Research in Astronomy and Astrophysics, 25:105010 (15pp), 2025 October
https://doi.org/10.1088/1674-4527/adf30b
CSTR: 32081.14.RAA.adf30b
© 2025. National Astronomical Observatories, CAS and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved. Printed in China.

A Halo-based Empirical Model for the Atomic Hydrogen Content

Zhixing Li[1], Hong Guo[2], and Yi Mao[1]
[1] Department of Astronomy, Tsinghua University, Beijing 100084, China; ymao@tsinghua.edu.cn
[2] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China; guohong@shao.ac.cn
Received 2025 April 22; revised 2025 July 3; accepted 2025 July 18; published 2025 September 3

Abstract

Atomic hydrogen (H I) gas, mostly residing in dark matter halos after cosmic reionization, is the fuel for star formation. Its relation to the properties of the host halo is the key to understand the cosmic H I distribution and to mock H I large scale structure accurately and conveniently for the large area H I intensity mapping surveys in the future. In this work, we propose a flexible empirical model of H I–halo relation. In this model, while the H I mass depends primarily on the mass of host halo, there is also secondary dependence on other halo properties. We apply our model to the observation data of the Arecibo Fast Legacy ALFA (ALFALFA) survey, and find it can successfully fit the cosmic H I abundance ($\Omega_{\rm H\,I}$), average H I–halo mass relation $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, and the H I clustering. The best fit of the ALFALFA data rejects the model with no secondary halo dependence of H I and the model with secondary dependence on the halo spin parameter (λ), and shows a strong dependence on halo formation time ($a_{1/2}$) and halo concentration ($c_{\rm vir}$). In an attempt to explain these findings from the perspective of hydrodynamical simulations, the IllustrisTNG simulation confirms the dependence of H I mass on secondary halo parameters. However, the IllustrisTNG results show a strong dependence on λ and a weak dependence on $c_{\rm vir}$ and $a_{1/2}$, and also predict a much larger value of H I clustering on large scales than observations. This interesting difference between simulation and observation calls for improvements in understanding the H I–halo relation from both theoretical and observational sides.

Key words: (cosmology:) large-scale structure of universe – radio lines: galaxies – ISM: clouds

1. Introduction

Atomic hydrogen (H I) gas is the fuel for the early formation of stars and galaxies. The H I gas is harbored mainly in the cold and dense regions of galaxies and dark matter halos after cosmic reionization, where it is self-shielded from the ultraviolet (UV) background with high recombination rates (see e.g., Krumholz et al. 2009; Diemer et al. 2018; Villaescusa-Navarro et al. 2018), making the H I emission a biased tracer of the dark matter distribution in the universe (Alonso & Ferreira 2015; Fialkov et al. 2020; Wang et al. 2021).

The 21 cm radiation originating from the hyperfine transition of atomic hydrogen serves as an innovative tool to study the large-scale structure of the universe (Abdalla & Rawlings 2005; Furlanetto et al. 2006). Given the large number of undetected H I sources in extragalactic surveys like HIPASS (Barnes et al. 2001) and Arecibo Fast Legacy ALFA (ALFALFA; Haynes et al. 2018), intensity mapping technology has been suggested and utilized in numerous 21 cm surveys as an effective method for investigating the large-scale structure without signal loss or the need to resolve individual sources. This approach allows for a shorter integration time and lower angular resolution to achieve reliable statistics on a large scale (see, for example, Bull et al. 2015). Based on this technology, several groundbreaking experimental 21 cm intensity mapping surveys have been carried out (Masui et al. 2013; Switzer et al. 2013; Wolz et al. 2022; Cunnington et al. 2023). Recently, Meerklass Collaboration et al. (2025) announced the first detection of the auto-power spectrum of H I signal with MeerKAT single-dish intensity maps at cosmological scales ($k < 0.3\,h\,{\rm Mpc}^{-1}$). The ongoing and next generation of surveys and telescopes, such as the BAO from Integrated Neutral Gas Observations (Battye et al. 2012), the Canadian Hydrogen Intensity Mapping Experiment (Newburgh et al. 2014), and especially the Square Kilometre Array (SKA; Dewdney et al. 2009; Square Kilometre Array Cosmology Science Working Group et al. 2020), are around the corner, in preparation for the detection of a 21 cm signal in the post-reionization universe.

To help form theoretical reference and generate realistic mocks for forthcoming surveys (Bagla et al. 2010; Villaescusa-Navarro et al. 2014; Seehars et al. 2016), it is urgent and necessary to find an efficient and accurate method to obtain H I large-scale structures. Hydrodynamic simulations are computationally expensive, so it is difficult to have large simulation boxes (e.g., see Villaescusa-Navarro et al. 2018). Semi-analytical models (SAMs) are faster but still require galaxy information or the need to build the galaxy catalog in the first place (Wang et al. 2012; Xie et al. 2017; Baugh et al. 2019; Chauhan et al. 2020;

Spinelli et al. 2020), and even the empirical framework of galactic-H I structure proposed by Guo et al. (2023) still requires the galaxy information like the star formation rate. Following the idea of the Halo Occupation Distribution (HOD), generating H I structures on top of *N*-body simulations is the most straightforward method. Large boxes of these dark matter-only simulations and simple scaling relation between H I and halo mass make it easy to make forecasts at cosmological scales. Therefore, it is of our great concern to grasp the connection between the H I gas and the mass of dark matter halos from various simulations and observations.

Based on hydrodynamical simulations and SAMs utilizing different galaxy formation models, several empirical formulae have been proposed to characterize this relation (see, for example, Villaescusa-Navarro et al. 2018; Baugh et al. 2019; Obuljen et al. 2019; Spinelli et al. 2020; Guo et al. 2023), although they are typically customized to match specific simulations or SAMs. There are similarities and debates among these studies. It is commonly agreed that as the halo mass increases, the H I mass increases rapidly for smaller halos due to cold accretion, while it increases more slowly with halo mergers for larger halos. Differences arise in the intermediate range of the relation. For halos within the range of $[10^{11}, 10^{13}]\,h^{-1}M_{\odot}$, it is generally assumed that H I mass increases steadily (Bagla et al. 2010; Villaescusa-Navarro et al. 2018), while recent research suggests a slight decrease in H I mass due to active galactic nucleus (AGN) feedback or virial shock-heating (Baugh et al. 2019; Chauhan et al. 2020; Zhang et al. 2022). The direct observational results conducted by Guo et al. (2020) (hereafter G20), which measures the H I–halo mass relation at $z \simeq 0$ by stacking the total H I signals from the ALFALFA (Giovanelli et al. 2005) survey for halos of varying masses derived from the galaxy group catalog (Lim et al. 2017), agree more with the later opinion.

Although previous work has shown that the total H I mass in a dark-matter halo is predominantly dictated by the halo mass (Villaescusa-Navarro et al. 2018; Baugh et al. 2019; Chauhan et al. 2020; Spinelli et al. 2020), the large scatter between the H I mass and the halo mass indicates that the H I mass is also affected by other baryonic processes and halo properties. To be more specific, the H I gas mass in and around a specific galaxy is influenced by processes such as star formation and various feedback mechanisms in the galaxy's baryon cycle, such as gas expulsion triggered by stellar and AGN feedback, as well as gas accretion from the intergalactic medium (e.g., see Tumlinson et al. 2017). All these baryonic processes muddle the relation between halo and H I mass.

One solution to reduce this scatter is to find some secondary dependence between H I and halo properties other than the halo mass. Previous works have noticed the secondary dependence between H I and the halo properties such as the spin parameter and the formation time (e.g., see Lagos et al. 2018; Chauhan et al. 2020; Spinelli et al. 2020). From observational results in G20, it is also discovered that the H I mass does not simply increase with the halo mass but also depends significantly on the halo richness, defined as the count of member galaxies by G20. A distinct peak is observed around $M_{\rm h} \sim 10^{11.5}\,h^{-1}M_{\odot}$, which is not attributed to the H I stacking errors, as this feature becomes more pronounced for halos with higher richness. This aligns with the results of Guo et al. (2017) (referred to as G17), indicating that H I-rich galaxies are more likely to be situated in halos with late formation times, based on spatial clustering analyses, as validated by Stiskalek et al. (2021).

However, the secondary dependence is not widely applied in the model of the relation. Several formulae have been proposed to describe this relation (e.g., see Bagla et al. 2010; Villaescusa-Navarro et al. 2014, 2018; Chauhan et al. 2020). Most of them use the halo mass as the only variable. However, the low H I bias measured by this work, as well as in G17, highlights the need to introduce the secondary dependence into the relation. A better H I–halo mass relation is the key to describe the large scale structures statistics, but the measured H I bias can reversely help to determine the relation and distinguish between different models of H I formations and H I–halo connection.

In this work, we present the measurement of H I clustering and H I bias using ALFALFA data. Then we propose an empirical formula for the H I–halo mass relation with secondary dependence of H I on halo properties and use three measured statistics measured from ALFALFA to constrain our formula. To validate the fitting results obtained in observation, we use a hydrodynamical simulation to further check the secondary dependence between H I and halo properties.

The rest of this paper is organized as follows. We introduce the observational measurements and our empirical model in Section 2, and present the results of fitting our model to the observation data in Section 3. These results are compared with the hydrodynamic simulations and the previous works in Section 4. We make the concluding remarks in Section 6. We leave technical details to Appendix A (on fitting prior and covariance matrix) and Appendix B (on model selection).

2. Measurements and Methodology

2.1. Observational Measurements

We use three sets of measurements to constrain the H I–halo mass relation—the average H I–halo mass relation ($\langle M_{\rm H\,I}|M_{\rm h}\rangle$), the cosmic H I abundance ($\Omega_{\rm H\,I}$) and the H I clustering measurements ($w_{\rm p}(r_{\rm p})$ and $\xi(s)$).

2.1.1. H I–Halo Mass Relation

In this work, we use the H I–halo mass relation $\langle M_{\rm H\,I}M_{\rm h}\rangle$ measured by G20 within the halo mass range of 10^{11}–$10^{14}\,h^{-1}M_{\odot}$, with a bin size of $\Delta \log M_{\rm h} = 0.25$

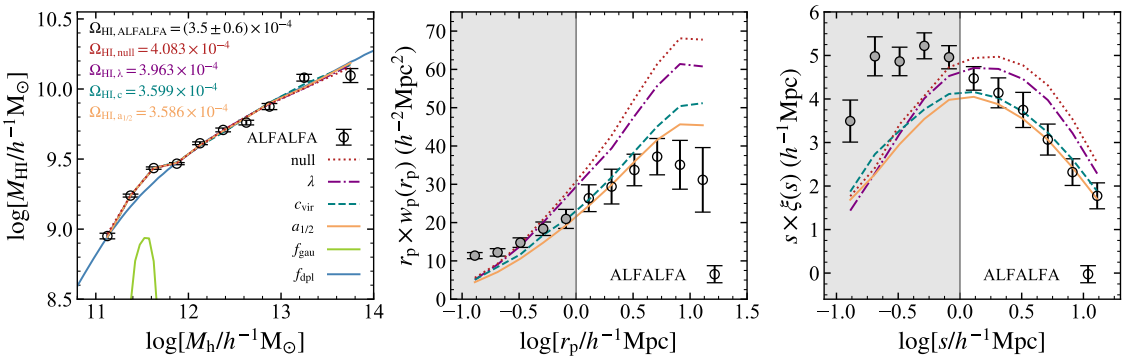


Figure 1. Best-fitting models to the observational results of ALFALFA including H I–halo mass average relation $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ (left panel), projected correlation function $w_{\rm p}(r_{\rm p})$ (middle panel) and correlation function $\xi(s)$ (right panel). Black circles represent the observational data used in our fitting with error bars estimated using the jackknife method, while colored lines signify the predictions using Equation (10), including halo spin parameter ("λ"), the halo concentration ("$c_{\rm vir}$"), and halo formation time ("$a_{1/2}$"). In particular, the red dotted line represents the fitting results without secondary dependence of the halo parameters according to Equations (7)–(9) ("null"). In the left panel, the light green line is the analytical Gaussian-like function ("$f_{\rm gau}$"), and the dark blue line is the double power law ("$f_{\rm dpl}$"). Data points in the gray-shaded regions in the middle and right panels are not used in fitting, because these regions signify the fine structures of galaxies where H I gas cannot be assigned simply as a function of halo mass.

(represented as open circles in the left panel of Figure 1). To account for the finite bin size of the halo mass in the observational data, the average H I mass within the halo mass bins defined by $M_{\rm h1} < M_{\rm h} < M_{\rm h2}$ is modeled as follows,

$$\langle M_{\rm H\,I}|M_{\rm h}\rangle_{\rm obs}\ |_{M_{\rm h1}}^{M_{\rm h2}} = \frac{\int_{M_{\rm h1}}^{M_{\rm h2}} \langle M_{\rm H\,I}|M_{\rm h}\rangle\, n(M_{\rm h})\, dM_{\rm h}}{\int_{M_{\rm h1}}^{M_{\rm h2}} n(M_{\rm h})\, dM_{\rm h}}, \quad (1)$$

where $n(M_{\rm h})$ represents the halo mass function. Using large halo mass bin sizes tends to smooth out the peak feature of the H I–halo mass relation at the lower mass end, particularly for halos of high richness (refer to Figure 2 of G20).

It should be noted that the lower limit of the halo mass in the observed H I–halo mass relation is about $10^{11}\,h^{-1}M_\odot$. According to G20, the fraction of H I mass contributed by halos below this limit is estimated to be approximately 30%, so there is no reason to ignore this part. Therefore, the behavior of $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ at lower masses is individually constrained by the H I abundance $\Omega_{\rm H\,I}$ mentioned in the following section.

2.1.2. Cosmic H I Abundance

The dimensionless cosmic H I abundance, $\Omega_{\rm H\,I}$, can be theoretically determined from $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ and the halo mass function as,

$$\Omega_{\rm H\,I} = \frac{1}{\rho_{\rm c}} \int \langle M_{\rm H\,I}|M_{\rm h}\rangle\, n(M_{\rm h})\, dM_{\rm h}, \quad (2)$$

where $\rho_{\rm c}$ is the critical density.

$\Omega_{\rm H\,I}$ describes the total amount of H I in the universe. We adopt the value of $\Omega_{\rm H\,I} = (4.55 \pm 0.29) \times 10^{-4}$ from Guo et al. (2023). It was obtained from the ALFALFA survey 100% sample (α.100) (Haynes et al. 2018) by integrating the Schechter function fitting to the measured H I mass function of α.100. We refer the readers to Jones et al. (2018) and Guo et al. (2023) for more details about the Schechter function and the measurement of H I abundance.

2.1.3. H I Clustering

Following G17, we measure the monopole and projected 2-Point Correlation Function (2PCF) using the ALFALFA 100 release data. It is indicated by previous works that the flux-limited feature in such extragalactic H I surveys will introduce a bias into the measurements (Papastergis et al. 2013; Guo et al. 2017; Obuljen et al. 2019). To correct this bias, we adopt the volume correction method proposed by Xu et al. (2016) and effective volume ($V_{\rm eff}$) (for more details, see the appendix of Martin et al. 2010) as our weighting scheme.

We employ the Landy–Szalay estimator (Landy & Szalay 1993) weighted by H I mass to evaluate the projected 2PCF $\xi(r_{\rm p}, r_\pi)$ in the redshift space. Here, r_π and $r_{\rm p}$ represent the separations of galaxy pairs along and perpendicular to the line-of-sight, respectively. The expression for $\xi(r_{\rm p}, r_\pi)$ is given by $\xi(r_{\rm p}, r_\pi) = ({\rm DD} - 2{\rm DR} + {\rm RR})/{\rm RR}$. The counts of galaxy pairs for data-data (DD), data-random (DR), and random-random (RR) are computed as follows,

$${\rm DD}(r_{\rm p}, r_\pi) = \sum_{(i,j)\in V_{ij}} \frac{M_{{\rm H\,I},i} M_{{\rm H\,I},j}}{V_{ij}}, \quad (3)$$

$${\rm DR}(r_{\rm p}, r_\pi) = \sum_{(i,j)\in V_{ij}} \frac{M_{{\rm H\,I},i} M_{{\rm H\,I},j}}{V_{ij}}, \quad (4)$$

$${\rm RR}(r_{\rm p}, r_\pi) = \sum_{(i,j)\in V_{ij}} \frac{M_{{\rm H\,I},i} M_{{\rm H\,I},j}}{V_{ij}}, \quad (5)$$

where $V_{ij} = \min(V_{\rm eff,i}, V_{\rm eff,j})$, with $V_{\rm eff,i}$ and $V_{\rm eff,j}$ being the effective volumes accessible to the ith and jth galaxies, respectively.

The projected 2PCF $w_{\rm p}(r_{\rm p})$ is then given by

$$w_{\rm p}(r_{\rm p}) = 2\int_0^{r_{\pi,\rm max}} \xi(r_{\rm p}, r_\pi)\, dr_\pi, \tag{6}$$

where $r_{\pi,\rm max}$ is fixed at $20\,h^{-1}$Mpc as in G17. Similarly, we also compute the monopole $\xi(s)$, with $s = \sqrt{r_{\rm p}^2 + r_\pi^2}$ being the separation between galaxy pairs in the redshift space.

Our galaxy sample comes from the 100% complete catalog of the ALFALFA survey (Haynes et al. 2018). The final sample includes 22330 galaxies with reliable H I mass measurements, covering 6518 deg^2 in the redshift range of $0.0025 < z < 0.06$. The random catalog is constructed similarly as in G17, with the weights assigned by randomly selecting from the full H I galaxy sample, and the distribution of redshifts and H I mass is kept the same as those of the real data. In our measurement, the logarithmic $r_{\rm p}$ bins are a constant width $\Delta \log r_{\rm p} = 0.2$ ranging from 0.13 to 12.92 h^{-1}Mpc (the same for s bins) and linear r_π bins of width $\Delta r_\pi = 2\,h^{-1}$Mpc from 0 to $20\,h^{-1}$Mpc. The error covariance matrices are estimated using the jackknife re-sampling method of 100 subsamples. The measurements of $w_{\rm p}(r_{\rm p})$ and $\xi(s)$ are shown as the open circles with error bars in the middle and right panels of Figure 1, respectively. In this work, we only use the H I clustering measurements above $1\,h^{-1}$Mpc (the white area in the figure) to avoid the scales dominated by the one-halo contribution.

In summary, for observational data, we have 1 data point in $\Omega_{\rm H\,I}$, 10 data points in $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, 6 data points in $w_{\rm p}(r_{\rm p})$ and 6 data points in $\xi(s)$. The total number of data points is 23.

2.2. A Halo-based Empirical Model

2.2.1. H I–Halo Mass Relation

The observed $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ (shown in the left panel Figure 1) is not a gradually increasing function of $M_{\rm h}$. Instead, it shows a noticeable bump feature around $M_{\rm h} \sim 10^{11.5}\,h^{-1}M_\odot$. This feature becomes more pronounced for halos with higher richness (more satellites within one halo). As outlined in G20, this bump is possibly due to the effects of virial shock heating and AGN feedback that occur in halos of approximately this mass range. Consequently, we suggest a flexible two-component model for $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, combining a log-normal distribution ($f_{\rm gau}$) at lower masses with a double power law distribution (referred to as DPL) $f_{\rm dpl}$, as follows.

$$f(M_{\rm H\,I}|M_{\rm h}) = f_{\rm gau}(M_{\rm h}) + f_{\rm dpl}(M_{\rm h}), \tag{7}$$

$$f_{\rm gau} = 10^{A_{\rm gau}\exp[-(\log(M_{\rm h}/M_{\rm gau})/\sigma_{\rm gau})^2]}, \tag{8}$$

$$f_{\rm dpl} = A_{\rm dpl} M_{\rm h}/[(M_{\rm h}/M_{\rm dpl})^{-\alpha} + (M_{\rm h}/M_{\rm dpl})^{\beta}], \tag{9}$$

where $A_{\rm gau}$, $M_{\rm gau}$, $\sigma_{\rm gau}$, $A_{\rm dpl}$, $M_{\rm dpl}$, α, and β are the model parameters. In particular, $A_{\rm gau}$, $M_{\rm gau}$, $\sigma_{\rm gau}$ are the parameters that determine the amplitude, position, and width of the bump feature (Gaussian component) respectively. Meanwhile, $A_{\rm dpl}$, $M_{\rm dpl}$ are used to determine the amplitude and transition point of the double power law, while α and β determine the slopes of the H I–halo mass relation at the low and high mass ends, respectively. Parameters except α can be properly constrained by the measurements of $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ in G20. The α parameter is individually constrained by the observational measurement of $\Omega_{\rm H\,I}$ due to the range of halo mass in the measured $\langle M_{\rm H\,I}|M_{\rm h}\rangle$. Such a model is flexible enough to fit the observational measurements, as well as the simulation predictions.

Then the additional dependence of $M_{\rm H\,I}$ on the halo assembly history is added to our model to account for the observed clustering measurements. Following G17, we investigate the secondary dependence of H I on three halo properties in this paper—formation time ($a_{1/2}$), concentration parameter ($c_{\rm vir}$) and spin parameter (λ). The halo formation time is defined as the epoch in which the halo reaches half of its peak mass over the entire history. $a_{1/2}$ is the scale factor corresponding to the halo formation time, i.e., $a_{1/2} = 1/(1 + z_{1/2})$. The concentration $c_{\rm vir}$ is the ratio between the halo virial radius and scale radius. The definition of spin parameter λ follows that of Peebles (1969). Our empirical model for the H I–halo relation is then formulated as

$$\langle M_{\rm H\,I}|M_{\rm h}, P\rangle = f(M_{\rm H\,I}|M_{\rm h}) \cdot P^{\gamma}, \tag{10}$$

where γ is the free model parameter and the parameter P represents the halo properties, either $a_{1/2}$, $c_{\rm vir}$ or λ. The γ parameter is constrained by the H I clustering measurements, as inspired by G17.

2.2.2. H I Clustering Model

Following the strategy of G17, we populate H I to the dark matter halos in the N-body simulations using Equation (10). $\Omega_{\rm H\,I}$ is obtained by summing up the H I mass in the whole simulation volume.

We use the dark matter halo catalog from the Small MultiDark simulation of Planck cosmology (SMDPL; Klypin et al. 2016), covering a volume of $400^3\,h^{-3}$Mpc^3 and assuming the cosmological parameters of $\Omega_m = 0.307$, $\Omega_b = 0.048$, $n_s = 0.96$, $h = 0.678$ and $\sigma_8 = 0.823$. The particle mass resolution is $9.6 \times 10^7\,h^{-1}M_\odot$, which is good enough to fully resolve the host halos of H I-rich galaxies. The halos are identified by the ROCKSTAR halo finder (Behroozi et al. 2013) and we only apply our model to the simulation output at $z = 0$.

For fair comparison, we construct mocks with the same geometry as the ALFALFA survey. To avoid the sample variance, there are 64 mock galaxy catalogs in total. We calculate the clustering measurements of each mock (with the RSD effect automatically included using halo peculiar

velocities), and use the average of the 64 mocks as our model prediction.

2.2.3. Model Fitting

To find the best-fit model parameters to the observational data described in Section 2.1, we apply the Markov Chain Monte Carlo (MCMC) technique, using the Bayesian inference tool `MultiNest` (Feroz et al. 2009). The likelihood surface is determined by χ^2,

$$\chi^2 = \chi_1^2 + \chi_2^2 + \chi_3^2, \tag{11}$$

$$\chi_1^2 = (\Omega_{\rm H\,I} - \Omega_{\rm H\,I}^*)^2/\sigma_{\Omega_{\rm H\,I}}^2, \tag{12}$$

$$\chi_2^2 = (\langle M_{\rm H\,I}|M_{\rm h}\rangle - \langle M_{\rm H\,I}|M_{\rm h}\rangle^*)^2/\sigma_{\langle M_{\rm H\,I}|M_{\rm h}\rangle}^2, \tag{13}$$

$$\chi_3^2 = (\boldsymbol{\xi}_{\rm all} - \boldsymbol{\xi}_{\rm all}^*)^T \boldsymbol{C}^{-1} (\boldsymbol{\xi}_{\rm all} - \boldsymbol{\xi}_{\rm all}^*), \tag{14}$$

where the data vector $\boldsymbol{\xi}_{\rm all} = [w_{\rm p}(r_{\rm p}), \xi(s)]$ (i.e., the combination of $w_{\rm p}(r_{\rm p})$ and $\xi(s)$), and $\boldsymbol{C}$ is the full error covariance matrix for $\xi_{\rm all}$. The quantity with (without) a superscript "*" is the one from the measurement (model). The cross-correlation between $w_{\rm p}(r_{\rm p})$ and $\xi(s)$ has been fully considered in $\boldsymbol{C}$. The corresponding correlation matrix is shown in Figure A1. We assume that the measurements of $\Omega_{\rm H\,I}$, $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, and $\xi_{\rm all}$ are independent of each other. Also, note that error $\sigma_{\langle M_{\rm H\,I}|M_{\rm h}\rangle}$ was estimated for the mean of H I mass for a given halo mass bin using the jackknife method and therefore was smaller than the error that would be estimated as the standard deviation of the total H I sample. With 23 data points and 8 model parameters, the final degrees of freedom (dof) for the model fitting is 15.

3. Results

3.1. Best-fit Models

To test the effect of halo assembly bias, we first fit the observational measurements without including any dependence on the secondary halo parameters (that is, $\gamma = 0$ in Equation (10)). The best-fitting model is shown as the red dotted lines in Figure 1 (indicated as "null"), with χ^2/dof of 52.74/16. (Note that in this case there are 7 free model parameters, and therefore the dof is 16.) Although it fits the relation of $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ well with the adopted functional form of Equation (7), the clustering measurements of $w_{\rm p}(r_{\rm p})$ and $\xi(s)$ are significantly overestimated. The predicted value of $\Omega_{\rm H\,I} = 4.08 \times 10^{-4}$ is also overestimated by 1σ, as indicated at the top of the left panel. As Reduced χ^2 (hereafter $\chi^2_{\rm re}$) is $52.74/16 \approx 3.30$, this model can be strongly rejected. Thus, it strongly supports the need to include the effect of halo assembly bias (Gao et al. 2005; Zhu et al. 2006; Wang et al. 2013a, 2013b) in the modeling of the H I content in halos, confirming the results of G17 and G20.

The best-fitting models for including λ, $c_{\rm vir}$, and $a_{1/2}$ are shown as dot–dashed, dashed, and solid lines in Figure 1, respectively. They are indistinguishable from each other in the fit to $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, as the average of H I–halo mass relation can always be fitted well with halo mass as the only variable. Large differences appear in the clustering prediction of $w_{\rm p}(r_{\rm p})$ and $\xi(s)$ by different models. "null"-model overestimates the H I clustering, as does the λ-model. The $a_{1/2}$ and $c_{\rm vir}$ models agree with each other, with the $a_{1/2}$ model showing slightly better agreement with the observation. These fitting results are fully consistent with those of G17. In G17, an extended abundance matching model by including λ as the secondary halo parameter also overestimates large-scale clustering measurements. The empirical relation for $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ proposed in this paper provides a more quantitative description between the H I mass and the halo mass.

To quantify the model selection, we introduce the reduced chi-square $\chi^2_{\rm re}$ and the evidence to evaluate the goodness of fit (GOF) (see more details in Appendix B). Both methods prefer the $a_{1/2}$ ($\chi^2_{\rm re} = 1.82$) and $c_{\rm vir}$ ($\chi^2_{\rm re} = 1.94$) models. The best-fitting $\chi^2_{\rm re}$ of λ-model is about 2.83, not much better than the "null"-model ($\chi^2_{\rm re} = 3.30$).

For the two components ($f_{\rm gau}$ and $f_{\rm dpl}$ in Equation (7)) in the proposed function $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, the individual contributions of each component to the best-fitting $a_{1/2}$ model are shown as the green and blue solid lines in the left panel of Figure 1, respectively. Most H I mass is contributed by the component of $f_{\rm dpl}$ and $f_{\rm gau}$ acts as a sharp peak around the halo mass of $10^{11.5}\,h^{-1}M_\odot$. As mentioned in G20, we believe that $M_{\rm gau}$ encodes the information about the physical processes involved, including virial shock-heating and AGN feedback for halos more massive than $M_{\rm gau}$.

It should be emphasized that our division of $f_{\rm gau}$ and $f_{\rm dpl}$ is different from the division of H I in central and satellite galaxies. $f_{\rm dpl}$ in our model dominates at all mass scales, while the total H I mass is dominated by central galaxies with $M_{\rm h} < 10^{13}\,h^{-1}M_\odot$ as shown in G20. It is tricky to model the contributions from central and satellite galaxies separately, since additional baryonic information such as star formation and feedback is also needed (Guo et al. 2021, 2023).

We can further quantify the scale-dependence of the clustering measurements using the H I bias, $b_{\rm H\,I}(r_{\rm p})$, defined as,

$$b_{\rm H\,I}(r_{\rm p}) = \sqrt{\frac{w_{\rm p}(r_{\rm p})}{w_{\rm p,dm}(r_{\rm p})}}, \tag{15}$$

where $w_{\rm p,dm}(r_{\rm p})$ is the projected 2PCF of the dark matter particle. It is measured in the same way as $w_{\rm p}$ by including the effect of RSD and integral constraint (e.g., see Guo et al. 2017). The resulting $b_{\rm H\,I}(r_{\rm p})$ measurements are shown in Figure 2. The best-fitting $a_{1/2}$-model has a large-scale H I bias of $b_{\rm H\,I} \sim 0.7$, while the "null"-model has a significantly larger bias of $b_{\rm H\,I}(r) \sim 0.87$. The observed value of H I bias is much smaller than the "null"-model prediction, which indicates that the neutral hydrogen at $z \simeq 0$ is distributed in the low-mass halos as well as late-forming (or low-concentration) halos.

Table 1
Best-fit Model Parameters[a] for Observational Measurements

P[b]	χ^2/dof[c]	$A_{\rm gau}$	$\log M_{\rm gau}$	$\sigma_{\rm gau}$	$\log A_{\rm dpl}$	$\log M_{\rm dpl}$	α	β	γ
null	52.74/16	$9.17^{+0.03}_{-0.27}$	$11.52^{+0.03}_{-0.00}$	$0.38^{+0.38}_{-0.01}$	$-1.93^{+0.06}_{-0.05}$	$11.45^{+0.11}_{-0.21}$	$0.09^{+0.17}_{-0.06}$	$0.71^{+0.02}_{-0.05}$	0
λ	42.41/15	$5.51^{+1.23}_{-0.68}$	$11.51^{+0.02}_{-0.01}$	$0.34^{+0.25}_{-0.07}$	$-5.64^{+1.34}_{-0.62}$	$11.66^{+0.08}_{-0.24}$	$0.01^{+0.16}_{-0.06}$	$0.81^{+0.02}_{-0.07}$	$-2.21^{+0.73}_{-0.34}$
$c_{\rm vir}$	29.08/15	$11.98^{-0.05}_{-0.84}$	$11.51^{+0.05}_{-0.03}$	$0.51^{+0.38}_{-0.08}$	$1.17^{+0.12}_{-0.79}$	$11.18^{+0.11}_{-0.08}$	$0.41^{+0.39}_{-0.26}$	$1.00^{+0.02}_{-0.09}$	$-3.12^{+0.86}_{-0.12}$
$a_{1/2}$	27.23/15	$10.60^{+0.30}_{-0.36}$	$11.51^{+0.03}_{-0.01}$	$0.45^{+0.32}_{-0.09}$	$-0.34^{+0.42}_{-0.24}$	$11.23^{+0.05}_{-0.15}$	$0.24^{+0.50}_{-0.15}$	$0.88^{+0.03}_{-0.05}$	$5.60^{+2.26}_{-1.12}$

Notes.
[a] The parameters $M_{\rm gau}$ and $M_{\rm dpl}$ are in units of $h^{-1}M_{\odot}$. Other parameters are unitless.
[b] The first model ("null") assumes no secondary dependence of halo parameters, i.e., γ is set to zero. Other models include secondary dependence of H I mass on halo spin parameter ("λ"), halo concentration ("$c_{\rm vir}$"), and halo formation time ("$a_{1/2}$"). The $a_{1/2}$-model shows the best agreement with the observation data.
[c] Note that the error for $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ was estimated for the mean of H I mass for a given halo mass bin using the jackknife method. Therefore χ^2 here is overestimated compared to the value it would be if the error would be estimated as the standard deviation of the sample of H I mass at individual halos.

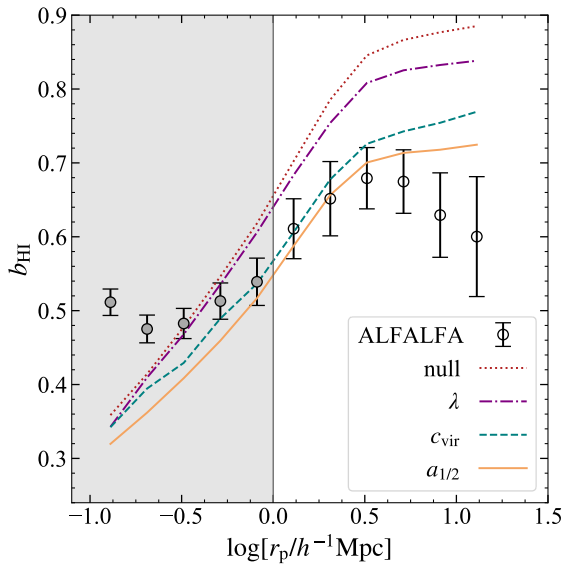


Figure 2. The H I bias $b_{\rm H\,I}(r)$ from the best-fitting models (with the same colors and line types as in Figure 1), in comparison with the observational results (black circles). Data points at $r_{\rm p} < 1\,h^{-1}$Mpc, as shown in the gray shaded region, are not used, due to small scale effects.

The H I bias can also be theoretically calculated and decomposed into contributions from different halo populations using halo models, as shown in Equation (16) (e.g., Villaescusa-Navarro et al. 2018; Chen et al. 2021), where all biases are assumed to be constant to reflect large-scale behavior.

$$b_{\rm HI} = \frac{\int dm\, n(m) b(m) \langle M_{\rm H\,I}(m)\rangle}{\int dm\, n(m) \langle M_{\rm H\,I}(m)\rangle}, \tag{16}$$

where $n(m)$ is the halo mass function, $\langle M_{\rm H\,I}(m)\rangle$ is the average H I–halo mass relation and $b(m)$ represents the halo bias dependent on the halo mass. With halo bias and mass function known, we can calculate H I bias using the average H I–halo mass relation (e.g., see Chen et al. 2021). However, the scatter of the H I–halo mass relation is so large that the distribution and the dependence on halo properties of H I can be substantially different within the same narrow halo mass bin. Figures 1 and 2 demonstrate that even with the same average H I–halo mass relation, the H I bias can vary considerably as long as secondary halo parameter is included. Therefore, it is advisable to include the secondary dependence of H I on halo properties into the equation and the H I bias can be described as Equation (17).

$$b_{\rm HI} = \frac{\int dm\, n(m,p) b(m,p) \langle M_{\rm H\,I}(m,p)\rangle}{\int dm\, n(m,p) \langle M_{\rm H\,I}(m,p)\rangle}, \tag{17}$$

where the p is the halo property like λ, $c_{\rm vir}$ or $a_{1/2}$.

Obuljen et al. (2019) measure the H I bias as $b_{\rm H\,I} = 0.878^{+0.022}_{-0.023}$, which is consistent with our null model here. If this were the case, there would be no need to introduce any secondary dependencies in the H I–halo mass model to reproduce the statistics of the H I large-scale structure. However, our next work in preparation indicates that the FKP weighting applied by their work leads to a biased and, more specifically, larger measurement of H I bias. Thus, we maintain that secondary dependence is a necessity that has been generally overlooked in previous works and models.

We show the best-fit parameters for different models in Table 1. Due to the observed low H I bias, the best-fitting model prefers to populate halos with lower bias values. As shown in Figure 4 of Sato-Polito et al. (2019), for halos of $M_{\rm h} < 10^{13}\,h^{-1}M_{\odot}$, the late-forming and low-$c_{\rm vir}$ halos would have lower bias relative to their early-forming and high-$c_{\rm vir}$ counterparts. Therefore, the best-fit value of γ in Table 1 is positive for the $a_{1/2}$-model and negative for the $c_{\rm vir}$-model.

The situation for the λ-model is slightly more complicated, since the relative bias between the high-λ and low-λ halos would switch at around $10^{11.5}\,h^{-1}M_{\odot}$. Given that our model only assumes a constant value of γ for the whole halo mass range, the negative value of best-fit γ for the λ-model indicates that the neutral hydrogen prefers to populate the low-spin

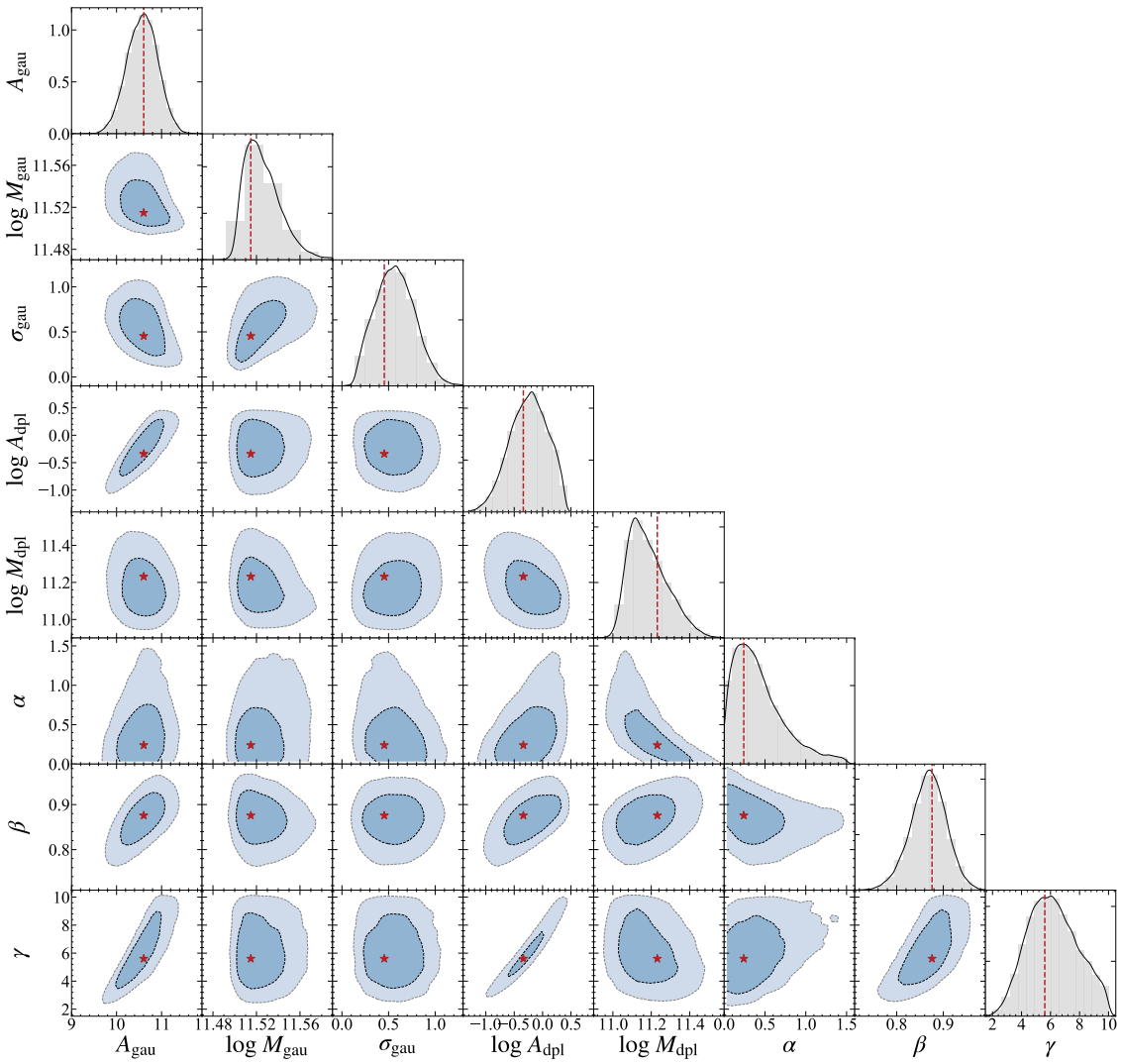


Figure 3. Posterior parameter distributions for the best-fit $a_{1/2}$-model that agrees best with the observations. We show the probability distribution function (PDF) of each parameter in gray histograms, and the correlation between different parameters with 68% (95%) confidence level in blue (light-blue) contours. The best-fit models are marked as red stars in contours and vertical red dashed lines in histograms, inferred by the maximum likelihood estimation (MLE). Note that "parameters" here refer to the coefficients in Equations (7)–(10), and should not be confused with the aforementioned "secondary parameters."

halos, which have lower bias for $M_{\rm h} > 10^{11.5}\,h^{-1}M_{\odot}$. However, the relative difference of $b_{\rm H\,I}$ between the null model and the $a_{1/2}$-model value is as large as $0.87/0.7–1 \approx 25\%$. For halos with $M_{\rm h} < 10^{12}\,h^{-1}M_{\odot}$ (which encompass most of the cosmic H I), the relative bias between the low-λ (or high-λ) halos and the ensemble average at the same mass is less than 10% (Sato-Polito et al. 2019). Thus, even if an $M_{\rm h}$-dependent γ parameter would be introduced to the λ-model, $b_{\rm H\,I}$ would not be reduced by 25% in the null model. As seen in Figure 2, the λ-model only has a slightly decreased large-scale bias compared to the null model. We will further discuss the effect of halo spin in the next section.

We show the posterior parameter distributions for the $a_{1/2}$-model in Figure 3. Most correlations between different parameters are weak, which implies that these parameters are mostly independent of each other. However, there are strong correlations between γ and $A_{\rm gau}$, and between γ and $A_{\rm dpl}$, as expected due to Equation (10).

Since γ can be constrained from the clustering measurements, these correlations imply that the H I–halo mass distribution $f(M_{\rm H\,I}|M_{\rm h})$ is not independent of the halo formation time distribution. This is also confirmed by the fact that for all coefficients (except for $M_{\rm gau}$) in Table 1, their values vary significantly among different models. It means that even if the observed $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ is well fit by different models, their H I mass distributions in each halo mass bin vary from each other. A similar effect is found in some HOD models, where the specific treatment of scatter in the stellar mass–halo mass relation significantly influences the predicted galaxy

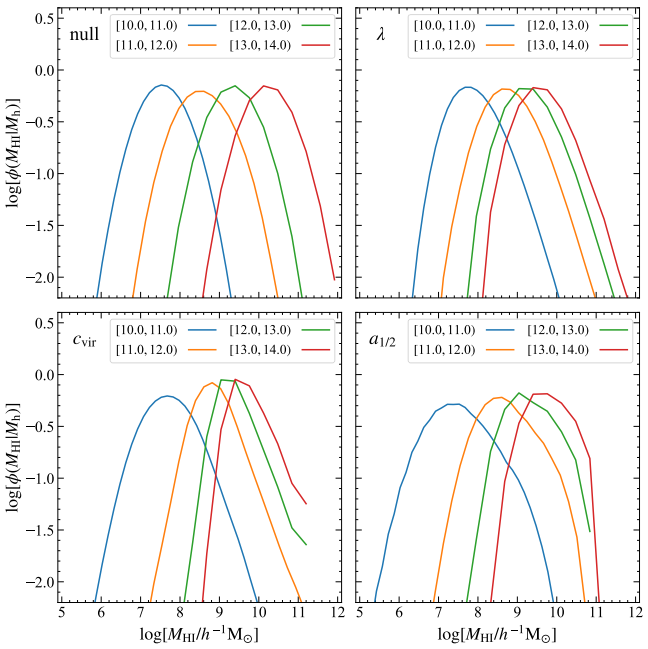


Figure 4. From left to right and top to bottom, this figure shows the PDF of H I mass and $\phi(M_{\rm H\,I}|M_{\rm h})$, for the null-model with a scatter $\delta = 0.5$ dex, λ-model, $c_{\rm vir}$-model and best-fit $a_{1/2}$-model for the SMDPL N-body simulation, in different halo mass bins, i.e., at $10.0 \leqslant \log(M_{\rm h}/h^{-1}M_\odot) < 11.0$ (blue), $11.0 \leqslant \log(M_{\rm h}/h^{-1}M_\odot) < 12.0$ (orange), $12.0 \leqslant \log(M_{\rm h}/h^{-1}M_\odot) < 13.0$ (green), and $13.0 \leqslant \log(M_{\rm h}/h^{-1}M_\odot) < 14.0$ (red), respectively. Different models generate different distributions, all of which look close to the standard log-normal distribution in the upper left subplot. Among these distributions, the best-fit $a_{1/2}$-model has the largest scatter, decreasing from 0.7 to 0.5 when $M_{\rm h}$ increases from $10^{10}\,h^{-1}M_\odot$ to $10^{14}\,h^{-1}M_\odot$.

correlation functions (Wang et al. 2013a). Such differences are reflected in halos below the lower limit of $10^{11}\,h^{-1}M_\odot$ for the observed H I–halo mass relation, causing the differences in the predictions of $\Omega_{\rm H\,I}$.

3.2. H I Mass Distributions

When fitting the H I–halo mass relation, we only use the information of average H I mass in each $M_{\rm h}$ bin, i.e., $\langle M_{\rm H\,I}|M_{\rm h}\rangle$, without assuming the probability distribution of H I mass in halos, $\phi(M_{\rm H\,I}|M_{\rm h})$. However, the secondary parameter dependence in Equation (10) actually encodes the information of $\phi(M_{\rm H\,I}|M_{\rm h})$. Thus, with the best-fit model parameters in Table 1, we can construct mock catalogs from the simulations and quantify $\phi(M_{\rm H\,I}|M_{\rm h})$.

We show in Figure 4 the distributions of $\phi(M_{\rm H\,I}|M_{\rm h})$ in different halo mass bins for the λ-model, $c_{\rm vir}$-model and also the best-fitting $a_{1/2}$-model. Additionally, H I distribution for null-model with a log-normal distributed scatter $\delta = 0.5$ is also presented as comparison, as such a scatter $\delta \approx 0.5$ is found in the IllustrisTNG (hereafter TNG) simulation. All of the other distributions look close to the log-normal ones, while some of them are increasingly skewed for massive halos. Compared with the artificially added scatter, the $c_{\rm vir}$-model and especially the $a_{1/2}$-model have a larger scatter, while the scatter of the λ-model is closest to 0.5 dex. This coincides with what we found in the TNG simulation, that the primary data of H I mass are more dependent on spin parameter λ than on $c_{\rm vir}$ or $a_{1/2}$ (see more details in Section 4.2), which also indicates possibly different feedback models in TNG simulation.

We adopt the best-fitting $a_{1/2}$-model and show in Figure 4 the distributions of $\phi(M_{\rm H\,I}|M_{\rm h})$ in different halo mass bins. While they look close to the log-normal distributions, the distributions are increasingly skewed for less massive halos. This is likely caused by the inclusion of the formation time dependence in $\phi(M_{\rm H\,I}|M_{\rm h})$.

In principle, $\phi(M_{\rm H\,I}|M_{\rm h})$ can be expanded as $\phi(M_{\rm H\,I}|M_{\rm h}, a_{1/2})$. However, in our model, $M_{\rm H\,I}$ is completely determined by $M_{\rm h}$ and $a_{1/2}$, without any scatter. The current observational data of ALFALFA are not large enough to fully constrain $\phi(M_{\rm H\,I}|M_{\rm h}, a_{1/2})$. We find from Figure 4 that the average scatter of $M_{\rm H\,I}$ in different $M_{\rm h}$ bins decreases from 0.7 to 0.5 when $M_{\rm h}$ increases from $10^{10}\,h^{-1}M_\odot$ to $10^{14}\,h^{-1}M_\odot$.

The measurements of $\phi(M_{\rm H\,I}|M_{\rm h})$ are sometimes referred to as the conditional H I mass function (CHMF). Nevertheless, our $M_{\rm H\,I}$ measurements are the total H I mass in halos rather than the H I mass of individual galaxies, as commonly used. Our model predictions can still be compared with the measurements of CHMF built from the galaxies (Li et al. 2022), providing more constraints on the H I content of galaxies.

Another application of $\phi(M_{\rm H\,I}|M_{\rm h})$ is the calculation of H I mass measured in different ways. It is important that the comparison of different studies is made with consistent definitions. There are three commonly used quantities—$\log\langle M_{\rm H\,I}\rangle$, $\langle \log M_{\rm H\,I}\rangle$, and the logarithm of the median values of $M_{\rm H\,I}$, in different halo mass bins. Although the stacking measurements H I of G20 are measured with $\langle M_{\rm H\,I}\rangle$, it is more common to compare the median $M_{\rm H\,I}$ in theoretical models. All of them can be calculated directly from $\phi(M_{\rm H\,I}|M_{\rm h})$. But for simplicity, we use the same mock catalog as in Figure 4. The errors of median $M_{\rm H\,I}$ are estimated from the 16th–84th percentile range.

We find that the logarithm of the median $M_{\rm H\,I}$ is very close to $\langle \log M_{\rm H\,I}\rangle$, but systematically smaller than $\langle M_{\rm H\,I}\rangle$ by about 0.3–0.5 dex, indicating a log-normal distribution for the H I mass. As expected, the measurement of $\langle M_{\rm H\,I}\rangle$ would be up-weighted by galaxies with larger $M_{\rm H\,I}$, causing the difference in the median values. However, their differences are still within the 1σ range of the median values, and there is no strong systematic trend of the difference with $M_{\rm h}$, as opposed to some theoretical models (e.g., Calette et al. 2021). We adopt the convention of taking the mean ($\langle M_{\rm H\,I}\rangle$) throughout this paper.

4. Comparison with Simulations

In an attempt to give a theoretical interpretation to our findings, we utilize the IllustrisTNG hydrodynamical simulation (hereafter

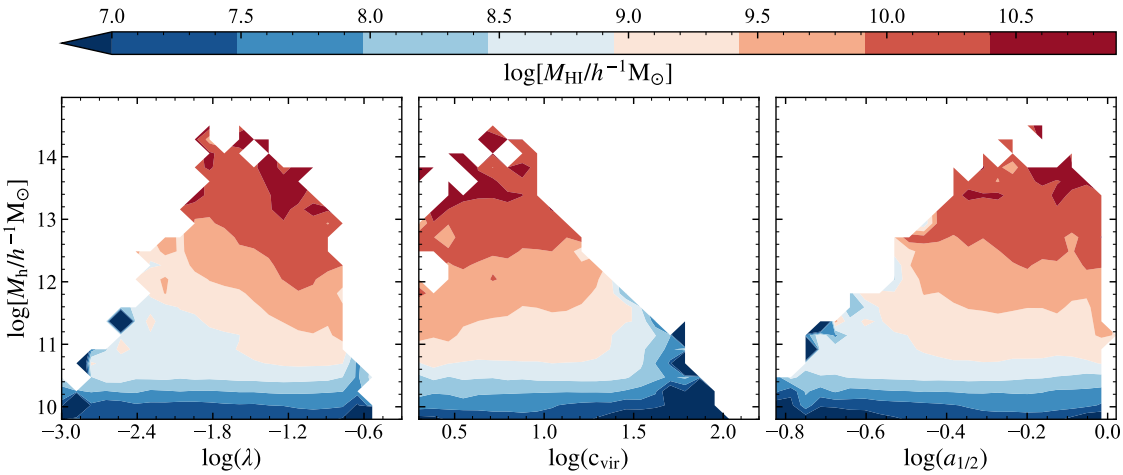


Figure 5. The joint dependence of H I mass on both halo mass and secondary halo parameters in TNG100 simulation at $z = 0$. We show the dependence on halo spin parameter λ (left), concentration $c_{\rm vir}$ (middle), and formation time $a_{1/2}$ (right). The H I mass is represented by the color in each panel, with the color bar shown on the top.

TNG) in this section to investigate the H I–halo mass relation and its dependence on other halo properties. We find some agreement and also inconsistency in H I–halo mass relation between the observation and TNG simulations. To verify the flexibility and universality of our empirical model, fittings of statistics are also performed, through which some interesting differences between simulation results and observational measurements are further confirmed.

4.1. The IllustrisTNG Simulation

The TNG simulation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2018; Springel et al. 2018) is a set of state-of-the-art hydrodynamical simulations, containing a wide range of physical processes of galaxy formation with the same Planck cosmology as SMDPL. In this work, we mainly adopt the simulation set of TNG100-1 with a box size of 110.7 Mpc and a dark matter particle mass resolution of $7.5 \times 10^6 M_\odot$. In addition, we also use the dark matter-only simulation of TNG100-1-Dark, to apply our H I–halo model.

Our modeling of the observational data in the previous section is based on halo catalogs generated with the ROCKSTAR halo finder, which is different from the original halos identified with SUBFIND (Springel et al. 2001) in TNG. For a fair comparison, we reprocess the snapshot data of TNG100-1 and TNG100-1-dark to generate halos and subhalos using the ROCKSTAR algorithm. The halo properties of formation time, concentration and spin are also provided by the ROCKSTAR finder. We adopt the post-processed H I mass of TNG subhalos in Diemer et al. (2018), where five different models for the atomic-to-molecular transition are presented. The results of all five models are similar (Diemer et al. 2019) and do not affect the conclusions. For simplicity, we employ the "K13" model, which is an analytical model to predict the fraction of molecular hydrogen H_2 and H I from the neutral gas surface density and metallicity (Krumholz 2013). We note that these H I masses were generated for the SUBFIND subhalos. Thus, we further cross-match the SUBFIND subhalos with our ROCKSTAR subhalos to obtain the final H I mass for each host halo in our catalog.

We additionally check the TNG300-1 simulation with a larger box size of 302.6 Mpc and a dark matter particle mass resolution of $5.9 \times 10^7 M_\odot$. Although mass resolution of TNG300-1 is not as fine as that of TNG100-1, the relatively large box size can avoid the effect of undersized boxes on statistics, especially on projected H I clustering. In the next sections, we primarily focus on the results and details of TNG100-1, such as the secondary dependence of H I on halo properties, as well as the predictions by fitting our model. For TNG300-1, the same fitting is performed, but the main focus is on the projected H I clustering w_p.

4.2. H I–Halo Relation in TNG

In Figure 5, we show the joint dependence of $M_{\rm H\,I}$ on $M_{\rm h}$ and secondary halo parameters in TNG100-1, including halo spin parameter λ, concentration $c_{\rm vir}$, and formation time $a_{1/2}$. For low-mass halos of $M_{\rm h} < 10^{10.8}\,h^{-1}M_\odot$, the dependence of $M_{\rm H\,I}$ on the secondary parameters is generally weak, as indicated by the flat contours in Figure 5. This means that for such low-mass halos, the H I mass is mainly determined by the host halo mass. For more massive halos, large H I content tends to be found in high-spin, low-concentration, and late-forming halos. The H I mass shows the strong dependence on λ and weak dependence on $c_{\rm vir}$ and $a_{1/2}$.

Qualitatively, the TNG result that the H I mass depends on both halo mass and the secondary parameters, which is

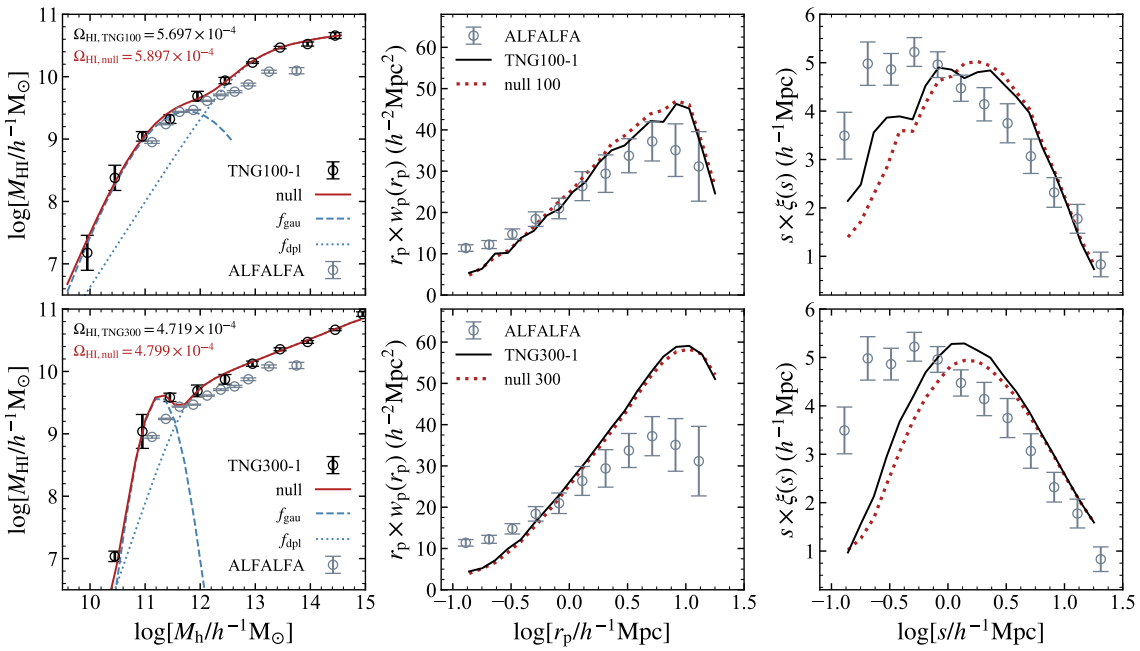


Figure 6. Statistics and predictions similar to Figure 1 obtained from TNG100-1 (top panels) and from TNG300-1 (bottom panels). (Left column) Best-fitting models to the TNG simulation results at $z = 0$ including H I–halo mass average relation $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ and $\Omega_{\rm H\,I}$. Black circles with error bars represent the results of $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ in the TNG simulation, and the red solid line represents the prediction by applying the "null"-model (i.e., without secondary dependence of halo parameters according to Equations (7)–(9)) to the corresponding TNG simulations. The blue dashed lines are the analytical Gaussian-like functions ("$f_{\rm gau}$"), and the blue dotted lines are the double power law ("$f_{\rm dpl}$"). For comparison, we show the measurements of ALFALFA with error bars (gray circles). (Middle and right columns) We show the projected correlation function $w_{\rm p}(r_{\rm p})$ (middle) and correlation function $\xi(s)$ (right) from the TNG300-1 and TNG100-1 simulations (black solid lines), and the predictions (red dotted line) by applying the "null"-model to the halo data of the TNG300-1 simulations and to the TNG100-1-Dark simulation, respectively.

consistent with our previous conclusion in the observational analysis. Nevertheless, the observations reject the model in which the H I mass depends on halo spin parameter λ, but show a strong dependence of $M_{\rm H\,I}$ on $a_{1/2}$ and $c_{\rm vir}$, as depicted in Figure 1. That is in conflict with the trend found in TNG.

In the left column of Figure 6, we show the H I–halo mass relation in the TNG100-1 (top) and TNG300-1 (bottom) simulations. We also display the measurements of the ALFALFA sample as the gray circles. The overall trend of increasing $M_{\rm H\,I}$ with $M_{\rm h}$ in the TNG simulations is found in a similar manner to that in the ALFALFA measurements. The low-mass halos of $M_{\rm h} < 10^{12}\,h^{-1}M_\odot$ in the TNG100-1 simulation have similar content of H I gas to those in the ALFALFA observation. On the other hand, $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ in the low-mass halos is much steeper in TNG300-1, and also less consistent with observational results of ALFALFA. For more massive halos, the total H I content is apparently over-predicted in both of the TNG simulations. As discussed in G20, this discrepancy results from the low-level AGN feedback in TNG, making over-abundant cold gas in the central galaxies.

In the middle and right columns of Figure 6, we show the H I clustering measurements, i.e., $w_{\rm p}(r_{\rm p})$ and $\xi(s)$, in the TNG simulations. Note that while the clustering measurements from the TNG100-1 simulation agree with the ALFALFA observations reasonably well, a fair comparison should be made between the ALFALFA (with the sample volume of $3.7\times10^6\,h^{-3}{\rm Mpc}^3$) and the TNG300-1 simulation (with a volume of $8.6\times10^6\,h^{-3}{\rm Mpc}^3$, or 2.3 times larger than the ALFALFA survey volume), as opposed to that between the ALFALFA and the TNG100-1 simulation (with a volume of $0.42\times10^6\,h^{-3}{\rm Mpc}^3$, or one-ninth of the ALFALFA survey volume), because of the integral constraint effect due to small volume size of a survey or simulation.

The initial conditions and post-processing of H I and H_2 models are exactly the same in the TNG100-1 and TNG300-1 simulations, but Figure 6 shows that the H I clustering amplitude is suppressed at large scales in the TNG100-1 simulation, compared to the TNG300-1 simulation, which results from the integral constraint effect, i.e., a systematic error by a constant offset. The small-scale (below $1\,h^{-1}$Mpc) difference between both simulations is likely limited by the resolution.

Regarding the comparison between the TNG300-1 and ALFALFA results, the middle and right panels of Figure 6 show that the H I clustering, especially the projected correlation function $w_{\rm p}(r_{\rm p})$, is significantly larger at large scales in TNG300-1 than in the ALFALFA observation. In other words, the TNG simulation over-predicts the H I bias.

We apply the "null"-model to the TNG100-1-Dark simulation, and obtain the best-fit model parameters by fitting its

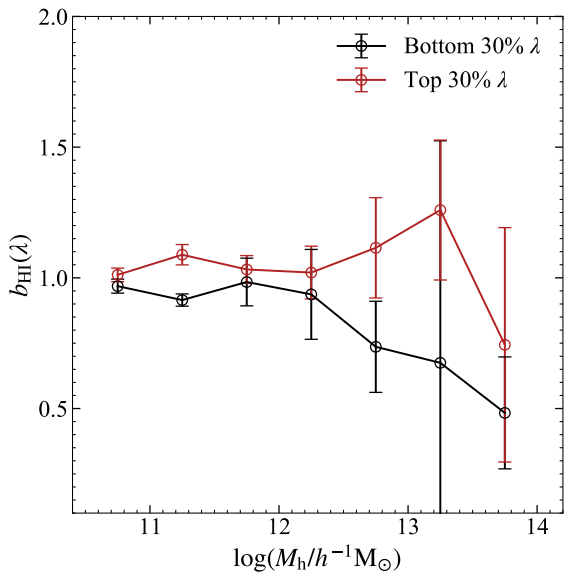


Figure 7. Secondary bias of H I on halo spin parameter λ and halo mass. The red circles with error bars represent H I bias with the top 30%λ, while black circles are that with bottom 30%λ.

prediction of H I abundance $\Omega_{\rm H\,I}$ and average $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ relation to the results of the TNG100-1 simulation. A similar fit is also done in TNG300 to demonstrate the flexibility of our model. The left column of Figure 6 shows that the average $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ relation can be well described by the null-model without secondary halo parameters, as found in the case of fitting to the observational data (see Figure 1). On the other hand, the predicted H I clustering from the "null"-model agrees well with the direct H I simulation results of TNG, as shown in the middle and right panels in Figure 6, which confirms the dependence of H I on spin parameter λ in TNG simulations. As shown by comparing the clustering between the "null"-model and the other secondary dependence models in Figure 1, the H I clustering is sensitive to both $c_{\rm vir}$ and $a_{1/2}$, and insensitive to λ. Thus, the dependence of H I mass on halo spin λ leads to the fact that halo properties are not needed to match the H I clustering. The fact that the H I clustering is insensitive to halo spin can also be explained as follows. The H I clustering is mostly contributed by halos of $10^{11}\,h^{-1}M_\odot < M_{\rm h} < 10^{12}\,h^{-1}M_\odot$, because the abundance of massive halos significantly decreases. However, the dependence of halo clustering (or halo bias) on λ is generally weak for halos in this mass range (see, e.g., Figure 4 of Sato-Polito et al. 2019). As a result, the H I clustering depends rather weakly on λ. This is also verified more directly in Figure 7, which gives the relative bias of H I on λ.

In sum, the TNG simulations show the strong dependence of H I mass on λ and weak dependence on $c_{\rm vir}$ and $a_{1/2}$, which are inconsistent with the findings in observations. Also, the H I clustering from the TNG simulations is significantly over-predicted at large scales, compared to the observations.

4.3. Comparison with Previous Work

The H I–halo mass relation has been recently studied extensively with different galaxy formation models (see, e.g., Bagla et al. 2010; Villaescusa-Navarro et al. 2018; Baugh et al. 2019; Chauhan et al. 2020; Spinelli et al. 2020; Spina et al. 2021; Zhang et al. 2022). We compare the measurements from ALFALFA and different fiducial models proposed by different simulations in Figure 8, including the SAMs of GALFORM (Baugh et al. 2019), SHARK (Chauhan et al. 2020), GAEA (Spinelli et al. 2020), L-GALAXIES (Zhang et al. 2022) and the analytical fitting formulae for TNG100-1 (Villaescusa-Navarro et al. 2018). For a fair comparison, the same models are also applied in SMDPL simulation and their parameters are set free to fit the real measurements (bottom panels of Figure 8).

As shown in the left upper panel of Figure 8, the H I mass in the models of Villaescusa-Navarro et al. (2018) and Spinelli et al. (2020) increases monotonically with the halo mass, while the H I–halo mass relation in other models has a non-monotonic feature. As discussed in G20, this non-monotonic feature is physically related to the virial shock-heating (Rees & Ostriker 1977; Silk 1977; White & Rees 1978) and AGN feedback (see e.g., Baugh et al. 2019; Chauhan et al. 2020). It is thus a necessary component of our empirical model. However, the decreasing of $M_{\rm H\,I}$ in the models of Baugh et al. (2019) and Chauhan et al. (2020) is too strong compared with the observation, while the monotonically increasing models of Villaescusa-Navarro et al. (2018) and Spinelli et al. (2020) would over-predict $M_{\rm H\,I}$ in the massive halos.

It is interesting to note that all of these different models have roughly consistent slopes at the low-mass end, with $M_{\rm H\,I}$ quickly increasing with $M_{\rm h}$. That is consistent with the three-stage scenario mentioned in Section 1, i.e., the H I–halo mass relations are dominated by the cold accretion mode in the halo mass ranges of $M_{\rm h} < 10^{12}\,h^{-1}M_\odot$, by the hot accretion mode in the range of $10^{12} < M_{\rm h} < 10^{13}\,h^{-1}M_\odot$, and by the mergers in the range of $M_{\rm h} > 10^{13}\,h^{-1}M_\odot$, respectively. Figure 8 suggests that the H I mass measurements in the halo mass range of 10^{11}–$10^{13}\,h^{-1}M_\odot$ can effectively distinguish these different theoretical models, as the cold gas content is very sensitive to the AGN feedback activities (Guo et al. 2022).

We note that the model of Chauhan et al. (2020) also includes the additional dependence of $M_{\rm H\,I}$ on the halo spin parameter. The scatter of H I–halo mass relation in the SHARK semi-analytic model (Lagos et al. 2018) is shown to be strongly correlated with λ, as in the case of TNG. As noted above, including the dependence on λ cannot resolve the tension of H I bias with observations.

We show the predicted values of cosmic H I abundance $\Omega_{\rm H\,I}$ at $z = 0$ in different models in the right column of Figure 8. In the bottom right panel, while most of these predictions show large offsets from the ALFALFA observation, our prediction

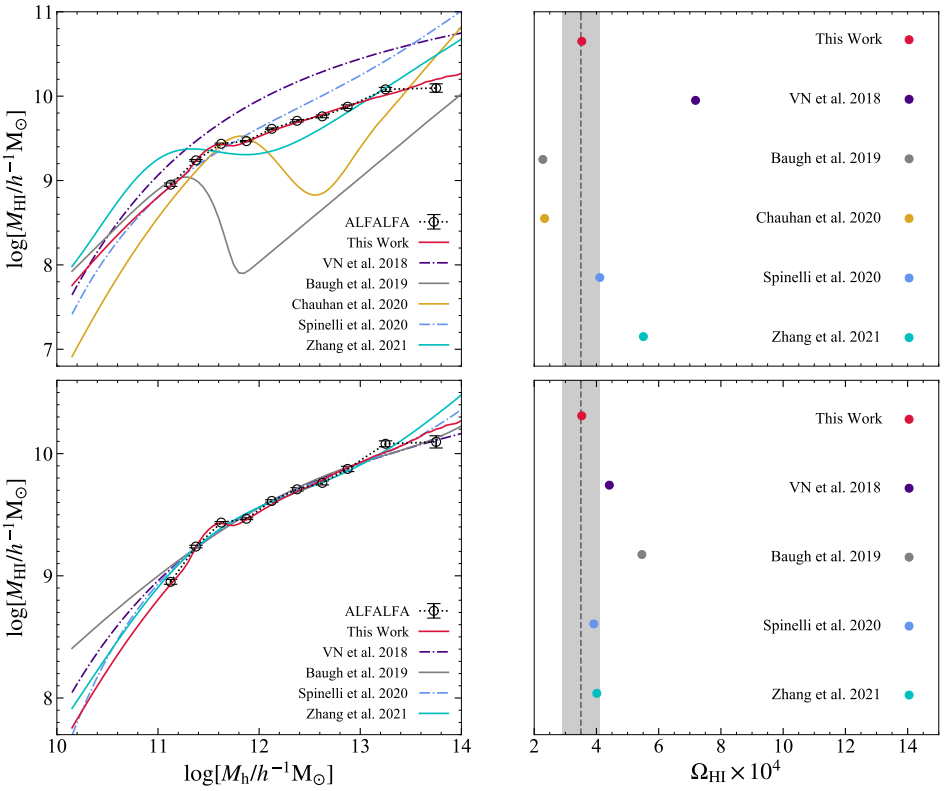


Figure 8. Comparison of different predictions from literature and our best-fitting model (top) and results of fitting the observations with formulas (bottom). The top left subplot shows the average H I–halo mass relation $\langle M_{\rm H\,I}|M_{\rm h}\rangle$. We show the SAMs from GALFORM (Baugh et al. 2019) (gray solid line), SHARK (Chauhan et al. 2020) (yellow solid line), GAEA (Spinelli et al. 2020) (blue dot–dashed line), L-GALAXIES (Zhang et al. 2022) (cyan solid line), and the analytical fitting formulae for TNG100-1 (Villaescusa-Navarro et al. 2018) (purple dot–dashed line), as well as our work (red solid line). We also show the ALFALFA observation result (G20) (black circles with error bars). In the top right subplot, the H I abundance is $\Omega_{\rm H\,I}$ at $z = 0$ for the same models. We also show the observation result (gray dashed vertical line, with light gray shaded area indicating the 1σ error) from Jones et al. (2018). In the bottom, the observations are fitted with the formulas mentioned above, except for Chauhan et al. (2020). In fairness, both $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ and $\Omega_{\rm H\,I}$ are taken into account when fitting through MCMC procedure.

matches best with the observational data. Note that the prediction of the GAEA model (Spinelli et al. 2020) shows the best agreement with our prediction, and Zhang et al. (2022) also fit well, which is due to the similar H I–halo mass relations at $M_{\rm h} < 10^{12}\,h^{-1}M_\odot$ that dominates the contribution to $\Omega_{\rm H\,I}$.

These comparisons suggest that our flexible, empirical model works well to accurately describe the H I distribution in the universe.

5. Discussions

According to the best-fit value of γ in Table 1, H I empirically prefers to reside in late-forming or low-$c_{\rm vir}$ halos. This agrees with the physical scenario that the H I-rich halos are caused by the late accretion of H I gas accompanied with the halo formation histories. The accreted H I gas is not yet consumed by the internal physical processes (Gao et al. 2005; Wechsler et al. 2006; Guo et al. 2017; Diemer et al. 2018; Guo et al. 2020). On the other hand, H I-rich halos could also result from early gas accretion, while the inherited angular momentum of the accreted gas from high-spin halos may prevent star formation by resisting gravitational collapse. This scenario has been suggested by some H I-extreme galaxies in observations (e.g., see Lutz et al. 2018). However, the negative best-fit value of γ suggests that neutral hydrogen preferentially resides in low-spin halos, which corresponds to the low bias of the H I-rich galaxies. This result contradicts the expected scenario mentioned above. In fact, if there is a strong dependence of H I mass and halo spin parameter, we should expect to find a much higher H I bias than observed because there is a very weak dependence of the halo bias on the spin parameter for low-mass halos of $M_{\rm h} < 10^{12}\,h^{-1}M_\odot$. Therefore, we conclude that the low H I clustering or bias obtained in this work disfavors the dependence between H I and the spin parameter empirically and theoretically.

In the TNG simulation, the clear dependence of H I content on the halo spin parameter is in fact caused by its AGN feedback mechanism. Quenched galaxies in TNG retain a large amount of H I gas in the kinetic AGN feedback mode. The weak kinetic AGN feedback only redistributes gas from the inner part to the outer part, keeping a roughly constant H I

reservoir. The halos with high spin parameter will further prevent the H I gas from falling into the center. However, such a scenario does not apply to other simulations. As shown in Liu et al. (2025), the efficient kinetic AGN feedback in the SIMBA simulation will directly remove the H I gas from the halo environment, effectively reducing the H I reservoir, which will lead to a much weaker dependence of H I gas on the halo spin parameter.

6. Conclusions

In this paper, we propose a flexible, empirical model for the H I–halo mass relation that successfully fits the observed cosmic H I abundance ($\Omega_{\rm H\,I}$), the average H I–halo mass relation $\langle M_{\rm H\,I}|M_{\rm h}\rangle$ and the H I clustering measurements for the ALFALFA sample in the redshift range of $0.0025 < z < 0.06$.

Our model consists of the primary dependence of H I mass on the mass of the host halo and the secondary dependence on other halo parameters, including halo spin parameter (λ), halo concentration ($c_{\rm vir}$), and halo formation time ($a_{1/2}$). We find that the observation data reject the "null"-model, which has no secondary parameters, with $\chi^2/{\rm dof} = 49.5/15$, and the λ-model, which has secondary dependence on halo spin parameter, with $\chi^2/{\rm dof} = 42.15/15$. We also find that the observed data strongly favor the $a_{1/2}$-model, which has the secondary dependence of H I mass on halo formation time, with a best-fitting $\chi^2/{\rm dof} = 27.23/15$. Moreover, the $c_{\rm vir}$-model, which has secondary dependence on halo concentration, is also in agreement with the observation data (but the agreement is slightly worse than the $a_{1/2}$-model).

In an attempt to explain these findings from the perspective of hydrodynamical simulations, we also investigate the H I–halo mass relation in the TNG simulation, and confirm that the total H I mass in halos shows strong dependence on the halo secondary parameters. However, the TNG simulation favors strong dependence of $M_{\rm H\,I}$ on λ, and weak dependence on $c_{\rm vir}$ and $a_{1/2}$. It also predicts the H I clustering with much larger values on large scales than the observations. These results of TNG are different from the findings in observations.

This interesting difference between the TNG simulation and the observation is related to the baryonic physics involved in the cold gas accretion and depletion adopted in the galaxy formation models. It demonstrates the importance of properly modeling the H I–halo mass relation in order to fully understand the H I distribution in the universe. Comparing our best-fitting model to the previous H I–halo mass relations in literature, we find substantial difference in both shape and amplitude, which reflects the large uncertainties in the modeling of H I gas. On the other hand, improving the accuracy of measurement and removing systematics of observations might also be necessary to understand the H I distribution.

An accurate H I–halo mass relation is essential for the construction of mock catalogs for future H I galaxy surveys as well as for the 21 cm intensity mapping surveys. While our model is built on measurements at the local universe, it has the potential to extend to study the evolution of the H I content at higher redshifts, which we leave to explore in future work.

Acknowledgments

This work is supported by National SKA Program of China (grant Nos. 2020SKA0110100, 2020SKA0110401), National Key R&D Program of China (grant Nos. 2018YFA0404502, 2018YFA0404503), the National Natural Science Foundation of China (NSFC, grant Nos. 11922305, 11833005, 11821303, and 12011530159), and the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-A02. We acknowledge the use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at the Shanghai Astronomical Observatory and acknowledge the Tsinghua Astrophysics High-Performance Computing platform at Tsinghua University for providing computational and data storage rsources that have contributed to the research results reported within this paper.

Appendix A
Prior and Assumption for Fitting

A.1. Covariance Matrix $\boldsymbol{C}$

Following Zehavi et al. (2004), we calculate the covariance matrix $\boldsymbol{C}$ of dispersion among 121 jackknife samples (Zehavi et al. 2004, 2011; Guo et al. 2017). The $\boldsymbol{C}$ is calculated as Equation (A1) and we present the corresponding correlation matrix, i.e., normalized covariance matrix, in Figure A1.

$$\boldsymbol{C}(\xi_i, \xi_j) = \frac{N-1}{N}\sum_{l=1}^{N}(\xi_i^l - \bar{\xi i})(\xi_j^l - \bar{\xi j}), \tag{A1}$$

where N is 121 in our work, and ξ_i denotes ξ or $w_{\rm p}$.

A.2. Parameter Prior and Data Likelihood

The prior distributions of the 8 parameters in Equation (10) are uniform, and the ranges are all presented in Table A1.

For the observational data, we assume that all distributions of data error or intrinsic scatter follow the Gaussian distribution. This assumption is verified for the individual H I mass in a certain halo mass interval, shown in Figure 8, while for measurements of H I clustering, this is the common assumption given various observational effects and the intrinsic scatter of data. Thus, the joint distribution should follow the χ^2 distribution, and the logarithm of data likelihood can be calculated as Equation (A2).

$$\chi^2 = (\boldsymbol{M} - \boldsymbol{D})^T \boldsymbol{C}_{\rm D}^{-1}(\boldsymbol{M} - \boldsymbol{D}), \tag{A2}$$

Table A1
Range of Uniform Prior Distribution for All Parameters

	$A_{\rm gau}$	$\log M_{\rm gau}$	$\sigma_{\rm gau}$	$\log A_{\rm dpl}$	$\log M_{\rm dpl}$	α	β	γ_λ[a]	$\gamma_{c_{\rm vir}}$	$\gamma_{a_{1/2}}$
Down	4.0	11	0.0001	−10.0	11.0	0.0	0.0	−4.0	−4.0	0.0
Up	9.0	12	3	0.0	12.0	1.0	1.5	0.0	0.0	10.0

Note.
[a] γ is set separately for different models with different halo properties, given that secondary dependence of H I mass on different halo properties has different inclination.

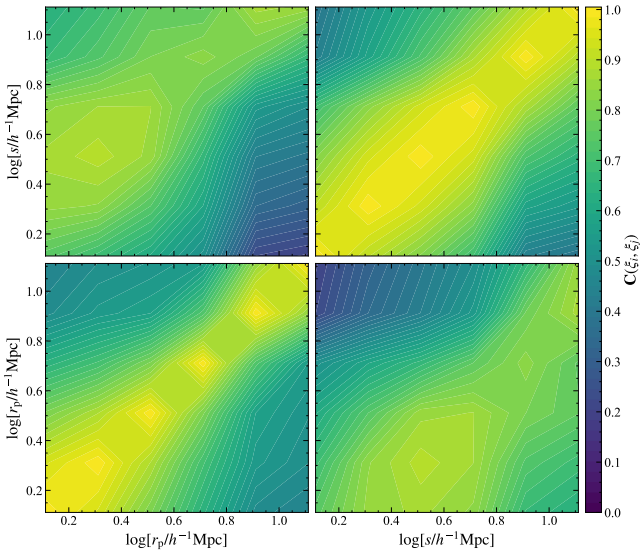


Figure A1. Correlation matrix, i.e., normalized error covariance matrix $\boldsymbol{C}$, calculated using the jackknife resampling method.

where $\boldsymbol{M}$ is the model predictions; $\boldsymbol{D} = [w_{\rm p}, \xi]$ represents the combined observational measurements, formed by concatenating the projected correlation function $w_{\rm p}$ and the redshift-space monopole ξ; $\boldsymbol{C}_{\rm D}$ is the full error covariance matrix for real data.

Appendix B
Goodness of Fit

Although the Bayesian approach has no statistics such as p value to quantify the GOF, we still make the model selections in this part based on the reduced χ^2 and logarithmic evidence, trying to verify the quality of the $a_{1/2}$-model.

B.1. Reduced χ^2

We choose the reduced χ^2 (noted as $\chi^2_{\rm red}$) in Equation (B1) to be our first method to evaluate the GOF (Andrae et al. 2010), given that a χ^2 distribution is assumed to be our likelihood distribution.

$$\chi^2_{\rm red} = \frac{\chi^2}{K}, \tag{B1}$$

where χ^2 is the data likelihood in this work (see Equation (A2)), and K is the dof. Generally speaking, the closer $\chi^2_{\rm red}$ is to 1, the better the fitting is.

It should be emphasized that the data errors of the H I–halo mass relation in G20 are the error on the mean calculated through jackknife resampling method, which are much smaller than the real standard deviation σ_i in A2. This leads to the overestimation for both of the χ^2 and $\chi^2_{\rm re}$, which means all of the models, especially the best-fit $a_{1/2}$-model, might be better than they seem now.

Table B1
Log-evidence and Reduced χ^2 For All Models

	Null	λ	$c_{\rm vir}$	$a_{1/2}$
$\log[E(M)]$[a]	−49.04	−44.48	−40.26	−37.12
$\chi^2_{\rm red}$[b]	3.30	2.83	1.94	1.82

Notes.
[a] The values of evidence are too small to present, so their logarithmic values are shown here.
[b] $\chi^2_{\rm red}$ is a commonly used statistic to judge the GOF for models, in the case that distributions of data errors are Gaussian. The closer to 1 the $\chi^2_{\rm red}$ is, the better the GOF.

B.2. Evidence

The evidence obtained from MCMC is the other way to select the best model. Evidence $E(M)$, also known as the marginal likelihood, quantifies the probability that data would be obtained if the adopted model were correct. Equation (B2) gives the definition, while Table B1 presents the evidences for all models mentioned in Section 3. It is obvious that the observational data have a preference for the $a_{1/2}$-model.

$$E(M) \equiv \int p(D|\theta, M)p(\theta)\, d\theta. \tag{B2}$$

ORCID iDs

Zhixing Li https://orcid.org/0000-0002-1287-6064
Hong Guo https://orcid.org/0000-0003-4936-8247
Yi Mao https://orcid.org/0000-0002-1301-3893

References

Abdalla, F. B., & Rawlings, S. 2005, MNRAS, 360, 27
Alonso, D., & Ferreira, P. G. 2015, PhRvD, 92, 063525
Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010, arXiv:1012.3754
Bagla, J. S., Khandai, N., & Datta, K. K. 2010, MNRAS, 407, 567
Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., et al. 2001, MNRAS, 322, 486
Battye, R. A., Brown, M. L., Browne, I. W. A., et al. 2012, arXiv:1209.1041
Baugh, C. M., Gonzalez-Perez, V., Lagos, C. d. P, et al. 2019, MNRAS, 483, 4922
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, ApJ, 803, 21
Calette, A. R., Rodríguez-Puebla, A., Avila-Reese, V., & Lagos, C. d. P. 2021, MNRAS, 506, 1507
Chauhan, G., Lagos, C. d. P., Stevens, A. R. H., et al. 2020, MNRAS, 498, 44
Chen, Z., Wolz, L., Spinelli, M., & Murray, S. G. 2021, MNRAS, 502, 5259
Cunnington, S., Li, Y., Santos, M. G., et al. 2023, MNRAS, 518, 6262
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEEP, 97, 1482
Diemer, B., Stevens, A. R. H., Forbes, J. C., et al. 2018, ApJS, 238, 33
Diemer, B., Stevens, A. R. H., Lagos, C. d. P., et al. 2019, MNRAS, 487, 1529
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fialkov, A., Barkana, R., & Jarvis, M. 2020, MNRAS, 491, 3108
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66
Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, AJ, 130, 2598
Guo, H., Li, C., Zheng, Z., et al. 2017, ApJ, 846, 61
Guo, H., Jones, M. G., Haynes, M. P., & Fu, J. 2020, ApJ, 894, 92
Guo, H., Jones, M. G., & Wang, J. 2022, ApJL, 933, L12
Guo, H., Jones, M. G., Wang, J., & Lin, L. 2021, ApJ, 918, 53
Guo, H., Wang, J., Jones, M. G., & Behroozi, P. 2023, ApJ, 955, 57
Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, ApJ, 861, 49
Jones, M. G., Haynes, M. P., Giovanelli, R., & Moorman, C. 2018, MNRAS, 477, 2
Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Krumholz, M. R. 2013, MNRAS, 436, 2747
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850
Lagos, C. d. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, MNRAS, 481, 3573
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Li, X., Li, C., Mo, H. J., Xiao, T., & Wang, J. 2022, ApJ, 941, 48
Lim, S. H., Mo, H. J., Lu, Y., Wang, H., & Yang, X. 2017, MNRAS, 470, 2982
Liu, K., Guo, H., Wang, S., et al. 2025, A&A, 693, A48
Lutz, K. A., Kilborn, V. A., Koribalski, B. S., et al. 2018, MNRAS, 476, 3744
Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, MNRAS, 480, 5113
Martin, A. M., Papastergis, E., Giovanelli, R., et al. 2010, ApJ, 723, 1359
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20
Meerklass Collaboration, Barberi-Squarotti, M., Bernal, J. L., et al. 2025, MNRAS, 537, 3632
Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, MNRAS, 477, 1206
Nelson, D., Pillepich, A., Springel, V., et al. 2018, MNRAS, 475, 624
Nelson, D., Springel, V., Pillepich, A., et al. 2019, ComAC, 6, 2
Newburgh, L. B., Addison, G. E., Amiri, M., et al. 2014, Proc. SPIE, 9145, 91454V
Obuljen, A., Alonso, D., Villaescusa-Navarro, F., Yoon, I., & Jones, M. 2019, MNRAS, 486, 5124
Papastergis, E., Giovanelli, R., Haynes, M. P., Rodríguez-Puebla, A., & Jones, M. G. 2013, ApJ, 776, 43
Peebles, P. J. E. 1969, ApJ, 155, 393
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541
Sato-Polito, G., Montero-Dorta, A. D., Abramo, L. R., Prada, F., & Klypin, A. 2019, MNRAS, 487, 1570
Seehars, S., Paranjape, A., Witzemann, A., et al. 2016, JCAP, 2016, 001
Silk, J. 1977, ApJ, 211, 638
Spina, B., Porciani, C., & Schimd, C. 2021, MNRAS, 505, 3492
Spinelli, M., Zoldan, A., De Lucia, G., Xie, L., & Viel, M. 2020, MNRAS, 493, 5434
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676
Square Kilometre Array Cosmology Science Working Group, Bacon, D. J., Battye, R. A., et al. 2020, PASA, 37, e007
Stiskalek, R., Desmond, H., Holvey, T., & Jones, M. G. 2021, MNRAS, 506, 3205
Switzer, E. R., Masui, K. W., Bandura, K., et al. 2013, MNRAS, 434, L46
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389
Villaescusa-Navarro, F., Viel, M., Datta, K. K., & Choudhury, T. R. 2014, JCAP, 2014, 050
Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ApJ, 866, 135
Wang, L., De Lucia, G., & Weinmann, S. M. 2013a, MNRAS, 431, 600
Wang, L., Weinmann, S. M., De Lucia, G., & Yang, X. 2013b, MNRAS, 433, 515
Wang, L., Weinmann, S. M., & Neistein, E. 2012, MNRAS, 421, 3450
Wang, Z., Chen, Y., Mao, Y., et al. 2021, ApJ, 907, 4
Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, MNRAS, 510, 3495
Xie, L., De Lucia, G., Hirschmann, M., Fontanot, F., & Zoldan, A. 2017, MNRAS, 469, 968
Xu, H., Zheng, Z., Guo, H., Zhu, J., & Zehavi, I. 2016, MNRAS, 460, 3647
Zehavi, I., Weinberg, D. H., Zheng, Z., et al. 2004, ApJ, 608, 16
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
Zhang, J., Motta, P., Novaes, C. P., et al. 2022, A&A, 664, A19
Zhu, G., Zheng, Z., Lin, W. P., et al. 2006, ApJL, 639, L5