

# Realistic mock observations of the sizes and stellar mass surface densities of massive galaxies in FIRE-2 zoom-in simulations

T. Parsotan <sup>1</sup>, <sup>1\*</sup> R. K. Cochrane <sup>2</sup>, C. C. Hayward <sup>3</sup>, D. Anglés-Alcázar, <sup>3,4</sup> R. Feldmann <sup>5</sup>,  
C. A. Faucher-Giguère <sup>6</sup>, S. Wellons <sup>6</sup> and P. F. Hopkins <sup>7</sup>

<sup>1</sup>Department of Physics, Oregon State University, 301 Weniger Hall, Corvallis, OR 97331, USA

<sup>2</sup>Harvard–Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA 02138, USA

<sup>3</sup>Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA

<sup>4</sup>Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA

<sup>5</sup>Institute for Computational Science, University of Zurich, Zurich CH-8057, Switzerland

<sup>6</sup>Department of Physics & Astronomy and CIERA, Northwestern University, 1800 Sherman Ave, Evanston, IL 60201, USA

<sup>7</sup>TAPIR, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 2020 December 1. Received 2020 November 28; in original form 2020 September 21

## ABSTRACT

The galaxy size–stellar mass and central surface density–stellar mass relationships are fundamental observational constraints on galaxy formation models. However, inferring the physical size of a galaxy from observed stellar emission is non-trivial due to various observational effects, such as the mass-to-light ratio variations that can be caused by non-uniform stellar ages, metallicities, and dust attenuation. Consequently, forward-modelling light-based sizes from simulations is desirable. In this work, we use the SKIRT dust radiative transfer code to generate synthetic observations of massive galaxies ( $M_* \sim 10^{11} M_\odot$  at  $z = 2$ , hosted by haloes of mass  $M_{\text{halo}} \sim 10^{12.5} M_\odot$ ) from high-resolution cosmological zoom-in simulations that form part of the Feedback In Realistic Environments project. The simulations used in this paper include explicit stellar feedback but no active galactic nucleus (AGN) feedback. From each mock observation, we infer the effective radius ( $R_e$ ), as well as the stellar mass surface density within this radius and within 1 kpc ( $\Sigma_e$  and  $\Sigma_1$ , respectively). We first investigate how well the intrinsic half-mass radius and stellar mass surface density can be inferred from observables. The majority of predicted sizes and surface densities are within a factor of 2 of the intrinsic values. We then compare our predictions to the observed size–mass relationship and the  $\Sigma_1 - M_*$  and  $\Sigma_e - M_*$  relationships. At  $z \gtrsim 2$ , the simulated massive galaxies are in general agreement with observational scaling relations. At  $z \lesssim 2$ , they evolve to become too compact but still star forming, in the stellar mass and redshift regime where many of them should be quenched. Our results suggest that some additional source of feedback, such as AGN-driven outflows, is necessary in order to decrease the central densities of the simulated massive galaxies to bring them into agreement with observations at  $z \lesssim 2$ .

**Key words:** radiative transfer – galaxies: formation.

## 1 INTRODUCTION

Observations of distant galaxies are crucial for understanding the physics orchestrating galaxy evolution and the assembly of galaxy structures (see Conselice 2014, for a review). The period around the peak of cosmic star formation, around  $1 \lesssim z \lesssim 3$ , is particularly important; at this epoch, stellar mass is building most rapidly (see the review by Madau & Dickinson 2014 and references therein), and measuring galaxy structure here can provide constraints on the drivers of high star formation rates. In particular, structures and morphologies can help distinguish between models of star formation (‘inside-out’ versus ‘outside-in’ growth; e.g. van Dokkum et al. 2010, 2015; Wuyts et al. 2012; Tacchella et al. 2016, 2018; Spilker et al. 2019; Whitney et al. 2019), determine the relative importance of *in situ* star formation as opposed to merger-driven mass assembly (Stott

et al. 2011; Newman et al. 2012; Huertas-Company et al. 2015; Hill et al. 2017, 2019) and discriminate between quenching mechanisms (Wu et al. 2018, 2020; Wang et al. 2019). However, characterization of the structures of high-redshift galaxies has historically been challenging, due to the small angular sizes of distant galaxies and the resolution limitations of ground-based telescopes.

Space-based imaging, notably the *Hubble Space Telescope* (HST), has been critical to the development of this field. Deep data, in particular from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), have the necessary combination of high angular resolution (of the order of 0.1–0.2 arcsec) and sensitivity to infer typical sizes of massive galaxies to  $z \sim 7$  (Allen et al. 2017; Hill et al. 2017). At low and intermediate redshifts ( $0 \lesssim z \lesssim 3$ ), more detailed analysis has been possible, and lower stellar mass galaxies can be studied. It is now well established that galaxy size correlates with properties such as stellar mass, star formation rate, and colour, and that empirical

\* E-mail: [parsotat@oregonstate.edu](mailto:parsotat@oregonstate.edu)

scaling relations evolve with redshift. More massive galaxies are, on average, larger than less massive ones, both in the local Universe (Shen et al. 2003; Lange et al. 2015) and at high redshift (Trujillo et al. 2004; Barden et al. 2005). At fixed stellar mass and redshift, star-forming galaxies are larger than their quiescent counterparts, at least out to  $z \sim 2$  (e.g. Toft et al. 2009; Williams et al. 2010; Barro et al. 2017; Whitaker et al. 2017). At high redshift, galaxies tend to be more compact (Ferguson et al. 2004; Daddi et al. 2005; Buitrago et al. 2008), with the most significant size evolution observed for galaxies classified as quiescent (e.g. Williams et al. 2010; Carollo et al. 2013; Mosleh et al. 2017). These various correlations are encapsulated in the evolving size–mass relations (e.g. van der Wel et al. 2014, though see Suess et al. 2019 for extensive discussion of the pitfalls of observational measurements of galaxy size).

Stellar surface density (e.g. within the innermost 1 kpc) is also observed to be correlated with various galaxy properties. Massive, quiescent galaxies tend to have higher stellar surface densities, with less dense galaxies displaying higher star formation rates, on average (Franx et al. 2008; Williams et al. 2010; Whitaker et al. 2017). These relations also evolve with redshift; at fixed stellar surface density, galaxies at higher redshifts are more highly star-forming (Franx et al. 2008).

As observations have provided a clearer view of the history of stellar mass assembly, simulations have attempted to explain observational results and use them to constrain their sub-grid models for key physical processes such as feedback from stars and massive black holes. One important question that has been explored is how AGN feedback leaves an imprint on the physical sizes of galaxies and on their central densities (Fan et al. 2008; Dubois et al. 2013; Ishibashi, Fabian & Canning 2013; Wellons et al. 2015; Genel et al. 2018; van der Vlugt & Costa 2019). Choi et al. (2018) recently explored this with two sets of simulations, one with and one without AGN feedback (though including stellar feedback). They showed that the galaxies simulated with AGN feedback showed a suppression of central cooling, resulting in lower stellar mass density in their cores. Similarly, Appleby et al. (2020) show that the X-ray black hole feedback implemented in the SIMBA cosmological hydrodynamical simulations (Davé et al. 2019) pushes dense gas outwards, lowering the central specific star formation rate. Zoldan et al. (2019) also argue that quasar-driven mechanical winds are required to reconcile simulations with observed galaxy sizes. Therefore, AGN feedback appears to be required not only to quench star formation in massive galaxies (e.g. Somerville & Davé 2015), but also to regulate their sizes and central densities. However, most current cosmological simulations rely on extensive tuning of sub-grid parameters to match observations, which limits their predictive power.

Another key limitation of using simulations to interpret observational results lies in the lack of observable predictions made by most simulations. For example, while studies such as Choi et al. (2018) compared the sizes of their galaxies to observationally derived relations between stellar mass and surface density, they typically do not fully forward-model their simulations for direct comparison with observations. Cosmological simulations do not, in general, fold the details of dust geometry into their output, and providing predictions for simulated galaxies with all possible observational set-ups (given the numerous variables, such as telescope, waveband, seeing, and instrument noise) would be impossible. However, interest in this field is growing, with accessible radiative transfer software (e.g. Jonsson 2006; Jonsson, Groves & Cox 2010; Dullemond et al. 2012; Camps & Baes 2015) enabling mock observables to be generated with relative ease (e.g. Hayward et al. 2014; Hayward & Smith 2015; Trayford

et al. 2017; Camps et al. 2018; Liang et al. 2018, 2019; Cochrane et al. 2019; Ma et al. 2019).

In this paper, we evaluate the extent to which stellar feedback alone can regulate the sizes and central densities of the most massive galaxies in the Feedback In Realistic Environments 2 (FIRE-2) cosmological ‘zoom-in’ simulations (Hopkins et al. 2018b)<sup>1</sup> presented in Anglés-Alcázar et al. (2017b). FIRE simulations include a variety of stellar feedback physics implemented explicitly in a multiphase interstellar medium (ISM), and have been shown to reproduce the size–mass relation at  $z = 0$  for  $M_* < 10^{10.5} M_\odot$  (El-Badry et al. 2016), the Kennicutt–Schmidt relation (Orr et al. 2018), and the mass–metallicity relation (Ma et al. 2016). In this work, we probe the limits of stellar feedback in the extreme environments of the inner kpc of massive galaxies ( $M_* \sim 10^{11} M_\odot$ ) at  $z = 1–3$ .

We build on the work performed by Price et al. (2017), who test how well the sizes and stellar masses of FIRE galaxies can be recovered using mock images. We note a few key differences between their work and ours here. First, while Price et al. (2017) made use of the MassiveFIRE suite of galaxies (Feldmann et al. 2016, 2017), simulated using the original FIRE module, we use updated FIRE-2 physics and a novel implementation of supermassive black hole (SMBH) accretion and growth, but neglect AGN feedback entirely. In this paper, we put particular emphasis on our projection of the simulations into ‘observer-space’, including the effects of dust attenuation. Price et al. (2017) applied a Calzetti et al. (2000) dust attenuation curve to individual stellar particles, so that the effective attenuation depended on the line-of-sight density of dust (or metals). We implement a more sophisticated model for dust attenuation and re-emission, via three-dimensional continuum radiative transfer, and also model projection effects. This enables us to simulate multiwavelength emission in a self-consistent manner, accounting for the geometry of the dust and star particles. Like Price et al. (2017), we generate broad-band images and convolve these with typical telescope point spread function. We then analyse the resultant mock observations in the same way as real data. This involves fitting each mock observation with a Sérsic profile, and deriving the effective radius, the mass-to-light ratio, and the stellar mass surface density. Price et al. (2017) tested the recovery of intrinsic FIRE galaxy sizes at  $z \sim 2$ . Here, we extend these tests to a wider range of redshifts ( $1.25 < z < 2.76$ ), and additionally test the recovery of the stellar mass surface density. Further extending the previous study, we make direct comparisons to the observationally derived scaling relations presented by van der Wel et al. (2014) and Barro et al. (2017).

The structure of the paper is as follows. In Section 2, we describe the FIRE-2 simulations and outline the creation of mock observations. In Section 3, we describe the methods used to measure stellar mass surface densities and effective radii and present the results of the analysis (with additional plots presented in the Appendix). In Section 4, we discuss the implications of our findings. We present our conclusions in Section 5.

## 2 A SAMPLE OF SIMULATED HIGH-REDSHIFT GALAXIES

### 2.1 Four massive, central galaxies from the FIRE-2 simulations

The FIRE project (Hopkins et al. 2014, 2018b) is a set of state-of-the-art hydrodynamical cosmological zoom-in simulations. One of the key motivations for these simulations was a more complete

<sup>1</sup><http://fire.northwestern.edu>

understanding of the role of stellar feedback in galaxy evolution. Stellar feedback is believed to regulate star formation and the masses of galaxies over time. In particular, it is needed to match observationally inferred gas consumption time-scales (e.g. Hopkins, Quataert & Murray 2011), galaxy stellar mass functions (e.g. Davidzon et al. 2017), and the stellar mass-to-halo mass relation (Moster et al. 2010; Behroozi, Wechsler & Conroy 2013; Moster, Naab & White 2013; Cochrane et al. 2018), as well as to explain the metal enrichment of the circumgalactic medium and intergalactic medium (e.g. Oppenheimer & Davé 2006; Muratov et al. 2017; Hafen et al. 2019).

The FIRE project reaches sufficient mass and force resolution to model various stellar feedback processes including supernovae, photoheating, stellar mass-loss from O- and AGB-stars, and radiation pressure (see Dale 2015) directly. The simulations do this explicitly via two main methods. The first is resolving the formation of giant molecular clouds (GMCs). Star formation in the FIRE simulations takes place in self-gravitating (according to the Hopkins, Narayanan & Murray 2013 criterion), self-shielding molecular gas (see Krumholz & Gnedin 2011) at high densities ( $n_{\text{H}} > 1000 \text{ cm}^{-3}$  in the simulations used in this paper). The second involves modelling mass, metal, energy, and momentum return using the predictions of stellar population synthesis (SPS) models, without explicit parameter tuning, which is necessarily applied in large-volume cosmological simulations. The details of the feedback mechanisms implemented are presented in Hopkins et al. (2018a). The simulations have been broadly successful at generating galactic winds self-consistently (Anglés-Alcázar et al. 2017a; Muratov et al. 2017) and reproducing observed galaxy properties, such as stellar masses, star formation histories, metallicities, morphologies, and kinematics (Hopkins et al. 2014; van de Voort et al. 2015; Ma et al. 2016; Feldmann et al. 2017; Sparre et al. 2017; Ma et al. 2018).

In this paper, we focus on the four central galaxies of the massive haloes simulated by Anglés-Alcázar et al. (2017b) using the FIRE-2 model (Hopkins et al. 2018b). The haloes were first simulated by Feldmann et al. (2016, 2017) using the original FIRE model (Hopkins et al. 2014), as part of the MASSIVEFIRE suite. Compared to FIRE, our new FIRE-2 simulations are run with a more accurate hydrodynamics solver (a mesh-free Godunov solver implemented in the GIZMO<sup>2</sup> code; Gaburov & Nitadori 2011; Hopkins 2015). They also feature improved treatments of cooling and recombination rates, gravitational softening, and numerical feedback coupling, and they adopt a higher density threshold for star formation (Hopkins et al. 2018a). Our simulations include a new treatment for the seeding and growth of SMBHs via gravitational torque-driven accretion (though no AGN feedback); see Anglés-Alcázar, Özel & Davé (2013), and Anglés-Alcázar et al. (2015, 2017b) for details. The mass resolution is  $3.3 \times 10^4 M_{\odot}$  for gas and star particles and  $1.7 \times 10^5 M_{\odot}$  for dark matter particles. We denote our simulated central galaxies using their halo names, A1, A2, A4, and A8. At  $z = 2$ , these haloes have masses of  $M_{\text{halo}} \sim 10^{12.5} M_{\odot}$  and host central galaxies with stellar masses of  $7 \times 10^{10}$  to  $3 \times 10^{11} M_{\odot}$  and a range of assembly histories. A detailed kinematic analysis of these galaxies was presented in Wellons et al. (2020).

## 2.2 Post-processing with SKIRT

The FIRE-2 simulations do not make direct predictions for observed emission. In order to make mock images of these galaxies, we must model the intrinsic stellar emission, and then the propagation of that emission between the source and the observer. To do this, we use

the radiative transfer methods (Steinacker, Baes & Gordon 2013) implemented in the Stellar Kinematics Including Radiative Transfer (SKIRT)<sup>3</sup> Monte Carlo radiative transfer code (Baes et al. 2011; Camps & Baes 2015). Our methods are detailed in Cochrane et al. (2019), where we presented a detailed analysis of the spatially resolved dust continuum emission in the central galaxies of haloes A1, A2, A4, and A8. We provide a brief description of the procedure here.

We assign spectral energy distributions (SEDs) to the stars in each galaxy according to their ages and metallicities, using STARBURST99 templates (Leitherer et al. 1999) (these templates are also used in the FIRE simulations themselves), using a Kroupa (2001) initial mass function (IMF). We model dust within the galaxy using a dust-to-metals mass ratio of 0.4 (Dwek 1998; James et al. 2002), assuming that dust is destroyed in gas particles with temperature  $> 10^6 \text{ K}$  (Draine & Salpeter 1979; Tielens et al. 1994). We use a Weingartner & Draine (2001) Milky Way dust prescription, which includes a mixture of graphite, silicate, and PAH grains. SKIRT then tracks the paths of photons through this model dust distribution, tracking dust absorption (and self-absorption), scattering, and re-emission.

We place detectors at five different angles with respect to the face-on galaxy to create mock observations at various inclinations. The inclinations are:  $0^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ , and  $180^\circ$ , where  $0^\circ$  is a face-on view of the halo, defined with respect to the angular momentum vector of the galaxy’s gas particles, and  $90^\circ$  is an edge-on view of the halo. This allows us to incorporate observational uncertainties that may arise due to viewing angle effects into our analysis.

We perform this post-processing on a subset of snapshots spanning the peak of cosmic star formation, when stellar mass is building up very rapidly (Madau & Dickinson 2014). The redshifts studied are  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.76$ .

## 3 MOCK OBSERVATIONS

### 3.1 Deriving the sizes and surface densities of FIRE-2 galaxies

We create mock photometric observations of each of the SKIRT outputs, at each inclination, using the Johnson B and Kron–Cousins  $R$  filter transmission functions.<sup>4</sup> We also produce rest-frame  $\sim 5000 \text{ \AA}$  images, to match the rest-frame wavelength at which galaxy sizes are inferred in observational studies (e.g. van der Wel et al. 2014). We convolve the resultant images with the *HST* WFC3/IR point spread function (PSF) that was acquired by Skelton et al. (2014). For simplicity, we use this PSF for all images, following Price et al. (2017). We then resample each image to the *HST* WFC3 drizzled pixel scale of 0.06 arcsec and insert it into a blank region of a random CANDELS *HST F160W* image (Skelton et al. 2014; Price et al. 2017). This process yields mock observed images with realistic correlated noise (see Fig. 1, for an example of the workflow).

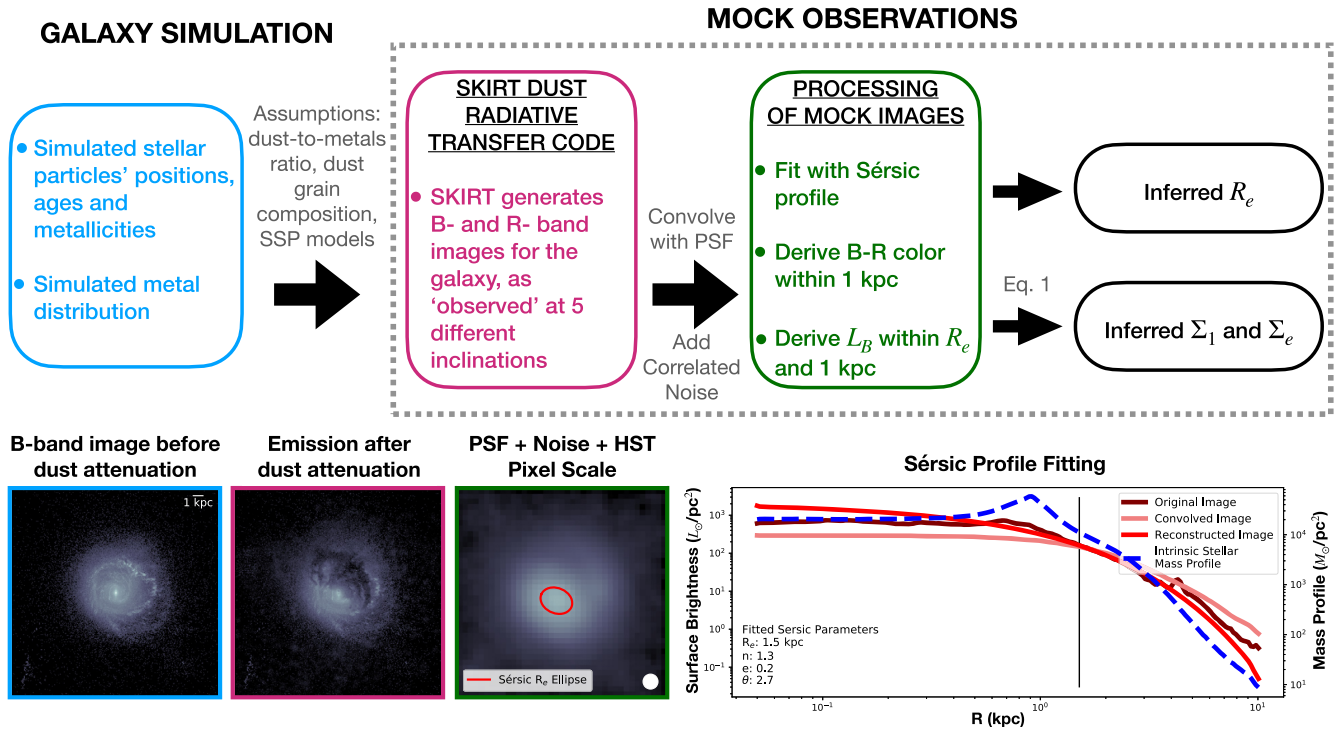
We then perform an analysis analogous to that used in observational studies. We fit a Sérsic profile (Sérsic 1968) to each mock  $B$ -band image using the STATMORPH python code (Rodríguez-Gomez et al. 2019). This procedure takes into account the PSF in order to fit for the intrinsic light distribution of the galaxy. We obtain the best-fitting ellipticity, angle of rotation, and the semimajor and semiminor axes. The effective radii ( $R_e$ ) quoted in this paper are the semimajor axes of the fitted Sérsic ellipses, to match the definition used by the observational work we compare to (van der Wel et al. 2014). Typically, the integrated Sérsic profiles recover  $\gtrsim 90$  per cent

<sup>2</sup><http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html>

<sup>3</sup><http://www.skirt.ugent.be>

<sup>4</sup><http://svo2.cab.inta-csic.es/theory/fps/>





**Figure 1.** The upper panel shows the workflow of this paper, and the lower panel shows an example of the process for an individual galaxy snapshot (galaxy A2, at  $z = 1.75$ , with a face-on orientation). The three images show the transformation of a  $B$ -band galaxy image from dust-free image (left), to dust-attenuated image produced using radiative transfer (centre), to the convolved image projected on to the *HST* 0.06 arcsec drizzled pixel scale (right; beam size shown in white). The right-hand image includes background correlated noise that would be observed in the rest-frame  $B$  band at the *HST* WFC/IR angular resolution. All three images have the same flux scale. The lower right-hand panel shows surface brightness profiles of the dust-unattenuated image (maroon) and the dust-attenuated, PSF-convolved image with noise (pink). The reconstructed surface brightness profile (red) is derived using Sérsic profile fits to the convolved image. The Sérsic profile fits are typically able to account for  $\gtrsim 90$  per cent of the light in the dust-attenuated, unconvolved image. The vertical black line shows the best-fitting stellar effective radius at this inclination, 1.5 kpc. The dashed blue line shows the stellar mass surface density profile derived from a two-dimensional projection of the stellar mass particle data.

of the light in the unconvolved SKIRT images. Comparisons of the surface brightness profiles of unconvolved  $B$ -band image to the best-fitting Sérsic profile show that the fit is also able to reconstruct the surface brightness profile (Fig. 1, bottom panel). In deriving the stellar effective radius in this way, we implicitly assume that there are no spatial variations in the mass-to-light ratio, consistent with the majority of observational analyses and in line with the results of Price et al. (2017). We discuss the limitations of this approach in Section 4.

We infer stellar mass surface densities from our synthetic images using well-established observational techniques. We follow the method outlined in Bell & de Jong (2001) to infer  $B$ -band mass-to-light ratios ( $M/L_B$ ) from observed  $B - R$  colour within 1 kpc and  $R_e$  apertures. We derive  $B - R$  colours using the methods established by Tacchella et al. (2015), which they show minimizes the effects of the PSF on the result. We use  $B$ -band and  $R$ -band Sérsic fits to calculate the flux at each of the two wavelengths, within a 1 kpc or  $R_e$  aperture. (Note that we repeated this procedure using  $R_{80}$ , the radius that contains 80 per cent of the galaxy's light, and found very similar results. This reflects the flat mass-to-light ratios seen in these simulated galaxies.) We then derive stellar mass within the aperture, using the  $B$ -band light (within an elliptical aperture with a semimajor axis of 1 kpc or the measured  $R_e$ ), the calculated  $B - R$  colour and an updated Bell & de Jong (2001) relation (see Appendix A). We repeat the process for galaxy images generated using different sky orientations to obtain an estimate of the standard deviation due to projection effects. This procedure enables us to calculate the stellar mass surface densities within 1 kpc and our measured  $R_e$  ( $\Sigma_1$  and  $\Sigma_e$ ,

respectively; see Cheung et al. 2012) that an observer would infer from the synthetic images.

The total stellar mass is derived directly from the simulation particle data, using a sphere of radius  $0.1R_{\text{vir}}$ , where  $R_{\text{vir}}$  is the virial radius of each galaxy. In principle, biases in recovering  $M_*$  would affect our comparisons with observations. However, a full investigation of this is beyond the scope of this paper, and various studies have found that  $M_*$  can be recovered within  $\sim 0.3$  dex (see e.g. Hayward & Smith 2015; Price et al. 2017; Carnall et al. 2018).

### 3.2 Recovery of intrinsic sizes and surface densities

Before embarking on the main analysis of this paper, we study how well the inferred effective stellar radii reflect the intrinsic half-mass radii calculated directly from the massive galaxy simulations (in three dimensions, using spherical shells). We find that the stellar effective sizes measured from the synthetic galaxy images tend to be slightly larger than the half-mass sizes calculated directly from the simulation particle data. This is the case for 13 of our 16 snapshots (see Fig. B1). Nevertheless, the majority (12/16) of our inferred sizes are within a factor of 2 of the intrinsic size, defined as the half-mass radius derived from the three-dimensional stellar particle data. The median values of  $\log_{10}(R_{e,\text{inferred}}/R_{e,\text{intrinsic}})$  are 0.15, 0.15, 0.32, and  $-0.02$  dex, at  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.75$ , respectively. Across all haloes, redshifts, and inclinations, the median offset is 0.17 dex, with a standard deviation of 0.20 dex. The

largest discrepancy between intrinsic and inferred galaxy size is seen at  $z = 2.25$ . This is driven by galaxy A4, which at this redshift is clumpy and quite obscured by dust (see Cochrane et al. 2019 for a more in-depth analysis of this amorphous morphology). This image is particularly difficult for STATMORPH to fit. This is also an issue for galaxy A8 at  $z = 1.75$ .

Intrinsic stellar mass surface densities are also fairly well recovered from mock observations for the majority of snapshots. The intrinsic stellar mass surface densities are acquired directly from the FIRE-2 simulations by calculating the total stellar mass within a given sphere, with a radius corresponding to 1 kpc or  $R_{e,\text{intrinsic}}$ , and then dividing by projected surface area (e.g.  $\Sigma_{e,\text{intrinsic}} = M(R_{e,\text{intrinsic}})/\pi R_{e,\text{intrinsic}}^2$ ). The median values of  $\log_{10}(\Sigma_{1,\text{inferred}}/\Sigma_{1,\text{intrinsic}})$  are 0.01, 0.07,  $-0.20$ , and 0.18 dex, at  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.75$ , respectively. Across all haloes, redshifts, and inclinations, the median offset between  $\Sigma_{1,\text{inferred}}$  and  $\Sigma_{1,\text{intrinsic}}$  is 0.04 dex, with a standard deviation of 0.37 dex. The corresponding median values of  $\log_{10}(\Sigma_{e,\text{inferred}}/\Sigma_{e,\text{intrinsic}})$  are  $-0.15$  dex,  $-0.15$  dex,  $-0.70$  dex, and 0.13 dex. Across all haloes, redshifts and inclinations, the median offset between  $\Sigma_{e,\text{inferred}}$  and  $\Sigma_{e,\text{intrinsic}}$  is  $-0.17$  dex, with a standard deviation of 0.52 dex. The discrepancy between the intrinsic and inferred surface densities of galaxy A4 at  $z = 2.25$  (see Section B, Fig. B2) is due to the same effects that extend  $R_e$  by a factor of  $\sim 3$ .

We also consider the uncertainties due to inclination effects explicitly (these uncertainties correspond to the size of the error bars,  $\sigma_{R_{e,\text{inferred}}}$ ,  $\sigma_{\Sigma_{e,\text{inferred}}}$ , and  $\sigma_{\Sigma_{1,\text{inferred}}}$ , shown in Figs B1 and B2). We first calculate the percentage uncertainties on the inferred radii ( $100 \times \sigma_{R_{e,\text{inferred}}}/R_{e,\text{inferred}}$ ), and derive the mean percentage uncertainty of the four haloes at each redshift. These are 20 per cent, 14 per cent, 19 per cent, and 24 per cent, at  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.75$ . Next, we repeat the process for the inferred stellar mass surface densities. The percentage uncertainties on  $\Sigma_1$  (i.e. mean of  $100 \times \sigma_{\Sigma_{1,\text{inferred}}}/\Sigma_{1,\text{inferred}}$ ) are 28 per cent, 30 per cent, 30 per cent, and 43 per cent, at  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.75$ . For  $\Sigma_e$ , the corresponding values are 37 per cent, 29 per cent, 31 per cent, and 63 per cent, at  $z = 1.25$ ,  $z = 1.75$ ,  $z = 2.25$ , and  $z = 2.75$ . As we will discuss in Section 3.3, such inclination effects will increase the scatter in observed relations relative to intrinsic ones.

### 3.3 Comparison to observational size–mass relations

In Fig. 2, we show our measurements of the four massive FIRE-2 galaxies on the size–mass plane, at each of the four redshifts studied. We overplot the size–mass relation derived by van der Wel et al. (2014), who also use rest-frame 0.5  $\mu\text{m}$  images. We find that the closest agreement between the massive FIRE-2 galaxies and the observationally derived relation occurs at high redshifts. At  $z = 2.76$ , two of the four haloes are broadly consistent with the late-type galaxy size–mass relation, and two are broadly consistent with the early-type relation. Note that, based on UVJ rest-frame colours, these FIRE-2 galaxies would be classed as star forming at all snapshots studied here. This is expected, since AGN feedback, which is believed to play a role in the quenching of galaxies, is not included in these simulations. At lower redshifts, the agreement worsens. By  $z = 1.25$ , all of the simulated galaxies are significantly offset below the observationally derived van der Wel et al. (2014) relations for both early and late-type galaxies.

One interesting feature of these results is the difference between intrinsic and inferred sizes. As noted in Section 3.2, the inferred sizes are generally within a factor of 2 of those calculated directly from the simulation data. Yet the empirical relations are fairly tight,

and, in some cases, the differences between intrinsic and observed sizes are larger than the scatter in the empirical relations. The morphology/viewing angle of the sources, as quantified by the error bars on each of the data points, contributes to this. The difference between intrinsic and inferred size could have implications for studies of the scatter in scaling relations, in particular for work that attempts to reproduce this scatter in simulations. Our results suggest that proper forward modelling of simulations into observational space is necessary for the scatter in scaling relations of simulated galaxies to be interpreted in a meaningful way.

### 3.4 Comparison to observational surface density–mass relations

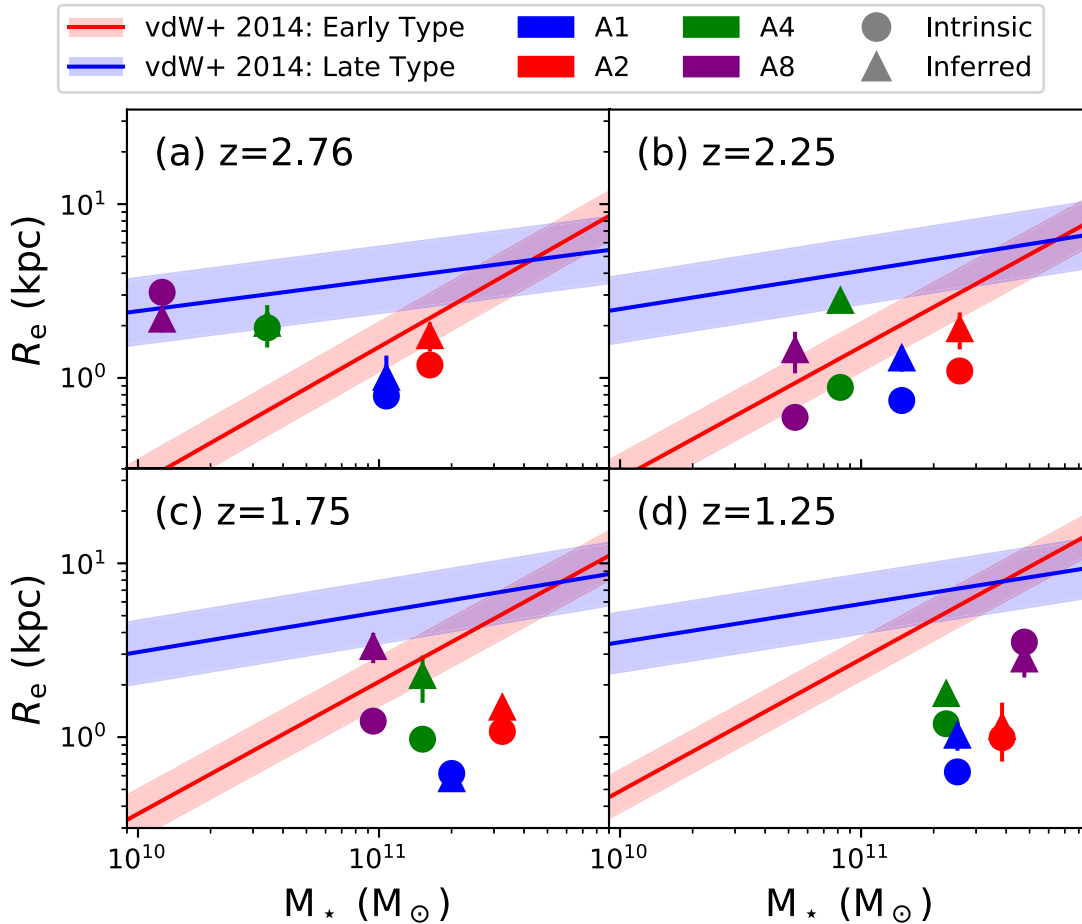
In Figs 3 and 4, we show the inferred stellar mass surface densities for each snapshot, as well as the intrinsic value taken directly from the simulation. Stellar mass surface densities are calculated within the central 1 kpc and  $R_e$  ( $\Sigma_1$  and  $\Sigma_e$ , respectively) for a number of observer inclinations. From Fig. 3, we see that at  $z = 2.76$  and  $z = 2.25$ , the inferred  $\Sigma_1$  shows consistency with the empirically derived relations of Barro et al. (2017) for all four galaxies. This is in line with the reasonable agreement found for the  $R_e - M_*$  relation. The measured 1 kpc surface densities are slightly larger than the intrinsic values. This is due to the overestimation of  $M/L_B$  for haloes A2, A4, and A8 at this redshift. At both  $z = 2.76$  and  $z = 2.25$ , all inferred surface densities remain consistent with one of the empirical relations. By  $z = 1.25$ , the intrinsic and inferred surface densities are too high for their stellar mass, compared to the observational relations, for all but halo A8.

In Fig. 4, we show the same relation, but with  $\Sigma_1$  replaced by  $\Sigma_e$ . We find similar behaviour to the  $R_e - M_*$  relation, as expected given that  $\Sigma_e$  depends on the measurement of  $R_e$ . Galaxies A4 and A8 show consistency with the star-forming  $\Sigma_e - M_*$  relation derived by Barro et al. (2017), and galaxies A1 and A2 lie within  $1\sigma$  of the quiescent relation. The consistency becomes worse with decreasing redshift, with  $R_e$  staying broadly constant at  $\sim 1$  kpc while stellar mass increases. By  $z = 1.25$ , all galaxies apart from A8 are too dense. At  $z = 1.25$ , each halo’s  $\Sigma_e$  is effectively the same as its  $\Sigma_1$ , with  $\Sigma_e$  differing from the empirical relation by a factor of  $\sim 10$ , except for halo A8 which has an  $R_e$  that is closer to the size–mass scaling relation (see Fig. 2d). We will discuss possible reasons for this in Section 4.

## 4 DISCUSSION

We have attempted to derive an observer’s view of the sizes and stellar mass surface densities of massive, intermediate redshift galaxies simulated using FIRE-2 physics. The haloes we study have masses  $M_{\text{halo}} \sim 10^{12.5} M_\odot$  and host central galaxies with stellar masses of  $\sim 10^{11} M_\odot$  at  $z = 2$ . These simulations include recipes for stellar feedback, implemented within a resolved, multiphase ISM. This is unlike many simulations that match observed central densities via implementations of AGN feedback alongside a much simplified, sub-grid ISM model. The unprecedented resolution of the FIRE-2 simulations enables us to probe the limits of stellar feedback in the extreme environments of the inner regions of massive galaxies.

We find that the sizes and surface densities of these simulated massive galaxies are generally within a factor of 2 of the intrinsic values, calculated directly from the simulations. Across all haloes and redshifts, the median offset between the inferred effective radius and the intrinsic half-mass radius, taken directly from the simulation data, is 0.17 dex, with inferred radii generally being slightly larger.



**Figure 2.** The stellar effective radius as a function of stellar mass, for each central galaxy at (a)  $z = 2.76$ , (b)  $z = 2.25$ , (c)  $z = 1.75$ , and (d)  $z = 1.25$ . We show the intrinsic half-mass radius (derived directly from the three-dimensional distribution of stellar mass within the simulation, using spherical shells), as well as the effective radius derived from our mock observations (defined as the semimajor axis of the ellipse that contains half of the total flux of the integrated best-fitting Sérsic model). Error bars are derived using the  $1\sigma$  uncertainty on the measurements using simulated galaxies with five sky orientations. We overplot the  $R_e - M_*$  scaling relations (van der Wel et al. 2014) for both late-type (blue) and early-type (red) galaxies, with shaded regions showing the  $1\sigma$  scatter. Values of  $R_e$  obtained from fits to the  $0.5 \mu\text{m}$  mock FIRE-2 observations fall below the late-type empirical relations for galaxies A1 and A2 at all redshifts. Our predictions for the observed sizes of galaxies A4 and A8 are in agreement with the late-type galaxy size–mass relationship at  $z = 2.75$ , but these galaxies become too compact at lower redshifts.

The standard deviation of the offsets is 0.20 dex. Both values are consistent with the results of Price et al. (2017), who perform similar analysis on FIRE galaxies, but without the detailed radiative transfer modelling that we perform, and find a systematic offset of  $\sim 0.1$  dex and a scatter of  $\sim 0.2$  dex. Across all haloes, redshifts, and inclinations, the median offset between  $\Sigma_{1,\text{observed}}$  and  $\Sigma_{1,\text{intrinsic}}$  is 0.04 dex, with a standard deviation of 0.37 dex. For  $\Sigma_e$ , the median offset is  $-0.17$  dex, and the standard deviation is 0.52 dex. While the median offsets are small, the scatter in the offsets is more substantial. This is a concern when considered along with the tightness of empirical relations such as the size–mass relation. We therefore stress the importance of forward-modelling simulations into observational space, for studies that make comparisons between simulated and observationally inferred scatter in scaling relations.

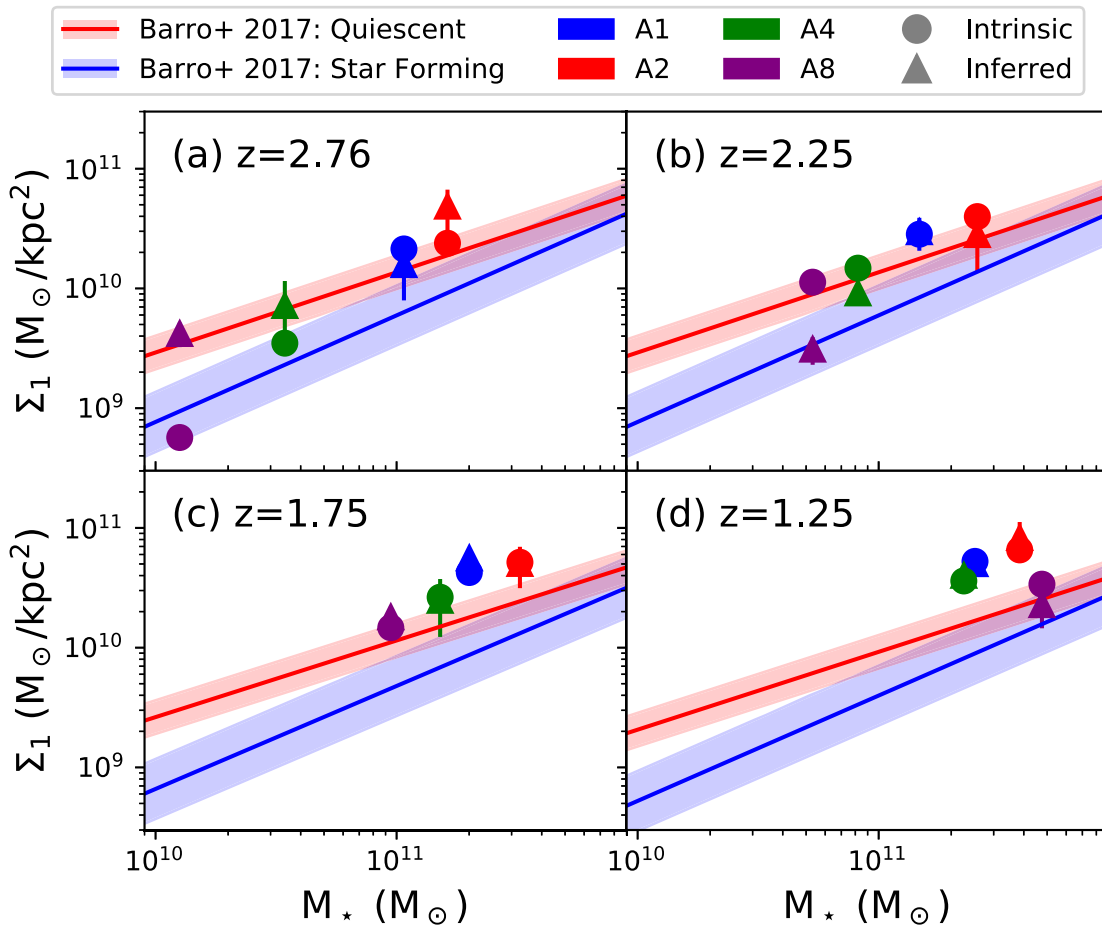
#### 4.1 Comparison to observational relations

Having forward-modelled the simulations into the observational plane, we make comparisons with the observationally derived size–mass relation from van der Wel et al. (2014) and the stellar mass–surface density relations from Barro et al. (2017). The key result

of this paper is that these massive galaxies are, in general, both too small and too dense compared to these empirical relations, with discrepancies increasing towards low redshift. While this is consistent with the study of massive FIRE galaxies performed by Wellons et al. (2020), less massive FIRE galaxies appear to have more realistic sizes (Garrison-Kimmel et al. 2018; Wellons et al. 2020). This could suggest that some piece of physics that is important for massive galaxies is missing from our simulations; possibilities of such additions will be discussed later in this section.

#### 4.2 Uncertainties in observational techniques

Before discussing possible improvements to the FIRE model, one important point is that the observational relations that we compare to are themselves uncertain. Inferring galaxy effective radii from observations can be difficult: both intrinsic uncertainties about the mass-to-light ratio and its constancy or radial dependence across the galaxy, and observational limitations such as the smearing effects of the PSF, limit the robustness of conclusions. Recently, Suess et al. (2019), argued that colour gradients bias the inference of half-mass



**Figure 3.** The stellar mass surface density, calculated within the central 1 kpc, as a function of stellar mass, for each central galaxy at (a)  $z = 2.76$ , (b)  $z = 2.25$ , (c)  $z = 1.75$ , and (d)  $z = 1.25$ . As in Fig. 2, the circles show the intrinsic values, derived directly from the simulation, and the triangles show the values inferred from our mock observations. Error bars are derived using the  $1\sigma$  uncertainty on the measurements using simulated galaxies with five sky orientations. We overplot the empirical  $\Sigma_1$ – $M_*$  scaling relations (Barro et al. 2017), with shaded regions showing the  $1\sigma$  dispersion. At high redshifts, the FIRE-2 galaxies show general agreement with the scaling relation for star-forming galaxies. However, by  $z = 1.75$ , the haloes have begun to diverge from the star-forming scaling relation and by  $z = 1.25$ , they lie a factor of 2 above the empirical relation for quiescent galaxies, with the exception of galaxy A8.

radii from half-light radii, driving the bulk of the apparent evolution of the size–mass relation. These colour gradients are dependent on a number of galaxy properties, including galaxy mass, size, surface density, and colour, and are not trivial to account for in observational studies. Suess et al. (2019) propose that spatially resolved SED modelling (e.g. dividing the galaxy into concentric annuli, which are fitted individually) can enable more robust inference of half-mass sizes from multiband imaging. This approach was adopted by Mosleh et al. (2017) in their study of the evolution of the sizes of star-forming and quiescent galaxies from  $z = 2$  to  $z = 0$ .

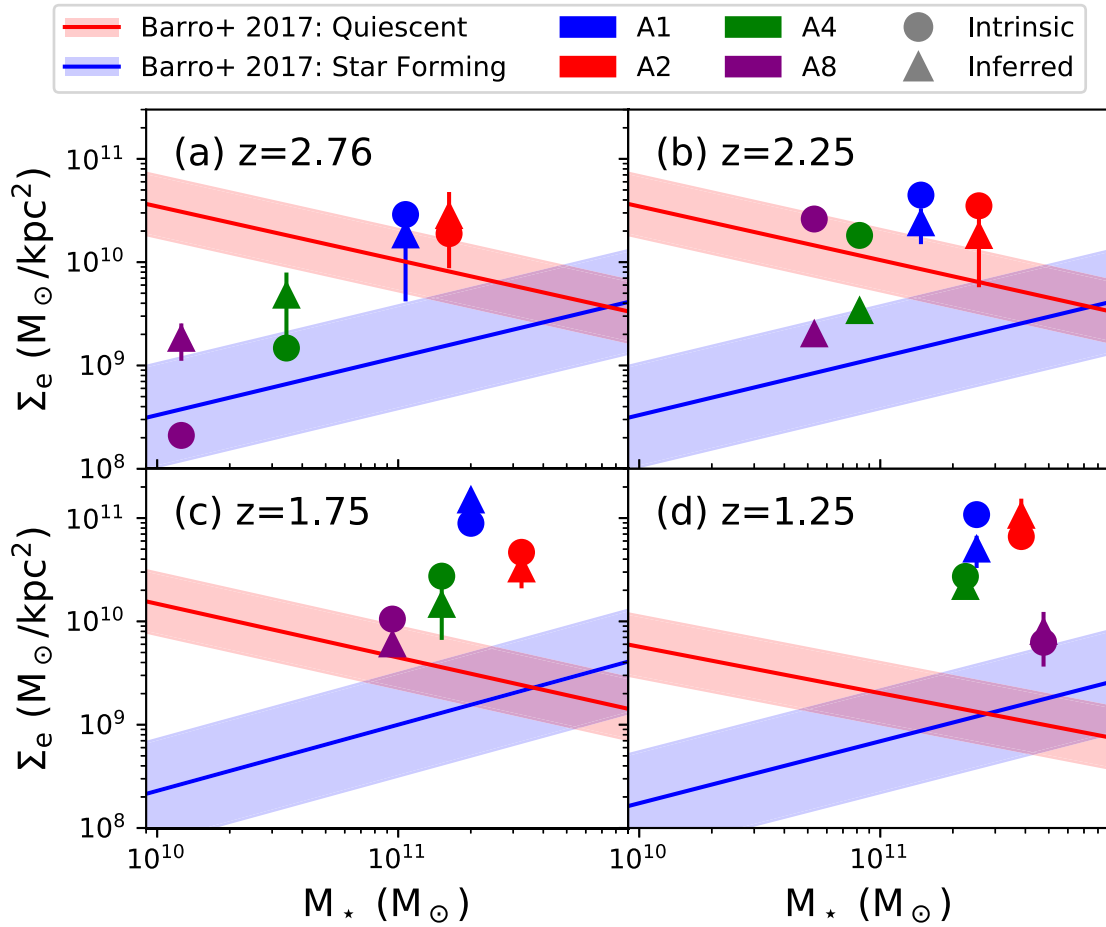
We have attempted to circumvent these observational uncertainties by casting our simulated galaxies into ‘observer space’ and making the same assumptions. Nevertheless, our method could be extended to derive stellar mass and effective radii in a more sophisticated manner. Price et al. (2017) estimate half-mass radii following the approach of Szomoru et al. (2013), which is also tested by Suess et al. (2019). The Szomoru et al. (2013) approach uses rest-frame  $u$ -band and  $g$ -band imaging to constrain possible mass-to-light-ratio gradients and construct colour-based stellar mass profiles. This approach yields half-mass radii that are, on average,  $\sim 25$  per cent smaller than rest-frame  $g$ -band half-light radii. While this detailed analysis is particularly important for galaxies with strong colour gradients, we

show in Appendix B that we are able to recover the intrinsic half-mass radii exceptionally well by simply using effective radii (perhaps because the FIRE-2 galaxies analysed here are broadly disc-like, with shallow colour gradients), and therefore adopt a simpler strategy.

Rather than infer total stellar mass, we opt to use the intrinsic stellar masses calculated directly from the FIRE-2 simulations. Price et al. (2017) derive this quantity by fitting SPS models to the mock photometry using the FAST code (Kriek et al. 2009). They show that stellar masses are recovered extremely well over a wide stellar mass range ( $10^{9.5} < M_*/M_\odot < 10^{11.25}$ ), with a median offset of  $\log_{10}(M_{*,\text{recovered}}/M_{*,\text{intrinsic}}) = -0.06$  dex and a scatter of order 0.1 dex over all projections. Thus, introducing stellar mass fitting into our methodology would likely only increase the scatter in our relations very slightly, and we opt to maintain simplicity in this work.

In this work, we have adopted simple techniques used in the majority of observational studies. Therefore, our results should be similarly susceptible to the biases that affect real observations; in short, if our simulation was well matched to the galaxies in the real Universe, we would expect our results to be wrong in the same way, and therefore match observations. Therefore, the lack of agreement between our synthetic observations and empirical relations strongly implies that there is some physics missing from the simulation. In





**Figure 4.** As Fig. 3, but with the stellar mass surface density calculated using the stellar effective radius  $R_e$ , rather than the central 1 kpc for the ‘inferred’ values and the 3D calculated half-mass radii for the ‘intrinsic’ values. Galaxies A4 and A8 are in agreement with the empirical scaling relation for star-forming galaxies at  $z = 2.76$ , and galaxies A1 and A2 lie on the relation for quiescent galaxies. With the exception of galaxy A8, the galaxies diverge from these relations with time, lying over an order of magnitude above them by  $z = 1.25$ .

the following subsection, we will speculate on where our simulation might be falling short of reality.

### 4.3 Possible physical causes of overcompactness

While the massive simulated galaxies presented in this paper appear to be more compact than observed galaxies of similar stellar mass at the same redshift, one important point to note is that less massive FIRE-2 galaxies do not suffer the same overcompactness (El-Badry et al. 2016). One likely reason for the overcompactness of the massive FIRE-2 galaxies is the lack of AGN feedback in our simulations. We know from observations that AGNs exert feedback on their host galaxies. It is seen directly via radio jets, observable in their strong radio synchrotron emission, and via X-ray bubbles and cavities (see the review by Fabian 2012). Recent years have also seen increasing amounts of direct observational evidence of ‘quasar mode’ feedback, including observations of high-velocity galactic outflows that cannot be attributed to starburst events (see e.g. Sturm et al. 2011; Rupke & Veilleux 2011; Cicone et al. 2014; Fiore et al. 2017). These outflows and their observational signatures have been modelled analytically and in idealized simulations (e.g. Faucher-Giguère & Quataert 2012; Zubovas & King 2012; Costa, Sijacki & Haehnelt 2014; Nims, Quataert & Faucher-Giguère 2015; Richings & Faucher-Giguère 2018a, b). Motivated by this, and by the need to explain a number of

empirical results including the sharp break in the stellar mass function at high masses and the quenching of massive galaxies, many galaxy formation simulations now include some form of AGN feedback (e.g. Springel, Di Matteo & Hernquist 2005a, b; Dubois et al. 2014, 2016; Hirschmann et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Weinberger et al. 2017, 2018; Davé et al. 2019).

A number of recent studies have shown that AGN feedback has an impact on galaxy sizes. As discussed in the introduction, Choi et al. (2018) perform two sets of cosmological hydrodynamical simulations, one without black holes or AGN feedback (no-AGN runs), and one with AGN feedback in the form of winds and X-ray radiation. The galaxies simulated with AGNs have larger half-mass radii at fixed stellar mass. In their simulations, AGN feedback quenches star formation, transforming compact blue galaxies into compact red ones. These quiescent galaxies have lower gas content than their star-forming counterparts in the no-AGN simulations, and subsequently undergo gas-poor mergers that lead to extended stellar envelopes. In addition, fast AGN-driven winds can ‘puff up’ the central region of a galaxy. Differences between the sizes of galaxies in the two simulations become apparent around  $z = 2$ , when *in situ* star formation becomes quenched. From around this time, galaxies with AGNs evolve more steeply in the mass–size plane than those without AGNs. By  $z = 1$ , around half of the galaxies with AGNs have become quenched, while those without AGNs remain star



forming. The quenched galaxies are clearly separated from the star-forming galaxies in the size–mass plane. Similar results are found by Dubois et al. (2016), who study lower-stellar-mass galaxies. They show that galaxies simulated with AGNs (both heating and jet mode feedback) display larger sizes than their no-AGN counterparts above  $M_* \sim 10^{10} M_\odot$ , with the differences increasing with stellar mass for both star-forming and quiescent galaxies, and order-of-magnitude differences by  $z = 0$ .

The galaxies with AGN feedback simulated by Choi et al. (2018) also show lower  $\Sigma_1$  values, with an offset of  $\sim 0.3$  dex from the no-AGN runs below  $z = 1$ , driven by gas and stellar mass-loss. This is due to high gas accumulation within the central region, with subsequent formation of dense stellar cores. Note, however, that their simulated quenched galaxies do still lie above observationally derived stellar mass–surface density relations. Dubois et al. (2016) show consistent results. No-AGN galaxies display cuspy centres, whereas massive galaxies with AGN are cored, with flatter central stellar mass densities and a less significant *in situ* stellar mass component.

These results suggest that the lack of AGN feedback within the FIRE-2 simulations could be one reason for the compact sizes and overdense cores of our galaxies. Our galaxies occupy similar parameter space in the size–mass and density–mass plane to those simulated by Choi et al. (2018) without AGN feedback at  $z \sim 1$  (their galaxies are well matched to ours, also having  $M_* \sim 10^{11} M_\odot$  at  $z = 2$ ). According to their results, our massive FIRE-2 galaxies should be quenched by around  $z = 1$ , rather than continuing to form stars as they do in our simulations. Encouragingly, Anglés-Alcázar et al. (2017b) showed that black holes transition to a rapid growth phase when the central stellar potential deepens and star formation becomes less bursty (see also Bower et al. 2017; Byrne et al. in preparation). This happens roughly at the time that galaxies exceed  $M_* \sim$  a few times  $10^{10} M_\odot$  and may correspond to the virialization of the inner CGM (Stern et al. 2020). At this stage, an additional source of feedback is required to regulate central densities. Future simulations should address the detailed balance between the higher central densities required for efficient black hole growth and the role of black hole feedback in suppressing central densities.

## 5 CONCLUSIONS

In this paper, we have explored the sizes and surface densities of simulated massive galaxies drawn from the FIRE-2 zoom-in simulations (Anglés-Alcázar et al. 2017b), which include black hole accretion but not AGN feedback. These simulations model various stellar feedback processes directly within a multiphase ISM. Thus, the sizes and surface densities of the simulated galaxies can be used to test the efficacy of the feedback model. We focus on the redshift range  $1 < z < 3$ , where stellar mass in the Universe is assembling most rapidly. We have modelled the observable sizes of four massive ( $M_* \sim 10^{11} M_\odot$  at  $z \sim 2$ ) galaxies, using radiative transfer techniques to include the reddening effects of a realistic dust distribution. We then convolved our images with typical filter profiles and an HST-like PSF, to create mock observations. From these mock observations, we attempted to derive physical properties, mirroring the attempts of observational studies. We base our stellar mass surface density measurements on well-established observational techniques, which convert an observed colour (in our case,  $B - R$ ) to a mass-to-light ratio (Bell & de Jong 2001). Sizes are derived using a popular Sérsic profile fitting package, which can successfully reconstruct surface brightness profiles. Our estimates of galaxy sizes and surface

densities are generally within a factor of 2 of the intrinsic quantities, which are inferred directly from the simulations.

With the goal of understanding the limitations of our AGN-free simulation, we have compared the inferred sizes of massive FIRE-2 galaxies to the empirical scaling relations derived by van der Wel et al. (2014) and Barro et al. (2017). While the simulated massive galaxies are relatively consistent with empirical size–mass (van der Wel et al. 2014) and surface density–mass scaling relations (Barro et al. 2017) at  $z \gtrsim 2$ , they significantly diverge from both relations by  $z = 1.25$ . Below  $z = 2$ , the simulated galaxies are too compact compared to observed galaxies at the same redshift, by up to a factor of 10. The simulated galaxies also become too dense towards low redshifts, with mass surface densities lying well above empirical relations. The most extreme offsets are seen for  $\Sigma_e$  (rather than  $\Sigma_1$ ), due to the added effects of the very small derived  $R_e$  values. Neither of these offsets can be attributed to purely observational effects, such as sky orientation.

The underpredicted sizes and stellar mass surface densities at  $z < 2$ , combined with the fact that less massive FIRE-2 galaxies have been shown to reproduce observationally expected sizes (El-Badry et al. 2016), indicate that there is some physics missing from these simulated massive galaxies. AGN feedback is expected to play a role in the sizes and star-formation rates of massive galaxies like these, and could be responsible for the discrepancies with observations. We will explore this possibility further with a new suite of simulations that include AGN feedback (Wellons et al., in preparation).

## ACKNOWLEDGEMENTS

We thank the anonymous reviewer for detailed suggestions that helped improve the content and clarity of the paper.

This work was initiated as a project for the Kavli Summer Program in Astrophysics held at the Center for Computational Astrophysics of the Flatiron Institute in 2018. The programme was co-funded by the Kavli Foundation and the Simons Foundation. We thank them for their generous support. TP would like to thank all of the participants of the summer programme including Marius Ramsøy, Ulrich Steinwandel, and Daisy Leung for stimulating discussion.

TP acknowledges funding from the Future Investigators in NASA Earth and Space Science and Technology (FINESST) Fellowship, NASA grant 80NSSC19K1610. RKC acknowledges funding from the John Harvard Distinguished Science Fellowship and thanks Sandro Tacchella for helpful discussions. The Flatiron Institute is supported by the Simons Foundation. DAA was supported in part by NSF grant AST-2009687. RF acknowledges financial support from the Swiss National Science Foundation (grant number 157591). CAFG was supported by the NSF through grants AST-1517491, AST-1715216 and CAREER award AST-1652522, by NASA through grant 17-ATP17-0067, by STScI through grant *HST*-AR-14562.001, and by a Cottrell Scholar Award from the Research Corporation for Science Advancement. SW was supported by the CIERA Postdoctoral Fellowship Program (Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University) and by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2001905.

The simulations were run using XSEDE (TG-AST160048, TG-AST140023), supported by NSF grant ACI-1053575, and Northwestern University’s compute cluster ‘Quest’, as well as using NASA HEC allocations SMD-16-7561 and SMD-17-1204. Support for PFH was provided by NSF Collaborative Research Grants 1715847 and 1911233, NSF CAREER grant 1455342, NASA grants 80NSSC18K0562, JPL 1589742. Numerical calculations were run

on the Caltech compute cluster ‘Wheeler,’ allocations FTA-Hopkins supported by the NSF and TACC, and NASA HEC SMD-16-7592. The data used in this work were, in part, hosted on facilities supported by the Scientific Computing Core at the Flatiron Institute. This research has made use of the SVO Filter Profile Service, supported from the Spanish MINECO through grant AyA2014-55216.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. Additional data including simulation snapshots, initial conditions, and derived data products are available at <https://fire.northwestern.edu/data/>.

## REFERENCES

- Allen R. J. et al., 2017, *ApJ*, 834, L11
- Anglés-Alcázar D., Özel F., Davé R., 2013, *ApJ*, 770, 5
- Anglés-Alcázar D., Özel F., Davé R., Katz N., Kollmeier J. A., Oppenheimer B. D., 2015, *ApJ*, 800, 127
- Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017a, *MNRAS*, 470, 4698
- Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017b, *MNRAS*, 472, L109
- Appleby S., Davé R., Kraljic K., Anglés-Alcázar D., Narayanan D., 2020, *MNRAS*, 494, 6053
- Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stavevski M., Valcke S., 2011, *ApJS*, 2, 22
- Barden M. et al., 2005, *ApJ*, 635, 959
- Barro G. et al., 2017, *ApJ*, 840, 47
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Bell E. F., de Jong R. S., 2001, *ApJ*, 550, 212
- Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, *MNRAS*, 465, 32
- Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, *ApJ*, 687, L61
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Camps P., Baes M., 2015, *Astron. Comput.*, 9, 20
- Camps P. et al., 2018, *ApJS*, 234, 20
- Carnall A., McLure R., Dunlop J., Davé R., 2018, *MNRAS*, 480, 4379
- Carollo C. M. et al., 2013, *ApJ*, 773, 112
- Cheung E. et al., 2012, *ApJ*, 760, 131
- Choi E., Somerville R. S., Ostriker J. P., Naab T., Hirschmann M., 2018, *ApJ*, 866, 91
- Cicone C. et al., 2014, *A&A*, 562, A21
- Cochrane R. K., Best P. N., Sobral D., Smail I., Geach J. E., Stott J. P., Wake D. A., 2018, *MNRAS*, 475, 3730
- Cochrane R. K. et al., 2019, *MNRAS*, 488, 1779
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Conselice C. J., 2014, *ARA&A*, 52, 291
- Costa T., Sijacki D., Haehnelt M. G., 2014, *MNRAS*, 444, 2355
- Daddi E. et al., 2005, *ApJ*, 631, L13
- Dale J. E., 2015, *New Astron. Rev.*, 68, 1
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davidzon I. et al., 2017, *A&A*, 605, A70
- Draine B. T., Salpeter E. E., 1979, *ApJ*, 231, 77
- Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, *MNRAS*, 433, 3297
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, *MNRAS*, 463, 3948
- Dullemond C. P., Juhasz A., Pohl A., Sereshfi F., Shetty R., Peters T., Commercon B., Flock M., 2012, *Astrophysics Source Code Library*, record ascl:1202.015
- Dwek E., 1998, *ApJ*, 508, 643
- El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, *ApJ*, 820, 131
- Fabian A., 2012, *ARA&A*, 50, 455
- Fan L., Lapi A., Zotti G. D., Danese L., 2008, *ApJ*, 689, L101
- Faucher-Giguère C.-A., Quataert E., 2012, *MNRAS*, 425, 605
- Feldmann R., Hopkins P. F., Quataert E., Faucher-Giguère C. A., Kereš D., 2016, *MNRAS*, 458, L14
- Feldmann R., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2017, *MNRAS*, 470, 1050
- Ferguson H. C. et al., 2004, *ApJ*, 600, L107
- Fiore F. et al., 2017, *A&A*, 601, A143
- Franx M., van Dokkum P. G., Förster Schreiber N. M., Wuyts S., Labbé I., Toft S., 2008, *ApJ*, 688, 770
- Gaburov E., Nitadori K., 2011, *MNRAS*, 414, 129
- Garrison-Kimmel S. et al., 2018, *MNRAS*, 481, 4133
- Genel S. et al., 2018, *MNRAS*, 474, 3976
- Grogin N. A. et al., 2011, *ApJS*, 197, 35
- Hafen Z. et al., 2019, *MNRAS*, 488, 1248
- Hayward C. C., Smith D. J. B., 2015, *MNRAS*, 446, 1512
- Hayward C. C. et al., 2014, *MNRAS*, 445, 1598
- Hill A. R. et al., 2017, *ApJ*, 837, 147
- Hill A. R., van der Wel A., Franx M., Muzzin A., Skelton R. E., Momcheva I., van Dokkum P., Whitaker K. E., 2019, *ApJ*, 871, 76
- Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Hopkins P. F., Quataert E., Murray N., 2011, *MNRAS*, 417, 950
- Hopkins P. F., Narayanan D., Murray N., 2013, *MNRAS*, 432, 2647
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C. A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Hopkins P. F. et al., 2018a, *MNRAS*, 477, 1578
- Hopkins P. F. et al., 2018b, *MNRAS*, 480, 800
- Huertas-Company M. et al., 2015, *ApJ*, 809, 95
- Ishibashi W., Fabian A. C., Canning R. E. A., 2013, *MNRAS*, 431, 2350
- James A., Dunne L., Eales S., Edmunds M. G., 2002, *MNRAS*, 335, 753
- Jonsson P., 2006, *MNRAS*, 372, 2
- Jonsson P., Groves B. A., Cox T. J., 2010, *MNRAS*, 403, 17
- Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
- Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, *ApJ*, 700, 221
- Kroupa P., 2001, *MNRAS*, 322, 231
- Krumholz M. R., Gnedin N. Y., 2011, *ApJ*, 729, 36
- Lange R. et al., 2015, *MNRAS*, 447, 2603
- Leitherer C. et al., 1999, *ApJS*, 123, 3
- Liang L., Feldmann R., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Hayward C. C., Quataert E., Scoville N. Z., 2018, *MNRAS*, 478, L83
- Liang L. et al., 2019, *MNRAS*, 489, 1397
- Ma X., Hopkins P. F., Faucher-Giguère C. A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, *MNRAS*, 456, 2140
- Ma X. et al., 2018, *MNRAS*, 478, 1694
- Ma X. et al., 2019, *MNRAS*, 487, 1844
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Mosleh M., Tacchella S., Renzini A., Carollo C. M., Molaeinezhad A., Onodera M., Khosroshahi H. G., Lilly S., 2017, *ApJ*, 837, 2
- Moster B. P., Somerville R. S., Maulbetsch C., Bosch F. C. V. D., Macci A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Muratov A. L. et al., 2017, *MNRAS*, 468, 4170
- Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, *ApJ*, 746, 162
- Nims J., Quataert E., Faucher-Giguère C.-A., 2015, *MNRAS*, 447, 3612
- Oppenheimer B. D., Davé R., 2006, *MNRAS*, 373, 1265
- Orr M. E. et al., 2018, *MNRAS*, 478, 3653
- Price S. H., Kriek M., Feldmann R., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., Barro G., 2017, *ApJ*, 844, L6

Richings A. J., Faucher-Giguère C.-A., 2018a, *MNRAS*, 474, 3673  
 Richings A. J., Faucher-Giguère C.-A., 2018b, *MNRAS*, 478, 3100  
 Rodriguez-Gomez V. et al., 2019, *MNRAS*, 483, 4140  
 Rupke D. S. N., Veilleux S., 2011, *ApJ*, 729, L27  
 Schaye J. et al., 2015, *MNRAS*, 446, 521  
 Sérsic J. L., 1968, *Atlas de Galaxias Australes*  
 Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, 343, 978  
 Skelton R. E. et al., 2014, *ApJS*, 214, 24  
 Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51  
 Sparre M., Hayward C. C., Feldmann R., Faucher-Giguère C. A., Muratov A. L., Kereš D., Hopkins P. F., 2017, *MNRAS*, 466, 88  
 Spilker J. S., Bezanson R., Weiner B. J., Whitaker K. E., Williams C. C., 2019, *ApJ*, 883, 81  
 Springel V., Di Matteo T., Hernquist L., 2005a, *MNRAS*, 361, 776  
 Springel V., Di Matteo T., Hernquist L., 2005b, *ApJ*, 620, L79  
 Steinacker J., Baes M., Gordon K., 2013, *ARA&A*, 51, 63  
 Stern J. et al., 2020, preprint (arXiv:2006.13976)  
 Stott J. P., Collins C. A., Burke C., Hamilton-Morris V., Smith G. P., 2011, *MNRAS*, 414, 445  
 Sturm E. et al., 2011, *ApJ*, 733, L16  
 Suess K. A., Kriek M., Price S. H., Barro G., 2019, *ApJ*, 885, L22  
 Szomoru D., Franx M., van Dokkum P. G., Trenti M., Illingworth G. D., Labbé I., Oesch P., 2013, *ApJ*, 763, 73  
 Tacchella S. et al., 2015, *ApJ*, 802, 101  
 Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack J. R., 2016, *MNRAS*, 458, 242  
 Tacchella S. et al., 2018, *ApJ*, 859, 56  
 Tielens A. G. G. M., McKee C. F., Seab C. G., Hollenbach D. J., 1994, *ApJ*, 431, 321  
 Toft S., Franx M., van Dokkum P., Schreiber N. M. F., Labbé I., Wuyts S., Marchesini D., 2009, *ApJ*, 705, 255  
 Trayford J. W. et al., 2017, *MNRAS*, 470, 771  
 Trujillo I. et al., 2004, *ApJ*, 604, 521  
 van der Vlugt D., Costa T., 2019, *MNRAS*, 490, 4918  
 van Dokkum P. G. et al., 2010, *ApJ*, 709, 1018  
 van Dokkum P. G. et al., 2015, *ApJ*, 813, 23  
 van de Voort F., Quataert E., Hopkins P. F., Kereš D., Faucher-Giguère C. A., 2015, *MNRAS*, 447, 140  
 van der Wel A. et al., 2014, *ApJ*, 788, 28  
 Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518  
 Wang E., Lilly S. J., Pezzulli G., Matthee J., 2019, *ApJ*, 877, 132  
 Weinberger R. et al., 2017, *MNRAS*, 465, 3291  
 Weinberger R. et al., 2018, *MNRAS*, 479, 4056  
 Weingartner J. C., Draine B. T., 2001, *ApJ*, 548, 296  
 Wellons S. et al., 2015, *MNRAS*, 449, 361  
 Wellons S., Faucher-Giguère C.-A., Anglés-Alcázar D., Hayward C. C., Feldmann R., Hopkins P. F., Kereš D., 2020, *MNRAS*, 497, 4051  
 Whitaker K. E. et al., 2017, *ApJ*, 838, 19  
 Whitney A., Conselice C. J., Bhatwadekar R., Duncan K., 2019, *ApJ*, 887, 113  
 Williams R. J., Quadri R. F., Franx M., van Dokkum P., Toft S., Kriek M., Labbé I., 2010, *ApJ*, 713, 738  
 Wu P.-F. et al., 2018, *ApJ*, 868, 37  
 Wu P.-F. et al., 2020, *ApJ*, 888, 77  
 Wuyts S. et al., 2012, *ApJ*, 753, 114  
 Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2019, *MNRAS*, 487, 5649  
 Zubovas K., King A., 2012, *ApJ*, 745, L34

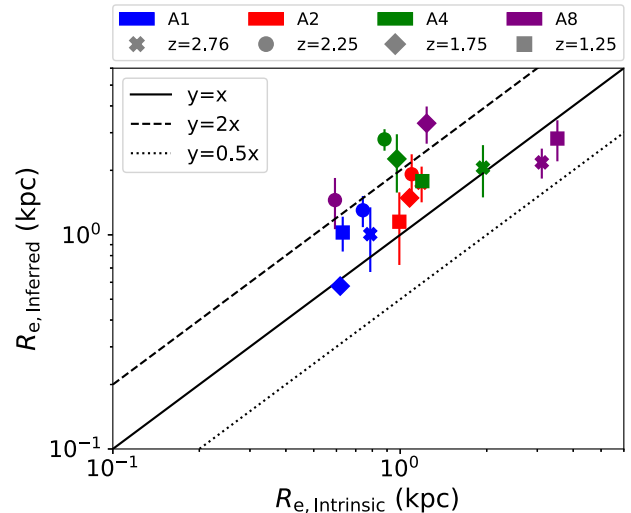
## APPENDIX A: MASS-TO-LIGHT RATIO VERSUS ( $B - R$ ) COLOUR

In Section 3.1, we describe the process of inferring stellar mass surface densities from our mock observations. In order to do this, we re-derived the Bell & de Jong (2001) relation (to update it to the AB magnitude system) using the `fspy` code (Conroy, Gunn & White 2009; Conroy & Gunn 2010), with a Kroupa (2001) IMF. We computed  $M/L_B$  for a modelled stellar population as a function of its  $B - R$  colour, for various e-folding time-scales. Following Bell & de Jong (2001), we acquired the  $M/L_B$  and  $B - R$  colours of each `fspy` track for each e-folding time-scale. The  $M/L_B$  and  $B - R$  colours of interest correspond to the period of time in which the synthetic stellar population is 12 Gyr in age. We fitted these points and derive the relationship:

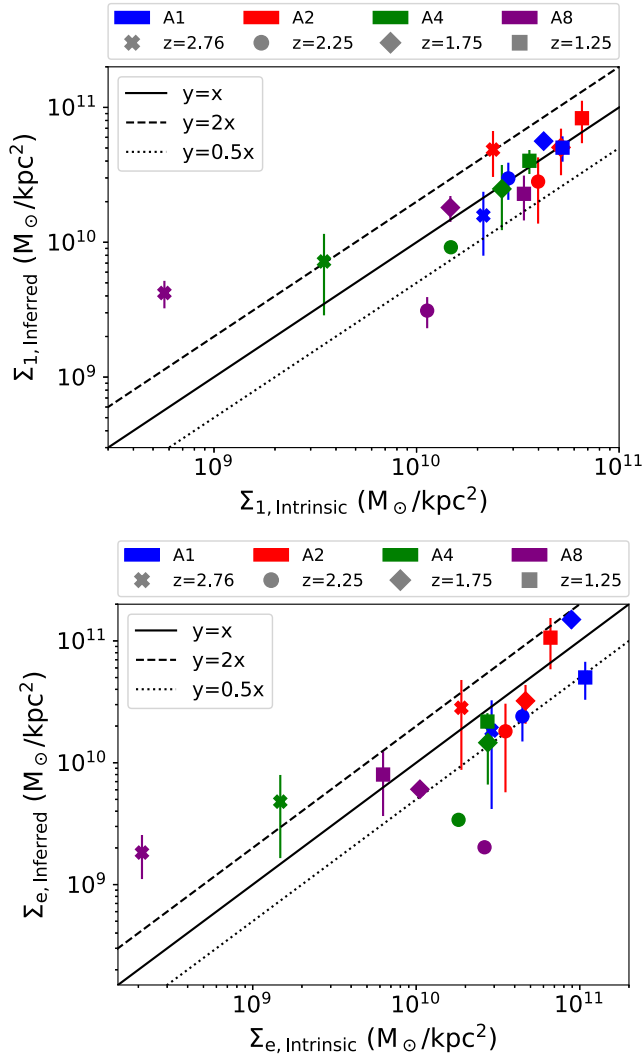
$$\log(M/L_B) = 1.04(B - R) - 0.29. \quad (\text{A1})$$

This new relationship is similar to that derived by Bell & de Jong (2001), with a change in the overall normalization due to the updated SSP models and the use of the AB magnitude system. We use this relation to infer mass-to-light ratios for our FIRE-2 galaxies from the  $B - R$  colours output by SKIRT. We are able to recover intrinsic mass-to-light ratios (using the stellar mass direct from the simulations) very well using this method (typically to within a factor of 2).

## APPENDIX B: COMPARISON OF INTRINSIC AND DERIVED SIZES AND SURFACE DENSITIES



**Figure B1.** The effective radii inferred from Sérsic fits to our synthetic images, against the intrinsic half-mass radii, measured directly from the simulations. The error bars on our observed sizes are the  $1\sigma$  uncertainties calculated using five different sky orientations. The solid black line shows the 1–1 relation, and the dashed/dotted lines show a factor of 2 offset from this relation. Inferred effective radii tend to be slightly larger than the intrinsic half-mass radii, but the majority of our estimates recover the intrinsic size to within a factor of 2.



**Figure B2.** The stellar mass surface density within the central 1 kpc (upper panel) and  $R_e$  (lower panel), inferred from our synthetic images, against the same quantities measured directly from the simulations. The error bars on our inferred surface densities are the  $1\sigma$  uncertainties calculated using five different sky orientations. The solid black line shows the 1–1 relation, and the dashed/dotted lines show a factor of 2 offset from this relation. The inferred stellar mass surface densities are, on average, lower than the intrinsic values, but the two values tend to agree within a factor of 2. This overall agreement is in part due to the offsets in M/L ratio and radius coincidentally cancelling one another out.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.