Giant low surface brightness galaxies in TNG100

Qirong Zhu,[1][*] Luis Enrique Pérez-Montaño ®,[2] Vicente Rodriguez-Gomez ®,[2]
Bernardo Cervantes Sodi ®,[2] Jolanta Zjupa,[3] Federico Marinacci ®,[4] Mark Vogelsberger ®[5]
and Lars Hernquist[6]

[1]*McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, 15213 PA, USA*
[2]*Instituto de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, Apdo. Postal 72-3, 58089 Morelia, Mexico*
[3]*Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, D-52425 Jülich, Germany*
[4]*Department of Physics and Astronomy 'Augusto Righi,' University of Bologna, via Gobetti 93/2, I-40129 Bologna, Italy*
[5]*Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[6]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*

ABSTRACT

Giant low surface brightness (GLSB) galaxies, such as Malin 1 and UGC 1382, contain the largest stellar discs known. GLSB galaxies also often contain large masses of neutral hydrogen (H I). However, these extreme galaxies' origin and properties remain poorly understood. Using the cosmological simulation IllustrisTNG 100, we identify and select a sample of ∼200 galaxies with extended ($R_{\rm H\,I} > 50$ kpc) and well-defined H I discs, ∼6 per cent of the total galaxies in the same stellar mass range ($10.2 < \log(M_*/{\rm M_\odot}) < 11.6$). This GLSB sample is heterogeneous, with mixed galaxy morphologies ranging from the most disc-dominated systems to massive ellipticals. These simulated GLSB galaxies are located in massive haloes ($V_{\rm max} > 150$ km s^{-1}) and their properties, such as total H I content, stellar disc parameters, star formation rate, and rotation curves, agree with observed GLSB galaxies. We construct a paired control sample to contrast with the GLSB galaxies. The GLSB galaxies tend to have large galaxy spin parameters (40 per cent larger) and larger *ex situ* stellar mass fractions than the paired control. We find evidence that aligned mergers promote the formation of extended discs and that isolated environments help the survival of those discs across cosmic time.

Key words: (cosmology:) dark matter – methods: numerical – galaxies: evolution.

1 INTRODUCTION

Freeman ([1970]) reported a constant central surface brightness (μ_0) in spiral galaxies, as well as in the disc components of S0s, with only a small spread in his sample. A constant μ_0 would suggest some common origin or regulatory processes in the galaxy discs. However, the apparent narrow distribution of μ_0 can be an artifact of selection biases in the limiting surface brightness due to sky brightness and instruments, as first noted by Arp ([1966]), and later demonstrated by Disney ([1976]) and Disney & Phillipps ([1983]). Discs with μ_0 substantially dimmer than the canonical value of 21.6 *B*-mag arcsec^{-2} (Freeman [1970]) could be missed by the surveys. But discs with μ_0 much brighter than 21.6 appear to be truly absent (e.g. Bosma & Freeman [1993]; Fathi [2010]), while the discs with μ_0 much dimmer than 21.6 have been discovered and are represented by a population of *low surface brightness (LSB) galaxies*.

Malin 1, a large galaxy characterized by its extended stellar disc, is one of the prime examples of an LSB galaxy. Before Malin 1, some S0/lenticular galaxies were found to be peculiar with faint outer stellar discs and large masses of neutral hydrogen (e.g. Gallagher [1979]; Hawarden et al. [1981]; Romanishin, Strom & Strom [1983];

Gottesman & Hawarden [1986]). But Malin 1 is definitive proof of the missing population of galaxies postulated by Disney ([1976]) and Disney & Phillipps ([1983]). In the two discovery papers (Bothun et al. [1987]; Impey & Bothun [1989]), it was unclear what morphological type Malin 1 belongs to. Decades later, we know that Malin 1 is a massive galaxy containing a stellar disc five times larger than that of the Milky Way (Boissier et al. [2016]), and its 'normal' bulge was not revealed until Barth ([2007]).

LSB galaxies are not exclusively low-mass dwarf galaxies, as they span the same range of physical parameters as the galaxies that define the conventional Hubble sequence (Schombert et al. [1992]; McGaugh, Bothun & Schombert [1995]; Honey et al. [2018]). Among the LSB galaxies, the 'crouching giants' (Disney & Phillipps [1987]) like Malin 1 are genuinely rare. These *giant low surface brightness (GLSB) galaxies* will be the focus of this work.

GLSB galaxies are critical, as these objects connect with other fields in galactic astronomy. A surprising fact about GLSB galaxies is their enormous neutral hydrogen mass (O'Neil et al. [2004]; Mishra et al. [2017]), which puts them among the most massive H I systems. These discs provide a rare opportunity to directly measure the mass distribution in the outer region of the haloes (Lelli, Fraternali & Sancisi [2010]). GLSB galaxies are also important in the context of star formation in the low-density environment (e.g. O'Neil, Oey & Bothun [2007]; Rahman et al. [2007]; Thilker et al. [2007]; Das, Boone &
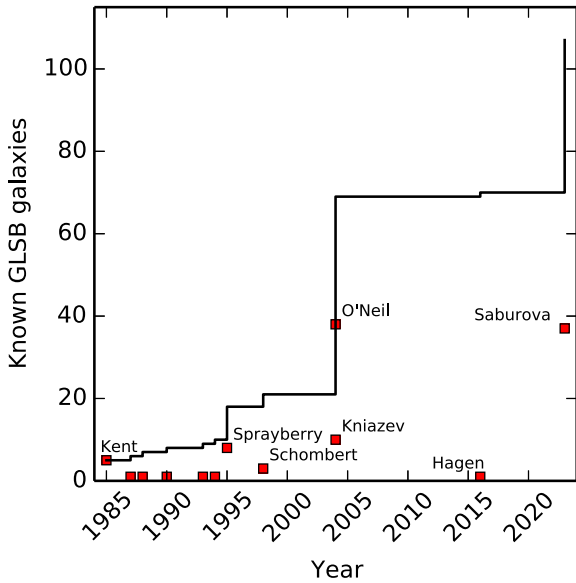
Figure 1. A summary of GLSB galaxy discoveries since 1985. Some authors are listed next to the discovery year, while the full list can be found in the main text. The total number of GLSB galaxies is 58 prior to 2004. Saburova et al. (2023) adds 37 new candidates, bringing the total number of GLSB galaxies up to 107 as of 2023.

Viallefond 2010), and may contain clues on how supermassive black hole coevolve with the host galaxies (Schombert 1998; Das et al. 2007, 2009; Mishra et al. 2015).

Fig. 1 summarizes a total of 107 GLSB galaxies from the literature as of year 2023 (Kent 1985; Bothun et al. 1987; Davies, Phillipps & Disney 1988; Bothun et al. 1990; McGaugh & Bothun 1994; Sprayberry et al. 1995; Schombert 1998; Kniazev et al. 2004; O'Neil et al. 2004; Hagen et al. 2016; Saburova et al. 2023). We have not found further follow-up work on most candidates in O'Neil et al. (2004). Special pipelines for searching for GLSB galaxies in these deep surveys might be necessary, as Kniazev et al. (2004) pointed out. Nevertheless, it is fair to state that a small sample size limits our current understanding of GLSB galaxies. The most recent addition of observed GLSB galaxies by Saburova et al. (2023) is based on visual inspecting deep Subaru Hyper Suprime-Cam data. We note that different definitions of GLSB galaxies are adopted by various authors.

While scarce, the extraordinary nature of GLSB galaxies has motivated substantial interest in their origin. Many mechanisms have been proposed over the years to explain the formation of GLSB galaxies, including:

(i) GLSB galaxies are extreme late-type galaxies, consuming gas to form stars at a slower rate than in normal galaxies (Bothun et al. 1987);

(ii) GLSB galaxies formed from rare (3σ) density peak within low-density environments (Hoffman, Silk & Wyse 1992);

(iii) They form in high spin dark matter haloes (Dalcanton, Spergel & Summers 1997);

(iv) Disc instabilities causing the material to migrate outwards (Noguchi 2001);

(v) Accretion of satellite galaxies (Peñarrubia, McConnachie & Babul 2006);

(vi) A result of head-on collisions as in Mapelli et al. (2008);

(vii) GLSB galaxies form in massive and rarefied dark matter haloes, hence shallower gravitational potential wells than normal galaxies (Kasparova et al. 2014).

While these formation scenarios are not mutually exclusive, it is unclear, which may play a dominant role. A pressing issue with GLSB galaxies is their potential incompatibility with the current hierarchical growth of galaxies in ΛCDM, characterized by frequent mergers (e.g. Fakhouri, Ma & Boylan-Kolchin 2010; Rodriguez-Gomez et al. 2015). Therefore, taken at face value, the un-evolved, dynamically fragile disc structure that survived many Gyrs without being disturbed could be difficult to reconcile with our current understanding of galaxies.

In the context of numerical simulations of galaxies, there are several important reasons to focus on rare galaxy populations such as GLSB galaxies. First, the sub-grid models of star/BH formation and feedback employed in the modern numerical simulations are calibrated against 'normal' galaxies. Therefore, GLSB galaxies serve as one critical subject to cross-validate these models. Examples of such cross-validations include jellyfish galaxies (Yun et al. 2019) and ultra-diffuse galaxies (Benavides et al. 2021). Secondly, the previous modellings of GLSB galaxies involve idealized set-ups of galaxy orbits and disc properties, which may not arise naturally within a fully cosmological context. Lastly, and perhaps most importantly, it is time to address a commonly held belief that mergers of GLSB galaxies with other galaxies pose severe challenges to the current framework of galaxy formation and evolution within ΛCDM (e.g. Boissier et al. 2016; Hagen et al. 2016; Saburova et al. 2021).

The rarity of GLSB galaxies implies that, to simulate these galaxies, a large cosmological box is required. The size of the cosmological box is merely necessary, not sufficient. Before the current generation of hydrodynamic simulations (see a recent review by Vogelsberger et al. 2020), previous efforts suffered from well-known overcooling problems and excessive angular momentum transfer, and failed to reproduce stellar discs as in late-type galaxies (e.g. Scannapieco et al. 2012). Recently, several studies (Di Cintio et al. 2019; Martin et al. 2019; Kulier et al. 2020; Pérez-Montaño et al. 2022) have been carried out to understand LSB galaxies in numerical simulations. In particular, using the large cosmological EAGLE simulation (Schaye et al. 2015), Kulier et al. (2020) find that a subset of mergers produce the most extensive LSB galaxies with stellar discs comparable in size to observed GLSB galaxies.

In previous work, Zhu et al. (2018) identified a galaxy with extended gas and stellar discs similar to Malin 1 in the Illustris TNG 100 simulation. The origin of the cold gas is from gas stripped from a pair of in-falling galaxies on to a more massive central. Much of the cold gas mass is also gained due to the cooling of the existing hot halo gas behind elongated cold gas streams resulting from the interactions between the galaxy pair. Mixing cold and hot gas accelerates the cooling process, producing a $10^{11} M_\odot$ cold gas disc with a spatial extent of over 200 kpc. The gas disc is rotationally supported, with a circular velocity higher than 400 km s^{-1}.

Is the galaxy in Zhu et al. (2018) just one exceptional object? Or is there a population of galaxies sharing similar properties in the simulation? This work explores the TNG100 data set looking for galaxies with similar properties regarding extended H I and stellar discs. The paper is organized as follows. We detail the methods of identifying and characterizing GLSB galaxies in Section 2 and present the main results in Section 3, concentrating on the galaxy properties. We then discuss the implications of our results, focusing on the formation history of GLSB galaxies in TNG100, compared

with the known formation scenarios in Section 4. Finally, a short list of the conclusions is summarized in Section 5.

2 METHODS

2.1 IllustrisTNG 100 simulation

The IllustrisTNG project (Pillepich et al. 2018b; Springel et al. 2018; Nelson et al. 2018; Naiman et al. 2018; Marinacci et al. 2018) consists of a series of cosmological MHD simulations using the AREPO code (Springel 2010; Weinberger, Springel & Pakmor 2020). In addition to gravity and magnetohydrodynamics, gas cooling, star formation and feedback, black hole (BH) growth and AGN feedback are all included as the essential ingredients. The IllustrisTNG simulations build upon the original Illustris project (Vogelsberger et al. 2014a, b; Genel et al. 2014; Sijacki et al. 2015) through a number of physical and numerical improvements (Weinberger et al. 2017; Pillepich et al. 2018a). In particular, the radio mode of AGN feedback has been replaced by a kinetic outflow launched from the BH particles, instead of the original radio bubbles (Sijacki et al. 2007, 2015; Vogelsberger et al. 2013). This particular modification is crucial to rectify the low gas fraction in the massive halos as reported by Genel et al. (2014).

We use the TNG100 simulation with a box of side length $75h^{-1}$ Mpc in this work. The baryonic mass resolution is 1.4×10^6 $M_\odot$ with a physical gravitational softening length for star and BH particles of 0.7 kpc at $z = 0$ and an adaptive softening for gas cells with a minimum of 0.185 co-moving kpc. We use galaxies from the SUBFIND catalogue, which already contains basic properties such as the total mass of each component: star, gas, BH, and dark matter, as well as photometries for star particles. A galaxy in our study is defined as a subhalo identified by the SUBFIND group finder (Springel et al. 2001; Dolag et al. 2009).

Additional important galaxy properties, if not contained in the SUBFIND catalogues, are provided by supplementary catalogues. In particular, halo spins are calculated with the same procedure as in Zjupa & Springel (2017). Galaxy merger histories are derived using the same method in Rodriguez-Gomez et al. (2015). For galaxy morphology, we adopt the kinematic decomposition in Marinacci, Pakmor & Springel (2014) and Genel et al. (2015), with a parameter $f(\epsilon_{>0.7})$[1] that denotes the mass fraction of disc stars. For more technical details, we refer to these papers. Additionally, we calculate the rotation curves based on their mass distributions as $v_{rot,i}^2 = GM_i(r)/r$, where $M_i(r)$ denotes the radial cumulative mass of each component.

2.2 GLSB and control samples selection

The GLSB candidate list is constructed as below. The extent of the H I distribution should be large, and we enforce $R_{HI} > 50$ kpc, where R_{HI} is measured where the surface mass density drops below 1 $M_\odot$ pc^{-2}, following the same definition as in e.g. Broeils & Rhee (1997) and Wang et al. (2016). Due to the H I size–mass relation (e.g. Wang et al. 2016), this naturally implies a large total H I mass. Based on the measured H I mass–size relation, the above choice translates into a total H I mass above 10^{10} $M_\odot$, consistent with the measurement of GLSB galaxies by Pickering et al. (1997) and O'Neil et al. (2004). These candidates are then verified based on their surface photometry.

Neither the H I properties nor the surface photometries are given in the main catalogue; therefore, we adopt a straightforward approach to these quantities.

2.2.1 Neutral hydrogen mass

The interstellar medium (ISM) in IllustrisTNG is treated using the Springel & Hernquist (2003) model. Feedback from supernovae is absorbed into an effective equation of state for the star-forming gas, with the kinetic effects of supernovae modelled with a galactic outflow. Therefore, the neutral hydrogen mass comes from both the star-forming and non-star-forming gas. We estimate the neutral gas fraction for each gas cell using two methods outlined in Marinacci et al. (2017). We briefly summarize the approach here.

The amount of neutral atomic hydrogen contained in each Voronoi cell is estimated according to

$$M_{HI,i} = (1 - f_{mol,i}) f_{neutr,i} X_i M_i, \tag{1}$$

with X_i the hydrogen mass fraction and M_i the total mass of the cell i. The neutral gas fraction $f_{neutr,i}$ and the molecular hydrogen mass fraction $f_{mol,i}$ are not directly found in the snapshots but need some special attention. In particular, for non-star-forming gas cells, we use the $f_{neutr,i}$ based on the gas cooling functions. For star-forming gas cells, all the cold gas is assumed to be neutral, therefore tied to the cold mass fraction x according to the definition in Springel & Hernquist (2003):

$$x = \frac{u_h - u}{u_h - u_c}, \tag{2}$$

where u_h, u_c, and u are the specific thermal energy of the hot phase, the cold phase, and the gas cell.

For cold gas, the mass in the molecular form needs to be subtracted from the total to get the neutral atomic hydrogen. Using a fitting formula by Leroy et al. (2008), the molecular fraction is given by

$$f_{mol,i} = \frac{(P/P_0)^\alpha}{(P/P_0)^\alpha + 1}, \tag{3}$$

where P is the cold gas pressure (total pressure in the cell multiplied by x), $P_0 = 1.7 \times 10^4$ K cm^{-3}, and a power-law index α of 0.8. We also have compared H I mass with the method by Bird et al. (2014), as used in Vogelsberger et al. (2014b), and found these two methods give overall consistent results with modest differences found mainly in the low-mass end. In Bird et al. (2014), a different estimate of molecular fraction based on n_H (Altay et al. 2011) is adopted. On average, the method in Bird et al. (2014) leads to 12 per cent larger H I mass at $M_{HI} = 10^9 M_\odot$ than that in Marinacci et al. (2017). At $M_{HI} = 10^{10} M_\odot$, the differences are ~5 per cent. We note that major uncertainties remain in the estimation of molecular gas masses from cosmological simulations (Diemer et al. 2018; Stevens et al. 2021), and the method by Leroy et al. (2008) would produce un-physical results when applied based on a cell-by-cell basis. Fortunately, the impact to R_{HI} would be minimal as the cold gas is mostly atomic anyway.

From visual inspection, some galaxies show strongly warped features (Semczuk et al. 2020), which renders any H I disc indiscernible. Therefore, we remove the objects from our candidate list if they do not contain a clear cold gas disc distribution. We compare the gas distribution with the derived R_{HI} value and remove those showing apparent differences. For instance, some galaxies undergoing close mergers can attain $R_{HI} > 50$ kpc, but they will not be considered GLSB candidates. Some galaxies may contain giant 'holes' (due to AGN kinetic feedback, see Nelson et al. 2021) in their H I images,

---

[1]Here, ϵ is defined for every star particle by $\epsilon \equiv J_z/J(E)$, with J_z the specific angular moment around the symmetry axis and $J(E)$ the maximum specific angular momentum possible given the specific binding energy E.
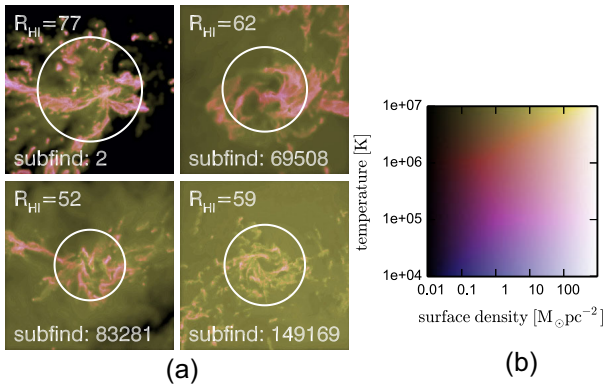
Figure 2. (a) Examples of galaxies excluded from the GLSB candidate sample due to the lack of a clear-defined H I disc. For these galaxies, the calculated $R_{\rm H\,I}$ values generally are inconsistent with the overall H I gas distributions. (b) Two-dimensional colour table that maps the gas surface density and temperature into the brightness and colour hue for the corresponding pixels.

and they are removed as well. Fig. 2 shows four galaxies in such categories. This additional requirement reduces the total number of candidates from 302 to 203.

2.2.2 Mock galaxy images and surface photometry fitting

The IllustrisTNG simulations output all the information necessary to conduct surface photometry studies. We create mock galaxy images with the procedure outlined in Torrey et al. (2015). The luminosity of each stellar particle follows the single stellar population (SSP) model of Bruzual & Charlot (2003). The radiation from each stellar particle is smoothed over its 24 nearest neighbours. Besides adaptive smoothing, no dust attenuation is applied, nor is any sky background added. Then the *U*-, *B*-, and *K*-band light is mapped into blue, green, and red channels.

With the mock images of each galaxy, we then apply a procedure to calculate the extended disc scale length. We restrict the fitting range to be brighter than 30 mag arcsec^{-2}, which is still deeper than most current and future surveys (Abraham & van Dokkum 2014). Fainter structures are thought to be contributed by minor mergers in the form of stellar streams (e.g. Johnston et al. 2008).

Photometric fittings for LSB galaxies differ significantly between different studies in the literature. For instance, Beijersbergen, de Blok & van der Hulst (1999) and Kniazev et al. (2004) find that the two-disc model describes some GLSB galaxies quite well. Saburova et al. (2018) used a 'bulge + disc' decomposition for UGC 1922. In Saburova et al. (2023), the authors fit a single exponential disc for the component between 25.1 and 27.6 mag arcsec^{-2} in *g* band to derive the disc properties. Hagen et al. (2016) uses a more complex three-component model for UGC 1382. In this work, we will use a two-component model following

$$\Sigma(r) = I_e \exp\left\{-b_n\left[\left(\frac{r}{r_e}\right)^{1/n} - 1\right]\right\} + I_{\rm out}\exp\left(-r/r_{\rm d}\right), \qquad (4)$$

as the default ('Sérsic + disc'). The impact of adopting alternative models are studied in Appendix A.

Surface photometry $\Sigma(r)$ is constructed using circular apertures without adding any additional sky background or noise. Major uncertainties come from the choice of an SSP and the coarse sampling of young star particles in the otherwise underlying smooth stellar

distribution. Therefore, the above procedure represents the most favourable condition to find the extended stellar discs among the entire simulated galaxy sample.

2.2.3 A paired control sample

Galaxies are intrinsically complex systems characterized by multiple attributes such as their stellar mass, gas mass, dark matter distribution, angular momentum, and assembly histories. The primary drivers for galaxy properties are the masses of individual mass components, such as stellar mass. Following Patton et al. (2016), we create a control sample, consisting of those galaxies sharing similar dark matter, stellar and *total* gas masses with the GLSB candidates, in order to study the differences between the two samples.

To this end, we construct a neighbour list of galaxies in the parameter space of $\{\log(M_{\rm dm}), \log(M_{\rm star}), \log(M_{\rm gas})\}$ using the l^2 norm as a distance metric. From the neighbour list output by SCIKIT-LEARN, we then search for the nearest neighbour of each GLSB candidate. We further require that the paired control sample only consists of unique members, i.e. containing no repeated entries. This additional requirement occasionally forces us to adopt the second or the third nearest neighbour for a few GLSB candidates. We end up with a control sample with the same number of galaxies ('paired control') as the GLSB candidates.

Additionally, we constructed a sample of TNG100 galaxies with their stellar mass within the range of the GLSB galaxies, with $10.2 < \log(M_*/M_\odot) < 11.6$. This leads to a sample of 3086 galaxies, which will be termed 'all galaxies'. The GLSB sample and its control sample are removed from the list. This sample is not controlled by any other galaxy properties, which represents 'normal' TNG100 galaxies. It is beneficial for us to gain insights into any special feature present in the GLSB and the paired control samples.

Also, we do not impose any constraints on whether or not the galaxy is the primary subhalo of any FOF group. Hence, satellites are included both in our GLSB and control samples. It is useful to note that the fraction of 'central' in the three samples are 0.85 (GLSB), 0.94 (paired control), and 0.79 (all galaxies), respectively. The requirement of pairs leads to a more selective sample for the paired control, such that the fraction of satellites is the least among the three.

3 RESULTS

3.1 Extended gas discs

The total number of GLSB candidates we identified at $z = 0$ is 203. Fig. 3 displays the face-on gas distribution of 16 GLSB galaxies. In each panel, the cold H I gas is shown in blue while the tenuous hot halo gas is in red. The size of the H I disc, $R_{\rm H\,I}$, is marked with a solid circle for each galaxy. The side length for each panel in Fig. 3 is 250 kpc. Both the visual impression and the $R_{\rm H\,I}$ are consistent with our goal of finding GLSB candidates hosting extended H I gas. The atomic gas is also mostly confined in a disc plane. For the galaxies shown here, there are noticeable asymmetries in the gas distribution outside their $R_{\rm H\,I}$. Those one-arm-shaped gas streams, as in SUBFIND 52622 and 506823 are telling features of the accretion origin of the cold gas.

Fig. 4 compares the H I properties of the GLSB and the control sample with the observed LSB galaxies with H I measurements from Matthews, van Driel & Monnier-Ragaigne (2001) and the HIX sample from Lutz et al. (2018). The size–mass relation derived by Wang et al. (2016) is shown in the solid thick line, using 500 nearby
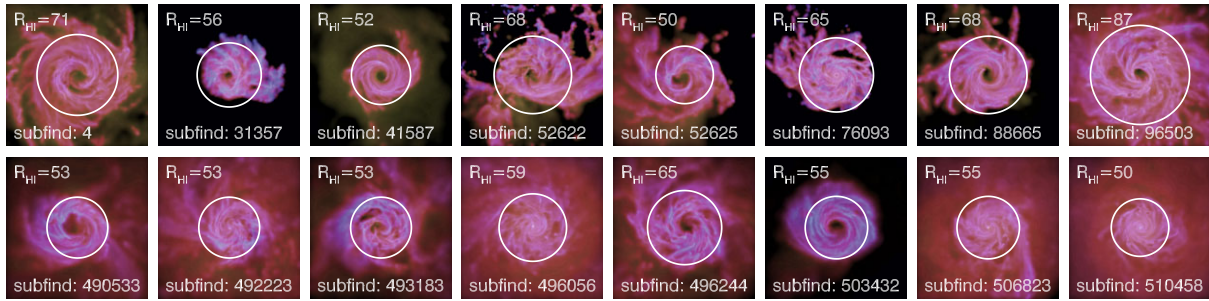
Figure 3. Face-on view of extended cold H I discs in GLSB candidates selected from TNG100. Only 16 out of 203 candidates are displayed here. The side length for each panel is 250 kpc. The size of H I disc is marked and specified for each galaxy. The same colour table is adopted here as in Fig. 2.
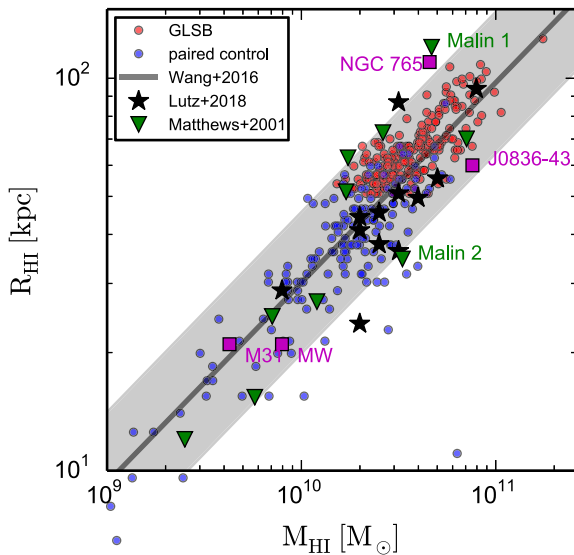

Figure 4 H I size–mass relation for the GLSB and paired control samples compared with observations. We include GLSB galaxies such as Malin 1 and Malin 2 reported by Matthews et al. (2001), the HIX sample by Lutz et al. (2018), together with the size–mass relation derived by Wang et al. (2016), where the shaded grey region corresponds to the 3σ uncertainties (0.18 dex). Some notable individual galaxies in Wang et al. (2016) are also included as well.

galaxies spanning across five orders of magnitudes in H I mass. We include the individual objects considered in Wang et al. (2016) as well, with NGC 765 (Portas et al. 2010), J0836-43 (Donley et al. 2006), MW (Kalberla & Kerp 2009), and M31 (Chemin, Carignan & Foster 2009).

The overall H I properties of both GLSB and the paired control sample agree with observations. The simulated galaxies follow the same relation given by Wang et al. (2016). The scatter is also confined within 3σ uncertainties (0.18 dex) from Wang et al. (2016), shown in the shaded grey region in Fig. 4. Not surprisingly, all the GLSB candidates are larger than 50 kpc in $R_{\rm H\,I}$, while the ones in the paired control sample are considerably smaller in $R_{\rm H\,I}$.

In Fig. 4, there is some overlap between the GLSB and the paired control sample. This is a consequence of requiring that the overall cold gas distribution in GLSB galaxies is in a well-defined disc. Galaxies with ongoing merger events are excluded from GLSB from visual inspection. Some of those H I rich galaxies end up in the paired control. Nevertheless, it is interesting to note that the largest

Table 1. Comparison of molecular fraction $f_{\rm mol} = M_{\rm H_2}/(M_{\rm H_2} + M_{\rm H\,I})$ of different galaxy samples.

Value name	GLSB (per cent)	Paired control (per cent)	All galaxies (per cent)
16th percentile	0.08	0.0	0.0
Median	0.3	0.5	0.8
84th percentile	0.9	1.8	5.2
Mean	0.4	1.1	2.4

H I masses are all found in the GLSB sample. The total H I mass in the GLSB sample spans from $2 \times 10^{10}\,{\rm M_\odot}$ to $\sim 10^{11}\,{\rm M_\odot}$. In contrast, the H I mass in the control sample extends below $10^{10}\,{\rm M_\odot}$.

Table 1 summarizes the molecular fraction for the GLSB, paired control, and 'all galaxies' samples. The molecular fraction based on the mid-plane pressure is known to be an underestimate (Diemer et al. 2018). Nevertheless, it is helpful to study the differences in molecular fractions. The median/average value is about half for the GLSB sample than the paired control. The distribution of the molecular fraction is also narrower for GLSB than the other. Given the larger H I mass in the GLSB sample, it appears that the density of H I is actually lower than the paired control.

Previously, O'Neil & Schinnerer (2004) and Cao et al. (2017) have reported lower molecular fractions in GLSB galaxies than their high surface brightness (HSB) counterparts. Interestingly, the attempts to detect CO in Malin 1 have been largely unsuccessful since Braine, Herpin & Radford (2000). Galaz et al. (2022) put an upper limit of 13 per cent on the molecular fraction for Malin 1. With a more sophisticated treatment of molecular hydrogen formation than this work, Diemer et al. (2019) reported that the molecular fraction of the Malin 1 analogue in Zhu et al. (2018) is between 3 per cent and 10 per cent, which is consistent with Galaz et al. (2022).

3.2 Extended stellar discs

Fig. 5 compares the galaxy size using r_{28} for three samples from TNG100, with r_{28} the 28 mag arcsec^{-2} isophote in B band. r_{28} is adopted by Kulier et al. (2020) and Saburova et al. (2023) to quantify the size of GLSB galaxies.

On average, the GLSB sample contains the largest galaxies indicated by r_{28}, with a median value of ~ 57 kpc. The paired control sample has a median r_{28} of 40 kpc. In contrast, the median r_{28} for 'all galaxies' is 23 kpc. This much smaller r_{28} for 'all galaxies' can be attributed to the mass function. Both the GLSB and paired control sample contain a higher fraction of massive galaxies than the 'all galaxies' sample.
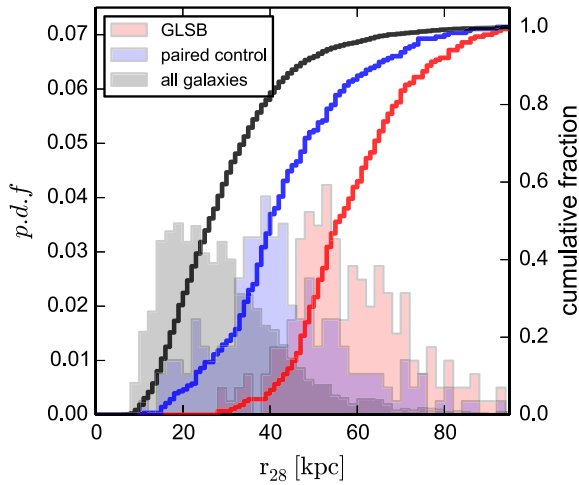
Figure 5. Distributions of r_{28} or the GLSB, paired control, and 'all galaxies' samples. The solid lines show the cumulative fractions. Overall, GLSB galaxies have the largest size, followed by the paired control and 'all galaxies' samples.

Are the GLSB galaxies larger due to an extended disc component? To see this, Fig. 6 shows the mock images of face-on and edge-on views of the same galaxies in the GLSB sample. These galaxies are selected from a sorted SUBFIND ID list, containing the first and the last eight entries. Visual impressions for the full GLSB sample are included in Appendix C.

Each galaxy is rotated according to the axis perpendicular to its H I disc. The side length of each image is also 250 kpc. In the top half, the limiting surface brightness in U, B, and K bands are set at 25, 25, and 22 mag arcsec^{-2}. In the blue band, the same 25 mag arcsec^{-2} is used for the operational definition of galaxy size, r_{25}.

In the lower half of Fig. 6, the limiting surface brightness of the same galaxies is extended down to three more magnitudes, mimicking the view from current deep imaging observations (e.g. Galaz et al. 2015; Saburova et al. 2023). However, the limiting surface brightness is still much brighter than the typical values for the stellar halo (e.g. Johnston et al. 2008; Cooper et al. 2010). The faint blue structure in the outer region in the upper half turns into extended structures at a much larger scale. The deep images reveal not only the faint spheroidals, but also faint but large discs, which would otherwise appear as sporadic blue dots in the shallow images. It is also evident that the disc is more preferentially associated with blue light.

Even with only the 16 galaxies shown here, it is clear that our GLSB sample is a heterogeneous population of different colours and different morphological types. In the shallow images, one can readily identify early-type galaxies, disc galaxies with red colours (Tacchella et al. 2019), and disc galaxies with blue colours. In the deep images, one can then find stellar discs within large early-type galaxies, e.g. SUBFIND 4 and SUBFIND 41587. On the other hand, one can also find extended stellar discs with little spheroidal components, such as SUBFIND 496056 and SUBFIND 496244. Recall the side length of each panel is 250 kpc. Therefore, the extent of the stellar discs is remarkable, which is intimately related to the large r_{28} of the GLSB sample shown in Fig. 5.

In the process of making the images, all the GLSB candidates are rotated in the coordinates defined by their H I disc. The edge-on view of the galaxies here also shows that the stellar discs are more or less in the same plane as the gas disc, which suggests that these two components are closely connected. In the literature, B-band images

are widely used to define galaxy morphologies. Working in B band reduces the impact of older stellar populations (Buta 2013). We can also see that the blue light in deep images of Fig. 6 is preferentially located in the disc structure. We will further examine this connection later.

Fig. 7 shows the azimuthally averaged radial profile for the same 16 GLSB candidates. For each galaxy, the scale length r_d and $\mu_0(B)$, the extrapolated B-band central surface brightness at $r = 0$, and σ, the root mean square error (RMSE) of the surface photometry fitting of the outer disc is listed as well.

The derived disc parameters for our GLSB sample are consistent with the observed GLSB galaxies. All of the galaxies here have $\mu_B(r = 0) > 23$ mag arcsec^{-2}. In addition, the size of the outer disc, r_d, is all larger than 10 kpc. If we adopt the definition of GLSB in Saburova et al. (2023), all galaxies in this figure can be classified as GLSB. The quality of the fitting varies for different galaxies. The fitting errors for some galaxies, such as SUBFIND 52622, 88665, and 503432, are small. Galaxies with large RMSE exhibit some downward break in their surface photometry at large galactic radii, e.g. SUBFIND 76093, 493183, and 506823. A broken-exponential profile (Erwin 2015) could offer better fits for these galaxies.

We apply the same photometric fitting procedure to the GLSB and the paired control samples. For some galaxies in GLSB sample, and for many objects in the paired control sample, it appears that a single component is dominant in the entire range. In these cases, the outer disc component is barely present at best, and a single Sérsic component could suffice. Therefore, we require the outer disc component to be at least 2 mag brighter than the inner Sérsic profile at r_{28}. This additional requirement allows us only to include galaxies with reliable measurements of outer discs. The total number of galaxies passing this cut is 99 and 171 for the two samples considered.

Fig. 8 compares the distribution of the stellar disc parameters from our above fitting procedure with previous measurements in the literature (van der Kruit 1987; de Jong & van der Kruit 1994; McGaugh & Bothun 1994; Sprayberry et al. 1995). We further use the definition of GLSB discs according to Sprayberry et al. (1995),

$$\mu_B(0) + 5 \log r_d > 27.0, \tag{5}$$

where $\mu_B(0) = \mu_B(r = 0)$. The units for $\mu_B(0)$ and r_d are mag arcsec^{-2} and kpc, respectively.

The outer disc of the GLSB galaxy sample falls into the same parameter space occupied by known GLSB galaxies (McGaugh & Bothun 1994; Sprayberry et al. 1995), as well as the gap between Malin 2 and the two extended discs, UGC 1382 (Hagen et al. 2016) and Malin 1. Only four galaxies in our GLSB sample are slightly on the left of the division according to equation (5). For the paired control sample, ~ 55 per cent (55) of galaxies in the paired control sample are consistent with GLSB discs, while the rest are not. Recently, Saburova et al. (2023) have compared the disc scale-length distributions in both the HSC field and the Eagle simulations and concluded that giant discs represent the large-sized end of the volume density distribution of normal-sized spirals. The distribution of the disc parameters for the paired control sample in Fig. 8 is consistent with this assessment.

The total number of large discs contributed by the two samples is thus 167 and 55, respectively. The latter number indicates that our selection criteria could be too strict such that some galaxies in the paired control sample could be safely deemed as GLSB instead. Nevertheless, most of the extended discs in Fig. 8 are contributed by the GLSB sample. Moreover, the majority of the GLSB sample indeed hosts extended discs. Therefore, we will use the same 203 galaxies in the GLSB and the paired control sample without further refinement for our following analysis.
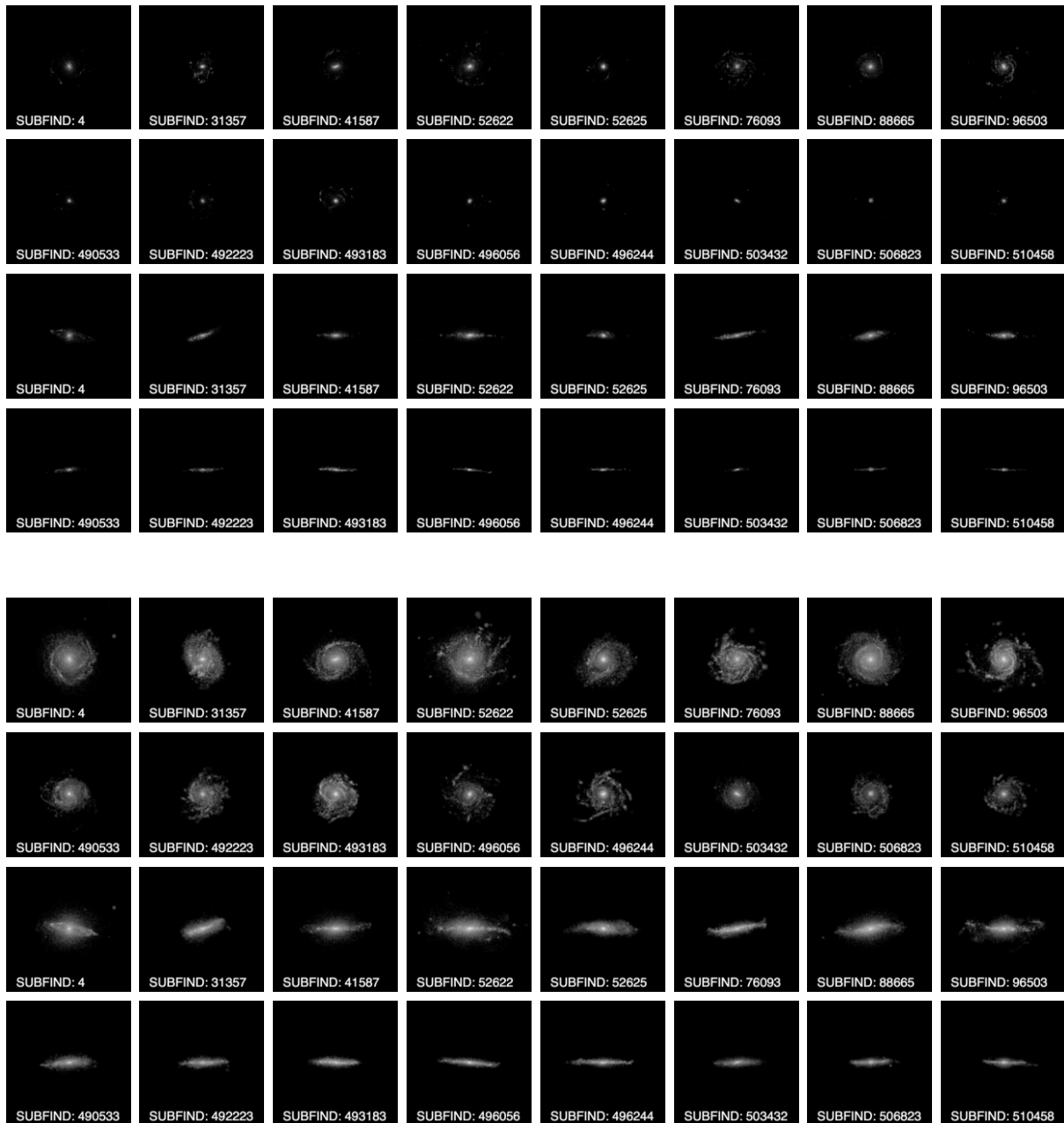
Figure 6. Mock images for 16 GLSB candidates (both face-on and edge-on view) in normal and deep imaging modes. The top panel adopts limiting surface brightness in U, B, and K bands at 25, 25, and 22 mag arcsec^{-2}, while the lower panel shows the same 16 galaxies but with three magnitudes deeper.

3.3 Main galaxy properties

Next, we study the main galaxy properties for our GLSB and control sample. We start with the total dark matter and total stellar mass in the top panel of Fig. 9. Both the samples follow the general trend of all TNG100 galaxies shown in the parameter space here. GLSB galaxies are mostly massive systems with total dark matter mass above 10^{12} M$_\odot$ and stellar mass above 10^{10} M$_\odot$. Hence, these are not in the regime of dwarf galaxies.

The centre panel compares the gas fraction, $f_{\mathrm{gas}} = M_{\mathrm{gas}}/(M_{\mathrm{gas}} + M_*)$, between the two samples. As the control sample is constructed by finding the closest neighbour measured in $\{\log(M_{\mathrm{DM}}), \log(M_{\mathrm{gas}}), \log(M_*)\}$ space, the overall agreement between the two distributions

in the centre panel shows that even only considering baryonic components, the two populations are close in the sense of their gas fractions. As a result, it is beneficial for us to focus on the differences in other galaxy properties instead of worrying about the impact of the major galaxy properties such as f_{gas}.

The lower panel of Fig. 9 displays the distribution of maximum rotational velocity V_{max} and $f_{\epsilon > 0.7}$, where $f_{\epsilon > 0.7}$ is the fraction of stars with $\epsilon > 0.7$. For every star particle, ϵ is the ratio between the specific angular momentum along the symmetry axis J_z and the maximum specific angular momentum possible at the specific binding energy E. V_{max} is a measure of the total mass distribution, while the latter is a good indicator of galaxy morphological type.
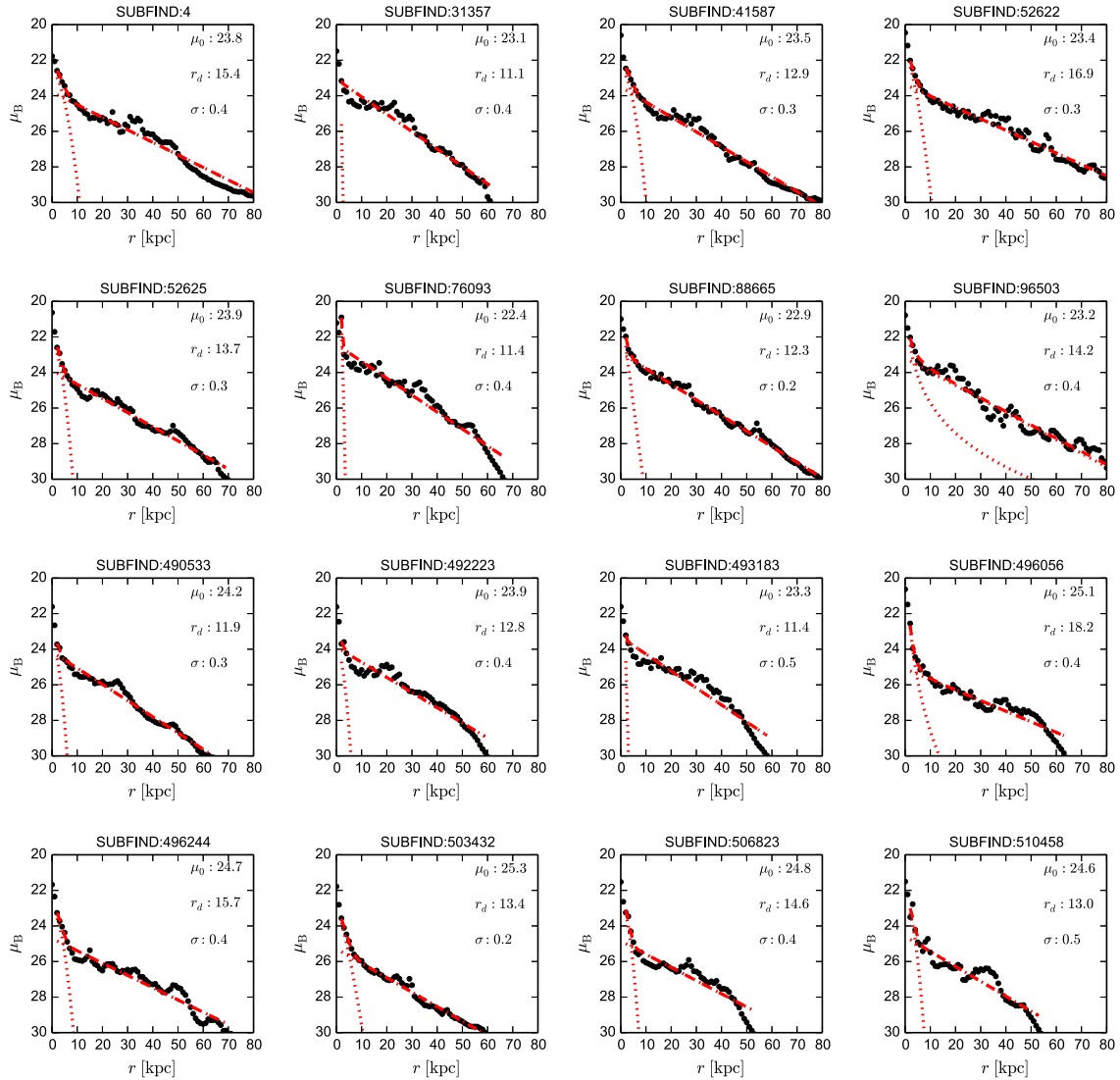
Figure 7. GLSB galaxy *B*-band surface photometry and fitting results. The total light follows a sum of an inner bulge and an outer exponential disc, as shown in the dotted and dashed red lines. The outer discs are extended ($r_d > 10$ kpc) with low central surface brightness $\mu_B(0) > 23$ mag arcsec^{-2}.

The entire TNG100 population is also shown here in the form of a heat map. We can find a general trend that $f_{\epsilon > 0.7}$ peaks at $V_{max} = 200$ km s^{-1}, and drops toward both higher and lower V_{max}.

The drop of $f_{\epsilon > 0.7}$ with $V_{max} \geq 200$ km s^{-1} reflects the overall transition from late-type galaxies to early-type galaxies as a function of the total mass. The minimum and maximum of V_{max} for the GLSB sample are 150 and 500 km s^{-1}, respectively. This mass span shows that the GLSB galaxies consist of both the most disc-dominated and the more massive early-type galaxies in TNG100. As a result, this figure confirms our earlier impression from the mock galaxy images that GLSB galaxies are a heterogeneous population with distinct galaxy morphologies.

The lack of any GLSB with $V_{max} < 150$ km s^{-1} is a result of a strict cut in $R_{\mathrm{H\,I}} = 50$ kpc. Some galaxies with smaller H I disc could also be classified as GLSB galaxies if their outer disc parameters satisfy equation (5) according to Sprayberry et al. (1995). Therefore our sample, as shown in Fig. 8 is incomplete if we only consider the definition of GLSB galaxies based on the H I disc. As a result, the

total number of GLSB galaxies in our sample should be interpreted as a *lower limit*.

We can probably gain a rough idea of how many GLSB galaxies could be missing due to our approach. In Pérez-Montaño et al. (2022), the mean effective *r*-band surface brightness is adopted as the defining property for LSB galaxies. Using the catalogue by Pérez-Montaño et al. (2022), we find that 92 out of 203 GLSB galaxies are also identified as LSB galaxies for their effective surface brightness. By comparison, only 37 in the paired control sample are recognized as LSB galaxies in both definitions. The effective *r* band $\mu(r)$ is 21.74 ± 1.08 for GLSB galaxies and 21.03 ± 1.15 mag arcsec^{-2} for the control. We note our method is able to identify the majority of GLSB galaxies.

3.4 Colour–magnitude diagram and star formation

It is now well-known that GLSB galaxies often contain HSB inner components. The inner disc of Malin 1 was reported by Barth (2007).
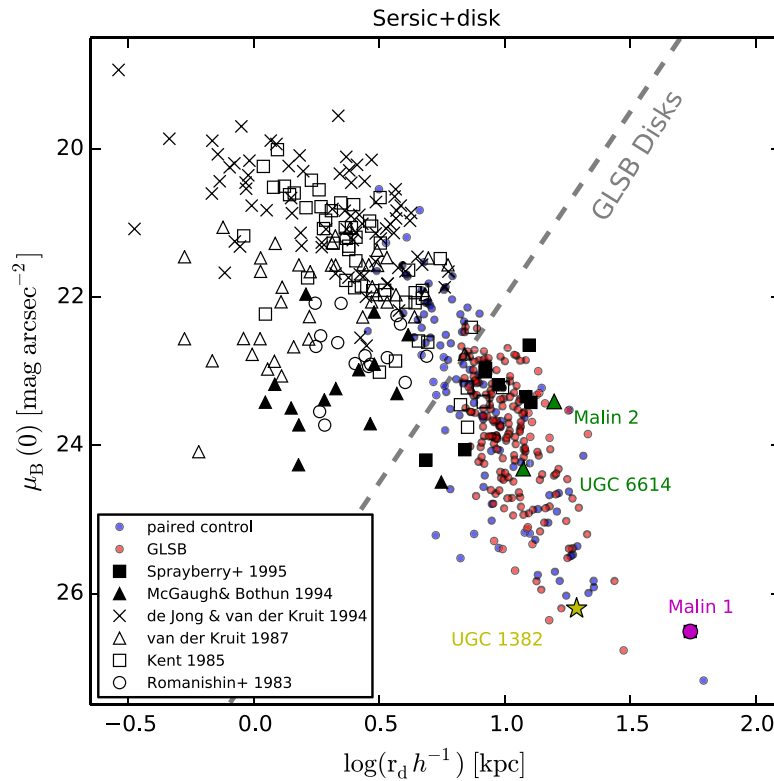
Figure 8. Distribution of the outer disc parameters from the photometric fitting in the *B* band. Previous measurements from the literature are also summarized here for comparison. According to the definition used in equation (5) (Sprayberry et al. 1995), the dashed line is adopted to separate GLSB from their HSB counterparts. The observed GLSB galaxies include those from McGaugh & Bothun (1994), Sprayberry et al. (1995) to UGC 1382 (Hagen et al. 2016), all to the right of the dashed line. For the GLSB sample, 167 galaxies out of 171 which can be reasonably fitted by the 'Sérsic + disc' profile (see Section 3.2 for details) are indeed GLSB discs. For the paired control sample, 55 objects out of 99 also show extended outer discs according to the photometric fitting, while the rest are consistent with an HSB disc.

UGC 1382 was classified as a normal early-type galaxy before its faint stellar disc was discovered (Hagen et al. 2016). More recently, Saburova et al. (2018) have presented a study of a 'Malin 1' cousin, UGC 1922. This galaxy contains a prominent bulge and an extended disc, similar to other known GLSB galaxies. Diffuse UV emission has been detected in some GLSB galaxies (e.g. O'Neil et al. 2007; Boissier et al. 2008, 2016; Hagen et al. 2016), which is a sign of current star formation activity. Nevertheless, a full picture of the overall location of GLSB galaxies in the galaxy colour–magnitude diagram is still lacking.

In the upper panel of Fig. 10, we show the GLSB and paired control samples in $u - r$ colour versus *r*-band magnitudes. Using the relation by Baldry et al. (2004) and Smethurst et al. (2015), we also plot the position of the green valley. Those above the green valley comprise the so-called red sequence and those below the blue cloud. As demonstrated by Nelson et al. (2018), one of the major improvements of TNG over the original Illustris model is more realistic galaxy colour bimodality.

The bottom panel of Fig. 10 summarizes the $u - r$ colour for GLSB, control, and 'all galaxies' samples. The differences between GLSB and the rest of the TNG100 galaxies are readily seen. The colour bimodality separates 'all galaxies' into two well-defined peaks. For GLSB galaxies, the peak for blue galaxies is absent, reflecting that few galaxies with $u - r < 1.5$ exist. And the colour distribution is closer to a single peak distribution around the red sequence with a tail into blue galaxies. The paired control shows

a similar distribution in galaxy colour as GLSB galaxies, but with slightly more blue galaxies. The peak for GLSB galaxies is slightly bluer than the paired control. We caution that Fig. 10 only shows $u - r$ colour for the entire galaxy, while there is a substantial colour gradient within each galaxy. As shown in Fig. 6, blue stars are preferentially found in the outer discs.

Many GLSB galaxies around the green valley are in transition between the blue and red galaxies, often still with ongoing star formation activities. Fig. 11 shows the distribution of star formation rate (SFR) and the total *B*-band luminosity. We also include the measurements from O'Neil et al. (2007) and Saburova et al. (2021) for comparison. Galaxies from O'Neil et al. (2007) are consistent with the majority of TNG100 galaxies (small dark dots), following a tight relation resembling the main sequence. Overall, galaxies included in O'Neil et al. (2007) are less massive than the ones in Saburova et al. (2021), primarily consisting of known GLSB galaxies. The upper part of the GLSB and control samples are consistent with the seven GLSB galaxies in Saburova et al. (2021), while a good fraction of the GLSB sample is characterized by SFR <1 M$_\odot$ yr^{-1}.

Both the GLSB and paired control samples deviate from an overall galaxy main sequence. Fig. 11 is also consistent with Fig. 10, that most of the two samples consist of galaxies in the green valley and the red sequence. We find no systematic differences between the two samples regarding their locations in the SFR–M_B despite their differences in the total H I mass. This can be explained by the fact that those gas cells with a density higher than the star formation
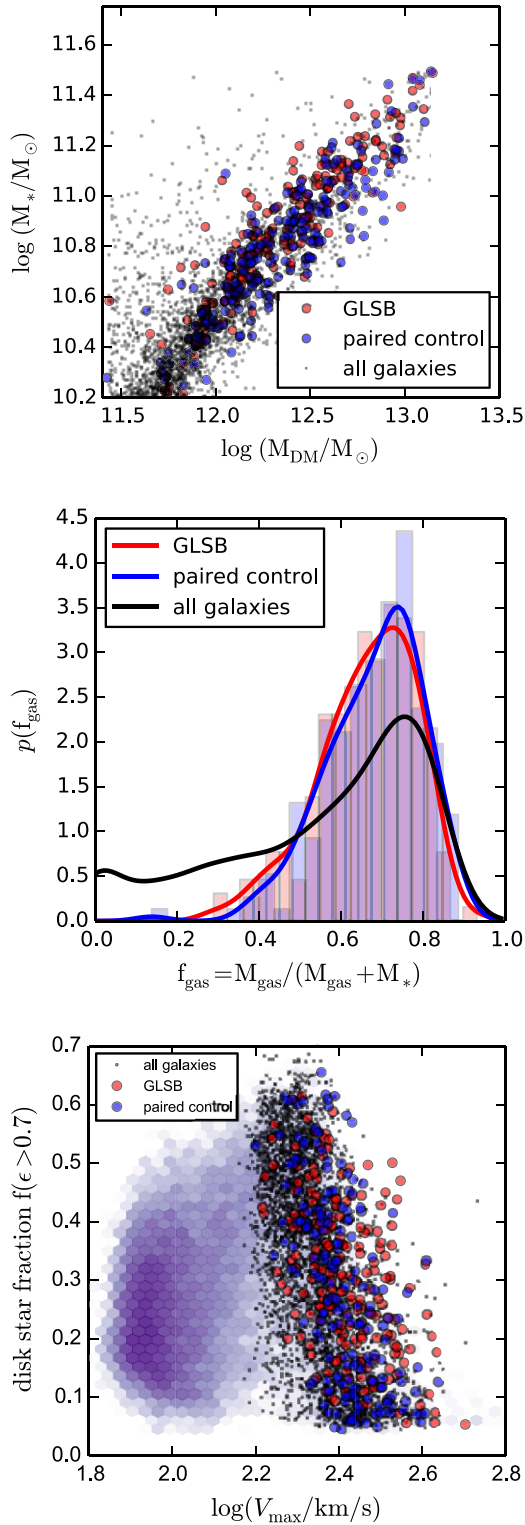
Figure 9. Top panel: Galaxy dark matter and stellar mass for GLSB, paired control, and 'all galaxies' samples. Middle panel: Distribution of gas fraction $f_{gas} = M_{gas}/(M_{gas} + M_*)$. Bottom panel: Disc mass fraction $f_{\epsilon > 0.7}$ as a function of maximum rotational velocity V_{max}. The dark hexagons show the overall distribution of TNG100 with the intensity corresponding to the number of galaxies in each hexagon.
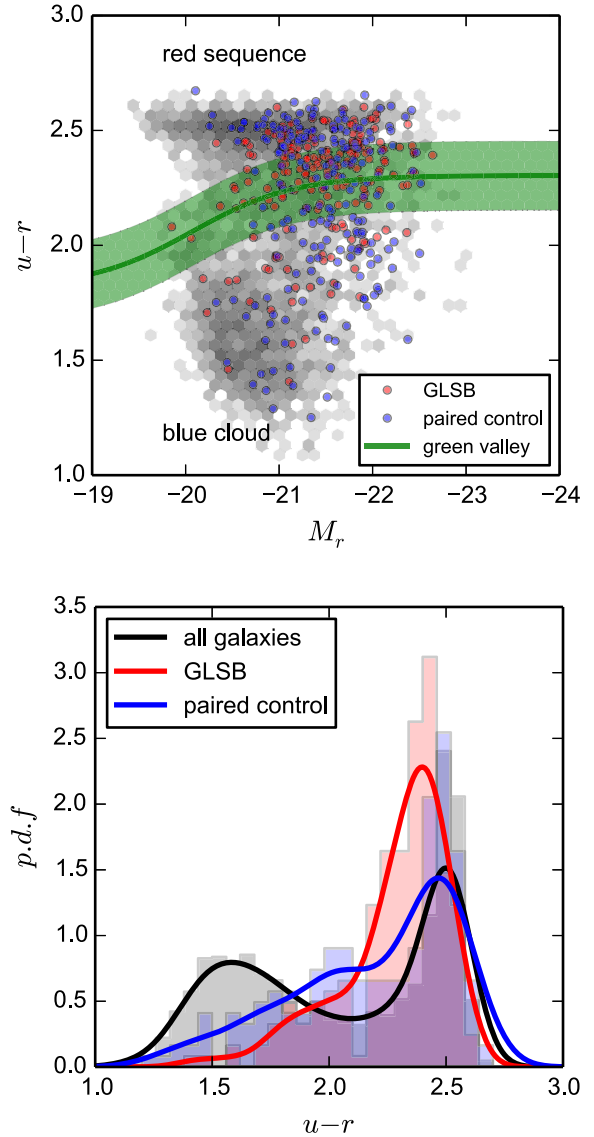
Figure 10. Upper panel: Galaxy $u - r$ colour and total r-band magnitude M_r relations for GLSB, paired control, and the rest of TNG100 galaxies. The observational definition of the green valley is adopted to separate the galaxies into three populations accordingly. GLSB and the control samples span the parameter space of red sequence and green valley, but with few galaxies in the blue cloud. Bottom panel: $u - r$ colour for the GLSB, paired control and 'all galaxies' at $z = 0$. Both the raw counts (in histograms) and the kernel density estimates (in smooth curves) are displayed here. The GLSB and paired control sample contain few blue galaxies.

threshold only contribute to star formation. Although GLSB galaxies generally contain more H I mass than the paired control, as long as the gas remains diffuse below this threshold, no significantly higher level of star formation activity is present.

3.5 Mass distribution

Owing to their extended H I distribution, GLSB galaxies provide a rare opportunity to measure the mass distributions up to large galactic radii. By re-analysing the H I maps of Malin 1 and NGC 7589 by Pickering et al. (1997), Lelli et al. (2010) derived new rotation curves for these two galaxies. Both galaxies show flat rotations curves above $200 \, km \, s^{-1}$. In contrast to the previous
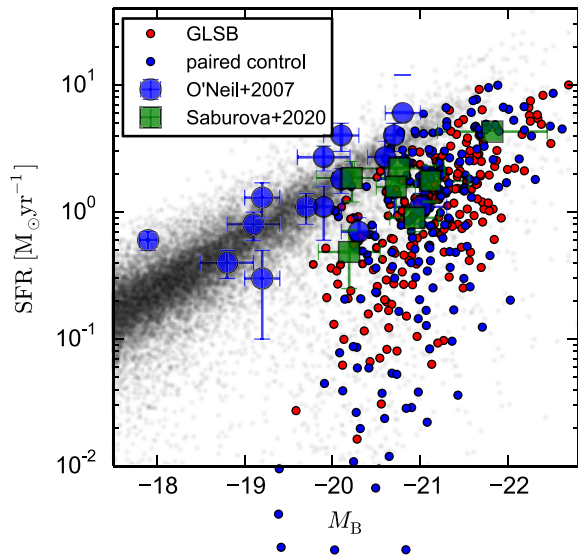
Figure 11. A comparison of total SFR versus the total *B*-band magnitudes for the GLSB and paired control samples. The rest of the TNG100 sample is shown in the small dark symbols. Observational measurements from O'Neil et al. (2007) and Saburova et al. (2021) are shown in large coloured symbols.

conclusion by Pickering et al. (1997), they find both galaxies contain steeply rising rotation curves in the inner region, which implies similarities between the inner GLSB component and normal galaxies.

In a photometric and spectroscopic study of Malin 2, Kasparova et al. (2014) performed the mass modelling for each component and found the dark matter halo to be best described with a large core radius of 27 kpc in a pseudo-isothermal profile. The inferred central dark matter density of Malin 2 ($0.003 \, \mathrm{M_\odot \, pc^{-3}}$) is almost a factor of 10 less than that of the Milky Way (e.g. see Read 2014). Kasparova et al. (2014) further argue that differences between GLSB galaxies and normal galaxies can be caused by the different dark matter halo properties. In particular, a massive and rarefied dark matter distribution with large-scale radius largely defines the scale of the giant disc. The total mass of Malin 2 is quite large, as shown in its rotation curve (see fig. 7 in Kasparova et al. 2014), which approaches $300 \, \mathrm{km \, s^{-1}}$ at radii ~ 100 kpc.

Mishra et al. (2017) presented the rotation curves of four GLSB galaxies derived from GMRT H I observations. The flat part of the rotation curve of the four galaxies ranges from 225 to $432 \, \mathrm{km \, s^{-1}}$. Measurements for yet another three GLSB galaxies can be found in a recent study by Saburova et al. (2021). For NGC 7589, UGC 1382, and UGC 6614, the rotational curves are overall flat with $V_{\mathrm{rot}} \gtrsim 200 \, \mathrm{km \, s^{-1}}$. To date, all the nine known GLSB galaxies are consistent with massive galaxies as indicated by their large H I rotational velocities.

In Fig. 12, we plot the rotation curves, in the form of enclosed mass $V_{\mathrm{rot}}^2 = GM_{\mathrm{en}}/r$, as a function of radius r for each component: stars, dark matter, and gas. For the total mass distribution shown in the last panel, the solid curve shows a median value of $240 \, \mathrm{km \, s^{-1}}$, while the shaded region corresponds to the 16th and 84th percentiles, at 190 and $300 \, \mathrm{km \, s^{-1}}$, respectively. This panel shows flat rotation curves with our GLSB sample, fully consistent with the nine measured GLSB galaxies we listed above. It is interesting to note that the rotation curves for the control sample do not deviate significantly from the

GLSB sample, which implies that the total mass distributions are similar.

The top two panels of Fig. 12 compare the mass distributions of stars and dark matter. Due to both the stellar and the dark matter components, the rotation curves are consistent between the GLSB and paired control samples. For the inner 20 kpc, the distribution of dark matter is virtually the same between the two. Finally, it is only in the gas distribution that we find differences in the mass distribution, as shown in the last panel. At $r = 80$ kpc, the GLSB sample contains larger gas masses, indicated by $V_{\mathrm{rot,gas}} \sim 60 \, \mathrm{km \, s^{-1}}$. For comparison, at the same distance, the control sample has a 30 per cent lower median $V_{\mathrm{rot,gas}}$, at $\sim 40 \, \mathrm{km \, s^{-1}}$. The differences in $V_{\mathrm{rot,gas}}$ become smaller at larger r, which indicates it is caused by different H I distributions of the two samples.

Unlike LSB dwarf galaxies, GLSB galaxies usually exhibit steep-rising rotation curves (Lelli et al. 2010; Saburova et al. 2021). Fig. 13 compares individual rotation curves with the observed GLSB galaxies. We include the long-slit spectroscopic observations from Saburova et al. (2018; UGC 1922), Saburova et al. (2019; UGC 1378), and Saburova et al. (2021; Malin1, Malin2, UGC1382, UGC6614, NGC7589). Since these measurements are optical, they offer much higher spatial resolution than currently available H I data, therefore being able to probe the central region. Our GLSB sample shows similar behaviour in terms of the rising of rotation curves in the inner 2 kpc as the observed ones.

3.6 Galaxy spin parameter

One of the most important properties of dark matter haloes is the halo spin, which is a fundamental concept to understand the formation of galaxy discs (Fall & Efstathiou 1980; Mo, Mao & White 1998). Galactic angular momentum originated from tidal torques acting on the dark matter haloes in the early Universe (Peebles 1969). The global spin parameter λ is found to follow a lognormal distribution in N-body dark matter haloes (e.g. Barnes & Efstathiou 1987; Bullock et al. 2001). For cosmological hydrodynamic simulations, Zjupa & Springel (2017) find that while the baryonic process radically changes the baryonic angular momentum, the dark matter component remains largely unaffected.

We compare the dimensionless spin parameter λ derived from the total matter content of the galaxy for the GLSB and control sample, according to

$$\lambda = \frac{J_{\mathrm{tot}}}{M_{\mathrm{tot}}} \frac{E_{\mathrm{tot}}^{1/2}}{G M_{\mathrm{tot}}^{3/2}} \tag{6}$$

and refer to λ as halo spin. This follows the definition by Peebles (1969), with M_{tot}, J_{tot}, E_{tot} the total mass, angular momentum, and total energy of a particular galaxy. The details of the methods of its computation with the TNG simulation remains the same as in Zjupa & Springel (2017).

The distributions of λ at $z = 0$ are shown in the top left-hand panel of Fig. 14, both in the histograms and as a smoothed kernel density estimation. The paired control sample closely follows the overall lognormal distribution, broadly consistent with 'all galaxies' in the same mass range. On the other hand, the GLSB sample has a distinctly larger spin parameter.

The median values for λ are 0.038 and 0.043 for 'all galaxies' and paired control. A two-sided KS test reports a *p*-value of 0.002, indicating notable differences in the control sample. On the other hand, the median value of λ is 0.060 for the GLSB sample. KS tests between GLSB and either of the other two samples can fully reject the hypothesis that the two samples are drawn from the same
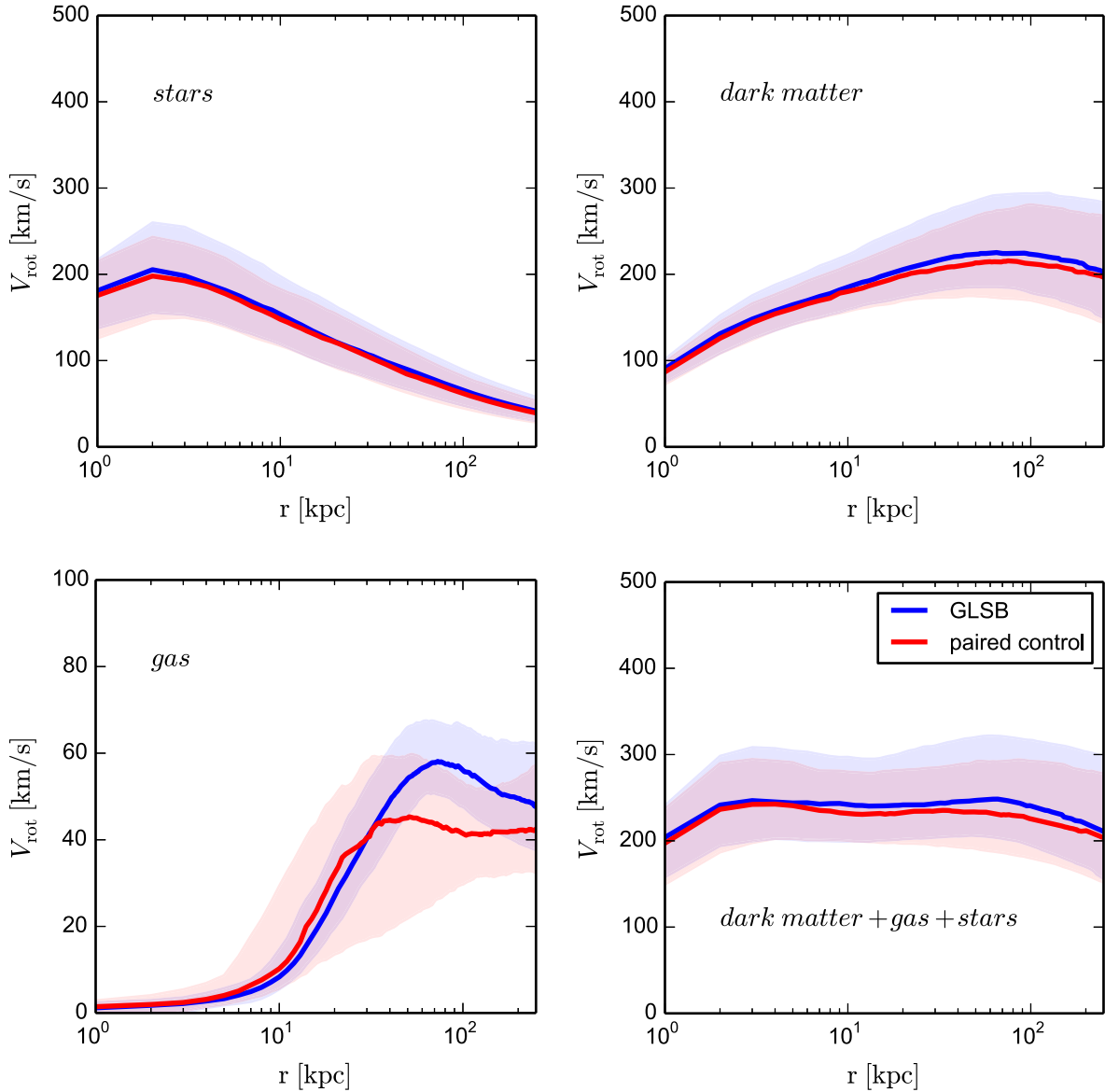
Figure 12. Rotation curves due to each mass component: dark matter, gas, and stars. The last panel is based on the total mass distribution. The GLSB galaxies and the paired control sample show little differences in their dark matter, stellar, and overall total mass distributions.

distribution, with a *p*-value $\ll 1$. While λ is ~ 40 per cent larger than the control sample, the lognormal nature of the distribution implies substantially more galaxies reside in the long tail of high-spin haloes.

We note that the above finding, which links the halo spin with galaxy surface brightness, is consistent with two recent studies, Kulier et al. (2020) and Pérez-Montaño et al. (2022), despite the differences in the definition of LSB. In particular, the galaxy mass ranges considered in these two works are substantially more extended than our GLSB sample. A strong correlation between halo spin and galaxy size (which is closely related to galaxy surface brightness) was also recently reported by Rodriguez-Gomez et al. (2022) for a large sample of TNG100 galaxies.

The other three panels in Fig. 14 are the distributions of λ at earlier redshifts, obtained from the main progenitors (obtained from merger trees, see Section 3.7) of our $z = 0$ galaxy samples, whereby

we include only progenitors with mass $>10^{10}$ M$_\odot$.[2] For the GLSB sample, the median value increases from 0.033 at $z = 3$, to 0.040 at $z = 2$, to 0.055 at $z = 1$, to 0.060 at redshift zero. In contrast, for the paired control sample, its median value of λ has increased from 0.033 at $z = 3$, to 0.036 at $z = 2$, to 0.039 at $z = 1$, to 0.043 at $z = 0$. It is evident that the differences in the spin parameter have emerged via a continuous and prolonged process. Interestingly, at $z = 2$ the differences in λ seem to converge between the samples, which is in good agreement with what it was found by Pérez-Montaño et al. (2022) where the authors found that the evolution of the spin

[2]The mass cut at 10^{10}M$_\odot$ for λ is intended for secure measurements with enough particles. With this condition imposed, a very tiny fraction (0 for GLSB, 1 for paired control, and 14 for 'all galaxies') of galaxies is without λ measurements at some earlier redshifts.
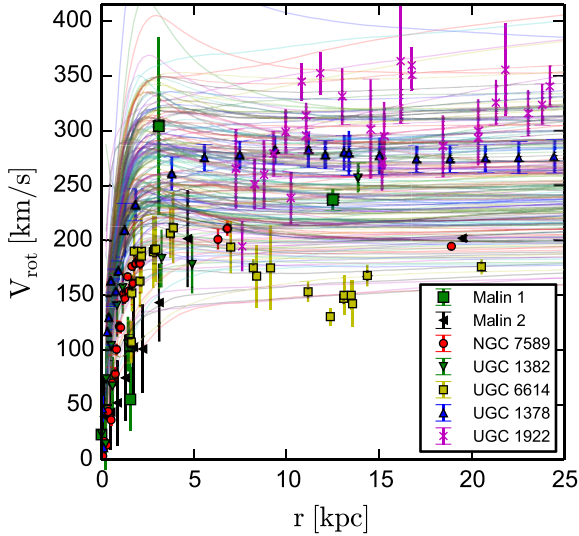
Figure 13. Rotation curves of individual galaxies from the GLSB sample, compared with recent observations (Saburova et al. 2018, 2019, 2021). Based on long-slit optical spectroscopy, these GLSB galaxies show steeply rising curves in their inner part, which is consistent with our data.
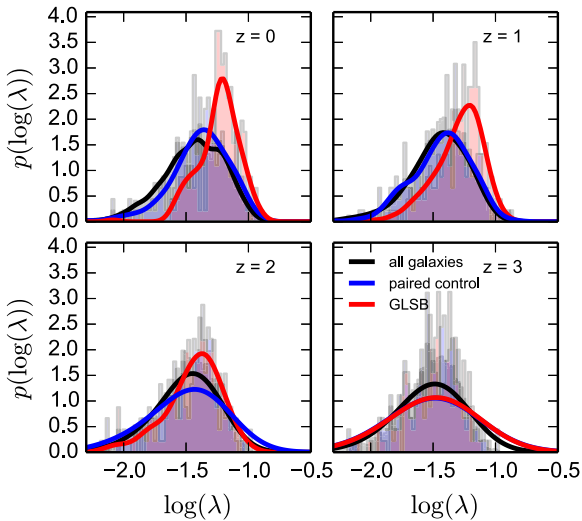

Figure 14. The distribution of halo dimensionless spin parameter λ for the GLSB sample, paired control sample, and 'all galaxies' in TNG100 at different redshifts. While the control sample closely follows the overall lognormal distribution, GLSB galaxies tend to reside in haloes with higher spin parameters. Going back in redshift, the difference starts to emerge since $z = 2$.

parameter between LSB and HSB galaxies exhibit a clear bifurcation at similar redshift.

3.7 Merger history

Merger trees provide a unique perspective on the formation histories of simulated galaxies. With the merger trees constructed using the method in Rodriguez-Gomez et al. (2015, 2016), we first created animations of each GLSB galaxy based on its gas distribution in each snapshot and visually inspected each animation. One of the common features which caught our attention is that the extended
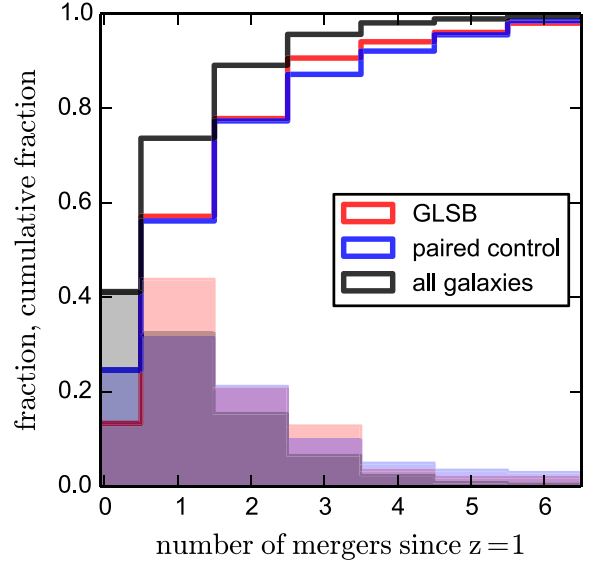

Figure 15. The distribution of the total number of major mergers since $z = 1$. The solid lines show the cumulative number of the underlying histograms. The distributions are asymmetric, with the peak at 0 or 1 and a long tail towards larger numbers of major mergers. Between the GLSB and paired control samples, we cannot rule out they are drawn from the same distributions.

H I gas at $z = 0$ can be visibly attributed to a galaxy merger or an accretion.

With the nomenclatures in Rodriguez-Gomez et al. (2015), we consider mergers to be major mergers, defined by a mass ratio $\mu > 1/4$. The mass ratio is based on the stellar masses of the two merging galaxies at the time when the secondary reached its maximum stellar mass. Fig. 15 compares the total number of major mergers since $z = 1$ for the GLSB and paired control samples. The rest of the TNG100 'all galaxies' are included as well.

The total number of major mergers follows a highly asymmetric distribution with its peak at 0 ('all galaxies') or 1 ('GLSB' and 'paired control'). According to the two-sample KS test, while we can reject the hypothesis that the total number of major mergers of the GLSB or control samples are drawn from the same distribution as the rest of the TNG100 galaxies, the same statement cannot be made between GLSB and paired control samples.

A notable fraction of galaxies in Fig. 15 registered no major mergers since $z = 1$. On the other hand, a significant number of galaxies had ≥ 2 major mergers. We also confirm that those extended discs without any major mergers are indeed long-lived galaxies based on the animations. Therefore GLSB galaxies, without being disturbed for a long time on a cosmological time-scale, e.g. since $z = 1$, do exist in ΛCDM, at least in the TNG simulations.

Fig. 16 further shows the details of the redshift of the last major merger. Overall, we find a broad distribution of redshift. For GLSB galaxies, a peak between $z = 0.2$ and 0.4 can be seen in the kernel density estimate. Translating redshift to look-back time, the GLSB sample contains many galaxies with major mergers completed 2–4 Gyr ago. For comparison, a peak at $z = 0.1$ can be identified for the paired control sample. The raw histogram of the latter has a pronounced peak at $z = 0$, indicating that some galaxies have just recently completed the merger process.

Fig. 17 compares the amount of stellar mass that was formed *ex situ*, in terms of the *ex situ* mass fraction f(exsitu). Here, *ex situ* stars are those formed outside the main galaxy but accreted subsequently
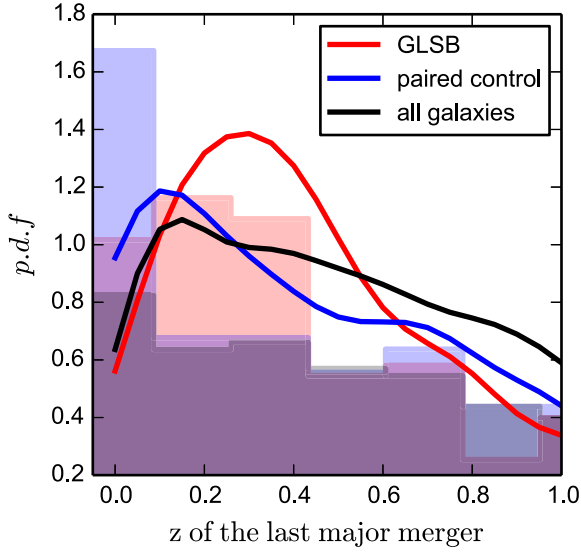
Figure 16. The redshift distribution of the last major merger for GLSB and control samples. The smooth curves are kernel density estimations, while the histograms are raw counts. For the GLSB sample, we find a peak between redshift 0.2 and 0.4 for the last major merger. For the paired control, there is a pronounced peak at $z \sim 0$, indicating a very recent/ongoing merger.
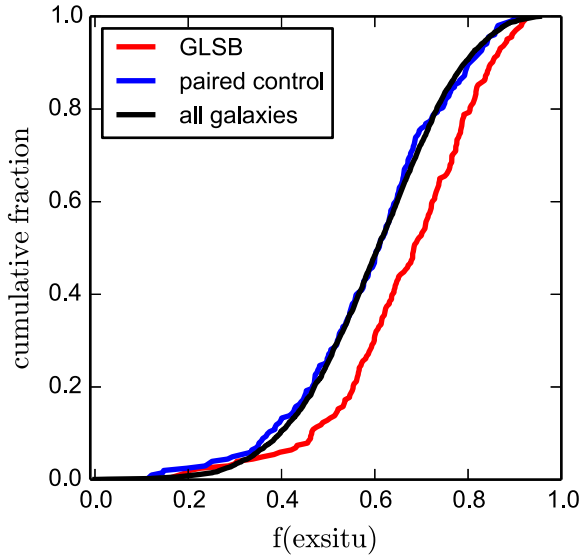

Figure 18. Tilt angle between the angular momentum of the gas and stars within 125 kpc from the galaxy centres. The GLSB sample shows a preference of $\theta < 10$ deg over the paired control and 'all galaxies' samples. Polar or counter-rotating gas configurations in the GLSB sample are rare. The dashed line indicates the cumulative distribution of purely random orientations between gas and stellar angular momenta. KS tests show that the GLSB sample statistically differs from the control sample.

had their last major merger more recently than HSB galaxies, with a major contribution of *ex situ* build-up.

Another feature of galaxy mergers is the orbital configuration of the two merging galaxies. Analysing the TNG100 simulation, Zeng, Wang & Gao (2021) find that major mergers with a spiral-in orbit leads to the formation of massive discs. Instead of the orbit parameters before the merger, we show the alignment of the angular momentum of gas and stellar component at $z = 0$, which is operationally more feasible for observations than the initial orbit configuration.

Fig. 18 compares the result of the tilt angle between the gas and stellar components for the GLSB and paired control samples. We also computed the result expected from a random orientation, assuming no correlation between the two angular momentum vectors, as shown in the dashed curve. One can see that the angular momenta of gas and stars in GLSB are so strongly aligned that ~ 80 per cent of the sample shows a tilt angle ≤ 10 deg and > 90 per cent with angle ≤ 20 deg. For comparison, while the paired control sample also shows a preference for small tilt angles, one can still find ≥ 45 deg or even counter-rotating configurations. In addition, the difference between the paired control and 'all galaxies' is minimal according to KS test. This figure indicates that while the TNG100 galaxies show some preferences towards the alignment of the two components, H I components are more aligned with the stellar angular momentum for the GLSB sample.

Figure 17. Stellar *ex situ* mass fraction for GLSB, paired control, and 'all galaxies'. The two control samples are in good agreement in the fraction of *ex situ* stars. On the other hand, the GLSB curve favours larger $f(exsitu)$, which is statistically different from the other two distributions.

3.8 Environment

The final galaxy property we examine before we move on to the discussion is the environment. Both mass and environment are recognized as drivers for galaxy evolution (Dressler 1980; Peng et al. 2010). Modern simulations such as Illustris are able to reproduce the

(Rodriguez-Gomez et al. 2016). In contrast, the *in situ* stars are those formed out of the condensed gas in the main galaxy. Visually, the *ex situ* mass fraction of the paired control sample agrees perfectly with 'all galaxies' in TNG100. On the other hand, GLSB galaxies consistently show a larger $f(exsitu)$. This impression is further verified with a two-sample KS test. The difference in $f(exsitu)$ is mostly driven by the number of mergers as shown in the previous figure. In Pérez-Montaño et al. (2022), with a different operational definition of LSB galaxies, the authors also find that LSB galaxies
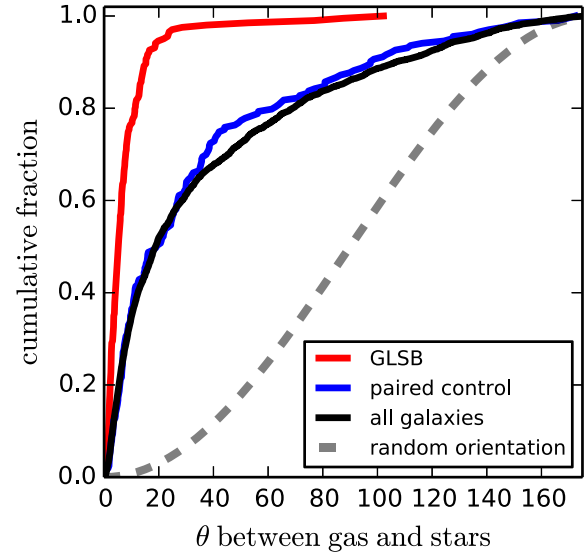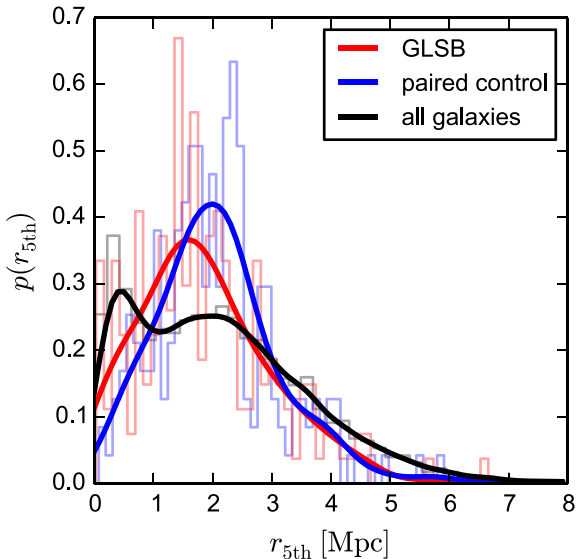
Figure 19. Distance to the fifth neighbour r_{5th} for the GLSB and two control samples. We only include massive galaxies with $M_r < -19.5$ in the distance calculation. Both the GLSB and the paired control samples show single-peaked distributions, while the 'all galaxies' sample contains a second peak of galaxies with $r_{5th} < 1$ Mpc.

observed rapid quenching process in massive galaxies in high-density environments (Vogelsberger et al. 2014a).

Bothun et al. (1993) reported a deficit of nearby neighbours for LSB discs on small scales below 2 Mpc, based on ~340 galaxies from the CfA redshift survey. Their interpretation is that the lack of tidal interactions prevents the low-density gas discs from rapidly evolving. Rosenbaum & Bomans (2004) found significant differences in galaxy densities around LSB and their HSB counterparts. This conclusion is drawn from the neighbour counting analysis of 16 123 SDSS galaxies. The fifth nearest neighbour distribution also shows differences between 2 and 5 Mpc for the two galaxy populations.

For GLSB galaxies specifically, Knezek (1993) investigated the environment for a flux-limited sample of galaxies in UGC with $M_{H I} > 10^{10} M_{\odot}$ and $D_{25} > 25$ kpc. While the GLSB galaxies trace the overall large-scale distribution of normal galaxies, a lack of nearby (within 1 Mpc) bright L^* neighbours was found. Reshetnikov, Moiseev & Sotnikova (2010) examined the nearby galaxy distribution for Malin 1, and found it in a relatively low-density region with a bright neighbour at a projected distance of 350 kpc. Junais et al. (2020) report that Malin 1 lies about 10 Mpc from the edge of its closest filament. Hagen et al. (2016) studied the local environment for UGC 1382 and concluded that it is located in a poor galaxy group, with only one bright neighbour within 500 kpc. Recently, Pérez-Montaño & Cervantes Sodi (2019) have also reported that the fraction of isolated central LSB galaxies is higher than that for HSB ones, and the density of their local environment is lower.

We also use nearest neighbours as a proxy for the environment. To this end, we calculate the distance to the fifth nearest neighbouring galaxies by only including those brighter than $M_r = -19.5$ following Vogelsberger et al. (2014a). Fig. 19 shows the distribution of r_{5th} for three samples.

At face value, Fig. 19 is in contrast to what is found by Bothun et al. (1993) and Rosenbaum & Bomans (2004). Namely, the paired control sample favours slightly more isolated environments than GLSB candidates. Both samples have few objects with $r_{5th} < 1$ Mpc. The peak is at 2 and 1.5 Mpc for the GSLB and paired control

samples. At >3 Mpc, the apparent differences disappear for the r_{5th} distribution.

This puzzle is resolved by noting that the construction of the control sample is quite different between this work and those in Bothun et al. (1993) and Rosenbaum & Bomans (2004). In particular, we control stellar, DM, and gas mass masses such that the control and the GLSB galaxies are paired. In doing so, galaxies in galaxy clusters are unlikely to be selected to be part of the control in the first place. This effect can be seen in the distribution for 'all galaxies' in the figure. In the absence of matching all three components, a peak at ~0.5 Mpc is present. And the differences between the GLSB and 'all galaxies' samples can be seen up to 6 Mpc. Interestingly, there is also a deficit of GLSB galaxies with $r_{5th} > 3$ Mpc.

As we use the total gas while constructing the paired sample, and the cold H I gas in GLSB galaxies is substantially more massive than the control, we essentially demand the control sample contain more mass in hot halo gas. As a result, we see some compensation effect here: the control sample would be slightly more isolated to retain more halo gas. The deficit of GLSB galaxies in the dense environment is broadly consistent with Malin 1, UGC 1382, and those studied in Knezek (1993).

4 DISCUSSION

4.1 Observational outlooks

Even today, surprising new features are being reported for Malin 1 (Junais et al. 2020; Saha et al. 2021). Our understanding of GLSB galaxies will undoubtedly be advanced in the coming decade. In particular, facilities like BlueMUSE (Richard et al. 2019) and Vera C. Rubin Observatory will play a pivotal role in this process.

Ongoing and future H I surveys (e.g. Yahya et al. 2015; Koribalski et al. 2020) will expand the volume of discovery compared to the previous generation of survey, such as HIPASS (Barnes et al. 2001) and ALFALFA Giovanelli et al. (2005). These surveys will be highly beneficial for GLSB discoveries as well. Additionally, they will test the connection between extended H I discs and the formation of GLSB discs we have explored in this work.

Saburova et al. (2023) predict that 13 000 GLSB galaxies exist within $z < 0.1$ in the full sky. The volume fraction of our sample is about twice their estimate. It is reasonable to expect a large number of GLSB galaxies await to be discovered in the coming era of Vera C. Rubin and SKA observatories. Thus, GLSB galaxies will be at the front of the unexplored low-surface brightness Universe.

4.2 Formation scenarios

In the previous section, we present our analysis of the basic properties of GLSB candidates. We find that our GLSB sample shows good agreement in terms of the H I mass, the parameters of the stellar disc, galaxy colour, and SFR with the observed GLSB galaxies. When comparing them with the paired control sample, we noticed significant differences in the halo spin, merger history, and angle between gas disc and the stellar components. However, we did not find significant differences in the dark matter mass distributions, nor their environments on ~Mpc scales. Since we have discussed each topic along the way in the text above, we focus on the formation scenarios here.

Bothun et al. (1987) and Impey & Bothun (1989) suggested the formation of Malin 1 as an un-evolved (or slowly evolving) disc at low redshift. This view is consistent with the large amount of available cold gas detected in Malin 1. In particular, given the relative isolated

environment of Malin 1, the gas disc could be long lived since $z \sim 2$ with a sustained low level of star formation due to the low surface density of the gas disc. Of course, this picture appears to be difficult to fit in hierarchical formation of galaxies, since discs can be fragile to frequent galaxy mergers and flybys (see, however, Sotillo-Ramos et al. 2023).

Looking at Fig. 15, there are certainly a few galaxies in our sample that can be described with the above picture. The relatively isolated environments of LSB galaxies also prevent them from any major mergers since $z = 1$, leading to the survival of a large gas disc already present. Nevertheless, the majority of the sample does not fit into this picture.

Hoffman et al. (1992) presented a formation scenario that GLSB galaxies consist of normal bulges but un-evolved extended discs as a result of the large-scale environment. Specifically, galaxies formed in rare 3σ density peaks in the void show similar inner structure as those formed in 1σ peaks within more clustered regions. For those galaxies formed in the void, their mass distribution in the outer regions is significantly more extended but the gas cooling time is much longer when compared to the dynamical time. The long cooling time (on the scale of Hubble time) thus prohibits efficient star formation within the extended gas disc over cosmic time.

The difficulty of this formation scenario to explain the entire GLSB population is highlighted in Fig. 19 where both the control and GLSB samples are located in a more isolated environment than the 'average' TNG100 galaxies. Moreover, we find our GLSB galaxies are rarely in the voids. Therefore, the large-scale environment alone is not the sole cause of GLSB formation.

Noguchi (2001) proposed a theory of forming extended stellar discs as a result of radial redistribution of matter due to bars formed in the HSB discs. In the numerical experiments, galaxy bulge and disc as well as the dark matter halo are followed as collisionless particles. In the model of moderate bulge fraction, a stellar bar forms within the stellar disc. In the later stages, it is observed that the initial disc is better characterized with a sum of two exponential discs. The outer disc shows a considerably larger disc scale length (a factor of two) and lower central surface brightness. The whole process is essentially internal and purely galactic, therefore having little dependence on the large-scale structure. Our GLSB sample contains both early-type and late-type galaxies, and we directly observe the build-up of gas discs from galaxy mergers. On the other hand, the spatial resolution of TNG100 could be too coarse to resolve this proposed process. TNG50, with a much higher resolution, can be used to explore further and test this formation mechanism (Zhao et al. 2020; Frankel et al. 2022). On the other hand, the formation scenario of the disc mass redistribution due to bars does not offer an immediate explanation for the large H I mass usually observed in GLSB galaxies.

Mapelli et al. (2008) presented a study of major mergers of two disc galaxies with small impact parameters using hydrodynamic simulations. The virial mass of the host and intruder are 4.9×10^{11} and $3.2 \times 10^{11} \mathrm{M_\odot}$, respectively. The host galaxy consists of a dark matter halo, a stellar bulge, and a disc as well as an exponential gas disc while the intruder is devoid of any baryons. Unlike previous numerical studies only involving collision-less dynamics, the inclusion of gas dynamics enables the authors to follow gas cooling and star formation in the merging process. As the adopted set-up was initially designed for Cartwheel-like galaxies, these simulations produce large stellar and gas ring 100–200 Myr after the close encounter with clear 'spokes' connecting the ring and the central stellar disc. Afterward, the stellar ring and the spokes slowly faded away, becoming indistinguishable from a giant disc structure. The

cold gas also settles into an extended but low surface density disc, akin to the observed GLSB galaxies.

The inconsistencies with the head-on collision theory and our results are readily seen in Figs 16 and 18. In particular, the timing for the last major merger is also significantly different. The time for the last major merger is registered at 2–4 Gyr ago. Additionally, we see a preferential in-spiral orbit, which results in a well-aligned gas disc with the stellar component. Therefore, we conclude that the recent (<1 Gyr) head-on collision is not the main mechanism to form GLSB galaxies in TNG..

Dalcanton et al. (1997) presented a general picture of disc formation within the frame of hierarchical structure formation. With the assumptions that the gas angular momentum distribution follows the same distribution of dark matter as a result of large scale tidal torques, and little angular momentum transportation occurs during the cooling and collapse of gas, the luminosity and the scale length of the disc is then directly determined by the halo spin parameter and total halo mass. Massive GLSB galaxies would naturally form in massive haloes with large λ.

Another scenario regarding of the formation of extended stellar discs such as in M31 is discussed in Peñarrubia et al. (2006). Satellite galaxies on prograde co-planar orbits are disrupted as they fall into the potential of MW/M31 size host galaxies. The debris form an extended exponential disc structure with a disc scale length as large as 30 kpc over the range of 30 and 200 kpc from the host. This mechanism works more efficiently for more massive dwarf galaxies due to stronger dynamical friction. The disc scale length is also strongly correlated with the initial stellar distribution within the dwarf galaxy. The more concentrated the initial stellar components are, the more compact the resultant debris discs.

We believe the mechanisms presented in Dalcanton et al. (1997), Peñarrubia et al. (2006), and an isolated environment could explain the GLSB sample in TNG100. Specifically, the distinct difference in the halo spin parameters between the GLSB and paired control samples is consistent with the predictions made by Dalcanton et al. (1997).

For the model in Peñarrubia et al. (2006), only stellar and dark matter components are followed in a static host potential, the more complex dynamics with gas and star formation are not considered. Based on the properties of UGC 1382, Hagen et al. (2016) have concluded that the extended gas disc in UGC 1382 is consistent with this formation scenario such that the total cold gas ($2 \times 10^{10} \mathrm{M_\odot}$) can be what is left over of one or a few disrupted dwarf galaxies.

The TNG simulations follow the gas dynamics, as well as gas cooling and star formation simultaneously. GLSB galaxies actually are among the most H I massive galaxies in the entire sample, suggesting some additional channel of cold gas supply. The reason is that any external galaxy alone is not able to provide the entire H I gas. In Zhu et al. (2018), using the thermal histories tracked by tracer particles (Genel et al. 2013), the authors find that a substantial fraction of the hot halo gas participated in the cooling process. The cooling occurred at the interface between the incoming cold gas delivered by the merger and the existing halo gas. This cooling process is distinct from classical cold accretion (Kereš et al. 2005) or normal cooling from shocked heated gas. The detail of this 'stimulated cooling' is beyond the scope of this study, and we reserve it for future work.

4.3 Comparison with other numerical studies

The formation of discs as a result of major mergers has been well studied over the years (e.g. Springel & Hernquist 2005; Athanassoula

et al. 2016; Peschken, Athanassoula & Rodionov 2017; Sparre & Springel 2017; Peschken, Łokas & Athanassoula 2020). However, the connection between mergers and the formation of GLSB galaxies has not been fully unveiled. There are a few recent numerical studies on the formation of LSB galaxies in a cosmological context that we can compare with.

Di Cintio et al. (2019) investigated LSB galaxy formation in the NIHAO simulation suite, with the stellar mass in the range between $10^{9.5}$ and 10^{10} M$_\odot$. They found that coplanar co-rotating mergers and aligned accretion of gas are able to reproduce galaxies with large H I reservoirs, extended effective radii and slowly rising rotation curves. On the other hand, perpendicular or misaligned merger/accretion events lead to galaxies with higher surface brightness. Our GLSB sample contains galaxies that are significantly more massive than Di Cintio et al. (2019), and more diverse in terms of galaxy morphology. Nevertheless, we note a similarity in terms of the formation mechanism of the extended stellar disc.

Kulier et al. (2020) analysed the EAGLE simulation to study the properties of LSB galaxies. The EAGLE simulation has a comparable volume to TNG100. LSB galaxies are found to be in more isolated environments, and in higher spin haloes. For a few galaxies, with extended stellar discs comparable to the observed GLSB galaxies, they are formed through mergers. While EAGLE uses SPH as its hydro solver and a completely different prescription for star formation and feedback, and AGN feedback (this is also true for NIHAO), we again find a very similar formation mechanism to form GLSB galaxies. While neutral gas was not considered in Kulier et al. (2020), there appears a trend of a deficiency of galaxies with $M_{H I} > 10^{10}$ M$_\odot$ by comparing the H I mass functions in Crain et al. (2017) and Diemer et al. (2019).

4.4 Caveats and uncertainties

By constructing a paired control sample, we are able to select those galaxies with stellar, total gas and DM mass close to the GLSB sample. By doing so, the samples are found to be located in a more isolated environment compared to the rest of TNG100 galaxies. And we are able to identify the differences in halo formation histories and halo spin parameters for GLSB galaxies. However, a more satisfying approach would be to select a control sample sharing similar 'initial conditions' with the GLSB galaxies. One way to achieve this is to apply the method in Stopyra et al. (2021), which introduces modifications to the formation history and the environment to the initial conditions in a controlled manner. Such future research may help us to establish some definitive causal relation between major mergers and GLSB galaxies.

One of the uncertainties is inherited from the star formation model of Springel & Hernquist (2003), which prevents us from directly contrasting the modeled star formation history with the recent observations of Young, Kuzio de Naray & Wang (2015, 2020). As much of the star formation is happening in the low-density environment for GLSB galaxies, a more explicit model that resolves the multiphase structure of the ISM such as FIRE-2 (Hopkins et al. 2018) and SMUGGLE (Marinacci et al. 2019; Kannan et al. 2020b, a) would be more suitable. The candidates in our GLSB sample from this study will be a good starting point to set up zoom-in re-simulations.

In our analysis, we have not applied any radiative transfer modelling in the post-processing, which would affect the B-band luminosities. Kulier et al. (2020) presented a detailed comparison between the galaxy photometries without dust extinction (our approach is similar) with the results using a Monte Carlo radiative transfer code. The mean *B*-band surface brightness is ∼0.22 mag arcsec^{-2} dimmer with radiative transfer than that obtained with star particles alone. Moreover, galaxies with lower mean surface brightness are less affected than those with higher surface brightness. Therefore, the addition of radiative transfer is unlikely to alter our results drastically, in particular, considering that the surface photometry is performed with face-on projection. Nevertheless, post-processing with multiband radiative transfer (e.g. Li et al. 2020) could be beneficial to study the dust emission (Rahman et al. 2007) as well as the molecular hydrogen in GLSB galaxies (Gerritsen & de Blok 1999; Das et al. 2010; Galaz et al. 2022).

We build our GSLB samples with extended H I discs. Given that most observed GLSB galaxies are indeed massive H I reservoirs (Bothun et al. 1987; O'Neil et al. 2004; Hagen et al. 2016; Saburova et al. 2023), it is a reasonable choice to start with the H I component. Nevertheless, we could miss a population of gas-poor GLSB galaxies. This caveat can be addressed by using only stars, by combining radiative transfer with a flexible multicomponent fitting.

In this study, we do not discuss the central BH in our sample nor compare the $M_{bh}-\sigma$ relation with the observed one (Subramanian et al. 2016). In Pérez-Montaño et al. (2022), we find that the high angular momentum of the configuration prevents the material to fall to the central regions and feed the black holes, showing also lower accretion rates. This leads to less massive BHs in LSB galaxies when compared with HSBs at fixed stellar mass.

We did compare the GLSB sample with the galaxies sharing the same ID in the original Illustris simulation. We could not associate any of the galaxies with extended H I discs anymore. Also, we confirm that the original simulation lacks a population of galaxies with extended H I discs in general. For the same galaxy in Zhu et al. (2018), it is a massive red elliptical. Considering one of the major advancement in TNG is an improved treatment of BH outflow, GLSB galaxies turn out to be a very sensitive test of the AGN feedback model.

Based on our comparisons of mass distributions in the GLSB and the control samples, we have not considered the formation channel discussed in Kasparova et al. (2014). The essence dark matter in TNG is collisionless, with its dynamics following the potential of both dark matter and the baryons. The adopted feedback processes, from supernovae or AGN outflow, are not likely to modify the distribution of dark matter significantly for this mass range. Therefore, the large-scale radius in Malin 2 found by Kasparova et al. (2014) and Saburova et al. (2021) is outside the model and parameter space covered by TNG.

5 CONCLUSIONS

Using the cosmological hydrodynamics simulation Illustris TNG100, we study the properties of GLSB galaxies and compare them with observed ones. We start with a sample of large H I discs, perform surface photometry fitting and quantifying the properties of the outer stellar disc. We then construct a paired control sample, which shares similar total dark matter, gas, and stellar mass, to contrast with the GLSB sample. Below are some of the most important findings from our study:

(i) We successfully identify a population of galaxies with (1) extended and regular H I distributions ($R_{H I} > 50$ kpc) and (2) confirm most of them have faint and large stellar discs, which share similar properties to the observed GLSB galaxies.

(ii) The GLSB sample is a heterogeneous sample with different galaxy morphologies ranging from the most disc-dominated systems to massive ellipticals. On average, the GLSB galaxies are found in haloes with $V_{max} > 150\,\mathrm{km\,s^{-1}}$, consisting of ~ 6 per cent of total galaxies in the same stellar mass range. They are found in the red sequence and cross the green valley in the galaxy colour–magnitude diagram.

(iii) The TNG100 GLSB sample is in very good agreement in terms of total H I mass, SFR, and galaxy rotation curves with observed GLSB galaxies. Moreover, the parameters of the simulated stellar discs are found in the current gap between extreme galaxies such as Malin 1/UGC 1382 and the rest of the known GLSB galaxies.

(iv) We have not identified differences in the dark matter mass distribution between the GLSB and the control samples.

(v) We detect significant differences in molecular factions, halo spin parameters, merger histories, and their *ex situ* stellar mass fraction between the GLSB and the control samples. The spin parameter λ is ~ 40 per cent larger for GLSB galaxies than for the paired control. All the GLSB galaxies have consistently larger *ex situ* mass fractions than the paired control or 'all galaxies'. We also find the preferential alignment of cold gas angular momentum with the stellar angular momentum in the GLSB sample such that the tilt angle is usually less than 10 deg.

(vi) The distance to the fifth nearest neighbour shows that GLSB galaxies are mostly located in somewhat isolated environments. However, the paired control sample follows a similar r_{5th} distribution. Therefore, the isolation alone does not guarantee that an extended disc will/can form.

(vii) We find that aligned galaxy mergers with a combination of relatively isolated environments could explain the formation of the extended discs in our sample. A few of the GLSB candidates experienced no major merger since $z = 1$, which contributes to the longevity of the disc structure.

Since GLSB galaxies are not in the calibration processes of tuning the parameters of TNG simulations, the results presented in this study can be regarded as genuine predictions by the model. Nevertheless, much remains to be learned about these intriguing objects. Our work is only a first step in this endeavour. So far, cosmological numerical simulations seem to have no problem reproducing them.

6 DATA AVAILABILITY

The data from the IllustrisTNG simulations used in this work are publicly available at the website https://www.tng-project.org (Nelson et al. 2019). The SUBFIND IDs for the candidate GLSB galaxies and the paired control sample will be publicly available.

REFERENCES

Abraham R. G., van Dokkum P. G., 2014, PASP, 126, 55
Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Arp H., 1966, ApJS, 14, 1
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Astropy Collaboration, 2022, ApJ, 935, 167
Athanassoula E., Rodionov S. A., Peschken N., Lambert J. C., 2016, ApJ, 821, 90
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić, Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Barnes J., Efstathiou G., 1987, ApJ, 319, 575
Barnes D. G. et al., 2001, MNRAS, 322, 486
Barth A. J., 2007, AJ, 133, 1085
Beijersbergen M., de Blok W. J. G., van der Hulst J. M., 1999, A&A, 351, 903
Benavides J. A. et al., 2021, Nature Astron., 5, 1255
Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313
Boissier S. et al., 2008, ApJ, 681, 244
Boissier S. et al., 2016, A&A, 593, A126
Bosma A., Freeman K. C., 1993, AJ, 106, 1394
Bothun G. D., Impey C. D., Malin D. F., Mould J. R., 1987, AJ, 94, 23
Bothun G. D., Schombert J. M., Impey C. D., Schneider S. E., 1990, ApJ, 360, 427
Bothun G. D., Schombert J. M., Impey C. D., Sprayberry D., McGaugh S. S., 1993, AJ, 106, 530
Braine J., Herpin F., Radford S. J. E., 2000, A&A, 358, 494
Broeils A. H., Rhee M. H., 1997, A&A, 324, 877
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240
Buta R. J., 2013, Galaxy Morphology. Springer Netherlands, Dordrecht, p. 1
Cao T.-W. et al., 2017, AJ, 154, 116
Chemin L., Carignan C., Foster T., 2009, ApJ, 705, 1395
Cooper A. P. et al., 2010, MNRAS, 406, 744
Crain R. A. et al., 2017, MNRAS, 464, 4204
Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659
Das M., Kantharia N., Ramya S., Prabhu T. P., McGaugh S. S., Vogel S. N., 2007, MNRAS, 379, 11
Das M., Reynolds C. S., Vogel S. N., McGaugh S. S., Kantharia N. G., 2009, ApJ, 693, 1300
Das M., Boone F., Viallefond F., 2010, A&A, 523, A63
Davies J. I., Phillipps S., Disney M. J., 1988, MNRAS, 231, 69P
de Jong R. S., van der Kruit P. C., 1994, A&AS, 106, 451
Di Cintio A., Brook C. B., Macciò A. V., Dutton A. A., Cardona-Barrero S., 2019, MNRAS, 486, 2535
Diemer B. et al., 2018, ApJS, 238, 33
Diemer B. et al., 2019, MNRAS, 487, 1529
Disney M. J., 1976, Nature, 263, 573
Disney M., Phillipps S., 1983, MNRAS, 205, 1253
Disney M., Phillipps S., 1987, Nature, 329, 203
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Donley J. L., Koribalski B. S., Staveley-Smith L., Kraan-Korteweg R. C., Schröder A., Henning P. A., 2006, MNRAS, 369, 1741
Dressler A., 1980, ApJ, 236, 351
Erwin P., 2015, ApJ, 799, 226
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189
Fathi K., 2010, ApJ, 722, L120
Frankel N. et al., 2022, ApJ, 940, 61
Freeman K. C., 1970, ApJ, 160, 811
Galaz G., Milovic C., Suc V., Busta L., Lizana G., Infante L., Royo S., 2015, ApJ, 815, L29
Galaz G. et al., 2022, ApJ, 940, L37
Gallagher J. S., 1979, AJ, 84, 1281
Genel S., Vogelsberger M., Nelson D., Sijacki D., Springel V., Hernquist L., 2013, MNRAS, 435, 1426
Genel S. et al., 2014, MNRAS, 445, 175
Genel S., Fall S. M., Hernquist L., Vogelsberger M., Snyder G. F., Rodriguez-Gomez V., Sijacki D., Springel V., 2015, ApJ, 804, L40
Gerritsen J. P. E., de Blok W. J. G., 1999, A&A, 342, 655
Giovanelli R. et al., 2005, AJ, 130, 2598
Gottesman S. T., Hawarden T. G., 1986, MNRAS, 219, 759
Hagen L. M. Z. et al., 2016, ApJ, 826, 210
Hawarden T. G., Longmore A. J., Goss W. M., Mebold U., Tritton S. B., 1981, MNRAS, 196, 175
Hoffman Y., Silk J., Wyse R. F. G., 1992, ApJ, 388, L13
Honey M., van Driel W., Das M., Martin J. M., 2018, MNRAS, 476, 4488
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Impey C., Bothun G., 1989, ApJ, 341, 89
Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, ApJ, 689, 936
Junais et al., 2020, A&A, 637, A21
Kalberla P. M. W., Kerp J., 2009, ARA&A, 47, 27
Kannan R., Marinacci F., Simpson C. M., Glover S. C. O., Hernquist L., 2020a, MNRAS, 491, 2088
Kannan R., Marinacci F., Vogelsberger M., Sales L. V., Torrey P., Springel V., Hernquist L., 2020b, MNRAS, 499, 5732
Kasparova A. V., Saburova A. S., Katkov I. Y., Chilingarian I. V., Bizyaev D. V., 2014, MNRAS, 437, 3072
Kent S. M., 1985, ApJS, 59, 115
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Knezek P. M., 1993, PhD thesis, Massachusetts Univ., Amherst
Kniazev A. Y., Grebel E. K., Pustilnik S. A., Pramskij A. G., Kniazeva T. F., Prada F., Harbeck D., 2004, AJ, 127, 704
Koribalski B. S. et al., 2020, Ap&SS, 365, 118
Kulier A., Galaz G., Padilla N. D., Trayford J. W., 2020, MNRAS, 496, 3996
Lelli F., Fraternali F., Sancisi R., 2010, A&A, 516, A11
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Li Y., Gu M. F., Yajima H., Zhu Q., Maji M., 2020, MNRAS, 494, 1919
Lutz K. A. et al., 2018, MNRAS, 476, 3744
Mapelli M., Moore B., Ripamonti E., Mayer L., Colpi M., Giordano L., 2008, MNRAS, 383, 1223
Marinacci F., Pakmor R., Springel V., 2014, MNRAS, 437, 1750
Marinacci F., Grand R. J. J., Pakmor R., Springel V., Gómez F. A., Frenk C. S., White S. D. M., 2017, MNRAS, 466, 3859
Marinacci F. et al., 2018, MNRAS, 480, 5113
Marinacci F., Sales L. V., Vogelsberger M., Torrey P., Springel V., 2019, MNRAS, 489, 4233
Martin G. et al., 2019, MNRAS, 485, 796
Matthews L. D., van Driel W., Monnier-Ragaigne D., 2001, A&A, 365, 1
McGaugh S. S., Bothun G. D., 1994, AJ, 107, 530
McGaugh S. S., Bothun G. D., Schombert J. M., 1995, AJ, 110, 573
Mishra A., Kantharia N. G., Das M., Srivastava D. C., Vogel S. N., 2015, MNRAS, 447, 3649
Mishra A., Kantharia N. G., Das M., Omar A., Srivastava D. C., 2017, MNRAS, 464, 2741
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Nelson E. J. et al., 2021, MNRAS, 508, 219
Noguchi M., 2001, MNRAS, 328, 353

O'Neil K., Schinnerer E., 2004, ApJ, 615, L109
O'Neil K., Bothun G., van Driel W., Monnier Ragaigne D., 2004, A&A, 428, 823
O'Neil K., Oey M. S., Bothun G., 2007, AJ, 134, 547
Patton D. R., Qamar F. D., Ellison S. L., Bluck A. F. L., Simard L., Mendel J. T., Moreno J., Torrey P., 2016, MNRAS, 461, 2589
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Peebles P. J. E., 1969, ApJ, 155, 393
Peñarrubia J., McConnachie A., Babul A., 2006, ApJ, 650, L33
Peng Y.-j. et al., 2010, ApJ, 721, 193
Pérez-Montaño L. E., Cervantes Sodi B., 2019, MNRAS, 490, 3772
Pérez-Montaño L. E., Rodriguez-Gomez V., Cervantes Sodi B., Zhu Q., Pillepich A., Vogelsberger M., Hernquist L., 2022, MNRAS, 514, 5840
Peschken N., Athanassoula E., Rodionov S. A., 2017, MNRAS, 468, 994
Peschken N., Łokas E. L., Athanassoula E., 2020, MNRAS, 493, 1375
Pickering T. E., Impey C. D., van Gorkom J. H., Bothun G. D., 1997, AJ, 114, 1858
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Portas A. M., Brinks E., Filho M. E., Usero A., Dyke E. M., Belles P. E., 2010, MNRAS, 407, 1674
Rahman N., Howell J. H., Helou G., Mazzarella J. M., Buckalew B., 2007, ApJ, 663, 908
Read J. I., 2014, J. Phys. G Nucl. Phys., 41, 063101
Reshetnikov V. P., Moiseev A. V., Sotnikova N. Y., 2010, MNRAS, 406, L90
Richard J. et al., 2019, preprint (arXiv:1906.01657)
Rodriguez-Gomez V. et al., 2015, MNRAS, 449, 49
Rodriguez-Gomez V. et al., 2016, MNRAS, 458, 2371
Rodriguez-Gomez V. et al., 2022, MNRAS, 512, 5978
Romanishin W., Strom K. M., Strom S. E., 1983, ApJS, 53, 105
Rosenbaum S. D., Bomans D. J., 2004, A&A, 422, L5
Saburova A. S., Chilingarian I. V., Katkov I. Y., Egorov O. V., Kasparova A. V., Khoperskov S. A., Uklein R. I., Vozyakova O. V., 2018, MNRAS, 481, 3534
Saburova A. S., Chilingarian I. V., Kasparova A. V., Katkov I. Y., Fabricant D. G., Uklein R. I., 2019, MNRAS, 489, 4669
Saburova A. S., Chilingarian I. V., Kasparova A. V., Sil'chenko O. K., Grishin K. A., Katkov I. Y., Uklein R. I., 2021, MNRAS, 503, 830
Saburova A. S., Chilingarian I. V., Kulier A., Galaz G., Grishin K. A., Kasparova A. V., Toptun V., Katkov I. Y., 2023, MNRAS, 520, L85
Saha K., Dhiwar S., Barway S., Narayan C., Tandon S., 2021, J. Astrophys. Astron., 42, 59
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schaye J. et al., 2015, MNRAS, 446, 521
Schombert J., 1998, AJ, 116, 1650
Schombert J. M., Bothun G. D., Schneider S. E., McGaugh S. S., 1992, AJ, 103, 1107
Semczuk M., Łokas E. L., D'Onghia E., Athanassoula E., Debattista V. P., Hernquist L., 2020, MNRAS, 498, 3535
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Smethurst R. J. et al., 2015, MNRAS, 450, 435
Sotillo-Ramos D., Donnari M., Pillepich A., Frankel N., Nelson D., Springel V., Hernquist L., 2023, MNRAS, in press
Sparre M., Springel V., 2017, MNRAS, 470, 3946
Sprayberry D., Impey C. D., Bothun G. D., Irwin M. J., 1995, AJ, 109, 558
Springel V., 2010, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., Hernquist L., 2005, ApJ, 622, L9
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2018, MNRAS, 475, 676
Stevens A. R. H. et al., 2021, MNRAS, 502, 3158
Stopyra S., Pontzen A., Peiris H., Roth N., Rey M. P., 2021, ApJS, 252, 28
Subramanian S., Ramya S., Das M., George K., Sivarani T., Prabhu T. P., 2016, MNRAS, 455, 3148
Tacchella S. et al., 2019, MNRAS, 487, 5416

Thilker D. A. et al., 2007, ApJS, 173, 538

Torrey P. et al., 2015, MNRAS, 447, 2753

van der Kruit P. C., 1987, A&A, 173, 59

van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

Vogelsberger M. et al., 2014a, MNRAS, 444, 1518

Vogelsberger M. et al., 2014b, Nature, 509, 177

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Rev. Phys., 2, 42

Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, MNRAS, 460, 2143

Weinberger R. et al., 2017, MNRAS, 465, 3291

Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32

Yahya S., Bull P., Santos M. G., Silva M., Maartens R., Okouma P., Bassett B., 2015, MNRAS, 450, 2251

Young J. E., Kuzio de Naray R., Wang S. X., 2015, MNRAS, 452, 2973

Young J. E., Kuzio de Naray R., Wang S. X., 2020, MNRAS, 493, 55

Yun K. et al., 2019, MNRAS, 483, 1042

Zeng G., Wang L., Gao L., 2021, MNRAS, 507, 3301

Zhao D., Du M., Ho L. C., Debattista V. P., Shi J., 2020, ApJ, 904, 170

Zhu Q. et al., 2018, MNRAS, 480, L18

Zjupa J., Springel V., 2017, MNRAS, 466, 1625

APPENDIX A: COMPARISON OF 'SÉRSIC + DISC' AND 'TWO-DISC' FITTINGS

In Section 2.2.2, the central part of the galaxy is modelled with a Sérsic profile while the outer part with a disc profile. In this section, we fit the face-on projected images in *B* band using a sum of two exponential profiles as

$$\Sigma(r) = I_{\mathrm{in}} \exp\left(-r/r_{\mathrm{d,in}}\right) + I_{\mathrm{out}} \exp\left(-r/r_{\mathrm{d,out}}\right), \tag{A1}$$

which involves two discs with scale-length $r_{\mathrm{d,in}}$ and $r_{\mathrm{d,out}}$. We then compare the outer disc parameters derived from these two models to assess the impacts of the underlying assumption of the galaxy components.
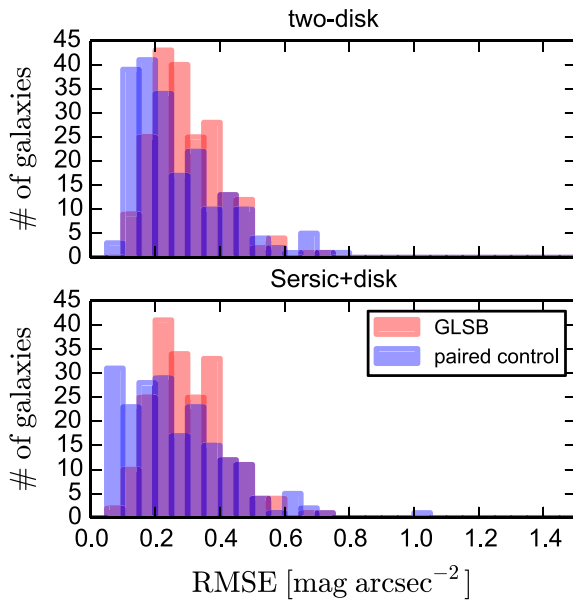

Figure A1. Distribution of root mean square error (RMSE) with to the 'two-disc' (upper) and 'Sérsic + disc' (bottom) fittings to the GLSB and paired control samples.
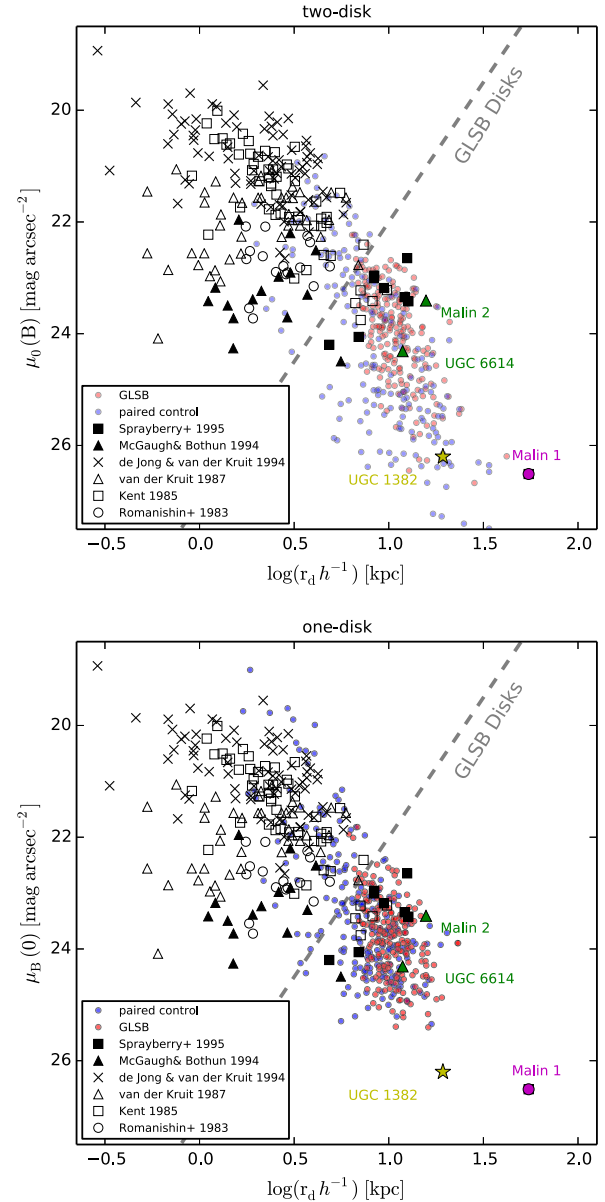
Figure A2. The same as Fig. 8, but with disc parameters derived using a 'two-disc' (upper) and a single exponential profile (bottom).

Fig. A1 shows the distribution of RMSE of the surface photometry fittings of the entire GLSB and paired control samples, which yields a reasonably good description of the face-on *B*-band profiles. With the 'Sérsic + disc' model, the median and the average of RMSE are 0.28 and 0.30 mag arcsec^{-2} for GLSB and 0.23 and 0.27 for the paired control. With 'two-disc,' the median and the average are 0.28 and 0.30 for GLSB, 0.23 and 0.26 for the paired control sample. The 'Sérsic + disc' slightly better fits the paired control galaxies than the 'two-disc' model. For the GLSB sample, the two models give the same quality of fittings regarding RMSE.

The top panel of Fig. A2 then compares the disc parameters derived from the 'two-disc' model with observations, which is very close to what is obtained in Fig. 8. The majority of the GLSB sample can be characterized with GLSB discs. For the paired control sample, the disc parameters span a much wider range, covering both the low- and high-surface brightness regimes.

We also fit a single exponential disc in the range of 25–28 mag arcsec^{-2}. This particular choice is inspired by the recent study by Saburova et al. (2023). The bottom panel of Fig. A2 displays the parameters of the outer disc from this exercise. Some of the faintest and the most extended discs from the 'Sérsic + disc' or 'two-disc' models are no longer present. The distribution of the disc parameters is more concentrated around the observed GLSB galaxies, with a gap to fill to reach UGC 1382 and Malin 1. Still, most galaxies in the GLSB sample are consistent with the definition of GLSB discs. For comparison, the paired control sample spans both the low- and high-surface brightness discs.

We believe the difference seen between 'two-disc' and 'one-disc' models is due to the contribution of any non-disc component to the faint discs. The 'two-disc' or 'Sérsic + disc' model subtracts the contribution from the inner galaxy quite well.

APPENDIX B: IMPACTS OF DIFFERENT PAIRED CONTROL SAMPLE CONSTRUCTIONS

Having a paired control sample is beneficial to remove the biases due to galaxy mass functions. We rely on a paired control sample using the nearest neighbour in [$\log(M_*)$, $\log M_{DM}$, $\log M_{gas}$]. This is motivated by the observed rotation curves for GLSB galaxies (e.g. Lelli et al. 2010; Saburova et al. 2021). Given the total mass budget in each component, such a choice allows us to study the internal mass distribution.

This section further explores the impacts of using different control parameters on the results. We construct two separate paired controls, one with stellar mass only $\log(M_*)$ and one with $\log(M_{DM})$.

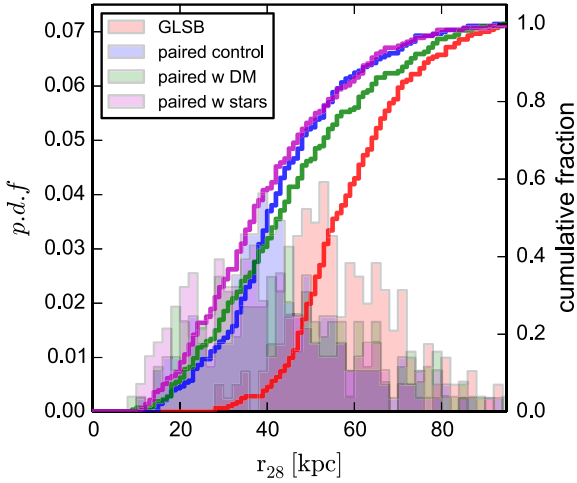**Figure B1.** The same as Fig. 5, but with the paired control sample constructed using different criteria. The 'paired with DM' only considers the dark matter mass, while the 'paired with stars' only takes the total stellar mass M_* in to account. The distributions of r_{28} in the three different paired controls are overall consistent with each other.
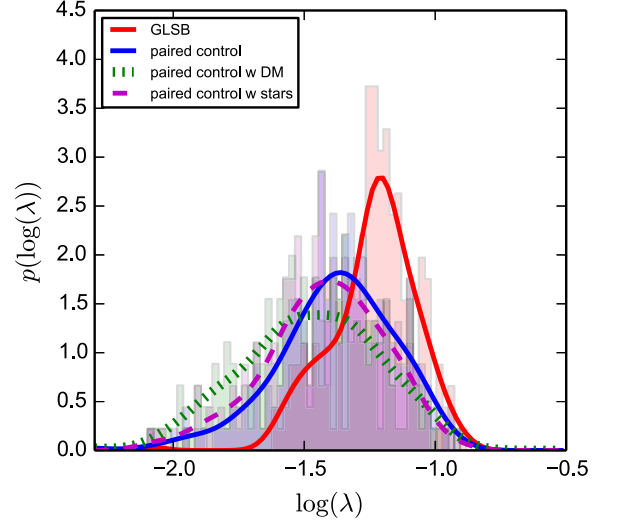
Figure B2. The distributions of halo spin parameters λ for the GLSB and three paired control samples.
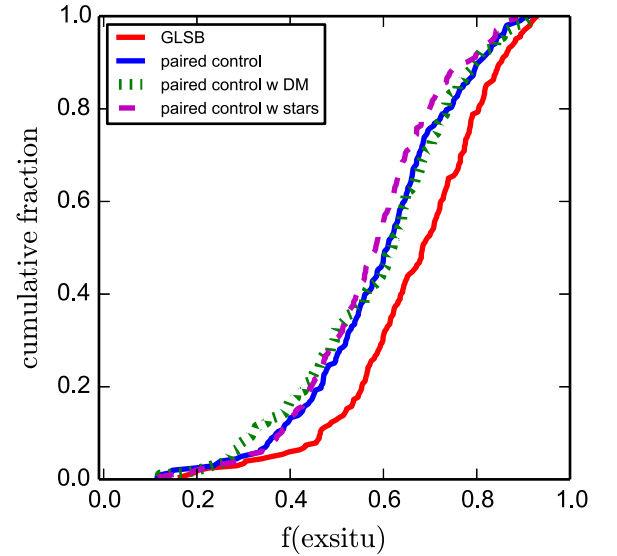
Figure B3. The distributions of *ex situ* stellar fractions for the GLSB and three paired control samples.

Fig. B1 compares the distribution of r_{28}, a proxy for galaxy size, between different samples. Despite being constructed with different parameters, three paired controls show consistent galaxy size distributions. The median values are 40, 43, and 37 kpc, respectively.
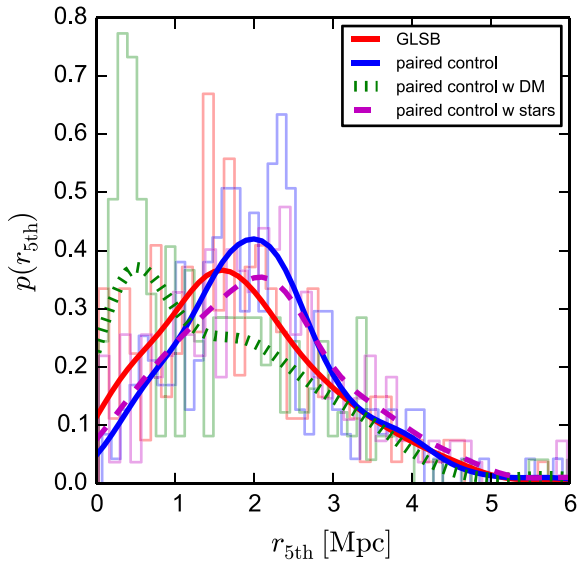
Figure B4. The distribution of r_{5th} for the GLSB and paired control samples. Unlike Figs B3 or B2, significant differences are present between the 'paired control with DM' and 'paired control' samples with a peak occurring at r_{5th} < 1 Mpc. On the other hand, the 'paired control with stars' agrees with the 'paired control' sample.

Fig. B2 shows the spin parameters for each sample at $z = 0$. Similarly in Fig. B3, the distribution of *ex situ* mass fractions are consistent between three paired controls.

Fig. B4 shows considerable differences in the environment. In particular, 'paired control with DM' peaks at 0.5 Mpc while 'paired control with stars' at $\sim$2 Mpc. This could be caused by the scatter in the $M_{DM} - M*$ relations, with different mass removal rates for DM and stars. Stars are less susceptible than DM to galactic tides. As we do not find GLSB galaxies in the dense environment, the differences we see in Fig. B4 should have minimum impact on our conclusions.

APPENDIX C: FULL GLSB CANDIDATE SAMPLE

Figs C1 and C2 show all 203 TNG100 galaxies in our final GLSB sample. The limiting surface brightness in U, B, and K are 28, 28, and 25 mag arcsec^{-2}, the same as in Fig. 6.

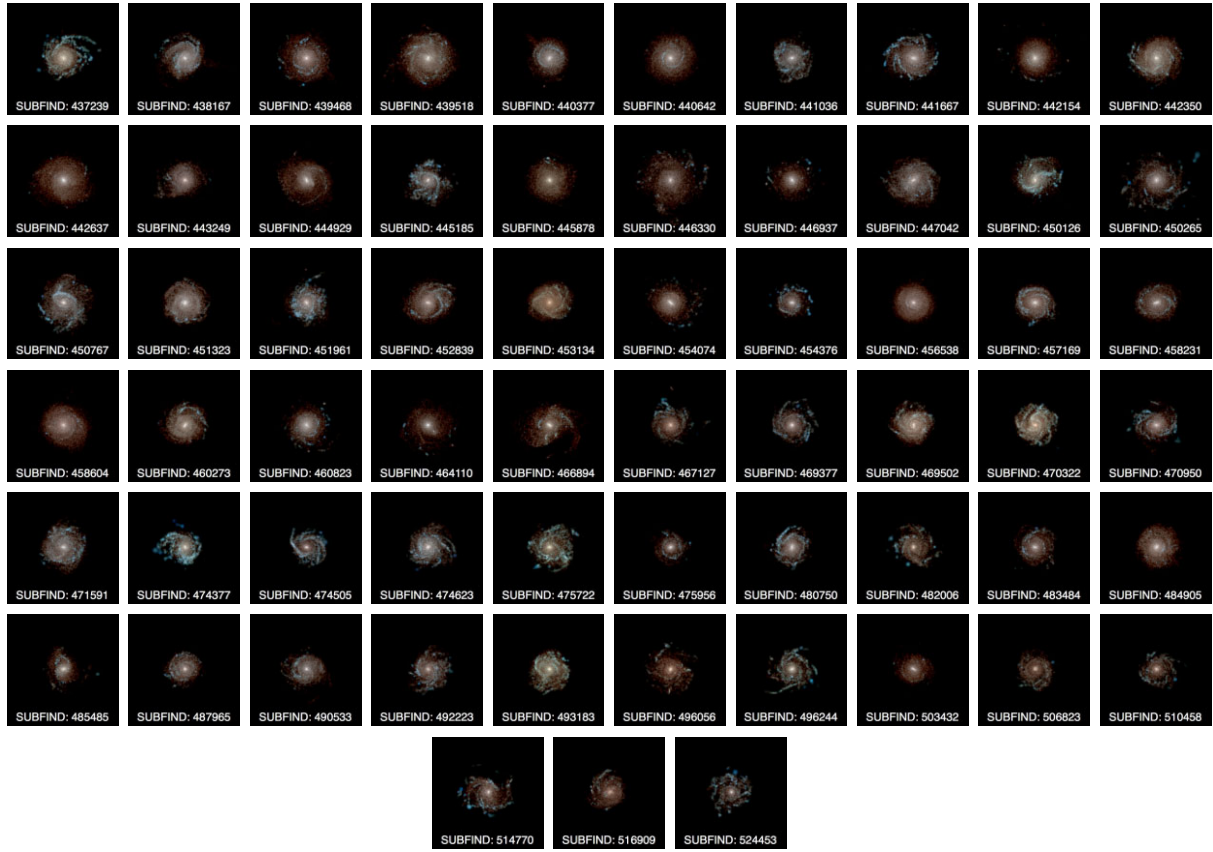**Figure C1.** Atlas of the TNG100 GLSB sample used in our study.

Figure C2. Continued. Atlas of the TNG100 GLSB sample used in our study.

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