Withstanding the cold: energy feedback in simulations of galaxies that include a cold interstellar medium

Chaikin, E.

1 Introduction

How did the galaxies in our Universe form? How did they evolve with time? And how did they acquire the properties that we can observe today? A proper attempt to tackle these questions requires a gruelling effort. The physics of galaxy formation and evolution includes highly non-linear, stochastic processes that occur on scales spanning billions of orders of magnitude – from the accretion of gas onto supermassive black holes (SMBHs) on scales of less than 10^{-6} pc in the centres of massive galaxies, to the sizes of the largest gravitationally bound objects in the Universe, galaxy clusters, whose radii can extend to a few Mpc. Owing to the sheer complexity of the physics of galaxies, numerical methods provide the most accurate approach to testing the theory of galaxy formation and validating it by comparing the predictions of the galaxy simulations to observational data.

Due to the rapid development of computational facilities in the last several decades, numerical simulations of galaxy formation have been undergoing striking growth ever since the first simulations including gas hydrodynamics and gravity were run (e.g. Katz, 1992). Although state-of-the-art numerical models are able to match observations for a wide range of galaxy properties, the theory is nowhere near complete. In particular, many uncertainties exist in modelling astrophysical processes that cannot be resolved by direct simulation, even at the highest resolution used by the most advanced simulations.

We will now review the main aspects of the theory of galaxy formation, starting at scales where the Universe appears homogeneous and isotropic, and then 'zooming in' – step by step – towards scales as small as the sizes of SMBHs.

1.1 Cosmological background

The spatial distribution of $\gtrsim 10^6$ observed galaxies inferred from large, wide-field galaxy surveys such as the Sloan Digital Sky Survey (SDSS; e.g. Alam et al., 2015) indicates that our Universe is statistically homogeneous on scales greater than 100 comoving Mpc. Observations of the cosmic microwave background radiation (CMB; e.g. Planck Collaboration et al., 2020) – the electromagnetic radiation produced when the Universe was young and hot – show that our Universe is isotropic (i.e. it appears statistically the same in all angular directions from us). These observational facts form the basis of the Cosmological Principle, which

1

states that the properties of the Universe are indistinguishable for all possible (comoving) observers when averaged over large enough scales. On scales greater than 100 comoving Mpc, the Universe can therefore be mathematically described as homogeneous and isotropic.

Another fundamental observational fact is that our Universe is not static. In 1929, Edwin Hubble measured the radial velocities of a sample of 24 galaxies in the local Universe and found that the observed galaxies recede from our own Galaxy with velocities proportional to their distance (Hubble, 1929). This was the first concrete observational evidence that our Universe is expanding with time, which we now know with great certainty. The expansion of a homogeneous and isotropic universe is governed by Einstein's field equations of General Relativity, which are the foundation of the theoretical models that describe the dynamics of our Universe.

It appears that only ≈ 15 per cent of the matter in the Universe is observable, while the remaining ≈ 85 per cent neither emits nor absorbs light and for this reason is known as *dark matter* (DM). The existence of this mysterious DM, which is thought to interact with ordinary, baryonic matter exclusively via gravity, was first hypothesized almost a century ago in the work of Zwicky (1933), who analysed the motion of galaxies in the Coma cluster. By using the virial theorem, Zwicky (1933) concluded that the amount of visible matter in the galaxy cluster was insufficient to explain the observed motion of the galaxies; and to be consistent with the observed velocity dispersion, it would require another, invisible-to-light kind of matter to be present within the cluster. Nowadays, evidence supporting the existence of DM comes from a wide spectrum of different observational probes: from the present-day rotation curve of our own Galaxy measured through the 21-cm line of the HI gas (e.g. Gunn et al., 1979) to the abundance of galaxy clusters detected in X-rays (e.g. Vikhlinin et al., 2009), and to the spectrum of temperature anisotropies of the CMB (e.g. Planck Collaboration et al., 2020), which traces the state of the Universe more than 13 Gyr ago (see e.g. Hu & Dodelson, 2002, for more details on the CMB). Moreover, we know that (most of) the DM needs to be *cold*. Here 'cold' means that regardless of which fundamental particles or elements constitute DM, DM has to move with velocities low enough compared to the speed of light so that already at early times in the Universe, DM could cluster and form gravitationally bound structures, which would later become sites for the formation of galaxies.

An even greater puzzle is the existence of *dark energy*, which is an unknown form of energy that manifests itself only on very large, intergalactic scales and makes our Universe expand with acceleration. The fact that the expansion of the Universe is accelerating was unknown for a long time, until Riess et al. (1998) and Perlmutter et al. (1999) measured distances to type-Ia supernovae in distant galaxies and showed that only the theoretical models including an accelerated cosmic expansion could explain the observed fluxes from those objects. The most

widely used and accepted model of dark energy is the *cosmological constant* Λ, which in Einstein's field equations is described by a constant term (multiplied by the space-time metric tensor), physically corresponding to a space filled with a constant-density field.

Remarkably, the theoretical model of the Universe that includes both the cosmological constant Λ and cold dark matter (CDM) and depends on only 6 free parameters, has been shown to consistently reproduce an ever-growing amount of diverse observational data (e.g. Planck Collaboration et al., 2020; Abbott et al., 2022), coming from both the local Universe and from the early Universe. In this model, which is known as the ΛCDM model, our Universe started out with a hot big bang and has been expanding ever since, with the current age of our Universe being ≈ 13.8 Gyr.

1.1.1 The ΛCDM model

The ΛCDM model is made up of four energy components: the cosmological constant Λ; cold dark matter; ordinary matter, which is also referred to as 'baryons'; and radiation, comprising all relativistic forms of energy. Mathematically, the universe in the ΛCDM model is described by the Friedmann-Robertson-Walker (FRW) metric,

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a(t)^2\left[\mathrm{d}\xi^2 + S_K(\xi)^2\mathrm{d}\Omega^2\right], \tag{1.1}$$

which assumes that the space is both homogeneous and isotropic. In the equation above, t is the cosmic time, ξ denotes the comoving radial coordinate, c is the speed of light, $a(t)$ is the cosmic scale factor, and $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin(\theta)^2\mathrm{d}\varphi^2$ is a solid-angle element. It is customary to normalize the scale factor such that the universe today has $a = 1$. The function $S_K(\xi)$ depends on the spatial curvature, K, of the four-dimensional space-time,

$$S_K(\xi) = \begin{cases} K^{-1/2}\sin(\sqrt{K}\xi), & \text{if } K > 0 \\ \xi, & \text{if } K = 0 \\ (-K)^{-1/2}\sinh(\sqrt{-K}\xi), & \text{if } K < 0 \end{cases}. \tag{1.2}$$

All distances in Eq. (1.1) scale with the cosmic scale factor a, which is itself a function of cosmic time, meaning that a universe described by this metric is, in general, not stationary. In the framework defined by the FRW metric, an expanding universe has the time derivative $\dot{a} > 0$. If the universe has always been expanding, which we believe is the case for our Universe, then there is a one-to-one relation between the scale factor a and cosmic time t. This implies that one can describe any moment in time using either a or t. Another common way to parametrize the time evolution is through the use of (cosmological) redshift,

$$z = \frac{1}{a} - 1\,. \tag{1.3}$$

1

The redshift has an advantage over a and t in that it can be directly measured (e.g. from emission lines in galaxy spectra). The moment of today is given by $z = 0$ corresponding to $a = 1$, while galaxies in the past had $z > 0$ or, equivalently, $a < 1$.

By solving the equations of General Relativity for the FRW metric, one can derive the two Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K c^2}{a^2}, \tag{1.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right), \tag{1.5}$$

where ρ is the mass density of the universe, P is the pressure, G is the gravitational constant, and each dot above a represents a derivative with respect to cosmic time t. The quantity

$$H \equiv \frac{\dot{a}}{a}, \tag{1.6}$$

is termed the *Hubble parameter*, and its value today,

$$H|_{a=1} \equiv H_0 \equiv h\,100\,\mathrm{km\ s^{-1}\ Mpc^{-1}}, \tag{1.7}$$

is known as the *Hubble constant*[1], named after Edwin Hubble who first measured it in 1929 (Hubble, 1929). Specifically, Hubble (1929) found a linear relationship between the radial velocity of observed galaxies v (relative to our Galaxy) and the distance to the observed galaxies d (also relative to our Galaxy), which contains H_0 as the constant of proportionality,

$$v = H_0\, d, \tag{1.8}$$

and estimated H_0 to be ~ 500 km s^{-1} Mpc^{-1}. The measurements of H_0 from the last decade converge to a lower value, ≈ 70 km s^{-1} Mpc^{-1} (e.g. Alam et al., 2017; Planck Collaboration et al., 2020; Khetan et al., 2021) with an average scatter within ± 4 km s^{-1} Mpc^{-1} depending on the type of observational probe used to obtain H_0.

According to Eq. (1.4), in the absence of spatial curvature ($K = 0$), the density of the (homogeneous) universe can be computed in a straightforward way if the Hubble parameter is known,

$$\rho_{\mathrm{crit}} = \frac{3H^2}{8\pi G}, \tag{1.9}$$

which for the present-day Universe gives $\rho_{\mathrm{crit},0} \equiv \rho_{\mathrm{crit}}(a = 1) \approx 1.88 \times 10^{-29}\, h^2$ g cm^{-3}. We call it the critical density, ρ_{crit}, as opposed to the density ρ that enters

[1] The dimensionless parameter h was originally introduced to 'factor out' the uncertainty on the value of H_0.

Eqs. (1.4) and (1.5); 'critical' in the sense that $\rho_{\rm crit}$ is the density for which the universe is exactly flat ($K = 0$). Remarkably, the measurements of the CMB (e.g. Planck Collaboration et al., 2020) together with the measurements of the baryon acoustic oscillations (e.g. Alam et al., 2017) indicate with high certainty that our Universe is consistent with the spatially flat FRW metric.

By writing $\rho = \Sigma_i \rho_i$, we can decompose the total mass density ρ entering Eqs. (1.4) and (1.5) into contributions from individual energy components[2] that permeate our Universe. These are the cosmological constant Λ, CDM, baryons, and radiation. Moreover, because the relationships between pressure P and density ρ for non-relativistic matter (CDM and baryons), radiation, and the cosmological constant, are, respectively, $P(\rho) = 0$, $P(\rho) = \rho c^2/3$, and $P(\rho) = -\rho c^2$, one can solve the Friedmann equations for each component separately and write down the component's density as a function a. By introducing the dimensionless density of each energy component i, $\Omega_i(a) = \rho_i(a)/\rho_{\rm crit,0}$, we can then write,

$$\Omega_{\rm b}(a) = \Omega_{\rm b,0}\, a^{-3}, \tag{1.10}$$

$$\Omega_{\rm cdm}(a) = \Omega_{\rm cdm,0}\, a^{-3}, \tag{1.11}$$

$$\Omega_{\rm r}(a) = \Omega_{\rm r,0}\, a^{-4}, \tag{1.12}$$

$$\Omega_{\Lambda}(a) = \Omega_{\Lambda,0}, \tag{1.13}$$

where we defined $\Omega_{i,0} \equiv \rho_i(a = 1)/\rho_{\rm crit,0}$, which is the present-day mass density of energy component i expressed in units of the critical density, and the subscripts b, cdm, r, and Λ stand for, respectively, baryons, CDM, radiation, and the cosmological constant. The combined density of baryons and CDM is commonly expressed as the total matter density, $\Omega_{\rm m} \equiv \Omega_{\rm b} + \Omega_{\rm cdm}$. Nowadays, the values of $\Omega_{i,0}$ are known with only a few per cent uncertainties. These values are $\Omega_{\rm b,0} \approx 0.05$, $\Omega_{\rm cdm,0} \approx 0.27$, $\Omega_{\rm r,0} \approx 5.5 \times 10^{-5}$ and $\Omega_{\Lambda,0} \approx 0.68$, which have been inferred from, among others, the CMB data acquired by the big space missions in the past decades such as WMAP (Bennett et al., 2003) and Planck (Planck Collaboration et al., 2014, 2020). Although the present-day value of $\Omega_{\rm r}$ is very small compared to the other energy components, $\Omega_{\rm r}$ increases fastest with decreasing a, meaning that at very early times ($a \ll 1$), the expansion of the Universe was dominated by radiation. The fact that the sum $\Sigma_i \Omega_{i,0} \approx 1$ indicates that our Universe is spatially flat or is very close to it.

1.1.2 Density perturbations and their growth in time

Of course, today's Universe around us is far from homogeneous, and the model discussed above is an approximation whose validity is justified only on very large

[2] We use the words 'mass' and 'energy' interchangeably because the energy density ε and mass density ρ are related as $\varepsilon = \rho c^2$, where c is the speed of light.

1

scales, and exclusively in the statistical sense. The theory predicts, however, that when the Universe was very young ($z \gtrsim 10^{10}$), it was in fact very close to being homogeneous. No galaxies existed back then; instead, the space was filled with small, 'primordial' density fluctuations ($\delta\rho/\rho \ll 1$), whose origin is still debated. The most widely accepted explanation is the theory of cosmic inflation, which predicts that shortly after the Big Bang, the Universe underwent a phase of rapid, exponential expansion, which blew quantum fluctuations of space-time to macroscopic scales. These primordial density fluctuations then became the seeds for the structure formation that would take place in the Universe at later times.

Mathematically, a density fluctuation can be described using the density contrast δ defined relative to the mean, background density of the Universe $\langle\rho\rangle$, $\delta \equiv \rho/\langle\rho\rangle - 1$. The fact that a density fluctuation is small therefore means $|\delta| \ll 1$. As long as $|\delta| \ll 1$, one can use the linear approximation of the Euler and Poisson equations to describe the evolution of δ. It is easy to demonstrate that in an expanding, matter-dominated universe a positive density fluctuation ($\delta > 0$) will grow in proportion to the cosmic scale factor

$$\delta(a) \propto a. \tag{1.14}$$

If the density contrast becomes comparable to unity, linear theory is no longer applicable. Fortunately, in this regime, the equations of hydrodynamics and gravity permit one exact analytical solution that describes the evolution of a spherically symmetric perturbation, which is known as the spherical collapse model. In the spherical collapse model, a spherically-symmetric, positive density fluctuation (a sphere with a uniform density that is higher than the density of the background) evolves in an otherwise homogeneous background. Initially, the spherical overdensity expands along with the homogeneous background but later the expansion of the perturbation begins to decelerate with respect to that of the background, which is due to the force of gravity pulling the perturbation inwards. Eventually, the expansion of the spherical overdensity halts, turns over, and the overdensity begins to collapse onto itself. The spherically-symmetric solution predicts that the perturbation will have collapsed into a singularity with infinite density. In reality, however, such a scenario is not possible because density perturbations are neither spherical nor homogeneous. What happens instead is that the perturbation will keep collapsing until it has reached virial equilibrium (i.e. the kinetic energy of the perturbation becomes equal to minus half of its potential energy). After virial equilibrium has been established, the perturbation will form a gravitationally bound structure, whose geometry in the first approximation can be described as spherical. We call such virialised density structures *dark-matter haloes*. Using the spherical collapse model and virial theorem, and assuming a flat ($K = 0$), matter-only universe, one can show that the mean density of a virialised, gravitationally-bound structure is $\Delta = 18\pi^2 \approx 178$ times greater than the density

of the homogeneous background. The (virial) mass of a spherically-symmetric halo is then

$$M_\Delta = \int_0^{r_\Delta} \mathrm{d}r\, 4\pi r^2 \rho(r) = \frac{4\pi}{3} r_\Delta^3 \rho_{\mathrm{crit}}\, \Delta\,, \tag{1.15}$$

where r_Δ is the radius of the sphere within which the density is Δ times greater than the critical density of the universe. Because $18\pi^2$ is close to the round number of 200, it became commonplace in cosmology to define the virial radius and virial mass of haloes using $\Delta = 200$. In the following text, any reference to halo mass or virial mass should be understood in the context of Eq. (1.15) with $\Delta = 200$, unless explicitly stated otherwise.

1.2 Baryons and galaxies

Measurements of element abundances in the early Universe, as well as the constraints inferred from the CMB temperature anisotropies, indicate that the energy density of our Universe consists of approximately 4 to 5 per cent of baryons. Although constituting a small fraction of the total energy density in the Universe, baryons are an essential building block of stars and galaxies. Unlike DM, baryons are collisional: they interact electromagnetically, experience pressure forces, and can dissipate energy. As we will discuss later, energy dissipation is a crucial step in creating conditions appropriate for star formation.

At very early times, the Universe was very hot, opaque, and ionized, and the baryons were strongly coupled to the photons. However, as the Universe expanded, it gradually became more dilute and its temperature decreased. At the times when the energy density of the Universe was dominated by radiation (redshifts $z \gtrsim 3000$), the age of the Universe and its temperature were related as

$$t \approx 1\,\mathrm{s} \left(\frac{T}{1\,\mathrm{MeV}}\right)^{-2}. \tag{1.16}$$

More precisely, the temperature T in the above equation is the temperature of the blackbody spectrum of the photons that permeated the Universe at time t. Having energies as high as $\sim$ MeV when the Universe was only ~ 1 s old, those photons kept the Universe ionized. However, by the time the age of the Universe reached ≈ 380 kyr, corresponding to redshift $z \approx 1100$, the temperature of the photons had dropped so much that they could no longer stop the protons and electrons from recombining into neutral particles. By the end of the recombination process, the Universe turned from nearly completely ionized to nearly completely neutral. Furthermore, this transition, in which the fraction of free electrons dropped enormously, allowed the photons to begin propagating almost freely through space, without undergoing regular scatterings on free electrons, as was the case before recombination. These free-streaming photons constitute the

cosmic microwave background that we observe today at the blackbody temperature of $T_{\rm CMB}(z=0) = 2.73$ K. The properties of the CMB reflect the properties of the Universe at the time when the CMB photons were released ($z \approx 1100$).

Following recombination the fraction of free electrons was very low but non-zero ($X_{\rm e} \sim 10^{-4}$) because not every electron was lucky enough to find a positively-charged particle and recombine. Although small, the free electron fraction of $X_{\rm e} \sim 10^{-4}$ was sufficient to keep the temperature of the gas of baryons coupled to the temperature of the photons, such that the gas temperature was evolving as $T_{\rm gas}(z) = T_{\rm CMB}(z) = 2.73(1+z)$ K. However, at redshifts lower than $z \approx 150$, the density of the baryons became so low that Compton scattering was no longer efficient at maintaining the coupling between the baryons and photons through the residual free-electron fraction. The baryonic gas started to cool adiabatically, which proceeded at a rate significantly faster than before, with the gas temperature evolving as $T_{\rm gas}(z) \propto (1+z)^2$. During that epoch, no galaxies existed yet, but DM haloes that had already formed from the collapse of density fluctuations were growing by accreting extra DM mass and assembling into more massive haloes through hierarchical merging (e.g. Lacey & Cole, 1993).

1.2.1 First galaxies

The formation of the first galaxies begins with the infall of gas towards the centres of DM haloes with virial mass $M_{200} \gtrsim 10^5$ M$_\odot$[3], and subsequent (molecular) cooling inside. When the gas enters the potential well of a DM halo, it can be shock-heated to the virial temperature of the halo, which is a measure of the depth of the halo's gravitational potential. Alternatively, the gas can flow into the halo along the cosmic web's filaments attached to the halo, in which case the change in the gas temperature due to the variations in the gravitational potential is more modest. Since stars are expected to form in cold, dense gas, the ability of the infalling, shock-heated gas to cool is essential in that in the absence of cooling, the initially hot gas would not be able to dissipate its energy, condense to the centre of the halo, and reach the high densities that are required for star formation.

Inside the halo, the cooling gas tends to settle into rotating, disc-like structures, which are a mere consequence of the fact that the inflowing gas has a non-zero angular momentum due to tidal torques from the nearby large-scale structures. Clouds of gas close to the centre of the halo eventually become sufficiently dense and cold, leading to the collapse initiated by their own gravity.

[3] The gas entering DM haloes with lower M_{200} is unable to cool and form stars. Also note that after the photons emitted by the first galaxies re-ionized the Universe and heated the gas to temperatures $> 10^4$ K, the halo's minimum mass for galaxy formation increased from $\sim 10^5$ M$_\odot$ to $\sim 10^8$ M$_\odot$ (e.g. Wise, 2019).

Mathematically, the mass of a spherically symmetric gas cloud that is expected to undergo a gravitational collapse has to be greater than the Jeans mass

$$M_{\rm J} = \left[\frac{10 k_{\rm B} T}{3(\gamma-1)\mu m_{\rm p} G}\right]^{3/2} \left(\frac{3}{4\pi\rho}\right)^{1/2}, \tag{1.17}$$

where $k_{\rm B}$ is the Boltzmann constant, μ is the mean molecular weight, $m_{\rm p}$ is the proton mass, γ is the ratio of specific heats of the gas, and T and ρ are the gas temperature and density, respectively. If the cooling time-scale t_c is much shorter than the free-fall time-scale,

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32 G\rho}}, \tag{1.18}$$

the collapse of the cloud can be approximated as isothermal. Eq. (1.17) then tells that as the density inside the cloud gets higher, its Jeans mass decreases. The latter implies that if there are some density perturbations on smaller scales within the collapsing cloud, their actual masses may eventually exceed their Jeans masses. The perturbations will then become gravitationally unstable and begin to collapse on their own. The process when a large collapsing gas cloud gets disintegrated into smaller clouds is called *fragmentation*.

At some later point, the inner regions of the smallest collapsing clouds reach such high densities that the photons released by the cooling gas become trapped inside. If the central region of a cloud is optically thick, its energy is no longer lost, and the cloud's further collapse can be roughly approximated as adiabatic. For an adiabatic collapse, both the temperature and the Jeans mass increase with the density. The moment when the Jeans mass of the optically thick core of the cloud exceeds the core's gravitational mass, the collapse of the core halts, with the core entering a hydrostatic equilibrium. From that moment on, such cores are known as protostars. The collapse is not stopped completely, however. While the density and temperature of the collapsing clouds increase towards the centre, their outskirts remain relatively cool and optically thin. As the surface of the hotter core radiates energy and the core accretes more matter from the outskirts, the core's (slow) contraction has to continue. The further contraction is accompanied by a further temperature increase in the centre. When the central temperature reaches a threshold of $\approx 10^7$ K, hydrogen fusion begins, and the protostar becomes a zero-age main-sequence star. The stars that are born from the fragmented gas that fell into the halo, together with all gas in between the stars, can be loosely defined as a *galaxy*.

1.2.2 Basics of stellar nucleosynthesis

Stars support themselves against gravitational collapse through nuclear fusion. The process of burning lighter elements into heavier ones releases energy, which

provides the necessary pressure support against gravity. To ignite a nuclear reaction, the stellar core has to reach a certain minimum temperature. Fusion of heavier elements generally requires higher temperatures and proceeds on shorter time-scales. When an element constituting the nuclear fuel is used up, gravity starts to dominate, which forces the star to contract and heat up. The contraction stops when the temperature of the core rises to a level that allows the burning of the heavier element, which is the product of the previous reaction(s).

Through the process of nuclear fusion, hydrogen is burnt into helium, and, subsequently, helium into carbon and oxygen. Because the maximum attainable temperature in the stellar core increases with the initial mass of the star, the cores of the stars that are less massive than $\approx 8\,M_\odot$ cannot satisfy the conditions required for the burning of carbon and oxygen, making them the heaviest elements produced in these stars. On the other hand, stars with initial masses[4] greater than $\approx 8\,M_\odot$ – or, for short, *massive stars* – do undergo further fusion, which includes the burning of carbon, oxygen, neon, silicon, as well as several other heavy elements. Ultimately, after a chain of nuclear-burning cycles, an iron core is formed, which is the end product of stellar nuclear fusion in massive stars. The fusion cannot proceed beyond iron because elements of the iron group have the highest binding energy per nucleon among all elements in the Universe, which means that by burning iron the star will not be able to generate energy to counteract gravity.

1.2.3 Stellar feedback and metal enrichment

Stars not only produce heavy elements in their interiors, but they also interact with the gas that surrounds them by emitting radiation and ejecting energy, momentum, and mass. This stellar 'feedback' plays a vital role in galaxy evolution. On small scales, stellar feedback can disperse the high-density gas within giant molecular clouds (GMCs) where the stars have originally formed; while on large, galactic scales, the cumulative effect of stellar feedback from an ensemble of stars may result in pushing the gas out of the galaxy, providing a mechanism to balance the cosmological accretion of gas onto the galaxy. Crucially, the impact of stellar feedback on both small and large scales allows galaxies to stabilize and modulate the rate at which they form stars. Indeed, observations have consistently reported that depletion times[5] in GMCs are a factor of $10-100$ greater than the clouds' free-fall times (e.g. Evans et al., 2009), which hints that some form of stellar feedback is necessary to obtain such low efficiency of star formation. Furthermore,

[4] The initial stellar mass is the mass of the star at the moment when the star has entered the zero-age main sequence.

[5] The depletion time of a (molecular) gas cloud is estimated by dividing the cloud's mass by its star formation rate. Physically, it gives the characteristic time-scale on which the cloud's mass is expected to be converted into stars.

multiple simulations of galaxy formation in the past decades have shown that without stellar feedback, simulated galaxies always form way too much stellar mass compared to observations (e.g. Schaye et al., 2010). Finally, metals in the circumgalactic medium (CGM) and intergalactic medium (IGM) such as carbon and oxygen are detected in absorption (e.g. Ellison et al., 2000; Tumlinson et al., 2011), indicating that the products of stellar nucleosynthesis had to find a way to escape the galaxy's interstellar medium (ISM) into the CGM and then IGM, and having stellar feedback is essential to be able to match these observed data (e.g. Oppenheimer & Davé, 2006; Muratov et al., 2017).

We will now discuss the three feedback processes that are important before the final stages of stellar evolution, starting shortly after the moment when the stars have reached the zero-age main sequence.

- When the radiation produced in the stellar interior is absorbed by metals in the outer layers of stars, it exerts an outward acceleration onto the absorbing gas, which – if strong enough – can cause the removal of stellar mass from the outer layers, thus creating a *stellar wind* (e.g. Cassinelli, 1979). Because of the very steep dependence of stellar luminosity on stellar mass, stellar winds are most important for the evolution of massive stars. Depending on the initial stellar mass, metallicity, and the evolutionary stage of the star, the mass loss rate through the stellar wind can vary between virtually zero and $\sim 10^{-3}$ $M_\odot\,s^{-1}$, with post-main-sequence stars experiencing the strongest outflows. The velocity of the wind can in certain cases reach values as high as ~ 3000 km s^{-1}, but can also remain at ~ 10 km s^{-1}. Since a stellar wind carries energy, mass, and momentum, it disperses and heats the gas in the vicinity of the star. Additionally, because stellar winds can be metal-rich, they affect the chemical composition of the gas local to the star.

- In addition to stellar winds, stars can inject momentum into the gas around them by direct radiation pressure that is coming from the stellar photons (e.g. Raskutti et al., 2017; Menon et al., 2023). This process is expected to be mediated by dust grains, whose opacity is high enough for stellar radiation to be frequently absorbed. The dust grains then transfer the absorbed photons' momentum to the nearby gas through collisions. Similar to stellar winds, radiation pressure is more important for more massive stars.

- Finally, stars can affect the surrounding gas through photoionization and photodissociation. This form of feedback is called HII regions. Massive stars generate a sizeable amount of photons with energies above 13.6 eV, which are capable of ionizing neutral hydrogen. As a result, massive stars live in bubbles of ionized gas, which are known as Strömgren spheres, whose (sharp) boundary is defined by the balance between the rate of ionizing photons produced by the massive star and the rate of recombination of pro-

1

tons and electrons into neutral hydrogen. The gas within an HII region has a temperature of $\sim 10^4$ K, corresponding to the sound speed of ~ 10 km s^{-1}. The gas inside the ionized bubble is overpressured relative to the neutral background, so the HII region expands with time with a velocity comparable to the sound speed.

An ample amount of energy, momentum, and mass (including metal mass), is also released at the end of stellar evolution of massive stars, which is what we will discuss next.

1.2.4 Core-collapse supernovae

The central density within iron cores, which develop inside massive stars by the end of stellar nucleosynthesis, exceeds 10^9 g cm^{-3}. At such high densities, a gas of free electrons is degenerate. Thanks to the Pauli exclusion principle, these free electrons produce (quantum) pressure support that stabilizes the iron core against gravitational collapse even in the absence of nuclear fusion. However, a system that is powered solely by degeneracy pressure cannot remain in equilibrium indefinitely. First, because the mass of the iron core is close to the Chandrasekhar mass[6], $M_{\rm Ch} \approx 1.4$ M$_\odot$, which means any extra layer of mass accreted to the outer boundary of the iron core can potentially destabilize the core. And second, because at densities $\gtrsim 10^9$ g cm^{-3} the inverse β process becomes efficient, through which free electrons are readily captured by iron nuclei. This leads to a diminishing number of free electrons, which weakens the electron pressure and further destabilizes the core. As a result of these two effects, the core begins to contract, which is accompanied by an increase in the core's temperature. When the central temperature reaches 10^{10} K, photons become so energetic that they start to disintegrate iron into helium nuclei. The process of disintegration consumes energy, rather than releasing it, which expedites the contraction and makes the core even more unstable. Ultimately, the core collapses onto itself, with the whole process taking less than a second to complete.

The core collapse triggers a massive explosion of the star, during which more than 10^{53} erg of (gravitational binding) energy is released in different forms. Around 99 per cent is liberated via neutrinos. About one per cent of energy is released kinetically, and the remainder, a fraction of less than one per cent, goes into radiation. The exploding star reaches a peak luminosity comparable to that of a moderate-mass galaxy ($L \sim 10^9$ L$_\odot$). This explosion is known as a *core-collapse supernova* (CC SN).

A CC SN releases ~ 10 M$_\odot$ of mass back into the ISM in the form of a blast-wave, and is expected to leave behind a remnant: a neutron star or a black hole

[6] The Chandrasekhar mass marks the maximum mass of a self-gravitating object at zero temperature that can be supported by the degeneracy pressure of its electrons.

(BH). Due to the high kinetic energy, the blastwave made of CC SN ejecta initially moves ballistically. The velocities that are developed during this stage reach 10^4 km s^{-1}. Since the typical temperature in the ISM is 10^4 K corresponding to the sound speed of ≈ 10 km s^{-1}, the ejecta moving at 10^4 km s^{-1} generate a strong shock. Geometrically, the SN ejecta resemble a bubble whose radius is given by the front of the (forward) shock and whose centre coincides with the position of the stellar remnant. This bubble expands with time, shocking, compressing, and sweeping up the ISM gas in front of it. Eventually (~ 1 kyr since the SN explosion; the exact time depends on the ISM density and composition), the amount of the ISM gas entrained within the blastwave becomes comparable to the initial mass of the SN ejecta. This is the moment when the blastwave begins to slow down, and its kinematic evolution needs to be described by the energy-conserving solution, as opposed to the ballistic solution. Taylor (1950) and Sedov (1959) showed that during the energy-conserving phase, the evolution of the bubble – the bubble's radius R as a function of time since the SN explosion t – depends solely on the explosion's kinetic energy, $E_{\rm SN}$, and the density of the ISM, ρ,

$$r(t) = \alpha_0 \left(\frac{E_{\rm SN} t^2}{\rho}\right)^{1/5} = 8.45 \text{ pc} \left(\frac{E_{\rm SN}}{10^{51} \text{ erg}}\right)^{1/5} \left(\frac{\rho}{0.1\, m_{\rm p} \text{ cm}^{-3}}\right)^{-1/5} \left(\frac{t}{1 \text{ kyr}}\right)^{2/5}, \tag{1.19}$$

where the order-of-unity dimensionless constant $\alpha_0 = 1.15167$ for the specific-heat ratio of $\gamma = 5/3$[7]. Most of the ISM material that is swept up by the propagating shock is concentrated in a thin layer immediately behind the shock front. As more and more material is being swept up, the gas behind the shock eventually gets very dense leading to significant radiative cooling. In consequence, the SN blastwave begins to lose its energy and move slower than predicted by Eq. (1.19). The evolution of the SN bubble is then described by the (radial) momentum-conserving solution, $r(t) \propto t^{1/4}$. Finally, the SN blastwave merges with the ISM when its radial velocity becomes comparable to the sound speed in the ISM.

Similar to stellar winds and radiation pressure feedback, the energy and momentum injected into the ISM by (numerous) CC SNe provide a way to stabilize the galaxy ISM and regulate star formation therein. However, the energies involved in CC SNe are in general greater than in the feedback processes discussed before, making CC SNe the dominant form of stellar feedback[8]. Indeed, theoretical models and numerical simulations have consistently shown that it is the cumulative effect of CC SN feedback from massive stars that is responsible for building up the hot phase of the ISM and generating highly mass-loaded galac-

[7] The Sedov-Taylor solution assumes that (i) the ISM is homogeneous, (ii) the gas in the ISM has negligible pressure, and (iii) the SN bubble is spherically symmetric.

[8] The high efficiency of CC SN feedback can be partly attributed to stellar winds and radiation pressure, which start dispersing high-density clumps within GMCs before the first CC SN goes off, thereby mitigating radiative energy losses in subsequent SN explosions (e.g. Agertz et al., 2013).

1

tic winds in relatively low- and moderate-mass galaxies (e.g. McKee & Ostriker, 1977; Scannapieco et al., 2008; Naab & Ostriker, 2017). Furthermore, because prior to the explosion, the interior of massive stars contains numerous products of nuclear fusion, ranging from He to Fe, feedback from CC SNe is important for the chemical enrichment of the gas, both within the ISM and beyond (e.g. Nomoto et al., 2006). Finally, SN remnants are major sources of cosmic rays, which can provide additional pressure support and heating mechanism for the gas within the galaxy, and may help drive galactic-scale winds (e.g. Pakmor et al., 2016).

1.2.5 Type-Ia supernovae

Stars with initial masses of less than $\approx 8\,\mathrm{M}_\odot$ cannot ignite carbon because their cores have insufficient mass to reach the required minimum temperature for carbon burning ($T \approx 5 \times 10^8$ K). As a consequence, such stars do not evolve towards SNe triggered by the collapse of the iron core[9]. Instead, they leave a compact remnant behind without an explosion after they have exhausted their nuclear reservoir and expelled their outer envelopes. These compact remnants are known as *white dwarfs*. White dwarfs have masses of $\sim 1\ \mathrm{M}_\odot$ and – similar to iron cores formed in massive stars – are supported against gravitational collapse by degeneracy pressure provided by free electrons. No white dwarf can exist with a mass greater than the Chandrasekhar mass, $M_{\mathrm{Ch}} \approx 1.4\ \mathrm{M}_\odot$.

Because a substantial fraction of stars are part of binary systems, white dwarfs commonly accrete matter from their companion stars, provided the separation between the companion stars and white dwarfs is small enough. If the accretion onto the white dwarf is substantial, the extra weight of the accreted matter exerted on the outer layers of the white dwarf will make it contract and heat up, such that eventually the conditions in the white dwarf's core may become favourable for carbon burning. If the carbon is ignited, the energy liberated from the carbon nuclear fusion will continue heating up the remnant. Yet, because the pressure of a degenerate gas is independent of temperature, the white dwarf will be unable to adapt to the released energy by expanding and cooling. It is important that the mass of the white dwarf is close to the Chandrasekhar mass, so that even at temperatures approaching 10^{10} K, the white dwarf remains degenerate. As a result, the thermonuclear fusion of carbon will proceed in a runaway fashion, ultimately leading to the explosion of the remnant[10]. Explosions of white

[9] More precisely, stars with initial masses in the transition mass range between massive and intermediate-mass stars, $\approx 8-10\ \mathrm{M}_\odot$, might not form an iron core either. However, they can still evolve toward an SN explosion by capturing free electrons onto their degenerate O+Ne+Mg cores. SNe triggered by the electron-capture reactions are known as *electron-capture SNe*.

[10] An alternative scenario that may lead to a thermonuclear explosion of a white dwarf is the merger of two white dwarfs that belong to the same binary system. The two white dwarfs should gradually coalesce and merge because of the loss of their orbital energy through gravitational waves.

dwarfs due to a thermonuclear runaway collapse are known as *type-Ia supernovae* (type-Ia SNe).

Similar to CC SNe, the kinetic energy that is received by the gas surrounding the exploded white dwarf is of the order of 10^{51} erg. However, different from CC SNe, a type-Ia SN can detonate on time-scales from $\sim 10^2$ Myr to up to many Gyr since the progenitor star was born. In contrast, massive stars that result in CC SNe explode on time-scales between ≈ 3 and 40 Myr since their birth, and for this reason, are predominantly found in galaxies that are actively forming stars. Besides energy feedback, type-Ia SNe also enrich gas with metals. In particular, type-Ia SNe produce about 50 per cent of iron in the Universe (e.g. Maoz & Graur, 2017).

1.2.6 Black holes

It is now widely accepted that galaxies harbour SMBHs in their centres. One of the strongest pieces of evidence in favour of their existence is the kinematics of stars around galaxy centres (e.g. Kormendy & Ho, 2013). Measuring the velocities of stars from their spectra in nearby, massive galaxies reveals that a gravitational mass greater than $\sim 10^7$ $M_\odot$, which does not emit light, has to be concentrated within $\lesssim 0.1$ pc. No objects composed of ordinary matter, like gas or stars, are currently known that can constitute such a high mass density. In contrast, SMBHs – which are a solution of the equations of General Relativity, have sizes much less than 0.1 pc, and can theoretically reach masses as high as 10^{10} $M_\odot$ – are perfect candidates to explain the high mass concentrations found in the centres of massive galaxies.

With the advent of the Event Horizon Telescope and its first two images of two SMBHs – first at the centre of the M87 galaxy in 2019 (Event Horizon Telescope Collaboration et al., 2019), and then at the centre of our own Galaxy in 2022 (Event Horizon Telescope Collaboration et al., 2022) – the consensus that most galaxies contain SMBHs has only solidified.

The formation of SMBHs is not yet well understood. Several theories exist that can explain the origin of BHs with masses of $\sim 10^2 - 10^5$ $M_\odot$ (e.g. Madau & Rees, 2001; Devecchi & Volonteri, 2009; Shang et al., 2010), which are usually referred to as 'seeds for SMBHs'. These seeds are expected to then grow into more massive BHs via merging with other BHs and/or by accreting surrounding gas. One possible route for creating seeds for SMBHs is the end products of Population III stars. Population III stars are the first stars that formed in the Universe. Because no metal enrichment had taken place prior to their birth (apart from an insignificant amount of metals produced during Big Bang Nucleosynthesis), the composition of their parent gas clouds was extremely metal-poor, which made the Population III stars more massive and more short-lived than typical massive stars observed in the present-day Universe. As a result, Population III stars are

1

expected to have left behind more massive BH remnants ($M_{\rm BH} \sim 10^2 - 10^3\,{\rm M}_\odot$), as opposed to the $M_{\rm BH} \sim 10\,{\rm M}_\odot$ BHs expected from massive stars born in the current epoch. An alternative to this first scenario is a direct collapse scenario in which a dense, metal-free gas cloud in the centre of a $\gtrsim 10^8\,{\rm M}_\odot$-mass halo undergoes a direct collapse into a $\sim 10^4 - 10^5\,{\rm M}_\odot$ BH. A third possibility for creating BH seeds is stellar collisions in star clusters proceeding in a runaway fashion, helped by the enhanced merging rates of BHs from individual stars in such dense environments. This scenario can explain the formation of $\sim 10^2 - 10^4\,{\rm M}_\odot$ BH seeds.

1.2.7 Active galactic nuclei

Compact ($r \lesssim 0.1$ kpc) regions in the centres of Milky Way-mass and more massive galaxies may occasionally appear unusually bright in various spectral bands, from radio to γ-ray energies. In the literature, centres of galaxies that appear too luminous to be explained by the stellar population alone, have received the name *active galactic nuclei* (AGN). The general consensus is that AGNs are caused by enormous releases of gravitational energy when the gas is being accreted onto SMBHs. It is also generally agreed that while stellar feedback regulates the evolution of low- and intermediate-mass galaxies, the evolution of massive galaxies is dominated by the energy feedback from AGN. In particular, AGN feedback provides a way to keep massive galaxies quenched, which allows models to reproduce the observed galaxy luminosity and mass functions (e.g. Bower et al., 2006); and is a critical heating mechanism that offsets gas cooling in the central regions of galaxy clusters, which is necessary to explain and reproduce the clusters' observed X-ray emission (e.g. Mittal et al., 2009).

The radiation emitted by AGNs is distinguished by (i) highly broad-band spectrum, spanning many orders of magnitude in frequencies; (ii) extremely broadened spectral lines, with line widths up to $\sim 10^4$ km s^{-1} (e.g. Kollatschny & Zetzl, 2013); and (iii) high bolometric luminosities, reaching values up to 10^{48} erg s^{-1} (e.g. Fiore et al., 2017). These and other features of AGN emission are produced by a combination of various physical phenomena operating together:

- SMBHs are surrounded by rapidly orbiting gas clouds, which can move with velocities as high as $\sim 10^4$ km s^{-1}. The Doppler effect applied to the velocity dispersion from these individual clouds leads to the lines' broad spectral widths observed in AGN spectra.

- SMBHs are expected to launch two collimated, relativistic jets when they accrete matter at a relatively low rate, with the jet efficiency increasing with the SMBH spin (e.g. Tchekhovskoy et al., 2010). Although the jets are created in the very compact region around the SMBH (whose size is comparable to the SMBH's gravitational radius, $\ll 1$ pc), they may extend out to distances as large as hundreds of kpc. Thanks to their prominent

synchrotron emission in the radio band, the jets form one of the strongest observational lines of evidence for AGN (e.g. Marshall et al., 2001).

- The AGN radiation at (ultra-)high frequencies is thought to be generated via inverse Compton scattering, whereby the ambient photons near the SMBH gain energy by interacting with relativistic jets of magnetized plasma ejected from the SMBH (e.g. Blandford et al., 2019).

- SMBHs are encircled by gaseous accretion discs formed from the inflowing gas that cannot easily lose its angular momentum. The inner edge of the accretion disc is located very close to the SMBH (just a few times the BH's gravitational radius). The emission from the accretion disc spans the optical and UV parts of AGN spectra. The properties and geometry of the accretion disc change depending on the rate with which the gas is being accreted onto the SMBH, which has consequences for the properties of AGN feedback. We will discuss this in more detail in §1.4.8.

1.3 Numerical simulations of galaxy formation

In the previous section, we reviewed the essential processes involved in the physics of galaxy formation and evolution. We will next discuss the motivation behind running numerical simulations of galaxy formation (§1.3.1), as well as touch upon the resolution and box size requirements for such simulations (§1.3.2). We will then review the numerical methods that are commonly employed to run simulations of galaxy formation (§1.3.3 and §1.3.4). Finally, in Section 1.4, we will describe the popular numerical algorithms for implementing essential astrophysical processes in a simulation of galaxy formation.

1.3.1 Motivation

Numerical simulations of galaxy formation provide us with a unique ability to test and improve our theoretical understanding of how galaxies form, interact, and evolve. They are of paramount importance because they give us the most accurate way of modelling non-linear galaxy evolution and the most reliable way to compare theory to observations. Unlike empirical or semi-analytical models of galaxy formation, which are also used to test galaxy formation physics (e.g. Lacey et al., 2016; Behroozi et al., 2019), numerical simulations rely on a much smaller number of (crude) approximations within the galaxy formation physics and, for this reason, are able to trace the highly non-linear dynamics and interactions of gas and stars in much more detail. The downside is, however, that numerical simulations are very expensive to run, and it is far from obvious how to tune the numerical model so that the simulated galaxies appear realistic and have properties matching observational data.

1

Numerical simulations can be exploited in various ways. If the simulation adopting a certain model of galaxy formation is unable to reproduce fundamental galaxy properties such as the galaxy stellar mass function and/or star formation main sequence, then the model can be ruled out and the physics underlying the model constrained. Conversely, for models that result in a realistic galaxy population, the simulated galaxies and their evolution can be thoroughly explored, providing insights into the physics of galaxy formation that cannot be inferred from observations or calculated analytically.

To give a few concrete examples: having a simulation with a series of outputs at consecutive redshifts, one can accurately trace the simulated galaxies throughout cosmic time to study the impact of galaxy-galaxy merges on the activity of AGN feedback (e.g. McAlpine et al., 2020), investigate how the properties of backsplash galaxies evolve during their passage through a galaxy cluster (e.g. Borrow et al., 2023), or determine the right conditions for cosmic gas accretion under which Milky Way-mass galaxies are likely to form stellar discs (e.g. Hafen et al., 2022; Semenov et al., 2023). These three examples constitute only a tiny fraction of the full list of applications of galaxy simulations. In principle, any aspect of galaxy formation and evolution physics can be tested with the use of numerical simulations.

Simulations of galaxy formation can constrain not only the galaxy formation physics but also the underlying cosmological model, on which the galaxy simulation is built. Specifically, if the simulated volume is sufficiently large, one can study the X-ray emission of hot gas in galaxy clusters, the effects of baryonic feedback on the large-scale structure, galaxy clustering, cosmic shear, the Sunyaev-Zel'dovich signal in the CMB maps, the equation of state of the dark energy (e.g. assuming that dark energy is not simply described by the cosmological constant Λ), alternative models of DM (e.g. assuming that DM particles can self-interact; see e.g. Correa et al., 2022), and even theories of modified gravity (i.e. beyond the theory of General Relativity; see e.g. Arnold et al., 2019).

Another notable application of numerical simulations is the development of a better understanding of individual physical processes that are incorporated into a galaxy formation model. This task can be accomplished by running two or more numerical simulations for the same initial conditions but with different values of some of the parameters of the model. Any systematic differences in the properties of the galaxy populations from the different simulations can then be attributed to the difference in the model parameters. Variations in model parameters may include (i) switching on and off AGN feedback, (ii) increasing or decreasing the strength of CC SN feedback, (iii) changing the form of the initial stellar mass function, or (iv) varying the seed mass of SMBHs (e.g. Schaye et al., 2010; Crain et al., 2015; Kugel et al., 2023).

1.3.2 Resolution and volume requirements

Ideally, in any simulation of galaxy formation, one would desire (i) to simulate volumes that are large enough to be statistically representative and (ii) to reach spatial resolutions high enough to resolve most of the baryonic processes that are essential for galaxy formation physics.

- In order to be statistically representative, many applications of galaxy simulations require that the simulated volume covers at least $\sim 50-100$ comoving Mpc per dimension[11], which is of the order of the scale at which the Universe starts to appear statistically homogeneous.
- For the resolution requirements, an example of a small-scale baryonic process that is crucial for galaxy evolution is gas accretion onto an SMBH (see §1.2.7). Given a non-rotating SMBH with a typical mass of $M = 10^7$ $\mathrm{M}_\odot$, the inner edge of its accretion disc can be approximated by the radius of the innermost stable circular orbit,

$$r_{\mathrm{isco}} = 6\frac{GM}{c^2} \approx 3\times10^{-6}\,\mathrm{pc}\left(\frac{M}{10^7\,\mathrm{M}_\odot}\right), \tag{1.20}$$

which, as can be seen, turns out to be more than 10 orders of magnitude smaller than the scale above which the Universe starts to appear homogeneous.

Simulating a cosmological volume of (hundreds of cMpc)^3 at sub-pc resolution is currently not feasible. First, running such a simulation on the largest supercomputer to date would take years in the best-case scenario to reach the redshifts of scientific interest. Second, the number of resolution elements would be so large that the simulation would not fit in the supercomputer's RAM.

At the present time, among the existing cosmological hydrodynamic simulations that have statistically representative volumes ($\sim 100^3$ cMpc^3), have relatively well-resolved galaxy internal structure, and were successfully run to redshift $z = 0$, the most numerically demanding ones are made up of $[(1-2)\times10^3]^3$ resolution elements for baryons and a similar number for DM (e.g. Dubois et al., 2014; Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019). These numbers translate[12] into a mass resolution of $\sim 10^6-10^7$ $\mathrm{M}_\odot$ or, equivalently[13], a spatial resolution of $\sim 10^2-10^3$ pc, which is clearly not sufficient to resolve individual stars, let alone accretion discs around SMBHs.

[11] For example, one of such applications can be to build a representative sample of relatively massive (with $z = 0$ stellar mass of $M_* \gtrsim 10^{10.5}\,\mathrm{M}_\odot$) galaxies to investigate how the star formation rate in these galaxies is tied to the AGN activity of their SMBHs.

[12] The (average) mass of resolution elements representing baryons is obtained by dividing the total baryonic mass contained in the simulated volume (based on the value of Ω_b) by the number of gas resolution elements in the simulation.

[13] See §1.3.4 and equation (1.32) for how mass and spatial resolutions of Lagrangian hydrodynamic schemes can be converted into one another.

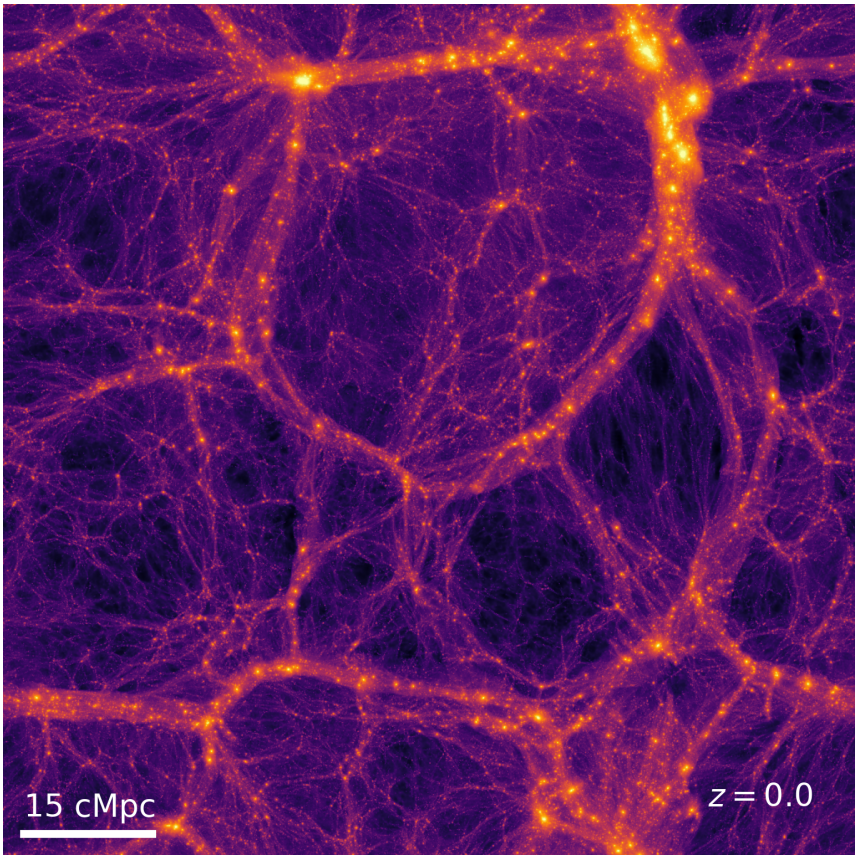

Figure 1.1: Distribution of dark matter in a simulation of galaxy formation at redshift $z = 0$. The size of the volume is $(100\ \mathrm{cMpc})^3$. The colour shows the log of the projected mass surface density of DM in a 10 cMpc-thick slab. Brighter colours correspond to higher densities.

Fig. 1.1 illustrates the typical scales in a modern simulation of galaxy formation; the volume of the simulation is equal to $(100\ \mathrm{cMpc})^3$. Shown is the distribution of DM at redshift $z = 0$, with the colour indicating the log of the projected mass surface density of DM in a slab of 10-cMpc width. Brighter colours correspond to higher surface densities of DM. DM haloes appear as the brightest, yellow spots, moderately bright elongated structures correspond to cosmic filaments, and the darkest regions indicate cosmic voids.

1.3.3 Simulating dark matter using gravity solvers

Because of the lack of high-performance computing machines ~ 40 years ago, the earliest simulations of galaxies did not model gas at all and only solved for the

evolution of DM, which was represented by just $\sim 10^3$ resolution elements (e.g. White, 1976; Aarseth et al., 1979). Since DM is collisionless, evolving DM without (explicitly) accounting for the baryons can tremendously simplify and expedite the calculations, and this was the only feasible approach at that time. Because the spatial distribution and dynamics of the large-scale structure are dictated largely by DM, while the effects of baryons start to dominate only at sub-Mpc scales, DM-only simulations have many important (cosmological) applications despite lacking the baryonic component. Nowadays, DM-only simulations are still popular, but are run in much bigger volumes than ~ 40 years ago (e.g. Springel et al., 2005b; Angulo et al., 2012).

Evolution in time

The evolution of non-relativistic matter that is collisionless and interacts exclusively via gravity is governed by the Poisson equation,

$$\nabla^2 \varphi = 4\pi G \rho \,, \tag{1.21}$$

which relates the gravitational potential of the matter φ to the density of the matter ρ. The solution for a point mass m at some position in space $\boldsymbol{r}_0$ – with the density field given by $\rho(\boldsymbol{r}) = m\,\delta(|\boldsymbol{r} - \boldsymbol{r}_0|)$ where $\delta(\boldsymbol{x})$ is the Dirac delta function – is the well-known, Newtonian potential, $\varphi(\boldsymbol{r}) = -G\,m/|\boldsymbol{r} - \boldsymbol{r}_0|$. To solve an arbitrarily complex gravitational problem, it is thus possible to discretize the mass density field into $N \gg 1$ point-like particles of masses m_i and write down the approximate solution to the Poisson equation as a superposition of the individual particles' potentials,

$$\varphi(\boldsymbol{r}) = \sum_{i=1}^{N} \varphi_i(\boldsymbol{r}) \equiv \sum_{i=1}^{N} \frac{-G\,m_i}{|\boldsymbol{r} - \boldsymbol{r}_i|} \,. \tag{1.22}$$

In order to advance a continuous distribution of collisionless matter, $\rho(\boldsymbol{r})$, in time, one, in general, has to solve the collisionless Boltzmann equation, which is never an easy task. However, for the case of N discrete particles representing the matter field $\rho(\boldsymbol{r})$, the particles' velocities $\boldsymbol{v}_i$ and positions $\boldsymbol{r}_i$ can be updated in a more straightforward way,

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} &= -\displaystyle\sum_{j \neq i}^{N} \nabla_i \varphi_j(\boldsymbol{r}_i) \,, \\ \dfrac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} &= \boldsymbol{v}_i \,. \end{cases} \tag{1.23}$$

Note that by discretizing the original, continuous density field into N particles, we implicitly adopted the Lagrangian approach to solving the equations of mo-

1

tion. Indeed, Eqs. (1.23) are the Euler-Lagrange equations for the Lagrangian

$$\mathcal{L}(\boldsymbol{r},\boldsymbol{v}) = \frac{1}{2}\sum_{i=1}^{N} m_i \boldsymbol{v}_i^2 - \sum_{1\leq i<j\leq N} m_i \varphi_j(\boldsymbol{r}_i). \tag{1.24}$$

There are different ways to time-discretize and time-integrate Eqs. (1.23). The most common ones are the second-order leapfrog integrator and higher-order Hermite integrators (for more details, see e.g. Dehnen & Read, 2011). Crucially, integrating these equations numerically introduces a finite time-step Δt, which characterizes the time resolution of the simulation. On the other hand, the (average) mass of the particles, by which the continuous density field is sampled, characterizes the mass resolution of the simulation (and, indirectly, the simulation's adaptive spatial resolution).

Solving Eqs. (1.23) numerically will inevitably lead to the $1/|\boldsymbol{r}_i - \boldsymbol{r}_j|^2$ divergence when the distance between two interacting particles, $|\boldsymbol{r}_i - \boldsymbol{r}_j|$, becomes very small. To prevent this numerical divergence, a *gravitational softening* parameter $\varepsilon_{\rm soft}$ is added to the gravitational potential sourced by all particles in the system: the exact potential of some particle i, φ_i, is replaced with its softened version,

$$\varphi_{i,\rm soft}(\boldsymbol{r}) = \frac{-G\,m_i}{(|\boldsymbol{r}-\boldsymbol{r}_i|^2 + \varepsilon_{\rm soft}^2)^{1/2}}. \tag{1.25}$$

When $|\boldsymbol{r}-\boldsymbol{r}_i| \gg \varepsilon_{\rm soft}$, $\varphi_{i,\rm soft}(\boldsymbol{r})$ asymptotes to the Newtonian, unsoftened potential, while at $|\boldsymbol{r}-\boldsymbol{r}_i| \ll \varepsilon_{\rm soft}$, $\varphi_{i,\rm soft}(\boldsymbol{r})$ converges to a finite value given by $-G\,m_i/\varepsilon_{\rm soft}$.

Another problem with the above algorithm is that solving Eqs. (1.23) requires $\mathcal{O}(N^2)$ evaluations of the potential (there are N particles in the system and the potential is evaluated $N-1$ times for each particle), which might become a bottleneck if the number of particles N is very large. For this reason, to solve the Poisson equation, modern simulations of galaxy formation employ a more efficient *Fast Multipole Method* (FMM) (Greengard & Rokhlin, 1987), in which the number of evaluations of the potential is reduced from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$ (for more details, see e.g. Schaller et al., 2023).

Boundary conditions

Cosmological simulations of galaxy formation (both DM-only simulations and ones including baryons) are run with periodic boundary conditions. The effects of periodicity are captured when solving the Poisson equation, which is commonly done by splitting the gravitational potential into two components: one for the short-range forces and the other for the long-range ones. The short-range component can be solved using the FMM method as explained above, while the long-range one, which captures the periodic effects, is solved using the Fourier transform. More specifically: (i) the N particles are first projected onto a periodic grid in real space; (ii) the Fast Fourier Transform (FFT) algorithm is used

to construct the density field in Fourier space; (iii) the Poisson equation for long-range forces is then solved in Fourier space, where it has a simple, algebraic form; (iv) finally, an inverse FFT is used to transform the solution of the Poisson equation from Fourier space back to the real space. The FFT algorithm automatically accounts for periodic effects.

Initial conditions

Initial conditions for modern cosmological simulations are generated at high redshift ($z \sim 100$) using a high-order Lagrangian perturbation theory (e.g. Hahn et al., 2021), which is more accurate than the linear Zel'dovich (1970) approximation. At $z \gg 10$, perturbation theory provides a very accurate solution for the evolution of density perturbations because the density contrast remains small (see §1.1.2). Inflation models predict that the post-inflation matter density field can be described by a Gaussian field with high accuracy, or in other words, the properties of the matter density field can be completely characterized by the field's power spectrum in Fourier space, $P(|\boldsymbol{k}|)$, where $\boldsymbol{k}$ is the wave vector. Perturbation theory is used to advance the matter density field to the initial redshift of the simulation. Individual particles are created at this redshift and positioned within the simulation domain such that the power spectrum of the discrete matter field they represent is equal to that predicted by perturbation theory[14]. Perturbation theory is also used to assign peculiar velocities to the created particles. After the simulation has started, the positions and velocities of the particles are continuously updated in time by solving Eqs. (1.23) for a long series of time-steps, until the simulation has reached the desired redshift. Modern simulations of galaxy formation require millions of time-steps to reach $z = 0$.

1.3.4 Simulating baryons using smoothed particle hydrodynamics

Simulating galaxy formation necessitates simulating baryons. Because baryons are collisional, to model the evolution of baryons it is necessary to solve the equations of hydrodynamics alongside the Poisson equation.

The equations of hydrodynamics are the Euler equations for the evolution of the density ρ, velocity $\boldsymbol{v}$, and internal energy u of an ideal gas,

[14] We note that discreetness errors in the power spectrum can be suppressed by perturbing particle masses rather than particle positions, to match $P(|\boldsymbol{k}|)$ predicted by perturbation theory (e.g. Hahn et al., 2021).

1

$$\begin{cases} \dfrac{\mathrm{D}\rho}{\mathrm{D}t} &= -\rho\,(\nabla\cdot\boldsymbol{v}) \\ \dfrac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} &= -\dfrac{\nabla P}{\rho} \\ \dfrac{\mathrm{D}u}{\mathrm{D}t} &= -\dfrac{P}{\rho}(\nabla\cdot\boldsymbol{v}), \end{cases} \tag{1.26}$$

where $\mathrm{D}/\mathrm{D}t = \partial/\partial t + (\boldsymbol{v}\cdot\nabla)$ is the Lagrangian derivative, and P is the gas pressure. To simplify this review, in the system of equations above, we neglected the gravity- and viscosity-related terms as well as the terms with the radiative cooling and heating rates, all of which are important for astrophysical gases and thus need to be accounted for in numerical simulations that evolve baryons. Without these extra terms, the entropy of the system does not change, and the internal energy u and pressure P are related as

$$P = A\,\rho^{\gamma} = (\gamma-1)\,\rho\,u, \tag{1.27}$$

where A is the adiabatic constant and γ is the adiabatic index.

One of the popular numerical methods to solve the equations of hydrodynamics is *smoothed particle hydrodynamics*[15] (SPH; for reviews, see e.g. Monaghan, 2005; Springel, 2010a). Similar to gravity solvers, the astrophysical gas is modelled as a collection of discrete particles. These gas particles represent fluid elements of constant mass[16] and move along with the fluid, making SPH a Lagrangian method. However, unlike gravity solvers, hydrodynamics requires continuity in order to estimate the density, velocity and pressure gradients that enter the hydrodynamic equations. The SPH methods handle this by introducing the kernel function $W(\boldsymbol{r}, h)$, which allows 'interpolating' any field stored in SPH particles to any point in space. The kernel function W needs to be normalized to unity, be spherically symmetric, and monotonically decrease with increasing distance from its centre. The simplest example of a kernel function is a Gaussian: $W(|\boldsymbol{r}|, h) = 1/(\sqrt{\pi}h)^3 \exp(-|\boldsymbol{r}|^2/h^2)$. The parameter h defines the characteristic size of the kernel and is called the *smoothing length*.

By applying an SPH kernel to a discrete set of particles, it becomes straightforward to compute the density at any location in the simulation $\boldsymbol{r}$, which is done as follows,

[15] Although in this thesis we exclusively use SPH, we note that there are good alternatives to SPH such as meshless solvers (e.g. Hopkins, 2015) and moving-mesh solvers (e.g. Springel, 2010b). In the case of mesh codes, the gas is discretized into cells, as opposed to particles used in SPH.

[16] The mass of individual fluid elements is constant in time only in the absence of astrophysical processes that redistribute the mass in the simulation between gas, stars and/or BHs, such as stellar chemical enrichment and gas accretion onto SMBHs.

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} m_i \, W(|\boldsymbol{r} - \boldsymbol{r}_i|, h), \tag{1.28}$$

where the sum is computed over the SPH particles representing the gas, which have positions $\boldsymbol{r}_i$ and masses m_i. Although correct, summing over all particles in the simulation in Eq. (1.28) would be numerically inefficient when N is very large. The solution is to choose an SPH kernel function $W(|\boldsymbol{r}|, h)$ that has finite support (i.e. $W(|\boldsymbol{r}|, h)$ becomes zero for sufficiently large but finite $|\boldsymbol{r}|$). Numerical simulations of galaxy formation usually adopt such SPH parameters that each gas particle 'sees' on average the ~ 50 gas particles that are closest to it[17].

Interpolation of any quantity A other than density (e.g. gas pressure or velocity) can be constructed with the help of the density field,

$$A(\boldsymbol{r}) = \sum_{i} A_i \frac{m_i}{\rho_i} \, W(|\boldsymbol{r} - \boldsymbol{r}_i|, h), \tag{1.29}$$

where we dropped the summation limits and introduced the short-hand notation $\rho_i = \rho(\boldsymbol{r}_i)$ and $A_i = A(\boldsymbol{r}_i)$. Using the interpolated versions of the gas density, velocity, and pressure, one can show that in SPH, the system of equations (1.26) acquires the form

$$\begin{cases} \dfrac{\mathrm{D}\rho_i}{\mathrm{D}t} &= \displaystyle\sum_j m_j (\boldsymbol{v}_i - \boldsymbol{v}_j) \cdot \nabla_i \, W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h) \\ \dfrac{\mathrm{D}\boldsymbol{v}_i}{\mathrm{D}t} &= -\displaystyle\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \nabla_i \, W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h) \\ \dfrac{\mathrm{D}u_i}{\mathrm{D}t} &= \dfrac{P_i}{\rho_i^2} \displaystyle\sum_j m_j (\boldsymbol{v}_i - \boldsymbol{v}_j) \cdot \nabla_i \, W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h), \end{cases} \tag{1.30}$$

which, despite discretization, exactly conserves mass, momentum, and energy. The conservation properties hold thanks to the Lagrangian nature of SPH, and are one of the attractive features of the method.

The above SPH equations are written for the case of constant entropy, whereas in reality entropy may increase – e.g. within shocks produced by SN explosions. In other words, Eqs. (1.30) are unable to capture the effects of shocks, which are of high importance in astrophysics. To circumvent this problem, an artificial viscosity term is added to the SPH equations for momentum and energy (Monaghan & Gingold, 1983; Monaghan, 1992, 1997), by means of which kinetic energy can be converted into heat, thus increasing the entropy. However, the added artificial viscosity has to vanish outside shocks or any other natural dissipative processes,

[17] The gas particles that are within an SPH kernel of some other particle i, are often referred to as '(SPH) neighbours of particle i'.

1

so that the entropy does not rise where there is no need for it. In SPH, this is realized by a switch function that makes use of the gas velocity divergence and curl to control the value of the artificial viscosity (for more details, see e.g. Cullen & Dehnen, 2010).

Eqs. (1.30) assume that the smoothing length h is the same for all particles in the system. This is a very restrictive requirement and numerical simulations abandoned it a long time ago. Instead, they require that each SPH particle's kernel contains the same mass. If the desired mass to be inside the kernel is $\langle M_{\rm ngb}\rangle$ – where the subscript 'ngb' stands for the 'neighbouring gas particles' that constitute this mass – then the smoothing length of some particle i, h_i, is related to $\langle M_{\rm ngb}\rangle$ as

$$\frac{4\pi\rho_i h_i^3}{3} = \langle M_{\rm ngb}\rangle, \tag{1.31}$$

where ρ_i is the density estimated within the particle's kernel. This constraint yields naturally adaptive spatial resolution in SPH, which increases (decreases) in regions of higher (lower) density. The smoothing length becomes a function of density and can vary from particle to particle, which introduces extra terms to Eqs. (1.30) that depend on the smoothing length and density (e.g. Price, 2012).

One can employ equation (1.31) to relate the (fixed) mass resolution and (adaptive) spatial resolution of an SPH scheme. The mass resolution is given by the average particle mass in the simulation, $\langle m\rangle$, while the spatial resolution can be approximated by the size of the smoothing length h. If gas particles have on average $\langle N_{\rm ngb}\rangle$ gas neighbours inside their kernels, then the desired mass within the kernel, $\langle M_{\rm ngb}\rangle$, and the average gas particle mass, $\langle m\rangle$, are related as $\langle m\rangle = \langle M_{\rm ngb}\rangle/\langle N_{\rm ngb}\rangle$. Then, at a given density ρ, the spatial resolution h can be calculated from the mass resolution $\langle m\rangle$ as,

$$h = \left(\frac{3}{4\pi}\frac{\langle N_{\rm ngb}\rangle\langle m\rangle}{\rho}\right)^{\frac{1}{3}} = 0.36\,{\rm kpc}\left(\frac{\langle m\rangle}{10^6\,{\rm M}_\odot}\right)^{\frac{1}{3}}\left(\frac{\langle N_{\rm ngb}\rangle}{48}\right)^{\frac{1}{3}}\left(\frac{\rho}{10\,m_{\rm p}{\rm cm}^{-3}}\right)^{-\frac{1}{3}}. \tag{1.32}$$

When integrating the system of equations (1.30) forward in time, it is essential to choose an appropriate size of the time-step Δt. The choice is made based on the Courant-Friedrichs-Lewy (CFL) condition, which limits the time-step Δt to be no greater than

$$\Delta t_{\rm CFL} = C_{\rm CFL}\frac{h}{c_{\rm sound}}, \tag{1.33}$$

where $c_{\rm sound}$ is the speed of sound[18] and $C_{\rm CFL}$ is a dimensionless parameter, which in most applications of SPH is of the order of 0.1. Physically, limiting

[18] In practice, in SPH schemes, $c_{\rm sound}$ is replaced with the signal velocity, which is a linear combination of $c_{\rm sound}$ and the relative (bulk) velocity between the particle to which the CFL criterion is applied and its SPH neighbours.

the time-step Δt by $\Delta t_{\rm CFL}$ means that a sound wave takes $1/C_{\rm CFL} \sim 10$ time-steps to travel across the particle kernel. This ensures that the hydrodynamical reaction of the gas to any external or internal disturbance is always well-resolved in time, which is crucial for the numerical stability and accuracy of the simulation.

1.4 Subgrid physics in simulations of galaxy formation

Modelling galaxy formation requires the implementation of a list of essential astrophysical processes on top of the hydrodynamics and gravity in the cosmological framework. Given the typical spatial resolution of state-of-the-art galaxy simulations (~ 0.5 kpc), many of these processes will be unresolved. Unresolved processes are implemented using so-called *subgrid* models. The goal of a subgrid model applied to a certain astrophysical process is to reproduce its effects on larger, resolved scales, even though the process itself may not be (adequately) resolved in the simulation. An example of a process modelled in the subgrid way is star formation, which occurs on unresolved, sub-pc scales but results in the distribution of stellar mass in massive haloes that may extend over tens of kpc. Subgrid models can be constructed either empirically or from first principles, and often depend on a number of free parameters that account for the numerical effects and need to be calibrated to produce a realistic galaxy population (e.g. Kugel et al., 2023).

An example of a galaxy produced by a simulation with a calibrated subgrid model is given in Fig. 1.2. The figure shows a simulated Milky Way-mass galaxy at $z = 0$, which is taken from the high-resolution simulation with the COLIBRE galaxy formation model. The mass resolution for gas is $m_{\rm gas} = 2.3 \times 10^5$ M$_\odot$. The galaxy resides in a $M_{200} = 1.4 \times 10^{12}$-M$_\odot$ halo and its stellar mass and stellar half-mass radius are equal to, respectively, $M_{\rm star} = 2.9 \times 10^{10}$ M$_\odot$ and 6 kpc. The figure displays the mass-weighted gas temperature (left column), gas mass surface density (middle column), and stellar mass surface density (right column). The top row shows the distributions of the three fields around and within the galactic halo, whose virial radius is indicated by the white dashed circle, while the bottom row shows a zoom-in onto the galaxy, which is located in the centre of the top panels. All fields are calculated in a slab with a width of 2 Mpc (top row) and 60 kpc (bottom row). The galaxy exhibits a prominent gaseous disc with spiral arms. The distribution of the stellar mass appears round and smooth because it is based on stellar particles of all ages, both young and old.

The implementation of the physical processes that are essential for simulating galaxies is described next.

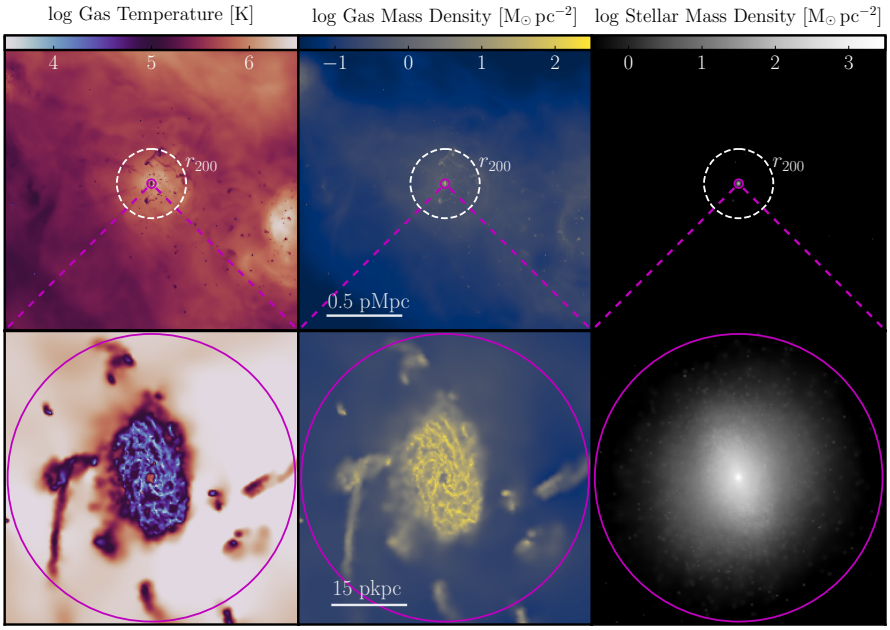

Figure 1.2: A Milky Way-mass galaxy at redshift $z = 0$ in a simulation of galaxy formation with a mass resolution of 2.3×10^5 $M_\odot$. Shown are the mass-weighted gas temperature (*left column*), gas mass surface density (*middle column*), and stellar mass surface density (*right column*). All fields are computed in a slab with a width of 2 Mpc (*top row*) and 60 kpc (*bottom row*). The bottom row shows a zoom-in region from the central part of the image in the top row, as illustrated by the pink circle and dashed lines. The white dashed circle in the top row indicates the virial radius of the halo. See text for details.

1.4.1 Simulating radiative cooling and heating processes

The prescription for radiative cooling and heating incorporates *fundamental* atomic and molecular processes. These do not depend on the numerical resolution and are implemented exactly as predicted by the theory. Specifically,

- The primary cooling channels that are followed in modern numerical simulations include (i) thermal bremsstrahlung, which has the dominant contribution to the cooling rates at temperatures $T \gtrsim 10^7$ K; (ii) collisional ionization (followed by radiative recombination) and excitation of both primordial elements and metals, which are important at intermediate temperatures, $10^4 \lesssim T \lesssim 10^7$ K; (iii) inverse Compton cooling off the CMB photons, which becomes relevant in gases that are ionized; and – if the simulation includes the gas in the cold phase – collisionally excited rotational and vi-

brational levels of molecules, such as H_2O and CO, which start to dominate at temperatures below $\sim 10^3$ K.

- The total heating rate includes contributions from processes such as (i) photoionization heating from the cosmic UV and X-ray background, (ii) photoelectric heating on dust grains, (iii) heating due to photodissociation of molecules, and (iv) heating by cosmic rays.

The total cooling and heating rates due to these processes are added to the right-hand side of the energy and momentum equations (1.30); the cooling (heating) as a consequence of adiabatic expansion (contraction) of gas is already incorporated into these hydrodynamic equations; and the heating due to hydrodynamic shocks is captured via the extra terms that depend on the artificial viscosity, which also need to be added to Eqs. (1.30). The gas is normally assumed to be in ionization equilibrium in the presence of a redshift-dependent, spatially uniform X-ray and UV cosmic background that is expected from active galactic nuclei and star-forming galaxies (e.g. Faucher-Giguère, 2020). Under the assumption of ionization equilibrium, the ionization degrees of individual elements are computed using algebraic equations, as opposed to solving the taxing system of partial differential equations.

In modern galaxy simulations, it is becoming customary to compute the cooling and heating rates based on individual elements' abundances. This approach is, unsurprisingly, more accurate than scaling the rates with only the mass fractions of hydrogen, helium, and the total metallicity of the gas. Modern simulations trace $10-15$ individual elements whose contributions to the cooling rates are most significant. These include H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe. The rates are calculated prior to running the simulation, by using a photoionization code. One of the most widely used photoionization codes is CLOUDY (Ferland et al., 2017), which incorporates a rich number of atomic and molecular processes and can compute the ionization, chemical, and thermal state of a gas cloud under various initial and boundary conditions. The standard input to CLOUDY includes gas density, composition, and, if present, the spectrum of the radiation background. By using CLOUDY, one can tabulate the cooling and heating rates of the gas as a function of density, temperature, metallicity, and redshift (e.g. Wiersma et al., 2009a; Ploeckinger & Schaye, 2020). The resulting tables are then loaded and interpolated at the start of the simulation. During the simulation, the correct rate is fetched for the given values of the four input parameters. Nearly all simulations of galaxy formation make use of pre-computed, look-up cooling tables (e.g. Schaye et al., 2015; Pillepich et al., 2018; Hopkins et al., 2018b; Davé et al., 2019) because of the high numerical efficiency combined with reasonable accuracy of this method.

While most simulations of galaxy formation follow more-or-less the same approach to modelling gas cooling rates at temperatures above $T \sim 10^4$ K, gas cool-

1

ing below $T \sim 10^4$ K may differ dramatically. The reason for this is twofold. First, modelling the cold phase requires incorporating a greater number of physical processes, among which are the (self-)shielding of gas, fine structure-line cooling, the formation of molecules and their effect on the cooling function, and the formation and evolution of dust and its interaction with the gas phase. Second, the cold phase is very dense, which leads to highly non-linear gas dynamics and very short time-steps reached in the simulation, resulting in a much greater computational expense. For these reasons, the developers of many previous cosmological simulations opted to *not* explicitly model the cold, dense gas. Instead, they imposed a so-called effective equation of state, $P(\rho) \propto \rho^{\gamma}$, where P and ρ are, respectively, the gas pressure and density, and γ is the polytropic index, usually set to 4/3 (see e.g. Schaye & Dalla Vecchia, 2008, for details). The justification of this approach is that physically, one expects cold, dense clouds to be in (local) pressure equilibrium with the hot, volume-filling phase of the ISM (e.g. Springel & Hernquist, 2003), so that the average ISM pressure can be approximated as a function of density only. Only recently have attempts been made to abandon the effective equation of state and begin modelling the gas cold phase more directly (e.g. Dubois et al., 2021; Feldmann et al., 2023).

1.4.2 Simulating star formation

The current state-of-the-art cosmological simulations lack the numerical resolution to follow the fragmentation of GMCs down to small scales at which stars are expected to form, $\lesssim 0.1$ kpc. This implies that the physics taking place on these scales is not resolved either. In order words, the process of star formation in a numerical simulation has to be implemented as a subgrid model.

Resolution elements representing stellar mass – which are standardly referred to as *star particles* or *stellar particles* – normally have the same mass as the gas particles. Clearly, a stellar particle with an (initial) mass of 10^6 $M_\odot$, which is typical for simulations of galaxy formation, is far more massive than the most massive stars in the Universe ($\lesssim 10^3$ $M_\odot$). Therefore, such a stellar particle physically represents a population of stars, as opposed to an individual star. In most galaxy simulations, a stellar particle is characterized by a single age, single metallicity, and initial stellar mass function. While ages and metallicities differ from particle to particle, the initial stellar mass function is in most cases assumed to be universal and is thus fixed throughout the simulation.

The numerical algorithm for star formation can be split into three parts: (i) a criterion for star formation that determines which gas in the simulation is eligible to be star-forming, (ii) how gas resolution elements that are star-forming compute their star formation rates, and (iii) how stellar particles are created from the star-forming gas.

(i) Physically, only cold and dense enough gas should be star-forming. So the

most basic criterion for star formation can be constructed by considering gas resolution elements whose temperatures are below a certain threshold temperature (e.g. $T \sim 10^3$ K) and/or whose densities are above some threshold density (e.g. $\rho \sim 10^2\, m_{\rm p}{\rm cm}^{-3}$). In the case of a threshold density, one also needs to keep in mind that the matter density of the Universe evolves with redshift as $(1+z)^3$. Thus, in order to not have star formation in intergalactic gas at $z \gg 10$, which would be unphysical, an additional constraint should be added, ensuring that only gas elements with sufficient density contrast (relative to the critical density of the Universe) can be star-forming. Other common classes of star-formation criteria include the gravitational instability criterion, in which gas elements become star-forming if half of (the absolute value of) their gravitational potential energy is greater than their kinetic energy due to thermal and turbulent motions; and the Jeans-instability criteria, in which a gas element is labelled as 'star-forming' if its mass is greater than the (thermal) Jeans mass given by Eq. (1.17).

(ii) As long as a gas element satisfies the star-formation criterion, it is assigned a star-formation rate. The most widely used algorithm to compute the star formation rate of a gas element is through the Schmidt (1959) law

$$\dot{m}_{\rm sf} = \varepsilon_{\rm sf} \frac{m_{\rm gas}}{t_{\rm ff}}\,, \tag{1.34}$$

in which $m_{\rm gas}$ is the mass of the gas element, the free-fall time $t_{\rm ff}$ is given by Eq. (1.18), and $\varepsilon_{\rm sf}$ is the star formation efficiency parameter. Observations of star formation in GMCs suggest that $\varepsilon_{\rm sf}$ is of the order of 1 per cent (e.g. Krumholz & Tan, 2007; Lee et al., 2016), so in numerical simulations $\varepsilon_{\rm sf} = 0.01$ is often the choice.

(iii) Owing to low resolution, many numerical simulations adopt a stochastic approach to creating stellar particles from the gas. One of the ways to implement stochastic star formation is as follows: given a star formation rate of a gas element, $\dot{m}_{\rm sf}$, and time-step Δt, one computes the probability that the gas element will be converted into a stellar particle in the time-step Δt as

$${\rm prob}_{\rm sf} = \min\left(1, \frac{\dot{m}_{\rm sf}}{m_{\rm gas}}\Delta t\right), \tag{1.35}$$

where $m_{\rm gas}$ is the mass of the gas element. After the value of ${\rm prob}_{\rm sf}$ has been calculated, a random number is drawn from a uniform distribution between 0 and 1, and if it is smaller than ${\rm prob}_{\rm sf}$, then a stellar particle is formed with an initial mass equal to $m_{\rm gas}$, while the parent gas element is removed from the simulation domain. The stellar particle will have the same position, peculiar velocity, and composition as the parent gas element. Note that since ${\rm prob}_{\rm sf} \ll 1$ for any reasonable time-step, the chance of a gas

1

element being converted into a stellar particle is very small at any given time-step, but gets higher at higher densities ρ because the free-fall time-scale scales with density as $t_{\rm ff} \propto 1/\sqrt{\rho}$.

1.4.3 Simulating stellar mass loss

Models for chemical enrichment determine how much mass of various chemical elements are produced by stars and what fraction of this mass is returned to the ISM.

Numerically, chemical enrichment is realized by specifying how the mass ejected by stellar particles depends on their internal properties, and how the ejected mass is subsequently distributed within the gas surrounding the stellar particles. The usual approach to determining the amount of ejected mass is through look-up tables, which is similar to how numerical simulations deal with gas cooling/heating rates. The tables for chemical enrichment provide net stellar yields per unit stellar mass for all individual elements that are traced in the simulation. These yields depend on the stellar age and metallicity. The tables themselves are generated based on separate, detailed calculations of the evolution of single stars for various initial masses and metallicities (e.g. Fishlock et al., 2014).

The tables from modern galaxy simulations normally include three channels of chemical enrichment:

- Chemical enrichment due to stellar mass loss in the asymptotic giant branch phase, which is relevant for stars with initial masses between approximately 1 and 8 $M_\odot$.
- Primordial elements and metals produced and ejected by massive stars ($\gtrsim 8$ $M_\odot$) in stellar winds and CC SNe.
- Primordial elements and metals produced and ejected in the thermonuclear collapse of white dwarfs: type-Ia SNe. The initial masses of stars that become white dwarfs are $\lesssim 8$ $M_\odot$.

For more details, see e.g. Wiersma et al. (2009b) and references therein.

At the start of a simulation, the tables are loaded and interpolated across stellar age and metallicity, and adjusted for the initial stellar mass function that is used in the simulation. For a stellar particle with a given age and metallicity at some time-step Δt, the tables determine how much mass this stellar particle will release in this time-step. The ejected mass is distributed among the $N_{\rm ngb}$ gas particles closest to the stellar particle. The value of $N_{\rm ngb}$ may vary strongly from simulation to simulation, depending on the scheme used to solve the gas hydrodynamics equations. For SPH methods (see §1.3.4), the typical average value of

$N_{\rm ngb}$ is between ≈ 50 and 100, depending on the form and extent of the SPH kernel (e.g. Price, 2012). If the masses of some chemical element 'elem' (e.g. hydrogen, carbon, oxygen, or iron) carried by the $N_{\rm ngb}$ closest gas particles are $m_{{\rm elem},i}$ at the beginning of the time-step, and the star particle releases some mass $\Delta m_{\rm elem}$ of this element, then during the time-step the masses $m_{{\rm elem},i}$ can be updated as

$$m'_{{\rm elem},i} = m_{{\rm elem},i} + \Delta m_{\rm elem} \frac{W(|\boldsymbol{r}_i|, h)}{\sum_{k=1}^{N_{\rm ngb}} W(|\boldsymbol{r}_k|, h)}, \tag{1.36}$$

where i is the index of the gas neighbour, W is the SPH kernel positioned at the stellar particle, $\boldsymbol{r}_i$ indicates the distance between the gas particle i and the stellar particle, and h is the SPH smoothing length carried by the stellar particle[19].

Note that in Eq. (1.36), we split the ejected mass $\Delta m_{\rm elem}$ among the gas neighbours using the weights $w_i = W(|\boldsymbol{r}_i|, h)/\sum_{k=1}^{N_{\rm ngb}} W(|\boldsymbol{r}_k|, h)$ that have the property $\sum_{k=1}^{N_{\rm ngb}} w_k = 1$, ensuring that the ejected mass is conserved. Although all numerical simulations tend to normalize the weights such that the mass in chemical enrichment is conserved (i.e. if a stellar particle loses Δm of its mass, the mass of the ambient ISM should increase by Δm), the exact form of the weights may differ between different simulations.

1.4.4 Simulating energy feedback from core-collapse supernovae

CC SNe release on average $\approx 10^{51}$ erg of energy in one event (e.g. Koo et al., 2020), which for a reasonable initial stellar mass function can be translated to $\approx 10^{51}$ erg per every ~ 100 $\mathrm{M}_\odot$ of stellar mass.

By assuming that each CC SN in a simulation releases 10^{51} erg of energy and all stars with zero-age main-sequence masses from $m_{\rm min}$ to $m_{\rm max}$ explode as CC SNe, one can calculate the total SN energy released by a stellar particle of initial mass m_* as

$$E_{\rm SN,tot} = 10^{51}\,{\rm erg}\, m_* \int_{m_{\rm min}}^{m_{\rm max}} \Phi(m)\,{\rm d}m, \tag{1.37}$$

where $\Phi(m)$ is the initial stellar mass function. For the Chabrier (2003) initial stellar mass function integrated within $(m_{\rm min}, m_{\rm max}) = (8, 100)$ $\mathrm{M}_\odot$, the total CC SN energy budget for each stellar particle is

$$E_{\rm SN,tot} = 1.18 \times 10^{55}\,{\rm erg} \left(\frac{m_*}{10^6\,\mathrm{M}_\odot}\right). \tag{1.38}$$

Each stellar particle in the simulation can then inject this energy into the gas surrounding it, which can be carried out either continuously (in multiple time-steps)

[19] Although stellar particles do not take part in hydrodynamics, they may still compute their SPH smoothing lengths based on the distribution of gas around them. The same algorithm is also suitable for BH particles, which we will discuss in §1.4.6.

1

or all at once (in a single time-step). The times when the SN energy injection takes place are counted from the moment when the star particle is born and normally fall into the time-interval from ≈ 3 to ≈ 40 Myr corresponding to the average lifetimes of $\sim 10^2$- and 10-$M_\odot$ stars, respectively. The SN energy can be injected either in thermal form (by increasing the internal energies of the gas particles), in kinetic form (by modifying the peculiar velocities of the gas particles), or both (e.g. Smith et al., 2018).

It might be tempting to distribute $E_{\rm SN,tot}$ among the gas neighbours in a way similar to how it is often done for chemical enrichment, using a weighting scheme similar to the one from Eq. (1.36). However, if the energy is injected thermally, at the typical resolution of galaxy simulations, this approach will lead to very inefficient SN feedback. This is because the injected energy will be smoothed over too much gas mass, leading to only a modest temperature increase and, consequently, a cooling rate that is unrealistically high (e.g. Dalla Vecchia & Schaye, 2012). As a consequence, the galaxies will form too much stellar mass and their morphologies will be implausible. In the literature, the aforementioned, numerically induced weakness of SN feedback is referred to as the 'overcooling problem' (e.g. Katz et al., 1996). The commonly used solutions are to inject (a part of) the SN energy kinetically (e.g. Hopkins et al., 2018a), distribute the energy within fewer gas neighbours, and/or inject the energy with a lower frequency in time but a higher amount per injection event (e.g. Dalla Vecchia & Schaye, 2012)

Additionally, many simulations boost the energy given by Eq. (1.38) by a factor between 1 and 4 in order to improve the agreement with observed galaxy properties such as the galaxy stellar mass function (e.g. Roca-Fàbrega et al., 2021). The use of the boost can be seen as a way to account for (i) the high energy losses due to enhanced cooling rates that are unavoidable at the typical resolution of galaxy simulations, (ii) the lack or underestimation of the effects of clustering of individual SNe in space and time, which otherwise – if well-resolved – would boost the SN momentum output (e.g. Gentry et al., 2017), and (iii) the presence of more-energetic hypernovae whose kinetic energies can be ~ 10 times greater than in normal CC SNe.

1.4.5 Simulating other forms of feedback from stars

The numerical implementation of the other forms of stellar feedback, such as feedback from type-Ia SNe, stellar winds, and radiation pressure, is not very different from that of CC SN feedback.

Namely, for a given feedback process, one derives how much energy and/or momentum should be released per unit mass and per unit time based on some theoretical and observational constraints (e.g. Gatto et al., 2017; Keller et al., 2022). The specific energy and/or momentum rate from the feedback process, which in general is a function of stellar age, metallicity, and initial stellar mass

function, is then scaled with the mass of the stellar particle that does the feedback (see Eq. 1.38 for how this is done for CC SN feedback), and with the size of the time-step during which the feedback is taking place. If the feedback is momentum-driven, the resulting momentum is injected into the gas phase by 'kicking' gas elements that are close to the stellar particle (i.e. modifying their peculiar velocities). If the feedback is energy-driven, then the choice has to be made whether the resulting energy is injected thermally and/or kinetically, similarly to how it is handled in the numerical implementation of CC SN feedback.

1.4.6 Seeding of supermassive black holes in galaxy simulations

In numerical galaxy formation, SMBHs are represented by black hole particles. The most common algorithm for creating (seeding) BH particles in galaxy simulations can be described as follows. At a given redshift z,

(i) Identify all haloes in the simulation that contain some minimum number of particles using the friends-of-friends finder[20];

(ii) For each identified halo whose mass is greater than a certain threshold mass (usually in the range from $\sim 10^8$ to 10^{11} $M_\odot$), find the gas resolution element with the deepest gravitational potential (or highest density);

(iii) If the halo does not yet harbour a BH particle, convert the gas element identified in step (ii) into a BH particle.

BH particles inherit most properties of their parent gas elements, including position, peculiar velocity, and mass. However, the inherited mass is used exclusively in gravity calculations, while the processes that are internal to the BHs, such as gas accretion onto BHs, are coupled to a different, initially much lower mass, which is called the *subgrid mass* (e.g. Springel et al., 2005c; Booth & Schaye, 2009). Using the subgrid mass in addition to the inherited gas-particle mass opens the possibility of modelling the initial stages of BH evolution even when the resolution of the simulation corresponds to the masses of SMBHs ($\gtrsim 10^6$ $M_\odot$). In other words, without a BH subgrid mass, all BH particles would become supermassive as soon as they are seeded, which would be unrealistic. The initial value of the BH subgrid mass is called the *seed mass*. In most simulations, the seed mass is a free parameter that needs to be calibrated such that the AGN feedback in the simulation is neither too strong nor too weak.

1.4.7 Simulating mass growth of supermassive black holes

BHs are collisionless particles: if we neglect AGN feedback, BH particles interact with the gas and stars exclusively via gravity. Once BHs have been seeded, they

[20] The friends-of-friends finder is a commonly used tool for identifying groups of particles representing haloes and galaxies in a numerical simulation.

1

start growing in mass. There are two primary channels through which BHs can acquire more mass. The first one is merging with other BH particles. Two BH particles can merge into a new, single BH particle if certain criteria are satisfied. The two common criteria that need to be satisfied simultaneously are:

- The separation between the two merging BH particles needs to be smaller than a certain threshold distance. The threshold distance is usually set to be of the order of the gravitational softening length (see §1.3.3 for details on the softening length).
- The smaller of the two BH particles has to be under a sufficiently strong gravitational influence of the more massive particle. This can be confirmed by comparing the relative velocity between the two BHs to the escape velocity from the larger BH, evaluated at the separation of the two particles.

When two BH particles merge, the properties of the more massive particle are updated, while the less massive particle is removed from the simulation.

The other channel of BH growth is the accretion of gas. BH particles accrete matter based on their instantaneous accretion rates, which are computed using the properties of the gas particles near the BHs. One of the standard ways to compute the accretion rate is to use the Bondi-Hoyle-Lyttleton formula,

$$\dot{M}_{\rm BH} = 4\pi G^2 \frac{M_{\rm BH}^2 \rho_{\rm gas}}{(v^2 + c_{\rm sound}^2)^{3/2}}, \tag{1.39}$$

where $M_{\rm BH}$ is the (subgrid) mass of the accreting BH particle, $\rho_{\rm gas}$ is the mass density of the gas around the BH, v is the relative velocity between the gas and BH, and $c_{\rm sound}$ is the local speed of sound. The above equation assumes that the gas flow onto an SMBH is spherically-symmetric, adiabatic, and in a steady state, and that the BH itself is a point-like particle. Some simulations in the past used to multiply the rate given by Eq. (1.39) by an ad-hoc factor $\alpha \gg 1$ (e.g. Springel et al., 2005a; Booth & Schaye, 2009; Vogelsberger et al., 2014), motivating it by the need to compensate for their low resolution, which did not allow to resolve the cold, high-density gas that BHs are expected to accrete.

For a given mass accretion rate, $\dot{M}_{\rm BH}$, one can compute the AGN bolometric luminosity, $L_{\rm bol}$, if the radiative efficiency parameter $\varepsilon_{\rm r}$ is known,

$$L_{\rm bol} = \varepsilon_{\rm r} \dot{M}_{\rm BH} c^2. \tag{1.40}$$

Physically, $\varepsilon_{\rm r}$ denotes the fraction of the accreted matter that escapes the gas surrounding the BH as radiation – as opposed to 'being swallowed' by the BH. The luminosity $L_{\rm bol}$ can be compared to the Eddington luminosity,

$$L_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p} c}{\sigma_{\rm T}}, \tag{1.41}$$

which is the maximum luminosity the BH can attain assuming a spherically-symmetric accretion, and where σ_T is the Thompson cross section for electron scattering. For L_{bol} greater than L_{Edd}, the radiation pressure on the accreted gas from the photons escaping the BH would be stronger than the gravitational force driving the gas inwards, toward the BH, which would slow down the accretion and reduce L_{bol} back to L_{Edd}. For this reason, many simulations do not allow the BH accretion rate, $\dot{M}_{BH}$, given by Eq. (1.39) to exceed the accretion at the Eddington luminosity,

$$\dot{M}_{Edd} = \frac{L_{Edd}}{\varepsilon_r c^2} = \frac{4\pi G M_{BH} m_p}{\varepsilon_r \sigma_T c}. \tag{1.42}$$

To guarantee that BH particles have the correct mass growth, they need to reside close to the minimum of the gravitational potential of the host galaxy. In the real Universe, this can happen naturally because SMBHs experience a gravitational drag force (also commonly referred to as 'dynamical friction') as they traverse the medium filled with background particles of much smaller mass (such as stars and putative DM), which causes the SMBHs to lose their energy and spiral in towards the centre of the galaxy (e.g. Ostriker, 1999). However, at the typical resolution of galaxy simulations ($\sim 10^6\,M_\odot$), the masses of many SMBHs will be comparable to the masses of background particles, which will strongly suppress the effects of dynamical friction. The common solution to help BH particles migrate towards the minimum of the galaxy's potential is as follows: every time-step, each BH particle in the simulation identifies the neighbouring particle within its kernel that has the lowest gravitational potential and repositions on top of this particle if the particle's potential is lower than at the location of the SMBH (e.g. Springel et al., 2005c; Booth & Schaye, 2009; Bahé et al., 2022).

1.4.8 Simulating AGN feedback

From both observational and theoretical perspectives, it is customary to separate energy feedback from BHs into two classes: the quasar (also known as 'thermal' and 'radiative') mode and radio (also known as 'kinetic' and 'jet') mode (e.g. Russell et al., 2013; Mocz et al., 2013). The quasar mode manifests itself when the BH is undergoing rapid growth by accreting gas at a high rate ($\gtrsim 0.01\dot{M}_{Edd}$), while the kinetic jet mode features at low accretion rates ($\lesssim 0.01\dot{M}_{Edd}$). Many numerical simulations implement both modes of AGN feedback (e.g. Vogelsberger et al., 2013; Dubois et al., 2014; Kaviraj et al., 2017; Pillepich et al., 2018; Davé et al., 2019), but some limit the AGN feedback to just one mode (e.g. Schaye et al., 2015, 2023), motivating it by the lack of numerical resolution to distinguish between the different modes and by the aim to reduce the overall complexity of the subgrid model for AGN feedback.

1

Theoretical models predict that at high accretion rates, the accretion disc around the BH can be approximated by a geometrically thin, optically thick disc that efficiently dissipates its energy in the form of radiation, with the radiative efficiency $\varepsilon_{\rm r}$ being of the order of 10 per cent (Shakura & Sunyaev, 1973). The radiation emitted by the disc exerts pressure on the nearby gas and dust, which can drive a strong wind. The wind will counteract the further supply of gas onto the BH, leading to self-regulation of the mass growth of the BH. On galactic scales, AGN-driven winds are observed to have velocities as high as a few 10^3 km $\rm s^{-1}$ in different phases (e.g. Rupke & Veilleux, 2013; Speranza et al., 2022); they can propagate to distances far beyond the host galaxy, ejecting the gas therefrom. When powerful enough, they can not only disrupt the ambient ISM but also quench the star formation in the host galaxy altogether.

Numerically, the quasar mode of AGN feedback can be implemented as follows. If a BH particle has a time-step Δt and its mass accretion rate is $\dot{M}_{\rm BH}$, then the energy that will be received by the gas around the BH in this time-step is

$$\Delta E_{\rm AGN} = \varepsilon_{\rm f} L_{\rm bol} \Delta t\,, \tag{1.43}$$

where the AGN bolometric luminosity $L_{\rm bol}$ is computed from $\dot{M}_{\rm BH}$ using Eq. (1.40) with a reasonable value of the radiative efficiency in the quasar mode (typically $\varepsilon_{\rm r} \sim 0.1$). The parameter $\varepsilon_{\rm f}$ indicates the fraction of the energy radiated around the SMBH that is coupled to the gas. In most cases, the coupling efficiency $\varepsilon_{\rm f}$ is set to around 0.1, which provides a good match to the observed scaling between masses of SMBHs and masses (or velocity dispersion) of stellar bulges (e.g. Booth & Schaye, 2009). Similarly to the implementation of energy feedback from CC SNe, numerical implementations of the radiative mode of AGN feedback have to assume how the AGN energy, $\Delta E_{\rm AGN}$, is distributed within the gas phase. One of the common approaches is to inject the energy thermally by heating gas elements within a certain radial distance from the accreting BH (e.g. Sijacki et al., 2007). In order to ensure that thermal AGN feedback remains efficient at most gas densities, gas particles should always receive ample amounts of energy from the AGN. Similar to the CC SN feedback discussed in §1.4.4, possible solutions include reducing the number of gas neighbours over which the energy is distributed in one time-step and/or performing energy injections at a lower frequency in time but with more energy accumulated into a single AGN-feedback event (e.g. Booth & Schaye, 2009). The critical parameter here is the ratio of the cooling time-scale of a heated gas element to the sound-crossing time-scale across the gas element. The ratio should be relatively high ($\gtrsim 10$) to prevent the injected energy from being quickly radiated away (e.g. Dalla Vecchia & Schaye, 2012).

At low accretion rates, the accretion of gas onto a BH can be described by a radiatively inefficient accretion flow or advection-dominated accretion flow (Narayan & Yi, 1995; Yuan & Narayan, 2014). In this case, the energy is liberated in terms of two collimated, relativistic jets, which are powered by the spin

of the SMBH. The jets inflate hot, buoyantly rising bubbles, and can extend in size to as far as hundreds of kpc. Their direction is often aligned with the spin of the BH's accretion disc, which in numerical simulations can be approximated by the average angular momentum of the gas elements close to the BH (e.g. Dubois et al., 2010; Davé et al., 2019). Alternatively, one can align the jets with the spin of the BH itself if the subgrid model computes the evolution of the BH spin (e.g. Huško et al., 2022). Numerically, jets are produced by applying a strong ($\sim 10^3 - 10^4$ km s^{-1}) velocity kick to gas elements around the BH within a sphere of a certain radius (e.g. Davé et al., 2019) or within a cylinder parallel to the spin of the BH (e.g. Dubois et al., 2010). The number of such discrete kick events per time-step depends on the jet power, which in most simulations is assumed to be proportional to the BH accretion rate $\dot{M}_{\rm BH}$.

1.5 Simulations of galaxy formation. Current status and future steps

Modern simulations of galaxies can be roughly split into four distinct groups: (i) simulations of individual galaxies in isolated environments, (ii) zoom-in simulations of individual haloes in a cosmological environment; (iii) large volume $(\sim 0.5 - 1$ cGpc$)^3$, low-resolution cosmological simulations; and (iv) cosmological simulations in moderate but statistically significant volumes $(\sim 25 - 100$ cMpc$)^3$ with a resolution high enough to enable studying properties of the ISM in individual galaxies.

(i) Simulations of individual galaxies in isolated environments are the most controlled tests of galaxy formation physics. Their main purpose is to test certain aspects of galaxy formation models: for instance, star formation and stellar feedback (e.g. Springel, 2000; Stinson et al., 2006; Marinacci et al., 2019; Smith et al., 2024). Unlike the other three categories of galaxy simulations, initial conditions for isolated-galaxy runs are non-cosmological. Such simulations do not commence at high redshifts ($z \sim 100$) where the gas density field is nearly homogeneous and no galaxies exist; instead, their initial conditions already contain a well-defined galaxy, which usually consists of rotationally supported discs of gas and stars, and, optionally, a stellar bulge. It is up to the simulation developer to decide the exact form for the initial morphology of the simulated object. Normally, the simulation domain contains only one galaxy[21], with the galaxy's dark-matter halo represented by an analytic gravitational potential (e.g. Springel et al., 2005a) that follows either Hernquist (1990) or Navarro, Frenk & White (1996) profile. The virial mass of the halo, as well as its spin parameter, needs to be

[21] A notable exception is the simulations of galaxy mergers in isolated environments (e.g. Springel et al., 2005a).

1

assumed before the start of the simulation. Other initial parameters include the mass of the stellar disc (and stellar bulge), the initial gas fraction, the initial gas metallicity, and the scale height of the stellar disc. In most cases, the simulation is run for $\sim 0.5-1$ Gyr, assuming a fixed redshift at all times (usually $z = 0$) and ignoring all cosmological effects.

(ii) Zoom-in simulations model the formation and evolution of the central galaxy in a single halo as well as its satellites. The basic structure of a zoom-in simulation is a high-resolution, zoomed-in, spherical region submerged into a low-resolution cosmological background. The zoomed-in region is positioned at the centre of the halo of interest, with its radius extending to a few times the $z = 0$ virial radius of the halo. Because only one (central) halo is modelled at a time, the resolution of zoom-in simulations is usually much higher than in any large cosmological-volume simulation, unless the zoomed-in halo is very massive (the virial mass $M_{200} \gtrsim 10^{14}$ $M_\odot$). Since for a fixed number of resolution elements, a smaller virial mass of the simulated halo implies higher resolution, zoom-in simulations applied to dwarf galaxies can be of particularly high resolution. Needless to say, this gain in numerical resolution is one of the most attractive features of zoom-in simulations, together with a more refined control over the initial conditions (compared to cosmological simulations from groups iii and iv). For instance, the FIRE2 simulations (Hopkins et al., 2018b) consist of zoom-in simulations of haloes with $z = 0$ virial masses from $\sim 10^9$ to 10^{12} $M_\odot$, corresponding to a gas-particle mass from 0.25×10^3 to 7×10^3 $M_\odot$. Besides FIRE2, the past decade has been rich with many other successful high-resolution zoom-in simulations of dwarf and Milky Way-mass galaxies; these include NIHAO (Wang et al., 2015), AURIGA (Grand et al., 2017), NEWHORIZON (Dubois et al., 2021), and LYRA (Gutcke et al., 2022).

(iii) The third group of simulations targets large cosmological volumes $(\sim 0.5-1$ cGpc$)^3$ at low spatial resolution. Although the internal structure of galaxies in such simulations is largely unresolved, they provide valuable information about the properties and evolution of the large-scale structure of the Universe, including a copious number of galaxy clusters that are found in such big volumes and that can thus be studied in detail. The largest hydrodynamical simulations from this group that were run to $z = 0$ are MAGNETICUM (Bocquet et al., 2016), BAHAMAS (McCarthy et al., 2017), MilleniumTNG (Pakmor et al., 2023), and FLAMINGO (Schaye et al., 2023).

(iv) Arguably the most numerically challenging is the fourth group, which comprises simulations at relatively high numerical resolution (resolution element masses of $\sim 10^5-10^7$ $M_\odot$ for gas) in somewhat smaller but still cosmo-

logically representative volumes ($\sim 25-100$ cMpc)^3. In essence, this group is a compromise between the second and third groups. The list of recent notable simulations from this group includes, among others, HorizonAGN (Dubois et al., 2014), Illustris (Vogelsberger et al., 2014), eagle (Schaye et al., 2015), MassiveBlack-II (Khandai et al., 2015), romulus25 (Tremmel et al., 2017), IllustrisTNG (Pillepich et al., 2018), and Simba (Davé et al., 2019). Furthermore, very recently the first ever cosmological-volume simulation with the explicit modelling of a multiphase ISM was run to $z = 0$ (FIREbox; Feldmann et al., 2023). Because of the relatively high resolution (6×10^4 $M_\odot$ for gas) and small time-steps readily reached in the dense gas, the simulated volume of FIREbox was limited to $\approx (22$ cMpc$)^3$.

1.5.1 Comparisons between simulated galaxies and observational data

The above simulations have been shown to reproduce a wide spectrum of galaxy properties and statistics that are observed directly and indirectly. These include

- Fundamental observables such as the galaxy stellar mass function and luminosity function at various redshifts;
- Galaxy morphology indicators such as the galaxy stellar mass-size relation and HI size versus HI mass relation;
- Mass-kinematics scaling relations such as the Tully-Fisher relation and the Magorrian relation;
- Galaxy chemical enrichment diagnostics such as galaxy stellar mass versus stellar metallicity and versus metallicity of the gas phase of the ISM;
- Properties of galaxy clusters such as radial profiles of gas temperature, pressure, metallicity, and X-ray luminosity;
- Diagnostics of galaxy evolution across cosmic time such as the cosmic star formation rate density as a function of redshift.

As an example, Fig. 1.3 illustrates how well modern simulations of galaxy formation can reproduce observational data for some of the properties from the above list. The simulation data are taken from the suite of eagle simulations (Schaye et al., 2015), where the curves and points coloured dark-blue, red, and green correspond to different simulations from the eagle set. The shaded regions highlight the scatter in the simulations. The observational data are displayed as filled and hollow grey data points, some of which have error bars. The left and right top panels show, respectively, the $z = 0.1$ galaxy stellar mass function (GSMF), and the specific star formation rate (sSFR) versus stellar mass for active

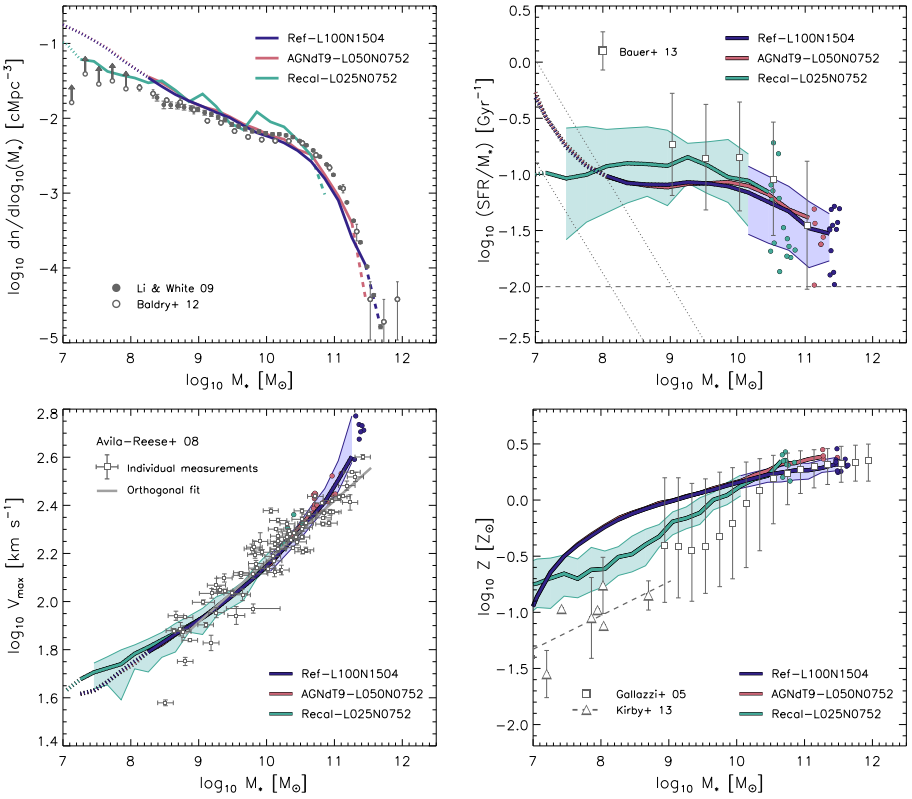

Figure 1.3: Comparison between the predictions of simulations (curves and points in dark-blue, red, and green) and observational data (grey data points). The simulation data are taken from the suite of EAGLE simulations (Schaye et al., 2015). Shown are the galaxy stellar mass function (*top left*), specific star formation rates versus stellar mass of active galaxies (*top right*), the maximum stellar rotational velocity versus galaxy stellar mass (*bottom left*), and stellar metallicity versus stellar mass (*bottom right*). All plots are shown at $z = 0.1$ and in all of them the EAGLE simulations exhibit good agreement with the observational data. See Schaye et al. (2015) for details.

galaxies at $z = 0.1$. The bottom panels show the maximum stellar rotational velocity (left) and stellar metallicity (right), both plotted versus galaxy stellar mass at $z = 0.1$. In all panels, we see good agreement between the predictions from the EAGLE simulations and the observational data. For further details on this comparison and discussion, see Schaye et al. (2015), from which Fig. 1.3 compiled.

Some of this agreement should not be considered unexpected because all models of galaxy formation contain free parameters and many of these free parameters are tuned such that the simulation matches a given set of observational data. For instance, by increasing the energy released per each CC SN explosion, one can reduce the average cosmic star formation rate, leading to a lower average stellar mass formed in the simulation by redshift $z = 0$. The true predictions of the simulation are therefore only those galaxy properties that have not been considered during the tuning of the subgrid parameters. For example, the EAGLE model was calibrated to reproduce the galaxy stellar mass function, galaxy sizes, and BH masses, all at $z \sim 0$ (Crain et al., 2015), which implies that in Fig. 1.3, all panels but the top left show the true predictions of EAGLE.

1.5.2 Challenges in numerical galaxy formation

Despite the tremendous success of numerical galaxy formation in the past decades, there are daunting issues and uncertainties that need to be addressed and resolved in the future with better numerical models and more powerful computational facilities:

- Arguably the biggest outstanding issue is the overall uncertainty in the subgrid modelling of all the intricate processes that are important for galaxy evolution but manifest themselves on numerically unresolved scales (e.g., star formation and energy feedback from stars and SMBHs). Ideally, one would desire to devise a subgrid model that is derived from the fundamental physics laws, has as few free parameters as possible, and yields a galaxy population whose properties not only reach an agreement with observational data but also converge when the numerical resolution is sufficiently high. Currently, all subgrid models are rather far from this ideal scenario, so there is a constant need to keep improving the subgrid models. The complexity of the subgrid models may be alleviated by the increase in numerical resolution, as it will make the treatment of the subgrid processes more direct.

- Another serious concern is the modelling of the ISM. As discussed in §1.4.1, most simulations of galaxy formation that have been run so far did not (directly) model the cold phase of the ISM. To our knowledge, the only two fairly large high-resolution simulations [simulated volume of $(\sim 20\ \mathrm{cMpc})^3$]

1

that did include the cold ISM are NewHorizon (Dubois et al., 2021) and FIREbox (Feldmann et al., 2023).

- In most galaxy simulations, radiative cooling and heating rates are calculated under the assumption of (collisional) ionization equilibrium (e.g. Vogelsberger et al., 2013; Schaye et al., 2015; Hopkins et al., 2018b; Dubois et al., 2021; Feldmann et al., 2023), for metals and/or primordial species. However, this assumption can break down if the ionization (or recombination) time-scale of a species exceeds the radiative cooling time-scale (e.g. Oppenheimer & Schaye, 2013), which is particularly important at gas temperatures $T \sim 10^5 - 10^6$ K where radiative cooling is the fastest. Furthermore, the effects of non-equilibrium cooling may alter, among others, the properties of galactic-scale molecular outflows and CO emission therefrom (e.g. Richings & Schaye, 2016).

- Finally, several processes that may play an important role under certain astrophysical conditions are often *completely* neglected in numerical simulations. This includes modelling the evolution and effects of magnetic fields as well as the injection and transport of cosmic rays. Both can be dynamically important and have an impact on the properties of the gas in the ISM and CGM (e.g. Pakmor & Springel, 2013; Salem et al., 2016; Hopkins et al., 2020).

1.6 This thesis

This thesis is built around the development of a new model of galaxy formation: colibre. The colibre simulations will be run in a broad range of volumes, from $\sim 50^3$ to $\sim 400^3$ cMpc^3, and spanning a wide range of gas resolutions, from $\sim 10^5$ to $\sim 10^7$ $M_\odot$ per particle. Compared to many previous simulations at similar resolutions and in similar volumes, the colibre model takes a big step forward by (i) directly modelling the multiphase structure of the ISM, allowing the gas to cool to as low as 10 K; (ii) treating the radiative cooling due to hydrogen and helium in non-equilibrium; (iii) increasing the number of dark matter particles by a factor of 4 relative to the initial number of gas particles[22]; and (iv) having more realistic subgrid models of numerous unresolved astrophysical processes, among which are star formation, the formation and evolution of dust, chemical enrichment, and feedback from CC SNe and AGN.

All simulations in this thesis were run with the astrophysical code swift (Schaller et al., 2023) and used the SPH scheme Sphenix (Borrow et al., 2022)

[22]The increase in the number of dark matter particles will improve the stellar kinematics and structural properties of simulated galaxies that are sampled with less than $\sim 10^3$ stellar particles (e.g. Ludlow et al., 2023).

to solve the equations of hydrodynamics. Each chapter of this thesis can be regarded as a step in the development of the COLIBRE model, with a particular focus on CC SN feedback in Chapters 3 and 4.

- In Chapter 2, we simulate isolated SN explosions in a homogeneous medium at ~ 1 pc resolution. These high-resolution hydrodynamic tests were created during the development of the algorithms for CC SN feedback that are employed in the COLIBRE model, but here we use them to solve a problem that is slightly outside the scope of the COLIBRE project. Specifically, using our high-resolution simulations, we show that the non-zero concentration of the radioactive isotope ^{60}Fe that has been consistently detected by different groups in samples from the surface of the Moon and from a ferromanganese crust on Earth, is very likely to have an SN origin. We demonstrate that the observed ^{60}Fe signal can be explained by two SNe detonated in the past 10 Myr within 100 pc from the Solar System, and that the velocity of the Solar System relative to the ISM is important for the interpretation of the signal.

- In Chapter 3, we run simulations of an isolated Milky Way-mass galaxy and small cosmological volumes to study different ways of selecting gas resolution elements for energy injection in CC SN feedback. We show that different algorithms for deciding which gas element receives the energy from the stellar particle in CC SN feedback, result in significant variations in galaxy star formation histories, morphologies, and mass outflow rates. Additionally, we introduce an isotropic algorithm for distributing energy in SN feedback around stellar particles, which is used in the next two chapters.

- In Chapter 4, we present a new thermal-kinetic model for CC SN feedback designed for cosmological simulations of galaxy formation that (may) include a cold ISM phase. The feedback model manifestly conserves energy, linear momentum, and angular momentum, and distributes SN energy in a statistically isotropic manner. We run simulations of isolated Milky Way-mass and dwarf galaxies to test and validate the model. We study the differences between the thermal and kinetic channels of energy injection and demonstrate that using the two channels together produces galaxies whose spatially resolved properties are closest to observations.

- Finally, in Chapter 5, we present the calibration of the COLIBRE model of galaxy formation. The prescriptions for SN and AGN feedback in the COLIBRE model are based on the results from Chapters 3 and 4. We show how various galaxy properties – such as galaxy stellar mass, size, and star formation rate – are affected by the choice of parameters of the COLIBRE prescriptions for AGN and SN feedback. We find the best-fitting values for

these parameters by computing the galaxy stellar mass function and size-stellar mass relation of simulated galaxies at $z = 0$ and matching them to observations, for which we use a machine-learning method. We show that the calibrated COLIBRE model not only reproduces the observed data used in the calibration, but also agrees with many other properties of observed galaxies to which the model was not calibrated, including the properties of cold gas.

References

Aarseth S. J., Gott J. R. I., Turner E. L., 1979, ApJ, 228, 664
Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, ApJ, 770, 25
Alam S., et al., 2015, ApJS, 219, 12
Alam S., et al., 2017, MNRAS, 470, 2617
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Arnold C., Leo M., Li B., 2019, Nature Astronomy, 3, 945
Bahé Y. M., et al., 2022, MNRAS, 516, 167
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bennett C. L., et al., 2003, ApJS, 148, 1
Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Borrow J., Schaller M., Bower R. G., Schaye J., 2022, MNRAS, 511, 2367
Borrow J., Vogelsberger M., O'Neil S., McDonald M. A., Smith A., 2023, MNRAS, 520, 649
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Cassinelli J. P., 1979, ARA&A, 17, 275
Chabrier G., 2003, PASP, 115, 763
Correa C. A., Schaller M., Ploeckinger S., Anau Montel N., Weniger C., Ando S., 2022, MNRAS, 517, 3045
Crain R. A., et al., 2015, MNRAS, 450, 1937
Cullen L., Dehnen W., 2010, MNRAS, 408, 669
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Dehnen W., Read J. I., 2011, European Physical Journal Plus, 126, 55
Devecchi B., Volonteri M., 2009, ApJ, 694, 302
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2010, MNRAS, 409, 985
Dubois Y., et al., 2014, MNRAS, 444, 1453
Dubois Y., et al., 2021, A&A, 651, A109
Ellison S. L., Songaila A., Schaye J., Pettini M., 2000, AJ, 120, 1175
Evans Neal J. I., et al., 2009, ApJS, 181, 321
Event Horizon Telescope Collaboration et al., 2019, ApJ, 875, L1
Event Horizon Telescope Collaboration et al., 2022, ApJ, 930, L12
Faucher-Giguère C.-A., 2020, MNRAS, 493, 1614
Feldmann R., et al., 2023, MNRAS, 522, 3831
Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Fiore F., et al., 2017, A&A, 601, A143
Fishlock C. K., Karakas A. I., Lugaro M., Yong D., 2014, ApJ, 797, 44
Gatto A., et al., 2017, MNRAS, 466, 1903

Gentry E. S., Krumholz M. R., Dekel A., Madau P., 2017, MNRAS, 465, 2471
Grand R. J. J., et al., 2017, MNRAS, 467, 179
Greengard L., Rokhlin V., 1987, Journal of Computational Physics, 73, 325
Gunn J. E., Knapp G. R., Tremaine S. D., 1979, AJ, 84, 1181
Gutcke T. A., Pakmor R., Naab T., Springel V., 2022, MNRAS, 513, 1372
Hafen Z., et al., 2022, MNRAS, 514, 5056
Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426
Hernquist L., 1990, ApJ, 356, 359
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., et al., 2018a, MNRAS, 477, 1578
Hopkins P. F., et al., 2018b, MNRAS, 480, 800
Hopkins P. F., et al., 2020, MNRAS, 492, 3465
Hu W., Dodelson S., 2002, ARA&A, 40, 171
Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168
Huško F., Lacey C. G., Schaye J., Schaller M., Nobels F. S. J., 2022, MNRAS, 516, 3750
Katz N., 1992, ApJ, 391, 502
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kaviraj S., et al., 2017, MNRAS, 467, 4739
Keller B. W., Kruijssen J. M. D., Chevance M., 2022, MNRAS, 514, 5355
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Khetan N., et al., 2021, A&A, 647, A72
Kollatschny W., Zetzl M., 2013, A&A, 549, A100
Koo B.-C., Kim C.-G., Park S., Ostriker E. C., 2020, ApJ, 905, 35
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
Kugel R., et al., 2023, MNRAS, 526, 6103
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C. G., et al., 2016, MNRAS, 462, 3854
Lee E. J., Miville-Deschênes M.-A., Murray N. W., 2016, ApJ, 833, 229
Ludlow A. D., Fall S. M., Wilkinson M. J., Schaye J., Obreschkow D., 2023, MNRAS, 525, 5614
Madau P., Rees M. J., 2001, ApJ, 551, L27
Maoz D., Graur O., 2017, ApJ, 848, 25
Marinacci F., Sales L. V., Vogelsberger M., Torrey P., Springel V., 2019, MNRAS, 489, 4233
Marshall H. L., et al., 2001, ApJ, 549, L167
McAlpine S., Harrison C. M., Rosario D. J., Alexander D. M., Ellison S. L., Johansson P. H., Patton D. R., 2020, MNRAS, 494, 5713
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McKee C. F., Ostriker J. P., 1977, ApJ, 218, 148
Menon S. H., Federrath C., Krumholz M. R., 2023, MNRAS, 521, 5160
Mittal R., Hudson D. S., Reiprich T. H., Clarke T., 2009, A&A, 501, 835
Mocz P., Fabian A. C., Blundell K. M., 2013, MNRAS, 432, 3381
Monaghan J. J., 1992, ARA&A, 30, 543
Monaghan J. J., 1997, Journal of Computational Physics, 136, 298
Monaghan J. J., 2005, Reports on Progress in Physics, 68, 1703
Monaghan J. J., Gingold R. A., 1983, Journal of Computational Physics, 52, 374
Muratov A. L., et al., 2017, MNRAS, 468, 4170
Naab T., Ostriker J. P., 2017, ARA&A, 55, 59
Narayan R., Yi I., 1995, ApJ, 452, 710
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nuclear Phys. A, 777, 424
Oppenheimer B. D., Davé R., 2006, MNRAS, 373, 1265
Oppenheimer B. D., Schaye J., 2013, MNRAS, 434, 1043

Ostriker E. C., 1999, ApJ, 513, 252
Pakmor R., Springel V., 2013, MNRAS, 432, 176
Pakmor R., Pfrommer C., Simpson C. M., Springel V., 2016, ApJ, 824, L30
Pakmor R., et al., 2023, MNRAS, 524, 2539
Perlmutter S., et al., 1999, ApJ, 517, 565
Pillepich A., et al., 2018, MNRAS, 473, 4077
Planck Collaboration et al., 2014, A&A, 571, A16
Planck Collaboration et al., 2020, A&A, 641, A6
Ploeckinger S., Schaye J., 2020, MNRAS, 497, 4857
Price D. J., 2012, Journal of Computational Physics, 231, 759
Raskutti S., Ostriker E. C., Skinner M. A., 2017, ApJ, 850, 112
Richings A. J., Schaye J., 2016, MNRAS, 458, 270
Riess A. G., et al., 1998, AJ, 116, 1009
Roca-Fàbrega S., et al., 2021, ApJ, 917, 64
Rupke D. S. N., Veilleux S., 2013, ApJ, 768, 75
Russell H. R., McNamara B. R., Edge A. C., Hogan M. T., Main R. A., Vantyghem A. N., 2013, MNRAS, 432, 530
Salem M., Bryan G. L., Corlies L., 2016, MNRAS, 456, 582
Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2008, MNRAS, 389, 1137
Schaller M., et al., 2023, arXiv e-prints, p. arXiv:2305.13380
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J., et al., 2010, MNRAS, 402, 1536
Schaye J., et al., 2015, MNRAS, 446, 521
Schaye J., et al., 2023, MNRAS, 526, 4978
Schmidt M., 1959, ApJ, 129, 243
Sedov L. I., 1959, Similarity and Dimensional Methods in Mechanics
Semenov V. A., Conroy C., Chandra V., Hernquist L., Nelson D., 2023, arXiv e-prints, p. arXiv:2306.13125
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Shang C., Bryan G. L., Haiman Z., 2010, MNRAS, 402, 1249
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Smith M. C., Sijacki D., Shen S., 2018, MNRAS, 478, 302
Smith M. C., et al., 2024, MNRAS, 527, 1216
Speranza G., et al., 2022, A&A, 665, A55
Springel V., 2000, MNRAS, 312, 859
Springel V., 2010a, ARA&A, 48, 391
Springel V., 2010b, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., Di Matteo T., Hernquist L., 2005a, MNRAS, 361, 776
Springel V., et al., 2005b, Nature, 435, 629
Springel V., Di Matteo T., Hernquist L., 2005c, ApJ, 620, L79
Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074
Taylor G., 1950, Proceedings of the Royal Society of London Series A, 201, 159
Tchekhovskoy A., Narayan R., McKinney J. C., 2010, ApJ, 711, 50
Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121
Tumlinson J., et al., 2011, Science, 334, 948
Vikhlinin A., et al., 2009, ApJ, 692, 1060
Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031
Vogelsberger M., et al., 2014, Nature, 509, 177
Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, MNRAS, 454, 83
White S. D. M., 1976, MNRAS, 177, 717

Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
Wise J. H., 2019, arXiv e-prints, p. arXiv:1907.06653
Yuan F., Narayan R., 2014, ARA&A, 52, 529
Zel'dovich Y. B., 1970, A&A, 5, 84
Zwicky F., 1933, Helvetica Physica Acta, 6, 110