a non-negligible fraction ($\sim$ 10%) of BCGs that assemble their stars extremely rapidly in the early universe. This counters the notion of late-time assembly for the entire population of BCGs, and especially counters the idea of treating the entire set of BCGs as a homogeneous population to begin with. What we have not addressed is the overall trend of later assembly with decreasing cluster mass and what this means for the BCG population. Using the same aforementioned timescale estimates, systems that we identify as highly over-dense regions that collapse at $z \sim 3$ should go on to form BCGs with stellar masses $M_* \gtrsim 10^{11}$ $M_\odot$ by $z \sim 1.5$. Therefore, our results suggest that there is a continuous population of bright, star-forming BCGs down to $z \sim 1.5$ – where the Hubble time is still short, allowing for the rapid assembly of substructure. Recent observations have indeed begun to undercover this population of star-forming galaxies (McDonald et al., 2016), especially with Webb et al. (2015) finding that the star formation rate in their BCG sample increases to $\sim 1000 - 3000$ $M_\odot\,\mathrm{yr}^{-1}$ as a function of increasing redshift. Similarly, the core of the recently investigated galaxy cluster SpARCS104922.6+564032.5 at $z = 1.71$ (Webb et al., 2017; Trudeau et al., 2019) appears qualitatively similar to our results from Section 4.3, and the scenario we propose above could explain the extended star formation and morphology within that system.

There is an additional connection of these high-redshift BCGs to galaxy clusters – one that could help ameliorate discrepancies in the determination of cosmological parameters. Specifically, the value of the density fluctuation power spectrum amplitude, σ_8, derived from galaxy cluster counts is known to disagree at $\sim 2\sigma$ with the value derived from the cosmic microwave background in the Planck mission (Ade et al., 2014; Douspis et al., 2018; Salvati et al., 2018). The number of massive clusters is highly sensitive to σ_8 and we expect these BCGs to preferentially exist in those clusters. In this study, through our use of the MDPL2 simulation, we use the Planck Collaboration XVI (2014) value of σ_8 which is higher than those derived from cluster counts. As it is, events such as SPT2349-56 are rare in this cosmology and ought to be rarer still if σ_8 were even slightly lower. A combination of surveying with the Wide-Field Infrared Survey Telescope (WFIRST) and follow-up confirmation with the JWST would provide a lower-bound on the number density, hence providing an additional constraint on σ_8 that could settle the issue. While we provide only an

approximate estimate of the number density of massive BCGs, a more detailed study could provide an exact constraint on the value.

One major caveat is that for the BCGs to be highly star-forming, they must not exhaust their gas supply, and they must not have their star formation halted by active galactic nuclei feedback. However, given that the regions we discovered at $z \gtrsim 3$ are highly over-dense, they undergo what could be considered an effectively *monolithic* formation scenario where the substructure gives rise to galaxies that are gas rich, which then rapidly merge to form the BCG. And only afterwards, after the newly formed BCG's gas content depletes due to feedback and consumption, would star formation quench and the galaxy age passively.

4.5 Conclusions

The cores of galaxy clusters are home to the most luminous galaxies in the universe – the brightest cluster galaxies (BCGs). These galaxies have unique properties, such as their velocity dispersion and luminosity profiles, that set them apart from other galaxies at the high-end of the galaxy luminosity function. The contemporary picture of their formation and growth scenario is that their stars are old and formed at high redshift ($z \gtrsim 4$) in separate individual galaxies that, at late times ($z \lesssim 1$), hierarchically assemble to form the massive galaxies we observe today. There are, however, recent observations of highly-overdense protoclusters at $z \gtrsim 4$ (Ishigaki et al., 2016; Miller et al., 2018; Jiang et al., 2018) that muddle this simple picture, since we expect these to go on to form the BCGs (Ito et al., 2019).

The hierarchical assembly picture of structure formation predicts the gradual build-up of objects through successive mergers of small objects. Under this theory, we expect the most-massive galaxies to be the youngest in terms of their assembly time. However, observations of massive ellipticals show a *downsizing* effect whereby the mass of the systems and the stellar ages are anti-correlated (Cimatti et al., 2006; Bower et al., 2006; Fontanot et al., 2009). In other words, the most-massive systems appear to have the oldest stellar populations. We propose a new paradigm wherein a similar downsizing effect occurs on the scale of galaxy clusters themselves. In our proposal, the cores of the most-massive clusters – the BCGs – assemble earlier than the cores of lower-mass clusters, on average. A subset of cluster-cores are assembled at very high redshift ($z \gtrsim 3$), with the probability of high-redshift assembly decreasing as a function of decreasing mass (at $z = 0$) of the clusters.

In order to determine the rapidity of assembly and growth of the BCGs, we studied the forward-evolution of a recently discovered, highly over-dense protocluster at $z \approx 4.3$, SPT2349-56 (Miller et al., 2018) using a non-cosmological hydrodynamic simulation informed by the observations. We found that the star formation peaks at $\sim 3000\,\mathrm{M_\odot\,yr^{-1}}$ very early in our simulation, and decays exponentially with a timescale of $\tau \sim 200$ Myr. By 1 Gyr, we found that the system remains at a stable star formation rate of $\sim 40\,\mathrm{M_\odot\,yr^{-1}}$. We found that the system assembles 90% of its mass at $t_{\mathrm{assembly}} \sim 370$ Myr after the initial condition and has a SDSS rest-frame $\mathrm{g-r}$ colour of $\mathrm{g-r} \approx 0.13$. In terms of redshift, 370 Myr corresponds to $z \sim 3.3$ and implies that fully formed, highly star-forming, blue BCGs should exist at redshifts $z \gtrsim 3.3$, given that there are observations of systems similar to SPT2349-56 above $z \gtrsim 4.3$ (Ishigaki et al., 2016; Jiang et al., 2018). We demonstrated that new observational tools such as the James Webb Space Telescope will be able to easily image such systems, given that we estimated their absolute magnitudes to be $M_{\mathrm{AB,F277W}} \sim -28.7$ in the F277W NIRCam band. Of course, these estimations depend on the assumptions we made when constructing our numerical experiment. The main sources of uncertainty are the gas fractions and the applicability of the abundance matching results (Behroozi et al., 2013a) to the $z \approx 4.3$ protocluster. However, all high-redshift galaxies are found with relatively high gas fractions and while we are near the boundary of applicability in the abundance matching results, we are still within the halo mass range where the results apply.

We expect that systems such as SPT2349-56 go on to form the cores of the most-massive clusters in the universe. We used the Multi-Dark Planck 2 simulation[15] – a child of the Multi-Dark Bolshoi simulations – to investigate the occurrence of highly over-dense assembly events in the early universe for all clusters with $z = 0$ masses $M_{\mathrm{vir}} \geqslant 5 \times 10^{14}\ \mathrm{M_\odot}$. Our analysis revealed that there is a clear trend of less-massive clusters having over-dense events occur at lower redshifts (compared to more-massive clusters), indicating that there is a *downsizing* effect. Specifically, we found that the fraction of events above $z \geqslant 3$ for a given bin of final cluster mass scales strongly with the final cluster mass, $f \sim M^{3.36}$. We also determined that $\sim 10\%$ of all the clusters we investigated begin to assemble their cores rapidly at high redshift, with a higher percentage of 16.4% for clusters with final masses above $M_{\mathrm{vir}} \geqslant 10^{15}\ \mathrm{M_\odot}$. Based on these estimates, we predict that there is a population of bright, blue BCGs above $z \gtrsim 1.5$ with a comoving number density of $n \approx 48\ \mathrm{cGpc^{-3}}$. Additionally, we predict

[15] A $(1\ \mathrm{cGpc}\,h^{-1})^3$ volume with 3840^3 dark matter elements.

that there is a similar star-forming BCG population that extends down to $z \sim 1.5$ – although those at lower redshift will be up to an order-of-magnitude less massive than their high-redshift counterparts, given the downsizing trend. At redshifts lower than $z \sim 1.5$, there is insufficient gas to support high star formation rates and those BCGs that begin assembling late will assemble through dry mergers, in keeping with the conventional picture. We emphasise the distinction between core assembly and assembly of the rest of the cluster, and that our results do not suggest that the entire cluster assembles at high-redshift. We expect protocluster cores to be embedded in an extended lower density (compared to the proto-core) galaxy distribution, with these galaxies eventually forming the satellite population of the assembling cluster.

Given that extraordinary infrared observational tools such as Wide-Field Infrared Space Telescope (WFIRST) and the JWST will launch in the upcoming decade, we expect the census of interesting astronomical objects to broaden significantly. Based on our arguments in this paper, we anticipate that some of those high-redshift objects will be the cores of the most-massive clusters in the universe – the BCGs. Not only are these interesting objects in their own right but their discovery and census would also help to constrain the discrepancy in the measurements of the cosmological parameter σ_8, given their rarity and association with the most-massive clusters.

4.6 Acknowledgements

This research was enabled in part by support provided by WestGrid and Compute/Calcul Canada. Some of the computations in this work were performed on facilities supported by the Scientific Computing Core at the Flatiron Institute. The Flatiron Institute is supported by the Simons Foundation. DR acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 534263]. DR and AB acknowledge support from NSERC (Canada) through the Discovery Grant program. CB acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 504189]. MHH acknowledges the receipt of a Vanier Canada Graduate Scholarship. We thank Chervin Laporte, Carlos Frenk, Julie Hlavacek-Larrondo, Laura Salvati, Axel Weiss, Mark Brodwin, Gabriella De Lucia, Michael McDonald, Belaid Moa, and Ondrea Clarkson for insightful discussions during the course of the study. We also thank the anonymous reviewer for their helpful comments which helped us improve our study.

The CosmoSim database used in this paper is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE, www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de). The Bolshoi simulations have been performed within the Bolshoi project of the University of California High-Performance Astro-Computing Center (UC-HiPACC) and were run at the NASA Ames Research Center.

Our analysis was performed using the Python programming language (Python Software Foundation, https://www.python.org). The following packages were used throughout the analysis: `h5py` (Collette, 2013), `numpy` (Oliphant, 2006; van der Walt et al., 2011), and `matplotlib` (Hunter, 2007). Part of the analysis was performed in the IPython environment (Perez & Granger, 2007). This research additionally made use of `astropy` (http://www.astropy.org) – a community-developed core Python package for Astronomy (Astropy Collaboration, 2013, 2018).

4.7 JWST observing strategy

To generate mock NIRCam images, we select an optimal observing strategy for our synthetic SPT2349-56 system assuming 10 ks of observing time. We used the MULTIACCUM MEDIUM 8 observing read-mode using Module B, in which each 10×1ks integration is divided non-destructively into 10×100s groups with 8×10s frames/group and 2×10s drop frames/group. This MULTIACCUM "up-the-ramp" observing strategy enables cosmic ray rejection, reduces the readout noise (roughly by the square root of the total number of frames) and increases the dynamic range by preventing saturation by bright sources. We generate point spread function (PSF) images for this observing strategy in each band pass using `pynrc` and the `WebbPSF` package (Perrin et al., 2012, 2014) and convolve with the idealised image. The 1σ AB surface brightness sensitivities determined by `pynrc` for the NIRCam detector with this observing strategy were $m_{\mathrm{AB,band}} = 27.38, 27.62, 28.52, 28.65$ AB mag arcsec^{-2} for F150W, F200W, F277W, F356W. These sensitivity estimates are based on characterization data for the detectors including readout and 1/f noise, dark current, and background levels equal to 1.2 times the minimum zodiacal light background. The

current development version of `pynrc` does not allow us to convert between calibrated and non-calibrated images (i.e., AB zero points, effective gain and exposure time), so we do not incorporate Poisson shot noise from the source light. For now, we naively model the total noise contributions as a single Gaussian process with standard deviation equal to the reported sensitivity in each band pass using our observing strategy.

Chapter 5

Conclusions

The brightest cluster galaxies (BCGs) are the most massive galaxies in the Universe and reside in the cores of massive galaxy clusters. Once thought to be the result of galaxies merging in the latter half of the history of the Universe, recent observations hint that some of these monstrous objects could have formed very early from highly overdense protoclusters – only a few billion years after the Big Bang. Understanding how that process occurs requires cosmological-scale hydrodynamics+gravity simulations that take into account a plethora of astrophysical processes. The main goals of my dissertation were twofold: (a) demonstrate that there are significant improvements to our current numerical methods and models that impact galaxy evolution and (b) that it is possible for the most massive galaxies observed to form rapidly in the early Universe.

5.1 Numerical methods

Starting with the fundamental hydrodynamics, I demonstrated in Chapter 2 that contemporary Lagrangian simulations lack the ability to properly simulate turbulence. Turbulence occurs when kinetic energy is injected into a fluid with small viscosity. For example, active galactic nuclei (AGN) feedback, stellar feedback, and galaxy merging all are astrophysical processes that inject kinetic energy into the gas that permeates the BCGs. If the kinetic energy is unable to thermalize on the scale of energy injection, it begins to break up the fluid into smaller and smaller eddies, until smallest scale is reached – the dissipation scale. The problem occurs when we try to simulate such a process: the resolution of all simulations are almost always above the dissipation scale. That leads to a build-up of kinetic energy just above the resolution scale, causing artificial kinetic energy in the simulated fluid. Additionally, Lagrangian

methods that do not allow mass transfer between particles suffer a similar problem. If mass cannot transfer between the gas particles, it is impossible for metals to mix due to turbulence on scales lower than the resolution.

The solution to both problems is to introduce additional mixing terms into the hydrodynamical equations, thermalizing the artificial kinetic energy and transporting metals between gas particles. These are sub-grid eddy mixing models, with the most popular being the constant-coefficient Smagorinsky model. I showed in Chapter 2, Section 2.1 that the constant-coefficient Smagorinsky model may not be the best option as it is over-diffusive in many cases, leading to too much dissipation and metal mixing. For the first time in Lagrangian hydrodynamics, I used a solution from Piomelli & Liu (1995), the dynamic Smagorinsky model, that is able to adjust the model parameter at simulation time to ensure that over-mixing does not occur. I incorporated the model into the hydrodynamical+gravity simulation code `GIZMO`. The dynamic model predicted much lower mixing on average, and led to changes in the metal enrichment histories of the gas in galactic environments, such as the circumgalactic medium (CGM).

I was not completely satisfied with the Smagorinsky model, whether dynamic or not, since (a) it assumes that the kinetic energy and metals mix isotropically on the resolution scale and (b) it ignores contributes in mixing due to compression of the gas. I showed a derivation following Clark et al. (1979) of why this should not be true in Chapter 2, Section 2.2. The movement of material on the smallest scales in simulation should actually be *anisotropic* and move in the direction of the velocity gradient tensor, and include contributions from compression in the flow. I incorporated this *gradient model* into `GIZMO` and added a dynamic procedure to calculate its model parameters. My results were surprising in that I found all of the mixing models provided a much more accurate solution compared to having no mixing in my simple hydrodynamical convergence tests, with none really being better than the other. However, when I applied the models to cosmological simulations I found that the gradient model predicted much tighter metal distribution functions in the gas of high-redshift ($z \gtrsim 3$) galaxies compared to all other models – exactly where BCGs begin forming! That exciting result demonstrates that we must incorporate sub-grid turbulent metal mixing models in order to trust our observations of metallicity in the early Universe. I additionally found that for the gradient model it did not matter whether a dynamic procedure was used or not. For that reason I recommend against using the dynamic procedure with the gradient model as it is computationally

expensive.

Not only is turbulence important for understanding our Universe, but those processes that drive turbulence – such as active galactic nuclei (AGN) feedback. AGN feedback receives considerable more attention in the literature and there are a plethora of models that mimic the astrophysics occurring below the resolution scale of our simulations. The goal Chapter 3 was to combine all of the most physically motivated sub-grid models available in order to see how they interact, and whether there is viability in combining them for future BCG studies.

I categorized the models into four aspects: seeding, accretion and growth, feedback, and dynamics.

For seeding, the majority of contemporary cosmological simulations use a method that artificially places supermassive black holes (SMBHs) into the centers of dark matter halos or galaxies that are forming in the simulation, in order to guarantee that each galaxy has a SMBH. I found that the most physically motivated alternative model is one that determines where SMBHs should form based on the local gas properties (Tremmel et al., 2017). I showed in Figure 3.2 that the new model drastically alters the formation history of SMBHs in the Universe by forming most of them in the early Universe, before $z \sim 4$. That has important implications for the formation of massive galaxies from protoclusters, as that is the exact time range when we expect the BCGs to begin forming.

For accretion and growth, I combined Bondi accretion and a torque accretion rate estimator method following Davé et al. (2019)[1]. Surprisingly, one of my main results is that kinetic energy constrained kinetic wind models for AGN feedback are not viable at the current resolutions of cosmological-scale simulations. The problem is that too much kinetic energy is injected in each event due to the large mass of the low-resolution gas particles. To fix the issue, I now constrain the momentum flux of the kinetic wind to $20\,L_{\rm BH}/c$ (Faucher-Giguère & Quataert, 2012; Davé et al., 2019). Another important result is that the torque accretion model does not seem viable when combining with the dynamical friction model I discussed in Section 3.1.4. For that reason, I recommend to use the Bondi model only when using the dynamical friction model until a solution can be found.

I introduced an entirely new feedback model in Chapter 3, Section 3.3 that involves three separate states that have three different feedback prescriptions. Each state covers a range of accretion rates: $\dot{M}_{\rm BH} \leqslant 0.03$, $0.03 < \dot{M}_{\rm BH} \leqslant 0.3$, and $0.3 < \dot{M}_{\rm BH}$. In

[1]Both models were already available in `GIZMO`.

the lowest accretion rate regime, I also incorporated a model for kinetic jet feedback based on a probabilistic model taking into account the mass budget of the system. My results in Chapter 3 are promising for having a well-calibrated model, the observed $M_{\rm BH}-M_*$ relationship holds and the galaxy stellar mass function matches observations well.

The last aspect that I changed was that for the dynamics of SMBHs. In the vast majority of cosmological simulations SMBHs are artificially pinned to the centers of their dark matter halos (or galaxies) at each timestep to ensure they do not fly away during a merger process. In some instances, they are simply pushed toward the center rather than pinned directly. I incorporated an alternative into `GIZMO`: the dynamical friction model of Tremmel et al. (2015) (following Chandrasekhar 1943). That model allows SMBHs to roam free in the simulations from their birthplaces, and accurately captures their in-fall into the centers of their hosts, as well as the dynamics during a merger process. I demonstrated in Figure 3.7 that the model reproduces expected in-fall rates of SMBHs in `GIZMO`, and is a viable option for future use.

5.2 Rapid massive galaxy evolution

The discovery of highly-overdense protoclusters with $\gtrsim 6$ or more highly starforming ($\sim 10^4$ $\mathrm{M_\odot\,yr^{-1}}$) galaxies at high redshift ($z \gtrsim 4$) introduced a shock in the extra-galactic community. How could such objects exist with so many galaxies? With such star formation rates? Why did our simulations not predict these objects? These are the questions that still plague us a few years after the discovery of the spectacular systems. I was interested in the fact that these objects are sufficiently dense that the galaxies within them should collapse rapidly into a massive galaxy. I asked: how massive would the object become, and over what timescale would the process occur?

To answer these questions, I designed custom initial conditions based on the observations of the protocluster SPT2349-56 (Miller et al., 2018) and used the `GIZMO` code and `MUFASA` galaxy evolution model (Davé et al., 2016, 2017) to study the forward-evolution of the system. In Chapter 4, I showed that the system collapses on a timescale of approximately $0.5-1$ Gyr, and reaches a maximum star formation rate of ~ 3000 $\mathrm{M_\odot\,yr^{-1}}$. The final stellar mass of the system was $M_* \sim 10^{12}$ and it visually resembled a BCG. Our work predicted that a BCG can form within ~ 1 Gyr in the early Universe from about $z \approx 4.4$ to $z \approx 2.7$, significantly earlier than previously thought. We predicted that the James Webb Space telescope would be able to resolve

these BCGs, possibly leading to an avenue to confirm our theory.

While one object is interesting, I asked myself what are the chances of finding more of these highly-overdense protoclusters? By investigating a large-volume dark-matter-only simulation, MultiDark Planck 2 (Riebe et al., 2013; Klypin et al., 2016), I was able to identify dark matter substructure at high redshift that undergoes similar, albeit not as intense, collapse as SPT2349-56. I found that those overdense collapse events were more likely to happen in the evolutionary history of the most massive clusters at $z = 0$. Specifically, I found that the probability of finding highly-overdense collapse at $z \gtrsim 3$ given a cluster mass $M_{\rm vir}|_{z=0}$ scales as $\sim M_{\rm vir}^{3.36}$. This is *downsizing* on the galaxy cluster scale, where the cores of the most massive galaxy clusters form first in the early Universe – producing massive BCGs earlier than expected. This result aligns with subsequent observations of fully mature galaxy clusters at high-redshift (Willis et al., 2020) and the wealth of high-redshift protoclusters discovered (Ishigaki et al., 2016; Jiang et al., 2018; Higuchi et al., 2019; Ito et al., 2019; Tilvi et al., 2020).

5.3 Summary

The field of galaxy evolution continues to evolve rapidly. At the beginning of my PhD studies I would not have expected that massive galaxies could have formed so early and with such star formation rates as we demonstrated in Chapter 4. The disagreement with previous numerical work on BCGs that I outlined in Section 4.1 shows that while numerical simulations are an important tool in our process of understanding galaxy evolution, they must be well-calibrated and converged. That is my goal with Chapters 2 & 3: I showed that we have a long way to go before we can completely trust our current models of the Universe, but there is a path.

Appendix A

Additional Information

The analysis code for the material I discuss in this dissertation are available upon request. The code for Section 2.1 already publicly available in `GIZMO` and I am in the process of preparing submission of the code for Section 2.2 to the public branch. The models in Chapter 3 may not be incorporated into the public version of `GIZMO`, but are available upon request.

I have tried to be as explicit and verbose as possible in my derivations and descriptions, I am sure much to the displeasure of a subset of my readers and reviewers.

Bibliography

Ade P. A. R., et al., 2014, Astronomy & Astrophysics, 571, A20

Ade P. A. R., et al., 2016, Astronomy & Astrophysics, 594, A13

Alexander D. M., Smail I., Bauer F. E., Chapman S. C., Blain A. W., Brandt W. N., Ivison R. J., 2005, Nature, 434, 738

Anders E., Grevesse N., 1989, Geochimica et Cosmochimica Acta, 53, 197

Anderson M. E., Bregman J. N., 2010, The Astrophysical Journal, 714, 320

Anglés-Alcázar D., Özel F., Davé R., 2013, Astrophysical Journal, 770

Anglés-Alcázar D., Özel F., Davé R., Katz N., Kollmeier J. A., Oppenheimer B. D., 2015, The Astrophysical Journal, 800, 127

Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017a, Monthly Notices of the Royal Astronomical Society: Letters, 472, L109

Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017b, Monthly Notices of the Royal Astronomical Society, 2853, 13

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, Monthly Notices of the Royal Astronomical Society, 426, 2046

Aragon-Salamanca A., Baugh C. M., Kauffmann G., 1998, Monthly Notices of the Royal Astronomical Society, 297, 427

Ardaneh K., Luo Y., Shlosman I., Nagamine K., Wise J. H., Begelman M. C., 2018, Monthly Notices of the Royal Astronomical Society, 479, 2277

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, Annual Review of Astronomy and Astrophysics, 47, 481

Astropy Collaboration 2013, Astronomy & Astrophysics, 558, A33

Astropy Collaboration 2018, The Astronomical Journal, 156, 123

Babul A., Balogh M. L., Lewis G. F., Poole G. B., 2002, Monthly Notices of the Royal Astronomical Society, 330, 329

Baes M., Camps P., 2015, Astronomy and Computing, 12, 33

Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, The Astrophysical Journal Supplement Series, 196, 22

Balarac G., Le Sommer J., Meunier X., Vollant A., 2013, Physics of Fluids, 25, 075107

Baldry I. K., et al., 2012, Monthly Notices of the Royal Astronomical Society, 421, 621

Balogh M. L., Babul A., Patton D. R., 1999, Monthly Notices of the Royal Astronomical Society, 307, 463

Barai P., Murante G., Borgani S., Gaspari M., Granato G. L., Monaco P., Ragone-Figueroa C., 2016, Monthly Notices of the Royal Astronomical Society, 461, 1548

Barnes D. J., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 1088

Bauer A., Springel V., 2012, Monthly Notices of the Royal Astronomical Society, 423, 2558

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, The Astrophysical Journal, 762, 109

Behroozi P. S., Wechsler R. H., Conroy C., 2013b, The Astrophysical Journal, 770, 57

Benhamadouche S., Arenas M., Malouf W., 2017, Nuclear Engineering and Design, 312, 128

Bennett J. S., Sijacki D., 2020, Monthly Notices of the Royal Astronomical Society, 499, 597

Benson A. J., 2010, Physics Reports, 495, 33

Benson A. J., Babul A., 2009, Monthly Notices of the Royal Astronomical Society, 397, 1302

Bentz M. C., Manne-Nicholas E., 2018, The Astrophysical Journal, 864, 146

Bertschinger E., 2011, GRAFIC-2: Multiscale Gaussian Random Fields for Cosmological Simulations, Astrophysics Source Code Library

Biernacki P., Teyssier R., Bleuler A., 2017, Monthly Notices of the Royal Astronomical Society, 469, 295

Bildfell C., Hoekstra H., Babul A., Mahdavi A., 2008, Monthly Notices of the Royal Astronomical Society, 389, 1637

Binney J., Tremaine S., 2011, Galactic Dynamics, 2nd editio edn. Princeton University Press, Princeton, New Jersey, USA

Bonaventura N. R., et al., 2017, Monthly Notices of the Royal Astronomical Society, 469, 1259

Bondi H., 1952, Monthly Notices of the Royal Astronomical Society, 112, 195

Bondi H., Hoyle F., 1944, Monthly Notices of the Royal Astronomical Society, 104, 273

Booth C. M., Schaye J., 2009, Monthly Notices of the Royal Astronomical Society, 398, 53

Bottrell C., Torrey P., Simard L., Ellison S. L., 2017a, Monthly Notices of the Royal Astronomical Society, 49, 1033

Bottrell C., Torrey P., Simard L., Ellison S. L., 2017b, Monthly Notices of the Royal Astronomical Society, 467, 2879

Bourne M. A., Sijacki D., 2017, Monthly Notices of the Royal Astronomical Society, 472, 4707

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, Monthly Notices of the Royal Astronomical Society, 370, 645

Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, Monthly Notices of the Royal Astronomical Society, 383, 93

Boylan-Kolchin M., Springel V., White S. D., Jenkins A., Lemson G., 2009, Monthly Notices of the Royal Astronomical Society, 398, 1150

Brook C. B., Stinson G., Gibson B. K., Shen S., Macciò A. V., Obreja A., Wadsley J., Quinn T., 2014, Monthly Notices of the Royal Astronomical Society, 443, 3809

Brough S., Couch W. J., Collins C. A., Jarrett T., Burke D. J., Mann R. G., 2008, Monthly Notices of the Royal Astronomical Society: Letters, 385, L103

Bruzual G., Charlot S., 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000

Bryan G. L., et al., 2014, The Astrophysical Journal Supplement Series, 211, 19

Buie E., Fumagalli M., Scannapieco E., 2020, The Astrophysical Journal, 890, 33

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, Monthly Notices of the Royal Astronomical Society, 321, 559

Burke C., Hilton M., Collins C., 2015, Monthly Notices of the Royal Astronomical Society, 449, 2353

Cadiou C., Dubois Y., Pichon C., 2019, Astronomy & Astrophysics, 621, A96

Calder A., et al., 2004, Computing in Science and Engineering, 6, 10

Camps P., Baes M., 2015, Astronomy and Computing, 9, 20

Carilli C. L., et al., 2010, The Astrophysical Journal, 714, 1407

Cerulo P., Orellana G. A., Covone G., 2019, Monthly Notices of the Royal Astronomical Society

Chabrier G., 2003, Publications of the Astronomical Society of the Pacific, 115, 763

Chandrasekhar S., 1943, The Astrophysical Journal, 97, 255

Chandrasekhar S., 1961, Hydrodynamic and hydromagnetic stability. Oxford University Press, London and New York

Choi E., Ostriker J. P., Naab T., Johansson P. H., 2012, The Astrophysical Journal, 754, 125

Choi E., Ostriker J. P., Naab T., Somerville R. S., Hirschmann M., Núñez A., Hu C.-Y., Oser L., 2017, The Astrophysical Journal, 844, 31

Christensen C. R., Davé R., Governato F., Pontzen A., Brooks A., Munshi F., Quinn T., Wadsley J., 2016, The Astrophysical Journal, 824, 57

Cielo S., Babul A., Antonuccio-Delogu V., Silk J., Volonteri M., 2018, Astronomy & Astrophysics, 617, A58

Cimatti A., Daddi E., Renzini A., 2006, Astronomy & Astrophysics, 453, L29

Ciotti L., Ostriker J. P., 1997, The Astrophysical Journal, 487, L105

Clark R. A., Ferziger J. H., Reynolds W. C., 1979, Journal of Fluid Mechanics, 91, 1

Clarkson O., Herwig F., 2020, Monthly Notices of the Royal Astronomical Society, 500, 2685

Colbrook M. J., Ma X., Hopkins P. F., Squire J., 2017, Monthly Notices of the Royal Astronomical Society, 467, 2421

Collette A., 2013, Python and HDF5. O'Reilly

Collins C. A., et al., 2009, Nature, 458, 603

Conroy C., Wechsler R. H., Kravtsov A. V., 2007, The Astrophysical Journal, 668, 826

Contini E., De Lucia G., Villalobos Á., Borgani S., 2014, Monthly Notices of the Royal Astronomical Society, 437, 3787

Cowie L. L., Binney J., 1977, The Astrophysical Journal, 215, 723

Crain R. A., McCarthy I. G., Schaye J., Theuns T., Frenk C. S., 2013, Monthly Notices of the Royal Astronomical Society, 432, 3005

Daddi E., et al., 2005, The Astrophysical Journal, 626, 680

Daddi E., et al., 2010, The Astrophysical Journal, 713, 686

Davé R., Finlator K., Oppenheimer B. D., 2006, Monthly Notices of the Royal Astronomical Society, 370, 273

Davé R., Oppenheimer B. D., Katz N., Kollmeier J. A., Weinberg D. H., 2010, Monthly Notices of the Royal Astronomical Society, 408, 2051

Davé R., Thompson R., Hopkins P. F., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3265

Davé R., Rafieferantsoa M. H., Thompson R. J., Hopkins P. F., 2017, Monthly Notices of the Royal Astronomical Society, 467, stx108

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, Monthly Notices of the Royal Astronomical Society, 486, 2827

De Lucia G., Blaizot J., 2007, Monthly Notices of the Royal Astronomical Society, 375, 2

De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, Monthly Notices of the Royal Astronomical Society, 366, 499

Dehnen W., Aly H., 2012, Monthly Notices of the Royal Astronomical Society, 425, 1068

Dekel A., Sari R., Ceverino D., 2009, The Astrophysical Journal, 703, 785

Di Mascio A., Antuono M., Colagrossi A., Marrone S., 2017, Physics of Fluids, 29, 035102

Diemer B., Kravtsov A. V., 2015, The Astrophysical Journal, 799, 108

Donzelli C. J., Muriel H., Madrid J. P., 2011, The Astrophysical Journal Supplement Series, 195, 15

Douspis M., Salvati L., Aghanim N., 2018, in 2nd World Summit: Exploring the Dark Side of the Universe. 25-29 June. p. 37 (arXiv:1901.05289), http://arxiv.org/abs/1901.05289

Dubinski J., 1998, The Astrophysical Journal, 502, 141

Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, Monthly Notices of the Royal Astronomical Society, 420, 2662

EHT-Collaboration 2019, The Astrophysical Journal, 875, L6

Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, The Astrophysical Journal, 136, 748

Elmegreen B. G., Scalo J., 2004, Annu. Rev. Astron. Astrophys, 42, 211

Emerick A., Bryan G. L., Low M.-M. M., Côté B., Johnston K. V., O'Shea B. W., 2018, The Astrophysical Journal, 869, 94

Escala I., et al., 2018, Monthly Notices of the Royal Astronomical Society, 474, 2194

Evoli C., Ferrara A., 2011, Monthly Notices of the Royal Astronomical Society, 413, 2721

Fabian A., 2012, Annual Review of Astronomy and Astrophysics, 50, 455

Fabian A. C., Nulsen P. E. J., 1977, Monthly Notices of the Royal Astronomical Society, 180, 479

Faucher-Giguère C. A., Quataert E., 2012, Monthly Notices of the Royal Astronomical Society, 425, 605

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, The Astrophysical Journal, 703, 1416

Federrath C., 2013, Monthly Notices of the Royal Astronomical Society, 436, 1245

Fielding D., Quataert E., McCourt M., Thompson T. A., 2017, Monthly Notices of the Royal Astronomical Society, 466, 3810

Fielding D., Quataert E., Martizzi D., 2018, Monthly Notices of the Royal Astronomical Society, 481, 3325

Finlator K., Davé R., 2008, Monthly Notices of the Royal Astronomical Society, 385, 2181

Fiore F., et al., 2017, Astronomy & Astrophysics, 601, A143

Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, Monthly Notices of the Royal Astronomical Society, 397, 1776

Frontiere N., Raskin C. D., Owen J. M., 2017, Journal of Computational Physics, 332, 160

Fryer C. L., Woosley S. E., Heger A., 2001, The Astrophysical Journal, 550, 372

Gaburov E., Nitadori K., 2011, Monthly Notices of the Royal Astronomical Society, 414, 129

Garnier E., Adams N., Sagaut P., 2009, Large Eddy Simulation for Compressible Flows. Scientific Computation, Springer Netherlands, Dordrecht, doi:10.1007/978-90-481-2819-8

Gaspari M., Temi P., Brighenti F., 2017, Monthly Notices of the Royal Astronomical Society, 466, 677

Gebhardt K., et al., 2000, The Astrophysical Journal, 543, L5

Genel S., 2016, The Astrophysical Journal, 822, 107

Genel S., et al., 2014, Monthly Notices of the Royal Astronomical Society, 445, 175

Germano M., Piomelli U., Moin P., Cabot W. H., 1991, Physics of Fluids A: Fluid Dynamics, 3, 1760

Gingold R. A., Monaghan J. J., 1977, Monthly Notices of the Royal Astronomical Society, 181, 375

Gingold R. A., Monaghan J. J., 1982, Journal of Computational Physics, 46, 429

Granato G. L., Ragone-Figueroa C., Domínguez-Tenreiro R., Obreja A., Borgani S., De Lucia G., Murante G., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1320

Grand R. J. J., et al., 2017, Monthly Notices of the Royal Astronomical Society, 207, 179

Greif T. H., Glover S. C. O., Bromm V., Klessen R. S., 2009, Monthly Notices of the Royal Astronomical Society, 392, 1381

Grete P., O'Shea B. W., Beckwith K., 2018, The Astrophysical Journal, 858, L19

Groves B., Dopita M. A., Sutherland R. S., Kewley L. J., Fischera J., Leitherer C., Brandl B., van Breugel W., 2008, The Astrophysical Journal Supplement Series, 176, 438

Guedes J., Callegari S., Madau P., Mayer L., 2011, The Astrophysical Journal, 742, 76

Gültekin K., et al., 2009, The Astrophysical Journal, 698, 198

Hafen Z., et al., 2019, Monthly Notices of the Royal Astronomical Society, 488, 1248

Hafen Z., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 3581

Hahn O., Abel T., 2011, Monthly Notices of the Royal Astronomical Society, 415, 2101

Hahn O., Teyssier R., Carollo C. M., 2010, Monthly Notices of the Royal Astronomical Society, 405, no

Hani M. H., Sparre M., Ellison S. L., Torrey P., Vogelsberger M., 2018, Monthly Notices of the Royal Astronomical Society, 475, 1160

Harris C. R., et al., 2020, Nature, 585, 357

Hernquist L., 1990, The Astrophysical Journal, 356, 359

Hernquist L., 1993, The Astrophysical Journal Supplement Series, 86, 389

Hernquist L., Katz N., 1989, The Astrophysical Journal Supplement Series, 70, 419

Higuchi R., et al., 2019, The Astrophysical Journal, 879, 28

Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, Monthly Notices of the Royal Astronomical Society, 442, 2304

Hopkins P. F., 2013, Monthly Notices of the Royal Astronomical Society, 428, 2840

Hopkins P. F., 2015, Monthly Notices of the Royal Astronomical Society, 450, 53

Hopkins P. F., 2017, Monthly Notices of the Royal Astronomical Society, 466, 3387

Hopkins P. F., Quataert E., 2010, Monthly Notices of the Royal Astronomical Society, 407, 1529

Hopkins P. F., Quataert E., 2011, Monthly Notices of the Royal Astronomical Society, 415, 1027

Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009, The Astrophysical Journal, 691, 1168

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, Monthly Notices of the Royal Astronomical Society, 445, 581

Hopkins P. F., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 800

Hosono N., Saitoh T. R., Makino J., 2016, The Astrophysical Journal Supplement Series, 224, 32

Hu C.-Y., Chiang C.-T., 2020, The Astrophysical Journal, 900, 29

Huang S., Katz N., Scannapieco E., Cottle J., Davé R., Weinberg D. H., Peeples M. S., Brüggen M., 2020, Monthly Notices of the Royal Astronomical Society, 497, 2586

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Iapichino L., Schmidt W., Niemeyer J. C., Merklein J., 2011, Monthly Notices of the Royal Astronomical Society, 414, 2297

Iapichino L., Viel M., Borgani S., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2529

Ishigaki M., Ouchi M., Harikane Y., 2016, The Astrophysical Journal, 822, 5

Ito K., et al., 2019, The Astrophysical Journal, 878, 68

Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F., 1999, The Astrophysical Journal Supplement Series, 125, 439

Jiang L., et al., 2018, Nature Astronomy, 2, 962

Kara R., ÇaÄ§lar M., 2018, Applied Mathematics and Computation, 322, 89

Karen Yang H.-Y., Reynolds C. S., 2016, The Astrophysical Journal, 829, 90

Katz N., 1992, The Astrophysical Journal, 391, 502

Kennicutt R. C., 1998, The Astrophysical Journal, 498, 541

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, Monthly Notices of the Royal Astronomical Society, 363, 2

Khani S., Waite M. L., 2015, Journal of Fluid Mechanics, 773, 327

Khatri R., Gaspari M., 2016, Monthly Notices of the Royal Astronomical Society, 463, 655

King A., Pounds K., 2015, Annual Review of Astronomy and Astrophysics, 53, 115

Kirkpatrick M. P., Ackerman A. S., Stevens D. E., Mansour N. N., 2006, Journal of the Atmospheric Sciences, 63, 526

Kleissl J., Kumar V., Meneveau C., Parlange M. B., 2006, Water Resources Research, 42, 1

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, The Astrophysical Journal, 740, 102

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, Monthly Notices of the Royal Astronomical Society, 457, 4340

Kormendy J., Ho L. C., 2013, Annual Review of Astronomy and Astrophysics, 51, 511

Kravtsov A. V., Borgani S., 2012, Annual Review of Astronomy and Astrophysics, 50, 353

Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astronomy Letters, 44, 8

Kritsuk A. G., Norman M. L., Padoan P., Wagner R., 2007, The Astrophysical Journal, 665, 416

Krumholz M. R., McKee C. F., Tumlinson J., 2009, The Astrophysical Journal, 699, 850

Landau L., Lifshitz E., 1987, Course of theoretical physics. Vol. 6: Fluid Mechanics. Pergamon Press

Lanson N., Vila J.-P., 2008a, SIAM Journal on Numerical Analysis, 46, 1912

Lanson N., Vila J.-P., 2008b, SIAM Journal on Numerical Analysis, 46, 1935

Laporte C. F. P., White S. D. M., 2015, Monthly Notices of the Royal Astronomical Society, 451, 1177

Laporte C. F. P., White S. D. M., Naab T., Gao L., 2013, Monthly Notices of the Royal Astronomical Society, 435, 901

Lauer T. R., Postman M., Strauss M. A., Graves G. J., Chisari N. E., 2014, The Astrophysical Journal, 797, 82

Lavoie S., et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 4141

Lecoanet D., et al., 2016, Monthly Notices of the Royal Astronomical Society, 455, 4274

Lee C. Y., Cant S., 2017, Flow, Turbulence and Combustion, 98, 155

Leitherer C., et al., 1999, The Astrophysical Journal Supplement Series, 123, 3

Li M., Li Y., Bryan G. L., Ostriker E. C., Quataert E., 2020, The Astrophysical Journal, 898, 23

Liang L., Durier F., Babul A., Davé R., Oppenheimer B. D., Katz N., Fardal M., Quinn T., 2016, Monthly Notices of the Royal Astronomical Society, 456, 4266

Lidman C., et al., 2012, Monthly Notices of the Royal Astronomical Society, 427, 550

Lidman C., et al., 2013, Monthly Notices of the Royal Astronomical Society, 433, 825

Lin Y., Mohr J. J., 2004, The Astrophysical Journal, 617, 879

Lochhaas C., Bryan G. L., Li Y., Li M., Fielding D., 2020, Monthly Notices of the Royal Astronomical Society, 493, 1461

Loubser S. I., Hoekstra H., Babul A., O'Sullivan E., 2018, Monthly Notices of the Royal Astronomical Society, 477, 335

Lucy L. B., 1977, The Astronomical Journal, 82, 1013

Lupi A., Haardt F., Dotti M., Fiacconi D., Mayer L., Madau P., 2016, Monthly Notices of the Royal Astronomical Society, 456, 2993

Lynden-Bell D., 1969, Nature, 223, 690

Madau P., Rees M. J., 2001, The Astrophysical Journal, 551, L27

Madau P., Haardt F., Dotti M., 2014, The Astrophysical Journal, 784, L38

Magorrian J., et al., 1998, The Astronomical Journal, 115, 2285

Martizzi D., Hahn O., Wu H.-Y., Evrard A. E., Teyssier R., Wechsler R. H., 2016, Monthly Notices of the Royal Astronomical Society, 459, 4408

McCarthy I. G., Babul A., Bower R. G., Balogh M. L., 2008, Monthly Notices of the Royal Astronomical Society, 386, 1309

McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, Monthly Notices of the Royal Astronomical Society, 465, 2936

McConnell N. J., Ma C.-P., 2013, The Astrophysical Journal, 764, 184

McDonald M., et al., 2016, The Astrophysical Journal, 817, 86

McNamara B., Nulsen P., 2007, Annual Review of Astronomy and Astrophysics, 45, 117

Mcconnell N. J., Ma C. P., Gebhardt K., Wright S. A., Murphy J. D., Lauer T. R., Graham J. R., Richstone D. O., 2011, Nature, 480, 215

Meneveau C., Katz J., 2000, Annual Review of Fluid Mechanics, 32, 1

Miller T. B., et al., 2018, Nature, 556, 469

Milosavljevic M., Merritt D., 2003, in AIP Conference Proceedings. American Institute of Physics, pp 201–210

Mitra S., Davé R., Finlator K., 2015, Monthly Notices of the Royal Astronomical Society, 452, 1184

Mo H. J., Mao S., White S. D. M., 1998, Monthly Notices of the Royal Astronomical Society, 295, 319

Mo H., van den Bosch F., White S., 2010, Galaxy Formation and Evolution. Cambridge University Press, Cambridge, doi:10.1017/CBO9780511807244, http://ebooks.cambridge.org/ref/id/CBO9780511807244

Monaghan J. J., 1989, Journal of Computational Physics, 82, 1

Monaghan J., 2000, Journal of Computational Physics, 159, 290

Monaghan J. J., 2002, Monthly Notices of the Royal Astronomical Society, 335, 843

Monaghan J. J., 2005, Reports on progress in physics, 68, 1703

Monaghan J. J., 2011, European Journal of Mechanics, B/Fluids, 30, 360

Morganti R., 2017, Frontiers in Astronomy and Space Sciences, 4, 1

Morris J. P., 1996, Publications of the Astronomical Society of Australia, 13, 97

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, Monthly Notices of the Royal Astronomical Society, 454, 2691

Murray N., Quataert E., Thompson T. A., 2005, The Astrophysical Journal, 618, 569

Muzzin A., et al., 2009, The Astrophysical Journal, 698, 1934

Naab T., Ostriker J. P., 2017, Annual Review of Astronomy and Astrophysics, 55, 59

Narayanan D., Bothwell M., Davé R., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1178

Nardini E., Risaliti G., Salvati M., Sani E., Imanishi M., Marconi A., Maiolino R., 2008, Monthly Notices of the Royal Astronomical Society: Letters, 385, L130

Navarro J. F., Frenk C. S., White S. D. M., 1997, The Astrophysical Journal, 490, 493

Neistein E., van den Bosch F. C., Dekel A., 2006, Monthly Notices of the Royal Astronomical Society, 372, 933

Nemmen R. S., Bower R. G., Babul A., Storchi-Bergmann T., 2007, Monthly Notices of the Royal Astronomical Society, 377, 1652

Netzer H., 2015, Annual Review of Astronomy and Astrophysics, 53, 365

Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nuclear Physics A, 777, 424

Nomura K. K., Post G. K., 1998, Journal of Fluid Mechanics, 377, 65

O'Sullivan E., et al., 2017, Monthly Notices of the Royal Astronomical Society, 472, 1482

Oliphant T. E., 2006, A guide to NumPy. Trelgol Publishing USA

Olsen K., Greve T. R., Narayanan D., Thompson R., Davé R., Rios L. N., Stawinski S., 2017, The Astrophysical Journal, 846, 105

Oppenheimer B. D., Davé R., 2006, Monthly Notices of the Royal Astronomical Society, 373, 1265

Oppenheimer B. D., Davé R., 2008, Monthly Notices of the Royal Astronomical Society, 387, 577

Oppenheimer B. D., Schaye J., 2013, Monthly Notices of the Royal Astronomical Society, 434, 1043

Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, Monthly Notices of the Royal Astronomical Society, 406, 2325

Oppenheimer B. D., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 2157

Oppenheimer B. D., Babul A., Bahé Y., Butsky I. S., McCarthy I. G., 2021, Universe, 7, 209

Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, The Astrophysical Journal, 725, 2312

O'shea B. W., Bryan G., Bordner J., Norman M. L., Abel T., Harkness R., Kritsuk A., 2004, in , Adaptive Mesh Refinement - Theory and Applications. Springer-Verlag, Berlin/Heidelberg, pp 341–349 (arXiv:0403044), doi:10.1007/3-540-27039-6_24, http://link.springer.com/10.1007/3-540-27039-6{_}24http://arxiv.org/abs/astro-ph/0403044

Ostriker J. P., Tremaine S. D., 1975, The Astrophysical Journal, 202, L113

Overzier R. A., 2016, The Astronomy and Astrophysics Review, 24, 14

Pan L., Scannapieco E., Scalo J., 2013, The Astrophysical Journal, 775, 111

Pasini T., et al., 2020, Monthly Notices of the Royal Astronomical Society, 497, 2163

Peebles P. J. E., 1968, The Astrophysical Journal, 153, 13

Perez F., Granger B. E., 2007, Computing in Science & Engineering, 9, 21

Perna M., Lanzuisi G., Brusa M., Cresci G., Mignoli M., 2017, Astronomy & Astrophysics, 606, A96

Perret V., Renaud F., Epinat B., Amram P., Bournaud F., Contini T., Teyssier R., Lambert J.-C., 2014, Astronomy & Astrophysics, 562, A1

Perrin M. D., Soummer R., Elliott E. M., Lallo M. D., Sivaramakrishnan A., 2012, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, 8442, 84423D

Perrin M. D., Sivaramakrishnan A., Lajoie C.-P., Elliott E., Pueyo L., Ravindranath S., Albert L., 2014, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, 9143, 91433X

Phillips G. J., Monaghan J. J., 1985, Monthly Notices of the Royal Astronomical Society, 216, 883

Pillepich A., et al., 2018, Monthly Notices of the Royal Astronomical Society, 475, 648

Pinto C., et al., 2015, Astronomy & Astrophysics, 575, A38

Piomelli U., Liu J., 1995, Physics of Fluids, 7, 839

Piomelli U., Cabot W. H., Moin P., Lee S., 1991, Physics of Fluids A: Fluid Dynamics, 3, 1766

Pipino A., Szabo T., Pierpaoli E., MacKenzie S. M., Dong F., 2011, Monthly Notices of the Royal Astronomical Society, 417, 2817

Planck Collaboration XIII 2015, Astronomy & Astrophysics, 594, A13

Planck Collaboration XVI 2014, Astronomy & Astrophysics, 571, A16

Poole G. B., Fardal M. A., Babul A., McCarthy I. G., Quinn T., Wadsley J., 2006, Monthly Notices of the Royal Astronomical Society, 373, 881

Pope S. B., 2000, Turbulent Flows, doi:10.1017/CBO9780511840531

Postman M., et al., 2012, The Astrophysical Journal Supplement Series, 199, 25

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, Monthly Notices of the Royal Astronomical Society, 338, 14

Prasad D., Sharma P., Babul A., 2015, The Astrophysical Journal, 811, 108

Prasad D., Sharma P., Babul A., 2018, The Astrophysical Journal, 863, 62

Price D. J., 2012a, Journal of Computational Physics, 231, 759

Price D. J., 2012b, Monthly Notices of the Royal Astronomical Society: Letters, 420, 33

Price D. J., Federrath C., 2010, Monthly Notices of the Royal Astronomical Society, 406, 1659

Ragone-Figueroa C., Granato G. L., Ferraro M. E., Murante G., Biffi V., Borgani S., Planelles S., Rasia E., 2018, Monthly Notices of the Royal Astronomical Society, 479, 1125

Rennehan D., Babul A., Hopkins P. F., Davé R., Moa B., 2019, Monthly Notices of the Royal Astronomical Society, 483, 3810

Rennehan D., Babul A., Hayward C. C., Bottrell C., Hani M. H., Chapman S. C., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4607

Reynolds C. S., McKernan B., Fabian A. C., Stone J. M., Vernaleo J. C., 2005, Monthly Notices of the Royal Astronomical Society, 357, 242

Riebe K., et al., 2013, Astronomische Nachrichten, 334, 691

Robertson B., Cox T. J., Hernquist L., Franx M., Hopkins P. F., Martini P., Springel V., 2006, The Astrophysical Journal, 641, 21

Roe P., 1981, Journal of Computational Physics, 43, 357

Romer A. K., Viana P. T. P., Liddle A. R., Mann R. G., 2001, The Astrophysical Journal, 547, 594

Ruggiero R., Lima Neto G. B., 2017, Monthly Notices of the Royal Astronomical Society, 468, 4107

Russell H. R., Fabian A. C., McNamara B. R., Broderick A. E., 2015, Monthly Notices of the Royal Astronomical Society, 451, 588

Ruszkowski M., Springel V., 2009, The Astrophysical Journal, 696, 1094

Sądowski A., Narayan R., McKinney J. C., Tchekhovskoy A., 2014, Monthly Notices of the Royal Astronomical Society, 439, 503

Sagaut P., 2006, Large Eddy Simulation for Incompressible Flows. Scientific Computation, Springer-Verlag, Berlin/Heidelberg, doi:10.1007/b137536

Sahlén M., et al., 2009, Monthly Notices of the Royal Astronomical Society, 397, 577

Salpeter E. E., 1955, The Astrophysical Journal, 121, 161

Salvati L., Douspis M., Aghanim N., 2018, Astronomy & Astrophysics, 614, A13

Sandage A., 1976, The Astrophysical Journal, 205, 6

Sarmento R., Scannapieco E., Pan L., 2016, The Astrophysical Journal, 834, 23

Scalo J., Elmegreen B. G., 2004, Annual Review of Astronomy and Astrophysics, 42, 275

Scannapieco E., Bildsten L., 2005, The Astrophysical Journal, 629, L85

Scannapieco E., Brüggen M., 2008, The Astrophysical Journal, 686, 927

Schaye J., et al., 2010, Monthly Notices of the Royal Astronomical Society, 402, 1536

Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521

Schmidt W., 2015, Living Reviews in Computational Astrophysics, 1, 64

Schmidt W., Federrath C., 2011, Astronomy & Astrophysics, 528, A106

Schmidt W., Niemeyer J. C., Hillebrandt W., Röpke F. K., 2006, Astronomy & Astrophysics, 450, 283

Schmidt W., et al., 2014, Monthly Notices of the Royal Astronomical Society, 440, 3051

Schmidt W., Engels J. F., Niemeyer J. C., Almgren A. S., 2016, Monthly Notices of the Royal Astronomical Society, 459, 701

Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016a, The Astrophysical Journal, 826, 200

Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016b, The Astrophysical Journal, 826, 200

Shankar F., et al., 2015, The Astrophysical Journal, 802, 73

Shen S., Wadsley J., Stinson G., 2010, Monthly Notices of the Royal Astronomical Society, 407, 1581

Shen S., Madau P., Aguirre A., Guedes J., Mayer L., Wadsley J., 2012, The Astrophysical Journal, 760, 50

Shen S., Madau P., Guedes J., Mayer L., Prochaska J. X., Wadsley J., 2013, The Astrophysical Journal, 765, 89

Shi Y., Zhu X. X., Ellero M., Adams N. A., 2013, Computers & Fluids, 82, 122

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, Monthly Notices of the Royal Astronomical Society, 380, 877

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, Monthly Notices of the Royal Astronomical Society, 452, 575

Silk J., Rees M. J., 1998, Monthly Notices of the Royal Astronomical Society, 324, 128

Simons R. C., et al., 2020, The Astrophysical Journal, 905, 167

Smagorinsky J., 1963, Monthly Weather Review, 91, 99

Smith B. D., et al., 2017, Monthly Notices of the Royal Astronomical Society, 466, 2217

Sokołowska A., Mayer L., Babul A., Madau P., Shen S., 2016, The Astrophysical Journal, 819, 21

Sokołowska A., Babul A., Mayer L., Shen S., Madau P., 2018, The Astrophysical Journal, 867, 73

Somerville R. S., Davé R., 2015, Annu. Rev. Astron. Astrophys, 53, 51

Song M., et al., 2016, The Astrophysical Journal, 825, 5

Spaans M., Silk J., 2006, The Astrophysical Journal, 652, 902

Springel V., 2000, Monthly Notices of the Royal Astronomical Society, 312, 859

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105

Springel V., 2010, Monthly Notices of the Royal Astronomical Society, 401, 791

Springel V., Hernquist L., 2002, Monthly Notices of the Royal Astronomical Society, 339, 289

Springel V., Hernquist L., 2003a, Monthly Notices of the Royal Astronomical Society, 339, 289

Springel V., Hernquist L., 2003b, Monthly Notices of the Royal Astronomical Society, 339, 312

Springel V., White S. D. M., 1999, Monthly Notices of the Royal Astronomical Society, 307, 162

Springel V., Di Matteo T., Hernquist L., 2005a, Monthly Notices of the Royal Astronomical Society, 361, 776

Springel V., et al., 2005b, Nature, 435, 629

Spyropoulos E. T., Blaisdell G. A., 1996, AIAA Journal, 34, 990

Stott J. P., et al., 2010, Astrophysical Journal, 718, 23

Stott J. P., Collins C. A., Burke C., Hamilton-Morris V., Smith G. P., 2011, Monthly Notices of the Royal Astronomical Society, 414, 445

Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, Monthly Notices of the Royal Astronomical Society, 471, 144

Suresh J., Rubin K. H., Kannan R., Werk J. K., Hernquist L., Vogelsberger M., 2017, Monthly Notices of the Royal Astronomical Society, 465, 2966

Tabor G., Binney J., 1993, Monthly Notices of the Royal Astronomical Society, 263, 323

Tacconi L. J., et al., 2010, Nature, 463, 781

Tacconi L. J., et al., 2013, The Astrophysical Journal, 768, 74

Tadaki K.-i., et al., 2019, Publications of the Astronomical Society of Japan, 71, 1

Taffoni G., Mayer L., Colpi M., Governato F., 2003, Monthly Notices of the Royal Astronomical Society, 341, 434

Taghinia J. H., Rahman M. M., Lu X., 2018, Energy and Buildings, 170, 47

Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, Monthly Notices of the Royal Astronomical Society, 414, 195

Tilvi V., et al., 2020, The Astrophysical Journal, 891, L10

Tomczak A. R., et al., 2014, The Astrophysical Journal, 783, 85

Toro E. F., 2009, Riemann Solvers and Numerical Methods for Fluid Dynamics, 3 edn. Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/b79761, http://link.springer.com/10.1007/b79761

Torrey P., et al., 2019, Monthly Notices of the Royal Astronomical Society, 000

Trebitsch M., et al., 2021, Astronomy & Astrophysics, 653, A154

Tremaine S. D., Richstone D. O., 1977, The Astrophysical Journal, 212, 311

Tremaine S., et al., 2002, The Astrophysical Journal, 574, 740

Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, Monthly Notices of the Royal Astronomical Society, 451, 1868

Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, Monthly Notices of the Royal Astronomical Society, 470, 1121

Tremmel M., et al., 2019, Monthly Notices of the Royal Astronomical Society, 483, 3336

Trudeau A., et al., 2019, Monthly Notices of the Royal Astronomical Society, 487, 1210

Tumlinson J., et al., 2013, The Astrophysical Journal, 777, 59

Tumlinson J., Peeples M. S., Werk J. K., 2017, Annual Review of Astronomy and Astrophysics, 55, 389

Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, The Astrophysical Journal Supplement Series, 192, 9

Urzay J., O'Brien J., Ihme M., Moin P., Saghafian a., 2013, Center for Turbulence Research, pp 123–139

Vazza F., Brunetti G., Gheller C., Brunino R., 2010, New Astronomy, 15, 695

Vazza F., Roediger E., Brüggen M., 2012, Astronomy & Astrophysics, 544, A103

Vazza F., Jones T. W., Brüggen M., Brunetti G., Gheller C., Porter D., Ryu D., 2017, Monthly Notices of the Royal Astronomical Society, 464, 210

Vazza F., Angelinelli M., Jones T. W., Eckert D., Brüggen M., Brunetti G., Gheller C., 2018, Monthly Notices of the Royal Astronomical Society: Letters, 481, L120

Virtanen P., et al., 2020, Nature Methods, 17, 261

Vitvitska M., Klypin A. A., Kravtsov A. V., Wechsler R. H., Primack J. R., Bullock J. S., 2002, The Astrophysical Journal, 581, 799

Vogelsberger M., et al., 2014, Monthly Notices of the Royal Astronomical Society, 444, 1518

Voit G. M., 2005, Advances in Space Research, 36, 701

Volonteri M., 2010, Astronomy and Astrophysics Review, 18, 279

Vreman A. W., 2004, Physics of Fluids, 16, 3670

Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, Monthly Notices of the Royal Astronomical Society, 387, 427

Wadsley J. W., Keller B. W., Quinn T. R., 2017, Monthly Notices of the Royal Astronomical Society, 471, 2357

Wang C., Ruszkowski M., Yang H.-Y. K., 2020, Monthly Notices of the Royal Astronomical Society, 493, 4065

Webb T. M. A., et al., 2015, The Astrophysical Journal, 814, 96

Webb T. M. A., et al., 2017, The Astrophysical Journal, 844, L17

Weinberger R., et al., 2017, Monthly Notices of the Royal Astronomical Society, 465, 3291

Werk J. K., et al., 2016, The Astrophysical Journal, 833, 1

Whiley I. M., et al., 2008, Monthly Notices of the Royal Astronomical Society, 387, 1253

White S. D. M., 1976, Monthly Notices of the Royal Astronomical Society, 174, 19

White S. D. M., Rees M. J., 1978, Monthly Notices of the Royal Astronomical Society, 183, 341

Williamson D., Martel H., Kawata D., 2016, The Astrophysical Journal, 822, 91

Willis J. P., et al., 2020, Nature, 577, 39

Wilson G., et al., 2009, The Astrophysical Journal, 698, 1943

Winsberg E., 2010, Science in the age of computer simulation. University of Chicago Press

Wittor D., Jones T., Vazza F., Brüggen M., 2017, Monthly Notices of the Royal Astronomical Society, 471, 3212

Zhuravleva I., et al., 2014, Nature, 515, 85

Zhuravleva I., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 4184

Zhuravleva I., Allen S. W., Mantz A., Werner N., 2018, The Astrophysical Journal, 865, 53

ZuHone J. A., Markevitch M., Ruszkowski M., Lee D., 2013, The Astrophysical Journal, 762, 69

Zubko V., Dwek E., Arendt R. G., 2004, The Astrophysical Journal Supplement Series, 152, 211

van de Voort F., Schaye J., Altay G., Theuns T., 2012, Monthly Notices of the Royal Astronomical Society, 421, 2809

van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22

van der Wel A., Holden B. P., Zirm A. W., Franx M., Rettura A., Illingworth G. D., Ford H. C., 2008, The Astrophysical Journal, 688, 48