First Light And Reionization Epoch Simulations (FLARES) – I. Environmental dependence of high-redshift galaxy evolution

Christopher C. Lovell, Aswin P. Vijayan, Peter A. Thomas, Stephen M. Wilkins, David J. Barnes, Dimitrios Irodotou and Will Roper

1 Centre for Astrophysics Research, School of Physics, Astronomy & Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
2 Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK
3 Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ABSTRACT

We introduce the First Light And Reionisation Epoch Simulations (FLARES), a suite of zoom simulations using the EAGLE model. We resimulate a range of overdensities during the Epoch of Reionization (EoR) in order to build composite distribution functions, as well as explore the environmental dependence of galaxy formation and evolution during this critical period of galaxy assembly. The regions are selected from a large (3.2 cGpc)³ parent volume, based on their overdensity within a sphere of radius 14 h⁻¹ cMpc. We then resimulate with full hydrodynamics, and employ a novel weighting scheme that allows the construction of composite distribution functions that are representative of the full parent volume. This significantly extends the dynamic range compared to smaller volume periodic simulations. We present an analysis of the galaxy stellar mass function (GSMF), the star formation rate distribution function (SFRF), and the star-forming sequence (SFS) predicted by FLARES, and compare to a number of observational and model constraints. We also analyse the environmental dependence over an unprecedented range of overdensity. Both the GSMF and the SFRF exhibit a clear double-Schechter form, up to the highest redshifts (z = 10). We also find no environmental dependence of the SFS normalization. The increased dynamic range probed by FLARES will allow us to make predictions for a number of large area surveys that will probe the EoR in coming years, carried out on new observatories such as Roman and Euclid.

Key words: galaxies: abundances – galaxies: evolution – galaxies: high-redshift.

1 INTRODUCTION

A goal of numerical galaxy evolution studies is to model a representative population of galaxies, resolving all of the relevant physics at the required scales, in order to provide a test bed for the study and interpretation of observed galaxies (Benson 2010). In order to achieve this it is necessary to simulate large volumes (in order to sample a representative volume of the universe) at high resolution (e.g. spatial, mass, and time; in order to resolve the internal physical processes within individual galaxies) and with all of the key physics included (such as full hydrodynamics, magnetic fields, etc.). Unfortunately this is not computationally feasible; compromises must be made with volume, resolution or choice of physics, depending on the scientific questions posed (for a review, see Somerville & Davé 2015).

The most common approach to obtain a representative population of galaxies is to simulate a large periodic cube, tens of Mpc across on a side. This approach has been used in a number of leading projects to simulate large volumes down to redshift zero, producing thousands of galaxies across a wide range of stellar masses. Projects such as EAGLE (Crain et al. 2015; Schaye et al. 2015), SIMBA (Davé et al. 2019), Illustris (Genel et al. 2014; Vogelsberger et al. 2014), Illustris-TNG (Nelson et al. 2018; Pillepich et al. 2018), Romulus (Tremmel et al. 2017), and Horizon-AGN (Dubois et al. 2014) have mass resolutions of order 10⁶ M⊙, sufficiently high to resolve the internal structure of galaxies. However, despite these large volumes, the rarest peaks of the overdensity distribution are still poorly sampled due to the lack of large-scale modes in constrained periodic volumes. Much larger volumes are required to sample the rare overdensities on large scales that are likely to evolve in to the most massive clusters by the present day. For example, the EAGLE simulation contains just 7, relatively low-mass clusters (M200,c > 10¹¹ M⊙) at z = 0 within the fiducial 100 Mpc volume (Schaye et al. 2015).

One means of overcoming the limitations of relatively small periodic volumes is to use much larger, dark matter-only simulations, with box lengths of order Gpc, as sources for zoom simulations. These use regions selected from the dark matter-only simulation as source initial conditions, and resimulate them at higher resolution with extra physics, such as full hydrodynamics (Katz & White 1993; Tormen, Bouchet & White 1997). This technique preserves the large-scale power and tidal forces by simulating the dark matter at low resolution outside the high-resolution region. A recent example is the C-EAGLE simulations, high-resolution hydrodynamic simulations of 30 clusters with a range of descendant masses (Bahé et al. 2017; Barnes et al. 2017b). These were selected from a parent dark matter simulation with volume (3.2 cGpc)³ (Barnes et al. 2017a).

This enormous volume contains 185 150 clusters (M200,c > 10¹² M⊙) and 1701 high-mass clusters (M200 > 10¹⁵ M⊙). The C-EAGLE zoom
The zoom technique can also be used to sample a range of overdensities, not just the peaks of the overdensity distribution. The GIMIC simulations (Crain et al. 2009) are one example of this approach; they picked five different regions of radius 20 cMpc at z = 1.5 from the Millennium simulation (Springel et al. 2005), with overdensities (−2, −1, 0, 1, 2)σ from the cosmic mean at z = 1.5. These were then resimulated at high resolution with full hydrodynamics. This not only allowed the investigation of the environmental effect of galaxy evolution, without having to simulate a whole periodic box, but also, by appropriately weighting each region according to its overdensity, the regions could be combined to produce composite distribution functions. These composite functions have much larger dynamic range than those obtained from smaller periodic boxes, and at much lower computational expense than running a large periodic volume.

In this paper, we use a similar approach to GIMIC to produce composite distribution functions of galaxy intrinsic properties, but focused on the Epoch of Reionization (EoR). The EoR approximately covers the first 1.2 billion years of the Universe’s history (4 ≤ z ≤ 15), from the birth of the first Population III stars, to when the majority of the intergalactic medium is ionized (Bromm & Yoshida 15), from the birth of the first Population III stars, to when the ground-based observatories, have discovered thousands of galaxies (2019). A number of surveys over the past 15 yr, with both space- and ground-based observatories, have discovered thousands of galaxies during this epoch (Beckwith et al. 2006; Warren et al. 2007; Grogin et al. 2011; Koekemoer et al. 2011; Wilkins et al. 2011; McCracken et al. 2012; Bouwens et al. 2015). Using intervening clusters as gravitational lenses has pushed the measurement of luminosity functions to even fainter magnitudes (Castellano et al. 2016; Livermore, Finkelstein & Lotz 2017; Atek et al. 2018; Ishigaki et al. 2018). Spectral energy distribution fitting has been used to characterise the intrinsic properties of these galaxies, measuring for example their stellar masses (e.g. Gonzalez et al. 2011; Duncan et al. 2014; Song et al. 2016; Stefon et al. 2017) and star formation rates (SFRs; e.g. Smit et al. 2012; Katsianis et al. 2017). However, we have yet to unambiguously detect Population III stars (Yoshida 2019), and the first stages of galaxy assembly are yet to be probed, particularly the seedling and early growth of super massive black holes (Smith, Bromm & Loeb 2017).

However, this situation may soon change with the introduction of a number of new observatories, each with unique capabilities for exploring the EoR. JWST will provide unprecedented sensitive imaging with NIRCam, and follow-up spectroscopy with NIRSpec and MIRI, to detect and characterize potentially the very first forming galaxies in the Universe (Gardner et al. 2006). In tandem, Roman and Euclid will produce wide-field surveys of the EoR (Laureijs et al. 2011; Spergel et al. 2015). These surveys will predominantly probe the bright end of the rest-frame UV Luminosity Function (UVLF), which is currently poorly constrained by periodic hydrodynamic simulations due to their small volume. They will also discover some of the most extreme galaxies, in terms of luminosity and intrinsic mass, in the observable Universe at these redshifts (Behroozi & Silk 2018). These observations will be important to constrain models of galaxy formation and evolution, but it is also possible to predict observed populations in advance and test the recovery of intrinsic parameters (Pforr, Maraston & Tonini 2012, 2013; Smith & Hayward 2015; Lower et al. 2020).

1Where σ is the rms mass fluctuation on the resimulation scale.
2 THE FLARE SIMULATIONS

We will now detail our simulations, including the EAGLE model, selection of the regions, the zoom resimulation technique, and our method for constructing composite distribution functions.

2 Project website available at https://flaresimulations.github.io/flares/.

Table 1. Variation of subgrid parameters between models.

| Simulation prefix | C_{\text{visc}} | \Delta T_{\text{MIN}} (K) |
|------------------|----------------|--------------------------|
| Ref              | 2\pi           | 10^{6.5}                 |
| AGNdT9           | 2\pi \times 10^2 | 10^5                     |

2.1 The EAGLE model

The EAGLE physics model is based on that developed for the OWLS project (Schaye et al. 2010), which is a heavily modified version of P-GADGET-3 (Springel et al. 2005), an N-body tree-PM SPH code. The hydrodynamics suite is collectively known as ‘ANARCHY’, (described in Appendix A of Schaye et al. 2015 and Schaller et al. 2015). In short, it consists of the Hopkins (2013) pressure–entropy SPH formalism, an artificial viscosity switch (Cullen & Dehnen 2010), an artificial conductivity switch (e.g. Price 2008), the Wendland (1995) C^2 smoothing kernel with 58 neighbours, and the Durier & Dalla Vecchia (2012) time-step limiter.

Radiative cooling, formation of stars, black hole seeding, and feedback from stars and black holes are all handled by subgrid models. Full details are provided in Schaye et al. (2015) and Crain et al. (2015). We use the AGNdT9 parameter configuration, which produces similar mass functions to the reference model but better reproduces the hot gas properties in groups and clusters (Barnes et al. 2017b). This is identical to that used in the C-EAGLE simulations, but differs from the fiducial reference simulation (see Table 1). It uses a higher value for \(C_{\text{visc}}\), which controls the sensitivity of the BH accretion rate to the angular momentum of the gas, and a higher gas temperature increase from AGN feedback, \(\Delta T\). A larger \(\Delta T\) leads to fewer, more energetic feedback events, whereas a lower \(\Delta T\) leads to more continual heating. These parameter changes impact the central black hole accretion, which has been shown to be efficient only at halo masses \(>10^{12} M_{\odot}\) (Bower et al. 2017). At \(z = 10\) no FLARES galaxies reside in such haloes, however, at \(z = 5\) a minority do \((<0.2\ \text{per cent})\), which may affect the early star formation histories of cluster galaxies (Bahé et al. 2017). We use an identical resolution to the fiducial EAGLE simulation, with gas particle mass \(m_g = 1.8 \times 10^6 M_{\odot}\), and a softening length of 2.66 c\text{kpc}.

2.2 Region selection

We use the same parent simulation as that used in the C-EAGLE simulations (Barnes et al. 2017a): a \(3.2\ \text{cGpc}^3\) dark matter-only simulation with a particle mass of \(8.01 \times 10^{10} M_{\odot}\), using a Planck Collaboration I (2014) cosmology. Fig. 2 shows a diagram of the box compared to the fiducial EAGLE reference volume, as well as the BLUETIDES simulation (Feng et al. 2016). The highest redshift snapshot available for this simulation is at \(z = 4.67\), which we use for our selection. Within this snapshot, we select spherical volumes that sample a range of overdensities. By taking a sufficiently large radius we can ensure that the density fluctuations averaged on that scale are linear, such that the distortion in the shape of the Lagrangian volume during the simulation will not be too extreme and that the ordering of the density fluctuations is preserved. The regions, and their overdensities, are given in Table A1.

To determine the density, we first distribute the mass on to a high resolution, \(3.2\ \text{cGpc}/1200 \sim 2.67\ \text{cMpc}\) cubic grid using a nearest grid point assignment scheme. We then find the density on larger scales by convolving the grid with a spherical top-hat filter of radius...
resimulation approach is that this can simply be achieved by running more simulations to increase the total simulated volume.

The selected regions are listed in Appendix A and the range of overdensities that each covers (evaluated at each point on the 2.67 cMpc grid enclosed by that volume) is shown in the lower panel of Fig. 4. We discuss how to combine the resimulations so as to obtain a representative sample of the whole universe in Section 2.4.

2.3 The resimulation method

Galaxies on the edge of the high resolution region will not be modelled correctly due to the presence of a pressureless boundary. To avoid this we resimulate a region 15 h⁻¹ cMpc in radius, and ignore all galaxies within 1 h⁻¹ cMpc of the edge of the sphere in post-processing. At higher redshift the Lagrangian high-resolution region can deform, but we found that it is close to spherical out to the highest redshifts considered in this work (κ = 10). Fig. 3 shows the dark matter distribution within the cut-out radius for a range of resimulations of differing overdensity, at κ = 4.7. We also show the fiducial periodic EAGLE volume to provide a visual comparison of the differing environments probed.

As in the standard EAGLE analysis, structures are first found using a Friends-Of-Friends (FOF; Davis et al. 1985) finder, then split into bound substructures using the SUBFIND algorithm (Springel et al. 2001). Their properties are then defined using those stellar particles within 30 pkpc of the location of the most tightly bound stellar particle. We limit our analysis to galaxies sampled by at least 50 star particles, which corresponds to a mass limit of approximately log₁₀(M/M⊙) ≥ 7.95.

2.4 Distribution function weighting

In this section, we describe how we combine our resimulations to obtain a statistically correct representation of the universal cosmological distribution of galaxies. As we show below in Section 3.2, distribution functions, such as the galaxy stellar mass function (GSMF), vary with the overdensity of the resimulated volume. Therefore, it is necessary to weight each resimulation to reproduce the correct distribution of those overdensities averaged over the whole universe, i.e. the cosmic mean.

As mentioned in Section 2.2, the overdensity within spherical top-hat regions of radius 14 h⁻¹ cMpc is sampled on a 2.67 cMpc grid; we label this sample δg. Since the grid sampling is finer than the size of the resimulation volume, each resimulation volume is associated with just under 2000 different values of δg. We show the distribution of those δg within each resimulation volume in the lower panel of Fig. 4. The most overdense regions, whilst containing a single highly overdense point, in fact contain points covering a range of overdensities. It is, therefore, important to account for this spread in sampled overdensity, rather than just using the central overdensity

3 Code provided at https://github.com/christopherlovell/DensityGridder.

4 We have tested and found that our results are insensitive to changes (±0.5 cMpc) in the size of this boundary region.

5 A number of galaxies identified by subfind are, on close inspection, ‘spurious’ structures, which manifest as an unrealistic ratio between the stellar, gas, or dark matter components (see McAlpine et al. 2016, for a discussion). These galaxies make up less than 0.1 per cent of all galaxies > 10⁹ M⊙ at z = 5, and are typically low mass. We use the following conditions to flag spurious galaxies: any subhalo with zero mass in the stellar, gas, or dark matter components. Once these galaxies have been identified, we remove them from the SUBFIND catalogues, and add their particle properties to the parent ‘central’ subhalo.
when determining the contribution from any particular resimulation volume.

The top panel of Fig. 4 contrasts the PDFs of $\delta_g$ for the whole box and for our resimulated sample. To generate the correct mean distributions, we divide into bins of overdensity as shown by the histogram in Fig. 4 (black solid line), then weight the resimulations appropriately to reproduce the cosmic distribution. Specifically, we do the following:

(i) The overdensity domain is split up into 50 bins of equal width in $\log_{10}(1 + \delta)$, $i = 1 \ldots N_\delta$. For each of these, it is possible to assign a weight, $w_{\text{true},i}$, in proportion to the fraction of $\delta_g$ that lie in that bin, such that $\sum_i w_{\text{true},i} = 1$.

(ii) Each resimulation, $j$, is similarly distributed over these overdensity bins with weights, $w_{ij}$, in proportion to the enclosed values of $\delta_g$. Thus $\sum_i w_{ij} = 1$.

(iii) The sample weight associated with each bin is $w_{\text{sample},i} = \sum_j w_{ij}$.

(iv) To obtain the correct universal average, we therefore have to weight each density bin by the ratio $r_i = w_{\text{true},i}/w_{\text{sample},i}$.

Ideally, we would associate each galaxy with the local value of $\delta_g$. However, for the purposes of simplicity in this paper, we give all galaxies within a particular resimulation equal weight – this will give some dispersion over the more correct method, which we will implement in a future paper.

(i) Hence, we adjust the contribution of each resimulation by a factor $f_j = \sum_i r_i w_{ij}$.

We note that

$$
\sum_j f_j = \sum_j \sum_i r_i w_{ij} = \sum_i r_i \sum_j w_{ij} = \sum_i r_i w_{\text{sample},i} = \sum_i w_{\text{true},i} = 1.
$$

These simulation weighting factors are listed in Table A1.

We further note that, at higher redshifts, the overdensities will evolve. Nevertheless, because even the most extreme perturbations are only mildly non-linear, we would expect that the ordering of the overdensities would largely be preserved. Hence, we use the same sampling at all redshifts. That also allows for a much more direct comparison of the evolution within each overdensity sample.

3 RESULTS

3.1 Galaxy number counts

We begin by examining the raw number counts of galaxies. Fig. 5 shows the cumulative distribution function of galaxies with stellar mass for both FLARES and the Reference periodic volume ($V = (100 \, \text{cMpc})^3$). We produce over ~20 times more $10^{10} M_\odot$ galaxies at $z = 5$ than obtained in the 100 cMpc periodic volume, despite the fact that the total high-resolution volume of all resimulated regions is only 50 per cent larger than the periodic volume. This confirms

---

6We tested using a greater number of bins and found that the quantitative weights did not change significantly.
C. C. Lovell et al.

Figure 4. Upper panel: the probability distribution function of sampled overdensities. The dashed black line shows a lognormal fit with the given parameters. The solid blue histogram shows the grid locations that lie within one of our resimulation volumes. The solid black histogram shows the distribution of our selected regions in overdensity, binned into 50 equal width bins, with the right y-axis showing only their number counts. Lower panel: the distribution of overdensities within each simulation volume. The vertical displacement is arbitrary. The cross shows the overdensity measured at the centre of the resimulated volume and the spread of values shows the overdensities within each volume evaluated at each point on the 2.67 cMpc grid.

that the first galaxies are significantly biased to higher overdensity regions.

3.2 The galaxy stellar mass function

The GSMF describes the number of galaxies per unit volume per unit stellar mass interval \( d \log_{10} M \)

\[
\phi(M) = N/Mpc^{-3} \, dex^{-1},
\]

and is commonly described using a Schechter function (Schechter 1976)

\[
\phi(M) \, d \log_{10} M = \ln(10) e^{-M/M^*} \left( \frac{M}{M^*} \right)^{\alpha + 1},
\]

which describes the high- and low-mass behaviour with an exponential and a power-law dependence on stellar mass, respectively. Recent studies have found that a double Schechter function can better fit the full distribution (e.g. the GAMA survey; Baldry, Glazebrook & Driver 2008)

\[
\phi(M) \, d \log_{10} M = \ln(10) e^{-M/M^*} \times \left[ \phi_1^* \left( \frac{M}{M^*} \right)^{\alpha_1 + 1} + \phi_2^* \left( \frac{M}{M^*} \right)^{\alpha_2 + 1} \right].
\]

The low-mass slope of the second Schechter function contributes to only a very narrow dynamic range. Above this range the exponential dominates, and below this the low-mass slope of the first Schechter function dominates. It is therefore poorly constrained by the binned data, and so as not to introduce further degrees of freedom into our fit we fix it at \( \alpha_2 = -1 \). We define the stellar mass \( M_\star \) as the total mass of all star particles, associated with the bound subhalo, within a 30 kpc aperture (proper) centred on the potential minimum of the subhalo.7

3.2.1 The cosmic GSMF

In this section, we present results for the universal GSMF, averaged within our (3.2 cGpc)3 box. This is obtained by combining the individual GSMFs from each of our resimulation volumes with appropriate weighting, as described in Section 2.4.

The top panel of Fig. 6 shows the GSMF for redshifts between \( z = 10 \rightarrow 5 \). We show differential counts in bins 0.2 dex in width (with 1σ poisson uncertainties). The solid lines show double-Schechter function fits at each integer redshift. The normalization increases with decreasing redshift, and the characteristic mass (or knee) of the distribution shifts to higher masses. This is more clearly seen in Fig. 7, which shows the evolution of the double-Schechter parameters with

7Two substructures within 30 kpc of each other are still identified as separate structures, and only the particles associated with each structure contributes to its aperture-measured properties.
Figure 6. Top: Redshift evolution of the FLARES composite GSMF. Points show binned differential counts with Poisson 1σ uncertainties from the simulated number counts. Solid lines show double-Schechter function fits, quoted in Table C2. The parameter evolution is shown in Fig. 7. Bottom: as for the top panel, but points show the counts from the periodic reference volume. The dashed lines show the double-Schechter fitted relation from FLARES. The coverage of the massive end in the periodic volume is poor.

Our composite GSMF significantly extends the dynamic range of the GSMF compared to the periodic volumes. To demonstrate, the bottom panel of Fig. 6 shows the FLARES double-Schechter fits, alongside the binned counts from the Reference periodic volume. At each redshift the maximum stellar mass probed is approximately an order of magnitude larger in FLARES. In fact, the periodic reference volume barely probes the exponential tail of the high-mass component of the GSMF. When fitting a double-Schechter to the binned Reference volume counts we found that the parameters of the high-mass component were completely unconstrained. However, it is clear from the bottom panel of Fig. 6 that the low-mass slope is consistent between the Reference volume and FLARES. We have also tested that this is the case for the (50 Mpc)³ AGNdT9 periodic volume. This provides evidence that our weighting method is accurately recovering the composite GSMF, without suffering from completeness bias.

In Fig. 8, we show the composite FLARES GSMF against a number of high-z observational constraints in the literature (Gonzalez et al. 2011; Duncan et al. 2014; Song et al. 2016; Stefanon et al. 2017; Bhatawdekar et al. 2019). These studies show a spread of ∼0.5 dex at z = 5, which highlights the difficulty of accurately measuring the GSMF at high redshift. The FLARES composite GSMF lies within this interstudy scatter, most closely following the relations derived by Song et al. (2016) up to z = 7. At z ≥ 8...
There is some disagreement over the normalization of the GSMF between different observational studies; however, FLARES is consistent up to $z = 9$. Observational constraints are limited to cluster lensing studies such as the Hubble Frontier Fields, which do not probe the high-mass end due to the limited volume probed, but can reach very lower stellar masses ($\sim 10^7 M_\odot$). The fits presented in Bhatawdekar et al. (2019) have a higher normalization than in FLARES over the accessible mass range, though they quote an uncertainty at $10^{8.5} M_\odot$ of $\sim 0.6$ dex at $z = 9$; FLARES lies within this uncertainty for the point sources, but is still in tension with the normalization for disc-like sources. There is good agreement with the low-mass slope for both sources.

We also compare in Fig. 8 to predictions from other galaxy formation models. The Feedback In Realistic Environments (FIRE) project performed zoom simulations of individual haloes with masses between $10^8$ and $10^{12} M_\odot$, which were then combined to provide a composite GSMF probing the low-mass regime (Ma et al. 2018). FLARES is consistent with FIRE at all redshifts where their mass range overlaps. Fig. 8 also shows both the 2015 and 2020 versions of L-GALAXIES. Both models are in reasonably good agreement at all redshifts shown, but tend to underestimate the number density of massive galaxies at $z = 5$ compared to both FLARES and the observations.

Yung et al. (2019b) presented results from the Santa Cruz (SAM; Somerville et al. 2015), which extends to a wide dynamic range. Whilst FLARES is consistent with this model for $z \leq 7$, at $z \geq 8$ the Santa Cruz model predicts a more power-law shape to the GSMF, with a lower normalization at the characteristic mass. This is in...
agreement with the observed flattening of the GSMF with increasing redshift.

3.2.2 Environmental dependence of the GSMF

Our zoom simulations of a range of overdensities not only allow us to construct a composite GSMF for the entire ($3.2\,\text{Gpc}^3$) volume, but also investigate the environmental effect on the GSMF. Section 2 demonstrates the wide range of environments probed, from extremely underdense void regions, to the most overdense high-redshift structures that are likely to collapse into massive, $>10^{15}\,\text{M}_\odot$ clusters by $z = 0$ (Chiang, Overzier & Gebhardt 2013; Lovell et al. 2018).

Fig. 9 shows the GSMF in bins of log-overdensity from $z = 5$–9. We use wider bins than previously ($0.4\,\text{dex}$) due to the lower galaxy numbers in each resimulation. As expected, higher overdensity regions have a higher normalization, $\sim +2\,\text{dex}$ above the lowest overdensity regions at $M_*/\text{M}_\odot = 10^{8.8}$ ($z = 5$). There is also an apparent difference in the shape as a function of log overdensity: lower overdensity regions exhibit a distribution that is more power-law like, whereas higher overdensity regions clearly show a double-Schechter like knee. This may be due to the higher number of galaxies in the overdense regions, better sampling the knee, but may also point to differing assembly histories for galaxies in different environments. We will explore the star formation and assembly histories more closely in future work.

The dependence of the GSMF on overdensity may explain the tension between the composite FLARES GSMF and other models at $z > 7$ seen in Fig. 8. Our much larger box allows us to sample extreme overdensities that are not present in smaller volumes. Observationally, the Song et al. (2016) results show a more power-law like form at $z = 8$. Double-Schechter forms of the GSMF at low-$z$ have been attributed to the contribution of a passive and star-forming population, each fit individually by a single Schechter function (Kelvin et al. 2014; Moffett et al. 2016), though this separation is not perfect (e.g. Ilbert et al. 2013; Tomczak et al. 2016). The robust double-Schechter shape measured in FLARES at $z \geq 8$ is therefore curious; we see in Appendix B that there is no significant passive population as a function of stellar mass. We therefore tentatively suggest that the tension may be due to the small volume probed observationally at these depths, which does not probe extreme environments that contribute significantly to the cosmic GSMF.

We do not fit each binned GSMF in log-overdensity as there are insufficient galaxies to provide a robust fit. However, we do provide fits to the normalization at a given stellar mass and redshift, in the following form:

$$\log_{10}\phi = m\left[\log_{10}(1 + \delta)\right] + c,$$

where $\log_{10}(1 + \delta)$ is the overdensity of the region. Table 2 shows these fits for bins $\pm 0.2\,\text{dex}$ wide centred at $\log_{10}(M_*/\text{M}_\odot) = [8.5, 9.7]$.

3.3 The star formation rate distribution function

The SFRF describes the number of galaxies per unit volume per unit star formation rate interval d$log_{10}\psi$, where $\psi$ is the SFR

$$\phi(\psi) = N/\text{Mpc}^{-3}\,\text{dex}^{-1},$$

We define the SFR as the sum of the instantaneous SFR of all star-forming gas particles, associated with the bound subhalo, within a 30 kpc aperture (proper) centred on the potential minimum of the subhalo.
The characteristic SFR, \( \psi > \psi \), are provided in Table C3. We also plot the parameter evolution with turnover between \( M \). The normalization of both components (\( \phi \psi / \phi \psi \)), increase with decreasing redshift. These trends are surprisingly similar to those seen for the equivalent parameters in the GSMF. The low-SFR normalization is slightly higher (\( \sim 0.3 \) dex) from \( z = 5 \) to 7. There is no prominent knee in the observed relations, and the exponential tail drops off at lower SFRs than in the simulations.

Table 2. Fits to the normalization, \( \log_{10}(\phi/Mpc^{-3} \text{dex}^{-1}) \), of the GSMF at different redshifts and masses (see Section 3.2.2).

| \( z \) | \( \log_{10}(M_\star/M_\odot) \) | \( m \) | \( c \) |
|-----|-----------------|-----|-----|
| 5   | 8.5             | 3.5 | -2.4|
| 7   | 8.5             | 4.4 | -3.2|
| 9   | 8.5             | 4.6 | -4.0|
| 5   | 9.7             | 4.8 | -3.6|
| 7   | 9.7             | 4.4 | -4.2|
| 9   | 9.7             | 4.0 | -4.9|

In Fig. 10, we plot the evolution of the FLARES composite SFRF. Points show binned differential counts with Poisson 1 \( \sigma \) uncertainties from the simulated number counts. Solid lines show double-Schechter function fits, quoted in Table C3.

3.3.1 The cosmic SFRF

In Fig. 10, we plot the evolution of the FLARES composite SFRF. We provide counts in bins 0.3 dex in width. There is a clear low-mass turnover between \( \sim 0.1 \) and 0.3 \( M_\odot \text{yr}^{-1} \), but above this the shape is well described by a double-Schechter function. Salim & Lee (2012) argue that a single-Schechter is inadequate to describe the SFRF, as they propose a ‘Saunders’ function that does not provide a good fit to the FLARES SFRF. We provide fits using the following parametrization:

\[
\phi(\psi) \text{dlog}_{10} \psi = \text{ln}(10) e^{-\psi/\psi^*} \times \left[ \phi_1^* \left( \frac{\psi}{\psi^*} \right)^{\alpha_1+1} + \phi_2^* \left( \frac{\psi}{\psi^*} \right)^{\alpha_2+1} \right]. \tag{8}
\]

We limit our fits to those galaxies with \( \psi > 0.5 M_\odot \text{yr}^{-1} \); these fits are provided in Table C3. We also plot the parameter evolution with redshift in Fig. 7. The characteristic SFR, \( \psi^* \), is offset by \( +10^8 \) to aid comparison with the GSMF characteristic mass, \( M_\star \).

The normalization of both components (\( \phi_1^* \); \( \phi_2^* \)), as well as the low-SFR slope (\( \alpha_1 \)), increase with decreasing redshift. These trends are surprisingly similar to those seen for the equivalent parameters in the GSMF. The low-SFR normalization is almost identical, as is the high-SFR normalization, with a small \( \sim +0.2 \) dex offset. The low-SFR slope \( \alpha_1 \) is shallower than that of the GSMF at the highest redshifts (\( z \geq 8 \)), but identical at lower redshifts. However, the evolution of the characteristic SFR is significantly flatter compared to that of the characteristic mass for the GSMF. This suggests a redshift-independent upper limit to the SFR. The strong correspondence between the shape of the GSMF and the SFRF may be the result of the tight star-forming sequence (SFS) relation at all redshifts (see Section 3.4).

This double-Schechter form of the SFRF is in some tension with observational constraints. Fig. 11 shows a comparison with UV derived relations from Smit et al. (2012) and Katsianis et al. (2017) (the latter using Bouwens et al. 2015 data). For low-SFRs, the observed normalization is slightly higher (\( \sim 0.3 \) dex) from \( z = 5 \) to 7. There is no prominent knee in the observed relations, and the exponential tail drops off at lower SFRs than in the simulations.

This double-Schechter form of the SFRF is in some tension with observational constraints. FLARES has a distinct double-Schechter shape, whereas the SC model appears as a single-schechter at \( z = 5 \), before evolving to a power law at \( z = 10 \). The BLUETIDES results (Wilkins et al. 2017) also show a similar power-law relation at \( z \geq 8 \), in tension with the prominent knee in flares. Both L-GALAXIES models show similar power-law like behaviour, though with lower normalization at the high-SFR end (Henriques et al. 2015, 2020), though in better agreement with the existing observational data at \( z = 6 \) compared to the Santa Cruz model and Flares.

The offset in normalization of the FLARES SFRF at high SFRs with the observations may be a selection effect due to highly dust-obscurity galaxies. These galaxies, with number densities of \( \sim 10^{-5} \, \text{Mpc}^{-3} \) at \( z \sim 2 \) (Simpson et al. 2014), will be missed in higher redshift rest-frame UV observations. We will perform a direct comparison with the UV luminosity function, including self-consistent modelling of dust attenuation, in Paper II, Vijayan et al. (in preparation). The offset may also be a modelling issue; EAGLE was not compared to high redshift observables during calibration, only to data at much lower redshifts (\( z = 0.1 \)) than those studied here (\( z \geq 5 \)). Improvements to the subgrid modelling at high redshift, particularly that of star formation feedback, may improve the agreement.

To investigate what effect our sampling of highly overdense regions has on the composite shape of the SFRF, we now look at the overdensity dependence of the SFRF.

3.3.2 Environmental dependence of the SFRF

Fig. 12 shows the SFRF for regions binned by their log-overdensity. There is almost no variation in the shape as a function of overdensity except for the highest overdensities, which show a more prominent double-Schechter knee in the high-SFR regime. This behaviour is identical to that seen for the GSMF. This may explain why the shape of the FLARES composite SFRF differs with those of other cosmological models. FLARES better samples the rare, high-density regions that contribute significantly to the high-SFR sequence (SFS) relation at all redshifts (see Section 3.4).
Figure 11. Evolution of the FLARES composite SFRF (coloured, solid lines), compared with observational constraints from UV data and other model predictions. Smit et al. (2012) derive SFRs from UVLF data, as do Katsianis et al. (2017) using Bouwens et al. (2015) data. Both are corrected to a Chabrier IMF using the conversion factors quoted in Kennicutt & Evans (2012). The Santa-Cruz SAM (Yung et al. 2019b, dashed line) and BLUETIDES simulation (Wilkins et al. 2017) show a different behaviour, with a power-law shape at higher redshifts, in contrast to the prominent knee seen in FLARES up to $z = 10$. Both L-GALAXIES models also show similar behaviour, though with lower normalization at the high-SFR end (Henriques et al. 2015, 2020).

these fits for bins ±0.2 dex wide centred at $\log_{10}(\psi/M_\odot \text{ yr}^{-1}) = [-0.5, 0.5]$. The normalization increases with increasing overdensity as expected. The trends with redshift are also broadly similar to those seen for the GSMF.9

3.4 The star-forming sequence

Observations at both high- and low-z suggest a tight relation between SFR and stellar mass, known as the ‘main sequence’, or SFS

(9)The only exception being the gradient of the GSMF relation at $M_*/M_\odot = 10^{10.7}$, which decreases with redshift, whereas the redshift dependence is positive for the SFRF at all SFRs.

(Brichmann et al. 2004; Noeske et al. 2007; Speagle et al. 2014). The SFS is typically parametrized as a linear relation

$$\log_{10}(\psi) = \alpha \log_{10}(M_*/M_\odot) + \beta.$$  (10)

Observations suggest that the normalization $\beta$ increases with redshift, whilst the slope $\alpha$ remains relatively constant (Daddi et al. 2007; Santini et al. 2009; Salmon et al. 2015).

There have been suggestions of a turnover in the SFS at high stellar masses, though the turnover mass, and its evolution with redshift, are less clear (Lee et al. 2015; Tasca et al. 2015; Santini et al. 2017). Such a turnover is necessary to explain the GSMF at low redshift; a single power law slope would lead to too many massive galaxies being formed (between $10^{10} < M_*/M_\odot < 10^{11}$; Leja et al. 2015).
The Flares SFRF between $z = 5$ and 9 split by binned log-overdensity. The binning is shown in the legend, along with the number of regions in each bin. Poisson 1σ uncertainties are shown for each bin from the simulated number counts. The normalization increases with increasing overdensity, and the maximum SFR increases.

Figure 12. The Flares SFRF between $z = 5$ and 9 split by binned log-overdensity. The binning is shown in the legend, along with the number of regions in each bin. Poisson 1σ uncertainties are shown for each bin from the simulated number counts. The normalization increases with increasing overdensity, and the maximum SFR increases.

Table 3. Fits to the normalization, $\log_{10}(\psi/\text{Mpc}^{-3} \text{dex}^{-1})$ of the SFRF at different redshifts and SFRs (see Section 3.3.2).

| $z$ | $\log_{10}(\psi/\text{M}_\odot \text{yr}^{-1})$ | $m$ | $c$ |
|-----|--------------------------------|-----|-----|
| 5   | -0.5                          | 3.0 | -2.0|
| 7   | -0.5                          | 3.2 | -2.2|
| 9   | -0.5                          | 3.5 | -2.6|
| 5   | 0.5                           | 3.8 | -2.8|
| 7   | 0.5                           | 4.4 | -3.4|
| 9   | 0.5                           | 4.5 | -4.0|

The turnover may be evidence for a change in the dominant channel of stellar mass growth, from smooth gas accretion to merger-driven growth.

The top panel of Fig. 13 shows the redshift evolution of the SFS in FLARES. In the bottom panel of Fig. 13, we also show the specific-star formation rate (sSFR) against $M_\star$ relation. To construct the median lines, we weight each galaxy in the sample by the appropriate factor for the overdensity of the resimulation volume, as described in Section 2.4.10 There is a clear trend of decreasing normalization with decreasing redshift, approximately 0.5 dex between $z = 10–5$.

There is some noise in the weighted relation at $z = 8$ for galaxies with $M_\star > 10^{9.5} \text{M}_\odot$; we checked, and found that this is due to a small number of galaxies in mean density regions above this mass limit with low SFRs, biasing the normalization down.

We have not excluded ‘passive’ galaxies from our measurement of the SFS. We present results for the SFS assuming different sSFR cuts in Appendix B, though note here that they make negligible difference to the relations at $z \geq 5$ for even the most liberal cuts.

There is a clear turnover in the FLARES SFS at high masses ($\sim 10^{9.3} \text{M}_\odot$). We account for this by fitting a piecewise-linear relation, with an upper and lower mass part, for stellar mass re-normalized at $10^{9.7} \text{M}_\odot$.

$$\log_{10}\psi = \alpha_1 \log_{10}(M_\star/10^{9.7} \text{M}_\odot) + \beta_1 \quad x \leq x_0$$

$$\log_{10}\psi = \alpha_2 \log_{10}(M_\star/10^{9.7} \text{M}_\odot) + \beta_2 \quad x \geq x_0,$$

where $\alpha_1$ is the low-mass slope, $\alpha_2$ is the high-mass slope, and $x_0$ is the turnover mass in log-solar masses. The normalization at the turnover, $\beta_0$, is then given by

$$\beta_0 = \beta_2 + \alpha_2 x_0$$

$$= \beta_1 + \alpha_1 x_0.$$  

We use the SCIPY implementation of non-linear least squares to perform the fit, combined with a non-parametric bootstrap approach for estimating parameter uncertainties. The bootstrap is implemented as follows: we select, with replacement, 10,000 times from the original data, each resample being the same size as the original data. We then fit each sample independently; parameter estimates are given by the median of the resampled fit distributions, and uncertainties are given as the 1σ spread in the distributions (unless otherwise stated). The parameter fits are quoted in Table C1.

Fig. 14 shows the redshift evolution of each parameter against observational constraints where available.11 There are few robust constraints; parameter estimates are given by the median of the resampled fit distributions, and uncertainties are given as the 1σ spread in the distributions (unless otherwise stated). The parameter fits are quoted in Table C1.

10In fact, as shown in Fig. 16, the environmental dependence is very weak and so the weighted relations are very similar to the unweighted ones.

11The high mass slope and turnover are poorly constrained at $z = 10$ so we omit them.
Figure 13. Top: Redshift evolution of the FLARES composite SFS. Solid lines show the weighted composite SFS for centrals + satellites, with the 16th–84th spread shaded. Bottom: as for the top panel, but showing the sSFR–stellar mass relation.

observational constraints at \( z > 6 \), so we show constraints down to \( z = 3 \) to provide context to the redshift evolution (Behroozi et al. 2013; Salmon et al. 2015; Schreiber et al. 2015; Shivaei et al. 2015; Santini et al. 2017), including the compilation of pre-2014 measurements from Speagle et al. (2014). These all represent single power-law measurements. For all observations we quote the approximate lower mass completeness limit for the whole fit in the legend. We also show a direct comparison of the fits to binned data from Santini et al. (2017) and Salmon et al. (2015) in Fig. 15 at \( z = 5–6 \).

The normalization is within the errors of the binned observations at these redshifts. The fitted normalization \( \beta \) is also within the spread of the fitted relations at these redshifts, and continues the apparent increasing normalization with increasing redshift from \( z = 3 \). We also show the inverse age of the universe (in Gyr); the fall in SFS normalization approximately follows the same relation, but slightly shallower.

Figure 14. Redshift evolution of the piecewise-linear fit to the SFS. Observational results are plotted where available in grey, from Behroozi, Wechsler & Conroy (2013), Speagle et al. (2014), Shivaei et al. (2015), Salmon et al. (2015), Schreiber et al. (2015), and Santini et al. (2017). The lower and upper mass completeness limits for these studies are quoted in the legend. Top: high- and low-mass slope, in orange and blue, respectively. Middle: normalization, \( \beta \), in orange. The inverse age of the universe in Gyr is shown in green; the normalization approximately follows the same relation, but with a slightly shallower evolution. Bottom: turnover mass in log-solar masses, in orange.

The slope of the observed relations shows considerable scatter spanning the range \( \sim 0.5–1.1 \). We suggest that this is due to the lower mass limit of these observations (quoted in the legend of Fig. 14). Since these studies fit a single power law, and assume a high lower mass completeness limit, \( (M_*/M_\odot) > 10^{10.3} \), the measured slope will be biased to shallower slopes. This can also be seen clearly in the binned relations in Fig. 15; both Santini et al. (2017) and Song et al. (2016) straddle the turnover mass in FLARES. Finally, this can also be seen in the redshift evolution of these studies. The observed slopes of Salmon et al. (2015), Behroozi et al. (2013), and Santini et al. (2017) all show a negative correlation with redshift. The lower mass completeness limit of these studies also increases with increasing redshift; as it increases, they tend to probe just the high-mass end of the SFS, rather than the steeper low-mass end. This
suggestions that many high redshift measures of the SFS, where the mass completeness does not extend to very low masses, are only probing the SFS at stellar masses above the turnover, and the measured slopes do not represent a universal relation for all masses.

The turnover mass shows a negative correlation with redshift, increasing from $\sim 10^{9.2}$ to $10^{9.6} M_\odot$ between $z = 9-5$. Ceverino, Klessen & Glover (2018) show no turnover in their FirstLight simulation results, but they do not probe above $10^{9.5}$ at $z = 6$, which is consistent with where we constrain the turnover. There are unfortunately no observational constraints on the turnover mass at $z > 3$. We note that the turnover mass is much lower than that measured in low-$z$ studies ($>10^{10} M_\odot$ at $z \leq 3$, Whitaker et al. 2014; Tasca et al. 2015).

3.4.1 Environmental dependence of the SFS

Fig. 16 shows the SFS for each region individually at $z = 5$, coloured by overdensity. The highest overdensities reach to higher stellar masses, as expected. However, there is no dependence on overdensity of either the normalization nor shape of the SFS, and we see this up to $z = 10$. Observationally, at $z \sim 2$ there is a similar lack of dependence on environment as measured between protocluster and field regions (Koyama et al. 2013, 2014; Shimakawa et al. 2017, 2018), though these authors do note some differences in dense subgroups in protocluster candidates (we leave an investigation of the small-scale overdensity dependence of the SFS to future work). However, at $z > 5$ Harikane et al. (2019) find a $5 \times$ enhancement in the SFR (at fixed stellar mass) of Lyman $\alpha$ emitters in protoclusters compared to the field, though they only probe the low-mass regime ($M_* < 10^9 M_\odot$). It is as yet unclear whether these galaxies represent the main SFS, or starbursts that lie above it. Harikane et al. (2019) show that dusty star-forming galaxies traced in the sub-mm are also spatially correlated with these structures (Geach et al. 2017), and lead to significant enhancements in the cosmic SFR density compared to the Madau & Dickinson (2014) relation. In FLARES, whilst the normalization of the SFS at $z = 5-6$ is low in the stellar mass regime $M_* < 10^9 M_\odot$ compared to observational constraints (Song et al. 2016; Santini et al. 2017), Fig. 16 shows that FLARES does produce a number of galaxies with SFRs at least $5 \times$ higher than on the main relation. The fraction of these ‘starburst’ galaxies is marginally higher in the overdense regions ($\sim 2$ per cent compared to $\sim 1$ per cent in the mean overdensity regions), which suggests starbursts may be biased to high-density regions. However, the relative and absolute numbers are small.

We leave a thorough exploration of the passive and star-bursting galaxy populations in FLARES to future work.

4 CONCLUSIONS

We have presented the first results from the FLARES simulations, resimulations with full hydrodynamics of a range of overdensities during the EoR ($z \geq 5$) using the EAGLE (Schaye et al. 2015) physics. We described our novel weighting procedure that allows the construction of composite distribution functions that mimic extremely large periodic volumes, significantly extending the dynamic range without incurring prohibitively large computational expense. To demonstrate
we presented results for the GSMF, the SFRF, and the SFS (SFS: SFR versus $M_\star$). Our findings are as follows:

(i) The FLARESGSMF exhibits a clear double-Schechter shape up to $z = 10$. Fits assuming this form show an increasing normalization, shallower low-mass slope and higher characteristic turnover mass with decreasing redshift. The GSMF is in good agreement with observational constraints at all redshifts up to $z = 8$, at which point there is some tension at the knee of the distribution. The normalization, and to a lesser extent the shape, of the GSMF shows a strong environmental dependence (i.e. bias).

(ii) The SFRF also exhibits a clear double-Schechter shape in the high-SFR regime. As for the GSMF, the normalization increases and the low-mass slope decreases with decreasing redshift; however, the characteristic turnover mass varies only weakly with redshift.

(iii) The SFS shows no obvious dependence on environment. The low-mass slope is relatively invariant with redshift, whereas the high-mass slope decreases with decreasing redshift. The characteristic turnover mass increases slowly with decreasing redshift, and the normalization decreases by about a factor of 3 between redshifts 10 and 5. There is reasonably good agreement with observational constraints at $z = 5$–6.

Upcoming space-based observatories, such as JWST, Euclid, and Roman will provide further probes of the GSMF and SFRF up to $z = 10$. The large volumes probed by Euclid and Roman in particular will provide stronger constraints on those extreme galaxies that populate the high-mass/high-SFR tails of each distribution. Our weighting scheme provides a means of testing the latest, high resolution hydrodynamic simulations against such constraints. We will also be able to test the impact of cosmic variance on these large surveys.

ACKNOWLEDGEMENTS

We wish to thank the anonymous referee for detailed comments and suggestions that improved this paper. We also wish to thank Scott Kay and Adrian Jenkins in particular for their invaluable help and support from the Royal Society under grant RGF/EA/181016. APV acknowledges the support of his PhD studentship from UK STFC DISCnet.

DATA AVAILABILITY

The data underlying this article (stellar masses and SFRs between $z = 5$–10) are available at flaresimulations.github.io/data. All of the codes used for the data analysis are public and available at github.com/flaresimulations.

REFERENCES

Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Bahe Y. M. et al., 2016, MNRAS, 456, 1115
Bahe Y. M. et al., 2017, MNRAS, 470, 4186
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Barnes D. J., Kay S. T., Henson M. A., McCarthy I. G., Schaye J., Jenkins A., 2017a, MNRAS, 465, 213
Barnes D. et al., 2017b, MNRAS, 471, 1088
Barrow K. S. S., Wise J. H., Norman M. L., O’Shea B. W., Xu H., 2017, MNRAS, 469, 4863
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Behroozi P., Silk J., 2018, MNRAS, 477, 5382
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benitez-Llambay A., 2015, py-sphviewer: Py-sphviewer v1.0.0
Benson J. A., 2010, Phys. Rep., 495, 33
Bhatwadekar R., Conselice C. J., Margele-Fentabent B., Duncan K., 2019, MNRAS, 486, 3805
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, MNRAS, 465, 32
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bromm V., Yoshida N., 2011, ARA&A, 49, 373
Castellano M. et al., 2016, ApJ, 823, L40
Ceverino D., Klessen R. S., Glover S. C. O., 2018, MNRAS, 480, 4842
Chabrier G., 2003, PASP, 115, 763
Chiang Y.-K., Overzier R., Gebhardt K., 2013, ApJ, 779, 127
Chiang Y.-K., Overzier R. A., Gebhardt K., Henriques B., 2017, ApJ, 844, L23
Clay S. J., Thomas P. A., Wilkins S. M., Henriques B. M. B., 2015, MNRAS, 451, 2692
Cooray A. et al., 2019, BAAS, 51, 48
Crain R. A. et al., 2009, MNRAS, 399, 1773
Crain R. A. et al., 2015, MNRAS, 450, 1937
Crain R. A. et al., 2017, MNRAS, 464, 4204
Cullen L., Dehnen W., 2010, MNRAS, 408, 669
Daddi E. et al., 2007, ApJ, 670, 156
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Di Matteo T., Croft R. A. C., Feng Y., Waters D., Wilkins S., 2017, MNRAS, 467, 4243
Dubois Y. et al., 2014, MNRAS, 444, 1453
Duncan K. et al., 2014, MNRAS, 444, 2960
Durier F., Dalla Vecchia C., 2012, MNRAS, 419, 1453
Eisenstein D. J., Udalski A., 2001, AJ, 122, 1
Fang J. J. et al., 2018, ApJ, 858, 100
Feng Y., Di Matteo T., Croft R., Tenneti A., Bird S., Battaglia N., Wilkins S., 2015, ApJ, 808, L17
Feng Y., Di Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Finlator K., Kneib J.-P., Cooray A., 2019, ARA&A, 57, 61
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Furlong M. et al., 2015, MNRAS, 450, 4486
Furlong M. et al., 2017, MNRAS, 465, 722
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Georgiev K., Karim A., Schaye J., Guiderdoni B., 2017, A&A, 600, L21
Gibson B., Moore B., 2000, MNRAS, 319, 1684
Gottl 10. The large volumes probed by Roman will provide further probes of the GSMF and SFRF up to Scott Kay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper. We also wish to thank the anonymous referee for detailed comments and suggestions that improved this paper. We also wish to thank Scott Kay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper. We also wish to thank Scott Kay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper. We also wish to thank Scott Kay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper. We also wish to thank Scott Kay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper. We also wish to thank ScottKay and Adrian Jenkins in particular for their invaluable help and suggestions that improved this paper.
Table A1. Regions selected from the parent volume for resimulation. We provide their positions within the parent volume, their overdensity $\delta$ as defined by equation (1), their rms overdensity $\sigma$, and weights, $f_j$, calculated as per Section 2.4.

| index | $(x, y, z)/(h^{-1} \text{cMpc})$ | $\delta$ | $\sigma$ | $f_j$ |
|-------|---------------------------------|----------|----------|-------|
| 0     | (623.5, 1142.2, 1525.3)         | 0.970    | 5.62     | 0.000027 |
| 1     | (524.1, 1203.6, 1138.5)         | 0.918    | 5.41     | 0.000196 |
| 2     | (542.2, 1709.6, 571.1)          | 0.852    | 5.12     | 0.000429 |
| 3     | (153.6, 1762.0, 531.3)          | 0.849    | 5.11     | 0.000953 |
| 4     | (39.8, 1686.1, 1850.6)          | 0.846    | 5.09     | 0.000444 |
| 5     | (847.6, 1444.0, 1062.6)         | 0.842    | 5.07     | 0.000828 |
| 6     | (1198.2, 135.5, 1375.3)         | 0.841    | 5.07     | 0.000666 |
| 7     | (1012.0, 1514.4, 1454.8)        | 0.839    | 5.06     | 0.001178 |
| 8     | (591.0, 359.6, 1610.2)          | 0.839    | 5.06     | 0.000265 |
| 9     | (746.4, 820.5, 945.2)           | 0.833    | 5.03     | 0.0001029|
| 10    | (1181.9, 1171.1, 974.1)         | 0.830    | 5.02     | 0.000387 |
| 11    | (38.0, 670.5, 47.0)             | 0.829    | 5.02     | 0.000719 |
| 12    | (1989.7, 368.7, 2076.5)         | 0.828    | 5.01     | 0.000668 |
| 13    | (1659.0, 1306.6, 760.8)         | 0.824    | 4.99     | 0.000488 |
| 14    | (57.8, 883.7, 2098.2)           | 0.821    | 4.98     | 0.001190 |
| 15    | (609.0, 2018.6, 115.7)          | 0.820    | 4.98     | 0.000757 |
| 16    | (122.9, 1124.1, 1304.8)         | 0.816    | 4.90     | 0.003738 |
| 17    | (1395.2, 415.7, 1575.9)         | 0.816    | 4.90     | 0.004678 |
| 18    | (128.3, 216.9, 258.4)           | 0.813    | 3.00     | 0.009359 |
| 19    | (1400.6, 1686.1, 806.0)         | 0.813    | 3.00     | 0.012324 |
| 20    | (699.4, 1760.2, 1725.9)         | 0.806    | 2.00     | 0.029311 |
| 21    | (1951.8, 2022.3, 1709.6)        | 0.806    | 2.00     | 0.027954 |
| 22    | (755.4, 1122.3, 867.5)          | 0.804    | 2.00     | 0.029311 |
| 23    | (516.9, 325.3, 603.6)           | 0.804    | 1.00     | 0.057876 |
| 24    | (937.9, 1382.5, 1077.1)         | 0.800    | 1.00     | 0.062099 |
| 25    | (1675.3, 1492.8, 1335.5)        | 0.800    | 1.00     | 0.074502 |
| 26    | (1270.5, 518.7, 862.0)          | 0.789    | 1.00     | 0.080377 |
| 27    | (242.2, 1881.3, 1624.7)         | 0.789    | 1.00     | 0.063528 |
| 28    | (1545.8, 1720.5, 1608.4)        | 0.789    | 1.00     | 0.058231 |
| 29    | (430.1, 296.4, 359.6)           | 0.789    | 1.00     | 0.034467 |
| 30    | (1733.1, 1097.0, 1060.8)        | 0.789    | 1.00     | 0.024216 |
| 31    | (1821.7, 947.0, 1431.3)         | 0.789    | 1.00     | 0.012087 |
| 32    | (1913.8, 1035.7, 45.2)          | 0.789    | 1.00     | 0.013127 |
| 33    | (2009.6, 2024.1, 1693.4)        | 0.789    | 1.00     | 0.064280 |
| 34    | (338.8, 934.3, 1646.4)          | 0.789    | 1.00     | 0.066277 |
| 35    | (1693.4, 914.5, 1977.1)         | 0.789    | 1.00     | 0.076001 |
| 36    | (778.9, 900.0, 1866.8)          | 0.789    | 1.00     | 0.076486 |
| 37    | (1790.9, 1239.7, 1335.5)        | 0.789    | 1.00     | 0.070408 |
| 38    | (2078.3, 77.7, 141.0)           | 0.789    | 1.00     | 0.062451 |
| 39    | (818.7, 110.2, 1628.3)          | 0.789    | 1.00     | 0.002721 |

APPENDIX B: THE IMPACT OF CUTTING PASSIVE GALAXIES FROM THE STAR-FORMING SEQUENCE

In Section 3.4, we showed the SFS assuming no cut for passive galaxies. We now briefly explore the impact of applying an evolving cut in specific star formation rate (sSFR), and how this impacts the SFS. We employ an sSFR cut that excludes those galaxies whose current star formation is insufficient to double the mass of the galaxy within twice the current age of the universe

\[
\text{sSFR} > \frac{1}{2 \times \text{age}} ,
\]

which leads to an evolving threshold for quiescence with redshift, shown in Fig. B1. Using this cut, we exclude 979 galaxies at $z = 5$ (out of a total of 32,824 with stellar masses above $10^8 \text{M}_\odot$).

Fig. B2 shows the SFS assuming this cut. There is almost no difference between this relation and that shown in Fig. 13. We tested using different thresholds (mass multiples of $\times 3^2$ and $\times 3$) and found that all our results are insensitive to the multiple of mass chosen. Observations typically use $UVJ$ colour to discriminate quiescent objects (e.g. Whitaker et al. 2011); at $z \sim 2$, this leads to a similar threshold for quiescence as an sSFR cut (Fang et al. 2018).

APPENDIX C: FITTED DISTRIBUTION FUNCTIONS

Tables C2 and C3 show double-Schechter fit parameters to the GSMF and SFRF. We use \texttt{fitdf}, a python module for fitting arbitrary distribution functions using Markov Chain Monte Carlo (MCMC). \texttt{fitdf} is built around the popular EMCEE package.
Figure C1. Posteriors from the GSMF fit at $z = 7$. $\alpha_2$ is fixed at $-1$ and is not shown.

Table C1. Best-fitting two-part piecewise-linear fits to the SFS.

| $z$ | $\alpha_0 + 9.7$ | $\alpha_1$ | $\alpha_2$ | $\beta$ |
|-----|-----------------|------------|------------|--------|
| 5   | 9.60            | 1.23       | 0.62       | 1.42   |
| 6   | 9.45            | 1.27       | 0.70       | 1.60   |
| 7   | 9.35            | 1.31       | 0.72       | 1.76   |
| 8   | 9.20            | 1.33       | 0.80       | 1.90   |
| 9   | 9.16            | 1.31       | 0.91       | 1.93   |
| 10  | --              | 1.24       | --         | --     |

where the subscript $i$ represents the bin of the property being measured, $N_{i,\text{obs}}$ is the inferred number of galaxies using the composite number density multiplied by the parent box volume, $N_{i,\text{exp}}$ is the expected number from the model, and $\sigma_i$ is the error estimate. Using this form, $\sigma$ can be explicitly provided from the resimulated number counts, $\sigma_i = N_{i,\text{obs}}/\sqrt{n_{i,\text{obs}}}$, where $n_{i,\text{obs}}$ is the number counts in bin $i$ from the resimulations.

A Poisson form of the likelihood is typically used for distribution function analyses in Astronomy due to the relatively small number of observations. Due to our resimulation approach we cannot use this form of the likelihood, since the number counts obtained from the composite approach, scaled to the size of the parent box volume, significantly underestimate the errors. Instead, we use a Gaussian form for the likelihood

$$\log(L) = -\frac{1}{2} \sum_i \frac{(N_{i,\text{obs}} - N_{i,\text{exp}})^2}{\sigma_i^2} + \log(\sigma_i^2), \quad (C1)$$

where the subscript $i$ represents the bin of the property being measured, $N_{i,\text{obs}}$ is the inferred number of galaxies using the composite number density multiplied by the parent box volume, $N_{i,\text{exp}}$ is the expected number from the model, and $\sigma_i$ is the error estimate. Using this form, $\sigma$ can be explicitly provided from the resimulated number counts, $\sigma_i = N_{i,\text{obs}}/\sqrt{n_{i,\text{obs}}}$, where $n_{i,\text{obs}}$ is the number counts in bin $i$ from the resimulations.

We use flat uniform priors in $\log_{10}(D^*)$, $\alpha_1$, $\log_{10}(\phi_1^*)$, and $\log_{10}(\phi_2^*)$. We fix $\alpha_2 = -1$ by setting a narrow top-hat prior around this value. We run chains of length $10^4$, then calculate the autocorrelation time, $\tau$, on these chains (Goodman & Weare 2010). We use $\tau$ to estimate the burn-in ($\tau \times 4$) and thinning ($\tau$) on our chains. Example posteriors for each parameter in a fit to the $z = 7$ GSMF are shown as a corner plot in Fig. C1.

Table C1 shows the piecewise-fits to the SFS; the fitting procedure is described in Section 3.4.

(v3.0, Foreman-Mackey et al. 2013). The code can be found at https://github.com/flaresimulations/fitDF.

(12) The chains for each fit are available at https://flaresimulations.github.io/flares/data.html.
Table C2. Best-fitting double-Schechter function parameter values for the GSMF. $\alpha_2$ is fixed at $-1$.

| $z$ | $M^*$ | $\log_{10}(\phi_1^*(\text{Mpc}^{-3} \text{dex}^{-1}))$ | $\log_{10}(\phi_2^*(\text{Mpc}^{-3} \text{dex}^{-1}))$ | $\alpha_1$ |
|-----|-------|-------------------------------------------------|-------------------------------------------------|-----------|
| 10  | 9.117$^{+0.041}_{-0.045}$ | $-6.557^{+0.188}_{-0.197}$ | $-4.87^{+0.065}_{-0.07}$ | $-3.542^{+0.193}_{-0.206}$ |
| 9   | 9.488$^{+0.036}_{-0.044}$ | $-6.372^{+0.116}_{-0.112}$ | $-4.832^{+0.056}_{-0.057}$ | $-3.073^{+0.076}_{-0.077}$ |
| 8   | 9.577$^{+0.039}_{-0.041}$ | $-5.904^{+0.081}_{-0.08}$ | $-4.565^{+0.059}_{-0.058}$ | $-2.833^{+0.065}_{-0.064}$ |
| 7   | 9.831$^{+0.039}_{-0.035}$ | $-5.443^{+0.051}_{-0.054}$ | $-4.374^{+0.052}_{-0.059}$ | $-2.515^{+0.033}_{-0.032}$ |
| 6   | 10.089$^{+0.029}_{-0.035}$ | $-5.057^{+0.047}_{-0.043}$ | $-4.159^{+0.046}_{-0.044}$ | $-2.293^{+0.019}_{-0.023}$ |
| 5   | 10.326$^{+0.039}_{-0.02}$ | $-4.686^{+0.023}_{-0.024}$ | $-3.942^{+0.033}_{-0.034}$ | $-2.111^{+0.012}_{-0.011}$ |

Table C3. Best-fitting double-Schechter function parameter values for the SFRF. $\alpha_2$ is fixed at $-1$.

| $z$ | SFR$^*$ | $\log_{10}(\phi_1^*(\text{Mpc}^{-3} \text{dex}^{-1}))$ | $\log_{10}(\phi_2^*(\text{Mpc}^{-3} \text{dex}^{-1}))$ | $\alpha_1$ |
|-----|-------|-------------------------------------------------|-------------------------------------------------|-----------|
| 5   | 1.402$^{+0.049}_{-0.067}$ | $-6.525^{+0.142}_{-0.123}$ | $-5.022^{+0.07}_{-0.069}$ | $-2.978^{+0.071}_{-0.074}$ |
| 5   | 1.359$^{+0.036}_{-0.044}$ | $-5.941^{+0.093}_{-0.093}$ | $-4.645^{+0.058}_{-0.058}$ | $-2.772^{+0.064}_{-0.06}$ |
| 5   | 1.433$^{+0.032}_{-0.028}$ | $-5.639^{+0.059}_{-0.06}$ | $-4.431^{+0.049}_{-0.058}$ | $-2.62^{+0.051}_{-0.045}$ |
| 5   | 1.633$^{+0.03}_{-0.027}$ | $-5.509^{+0.052}_{-0.057}$ | $-4.186^{+0.036}_{-0.04}$ | $-2.482^{+0.036}_{-0.038}$ |
| 5   | 1.684$^{+0.015}_{-0.015}$ | $-5.059^{+0.041}_{-0.039}$ | $-3.907^{+0.024}_{-0.026}$ | $-2.307^{+0.026}_{-0.025}$ |
| 5   | 1.755$^{+0.011}_{-0.012}$ | $-4.68^{+0.033}_{-0.033}$ | $-3.644^{+0.02}_{-0.02}$ | $-2.139^{+0.02}_{-0.019}$ |

This paper has been typeset from a TeX/LaTeX file prepared by the author.