The sensitivity of stellar feedback to IMF averaging versus IMF sampling in galaxy formation simulations

Matthew C. Smith

1 Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

ABSTRACT

Galaxy formation simulations frequently use Initial Mass Function (IMF) averaged feedback prescriptions, where star particles are assumed to represent single stellar populations that fully sample the IMF. This approximation breaks down at high mass resolution, where stochastic variations in stellar populations become important. We discuss various schemes to populate star particles with stellar masses explicitly sampled from the IMF. We use Monte Carlo numerical experiments to examine the ability of the schemes to reproduce an input IMF in an unbiased manner while conserving mass. We present our preferred scheme which can easily be added to pre-existing star formation prescriptions. We then carry out a series of high resolution isolated simulations of dwarf galaxies with supernovae, photoionization and photoelectric heating to compare the differences between using IMF averaged feedback and explicitly sampling the IMF. We find that if supernovae are the only form of feedback, triggering individual supernovae from IMF averaged rates gives identical results to IMF sampling. However, we find that photoionization is more effective at regulating star formation when IMF averaged rates are used, creating more, smaller H\textsc{ii} regions than the rare, bright sources produced by IMF sampling. We note that the increased efficiency of the IMF averaged feedback versus IMF sampling is not necessarily a general trend and may be reversed depending on feedback channel, resolution and other details. However, IMF sampling is always the more physically motivated approach. We conservatively suggest that it should be used for star particles less massive than \( \sim 500 \, M_{\odot} \).

Key words: galaxies: formation, galaxies: evolution, methods: numerical

1 INTRODUCTION

Substantial progress in understanding the formation and evolution of galaxies has been made over the last few decades by accounting for the role played by stellar feedback (see e.g. the reviews of Somerville & Davé 2015; Naab & Ostriker 2017, and the references therein). Stars interact with the gas inside galaxies and beyond via a variety of complex processes, including supernovae (SNe), stellar winds and radiation (providing photoionization, photodissociation, photoheating and radiation pressure). They can therefore act to influence the thermal and kinetic state of the interstellar medium (ISM), regulate star formation, enrich gas with metals, drive galactic outflows (carrying metals out into the circumgalactic medium (CGM)) and more besides. As the preferred terminology, “feedback”, would suggest, these processes represent a back-reaction on the gas from which stars themselves originate. Thus, numerical simulations of galaxies must account for the gaseous and stellar components, as well as their interactions, in a self-consistent manner.

The “star particle” has been a nearly ubiquitous feature of hydrodynamic simulations of galaxy formation for nearly three decades (see early examples in e.g. Katz 1992; Katz et al. 1996; Navarro & White 1993; Mihos & Hernquist 1994). In the majority of cases, the star particle does not represent individual stars or even star clusters, per se. It instead traces an underlying collisionless stellar fluid in much the same way that a dark matter particle traces the dark matter fluid. Unlike the dark matter component, however, the stellar component must interact with the gas through processes in addition to gravity. Star particles can be created from gas mass to represent the process of star formation. Stellar feedback can be directly tied to the star particle (as in Katz 1992, one of the first examples of stellar feedback being explicitly included in a simulation). However, when the mass resolution is low, some or all of the stellar feedback may not be directly associated with star particles but instead associated with the star forming gas. For example,
Springel & Hernquist (2003) presents models which treat unresolved SN feedback by modifying the equation of state of star forming gas and launching stellar feedback driven winds from the gas without directly involving a star particle. There is no inconsistency in this context as the star particle does not explicitly represent the location of the stellar mass, which is in some sense smoothed over the ensemble of particles. Such “diffuse” stellar feedback models are frequently used in modern large volume cosmological simulations, usually in combination with star particle centred feedback to a greater or lesser extent (Vogelsberger et al. 2013, 2014; Dubois et al. 2014, 2016; Davé et al. 2016, 2019; Pillepich et al. 2018). Once the mass resolution of the star particle approaches \( \sim 10^4 M_\odot \), or better, stellar feedback is usually tied directly to star particles and is modelled in a more explicit manner (for a small sample of contemporary approaches see e.g. Hopkins et al. 2014; Hopkins et al. 2018; Ceverino et al. 2014; Kimm et al. 2015; Agertz & Kravtsov 2015; Marinacci et al. 2019). This is appropriate because the lower particle masses begin to allow small scale spatial and temporal clustering of feedback processes to be resolved. Regardless of the approach taken, a link between the mass in stars and the resulting stellar feedback budget must be made. Because individual stars are not resolved in the schemes described above, this relationship is obtained by averaging over the statistical distribution of stellar masses.

This distribution is given by the initial mass function (IMF). Combining an IMF with a set of stellar evolution models can produce the net feedback properties (e.g. SN rate, luminosity in various bands relevant to feedback, wind power and mass loss rate etc.) for a single stellar population (SSP) under the assumption that the IMF is fully sampled. Then, the feedback budget can be determined for a star particle in a simulation (typically with a dependence on its age and metallicity via lookup tables), rescaled by its mass. We will refer to this as IMF averaged feedback throughout this work. It is important to note that even when IMF averaged SN rates are used, SNe can still be modelled as discrete events by stochastically sampling the rates (see e.g. Stinson et al. 2010; Hopkins et al. 2014; Kimm et al. 2015; Revaz et al. 2016; Smith et al. 2018). Consistent evidence for a universal IMF has been amassed from observations of large populations of stars, featuring a steep power law at high masses (Salpeter 1955) with a knee at \( \sim 1 M_\odot \) (Kroupa et al. 1993; Kroupa 2001; Chabrier 2003). Potential deviations away from the universal IMF can arise in two forms. Firstly, there has been some evidence that the IMF may vary systematically between galaxies (Hoversten & Glazebrook 2008; van Dokkum & Conroy 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012; Kalirai et al. 2013; La Barbera et al. 2013; Geha et al. 2013), with possible correlations with central velocity dispersion and/or metallicity. However, the magnitude of this phenomenon remains unclear. Nonetheless, if the IMF does indeed vary systematically, this effect is conceptually simple to include in IMF averaged feedback schemes. For example, in cosmological zoom-in simulations of Milky Way-like galaxies, Gultek & Springel (2019) use an IMF that varies as a function of metallicity to adjust SN rates and metal enrichment from AGB winds on-the-fly.

Alternatively, deviations could arise due to variations of the IMF on small scales or as a result of undersampling a universal IMF in small populations. The total distribution of stars in a galaxy is a composite of the various individual star forming regions. Thus, an integrated galactic IMF (IGIMF) arises from the combination of the IMF within individual star clusters and the cluster mass function (Kroupa & Weidner 2003). If a common universal IMF is well traced within each cluster independent of cluster mass, then the IGIMF will be identical to the universal IMF. If the IMF varies from cluster to cluster the shape of the IGIMF will change. In particular, if the high-mass cut-off is a function of cluster mass, then the IGIMF slope will be steeper than the universal IMF. There is evidence that massive stars are rarer in low-SFR environments (Meurer et al. 2009; Lee et al. 2009, 2016; Gunawardhana et al. 2011), which could be a consequence of this scenario. It has been posited that this could arise due differences in how star formation proceeds in clusters of different mass, leading to an intrinsic relationship between cluster mass and high-mass cutoff (Kroupa & Weidner 2003; Weidner & Kroupa 2006; Weidner et al. 2010, 2013). However, it has also been argued that no such deterministic link exists and that observations are consistent with a stochastically sampled uniform IMF (Elmegreen 2006; Corbelli et al. 2009; Calzetti et al. 2010; Fumagalli et al. 2011; Andrews et al. 2013, 2014). These effects cannot be captured a priori with IMF averaged feedback as they emerge as a result of stochastic effects arising from undersampling of the IMF in a given small population of stars.

The alternative is to populate star particles at their birth with an inventory of stars by explicitly sampling the IMF. Feedback budgets can then be based on the individual stars residing in the star particle. When star particles are massive enough, this will converge with the IMF averaged approach because the IMF will be well sampled within the particle. Taking the opposite limit yields the modelling of individual stars. The assumption that a stellar population samples the IMF well only holds for population masses above \( \sim 10^2 M_\odot \) (Carigi & Hernandez 2008; Revaz et al. 2016), below which stochastic effects will begin to emerge. Explicit sampling of the IMF allows issues related to undersampling of the IMF (as described in the previous paragraph) to be captured. It also allows the inhomogeneous distribution of stellar feedback among stars (varying luminosities, mass loss, SN energy injection etc.) to be resolved. Various schemes for explicit IMF sampling have been presented and used in simulations of individual GMCs, patches of discs (stratified boxes) and entire galaxies (see e.g. Gatto et al. 2017; Sormani et al. 2017; Genn et al. 2018; Hu et al. 2017; Hu 2019; Fujimoto et al. 2018; Emerick et al. 2019; Applebaum et al. 2020; Gultek et al. 2021, some of which we will discuss in greater detail in Section 2).

Su et al. (2018) estimate the effect of using stochastically sampled stellar masses compared to IMF averaged feedback in cosmological zoom-in simulations. They find that as long as SNe are modelled as discrete events (even if sampled from IMF averaged rates), stochastic variation of the stellar content of star particles does not produce a significant departure from simulations run with IMF averaged feedback. However, they do not explicitly perform IMF sampling, instead using a toy model to modulate their IMF averaged feedback (OB winds, luminosities and SN rates). In Section 4 we will discuss the validity of such an approach. Grudić & Hopkins (2019) use a similar technique in simulations of individual GMCs, but find that approximating the inhomoge-
neous distribution of UV luminosities among stars, leading to the presence of rare, bright sources, results in lower star formation efficiencies compared to using IMF averaged feedback. In non-cosmological simulations of dwarfs, Applebaum et al. (2020) find that when SN feedback and $H_2$ dissociating radiation are linked to explicitly sampled stellar masses rather than being based on IMF averaged rates (that still discretize SNe), feedback is moderately less efficient at regulating SFRs and the mass of cold gas. We will discuss their findings in more detail in Section 4.

This work is laid out as follows. Section 2 contains an extended discussion of the details of explicit IMF sampling schemes. We shall use some simple Monte Carlo numerical experiments to demonstrate the advantages and disadvantages of several schemes, populating star particles of various masses. In Section 3 we use high resolution non-cosmological simulations of dwarf galaxies to directly compare the use of IMF averaged feedback to explicit IMF sampling. In Section 4 we discuss our findings and their consequences in greater detail, as well as providing a comparison to some other relevant works. Section 5 presents our conclusions. Appendix A demonstrates that our simulations are robust to stochastic effects by rerunning the early stages of a subset of our simulations with randomly perturbed initial conditions and different random number generator seeds.

2 POPULATING STAR PARTICLES FROM AN IMF

2.1 Requirements of an IMF sampling scheme

One of the main reasons for paying close attention to details of the output IMF in a galaxy formation simulation is its relationship to the feedback budget. Since the vast majority of stellar feedback arises from comparatively rare, massive stars, subtle changes to the distribution of stellar masses can potentially influence the evolution of the galaxy. When designing a scheme to populate star particles created in a galaxy formation simulation with stellar masses drawn from an IMF, there are two main requirements that are often in tension with each other. The scheme must attempt to reproduce the input IMF as closely as possible, but it should also conserve stellar mass. Problems arise because a sequence of discrete stellar masses drawn from an IMF is in general unlikely to sum exactly to a previously specified value (e.g. the mass of a star particle), with the last sampled mass overshooting the target. What a scheme does in this scenario determines how closely it prioritizes mass conservation vs. reproducing the IMF.

One possible solution is to accept the last draw and source more mass to make up the difference. Such an approach is trivial to implement if the initial star particle mass is smaller than the mass of the gas resolution element from which it is formed. Otherwise, the mass reservoir can be ‘topped up’ by taking mass from other nearby gas particles/cells (either by merging or partially draining them, see e.g. Hirai et al. 2020; Gutcke et al. 2021). If done in combination with the standard method for forming star particles (stochastically sampling the SFRs of individual gas particles/cells independently to trigger the spawning of or conversion to a star particle), care must be taken to avoid or minimize inconsistencies between the expected SFR of the gas and the rate at which stellar mass is created (e.g. Hu 2019 avoids this issue by exchanging mass between star particles after they have been created). However, in this work we will avoid any additional transfer of mass to particles and focus on methods to populate a star particle of fixed mass with stars, presenting a scheme that can be easily incorporated into any existing implementation of star formation.

Before proceeding, we will comment on why it is important that an input IMF is accurately reproduced in a galaxy formation simulation. In nature, the IMF is an emergent property of the process of star formation, arising from the small scale physics governing the fragmentation and gravitational collapse of turbulent, star forming gas and its interaction with feedback. By contrast, in galaxy formation simulations it is impossible to resolve the physical processes that give rise to the IMF. An IMF must therefore be imposed upon the simulation as a sub-grid model, either in an explicit manner by sampling from it on-the-fly or by taking IMF averaged approaches to stellar feedback. The chosen IMF may be derived empirically from observations or it can be based on theoretical expectations. It can be a fixed, universal IMF or it could take a more complicated form, being allowed to vary with galactic or even local properties. Regardless of the form adopted, once chosen, the input IMF encodes the unresolved physics of small scale star formation that cannot be captured in the simulation. Therefore, it is imperative that it is not subsequently biased by numerical issues originating from the implementation. As an obvious example, when the star particle mass resolution begins to approach (or even drop below) the upper mass cut of the IMF, the resulting distribution of stellar masses is very vulnerable to being biased away from input IMF (as we shall show below). Such a bias must be avoided as much as possible, since it is obvious that the distribution should be independent of the mass resolution as this is entirely numerical.

A more subtle problem can occur in regions of low SFR. When a star particle is created and is populated from the IMF, it is possible that there is not enough mass in the local star forming cloud (let alone the star particle) to satisfy a draw from the input IMF. It is tempting under these circumstances to simply discard the draw, since it is clear that such a massive star could not form in this environment. However, this also leads to a biasing of the input IMF that has its basis in numerics not physics. If the resolved physics of the simulation frequently provides situations where there is not enough gas mass available to accurately sample from the input IMF, this suggests that the input IMF is inappropriate. A better IMF should be chosen that encodes either an empirical or theoretical model for how star formation proceeds in low SFR environments. If draws are discarded, then the resulting distribution of stellar masses is simply a flawed realisation of the input IMF rather than reflecting any resolved physics. One could imagine that in the scenario described above, a different output from the random number generator could have resulted in the massive star being formed later in the evolution of the star forming cloud, when there was sufficient mass available. It is also possible that the manner in which the cloud assembled is sensitive to resolution. The final distribution of stellar masses averaged over many star particles should have no such dependence on numerics since this introduces a bias with no physical basis. Additionally, we note that it is entirely possible that a low mass galaxy does not form enough stars to fully populate
the IMF. This does not necessarily indicate that the input IMF is inappropriate, rather that multiple realisations of the simulation should be performed to assess the impact of stochasticity.

In the previous discussion, it is clear that the majority of cases in which the output IMF can be biased are due to issues of mass conservation. While we have described why simply throwing away a draw from the IMF because of the lack of stellar mass must be avoided, it is also apparent that simply accepting every draw would result in a net overproduction of stars because the target mass will always be exceeded. An IMF sampling scheme must therefore balance the competing demands of mass conservation and IMF preservation. The compromise usually takes the form of an inconsistency between the target mass and the mass of the samples (which we will refer to as the assigned mass) on a particle by particle basis, but a consistency when averaged over many particles. The particle level inconsistency may be eliminated after the fact by a transfer of mass (as mentioned at the beginning of this section) or can simply be left as a mismatch between the dynamical and assigned mass of the particle (the approach we adopt in this work). Because the exact N-body interactions between stars cannot usually be treated in a galaxy formation simulation (the use of softened gravitational forces often being adopted) this inconsistency is not of much concern from a dynamical perspective. The mass inconsistency can also be resolved conceptually if low mass stars (which do not contribute substantially to the feedback budget) are not tracked explicitly. The mass inconsistency can then be thought of as a simple redistribution of the low mass stars between star particles, as long as the overall inconsistency sums to zero over many particles (e.g. Applebaum et al. 2020 use this philosophy). This has no practical impact on the operation of the scheme, other than to rationalise the discrepancy. Additionally, it obviously does not work in the case where the combined mass of the explicitly tracked massive stars exceeds the dynamical mass. Regardless, if the inconsistency is too large there may not be sufficient mass to return via stellar winds and SN ejecta as required by the assigned stellar masses. We will briefly touch on this subject later in this work.

Finally, we note that we have been vague as to distinguishing between a truly universal IMF versus an IGIMF arising from an IMF that varies between clusters, as described in Section 1. For example, if stars born in clusters follow the shape of a universal IMF, but properties of the cluster (most importantly the mass) impose a high mass cutoff then the IGIMF (which is composed of the sum of the IMFs of the individual clusters) will be steeper than the canonical IMF (Kroupa & Weidner 2003). This means that the IGIMF will have a dependence on the mass function of the star forming regions with galaxy properties (mass, metallicity etc.) playing a key role. However, it is important to note that star particles should not be conflated with physical star clusters. Unless the adopted star formation prescription produces a distribution of star particle masses such that the mass function is an emergent property of the simulation (and is believed to be an accurate representation of the true cluster mass function), the star particle mass has no physical meaning. In almost all cases the star particle mass is set uniformly as a parameter of the simulation or as a consequence of the gas resolution. This means that truncation of the upper end of a universal IMF based on the star particle mass is unphysical. If the use of a particular form of a cluster mass dependent IMF (and resulting IGIMF) is required, a more complex sub-grid model is needed.

### 2.2 Methods for explicit IMF sampling

We will now explore several methods for populating star particles with stellar masses from the IMF. As explained in the previous paragraph, we do not consider methods that account for the truncation of the high mass end of the IMF as a function of the desired total sample mass (the target mass) (e.g. the ‘sorted sampling’ method of Weidner & Kroupa 2006) as in general star particles do not accurately represent stellar clusters. Three simple methods of sampling from the IMF to reach a target mass are ‘stop before’, ‘stop after’ and ‘stop nearest’ (see Haas & Anders 2010). These terms refer to what is done with the last drawn stellar mass which will inevitably exceed the desired target.

‘Stop before’ sampling discards the last drawn stellar mass. It guarantees that the total drawn stellar mass (which we refer to as the assigned mass, \(m_{\text{asn}}\)) will not exceed the mass of the star particle. However, this means that the sum of assigned masses over all particles will inevitably show a deficit relative to the total dynamical mass of star particles formed. It will also create a bias towards low mass stars since a high mass star is more likely to exceed the target mass.

The output IMF will therefore deviate downwards from the input IMF, being finally truncated at the star particle mass. As the star particle mass is increased, this bias is reduced. Likewise, the fractional deficit of the total assigned mass to the total particle mass will reduce because the relative contribution of the last draw is reduced. ‘Stop after’ sampling takes the opposite approach, always keeping the last drawn star. By avoiding discarding any draw, this method guarantees that the distribution of assigned stellar masses perfectly follows the shape of the input IMF. The downside is that this comes at the cost of normalizing the IMF upwards, always assigning more stars to particles than the available stellar mass. Again, the relative impact of this bias is reduced as the star particle mass is increased. It should also be noted that any scheme which accepts the last drawn star and then makes the sampled and dynamical masses consistent by taking additional mass from nearby gas after the decision to create a particle has already been made will suffer a similar bias, artificially inflating the local SFR.

A compromise between these two methods is the ‘stop nearest’ approach. With this scheme, the last drawn stellar mass is kept if the total assigned mass is closer to the target mass with its inclusion than without. The motivation is that the cases of the last drawn mass being discarded will be counterbalanced by the occasions it is kept. This is true for large star particle masses, where the size of the gap between the penultimate drawn mass and the target mass is insignificant compared to the target mass. However, as we shall demonstrate, this method suffers from the same biases as the ‘stop before’ scheme (albeit to a lesser extent) when the star particle mass is smaller.

A common deficiency of these three methods is that the populations of individual particles are completely independent of each other. This is only appropriate if the star particle represents an independent single stellar population.
(SSP) (e.g. as formed in a star cluster). It is inappropriate for small star particle masses since the ensemble of star particles represent the stellar population together. However, as discussed earlier, it is usually impossible at the resolutions of a galaxy formation simulation for a realistic population of stars to be an emergent feature of the local physical environment. Thus, in order to accurately recover an input IMF, each star particle cannot be populated in isolation. Instead, information about the distribution of stellar masses in previous star particles must be taken into account to correct for biases. While this necessarily involves some form of ‘action at a distance’, it is important to note that this was already implied by requiring that ensembles of stars follow an enforced IMF together rather than setting individual stellar masses independently from resolved local conditions. We implement this approach in a scheme we refer to as ‘adjusted target’. Fig. 1 contains a flowchart detailing the algorithm. The scheme is similar to ‘stop after’ except that the target mass is not in general the same as the particle mass. Instead, the target mass for a given particle is adjusted based on how far the previous star particle overshot its target. This allows us to keep the total assigned mass consistent with the total stellar mass formed across many particles, but to simultaneously perfectly reproduce the input IMF because we never discard a draw. This scheme is very similar to that presented in Hu et al. (2017) with the difference that we do not adjust the dynamical mass of particles to enforce consistency between assigned mass and dynamical mass for each particle individually. This avoids unphysical mass transfer over potentially arbitrary distances. Similarly, there is no inter-particle mass exchange so it has the advantage that it can be trivially implemented on top of any pre-existing implementation of star formation with a minimum of additional coding.

When a star particle is populated, we draw stellar masses from the input IMF and assign them to the star particle until the total assigned mass exceeds the target. The last draw is kept and the star particle is now finished with. The ‘offset mass’ is equal to the sum of masses assigned to the particle minus the target. The target mass of the next particle to be populated is set to its particle mass minus the offset mass from the previous particle. The target mass will always be equal to or less than the particle mass. However, while the assigned stellar mass will always exceed the target mass it will not always exceed the particle mass. Some particles will have an excess of assigned stellar mass while others will have a deficit. Averaged across multiple star particles the total assigned mass will equal the total star particle dynamical mass.

Without modification, the algorithm presented in Fig. 1 accounts for circumstances where the target mass is negative or zero. This occurs if the last assigned stellar mass of the previous particle resulted in a substantial overshoot of the target. When the target mass is negative or zero, the algorithm will draw no stellar masses for the star particle and the offset mass is reduced by the mass of the star particle, resulting in a higher target mass for the next star particle. This feature of the adjusted target scheme means that it is possible to include stellar masses that are larger than the fiducial star particle mass, simply resulting in a particle that has more assigned mass than dynamical mass and some particles that compensate by having no assigned mass. For example, consider the extreme example where a 100 $M_\odot$ star is the first draw for a 20 $M_\odot$ star particle and, for the sake of simplicity, the current offset mass is 0 $M_\odot$ (meaning that the target mass for this star particle is also 20 $M_\odot$). The draw of the 100 $M_\odot$ star is accepted and assigned to the star particle. The target mass has been exceeded, so sampling is now concluded for this star particle. The offset mass is now 80 $M_\odot$, because the target was overshot by that amount. The algorithm will produce target masses for the next four 20 $M_\odot$ particles that are less than or equal to zero. They will therefore have no assigned stellar mass. The resulting total assigned mass equals the total dynamical mass across the five particles, at the cost that each particle has an inconsistency between its individual assigned and dynamical masses. As mentioned above, because we cannot anyway resolve the dynamics of individual stars this is largely irrelevant. It is only a problem if there is not enough mass available to be returned through feedback (discussed below). In practice, cases such as this are rare because high mass stars are proportionally less likely to be drawn. However, as we shall demonstrate, failing to allow for their formation biases the IMF and reduces the overall feedback budget. The target mass usually stays close to the particle mass (unless

---

**Figure 1.** A schematic illustration of our ‘adjusted target’ star particle populating scheme. Stellar masses are drawn from the input IMF until a target mass is exceeded. The last drawn mass is always kept, ensuring that the output IMF is unbiased. The target mass is not in general equal to the dynamical mass of the star particle, but compensates for the amount that the target was overshot when sampling the previous star particle (the offset mass). Thus, some star particles will have more assigned stellar mass than their dynamical mass while the reverse is true for others. Across an ensemble of star particles the total assigned mass is equal to the total dynamical mass and the input IMF is perfectly reproduced.
the particle mass is considerably smaller than typical stellar mass), only moving substantially if a very massive star is drawn.

As described here, our algorithm only ensures the input IMF is reproduced over the sequence of sampled star particles. It does not account for spatial information, so it does not guarantee that a sub-volume of a simulation will contain a population of stars that reproduce the IMF. In principle, multiple concurrent realisations of the underlying “book-keeping” can be used. For example, in a cosmological simulation it would make sense to define the “next star particle” as the next formed in the same galaxy, maintaining separate offset masses for individual systems, ensuring that the IMF is maintained in each. For a non-cosmological simulation the galaxy could likewise be divided into separate regions where the algorithm is applied independently, maintaining the IMF locally. However, in practice we find that this is unnecessary since it is extremely unlikely that applying the algorithm globally results in, for example, all the massive stars being formed at one side of the galaxy.

A regional division may be necessary if an input IMF that has a dependence on local conditions is used such that IMF varies strongly within a galaxy at a given moment in time. If the IMF varies slowly as a function of time (compared to the time-scale on which \( m_{\text{offset}} \) fluctuates, \( \sim m_{\text{part}}/M_\odot \)) and remains roughly uniform within a galaxy then this is not necessary. It should also be noted that our scheme as proposed does not permit the spatial clustering of stars as a function of their mass to be directly specified, with correlations simply arising from completely random sampling of the IMF, although this is unlikely to be of much concern at the resolution of global galaxy simulations.

Finally, we will briefly mention some other sampling methods that have been used in the literature that do not populate a star particle with a sequence of drawn stellar masses. In stratified box simulations, Gatto et al. (2017) populate sink particles with stellar masses drawn from the IMF. They only consider massive stars, sampling once from a mass range of \( 9 - 120 M_\odot \) every time \( 120 M_\odot \) is accreted onto the sink and assume the remaining mass forms lower mass stars. However, this scheme can produce an overall deficit of massive stars because sink particle masses are not in general a multiple of \( 120 M_\odot \). Geen et al. (2018) adopt a similar approach in simulations of GMCs. Sormani et al. (2017) present a scheme which instead divides the IMF into a number of bins and populates particles by performing Poisson sampling to determine the number of stars in each bin based on the expected number. This has the advantage of reproducing the IMF in an unbiased, albeit discretized, form across an ensemble of particles, at the cost of a discrepancy between the sampled and dynamical mass within a given particle. No additional correction is necessary to ensure the discrepancy sums to zero over a population of particles (in contrast to the adjusted target scheme), but the individual discrepancies can in theory be much larger than the previously described schemes (although increasingly large discrepancies represent more unlikely outcomes of the sampling). The discretized nature of the Sormani et al. (2017) scheme means that it incurs a penalty to computational cost when sampling and increasing memory requirements when the IMF is sampled at finer granularity (i.e. smaller mass bins), a problem not faced by other schemes in this section which allow for a continuum of stellar masses.

### 2.3 Monte Carlo tests of IMF sampling methods

We now present some idealized Monte Carlo tests of the four main IMF sampling schemes described in the previous section (stop before, stop after, stop nearest and adjusted target) to demonstrate their properties when used to populate star particles. Throughout this work we will use a Kroupa (2001) IMF with minimum and maximum stellar masses of \( m_{\text{min}} = 0.08 M_\odot \) and \( m_{\text{max}} = 100 M_\odot \), respectively. Our findings will hold in a qualitative sense for any reasonable input IMF. We will draw samples from the IMF for a total mass budget of \( 10^5 M_\odot \). We have confirmed that this is a sufficiently large amount of mass that it can be populated almost perfectly with stellar masses drawn from our input IMF (no matter what method is used) if no restrictions are applied. However, we will instead subdivide this mass reservoir into particles of mass \( m_{\text{part}} \) and use the four IMF sampling methods to populate them. This is representative of how star particles would be populated in a galaxy formation simulation. We will use various values of \( m_{\text{part}} \) to examine the resolution dependence of the four schemes.

In Fig. 2 we compare the total distribution of individual stellar masses across the various star particles to the analytic expectation from the input IMF. Specifically, we show the number of stars per logarithmic mass bin, \( \xi_L = dN/d\log_10 m_\star \). In the top panels we show \( \xi_L \) directly while in the bottom panels we show its ratio to the analytic expectation from the input IMF. The general trends are as follows. The stop before method results in a suppression of the high-mass end of the IMF while boosting the low-mass end, as expected. The maximum stellar mass that can be accepted with the stop before method is equal to the star particle mass, since any higher mass will carry it over the total. Even then, this is only true when it is the first mass drawn for the particle. In general, because a more massive star is more likely to carry the total assigned mass over the star particle mass than a less massive star, low mass draws are more likely to be accepted. When \( m_{\text{part}} = 20 M_\odot \), the shape of the output IMF is very distorted. There is a \( \sim 15\% \) enhancement of \( \xi_L \) for stars less than \( \sim 0.5 M_\odot \), followed by a sharp drop at higher masses with a net suppression above \( 3 M_\odot \). The biasing effect is less severe as \( m_{\text{part}} \) is increased because the relative importance of the final drawn stellar mass is reduced. However, there is still a strong suppression of the high-mass end of the IMF when \( m_{\text{part}} = 100 M_\odot \) and it is still noticeable when \( m_{\text{part}} = 500 M_\odot \).

The stop after method always perfectly reproduces the shape of the input IMF (modulo some noise at the high-mass end) because no draw is ever discarded. However, this always results in an overshoot of the particle mass meaning that the absolute number of stars is biased upwards. When \( m_{\text{part}} = 20 M_\odot \), this manifests as an enhancement of \( \xi_L \) by an average of 16% across the whole mass range. Again, this bias reduces as the particle mass increases, with an enhancement of 3.9% and 0.8% when \( m_{\text{part}} \) is increased to \( 100 M_\odot \) and \( 500 M_\odot \), respectively.

The stop nearest method suffers from the same issues as the stop before method, albeit to a lesser extent. In this method, a draw will be rejected if the drawn mass is more
As described in previous sections, IMF sampling schemes must balance preserving the input IMF with the two additional issues of minimizing the inconsistency between the assigned and dynamical masses for each individual star particle and conserving mass across the population of star particles. First, we shall examine the relative mass inconsistency, \( m_{\text{assign}} / m_{\text{part}} \), for the individual star particles populated in our Monte Carlo experiments. Fig. 3 shows PDFs for this ratio for our four trialled sampling methods with \( m_{\text{part}} = 20 \, M_\odot, 100 \, M_\odot \) and \( 500 \, M_\odot \). All distributions peak at a value of 1 i.e. where the assigned mass is consistent with the particle mass. The stop before has a tail that extends to lower values of the ratio i.e. representing particles that have less assigned mass than their particle mass. The reverse is true for stop after. These trends are an obvious result of the construction of the schemes. The stop before scheme can assign no more mass than \( m_{\text{part}} \) and no less mass than \( \text{MAX} (m_{\text{part}} - m_{\text{max}}, 0) \). Equivalently, the stop after scheme can assign no less mass than \( m_{\text{part}} \) and no more mass than \( m_{\text{part}} + m_{\text{max}} \). These limits can be seen in the PDFs and result in a narrowing of the distribution as \( m_{\text{part}} \) is increased.
The stop nearest method results in the tightest distribution of $m_{\text{assigned}}/m_{\text{part}}$ about unity of the four schemes. It is possible for the assigned mass to be below or above the particle mass. The minimum mass that can be assigned is $\text{MIN} \left( m_{\text{part}} - \frac{1}{2}m_{\text{max}}, 0 \right)$ while the maximum mass that can be assigned is $\text{MAX} \left( \frac{1}{2}m_{\text{max}}, m_{\text{part}} + \frac{1}{2}m_{\text{max}} \right)$. Again, the distribution tightens in a relative sense as $m_{\text{part}}$ increases because the impact of the last drawn stellar mass is reduced. The adjusted target scheme has the largest spread in $m_{\text{assigned}}/m_{\text{part}}$ of the four schemes. It too can assign more or less stellar mass than the particle mass. Its minimum assignable mass is the same as the minimum for the stop before scheme and its maximum assignable mass is the same as the maximum for the stop after scheme. It can be seen in Fig. 3 that the wings of the adjusted target PDFs trace those of the stop before and stop after PDFs. However, when $m_{\text{part}} = 20 M_{\odot}$, adjusted target produces more particles with no assigned mass (relative to stop before and stop nearest) because there will occasionally be a series of $m_{\text{assigned}} = 0$ particles to compensate for a massive star being drawn previously (as described in the previous section).

Fig. 4 shows how the choice of sampling scheme affects the ratio between the total assigned mass for the whole ensemble of particles, $M_{\text{assigned}} = \Sigma_j m_{\text{assigned},j}$, and the true mass budget, $M_{\text{true}} = 10^7 M_{\odot}$, for various values of $m_{\text{part}}$. As expected, the stop before scheme produces a deficit of assigned stellar masses (because it always throws away the last draw) while stop after correspondingly produces an excess (because it always keeps the last draw). This is most apparent at low values of $m_{\text{part}}$, but a small bias is still evident at $m_{\text{part}} = 500 M_{\odot}$. Stop nearest has a marginal mass deficit at the lowest $m_{\text{part}}$ but recovers the true mass by $m_{\text{part}} = 100 M_{\odot}$. The adjusted target always recovers the true mass budget even with $m_{\text{part}} = 20 M_{\odot}$. Of particular relevance to galaxy formation simulations is whether the total stellar feedback budget is recovered. Because the majority of the budget is generated by massive stars, the recovery of the amount of feedback per unit stellar mass from the input IMF is sensitive to how closely a sampling scheme reproduces the high-mass end of the IMF. In Fig. 4 we also show how the number of SN progenitors generated by our Monte Carlo experiments compares to the expected number. We assume that stellar masses in the range $8 - 35 M_{\odot}$ will produce core collapse SNe (we ignore Type Ia SNe). Again, as expected, the stop before and stop after schemes produce too few or too many SN progenitors, respectively, relative to the analytic expectation. The stop nearest scheme also produces too few SN progenitors with deficits of 23.8%, 7.8% and 3.8% for $m_{\text{part}} = 20 M_{\odot}, 50 M_{\odot}$ and $100 M_{\odot}$, respectively. It converges onto the correct number of progenitors within 1% for $500 M_{\odot}$. By contrast, the adjusted target scheme gives the correct number of SN progenitors within 0.5% or better across all the values of $m_{\text{part}}$ we test. Note that if the maximum core collapse SNe progenitor mass is increased, the biasing of the feedback budget will be worse. Likewise, feedback channels that depend strongly on the most massive stars (e.g. ionizing radiation) will show similar biases.

An examination of Fig. 2 and 4 may suggest a simpler version of our adjusted target scheme. Since the stop after scheme correctly recovers the shape of the IMF but has a constant offset across the entire mass range, it is possible to recover the normalisation by choosing a single target mass for all particles (lower than the particle dynamical mass) in advance to compensate for the overshoot, rather than correcting from particle to particle? We would set $m_{\text{target}} = \langle \xi_L, \text{analytic} \rangle m_{\text{part}}$, where we have averaged the expected and output mass distributions from the stop after Monte Carlo experiment (Fig. 2) across the stellar mass range. This does indeed correctly compensate for the overshoot and restore the correct IMF normalisation. However, it has several drawbacks. Firstly, while the normalisation is correct across many particles, over several consecutive particles the discrepancy between the dynamical mass and the assigned mass is more significant than our adjusted target scheme. This is because this scheme produces a mass discrepancy resulting from uncorrelated over- and undershoots, meaning that the ratio between the total dynamical and assigned masses is a random walk about unity.\(^2\) By contrast,\(^2\) Note that the decision not to correlate over- and undershoots from one particle to another does not avoid any perceived ‘action at a distance’. This effect is simply an implicit part of the renormalisation procedure, rather than being more explicit in the case of the adjusted target scheme. Regardless, as previously men-

\(^2\) Note that the decision not to correlate over- and undershoots from one particle to another does not avoid any perceived ‘action at a distance’. This effect is simply an implicit part of the renormalisation procedure, rather than being more explicit in the case of the adjusted target scheme. Regardless, as previously men-
by design the adjusted target scheme always pushes this ratio towards unity, resulting in smaller amplitude deviations (e.g. by a factor of ~3 for trials of 10 consecutively populated 20 M\_\odot particles). This quality may be important for regions with low SFRs in order to avoid significant under- or overestimations of the feedback budget.

Secondly, it is clear from Fig. 2 and 4 that the magnitude of the correction factor \((\xi_{\text{analytic}}/\xi_t)\) depends on particle mass. It also depends on the IMF shape, as well as the lower and upper bounds. Thus, a new correction factor must be obtained whenever these parameters are varied, either by performing another Monte Carlo experiment or by obtaining some functional form of this dependence. By contrast, our adjusted target scheme is independent of both the particle mass and the details of the IMF. Thus, it adaptively copes with variable IMFs with no modification (so long as the IMF does not change significantly between consecutive particles, instead varying smoothly with time as, for example, local metallicity gradually changes). The simpler scheme would require additional calculation of the correction factor from functional forms or lookup tables on-the-fly to treat an IMF that varied during a simulation. This is also likely to increase the level of noise mentioned in the previous paragraph as the scheme relies on compensating over a larger ensemble of particles. Additionally, even if a functional form for the dependence of the correction factor on star particle mass is obtained, the simplified scheme requires the star particle mass to be uniform throughout the simulation. This is not the case for all implementations of star formation. Many Lagrangian codes convert entire gas particles to star particles, meaning that a scatter in the gas particle mass results in a scatter in star particle mass. While gas particle masses may be uniform in the initial conditions, mass return from feedback processes or particle splitting/merging to achieve variable resolution will impart a mass scatter. As another example, in the pseudo-Lagrangian AREPO code (more details can be found in the next section), (de-)refinement operations keep the gas cell mass to within a factor of 2 of a predetermined target. Converting cells into star particles thus generates a scatter in the particle mass. Our adjusted target scheme adaptively compensates for varying particle mass.

To summarise, the step before, stop after and stop nearest sampling methods all fail to reproduce the input IMF except at large star particle masses. This has direct consequences for the stellar feedback budget, resulting in substantial under- or overestimation of the number of SN progenitors when \(m_{\text{part}}\) is less than \(\sim 100 M_\odot\). By contrast, the adjusted target scheme does not bias the input IMF in any way, produces a total assigned mass that is consistent with the total star particle mass and recovers the correct feedback budget. The sole advantage of the stop nearest method is that it is the best scheme for minimizing the discrepancy between the assigned and dynamical mass of an individual particle. However, this discrepancy is not particularly important as galaxy formation simulations cannot typically resolve exact N-body dynamics and the much larger inconsistency with the input IMF and the total stellar mass budget is a far worse penalty. Above \(m_{\text{part}} = 500 M_\odot\), the schemes largely converge, so in this mass range the stop nearest approach may be adopted simply because it can be implemented locally to each task in a parallel computation. However, implementing our improved scheme is not substantially more complex and carries negligible additional computational penalty. Our scheme can be integrated into any pre-existing star formation scheme that uses star particles with a minimum of effort because it does not require additional modification of the mass of star particles or the transfer of mass between star particles.

### 3 Explicit IMF Sampling vs. IMF Averaging in Simulations

In this section we will study the impact on galaxy evolution simulations of using explicit IMF sampling to determine the stellar feedback budgets of star particles as opposed to the more common IMF averaged approach. We will present simulations of an idealized isolated dwarf galaxy with a baryonic mass resolution of \(20 M_\odot\). We include stellar feedback in the form of SNe, photoionization and photoheating in HI regions and photoelectric heating of dust grains. Detailed descriptions of these models can be found in Smith et al. (2020) (hereafter Paper I), but we will summarise the salient details in the following section. For convenience, we will denote simulations with SNe, ionizing radiation and photoelectric heating by the abbreviations SN, PI and PE, respectively. Simulations with all feedback channels switched on are denoted as SN-PI-PE while NoFB is adopted when no feedback is used. We will refer to simulations that explicitly sample the IMF by the abbreviation ‘IMFsam’ and simulations that use IMF averaged stellar feedback as ‘IMFav’. The IMFsam simulations were originally presented in Paper I.

#### 3.1 Numerical Methods

We use the moving-mesh code AREPO (Springel 2010; Pakmor et al. 2016) along with our own sub-grid models for star formation and stellar feedback (which are described in more detail in Paper I). We use the GRACKLE chemistry and cooling library\(^7\) (Smith et al. 2017) in its primordial six-species non-equilibrium mode, along with tabulated metal cooling, ionization and heating from a meta-galactic UV background (Haardt & Madau 2012) and the self-shielding prescription of Rahmati et al. (2013). Gas cells can have a non-zero SFR when their Jeans mass drops below 8 times the cell mass. The SFR is then given by a simple Schmidt law, \(\dot{\rho}_\star = \epsilon_{\text{SF}} \rho / t_f\), where \(\rho\) is the gas density, \(t_f = \sqrt{5\pi/32G\rho}\) is the local free-fall time and we adopt a fixed efficiency of \(\epsilon_{\text{SF}} = 0.02\), motivated by observed efficiencies in dense gas (see e.g. Krumholz & Tan 2007, and references therein). The SFR is then stochastically sampled to convert gas cells into collisionless star particles.

We include photoelectric heating of dust grains by a spatially varying far-UV (FUV) field generated by star particles. The FUV energy density at each location in the domain is calculated using the gravity tree to sum the fluxes from sources, using a local approximation for attenuation by dust which is valid in dust-poor systems. We approximate the dust-to-gas ratio as a function of metallicity using the broken

\[^7\]https://grackle.readthedocs.io
power-law of Rémy-Ruyer et al. (2014). H II regions around ionizing sources are included using a novel anisotropic overlapping Strömgren type approximation (first presented in Paper I). The balance between the ionizing photon luminosity and the recombinantion rate is calculated in independent angular pixels around sources to determine the extent of an H II. This helps mitigate the mass-biasing error encountered by previous methods. The algorithm accounts for H II regions from multiple sources overlapping. If a cell is tagged as belonging to an H II region, it is immediately heated to 10^4 K and is forbidden from cooling below that temperature while it remains tagged. SN feedback is included with the scheme first presented in Smith et al. (2018). This injects mass, metals, energy and momentum into the gas cell containing a star particle and its immediate neighbours (those that share a face with the host cell). The scheme ensures an isotropic injection of feedback quantities which is non-trivial in a Lagrangian code. We use the SN scheme in its mechanical feedback mode which compensates for missing momentum when the Sedov-Taylor phase of a SN remnant is unresolved. In practice, at the resolution we adopt in this work the majority of SNe are well resolved and we find that we achieve similar results when a simple thermal dump of SN energy is used. In this work, we do not include stellar winds.

When the explicit IMF sampling scheme is used (IMFsam), star particles are populated with an inventory of stellar masses at the moment of creation using the adjusted target scheme described in Section 2. Rejection sampling is used to draw stellar masses from the IMF. We use a single value of the offset mass for the whole galaxy at any one time, rather than concurrent versions for different regions of the galaxy (see the discussion in Section 2.2). We do not explicitly record the masses of stars less than 5 M☉ since they would contribute negligible feedback (as we do not include winds from AGB stars), although they are taken into account for the purposes of determining the total stellar mass assigned to the star particle. For stars more massive than 5 M☉, we then use lookup tables as a function of mass in order to determine quantities relevant to feedback. We make the simplifying assumption in this work that all stars have a metallicity of 0.1 Z⊙, the initial metallicity of our initial conditions. None of the quantities used have a strong dependence on metallicity relative to the deviation from the initial metallicity that we see in this work, so this approximation is reasonable. We obtain the lifetime of the star from the PARSEC grid of stellar tracks (Bressan et al. 2012). The FUV and ionizing photon luminosities of the star are derived from the OSTAR2002 grid of stellar models Lanz & Hubeny (2003) as compiled by Emerick et al. (2019), making the approximation that they are fixed at their main sequence values throughout the life of the star. The net luminosity of a star particle is the sum of all extant stars assigned to it. When a star reaches the end of its life it ceases to contribute radiation. Star particles with extant massive stars have their time-steps limited to 0.1 Myr, although in practice their time-step is usually much smaller as constrained by other criteria. If a star in the mass range 8 – 35 M☉ reaches the end of its life, a SN is triggered, resolved from the host star particle. The ejecta mass and metallicity depends on the progenitor mass, based on Chieffi & Limongi (2004). We use a constant SN energy of 10^{51} ergs for all SNe. When SN feedback is switched off, we still return ejecta when a SN progenitor reaches the end of its life but do not inject the energy.

For simulations using IMF averaged feedback quantities (IMF sam) the sampling procedure is not performed. Instead, lookup tables of FUV and ionizing photon luminosities, and SN rates per unit stellar mass as a function of population age are used. For consistency, these tables are derived using the same relationships between individual stellar mass and feedback quantities that are used in the IMFsam schemes. We begin by populating a 10^8 M⊙ SSP with stellar masses drawn from the IMF, the high mass budget guaranteeing that the IMF is very well sampled. Lifetimes and luminosities are obtained in a similar manner to the on-the-fly approach used in the IMFsam schemes. We then integrate through the lifetime of the SSP, recording how the net FUV and ionizing photon luminosities decrease as stars reach the end of their lives. We also obtain SN rates by recording when stars in the 8 – 35 M⊙ mass range die. We then normalize these luminosities and rates by the initial mass of the SSP. When a simulation is performed, star particles are assigned their luminosities using these lookup tables based on their age. The expected number of SNe that will be produced by a star particle in a given time-step, ∆t, is

\[ \tilde{N}_{\text{SN}} = \bar{n}_{\text{SN}} \left( t_{\text{part}} \right) m_{\text{part}, 0} \Delta t, \]

where \( \bar{n}_{\text{SN}} \left( t_{\text{part}} \right) \) is the SN rate per unit mass obtained from the lookup tables as function of star particle age and \( m_{\text{part}, 0} \) is the initial star particle mass at the point of creation, before mass loss due to feedback. The SN rate is normalized by \( m_{\text{part}, 0} \) rather than \( m_{\text{part}} \) in order to recover the normalization used when creating the rate tables from the SSP. In other words, SNe are independent events representing an implicit sampling of the IMF and so the chances of a SN being generated from a star particle should not be reduced because a previous SN occurred and reduced the mass of the particle. Likewise, luminosities obtained from the lookup tables are normalized by the initial particle mass, not the current mass. In constructing the IMF averaged luminosities we have already explicitly linked the radiation output to the initial mass of the SSP and it is therefore independent of the SN events that have occurred in the star particle and the ensuing mass loss. The number of SNe that occur in a given time-step is then determined by drawing from a Poisson distribution with a mean \( \lambda = \tilde{N}_{\text{SN}} \). In Section 4 we will discuss in more detail why Poisson sampling is the appropriate procedure. If a SN occurs, we map the age of the star particle back onto the mass of the progenitor to obtain the ejecta mass and metallicity. Like the IMFsam schemes, star particles have a maximum time-step of 0.1 Myr enforced. This ensures the changes of luminosity are time-resolved, but also guarantees that \( \tilde{N}_{\text{SN}} \ll 1 \). This is necessary to ensure that
SNe are individually time-resolved, which is important to accurately capture clustering effects.

It is possible that a star particle may not have sufficient mass to return as ejecta for a triggered SN event. For the IMFsam scheme this is a result of the potential discrepancy between the assigned and dynamical mass of the star particle. For the IMFav scheme this is because the stochastic sampling of the rates means that it is possible, albeit unlikely, for a single star particle to produce more SNe than its mass should allow. If there is insufficient mass in a star particle to return as ejecta, we instead return as much as possible (i.e. the mass of the particle) and remove the particle. As reported in Paper I, this is a relatively rare occurrence and results in an overall reduction of the total ejecta across all SNe by 5.1% for a 20 $M_\odot$ star particle. Only 2% of SNe have their ejecta mass reduced by more than 30% and none have their ejecta reduced by more than 36%. This may be of concern if a detailed study of chemical enrichment is of interest (ignoring uncertainties in yields) but for this work we find this level of deviation from the tabulated yields acceptable. It is also possible that in the event of an entire star particle being removed due to a lack of ejecta mass another massive star hosted in the particle is prematurely deleted before it has reached the end of its life. This occurs in less than 0.1% of the SN events in our simulations.

### 3.2 Simulation details

For our idealized dwarf galaxy we use the ‘fiducial’ initial conditions from Paper I, generated using the MAKE-NEW-DISK code (Springel et al. 2005). The system has a virial mass of $10^{10} M_\odot$. There is a $6.825 \times 10^7 M_\odot$ gas disc with an exponential radial density profile with a scale length of 1.1 kpc. The vertical structure is set to achieve hydrostatic equilibrium at an initial temperature of $10^4$ K. The gas is initialized with a metallicity of 0.1 $Z_\odot$. We do not include a circumgalactic medium (CGM). There is also an initial stellar disc with a mass of 9.75 $\times 10^7 M_\odot$ and the same radial density profile as the gas disc. The vertical structure is Gaussian with a scale height of 0.7 kpc. Star particles present in the initial conditions do not contribute stellar feedback. The rest of the mass of the system is in the form of a live, spherically symmetric dark matter halo. This follows a Hernquist (1990) density profile chosen to provide a close match to a Navarro et al. (1997) profile with a concentration parameter, c, of 15 and a spin parameter, $\lambda$, of 0.04. Gas cells and star particles have a mass of 20 $M_\odot$. (deregiment and refinement operations keep the gas cells within a factor of 2 of this target mass), while dark matter particles have a mass of 1640 $M_\odot$. Gravitational softening lengths are fixed at 1.75 pc and 20 pc for star and dark matter particles, respectively. Gas cells use adaptive softening lengths down to a minimum of 1.75 pc. The initial conditions are relaxed for 100 Myr with cooling but without star formation while initial turbulence is driven with a pseudo-SN feedback scheme (described in detail Paper I). This is to avoid a rapid vertical collapse of the disc when the simulations are started.

### 3.3 Results

Fig. 5 shows the SFR for the simulations, averaged over 10 Myr. Dashed lines show results from simulations with IMF averaged feedback (IMFav) while solid lines show simulations with explicit IMF sampling (IMFsam). This convention is used throughout this work. Without feedback, the SFR rises rapidly and is only limited by the supply of gas (we only perform an IMFsam NoFB simulation). When feedback is included, the results are qualitatively the same between IMFav and IMFsam simulations. A detailed discussion about how the various feedback schemes regulate the SFR can be found in Paper I. As found in that paper, photoelectric heating by itself is inefficient, providing little suppression of star formation. It is slightly more efficient in the IMFsam simulation, reducing the peak SFR reached, but this is an extremely marginal effect. SN feedback is able to suppress the SFR by almost two orders of magnitude on average. It does so in a very bursty manner. The IMFav and IMFsam produce nearly identical results. Differences between the two are well within the margin of numerical noise arising from stochastic sampling/triggering of SNe. We confirm this in Appendix A, where we repeat the early stages of these simulations four additional times with different seeds for the random number generators and with randomly perturbed initial conditions.

Photoionization feedback is also able to regulate SFRs (see Paper I for a detailed discussion of why this occurs). It produces a smoother SF history as it disrupts star forming regions locally in a more gentle manner. When averaged over the last 500 Myr of the simulation (allowing the initial transient to settle), the IMFsam simulation regulates the SFR to roughly the same degree as the SN feedback simulations. However, the IMFav reduces the SFR by a factor of 2.4 relative to the IMFsam simulation. When all feedback channels are turned on (SN-PI-PE) with explicit IMF sampling, SFRs are reduced by ~ 40% relative to the SN or photoionization only simulations, but this is small compared to the initial reduction from the no feedback case. When the IMF averaged values are used instead, switching on all the feedback results in a reduction of 26% relative to the IMFav PI simulation but 73% relative to the IMFsam SN simulation. Photoionization is therefore the dominant regu-
Figure 6. Outflow rates across thin slabs parallel to the disc at 1 kpc above and below the disc midplane. This top panel shows the absolute rates while the bottom panel normalises by the SFR (as in Fig. 5) to produce mass loading factors. In line with the findings of Paper I, the addition of photoionization feedback diminishes SN-driven winds in these simulations. This effect is more pronounced for the IMF averaged feedback case.

Fig. 6 shows the mass outflow rate measured through parallel slabs 1 kpc above and below the disc midplane. This is measured as:

\[
\dot{M}_{\text{out}} = \sum_i \frac{\dot{M}_{\text{out},i}}{\Delta \bar{\rho}},
\]

(2)

where the summation is carried out over all gas cells with a positive outflow velocity, \(v_{\text{out}}\), perpendicular to the disc plane located within a slab of thickness \(\Delta z = 100\) pc. Fig. 6 shows both the absolute rates and mass loading factors. The latter is obtained by normalising the absolute rates by the SFR averaged over the preceding 10 Myr as shown in Fig. 5. The results for the IMFsam simulations were presented and discussed at length in Paper I. We refer the reader to that work for a detailed explanation of the impact of various combinations of feedback channels on outflow rates, but give the salient details here. In the absence of efficient feedback (i.e. the NoFB or PE simulations) there is a small flow of gas outwards through 1 kpc, arising from the settling of the idealised initial conditions (note that there is no CGM). The addition of photoionization feedback leads to a small enhancement of this outward flow due to additional thermal support and the momentum input from expanding HII regions. The impact on the apparent mass loading factor is large due to the significant reduction in the SFR. However, because the outflow is only a minor enhancement of a flow that existed in the NoFB simulation, caution should be adopted before interpreting the mass loading factor as indicating the strength of a feedback driven wind in this case.

SN feedback alone leads to strong bursts of outflows with mass loading factors of 10–100 for both IMFav and IMFsam simulations. However, combining all the feedback channels (SN-PI-PE) leads to a suppression of outflow rates by roughly an order of magnitude. As demonstrated in Paper I, this is due to a reduction in the clustering of SNe in both space and time by the pre-SN photoionization feedback. This effect is significantly more pronounced for the IMFav case, with essentially no enhancement of absolute outflow rates relative to the simulations without SNe and a significant deficit relative to the equivalent IMFsam run. The reduction in absolute outflow rate is mainly a result of the lower overall SFR in these runs, leading to a correspondingly lower SN budget to drive outflows. We might also expect a reduction in the loading factor due to even more significant reduction in SN clustering, but it is comparable between IMFav and IMFsam. Again, this is because in the absence of efficient SN feedback the mild outflow we see across 1 kpc is only weakly related to stellar feedback, so the variation of the mass loading factor is dominated by the differing SFR. Note that the mass loadings for the SN, PI and PE simulations were presented and discussed at length in Paper I. We refer the reader to that work for a detailed explanation of the impact of various combinations of feedback channels on outflow rates, but give the salient details here. In the absence of efficient feedback (i.e. the NoFB or PE simulations) there is a small flow of gas outwards through 1 kpc, arising from the settling of the idealised initial conditions (note that there is no CGM). The addition of photoionization feedback leads to a small enhancement of this outward flow due to additional thermal support and the momentum input from expanding HII regions. The impact on the apparent mass loading factor is large due to the significant reduction in the SFR. However, because the outflow is only a minor enhancement of a flow that existed in the NoFB simulation, caution should be adopted before interpreting the mass loading factor as indicating the strength of a feedback driven wind in this case.
**Figure 8.** Face-on and edge-on projections of the gas discs at 1 Gyr. *NoFB* and *PE* simulations are not shown. SN feedback alone produces a thick and highly disordered disc, with large transient cavities caused by SN superbubbles. When ionizing radiation is the only source of feedback the discs are more ordered and the large holes are not evident. The *SN-PI-PE IMFav* simulation is qualitatively the same as the *PI IMFav* run, but when explicit IMF sampling is used small, transient SN-driven cavities appear marginally more frequently.

*SN-PI-PE* simulations (with both IMF schemes) are all consistent with the range provided by observations of outflows from dwarfs (see e.g. McQuinn et al. 2019), although strong mass loadings from low mass galaxies are often required by theory (again, we refer the reader to the discussion in Paper I).

Fig. 7 shows PDFs for the ambient density where star particles are created and where SNe occur. We exclude the first 400 Myr to ignore the initial transient phase of the simulation. Our Jeans mass based star formation criteria means that the onset of star formation occurs between $\sim 20-100$ cm$^{-3}$. Simulations without feedback or with photoelectric heating alone peak at around a density of $10^4$ cm$^{-3}$. In simulations where SN feedback is switched off, we still record where SNe would occur either based on a SN progenitor reaching the end of its life (in the *IMFsam* simulations) or by sampling the SN rates (in the *IMFav* simulations). As noted in Section 3.1 we also return ejecta mass but do not add SN energy as gas continues to collapse beyond the onset of star formation. The difference between the *IMFav* and *IMFsam* photoelectric heating runs is simply a result of the different gas fractions in the disc due to the offset in SFR peaks in the first 400 Myr, essentially producing an offset of the disc evolution in time. When SN feedback is used alone, the PDF of star formation ambient densities is broadened. This is because the feedback alters the distribution of dense gas by driving turbulence and disrupting collapsing clouds. *IMFav* and *IMFsam* give identical PDFs. When ionizing radiation is included the peak of the distribution is moved to lower densities and the maximum density reached is reduced. This is because the radiation is able to halt the collapse of dense clouds and disrupt them earlier than the SN feedback (which is delayed by the lifetime of its progenitors). It can be seen that this effect is significantly more pronounced in the *IMFav* than the *IMFsam* simulations, a key difference between the two methods.

In the absence of feedback or with photoelectric heating
Figure 9. The surface density of \( \text{H}_\text{II} \) divided by the total hydrogen density at 1 Gyr, shown face-on and edge-on. \( \text{NoFB} \) and \( \text{PE} \) simulations are not shown. Generally, the ionized regions trace the low density gas. Patches of high ionization fraction can be seen embedded in dense gas in the simulations with photoionization feedback. These are \( \text{H}_\text{II} \) regions around ionizing sources. The \( \text{IMF}_{\text{av}} \) simulations produces many small \( \text{H}_\text{II} \) regions scattered throughout the disc compared to the \( \text{IMF}_{\text{sam}} \) approach which features rarer, but larger \( \text{H}_\text{II} \) regions.

alone, the PDFs of the ambient density where SN progenitors die reflects the PDFs of star particle birth densities with a broadening towards lower densities. This lower density tail arises because of runaway gas consumption in star forming clouds, dropping the local density, as well as being caused by star particles drifting out of their (now very compact) birth clouds. When SN feedback is used alone, the PDF of SN site densities spans roughly ten orders of magnitude in density. The first SNe to occur in a star forming cloud explode in the dense gas of star forming regions. They are able to disperse these dense clouds and successive SNe contribute to the creation of a superbubble, with each subsequent SN occurring in lower density gas. This gives rise to the broad range of ambient densities. Again, the simulation using IMF averaged SN rates and that explicitly sampling the IMF for SN progenitors produce essentially identical PDFs. When ionizing radiation is included (either on its own or with the other feedback channels) almost no SNe occur in star forming gas. The radiation is able to clear dense gas prior to SNe occurring. The PDFs are similar for the \( \text{IMF}_{\text{av}} \) and \( \text{IMF}_{\text{sam}} \) simulations. The low density tail does not extend as far as the SN only simulations. This is because the creation of superbubbles is inhibited by the ionizing radiation acting to reduce the clustering of SNe, as shown in greater detail in \textit{Paper I}. The \( \text{SN-PI-PE \ IMF}_{\text{sam}} \) PDF extends to slightly lower densities than the \( \text{IMF}_{\text{av}} \) equivalent because the impact of the radiation on clustering is not as strong in this case, allowing some SNe to occur in larger bubbles.

Fig. 8 shows face-on and edge-on projections of the gas disc after 1 Gyr. Simulations \( \text{NoFB} \) and \( \text{PE} \) are not shown, but equivalent plots can be found in \textit{Paper I}. In these simulations a large proportion of the gas has been consumed by 1 Gyr, leaving isolated, extremely compact knots of dense gas within a low density ambient medium. The gas discs in those simulations are extremely thin. For the simulations that we do show in Fig. 8, the \( \text{IMF}_{\text{av}} \) and \( \text{IMF}_{\text{sam}} \) simulations are qualitatively similar. When SNe are the only source of feedback, the disc is thick and highly disordered. SNe su-
The mass of gas tagged as belonging to an H\textsc{ii} region normalized by the mass in young stars. The instantaneous value (shown in the paler colours) is very noisy so we also plot a 50 Myr moving average (the bold colours). The IMF\text{av} simulations produce almost twice as much photoionized gas per unit stellar mass as the IMF\text{sam} equivalents, because sources are typically embedded in lower density gas.

Fig. 8 shows the ratio between ionized and total hydrogen surface densities, face-on and edge-on. It is therefore a form of projected ionization fraction. With reference to Fig. 8, it can be seen that ionized regions largely trace the diffuse gas. However, in the simulations with photoionization feedback there are small regions of high ionization fraction embedded in dense gas. These are H\textsc{ii} regions around ionizing sources. There is a qualitative difference in the distribution and size of H\textsc{ii} regions between the IMF\text{av} and IMF\text{sam} simulations. The former simulations contain many small H\textsc{ii} regions, speckled throughout the disc. The IMF\text{sam} simulations produce fewer, but larger H\textsc{ii} regions. This difference is not a transient effect and persists throughout the course of the simulations.

The reason for this difference lies in the differing discretization of the ionizing sources when IMF averages are used as opposed to explicitly tracking the emission from individual massive stars. With the former approach, every star particle of the same age emits the same ionizing flux. However, the IMF\text{sam} scheme correctly ties the origin of ionizing photons to comparatively rare, bright sources. The total luminosity of ionizing radiation per unit stellar mass is the same in both methods, by construction. However, the distribution of this luminosity between sources makes a significant difference to the ability to photoionize gas. Despite having approximately half the average global SFR, the IMF\text{av} simulations with photoionization feedback keep roughly the same mass of gas photoionized as the IMF\text{sam} simulations. This is illustrated in Fig. 10 where we plot the ratio between the photoionized mass (tagged as belonging to an H\textsc{ii} region by our sub-grid model) and the mass in star particles younger than 10 Myr. The pale lines show the instantaneous value at every output time (5 Myr) but as this is extremely noisy we also plot a 50 Myr moving average. Fig. 10 demonstrates that stars are almost twice as efficient at creating H\textsc{ii} regions when the IMF averaged luminosities are assigned to star particles.

The mass of gas in an H\textsc{ii} region is determined by the balance between photoionization and recombination. It is therefore linearly proportional to the ionizing photon rate and inversely proportional to the square of the ambient density. The net ionizing photon rate per unit stellar mass is the same regardless of the discretization method adopted, so Fig. 10 implies that ionizing sources are typically embedded in moderately denser gas in the IMF\text{sam} case. Indeed, Fig. 7 indicates this to be the case. The ionizing photon budget is dominated by rare, massive stars. With the IMF averaging approach, every star particle is an ionizing source. A 20 M\text{☉} star particle will initially produce photoionizing photons at a rate of approximately $10^{48}$ s$^{-1}$, this rate only beginning to decrease after approximately 3 Myr, corresponding to the lifetime of the most massive stars. An ionizing photon rate of $10^{48}$ s$^{-1}$ corresponds to a roughly 16 M\text{☉} star. Once even a small quantity of gas is photoionized, the resulting D-type expansion allows the source to grow an H\textsc{ii} region. It will be able to photoionize a 20 M\text{☉} gas cell as long the gas density is less dense than 216 cm$^{-3}$. We find that for the IMF\text{av} SN-PI-PE simulation 53% of the total ionizing luminosity is emitted by sources for which the resulting H\textsc{ii} region is resolvable by at least one cell at birth. Despite half the ionizing photons being ‘wasted’ by under-resolution, this feedback is still highly efficient. Because every star particle immediately begins emitting radiation upon creation, as soon as a star forming region creates a single star particle the further collapse of the cloud can begin to be arrested. This effect leads to the shifting of the birth density PDF towards the SF threshold in Fig. 7.

By contrast, even though explicit IMF sampling leads to far brighter sources they are significantly rarer. Only 8% of 20 M\text{☉} star particles are assigned at least one star more massive than 16 M\text{☉} when the IMF\text{sam} scheme is used. This means that on average a proportionally larger amount of stellar mass has to be created before an ionizing source capable of affecting the star forming cloud is born, resulting in a higher effective cloud-scale star formation efficiency. This has the added effect that star forming clumps can collapse to higher densities, as seen in Fig. 7. This means that the sources are more likely to be born in higher density gas, resulting in a lower total H\textsc{ii} region mass for the same net ionizing luminosity per unit stellar mass. It is important to note that despite being embedded in typically higher density gas, the higher luminosity sources produced by the IMF\text{sam} approach means that 85% of the total ionizing luminosity is associated with resolved H\textsc{ii} regions (by one cell) at birth, compared to 53% (i.e. the opposite trend to what would be expected if resolution effects were driving the difference between the approaches). In summary, while the IMF\text{av} simulations produce many small H\textsc{ii} regions around low brightness sources, the IMF\text{sam} simulations produce fewer, but...
larger, H\textsc{ii} regions around massive stars. For the reasons explained above, the latter scenario is less efficient at regulating star formation.

4 DISCUSSION

We now discuss our findings in greater detail and place them in the context of some other relevant works. In Section 4.1 we review best practices for schemes that trigger discrete SNe from IMF averaged rates. We show that if SNe are the only form of feedback, this will yield identical results to explicit IMF sampling, which is what we see in our simulations. In Section 4.2 we discuss our contrasting result that the photoionization feedback is in fact sensitive to the choice of method. We consider what these findings might mean for other non-SN feedback channels and how this sensitivity depends on resolution. We also explore whether a simpler toy model can capture the effects produced by explicit IMF sampling.

4.1 Insensitivity of supernova feedback to IMF averaging vs IMF sampling

4.1.1 Best practices for discretizing IMF averaged SN rates

Mac Low & McCray (1988) showed that as long as the number of SNe a star particle will produce over its lifetime is more than $\sim 10$ then the injection of SN feedback can be modelled as a continuous injection of energy. However, at higher resolution resolving the effects of discrete SNe becomes important (see e.g. Su et al. 2018; Applebaum et al. 2020). Keller & Kruijssen (2020) also demonstrate that simply injecting the entire feedback budget of the star particle in one event also deviates from simulations that account for a spread of SN events as a function of time. The impact of SN feedback is highly dependent on the clustering properties of the SNe (Sharma et al. 2014; Yadav et al. 2017; Gentry et al. 2017, 2019; Fielding et al. 2017, 2018; El-Badry et al. 2019 and Paper I). It is therefore important that if explicit IMF sampling is not used SNe must still be modelled as individual events sampled from the IMF averaged SN rates (if the resolution is high enough to detect the effects).

After an expected number of SNe, $N_{SN}$, that will occur in a given time-step has been determined from the SN rate (in general a function of star particle age and metallicity), initial particle mass and time-step size (as described in Section 3.1), there are two main approaches to converting these rates into discrete SN events. In the first approach, Bernoulli trials can be carried out every time-step to determine whether a SN occurs. This essentially represents the flipping of a biased coin. If $N_{SN} \leq 1$, then a Bernoulli trial is carried out with a probability of success equal to $N_{SN}$ to decide whether a single SN occurs. If $N_{SN} > 1$, then the number of SNe that occur is at least equal to the integer part of $N_{SN}$ with a Bernoulli trial carried out with a probability equal to the fractional part of $N_{SN}$ to determine whether an additional SN is added (see e.g. Stinson et al. 2010 and the RIMFS scheme of Revaz et al. 2016). This approach enforces a relatively smooth sampling of the rates, implicitly assuming that SN events are spread out evenly through whatever interval was used to determine the SN rates in the first place. When star particle masses are small, it also implicitly assumes that massive stars are distributed evenly across the particles which is essentially equivalent to assuming that the IMF is well sampled in each particle. In other words, these schemes capture the mean SN rates but do not capture the noise that emerges when considering a truly random sampling of an IMF in a single star particle.

Sampling from the Poisson distribution captures this noise. The probability of $N_{SN}$ occurring given $N_{SN}$ is

$$P_{Poiss}(k = N_{SN}) = \frac{N_{SN}^k e^{-N_{SN}}}{k!}.$$  

Sampling from this distribution will produce a number of SN events that are scattered about $N_{SN}$, capturing the noise, in contrast to the Bernoulli trial method.

It is worth pointing out that there is a great deal of confusion in the literature between conducting Bernoulli trials and sampling the binomial distribution, with the latter frequently being used when referring to the former. The binomial distribution describes the probability of achieving a given number of successes from a finite number of independent Bernoulli trials that have the same probability of success. In the limit that the number of trials is one, the binomial distribution converges to the Bernoulli distribution, but this is a trivial statement. Alternatively, if multiple time-steps all have the same SN rate then the total number of SN produced across all the time-steps will represent a draw from the binomial distribution. But one cannot “sample from the binomial distribution” to produce a number of SNe given an expected number of SNe for an individual star particle. This confusion also frequently leads to the statement that Bernoulli trial and Poisson sampling schemes are generally equivalent, because the binomial and Poisson distributions converge in the limit of many trials with low probability of success. This is true across many time-steps if $N_{SN}$ is sufficiently small and the SN rate does not change between each time-step. However, within each of the small time-steps the Bernoulli trial and the Poisson draw are only equivalent when $N_{SN}$ is very small such that $P_{Poiss}(N_{SN} > 1) \rightarrow 0$, which is a subtly different convergence criterion. Thus, both schemes may produce the same number of SNe averaged over some long timescale, but the temporal clustering will be different, unless this criteria is met. In practice, this limit is frequently desirable because in order to correctly capture the effects of SN clustering we wish to treat SNe individually. However, it is still more formally correct to draw from a Poisson distribution (technically allowing for the possibility that more than one SNe occur in a given time-step) but impose a time-step limiter to ensure that $N_{SN}$ is very small such that it is extremely unlikely that more than one SNe occurs. That is the approach we take in this work.

6 Which may or may not be desirable, depending on the manner in which one believes the IMF to be sampled in nature. Nonetheless, it is inconsistent with the form of explicit IMF sampling presented in the majority of this work.

7 This is because Bernoulli trials are sometimes referred to as binomial trials (as distinct from, but related to, the binomial distribution). Anecdotal evidence gathered by tracing such references back to the details of published methods papers or publicly available code, where available, suggests that references to sampling from a binomial distribution most likely always refer to a Bernoulli trial scheme. For clarity, we will consistently refer to Bernoulli trials in this work, even when citing authors that prefer the alternate phrasing.
produce a different clustering of SNe than the explicit sampling scheme. As we have already argued, this should not be the case if the schemes are constructed in a consistent manner. They carry out a Monte Carlo experiment to compare the timing of SNe produced from star particles that are populated with stars from the IMF as opposed to those that sample the IMF averaged SN rates. With their adopted particle mass of $420 \, M_\odot$ and time-step of 1 Myr, they find that approximately 25% of particles will experience at least one time-step in which more than one SN occurs when the explicit IMF sampling is used. When they use Bernoulli trials to sample the rates this never happens (because $N_{SN}$ is always less than one). Applebaum et al. (2020) note that they could possibly have seen different results if they had adopted Poisson sampling instead of their Bernoulli trial scheme.

When we repeat their experiment with Poisson sampling instead of Bernoulli trials, we find that the fraction of particles experiencing multiple SNe in a time-step is identical to the explicitly sampled IMF particles, to within 0.3%. Furthermore, we demonstrate in Fig. 11 that the SN rates that emerge from the particles that are populated from the IMF are described very well by Poisson distributions. For a given 1 Myr time-step, we plot the expected number of SNe, $\bar{N}_{SN}$, and the probability that a particle will produce one, two or three SNe in that time-step, assuming a Poisson distribution with a mean of $\bar{N}_{SN}$. We over-plot (in dashed lines) the fraction of star particles from our Monte Carlo experiment (which total $10^7 \, M_\odot$) that produce that number of SNe in that time-step. It can be seen that the two agree very well (although the $N_{SN} = 3$ fractions are very noisy because of the rarity of such an event). Indeed, we find that the number of SNe in a time-step are well described by Poisson distributions whatever the choice of particle mass and time-step, in contrast to a Bernoulli trial scheme.

With star particles of mass $420 \, M_\odot$, a time-step of 1 Myr is too large for the Bernoulli trial approach to converge with the more consistent Poisson distribution. However, the discrepancy still appears to be relatively small. It is possible that the use of a ‘blast-wave feedback’ model (Stinson et al. 2006) exacerbates the effect since the adopted sub-grid evolution of the SN remnant (which affects the length of time for which cooling is shut off in this model) is different for a single injection of the energy of multiple SNe compared to spreading those events into multiple independent injections. Regardless, if the SN clustering is the primary cause of the difference between the quantized feedback and explicitly sampled IMF schemes, then it results from an inconsistent implementation of the SN rate sampling. We suspect that the difference is more likely to be caused by the first explanation (differences in the distribution of LW sources). Whatever the cause, we feel that the results of Applebaum et al. (2020) should only be interpreted as reinforcing the need for explicit IMF sampling when pre-SN feedback is included. They do not demonstrate that the stochastic sampling of IMF averaged SN rates is an inferior approach to explicit IMF sampling in the absence of pre-SN feedback channels.

The fact that the distribution and timing of SN events among star particles is identical when we use quantized feedback or explicit IMF sampling also shows that the stellar mass implied by the SN events is consistent with the dynamical mass to the same degree in both cases. Both quantized sampling of SN rates and explicit IMF sampling allow the
fraction of star particle mass occupied by low mass stars to vary, by construction. Some particles will have more massive stars per unit stellar mass than the IMF average, while others have correspondingly fewer. However, in the case of quantized sampling of SN rates, the inventory of the star particle is not known until a SN occurs, indicating that a SN progenitor implicitly existed in the particle from its birth. This means that it is impossible for the level of non-SN feedback (which will also be pre-SN) generated by the star particle to make consistent with the stellar inventory that is ‘discovered’ once SNe have occurred. This can be thought of as an inconsistency between the implied distributions of stellar masses responsible for the SN and non-SN feedback (see Applebaum et al. 2020, fig. 3 for an illustration of this effect). With explicit IMF sampling, the stellar inventory is known from the birth of the particle, so the level of pre-SN feedback can be made consistent with the number and timing of the SNe produced by the particle. In the absence of non-SN feedback, quantized SN feedback and explicit IMF sampling are equivalent.

To summarise this section, we find that if SNe are the only source of stellar feedback, triggering discrete, individual SNe from IMF averaged rates gives identical results to explicit sampling of the IMF at the moment of a star particle’s creation (assuming the sampling can be carried out without biasing the IMF, as described in Section 2.2). We caution that this is only the case when the time-step in which the sampling is carried out is sufficiently small and that, in general, sampling from a Poisson distribution is the correct procedure. We have demonstrated this with our simulations shown in Section 3.3 but have described in this section why this must necessarily be true for any consistently constructed scheme. Therefore, if SNe are the only form of stellar feedback considered in a simulation nothing is gained by adopting an explicit IMF sampling scheme, which carries with it a penalty in terms of code complexity and (potentially) memory requirements.

4.2 Sensitivity of non-supernova feedback to IMF sampling

4.2.1 Under what circumstances does the sensitivity arise?

SN feedback is unique among stellar feedback channels in that it is composed of discrete, instantaneous (relative to other astrophysical timescales) events. All other forms of stellar feedback (e.g. radiation or stellar winds) are more continuous in nature, coupling to gas over an extended period of time. The degree to which the impact of the feedback is sensitive to the details of IMF sampling depends on the resolution in two ways. Firstly, does explicit IMF sampling produce significant enough inhomogeneities from particle to particle for results to deviate from an IMF averaged approach? The larger the star particle mass, the smaller the spread in feedback properties. Secondly, is the gas resolution high enough that the difference causes a resolvable effect? This also depends on other simulation details, such as the star formation prescription.

In the simulations we presented in this work, the predominant difference between the IMFav and IMFSam simulations was the efficiency of the photoionization feedback.
discrete feature at $s_{\text{part}} = 0$ which cannot be meaningfully represented in our logarithmic PDFs. We therefore include an inset figure that shows the probability that a particle has an ionizing luminosity of zero. Thus the integral under the PDF and the value of corresponding bar in the inset plot sum to unity together.

As expected, Fig. 12 shows that the spread around the IMF averaged specific luminosity reduces as the particle mass increases. This is because the IMF is more completely sampled within the particle for a larger particle mass i.e. $s_{\text{part}}$ converges with $s_{\text{IMFav}}$ as $m_{\text{part}}$ goes to infinity. It is important to note that this does not mean that inhomogeneities are not important, rather that a simulation with a large particle mass is incapable of resolving them. If the particle mass is large enough that inhomogeneities vanish, then using IMF averaged values will not result in any further loss of fidelity in the simulation and so this simpler approach can be adopted. With $m_{\text{part}} = 10^4 M_\odot$, the largest particle mass we consider, each particle samples the IMF relatively well, so $s_{\text{part}}$ is very tightly distributed about the IMF average, deviating by a factor of 2 at most. However, as the particle mass decreases, the spread about the IMF averaged value is no longer insignificant. It already begins to span a factor of a few when $m_{\text{part}} = 10^2 M_\odot$. The distribution continues to become broader and flatter as the particle mass as further decreased. A sharp cutoff is apparent at the low end of the PDF. As mentioned before, this originates from particles that are assigned no ionizing sources. The smallest non-zero $s_{\text{part}}$ corresponds to the emission from a single $7 M_\odot$ in the particle, so this cutoff occurs at higher values of $s_{\text{part}}$ for lower $m_{\text{part}}$. Correspondingly, the fraction of particles that produce no ionizing radiation (the inset panel) increases with decreasing $m_{\text{part}}$, such that when $m_{\text{part}} = 20 M_\odot$ only approximately a quarter of particles produce radiation. However, even if particles do emit radiation, when $s_{\text{part}}$ is more than factor of a few below the IMF averaged value this emission is negligible.

The top end of the PDFs represent the rare bright sources. It can be seen that with $m_{\text{part}} = 300 M_\odot$, some particles can have a specific ionizing luminosity an order of magnitude larger than the IMF averaged value. The maximum physically obtainable specific ionizing luminosity for the stellar mass range we consider corresponds to emission from a $100 M_\odot$ star. This is 18.47 times the IMF averaged value.\footnote{We confirmed in test simulations for Paper I that our results are relatively insensitive to dropping this maximum stellar mass to 50 $M_\odot$, which corresponds to a specific ionizing luminosity relative to the IMF average of 9.65.} This limit is indicated with a dashed line in Fig. 12. Low mass particles can exceed this limit when they are assigned more stellar mass than their dynamical mass (a natural consequence of our IMF sampling scheme which is not particularly problematic, as discussed in Section 2). Note that this does not result in an unphysically bright source in the simulation or affect the clustering of bright sources. It simply indicates that a more physically correct determination of their specific ionizing luminosity should account for the particles that have a mass discrepancy in the opposite sense, created by the adjusted target IMF sampling scheme to ensure mass consistency over multiple particles.

Fig. 12 demonstrates quantitatively how the ability to resolve inhomogeneities in the strength of ionizing sources varies as a function of particle mass. Similar scalings will result from any feedback source and are relatively easy to determine. However, the impact of the inhomogeneities on the outcome of the simulation are harder to predict. By averaging over the IMF, one gains more sources that have the IMF averaged feedback strength at the cost of losing the strongest sources. In our simulations, using IMF averaged ionizing photon rates results in more efficient feedback. This is because a significant number of sources with the IMF averaged rate are able to start forming resolvable H\textsc{ii} regions and sources are typically present earlier in the life of a star forming cloud, so the penalty incurred by losing the brightest sources is outweighed by benefits of increasing the overall number of ionizing sources. This positive impact on the overall feedback strength of IMF averaging will not necessarily occur in all scenarios. If the IMF averaged ionizing photon rate was not enough to form resolvable H\textsc{ii} regions then the reverse could occur since the simulation would contain no sources bright enough to have an impact. Note also that in simulations of an individual GMC at a higher resolution than us, Grudič & Hopkins (2019) found the opposite trend (i.e. more discretization of ionizing sources leads to more effective feedback). Thus, it is difficult to predict the impact of IMF averaging vs sampling in a given simulation. However, we would suggest that as the latter is arguably more physical it should in general be adopted if there will be significant particle to particle inhomogeneities.

These arguments apply to all forms of non-SN stellar feedback, not just ionizing radiation. We possibly see the reverse trend affecting our photoelectric heating feedback in our simulations than affects photoionization feedback. Photoelectric heating is very ineffective in our simulations of dwarf galaxies, as can be seen in Fig 5. This is because of the low dust-to-gas ratio (as discussed in Paper I). The simulations with photoelectric heating alone have similar SFRs to the simulation without anyfeedback. Nonetheless, the peak SFR in the IMFsam version is marginally lower than the IMFav equivalent, possibly indicating that IMF averaging weakens the photoelectric heating feedback, in contrast to the photoionization feedback which is strengthened. This could occur if only the brightest sources were able to cause sufficient heating to have an impact and that spreading their luminosity among multiple sources (via IMF averaging) renders heating ineffective. However, we caution that the difference is very slight and these galaxies are already in an unphysical regime, having fragmented into very dense clumps and reached a very high SFR due to ineffective feedback.

\subsection{4.2.2 Can a simplified model be used to replicate the effects of explicit IMF sampling?}

Su et al. (2018) examines how discretizing stellar feedback affects dwarf galaxies, in this case with cosmological zoom-in simulations with a mass resolution of 250 $M_\odot$. In common with the other works previously mentioned, they find that modelling SNe as a continuous injection of feedback, rather than as discrete events, substantially weakens its impact. Yet, in contrast to us, they find that the efficiency of continuous non-SN feedback mechanisms (e.g. radiation and stellar
winds) are largely unaffected by accounting for the effects of IMF sampling as opposed to their default IMF averaged rates. This could indicate that the impact of the particle to particle inhomogeneities do not result in a resolvable effect at their mass resolution. However, they do not actually perform explicit IMF sampling but instead attempt to replicate the effects via a toy model. At the point of creation, a star particle is assigned a number of O stars, $N_0$, drawn from a Poisson distribution with an expectation value of $\langle N_0 \rangle = m_{\text{part}}/100M_\odot$. Their feedback schemes then operate as in the fiducial case, but all IMF averaged rates that are linked to massive stars (photoionization, photoelectric heating, UV radiation pressure, OB star winds and core-collapse SN rates) are multiplied by a factor $N_0/\langle N_0 \rangle$. When a SN occurs, $N_0$ is reduced by one. It can therefore be seen that the strength of feedback will vary from star particle to particle according to the number of assigned sub-grid O stars, but the net rates over multiple particles will maintain the IMF average. This scheme is much simpler than explicitly sampling, assigning individual stars to particles and looking up their individual feedback budgets. The downside is that the approximation does not completely capture the correct behaviour, for a number of reasons. As we have demonstrated in the previous section, stochastically triggering SNe from IMF averaged SN rates already captures the correct clustering properties, as would be produced by explicit IMF sampling. Thus, boosting the IMF averaged SN rates will result in an additional over-enhancement of clustering.

A potentially more problematic issue with the approximation is that it assumes massive stars are uniform in terms of their feedback budget, as the authors caution. A large amount of the scatter in the specific ionizing luminosity of a star particle (shown in Fig. 12) originates not just from the variation in the number of massive stars assigned to a star particle but also from the strong mass dependence of the ionizing luminosity produced by individual stars. In other words, there is a significant variation in luminosity among OB stars. Fig. 13 shows the results of a similar Monte Carlo experiment, where we assign a $t_{\text{part}} = 0$ ionizing photon rate to $250 M_\odot$ star particles either by populating them with stars sampled from the IMF, as usual, or by using the toy model of Su et al. (2018). It can be seen that the toy model is not a good approximation to the correct distribution, with a much narrower range of ionizing photon rates, producing far more stars close to the IMF average value and lacking the brightest sources.

Additional inconsistencies in the toy model arise as the star particles age. The multiplicative factor that modulates the feedback rates is reduced by $1/(N_0)$ when a SN occurs, regardless of when it occurs. An early SN indicates a more massive progenitor, which in turn means a larger drop in the ionizing photon budget for the particle. This decrease in luminosity as the most massive stars die is already accounted for in the time evolution of the IMF averaged rates. Another direct consequence of the link between lifetime, stellar mass and photon budget is that the earliest SNe should occur in regions that have been exposed to the highest ionizing flux, an effect not captured by the toy model. All of these issues taken together, but in particular that shown in Fig. 13, suggest that IMF sampling effects likely cannot be replicated by simply modulating the IMF averaged rates.

5 SUMMARY AND CONCLUSIONS

It is a common practice in simulations of galaxies to treat stellar mass in a homogenised manner, with star particles producing IMF averaged feedback. With the advent of simulations with increasingly higher baryonic mass resolution, we have examined the extent to which this approximation is valid and under what circumstances it breaks down. We began by exploring methods of populating star particles with inventories of stars drawn from the IMF. Because of the challenges of filling an arbitrary mass budget with a discrete set of randomly drawn stellar masses, the goals of conserving mass (globally and locally) and faithfully reproducing the input IMF are often in tension. We argued that because the IMF is almost never an emergent property because of resolution limits and missing physics, an input IMF must be provided and reproduced accurately by the sampling scheme. This IMF may be fixed or it can vary based on some sub-grid prescription, but it should not be biased by numerical issues such as the choice of star particle mass, since this will

---

9 As we have already discussed at length, this is true only if the sampling is carried out consistently. Su et al. (2018) use Bernoulli trials. The combination of their particle mass and typical time-step yields $N_{\text{SN}} \sim 10^{-5}$, which is sufficiently small for Bernoulli trials and Poisson sampling to converge.

---

Figure 13. Similar to Fig. 12, the distribution of the specific ionizing luminosity, normalized by the IMF average, for a $10^7 M_\odot$ population of star particles. In this plot we show the results for 250 M⊙ particles with their luminosity obtained by explicitly populating the particles from the IMF ($\text{IMF}_{\text{sam}}$) or by using the toy model of Su et al. (2018) to modulate the IMF averaged rates ($\text{IMF}_{\text{toy}}$). The toy model results in a much narrower distribution than that which arises from sampling the IMF directly because it does not account for the substantial variation in the luminosity of massive stars as a function of mass. Note that the toy model forces the particles to adopt one of a discrete set of luminosities, so plotting this as a PDF is sub-optimal but we do so to enable easy comparison to Fig. 12.
 IMF averaging versus IMF sampling

In the second part of this work, we carried out isolated simulations of an $M_{\text{vir}} = 10^{10} M_\odot$ dwarf galaxy with a baryonic mass resolution of $20 M_\odot$. We included stellar feedback in the form of core-collapse SNe, photoionizing radiation and photoelectric heating, treated with the new sub-grid models described in Paper I and implemented in the AREPO code. We compared two sets of simulations that either used IMF averaged rates for feedback or explicitly populated particles with discrete stars (using the adjusted target method). Our key findings are as follows:

- If SNe are the only source of stellar feedback, triggering individual SNe via Poisson sampling of IMF averaged rates yields identical results to explicitly sampling stellar masses from the IMF. This is because the distribution of the SNe in space and time is essentially perfectly reproduced by a stochastic sampling of the rates as long as the time resolution is sufficiently high.
- The impact of ionizing radiation is overestimated in our simulations when IMF averaged rates are adopted, particularly its ability to regulate SFRs. Approximately twice the mass of gas is photoionized per unit stellar mass when the IMF averaged rates are used compared to the explicit IMF sampling equivalent simulations because sources are typically embedded in lower density gas. When IMF averaging is used every star particle immediately emits ionizing photons, meaning that clumps of gas begin to have their collapse disrupted as soon as star formation begins. This results in the production of many small H$\text{II}$ regions. By contrast, IMF sampling correctly produces brighter but, crucially, rarer sources. This discretization means that on average a larger mass of stars must be formed before a significant ionizing source appears. It also means that these sources are typically formed later in the evolution of a star forming clump and are thus embedded in higher density gas. This results in less efficient feedback and fewer but larger H$\text{II}$ regions. It is important to note that this trend is not driven by an inability to resolve H$\text{II}$ regions (see Section 3.3 for details).

The scenario we see in these simulations is not necessarily ubiquitous. The degree to which the strength of IMF averaged feedback deviates from the explicitly sampled case depends on a number of factors. Firstly, the star particle must be of a sufficiently small mass such that appreciable inhomogeneities are apparent across the population of star particles. This will vary between different feedback channels. Secondly, the impact of these inhomogeneities must itself be resolvable in the simulation. IMF averaging may also potentially decrease the effectiveness of feedback under certain conditions. Feedback strength goes up in our simulations when we use IMF averaging because the resultant ionizing luminosity remains high enough to have a resolvable impact. We thus increase the number of effective sources of feedback, despite the penalty of losing the strongest sources. However, if a feedback channel is completely dependent on extremely rare, strong sources (e.g. the most massive O stars), then IMF averaging may result in a net reduction in the feedback strength.

Given the complex dependence on resolution, the details of sub-grid models and the highly non-linear behaviour of stellar feedback, it is difficult to predict a priori whether there will be a difference between an IMF averaged feedback scheme and an IMF sampled scheme in any given simulation. However, the IMF sampling approach is always the more physically motivated approach. In other words, the best an IMF averaged scheme can achieve is to give results that converge with an IMF sampling approach. Even then, it will only converge when the effects of sampling are unresolvable and thus no more fidelity can be gained by explicitly populating star particles with discrete stars. Therefore, rather than trying to estimate how much of an effect choosing IMF averaging over sampling will have on the resolved feedback a priori, we suggest that the choice should simply be based on the level of inhomogeneities between particles for the chosen particle mass. This varies between different feedback channels, but based on the variation in the specific ionizing luminosity between particles (see Fig. 12), we very conservatively suggest that explicit IMF sampling should be used when the particle mass is less than $\sim 500 M_\odot$. 
In this work, we have used a single, fixed IMF (Kroupa 2001) in our tests. However, as we discussed, our techniques and findings are applicable to any IMF, including those that can vary based on local properties. Variable IMFs can already be adopted by IMF averaged feedback schemes by adding additional dimensions to lookup tables, but they can be treated in a very natural fashion by explicit sampling schemes, allowing very fine grained control of stellar populations on-the-fly. This would be an interesting avenue of future research. We have also only considered single stars. The sampling scheme could be extended to explicitly account for binary systems. The inclusion of binary stellar evolution results in a modest increase in the total ionizing photon budget and extends the production of ionizing radiation to later times (see e.g. Eldridge & Stanway 2009), as well as introducing an additional population of late-time core-collapse SNe (Zapartas et al. 2017). It would also allow feedback channels that depend on binary systems (e.g. Type Ia SNe, jets from High Mass X-ray Binaries and OB runaways etc.) to be treated in a self-consistent manner. Thus, using explicit IMF sampling in galaxy formation simulations not only allows us to capture the inhomogeneous distribution of stellar feedback sources but may also provide a more direct link to models of stellar evolution than IMF averaged approaches.

ACKNOWLEDGEMENTS

The author is grateful to Elaad Applebaum, John Forbes and Chia-Yu Hu for helpful discussions and to Rachel Somerville, Greg Bryan and the anonymous reviewer for comments which improved this manuscript. The work of M.C.S. was supported by a grant from the Simons Foundation (CCA 668771, L.E.H.). The simulations were run on the Flatiron Institute's research computing facilities (Iron and Popeye compute clusters), supported by the Simons Foundation (CCA 668771, L.E.H.).

DATA AVAILABILITY STATEMENT

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

REFERENCES

Agertz O., Kravtsov A. V., 2015, ApJ, 804, 18
Andrews J. E., et al., 2013, ApJ, 767, 51
Andrews J. E., et al., 2014, ApJ, 793, 4
Applebaum E., Brooks A. M., Quinn T. R., Christensen C. R., 2020, MNRAS, 492, 8
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Calzetti D., Chabrier G., et al., 2000, ApJ, 533, 682
Carigi L., Krumholz M. R., 2016, ApJ, 823, 50
Cappellari M., et al., 2012, Nature, 484, 485
Carigi L., Hernandez X., 2008, MNRAS, 389, 582
Ceverino D., Klypin A., Kimm T., Tissera P., Springel V., 2019, MNRAS, 482, 2244
Chabrier G., 2003, PASP, 115, 763
Chieffi A., Limongi M., 2004, ApJ, 608, 405
Comrey C., van Dokkum P. G., 2012, ApJ, 760, 71
Corbelli E., Verley S., Elmegreen B. G., Giovanardi C., 2009, A&A, 495, 479
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davé R., Angleés-Alcazar D., Narayan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Dubois Y., et al., 2014, MNRAS, 444, 1453
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Weller C., Volonteri M., 2016, MNRAS, 463, 3948
El-Badry K., Ostriker E. C., Kim C.-G., Quataert E., Weisz D. R., 2019, MNRAS, 490, 1961
Eldridge J. J., Stanway E. R., 2009, MNRAS, 400, 1019
Elmegreen B. G., 2006, ApJ, 648, 572
Emerick A., Bryan G. L., Mac Low M.-M., 2019, MNRAS, 482, 1304
Fielding D., Quataert E., Martizzi D., Faucher-Giguère C.-A., 2017, MNRAS, 470, L29
Fielding D., Quataert E., Martizzi D., 2018, MNRAS, 481, 3325
Fujimoto Y., Krumholz M. R., Tachibana S., 2018, MNRAS, 480, 4025
Fumagalli M., da Silva R. L., Krumholz M. R., 2011, ApJ, 741, L26
Gatto A., et al., 2017, MNRAS, 466, 1903
Geen S., Watson S. K., Rosdahl J., Bieri R., Klessen R. S., Henribe P., 2018, MNRAS, 481, 2548
Geha M., et al., 2013, ApJ, 771, 29
Genel S., et al., 2019, ApJ, 871, 21
Gentry E. S., Krumholz M. R., Dekel A., Madau P., 2017, MNRAS, 465, 2471
Gentry E. S., Krumholz M. R., Madau P., Lui P., 2019, MNRAS, 483, 3664
Grudic M. Y., Hopkins P. F., 2019, MNRAS, 488, 2970
Gunawardhana M. L. P., et al., 2011, MNRAS, 415, 1647
Gutcke T. A., Springel V., 2019, MNRAS, 482, 118
Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, MNRAS, 501, 5597
Haardt F., Madau P., 2012, ApJ, 746, 125
Haas M. R., Anders P., 2010, A&A, 512, A79
Hernquist L., 1990, ApJ, 356, 359
Hirai Y., Fujii M. S., Saitoh T. R., 2020, arXiv e-prints, p. arXiv:2005.12906
Hopkins P. F., Kereš D., Onorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Hopkins P. F., et al., 2018, MNRAS, 480, 800
Hoversten E. A., Glazebrook K., 2008, ApJ, 675, 163
Hu C.-Y., 2019, MNRAS, 483, 3363
Hu C.-Y., Naab T., Glover S. C. O., Walch S., Clark P. C., 2017, MNRAS, 471, 2151
Kalirai J. S., et al., 2013, ApJ, 763, 110
Kata N., 1992, ApJ, 391, 502
Kata N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Keller B. W., Kruissee J. M. D., 2020, arXiv e-prints, p. arXiv:2004.03608
Keller B. W., Wadsley J. W., Wang L., Kruissee J. M. D., 2019, MNRAS, 482, 2244
Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900
Kroupa P., 2001, MNRAS, 322, 231
Kroupa P., Weidner C., 2003, ApJ, 598, 1076
Kroupa P., Tout C. A., Gilmore G., 1993, MNRAS, 262, 545
Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
La Barbera F., Ferreras I., Vazdekis A., de la Rosa I. G., de la Rosa I., 2013, MNRAS, 433, 3127
Lanza T., Hubeny I., 2003, ApJS, 146, 417
Lee J. C., et al., 2009, ApJ, 706, 599
Lee J. C., Veilleux S., McDonald M., Hilbert B., 2016, ApJ, 817, 177
Mac Low M.-M., McCray R., 1988, ApJ, 324, 776

MNRAS 000, 000–000 (0000)
APPENDIX A: ROBUSTNESS TO STOCHASTICITY

Simulations of galaxy evolution are inherently chaotic to some degree, meaning that small perturbations introduced by, for example, seemingly minor differences in initial conditions, choice of random number generator seed value, floating-point round-off and non-deterministic behaviour of parallelised codes can lead to measurable large scale differences in the outcome (Keller et al. 2019; Genel et al. 2019). In order to approximately assess how the differences between our various feedback schemes compare to the magnitude of this stochastic uncertainty, we perform some re-simulations of the early stages of the SN and PI simulations with both the IMFav and IMFsam schemes. For each fiducial simulation, we carry out four additional re-simulations. In each re-simulation we use different seed values for random number generators. We also perturb the position of every gas cell, star particle and dark matter particle in the initial conditions by moving it 0.1 pc in a random direction. The SFRs for these simulations can be seen in Fig. A1, with the fiducial simulations shown in the bold colours and the re-simulations shown in a lighter shade.

We were limited by computational expense from completely re-simulating the full 1 Gyr of our fiducial simulations, but it can be seen that for the 450 Myr re-simulated the IMFav and IMFsam SN simulations are consistent with each other within the range of stochastic scatter. We are therefore confident in our assertion that IMF averaging and IMF sampling give essentially identical results if SNe are the only source of feedback. The PI re-simulations have a very tight scatter about the fiducial simulations, such that the magnitude of the stochastic uncertainty is much smaller than the difference between the IMFav and IMFsam schemes.