

**MIGHTEE-HI: Evolution of HI Scaling Relations of Star-forming Galaxies at  $z < 0.5$ \***

Francesco Sinigaglia<sup>1,2,3,4</sup> , Giulia Rodighiero<sup>1,2</sup> , Ed Elson<sup>5</sup> , Mattia Vaccari<sup>5,6,7,8</sup> , Natasha Maddox<sup>9</sup> ,  
Bradley S. Frank<sup>6,7,10,11</sup> , Matt J. Jarvis<sup>5,12</sup> , Tom Oosterloo<sup>13,14</sup> , Romeel Dave<sup>5,15,16</sup> , Mara Salvato<sup>17</sup> ,  
Maarten Baes<sup>18</sup> , Sabine Bellstedt<sup>19</sup> , Laura Bisigello<sup>1,2</sup> , Jordan D. Collier<sup>6,7,20,21</sup> , Robin H. W. Cook<sup>19</sup> ,  
Luke J. M. Davies<sup>19</sup> , Jacinta Delhaize<sup>11</sup> , Simon P. Driver<sup>19</sup> , Caroline Foster<sup>22</sup> , Sushma Kurapati<sup>11</sup> ,  
Claudia del P. Lagos<sup>19,23</sup> , Christopher Lidman<sup>24,25</sup> , Pavel E. Mancera Piña<sup>13,14</sup> , Martin J. Meyer<sup>19,23</sup> ,  
K. Moses Mogotsi<sup>26,27</sup> , Hengxing Pan<sup>5,12</sup> , Anastasia A. Ponomareva<sup>12</sup> , Isabella Prandoni<sup>8</sup> ,  
Sambatriniaina H. A. Rajohnson<sup>11</sup> , Aaron S. G. Robotham<sup>19</sup> , Mario G. Santos<sup>5,10</sup> , Srikrishna Sekhar<sup>5,6,7,28</sup> ,  
Kristine Spekkens<sup>29</sup> , Jessica E. Thorne<sup>19</sup> , Jan M. van der Hulst<sup>13</sup> , and O. Ivy Wong<sup>19,21</sup>

<sup>1</sup> Department of Physics and Astronomy, Università degli Studi di Padova, Vicolo dell'Osservatorio 3, I-35122, Padova, Italy; [francesco.sinigaglia@phd.unipd.it](mailto:francesco.sinigaglia@phd.unipd.it)

<sup>2</sup> INAF—Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy

<sup>3</sup> Instituto de Astrofísica de Canarias, Calle via Láctea s/n, E-38205, La Laguna, Tenerife, Spain

<sup>4</sup> Departamento de Astrofísica, Universidad de La Laguna, E-38206, La Laguna, Tenerife, Spain

<sup>5</sup> Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa

<sup>6</sup> The Inter-University Institute for Data Intensive Astronomy (IDIA), Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa

<sup>7</sup> The Inter-university Institute for Data Intensive Astronomy (IDIA), Department of Physics and Astronomy, University of the Western Cape, 7535 Bellville, Cape Town, South Africa

<sup>8</sup> INAF—Istituto di Radioastronomia, via Gobetti 101, I-40129, Bologna, Italy

<sup>9</sup> University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany

<sup>10</sup> South African Radio Astronomy Observatory, 2 Fir Street, Black River Park, Observatory, 7925, South Africa

<sup>11</sup> Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

<sup>12</sup> Oxford Astrophysics, Denys Wilkinson Building, University of Oxford, Keble Road, Oxford, OX1 3RH, UK

<sup>13</sup> Kapteyn Astronomical Institute, University of Groningen, Landleven 12, NL-9747 AD Groningen, The Netherlands

<sup>14</sup> ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, NL-7991 PD Dwingeloo, The Netherlands

<sup>15</sup> SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

<sup>16</sup> South African Astronomical Observatories, Observatory, Cape Town 7925, South Africa

<sup>17</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, D-857498, Garching, Germany

<sup>18</sup> Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium

<sup>19</sup> International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

<sup>20</sup> School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

<sup>21</sup> CSIRO Astronomy and Space Science, P.O. Box 1130, Bentley, WA 6102, Australia

<sup>22</sup> School of Physics, University of New South Wales, Sydney, NSW 2052, Australia

<sup>23</sup> ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>24</sup> Centre for Gravitational Astrophysics, College of Science, The Australian National University, ACT 2601, Australia

<sup>25</sup> The Research School of Astronomy and Astrophysics, The Australian National University, ACT 2601, Australia

<sup>26</sup> South African Astronomical Observatory, P.O. Box 9, Observatory, 7935, Cape Town, South Africa

<sup>27</sup> Southern African Large Telescope, P.O. Box 9, Observatory, 7935, Cape Town, South Africa

<sup>28</sup> National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA

<sup>29</sup> Department of Physics and Space Science, Royal Military College of Canada, P.O. Box 17000, Station Forces, Kingston, Ontario, K7K 7B4, Canada

Received 2022 June 8; revised 2022 July 29; accepted 2022 July 29; published 2022 August 12

**Abstract**

We present the first measurements of HI galaxy scaling relations from a blind survey at  $z > 0.15$ . We perform spectral stacking of 9023 spectra of star-forming galaxies undetected in HI at  $0.23 < z < 0.49$ , extracted from MIGHTEE-HI Early Science data cubes, acquired with the MeerKAT radio telescope. We stack galaxies in bins of galaxy properties (stellar mass  $M_*$ , star formation rate SFR, and specific star formation rate sSFR, with  $\text{sSFR} \equiv M_*/\text{SFR}$ ), obtaining  $\gtrsim 5\sigma$  detections in most cases, the strongest HI-stacking detections to date in this redshift range. With these detections, we are able to measure scaling relations in the probed redshift interval, finding evidence for a moderate evolution from the median redshift of our sample  $z_{\text{med}} \sim 0.37$  to  $z \sim 0$ . In particular, low- $M_*$  galaxies ( $\log_{10}(M_*/M_\odot) \sim 9$ ) experience a strong HI depletion ( $\sim 0.5$  dex in  $\log_{10}(M_{\text{HI}}/M_\odot)$ ), while massive galaxies ( $\log_{10}(M_*/M_\odot) \sim 11$ ) keep their HI mass nearly unchanged. When looking at the star formation activity, highly star-forming galaxies evolve significantly in  $M_{\text{HI}}$  ( $f_{\text{HI}}$ , where  $f_{\text{HI}} \equiv M_{\text{HI}}/M_*$ ) at fixed SFR (sSFR), while at the lowest probed SFR (sSFR) the scaling relations show no evolution. These findings suggest a scenario in which low- $M_*$  galaxies have experienced a strong HI depletion during the last  $\sim 5$  Gyr, while massive galaxies have undergone a significant HI replenishment through some accretion mechanism, possibly minor mergers. Interestingly, our results are in good agreement with the

\* Released on July 29, 2022.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

predictions of the SIMBA simulation. We conclude that this work sets novel important observational constraints on galaxy scaling relations.

*Unified Astronomy Thesaurus concepts:* [Galaxy formation \(595\)](#); [Galaxy evolution \(594\)](#); [Emission line galaxies \(459\)](#); [Astronomical methods \(1043\)](#); [Radio astronomy \(1338\)](#)

## 1. Introduction

In the modern paradigm of galaxy formation and evolution, the baryon cycle of galaxies can be investigated through the parameterization of scaling relations linking their physical properties at different cosmic times.

In this context, the neutral atomic hydrogen (HI) constitutes the fundamental component for  $H_2$  assembly and therefore represents the raw fuel for star formation. Unveiling HI scaling relations in galaxies is thus a task of paramount importance to understand how the availability of fresh cold gas regulates star formation and galaxy evolution.

Tight relations at  $z \sim 0$  between the HI content of star-forming galaxies and their stellar mass ( $M_*$ , Huang et al. 2012; Maddox et al. 2015), star formation rate (SFR, Feldmann 2020), and disk size (e.g., Wang et al. 2016, and references therein), among others, have been revealed by large-scale HI galaxy surveys, such as the HI Parkes All-Sky Survey (HIPASS; Barnes et al. 2001), the Arecibo Legacy Fast ALFA Survey (ALFALFA; Giovanelli et al. 2005), and the GALEX Arecibo SDSS Survey (GASS; Catinella et al. 2010).

Nonetheless, the HI content of galaxies has been so far included in the global scaling relation picture only at very low redshift ( $z < 0.15$ ), due to the intrinsic faintness of the 21 cm hyperfine transition emission line and, hitherto, the limited sensitivity of radio telescopes.

A few blind deep observational efforts—e.g., the Blind Ultra-Deep HI Environmental Survey (BUDHIES; Verheijen et al. 2007), the COSMOS HI Large Extragalactic Survey (CHILES; Hess et al. 2019), and the Arecibo Ultra-Deep Survey (AUDS; Hoppmann et al. 2015)—have reported sparse detections at  $z > 0.15$ . However, their samples are not large enough to constrain scaling relations.

In the build-up toward the the Square Kilometre Array (SKA), spectral line stacking (e.g., Zwaan 2000) can be adopted as an alternative cheaper observational technique to direct detection, performing an average HI mass ( $M_{HI}$ ) detection of a given galaxy sample. Stacking has been proven to be a powerful tool to investigate many different aspects of galaxy evolution, among which are the presence and abundance of HI in galaxy clusters (e.g., Lah et al. 2009; Healy et al. 2021), scaling relations (e.g., Brown et al. 2017), the  $M_{HI}$  content of active galactic nucleus (AGN) host galaxies (Geréb et al. 2015), and the redshift evolution of the HI cosmic density parameter  $\Omega_{HI}$  (e.g., Delhaize et al. 2013; Rhee et al. 2013; Chowdhury et al. 2020, and references therein). In particular, Rhee et al. (2013, 2016, 2018) reported tentative detections ( $\lesssim 3\sigma$ ) at  $z \sim 0.2$ ,  $z \sim 0.37$ , and  $z \sim 0.32$ , respectively. Bera et al. (2019) presented a strong  $\sim 7.3\sigma$  detection at  $z \sim 0.34$ . Chowdhury et al. (2020, 2021, 2022a, 2022b) reported clear  $\sim 4.5\sigma$ ,  $\sim 5\sigma$ ,  $\sim 5.6\sigma$  and  $\sim 5.2\sigma$ , and  $7.1\sigma$  detections with stacking at  $z \sim 1$ ,  $z \sim 1.3$ ,  $z \sim 1$  and  $z \sim 1.3$ , and  $z \sim 1$ , respectively.

In this Letter, we perform spectral stacking on MIGHTEE-HI (see Section 2 for a description of the MIGHTEE survey) data cubes to study HI scaling relations out to  $z \sim 0.5$  and report the first direct measurement of such relations at  $z > 0.15$ . In particular, we focus here on the relations between stellar

mass, HI mass, and star formation activity to study how such relations linking key galaxy parameters evolve at  $z < 0.5$ . Moreover, we compare our stacked results with simulated results from the SIMBA cosmological simulation (Davé et al. 2019).

The paper is structured as follows. In Section 2 we introduce the MIGHTEE-HI survey and our HI data. In Section 3 we present our galaxy sample. In Section 4 we summarize the basic principles of the stacking procedure we adopt throughout the paper. Section 5 presents the main results of this work and our interpretation of them. We conclude in Section 6.

We assume a spatially flat ( $\Omega_k = 0$ )  $\Lambda$ CDM Cosmology, with cosmological parameters from the latest Planck collaboration results (Planck Collaboration et al. 2020), i.e.,  $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m = 0.315$ , and  $\Omega_\Lambda = 0.685$ .

## 2. HI Data from MIGHTEE

The MeerKAT International GigaHertz Tiered Extragalactic Exploration Large Survey Program (MIGHTEE; Jarvis et al. 2016) is a survey, conducted with the MeerKAT radio interferometer (e.g., Jonas & MeerKAT Team 2016), observing four deep, extragalactic fields (COSMOS, XMMLSS, ECDFS, and ELAIS-S), characterized by a wealth of multiwavelength data made available by past and ongoing observational efforts. MeerKAT is the SKA precursor located in South Africa and comprises 64 offset Gregorian dishes (13.5 m diameter main reflector and 3.8 m subreflector), equipped with receivers in the UHF band ( $580 < \nu < 1015 \text{ MHz}$ ),  $L$  band ( $900 < \nu < 1670 \text{ MHz}$ ), and  $S$  band ( $1750 < \nu < 3500 \text{ MHz}$ ).

The MeerKAT data were acquired in spectral and full Stokes mode, thereby making MIGHTEE a spectral line, continuum, and polarization survey. In this paper, we make use of the Early Science HI spectral line data from MIGHTEE, presented in Maddox et al. (2021); B. Frank et al. 2022 (in preparation). The observations were conducted between mid 2018 and mid 2019 and targeted  $\sim 3.5 \text{ deg}^2$  in the XMMLSS field and  $\sim 1.5 \text{ deg}^2$  in the COSMOS field. These observations were performed with the full array (64 dishes) in the  $L$  band, using the 4 k correlator mode (209 kHz, corresponding to  $52 \text{ km s}^{-1}$  at  $z = 0.23$  and  $56 \text{ km s}^{-1}$  at  $z = 0.49$ ). The MIGHTEE-HI Early Science visibilities were processed with the `ProcessMeerKAT` calibration pipeline. The pipeline is `Casa` based and performs standard calibration routines and strategies for the spectral line data such as flagging, bandpass, and complex gain calibration. The continuum subtraction was done in both the visibility and image domain using standard `Casa` routines `UVSUB` and `UVCONTSUB`. Residual visibilities after continuum subtraction were imaged using `Casa`'s task `TCLEAN` with Briggs' weighting (`ROBUST = 0.5`). Eventually, median filtering was applied to the resulting data cubes to reduce the impact of errors due to continuum subtraction. A full description of the data reduction strategy and data quality assessment will be presented in B. Frank et al. 2022 (in preparation).

Out of the full MIGHTEE-HI data set, we make use of MIGHTEE-HI data cubes covering the COSMOS field. Our analysis is limited to the redshift interval  $0.23 < z < 0.49$ . At

**Table 1**

Summary of the Details of MIGHTEE-H I Data Presented in Section 2 and Used in This Paper

MIGHTEE-H I Data	
Survey Parameter	Value
Field	COSMOS
Area	1.5 deg <sup>2</sup>
Integration time	16 hr
Frequency resolution	209 kHz
Velocity resolution	52 km s <sup>-1</sup> at $z = 0.23$
Frequency range	0.950–1.050 GHz
Velocity range	68,952–146,898 km s <sup>-1</sup>
Beam (FWHM)	14''5 × 11''0

these redshifts, MIGHTEE-H I data are found to have well-behaved Gaussian noise, with the median HI noise rms increasing with decreasing frequency, from 85  $\mu\text{Jy beam}^{-1}$  at  $\nu \sim 1050$  MHz to 135  $\mu\text{Jy beam}^{-1}$  at  $\nu \sim 950$  MHz (B. Frank et al. 2022, in preparation). We exclude the spectral bands at  $0.09 < z < 0.23$  and  $z > 0.49$  from our analysis, as they are characterized by strong RFI features (Maddox et al. 2021; B. Frank et al. 2022, in preparation). A first preliminary unguided visual source finding reported no direct HI detections at  $z > 0.23$ . A summary of the features of the MIGHTEE-H I data used in this work is provided in Table 1.

### 3. Sample Selection

We select star-forming galaxies at redshift  $0.23 < z < 0.49$  in the COSMOS field (Scoville et al. 2007), with spectroscopic redshift information available.

We start by considering the latest COSMOS photometric sample publicly available as part of the COSMOS2020 data release (Weaver et al. 2021). This data set includes estimates of the derived galaxy properties (we consider  $M_*$  and SFR in this work) obtained through spectral energy density (SED) fitting. Then, we select star-forming galaxies, according to a color-color  $NUV - r/r - J$  plane selection. In particular, quiescent galaxies are defined as those having  $M_{NUV} - M_r > 3(M_r - M_J) + 1$  and  $M_{NUV} - M_r > 3.1$ , while the remaining galaxies are flagged as star-forming (see Laigle et al. 2016). This defines our parent sample.

Out of the parent sample, we extract only galaxies having a spectroscopic counterpart. This is done by cross-matching the parent sample with a list of spectroscopic redshifts assembled by querying publicly available catalogs from spectroscopic surveys of the COSMOS field (M. Salvato et al. 2022, in preparation.) and spectroscopic redshifts acquired by the DEVILS survey (Davies et al. 2018). Because the SED fitting was performed by adopting the photometric redshift as estimates for redshift, we checked for outliers in the photometric redshifts determination.<sup>30</sup> We find that the outliers constitute  $\lesssim 5\%$  of our sample and excluded them.

Moreover, we explicitly cross-checked the accuracy of the SFR determination, which can induce substantial deviations when obtained through SED fitting in the absence of FIR-band data in the case of dust-attenuated galaxies. In particular, where it applies, we compare our SFR estimates with the ones obtained relying on independent FIR measurements (not used

in the SED fitting mentioned above) presented in Jin et al. (2018). We find that there are no systematic offsets and that the two different SFR estimates are in general consistent within a scatter of  $\sim 0.4$  dex, comparable to IR-based uncertainties on the SFR. On the other hand, weakly and nonattenuated galaxies should have an accurate SFR estimate due to the richness of photometric bands in the COSMOS2020 catalog, provided that photo- $z$  outliers are excluded. We also checked for the presence of AGNs in our sample by cross-matching it with the radio-selected AGN catalog built by Smolčić et al. (2017). AGNs constitute a  $\sim 3.5\%$  fraction, and we do not explicitly remove them from our sample. Their impact on our stacked results will be addressed in future publications.

The COSMOS photometric sample has been shown to be 90% complete in  $M_*$  down to  $\log_{10}(M_*/M_\odot) \sim 8.5$  at  $z \lesssim 0.5$  (e.g., Laigle et al. 2016). When comparing the  $M_*$  distributions (Figure 1, left panel), the spectroscopic sample appears to be faithfully sampling the parent photometric sample at  $\log_{10}(M_*/M_\odot) \gtrsim 9.5$ , while the two distributions differ at  $\log_{10}(M_*/M_\odot) < 9.5$ . However, one should take into account that our spectroscopic sample is the result of the combination of several different surveys, having different targets and performed with different survey strategies. Hence, the resulting incompleteness at small  $M_*$  is not surprising. Nonetheless, as long as the galaxies of the spectroscopic sample at a given  $M_*$  are representative of the star-forming galaxy population of the photometric sample at the same  $M_*$ , the impact of the incompleteness of the sample is minimized by the fact that we group galaxies into bins.

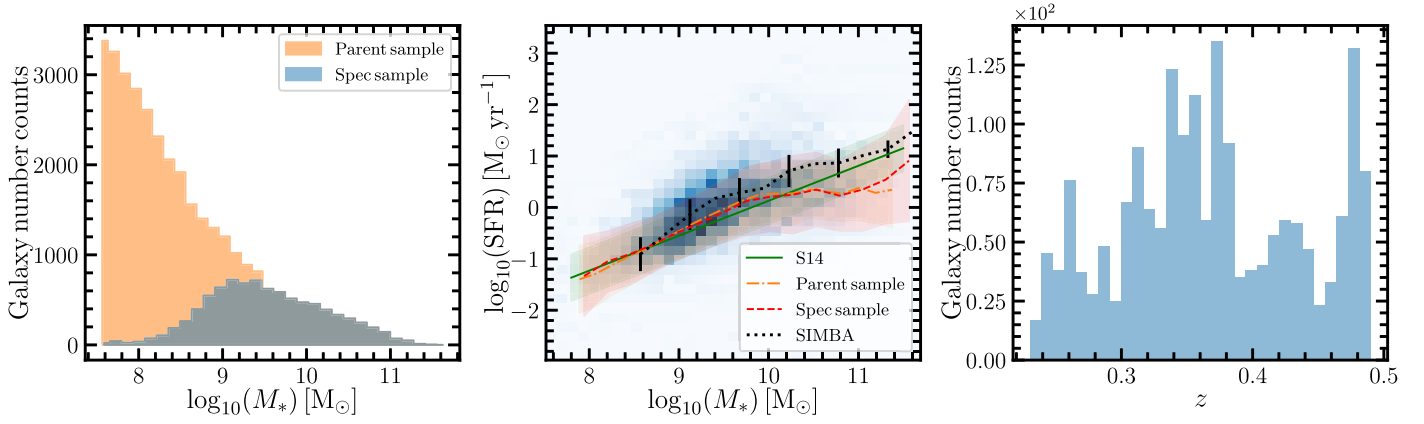
As a cross-check, we compare the spectroscopic and photometric samples in the SFR- $M_*$  plane (Figure 1, central panel). We group galaxies into bins of  $M_*$ , and plot the trend connecting the average SFR in different bins (red dashed for our sample, orange dashed-dotted for the parent photometric sample), with uncertainty given by the standard deviation, reported as shaded areas. The spectroscopic sample is in excellent agreement with the photometric sample, and both span well the main-sequence parameterization provided by Speagle et al. (2014) across its dispersion (above and below the mean relation), at least at  $\log_{10}(M_*/M_\odot) \gtrsim 8.5$ . Therefore, in the remainder of the paper, we assume our sample is not biased by selection effects and include in our analysis all the galaxies down to  $\log_{10}(M_*/M_\odot) \sim 8.5$ . Our final spectroscopic sample consists of 9023 sources.

### 4. Stacking Procedure

We adopt a standard spectral line stacking procedure throughout the paper (see, e.g., Healy et al. 2019). We obtain HI spectra by extracting cubelets around individually undetected galaxies from the full data cubes with suitable apertures ( $3 \times$  the FWHM of the beam on the image plane,  $\pm 1500$  km s<sup>-1</sup> along the velocity axis) and integrating them over angular coordinates. We choose these apertures to ensure that the whole flux emitted by galaxies is included in the cubelets. The angular aperture corresponds to  $\sim 130$  physical kpc at  $z = 0.23$  (i.e., minimum aperture), larger than the typical HI disk size. This breaks the degeneracy between underestimating and overestimating the flux depending on cubelet aperture, leaving us only with the problem of subtracting the flux contamination by nearby sources. Optical coordinates and spectroscopic redshift measurements are used to define the center of the cubelets. Each spectrum at observed frequency

<sup>30</sup> Outliers are defined as those galaxies having  $|z_{\text{spec}} - z_{\text{phot}}| > 0.15 \times (1 + z_{\text{spec}})$ .





**Figure 1.** Left:  $\log_{10}(M_*)$  distribution of our spectroscopic galaxy sample (blue) and of the parent photometric sample (orange). Center: distribution of our sample (blue colormap histogram) on the  $\log_{10}(\text{SFR}) - \log_{10}(M_*)$  plane, with the main-sequence parameterization from Speagle et al. (2014), average trends from our sample, the parent photometric sample, and the SIMBA simulation overlaid as green solid, red dashed, orange dashed-dotted, and black dotted lines, respectively. Red, orange, and green shaded regions indicate uncertainties on the curves of the same colors. Right: redshift distribution of our spectroscopic galaxy sample.

$\nu_{\text{obs}}$  is then de-redshifted to its rest-frame frequency  $\nu_{\text{rf}}$  through  $\nu_{\text{rf}} = \nu_{\text{obs}}(1+z)$  and converted to units of velocity as  $v = cz$ . Furthermore, spectra are resampled to a reference spectral template to ensure that all the spectra are binned in the same manner in the spectral direction. We convert spectra from units of flux to units of  $M_{\text{HI}}$  (per velocity channel) (e.g., Zwaan et al. 2001):

$$M_{\text{HI}}(v) = (2.356 \times 10^5) D_L^2 S(v) (1+z)^{-1} M_{\odot} \text{ km}^{-1} \text{ s},$$

where  $D_L$  is the luminosity distance of the galaxy in units of Mpc,  $S(v)$  is the 21 cm spectral flux density in units of Jy, and  $(1+z)^{-1}$  is a correction factor accounting for the flux reduction due to the expansion of the universe. Lastly, we coadd all the spectra together. The stacked spectrum can then be expressed as

$$\langle M_{\text{HI}}(v) \rangle = \frac{\sum_{i=0}^{n_{\text{gal}}} M_{\text{HI},i}(v) \times w_i \times f_i}{\sum_{i=0}^{n_{\text{gal}}} w_i \times f_i^2}, \quad (1)$$

where  $n_{\text{gal}}$  is the number of coadded spectra, and  $f_i$  and  $w_i$  are the average primary beam transmission and the weight assigned to each source. In the standard unweighted case,  $w_i = 1$  and  $\sum_i w_i = n_{\text{gal}}$ . This equation implements primary beam correction following the procedure detailed in Geréb et al. (2013). Throughout the paper, we assume the Fabello et al. (2011) weighting scheme ( $w_i = 1/\sigma_{\text{rms},i}^2$ ). The  $1\sigma$  noise uncertainty (in units  $M_{\text{HI}}$ ) is evaluated by computing the rms of the noisy channels  $\sigma_{\text{rms}}$  of the stacked spectrum, i.e., those channels outside the spectral interval integrated to compute  $M_{\text{HI}}$ .

To further confirm the legitimacy of our detection, we also generate a reference spectrum obtained by stacking noise spectra (one noise spectrum per galaxy) extracted at randomized positions. The positions of the noise spectra are obtained by adding a fixed angular offset to the center of each galaxy cubelet in a random direction and defined over the same spectral range as the corresponding galaxy cubelet. The angular offset ( $100''$ ) is chosen to guarantee that the reference spectrum is extracted close to the galaxy spectrum, although without overlaps. Also, we double-check that the reference spectrum of each galaxy has no overlaps with other known optical galaxies, and reject it and draw a new one if there is any overlap.

We compute the integrated signal-to-noise ratio (S/N) of the final stacked spectrum as

$$S/N = \sum_i^{N_{\text{ch}}} \langle S_i \rangle / (\sigma_{\text{rms}} \sqrt{N_{\text{ch}}}), \quad (2)$$

where  $\langle S_i \rangle$  is the stacked spectrum and  $N_{\text{ch}}$  is the number of channels over which the integration is performed (e.g., Healy et al. 2019). We estimate uncertainties on the stacked spectrum by applying jackknife resampling to the galaxy sample, eliminating one galaxy at a time.

We address the problem of flux contamination due to source confusion using detailed MeerKAT-like simulated data cubes. In particular, we use the Obreschkow & Meyer (2014) flux-limited mock galaxy catalog, based on the SKA Simulated Skies semianalytic simulations ( $S^3$ -SAX) to inject galaxies with realistic HI masses and clustering into a blank synthetic data cube matching the same angular and spectral size as our observations (see also Elson et al. 2016, 2019). Then, following Elson et al. (2016), we decomposed the spectrum extracted for each target galaxy into contributions from the actual target and contributions from nearby contaminating galaxies. In this way, we could accurately calculate that the level of contamination is  $\lesssim 10\%$  in all the studied cases and subtract a fixed 10% contribution from the output  $\langle M_{\text{HI}} \rangle$  from stacking.

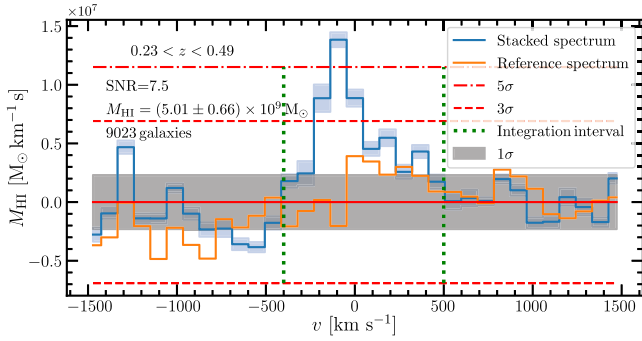
## 5. Results and Discussion

In this section we present the main results of this paper. We first discuss the results yielded by stacking in the  $0.23 < z < 0.49$  redshift range and then compare them with known scaling relations at mean redshift  $z \sim 0$ .

### 5.1. Stacking at $0.23 < z < 0.49$ on MIGHTEE-HI Data

We first produce a global stack using the full galaxy sample. Figure 2 shows the resulting stacked spectrum. As anticipated, we detect the signal at  $\gtrsim 5\sigma$ . This represents the strongest detection obtained to date with stacking in this redshift interval. The corresponding  $\langle M_{\text{HI}} \rangle$  measurement and integrated S/N are reported in Figure 2.

In Figure 3 we present the stacked spectra obtained in bins of  $M_*$  (top row), SFR (mid row), and specific star formation rate (sSFR) (bottom row). Here, we fix empirically the bin limits to



**Figure 2.** Stacked  $M_{\text{HI}}$  spectrum obtained in the full  $0.23 < z < 0.49$  redshift range we investigate in this paper. The blue solid line and the blue shaded region show the stacked spectrum and the associated jackknife uncertainties, respectively. The orange solid line and the gray shaded area indicate the reference spectrum (see text) and  $\sigma_{\text{rms}}$ , respectively. The red dashed and dashed-dotted horizontal lines mark the  $3\sigma_{\text{rms}}$  and  $5\sigma_{\text{rms}}$  noise levels, respectively. The green vertical line shows the velocity integration limits in the  $M_{\text{HI}}$  computation.

find a compromise between having enough sources per bin to claim a detection and dissecting the  $\log_{10}(\text{SFR}) - \log_{10}(M_*)$  in meaningful intervals. Our scaling relations are evaluated at the median redshift of our sample  $z_{\text{med}} \sim 0.37$  (see Figure 1, right panel, for the redshift distribution of our sample). To address the potential impact of the fact that the redshift distributions of the subsamples defined in different galaxy properties bins may peak at redshifts higher/lower than  $z \sim 0.37$  due to selection effects, we compute the median redshift of each subsample obtained with the aforementioned property cuts. The results we obtain are

1.  $M_*$  bins:  $z_{\text{med}} = \{0.36, 0.37, 0.36\}$ ;
2. SFR bins:  $z_{\text{med}} = \{0.35, 0.37, 0.38\}$ ;
3. sSFR bins:  $z_{\text{med}} = \{0.36, 0.36, 0.38\}$ .

We notice that galaxies in all the bins have distribution peaking around the global median redshift  $z_{\text{med,tot}} \sim 0.37$ , with maximum percentage deviation  $\Delta z_{\text{max}}/z_{\text{med,tot}} \sim 5\%$ , i.e., a negligible effect. As a result, we regard our results not to be affected by redshift biases.

In this case, we detect signal at  $\gtrsim 5\sigma$  in six bins and at  $\gtrsim 3\sigma$  in the remaining three bins. The corresponding  $\langle M_{\text{HI}} \rangle$  measurements are reported inside the panels in Figure 3 together with the resulting integrated S/N and the number of coadded spectra. We notice that some stacks present nonnegligible negative mass structures at  $|v| > 500 \text{ km s}^{-1}$ , where only random noise should be present. The origin of these features may be due to continuum oversubtraction or residual RFI. To test the possible effect of RFI, we repeated the stacking procedure after carefully flagging the frequency bands affected by RFI (B. Frank et al. 2022, in preparation.) and conservatively excluding the galaxies lying in such regions. This is done to mitigate the possible negative impact of badly behaved spectra, whose importance was not already suppressed by the weighting scheme. After this operation, the resulting stacked spectra no longer present outlier negative mass structures, and the  $M_{\text{HI}}$  estimates are in excellent agreement with the ones obtained with the full sample, indicating that our results are not biased by this issue.

To fully account for these features, in the three worst cases (central and right panels in the top row, central panel in the central row of Figure 3) we include an additional uncertainty

term in Equation 2, replacing  $\sigma_{\text{rms}}$  with  $\sigma = \sigma_{\text{rms}} + \sigma_{\text{dip}}$ . We estimate  $\sigma_{\text{dip}}$  empirically in such a way as to reduce the statistical significance of the negative structures below the detection threshold (conservatively set at  $2.5\sigma$ ). The final S/N shown in Figure 3 is the conservative estimate obtained after this additional operation, while previously the calculation based only on  $\sigma_{\text{rms}}$  returned S/N  $> 5$  in all three cases.

## 5.2. Scaling Relations at $z \sim 0$

We adopt as fiducial observational results at  $z \sim 0$  the findings presented in Guo et al. (2021, G21 hereafter). Therein, the authors investigate the interdependence of  $M_{\text{HI}}$ ,  $M_*$ , and SFR, among others, performing spectral stacking on cross-matched ALFALFA (Haynes et al. 2018) and Sloan Digital Sky Survey (SDSS, York et al. 2000) data. Results are in close agreement with previous results (e.g., Brown et al. 2017; Catinella et al. 2018), where applicable. Furthermore, the scaling relations are assessed separately for star-forming and quenched galaxies. Therefore, within this work we make comparisons only with the star-forming galaxy results.

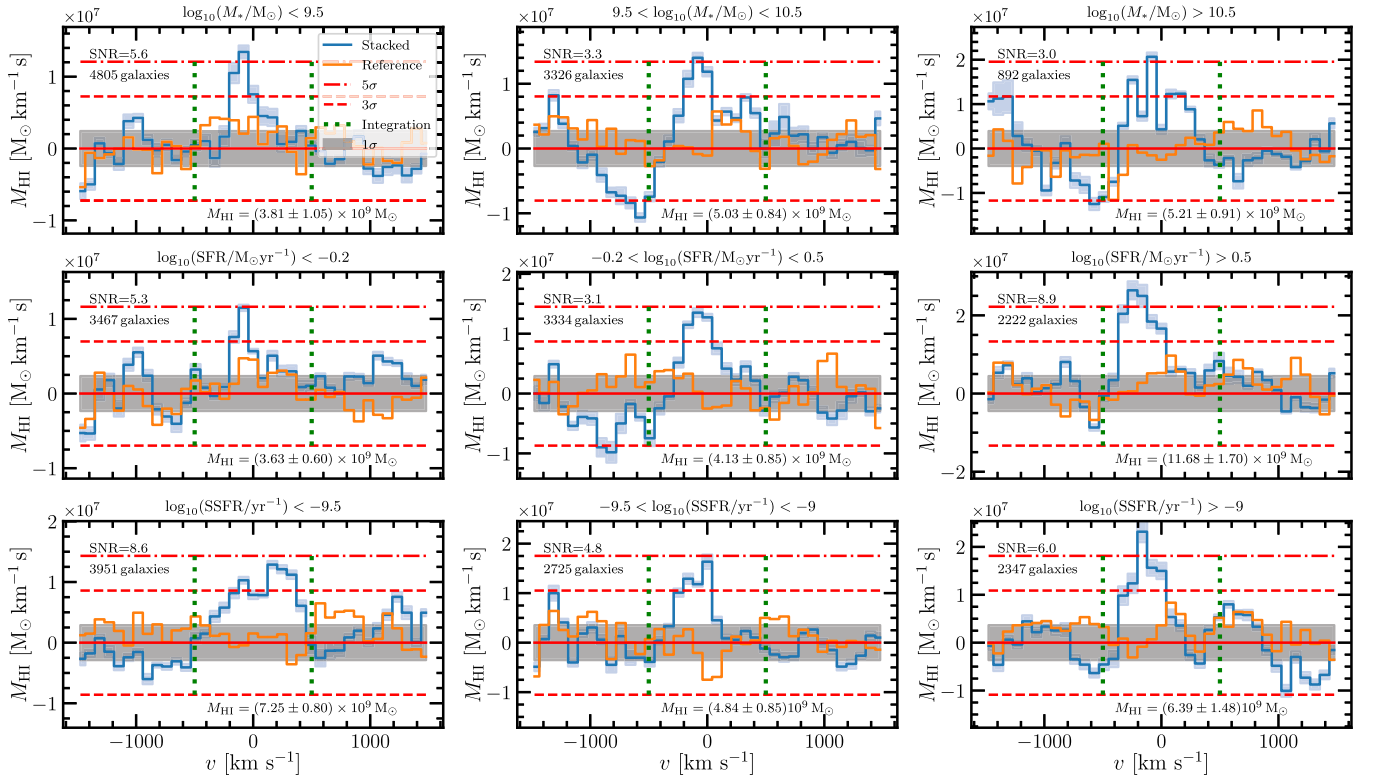
### 5.3. SIMBA: Reference Cosmological Hydro Simulation

SIMBA is a cosmological hydrodynamic simulation run with the GIZMO meshless finite-mass hydrodynamics, employing  $N_{\text{dm}} = 1024^3$  dark matter particles and  $N_{\text{gas}} = 1024^3$  gas elements in a  $V = (100 \text{ Mpc } h^{-1})^3$  comoving volume. The SIMBA fiducial model adopts and updates star formation and feedback subgrid prescriptions used in the MUFASA simulation (Davé et al. 2016) and introduces the treatment of black hole growth and accretion from cold and hot gas. Moreover, models for on-the-fly dust production, growth, and destruction, and HI and  $\text{H}_2$  abundance computation are implemented (see Davé et al. 2019 and references therein for details).

In this work, we make use of SIMBA galaxy catalogs at  $z = 0$  and  $z = 0.4$  and select only star-forming galaxies with  $\log_{10}(M_*) > 8.5$ , consistent with our observational data, by separating the star-forming and quenched populations in the  $M_*$ -SFR plane. In particular, we flag as star-forming those galaxies with  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) > \log_{10}(M_*/M_{\odot}) - 10.3$  and  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) > 0.9 \times \log_{10}(M_*/M_{\odot}) - 9.2$  at  $z = 0$  and  $z = 0.4$ , respectively. These equations are obtained from a  $3\sigma$  cut below the Speagle et al. (2014) parameterizations at the corresponding redshifts, slightly modified to account for the mismatch between the SIMBA and the observed main sequence and to separate well star-forming and quiescent galaxies.

### 5.4. Comparison of Scaling Relations

Figure 4 shows a comparison between our resulting scaling relations at  $0.23 < z < 0.49$ , the findings of G21 at  $z \sim 0$  and the theoretical results obtained by the SIMBA simulation at  $z = 0$  and  $z = 0.4$ . The top-left panel shows the  $\log_{10}(M_{\text{HI}}) - \log_{10}(M_*)$  relation, the top-right panel shows the  $\log_{10}(f_{\text{HI}}) - \log_{10}(M^*)$  relation, the bottom-left panel shows the  $\log_{10}(M_{\text{HI}}) - \log_{10}(\text{SFR})$  relation, and the bottom-right panel shows the  $\log_{10}(f_{\text{HI}}) - \log_{10}(\text{sSFR})$  relation. The SIMBA curves were generated by breaking the full galaxy sample into bins of  $M_*$  (top-left panel), or SFR (bottom-left panel), or sSFR (bottom-right panel), and computing the average and standard deviation in each bin.



**Figure 3.** Stacked  $M_{\text{HI}}$  spectra obtained in different intervals of  $M_*$  (top row), SFR (mid row), and sSFR (bottom row). The title of each panel indicates the corresponding interval of galaxy properties. Lines and colors follow the same scheme as Figure 2.

We fit a linear model  $\log_{10}(Y) = \alpha \log_{10}(X) + \beta$  to our data at  $z \sim 0.37$ . We compute the mean values for the fitting parameters and the associated uncertainties (68% confidence interval) through parametric bootstrap resampling of our data, drawing 10,000 samples and fitting the model with a least-squares minimization to each sample. The resulting best-fitting coefficients are

1.  $\log_{10}(M_{\text{HI}}) = \alpha \log_{10}(M_*) + \beta$ :  $\alpha = 0.08^{+0.07}_{-0.07}$  ( $\sim 4.8\sigma$  difference from  $z \sim 0$ ),  $\beta = 8.86^{+0.75}_{-0.74}$ ;
2.  $\log_{10}(f_{\text{HI}}) = \alpha \log_{10}(M^*) + \beta$ :  
 $\alpha = -0.92^{+0.07}_{-0.07}$ ,  $\beta = 8.88^{+0.72}_{-0.75}$ ;
3.  $\log_{10}(M_{\text{HI}}) = \alpha \log_{10}(\text{SFR}) + \beta$ :  
 $\alpha = 0.26^{+0.04}_{-0.04}$ ,  $\beta = 9.65^{+0.44}_{-0.43}$ ;
4.  $\log_{10}(f_{\text{HI}}) = \alpha \log_{10}(\text{sSFR}) + \beta$ :  
 $\alpha = 0.40^{+0.07}_{-0.07}$ ,  $\beta = 3.56^{+0.69}_{-0.68}$ .

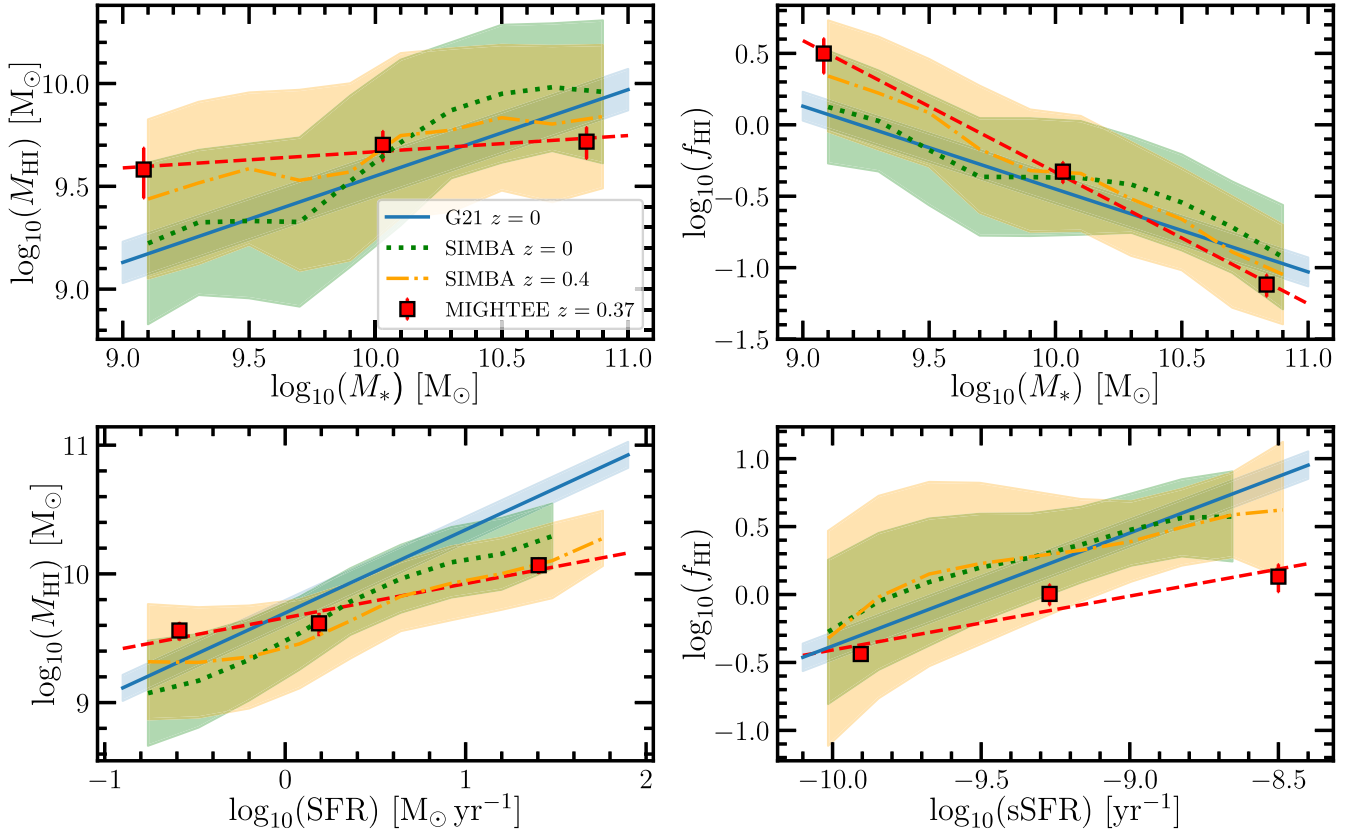
The  $\log_{10}(M_{\text{HI}}) - \log_{10}(M_*)$  relation (Figure 4, top-left panel) is found to have comparable normalization at the two redshifts at  $\log_{10}(M_*/M_{\odot}) \sim 10.3$ , but different slopes, the relation at  $z=0$  has a steeper slope than that at  $z \sim 0.37$ . In particular, low- $M_*$  galaxies ( $\log_{10}(M_*/M_{\odot}) \sim 9$ ) are  $\sim 0.5$  dex more HI-rich at  $z \sim 0.37$  than at  $z \sim 0$ , while massive galaxies ( $\log_{10}(M_*/M_{\odot}) \sim 11$ ) at  $z \sim 0.37$  and  $z \sim 0$  converge to similar  $M_{\text{HI}}$  values within 0.25 dex (and compatible within uncertainties). The same result can be visualized in terms of the HI

fraction in the  $\log_{10}(f_{\text{HI}}) - \log_{10}(M_*)$  plane (Figure 4, top-right panel).

In terms of star formation properties, the SFR displays a weaker correlation with  $M_{\text{HI}}$  than at  $z \sim 0$  (bottom-left panel). Galaxies having  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) \lesssim 0.5$  show no  $M_{\text{HI}}$  evolution. Instead, galaxies at  $z \sim 0$  with  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) > 0.5$  feature an excess in  $M_{\text{HI}}$  of up to  $\sim 0.5$  dex in  $\log_{10}(M_{\text{HI}})$  at fixed  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) \sim 1.4$  with respect to galaxies at  $z \sim 0.37$ . Lastly, the  $f_{\text{HI}} - \text{sSFR}$  relations at  $z \sim 0$  and  $z \sim 0.37$  converge at  $\log_{10}(\text{sSFR}/\text{yr}^{-1}) \sim -10$ , while the galaxies at  $z \sim 0.37$  at (fixed) larger sSFR have systematically smaller  $f_{\text{HI}}$  than at  $z \sim 0$  (up to  $\sim 0.8$  dex).

### 5.5. Discussion

Our findings suggest evidence for moderate evolution of scaling relations from  $z \sim 0.37$  to  $z \sim 0$ . Low- $M_*$  galaxies have depleted their HI reservoirs over the last 4 Gyr. The long HI depletion time due to star formation for these galaxies,  $\tau_{\text{dep,HI}}(\dot{M}_* = \log_{10}(M_*) \sim 9.1) = M_{\text{HI}}(\dot{M}_*)/\text{SFR}(\dot{M}_*) \sim 15$  Gyr, suggests that star formation depletes only  $\Delta M_{\text{HI}} \sim 1 \times 10^9 M_{\odot}$  (i.e.,  $\sim 0.15$  dex) over the last 4 Gyr. As a result, we argue that star formation alone could not be able to account for the observed reduction in  $M_{\text{HI}}$  from  $z \sim 0.37$  to  $z \sim 0$  and that another HI depletion/removal mechanism may be acting in parallel to star formation. On the other hand, massive galaxies have experienced an efficient replenishment of the HI content in their disks, counteracting the HI depletion due to star formation and feedback processes. In particular, assuming the  $\text{H}_2$  formation rate and the SFR are in equilibrium, and that the  $\text{H}_2$  depletion time  $\tau_{\text{H}_2}$  at  $z < 0.5$  is of order  $0.1 < \tau_{\text{H}_2} < 1$  Gyr (e.g., Tacconi et al. 2018 and references therein), there must be an efficient HI accretion mechanism fueling massive galaxies (see, e.g., Sancisi et al. 2008).



**Figure 4.** Star-forming galaxy HI scaling relations:  $\log_{10}(M_{\text{HI}}) - \log_{10}(M_*)$  (top left),  $\log_{10}(f_{\text{HI}}) - \log_{10}(M_*)$  (top right),  $\log_{10}(M_{\text{HI}}) - \log_{10}(\text{SFR})$  (bottom left), and  $\log_{10}(f_{\text{HI}}) - \log_{10}(\text{sSFR})$ . Our H I-stacking results at median redshift  $z \sim 0.37$  are displayed as red squares. Uncertainties on red squares along the x-axis, not shown for the sake of clarity of visualization, correspond to the width of the bins. The red dashed line represents the fit to our data. Our reference results at  $z \sim 0$  (G21, Guo et al. 2021) are plotted as a blue solid line. Green dotted and yellow dashed-dotted lines show the predictions of the SIMBA cosmological hydrodynamic simulation at  $z = 0$  and  $z = 0.4$ , respectively, for comparison. Shaded regions indicate  $1\sigma$  uncertainties.

This is also consistent with the HI depletion time of massive galaxies that we are able to derive in this work:  $\tau_{\text{dep, HI}}(\dot{M}_* = \log_{10}(M_*) \sim 10.75) = M_{\text{HI}}(\dot{M}_*) / \text{SFR}(\dot{M}_*) \sim 1 - 1.2$  Gyr. The nature of the HI growth into galaxies is still debated, with coplanar accretion from cosmic flows being favored by observations of Mg II absorbers in quasars down to  $z \sim 0.2$  and statistical arguments (e.g., Bouché et al. 2013; Peng & Renzini 2020, and references therein). Theoretical predictions indicate that at  $z < 0.5$ , the main cold gas accretion mode onto galaxies is not cosmological accretion (as suggested for accretion at high  $z$  by results based on observations; e.g., Conselice et al. 2013) or the galactic fountain mechanisms (e.g., Fraternali 2017), but mergers (e.g., Sánchez Almeida et al. 2014; Padmanabhan & Loeb 2020). Furthermore, simulations tell that most of the HI in the universe is already contained in galaxy disks at the probed redshifts (e.g.,  $\sim 97\%$  at  $z = 0$  in Villaescusa-Navarro et al. 2018), thus making mergers a potential efficient HI transfer mechanism. We speculate that minor mergers between low- $M_*$  HI-rich galaxies and massive galaxies could be the mechanism that mainly refills the latter of HI. Intriguingly, Jackson et al. (2022) recently found evidence in observations that minor mergers play a major role in the formation of HI-rich massive disk galaxies at  $z \sim 0$ , supporting our proposed scenario. On the other hand, Di Teodoro & Fraternali (2014) argue that cold gas transfer through minor mergers at  $z \sim 0$  is not able to sustain star formation, even under stringent assumptions. This is however not in direct contrast with our findings, as low- $M_*$  galaxies are found to contain much less HI at  $z \sim 0$  than at  $z \sim 0.37$ . In any case, our conclusions do not exclude the scenario

in which smooth accretion from the intergalactic medium is the dominant cold gas accretion onto galaxies.

The decrease in  $M_{\text{HI}}$  observed in highly star-forming galaxies with  $\log_{10}(\text{SFR}/(M_{\odot} \text{ yr}^{-1})) > 0.5$  from  $z \sim 0.37$  to  $z \sim 0$  suggests that HI features a stronger correlation with star formation at the latter redshift. Interesting insights are offered by the  $f_{\text{HI}} - \text{sSFR}$  relation. In fact, making bins in sSFR corresponds to binning the  $\text{SFR} - M_*$  plane (central panel, Figure 1) with bin limits being diagonal lines in the  $\log_{10}(\text{SFR}) - \log_{10}(M_*)$  plane. In particular, the three bins roughly correspond to galaxies in the lower (and below), central, and upper (and above) parts of the main sequence, from lower to higher sSFR, respectively. This suggests that the galaxies at fixed sSFR lying above the main sequence are the ones experiencing a larger increase of their  $f_{\text{HI}}$  over the last  $\sim 5$  Gyr.

To develop a more complete picture on the SFR evolution, we would need to include also the  $\text{H}_2$  scaling relations in our framework, which goes, however, beyond the scope of the paper. We leave such a study for future work.

However, we notice that HI replenishment in massive galaxies is not able to supply enough gas to fully sustain star formation and, hence, prevent the observed reduction in the SFR. We speculate that the main reason for this could be that fresh HI accretes onto the outer part of the disk and takes a significant amount of time to migrate toward the region within the optical radius—where the bulk of star formation takes place—due to galaxy angular momentum (Peng & Renzini 2020).



Interestingly, our findings are in good agreement with the predictions of the SIMBA simulation employing the full baryon physics model. The only significant discrepancy is found in the  $\log_{10}(f_{\text{HI}}) - \log_{10}(\text{sSFR})$  plane and is due to the SIMBA main sequence at  $z = 0.4$  having a systematic positive SFR offset with respect to the observed main sequence (both the Speagle et al. 2014 parameterization and, even more, the observed sample), especially at large  $M_*$ . The immediate aim of this comparison is to provide a minimum contextualization of our observational results into a theoretical scenario. However, the agreement between theory and observations offers the unique advantage of being able to use SIMBA as a benchmark to study which is the driving process that determines the observed scaling relations, and to compare it more thoroughly to our data to better constrain models. This goes beyond the scope of the paper, and we leave it for future work. The full SIMBA suite also includes other runs with only partial modeling of feedback and baryon processes. Davé et al. (2020) find that the most crucial phenomena to be modeled to reproduce the high-mass end of the HI (and H<sub>2</sub>) mass function are AGN, X-ray, and jet feedback. We plan to perform an in-depth investigation of the impact of these aspects on scaling relations in forthcoming publications, comparing our findings with other cosmological simulations and semianalytical models too.

## 6. Conclusions

In this paper, we have performed stacking of 9023 21 cm undetected star-forming galaxy spectra extracted from MIGH-TEE-HI data cubes at  $0.23 < z < 0.49$  in the COSMOS field. In particular, we have subdivided the full sample into galaxy properties subsets with the aim of directly measuring for the first time HI scaling relations at a median redshift  $z_{\text{med}} \sim 0.37$ .

We find moderate evolution of the probed scaling relations from  $z \sim 0.37$  to  $z \sim 0$ , with no significant evolution in the  $\log_{10}(M_{\text{HI}}) - \log_{10}(M_*)$  and  $\log_{10}(f_{\text{HI}}) - \log_{10}(M_*)$  relations at  $\log_{10}(M_*/M_{\odot}) \gtrsim 10$ , implying the necessity of an efficient HI replenishment mechanism in massive galaxies. The  $\log_{10}(M_{\text{HI}}) - \log_{10}(\text{SFR})$  and  $\log_{10}(f_{\text{HI}}) - \log_{10}(\text{sSFR})$  relations evidence how highly star-forming galaxies evolve significantly in SFR (sSFR) at fixed  $M_{\text{HI}} (f_{\text{HI}})$ , while the evolution of scaling relations at lower SFR (sSFR) is milder. We argue that the aforementioned HI replenishment mechanism is not able to prevent star formation quenching in massive galaxies. We will further investigate these aspects in forthcoming publications.

We argue that future MIGHTEE-HI data beyond the Early Science data set will allow us to strengthen the statistical significance of the results, as they will enlarge the footprint at  $0.23 < z < 0.49$  from the  $\sim 1.5 \text{ deg}^2$  of the COSMOS field used in this paper, to  $\sim 20 \text{ deg}^2$  of the final data release.

The authors warmly thank the anonymous referee for the constructive review and helpful comments they offered. The authors also thank Alvio Renzini for insightful comments and discussions. F.S. acknowledges the support of the doctoral grant funded by the University of Padova and by the Italian Ministry of Education, University and Research (MIUR). G.R. acknowledges the support from grant PRIN MIUR 2017-20173ML3WW\_001. M.V. acknowledges financial support from the South African Department of Science and Innovation's National Research Foundation under the ISARP

RADIOSKY2020 Joint Research Scheme (DSI-NRF grant No. 113121) and the CSUR HIPPO Project (DSI-NRF grant No. 121291). N.M. acknowledges support of the LMU Faculty of Physics. H.P. acknowledges support from the South African Radio Astronomy Observatory (SARAO). S.H.A.R. is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation. C.F. is the recipient of an Australian Research Council Future Fellowship (project number FT210100168) funded by the Australian Government. A.A.P. acknowledges support of the STFC consolidated grant ST/S000488/1 and from the Oxford Hintze Centre for Astrophysical Surveys, which is funded through generous support from the Hintze Family Charitable Foundation. I.P. acknowledges financial support from the Italian Ministry of Foreign Affairs and International Cooperation (MAECI grant No. ZA18GR02) and the South African Department of Science and Technology's National Research Foundation (DST-NRF grant No. 113121) as part of the ISARP RADIOSKY2020 Joint Research Scheme. J.M.v.d.H. acknowledges support from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) / ERC grant Agreement nr. 291531 (HISTORYNU). The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge the use of the ilifu cloud computing facility—www.ilifu.ac.za, a partnership between the University of Cape Town, 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 the Inter-University Institute for Data Intensive Astronomy (IDIA a partnership between the University of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory), the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA).

## ORCID iDs

Francesco Sinigaglia  <https://orcid.org/0000-0002-0639-8043>

Giulia Rodighiero  <https://orcid.org/0000-0002-9415-2296>

Ed Elson  <https://orcid.org/0000-0001-9359-0713>

Mattia Vaccari  <https://orcid.org/0000-0002-6748-0577>

Natasha Maddox  <https://orcid.org/0000-0001-8312-5260>


Bradley S. Frank  <https://orcid.org/0000-0003-3599-1521>

Matt J. Jarvis  <https://orcid.org/0000-0001-7039-9078>

Tom Oosterloo  <https://orcid.org/0000-0002-0616-6971>

Romeel Davé  <https://orcid.org/0000-0003-2842-9434>

Mara Salvato  <https://orcid.org/0000-0001-7116-9303>

Maarten Baes  <https://orcid.org/0000-0002-3930-2757>

Sabine Bellstedt  <https://orcid.org/0000-0003-4169-9738>

Laura Bisigello  <https://orcid.org/0000-0003-0492-4924>

Jordan D. Collier  <https://orcid.org/0000-0002-2326-7432>

Robin H. W. Cook  <https://orcid.org/0000-0002-3238-8359>

Luke J. M. Davies  <https://orcid.org/0000-0003-3085-0922>

Jacinta Delhaize  <https://orcid.org/0000-0002-6149-0846>

Simon P. Driver  <https://orcid.org/0000-0001-9491-7327>

Caroline Foster  <https://orcid.org/0000-0003-0247-1204>



Sushma Kurapati  <https://orcid.org/0000-0001-6615-5492>  
 Claudia del P. Lagos  <https://orcid.org/0000-0003-3021-8564>  
 Christopher Lidman  <https://orcid.org/0000-0003-1731-0497>  
 Pavel E. Mancera Piña  <https://orcid.org/0000-0001-5175-939X>  
 Martin J. Meyer  <https://orcid.org/0000-0002-2838-3010>  
 K. Moses Mogotsi  <https://orcid.org/0000-0002-5136-7983>  
 Hengxing Pan  <https://orcid.org/0000-0002-9160-391X>  
 Anastasia A. Ponomareva  <https://orcid.org/0000-0003-4100-0173>  
 Isabella Prandoni  <https://orcid.org/0000-0001-9680-7092>  
 Sambatriniaina H. A. Rajohnson  <https://orcid.org/0000-0001-8461-4701>  
 Aaron S. G. Robotham  <https://orcid.org/0000-0003-0429-3579>  
 Mario G. Santos  <https://orcid.org/0000-0003-3892-3073>  
 Srikrishna Sekhar  <https://orcid.org/0000-0002-8418-9001>  
 Kristine Spekken  <https://orcid.org/0000-0002-0956-7949>  
 Jessica E. Thorne  <https://orcid.org/0000-0002-7921-0785>  
 Jan M. van der Hulst  <https://orcid.org/0000-0002-9316-763X>  
 O. Ivy Wong  <https://orcid.org/0000-0003-4264-3509>

## References

- Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., et al. 2001, *MNRAS*, **322**, 486
- Bera, A., Kanekar, N., Chengalur, J. N., & Bagla, J. S. 2019, *ApJL*, **882**, L7
- Bouché, N., Murphy, M. T., Kacprzak, G. G., et al. 2013, *Sci*, **341**, 50
- Brown, T., Catinella, B., Cortese, L., et al. 2017, *MNRAS*, **466**, 1275
- Catinella, B., Schiminovich, D., Kauffmann, G., et al. 2010, *MNRAS*, **403**, 683
- Catinella, B., Saintonge, A., Janowiecki, S., et al. 2018, *MNRAS*, **476**, 875
- Chowdhury, A., Kanekar, N., Chengalur, J. N., Sethi, S., & Dwarakanath, K. S. 2020, *Natur*, **586**, 369
- Chowdhury, A., Kanekar, N., Das, B., Dwarakanath, K. S., & Sethi, S. 2021, *ApJL*, **913**, L24
- Chowdhury, A., Kanekar, N., & Chengalur, J. N. 2022a, *ApJL*, **931**, L34
- Chowdhury, A., Kanekar, N., & Chengalur, J. N. 2022b, arXiv:2207.00031
- Conselice, C. J., Mortlock, A., Bluck, A. F. L., Grützbauch, R., & Duncan, K. 2013, *MNRAS*, **430**, 1051
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, **486**, 2827
- Davé, R., Crain, R. A., Stevens, A. R. H., et al. 2020, *MNRAS*, **497**, 146
- Davé, R., Thompson, R., & Hopkins, P. F. 2016, *MNRAS*, **462**, 3265
- Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2018, *MNRAS*, **480**, 768
- Delhaize, J., Meyer, M. J., Staveley-Smith, L., & Boyle, B. J. 2013, *MNRAS*, **433**, 1398
- Di Teodoro, E. M., & Fraternali, F. 2014, *A&A*, **567**, A68
- Elson, E. C., Baker, A. J., & Blyth, S. L. 2019, *MNRAS*, **486**, 4894
- Elson, E. C., Blyth, S. L., & Baker, A. J. 2016, *MNRAS*, **460**, 4366
- Fabello, S., Catinella, B., Giovanelli, R., et al. 2011, *MNRAS*, **411**, 993
- Feldmann, R. 2020, *CmPhy*, **3**, 226
- Fraternali, F. 2017, in *Gas Accretion onto Galaxies*, ed. A. Fox & R. Davé, Vol. 430 (Berlin: Springer), 323
- Geréb, K., Morganti, R., Oosterloo, T. A., Guglielmino, G., & Prandoni, I. 2013, *A&A*, **558**, A54
- Geréb, K., Morganti, R., Oosterloo, T. A., Hoppmann, L., & Staveley-Smith, L. 2015, *A&A*, **580**, A43
- Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, *AJ*, **130**, 2598
- Guo, H., Jones, M. G., Wang, J., & Lin, L. 2021, *ApJ*, **918**, 53
- Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, *ApJ*, **861**, 49
- Healy, J., Blyth, S. L., Elson, E., et al. 2019, *MNRAS*, **487**, 4901
- Healy, J., Blyth, S. L., Verheijen, M. A. W., et al. 2021, *A&A*, **650**, A76
- Hess, K. M., Luber, N. M., Fernández, X., et al. 2019, *MNRAS*, **484**, 2234
- Hoppmann, L., Staveley-Smith, L., Freudling, W., et al. 2015, *MNRAS*, **452**, 3726
- Huang, S., Haynes, M. P., Giovanelli, R., & Brinchmann, J. 2012, *ApJ*, **756**, 113
- Jackson, R. A., Kaviraj, S., Martin, G., et al. 2022, *MNRAS*, **511**, 607
- Jarvis, M., Taylor, R., Agudo, I., et al. 2016, *MeerKAT Science: On the Pathway to the SKA (Trieste: PoS)*, 6
- Jin, S., Daddi, E., Liu, D., et al. 2018, *ApJ*, **864**, 56
- Jonas, J. & MeerKAT Team 2016, *MeerKAT Science: On the Pathway to the SKA (Trieste: PoS)*
- Lah, P., Pracy, M. B., Chengalur, J. N., et al. 2009, *MNRAS*, **399**, 1447
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24
- Maddox, N., Hess, K. M., Obreschkow, D., Jarvis, M. J., & Blyth, S. L. 2015, *MNRAS*, **447**, 1610
- Maddox, N., Frank, B. S., Ponomareva, A. A., et al. 2021, *A&A*, **646**, A35
- Obreschkow, D., & Meyer, M. 2014, arXiv:1406.0966
- Padmanabhan, H., & Loeb, A. 2020, *MNRAS*, **496**, 1124
- Peng, Y.-j., & Renzini, A. 2020, *MNRAS*, **491**, L51
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Rhee, J., Lah, P., Briggs, F. H., et al. 2018, *MNRAS*, **473**, 1879
- Rhee, J., Lah, P., Chengalur, J. N., Briggs, F. H., & Colless, M. 2016, *MNRAS*, **460**, 2675
- Rhee, J., Zwaan, M. A., Briggs, F. H., et al. 2013, *MNRAS*, **435**, 2693
- Sánchez Almeida, J., Elmegreen, B. G., Muñoz-Tuñón, C., & Elmegreen, D. M. 2014, *A&ARv*, **22**, 71
- Sancisi, R., Fraternali, F., Oosterloo, T., & van der Hulst, T. 2008, *A&ARv*, **15**, 189
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Smolčić, V., Delvecchio, I., Zamorani, G., et al. 2017, *A&A*, **602**, A2
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, **214**, 15
- Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, *ApJ*, **853**, 179
- Verheijen, M., van Gorkom, J. H., Szomoru, A., et al. 2007, *ApJL*, **668**, L9
- Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, *ApJ*, **866**, 135
- Wang, J., Koribalski, B. S., Serra, P., et al. 2016, *MNRAS*, **460**, 2143
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, **258**, 11
- York, D. G., Adelman, J. A., John, E. J., et al. 2000, *AJ*, **120**, 1579
- Zwaan, M. A. 2000, PhD thesis, Rijksuniversiteit Groningen
- Zwaan, M. A., van Dokkum, P. G., & Verheijen, M. A. W. 2001, *Sci*, **293**, 1800