The origin of ultra diffuse galaxies: stellar feedback and quenching

T. K. Chan,[1]★ D. Kereš,[1] A. Wetzel,[2,3,4]† P. F. Hopkins,[2] C.-A. Faucher-Giguère,[5] K. El-Badry,[6] S. Garrison-Kimmel[2] and M. Boylan-Kolchin[7]

[1]*Department of Physics, Center for Astrophysics and Space Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[2]*TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[3]*The Observatories of the Carnegie Institution for Science, Pasadena, CA 91101, USA*
[4]*Department of Physics, University of California, Davis, CA 95616, USA*
[5]*Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[6]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA*
[7]*Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712-1205, USA*

ABSTRACT

We test if the cosmological zoom-in simulations of isolated galaxies from the FIRE project reproduce the properties of ultra diffuse galaxies (UDGs). We show that outflows that dynamically heat galactic stars, together with a passively aging stellar population after imposed quenching, naturally reproduce the observed population of red UDGs, without the need for high spin haloes, or dynamical influence from their host cluster. We reproduce the range of surface brightness, radius, and absolute magnitude of the observed red UDGs by quenching simulated galaxies at a range of different times. They represent a mostly uniform population of dark matter-dominated dwarf galaxies with $M_* \sim 10^8 \, M_\odot$, low metallicity, and a broad range of ages; the more massive the UDGs, the older they are. The most massive red UDG in our sample ($M_* \sim 3 \times 10^8 \, M_\odot$) requires quenching at $z \sim 3$ when its halo reached $M_h \sim 10^{11} \, M_\odot$. Our simulated UDGs form with normal stellar-to-halo ratios and match the central enclosed masses and the velocity dispersions of the observed UDGs. Enclosed masses remain largely fixed across a broad range of quenching times because the central regions of their dark matter haloes complete their growth early. If our simulated dwarfs are not quenched, they evolve into bluer low surface brightness galaxies with M/L similar to observed field dwarfs. While our simulation sample covers a limited range of formation histories and halo masses, we predict that UDG is a common, and perhaps even dominant, galaxy type around $M_* \sim 10^8 \, M_\odot$, both in the field and in clusters.

Key words: galaxies: evolution – galaxies: haloes – galaxies: kinematics and dynamics – galaxies: structure – dark matter.

1 INTRODUCTION

Low surface brightness galaxies (LSBs) with large effective radii were detected and studied by many authors over the past several decades (Impey, Bothun & Malin 1988; Bothun, Impey & Malin 1991; Dalcanton et al. 1997; Caldwell 2006; McConnachie et al. 2008). van Dokkum et al. (2015a) sparked a recent interest in LSBs by finding many LSBs in the Coma cluster, named ultra diffuse galaxies (UDGs), with absolute magnitudes comparable to those of dwarf galaxies ($M_g \sim -14$), but with effective radii as large as the Milky Way (MW) ($\sim 4 \, kpc$) and surface brightnesses of

$\sim 25 \, mag \, arcsec^{-2}$. They appear spheroidal and red, indicating old stellar populations. Since then, a large number of UDGs have been discovered in the Coma cluster (Koda et al. 2015), the Virgo cluster (Caldwell 2006; Mihos et al. 2015), the Fornax cluster (Muñoz et al. 2015), clusters with $z \sim 0.044-0.063$ (van der Burg, Muzzin & Hoekstra 2016), the Abell 2744 cluster (Janssens et al. 2017), the Abell S1063 clusters (Lee et al. 2017), the Pisces–Perseus supercluster (Martínez-Delgado et al. 2016), the M77 group (Trujillo et al. 2017), the elliptical galaxy NGC 5485 (Merritt et al. 2016), three nearby isolated groups (Román & Trujillo 2017b) and the Hickson Compact Group 95 (Shi et al. 2017).

Because of their large effective radii and low inferred stellar masses, van Dokkum et al. (2015a) proposed that UDGs are 'failed' L⋆ galaxies initially forming in relatively massive haloes that were quenched at $z \sim 2$. This hypothesis is supported by the stellar

velocity dispersion and the number of globular clusters (GCs) of a massive UDG, Dragonfly 44, in the Coma cluster for which van Dokkum et al. (2016, 2017) inferred a total halo mass $\sim 10^{12}$ M$_\odot$.

However, from recent observations of the GC systems of other UDGs, Beasley et al. (2016); Beasley & Trujillo (2016); Peng & Lim (2016) argued that the UDGs are 'failed' dwarf galaxies. By measuring the velocity dispersion of the GC system in a UDG (VCC 1287 in Virgo Cluster), Beasley et al. (2016) inferred a dynamical mass of 4.5×10^9 M$_\odot$ within 8.1 kpc. By comparing its dynamical mass with numerical simulations, they estimated its halo mass $M_{200} = (8 \pm 4) \times 10^{10}$ M$_\odot$, comparable to a dwarf galaxy halo. By measuring the number of GCs in Coma UDGs and assuming the correlation between GC number and halo mass, Amorisco et al. (2018) found most of the Coma UDGs reside in dwarf haloes. Similar conclusions were reached with measurements of GC specific frequencies of UDGs (Beasley et al. 2016; Beasley & Trujillo 2016; Peng & Lim 2016). Furthermore, Román & Trujillo (2017a) revealed that the spatial distribution of UDGs in a galaxy cluster resembles the distribution of dwarf galaxies rather than L⋆ galaxies. Based on these measurements, Beasley & Trujillo (2016) argued that UDGs are quenched galaxies that inhabit Large Magellanic Cloud (LMC)-sized haloes and quench their star formation at $z \sim 3$. In this scenario, cluster UDGs have to be quenched for more than 10 Gyr. Sifón et al. (2018) used weak gravitational lensing to show that the average virial mass of 784 UDGs in 18 clusters is $m_{200} \leq 10^{11.80}$ M$_\odot$, consistent with dwarf halo masses but leaving a possibility of the most massive UDGs to be hosted in MW-mass haloes.

Yozin & Bekki (2015) similarly argued that UDGs have dwarf progenitors, but they quenched at much later times. They simulated interactions between a cluster and an infalling diffuse dwarf galaxy at $z \sim 2$, and showed that the harsh cluster environment can rapidly halt any ongoing star formation. Their initial conditions assumed the infalling dwarf was hosted in a high spin halo, allowing the galaxy to be much more diffuse than normal galaxies even before interacting with the host cluster. Following this line of thought, Amorisco & Loeb (2016) proposed that UDGs are the high spin tail of the dwarf galaxy population, so they are diffuse even without interacting with the cluster. They predicted there should also be a field population of UDGs but with possibly different morphologies and colours. Rong et al. (2017) supported this hypothesis by finding that UDGs in their cosmological simulations reside in high spin haloes.

Upon finding UDGs with bluer colour far from clusters, Román & Trujillo (2017b) and Trujillo et al. (2017) suggested that red UDGs in clusters might be initially LSB diffuse dwarf galaxies born in the field that are later processed in groups and ultimately accreted into galaxy clusters.

In this paper, we use cosmological zoom-in simulations from the FIRE simulation to study the effects of stellar feedback and quenching on the progenitors of UDGs. Stellar feedback is known to shape dark matter (DM) profiles, creating large cores in the DM distribution of dwarf galaxy haloes (Navarro, Eke & Frenk 1996; El-Zant, Shlosman & Hoffman 2001; Gnedin et al. 2004; Read & Gilmore 2005; Governato et al. 2010; Governato et al. 2012; Macciò et al. 2012; Peñarrubia et al. 2012; Pontzen & Governato 2012; Teyssier et al. 2013; Di Cintio et al. 2014; Pontzen & Governato 2014; Chan et al. 2015; Tollet et al. 2016). Feedback can also drive significant radial migrations of stars via two processes: (1) inflowing/outgoing gas clouds can form stars that inherit the velocities of the gas clouds and continue migrating within their first 100 Myr and (2) feedback-driven gas outflows modify central gravitational potential and trans-

fer energy to stars non-adiabatically (in the same manner as in DM core creation; see El-Badry et al. 2016). Through these processes, feedback expands galaxies into diffuse spheroids, producing large effective radii (El-Badry et al. 2016) and large axial ratios (Wheeler et al. 2017) simultaneously. Effects of stellar feedback on both DM and stellar distributions peak at $M_* \sim 10^8$ M$_\odot$, which is also a typical mass of the observed UDGs.

Using cosmological simulations of isolated galaxies with stellar feedback, Di Cintio et al. (2017) recently also showed that feedback can produce extended stellar profiles similar to observed UDGs. Our study differs both in the stellar feedback model and in the inclusion of the effect of quenching, which has significant effects on the formation of red UDGs (see Section 4.4 for a comparison of their findings with our work).

In Section 2, we describe the simulation methodology, the suite of simulations used in this paper and the method for mock observations with GALFIT. In Section 3, we show how radius, surface brightness and other properties of simulated galaxies change with quenching time. In Section 4, we discuss the structural properties of our dwarfs, the connections to the formation scenarios discussed in the literature as well as the implications for the properties of field dwarf galaxies. Finally, we present our conclusions in Section 5.

2 METHOD

2.1 Simulation code and setup

Our simulations utilize the GIZMO[1] code (Hopkins 2015) in the mesh-free Lagrangian finite mass (MFM) mode for hydrodynamics. GIZMO uses an updated version of the PM+Tree algorithm from GADGET-3 (Springel 2005) to calculate gravity and adopts fully conservative adaptive gravitational softening for gas (Price & Monaghan 2007). We employ the zoom-in technique to reach high resolutions in a cosmological environment and evolve simulations from $z = 99$ to $z = 0$.

Gas cooling is calculated with the tabulated cooling rates from CLOUDY (Ferland et al. 2013) for $T = 10-10^{10}$ K, including atomic, metal-line, and molecular cooling. We apply the redshift-dependent ultraviolet background model from Faucher-Giguère et al. (2009) that ionizes and heats gas in an optically thin approximation and use an approximate prescription to account for self-shielding of dense gas.

Star formation and stellar feedback are implemented using the FIRE-2 algorithm (Hopkins et al. 2017), which is an updated version of the FIRE feedback scheme from Hopkins et al. (2014). Briefly, stars form in self-gravitating molecular gas at densities $n_H \geq 1000$ cm^{-3}, with 100 per cent instantaneous efficiency per free fall time. Stellar feedback physics implemented includes stellar winds, radiation pressure from young stars, Type II and Type Ia supernovae, photoelectric heating, and photoheating from ionizing radiation. We calculate the energy, momentum, mass, and metal return according to the STARBURST99 stellar population synthesis model (Leitherer et al. 1999) and Kroupa IMF (Kroupa 2002). Full details of the implementation of gas and gravitational physics are provided in Hopkins et al. (2017).

All simulations analysed in this work are a part of the FIRE-2 simulation suite of the FIRE project.[2] Most are based on the initial conditions previously explored with FIRE-1 models in Hopkins

[1] http://www.tapir.caltech.edu/~phopkins/Site/GIZMO
[2] http://fire.northwestern.edu

Table 1. Simulation details. M_h^0 and M_*^0 are the halo mass and stellar mass (within $0.2R_{vir}$) of the largest halo in the zoom-in region at $z = 0$; R_{vir} is the virial radius; m_b is the mass of a gas particle in the simulation; m_{dm} is the mass of a DM particle in the simulation. ϵ_b is the minimum gravitational softening of a gas particle; ϵ_{dm} is the Plummer equivalent gravitational softening of a DM particle. All simulations are a part of the FIRE-2 simulation suite (Hopkins et al. 2017). The initial conditions for **m11q** are from Kim et al. (2014), while **m11z** and **m11c** are from Chan et al. (2015). **m11a**, **m11b**, and **m11f** are newly targeted haloes in the mass range relevant for UDGs. Note that Hopkins et al. (2017) presented higher resolution runs of **m10z** and **m11c**, which we discuss in Appendix C.

Name	M_h^0 ($M_\odot$)	M_*^0 ($M_\odot$)	R_{vir} (kpc)	m_b ($M_\odot$)	ϵ_b (pc)	m_{dm} ($M_\odot$)	ϵ_{dm} (pc)
m10z	3.5e10	5.3e7	85	2.1e3	0.43	1.0e4	43
m11a	4.1e10	1.2e8	90	2.1e3	0.43	1.0e4	43
m11b	4.4e10	1.1e8	92	2.1e3	0.43	1.0e4	43
m11q	1.2e11	1.0e9	1.4e2	7.1e3	0.42	3.5e4	14
m11c	1.5e11	2.0e9	1.4e2	1.7e4	0.86	8.3e4	86
m11f	4.9e11	2.7e10	2.1e2	1.7e4	0.86	8.3e4	86

et al. (2014) and Chan et al. (2015)[3,4] while several additional galaxies were specifically simulated to explore the relevant mass scale of UDGs. Our sample includes all FIRE-2 galaxies with $z = 0$ stellar mass $5 \times 10^7 - 2 \times 10^9$ $M_\odot$ and one additional higher mass galaxy, **m11f**, selected to explore the UDG formation with quenching at high redshift. All satisfy the absolute magnitude range of the observed UDGs in at least one simulation snapshot in the redshift interval $z = 0 - 3$ when post-processed with our mock observations. Here, we focus on properties of stellar population in our simulations; their gas properties were explored in El-Badry et al. (2018). Parameters of the simulated galaxies are listed in Table 1.

The galaxies we examine are isolated field dwarfs with $M \sim 10^{h\,10-11}$ $M_\odot$ at $z = 0$, where the effects of stellar feedback on the underlying density distributions are *large* (see e.g. Chan et al. 2015, El-Badry et al. 2016, Tollet et al. 2016; but also Oñorbe et al. 2015 and Fitts et al. 2017 for the effect in lower mass haloes). Our simulations were run in a 'standard' flat ΛCDM cosmology with the following cosmological parameters: $\Omega_0 \approx 0.27$, $\Lambda \approx 0.73$, $\Omega_b \approx 0.045$, and $h \approx 0.7$.

We note that our simulated haloes have a normal distribution of spin parameters. We measured the spin parameters of our DM haloes (in default runs with full hydrodynamics and feedback) and found that at $z = 0$ all except one are within 1σ of the measured spin parameter distribution from Macciò, Dutton & van den Bosch (2008) with values $\lambda \sim 0.02 - 0.035$. The exception is **m11b** whose spin parameter is about 2σ above the mean with spin parameter $\lambda = 0.077$.

2.2 Simulation analysis and mock observations

All of our isolated dwarf galaxies remain gas-rich and star-forming until $z = 0$. This is in contrast to UDGs observed in galaxy clusters, many of which are quenched, likely because of the environmental

effect of the clusters. To mimic the quenching of star formation that likely occurs during the infall into the cluster environment, we artificially stop each galaxy's star formation at the assumed quenching time $t = t_q$ (measured as cosmic time starting from the big bang) and passively evolve its stellar populations to $z = 0$. The minimum t_q we consider is 2 Gyr (i.e. $z \sim 3.5$) since all of the simulated galaxies in this work have sizes smaller than ~ 1 kpc at earlier times and would therefore not satisfy our observationally motivated UDG selection (see Section 3).

We assume that stellar morphology has not changed since t_q. Even if quenching processes remove galactic gas (e.g. via ram pressure), this is a reasonable assumption as long as the galaxy is DM dominated and stays far from the cluster centre where cluster tidal interactions are important. In Appendix B, we study dynamical effects of gas removal and show that they tend to slightly decrease surface brightness but do not affect any of our conclusions. We do not attempt to account for other possible cluster interactions (e.g. tidal disruptions, galaxy harassment, etc.), which would require a full scale galaxy cluster simulation that is beyond the scope of our paper. In other words, our UDG candidates are simulated dwarf galaxies at t_q, but with their stellar populations artificially aged to $z = 0$.

In order to compare our simulated galaxies with observations, we produce mock images at $z = 0$ for $t_q \sim 2 - 14$ Gyr ($z_q \sim 3.4 - 0$) with passively evolved stellar populations and perform mock observations to estimate their g-band surface brightnesses $\mu(g)$, effective radii r_{eff}, and $g - i$ colours. We follow the steps in van Dokkum et al. (2015a) closely for a more direct comparison. The galaxy images are initially centred on the haloes of their main progenitors identified with the Amiga Halo Finder (AHF) (Knollmann & Knebe 2009), which uses an adaptive mesh refinement hierarchy (Knebe, Green & Binney 2001) to define centres. Halo centres may not coincide with galaxy centres during ongoing mergers or instabilities, however, so we relocated centres with a χ^2 minimization on galaxy images with GALFIT (Peng et al. 2002). To calculate enclosed masses (e.g. in Figs 6 and 7), we applied the two-step procedure described in Chan et al. (2015) to centre on the stellar distribution of the galaxy. We use AHF and virial overdensities from Bryan & Norman (1998) to calculate virial mass M_h, virial radius R_{vir} and M_*, the total stellar mass enclosed within $0.2R_{vir}$.

van Dokkum et al. (2015a) inferred axial ratios and effective radii from combined $g + i$ band images, and surface brightnesses from g-band images. To follow their procedure, we generated a table of SDSS (Sloan Digital Sky Survey) g- and i-band luminosities for stellar populations of different ages and metallicities[5] with the Flexible Stellar Population Synthesis model (FSPS) (Conroy, Gunn & White 2009), assuming the latest Padova stellar evolution model (Marigo & Girardi 2007; Marigo et al. 2008) and the Kroupa initial mass function (Kroupa 2002). The luminosity of each stellar particle is interpolated from the table according to their stellar ages, masses and metallicities. Then, for each simulation output we project the $g + i$ band luminosities over a 40 kpc × 40 kpc (60 kpc × 60 kpc for **m11f**) region on to 1000^2 uniform mesh, and generate mock galaxy images. We explore the effects of lowering the image resolution (down to 100 × larger pixels, i.e. 100 × 100 pixels per image) in

---

[3]The FIRE-1 runs corresponding to **m10z**, **m11q** and **m11c** were named as **m10h573**, **m11** and **m11h383**, respectively.

[4]We note that our values of m_b and m_{dm} for the runs included in Chan et al. (2015) differ from values in Table 1 in that paper owing to their omission of the factor of $h = 0.7$. This omission did not affect any of the results quoted in that paper.

[5]Although we do not use the Canada–France–Hawaii Telescope (CFHT) MegaCam filters, which the observations in van Dokkum et al. (2015a) are performed with, the difference between the SDSS and CFHT filters are negligible and would not affect our main results.
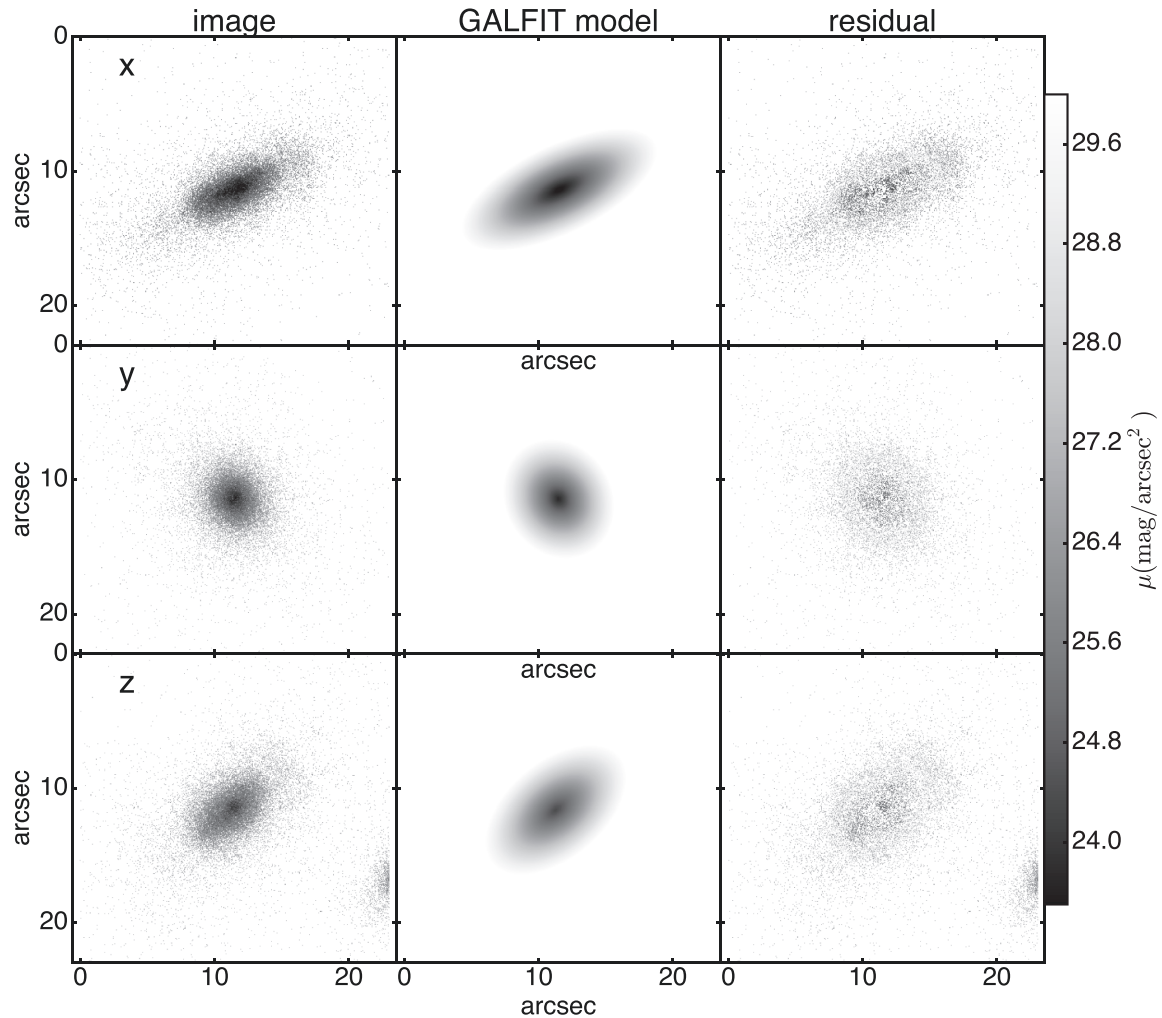
Figure 1. The mock *g*-band image of **m11b** ($M_* \sim 10^8\,\mathrm{M}_\odot$) quenched at a cosmic time $t_\mathrm{q} = 6$ Gyr and passively aged stellar population from t_q to $z = 0$, the best-fitting GALFIT model and the residuals from the fit. Each panel spans 16 kpc × 16 kpc (23 arcmin × 23 arcmin if we place the galaxy at the distance of the Coma cluster) and is shaded according to surface brightness. The top, middle, and bottom panels show images viewed along *x, y,* and *z* direction, respectively.

Appendix D and found that this does not significantly affect our results.[6]

We also generate images with *g*-band luminosities to estimate *g*-band surface brightness. We do not account for any dust attenuation because we assume all gas is removed immediately after the infall (we briefly discuss the dust attenuation effects at $z = 0$ in Section 4.4). The left-hand panel of Fig. 1 shows the processed *g* + *i* band image of **m11b** at $z = 0$ with $t_\mathrm{q} = 6$ Gyr (i.e. passively evolved from $z = 1$), viewed along three perpendicular directions.

To estimate structural parameters from the mock images, we fit them with the Sérsic profiles using GALFIT (Peng et al. 2002, 2010), similar to the techniques used in other UDG observations (e.g. Koda et al. 2015; Mihos et al. 2015). We allow n_s to vary in our fits to increase convergence (Koda et al. 2015; Mihos et al. 2015). Our galaxies have $n_s = 0.8 \pm 0.4$, close to the $n_s = 1$ profiles, used in van Dokkum et al. (2015a). We have compared central surface brightness of our galaxies obtained with the $n_s = 1$ fits to those

with variable n_s and found only minor differences. Our fits do not account for sky noise: we have tested adding sky noise to our images in Appendix A and again found very little difference in the inferred properties of our galaxies (see Fig. A1). The middle panels of Fig. 1 show the GALFIT models and the right-hand panels show the residuals.

3 RESULTS

3.1 Effective radius and surface brightness

We first define UDGs based on the van Dokkum et al. (2015a) sample: (1) $\mu(g) \gtrsim 23.5$ mag arcsec^{-2} and (2) $r_\mathrm{eff} \gtrsim 1.25$ kpc. This definition is also similar to the selection criteria in Román & Trujillo (2017a). Koda et al. (2015) and Mihos et al. (2015) also use the similar r_eff limit to define UDGs, but require different μ cutoffs according to the bands they use.[7]

---

[6]This is because GALFIT attempts to fit for average surface brightness within each elliptical ring.

[7]Although our UDG definition does not include limits on the total magnitude, we may impose magnitude cuts to match specific observation samples in the following, e.g. Table 2.

Fig. 2 shows the central g-band surface brightness $\mu(g, 0)$, defined as the surface brightness of the fitted Sérsic profile at the centre, as a function of the effective radius of our galaxies, compared to the observed red UDGs and dwarf galaxies associated with galaxy clusters. We 'observe' each galaxy along three perpendicular viewing angles for different assumed quenching times, and label the results with different symbols accordingly in Figs 2 and 3. The choice of viewing angle only mildly affects the results, consistent with expectations for roughly spheroidal geometries. The differences in surface brightness between different viewing angles are in general smaller than 1 mag arcsec^{-2}, except for very irregular geometries during mergers.

In Fig. 2, we show that μ and r_{eff} of **m10z**, **m11a**, **m11b**, and **m11q** are a good match to the observed red UDGs for $t_q \sim 6-13$ Gyr. Later quenching times allow these galaxies to form more stars, while strong stellar feedback increases their r_{eff}. **m11c** and **m11f** agree with the red UDGs in van Dokkum et al. (2015a) only for very early quenching times ($t_q \sim 2-4$ Gyr) but have much higher surface brightness if their star formation proceeds to later times. For early quenching times these two galaxies are therefore valid progenitors of UDGs according to our criteria stated in Section 3.1. If we, in addition to our standard criteria, consider the absolute magnitude range of van Dokkum et al. (2015a), **m11f**, can only reproduce the bright end of the UDGs for $t_q \lesssim$ Gyr, owing to its large g-band magnitude if quenched much later (see Figs 3 and 4). Overall, the luminosities, effective radii, and colours of stellar populations formed by $z \sim 0-3$ in dwarfs simulated with the FIRE-2 model are consistent with those of observed UDGs, but galaxies forming in more massive haloes (at $z = 0$) require earlier quenching times.

Fig. 3 shows effective radius, r_{eff}, as a function of absolute g-band magnitude, M_g, for the FIRE-2 dwarfs and the observed red UDGs from van Dokkum et al. (2015a). We also show lines indicating the average surface brightness within the effective radius.[8]

All galaxies can roughly match the parameter space of the observed UDGs for a wide range of quenching times, except **m11c** and **m11f**, our two most massive systems. These galaxies meet our UDG criteria only for snapshots with $t_q < 3$ Gyr (excluding some occasional contraction periods and minor mergers) and represent more massive UDGs. At larger t_q, their surface brightnesses are higher than the observed UDGs in the van Dokkum et al. (2015a) sample.

3.2 Effects of quenching time

In this section, we explore how properties of our simulated galaxies at $z = 0$ depend on quenching time. We explore the range of the allowed quenching times by matching the observed properties of our simulated UDGs, e.g. r_{eff}, M_g, and μ_g, to the observations.

We plot the properties of our galaxies as a function of quenching time in Fig. 4 and compare them with the median values from the observations (van Dokkum et al. 2015a), shown with horizontal dashed lines. The grey regions show the minimum and maximum observed values for r_{eff}, M_g, and μ_g. We also show the evolution of axial ratio, $g - i$ colour, stellar mass M_*, stellar age Age$_*$, and metallicity [Fe/H].

Absolute g-band magnitudes and mass weighted stellar ages drop with quenching time for all galaxies because later quenching im-

---

[8] Note that the difference between average surface brightness within r_{eff} and central surface brightness, $\langle \mu_g \rangle_e - \mu(g, 0)$, is generally small because of relatively flat profiles, but it can occasionally reach up to ~0.5 mag.

plies a shorter passive evolution period and a longer time during which a galaxy can form stars. The typical axial ratios of our simulated galaxies are ~ 0.8 since our galaxies are usually spheroidal owing to continuous stellar feedback, which prevents formations of prominent discs in dwarfs (Wheeler et al. 2017; El-Badry et al. 2018). Our most massive galaxy builds a stellar disc at late times, which shows as a fast drop in axial ratio.

Finally, the $g - i$ colours of all but our most massive dwarfs are approximately $0.75-0.85$ for $t_q \lesssim 10-11$ Gyr, consistent with observations of UDGs. The slow change in colours is caused by the interplay between increasing metallicities and decreasing mean stellar ages as we increase t_q. These competing effects mostly cancel out and prevent strong changes in the overall colours of the galaxies until $t_q \sim 10$ Gyr for lower mass dwarfs and $t_q \sim 12$ Gyr for our higher mass dwarfs. This implies that mean stellar ages and metallicities of UDGs cannot be determined with $g - i$ colours alone.

Although we cannot infer precise quenching times of observed UDGs from their $g - i$ colours only, effective radius, surface brightness, and g-band magnitude can provide tighter constraints. **m10z**, **m11a**, and **m11b** correspond to red UDGs if $t_q \gtrsim 5$ Gyr. The more massive dwarfs **m11q** and **m11c** must have $t_q \lesssim 6$ Gyr and $t_q \lesssim 3$ Gyr, respectively, in order to match the g-band magnitude of the sample in van Dokkum et al. (2015a).

Stellar age and metallicity measurements also give useful constraints on quenching time. Kadowaki, Zaritsky & Donnerstein (2017) found low stellar metallicity ([Fe/H] $\lesssim -1.5$) from the stacked spectrum of Coma UDGs. Gu et al. (2017) analysed the optical spectra of three of the brightest Coma UDGs and found they are metal poor ([Fe/H] $= -0.8^{+0.5}_{-0.5} - 1.3^{+0.4}_{-0.4}$) and old ($7.9^{+3.6}_{-2.5} - 9.1^{+3.9}_{-5.5}$ Gyr). Using optical through near-infrared SED fitting, Pandya et al. (2018) inferred a UDG in the Virgo cluster (VCC 1287) to be metal poor ([Fe/H] $\lesssim -1.0$) and old ($\gtrsim 9$ Gyr). These data and Fig. 4 together imply an early quenching scenario for cluster UDGs ($t_q \lesssim 6$ Gyr).

3.3 Characteristic properties of simulated red UDGs

Table 2 lists the properties of our simulated galaxies, for a range of quenching times during which they match the observed range of r_{eff} and $\mu(g, 0)$. We also list the median values of the observed UDGs (van Dokkum et al. 2015a). When the g-band magnitude of our simulated galaxy matches the median of the observed sample at a certain quenching time, the galaxy also yields a close match in effective radius, central surface brightness, and colour, suggesting that our simulated galaxies are good analogues of the observed UDGs. While our galaxies are slightly less spheroidal than UDGs, we expect they would be rounder if the dynamical effect of gas removal were taken into account (shown in Appendix B). Simulated and observed UDGs have similar stellar masses, largely determined by the absolute magnitude selection of the sample, since old stars have approximately constant stellar mass-to-light ratios.

With earlier quenching times, stars are older at $z = 0$ and have lower metallicities because of the evolution of the galaxy mass-metallicity relation (e.g. Zahid et al. 2013; Ma et al. 2016). Metallicities of UDGs could potentially be used to constrain quenching times of UDGs. However, for our simulated sample, typical metallicities grow very slowly with later quenching times because the metallicity evolution is offset by larger stellar masses of the galaxies that satisfy the UDG selection for early quenching times (typically galaxies with $z = 0$ halo mass $\gtrsim 10^{11}$ M$_\odot$). A larger sample
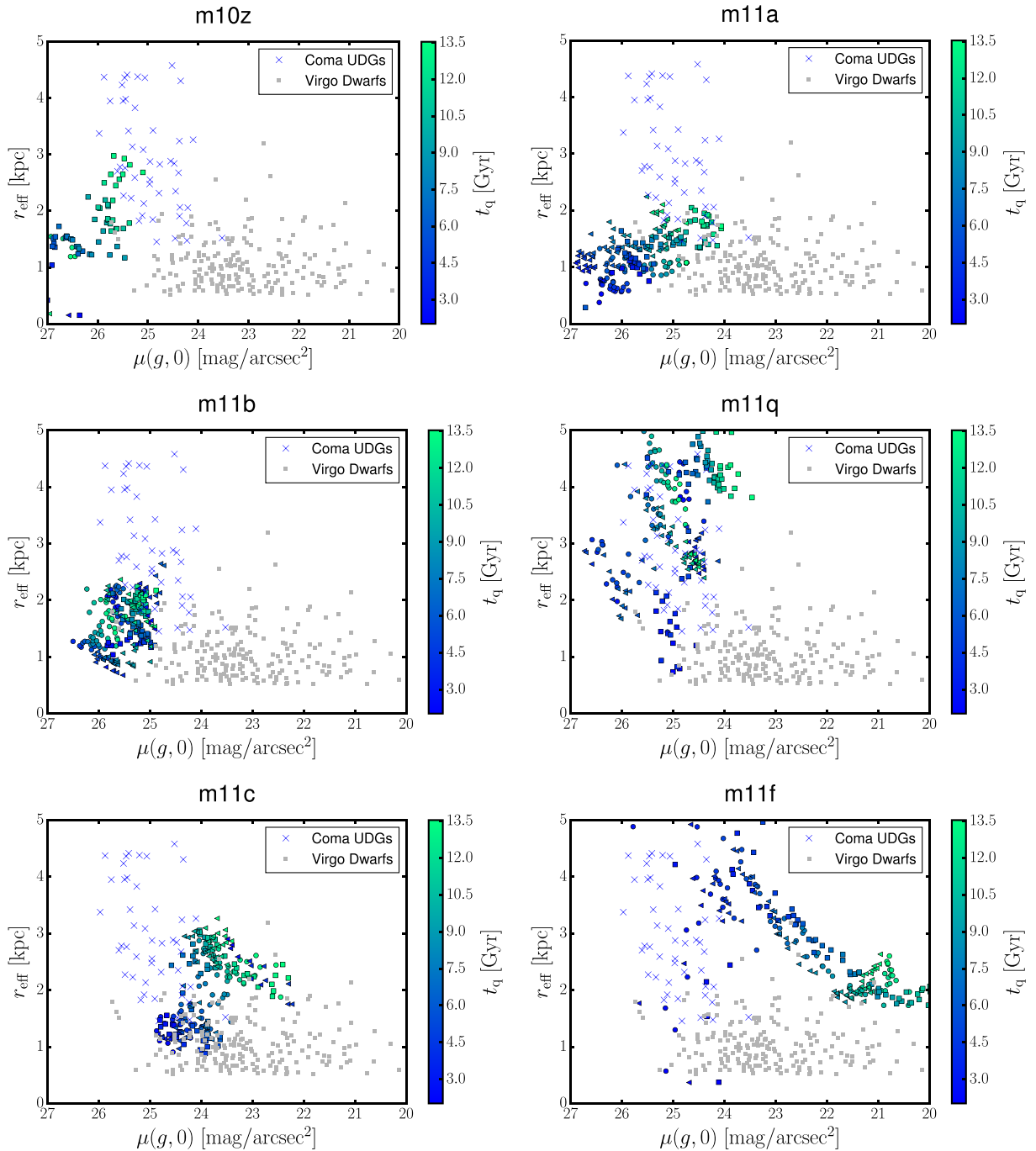
Figure 2. Central *g*-band surface brightness of our simulated galaxies plotted against their effective radius. The colours of points represent quenching times, at which we artificially stop their star formation and passively evolve their stellar population to $z = 0$ according to FSPS. The styles show different viewing angles (*squares*: along *x*-axis; *triangles* : along *y*-axis; *circles*: along *z*-axis). Each panel represents a single simulated galaxy. We also show the observed values of early-type galaxies in the Virgo cluster (Gavazzi et al. 2005) and UDGs in the Coma cluster (van Dokkum et al. 2015a). We find that all of these simulated galaxies have sizes and surface brightnesses that are consistent with observed UDGs, depending on the quenching time that we assume.

of simulated galaxies is needed to explore the metallicity trends in detail.

Observationally, long exposure spectroscopic studies combined with stellar population modelling are necessary to determine dynamical masses, stellar ages, and metallicities of UDGs and potentially constrain their origin and quenching times (e.g. Makarov et al. 2015; Gu et al. 2017; Kadowaki et al. 2017; Pandya et al.

2018). LSB sensitive instruments such as the Keck Cosmic Web Imager (Martin et al. 2010) should be helpful in extending such studies to a larger number of objects. Overall, for the observed range of surface brightnesses and magnitudes of UDGs, our model predicts a uniform population of galaxies in terms of their stellar masses and $g - i$ colours but with a broad range of average stellar ages.
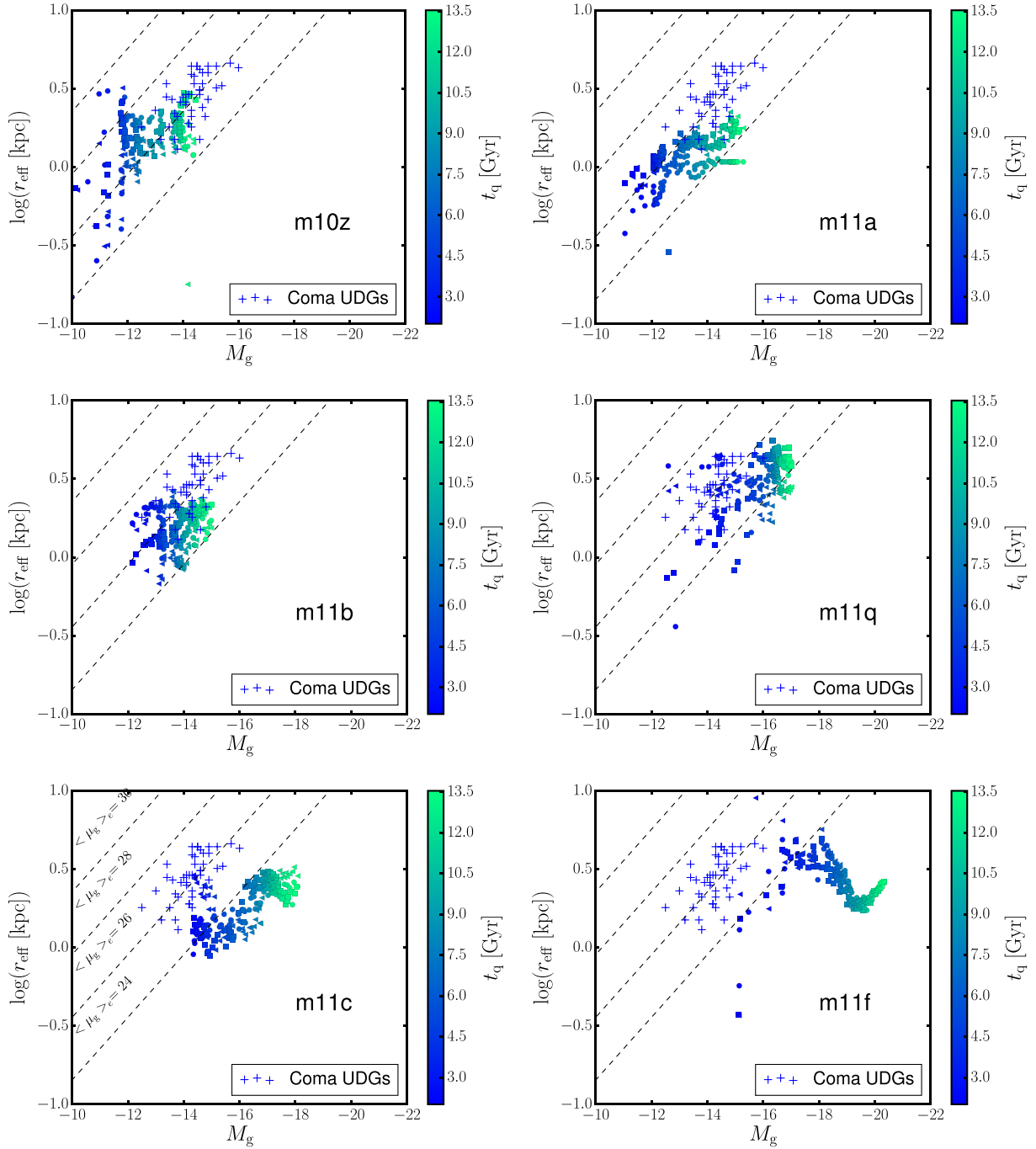
Figure 3. Absolute *g*-band magnitude (in AB) plotted against effective radius. Dashed lines show constant *g*-band surface brightnesses within effective radii, $\langle \mu_g \rangle_e$(mag arcsec^{-2}). The colour of each point represents the assumed quenching time, whereas point styles show different orientations, as in Fig. 2. Cross symbols represent the observed UDGs from van Dokkum et al. (2015a). Our simulated galaxies can match the range of observed surface brightness and magnitudes at some quenching times, but more massive UDGs have to quench very early to match the range of observed magnitudes.

3.4 UDGs and their haloes

The M_*/M_h row in Table 2 shows the stellar-to-halo mass ratios measured at the range of quenching times when these galaxies satisfy red UDG criteria. The upper panel of Fig. 5 shows the ratios as a function of quenching time. While the ratios are small, they do not deviate much from the dwarfs with similar masses in the Local Group (LG; McConnachie 2012). It is reasonable to expect that DM haloes of UDGs cannot grow significantly after t_q due to cluster influence. Furthermore, the outskirts of UDG haloes will likely be stripped or modified following the infall into a cluster. Stellar-to-total mass (stellar + DM) ratio within effective radius, given in the last row of Table 2, is thus a more robust quantity. This ratio shows that for quenching times when our galaxies are analogues to UDGs, their central regions are strongly DM dominated.

Recently, Beasley et al. (2016) claimed that one of the UDGs in the Virgo cluster, VCC 1287, has a stellar fraction $\sim 3.5 \times 10^{-4}$,
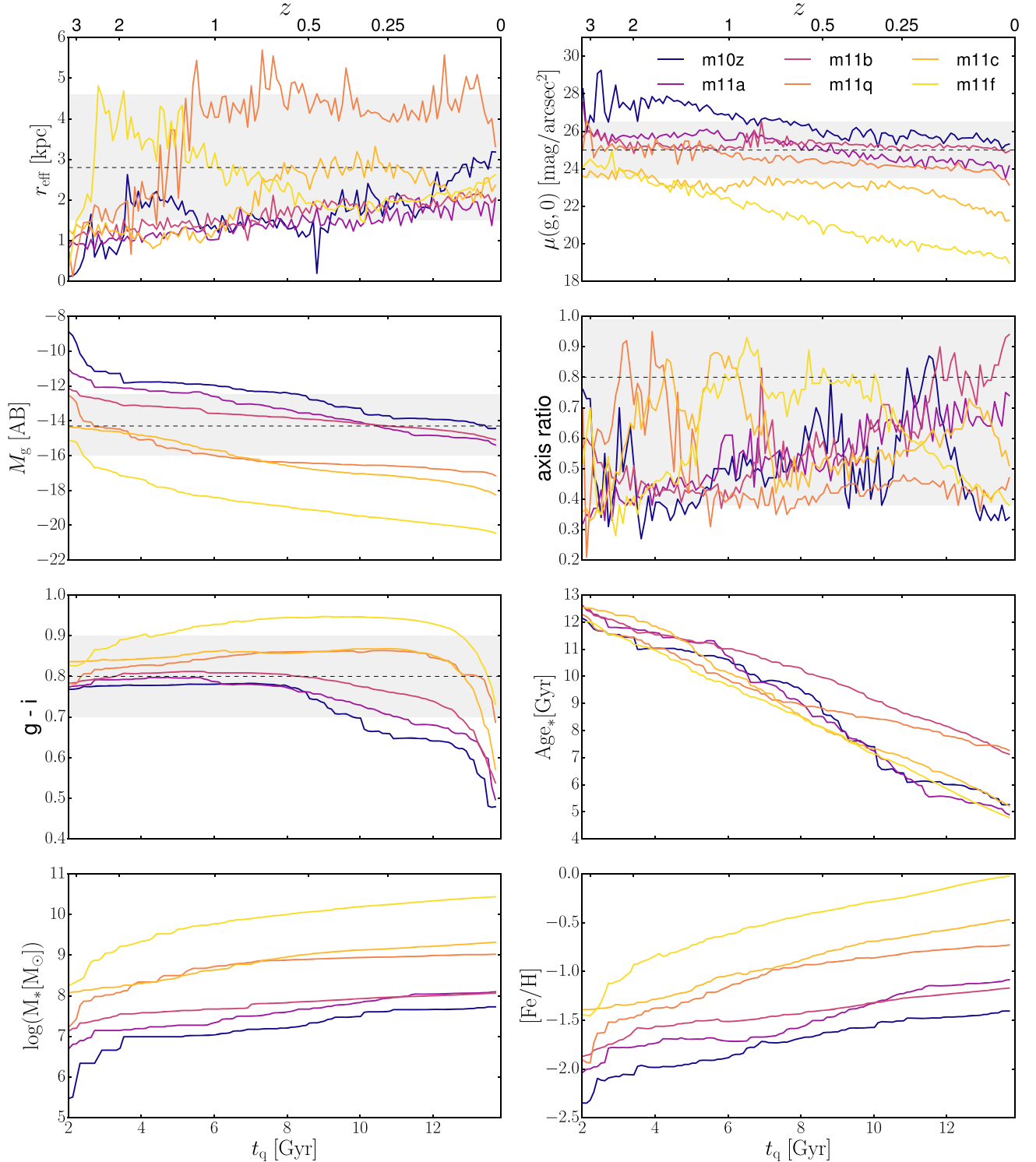
Figure 4. Effective radius r_{eff}, central g-band surface brightness $\mu(g, 0)$, absolute g-band magnitude M_g, axial ratio, $g - i$ colour, (mass-weighed) stellar age Age$_*$, stellar mass within $0.2R_{\mathrm{vir}}$, M_*, as functions of quenching time. We use GALFIT to determine central surface brightness from the g-band images and effective radii and axial ratios from $g + i$ images. All other quantities are obtained directly from star particles without using a fit. All quantities are measured as viewed along x-axis. Horizontal dashed lines show the median values of the observed UDGs (van Dokkum et al. 2015a) while horizontal shaded regions show their ranges.

much lower than 'normal' dwarf galaxies of similar masses whose stellar fractions are $\sim 10^{-2}$ to 10^{-3}. They estimated its total halo mass using the velocity dispersion and the number of GCs. While at face value this is lower than the stellar fractions of our simu-

lated dwarfs, they do not directly measure the mass at the virial radius; instead, they infer it assuming a density profile, which introduces significant uncertainty. However, their observations do directly constrain the mass within 8.1 kpc with the 'trace mass

Table 2. Characteristic properties of simulated UDGs. The properties of the simulated galaxies are extracted for quenching times when they satisfy UDG selection criteria from Section 3.1 and when, in addition, their M_g falls within the observed range of red UDGs in van Dokkum et al. (2015a), $M_g = [-16.0, -12.5]$. The values presented in large numbers are determined at the t_q for which the g-band magnitude is the closest to the median observed magnitude $M_g = -14.3$, while the small numbers show the maximum and minimum during the range of t_q described above. Effective radius, r_{eff} and central surface brightness, $\mu(g, 0)$ are determined by GALFIT. The absolute magnitude, M_g and colour, $g - i$, are determined directly from star particles. Stellar age and metallicity are mass weighted. Stellar mass is measured within $0.2R_{vir}$. Next, we show the range of halo masses, $M_{h, q}$, and the stellar-to-halo mass ratio, $M_*/M_{h, q}$. $M_{1/2, obs}$ is the total stellar plus DM mass within de-projected half-light radius ($r_{1/2} = r_{eff} \times \sqrt{b/a} \times 4/3$, where b/a is axial ratio from GALFIT), whereas $M_{1/2}$ is the total stellar plus DM mass within 3D half stellar mass radius. In addition, we show the ratio of the stellar mass to stellar plus DM mass, $f_{1/2, *}$, within the de-projected half-light radius $r_{1/2}$. All of the masses and mass ratios are measured at t_q (i.e. we assume that structural properties of galaxies and haloes remain fixed after t_q). The second column ('Observed') shows properties of observed UDGs. Given the constraints on effective radius, surface brightness, and g-band luminosity/magnitude, we predict colours, ages, metallicities, and stellar-to-halo mass ratios. All quantities are measured as viewed along x-axis.

	Observed		m10z		m11a		m11b		m11q		m11c		m11f	
t_q (Gyr)	–	$^{4.6}_{1.5}$	13.4	$^{13.5}_{8.2}$	10.4	$^{13.5}_{5.7}$	10.7	$^{13.5}_{2.6}$	2.7	$^{6.0}_{2.3}$	2.0	$^{6.5}_{2.0}$	2.2	$^{2.4}_{2.2}$
r_{eff} (kpc)	2.8	$^{4.6}_{1.5}$	2.9	$^{3.2}_{1.4}$	1.4	$^{1.4}_{1.3}$	2.0	$^{2.0}_{1.3}$	1.7	$^{4.0}_{1.3}$	1.3	$^{1.5}_{1.3}$	1.6	$^{2.5}_{1.6}$
$\mu(g, 0)$	25.0	$^{26.5}_{23.5}$	25.67	$^{25.67}_{26.3}$	24.42	$^{23.76}_{25.96}$	25.14	$^{24.92}_{25.77}$	25.98	$^{25.47}_{26.01}$	23.74	$^{23.4}_{23.74}$	26.24	$^{24.87}_{26.24}$
M_g[AB]	−14.3	$^{-16.0}_{-12.5}$	−14.3	$^{-14.4}_{-12.5}$	−14.3	$^{-15.2}_{-12.5}$	−14.3	$^{-15.0}_{-12.7}$	−14.3	$^{-16.0}_{-13.8}$	−14.3	$^{-16.0}_{-14.3}$	−15.2	$^{-16.0}_{-15.2}$
$g - i$	0.8	$^{0.7}_{0.9}$	0.54	$^{0.51}_{0.77}$	0.72	$^{0.59}_{0.79}$	0.76	$^{0.59}_{0.79}$	0.81	$^{0.85}_{0.8}$	0.84	$^{0.86}_{0.84}$	0.83	$^{0.83}_{0.83}$
Age$_*$ (Gyr)	–		5.4	$^{5.4}_{9.3}$	6.8	$^{5.1}_{11.0}$	8.7	$^{7.2}_{12.2}$	11.6	$^{9.9}_{11.8}$	12.6	$^{9.9}_{12.6}$	12.1	$^{11.9}_{12.1}$
[Fe/H]	–		−1.41	$^{-1.41}_{-1.67}$	−1.26	$^{-1.1}_{-1.7}$	−1.28	$^{-1.17}_{-1.77}$	−1.54	$^{-1.15}_{-1.71}$	−1.39	$^{-1.02}_{-1.39}$	−1.45	$^{-1.37}_{-1.45}$
M_* (10^8 M$_\odot$)	–		0.53	$^{0.55}_{0.19}$	0.8	$^{1.22}_{0.21}$	0.94	$^{1.16}_{0.26}$	1.2	$^{5.58}_{0.74}$	1.36	$^{5.76}_{1.36}$	2.77	$^{6.12}_{2.77}$
$M_{h, q}$ (10^{10} M$_\odot$)	–		3.5	$^{3.5}_{3.4}$	3.8	$^{4.1}_{2.9}$	4.0	$^{4.4}_{1.3}$	3.4	$^{9.3}_{3.1}$	2.2	$^{11.4}_{2.2}$	5.5	$^{9.7}_{5.5}$
$10^3 M_*/M_{h, q}$	–		1.53	$^{1.59}_{0.55}$	2.04	$^{2.94}_{0.72}$	2.36	$^{2.65}_{1.92}$	3.75	$^{6.09}_{1.71}$	6.14	$^{4.93}_{6.14}$	4.91	$^{5.94}_{4.91}$
$M_{1/2, obs}$ (10^8 M$_\odot$)	–		4.28	$^{4.98}_{2.29}$	3.41	$^{2.5}_{2.16}$	7.65	$^{9.3}_{2.6}$	2.1	$^{21.54}_{1.55}$	3.68	$^{3.0}_{3.68}$	13.66	$^{14.25}_{13.66}$
$M_{1/2}$ (10^8 M$_\odot$)	–		8.83	$^{9.53}_{4.67}$	4.76	$^{4.66}_{2.82}$	7.68	$^{9.92}_{4.87}$	20.48	$^{32.83}_{52.56}$	10.06	$^{30.13}_{10.06}$	25.1	$^{26.0}_{25.1}$
$10 \times f_{1/2, *}$	–		0.39	$^{0.4}_{0.26}$	0.86	$^{1.2}_{0.37}$	0.6	$^{0.58}_{0.32}$	0.27	$^{0.99}_{0.05}$	0.91	$^{0.73}_{0.91}$	0.64	$^{0.83}_{0.64}$

estimator' (Watkins, Evans & An 2010) using GCs as tracers; we therefore compare their estimate to the enclosed masses of our simulated galaxies measured at this radius.

Fig. 6 shows the enclosed stellar and DM mass within 8.1 kpc at different t_q. We do not include the gas mass as we are comparing to a red cluster UDG that has likely lost its interstellar gas. The enclosed masses of our simulated dwarfs are roughly constant over 10 Gyr, and our lower mass haloes **m10z, m11a,** and **m11b** can match the measured enclosed mass in VCC 1287, while having normal stellar fractions. Therefore, our simulated red UDGs formed as regular dwarf galaxies. However, their growths stopped at early times while outskirt of isolated haloes can continue growing until much later. We may wrongly infer much larger masses of their host haloes from the measurements of the central regions of UDGs and the mass profiles of isolated haloes.

The enclosed masses within the inner 8.1 kpc of several of our dwarfs match the value for VCC 1287 for long periods of time, during which their total halo masses change significantly. For example, the halo mass of **m10z** grew by a factor of 6 (see Fig. 5) while the mass within 8.1 kpc remained consistent with VCC 1287. This constancy (lack of growth) of the inner DM profile is typical in ΛCDM: most 'growth' in low-mass haloes at late cosmic times occurs because of a drop in the reference density, not because of a change in the mass enclosed within a fixed physical radius ('pseudo-evolution', Diemer, More & Kravtsov 2013; van den Bosch et al. 2014; Wetzel, Tollerud & Weisz 2015). The exception is **m11f**, which forms the most massive halo in our sample and whose stellar mass grew rapidly at late times, contributing a significant fraction of mass within 8.1 kpc and contracting the underlying DM profile (see e.g. Chan et al. 2015), although the overall effect is small as shown in Fig. 6.

Given the board range of possible quenching times and the constancy of inner halo mass, one cannot accurately estimate their total halo mass and stellar-to-halo mass ratios from their enclosed masses without knowing their quenching times. If we determine their total halo mass by comparing their inner mass and the $z = 0$ halo mass profile of an isolated halo, the estimated halo mass may be several times larger than the true value.
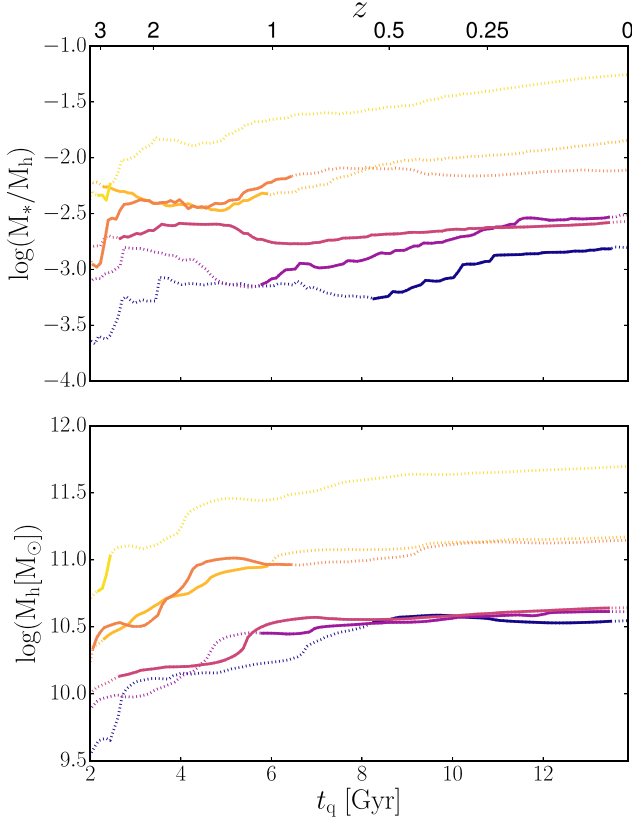
4 DISCUSSION

4.1 UDG mass estimates

While recent observations have constrained the masses of the inner haloes of several UDGs, the fraction of known UDGs with accurate measurements for the inner halo mass is very small. For those few systems with their inner masses measured, moreover, one must assume a density profile and extrapolate to infer a total halo mass. Together, these complications have led to two different views for the characteristic mass of UDG host haloes.

van Dokkum et al. (2015a, 2016) suggested that UDGs are 'failed' L⋆ galaxies, whereas Yozin & Bekki (2015), Amorisco & Loeb (2016), Beasley et al. (2016), Beasley & Trujillo (2016) and Peng & Lim (2016) argued they are 'failed' *dwarf* galaxies, based on the inferred ratios of stellar-to-halo mass. Both camps support their claims with the enclosed masses of UDGs inferred from velocity dispersions and the numbers of GCs.
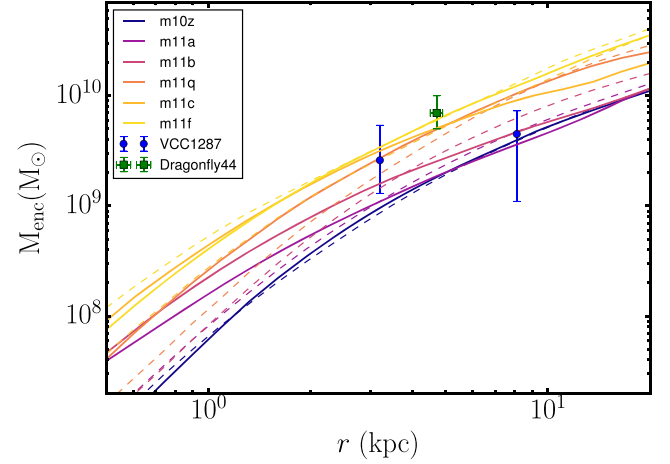
In the previous section, we showed that simulated UDGs can form in haloes with $M_h \sim 3-15 \times 10^{10}$ M$_\odot$ and that their central enclosed masses alone are not be a good indicator of their host halo mass at the time of quenching. Here, we compare the velocity dispersions and masses of our simulated UDGs to the several observed examples

Figure 7. Cumulative mass profiles of our simulated haloes including stars and DM, compared to the observations of VCC 1287 (*blue circle*) (Beasley et al. 2016) and Dragonfly 44 (*green square*) (van Dokkum et al. 2016). Solid (dashed) lines show the mass profiles at the earliest (latest) quenching time when simulated galaxies match the UDG selection criteria and the range of absolute magnitudes from van Dokkum et al. (2015a) (see Table 2). While our dwarf galaxies/haloes match the outer measurement of VCC 1287, our most massive galaxy/halo **m11f** in the sample is a good match for Dragonfly 44.

and critically examine the methodology used to infer masses from observations.

In Table 3, we show the range of velocity dispersion as seen along three perpendicular directions and the average values of each of our UDG analogues. The line-of-sight velocity dispersion of our galaxies ranges from ∼20 to 50 km s^{-1}. Our intermediate mass dwarfs provide a good match to $\langle \sigma \rangle = 33^{+16}_{-10}$km s^{-1} measured for VCC 1287 (Beasley et al. 2016) and our most massive UDG provides a match to $\langle \sigma \rangle = 47^{+8}_{-6}$km s^{-1} measured for Dragonfly 44 (van Dokkum et al. 2016).

With the measured velocity dispersion and effective radius, van Dokkum et al. (2016), Beasley et al. (2016), and Beasley & Trujillo (2016) inferred the enclosed mass within stellar half-light radius using an equation first presented in Wolf et al. (2010):

$$M_{1/2} \simeq 9.3 \times 10^5 \left(\frac{\langle \sigma_{\mathrm{los}}^2 \rangle}{\mathrm{km}^2\ \mathrm{s}^{-2}} \right) \left(\frac{r_{\mathrm{eff}}}{\mathrm{kpc}} \right) \mathrm{M}_{\odot}, \qquad (1)$$

where $M_{1/2}$ is the total mass within the 3D half-light radius and $\langle \sigma_{\mathrm{los}}^2 \rangle$ is the square of the line-of-sight velocity dispersion. Strictly speaking this relation is only valid for velocity dispersion dominated spherically symmetric systems but in practice, it is applied to estimate masses of a variety of dwarf galaxies (see discussions in González-Samaniego et al. 2017 for details).

We have applied this equation to the line-of-sight velocity dispersion and circularized r_{eff} for our dwarf galaxies and compared them to the actual enclosed mass within the 3D half-light radii.[9] When used directly with r_{eff} from GALFIT, this approach tends to over-predict the mass within the 3D effective radius and it shows large variations between different sight-lines. We get a better agreement and no systematic offset when we use the same approach after the

---

[9]To get the 3D half-light radius we follow the observational approach and estimate it from the circularized, de-projected effective radius, $r_{1/2} = r_{\mathrm{eff}} \times \sqrt{b/a} \times 4/3$, where r_{eff} and the axial ratio, b/a, are calculated by GALFIT.

Figure 5. (*Upper*) stellar fraction and (*lower*) halo mass as functions of quenching time for our sample of simulated galaxies. Solid lines show the quenching times when they satisfy the UDG criteria and dotted lines when they do not. The colour scheme follows Fig. 4. Our UDGs have 'normal' stellar fraction for they stellar and halo mass range (∼10^{-2}to 10^{-3}).
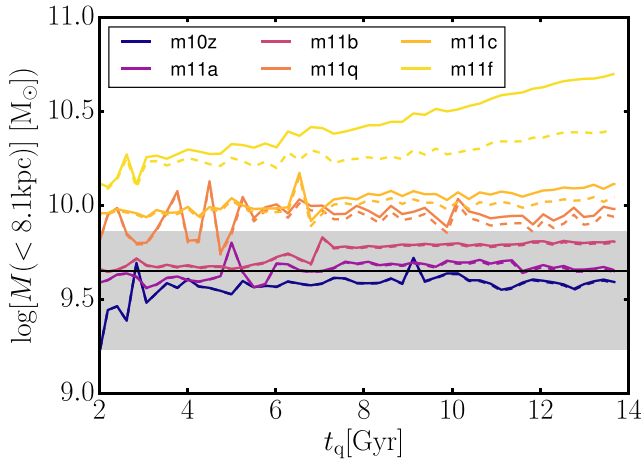
Figure 6. Enclosed stellar plus DM (solid) and DM only (dashed) masses of our simulated haloes within 8.1 kpc as a function of t_{q} compared to VCC 1287 (Beasley et al. 2016) (black horizontal line; shaded region indicates the uncertainties in the observation). The colours follow Fig. 4. The enclosed masses of **m10z**, **m11a**, and **m11b** roughly match the observed value for a broad range of t_{q}.

Table 3. Line-of-sight velocity dispersion $\langle\sigma\rangle$ $(=\sqrt{\langle\sigma_{\rm los}^2\rangle})$ calculated at $t_{\rm q}$ when galaxies satisfy the UDG criteria and their g-band absolute magnitude is closest to $M_g \sim -14.3$, the average absolute magnitude of UDGs in van Dokkum et al. (2015a). Velocity dispersions are measured in a cylindrical aperture with radius of 20 kpc, which includes vast majority of stars of each simulated UDG. The larger fonts show values averaged over three perpendicular directions, whereas the smaller fonts represent the maximum and minimum values of velocity dispersion as measured from three different perpendicular directions. For reference, $\langle\sigma\rangle$ of VCC 1287 is 33^{+16}_{-10} km s^{-1} (Beasley et al. 2016) and of Dragonfly 44 is 47^{+8}_{-6} km s^{-1} (van Dokkum et al. 2016).

	m10z		m11a		m11b		m11q		m11c		m11f	
$\langle\sigma\rangle$ (km s^{-1})	21.4	$^{22.6}_{20.1}$	21.6	$^{24.4}_{19.5}$	23.8	$^{24.1}_{23.1}$	31.2	$^{36.6}_{27.0}$	29.5	$^{33.4}_{27.2}$	47.7	$^{54.2}_{39.0}$

gas removal and subsequent relaxation (see the Appendix B), which tends to make galaxies smoother. Furthermore, when we apply this equation to the actual 2D half-mass radii (instead of half-light) we recover the actual enclosed mass within the 3D half-mass radius to within 15 per cent when averaged over three orthogonal projections (consistent with the tests of this mass estimator on a lower mass FIRE-2 simulations by González-Samaniego et al. 2017). We therefore conclude that velocity dispersion and effective radius can indeed provide reasonable estimate of the enclosed mass but a larger number of measured systems are needed to reach a robust measure of the typical enclosed masses of UDGs.

In Fig. 7, we present the enclosed mass profiles of our simulated haloes, including their stellar and DM components.[10] Profiles are compared to the inferred values from the observations of Beasley et al. (2016) and van Dokkum et al. (2016). The outer point from Beasley et al. (2016), already studied in Fig. 6, is well matched by several of our lower mass dwarf galaxies suggesting that it is indeed a 'quenched dwarf'.

The innermost point of VCC 1287 was calculated by the same authors with equation (1), with an additional assumption that GC velocity dispersion represents stellar velocity dispersion. This suggests a higher halo mass and it can be matched by four of our highest mass haloes **m11b**, **m11q**, **m11c**, and **m11f**. We note that using GC velocity dispersion instead of stellar velocity dispersion could be problematic because they only used 7 GCs, one of which is $r \gtrsim 3r_{\rm eff}$ away from the galactic centre. However, even with this limitation our **m11b** galaxy (with a $z = 0$ halo mass of 4.4×10^{10} M$_\odot$) can match both mass measurements, suggesting that VCC 1287 formed in a dwarf mass halo. Future spectroscopic studies with long integration times are needed to constrain its actual stellar kinematics.

The mass measurement of Dragonfly 44 came directly from stellar kinematics and equation (1), and it is matched by the largest galaxy in our sample, **m11f**. **m11c**'s mass profile provides a marginal match, but only for the latest quenching time for which it is still identified as a UDG. At the quenching times when this galaxy satisfies the UDG criteria, its halo mass is only $\sim 0.5-1 \times 10^{11}$ M$_\odot$, implying its total mass today $\sim 1 \times 10^{11}$ M$_\odot$, assuming that its growth is insignificant after its infall.

The number of GCs in Dragonfly 44 is 94^{+25}_{-20}, much higher than inferred from its stellar mass and luminosity (van Dokkum et al. 2016). Interestingly, according to relations in Harris, Harris & Alessi (2013), haloes with ~ 94 GCs have masses similar to that of **m11f** at $z = 0$. If all of the GCs in this system formed at very early times and the galaxy's stellar mass growth was stopped at high

redshifts, one should actually expect a very high number of GCs despite its low stellar and halo mass at quenching. This is because galaxies used for the Harris et al. (2013) relation continue growing their stellar and halo masses to much later times, unlike quenched UDGs. This may also explain the finding of Amorisco et al. (2018) that the ratios between the number of GCs and stellar mass are much higher in some UDGs than the galaxies studied in Harris et al. (2013).

We have also checked the more massive galaxies presented in Hopkins et al. (2017), **m12z**, **m12i**, and **m12c**, with $z = 0$ halo masses 8.7×10^{11}, 1.3×10^{12}, and 1.4×10^{12} M$_\odot$, respectively. They can only match the absolute magnitude and effective radius of Dragonfly 44 for $t_{\rm q} \lesssim 2$ Gyr, i.e. $z_{\rm q} \gtrsim 3.3$. At those times, their halo masses were around 10^{11} M$_\odot$ and their inner masses matched Dragonfly 44. Hence, Dragonfly 44 can form in a halo similar to the MW's progenitor, but it has to be quenched very early. This means that at the time of quenching, Dragonfly 44 is not hosted by a highly 'overmassive' DM halo. Its central halo mass, which appears unusually high for its estimated stellar mass, can be explained by its early formation time.

While forming a massive UDG requires very early quenching, we note that the Coma cluster was not fully formed so early in the structure formation process: the Coma cluster progenitor at $t = 2$ Gyr is expected to have less than 5 per cent of its present-day halo mass (Li et al. 2007). However, it is possible that the most massive UDGs ($M_\star \gtrsim 2 \times 10^8$ M$_\odot$) were quenched in group-mass progenitors that existed at high redshift. Indeed, semi-empirical constraints suggest that the majority of quiescent low-mass satellites in galaxy clusters today quenched in a group (Wetzel et al. 2013). The exact mechanism and the feasibility of such a scenario have to be explored with cosmological simulations that follow formations of galaxy groups or clusters.

Relatively early quenching of massive UDGs is also suggested by spectroscopic observations (Makarov et al. 2015; Gu et al. 2017; Kadowaki et al. 2017). In particular, Gu et al. (2017) inferred the stellar age and metallicity of Dragonfly 44 to be $8.9^{+4.3}_{-3.3}$ Gyr and [Fe/H] $= -1.3^{+0.4}_{-0.4}$, respectively. Our **m11f** simulation has similar metallicity at $t_{\rm q}$ when it is a UDG but our age estimate is close to the upper range suggested by observations (see Table 2). More massive simulated haloes result in even older stellar ages for the allowed range of quenching times. However, we stress that uncertainties in age measurement are too large for an accurate determination of the quenching time of Dragonfly 44.

Our analysis therefore suggests that galaxies like Dragonfly 44 formed early ($z \sim 3$), in $\sim 10^{11}$ M$_\odot$ haloes and stopped growing after their infall into a cluster. These objects could continue accreting mass, forming stars and reaching much higher luminosities and halo masses by $z = 0$ if they did not fall into a cluster. But due to the cluster influence, they should have stellar and halo masses significantly lower than those hosting L$\star$ galaxies at $z = 0$, since

---

[10]While gas can be a non-negligible component in our galaxies, we leave it out of the enclosed mass calculation as we are comparing to red UDGs in clusters.

their formation and quenching likely took place in much lower mass haloes. Overall, our simulations suggest that a majority of UDGs in clusters form in haloes that span a relatively broad range of halo masses (a few $\times 10^{10-11}$ M$_\odot$), with more massive UDGs forming in more massive haloes.

4.2 Galaxy expansion

One of the distinctive features of UDGs is their diffuseness. While we have already shown that stellar feedback leads to large effective radii and quenching of star formation can redden their colours, a large number of UDGs have been discovered in clusters, leading to a natural question: can satellite galaxies be further puffed-up with tidal heating and ram-pressure stripping? We explore the dynamical effects of gas removal in the Appendix B and show that this mildly increases the size and reduces the surface brightness further.

Without invoking clusters, Yozin & Bekki (2015), Amorisco & Loeb (2016), and Rong et al. (2017) proposed that UDGs are diffuse because their progenitors have larger angular momenta compared to normal galaxies. In other words, they are the *high spin tail* of the galaxy population. Yet, high spin galaxies are also more likely to resemble disc-like (Yozin & Bekki 2015) rather than spheroidal structures as observed in UDGs.[11] We note (as discussed in Section 2.1) that most of our simulated UDGs have normal spin parameters and we find no clear differences between the UDGs forming in low and high spin haloes.[12]

Observations also seem to contradict the 'high spin tail' scenario. For example van Dokkum et al. (2016) showed that Dragonfly 44 is dispersion dominated with no evidence of rotation and radial variations in the velocity dispersion. Similarly, Burkert (2017) argued that from their axial ratios, UDGs are unlikely puffed-up disc galaxies, but are instead similar to dwarf spheroidal galaxies.

Tidal stirring provides a possible pathway from gas-rich dwarf irregulars with rotational support to gas-free dwarf spheroidals through repeated tidal interactions with a massive host galaxy (Mayer et al. 2001; Klimentowski et al. 2007, 2009; Łokas et al. 2015). But this mechanism may not explain the abundance of UDGs with high axial ratios observed in a variety of environments, from cluster, cluster outskirt, and group, e.g. Martínez-Delgado et al. (2016), van der Burg et al. (2016, 2017), and Román & Trujillo (2017a,b). Furthermore, such mechanism might not be needed to transform most of the dwarf irregulars as a large fraction of the isolated dwarfs are likely dispersion dominated (Wheeler et al. 2017).

Stellar feedback can produce diffuse and spheroidal stellar distribution independent of host interactions. Through stellar migration and dynamical heating, feedback can decrease both the surface brightness and ellipticity of a dwarf galaxy while simultaneously increasing its effective radius (Chan et al. 2015; El-Badry et al. 2016). One generic effect of stellar feedback in dwarfs with $M_h \sim 10^{10-11}$ M$_\odot$ is a cored stellar profile (Read & Gilmore 2005; Stinson et al. 2013; El-Badry et al. 2016), implying a flat central light profile and low Sérsic index.[13] A flat surface brightness profile has been observed in one of the biggest UDGs in the Coma cluster

(van Dokkum et al. 2015b) and low Sérsic indices ($\sim 0.8-0.9$) in UDGs were reported in various observations (e.g. Koda et al. 2015; Román & Trujillo 2017b).

If feedback is the major driver of their diffuseness, it is very natural to expect an abundant population of diffuse galaxies far from cluster centres and even in the field, which we will discuss in Section 4.4. Román & Trujillo (2017a) and Martínez-Delgado et al. (2016) found galaxies with large $r_{\rm eff}$ and LSB even in underdense regions, so cluster interactions are likely not an essential factor for their diffuseness, consistent with the scenario where feedback plays the dominant role in shaping the UDGs.

Furthermore, for feedback-driven radial migration, old stars experienced more feedback episodes than young stars, so older stars will migrate outside and young stars remain near the centre. We should therefore expect mixed or even inverted age gradients in UDGs (El-Badry et al. 2016), i.e. stars far from the centre might be older than the stars at the centre, which could be observed in the future.

4.3 Gas removal and quenching

In order to quench dwarf galaxies, their gas supply needs to be truncated and their ISM gas also needs to be largely removed (or consumed) in order to stop their star formation. But the exact mechanism of gas removal from UDGs is still an open question. Tidal stripping is one possible mechanism, but it tends to remove weakly bound stars near the edges and reduce the sizes of the galaxies, making them more compact rather than more diffuse (e.g. Read et al. 2006). Mowla et al. (2017) did not find any signature of tidal stripping out to 4 $r_{\rm eff}$ in the Coma UDGs. However, Venhola et al. (2017) found elongated and distorted shapes of the largest UDGs ($r_{\rm eff} > 3$ kpc) in the Fornax cluster, which may indicate the effect of tidal stripping, but the total contribution is unclear. Simulating the interaction between a cluster and a high spin dwarf galaxy, Yozin & Bekki (2015) showed ram pressure stripping can efficiently remove the gas and quench the dwarf galaxy if it falls in at $z \sim 2$.

Our model does not specify the gas removal mechanism but assumes that feedback-expanded dwarfs have their gas reservoir removed along with the truncation of their gas supply (so-called strangulation) as they fall into clusters, enabling them to quench their star formation and turn into red UDGs. This can occur because hot cluster environment shuts down gas accretion in infalling satellite galaxies (e.g. Kereš et al. 2005, 2009; Simha et al. 2009; van de Voort et al. 2017) while their gas reservoir can be either removed by ram pressure of hot gas or by a feedback episode shortly after infall. Exact nature of gas removal and prevention of further gas accretion will be explored in future work.

In both the major text and Appendix B, we assume instant quenching, since we expect a short quenching time-scale in the cluster environments. Yozin & Bekki (2015) showed the cluster can quench the dwarf galaxy within 2 Gyr with ram pressure stripping. Wetzel et al. (2015) and Fillingham et al. (2015) constrained the quenching times of similarly low-mass ($M_* \sim 10^8$ M$_\odot$) LG satellites to be $\lesssim 3$ Gyr. Quenching in more massive clusters should occur even faster and more efficiently.

4.4 Implications for blue dwarf galaxies

Without accounting for quenching, none of our simulated galaxies end up as red UDGs at the present. However, even without quenching, at $z = 0$ three of our simulated galaxies with

---

[11]Yozin & Bekki (2015) found their galaxies have high axial ratios but they only considered face-on images.

[12]We note that unlike Dragonfly 44, two of our galaxies **m11b** and **m11f** do develop clear stellar discs at late times (without quenching). However, at those times they are not identified as red UDGs.

[13]The average Sérsic indexes of our galaxies are 0.8 ± 0.4.

$M_* \sim 10^8$ M$_\odot$ (**m10z**, **m11a**, **m11b**) have large effective radii, LSBs, and $M_g > -16$ (as shown in Fig. 4), i.e. they satisfy most of the UDG criteria. These diffuse galaxies are much bluer than the red UDGs in van Dokkum et al. (2015a) ($g - i < 0.7$). This implies that there should be a significant population of blue UDG-like dwarfs in the field and at the cluster outskirts with $\mu(g, 0) > 23.5$ mag arcsec^{-2} and $r_{\text{eff}} > 1$ kpc. According to Fig. 4, these blue UDGs have young stellar ages and are spheroidal (axial ratio ~ 0.6) and typically dispersion supported (as shown in El-Badry et al. 2018).

Indeed, there are 'blue' UDGs observed far from galaxy clusters or even in groups, e.g. Martínez-Delgado et al. (2016), Merritt et al. (2016), Román & Trujillo (2017b), Trujillo et al. (2017), and Shi et al. (2017). Pandya et al. (2018) studied one of the UDGs in the sample of Martínez-Delgado et al. (2016) and found younger stellar populations (~ 3 Gyr) and higher metallicity ([$Z/Z_\odot$] ~ -0.6) than UDGs in cluster environments (e.g. Gu et al. 2017).

A similar population of bluer UDGs was recently observed by Román & Trujillo (2017a) outside of the over dense region of the galaxy cluster Abell 168. Compared to UDGs near cluster centres, UDGs in lower density regions have similar effective radii and surface brightnesses, but higher luminosities, bluer colours, and slightly higher stellar masses. Based on these properties, Román & Trujillo (2017a) suggested that blue UDGs could be a LSB extension of regular dwarf galaxies. These observations also showed that unlike other low-redshift galaxies, the stellar mass distribution of UDGs peaks at 10^8 M$_\odot$, coincident with the mass range of the most efficient dynamical effect of stellar feedback (e.g. Governato et al. 2012; Di Cintio et al. 2014; Chan et al. 2015; El-Badry et al. 2016; Tollet et al. 2016) and in a good agreement with our predictions for the properties of field UDGs. It is therefore clear that a large population of galaxies can remain diffuse in the field, owing to the effects of stellar feedback (also suggested in Di Cintio et al. 2017).

Galaxies that, without quenching, reach much higher stellar masses by $z = 0$ (e.g. **m11q**, **m11c**, and **m11f**) are too bright to be included in van Dokkum et al. (2015a) sample. However one of them, **m11q**, is still relatively diffuse at $z = 0$ and could be an example of a more massive but more rare population of UDGs with $M_g < -16$, $\mu(g, 0) \sim 23$–23.5, $r_{\text{eff}} \sim 1$–5 kpc and $g - i < 0.8$. Indeed, several observations, e.g. Mihos et al. (2015), Román & Trujillo (2017b), found brighter examples of UDGs.

Finally, all of our galaxies are simulated as field dwarfs in cosmological simulations without the influence of a cluster, so at $z = 0$ they should resemble field dwarf galaxies. El-Badry et al. (2016) showed that the effective radii of galaxies in FIRE simulations agree with those of the observed galaxies in NASA-Sloan Atlas (Blanton et al. 2011), resembling both the trend and scatter of the sample. In Fig. 8, we compare properties of our simulated galaxies with the sample of nearby field dwarfs from Jansen et al. (2000). We consider our galaxies at $z \sim 0$–0.1 (to account for the star formation and feedback driven size variations) without passive aging, and, following the observations, measure their surface brightnesses $\mu(\text{B}, r_{1/2})$ at 2D half-light radii $r_{1/2}$ in B band.

Given that the observations are in B band, we calculate both the attenuated and unattenuated luminosities of star particles and estimate galactic luminosity, effective radii, and surface brightness for both cases[14] and show results in Fig. 8. For low-mass dwarfs, attenuated and unattenuated values are almost the same so we only
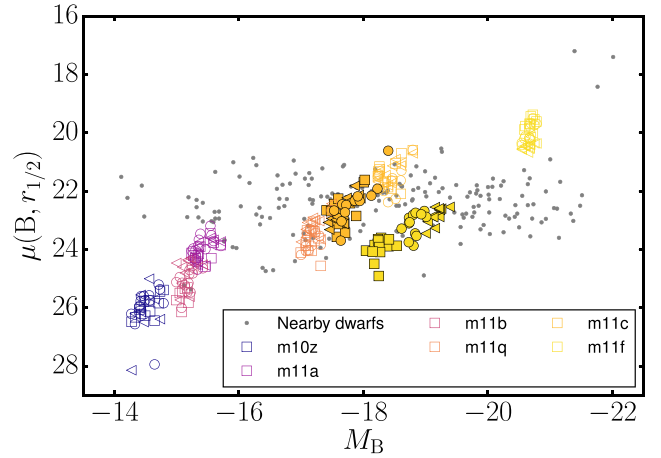
Figure 8. B-band surface brightness, $\mu(B, r_{1/2})$, at a half-light radius $r_{1/2}$ as a function of B-band absolute magnitude, M_B, of our simulated galaxies for a number of simulation output times in the redshift range $z \sim 0$–0.1, compared to nearby galaxies from Jansen et al. (2000). Empty symbols represent unattenuated values whereas solid symbols represent attenuated values. We show different lines of sight with different symbols, in the same manner as Fig. 2.

show attenuated values for our higher mass dwarfs, $M_* \sim 10^9$ M$_\odot$, where differences are significant. The figure shows that our galaxies provide a reasonable match to the observed nearby field dwarfs, although simulated $M_* \sim 10^8$ M$_\odot$ dwarfs tend to be lower surface brightness than the dimmest dwarfs in the observed sample. We note that potential complex selection effects in the observed sample are not taken into account. Jansen et al. (2000) noted that the relative completeness at a given luminosity of their sample, especially at LSB, is not well characterized, leaving out a potentially large population of LSB dwarfs such as the ones in our simulations. Much lower surface brightness galaxies indeed exist in the underdense environment near clusters (e.g. Román & Trujillo 2017a).

In Fig. 9, we compare simulated dwarf galaxies with observed dwarfs in the LG and nearby regions (McConnachie 2012), who use different photometric bands and probe galaxies to lower magnitudes than the Jansen et al. (2000) sample. Following McConnachie (2012), we measure the mean surface brightness within the circular isophote defined by the half-light radius. We do not consider passive aging and attenuation, and show the results at $z = 0$–0.1 to account for the occasional bursts of star formation. The figure shows three of our simulated galaxies whose absolute magnitude overlaps with McConnachie (2012) sample. Our galaxies resemble the trend of the higher mass end of the nearby dwarfs, although simulated galaxies have somewhat lower surface brightnesses. Field dwarfs are slightly higher surface brightness (i.e. more compact) than UDGs of the same absolute magnitude. At V-band magnitude $M_V \sim -14.5$ to -15.5, most of the field dwarfs are diffuse with $r_{\text{eff}} \gtrsim 1$ kpc and have V-band effective surface brightnesses $\langle \mu_V \rangle_e \sim 21$–$23.5$ mag arcsec^{-2}, while our simulated dwarfs have $\langle \mu_V \rangle_e \sim 22.5$–$25$ mag arcsec^{-2}.

While our current sample is too small for a detailed statistical comparison, it does not produce any high surface brightness ana-

---

[14]We assume the gas to dust ratio from Bouchet et al. (1985) scaled by metallicity, the SMC-like dust extinction curves from Pei (1992) and use

the method from Hopkins et al. (2005) to calculate the dust attenuation of stellar light.
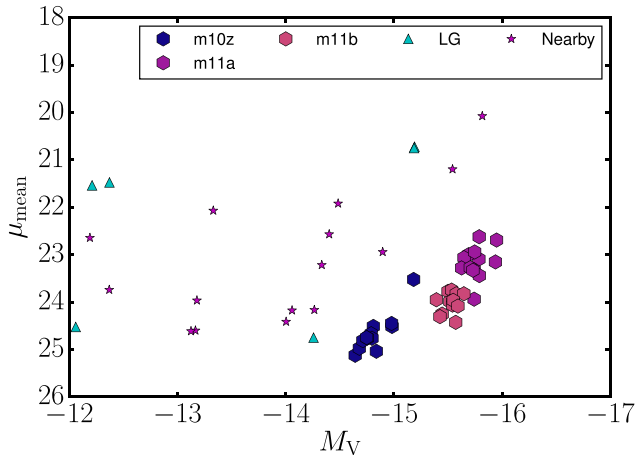
Figure 9. Mean *V*-band surface brightness within the effective radius, μ_{mean}, viewed along *x* direction as a function of *V*-band absolute magnitude M_V for three of our simulated galaxies at $z\sim 0-0.1$, compared to the LG and nearby galaxies from McConnachie (2012). The nearby galaxies are defined as isolated galaxies that do not belong to any major galaxy grouping but are still within 3 Mpc of the Sun.

logues in the relevant mass range $M_* \sim 10^7 - 10^9$ M$_\odot$, which could indicate excessive expansion by feedback. We note, however, that the observational sample is limited by the small survey volume and multiple examples of isolated galaxies with much lower surface brightness (typically by 2–3 mag arcsec^{-2}) than the McConnachie (2012) sample exist (e.g. Dalcanton et al. 1997; Bellazzini et al. 2017; Román & Trujillo 2017b) outside of LG. In the nearby Universe, deep optical follow-ups of H<small>I</small> detected objects (already tested in e.g. Tollerud et al. 2015) or hostless transients due to novae or SNe (Conroy & Bullock 2015), or a large CCD survey (e.g. Dalcanton et al. 1997) could be used to uncover an even larger number of blue UDGs.

Our results suggest that UDG-type surface brightness is the dominant outcome of galaxy formation in low-mass haloes that host galaxies with stellar masses $\lesssim 10^8$ M$_\odot$ ($M_{\mathrm{h}} \lesssim 5 \times 10^{10}$ M$_\odot$), with a caveat that our simulated sample currently contains only a small number of galaxies. While our lower mass examples appear to have lower surface brightness than local observed dwarfs, we note that the full population of low-redshift field galaxies has not yet been properly characterized at very LSB, and that typical observed samples are biased towards high surface brightness. For example, Huang et al. (2012) found that about half of the H<small>I</small> selected dwarf galaxies in their survey, which have typically LSB, do not have a counterpart in the SDSS spectroscopic survey, suggesting that current surveys miss a significant fraction of such objects. More careful analysis of these sources revealed that a large fraction of these galaxies have properties similar to blue UDGs (Leisman et al. 2017). Our simulated galaxies that remain diffuse at $z = 0$ with $\mu(g, 0) > 23.5$ mag arsec^{-2}, (**m10z, m11a** and **m11b**), are all gas rich with corresponding gas fractions, $f_{\mathrm{gas}} = m/(m + m_*)$, of 0.57, 0.55, and 0.9, respectively and could represent such H<small>I</small>-rich UDGs.

Di Cintio et al. (2017) proposed a formation scenario for field UDGs very similar to ours: feedback-driven gas outflows affect stellar profiles in the central regions in the same way as DM core formation, and produce diffuse LSB dwarf galaxies. They considered isolated galaxies in cosmological simulations in the Numerical Investigation of a Hundred Astrophysical Objects (NIHAO)

project (Wang et al. 2015), showing that their dwarf galaxies with $M_{\mathrm{h}} \sim 10$ M$_\odot$ at $z = 0$ can match the surface brightness of observed UDGs, similar to our finding for the field UDG population, despite differences in stellar feedback models and hydrodynamical methods.[15]

Given the burstiness of star formation and resulting outflows in FIRE simulations of dwarf galaxies (Muratov et al. 2015), our simulations make specific predictions for blue UDGs formed by stellar feedback. They should have (a) a range of sizes at fixed stellar mass depending on where they are in their burst cycles, (b) mixed or even inverted age and metallicity gradients (El-Badry et al. 2016), and (c) sizes and velocity dispersions that correlate with their recent star formation history (El-Badry et al. 2017).

5 CONCLUSIONS

We study the origin of UDGs using FIRE-2 cosmological simulations of field dwarf galaxies with halo masses $M_{\mathrm{h}}(z = 0) \sim 10^{10-11}$ M$_\odot$. Our earlier work with the FIRE simulations (El-Badry et al. 2016; Chan et al. 2015) showed that in this halo mass range, stellar feedback and associated changes in the gravitational potential are the most effective in dispersing both DM and stellar populations in the inner halo. In addition, newly formed stars can inherit the velocity of the star forming gas cloud pushed by an outflow episode and further expand the stellar population. Here, we show that these mechanisms lead to a diffuse quasi-spherical stellar distribution with surface brightness and overall properties comparable to observed UDGs.

We then assume that star formation and growth of progenitors of UDGs stop (i.e. galaxies 'quench') during infall into a cluster of galaxies as a combination of tidal and gas stripping processes prevents fresh gas supply and removes the existing gas. To mimic this quenching, we artificially stop star formations of UDG progenitors at a cosmic time t_{q} and passively evolve their stars to $z \sim 0$ according to a stellar population synthesis model (FSPS; Conroy et al. 2009). Finally we generate synthetic images and use GALFIT (Peng et al. 2002) to estimate their central surface brightness and effective radii. Our main findings are summarized below:

All of our simulated galaxies with $M_* \sim 10^7 - 10^8$ M$_\odot$ are diffuse and spatially extended ($\mu(g) \gtrsim 23.5$ mag arcsec^{-2} and $r_{\mathrm{eff}} \gtrsim 1.25$ kpc). These galaxies are typically hosted in haloes with $M_{\mathrm{h}} \sim 3 \times 10^{10} - 1 \times 10^{11}$ M$_\odot$ at their quenching times.

The dynamical effects of stellar feedback produce UDGs even without taking into account cluster influence. Gas removal can help to further expand the galaxies, as shown in Appendix B but this effect is likely of secondary importance.

DM haloes of our simulated galaxies have a typical distribution of spin parameters, suggesting that formation of UDGs does not require high spin haloes.

Red UDGs require quenching of star formation. Our simulations indicate that typical UDGs are dwarf galaxies quenched over a wide range of times ($t_{\mathrm{q}} > 3$ Gyr). Simulated analogues of observed red UDGs that form in haloes $M_{\mathrm{h}} \sim 2 \times 10^{10} - 1 \times 10^{11}$ M$_\odot$ haloes can be quenched over a broad time interval $t_{\mathrm{q}} \sim 5 - 11$ Gyr, i.e. redshift range $z \sim 0.3 - 2$.

---

[15]While Di Cintio et al. (2017) only studied blue field UDGs with simulated isolated or central galaxies fully evolved to $z = 0$, we also account for the effects of quenching, which enables us to more directly address the formation mechanism of red UDGs, commonly observed in clusters.

The most massive red UDGs ($M_\star \gtrsim 2 \times 10^8$ M$_\odot$) in our simulations require earliest quenching. Our higher mass haloes can host red UDGs if their star formation and growth is quenched at very early times, $t_q \sim 3-5$ Gyr, i.e. at $z_q \sim 2-3$. At the time of quenching, the host halo mass of our most massive simulated UDG, similar to e.g. Dragonfly 44, is around 10^{11} M$_\odot$.

Colours of red UDGs are approximately independent of quenching time as galaxies quenched later (i.e. with a younger stellar population) typically have higher metallicity. This prevents an accurate estimation of quenching time from $g - i$ colour.

Galaxies with $M_* \lesssim 10^8$ M$_\odot$ remain diffuse even at $z = 0$ but have relatively blue colours. We predict that diffuse galaxies with bluer colours ($g - i < 0.8$) are prevalent in the field. Our galaxies at $z = 0$ match the magnitude–surface brightness relations of some samples of nearby galaxies, but have lower surface brightness than the LG sample from McConnachie (2012). While our sample is small and statistics are limited, this raises an interesting prospect that there is a large number of undiscovered LSB galaxies in the Universe.

Given the UDG formation process in our simulations, the size and velocity dispersion of 'blue' UDGs at a fixed mass should correlate with their recent star formation history.

Our galaxies are not hosted in overmassive haloes ($M_*/M_h \lesssim 10^{-4}$) at quenching, but instead have a stellar-to-halo mass ratios similar to observed dwarfs, $M_*/M_h \sim 10^{-3}$.

The enclosed masses of our simulated galaxies can match the measured masses of observed UDGs in clusters, if we assume that the growth of their central enclosed mass stopped when they were quenched. Even if such haloes had evolved in isolation outside of clusters, we showed that in most cases their enclosed DM mass on these scales would remain unchanged, owing to halo growth largely by 'pseudo-evolution'. Owing to a broad range of formation times of these objects, inferred halo masses at $z = 0$ can therefore lead to misleading conclusions about their host halo masses. This is especially important for the most massive UDGs whose dense central regions formed at very early times.

ACKNOWLEDGEMENTS

REFERENCES

Amorisco N. C., Loeb A., 2016, MNRAS, 459, L51
Amorisco N. C., Monachesi A., Agnello A., White S. D. M., 2018, MNRAS, 475, 4235
Beasley M. A., Trujillo I., 2016, ApJ, 830, 23
Beasley M. A., Romanowsky A. J., Pota V., Navarro I. M., Martinez Delgado D., Neyer F., Deich A. L., 2016, ApJ, 819, L20
Bellazzini M., Belokurov V., Magrini L., Fraternali F., Testa V., Beccari G., Marchetti A., Carini R., 2017, MNRAS, 467, 3751
Blanton M. R., Kazin E., Muna D., Weaver B. A., Price-Whelan A., 2011, AJ, 142, 31
Bothun G. D., Impey C. D., Malin D. F., 1991, ApJ, 376, 404
Bouchet P., Lequeux J., Maurice E., Prevot L., Prevot-Burnichon M. L., 1985, A&A, 149, 330
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Burkert A., 2017, ApJ, 838, 93
Caldwell N., 2006, ApJ, 651, 822
Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C.-A., Quataert E., 2015, MNRAS, 454, 2981
Conroy C., Bullock J. S., 2015, ApJ, 805, L2
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Dalcanton J. J., Spergel D. N., Gunn J. E., Schmidt M., Schneider D. P., 1997, AJ, 114, 635
Di Cintio A., Brook C. B., Macciò A. V., Stinson G. S., Knebe A., Dutton A. A., Wadsley J., 2014, MNRAS, 437, 415
Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, MNRAS, 466, L1
Diemer B., More S., Kravtsov A. V., 2013, ApJ, 766, 25
El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131
El-Badry K., Wetzel A. R., Geha M., Quataert E., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2017, ApJ, 835, 193
El-Badry K. et al., 2018, MNRAS, 473, 1930
El-Zant A., Shlosman I., Hoffman Y., 2001, ApJ, 560, 636
Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Ferland G. J. et al., 2013, Rev. Mex. Astron. Astrofis. 49, 137
Fillingham S. P., Cooper M. C., Wheeler C., Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., 2015, MNRAS, 454, 2039
Fitts A. et al., 2017, MNRAS, 471, 3547
Gavazzi G., Donati A., Cucciati O., Sabatini S., Boselli A., Davies J., Zibetti S., 2005, A&A, 430, 411
Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16
González-Samaniego A., Bullock J. S., Boylan-Kolchin M., Fitts A., Elbert O. D., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., 2017, MNRAS, 472, 4786
Governato F. et al., 2010, Nature, 463, 203
Governato F. et al., 2012, MNRAS, 422, 1231
Gu M. et al., 2018, ApJ, 859, 37
Harris W. E., Harris G. L. H., Alessi M., 2013, ApJ, 772, 82
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., Hernquist L., Martini P., Cox T. J., Robertson B., Di Matteo T., Springel V., 2005, ApJ, 625, L71
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., 2017, preprint (arXiv:1702.06148)

Huang S., Haynes M. P., Giovanelli R., Brinchmann J., Stierwalt S., Neff S. G., 2012, AJ, 143, 133

Impey C., Bothun G., Malin D., 1988, ApJ, 330, 634

Jansen R. A., Franx M., Fabricant D., Caldwell N., 2000, ApJS, 126, 271

Janssens S., Abraham R., Brodie J., Forbes D., Romanowsky A. J., van Dokkum P., 2017, ApJ, 839, L17

Kadowaki J., Zaritsky D., Donnerstein R. L., 2017, ApJ, 838, L21

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kim J.-h. et al., 2014, ApJS, 210, 14

Klimentowski J., Łokas E. L., Kazantzidis S., Prada F., Mayer L., Mamon G. A., 2007, MNRAS, 378, 353

Klimentowski J., Łokas E. L., Kazantzidis S., Mayer L., Mamon G. A., 2009, MNRAS, 397, 2015

Knebe A., Green A., Binney J., 2001, MNRAS, 325, 845

Knollmann S. R., Knebe A., 2009, ApJS, 182, 608

Koda J., Yagi M., Yamanoi H., Komiyama Y., 2015, ApJ, 807, L2

Kroupa P., 2002, Science, 295, 82

Lee M. G., Kang J., Lee J. H., Jang I. S., 2017, ApJ, 844, 157

Leisman L. et al., 2017, ApJ, 842, 133

Leitherer C. et al., 1999, ApJS, 123, 3

Li Y., Mo H. J., van den Bosch F. C., Lin W. P., 2007, MNRAS, 379, 689

Łokas E. L., Semczuk M., Gajda G., D'Onghia E., 2015, ApJ, 810, 100

Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140

Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940

Macciò A. V., Stinson G., Brook C. B., Wadsley J., Couchman H. M. P., Shen S., Gibson B. K., Quinn T., 2012, ApJ, 744, L9

Makarov D. I., Sharina M. E., Karachentseva V. E., Karachentsev I. D., 2015, A&A, 581, A82

Marigo P., Girardi L., 2007, A&A, 469, 239

Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883

Martin C., Moore A., Morrissey P., Matuszewski M., Rahman S., Adkins S., Epps H., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham, p. 77350M

Martínez-Delgado D. et al., 2016, AJ, 151, 96

Mayer L., Governato F., Colpi M., Moore B., Quinn T., Wadsley J., Stadel J., Lake G., 2001, ApJ, 547, L123

McConnachie A. W., 2012, AJ, 144, 4

McConnachie A. W. et al., 2008, ApJ, 688, 1009

Merritt A., van Dokkum P., Danieli S., Abraham R., Zhang J., Karachentsev I. D., Makarova L. N., 2016, ApJ, 833, 168

Mihos J. C. et al., 2015, ApJ, 809, L21

Mowla L., van Dokkum P., Merritt A., Abraham R., Yagi M., Koda J., 2017, ApJ, 851, 27

Muñoz R. P. et al., 2015, ApJ, 813, L15

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

Pandya V. et al., 2018, ApJ, 858, 29

Pei Y. C., 1992, ApJ, 395, 130

Peñarrubia J., Pontzen A., Walker M. G., Koposov S. E., 2012, ApJ, 759, L42

Peng E. W., Lim S., 2016, ApJ, 822, L31

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Pontzen A., Governato F., 2014, Nature, 506, 171

Price D. J., Monaghan J. J., 2007, MNRAS, 374, 1347

Read J. I., Gilmore G., 2005, MNRAS, 356, 107

Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006, MNRAS, 366, 429

Román J., Trujillo I., 2017a, MNRAS, 468, 703

Román J., Trujillo I., 2017b, MNRAS, 468, 4039

Rong Y., Guo Q., Gao L., Liao S., Xie L., Puzia T. H., Sun S., Pan J., 2017, MNRAS, 470, 4231

Shi D. D. et al., 2017, ApJ, 846, 26

Sifón C., van der Burg R. F. J., Hoekstra H., Muzzin A., Herbonnet R., 2018, MNRAS, 473, 3747

Simha V., Weinberg D. H., Davé R., Gnedin O. Y., Katz N., Kereš D., 2009, MNRAS, 399, 650

Springel V., 2005, MNRAS, 364, 1105

Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, MNRAS, 428, 129

Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, MNRAS, 429, 3068

Tollerud E. J., Geha M. C., Grcevich J., Putman M. E., Stern D., 2015, ApJ, 798, L21

Tollet E. et al., 2016, MNRAS, 456, 3542

Trujillo I., Roman J., Filho M., Sánchez Almeida J., 2017, ApJ, 836, 191

van de Voort F., Bahé Y. M., Bower R. G., Correa C. A., Crain R. A., Schaye J., Theuns T., 2017, MNRAS, 466, 3460

van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713

van der Burg R. F. J., Muzzin A., Hoekstra H., 2016, A&A, 590, A20

van der Burg R. F. J. et al., 2017, A&A, 607, A79

van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015a, ApJ, 798, L45

van Dokkum P. G. et al., 2015b, ApJ, 804, L26

van Dokkum P. et al., 2016, ApJ, 828, L6

van Dokkum P. et al., 2017, ApJ, 844, L11

Venhola A. et al., 2017, A&A, 608, A142

Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, MNRAS, 454, 83

Watkins L. L., Evans N. W., An J. H., 2010, MNRAS, 406, 264

Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336

Wetzel A. R., Tollerud E. J., Weisz D. R., 2015, ApJ, 808, L27

Wheeler C. et al., 2017, MNRAS, 465, 2420

Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Muñoz R. R., Simon J. D., Avedo F. F., 2010, MNRAS, 406, 1220

Yozin C., Bekki K., 2015, MNRAS, 452, 937

Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, ApJ, 771, L19

APPENDIX A: galfit WITH SKY BACKGROUND

Realistic galaxy image contains sky background in addition to galactic light, so a sky subtraction is required. A proper subtraction is not trivial for galaxies with surface brightnesses comparable to the background, e.g. UDGs. To gauge the potential impact of sky background on the estimated properties of UDGs, we generate simulated galaxy images along with stochastic sky backgrounds whose average surface brightness is $\sim$26 mag arcsec^{-2}. Then we estimate their central surface brightness and effective radius with a two-component fit, assuming Sérsic profiles for galaxies and tilted flat planes for sky backgrounds. Fig. A1 shows that the differences in central surface brightness and effective radius with and without sky backgrounds are small, even when the sky is, on average, brighter than the galaxy, illustrating the robustness of the fitting. The fitted r_{eff} differs from Fig. 4 because here we use g-band instead of $g + i$ images.
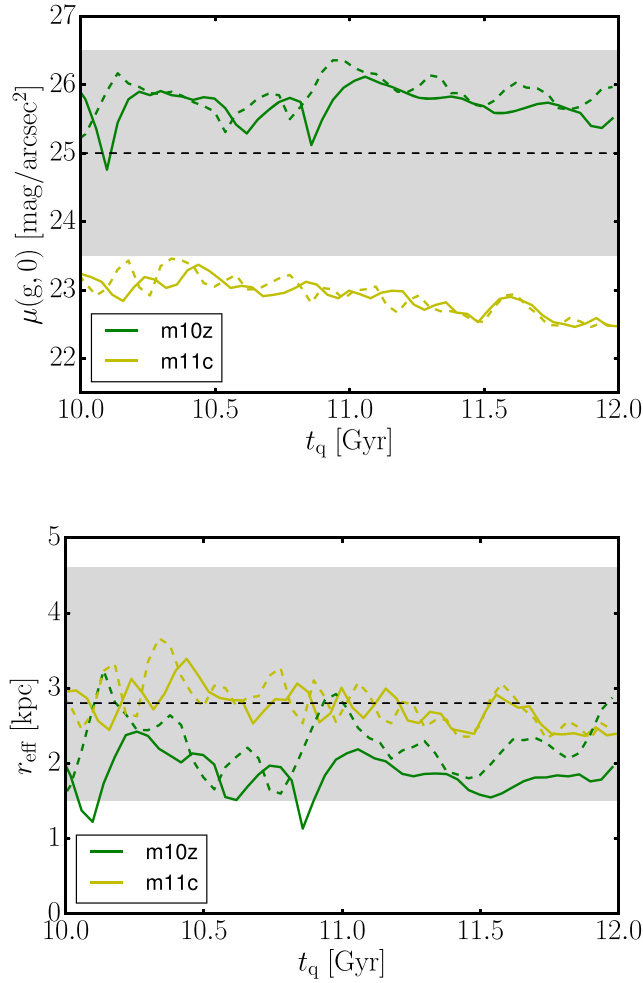
Figure A1. Time evolution of central *g*-band surface brightnesses $\mu(g, 0)$ and effective radii of two of our simulated galaxies with a sky background (dashed) and without (solid) in their *g*-band images. The sky background is a random noise with the averaged surface brightness around 26 mag arcsec^{-2}. $\mu(g, 0)$ is obtained with GALFIT. Sky background and galaxy are fitted simultaneously. Stars are passively aged to $z = 0$ and dust attenuation is not considered. Including a typical sky background has only a small effect on the estimated values.

APPENDIX B: DYNAMICAL EFFECT OF GAS REMOVAL

Most of the observed red UDGs are detected in galaxy clusters and were probably quenched through interactions with the host cluster. They do not show signatures of tidal interaction (Mowla et al. 2017) but it is possible that their gas was removed by ram pressure stripping. To simulate the effect of gas removal on a dwarf galaxy, we test a simple toy model introduced in El-Badry et al. (2017): all gas particles instantaneously receive 1000 km s^{-1} velocity boosts at a given infall/stripping time. Then we evolve the galaxy to $z = 0$ (by continuing the run in a fully cosmological environment) and estimate its properties with GALFIT as described in the main text.

Since the gas velocity after the kick is much higher than the escape velocity, all of the gas is quickly removed and the galaxy is quenched after around 100 Myr. El-Badry et al. (2016) tested this method and concluded that the effect is almost identical to instantaneously taking out all of the gas particles. Fast moving particles also affect the surrounding gas so the galaxy can never
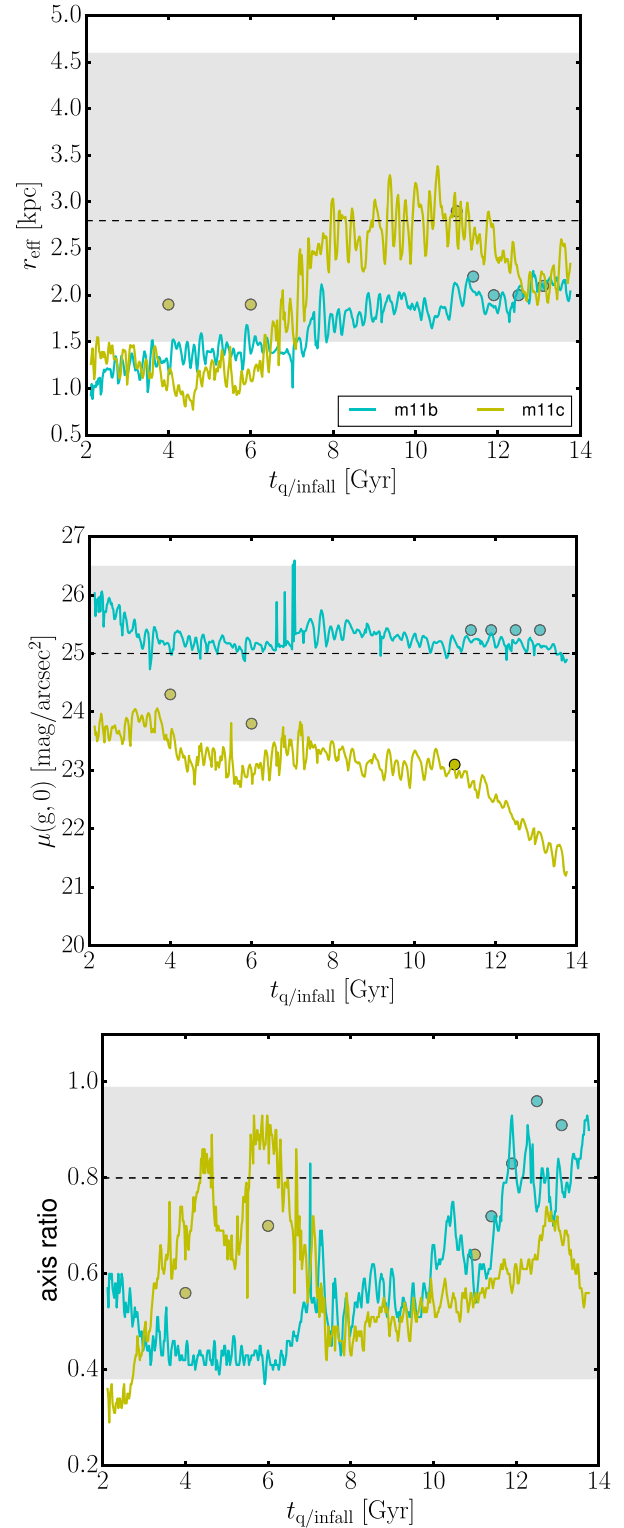
Figure B1. Evolution of the effective radius, central surface brightness, and axial ratio for two of our simulated galaxies. Lines show values from passive evolution scenario described in the main text, whereas points show values from gas-stripped **m11b** and **m11c** dynamically evolved to $z = 0$. Lines are functions of quenching time t_q, while individual points indicate several different times for which we apply our ram pressure approximation, mimicking the effect of hot cluster gas during the infall. Galaxies are typically quenched shortly ($\sim$100 Myr) after gas stripping.
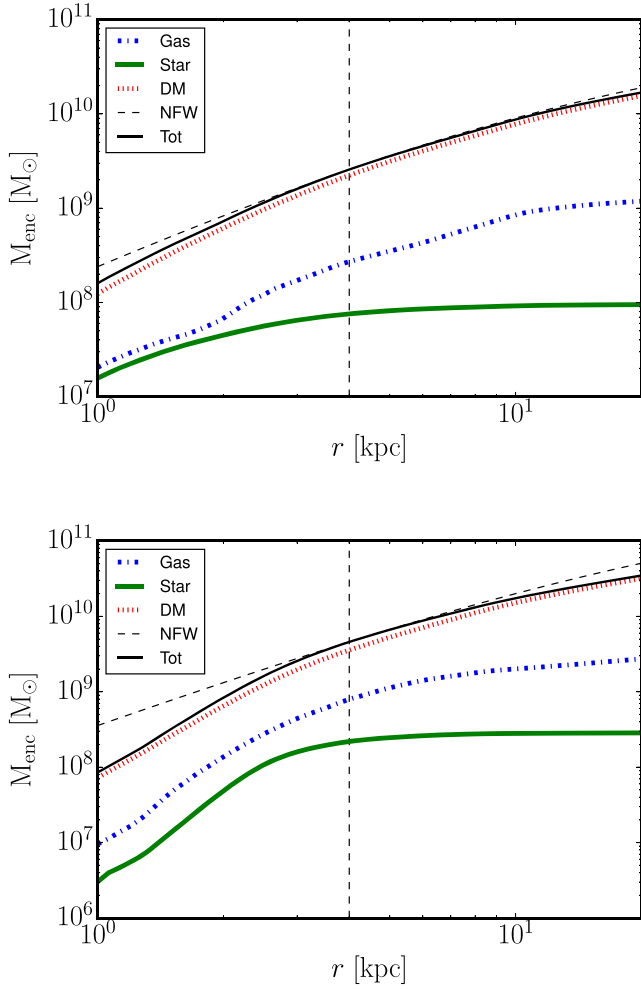
Figure B2. Cumulative gas (blue; dash–dotted), star (green; solid), DM (red; dotted), and total (black; solid) mass profiles of **m11b** at $t = 11$ Gyr (Upper) and **m11c** at $t= 5$ Gyr. Dashed vertical lines show the effective radii and black dashed curved lines show the NFW profiles whose enclosed masses within the effective radii match our haloes. DM dominates enclosed mass throughout the halo for both galaxies.

accrete new gas and gets quenched. While in our default approach we only passively quench star formation and do not allow star particles to move after quenching, here they can freely move and adapt to a new and shallower gravitational potential.

Fig. B1 shows that the effect of gas removal is small, since dynamical relaxation after gas stripping induces only a slight increase in size and a slight drop in central surface brightness, while there is no clear systematic effect on the axial ratio. The additional dynamical effects of gas stripping (compared to the fiducial model) therefore only help making our simulated galaxies slightly more diffuse in the relevant mass range. These results also show that galaxy sizes do not evolve much after ram pressure stripping and remain largely 'frozen' in time motivating our approach of passively evolving galaxies after the quenching time. Overall both galaxies stay within or outside of the observational range of UDGs even when gas stripping is applied.

Relatively weak effect of gas removal is not entirely surprising given that galaxies are already largely spheroidal and dark matter dominated. While the gas dominates baryonic component, its gravitational influence is much weaker than that from the dark matter. This is illustrated in Fig. B2 where we show the enclosed mass profiles for stars, gas, and dark matter in the inner halo for the two galaxies for which we apply quenching. Profiles are shown at characteristic times for which passively evolved counterparts correspond to the observed red UDGs. Dark matter dominates at all radii.

APPENDIX C: EFFECT OF RESOLUTION

To evaluate the effect of resolution, we compare two of our galaxies, **m10z** and **m11c**, with their higher resolution versions, presented in Hopkins et al. (2017) (these have particle masses $m_{gas} = 260$ M$_\odot$ and $m_{gas} = 2.1 \times 10^3$ M$_\odot$, respectively). The high-resolution versions have particle masses eight times smaller and softening lengths twice shorter than those shown in Table 1. While their halo masses do not change with resolution, their stellar masses at $z = 0$ drop by 40 per cent and 60 per cent in **m10z** and **m11c**, respectively.

Fig. C1 shows the difference in central surface brightness, effective radius, and g-band magnitude between two resolutions. The sizes of the galaxies are not sensitive to resolution, but their g-band magnitudes decrease by $\sim$1, as expected from their smaller stellar masses. Their surface brightness also drops accordingly. These changes are consistent with what we expect from stellar mass difference between lower and higher resolution galaxies. We therefore conclude that while resolution can affect the stellar masses of our simulated galaxies, the surface brightness and size at a given stellar mass are largely not affected by resolution.
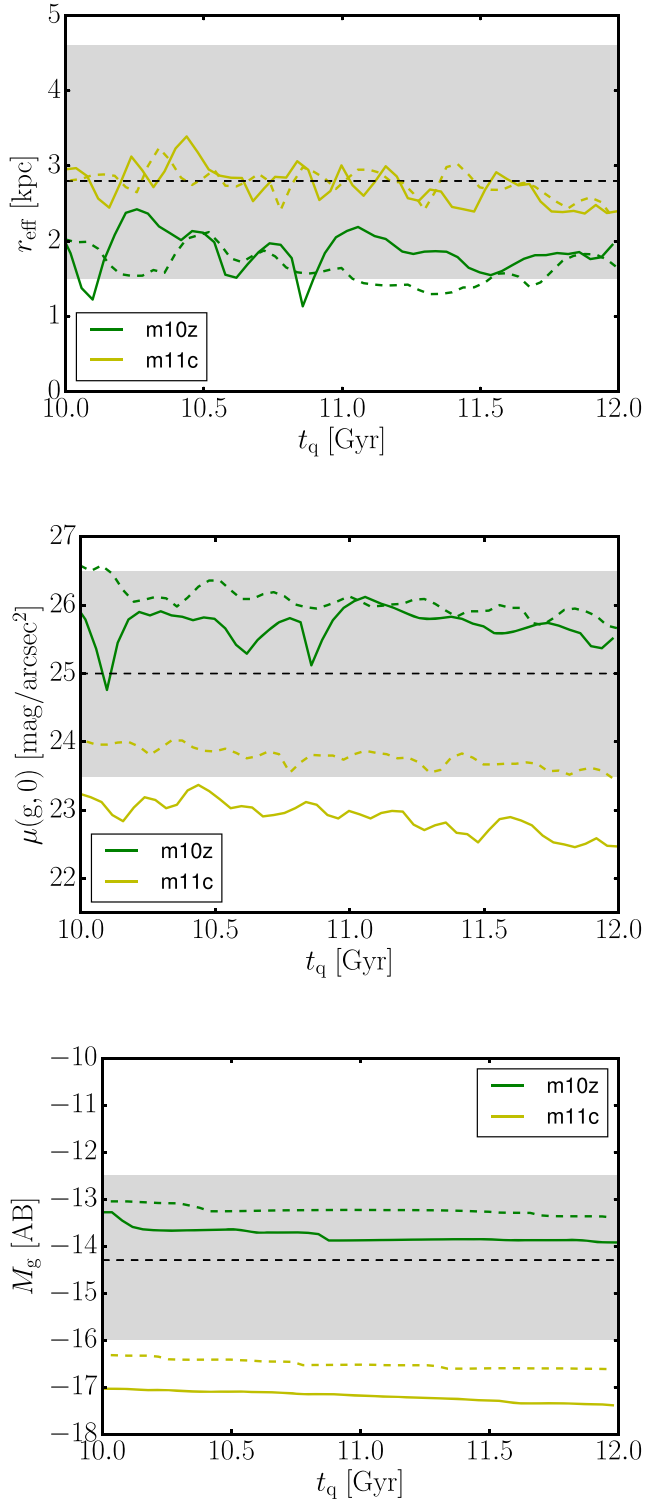
Figure C1. Time evolution of central *g*-band surface brightness $\mu(g, 0)$ and effective radius for the two simulated galaxies with a higher resolution (dashed) and with the fiducial resolution (solid). Values are obtained from *g*-band images using GALFIT.

APPENDIX D: GALFIT MODELLING WITH DIFFERENT RESOLUTION IMAGES

Stellar particles in simulated galaxies have well determined positions and represent a population of stars with 'radii' (i.e. gravitational softening) that is typically much finer than typical resolution from observations. To mimic a range of point spread functions (PSF) of different telescopes and a range of distances at which one observes UDGs, we vary the pixel size of our GALFIT images (i.e. the resolution of the 2Dprojection of stellar properties) of our **m11b** galaxy from 40 to 400 pc, and show results in Fig. D1. This
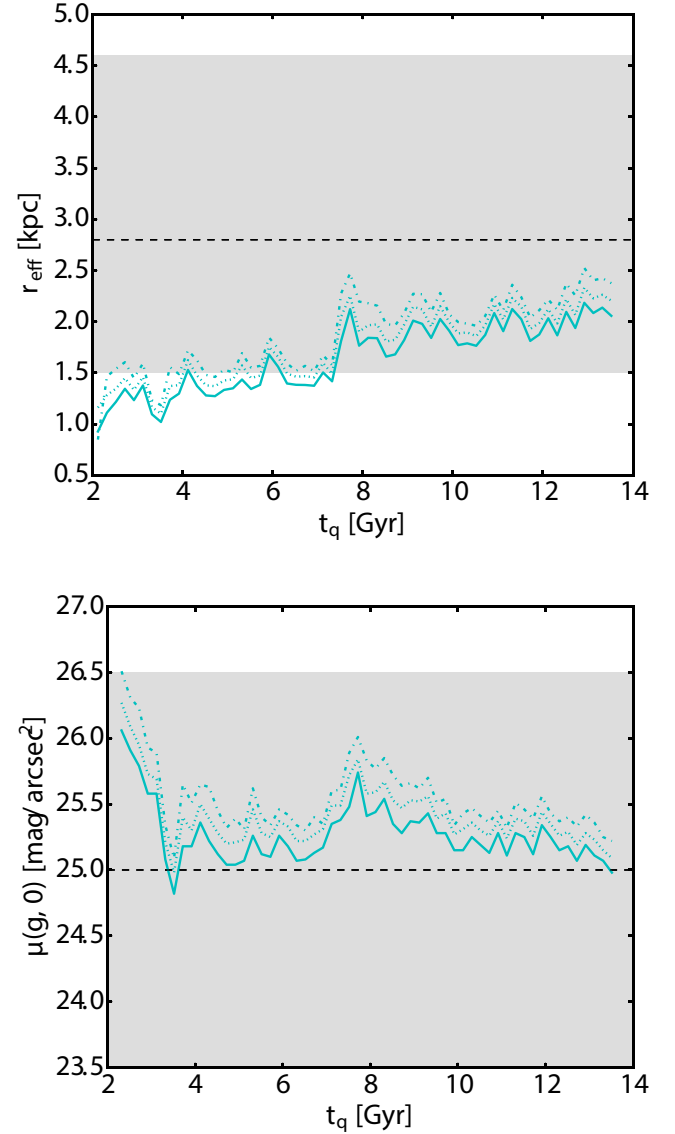
Figure D1. Time evolutions of the effective radius and surface brightness of **m11b** as calculated by GALFIT from 40 kpc × 40 kpc *g*-band images with uniform-size pixels each with 400 pc (dash–dotted), 200 pc (dot), and 40 pc (solid) on a side (i.e. 100× 100, 500 × 500,and 1000 × 1000 pixels).

approximately spans the range of PSF between the *Hubble Space Telescope* and the Canada–France–Hawaii Telescope (CFHT) at the distance of the Coma cluster (i.e. 0.08–0.8 arcsec for the assumed distance of 100 Mpc). For the tested range, there is a small systematic increase in effective radii and a slight decrease in surface brightness in low-resolution images, but the change is smaller than the short-term time variations of these properties.

This paper has been typeset from a TEX/LATEX file prepared by the author.