# MIGHTEE-H I : the M_HI – M* relation over the last billion years

Pan, H.; Jarvis, M.J.; Santos, M.G.; Maddox, N.; Frank, B.S.; Ponomareva, A.A.; … ; Collier, J.D.

# MIGHTEE-H I: the $M_{HI} - M_\star$ relation over the last billion years

Hengxing Pan [1,2]★ Matt J. Jarvis, [1,2] Mario G. Santos, [1,3] Natasha Maddox, [4,5] Bradley S. Frank, [3,6,7]
Anastasia A. Ponomareva, [2] Isabella Prandoni, [8] Sushma Kurapati, [7] Maarten Baes, [9]
Pavel E. Mancera Piña, [10] Giulia Rodighiero, [11,12] Martin J. Meyer, [13] Romeel Davé, [14]
Gauri Sharma, [1,15,16] Sambatriniaina H. A. Rajohnson, [7] Nathan J. Adams, [17]
Rebecca A. A. Bowler, [17] Francesco Sinigaglia, [11,12] Thijs van der Hulst, [18] Peter W. Hatfield, [2]
Srikrishna Sekhar [6,19] and Jordan D. Collier[6,20,21]

*Affiliations are listed at the end of the paper*

## ABSTRACT

We study the $M_{HI} - M_\star$ relation over the last billion years using the MIGHTEE-H I sample. We first model the upper envelope of the $M_{HI} - M_\star$ relation with a Bayesian technique applied to a total number of 249 H I-selected galaxies, without binning the data sets, while taking account of the intrinsic scatter. We fit the envelope with both linear and non-linear models, and find that the non-linear model is preferred over the linear one with a measured transition stellar mass of $\log_{10}(M_\star/M_\odot) = 9.15 \pm 0.87$, beyond which the slope flattens. This finding supports the view that the lack of H I gas is ultimately responsible for the decreasing star formation rate observed in the massive main-sequence galaxies. For spirals alone, which are biased towards the massive galaxies in our sample, the slope beyond the transition mass is shallower than for the full sample, indicative of distinct gas processes ongoing for the spirals/high-mass galaxies from other types with lower stellar masses. We then create mock catalogues for the MIGHTEE-H I detections and non-detections with two main galaxy populations of late- and early-type galaxies to measure the underlying $M_{HI} - M_\star$ relation. We find that the turnover in this relation persists whether considering the two galaxy populations as a whole or separately. We note that an underlying linear relation could mimic this turnover in the observed scaling relation, but a model with a turnover is strongly preferred. Measurements on the logarithmic average of H I masses against the stellar mass are provided as a benchmark for future studies.

**Key words:** methods: statistical – galaxies: fundamental parameters – radio lines: galaxies.

## 1 INTRODUCTION

The relation between the mass of neutral atomic hydrogen (H I) gas and stars in galaxies reveals the connection of star-forming activity to their raw fuel, but this relation is not straightforward due to complex physical processes involved in the course of galaxy evolution. A comprehensive and accurate measurement of this relation is required to illuminate their interplay, and thus to help us better understand the physics of galaxy formation and evolution.

Over the past few decades, the direct detection of emission lines from the neutral hydrogen component of galaxies has been limited to the local Universe, or massive H I systems, by the sensitivity of modern radio instruments, such as Parkes and Arecibo telescopes. None the less, several studies have been conducted to investigate the H I and stellar mass relation, benefiting from the H I Parkes All-Sky Survey (Barnes et al. 2001) and the Arecibo Legacy Fast ALFA (ALFALFA) survey (Giovanelli et al. 2005).

In particular, exploring the upper envelope of the H I and stellar mass ($M_{HI} - M_\star$) relation has been one of the main means used to enlighten the processes of gas consumption and star formation. For example, both Huang et al. (2012) and Maddox et al. (2015) have found an upper limit for H I mass as a function of the stellar mass at high masses for H I-selected samples. Maddox et al. (2015) and Parkash et al. (2018) suggest that this upper limit can be explained by a stability model in which the large halo spin of disc galaxies can stabilize the H I disc and prevent it from collapsing and forming stars (Obreschkow et al. 2016). In this scenario, the highest spin parameter is restrained by the amount of gas infall and tidal torque that haloes can experience during the proto-galactic growth, and therefore the gas fraction is linked to the specific angular momentum of galaxies in general (Zoldan et al. 2018; Mancera Piña et al. 2021b). As such, this could also be related to the position of the galaxies with respect to the cosmic web, the filaments of which are presumably the source of the infalling gas (see e.g. Tudorache et al. 2022). On the other hand, the underlying $M_{HI} - M_\star$ relation has been investigated by Catinella et al. (2010) and Parkash et al. (2018) for example, based on stellar mass-selected samples as H I-selected samples tend to exclude H I-poor galaxies resulting in measurements of high average (or median) H I masses. They conclude that the observed flatness in the underlying $M_{HI} - M_\star$ relation is due to the increasing fraction of gas poor early-type galaxies (EGTs). All these studies indicate an increase in the H I mass with stellar mass, but diverge at the high mass end owing to the effect of sample selection, limited statistics, or both.

★ E-mail: hpan@uwc.ac.za

Furthermore, there have been discoveries of flattening of the star formation rate (SFR)-$M_{\star}$ relation at high stellar masses from the local to high-z Universe (e.g. Noeske et al. 2007; Erfanianfar et al. 2015; Johnston et al. 2015; Leslie et al. 2020; Fraser-McKelvie et al. 2021). The mechanisms responsible for this flattening remain under debate (e.g. Gavazzi et al. 2015; Tacchella et al. 2018; Popesso et al. 2019; Cook et al. 2020; Feldmann 2020), and can be broadly summarized as the lack of H I gas versus the low conversion efficiency from H I to stars, through the molecular hydrogen ($H_2$) phase. A thorough investigation into the link between H I and the stellar mass can help to disentangle these two possible causes. Noticeably, the flattening of the SFR-$M_{\star}$ relation towards high masses resembles the upper limit found on the $M_{H\,\textsc{i}} - M_{\star}$ relation.

At high redshifts, stacking approaches (e.g. Delhaize et al. 2013; Rhee et al. 2013; Healy et al. 2019; Chowdhury et al. 2020; Guo et al. 2021; Bera et al. 2022, 2023; Sinigaglia et al. 2022) have been developed to break the barrier of the sensitivity limitation. However, in these stacking processes, one only measures the average properties of galaxies bearing the consequence of losing information about their intrinsic scatter, which is a key parameter to describe the shape of the distribution of H I masses, and hence the strength of the correlation between H I and a second galaxy property, such as the stellar mass.

In addition, only the arithmetic operations (e.g. average) of the H I fluxes are allowed for these stacking practices as the logarithmic operation cannot be done for negative fluxes that are influenced by the background noise, although an arithmetic average would be sufficient if we were just interested in measuring the cosmic H I density. In terms of scaling relations, there are notable differences in the means and standard deviations between arithmetic and logarithmic operations of H I masses mostly due to the different contribution of the low mass samples (Rodríguez-Puebla et al. 2020; Saintonge & Catinella 2022). For H I-selected samples, the logarithmic average (or median) can adequately trace the main distribution, and is preferred in the literature (Cortese et al. 2011; Huang et al. 2012; Maddox et al. 2015; Parkash et al. 2018). Therefore, it will add further complexities to a fair comparison of measured scaling relations between stacked samples and direct detections, based on different statistics. Above all, these approaches require binning the data sets in a second galaxy property, and it could be troublesome to determine the binning width when the sample size is small.

With the MeerKAT radio telescope and the future SKA, we are now entering a new era of radio astronomy. The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE; Jarvis et al. 2016) is one of the large survey projects that is underway with MeerKAT, and will cover 20 square degrees over the four best-studied extragalactic fields observable from the southern hemisphere to $\mu$Jy sensitivity at GHz frequencies. The MIGHTEE-H I Early Science project has already allowed us to reach $M_{H\,\textsc{i}} \lesssim 10^7\,M_{\odot}$ in the local Universe, and $M_{H\,\textsc{i}} \sim 10^9\,M_{\odot}$ up to $z = 0.084$, with higher H I-mass galaxies observable out to the lower-frequency end of the L-band window at $z \sim 0.6$ (Maddox et al. 2021). Furthermore, the combination of this depth and high spatial resolution, resulting in essentially no source confusion compared to single-dish surveys, and the depth of the ancillary data, allowing us to reach stellar masses of $\sim 10^6\,M_{\star}$, makes this a unique data set to investigate the $M_{H\,\textsc{i}} - M_{\star}$ scaling relation over a large baseline in H I and stellar mass.

In this paper, we first use a Bayesian technique for modelling the upper envelope of the $M_{H\,\textsc{i}} - M_{\star}$ scaling relation consistently without binning the data sets from the MIGHTEE-H I catalogue, while taking account of the intrinsic scatter, based on our previous work (Pan et al. 2020, 2021). This technique employs fluxes of H I emission line as measurables that can naturally account for the thermal noise from

**Table 1.** Key parameters of the MIGHTEE-H I Early Science data.

| | |
|---|---|
| Survey area | $\sim 1.5$ deg$^2$ (COSMOS) |
| | $\sim 3 \times 1.2$ deg$^2$ (XMMLSS) |
| Integration time | $\sim 16$ h (COSMOS) |
| | $3 \times 12$ h (XMMLSS) |
| Velocity resolution | 44.11 km s$^{-1}$ at $z = 0$ |
| Synthesized beam | $14.5'' \times 11''$ (COSMOS) |
| | $12'' \times 10''$ (XMMLSS) |
| $3\sigma$ H I column density sensitivity | $4.05 \times 10^{19}$ atoms cm$^{-2}$ (COSMOS) |
| | $9.83 \times 10^{19}$ atoms cm$^{-2}$ (XMMLSS) |

the radio receiver on the linear scale, while the intrinsic scatter of galaxy properties may be better described on the logarithmic scale. This is non-trivial as the propagation of uncertainties measured from the linear to logarithmic scale must rely on an approximation, which breaks down when the signal to noise ratio is low. If we instead measure the scaling relation in flux space, this issue does not exist. Our approach can also mitigate low number statistics without the binning strategy. We note that this technique can be easily adjusted and applied to measure other H I-scaling relations and the H I mass function (HIMF) directly above or below the detection threshold. We then create mock MIGHTEE-H I galaxies to quantify the H I-selection effect and measure the underlying $M_{H\,\textsc{i}} - M_{\star}$ relation by dividing the mock samples into late-type galaxies (LGTs) and EGTs.

This paper is organized as follows. We describe our MIGHTEE-H I and the ancillary data in Section 2, and the Bayesian technique in Section 3. We present our main results in Section 4, and conclude in Section 5. We use the standard $\Lambda$CDM cosmology with a Hubble constant $H_0 = 67.4$ km s$^{-1}$ Mpc$^{-1}$, total matter density $\Omega_{\mathrm{m}} = 0.315$, and dark energy density $\Omega_{\Lambda} = 0.685$ (Planck Collaboration VI 2020), and AB magnitudes throughout.

## 2 DATA

### 2.1 MIGHTEE-H I

MIGHTEE-H I is the H I emission project within the MIGHTEE survey, and is described in detail by Maddox et al. (2021). The MIGHTEE-H I Early Science data were collected between mid-2018 and mid-2019, in L-band ($900 < \nu < 1670$ MHz), with a spectral resolution of 208 kHz over two well-studied fields: COSMOS and XMMLSS. The visibilities were processed with the IDIA Pipeline:[1] processMeerKAT. This pipeline does full-polarization calibration on MeerKAT data including automated flagging. The spectral line imaging was carried out using the CASA task TCLEAN (robust = 0.5), and the continuum subtraction was undertaken in both the visibilities and imaging planes using the standard CASA routines UVSUB and UVCONTSUB. The effect of direction-dependent artefacts was reduced by a per-pixel median filtering operation. The full data reduction pipeline for MIGHTEE-H I is described by Frank et al. (in preparation). Key parameters of the processed Early Science data are listed in Table 1. We note that this combination of depth and resolution is unique for a blind H I survey.

### 2.2 H I flux

We employ the Cube Analysis and Rendering Tool for Astronomy (Comrie et al. 2021) for visual source finding, then we extract a

---

[1] https://idia-pipelines.github.io

cubelet centred on all detected sources. We smooth the cubelet, and clip it at a $3\sigma$ threshold as a mask for removing the noise, then apply the mask to the cubelet with original resolution. We then clip out by hand any remaining noise peaks and integrate the flux densities over the frequency channels to make moment-0 maps. The total flux is calculated by summing all flux densities over the spatial pixels. We obtain uncertainties on the bright and faint sources varying from 5 per cent to 20 per cent of their H I fluxes roughly (see Ponomareva et al. 2021, for full details).

The H I mass under the optically thin gas assumption can be calculated via

$$M_{\rm H I} = 2.356 \times 10^5 D_L^2 (1+z)^{-1} S, \tag{1}$$

where $M_{\rm H I}$ is the H I mass in solar masses, $D_L$ is the luminosity distance in Mpc, and $S$ is the integrated flux in Jy km s$^{-1}$ (Meyer et al. 2017). We note that our technique works on the H I flux space directly rather than the mass space, and this equation is only needed for our technique to predict the flux $S$ when the H I mass is modelled by the $M_{\rm H I} - M_\star$ relation in Section 3.2.

## 2.3 Ancillary data

All MIGHTEE fields are covered by various multiwavelength photometric and spectroscopic surveys ranging from X-ray to far-infrared bands (e.g. Cuillandre et al. 2012; McCracken et al. 2012; Jarvis et al. 2013; Aihara et al. 2017, 2019). We measure the magnitudes of the sample galaxies by extracting the flux within an elliptical aperture defined in the $g$-band, and we apply this aperture to the $urizYJHK_s$-bands. Based on independent, repeat measurements of several galaxies, the photometry is accurate to $\sim$0.015 mags. We then employ the spectral energy distribution (SED) fitting code LePhare (Ilbert et al. 2006) for deriving the stellar properties of the galaxies, such as stellar mass, stellar age, and SFR, and the uncertainty on the stellar mass is conservatively assumed to be $\sim$0.1 dex, due to assumptions made on galaxy metallicity, star formation history, initial mass function (IMF), and so on in the SED fitting process (Adams et al. 2021; Maddox et al. 2021). In particular, the star formation histories use Bruzual & Charlot (2003) stellar synthesis models including templates with either a constant star formation history or an exponential star formation history. For the exponential star formation history, there are a few different characteristic time-scales for the exponent ($\tau$) ranging from $\tau = [0.1, 0.3, 1, 2, 3, 5, 10, 15, 30]$ Gyrs. For each template, there are also 57 different ages from 0.01 Gyr up to the age of the universe.

Supplemented with such a rich ancillary data set, we are in an excellent position to study H I galaxies from various perspectives, and understand the links between H I gas and other key galaxy properties such as colour, SFR, and stellar mass, in order to gain a complete picture of the diverse galaxy populations as they evolve across the cosmic time.

## 2.4 Morphological classification

We classify our H I detections into four samples based on their optical morphology. In total, we have a sample of 276 H I detections. By removing the objects outside the deep imaging footprint of the Hyper SuprimeCam Subaru Strategic Program (Aihara et al. 2017) and the ones without a classified type, we have a total number of 249 galaxies including 161 spiral galaxies (SP), 64 irregulars (IR), 15 mergers (ME), and nine elliptical galaxies (ET). Details of the galaxy morphology classification are described in Rajohnson et al. (2022). We note that many of the galaxies classified as irregular

are in fact very low mass, and thus could alternatively be classified as dwarf galaxies. We do not differentiate early and late-stage ME due to their small sample size, and label all of them as ME in our analysis.

In Fig. 1, we show the colour- and SFR-stellar mass diagrams colour-coded by their morphologies in the left- and right-hand panels, respectively. We draw the upper boundary of the green valley galaxies from Schawinski et al. (2014) as the grey dashed line on the colour-stellar mass diagram, and ridge lines of the main sequence (MS) of star-forming galaxies from Peng et al. (2010) and Speagle et al. (2014) as grey and blue dashed lines, respectively. These demonstrate that our H I-selected sources are mostly settled in the blue cloud and green valley, and largely distributed above the MS of star-forming galaxies, with very few being red (or passive) galaxies. The irregular and SP dominate at the low and high mass ends, respectively, with considerable overlapping area at the intermediate mass range. This feature motivates us to separate the H I-selected galaxies based on their morphology, and investigate the dominant sample of spirals, in addition to the H I-selected sample as a whole.

It is worth noting that the SFR-$M_\star$ relation predominantly follows a power law at low stellar masses, and flattens at $M_\star \gtrsim 10^{10} \, M_\odot$ as expected for the MS galaxies (e.g. Lee et al. 2015; Schreiber et al. 2015; Saintonge et al. 2016; Leslie et al. 2020), even with our limited sample size from the Early Science data.

# 3 BAYESIAN ANALYSIS

## 3.1 Bayesian framework

Our technique is established on Bayes' theorem

$$\mathcal{P}(\Theta|D, H) = \frac{\mathcal{L}(D|\Theta, H)\Pi(\Theta|H)}{\mathcal{Z}(D|H)}, \tag{2}$$

where $\mathcal{P}$ is the posterior distribution of the model parameters $\Theta$, given the available data $D$ and a model $H$. $\mathcal{L}$ is the likelihood of the data $D$ given parameter values and the model, and $\Pi$ is the prior knowledge of our prejudices about the values of the model parameters. $\mathcal{Z}$ is the Bayesian evidence, which can be thought of as a normalization factor and can be expressed as an integral of $\mathcal{L}$ and $\Pi$ over a $n$-dimensional parameter space $\Theta$,

$$\mathcal{Z}(D|H) = \int \mathcal{L}(D|\Theta, H)\Pi(\Theta|H)d^n\Theta, \tag{3}$$

and in addition it crucially facilitates model selection between different models when their evidences are compared quantitatively, as the evidence is the probability of the data given a model after all the free parameters are marginalized over. The difference in the log-evidence, $\Delta \ln(\mathcal{Z}_B) = \ln(\mathcal{Z}_B) - \ln(\mathcal{Z}_A)$, known as the Bayes factor, is commonly used to interpret how much better Model B is compared to A, providing a fair way of discriminating between models with different numbers of parameters by penalizing models that are too flexible. We follow the criteria in Malefahlo et al. (2021), where $\Delta \ln(\mathcal{Z}) < 1$ is 'not significant', $1 < \Delta \ln(\mathcal{Z}) < 2.5$ is 'significant', $2.5 < \Delta \ln(\mathcal{Z}) < 5$ is 'strong', and $\Delta \ln(\mathcal{Z}) > 5$ is 'decisive'.

We use Multinest (Feroz, Hobson & Bridges 2009; Buchner et al. 2014), an efficient and robust Bayesian inference tool for cosmology and particle physics, to sample the parameter space and explore the full posterior distribution for parameter estimation and the evidence for Bayesian model comparison.
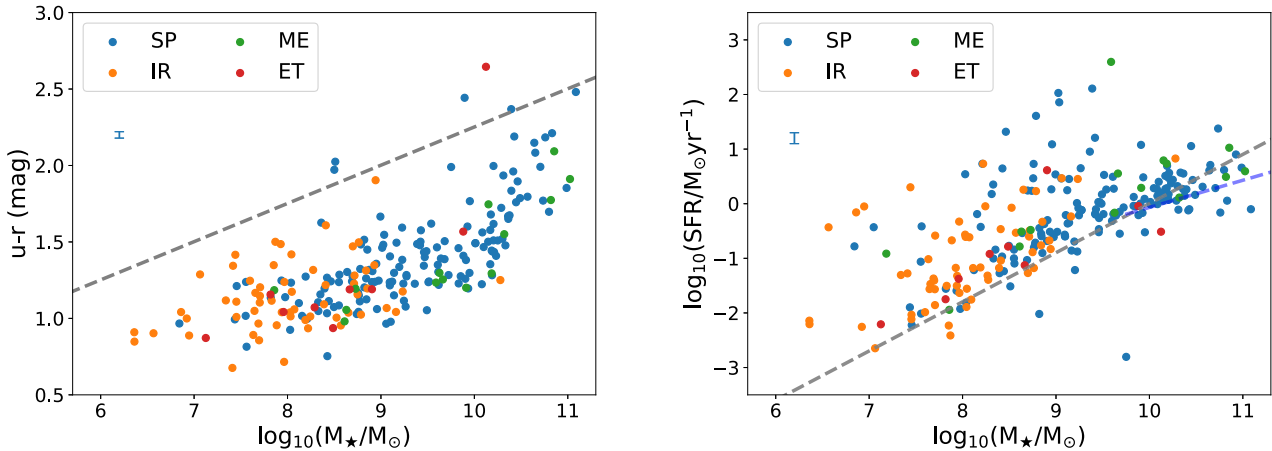
**Figure 1.** Colour (left-hand panel) and SFR (right-hand panel) against the stellar mass. Measurements are colour-coded by the morphologies of galaxies classified as SP, IR, ME, and ET. The grey dashed line in the left-hand panel is the upper boundary of the green valley galaxies from Schawinski et al. (2014), while in the right-hand panel the grey and blue dashed lines show the ridge lines of the MS of star-forming galaxies from Peng et al. (2010) and Speagle et al. (2014), who used the same IMF (Chabrier 2003) and stellar population synthesis models (Bruzual & Charlot 2003), as used in this paper. $1\sigma$ uncertainties on the galaxy $u-r$ colour and SFR are illustrated with blue error bars in the top left-hand corners, respectively.

### 3.2  $M_{HI} - M_\star$ models

We fit two $M_{HI} - M_\star$ models: linear and non-linear models to the data in their logarithmic space, given that both have been used previously (e.g. Maddox et al. 2015; Parkash et al. 2018). First, we model the logarithmic average of $M_{HI}$ as a linear function of $\log_{10}(M_\star)$ as follows:

$$\langle \log_{10}(M_{HI}) \rangle = \alpha[\log_{10}(M_\star) - 10] + \beta, \tag{4}$$

where $\alpha$ and $\beta$ are the free parameters corresponding to the slope, and intercept at $M_\star = 10^{10}\,M_\odot$. We note this single power law relation as '**Model A**'.

For the non-linear relation, we use the double power-law relation:

$$\langle \log_{10}(M_{HI}) \rangle = \log_{10}\left( \frac{M_0}{\left(\frac{M_\star}{M_{tr}}\right)^a + \left(\frac{M_\star}{M_{tr}}\right)^b} \right), \tag{5}$$

where $M_{tr}$, $M_0$, $a$, and $b$ are the free parameters to be fitted for. $M_{tr}$ indicates the transition stellar mass, and $M_0$ is a value along the ordinate at $M_\star = M_{tr}$ where we have $\langle \log_{10}(M_{HI}) \rangle = \log_{10}(M_0/2)$; $a$ and $b$ determine the low- and high-mass slopes of the scaling relation. We denote this double power law relation as '**Model B**', i.e. our non-linear model. When $a = b$, equation (5) is equivalent to equation (4).

### 3.3  Likelihood

The relationship between H I and stellar mass of galaxies cannot be fully described by a single variable function, no matter which model we use. We actually require a bivariate distribution function to capture the whole picture of the $M_{HI} - M_\star$ relation. Without loss of generality, if we assume the Model A or B supplemented with an intrinsic scatter $\sigma_{HI}$ is good enough to describe this relation for our relatively small sample, then the probability of having a H I mass ($M_{HI}$) at a given stellar mass ($M_\star$) follows,

$$P(M_{HI}|M_\star) = \frac{1}{\sqrt{2\pi}\sigma_{HI}} e^{-\frac{1}{2}\left(\frac{\log_{10}(M_{HI}) - \langle \log_{10}(M_{HI}) \rangle}{\sigma_{HI}}\right)^2}. \tag{6}$$

We take the intrinsic scatter $\sigma_{HI}$ as an additional free parameter for our Models A and B. This Gaussian form of distribution function can be replaced with a Schechter function or any other forms if required.

With this conditional H I mass distribution, the probability of having a measured flux, $S_m$, for a single source can be expressed as

$$P(S_m|M_\star) = \int dM_{HI} P(M_{HI}|M_\star) P_n(S_m - S(M_{HI})), \tag{7}$$

where $P_n$ follows the noise distribution of Normal(0, $\sigma_n$), and $S(M_{HI})$ is given by the equation (1).

The likelihood of all the sources having the measured fluxes, given the model and known stellar masses, is given by

$$\mathcal{L} \propto \prod_{\text{source}} P(S_m|M_\star, \text{Model A}(\alpha, \beta, \sigma_{HI}),$$
$$\text{or } |M_\star, \text{Model B}(a, b, \sigma_{HI}, M_0, M_{tr})). \tag{8}$$

By maximizing equation (8), we obtain the best-fitting $M_{HI} - M_\star$ relation with an estimate of the intrinsic scatter for the given sample.

### 3.4  Priors

Priors are our background knowledge of the model parameters, and thus define the sampled parameter space. A uniform prior distribution provides an equal weighting of the input parameter space, and is preferred in general if this prior distribution is not known well. We assign uniform prior probability distributions to $\alpha$, $\beta$, and $\sigma_{HI}$ for our Model A. For Model B, we assign uniform distributions to $a$, $b$, and $\sigma_{HI}$, and adopt uniform logarithmic priors for $M_0$ and $M_{tr}$. All of these priors are listed in Table 2.

## 4  RESULTS

In this section, we investigate the relation between H I mass and stellar mass based on our H I-selected sample with the Bayesian method outlined in Section 3 and mock samples. In Section 4.1, we first use the full sample to maximize the baseline in stellar mass and H I mass for modelling the upper envelope, and then consider the morphologically classified SP as a separate population in order

**Table 2.** Priors of the Models A and B for the H I and stellar mass relation.

| Model | Parameter | Prior probability distribution |
|---|---|---|
| A | $\alpha$ | uniform $\in [-2.5, 2.5]$ |
| | $\beta$ | uniform $\in [7, 12]$ |
| | $\sigma_{\rm H I}$ | uniform $\in [0, 2]$ |
| B | $a$ | uniform $\in [-2, 0]$ |
| | $b$ | uniform $\in [-0.5, 1.5]$ |
| | $\sigma_{\rm H I}$ | uniform $\in [0, 2]$ |
| | $\log_{10}(M_0)$ | uniform $\in [7, 12]$ |
| | $\log_{10}(M_{\rm tr})$ | uniform $\in [7, 12]$ |

to compare with previous studies. In Section 4.2, we create mock samples to model the underlying $M_{\rm H I} - M_\star$ relation with two main galaxy populations based on the measured upper envelope of the $M_{\rm H I} - M_\star$ relation from the MIGHTEE-H I catalogue.

## 4.1 Modelling the upper envelope of the $M_{\rm H I} - M_\star$ relation

### 4.1.1 The whole sample

We show the H I and stellar mass distribution for the complete H I-selected sample in the top panels of Fig. 2. The best-fitting lines for our linear Model A and non-linear Model B are shown as the blue lines in the left- and right-hand panels, respectively. The $1\sigma$ statistical scatter, predominantly due to the H I flux uncertainties and our limited sample size are denoted by blue shaded areas, while the total (statistical plus intrinsic) scatter in the H I mass distribution around the stellar mass are shown by green shaded regions. We find that the non-linear model is decisively preferred over the linear model with a Bayes factor of $\Delta \ln(\mathcal{Z}_\mathcal{B}) = 6.16^{+0.07}_{-0.07}$ for the full sample at $0 < z < 0.84$, and list the best-fitting parameters in Table 3.

The agreement between the full MIGHTEE-H I sample and the spectroscopic ALFALFA–SDSS galaxy sample of Maddox et al. (2015) is excellent for our non-linear Model B, with most parts of the Maddox et al. (2015) relation (denoted by grey dashed line) falling within the $1\sigma$ statistical uncertainties of our data (blue shaded area). Compared to ALFALFA, the deficit at the high H I mass ($M_{\rm H I} \gtrsim 10^{10} \, {\rm M}_\odot$) end seen in both panels suggests we detected fewer H I galaxies at these masses in the MIGHTEE-H I Early Science data. This is due to the limited volume surveyed thus far, which precluded us from finding the rarer, high H I–mass systems in the current area, and we will require the full MIGHTEE survey, where the survey volume will reach 20 deg$^2$, to fully explore this region. Our results are in excellent agreement with the SIMBA simulation (Davé et al. 2019), where we include the MS galaxies defined as specific SFR (sSFR) $> 10^{-1.8 + 0.3z}$ Gyr$^{-1}$ on top of the H I selection that is identical to the MIGHTEE-H I selection (blue dashed circles). This sSFR selection cuts out many red galaxies, and the H I-selected SIMBA sample shows no obvious systematic difference in typical sSFR compared to the MIGHTEE-H I-selected sample. However, we do find a deviation of the H I masses between the H I-selected MIGHTEE sample and the H I-selected SIMBA sample without excluding red galaxies (red dashed circles). This deviation suggests that SIMBA overestimates the amount of H I gas in the massive dead red galaxies as these galaxies seem to have a moderate amount of H I gas and would have been detected by MIGHTEE-H I, thus weighting down the average H I mass at the massive end if they were present. The statistical significance of this difference is, however quite low, and would need to be investigated with a larger sample. Given that the H I column

density is typically very low in red galaxies, another possibility is that our MIGHTEE-H I Early Science observation could miss some of those objects.

Compared to the measurement from Maddox et al. (2015), Model A would suggest an excess of H I-rich galaxies at the low-mass end, but this is due to the model being a poor description of the data and this excess disappears when we use a more flexible non-linear model to fit for the data (in the right-hand panel). The low-number statistics also plays a role at the low mass end as the statistical error, indicated by the blue area, increases. However, we note that the intrinsic scatter in the relation still dominates. The global difference between our linear Model A and the binned median H I masses in Maddox et al. (2015) also suggests the limitation of a simple linear modelling, due to the complex nature of the upper bound in the $M_{\rm H I} - M_\star$ relation. None the less, Model A is consistent with the H I-selected sample in Parkash et al. (2018) at the high-mass end but a higher detection of rate of H I galaxies at the low-mass end suggests that the H I-selected sample in Parkash et al. (2018) is less complete at $M_{\rm H I} < 10^9 \, {\rm M}_\odot$.

For Model B, we find a transition stellar mass of $\log_{10}(M_\star/{\rm M}_\odot) = 9.15 \pm 0.87$, which breaks the upper envelope of the $M_{\rm H I} - M_\star$ relation into two regions. At the high mass end, the measured slope (indicated by the parameter $b$, is much flatter than that at the low mass end (indicated by the parameter $a$). This finding is consistent with the steeper slope of Model A measured from the full H I sample with respect to the spiral-only galaxies, which tend to be massive systems (see also Section 4.1.2), and is also in line with the upper envelope of H I mass fraction found by Maddox et al. (2015). Thus, we confirm that the H I gas fraction decreases as a function of stellar mass at $M_\star \gtrsim 10^9 \, {\rm M}_\odot$ with the MIGHTEE-H I Early Science data. This trend is similar to the galaxy MS, where the SFR-$M_\star$ relation is linear up to a critical mass of $\sim 3 \times 10^{10} \, {\rm M}_\odot$, and then flattens out towards higher masses (e.g. Erfanianfar et al. 2015). It is also interesting to note that a similar curvature has been suggested for the baryonic specific angular momentum–baryonic mass relation with the slope change occurring at $\sim 10^9 \, {\rm M}_\odot$ (e.g. Kurapati et al. 2018; Kurapati, Chengalur & Verheijen 2021), albeit whether this break is real is still debated (Mancera Piña et al. 2021a, b).

The triangular and diamond symbols in Fig. 2 denote ME and ET, respectively. We see the ET predominantly lie below the model fits, while the ME are randomly distributed around the best-fitting models. Thus, it shows that a lower fraction of H I gas is detected in ETs compared to other types of galaxies from the H I-selected sample as we might expect. However, we do not draw strong conclusions about this given their small number in our sample.

We note that our H I-selected sample is flux-limited, and thus exhibits a selection bias against galaxies with low H I masses, which becomes more severe going to higher redshift as seen from the colour-coded symbols. We return to this in Section 4.2.

In Fig. 3, we show the H I mass as a function of SFR. We first replace the stellar mass with SFR in our models, then fit the Model A (dashed blue line) as $\langle \log_{10}(M_{\rm H I}) \rangle = 0.4 \log_{10}({\rm SFR}) + 9.44$ and the Model B (solid blue line), and observe a moderate flattening feature at the high SFR end with a measured transition SFR of $\log_{10}({\rm SFR/M}_\odot {\rm yr}^{-1}) = 0.79 \pm 0.53$, over which the statistical uncertainties are large due to only a few highest-SFR/bursty spirals. We also find that the majority of our H I-selected galaxies are able to support their star formation activity given a sufficient H I fuel supply, with the H I depletion times, $M_{\rm H I}/{\rm SFR}$, varying in the range of 1–100 Gyr. This suggests that the correlation between SFR and the H I mass is consistent with being almost linear across the
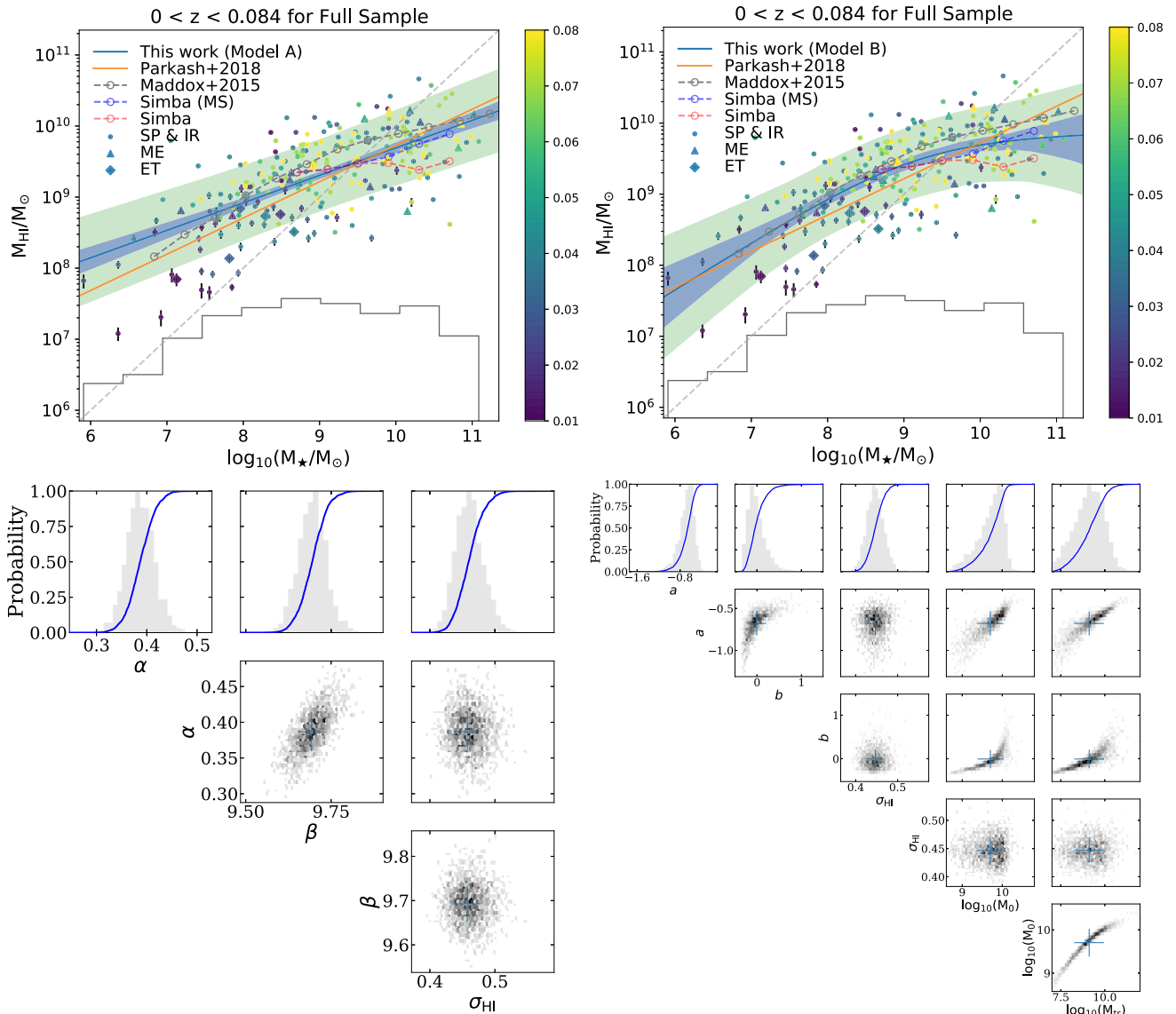
**Figure 2.** Observed $M_{\rm HI} - M_\star$ relation (top) and its posterior parameters (bottom) with the best-fitting Model A (left-hand panel) and B (right-hand panel) at $0 < z < 0.084$ for the full H I selected sample from MIGHTEE-H I. Top row: the dots are spiral and irregular galaxies while triangles and diamonds correspond to ME and ET. They are colour-coded by redshift. The blue shaded areas are the statistical uncertainties, and the green ones are the intrinsic scatters added to the statistical uncertainties. The orange line is the H I-selected sample from Parkash et al. (2018). The grey dashed circles are the measurements from ALFALFA (Maddox et al. 2015). The blue dashed circles are the MS galaxies from the SIMBA simulation (Davé et al. 2019) while the red ones are their full H I samples. The diagonal light grey dashed line is the one-to-one relation. The black line at the bottom of the top panels indicates the normalized distribution of stellar mass. Bottom row: the grey histograms are the (one- or two-dimensional) marginal posterior probabilities. The blue curves are the cumulative distributions. The blue crosses in the two-dimensional posteriors are the set of parameters with the maximum likelihood, and the $1\sigma$ error bars error are estimated in the one-dimensional marginal posterior space.

entire H I mass range on the logarithmic scale, and the shortage of H I gas is likely ultimately responsible for the decreasing SFR towards the higher stellar masses, although we notice a slightly larger intrinsic scatter of $\sim 0.48$ dex for this relation compared to the $\sim 0.44$ dex for the $M_{\rm HI} - M_\star$ relation. The lower turnover mass of $\log_{10}(M_\star/M_\odot) = 9.15 \pm 0.87$ for the $M_{\rm HI} - M_\star$ relation against $\log_{10}(M_\star/M_\odot) \sim 10$ for the SFR-$M_\star$ relation signifies that the loss of H I gas supply at high masses may not immediately reflected on the quenching of star formation, albeit with large statistical uncertainties.

### 4.1.2 SP

In this section, we consider the population of morphologically classified SP. In Fig. 4, we show the observed $M_{\rm HI} - M_\star$ relation and the posterior parameters for the H I-selected SP from the MIGHTEE-H I catalogue at $0 < z < 0.084$, with our best-fitting Models A and B in the left- and right-hand top panels, respectively. Based on the best fits in Table 3 and the returned Bayesian evidences, we see that the data are much less in favour of the non-linear model over the linear one with $\Delta \ln(\mathcal{Z}_{\mathcal{B}}) = 1.44 \pm 0.04$, which is significant but not decisive or strong. We also find that the posterior distributions of

**Table 3.** The best-fitting parameters of the observed $M_{HI} - M_\star$ relation with Models A and B for our MIGHTEE-H I-selected samples at $0 < z < 0.84$. The values listed are those with the maximum likelihood from our fitting.

| Model | Parameter | Spirals | Full sample |
|-------|-----------|---------|-------------|
| A | $\alpha$ | $0.278 \pm 0.040$ | $0.387 \pm 0.027$ |
|   | $\beta$ | $9.649 \pm 0.043$ | $9.693 \pm 0.039$ |
|   | $\sigma_{HI}$ | $0.44 \pm 0.02$ | $0.46 \pm 0.02$ |
| B | $a$ | $-0.523 \pm 0.164$ | $-0.672 \pm 0.133$ |
|   | $b$ | $0.022 \pm 0.429$ | $-0.035 \pm 0.194$ |
|   | $\sigma_{HI}$ | $0.440 \pm 0.024$ | $0.435 \pm 0.020$ |
|   | $\log_{10}(M_0)$ | $9.869 \pm 0.365$ | $9.771 \pm 0.327$ |
|   | $\log_{10}(M_{tr})$ | $9.52 \pm 1.56$ | $9.15 \pm 0.87$ |
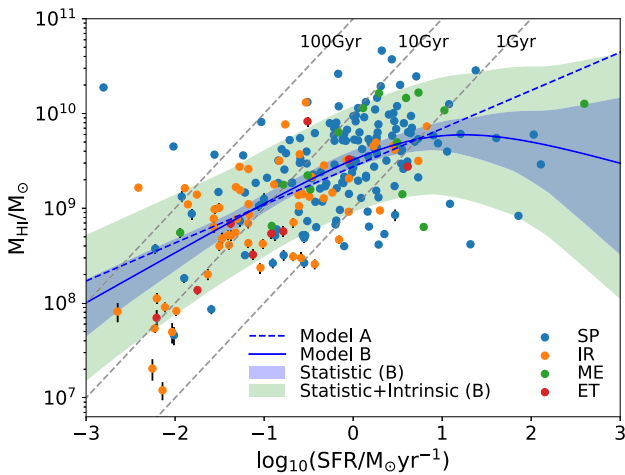


**Figure 3.** $M_{HI}$ as a function of the SFR for the full H I-selected sample from the MIGHTEE-H I catalogue at $0 < z < 0.084$. Measurements in dots are colour-coded by their morphologies. The dashed and solid blue lines are the best-fitting of Models A and B, respectively. The blue and green shaded areas are the statistical uncertainties and intrinsic scatters added to the statistical uncertainties for the Model B. The dashed grey lines are the time scales for depletion of the H I gas, defined as $M_{HI}$/SFR.

Model B for spirals are not as well-converged as for the full sample in Fig. 2.

For both models, we find a systematically higher detection of H I gas than what Parkash et al. (2018) found at $M_\star \gtrsim 10^9 \, M_\odot$ from a stellar mass-selected spirals. This is likely to be the result of different selection effects, with the Parkash et al. (2018) spiral galaxy sample being $M_\star$-selected and the MIGHTEE-H I sample being H I-selected. The latter tends to be populated by higher H I mass objects at any given stellar mass. It implies that there still exists a significant fraction of H I-poor SP to be picked up by a deeper H I survey. We measure an intrinsic scatter of $0.44 \pm 0.03$ dex for both models, which is roughly consistent with the 0.4 dex obtained from the stellar mass-selected spirals in Parkash et al. (2018).

To compare with the SIMBA spirals, we select galaxies in SIMBA with fraction of kinetic energy ($\kappa_{rot}$) $> 0.7$ (Sales et al. 2012), and denote their median H I masses against the stellar masses as red dashed circles. We then use the same criterion of sSFR $>10^{-1.8 + 0.3z}$ Gyr$^{-1}$ to exclude the red SP, and show their $M_{HI} - M_\star$ relation as blue dashed circles. Overall, we find good agreement between MIGHTEE-H I and SIMBA MS spirals for the $M_{HI} - M_\star$ relation, and notice a lower detection of the average H I mass for the whole SIMBA spiral sample at the most massive end. This trend is similar

to what we found in Section 4.1.1 for the full MIGHTEE-H I sample, and indicates that there are probably too many red SP that have non-negligible amount of H I gas in SIMBA.

The best-fitting transition stellar mass for our Model B for the H I-selected spirals is $\log_{10}(M_\star/M_\odot) = 9.52 \pm 1.56$, which is higher than the $\log_{10}(M_\star/M_\odot) = 9.15 \pm 0.87$ for the full sample, but has much larger statistical uncertainties due to reduced number of sources and a narrower range of $M_\star$. This trend roughly corresponds to the difference of the best-fitting slopes using Model A between these two samples, where the spirals have an obvious shallower slope of 0.278 compared to the 0.387 for the full sample, although the corresponding intrinsic scatters are similar.

The distinction of our best-fitting models between spirals and the full H I-selected sample is a strong indication of very different gas processes between the spiral and lower-mass irregular galaxies, since the IR dominate over other types of galaxies except the spirals in our catalogue. Indeed, given the stability model for disc galaxies in Obreschkow et al. (2016), the halo spin parameter for the SP can limit the maximum H I gas supply, and there seems to be no such a limitation for the lower-mass galaxies as their discs become unstable when the velocity dispersion reaches similar to rotation velocity. However, we cannot distinguish whether the different slopes are due to galaxy mass or morphology with the current sample, and this should be better explored with the full MIGHTEE survey. Stacking on the SP to lower stellar mass and other types of galaxies at higher stellar mass will also help to further clarify this difference.

### 4.2 Modelling the underlying $M_{HI} - M_\star$ relation

Thus far, we have discussed the $M_{HI} - M_\star$ relation based on the H I-selected sample from MIGHTEE-H I Early Science data. The H I selection essentially means that we are exploring the upper envelope in H I mass and therefore constraining the maximum amount of H I as a function of stellar mass. To fully assess how the H I selection influences our results, in this section we create mock samples of all the galaxies above and below the MIGHTEE-H I detection threshold with two main galaxy populations: LTGs and ETGs. This also allows us to check how the limited volume and the flux-density limit selection of our sample may influence the results presented in the previous section when determining whether there is any evidence for a transition in the upper envelope of the stellar-mass to H I-mass relation.

#### 4.2.1 The whole sample

We create mock MIGHTEE-H I galaxies with and without a break in the underlying $M_{HI} - M_\star$ relation when considering the LTGs and ETGs as our whole sample, as shown in Fig. 5. We first employ the galaxy stellar mass function (GSMF) from Driver et al. (2022) and generate galaxy samples across the redshift range of the MIGHTEE-H I observations, but with 10 times the survey area to reduce the random error in this process. We then find the best-fitting parameters of the underlying models for the $M_{HI} - M_\star$ relation such that our mock H I distribution matches our modelled upper envelope described by the average H I mass $\langle \log_{10}(M_{HI}) \rangle$ and the global intrinsic scatter $\sigma_{HI}$, above the H I detection threshold in each MIGHTEE-H I field. Considering the large mock survey area and the fact that no measurement uncertainty is introduced for mock data, we first bin the sample in $M_\star$ and obtain the best-fitting parameters by minimizing the difference between observed $\langle \log_{10}(M_{HI}) \rangle$ from our Model B for MIGHTEE-H I (derived in Table 3) and for our mock sample as a function of $M_\star$, along with their $\sigma_{HI}$, based
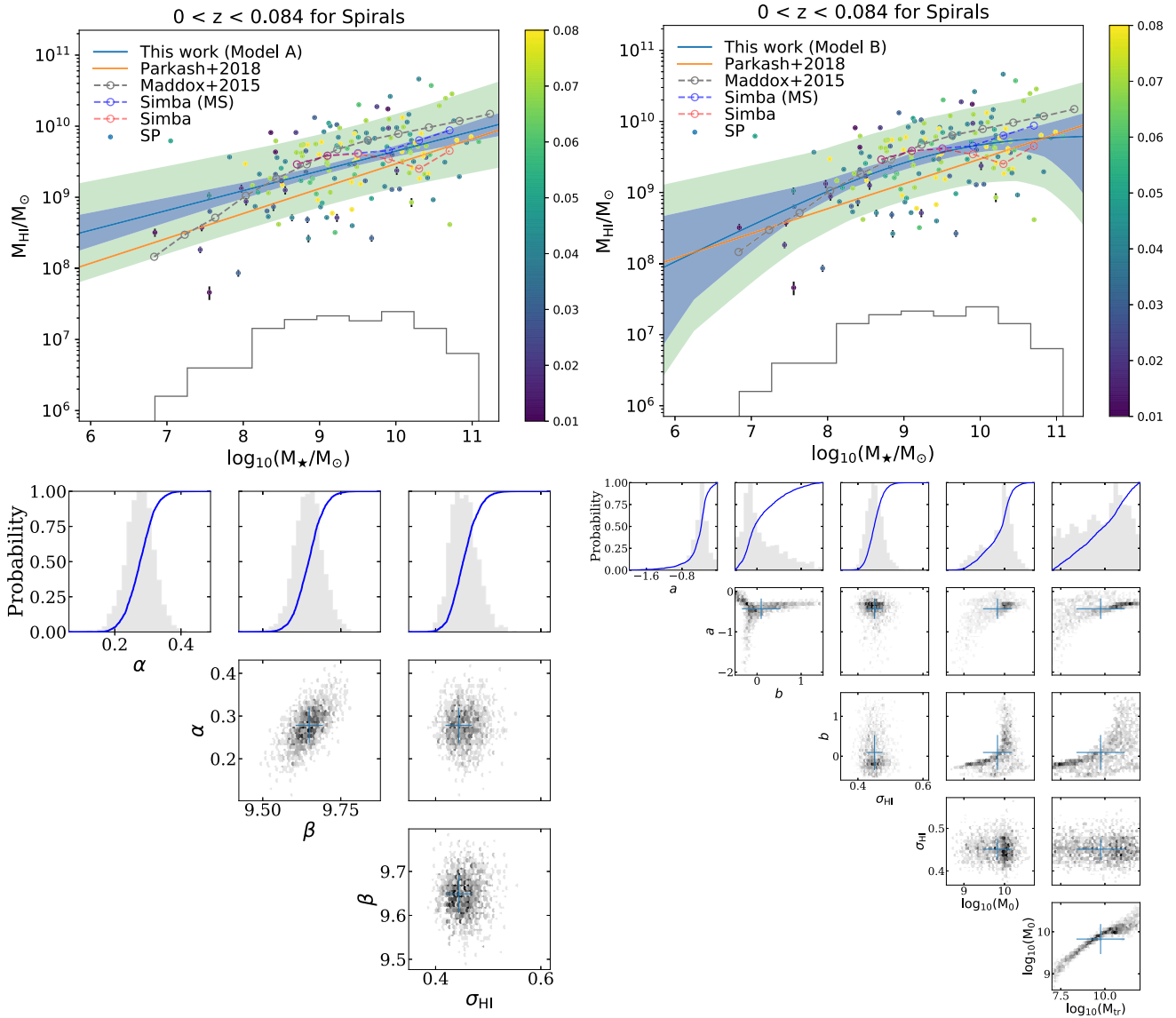
**Figure 4.** Observed $M_{HI} - M_{\star}$ relation (top) and its posterior parameters (bottom) for SP with the best-fitting Model A (left-hand panel) and B (right-hand panel) at $0 < z < 0.084$ from the MIGHTEE-H I catalogue. Top row: the dots are SP colour-coded by redshift. The blue shaded areas are the statistical uncertainties, and the green ones are the intrinsic scatters added to the statistical uncertainties. The orange line is the stellar mass-selected spirals from Parkash et al. (2018). The grey dashed circles are the measurements from ALFALFA (Maddox et al. 2015). Bottom row: the grey histograms are the (one- or two-dimensional) marginal posterior probabilities. The blue curves are the cumulative distributions. The blue crosses in the two-dimensional posteriors are the set of parameters with the maximum likelihood, and the $1\sigma$ error bars are estimated in the one-dimensional marginal posterior space.

on the same Bayesian framework. In general, the fitting for a two-dimensional distribution to another one requires binning the data in the two-dimensional space. However, our MIGHTEE-H I sample size is relatively small and the distribution of $M_{HI}$ against $M_{\star}$ can be well-described by a scaling relation and an associated scatter as shown in the previous section, therefore such a fitting in our case can be achieved by constraining the $\langle \log_{10}(M_{HI}) \rangle$ and $\sigma_{HI}$, which are the key elements that delineate a two-dimensional distribution in a simple way. We list the best-fitting parameters in Table 4.

In Fig. 5, we show the best-fitting results in blue lines with our Models A and B in the left- and right-hand panels, respectively. The filled and empty symbols represent the 'detections' and non-detections, respectively. Although it appears that both models can

mimic a broken $M_{HI} - M_{\star}$ relation for our MIGHTEE-H I observation, the relative evidence between Models A and B for fitting the underlying $M_{HI} - M_{\star}$ relation is $\ln(\mathcal{Z}_{\mathcal{B}})$ - $\ln(\mathcal{Z}_{\mathcal{A}}) = 3.9 \pm 0.3$, which strongly implies a non-linear underlying $M_{HI} - M_{\star}$ relation over the linear one. The black circles are the observed $\langle \log_{10}(M_{HI}) \rangle$ for our mock samples with the underlying Models A and B in the left- and right-hand panels, respectively. It is clear that the 'observed' broken relation of the Model B mock sample agrees with the data (the dashed black line) better than the 'observed' broken relation of the Model A mock sample especially at around the transition mass range. The non-linear model also demonstrates a better agreement than the linear model for the averaged H I mass with Parkash et al. (2018), based on their $M_{\star}$-selected sample.
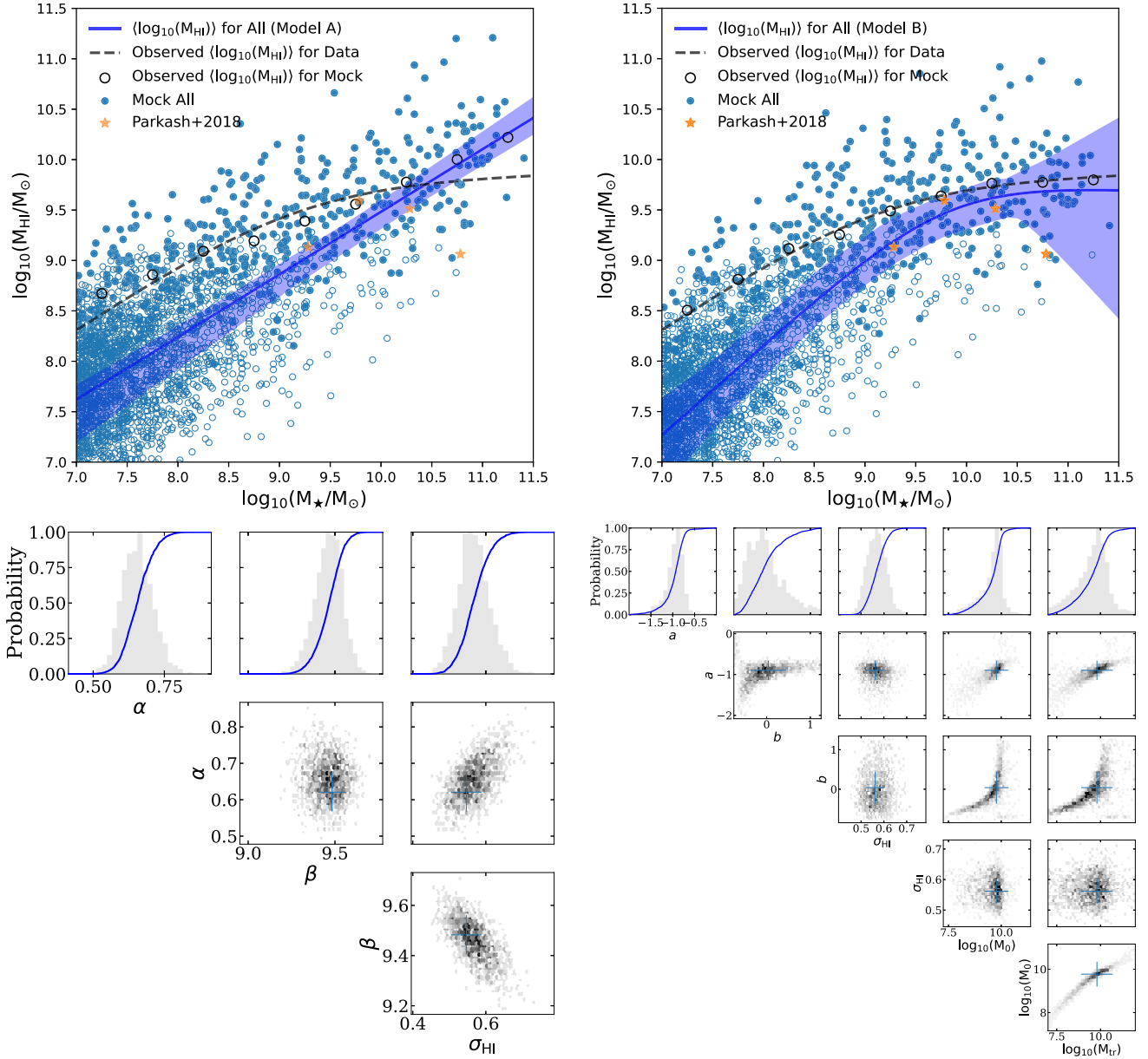
**Figure 5.** Top row: $M_{HI}$ as a function of the stellar mass for the whole mock MIGHTEE-H I sample at $0 < z < 0.084$, with Model A (left-hand panel) and B (right-hand panel) fitted for the underlying $M_{HI} - M_\star$ relation. The blue dots with the filled and empty ones are the detections and non-detections. The dashed black lines in the left- and right-hand panels are the observed $\langle \log_{10}(M_{HI}) \rangle$ from Model B for our MIGHTEE-H I observation while the black circles are 'observed' $\langle \log_{10}(M_{HI}) \rangle$ for our mock sample. The solid blue lines are the best-fitting of Models A and B in the left- and right-hand panels for the whole underlying sample with the blue shaded areas being the $1\sigma$ statistical uncertainties. The orange stars are the stellar mass-selected sample from Parkash et al. (2018). The blue dots are down-sampled for presentation. Bottom row: the grey histograms are the (one- or two-dimensional) marginal posterior probabilities. The blue curves are the cumulative distributions. The blue crosses in the two-dimensional posteriors are the set of parameters with the maximum likelihood, and the $1\sigma$ error bars are estimated in the one-dimensional marginal posterior space.

We also perform a two-dimensional Kolmogrov–Smirnov (KS) test[2] (Peacock 1983; Fasano & Franceschini 1987) to check consistency of distributions for the mock 'detections' from our underlying models and the MIGHTEE-H I detections. If the KS statistic is large, the $p$-value will be small, and it can be taken as evidence against the null hypothesis that the two distributions are identical.

We accept a confidence level of 95 per cent, which means that we reject the null hypothesis if the $p$-value is less than 0.05. We find that the $p$-value is 0.129 for the mock detections from our Model B against our MIGHTEE-H I data, and the $p$-value is 0.007 for the mock detections from the underlying Model A. These $p$-values suggests that the non-linear underlying model is doing a better job in mocking the MIGHTEE-H I detections than the linear one, which is consistent with what we see from the Bayesian evidences.

[2]https://github.com/syrte/ndtest

**Table 4.** The best-fitting parameters of the underlying $M_{\rm HI} - M_\star$ relation with Models A and B for our mock MIGHTEE-H I samples at $0 < z < 0.84$. The values listed are those with the maximum likelihood from our fitting.

| Model | Parameter | The whole sample | LGTs |
|-------|-----------|------------------|------|
| A | $\alpha$ | $0.621 \pm 0.053$ | $0.678 \pm 0.052$ |
| | $\beta$ | $9.48 \pm 0.075$ | $9.54 \pm 0.08$ |
| | $\sigma_{\rm HI}$ | $0.55 \pm 0.04$ | $0.56 \pm 0.04$ |
| | $a$ | $-0.896 \pm 0.188$ | $-0.941 \pm 0.203$ |
| | $b$ | $0.040 \pm 0.404$ | $-0.001 \pm 0.419$ |
| B | $\sigma_{\rm HI}$ | $0.562 \pm 0.042$ | $0.56 \pm 0.042$ |
| | $\log_{10}(M_0)$ | $9.77 \pm 0.51$ | $9.74 \pm 0.62$ |
| | $\log_{10}(M_{\rm tr})$ | $9.8 \pm 0.88$ | $9.69 \pm 1.03$ |

### 4.2.2 LGTs

We now divide the whole galaxy sample into LTGs and ETGs based on their known relative fractions as a function of stellar mass (Rodríguez-Puebla et al. 2020), and we assign the H I masses with a double power law for ETGs (Calette et al. 2018). The LTGs dominate the sample in our mass range of interest ($7 < \log_{10}(M_\star/\rm M_\odot) < 11$), and we only consider tuning our Models A and B for LTGs in order to find the best-fitting parameters that make our mock H I distribution satisfy the observed $\langle \log_{10}(M_{\rm HI}) \rangle$ and the global intrinsic scatter $\sigma_{\rm HI}$. We also list the best-fitting parameters of the intrinsic $M_{\rm HI} - M_\star$ relation for LTGs in Table 4.

In Fig. 6, we show the mock galaxy samples, along with the best-fitting intrinsic $M_{\rm HI} - M_\star$ relation (solid blue line) for LTGs with our Models A and B in the left- and right-hand panels. The observed $M_{\rm HI} - M_\star$ relation with Model B from our MIGHTEE-H I observation at $0 < z < 0.084$ are shown by dashed black lines in both panels. The blue and red dots are LTGs and ETGs, with the filled and empty symbols representing the 'detections' and non-detections. The black dots are the average of $\log_{10}(M_{\rm HI})$ with the error bar being the intrinsic scatter for all samples at a given stellar mass bin (also shown in Table 5), and are in good agreement with the median of $\log_{10}(M_{\rm HI})$ for the stellar mass-selected sample in Parkash et al. (2018). Compared to Brown et al. (2015) (grey stars), our measured average H I masses appear to be systematically lower, which actually is not surprising as we are using the logarithmic average of H I masses $\langle \log_{10}(M_{\rm HI}) \rangle$ against the arithmetic average of H I masses $\log_{10}(\langle M_{\rm HI} \rangle)$ in Brown et al. (2015). We find that the H I selection lifts up the 'observed' $M_{\rm HI} - M_\star$ relation but the turnover feature on this relation persists to a large degree, albeit with a weaker break on the observed $M_{\rm HI} - M_\star$ relation. In other words, the H I sample selection has a stronger impact on the lower mass end for the $M_{\rm HI} - M_\star$ relation as the dwarf galaxies are most sensitive to our detection floor, and hence the underlying average H I masses have a sharper bend than the observed average H I masses as a function of the stellar mass. Although the ETGs can downweight the average H I mass at the high mass end, it is obvious that the LTGs alone show an intrinsic turnover on the $M_{\rm HI} - M_\star$ relation as indicated by the solid blue line.

By marginalizing over the stellar mass in Fig. 6, we can also obtain a best-fitting HIMF shown by the black line in the bottom right-hand panel. The HIMF constructed by our approach is in good agreement with Jones et al. (2018) and Butcher et al. (2018) across a wide range of H I masses, except for the highest mass end ($\log_{10}(M_{\rm HI}/\rm M_\odot) \gtrsim$ 10) due to the unsophisticated modelling (e.g. the assumption of a symmetric H I distribution at a given stellar mass for each population) for the bivariate H I-stellar mass distribution. We refer readers to a more detailed approach to measuring the first MeerKAT HIMF by

Ponomareva et al. (2023). None the less, the number of the most H I-massive galaxies is several orders of magnitude smaller than that of the less H I-massive galaxies, therefore their impact on the logarithmic average of H I masses is limited.

To demonstrate that the break measured in this paper is not an artefact of the H I selection, we also build mock MIGHTEE-H I samples with no break in the underlying $M_{\rm HI} - M_\star$ relation for LTGs, which is described by a single power law (i.e. our Model A). The mock samples are shown in the left-hand panel of Fig. 6. The solid blue line is the best-fitting of Model A for LTGs with the blue shaded areas being the $1\sigma$ statistical uncertainties. By comparing the black circles with the black dashed lines, we find a similar result as in the previous section that the 'observed' broken relation of the Model B mock LTGs with ETGs in the right-hand panel agrees with the data better than that of the Model A mock LTGs with ETGs in the left-hand panel especially at around the transition mass range. The relative evidence between Models A and B for fitting the underlying $M_{\rm HI} - M_\star$ relation is $\ln(\mathcal{Z}_{\mathcal{B}}) - \ln(\mathcal{Z}_{\mathcal{A}}) = 4.4 \pm 0.4$, which is slightly larger than the Bayes factor of $3.9 \pm 0.3$ when the LTGs and ETGs are considered as a whole. In other words, our Bayesian analysis suggests that modelling the LTGs with a break $M_{\rm HI} - M_\star$ relation is favoured by our MIGHTEE-H I observations. We note that our analysis shows that the actual position of the break and the slope of the relation below and above the break are challenging to constrain, as can be inferred from the error bars in Table 4. The full MIGHTEE survey, and a combination of MIGHTEE with other H I surveys, such as Looking At the Distant Universe with the MeerKAT Array (Blyth et al. 2016) will provide much stronger constraints.

We also note that there are other selection effects, such as the limited volume meaning we are susceptible to different environments, which may impact on our measurement of the underlying $M_{\rm HI} - M_\star$ relation. However, these effects are likely to be sub-dominant with respect to the flux-limited nature in our sample, and hard to be quantified without the help of large numerical simulations including dwarf H I galaxies with their masses down to $\log_{10}(M_{\rm HI}/\rm M_\odot) \sim$ 7. We also cannot create a mock spiral galaxy population with our approach to assess their intrinsic $M_{\rm HI} - M_\star$ relation due to the uncertain correlation between their morphology, stellar, and gas components. We look forward to seeing these aspects in the future hydro-dynamical and semi-analytic galaxy simulations.

## 5 CONCLUSIONS

We have developed a Bayesian technique that allows us to measure the $M_{\rm HI} - M_\star$ relation above or below the detection threshold in a unified way while taking into account its intrinsic scatter without binning the data sets. We implement this technique with the MIGHTEE-H I Early Science data, and highlight our main results as follows:

(i) We model the upper envelope of the $M_{\rm HI} - M_\star$ relation down to $M_{\rm HI} \sim 10^7 \, \rm M_\odot$, and up to z = 0.084 using a H I-selected sample of 249 galaxies. We use a double power-law model to fit our data, and find this non-linear model is preferred by the data over the linear model, with a transition stellar mass of $\log_{10}(M_\star/\rm M_\odot) =$ $9.15 \pm 0.87$, which roughly corresponds to the break in the stellar mass of $M_\star \sim 10^9 \, \rm M_\odot$ found by Maddox et al. (2015). Beyond this transition (or break) stellar mass, the slope of $M_{\rm HI} - M_\star$ relation flattens.

(ii) We also examine the corresponding SFR–$M_{\rm HI}$ relation and find that it is almost linear across the whole H I mass range, albeit
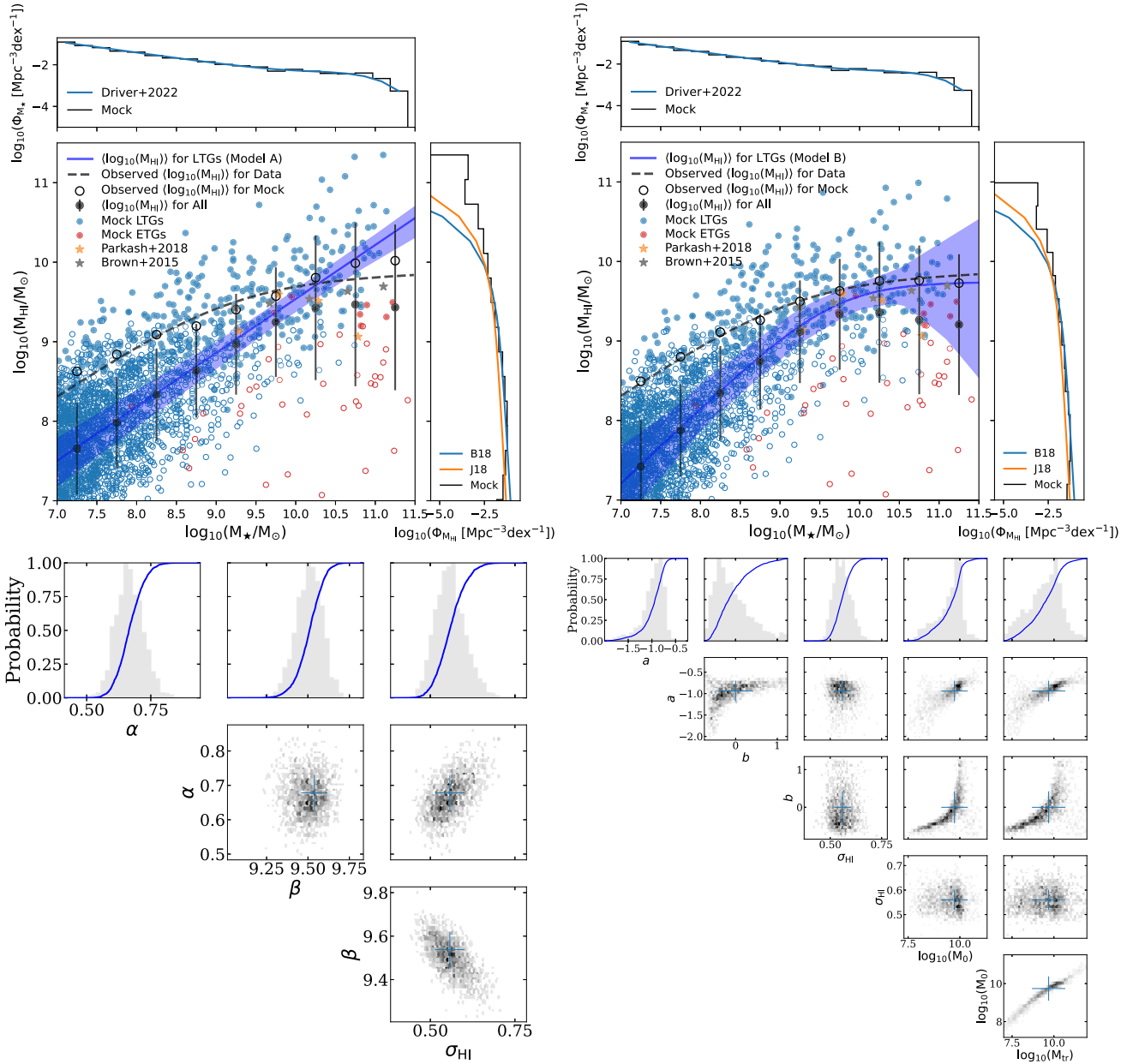
**Figure 6.** Top row: $M_{HI}$ as a function of the stellar mass for the mock MIGHTEE-H I samples for LTGs and ETGs at $0 < z < 0.084$, with Model A (left-hand panel) and B (right-hand panel) fitted for the underlying $M_{HI} - M_\star$ relation for LTGs. The blue and red dots are LTGs and ETGs, with the filled and empty ones being the detections and non-detections. The dashed black lines in the left- and right-hand panels are the observed $\langle \log_{10}(M_{HI}) \rangle$ from Model B for our MIGHTEE-H I observation while the black circles are 'observed' $\langle \log_{10}(M_{HI}) \rangle$ for our mock sample. The solid blue lines are the best-fitting of Models A and B in the left- and right-hand panels for LTGs with the blue shaded areas being the $1\sigma$ statistical uncertainties. The black dots are the average of $\log_{10}(M_{HI})$ with the error bar being the intrinsic scatter for all samples at given stellar mass bins. The orange and grey stars are the stellar mass-selected samples from Parkash et al. (2018) and Brown et al. (2015). The upper panel shows the GSMF from Driver et al. (2022) that we use to mock our galaxy samples, while the bottom right-hand panel shows the HIMF, which is the result of marginalizing over the $M_\star$ axis. The HIMFs in Jones et al. (2018) for the ALFALFA survey and Butcher et al. (2018) for the NIBLES survey are shown in blue and orange lines. The colour-coded dots are down-sampled for presentation. Bottom row: the grey histograms are the (one- or two-dimensional) marginal posterior probabilities. The blue curves are the cumulative distributions. The blue crosses in the two-dimensional posteriors are the set of parameters with the maximum likelihood, and the $1\sigma$ error bars are estimated in the one-dimensional marginal posterior space.

with a large scatter of ∼0.48 dex. Combined with the flattening feature on the $M_{HI} - M_\star$ relation, this supports the hypothesis that the shortage of H I gas supply is likely ultimately responsible for the quenching of the star formation activity observed in massive MS galaxies.

(iii) By separating our full sample into spirals, IR, ME, and ETs, we find the H I sample is dominated by the spirals at the high mass end, and by the IR at the low mass end. These two type of galaxies exhibit significantly different slopes for the $M_{HI} - M_\star$ relation, and are likely to be responsible for the detected transition

**Table 5.** Logarithmic average of $M_{H\,I}$ and its intrinsic scatter as a function of stellar mass for the mock MIGHTEE-H I LTGs and ETGs including detections and non-detections.

| $\log_{10}(M_\star)$ | $0.0 < z < 0.084$ | |
|---|---|---|
| | $\langle\log_{10}(M_{H\,I})\rangle$ | $\sigma_{H\,I}(M_\star)$ |
| 7.25 | 7.42 | 0.58 |
| 7.75 | 7.88 | 0.58 |
| 8.25 | 8.35 | 0.59 |
| 8.75 | 8.74 | 0.6 |
| 9.25 | 9.12 | 0.65 |
| 9.75 | 9.33 | 0.7 |
| 10.25 | 9.36 | 0.89 |
| 10.75 | 9.27 | 0.93 |
| 11.25 | 9.21 | 0.89 |

stellar mass from the full sample, although we cannot rule out a pure mass dependence. In addition, we find that the ETs show a lower fraction of H I mass than other types from the H I-selected sample, and the highest mass galaxies show a higher fraction of H I mass than predicted by hydrodynamic simulations (Davé et al. 2019), although small number statistics prohibits a strong statement about the H I characteristics of ET and the most massive ones.

(iv) We created mock galaxies above and below the MIGHTEE-H I detection threshold with two broad galaxy populations of LTGs and ETGs to measure the underlying $M_{H\,I} - M_\star$ relation over the last billion years. We find that the H I selection can lift the 'observed' $M_{H\,I} - M_\star$ relation on average but the turnover feature on this relation is largely immune to this effect, albeit with a weaker break, regardless of whether the two galaxy populations are taken as a whole or separately in their intrinsic bivariate distribution of H I and stellar masses.

(v) We fit a linear underlying $M_{H\,I} - M_\star$ scaling relation (i.e. Model A) to the observed relation from our MIGHTEE-H I observation in addition to the non-linear underlying relation (i.e. Model B). Although both models can mimic a broken $M_{H\,I} - M_\star$ relation for our MIGHTEE-H I observation, the Bayesian evidence suggests that the non-linear model is strongly favoured by our data over the linear one. This fact indicates that a careful analysis is needed to establish whether the observed knee in the $M_{H\,I} - M_\star$ scaling relation is real or not.

(vi) We also find that the evidence for a break in the intrinsic underlying $M_{H\,I} - M_\star$ relation of LTGs is stronger than the evidence for a break in the upper envelope of spirals, demonstrating that the underlying break is stronger than the observed break for the same/similar galaxy samples. The evidence for a break is also stronger for LTGs and ETGs when modelled separately than as a whole sample in the underlying relation.

Taken together, our new analysis using the MIGHTEE-H I Early Science data agrees with the results presented in Maddox et al. (2015), where they also found an upper envelope in the amount of H I that a galaxy can retain is dependent on its stellar mass, and we find that this is likely to be related to the morphology of the galaxy. A direct cause of this result could be the tight link between specific angular momentum (or halo spin parameter) and the gas fraction (Obreschkow et al. 2016; Kurapati et al. 2021; Mancera Piña et al. 2021a, b; Hardwick et al. 2022; Romeo, Agertz & Renaud 2022) for rotation-dominated galaxies. Interestingly, the transition mass that we find using our double-power law (Model B) to describe the upper envelope in the $M_{H\,I} - M_\star$ relation corresponds to the $M_{H\,I}/M_\star$

ratio at which we find that the spin axis of the galaxy to flip from aligned to mis-aligned from its nearest filament, using a subset of the same data (Tudorache et al. 2022). Given that Maddox et al. (2015) suggest that at $M_\star > 10^8$ M$_\odot$, galaxies with higher H I fractions sit in haloes with higher spin parameters, which can work to stabilize H I discs, the spin parameter may in turn be related to their proximity to a filament, along which the gas flows in towards the galaxy (e.g. Codis et al. 2018). Given the limited statistics available in Tudorache et al. (2022) and this study, we cannot decisively investigate these multidimensional trends, however, with the full MIGHTEE survey such an analysis would be within reach.

## DATA AVAILABILITY

The MIGHTEE-H I spectral cubes and source catalogue will be released as part of the first data release of the MIGHTEE survey.

## REFERENCES

Adams N. J., Bowler R. A. A., Jarvis M. J., Häußler B., Lagos C. D. P., 2021, MNRAS, 506, 4933

Aihara H. et al., 2017, PASJ, 70

Aihara H. et al., 2019, PASJ, 71, 114

Barnes D. G. et al., 2001, MNRAS, 322, 486

Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2022, ApJ, 940, L10

Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2023, ApJ, 950, 8

Blyth S. et al., 2016, Proc. MeerKAT Science: On the Pathway to the SKA. PoS, Stellenbosch, p. 4

Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, MNRAS, 452, 2479

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buchner J. et al., 2014, A&A, 564, A125

Butcher Z., Schneider S., van Driel W., Lehnert M. D., 2018, A&A, 619, A89

Calette A. R., Avila-Reese V., Rodríguez-Puebla A., Hernández-Toledo H., Papastergis E., 2018, RMxAA, 54, 443

Catinella B. et al., 2010, MNRAS, 403, 683

Chabrier G., 2003, PASP, 115, 763

Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, Nature, 586, 369

Codis S., Jindal A., Chisari N. E., Vibert D., Dubois Y., Pichon C., Devriendt J., 2018, MNRAS, 481, 4753

Comrie A. et al., 2021, CARTA: Cube Analysis and Rendering Tool for Astronomy, Astrophysics Source Code Library, record ascl:2103.031

Cook R. H. W., Cortese L., Catinella B., Robotham A., 2020, MNRAS, 493, 5596

Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, MNRAS, 415, 1797

Cuillandre J.-C. J. et al., 2012, in Peck A. B., Seaman R. L., Comeron F., eds, Proc. SPIE Conf. Ser.Vol. 8448, Observatory Operations: Strategies, Processes, and Systems IV. SPIE, Bellingham, p. 84480M

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398

Driver S. P. et al., 2022, MNRAS, 513, 439

Erfanianfar G. et al., 2015, MNRAS, 455, 2839

Fasano G., Franceschini A., 1987, MNRAS, 225, 155

Feldmann R., 2020, The link between star formation and gas in nearby galaxies. Commun. Phys., 3, 226

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Fraser-McKelvie A. et al., 2021, MNRAS, 503, 4992

Gavazzi G. et al., 2015, A&A, 580, A116

Giovanelli R. et al., 2005, AJ, 130, 2598

Guo H., Jones M. G., Wang J., Lin L., 2021, ApJ, 918, 53

Hardwick J. A., Cortese L., Obreschkow D., Catinella B., 2022, MNRAS, 516, 4043

Healy J., Blyth S.-L., Elson E., van Driel W., Butcher Z., Schneider S., Lehnert M. D., Minchin R., 2019, MNRAS, 487, 4901

Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, ApJ, 756, 113

Ilbert O. et al., 2006, A&A, 457, 841

Jarvis M. et al., 2016, Proc. MeerKAT Science: On the Pathway to the SKA. PoS, Stellenbosch, p. 6

Jarvis M. J. et al., 2013, MNRAS, 428, 1281

Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, MNRAS, 453, 2540

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS, 477, 2

Kurapati S., Chengalur J. N., Pustilnik S., Kamphuis P., 2018, MNRAS, 479, 228

Kurapati S., Chengalur J. N., Verheijen M. A. W., 2021, MNRAS, 507, 565

Lee N. et al., 2015, ApJ, 801, 80

Leslie S. K. et al., 2020, ApJ, 899, 58

McCracken H. J. et al., 2012, A&A, 544, A156

Maddox N. et al., 2021, A&A, 646, A35

Maddox N., Hess K. M., Obreschkow D., Jarvis M. J., Blyth S. L., 2015, MNRAS, 447, 1610

Malefahlo E. D., Jarvis M. J., Santos M. G., White S. V., Adams N. J., Bowler R. A. A., 2021, MNRAS, 509, 4291

Mancera Piña P. E., Posti L., Fraternali F., Adams E. A. K., Oosterloo T., 2021a, A&A, 647, A76

Mancera Piña P. E., Posti L., Pezzulli G., Fraternali F., Fall S. M., Oosterloo T., Adams E. A. K., 2021b, A&A, 651, L15

Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, PASA, 34, e052

Noeske K. G. et al., 2007, ApJ, 660, L43

Obreschkow D., Glazebrook K., Kilborn V., Lutz K., 2016, ApJ, 824, L26

Pan H., Jarvis M. J., Allison J. R., Heywood I., Santos M. G., Maddox N., Frank B. S., Kang X., 2020, MNRAS, 491, 1227

Pan H., Jarvis M. J., Ponomareva A. A., Santos M. G., Allison J. R., Maddox N., Frank B. S., 2021, MNRAS, 508, 1897

Parkash V., Brown M. J. I., Jarrett T. H., Bonne N. J., 2018, ApJ, 864, 40

Peacock J. A., 1983, MNRAS, 202, 615

Peng Y.-j. et al., 2010, ApJ, 721, 193

Planck Collaboration VI, 2020, A&A, 641, A6

Ponomareva A. A. et al., 2021, MNRAS, 508, 1195

Ponomareva A. A. et al., 2023, MNRAS, 522, 5308

Popesso P. et al., 2019, MNRAS, 483, 3213

Rajohnson S. H. A. et al., 2022, MNRAS, 512, 2697

Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, MNRAS, 435, 2693

Rodríguez-Puebla A., Calette A., Avila-Reese V., Rodriguez-Gomez V. et al., 2020, PASA, 37, e024

Romeo A. B., Agertz O., Renaud F., 2022, MNRAS, 518, 1002

Saintonge A. et al., 2016, MNRAS, 462, 1749

Saintonge A., Catinella B., 2022, ARA&A, 60, 319

Sales L. V., Navarro J. F., Theuns T., Schaye J., White S. D. M., Frenk C. S., Crain R. A., Vecchia C. D., 2012, MNRAS, 423, 1544

Schawinski K. et al., 2014, MNRAS, 440, 889

Schreiber C. et al., 2015, A&A, 575, A74

Sinigaglia F. et al., 2022, ApJ, 935, 9

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Tacchella S. et al., 2018, ApJ, 859, 56

Tudorache M. N. et al., 2022, MNRAS, 513, 2168

Zoldan A., De Lucia G., Xie L., Fontanot F., Hirschmann M., 2018, MNRAS, 481, 1376

[1] *Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa*

[2] *Astrophysics, University of Oxford, Denys Wilkinson Buiding, Keble Road, Oxford OX1 3RH, UK*

[3] *South African Radio Astronomy Observatory (SARAO), 2 Fir Street, Observatory 7925, South Africa*

[4] *School of Physics, H.H. Wills Physics Laboratory, Tyndall Avenue, University of Bristol, Bristol BS8 1TL, UK*

[5] *University Observatory, Faculty of Physics, Ludwig-Maximilians Universität, Scheinerstr. 1, 81679 Munich, Germany*

[6] *The Inter-University Institute for Data Intensive Astronomy (IDIA), University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*

[7] *Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*

[8] *INAF – Istituto di Radioastronomia, Via P. Gobetti 101, 40129 Bologna, Italy*

[9] *Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium*

[10] *Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*

[11] *Dipartimento di Fisica e Astronomia, Università di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova, Italy*

[12]*INAF–Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*

[13]*International Centre for Radio Astronomy Research (ICRAR), The University of Western Australia, 35 Stirling Hwy, Perth, WA 6009, Australia*

[14]*Institute for Astronomy, Royal Observatory, University of Edinburgh, Edinburgh EH9 3HJ, UK*

[15]*Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy*

[16]*INFN Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy*

[17]*Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK*

[18]*Kapteyn Astronomical Institute, University of Groningen, Landleven 12, NL-9747AD Groningen, the Netherlands*

[19]*National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro NM 87801, USA*

[20]*School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia*

[21]*CSIRO Astronomy and Space Science, PO Box 1130, Bentley, WA 6102, Australia*

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