

# HI asymmetries in spatially resolved SIMBA galaxies

Nadine A. N. Hank<sup>1</sup>,<sup>\*</sup> Marc A. W. Verheijen<sup>1</sup>, Sarah-L. Blyth<sup>2</sup>, Romeel Davé<sup>3,4</sup>,  
Kyle A. Oman<sup>5,6</sup>, Nathan Deg<sup>7</sup> and Marcin Glowacki<sup>3,8,9</sup>

<sup>1</sup>*Kapteyn Astronomical Institute, University of Groningen, Landleven 12, NL-9747 AD Groningen, the Netherlands*

<sup>2</sup>*Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*

<sup>3</sup>*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*

<sup>4</sup>*Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa*

<sup>5</sup>*Institute for Computational Cosmology, Physics Department, Durham University, South Road, Durham DH1 3LE, UK*

<sup>6</sup>*Centre for Extragalactic Astronomy, Physics Department, Durham University, South Road, Durham DH1 3LE, UK*

<sup>7</sup>*Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada*

<sup>8</sup>*International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia*

<sup>9</sup>*Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, Cape Town 7701, South Africa*

Accepted 2025 May 22. Received 2025 May 7; in original form 2024 December 29

## ABSTRACT

We present a study of the neutral atomic hydrogen (HI) content of spatially resolved, low-redshift galaxies in the SIMBA cosmological simulations. We create synthetic HI data cubes designed to match observations from the Apertif Medium-Deep HI imaging survey, and follow an observational approach to derive the HI size–mass relation. The HI size–mass relation for SIMBA is in broad agreement with the observed relation to within 0.1 dex, but SIMBA galaxies are slightly smaller than expected at fixed HI mass. We quantify the HI spectral ( $A_{\text{flux}}$ ) and morphological ( $A_{\text{mod}}$ ) asymmetries of the galaxies and motivate standardizing the relative spatial resolution when comparing  $A_{\text{mod}}$  values in a sample that spans several orders of magnitude in HI mass. Galaxies are classified into three categories (isolated, interacted, or merged) based on their dynamical histories over the preceding  $\sim 2$  Gyr to contextualize disturbances in their HI reservoirs. We determine that the interacted and merged categories have higher mean asymmetries than the isolated category, with a larger separation between the categories'  $A_{\text{mod}}$  distributions than between their  $A_{\text{flux}}$  distributions. For the interacted and merged categories, we find an inverse correlation between baryonic mass and  $A_{\text{mod}}$  that is not observed between baryonic mass and  $A_{\text{flux}}$ . These results, coupled with the weak correlation found between  $A_{\text{flux}}$  and  $A_{\text{mod}}$ , highlight the limitations of only using  $A_{\text{flux}}$  to infer the HI distributions of spatially unresolved HI detections.

**Key words:** galaxies: evolution – galaxies: interactions – galaxies: ISM – radio lines: galaxies.

## 1 INTRODUCTION

It has long been observed that peculiar and disturbed morphologies are prevalent in the stellar (e.g. Arp 1966) and/or gas (e.g. Baldwin, Lynden-Bell & Sancisi 1980; Hibbard et al. 2001) distributions of galaxies. These asymmetries can arise as observational signatures of the myriad physical processes galaxies undergo over cosmic time, which also drive changes in other galaxy properties [e.g. colour, stellar mass, gas content, star formation rate (SFR), etc.]. A number of these processes are associated with different environments and can have a significant impact on a galaxy's neutral atomic hydrogen (HI) gas content, since it is susceptible to both gravitational and hydrodynamic forces. In low-density environments, gas accretion from the cosmic web (e.g. Bournaud et al. 2005; Sancisi et al. 2008) and accretion of low-mass companion galaxies (Di Teodoro & Fraternali 2014) can replenish the gas in a galaxy, while fly-by interactions (Mapelli, Moore & Bland-Hawthorn 2008) can remove gas from the outskirts of a galaxy. In high-density environments,

processes such as ram-pressure stripping (Gunn & Gott 1972), viscous stripping (Nulsen 1982), and galaxy harassment (Moore et al. 1996) deplete a galaxy's gas reservoir. Thus, studying HI asymmetries, which have also been observed in the integrated HI spectra (Peterson & Shostak 1974; Haynes et al. 1998; Espada et al. 2011) and the HI kinematics (Swaters et al. 1999; Noordermeer, Sparke & Levine 2001; Chemin et al. 2006; van Eymeren et al. 2011a) of galaxies, could enhance our understanding of the evolutionary processes that these galaxies are undergoing.

One of the first studies focusing on HI asymmetries was undertaken by Peterson & Shostak (1974), who examined the global HI profiles of 23 galaxies with peculiar optical morphologies. They found a general correlation between the degree of asymmetry in the HI profiles of these galaxies and the degree of asymmetry observed in their optical images, and speculated that the asymmetries were induced by tidal interactions or the presence of warps in the HI discs. As HI synthesis observations of nearby galaxies became available, Baldwin et al. (1980) were the first to highlight that large-scale morphological asymmetries are not just a feature observed in stellar discs after studying the spatial HI distributions of  $\sim 20$  spiral galaxies. They reserved the term 'lopsided' for any galaxy

\* E-mail: [nhank@astro.rug.nl](mailto:nhank@astro.rug.nl)

whose H I distribution is disproportionately skewed towards one side of the galaxy, and proposed a kinematic model in which the observed lopsidedness is connected to off-centre elliptical orbits. Using data from a range of single-dish H I surveys, Richter & Sancisi (1994) visually examined the H I profiles of  $\sim 1700$  disc galaxies and found that at least 50 per cent displayed strong or moderate asymmetric H I profiles, suggesting that asymmetries are far more common than previously assumed. Since their sample consisted mainly of field galaxies to minimize the inclusion of interacting systems, the authors inferred that asymmetries are long-lived, as opposed to frequently induced signatures. A similar result was found by Haynes et al. (1998), who used the ratio of the integrated flux in two halves of the global H I profiles to quantify asymmetry in 104 isolated galaxies, and found that  $\sim 50$  per cent of their sample was considerably asymmetric. Matthews, van Driel & Gallagher (1998) found an even higher occurrence of asymmetry, albeit in a much smaller sample of 30 late-type spiral galaxies, with  $\sim 77$  per cent displaying some degree of asymmetry in their H I profiles. Although many of these studies were carried out using targeted galaxy samples, which tend to suffer from sample selection limitations, they remain crucial to our wider understanding of the ubiquity of H I asymmetries. Nevertheless, untargeted H I surveys are essential to get a more unbiased view of H I asymmetries in the general galaxy population.

The latest generation of untargeted H I surveys on the Square Kilometre Array (SKA) precursor and pathfinder telescopes has been made possible by technological advances in the radio wavelength domain over the last two decades. Consequently, significant improvements in observing capabilities have been made. These surveys, which are arranged in a three-tiered ‘wedding cake’ configuration in terms of their depth and sky area, will enable detailed studies exploring the evolution of H I in galaxies using much larger samples than were previously available. Shallow surveys such as the Wide-field ASKAP L-band Legacy All-sky Blind survey (WALLABY, Koribalski et al. 2020) with the Australian SKA Pathfinder (ASKAP, Johnston et al. 2008; Hotan et al. 2021) and the Shallow Northern-sky Survey (Adams & van Leeuwen 2019) with the APERTure Tile In Focus (Apertif, Verheijen et al. 2008; van Cappellen et al. 2022) phased array feed system on the Westerbork Radio Synthesis Telescope (WSRT) form the bottom tier and will spatially resolve hundreds to thousands of nearby galaxies in H I. The middle tier consists of the H I emission project within the MeerKAT International GHz Tiered Extragalactic Exploration (arvis et al. 2016; Maddox et al. 2021) survey and the Medium-Deep H I imaging Survey (MDS, Verheijen et al. 2009) with Apertif, which will observe local volumes to greater depth. The top tier of this arrangement is occupied by the Deep Investigation of Neutral Gas Origins (Meyer 2009) survey with ASKAP and the Looking At the Distant Universe with the MeerKAT Array (Blyth et al. 2016) survey, which will provide direct H I detections out to  $z < 1.4$ . Other significant surveys aimed at directly detecting H I in galaxies beyond  $z = 0.15$  include the Blind Ultra-Deep H I environmental Survey (BUDHIES, Jaffé et al. 2013) with the WSRT, which observed H I in two  $z \sim 0.2$  galaxy clusters, the HIGHz Arcicbo survey (Catinella & Cortese 2015), which targeted massive, relatively isolated galaxies at similar redshifts to BUDHIES, and the COSMOS H I Large Extragalactic Survey (Fernández et al. 2013), which utilized the increased frequency coverage of the upgraded Karl G. Jansky Very Large Array and more than 1000 h of observing time to push the redshift limit of H I imaging studies out to  $z = 0.45$ . Collectively, these surveys will enable much larger resolved galaxy samples at both low and high redshift to be studied for the first time. As a result, there has been

renewed interest in H I asymmetries in the lead up to these new data sets.

Espada et al. (2011) utilized the asymmetry index introduced in Haynes et al. (1998) to study the distribution of spectral asymmetry in a sample of extremely isolated disc galaxies from the Analysis of the interstellar Medium of Isolated GALaxies (AMIGA, Verdes-Montenegro et al. 2005) project, and found that the distribution could be described by a half-Gaussian with a mean of 1.0 and a standard deviation of  $\sigma = 0.13$ . They classified a galaxy as asymmetric if the measured value exceeded the  $2\sigma$  threshold of their sample, and found that only 9 per cent of their sample met this criterion. Subsequent studies using the same asymmetry metric have adopted the Espada et al. (2011)  $2\sigma$  and  $3\sigma$  thresholds to calculate the fractions of highly asymmetric global H I profiles in their samples. For example, Scott et al. (2018) used the  $3\sigma$  threshold and measured asymmetry fractions of 16 per cent and 26 per cent for galaxies with H I detections in the Virgo and A1367 clusters, respectively. Bok et al. (2019) compared samples of close galaxy pairs and isolated galaxies selected from the Arcicbo Legacy Fast ALFA (Haynes et al. 2018) survey, and found an enhanced asymmetry fraction in their close pair sample (27 per cent) relative to their isolated sample (18 per cent). Watts et al. (2020a) analysed the asymmetries of central and satellite galaxies from the extended GALEX Arcicbo SDSS Survey (Catinella et al. 2018) and found that satellite galaxies are generally more asymmetric than central galaxies. These results suggest that a correlation exists between the environment and the incidence of asymmetry. However, it remains unclear which astrophysical processes are primarily responsible for these asymmetries.

A number of recent studies have introduced additional techniques to quantify H I profile asymmetries (e.g. Deg et al. 2020; Reynolds et al. 2020; Yu, Ho & Wang 2020). The use of these new methods in combination with the widely used asymmetry index of Haynes et al. (1998) could offer valuable improvements in characterizing the H I profiles that will be available in large numbers. However, maximizing the amount of information obtained from H I profiles requires more knowledge of the underlying H I distributions, since the shape of the H I profile depends largely on the internal kinematics of the galaxy and to a lesser extent the spatial H I distribution.

There are comparatively fewer quantitative asymmetry studies using H I images of galaxies due to the lack of spatially resolved H I observations available relative to H I spectra. However, various methods that were developed for optical and near-infrared data have been adopted in H I studies in recent years. One such method is the use of the normalized amplitude of the  $m = 1$  Fourier component (see Jog & Combes 2009, for a comprehensive review on this method). This technique was applied by Angiras et al. (2006, 2007) to investigate morphological H I asymmetries in the Eridanus group and the Ursa Major volume. They measured asymmetric fractions of  $\sim 17$ – $27$  per cent and found that the average asymmetry values measured in the outer regions of the H I in the Eridanus sample was greater than the Ursa Major sample. Subsequently, van Eymeren et al. (2011b) used this method on a galaxy sample of from the Westerbork H I survey of Irregular and SPiral galaxies (WHISP, van der Hulst, van Albada & Sancisi 2001) galaxies and found a continued increase in morphological asymmetry with radius for 20 per cent of galaxies in their sample.

Another method is the use of the concentration, asymmetry, smoothness, Gini, and  $M_{20}$  non-parametric measurements (Abraham, van den Bergh & Nair 2003; Conselice 2003; Lotz, Primack & Madau 2004). Holwerda et al. (2011a, b) were the first to apply these measurements to H I observations from The H I Nearby Galaxy Survey (Walter et al. 2008) and WHISP, and found that the asymmetry

parameter was the most promising at identifying interacting galaxies in their sample. Following from this, Giese et al. (2016) used simulated galaxies in order to better understand the observational effects on these non-parametric measurements. They observed that asymmetry was the ideal parameter to identify galaxies with lopsided H I distributions, and determined a set of reliable resolution and inclination limits for the use of these parameters in upcoming large H I surveys. Recent studies have also explored the use of these non-parametric measurements with machine learning techniques to identify morphologically disturbed galaxies in the WALLABY Pilot Survey (e.g. Holwerda et al. 2023, 2025).

Wang et al. (2013) studied 23 nearby galaxies with high H I mass fractions and compared them with a control sample of 19 galaxies. They found little difference in the overall asymmetry distributions of the populations, but noted that the outermost regions of the H I-rich galaxies tended to be more asymmetric. Lelli, Verheijen & Fraternali (2014) developed a modified outer asymmetry parameter that is more sensitive to low surface brightness regions on the outskirts of galaxies. They found that this modified asymmetry parameter was able to distinguish between samples of starburst dwarf and irregular galaxies.

More recently, Reynolds et al. (2020) investigated the spectral, morphological, and kinematic H I asymmetries in  $\sim 140$  galaxies from three different surveys. They observed a slight trend of increasing incidence of asymmetry with increasing environment density, but found no correlation with H I mass. They also found weak to moderate correlations between various asymmetry parameters, suggesting that the spectral asymmetry parameters are sufficient for spatially unresolved data, when the other parameters cannot be used. However, they concluded that a larger sample was required to confirm these trends.

In addition to these observational studies, a number of recent studies have used large-scale cosmological hydrodynamical simulations to investigate H I asymmetries. These simulations offer significantly larger samples of spatially resolved galaxies, spanning a range of properties and cosmic environments, than are currently available with H I surveys. Furthermore, they have the added benefit of providing detailed information about the galaxies' environments, their merger histories, and their progenitors at earlier cosmic epochs. Thus, simulations can offer important insights into the underlying astrophysical processes driving asymmetries.

Watts et al. (2020b) utilized the ILLUSTRISTNG simulations (Nelson et al. 2018) to examine the influence of environment on the profile asymmetries of  $\sim 10\,500$  galaxies, using halo mass as a proxy for environment. They observed a higher incidence of profile asymmetry among the satellite galaxy population than the central galaxy population, and found that this is driven by satellite galaxies residing in more massive haloes analogous to large galaxy groups. Using  $\sim 3500$  galaxies from the EAGLE simulations (Schaye et al. 2015), Manuwal et al. (2022) also found that the H I profiles of satellite galaxies are generally more asymmetric than those of central galaxies, and determined that ram-pressure and tidal stripping are the strongest drivers of satellite asymmetry. Glowacki et al. (2022) used the SIMBA simulations (Davé et al. 2019) to investigate trends between profile asymmetries and galaxy properties for  $\sim 4000$  galaxies. They found that the strongest correlation is between H I mass and profile asymmetry, noting that the occurrence of profile asymmetries decreases with increasing H I mass.

Bilimogga et al. (2022) created mock H I observations of  $\sim 190$  galaxies from the EAGLE simulations to investigate the combined effect of observational effects on spectral and morphological H I asymmetries. They established various limits required for the reso-

lution, H I column density sensitivity and signal-to-noise ratio of an observation to ensure robust asymmetry measurements. They also found no correlation between the spectral and morphological H I asymmetries in their sample, seemingly in contradiction with the findings of Reynolds et al. (2020). Given that these studies both utilized relatively small ( $N < 200$ ) samples, these conflicting results illustrate that a much larger sample is required to assess whether a correlation exists.

While various studies have recognized cosmological simulations as a useful tool to investigate H I asymmetries and their limitations, research has yet to capitalize on the quantity of galaxies with spatially resolved H I that is available with cosmological simulations. In this paper, we examine the H I spectral and morphological asymmetries of  $\sim 1100$  spatially resolved, low-redshift galaxies from the SIMBA cosmological simulations to understand and contextualize the state of disturbances in the H I reservoirs of galaxies. Here, we consider the dynamical histories of galaxies over the preceding  $\sim 2$  Gyr, with a focus on investigating whether galaxy–galaxy interactions and mergers are the physical drivers of the H I asymmetries observed in our sample at  $z \sim 0$ . We emphasize that our aim is not to determine the visibility time-scales over which these H I asymmetries persist or to estimate merger rates using H I data, but to compare and contrast the distributions of H I spectral and morphological asymmetries of galaxies with different dynamical histories. In addition, we explore possible correlations between H I asymmetry and baryonic mass in our simulated sample.

SIMBA, which was calibrated to match the galaxy stellar mass function at  $z = 0$ , successfully reproduces a number of galaxy observations at low redshifts (e.g. Davé et al. 2019; Li, Narayanan & Davé 2019; Thomas et al. 2019; Appleby et al. 2020; Glowacki, Elson & Davé 2020; Cui et al. 2021). Most importantly for this work, SIMBA has been shown to reproduce the atomic and molecular gas contents of galaxies and shows good agreement with the H I mass function at  $z = 0$  (Davé et al. 2020). We use the MARTINI package (Oman 2024) to generate synthetic H I data cubes designed to match observations from the MDS. In doing so, we aim to enable a direct comparison between the H I content of our simulated sample and that of the observed galaxy populations from the MDS, which will be explored in a future study.

This paper is structured as follows. In Section 2, we describe the key characteristics of the SIMBA simulations, the galaxy sample selection, the creation of mock H I data cubes, and the comparison of our sample to the observed H I size–mass relation. Section 3 details the parameters used in this work to quantify spectral and morphological asymmetries, and our justification for standardizing the relative spatial resolution across our sample when measuring and comparing morphological asymmetries. In Section 4, we present our main results, and compare our findings with existing studies in Section 5. Finally, our conclusions are presented in Section 6. Throughout this paper, we use  $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2 DATA

### 2.1 The SIMBA simulations

The SIMBA simulations (Davé et al. 2019) consist of a suite of cosmological hydrodynamical simulations of galaxy formation. These simulations solve for the co-evolution of galaxies, black holes, and intergalactic gas using a branched version of GIZMO (Hopkins 2015), a gravity + hydrodynamics solver based on GADGET-2 (Springel 2005), in its meshless finite-mass variant. In this study, we use the fiducial simulation (hereafter SIMBA-100), which evolves

a representative  $(100 h^{-1} \text{ Mpc})^3$  comoving volume initialized with  $1024^3$  dark matter particles and  $1024^3$  gas elements down to  $z = 0$ . The resulting mass resolution is  $9.6 \times 10^7 M_\odot$  for dark matter particles and  $1.82 \times 10^7 M_\odot$  for gas elements. We supplement this with the high-resolution volume (hereafter SIMBA-25), which has identical input physics to SIMBA-100 but has  $8 \times$  the mass resolution ( $1.2 \times 10^7 M_\odot$  for dark matter particles and  $2.28 \times 10^6 M_\odot$  for gas elements). This enables us to include lower mass galaxies in our sample. We refer the reader to Davé et al. (2019) for a comprehensive description of the simulations, and focus here on the key aspects of SIMBA and those most relevant to this study.

SIMBA follows a  $\Lambda$ -cold dark matter cosmology with parameters based on the Planck Collaboration XIII (2016) results:  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ ,  $\Omega_b = 0.048$ ,  $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\sigma_8 = 0.82$ , and  $n_s = 0.97$ . The initial conditions are generated at a starting redshift of  $z = 249$  using the MUSIC code (Hahn & Abel 2011). All simulation runs output 151 snapshots from  $z = 20 \rightarrow 0$ , with a snapshot spacing of  $\sim 230 \text{ Myr}$  between  $z = 0$  and 0.1.

As the next generation of the MUFASA simulations (Davé, Thompson & Hopkins 2016), SIMBA builds on the work of its predecessor by incorporating a unique two-mode accretion model for black holes and implementing various improvements to the subgrid prescriptions for star formation and feedback from active galactic nuclei. For cold gas ( $T < 10^5 \text{ K}$ ), black hole growth is implemented via a torque-limited accretion model (Hopkins & Quataert 2011; Anglés-Alcázar et al. 2017), where instabilities in the cold gaseous disc give rise to gravitational torques which drive gas inflows, while Bondi accretion (Bondi 1952) is adopted for hot gas ( $T > 10^5 \text{ K}$ ). The accretion energy drives feedback that results in the quenching of galaxies, with a kinetic subgrid model for black hole feedback along with X-ray energy feedback included in SIMBA.

Radiative cooling and photoionization heating are incorporated using the GRACKLE-3.1 library (Smith et al. 2017), which includes metal cooling and non-equilibrium evolution of primordial elements. The chemical enrichment model follows 11 elements (H, He, and nine metals), with enrichment tracked from Type Ia and Type II supernovae as well as asymptotic giant branch stars. A Chabrier (2003) initial mass function is used to compute stellar evolution.

The total cold ( $T < 10^5 \text{ K}$ ) neutral ( $\text{H I} + \text{H}_2$ ) gas fraction is calculated self-consistently during the simulation runs by accounting for self-shielded gas on-the-fly, in lieu of applying self-shielding in the post-processing stage, as was the case in MUFASA (Davé et al. 2017). This is done using the prescription described in Rahmati et al. (2013), where the metagalactic ionizing flux strength is attenuated based on the gas density, and assuming a spatially uniform ionizing background as given in Haardt & Madau (2012).

The  $\text{H}_2$  mass fraction for each gas element is given by

$$f_{\text{H}_2} = 1 - 0.75 \frac{s}{1 + 0.25s}, \quad (1)$$

with

$$s = \frac{\ln(1 + 0.6\chi + 0.1\chi^2)}{0.0396Z\Sigma}, \quad (2)$$

where  $Z$  is the metallicity (in solar units) of the gas element,  $\chi$  is the radiation field strength as a function of metallicity (Krumholz & Gnedin 2011), and  $\Sigma$  is the local column density (in  $M_\odot \text{ pc}^{-2}$ ), which is estimated using the Sobolev approximation and modified to account for subresolution clumping (see full discussion in Davé et al. 2016). Thereafter, the  $\text{H I}$  mass fraction per gas element is determined by subtracting the  $\text{H}_2$  fraction from the total cold neutral gas fraction.

Only gas elements above a hydrogen number density threshold of  $n_{\text{H}} > 0.13 \text{ H atoms cm}^{-3}$  and with  $f_{\text{H}_2} > 0$  are eligible for star formation. Gas above this density is considered the interstellar medium (ISM) and is artificially pressurized by imposing  $T = 10^4 (n_{\text{H}}/0.13)^{1/3} \text{ K}$  in order to resolve the Jeans mass (see Davé et al. 2016). SIMBA's star formation prescription follows an  $\text{H}_2$ -based Schmidt (1959) law and the SFR is given by

$$\text{SFR} = \frac{0.02\rho f_{\text{H}_2}}{t_{\text{dyn}}}, \quad (3)$$

where  $\rho$  is the gas density and  $t_{\text{dyn}} = 1/\sqrt{G\rho}$  is the dynamical time. Star particles are spawned stochastically from eligible gas elements, where a gas element is converted into a star particle of the same mass and metallicity as the parent gas element.

## 2.2 Galaxy finding and sample selection

During the simulation runs, haloes are identified on-the-fly using a 3D friends-of-friends (FoF) halo finder built into GIZMO, with a linking length equal to 0.2 times the mean interparticle separation. Galaxies are identified separately using a 6D FoF galaxy finder, with a spatial linking length of 0.0056 times the mean interparticle separation (corresponding to twice the minimum softening length) and a velocity linking length equal to the local velocity dispersion. This is applied to all stars, black holes, and dense gas elements with  $n_{\text{H}} > 0.13 \text{ H atoms cm}^{-3}$ . The minimum stellar mass of a resolved galaxy is taken to be 32 star particle masses, above which galaxy properties can be reliably computed. This limit corresponds to  $M_{*,\text{min}} = 7.3 \times 10^7 M_\odot$  for SIMBA-25 and  $M_{*,\text{min}} = 5.8 \times 10^8 M_\odot$  for SIMBA-100.

$\text{H I}$  is assigned to galaxies in a separate step, instead of only using the gas elements assigned via the 6D FoF galaxy finder. This decision is motivated by the presence of  $\text{H I}$  in more diffuse, self-shielded gas which typically extends beyond the stellar radius of the galaxy (see discussions in Davé et al. 2019, 2020). The  $\text{H I}$  fraction is computed as described in Section 2.1, except all gas elements within a galaxy's halo with  $n_{\text{H}} > 0.001 \text{ H atoms cm}^{-3}$  are considered. The  $\text{H I}$  is then assigned via proximity to the galaxy to which these gas elements are most gravitationally bound. This allows for the  $\text{H I}$  residing in gas outside of the star-forming region of the galaxy to be included, while ensuring that no  $\text{H I}$  is double-counted.

Haloes and galaxies are cross-matched in post-processing using CAESAR,<sup>1</sup> a PYTHON-based package that builds on the YT simulation analysis toolkit (Turk et al. 2011). CAESAR outputs this information, in addition to pre-computed physical properties and particle lists of haloes and galaxies, in the form of a standalone Hierarchical Data Format (HDF5) catalogue for each simulation snapshot.

For each volume, we use the snapshot at  $z = 0.0167$  and its associated CAESAR catalogue to select our sample. This snapshot is chosen because it is closest in redshift to A262 ( $z = 0.0163$ , Choque-Challapa et al. 2021), which is one of the galaxy clusters observed in the MDS that we aim to compare our mock sample with in a future study. We impose a lower stellar mass limit of  $M_* \geq 7.3 \times 10^7 M_\odot$  to SIMBA-25 and  $M_* \geq 5.8 \times 10^8 M_\odot$  to SIMBA-100, which corresponds to the minimum stellar mass of a resolved galaxy in each volume. We also impose a lower  $\text{H I}$  mass limit of  $M_{\text{H I}} \geq 2.28 \times 10^8 M_\odot$  to SIMBA-25 and  $M_{\text{H I}} \geq 1.82 \times 10^9 M_\odot$  to SIMBA-100, to ensure that the  $\text{H I}$  content of a galaxy is composed of at least 100 gas elements. We apply this  $\text{H I}$  mass limit to the  $\text{H I}$  mass that is confined to the

<sup>1</sup><https://caesar.readthedocs.io/en/latest/>

dense ISM gas of a galaxy as opposed to the galaxy’s total HI mass provided by the CAESAR catalogue. This decision is motivated by the method used to generate synthetic HI data cubes in this work. MARTINI (described further in Section 2.3) is used to extract gas elements in a spherical aperture that is centred on the minimum of a galaxy’s gravitational potential when loading a galaxy’s HI content into a data cube. However, the method SIMBA uses to assign HI to galaxies means that it is possible for a massive galaxy to have HI in its surrounding environment that has been assigned to it, but no HI located in the central, star-forming region of the galaxy. This will result in an empty data cube if the galaxy’s HI content lies entirely outside MARTINI’s extraction aperture. One solution would be to increase the aperture used, but this slows down the cube creation considerably, increases the likelihood of including HI emission from neighbouring galaxies in the foreground/background, and does not guarantee that all of the galaxy’s HI content will be contained in the data cube. Instead, by imposing the HI cut to the HI mass in the central region of the galaxy, we ensure that a significant fraction of a galaxy’s HI is still associated with the galaxy itself. We also performed consistency checks to ensure that the extracted cubes were large enough to encompass the full HI distributions of the galaxies (see Section 2.4).

The specified stellar and HI mass cuts are applied to each volume, resulting in an initial sample of 669 galaxies from SIMBA-25 and 5914 galaxies from SIMBA-100. Due to the larger effective volume of SIMBA-100 and limited computing resources, we reduce the sample size of this volume to match that of SIMBA-25 by randomly selecting 669 galaxies from the 5914 galaxies such that the HI mass distribution of the downsized sample reproduces the distribution of SIMBA-100’s initial sample. None the less, our starting sample of 1338 galaxies (and the final 1157 galaxies after the sample refinement described in Section 2.4) is  $> 5\times$  larger than previous studies investigating HI morphological asymmetries (e.g. Reynolds et al. 2020; Bilimogga et al. 2022) and sufficient for the objectives of this study. Our procedure for generating mock HI data cubes for the starting sample is described in the following section.

### 2.3 Synthetic HI data cubes

MARTINI<sup>2</sup> (Oman 2019, 2024) is a PYTHON package developed for creating synthetic, spatially, and kinematically resolved HI data cubes from smooth particle hydrodynamical simulations. It includes instrument-specific submodules for the data cube, beam, noise, source, and spectral model, which can be configured to reproduce mock interferometric HI observations of galaxies. The methodology used to create HI data cubes with MARTINI is described in Oman et al. (2019). Here, we provide a summary that is specific to our sample of galaxies.

In MARTINI, we specify a fixed aperture with a radius of 100 kpc to extract the galaxy of interest and the region surrounding it from the simulation snapshot. The specific angular momentum vector of the target galaxy’s HI disc,  $L_{\text{HI}}$ , is calculated using the central 1/3 of the HI disc (by mass). The disc plane is then oriented in the  $y$ - $z$  plane by aligning the  $x$ -axis with the direction of  $L_{\text{HI}}$ . Each galaxy

is given a randomly selected azimuthal rotation (around  $L_{\text{HI}}$ ) and inclination angle with respect to the line of sight. For all galaxies in our sample, the kinematic position angle is set to  $270^\circ$ , measured counterclockwise from North.

Galaxies are placed in the *Hubble* flow at a distance of 71.86 Mpc, the adopted distance to A262 (Choque-Challapa et al. 2021), resulting in an angular scale of  $\sim 348$  pc arcsec<sup>-1</sup>. Each data cube, which is centred on the minimum of the target galaxy’s gravitational potential, is generated with  $175 \times 175$  pixels and a pixel size of  $5$  arcsec  $\times$   $5$  arcsec. A velocity channel spacing of  $7.86$  km s<sup>-1</sup> is used and 128 velocity channels are included to ensure that the full extent of the HI emission of HI massive galaxies is contained within the data cubes. The data cubes are convolved to a 15-arcsec FWHM circular Gaussian beam, which corresponds to  $\sim 1.74$  kpc pixel<sup>-1</sup> at the distance of A262. These velocity and angular resolutions are chosen to match the observational parameters of the MDS data cubes. The unit of flux in the data cubes is Jy beam<sup>-1</sup>. For the purposes of this study, no instrumental noise is added to the data cubes.

### 2.4 Sample refinement

Because we extract all gas elements within 100 kpc, the HI of a neighbouring galaxy will be included in the target galaxy’s data cube if the galaxies are within 100 kpc and  $(128/2) \times 7.86$  km s<sup>-1</sup> of each other. We use the OBJECTS task from GIPSY<sup>3</sup> (van der Hulst et al. 1992; Vogelaar & Terlouw 2001) to remove the HI emission from companion galaxies to prevent this extraneous HI from contributing to the HI mass and the asymmetry parameters measured for the galaxies in our sample. For each data cube, OBJECTS was set to identify all independent structures above a column density threshold of  $1 \times 10^{19}$  cm<sup>-2</sup> over 25 km s<sup>-1</sup> and present in at least five channels. The largest 3D structure identified by OBJECTS that is closest to the centre of the data cube (spatially and spectrally) was used as a mask to separate the target galaxy from all other detected objects in the data cube. For each galaxy, the masked data cube was then used to measure the HI mass using the formula

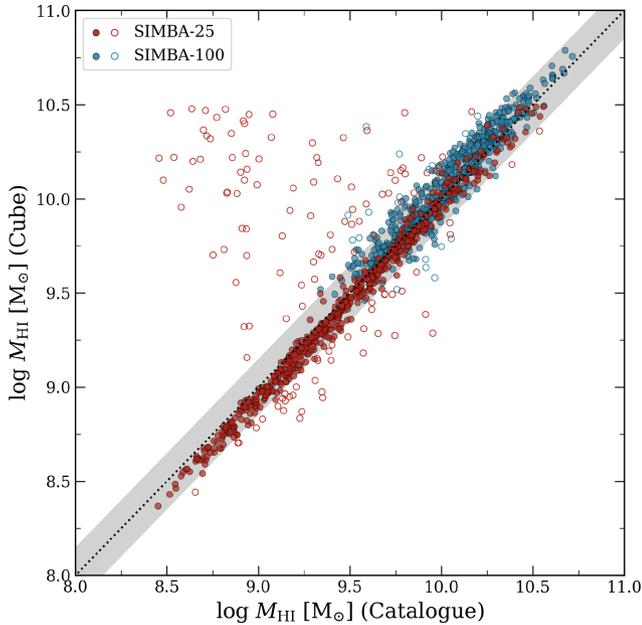
$$\left(\frac{M_{\text{HI}}}{M_{\odot}}\right) \simeq \frac{2.35 \times 10^5}{(1+z)} \left(\frac{D}{\text{Mpc}}\right)^2 \left(\frac{S}{\text{Jy km s}^{-1}}\right), \quad (4)$$

where  $z$  is the redshift of the galaxy (set to  $z = 0.0163$ ),  $D$  is the proper distance (71.86 Mpc), and  $S = \int S_{\nu} d\nu$  is the integrated HI line flux density measured in the data cube (Meyer et al. 2017).

In Fig. 1, we compare the measured HI masses from our masked data cubes to the expected HI masses from the CAESAR catalogues. For both volumes’ samples, the majority of galaxies are scattered within  $\sim 0.15$  dex of the 1:1 relation, illustrating that SIMBA’s approach to assigning HI to galaxies yields reasonable HI masses that are in agreement with the HI masses measured from the mock data cubes. This also indicates that most of the HI content of the galaxies is successfully recovered after masking the data cubes. However, a considerable number of galaxies, mostly from the SIMBA-25 sample, have much higher ( $\geq 0.5$  dex) measured HI masses than expected. Upon inspecting their cubes, we find that most of these outliers are satellite galaxies whose central galaxy has remained in the data cube due to connecting bridges of HI between the galaxies. A higher column density threshold of  $3.5 \times 10^{19}$  cm<sup>-2</sup> was used to test if the target galaxies could be disconnected from their larger companions, but this threshold was either unsuccessful in separating the galaxies or resulted in considerable amounts of low column density HI being

<sup>2</sup><https://github.com/kyleaoman/martini>, version 1.5 (git commit c455f17). Subsequent versions of MARTINI released after the analysis for this work was carried out include some minor changes that would slightly affect our measurements. However, we have checked a subset of galaxies with code versions 2.0.0 and 2.0.6 and find that the changes are minimal and do not impact our conclusions.

<sup>3</sup><https://www.astro.rug.nl/~gipsy>



**Figure 1.** Comparison between the H I mass measured from the masked mock data cube and the expected H I mass from the CAESAR catalogue for the SIMBA-25 and SIMBA-100 galaxies (colour-coded according to the legend). The dotted line shows the 1:1 relation, with the shaded region representing a 0.15 dex buffer from the relation. Galaxies that deviate from the 1:1 relation by more than  $3\sigma_{\text{MAD}}$  are plotted as open circle markers and have been excluded from the samples.

masked out, causing galaxies to scatter downwards from the 1:1 relation. It is also worth noting that the SIMBA-100 sample appears slightly offset upwards from the 1:1 relation by  $\sim 0.1$  dex, whereas the SIMBA-25 sample lies slightly below the 1:1 relation. Although a higher column density threshold would shift the SIMBA-100 galaxies closer to the 1:1 relation, this would result in the SIMBA-25 sample shifting further away from the relation since the same threshold is used to ensure that both samples are treated consistently.

We opted to remove the significantly outlying galaxies from both samples for a number of reasons. First, these galaxies will result in artificially high H I asymmetries that are inflated due to the presence of multiple sources in the data cubes and not the physical processes disturbing the H I discs. Second, when investigating the dynamical histories of galaxies (see Section 4 for more details), we consider the fractional changes in a galaxy’s baryonic mass (calculated using the stellar and H I masses) over several snapshots up to a redshift of  $z \sim 0.2$ . However, we do not generate cubes at these snapshots and rely solely on the masses provided in the CAESAR catalogues. By restricting our samples to galaxies with measured H I masses that are in agreement with the catalogue H I masses, we reduce the likelihood of under- or overestimating the H I masses (and therefore the baryonic masses) at previous snapshots.

To identify outliers in each sample, we use the robust estimator  $\sigma_{\text{MAD}} = 1.4826\text{MAD}$ , where MAD is the median absolute deviation, because it is less sensitive to the presence of outliers than the standard deviation around the mean. Galaxies that deviate from the 1:1 H I mass relation by more than  $3\sigma_{\text{MAD}}$  (shown as open circle markers in Fig. 1) are excluded from the samples. We further remove galaxies that cannot be reliably traced in the snapshots used to determine their dynamical histories (see Section 4). These cuts remove a total of 181 galaxies, resulting in the final 513 galaxies from SIMBA-25 and 644 galaxies from SIMBA-100 which are studied in the remainder of this

work. The H I total intensity (moment-0) maps and global H I profiles are obtained from the masked cubes by integrating along the spectral axis and by summing the flux in each channel, respectively.

Fig. 2 shows the distributions of the H I mass, stellar mass, baryonic mass ( $M_{\text{baryon}} = M_{\star} + 1.33M_{\text{H I}}$ , where the factor of 1.33 accounts for the presence of helium), H I gas fraction ( $M_{\text{H I}}/M_{\star}$ ), H I diameter, and specific SFR ( $\text{sSFR} = \text{SFR}/M_{\star}$ ) for the final galaxy samples. The stellar masses and SFRs are taken from the CAESAR catalogues at  $z = 0.0167$ . All properties relating to the H I mass are obtained using the masked H I data cubes.

## 2.5 Data verification

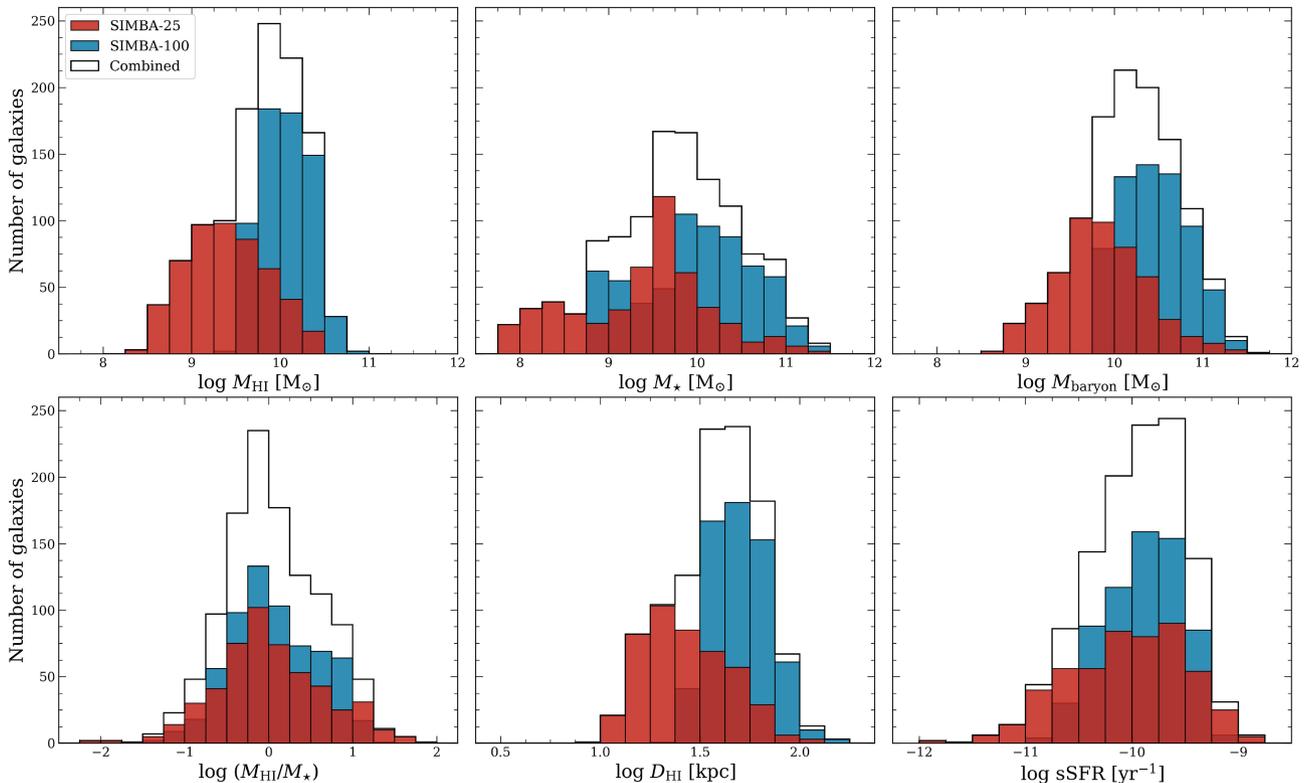
Before proceeding further, we quantify the spatial distributions of H I in our samples using the H I radius,  $R_{\text{H I}}$ , and assess SIMBA’s ability to reproduce the well-studied H I size–mass relation. Observationally, it has been shown that a tight relation exists between the total H I mass of a galaxy and the diameter of its H I disc (e.g. Broeils & Rhee 1997; Verheijen & Sancisi 2001; Begum et al. 2008; Lelli, McGaugh & Schombert 2016; Wang et al. 2016; Naluminsa, Elson & Jarrett 2021; Rajohnson et al. 2022). The relation can be described by

$$\log\left(\frac{D_{\text{H I}}}{\text{kpc}}\right) = m \log\left(\frac{M_{\text{H I}}}{M_{\odot}}\right) + c, \quad (5)$$

where  $D_{\text{H I}}$  is the diameter of the galaxy measured at an H I surface mass density of  $1 M_{\odot} \text{pc}^{-2}$ . The H I size–mass relation holds irrespective of morphological type and galaxy size, suggesting that the mean H I surface density remains approximately constant and that galaxies likely evolve along the relation.

Several studies utilizing hydrodynamical simulations (Bahé et al. 2016; Marinacci et al. 2017; Diemer et al. 2019; Stevens et al. 2019; Gensior et al. 2024) and semi-analytic models (Obreschkow et al. 2009; Wang et al. 2014; Lutz et al. 2018) have found remarkable success in reproducing the observed relation with simulated data at low redshift, albeit with some variations in the scatter, slope, and intercept. Since SIMBA dynamically tracks H I instead of applying a post-processing correction, and does not fine-tune any parameters to reproduce observed H I scaling relations, the robustness of the H I size–mass relation makes it an essential benchmark to evaluate how the subgrid prescriptions and feedback mechanisms implemented in SIMBA affect the H I content of galaxies.

We follow an observational approach by making use of the moment-0 maps to derive azimuthally averaged radial H I surface density profiles. This is done with the ELLINT task from GIPSY, which integrates the moment-0 map in concentric ellipses. The position and inclination angles of the ellipses are set to the values used when generating the data cubes in Section 2.3. Each radial profile is deprojected to the face-on equivalent and then scaled by the total H I mass of the galaxy. To get  $R_{\text{H I}}$ , we interpolate the profiles and measure the radius at which the H I surface mass density ( $\Sigma_{\text{H I}}$ ) is equal to  $1 M_{\odot} \text{pc}^{-2}$  ( $\sim 1.249 \times 10^{20} \text{cm}^{-2}$ ). It should be noted that this procedure is impacted by beam smearing. If the minor axis of a galaxy’s H I disc is not sufficiently resolved, this will lead to the ‘smearing’ out of H I emission and a seemingly rounder H I disc. A galaxy’s inclination angle is used to determine the ellipticity of the ellipses used to derive its radial profile, as well as the correction factor to deproject the profile to represent face-on values. If the outermost ellipse does not enclose the full extent of the rounder H I disc due to beam smearing, this will lead to underestimating  $R_{\text{H I}}$  because the total H I mass of the galaxy is used to scale the profile to surface mass densities. Alternatively, if an edge-on galaxy has a



**Figure 2.** Sample properties for the final sample of SIMBA-25 and SIMBA-100 galaxies, following the same colour coding as in Fig. 1. Top row: distributions of the measured H I mass, stellar mass, and baryonic mass (calculated as  $M_{\text{baryon}} = M_{\star} + 1.33M_{\text{HI}}$ ). Bottom row: distributions of the H I gas fraction, H I diameter, and sSFR ( $\text{sSFR} = \text{SFR}/M_{\star}$ ). In all panels, the step histogram represents the combined sample. The stellar masses and sSFRs are taken from the CAESAR catalogue. All other properties related to the H I content make use of the H I from the masked cubes. The H I diameter distributions only include galaxies that are resolved by  $\geq 2$  beams across the minor axis (see Section 2.5).

warp in its H I disc or extraplanar H I that falls outside the outermost ellipse,  $R_{\text{HI}}$  will also be underestimated for the same reason. Thus, to mitigate the effects of beam smearing, we restrict our analysis in this section to galaxies that are resolved by  $\geq 2$  beams across the minor axis, leaving us with 456 galaxies from SIMBA-25 and 617 galaxies from SIMBA-100.

Fig. 3 shows the median radial profiles, normalized to  $R_{\text{HI}}$ , for the SIMBA-25 (red) and SIMBA-100 (blue) samples. To investigate if any differences between the median profiles of the two samples is a result of their different H I mass ranges, we also show the median radial profile of SIMBA-25 galaxies with  $M_{\text{HI}} \geq 10^{9.5} M_{\odot}$ , where the two samples overlap. For an observational comparison, we include the median radial profile of a subsample of galaxies from Wang et al. (2016) that are resolved by  $\geq 3$  beams across the major axis (see their fig. 2 for the H I surveys used to compile the sample). For all samples, the median values are computed in bins of  $0.1 \times R/R_{\text{HI}}$ . The data for the high-H I mass SIMBA-25 subsample and SIMBA-100 have been slightly offset horizontally for visualization purposes. The error bars and shaded area indicate the interquartile range (i.e. the region between the 25th and 75th percentiles) of each sample.

Qualitatively, the median profiles of SIMBA-25 and SIMBA-100 galaxies broadly agree with the median profile from observed data in terms of their shapes, but there are apparent differences between the simulated and observed samples in the innermost regions of the H I discs. SIMBA-100 galaxies have systematically higher central H I surface densities than observed in the local Universe, with a steeper decline in surface density beyond  $R_{\text{HI}}$ . Galaxies from SIMBA-25 also display slightly higher central H I surface densities in the inner

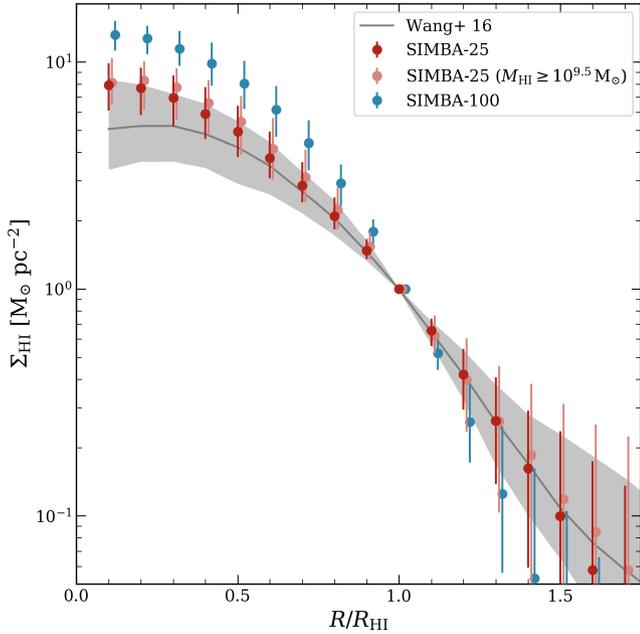
regions ( $R \lesssim 0.5R_{\text{HI}}$ ), but the median profile falls within the overlap of the interquartile ranges of the radial profiles from SIMBA-25 and the observational sample, and largely agrees with observations at larger radii ( $R \gtrsim R_{\text{HI}}$ ). While the median radial profile of the high-H I mass SIMBA-25 subsample has marginally higher H I surface densities than that of the full SIMBA-25 sample, the profiles are generally consistent. However, despite having similar H I mass ranges, the median profiles of the high-H I mass SIMBA-25 subsample and SIMBA-100 are still offset from each other, indicating that the difference in the H I mass ranges of SIMBA-25 and SIMBA-100 is not the primary cause of the offsets between the two simulated samples.

The observed differences between the median profiles of the SIMBA-25 and SIMBA-100 samples are possibly due to the difference in the numerical resolutions of the two volumes. As mentioned, SIMBA incorporates a self-shielding approximation from Rahmati et al. (2013) that depends on the local gas density in order to compute the H I fractions. It is likely that the higher particle numbers for galaxies in SIMBA-25 relative to galaxies of similar H I mass in SIMBA-100 results in more self-shielding in the former volume, and thus more H I in the outskirts of the H I discs.

We now focus on investigating the H I size–mass relation for SIMBA-25 and SIMBA-100. We correct the measured H I diameters for beam smearing effects using

$$D_{\text{HI}} = \sqrt{D_{\text{HI},0}^2 - b_{\text{maj}} \times b_{\text{min}}}, \quad (6)$$

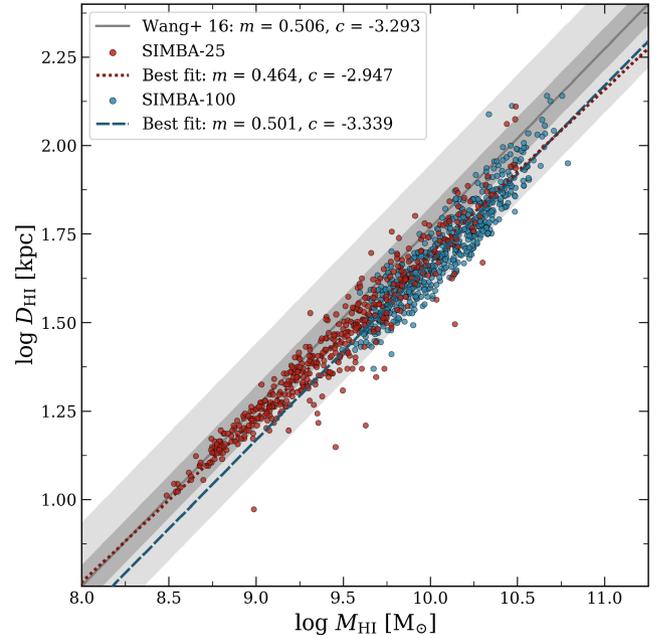
where  $D_{\text{HI},0}$  ( $\equiv 2R_{\text{HI}}$ ) is the uncorrected H I diameter, and  $b_{\text{maj}}$  and  $b_{\text{min}}$  are the major and minor axes of the beam, respectively. The corrected H I diameters are used to construct the H I size–mass



**Figure 3.** Median azimuthally averaged H I surface density profiles for galaxies from SIMBA-25, the high-H I mass ( $M_{\text{HI}} \geq 10^{9.5} M_{\odot}$ ) subsample of SIMBA-25, and SIMBA-100 (colour-coded as indicated in the legend). The median radial profiles for the high-H I mass SIMBA-25 subsample and SIMBA-100 have been slightly offset horizontally for easier viewing. The solid line shows the median radial profile for a combined sample of resolved galaxies from Wang et al. (2016, we refer the reader to their paper for further details on the H I surveys used to compile the sample). The error bars and the shaded region indicate the interquartile range of the three samples.

relation, which is shown in Fig. 4 along with the individual data points of each sample. The red dotted and blue dashed lines correspond to the best linear fits to the SIMBA-25 and SIMBA-100 data, respectively, with the fit parameters quoted in the figure legend. The solid grey line shows the best fit from Wang et al. (2016) for comparison, with the dark and light grey shaded regions representing the  $1\sigma$  and  $3\sigma$  scatter of their fit. Overall, both SIMBA samples reproduce the observed H I size–mass relation reasonably well and the majority of galaxies lie within  $3\sigma$  of the relation. The SIMBA-25 sample produces a slope that is slightly shallower than observed (0.464 compared to 0.506 in Wang et al. 2016, i.e. an  $\sim 8$  per cent difference), resulting in galaxies that have marginally smaller H I discs than expected at the high-H I mass end ( $M_{\text{HI}} \geq 10^{9.5} M_{\odot}$ ). The SIMBA-100 sample matches the observed slope remarkably well to within 1 per cent, but the H I discs of these galaxies are also smaller than expected given their H I masses. It is not surprising that the H I discs in SIMBA-100 are generally more compact than observed, given the median radial profile of the SIMBA-100 sample shown in Fig. 3. Nevertheless, the offset between the observed relation from Wang et al. (2016) and that obtained for SIMBA-100 is relatively small ( $< 0.1$  dex at  $M_{\text{HI}} = 10^{10} M_{\odot}$ ). Despite the differences in the best linear fits obtained for SIMBA-25 and SIMBA-100, we find that when only fitting the SIMBA-25 galaxies with  $M_{\text{HI}} \geq 10^{9.5} M_{\odot}$ , the slope and intercept measured are consistent with those of SIMBA-100 within the uncertainties.

We find that the rms scatter around the relation is 0.06 and 0.05 dex for SIMBA-25 and SIMBA-100, respectively, which is comparable to the observed scatter of  $\sim 0.06$  dex from Wang et al. (2016). In Section 5.1.1, we further discuss the H I size–mass relations of the SIMBA samples in the context of a new study by Wang et al. (2024),



**Figure 4.** The H I size–mass relation for SIMBA-25 and SIMBA-100 galaxies, following the same colour coding as in Fig. 1. The solid line indicates the best fit from Wang et al. (2016), with the  $1\sigma$  and  $3\sigma$  scatter of the observational fit shown by the dark and light shaded regions. The dotted and dashed lines represent the best fitting linear relations for SIMBA-25 and SIMBA-100 galaxies, respectively. For all fits shown, the slope and intercept are given in the legend.

who find a slightly shallower relation when using new single-dish data combined with archival H I observations.

### 3 METHODOLOGY

#### 3.1 Quantitative asymmetry measures

A number of methods exist to quantify irregular gas distributions. Here, we focus on the quantitative parameters used in this work to characterize asymmetries in the global H I profiles and the moment-0 maps of galaxies.

##### 3.1.1 Spectral asymmetries

One of the most widely used methods of quantifying asymmetries in the global H I profiles of galaxies was first proposed by Haynes et al. (1998). In this method, the spectral asymmetry is calculated using the ratio of areas under the profile at velocities less than, and greater than, the systemic velocity ( $v_{\text{sys}}$ ):

$$A_{\text{ID}} = \frac{S_H}{S_L} = \frac{\int_{v_{\text{sys}}}^{v_h} S_v dv}{\int_{v_l}^{v_{\text{sys}}} S_v dv}, \quad (7)$$

where  $S = \int S_v dv$  is the integrated flux density contained within the specified velocity bounds and  $v_l$  ( $v_h$ ) corresponds to the lower (upper) velocity measured at either 20 per cent or 50 per cent of the peak flux density. In the literature, the final spectral asymmetry measure,  $A_{\text{ID}}$ , is typically expressed as a number greater than 1, and so the inverse ratio  $S_L/S_H$  is used instead if  $S_H/S_L < 1$ . With this definition, a value of 1 implies a perfectly symmetric global H I profile and larger values indicate a departure from symmetry.

We opt to use the equivalent definition used in Matthews et al. (1998), which expresses the spectral asymmetry as a value between

0 and 1 by considering the relative flux in the two halves of the profile:

$$A_{\text{flux}} = \frac{|S_H - S_L|}{S_H + S_L} = \frac{\left| \int_{v_{\text{sys}}}^{v_h} S_v dv - \int_{v_l}^{v_{\text{sys}}} S_v dv \right|}{\int_{v_l}^{v_h} S_v dv}. \quad (8)$$

We use  $v_{l,20}$  and  $v_{h,20}$  as the velocity limits to ensure that we capture as much of the global profile as possible in the asymmetry measurement and to be consistent with previous studies (e.g. Espada et al. 2011; Deg et al. 2020; Watts et al. 2020a). The profile is linearly interpolated to determine  $v_{l,20}$  and  $v_{h,20}$  since the limits may fall within a velocity channel. For double-horned profiles, each peak and its velocity bounds are determined separately by searching for local maxima to the left and right of the central velocity of the data cube (as an initial estimate of  $v_{\text{sys}}$ ). If the global profile crosses the 20 percent flux level at multiple velocities to the left (or right) of  $v_{\text{sys}}$ , the innermost crossing is taken as the velocity limit. We then define  $v_{\text{sys}}$  to be the midpoint of the velocity width at the 20 percent flux level, i.e.  $v_{\text{sys}} = (v_{l,20} + v_{h,20})/2$ . The minimum number of gas elements required for the construction of the mock data cubes ensures that particle shot noise does not impact the peak flux measurements in double-horned profiles.

### 3.1.2 Morphological asymmetries

The subject of characterizing spectral asymmetries in H I has been investigated in greater depth than that of 2D morphological asymmetries due to the availability of data. The number of global H I profiles, which can be obtained using single-dish radio telescopes, greatly exceeds the number of spatially resolved H I maps, which require observations with radio interferometers. As such, a number of early studies investigating H I asymmetries in spatially resolved data adopted methods that were first introduced in optical studies. One such example is the 2D rotational asymmetry parameter. First introduced in Schade et al. (1995) and revisited in Conselice, Bershad & Jangren (2000), the parameter is expressed as

$$A = \frac{\sum_{i,j} |I(i,j) - I_{180}(i,j)|}{2 \sum_{i,j} |I(i,j)|}, \quad (9)$$

where  $I(i,j)$  is the pixel flux value in the galaxy image, and  $I_{180}(i,j)$  is the pixel flux value in the galaxy image rotated  $180^\circ$  about a specified rotation point. This statistic has been frequently used to study disturbed optical morphologies (e.g. Conselice 2003; Lotz et al. 2004; Bluck et al. 2012; Abruzzo et al. 2018; Bignone et al. 2020; Thorp et al. 2021) and has more recently been extended to H I studies (e.g. Holwerda et al. 2011a, b; Giese et al. 2016; Reynolds et al. 2020). Possible values of  $A$  range from 0 to 1, where 0 indicates a perfectly symmetric distribution.

Due to the fact that  $A$  is, by definition, a flux-weighted measurement, the parameter is often dominated by asymmetric features in the bright central regions of a galaxy. Fainter features on the outskirts of galaxies, such as tidal tails, tend to have negligible contributions when calculating  $A$  due to their low surface brightness. This is because the current formulation of  $A$  normalizes the difference between the original image and its rotated counterpart by the sum of the flux in the original image. For this reason, Lelli et al. (2014) introduced a new statistic that is more sensitive to the outer regions of the galaxy, given by:

$$A_{\text{mod}} = \frac{1}{N} \sum_{i,j} \frac{|I(i,j) - I_{180}(i,j)|}{|I(i,j) + I_{180}(i,j)|}, \quad (10)$$

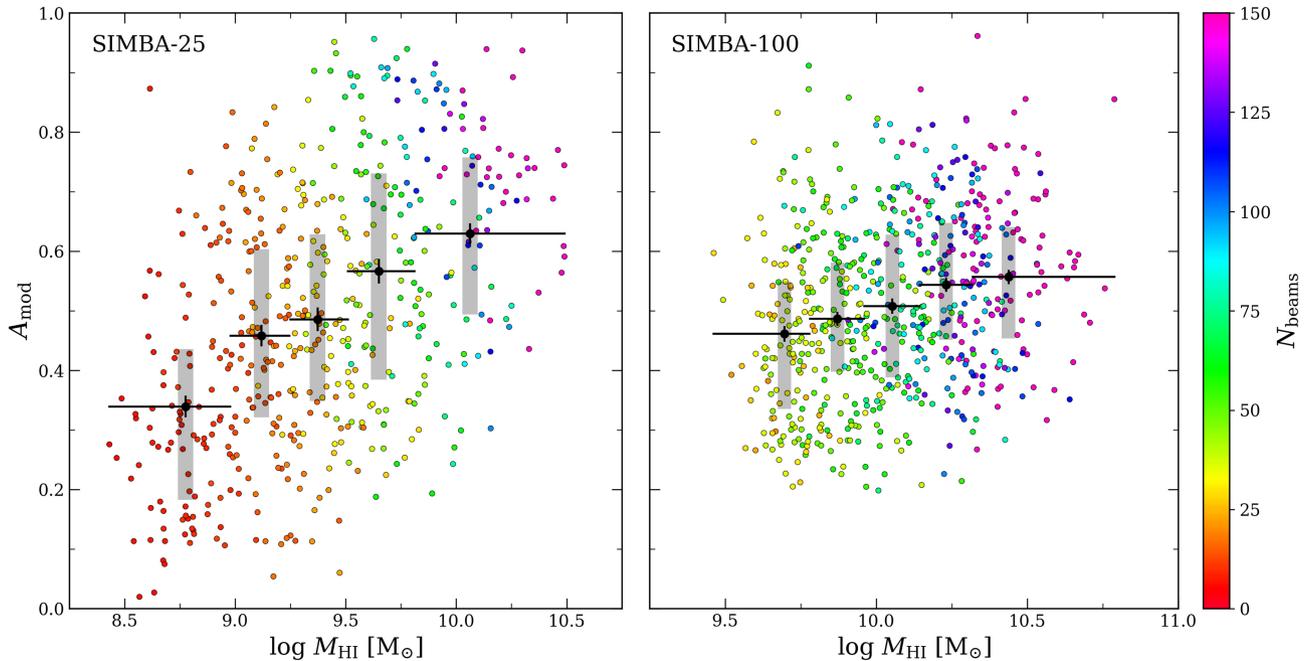
where  $N$  is the total number of non-zero pixels in the additive image. With this formulation, the contribution of low column density asymmetries is upweighted because the residual image is now normalized to the local flux density. The modified asymmetry parameter, hereafter  $A_{\text{mod}}$ , has been shown to reliably quantify the asymmetry in the outskirts of the H I disc where the effects of external mechanisms are likely to be observed (e.g. Bilimogga et al. 2022; Deb, Verheijen & van der Hulst 2023; Roberts et al. 2023). Throughout this work, we use  $A_{\text{mod}}$  to quantify the morphological H I asymmetries and use the centre of the moment-0 map (i.e. the minimum of the galaxy's gravitational potential) as the point around which we rotate the map. It is worth noting that our choice of rotation centre will lead to large  $A_{\text{mod}}$  values if there is a significant offset between the galaxy's H I disc and its potential minimum.

Bilimogga et al. (2022) investigated the impact of various observational effects, such as the spatial resolution, signal-to-noise ratio, and column density threshold implemented when measuring  $A_{\text{mod}}$  from mock H I observations. They found that the majority of  $A_{\text{mod}}$  values measured using a column density threshold of  $5 \times 10^{19} \text{ cm}^{-2}$  are within  $\pm 0.1$  of their intrinsic  $A_{\text{mod}}$  value irrespective of the spatial resolution. Although we are sensitive to lower column densities in our samples, we adopt the proposed threshold of  $5 \times 10^{19} \text{ cm}^{-2}$  because the  $3\sigma$  column density sensitivity of the MDS is  $\sim 6 \times 10^{19} \text{ cm}^{-2}$  over  $25 \text{ km s}^{-1}$  and we will not recover the H I emission below this limit in the observations. The values of all pixels below the column density threshold are set to zero in the moment-0 maps when computing  $A_{\text{mod}}$ . No cuts based on signal-to-noise ratio are necessary since our simulated moment-0 maps are noise-free. In Section 3.2, we examine the effect of resolution on  $A_{\text{mod}}$  in greater detail.

### 3.2 The effect of resolution on spatially resolved asymmetries

Numerous studies have explored how spatial resolution impacts morphological asymmetry parameters, both in simulated (Giese et al. 2016; Thorp et al. 2021; Bilimogga et al. 2022; Deg et al. 2023) and observational (Reynolds et al. 2020; Roberts et al. 2023) data, and have found that degrading the resolution of an observation leads to a decrease in the measured asymmetry due to the 'smoothing' out of asymmetric features. Although high-resolution interferometric observations are desired for studying the detailed H I morphologies of galaxies, it is also common practice to smooth observations to larger beam sizes to improve the signal-to-noise ratio and probe the lower column density outskirts of the H I discs. Additionally, the majority of spatially resolved H I detections from ongoing and future untargeted H I surveys will be marginally resolved ( $\leq 3$  beams across the major axis). It is crucial to strike a balance between determining a minimum resolution threshold that is not so strict that it significantly reduces one's sample size, while ensuring that galaxies are sufficiently well resolved such that the measured asymmetry is still representative of the intrinsic asymmetry. As such, the minimum resolution thresholds adopted in the studies mentioned above are somewhat subjective and vary across the studies.

We approach the subject of resolution effects from a slightly different angle, by examining whether the differences in spatial resolution across a sample can result in a biased relationship between H I mass and  $A_{\text{mod}}$ . Glowacki et al. (2022) observed a weak trend of decreasing spectral asymmetry with increasing H I mass, but this has not been investigated with morphological asymmetry measurements. To quantify how well resolved a galaxy is, we sum up the total number of pixels above  $5 \times 10^{19} \text{ cm}^{-2}$  in the moment-0 map, and divide by the number of pixels in the 15 arcsec circular beam to get the number of beams in the area of the galaxy,  $N_{\text{beams}}$ . We use this approach



**Figure 5.** Scatter plot of morphological asymmetry ( $A_{\text{mod}}$ ) as a function of H I mass for SIMBA-25 (left) and SIMBA-100 (right) galaxies. Here,  $A_{\text{mod}}$  is calculated using the  $15 \text{ arcsec} \times 15 \text{ arcsec}$  resolution moment-0 maps with an applied column density threshold of  $5 \times 10^{19} \text{ cm}^{-2}$ . The colour of individual data points shows the number of beams in the area of the galaxy. The large circle markers indicate the mean  $A_{\text{mod}}$  value in bins of increasing H I mass, with the bin widths shown by the horizontal error bars. The vertical error bars show the standard error on the mean ( $1\sigma/\sqrt{N}$ ), where  $\sigma$  is the standard deviation and  $N$  is the number of galaxies in the bin. The shaded regions show the interquartile (25th–75th percentile) range.

instead of calculating the number of beams across the major axis to account for the fact that a face-on galaxy has more resolution elements in its area than an edge-on galaxy with the same measured major axis.

In Fig. 5, we show  $A_{\text{mod}}$  as a function of H I mass for the SIMBA-25 (left panel) and SIMBA-100 (right panel) galaxies. The data points are colour-coded by  $N_{\text{beams}}$  calculated at 15 arcsec resolution. Each sample is divided into five bins of increasing H I mass, with varying bin widths to ensure that each bin in the respective sample has approximately the same number of galaxies. The mean  $A_{\text{mod}}$  value of each bin is illustrated by the black circle markers, with the vertical error bars and shaded regions representing the standard error on the mean ( $1\sigma/\sqrt{N}$ ) and the interquartile range, respectively. The horizontal error bars represent the bin widths.

Despite considerable scatter within each bin, shown by the spread in the individual data points as well as the interquartile range, it is evident that  $A_{\text{mod}}$ , when measured at a fixed resolution, tends to increase with increasing H I mass. A Spearman rank test of H I mass and  $A_{\text{mod}}$  gives a correlation coefficient of  $r_s = 0.46$  ( $p$ -value  $< 0.001$ ) for SIMBA-25 and  $r_s = 0.25$  ( $p$ -value  $< 0.001$ ) for SIMBA-100, suggesting a weak but statistically significant correlation. This trend is less pronounced in the SIMBA-100 sample, which could be partly due to the compactness of the H I discs in the sample (see Section 2.5) as well as the fact that the SIMBA-100 sample spans a smaller H I mass range than the SIMBA-25 sample.

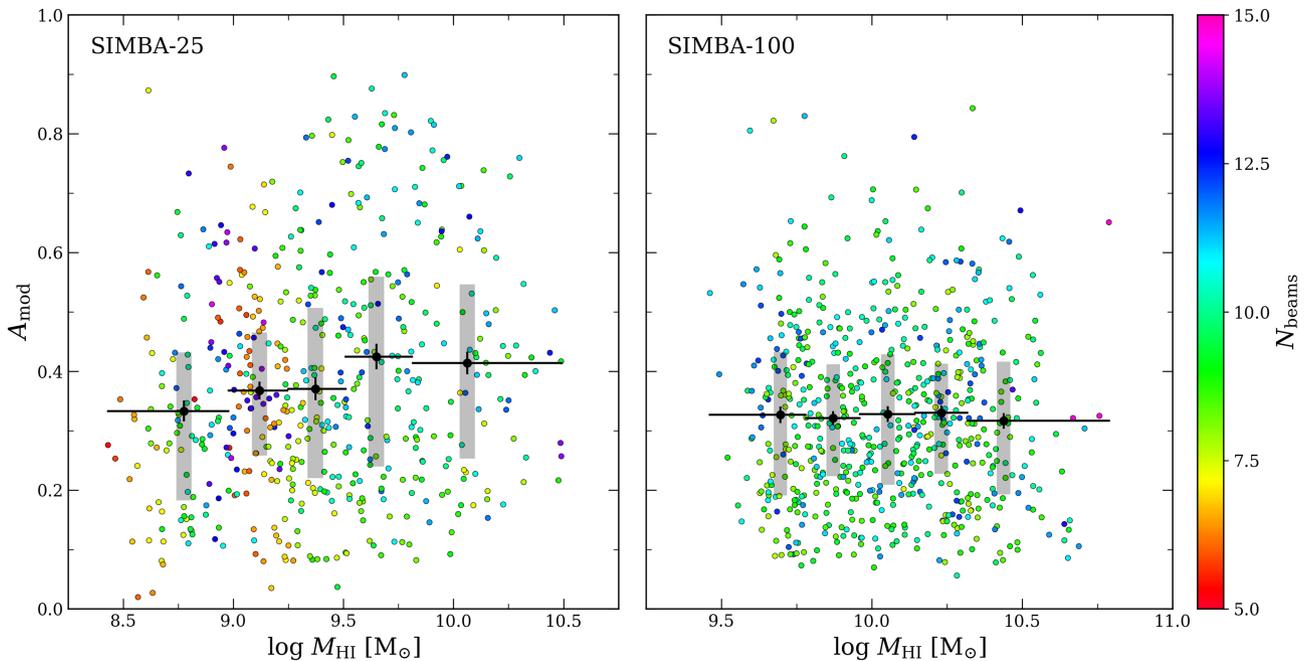
From the colour gradient observed in both panels of Fig. 5, it is apparent that more massive H I discs are more resolved. This is not surprising since we have shown that both SIMBA samples adhere to the H I size–mass relation. However, this raises the question of whether the positive relation between H I mass and  $A_{\text{mod}}$  is a consequence of resolution. The median  $N_{\text{beams}}$  of galaxies in the lowest and highest mass bins of the SIMBA-25 sample are  $\sim 9.5$  and  $\sim 104.4$ , respectively.

In the case of SIMBA-100, the median  $N_{\text{beams}}$  of galaxies in the lowest and highest mass bins are  $\sim 37.4$  and  $\sim 159.1$ , respectively. This means that the highest H I mass galaxies can be 4–11 times more resolved than the lowest H I mass galaxies. Furthermore, we do not find any galaxies in the SIMBA-100 sample with  $A_{\text{mod}} \leq 0.2$ , and those galaxies with low  $A_{\text{mod}}$  values in the SIMBA-25 sample tend to be resolved by less than  $\sim 25$  beams.

To assess the dependence of this observed trend on resolution, we reduce the number of resolution elements across the moment-0 maps by smoothing to larger circular beam sizes of 30, 45, 60, 90, and 120 arcsec. We calculate  $N_{\text{beams}}$  at each resolution in the same manner as before, by summing the pixels in the moment-0 map above  $5 \times 10^{19} \text{ cm}^{-2}$  and dividing by the pixel area of the beam. Throughout this exercise, the pixel size is fixed at  $5 \text{ arcsec} \times 5 \text{ arcsec}$ .<sup>4</sup> Although we become more sensitive to lower column densities when degrading the resolution, we maintain the same column density threshold to ensure a consistent comparison. We identify the resolution at which a galaxy is resolved by approximately 10 beams and recalculate  $A_{\text{mod}}$  at this resolution. This is done instead of enforcing that a galaxy be resolved by at least 10 beams, as the latter results in a broader distribution of  $N_{\text{beams}}$  for both samples. With this approach, all galaxies are resolved by at least five beams, but no more than 15 beams. In Appendix A, we show the moment-0 maps of a few galaxies, at the initial 15 arcsec resolution and smoothed to the resolution at which  $N_{\text{beams}} \approx 10$ , along with their global H I profiles.

Fig. 6 shows  $A_{\text{mod}}$  as a function of H I mass for both samples when  $N_{\text{beams}}$  has been standardized. The colour gradient observed in

<sup>4</sup>We tested whether downsampling the pixels (by increasing their size) in the smoothed data affects the  $A_{\text{mod}}$  values and found no systematic difference, despite slightly increased scatter at low  $A_{\text{mod}}$  values.



**Figure 6.** Same as in Fig. 5, except  $A_{\text{mod}}$  is calculated using the resolution at which a galaxy is resolved by approximately 10 beams in its area. The colour of individual data points shows the number of beams in the area of the galaxy at the resolution used when calculating  $A_{\text{mod}}$ .

Fig. 5 is largely absent in Fig. 6 as galaxies now have similar  $N_{\text{beams}}$  by construction. There is a small vertical band of galaxies with  $N_{\text{beams}} \geq 12$  around  $M_{\text{HI}} \approx 10^9 M_{\odot}$  in the SIMBA-25 sample, but these galaxies have a range of  $A_{\text{mod}}$  values and are not preferentially more asymmetric despite being marginally more resolved. Degrading the resolution of the moment-0 maps results in a clear shift towards lower  $A_{\text{mod}}$  values at higher H I masses in both samples. The low  $A_{\text{mod}}$  ‘floor’ of the SIMBA-100 sample that is present at  $A_{\text{mod}} \approx 0.2$  in the right panel of Fig. 5 is not observed in Fig. 6 and a larger fraction of galaxies from both samples populate the low  $A_{\text{mod}}$  region of the figure. Reassuringly, we still observe highly asymmetric galaxies ( $A_{\text{mod}} \geq 0.6$ ) in both samples, which gives us confidence that the galaxies are sufficiently resolved and that asymmetric features have not been completely washed out by degrading the resolution.

There is no discernible trend between  $A_{\text{mod}}$  and H I mass in the SIMBA-100 sample when galaxies are resolved by approximately the same number of beams. We find that the trend persists in the SIMBA-25 sample according to the Spearman rank test ( $r_s = 0.15$ ,  $p$ -value  $< 0.001$ ), but it is considerably weaker than previously observed, indicating that the previous trend between  $A_{\text{mod}}$  and H I mass was exaggerated due to the effect of resolution. If we only consider the H I mass range  $10^{9.5} M_{\odot} \leq M_{\text{HI}} \leq 10^{10.5} M_{\odot}$ , where the two samples overlap, we find no correlation between  $A_{\text{mod}}$  and H I mass in the SIMBA-25 sample ( $r_s = 0.03$ ,  $p$ -value = 0.67). This suggests that the conflicting results between the two samples is due to the limited H I mass range of the SIMBA-100 sample.

Overall, these findings demonstrate that resolution effects can still persist if only a minimum resolution threshold is applied, and that failing to account for these effects can produce misleading results. If a sample spans a large H I mass range, one must either bin galaxies into different groups based on  $N_{\text{beams}}$  or H I mass to compare  $A_{\text{mod}}$  values in a fair and consistent manner, or degrade the resolution of galaxies as has been done in this work. We opt for this approach because it allows us to utilize the full sample when investigating correlations between  $A_{\text{mod}}$  and other galaxy properties. In the remainder of this

work, we use the  $A_{\text{mod}}$  values measured at the resolution at which  $N_{\text{beams}} \approx 10$ .

#### 4 PHYSICAL DRIVERS OF H I ASYMMETRY

As previously mentioned, disturbances in the H I reservoirs of galaxies can be caused by various physical processes that lead to the depletion or replenishment of a galaxy’s H I content. A number of studies utilizing cosmological simulations have focused on the dependence of environment on spectral asymmetries (e.g. Watts et al. 2020b; Manuwal et al. 2022) and have found that while satellite galaxies, as a population, are typically more asymmetric than central galaxies, environmental processes are not the sole driver of the observed asymmetries in the global H I profiles.

We take advantage of the ability to trace a galaxy’s recent dynamical history in the simulations to investigate galaxy–galaxy interactions and mergers as the possible origins of H I asymmetries in our samples. We consider a total of 11 snapshots between the redshift range  $0.0167 \leq z \leq 0.192$ , corresponding to a period of  $\sim 2.2$  Gyr. Previous estimates of the observability time-scales of tidal features and disturbed morphologies range from  $\sim 0.2$ –2 Gyr, where the time-scale depends on the mass ratio, gas fractions, and relative orientations of the interacting galaxies (e.g. Lotz et al. 2010a, b; Holwerda et al. 2011c; Whitney et al. 2021). Thus, we do not expect that any signatures of an interaction or merger that took place beyond  $z \sim 0.2$  will be sustained until  $z = 0.0167$ , because a galaxy will likely have had enough time to regularize its H I disc over multiple revolutions.

We use CAESAR to track the galaxies in our two samples across the snapshots of interest. For each target galaxy in the  $z = 0.0167$  snapshot, its most massive progenitor in the previous snapshots is identified as the galaxy that contains the most shared star particle IDs. For each snapshot, we then obtain the stellar mass, H I mass and baryonic mass ( $M_{\text{baryon}} = M_{\star} + 1.33M_{\text{HI}}$ ) of the target galaxy

to determine its mass growth history, as well as the 3D separations ( $r_{\text{sep}}$ ) between the target galaxy and its neighbouring galaxies.

To identify galaxies that have undergone a merger within the specified time period, we compute the fractional variation in a galaxy's stellar and baryonic mass using

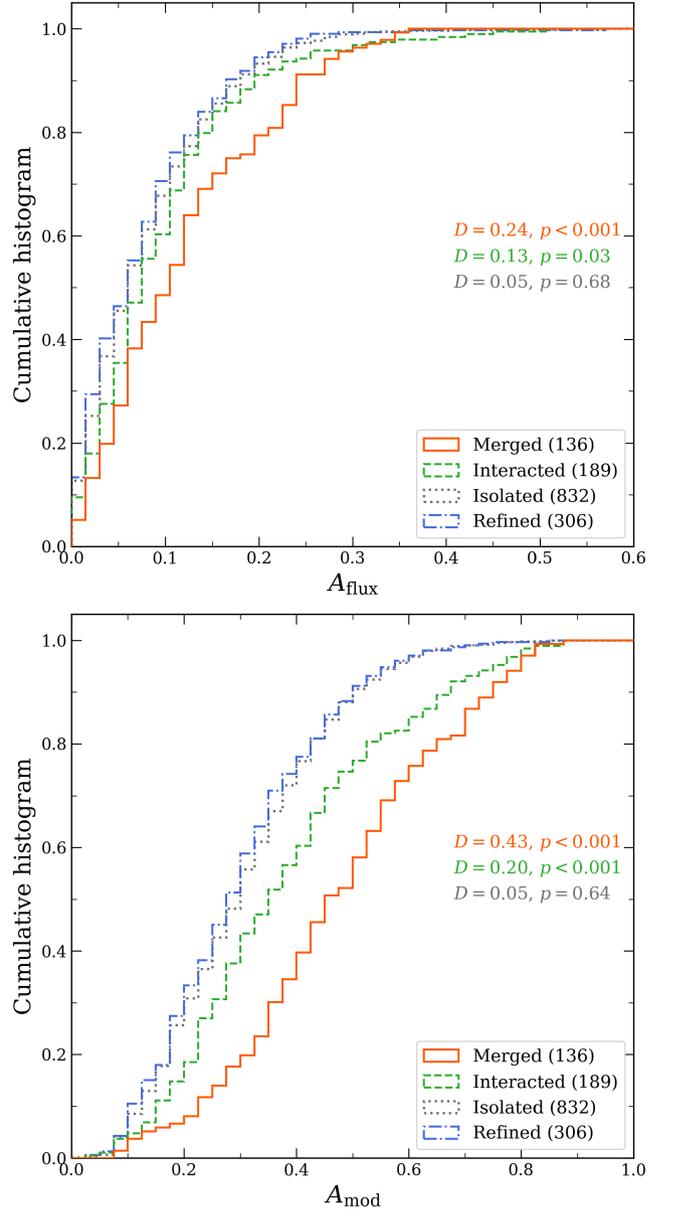
$$\mathcal{R} = \frac{M_t - M_{t-1}}{M_{t-1}}, \quad (11)$$

where  $M_t$  is the galaxy's mass at a given snapshot and  $M_{t-1}$  is its mass at the previous snapshot. We define a merger as a fractional increase in the stellar and baryonic mass ( $\mathcal{R}_*$  and  $\mathcal{R}_{\text{baryon}}$ , respectively) of  $\geq 5$  per cent. Our reason for considering  $\mathcal{R}_*$  in addition to  $\mathcal{R}_{\text{baryon}}$  is to reduce misidentifications at the low-mass end, where HI-rich dwarf galaxies may experience large jumps in their baryonic mass due to gas inflows and be incorrectly identified as mergers. Following a similar approach as Rodríguez Montero et al. (2019), we estimate the fractional stellar mass increase that can be attributed to star formation ( $\mathcal{R}_{\text{SF}}$ ) using the SFRs provided in the CAESAR catalogues and the time difference between the two snapshots where the mass increase occurs, and require that  $\mathcal{R}_* - \mathcal{R}_{\text{SF}} \geq 5$  per cent to ensure that the observed stellar mass increase cannot be explained solely by enhanced star formation. Lastly, we require that prior to the target galaxy's mass increase, the closest neighbouring galaxy must have  $r_{\text{sep}} \leq 50$  kpc, and that the baryonic mass ratio between the neighbouring and the target galaxy ( $M_{\text{baryon}}$  ratio) be  $\geq 5$  per cent.

For recent close encounters, we also consider interactions between galaxies with  $M_{\text{baryon}}$  ratio  $\geq 5$  per cent, but adopt a minimum  $r_{\text{sep}}$  of 100 kpc. Although interaction-driven effects such as enhanced star formation have been observed in close galaxy pairs with projected separations of  $\sim 150$  kpc, a stellar mass ratio  $\geq 10$  per cent is typically used in the existing literature (e.g. Patton et al. 2013; Moreno et al. 2021). We opt for 100 kpc as a compromise for the lower  $M_{\text{baryon}}$  ratio adopted in this work, which is chosen to be consistent with how mergers are identified.

With these criteria, we identify 136 galaxies (78 and 58 from SIMBA-25 and SIMBA-100, respectively) that have undergone a merger, 189 galaxies (106 and 83) that have had a close interaction, and 832 galaxies (329 and 503) that have remained relatively isolated over the preceding  $\sim 2.2$  Gyr. For simplicity, we hereafter refer to these categories as merged, interacted, and isolated, respectively. Based on our criteria, it is possible that galaxies in the isolated category may have experienced interactions with neighbouring galaxies that meet the minimum  $r_{\text{sep}}$  condition but not the  $M_{\text{baryon}}$  ratio condition, or vice versa. To examine if these selection criteria impact our results, we define a refined isolated category consisting of galaxies that have had no neighbouring galaxy (regardless of mass) within 100 kpc and no significant neighbouring galaxy (with  $M_{\text{baryon}}$  ratio  $\geq 5$  per cent) within 500 kpc over the considered time period. The unmasked moment-0 maps of the refined isolated galaxies were also visually inspected to ensure the absence of extraneous HI emission. Of the 832 isolated galaxies, 306 galaxies (94 and 212) belong to the refined isolated category.

For the remainder of this work, we have combined the samples from SIMBA-25 and SIMBA-100 to increase the number of galaxies per category for better statistics, and to enable us to investigate trends between HI asymmetry and baryonic mass. In the following three sections, we study the distributions of the asymmetry parameters for the three categories, how HI spectral and morphological asymmetries compare, and whether HI asymmetry is correlated with baryonic mass.



**Figure 7.** Comparison between the cumulative histograms of the spectral asymmetry ( $A_{\text{flux}}$ , top) and morphological asymmetry ( $A_{\text{mod}}$ , bottom) for galaxies that have undergone a merger (mass ratios  $> 1:20$ ) within the last  $\sim 2.2$  Gyr (solid lines), galaxies that have experienced an interaction (dashed lines) and galaxies that have remained relatively isolated (dotted lines) within the same time period. In both panels, the refined isolated category's cumulative histogram is shown by the dash-dotted histogram. The results of the two-sample KS tests comparing the isolated, interacted and merged categories to the refined isolated category are given in both panels.

#### 4.1 The relationship between HI symmetry and recent dynamical history

In Fig. 7, we show the cumulative distributions of  $A_{\text{flux}}$  (top panel) and  $A_{\text{mod}}$  (bottom) for galaxies in the isolated, refined isolated, interacted, and merged categories (grey dotted, blue dash-dotted, green dashed, and orange solid histograms, respectively). For ease of comparison, we provide the median and interquartile range of  $A_{\text{flux}}$  and  $A_{\text{mod}}$  for each category in Table 1. Before comparing the

**Table 1.** The median ( $P_{50}$ ) and interquartile range ( $P_{75} - P_{25}$ ) of  $A_{\text{mod}}$  and  $A_{\text{flux}}$  measured for the different categories.

	Aflux		Amod	
	$P_{50}$	$P_{75} - P_{25}$	$P_{50}$	$P_{75} - P_{25}$
Refined	0.065	0.092	0.289	0.212
Isolated	0.067	0.097	0.307	0.217
Interacted	0.082	0.088	0.364	0.266
Merged	0.106	0.126	0.471	0.265

cumulative distributions of different categories, it is important to note the differences in the range of values of  $A_{\text{flux}}$  and  $A_{\text{mod}}$ . Although both asymmetry parameters can theoretically have a maximum value of 1, we do not measure an  $A_{\text{flux}}$  value  $> 0.6$  in our sample, while  $A_{\text{mod}}$  covers a slightly larger range of values from 0 – 0.9. It is also apparent that the distributions of  $A_{\text{flux}}$  and  $A_{\text{mod}}$  are different. The cumulative distributions of  $A_{\text{flux}}$  for all categories are positively skewed, indicating that a relatively large fraction of galaxies have low spectral asymmetry and a small fraction of galaxies have high spectral asymmetry, whereas the  $A_{\text{mod}}$  distributions are generally broader and do not exhibit the same behaviour. The interquartile range of  $A_{\text{flux}}$  for all categories is also smaller than that of  $A_{\text{mod}}$  by a factor of  $\sim 2 - 3$ , confirming the larger spread in  $A_{\text{mod}}$  compared to  $A_{\text{flux}}$ .

Starting with the  $A_{\text{flux}}$  distributions of the different categories, we find that merged galaxies have the highest mean asymmetry, followed by interacted galaxies, with refined isolated galaxies having the lowest mean asymmetry and smallest standard deviation (Table 1). This trend is emphasized in the figure, where the cumulative histogram of the merged category has a shallower slope and a lower cumulative fraction than that of the refined isolated category at a given  $A_{\text{flux}}$  value, confirming that the merged category is, on average, more asymmetric. The differences between the cumulative histograms of the refined isolated and interacted categories are subtle and only become apparent at  $A_{\text{flux}} > 0.2$ , where the interacted category has an extended tail towards high  $A_{\text{flux}}$  values. We observe a changeover at  $A_{\text{flux}} \approx 0.32$ , where the interacted category has a lower cumulative fraction than the merged category at high  $A_{\text{flux}}$  values. However, this changeover occurs at a cumulative fraction of  $\sim 0.95$ , indicating that the interacted category is only more asymmetric than the merged category for a small fraction ( $< 5$  per cent) of the most asymmetric global HI profiles. We find that the isolated category is only slightly more asymmetric than the refined isolated category, but has a comparable standard deviation and slightly smaller standard error on the mean due to the larger sample size.

As mentioned previously, it has become common practice to characterize the prevalence of spectral asymmetries in a sample by determining the fraction of galaxies with measured asymmetry values above a certain threshold. Recent studies utilizing the  $A_{1D}$  index from Haynes et al. (1998) adopted threshold values of 1.26 and 1.39, corresponding to deviations greater than  $2\sigma$  and  $3\sigma$  from the mean of the asymmetry distribution of isolated AMIGA galaxies as determined by Espada et al. (2011). It is possible to get the equivalent thresholds expressed as a number between 0 and 1 using the conversion  $A_{\text{flux}} = (A_{1D} - 1)/(A_{1D} + 1)$ , which gives us limits of  $A_{\text{flux}} > 0.12$  and  $> 0.16$ . We use these values to calculate the relative fractions of asymmetric global HI profiles in the different categories and estimate the uncertainties in the fractions using Poisson error propagation. The results are quoted in Table 2. The asymmetric fractions of galaxies with  $A_{\text{flux}} > 0.12$  ( $> 0.16$ ) in the refined isolated, isolated, interacted, and merged categories

**Table 2.** Comparison of the relative fractions of galaxies with  $A_{\text{flux}}$  values above a certain threshold in each category.

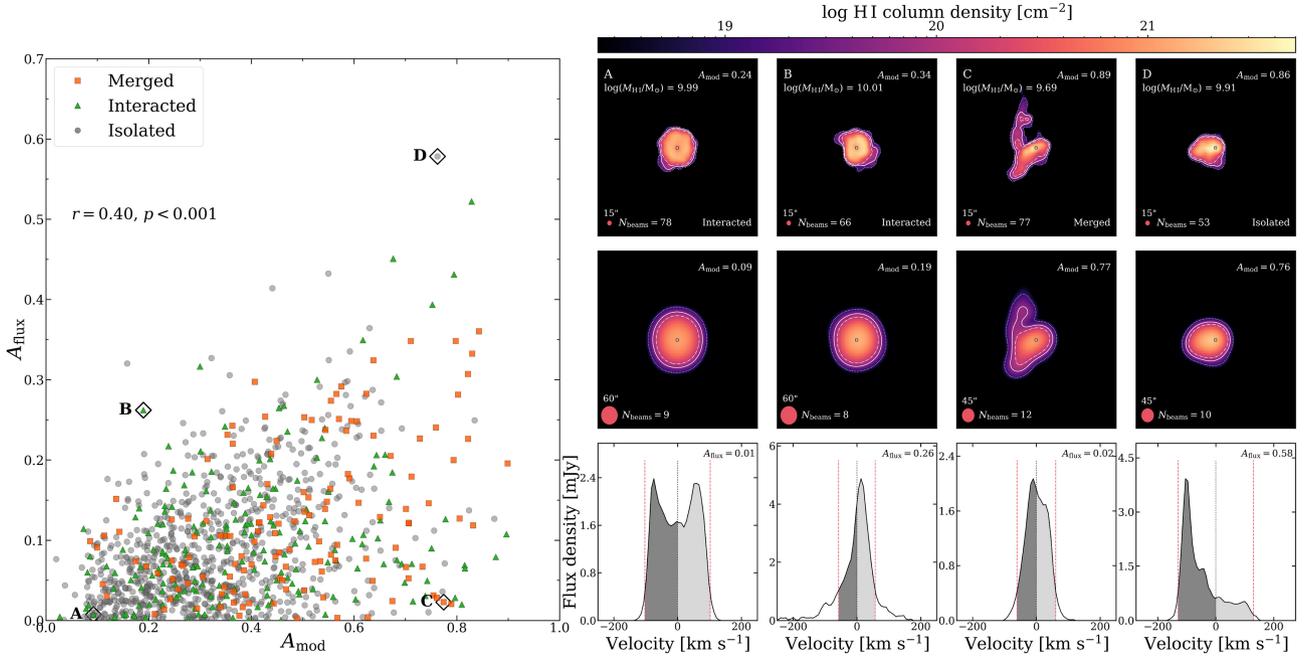
	$N$	$A_{\text{flux}} > 0.12$	$A_{\text{flux}} > 0.16$
		(per cent)	(per cent)
Refined	306	$26 \pm 3$	$14 \pm 2$
Isolated	832	$28 \pm 2$	$15 \pm 1$
Interacted	189	$33 \pm 4$	$17 \pm 3$
Merged	136	$46 \pm 7$	$29 \pm 5$

are 26 per cent (14 per cent), 28 per cent (15 per cent), 33 per cent (17 per cent), and 46 per cent (29 per cent), respectively. Isolated galaxies have comparable, but slightly higher, asymmetric fractions to the refined isolated galaxies. In contrast, the merged category has significantly higher fractions than both isolated categories, and the interacted category has asymmetric fractions that are consistent with both isolated categories, within the uncertainties. We compare these results with the existing literature in Section 5.1.2.

The differences between the categories become even more pronounced when considering the cumulative distributions of  $A_{\text{mod}}$ . At low  $A_{\text{mod}}$  values, the interacted category's cumulative histogram is narrowly offset from both isolated categories and has a similar slope, before diverging at  $A_{\text{mod}} \approx 0.3$ . In comparison, the merged category's cumulative histogram is more clearly separated and systematically offset towards higher asymmetry than both the isolated and interacted categories. Again, the cumulative histograms of the isolated and refined isolated categories are marginally offset from each other.

To quantify the differences between the categories' distributions of  $A_{\text{flux}}$  and  $A_{\text{mod}}$ , we perform two-sample Kolmogorov–Smirnov (KS) tests comparing the isolated, interacted, and merged categories to the refined isolated category, with the results quoted in the panels of Fig. 7. Despite the small offsets observed between the cumulative distributions of the isolated and refined isolated categories, given the  $p$ -values we cannot reject the null hypothesis that they have the same distribution. We also perform KS tests between the refined isolated category and the residual subset of the isolated category, and still obtain  $p$ -values  $> 0.1$ . These results suggest that including galaxies that have had close encounters with companion galaxies below our specified baryonic mass threshold, or encounters with more massive galaxies beyond  $r_{\text{sep}} \geq 100$  kpc, has a negligible impact on the  $A_{\text{flux}}$  and  $A_{\text{mod}}$  distributions of galaxies in the isolated category. To avoid redundancy, the remaining results do not include the refined isolated category since the category's asymmetry distributions are statistically indistinguishable from that of the isolated category, and the latter has a larger sample size.

We find that the distributions of  $A_{\text{mod}}$  and  $A_{\text{flux}}$  of merged galaxies are statistically different from those of the isolated galaxies. For both the interacted and merged categories, the KS test statistic  $D$  returned for  $A_{\text{mod}}$  is greater than that for  $A_{\text{flux}}$ , confirming that there is a larger distinction between the  $A_{\text{mod}}$  distributions of different categories than there is for the  $A_{\text{flux}}$  distributions. Furthermore, while the  $A_{\text{mod}}$  distribution of interacted galaxies is statistically different to that of the isolated category, this is not the case for  $A_{\text{flux}}$ . Therefore, it seems that the HI morphologies of the interacted galaxies are typically more asymmetric than the HI morphologies of the isolated galaxies, but their global HI profiles do not show the same clear differences. Similarly, we also find a statistically significant difference between the  $A_{\text{mod}}$  distributions of the interacted and merged categories ( $p$ -value  $< 0.001$ ), but no strong difference between their  $A_{\text{flux}}$  distributions ( $p$ -value = 0.03).



**Figure 8.** Left: spectral asymmetry ( $A_{\text{flux}}$ ) plotted against morphological asymmetry ( $A_{\text{mod}}$ ) for the isolated (circles), interacting (triangles), and merged (squares) galaxies in our sample. The Spearman correlation coefficient and associated  $p$ -value using the full sample are quoted in the panel. Right: the moment-0 maps (top and middle rows) and global H I profiles (bottom) of four galaxies located in different regions of the left panel. The moment-0 maps are shown at 15 arcsec resolution (top row) and the resolution at which  $N_{\text{beams}} \approx 10$  (middle). Each panel showing the moment-0 maps is 200 kpc across and centred on the potential minimum of the galaxy, which is shown by a cross marker. The dash-dotted, solid, and dotted contours are at column densities of 1, 5, and  $12.5 \times 10^{19} \text{ cm}^{-2}$ . In the bottom row, the vertical dashed lines represent  $v_{l,20}$  and  $v_{h,20}$ , and the dotted line represents  $v_{\text{sys}}$ . Galaxies A (ID: 40) and C (ID: 1997) are from SIMBA-25 and galaxies B (ID: 3435) and D (ID: 34668) are from SIMBA-100. The logarithm of the H I mass and the category of each galaxy are given in the top panels. The measured  $A_{\text{mod}}$  and  $A_{\text{flux}}$  values are quoted in the top-right corner of the relevant panel.

## 4.2 Comparison between asymmetry measures

We turn to a direct comparison of  $A_{\text{flux}}$  and  $A_{\text{mod}}$  on a galaxy-by-galaxy basis to investigate the differences in the distributions of the two parameters in more detail and explore the correlation between the parameters. Because the majority of H I detections from upcoming H I surveys will be spatially unresolved but spectrally resolved, understanding whether a correlation exists between spectral and morphological asymmetries will enable us to determine if a galaxy’s global H I profile can be used to infer the state of its underlying H I disc. Recent studies by Reynolds et al. (2020) and Bilimogga et al. (2022) have examined the correlations between various spectral and morphological H I asymmetries and produced conflicting results. The tension between these studies’ findings could be due to numerous factors, such as the parameters used to quantify the asymmetries, the differences in the mass ranges of the samples and the relatively small ( $N < 200$ ) sample sizes, as well as the treatment of observational effects.

The left panel of Fig. 8 shows  $A_{\text{flux}}$  against  $A_{\text{mod}}$  for isolated, interacting, and merged galaxies (grey circle, green triangle, and orange square markers, respectively). For comparison, the right half of Fig. 8 shows the moment-0 maps and global H I profiles of four galaxies located in different regions of the scatter plot (see the figure caption for more details). While the merged and interacting categories were previously found to have higher mean asymmetries than the isolated category, this does not translate to any category occupying a specific region of the figure as there is considerable overlap between the three categories at various values of  $A_{\text{flux}}$  and

$A_{\text{mod}}$ . Remarkably, there are still isolated galaxies with high  $A_{\text{flux}}$  and/or high  $A_{\text{mod}}$  values. We discuss possible explanations for this in Section 5.1.2.

It is not surprising that the majority of galaxies from all three categories fall below  $A_{\text{flux}} = 0.16$ , given the results in Table 2. However, it is apparent that these galaxies with low  $A_{\text{flux}}$  values can still have a large range of  $A_{\text{mod}}$  values. On the other hand, the majority of galaxies with high  $A_{\text{flux}}$  values  $> 0.3$  tend to have high  $A_{\text{mod}}$  values  $> 0.6$ . This suggests that galaxies with highly asymmetric global H I profiles are likely to have disturbed H I discs, but a symmetric global profile is not necessarily indicative of a symmetric H I disc. Because the shape of a galaxy’s global H I profile is dependent on both the kinematics of the galaxy and the total mass distribution of H I, it is possible that distortions in both properties may conspire to produce a symmetric global H I profile. Furthermore, the shape of the global H I profile is also affected by viewing angle and inclination, which will in turn affect the measured asymmetry (e.g. Deg et al. 2020). It is also apparent that the top-left region of the left panel of Fig. 8, where galaxies with low  $A_{\text{mod}}$  and high  $A_{\text{flux}}$  values would reside, is virtually empty compared to the other regions of the figure. In general, galaxies with low  $A_{\text{mod}}$  values tend to have relatively symmetric global H I profiles, indicating that these galaxies must also have fairly regular kinematics. A Spearman rank test performed using the full sample gives a correlation coefficient of  $r_s = 0.40$  ( $p$ -value  $< 0.001$ ), indicating a weak, but statistically significant, correlation between  $A_{\text{flux}}$  and  $A_{\text{mod}}$ . We compare these results with previously published studies in Section 5.1.2.

### 4.3 The relationship between HI symmetry and baryonic mass

Having analysed the cumulative distributions of  $A_{\text{mod}}$  and  $A_{\text{flux}}$  for our sample, along with the correlation between the two asymmetry parameters, we now explore possible trends between HI asymmetry and baryonic mass. Previous studies have investigated correlations between HI spectral asymmetry and stellar mass (as a proxy for halo mass), albeit with inconclusive results on whether a mass dependence exists (e.g. Reynolds et al. 2020; Watts et al. 2020a; Glowacki et al. 2022; Manuwal et al. 2022). We instead use the baryonic mass since it includes the gas mass, which is an important consideration for low-mass dwarf galaxies that are typically gas-dominated. Furthermore, it has been well established that a galaxy’s baryonic mass and its total dynamical mass are closely correlated, as demonstrated by the baryonic Tully–Fisher relation (e.g. McGaugh et al. 2000; Verheijen 2001; Ponomareva et al. 2021; Gogate et al. 2023).

Fig. 9 shows  $A_{\text{flux}}$  (top row) and  $A_{\text{mod}}$  (bottom) as a function of baryonic mass, adopting the same marker and colour scheme used in Fig. 8 for the three categories. To tease out any trends that may be hidden by the large scatter in the left panels, the right panels display the median asymmetries of each category in bins of baryonic mass, with varying bin widths to ensure that the total number of galaxies per bin remains approximately constant. The vertical error bars and shaded regions depict the standard error on the median and the interquartile range, respectively.

From the left panels and our sample’s baryonic mass distribution (top-right panel of Fig. 2), it is evident that the sample suffers from incompleteness at the low-mass end where there are only galaxies from the SIMBA-25 sample. We fit a single Schechter (1976) function to SIMBA-25’s baryonic mass function (BMF, see Appendix B) and find that the sample is complete down to  $M_{\text{baryon}} \approx 10^{9.5} M_{\odot}$ , before deviating ( $\geq 0.25$  dex) from the best-fitting Schechter function. Since it is unclear how incompleteness at the low-mass end may impact the reliability of any observed trends between HI asymmetry and baryonic mass in this mass regime, only galaxies with  $M_{\text{baryon}} \geq 10^{9.5} M_{\odot}$  are considered when plotting the median trends. We discuss possible reasons for the apparent absence of highly asymmetric galaxies in the low baryonic mass regime in Section 5.2.

As can be seen from the top-right panel of Fig. 9, merged galaxies have the highest median  $A_{\text{flux}}$  values in all mass bins, while interacted galaxies have comparable or marginally higher median  $A_{\text{flux}}$  values than isolated galaxies, supporting the results from Section 4.1. However, we find no trend between  $A_{\text{flux}}$  and baryonic mass for any category. In contrast, the bottom-right panel of Fig. 9 shows a clear inverse trend of decreasing  $A_{\text{mod}}$  with increasing baryonic mass for the interacted and merged categories. There are hints of an inverse correlation between  $A_{\text{mod}}$  and baryonic mass for the isolated category, but the trend is much flatter than the trends observed for the interacted and merged categories. These results suggest that larger galaxies with deeper potential wells, and thus greater gravitational restoring forces, tend to be relatively more symmetric, while the HI discs of lower mass galaxies are more susceptible to being distorted by galaxy–galaxy interactions and mergers. We discuss the differences in trends for  $A_{\text{flux}}$  and  $A_{\text{mod}}$  and the implications for upcoming HI surveys in Section 5.3.

## 5 DISCUSSION

In the following sections, we place the findings presented in this work into the broader context of the existing literature, provide possible explanations for unusual results, and discuss implications for future work.

### 5.1 Comparison with previous work

#### 5.1.1 The HI size–mass relation

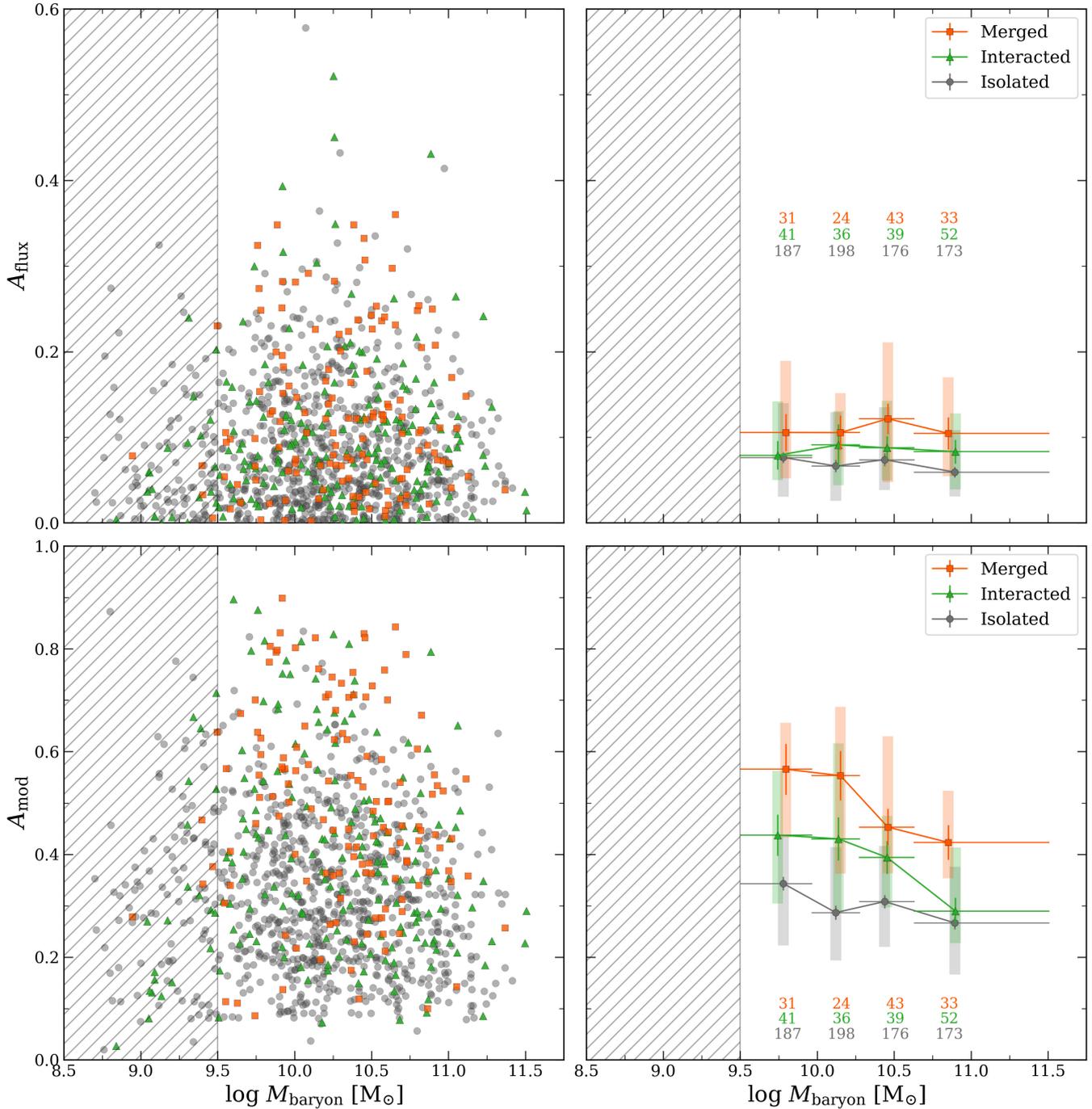
As outlined in Section 2.5, we find that the HI size–mass relations for our SIMBA-25 and SIMBA-100 samples broadly match the observed relation from Wang et al. (2016), but galaxies with  $M_{\text{HI}} \gtrsim 10^{9.5} M_{\odot}$  tend to fall below the relation. Despite masking the data cubes, we are still sensitive to HI emission down to  $\sim 3 \times 10^{18} \text{ cm}^{-2}$  (i.e. an order of magnitude lower than the  $3\sigma$  column density sensitivity of most completed interferometric HI surveys). Coupled with the fact that the data cubes are noise-free, this means that our HI measurements include faint HI emission that would typically be missed in real interferometric observations. Furthermore, the inclusion of faint HI has a larger impact on the measured HI mass than  $R_{\text{HI}}$ , since the latter is by definition measured at the radius where  $\Sigma_{\text{HI}} = 1 M_{\odot} \text{ pc}^{-2}$ , which corresponds to a column density of  $\sim 1.249 \times 10^{20} \text{ cm}^{-2}$ .

More recently, Wang et al. (2024) have combined single-dish data from the Five-hundred-metre Aperture Spherical radio Telescope and the Green Bank Telescope with archival interferometric data from the Very Large Array to recover missing HI flux from the latter due to the lack of short baseline spacings. With updated HI measurements for 13 galaxies, they find that the galaxies shift rightwards by varying amounts in the HI size–mass plane (see their fig. 8), resulting in a slightly shallower relation and more galaxies that lie  $\geq 1\sigma$  below the previously established relation from Wang et al. (2016). While a larger sample is needed to determine whether the HI size–mass relation is indeed shallower, these results tentatively suggest that SIMBA’s relation agrees more with observations that include the diffuse HI component.

#### 5.1.2 The relative fractions of asymmetric galaxies

As already mentioned, various studies have compared the prevalence of asymmetric global HI profiles in different galaxy populations in an effort to infer the probable external mechanisms responsible for the observed asymmetries. Our result of a higher fraction of galaxies with  $A_{\text{flux}} > 0.12$  in the merged category relative to the isolated category (46 per cent versus 28 per cent, see Table 2) qualitatively supports the findings of Bok et al. (2019), but is in conflict with Zuo et al. (2022), who found no significant difference in the levels of spectral asymmetry between their merger and control samples. It should be noted that Zuo et al. (2022) use the curve of growth approach introduced in Yu et al. (2020), as well as different velocity limits, to quantify spectral asymmetries. Furthermore, the experimental samples of both studies consist of galaxies in different merging stages and their control samples have different selection criteria, making a direct comparison between our results and theirs difficult.

Another point of tension between our results and the literature is the relatively high fraction of asymmetric global HI profiles in our isolated category. For example, Espada et al. (2011) found that only 2 per cent of their isolated AMIGA sample exceeded their  $3\sigma$  threshold, whereas 15 per cent of galaxies in our isolated category are found to have  $A_{\text{flux}} > 0.16$ . Espada et al. (2011) do not mention the mass range of their sample, but the lack of a correlation between  $A_{\text{flux}}$  and baryonic mass presented in this work suggests that this is not the main reason for any discrepancy between our results and theirs. However, one possible reason could again be the absence of noise in our data cubes. While it has been shown that noise will artificially inflate spectral asymmetries in low signal-to-noise data (e.g. Deg et al. 2020; Watts et al. 2020a; Bilimogga et al. 2022), it could also lead to a decrease in the measured asymmetry if one half



**Figure 9.** H I spectral asymmetry ( $A_{\text{flux}}$ , top) and morphological asymmetry ( $A_{\text{mod}}$ , bottom) as a function of baryonic mass. The left panels show the individual data points from the isolated (circles), interacted (triangles), and merged (squares) categories. In the right panels, the markers indicate the median H I asymmetry of each category in bins of increasing baryonic mass, with the bin widths shown by the horizontal error bars. The vertical error bars and shaded regions show the standard error on the median and the interquartile (25th–75th percentile) range, respectively. Galaxies that lie in the hatched region (with  $M_{\text{baryon}} < 10^{9.5} M_{\odot}$ ) are excluded from the binned median trends. The number of galaxies in each category is quoted per bin.

of a very lopsided global H I profile has flux density values that are comparable to the noise level of an observation. Similarly to what was noted by Watts et al. (2020b, see their figs 10 and 11), we find a number of asymmetric global H I profiles in the isolated category where one half of the profile consists entirely of faint emission. An example of this is shown in the bottom-right panel of the third row of Fig. 8. If a given observation is not sensitive enough to detect the low flux density half of such a global profile, this could lead to

an underestimation of the intrinsic spectral asymmetry as the profile would likely be mischaracterized as a single-peaked profile. It is also worth noting that the AMIGA data were taken with large single-dish telescopes with relatively small primary beams, thereby suppressing the sensitivity to any asymmetries present in the H I disc outskirts of the targeted galaxies.

With respect to H I morphological asymmetries, spatially asymmetric, diffuse, low column density H I could also exist below the

column density sensitivity of interferometric observations, leading to a lower rate of morphological asymmetries measured in observational data. The presence of this diffuse HI in our simulated isolated galaxies could be due to the accretion of cold gas from the cosmic web or feedback-driven outflows, as Manuwal et al. (2022) found that EAGLE galaxies with asymmetric global profiles are subject to stronger outflows and more elevated accretion rates than those with symmetric global profiles. On the other hand, it is possible that SIMBA and other cosmological simulations may overproduce the amount of diffuse HI in galaxies, and that this results in more asymmetric HI distributions, since the diffuse HI is more easily disturbed by internal and external mechanisms. However, deeper HI observations for a statistical sample are crucial to determine if the excess diffuse HI predicted by simulations is in fact present but hidden in the noise of current observations, or if there is further tension between simulations and observations (see e.g. Marasco et al. 2025).

### 5.1.3 The correlation between HI spectral and morphological asymmetry

Our result of a weak correlation is broadly in agreement with the findings of Reynolds et al. (2020), but is in conflict with those of Bilimogga et al. (2022), who found no correlation between  $A_{\text{flux}}$  and  $A_{\text{mod}}$  for their sample of EAGLE galaxies. Since our mock samples were generated in similar manners and we adopt the same rotation centre and column density threshold used in Bilimogga et al. (2022) when calculating  $A_{\text{mod}}$ , it is unclear why our results disagree. One possible reason for this tension could be that we measure  $A_{\text{mod}}$  at the resolution at which  $N_{\text{beams}} \approx 10$ , whereas they measure  $A_{\text{mod}}$  at a fixed spatial resolution of 56 arcsec (corresponding to a physical scale of  $\sim 4.62$  kpc at the distance of 17 Mpc). To test this, we recalculate the Spearman rank test using the  $A_{\text{mod}}$  values calculated at a fixed spatial resolution of 15 arcsec, which corresponds to a physical scale of  $\sim 5.22$  kpc, and obtain a marginally weaker correlation coefficient of  $r_s = 0.34$  ( $p$ -value  $< 0.001$ ). Thus, our resolution approach in this work has a slight impact on the correlation coefficient measured, but is likely not the reason for the discrepancy between our results and those of Bilimogga et al. (2022). Another reason could be that the differences in how various feedback mechanisms are implemented in SIMBA and EAGLE lead to overall different spatial distributions of HI in the simulations. However, a full investigation into the cause of this conflict is beyond the scope of this study.

## 5.2 On the absence of asymmetric galaxies at low baryonic masses

While investigating the relationship between HI asymmetry and baryonic mass in Section 4.3, we noted the apparent lack of highly asymmetric galaxies below the sample’s mass completeness limit of  $M_{\text{baryon}} \approx 10^{9.5} M_{\odot}$ . This is somewhat counter-intuitive since low-mass galaxies are expected to host more asymmetric HI discs due to their shallower potential wells. A number of studies have observed that highly asymmetric galaxies tend to have lower HI gas fractions compared to their symmetric counterparts (e.g. Watts et al. 2020a; Glowacki et al. 2022). With this in mind, we checked whether low-mass galaxies below the completeness limit are preferentially more HI-rich than the rest of the sample as a possible explanation and found that this is not the case.

We suspect that the absence of high HI asymmetries at the low-mass end is largely a consequence of the HI mass cuts imposed during our sample selection. As mentioned in Section 2.2, we applied

the same HI mass cuts to both the expected HI mass provided by CAESAR and the HI mass that is confined to the star-forming region of the galaxy, to ensure that the inner parts of the HI discs were sufficiently well-resolved. Thus, it seems plausible that galaxies with HI masses close to the HI mass threshold may have truncated, relatively symmetric HI discs with the vast majority of the HI centrally concentrated within the stellar disc. Additionally, the incompleteness of the sample with respect to the BMF, which likely impacts the sampling of galaxies from the different categories, also hinders our ability to make any firm conclusions on the trends observed below the completeness limit.

## 5.3 Implications for future studies

Throughout Section 4, we observed a more distinct separation between the distributions of  $A_{\text{mod}}$  for galaxies with different dynamical histories compared to  $A_{\text{flux}}$ , which systematically underestimated the degree of asymmetry in the galaxies’ HI discs. As previously stated, the shape of a galaxy’s global HI profile is a convolution of the galaxy’s kinematics and the spatial distribution of its HI disc, meaning that the profile’s shape (and thus the measured  $A_{\text{flux}}$  value) can vary significantly depending on the galaxy’s inclination and how an asymmetric feature is oriented relative to the line of sight (e.g. Manuwal et al. 2022; Deg et al. 2023). It is also worth noting that  $A_{\text{flux}}$  will not adequately characterize asymmetries in a global profile if said asymmetries, such as variations in the slopes of the profile edges, do not result in differences in the areas of the two profile halves. Thus, in the context of ongoing and future untargeted HI surveys where the majority of direct detections will be spatially unresolved, these results suggest that  $A_{\text{flux}}$  should be considered a lower limit on the intrinsic HI asymmetry, provided that the data used are of sufficient quality in terms of their signal-to-noise ratio and spectral resolution (see e.g. Deg et al. 2020; Watts et al. 2020a). In particular, it must be emphasized that a low  $A_{\text{flux}}$  value cannot be used to infer that an HI disc is undisturbed.

Regarding future studies with access to large samples of spatially resolved HI observations, we have shown that failing to account for differences in spatial resolution can introduce unexpected biases. To fairly compare galaxies within a survey spanning a wide range of HI masses, it is important to match the relative spatial resolution across the sample when calculating morphological asymmetries. Alternatively, another approach would be to group galaxies into subcategories based on their spatial resolution to avoid reducing the resolution of the entire sample to that of the least spatially resolved galaxy. However, with the second approach, there is the risk of further dividing what may already be a limited sample. We have also demonstrated that baryonic mass is another crucial factor to consider when comparing the HI morphological asymmetries obtained from data sets across different surveys. If one wishes to investigate the trends between HI asymmetries and environment, care must be taken to match samples in terms of their masses.

Lastly, it is important to acknowledge that while  $A_{\text{mod}}$  is generally better at characterizing disturbed HI morphologies, the overlap of different categories at various  $A_{\text{mod}}$  values indicates that  $A_{\text{mod}}$  on its own is insufficient to deduce the most probable mechanism responsible for the observed HI asymmetry of a galaxy. As such, applying a threshold in  $A_{\text{mod}}$  would not result in a clear separation between our galaxy categories. However, incorporating an analysis of the HI kinematics could assist in the interpretation of HI asymmetries. For example, Serra et al. (2024) recently presented a detailed investigation of dwarf galaxy NGC 1427A in the Fornax cluster, in which they convincingly argue that the complex HI morphology and

kinematics of the system results from ram pressure being exerted by the intracluster medium in the aftermath of a tidal interaction (or merger) between two galaxies. Such an analysis is beyond the scope of this study, but it is a compelling direction worth pursuing in our future work with the MDS.

## 6 CONCLUSIONS

We have used the SIMBA cosmological simulations to investigate the relationship between the recent dynamical histories and HI spectral and morphological asymmetries of  $\sim 1100$  spatially resolved galaxies at low redshift. We created synthetic HI data cubes designed to match HI observations from the MDS to enable the extraction of HI data products in an observational manner, and quantified asymmetries in the global HI profiles and moment-0 maps using the flux ratio asymmetry ( $A_{\text{flux}}$ ) and the modified asymmetry parameter ( $A_{\text{mod}}$ ), respectively. We placed galaxies into three main categories (isolated, interacted, or merged) based on their dynamical histories over the preceding  $\sim 2$  Gyr, and defined a refined isolated category to check the robustness of our results for the isolated category. We then compared the HI asymmetry distributions of the various categories and explored trends between asymmetry and baryonic mass. Our findings are summarized as follows:

(i) Galaxies from SIMBA follow the observed HI size–mass relation from Wang et al. (2016) relatively well, but tend to have slightly smaller HI diameters at  $M_{\text{HI}} \gtrsim 10^{9.5} M_{\odot}$ . This likely owes to the absence of noise in the mock data cubes, which can have a larger impact on the measured HI mass than the HI diameter when more diffuse HI is recovered in the disc outskirts (see e.g. Wang et al. 2024).

(ii) A spurious trend between HI mass and  $A_{\text{mod}}$  is introduced when measuring  $A_{\text{mod}}$  at a fixed resolution, which is largely a secondary effect of the HI size–mass relation. This trend disappears when  $A_{\text{mod}}$  is measured using the moment-0 map smoothed to the spatial resolution at which a galaxy is resolved by approximately 10 beams in its area. This highlights that the relative spatial resolution must be taken into account when comparing  $A_{\text{mod}}$  values in a sample that spans a wide range of HI masses.

(iii) The cumulative distributions for both  $A_{\text{flux}}$  and  $A_{\text{mod}}$  show that the refined isolated and isolated categories are statistically indistinguishable from one another, indicating that our results are insensitive to the chosen isolation selection criteria.

(iv) Galaxies in the interacted category are, on average, more asymmetric than galaxies in the isolated category, but a systematic difference between the two categories' asymmetry distributions is only found with  $A_{\text{mod}}$  and not  $A_{\text{flux}}$ . Galaxies in the merged category are generally even more asymmetric, with statistically different asymmetry distributions compared to both the isolated and interacted categories. These results suggest that galaxy mergers and interactions are likely responsible for the higher mean asymmetries observed in these categories.

(v) For the interacted and merged categories, an inverse correlation exists between baryonic mass and  $A_{\text{mod}}$ , suggesting that massive galaxies are typically more symmetric, whereas lower mass galaxies are more prone to having their HI discs distorted from interactions and mergers. This trend is not seen between baryonic mass and  $A_{\text{flux}}$  and is likely due to  $A_{\text{flux}}$  underestimating the degree of asymmetry of a galaxy's HI disc compared to  $A_{\text{mod}}$ .

Overall, these results showcase the importance of spatially resolved HI observations and the need for caution when relying solely on the global HI profiles of galaxies to study HI asymmetries

in relation to other galaxy properties. Looking to the future, a natural progression of this work will be to incorporate kinematic asymmetries and expand our analysis to observations from the MDS. This is a rich data set with HI detections across a range of cosmic environments, from voids to filaments and clusters, allowing us to investigate the average gas content of various galaxy populations and how different environmental mechanisms impact the HI reservoirs of galaxies.

## ACKNOWLEDGEMENTS

The authors thank the anonymous referee for constructive comments and suggestions. We acknowledge helpful discussions with the ASymba team, N. Hatamkhani, J. Healy, and R. Talens. We also thank J. Wang for providing the observational data used in Fig. 3.

NH acknowledges support from the National Research Foundation of South Africa (grant no. 120224) and the Leids Kerkhoven–Bosscha Fonds (LKBF). KAO acknowledges support by the Royal Society through Dorothy Hodgkin Fellowship DHF/R1/231105. MG is supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (DP210102103), and through UK Science and Technology Facilities Council (STFC) grant ST/Y001117/1. MG acknowledges support from the Inter-University Institute for Data Intensive Astronomy (IDIA). IDIA is a partnership of the University of Cape Town (UCT), the University of Pretoria, and the University of the Western Cape.

We acknowledge the use of the ilifu cloud computing facility ([www.ilifu.ac.za](http://www.ilifu.ac.za)), a partnership between UCT, the University of the Western Cape, the University of Stellenbosch, Sol Plaatje University, the Cape Peninsula University of Technology, and the South African Radio Astronomy Observatory. The ilifu facility is supported by contributions from IDIA, the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA). This work has made use of the following PYTHON packages: ASTROPY (Astropy Collaboration 2018), MATPLOTLIB (Hunter 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), and SCIPY (Virtanen et al. 2020). This work has also made use of NASA's Astrophysics Data System. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

## DATA AVAILABILITY

The SIMBA simulation snapshots and accompanying CAESAR catalogues are publicly available at <http://simba.roe.ac.uk>. The data underlying this work will be made available upon reasonable request to the corresponding author.

## REFERENCES

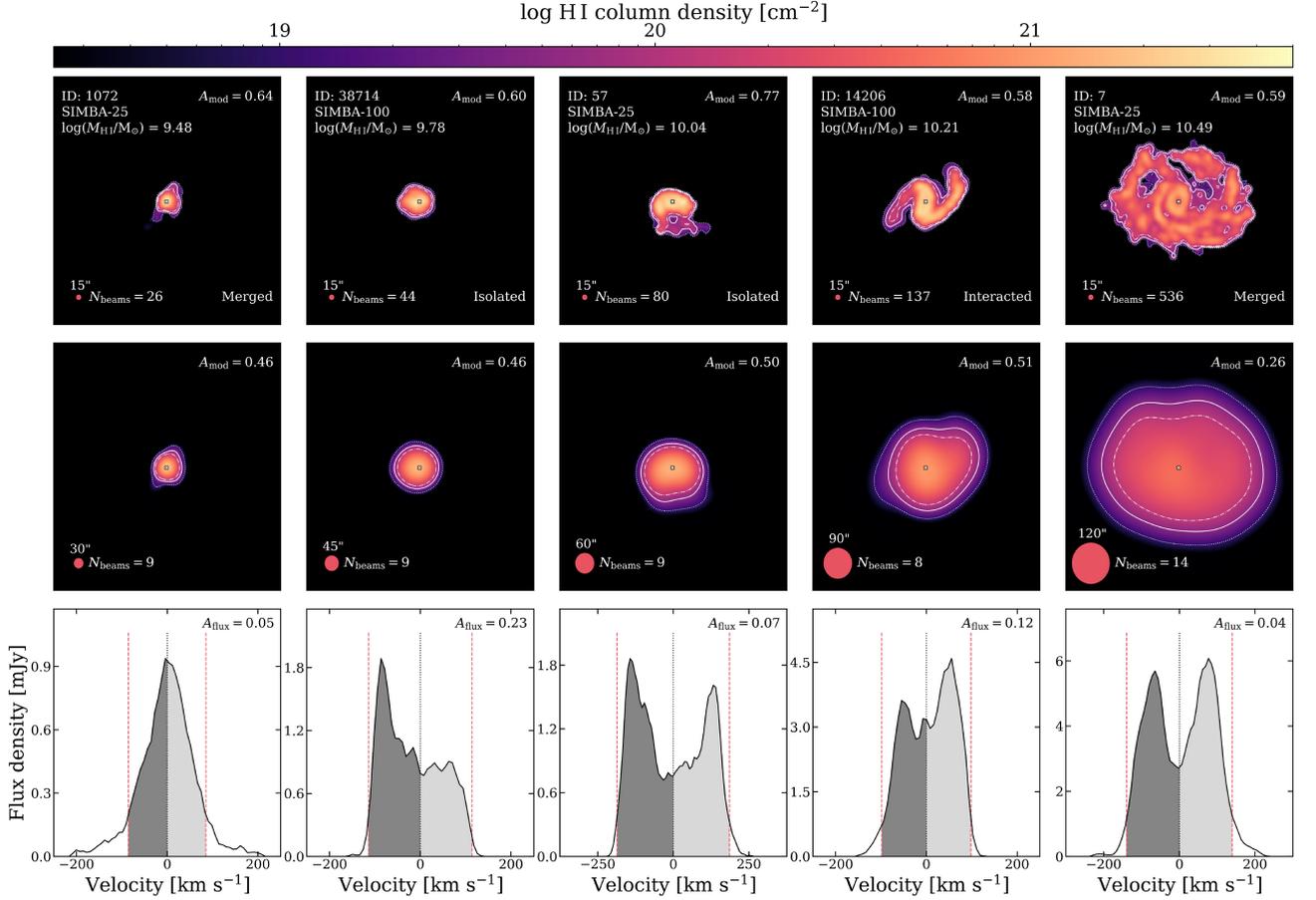
- Abraham R. G., van den Bergh S., Nair P., 2003, *ApJ*, 588, 218  
 Abruzzo M. W., Narayanan D., Davé R., Thompson R., 2018, preprint (arXiv:1803.02374)  
 Adams E. A. K., van Leeuwen J., 2019, *Nat. Astron.*, 3, 188  
 Angiras R. A., Jog C. J., Dwarakanath K. S., Verheijen M. A. W., 2007, *MNRAS*, 378, 276  
 Angiras R. A., Jog C. J., Omar A., Dwarakanath K. S., 2006, *MNRAS*, 369, 1849  
 Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017, *MNRAS*, 464, 2840  
 Appley S., Davé R., Kraljic K., Anglés-Alcázar D., Narayanan D., 2020, *MNRAS*, 494, 6053  
 Arp H., 1966, *ApJS*, 14, 1

- Astropy Collaboration, 2018, *AJ*, 156, 123
- Bahé Y. M. et al., 2016, *MNRAS*, 456, 1115
- Baldwin J. E., Lynden-Bell D., Sancisi R., 1980, *MNRAS*, 193, 313
- Begum A., Chengalur J. N., Karachentsev I. D., Sharina M. E., Kaisin S. S., 2008, *MNRAS*, 386, 1667
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJ*, 585, L117
- Bignone L. A., Pedrosa S. E., Trayford J. W., Tissera P. B., Pellizza L. J., 2020, *MNRAS*, 491, 3624
- Bilimogga P. V., Oman K. A., Verheijen M. A. W., van der Hulst T., 2022, *MNRAS*, 513, 5310
- Bluck A. F. L., Conselice C. J., Buitrago F. o., Grützbauch R., Hoyos C., Mortlock A., Bauer A. E., 2012, *ApJ*, 747, 34
- Blyth S. et al., 2016, in Taylor A. R., Camilo F., Leeuw L., Moodley K. eds, Proc. Sci. Vol. 277, MeerKAT Science: On the Pathway to the SKA. SISSA, Trieste, PoS#4
- Bok J., Blyth S. L., Gilbank D. G., Elson E. C., 2019, *MNRAS*, 484, 582
- Bondi H., 1952, *MNRAS*, 112, 195
- Bournaud F., Combes F., Jog C. J., Puerari I., 2005, *A&A*, 438, 507
- Broeils A. H., Rhee M. H., 1997, *A&A*, 324, 877
- Catinella B. et al., 2018, *MNRAS*, 476, 875
- Catinella B., Cortese L., 2015, *MNRAS*, 446, 3526
- Chabrier G., 2003, *ApJ*, 586, L133
- Chemin L. et al., 2006, *MNRAS*, 366, 812
- Choque-Challapa N., Aguerrí J. A. L., Mancera Piña P. E., Peletier R., Venhola A., Verheijen M., 2021, *MNRAS*, 507, 6045
- Conselice C. J., 2003, *ApJS*, 147, 1
- Conselice C. J., Bershadly M. A., Jangren A., 2000, *ApJ*, 529, 886
- Cui W., Davé R., Peacock J. A., Anglés-Alcázar D., Yang X., 2021, *Nat. Astron.*, 5, 1069
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieerantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, *MNRAS*, 497, 146
- Davé R., Rafieerantsoa M. H., Thompson R. J., Hopkins P. F., 2017, *MNRAS*, 467, 115
- Davé R., Thompson R., Hopkins P. F., 2016, *MNRAS*, 462, 3265
- Deb T., Verheijen M. A. W., van der Hulst J. M., 2023, *A&A*, 676, A118
- Deg N., Blyth S. L., Hank N., Kruger S., Carignan C., 2020, *MNRAS*, 495, 1984
- Deg N., Perron-Cormier M., Spekkens K., Glowacki M., Blyth S. L., Hank N., 2023, *MNRAS*, 523, 4340
- Di Teodoro E. M., Fraternali F., 2014, *A&A*, 567, A68
- Diemer B. et al., 2019, *MNRAS*, 487, 1529
- Eckert K. D., Kannappan S. J., Stark D. V., Moffett A. J., Berlind A. A., Norris M. A., 2016, *ApJ*, 824, 124
- Espada D., Verdes-Montenegro L., Huchtmeier W. K., Sulentic J., Verley S., Leon S., Sabater J., 2011, *A&A*, 532, A117
- Fernández X. et al., 2013, *ApJ*, 770, L29
- Gensior J. et al., 2024, *MNRAS*, 531, 1158
- Giese N., van der Hulst T., Serra P., Oosterloo T., 2016, *MNRAS*, 461, 1656
- Glowacki M., Deg N., Blyth S. L., Hank N., Davé R., Elson E., Spekkens K., 2022, *MNRAS*, 517, 1282
- Glowacki M., Elson E., Davé R., 2020, *MNRAS*, 498, 3687
- Gogate A. R., Verheijen M. A. W., van der Hulst J. M., Jaffé Y. L., 2023, *MNRAS*, 519, 4279
- Gunn J. E., Gott J. Richard I., 1972, *ApJ*, 176, 1
- Haardt F., Madau P., 2012, *ApJ*, 746, 125
- Hahn O., Abel T., 2011, *MNRAS*, 415, 2101
- Haynes M. P. et al., 2018, *ApJ*, 861, 49
- Haynes M. P., Hogg D. E., Maddalena R. J., Roberts M. S., van Zee L., 1998, *AJ*, 115, 62
- Hibbard J. E., van Gorkom J. H., Rupen M. P., Schiminovich D., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H. eds, ASP Conf. Ser. Vol. 240, Gas and Galaxy Evolution. Astron. Soc. Pac., San Francisco, p. 657
- Holwerda B. W. et al., 2023, *MNRAS*, 521, 1502
- Holwerda B. W. et al., 2025, *Publ. Astron. Soc. Aust.*, 42, e028
- Holwerda B. W., Pirzkal N., Cox T. J., de Blok W. J. G., Weniger J., Bouchard A., Blyth S. L., van der Heyden K. J., 2011c, *MNRAS*, 416, 2426
- Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S. L., van der Heyden K. J., Elson E. C., 2011a, *MNRAS*, 416, 2401
- Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S. L., van der Heyden K. J., Elson E. C., 2011b, *MNRAS*, 416, 2415
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Hopkins P. F., Quataert E., 2011, *MNRAS*, 415, 1027
- Hotan A. W. et al., 2021, *Publ. Astron. Soc. Aust.*, 38, e009
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jaffé Y. L., Poggianti B. M., Verheijen M. A. W., Deshev B. Z., van Gorkom J. H., 2013, *MNRAS*, 431, 2111
- Jarvis M. et al., 2016, in Taylor A. R., Camilo F., Leeuw L., Moodley K. eds, Proc. Sci. Vol. 277, MeerKAT Science: On the Pathway to the SKA. SISSA, Trieste, PoS#6
- Jog C. J., Combes F., 2009, *Phys. Rep.*, 471, 75
- Johnston S. et al., 2008, *Exp. Astron.*, 22, 151
- Koribalski B. S. et al., 2020, *Ap&SS*, 365, 118
- Krumholz M. R., Gnedin N. Y., 2011, *ApJ*, 729, 36
- Lelli F., McGaugh S. S., Schombert J. M., 2016, *AJ*, 152, 157
- Lelli F., Verheijen M., Fraternali F., 2014, *MNRAS*, 445, 1694
- Li Q., Narayanan D., Davé R., 2019, *MNRAS*, 490, 1425
- Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010a, *MNRAS*, 404, 575
- Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010b, *MNRAS*, 404, 590
- Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163
- Lutz K. A. et al., 2018, *MNRAS*, 476, 3744
- Maddox N. et al., 2021, *A&A*, 646, A35
- Manuwal A., Ludlow A. D., Stevens A. R. H., Wright R. J., Robotham A. S. G., 2022, *MNRAS*, 510, 3408
- Mapelli M., Moore B., Bland-Hawthorn J., 2008, *MNRAS*, 388, 697
- Marasco A. et al., 2025, *A&A*, 697, A86
- Marinacci F., Grand R. J. J., Pakmor R., Springel V., Gómez F. A., Frenk C. S., White S. D. M., 2017, *MNRAS*, 466, 3859
- Matthews L. D., van Driel W., Gallagher J. S. L., 1998, *AJ*, 116, 1169
- McGaugh S. S., Schombert J. M., Bothun G. D., de Blok W. J. G., 2000, *ApJ*, 533, L99
- Meyer M., 2009, in Heald G., Serra P. eds, Proc. Sci. Vol. 89, Panoramic Radio Astronomy: Wide-field 1–2 GHz Research on Galaxy Evolution. SISSA, Trieste, PoS#15
- Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, *Publ. Astron. Soc. Aust.*, 34, 52
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, 379, 613
- Moreno J. et al., 2021, *MNRAS*, 503, 3113
- Naluminsa E., Elson E. C., Jarrett T. H., 2021, *MNRAS*, 502, 5711
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Noordermeer E., Sparke L. S., Levine S. E., 2001, *MNRAS*, 328, 1064
- Nulsen P. E. J., 1982, *MNRAS*, 198, 1007
- Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009, *ApJ*, 698, 1467
- Oman K. A., 2019, Astrophysics Source Code Library, record ascl:1911.005
- Oman K. A., Marasco A., Navarro J. F., Frenk C. S., Schaye J., Benítez-Llambay A., 2019, *MNRAS*, 482, 821
- Oman K., 2024, *J. Open Source Softw.*, 9, 6860
- Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, *ApJ*, 759, 138
- Patton D. R., Torrey P., Ellison S. L., Mendel J. T., Scudder J. M., 2013, *MNRAS*, 433, L59
- Peterson S. D., Shostak G. S., 1974, *AJ*, 79, 767
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Ponomareva A. A. et al., 2021, *MNRAS*, 508, 1195
- Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, *MNRAS*, 430, 2427
- Rajohnson S. H. A. et al., 2022, *MNRAS*, 512, 2697
- Reynolds T. N., Westmeier T., Staveley-Smith L., Chauhan G., Lagos C. D. P., 2020, *MNRAS*, 493, 5089
- Richter O. G., Sancisi R., 1994, *A&A*, 290, L9
- Roberts I. D. et al., 2023, *A&A*, 675, A78
- Rodríguez Montero F., Davé R., Wild W., Anglés-Alcázar D., Narayanan D., 2019, *MNRAS*, 490, 2139
- Salucci P., Persic M., 1999, *MNRAS*, 309, 923

- Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, *A&AR*, 15, 189
- Schade D., Lilly S. J., Crampton D., Hammer F., Le Fevre O., Tresse L., 1995, *ApJ*, 451, L1
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schechter P., 1976, *ApJ*, 203, 297
- Schmidt M., 1959, *ApJ*, 129, 243
- Scott T. C., Brinks E., Cortese L., Boselli A., Bravo-Alfaro H., 2018, *MNRAS*, 475, 4648
- Serra P. et al., 2024, *A&A*, 690, A4
- Smith B. D. et al., 2017, *MNRAS*, 466, 2217
- Springel V., 2005, *MNRAS*, 364, 1105
- Stevens A. R. H., Diemer B., Lagos C. d. P., Nelson D., Obreschkow D., Wang J., Marinacci F., 2019, *MNRAS*, 490, 96
- Swaters R. A., Schoenmakers R. H. M., Sancisi R., van Albada T. S., 1999, *MNRAS*, 304, 330
- Thomas N., Davé R., Anglés-Alcázar D., Jarvis M., 2019, *MNRAS*, 487, 5764
- Thorp M. D., Bluck A. F. L., Ellison S. L., Maiolino R., Conselice C. J., Hani M. H., Bottrell C., 2021, *MNRAS*, 507, 886
- Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, *ApJS*, 192, 9
- van Cappellen W. A. et al., 2022, *A&A*, 658, A146
- van der Hulst J. M., Terlouw J. P., Begeman K. G., Zwitser W., Roelfsema P. R., 1992, in Worrall D. M., Biemesderfer C., Barnes J. eds, ASP Conf. Ser. Vol. 25, *Astronomical Data Analysis Software and Systems I*. Astron. Soc. Pac., San Francisco, p. 131
- van der Hulst J. M., van Albada T. S., Sancisi R., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H. eds, ASP Conf. Ser. Vol. 240, *Gas and Galaxy Evolution*. Astron. Soc. Pac., San Francisco, p. 451
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- van Eymeren J., Jütte E., Jog C. J., Stein Y., Dettmar R. J., 2011a, *A&A*, 530, A29
- van Eymeren J., Jütte E., Jog C. J., Stein Y., Dettmar R. J., 2011b, *A&A*, 530, A30
- Verdes-Montenegro L., Sulentic J., Lisenfeld U., Leon S., Espada D., Garcia E., Sabater J., Verley S., 2005, *A&A*, 436, 443
- Verheijen M. A. W., 2001, *ApJ*, 563, 694
- Verheijen M. A. W., Oosterloo T. A., van Cappellen W. A., Bakker L., Ivashina M. V., van der Hulst J. M., 2008, in Minchin R., Momjian E. eds, AIP Conf. Ser. Vol. 1035, *The Evolution of Galaxies Through the Neutral Hydrogen Window*. Am. Inst. Phys., New York, p. 265
- Verheijen M. A. W., Sancisi R., 2001, *A&A*, 370, 765
- Verheijen M., Oosterloo T., Heald G., van Cappellen W., 2009, in Heald G., Serra P. eds, Proc. Sci. Vol. 89, *Panoramic Radio Astronomy: Wide-field 1–2 GHz Research on Galaxy Evolution*. SISSA, Trieste, PoS#10
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Vogelaar M. G. R., Terlouw J. P., 2001, in Harnden F. R. J., Primini F. A., Payne H. E. eds, ASP Conf. Ser. Vol. 238, *Astronomical Data Analysis Software and Systems X*. Astron. Soc. Pac., San Francisco, p. 358
- Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt Robert C. J., Thornley M. D., Leroy A., 2008, *AJ*, 136, 2563
- Wang J. et al., 2013, *MNRAS*, 433, 270
- Wang J. et al., 2014, *MNRAS*, 441, 2159
- Wang J. et al., 2024, *ApJ*, 973, 15
- Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, *MNRAS*, 460, 2143
- Watts A. B., Catinella B., Cortese L., Power C., 2020a, *MNRAS*, 492, 3672
- Watts A. B., Power C., Catinella B., Cortese L., Stevens A. R. H., 2020b, *MNRAS*, 499, 5205
- Whitney A., Ferreira L., Conselice C. J., Duncan K., 2021, *ApJ*, 919, 139
- Yu N., Ho L. C., Wang J., 2020, *ApJ*, 898, 102
- Zuo P., Ho L. C., Wang J., Yu N., Shangguan J., 2022, *ApJ*, 929, 15

## APPENDIX A: MOMENT-0 MAPS AND GLOBAL H I PROFILES

Fig. A1 shows the moment-0 maps (top and middle rows) and global H I profiles (bottom row) of five galaxies, arranged from left to right in order of increasing H I mass. For each galaxy, the moment-0 maps are shown at 15 arcsec resolution (top panel) and the resolution at which  $N_{\text{beams}} \approx 10$  (middle). The value of  $N_{\text{beams}}$  is given in the bottom-left corner of the panels in the top and middle rows next to the ellipse showing the size of the beam. Each panel presenting the moment-0 maps is 250 kpc across and centred on the minimum of the galaxy's gravitational potential, which is shown by a white cross. The white dash-dotted, solid, and dotted contours are at column densities of 1, 5, and  $12.5 \times 10^{19} \text{ cm}^{-2}$ . For each global H I profile, the dashed pink lines represent  $v_{l,20}$  and  $v_{h,20}$ , and the black dotted line represents  $v_{\text{sys}}$ . The galaxy ID, volume, logarithm of the H I mass, and the category are given in the top panels. The measured  $A_{\text{mod}}$  and  $A_{\text{flux}}$  values are quoted in the top-right corner of the relevant panel.



**Figure A1.** Moment-0 maps (top and middle rows) and global HI profiles (bottom row) of five galaxies in our sample. See the text of Appendix A for more details.

## APPENDIX B: DETERMINING THE INCOMPLETENESS AT LOW BARYONIC MASSES

One way to assess the incompleteness of a sample is to fit a model function to the data and establish where the fit deviates significantly. Numerous studies (see e.g. Salucci & Persic 1999; Bell et al. 2003; Papastergis et al. 2012; Eckert et al. 2016) have shown that the BMF is well described by a single Schechter function, given by

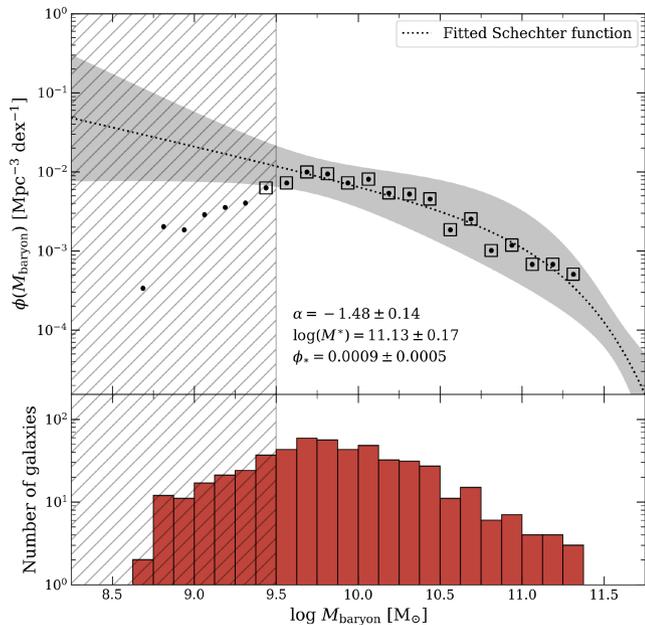
$$\phi(M_{\text{baryon}}) = \ln(10)\phi_* \left(\frac{M_{\text{baryon}}}{M^*}\right)^{(\alpha+1)} \exp\left(-\frac{M_{\text{baryon}}}{M^*}\right), \quad (\text{B1})$$

where  $M^*$  is the characteristic ‘knee’ mass above which the number of galaxies declines exponentially,  $\alpha$  sets the slope of the power law at the low-mass end, and  $\phi_*$  is the normalization constant.

We use the SIMBA-25 sample to determine the completeness limit, since the incompleteness is most apparent at the low-mass end of the

baryonic mass distribution, where there are only galaxies from this sample. Additionally, attempting to fit the entire sample introduces complications because the two samples are taken from different volumes.

A non-linear least-squares fitting routine is used to fit equation (B1) to the mid- to high-mass end of SIMBA-25’s BMF, and the uncertainties are computed from the covariance matrix. The best-fitting Schechter function and the  $3\sigma$  uncertainty of the fit are shown in Fig. B1 as the black dotted line and grey shaded region, respectively. For transparency, we include the parameters of the best fit in the figure, but stress that we are not attempting to compare the results of the fit to those found in the literature. We estimate the baryonic mass completeness limit to be  $M_{\text{baryon}} = 10^{9.5} M_{\odot}$  as the BMF and the best Schechter fit are in broad agreement down to this mass, before the data diverge by more than 0.25 dex from the fit (shown by the hatched region in the figure).



**Figure B1.** The BMF of the SIMBA-25 sample (closed circle markers). The open square symbols represent the data points used to fit the single Schechter function. The parameters of the best fit, which is shown by the dotted line, are quoted in the top panel. The shaded region shows the  $3\sigma$  uncertainty of the fit. The histogram of the baryonic mass distribution is shown in the bottom panel. The hatched region indicates where the sample is incomplete.

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