Neutral Hydrogen (HI) in cosmological simulations

Tutors: Bernhard Vos Ginés & Violeta González Pérez

A thesis presented for the degree in

Bsc. Physics

Facultad de Ciencias

Universidad Autónoma de Madrid

September 11, 2023

Abstract

We study the distribution and clustering of neutral Hydrogen (HI) in the Universe which acts as a tracer of gas and can help constrain current cosmological models. For this study, we have used four theoretical prescriptions from the literature relating HI mass to halo mass. We have analyzed their predictions and the physical processes that govern their evolution. We then study the amount of HI mass available in UNIT simulations by making use of the Semi-Analytical galaxy evolution code (SAGE) [Croton et al., 2016]. We have analysed two different models which are presented in Zoldan et al. [2016] and make use of SAGE parameters. To study these models in more depth we obtained the mean HI density parameter for both. From there, we extended the analysis by calculating the HI density contrast, which is dependent on the position of the neutral hydrogen in the simulation, for both models. Finally, we explore their clustering via the two-point correlation function for both large scales (5-150 Mpc/h) and small scales (0.1-10 Mpc/h).

1 Introduction

To understand the Universe's structures as they are in the current time it is important to create successful models that imitate their evolution. These models, however, must match observational results as closely as possible and at all epochs to accurately understand the formation of structures we discern nowadays. The synergy between modelization and observational results requires of great amounts of data to accurately constrain these models.

These efforts have led to the construction of Earth-based and space telescopes such as the Very Large Telescope (VLT) [Rouan, 2011] and Hubble Space Telescope (HST) Lallo [2012], among many others, which have allowed astronomers to study the Universe up to when it was about 1 billion years old. On the other hand, radiation from the Cosmic Microwave Background (CMB) which decoupled from matter when the Universe was around 380,000 years old have allowed the study of these early times. The challenge comes in the attempt to connect these two periods due to the lack of observations [Pritchard and Loeb, 2012].

Among other approaches, the study of the redshifted 21cm HI line might be a viable method to cover the observational gap. This line is produced by the hyperfine splitting due to the electron-proton interaction in neutral Hydrogen [Pritchard and Loeb, 2012]. The usefulness of this transition line resides in the ubiquitousness of neutral Hydrogen. Furthermore, the 21cm radiation is unlikely to be absorbed by gas due to the low probability of the transition and is transparent to dust which cannot absorb such wavelengths. It is also not necessary to resolve galaxies with SKA Intensity Mapping technique which makes the task of reaching high redshifts much simpler [Bharadwaj et al., 2001]. Therefore, HI can be used as a tracer of gas and can help us understand and constrain the current cosmological models.

From HI intensity measurements the obtention of HI mass is direct Dickey [1990]. This is why developing models of $M_{\rm HI}$ is vital since it can be easily constrained with observations. This way, we can understand which parameters affect the evolution of neutral Hydrogen and to which extent they are important.

To obtain this data, interferometry surveys are needed. Some such as MeerKAT [Santos et al., 2017], from which results have already been obtained, and the Square Kilometer Array (SKA) [Bacon et al., 2020] are currently underway.

2 HI -Halo mass relation

In this section we will study the HI mass as a function of host halo masses by considering several prescriptions. Four of these are functions of halo mass and two of them are models which are based on semi-analytic observables. We now proceed to explain both methods in more detail.

2.1 Theoretical HI prescriptions

The distribution of neutral Hydrogen as a function of halo mass has been extensively studied in the literature. In this section, we have selected four theoretical descriptions

that have been proposed and we examine each of them in detail. Specifically, the prescriptions selected are those suggested by Padmanabhan and Kulkarni [2017], Bagla et al. [2010], Baugh et al. [2018] and Spinelli et al. [2020].

2.1.1 Prescription 1: Padmanabhan and Kulkarni [2017]

The HI -halo mass (HIHM) relation proposed is based on the stellar-mass-halo-mass relation (SHM). They postulate that the HIHM relation can be modelled to be similar to the SHM at $z \approx 0$. The physical basis for the choice of this parametrisation arises in the Schechter form description of the stellar mass and luminosity distributions. The Schechter form is a model which attempts to predict the number of stars or galaxies as a function of Stellar Mass or Luminosity [Weigel et al., 2016]. The assumption made by Padmanabhan and Kulkarni [2017] is that, as a consequence of the Schechter form description, at $z \approx 0$ both HIHM and the SHM can be modelled similarly. Even though this is an assumption made in [Padmanabhan and Kulkarni, 2017], some information about the link between both can be found in the following literature: Lagos et al. [2011a], Fu et al. [2013] and Bera et al. [2022].

Since this description is valid for $z \approx 0$, modifications have to be made via the use of a redshift- and mass-dependent Navarro–Frenk–White profile. These profiles describe the dark matter density as a function of the virial radius for a dark matter halo [Navarro et al., 1996].

The parametrisation adopted is then constrained using a combination of data from radio emission at low redshift observations and optical/UV emission at high redshift observations. Results are then compared to hydrodynamical model findings. Therefore, the HI - halo mass relation suggested is based on the form of the SHM function introduced by Moster et al. [2012] and is as follows:

$$M_{\rm HI} = 2N_1 M_{\rm h} \left[\left(\frac{M_{\rm h}}{M_1} \right)^{-b_1} + \left(\frac{M_{\rm h}}{M_1} \right)^{y_1} \right]^{-1} \tag{1}$$

where M_h is the halo mass and the rest of the parameters are written as:

$$\begin{aligned} log_{10}(M_1) &= log_{10}(M_{10}) + \frac{z}{z+1} M_{11} \\ N_1 &= N_{10} + \frac{z}{z+1} N_{11} \\ b_1 &= b_{10} + \frac{z}{z+1} b_{11} \\ y_1 &= y_0 + \frac{z}{z+1} y_{11} \end{aligned} \tag{2}$$

The best-fitting values for the following free parameters are $M_{10} = 4.58 \pm 0.19 \times 10^{11} \mathrm{M}_\odot$, $N_{10} = (9.89 \pm 4.89) \times 10^{-3}$, $b_{10} = 0.90 \pm 0.39$ and $y_{10} = 0.74 \pm 0.03$ for $z = 0$. The rest of the parameters govern the HI evolution for $z > 0$ and their best-fitting values are $M_{11} = 1.56^{+0.53}_{-2.70}$, $N_{11} = 0.009^{+0.06}_{-0.001}$, $b_{11} = -1.08^{+1.52}_{-0.08}$ and $y_{11} = 4.07^{+0.39}_{-2.49}$.

The HI -Halo Mass relation introduced by Padmanabhan and Kulkarni [2017] has been represented in Fig.1. As can be seen, there is HI suppression at both the low-mass and high-mass ends with a maximum near $log_{10}(M_{\rm h}) = 12.5 M_\odot h^{-1}$. Possible explanations for the HI suppression are the high efficiency of supernovae feedback at

low masses, since the effect of gas expulsion and re-heating is more notable at these mass values. For the higher masses, a possible explanation is the high efficiency of AGN-feedback, since the expulsion of highly energetic radiation and outflows removes HI from the halo [Moster et al., 2010].

2.1.2 Prescription 2: Bagla et al. [2010]

The next prescription studied has been derived from considering gravity-only N-body simulations, specifically using the TREEPM code as described by Bagla [2002], Bagla and Ray [2003] and Khandai and Bagla [2009]. The cosmological parameters arise from the best-fit model for a WMAP-5 cosmology. The results from the simulation are combined with an ansatz that assigns HI to dark matter haloes. The ansatz used is chosen to provide results that are consistent with theoretical models and observational constraints. The ansatz imposes a lower cut-off with gas in haloes whose circular velocity is lower than $v_{\rm min} = 30{\rm kms}^{-1}$ being assigned no HI mass [1]. For the higher halo mass end, it has been found that haloes whose masses greatly exceed $M_{\rm h} = 10^{11} M_{\odot}$ do not host significant amounts of HI . Therefore, the HI assignment is such that the fraction of gas in neutral form is negligible for very massive haloes. This is done with a cut-off at circular velocities of $v_{\rm max} = 200{\rm kms}^{-1}$. Since we are working with halo masses and the mass within the virial radius is dependent on the peculiar velocity, both these cut-offs in peculiar velocities are converted to maximum and minimum masses via:

$$M_{\rm vir} \approx 10^{10}{\rm M}_{\odot} \left(\frac{\rm v_{circ}}{60{\rm kms}^{-1}}\right) \left(\frac{1+{\rm z}}{4}\right)^{-\frac{3}{2}} \tag{3}$$

As stated in their study, they are the first to consider enough resolution to resolve the smallest haloes which may contain significant amounts of HI and at the same time consider volumes which are big enough to obtain conclusions about the large-scale distribution.

From the different prescriptions suggested in Bagla et al. [2010], and following Avila et al. [2021], we only consider the following relation:

$$M_{\rm HI} = \begin{cases} \dfrac{f_2 M_{\rm h}}{1+\left(\dfrac{M_{\rm h}}{M_{\rm max}}\right)^2}, & \text{if } M_{\rm h} > M_{\rm min} \\ \\ 0, & \text{if } M_{\rm h} < M_{\rm min} \end{cases} \tag{4}$$

where the values for the parameters are redshit-dependent. We consider $f_2 = 0.0159$, to match a neutral hydrogen mean number density of $\Omega_{\rm HI} = 0.001$. As for other parameters, the minimum mass has a value of $M_{\rm min} = 1.92 \times 10^9 {\rm M}_{\odot}$ and the maximum mass a value of $M_{\rm max} = 5.68 \times 10^{11} {\rm M}_{\odot}$ [Avila et al., 2021]. These values are obtained by using Eq. 3 from Bagla et al. [2010] and imposing their cut-offs for circular velocities.

The HIHM relation introduced is represented in Fig.1. As can be seen there is also HI suppression at both the low- and high-mass ends due to supernovae feeback and AGN feedback respectively. In this case, the maximum is located at $log_{10}(M_{\rm h}) \approx 12.0$.

[1] This is true for $z \approx 3$. For lower redshifts, higher circular velocity values are expected but they have kept it as a fixed value

2.1.3 Prescription 3: Baugh et al. [2018]

The parametrisation introduced by Baugh et al. [2018] is based on physical predictions made by a combination of semi-analytical models. GALFORM [Cole et al., 2000], Bower et al. [2006], Baugh et al. [2005], Gonzalez-Perez et al. [2014a], Lacey et al. [2016]) is a semi-analytic model for the creation and evolution of galaxies and uses background cosmological models for its functioning. In earlier GALFORM versions only the total cold gas mass was simulated, however, its complicated dependence on halo-mass prevents the making of assumptions for HI . A physical model was needed to find the dependence between HI and halo mass, an extension which was carried out by Lagos et al. [2011b] whose treatment of star formation allows for the distinction between molecular and atomic Hydrogen content in galaxies. For the cosmological background needed, the P-Millennium N-body simulation was used. It is one of the Millennium series of simulations of dark matter structure formation in cosmological-scale volumes whose cosmological parameters were updated to use Planck values [Ade et al., 2014].

The physical predictions made by these simulations are parametrised by:

$$\frac{M_{\rm HI}}{M_{\rm h}} = A_1 exp\left[-\left(\frac{M_{\rm h}}{M_{\rm break}}\right)^{\alpha}\right] \times \left(\frac{M_{\rm h}}{10^{10}{\rm h}^{-1}{\rm M}_{\odot}}\right)^{\beta} + A_2, \tag{5}$$

where the values used for the parameters are $A_1 = 0.007$, $A_2 = 1.5 \times 10^{-4}$, $M_{\rm break} = 10^{11.5}{\rm h}^{-1}{\rm M}_{\odot}$, $\alpha = 1.5$, $\beta = 0.2$. The choice of values arises from the Gonzalez-Perez et al. [2014b] recalibration. This parametrisation assumes that for low halo masses where $M_{\rm h} << M_{\rm break}$, the proportionality $M_{HI} \propto M_{\rm h}^{1+\beta}$ holds. For large halo masses, $M_{\rm h} >> M_{\rm break}$ the proportionality assumed is $M_{\rm HI} \propto M_{\rm h}$. The HI mass dependence on halo mass can be observed in Fig.1.

These assumptions arise from their findings from the recalibrated models. They show that for dark matter halo masses between $10^9{\rm h}^{-1}{\rm M}_{\odot} \lessapprox {\rm M}_{\rm h} \lessapprox 10^{11}{\rm h}^{-1}{\rm M}_{\odot}$ the median HI mass of central galaxies depends on their host dark matter halo mass. Specifically, the scaling goes as $M_{\rm HI} \propto M_{\rm h}^{1.5}$ whereas in the parametrisation suggested by Baugh et al. [2018] they assume a proportionality of $M_{\rm HI} \propto M_{\rm h}^{1.2}$ for these low masses. This discussion has only taken into account central galaxies, which is an accurate approximation since HI content in dark matter haloes is dominated by central galaxies at these low dark matter halo masses.

At host dark matter halo masses greater than $10^{11}{\rm h}^{-1}{\rm M}_{\odot}$, there is a strong break which is due to AGN feedback. Thus, the HI mass in central galaxies above this host halo mass experiences a significant drop. It is important to note that α determines the sharpness of the break produced at $M_{\rm break}$. The reason for this is that gas cooling is suppressed which implies that the amount of cold gas in central galaxies in these haloes also decreases. Therefore, the galaxy only acquires cold gas via galaxy mergers.

For greater halo masses, $M_{\rm h} > 10^{12}{\rm h}^{-1}{\rm M}_{\odot}$, it is found that satellite galaxies are dominant [Baugh et al., 2018]. As can be seen in Fig.1, the slope of the curve is very similar at low and high values of dark matter halo mass.

2.1.4 Prescription 4: Spinelli et al. [2020]

The fourth, and last prescription studied is presented in Spinelli et al. [2020]. They have used the outputs from the GAlaxy Evolution and Assembly (GAEA) semi-analytic model described in several works such as De Lucia et al. [2014] or Hirschmann et al. [2016].

This model has been run on the Millenium I and II simulations, two large-scale dark matter cosmological simulations which are based on WMAP1 cosmology. It has not been updated with Planck [Ade et al., 2014] values since they expect that the differences will not have a major impact in the predictions. A lower cut on stellar mass ($M_{\rm s}$) for the galaxies has also been applied, lowering the usual thresholds used for GAEA, which allows the study of lower HI masses. These mass thresholds are $M_{\rm s} > 10^{6}{\rm M}_{\odot}$ for Millenium I and $M_{\rm s} > 10^{8}{\rm M}_{\odot}$ for Millenium II.

The prediction made for the HI -halo mass relation is based on the parametrisation by Baugh et al. [2018], however, they have included a cut-off at lower halo mass and re-calibrated the parameters. Thus, they suggest the following HIHM:

$$M_{\rm HI} = M_{\rm h}\left[a_1\left(\frac{M_{\rm h}}{10^{10}{\rm h}^{-1}{\rm M}_{\odot}}\right)^{\beta} \times e^{-\left(\frac{M_{\rm h}}{M_{\rm break}}\right)^{\alpha}} + a_2\right] \times e^{-\left(\frac{M_{\rm min}}{M_{\rm h}}\right)^{\gamma}} \quad (6)$$

where a_1, a_2, α, β, $M_{\rm break}$ and $M_{\rm min}$ are free parameters. The value for γ is kept fixed at $\gamma = 0.5$ since they find it best describes their predictions. Similarly to Baugh et al. [2018], this parametrisation assumes that for low halo masses $M_{\rm HI} \propto M_{\rm h}^{1+\beta}$ and for large halo masses $M_{\rm HI} \propto M_{\rm h}$. However, the mass break occurs at a slightly higher value, at $M_{\rm break} \approx 10^{12}{\rm h}^{-1}{\rm M}_{\odot}$. As for the parameters, their values for $z = 1$, which is the closest redshift to the redshift used in UNIT simulations, are the following: $a_1 = 3.8 \times 10^{-3}$, $a_2 = 1.6 \times 10^{-3}$, $\alpha = 0.24$, $\beta = 1.70$, $log_{10}M_{\rm break} = 8.30$ and $log_{10}M_{\rm min} = -1.3$

Similarly to the results found in Baugh et al. [2018], for lower halo masses the central galaxies dominate, with a break ocurring at $M_{\rm break} \approx 10^{12}{\rm h}^{-1}{\rm M}_{\odot}$ although this break is not as pronounced as in Baugh et al. [2018]. This occurs when AGN feedback becomes more efficient. For higher halo masses, $M_{\rm break} \gtrapprox 10^{12,5}{\rm h}^{-1}{\rm M}_{\odot}$, the satellites are found to be the dominant contribution [Spinelli et al., 2020] and the amount of HI increases with host halo mass. The slope for this increase, however, is lower than the slope for low halo masses.

2.2 Semi-analytical HI prescriptions

Understanding the distribution, formation and evolution of galaxies is vital to develop consistent galaxy models in the field of cosmology.

In this semi-analytic approach we have used a combination of the "Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys" (UNIT) [Chuang et al., 2019], "Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement" algorithm (ROCKSTAR) [Behroozi et al., 2012] and the 'Semi-Analytic Galaxy Evolution" (SAGE) code [Croton et al., 2016] to obtain the cold gas mass, among other observables, for a distribution of dark matter haloes.

The UNIT simulations are based on four N-body dark matter simulations, although only one is used in this work, which contain 4096^3 particles of mass $1.25 \times 10^{9}{\rm h}^{-1}{\rm M}_{\odot}$.

It also utilizes the following parameters: $\Omega_{\rm m,0} = 1 - \Omega_{\Lambda,0} = 0.3089$, $h = 0.6774$, $n_{\rm s} = 0.9667$, and $\sigma_8 = 0.8147$, which have been obtained from Table 4 in Planck Collaboration [2016]. Each simulation has a comoving volume of $V_{\rm box} = 1.0{\rm h}^{-3}{\rm Gpc}^3$. The UNIT simulations output a distribution of dark matter particles.

To generate dark matter haloes, the ROCKSTAR [Behroozi et al., 2012] halo finder is applied in UNIT. These dark matter haloes are then populated with galaxies via SAGE. The haloes are 'seeded' with hot gas that is then allowed to evolve as described in Croton et al. [2016].

SAGE models output, among other properties, the evolution of cold gas in galaxies. However, it does not specifically model the amount of that cold gas which can be found in HI form. Therefore, this has to be calculated a posteriori. For such calculation we have used two different assignment models which are described in the following sections.

In our work, we have used a snapshot at redshift $z = 1.372$ and, following Avila et al. [2021], we have imposed the following cut-offs: $M_\star > 10^9{\rm M}_\odot/{\rm h}$, $M_{\rm CG} > 10^2{\rm M}_\odot/{\rm h}$, $R_{\rm disk} \neq 0$ and $M_{\rm h} > 10^{10.5} M_\odot/h$.

2.2.1 Model 1

From the cold gas mass obtained, we now calculate HI mass following the assumption that all cold gas is neutral. Therefore, the amount of Hydrogen can be estimated from its mass fraction and thus discerned from the other possible elements. To distinguish atomic and molecular Hydrogen, the molecular-to-atomic ratio can be used. For this model, and following Zoldan et al. [2016], a constant molecular-to-atomic ratio is used. Therefore, the HI mass obtained is as follows:

$$M_{\rm HI} = f_{\rm H} \left(1 - \frac{R_{\rm mol}}{R_{\rm mol} + 1}\right) M_{\rm CG}, \tag{7}$$

where $M_{\rm CG}$ is the cold gas mass in the disk. From Big Bang Nucleosynthesis (BBN), the HI mass fraction is $f_{\rm H} = 0.75$, and the molecular-to-atomic ratio is defined as $R_{\rm mol} = M_{\rm H_2}/M_{\rm HI} = 0.4$ (Zwaan et al. [2005]; Keres et al. [2003]). As can be seen, this $R_{\rm mol}$ yields a linear relationship between cold gas mass and HI mass.

We are now interested in the HI mass that is hosted in dark matter haloes as a function of halo mass. To study this dependence, we have used the HI mass calculated with model 1 from the cold gas mass present in each galaxy. Then, the average HI mass was calculated by subdividing the range of host halo masses into bins and, having assigned each halo to its corresponding bin, averaging the HI masses obtained.

In Fig.1 we have shown this average HI mass as a function of host halo mass. Furthermore, we have shown as a 2-dimensional histogram the HI mass predicted by model 1 as a function of host halo mass. This is commented further in section 2.3.

2.2.2 Model 2

Similarly to the previous model, the HI mass is calculated from SAGE cold gas mass but is now also dependent on two other parameters: disk scale length and disk stellar mass that make the molecular-to-atomic ratio to differ. The same Hydrogen mass fraction as before is used to obtain the total Hydrogen mass but the molecular-to-atomic ratio

in this model differs. The ratio used here is also described in Zoldan et al. [2016] and has the following shape:

$$R_{\rm mol} = \left[3,44\ R_{\rm c}^{-0.506} + 4.82\ R_{\rm c}^{-1,054}\right]^{-1} \tag{8}$$

$$R_{\rm c} = \left[K\ r_{\rm disk}^{-4}\ M_{\rm CG} \times (M_{\rm CG} + 0.4\ M_{\star,\,\rm disk})\right]^{0.8} \tag{9}$$

As can be seen, $R_{\rm c}$ is dependent on the disk size $r_{\rm disk}$, cold gas mass in the disk $M_{\rm CG}$, and the disk stellar mass $M_{\star,\,\rm disk}$. These relationships were found by Obreschkow and Rawlings [2009] based on the HI and H_2 gas profiles of disk galaxies studied and modelled in Blitz and Rosolowsky [2006].

The same calculation that was described above for model 1 is performed to obtain the average $M_{\rm HI}$ as a function of halo mass. Thus, the average HI mass as a function of host halo mass using model 2 is shown in Fig.1.

2.3 Discussion

This section has been dedicated to the study of the results obtained for the HI - halo mass relation. As discussed, we have followed two distinct approaches: a semi-analytical method using galaxy observables outputted by SAGE; and a theoretical one studying the models from the literature.

From the semi-analytical method we obtained the cold gas mass, among other parameters, for each one of the galaxies in the simulation. Then we applied two different models to obtain the neutral Hydrogen mass.

We represent the HI mass-halo mass relation for model 1 via a 2-dimensional histogram in Fig.1. This histrogram is shown as a heatmap where each pixel represents the amount of objects that there are in the simulation with that halo mass and that HI mass. This is shown in logarithmic scale and having used model 1 for the calculation of HI . For low halo masses we observe a clear area where the probability of locating a galaxy with that HI mass is certainly higher. This trend is, however, lost at medium and high halo masses.

We observe two dotted lines which represent the average HI mass for models 1 (dark purple) and 2 (black). We can also observe the standard deviation for model 1. For clarity purposes, only the error for model 1 is shown but the standard deviation for model 2 has been calculated and has been found to be equivalent.

The results for both models are quite similar. However, model 1, which only depends on cold gas mass, consistently predicts higher HI values. The difference for the predicted HI masses between the models is approximately constant, albeit the difference is a bit bigger for halo masses in the lower end. This difference seems to reach its highest point at around $log_{10}(M_{\rm h}) \approx 11.5$. At higher halo masses this difference remains small and quite constant.

In this figure we have also shown the HIHM relation for the four theoretical models considered. For low halo masses they all predict low HI masses whose increase follow power laws. It can be seen that, although the starting points differ slightly, the steepness of the curves is quite similar. These low values are due to supernovae feedback whose effect is more notable in the lower halo mass end.

However, in the high-mass end both Padmanabhan and Kulkarni [2017] and Bagla et al. [2010] predict a decreasing amount of HI . This decrease is steeper for Padmanab-

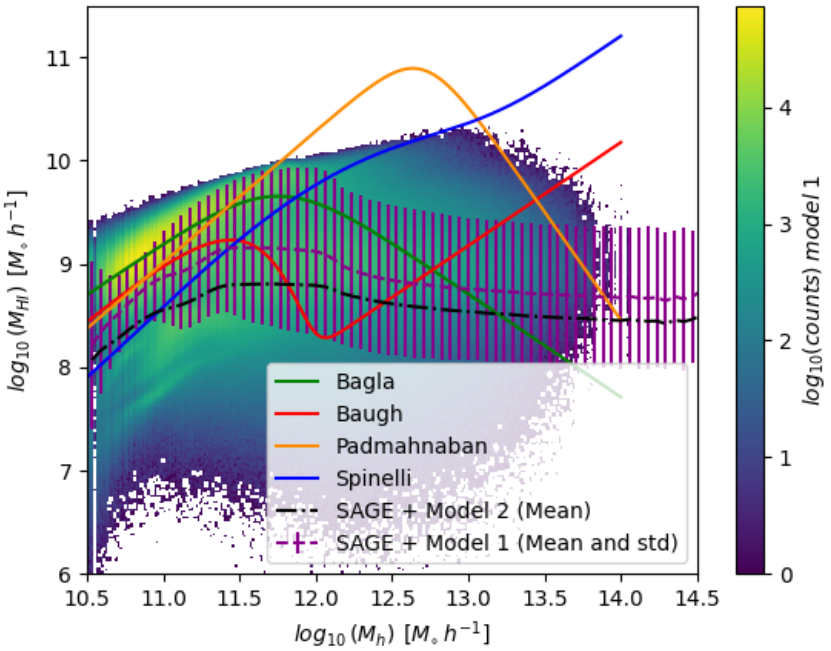

Figure 1: The HI mass as a function of host dark matter halo mass is presented here. The $M_{\rm HI}$ obtained from model 1 has been represented as a heatmap with logarithmic scaling. Furthermore, the mean $M_{\rm HI}$ and standard deviation found from model 1 are represented in purple. The mean $M_{\rm HI}$ for model 2 is represented as a black dotted line. The results from the theoretical prescriptions described in section 2.1 have also been represented. Padmanabhan and Kulkarni [2017] can be found in orange, Bagla et al. [2010] in green, Baugh et al. [2018] in red and Spinelli et al. [2020] in blue.

han and Kulkarni [2017] than Bagla et al. [2010]. On the other hand, both Baugh et al. [2018] and Spinelli et al. [2020] predict an increase in HI mass for high halo masses. This difference between the models is due to the fact that both Baugh et al. [2018] and Spinelli et al. [2020] have considered subhaloes which means that one dark matter halo may host more than one galaxy. This is relevant because at higher halo masses satellite galaxies begin to dominate as is shown, for example, in Baugh et al. [2018]. Since satellites begin to dominate, we expect an increase in HI for high halo masses in models which do consider subhaloes, as is the case of Baugh et al. [2018] and Spinelli et al. [2020]. On the other hand, for models which do not consider subhaloes we expect a decrease in HI mass with increasing halo mass due to the onset of AGN feedback. This is the case of Padmanabhan and Kulkarni [2017] and Bagla et al. [2010].

For haloes of medium masses, both Baugh et al. [2018] and Spinelli et al. [2020] show a slight decrease in HI mass before continuing their increase. Baugh et al. [2018] has shown that this decrease is due to the onset of AGN feedback which soon is dominated by satellite galaxies.

3 HI density

In the previous section we have described the method used for obtaining the HI mass for different dark matter haloes as a function of halo mass $M_{\rm h}$. We are now interested in how the HI mass found with models 1 and 2 is distributed in space, but first, we must calculate the total amount of HI . This is done by finding the average density normalized to the critical density for both models:

$$\Omega_{\rm HI}^{\rm avg} = \frac{\rho_{\rm HI}^{\rm avg}}{\rho_{\rm crit}} = \frac{1}{\rho_{\rm crit}} \frac{\sum M_{\rm HI}}{V_{\rm box}} = \begin{cases} 1.49 \times 10^{-3} \\ 7.49 \times 10^{-4} \end{cases} \tag{10}$$

where $\rho_{\rm crit} = 7.09 \times 10^{10} {\rm h}^2 {\rm M}_\odot / {\rm Mpc}^3$ is the critical density of the universe and $V_{\rm box}$ is the total volume of the simulation box. The upper value of Eq. 10 has been found using Eq. 7 (model 1) and the bottom value with Eq. 8 (model2).

To find the distribution in space of HI , we must now calculate the density as a function of position. This is done by dividing the simulation volume into smaller cubes of side $L_{\rm cell} = 5$Mpc/h and calculating the density in each. This results in $\left(\frac{L_{\rm box}}{L_{\rm cell}}\right)^3 = 200^3$ bins. The python library SciPy [Virtanen et al., 2020] is used, specifically the function *scipy.stats.binned_statistic_dd* where the statistic used was our own density function. This density is calculated for both models. For the sake of representation, we have shown in Fig.2 5Mpc/h wide slices of the density contrast that were taken on the x-y plane. The density contrast is defined as:

$$\delta = \frac{\rho_{\rm HI} - \rho_{\rm HI}^{\rm avg}}{\rho_{\rm HI}^{\rm avg}} \tag{11}$$

In the Fig.2, the density contrast calculated for both models is shown, as well as the difference between the two models. For visual purposes, we have represented $log_{10}(2+\delta)$ for models 1 and 2.

Both models show very similar large scale structures, and, as such, the plot showing the difference is quite homogeneous. There are, however, some regions where the difference between the models is greater. These positions coincide with the overdensities. Therefore, the difference is enhanced in high density regions. This is expected because the percentage error is approximately the same across the whole simulation independently of the density. This means that in overdense regions the density difference must be higher.

4 HI Clustering

In the previous section we have studied the HI density as a function of position. HI lies in galaxies which are placed in dark matter haloes and thus, studying the clustering of HI can give us information about how this gas and galaxies distribute on their host dark matter haloes. The study of clustering is done with statistical tools such as correlations functions, specifically with the 2-point correlation function (2PCF) which is the most widely used. The 2PCF indicates the probability of there being an overdensity at a certain distance when compared to a random Poisson distribution [Storey-Fisher and Hogg, 2021]. There are several estimators to calculate the 2PCF with one of the most

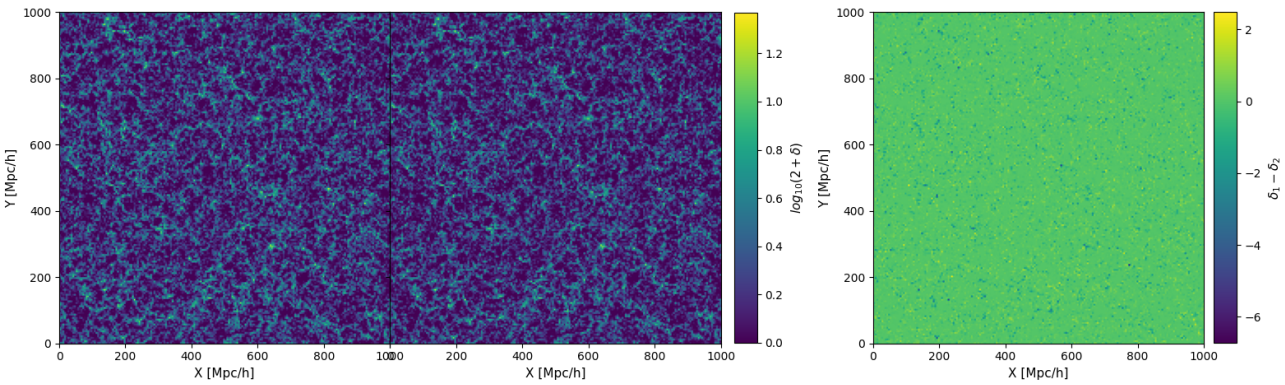

Figure 2: The left and center panels show the HI density contrast (δ) of a 5Mpc/h wide slice taken on the x-y plane. The distribution of HI shown on the left panel was calculated using model 1, whereas the one shown on the center panel was calculated using model 2. To increase the contrast, the $log_{10}(2+\delta)$ was represented in both panels. Finally, the right panel shows the difference between the density contrast of the two models.

common being the Natural Estimator [Peebles and Hauser, 1974]:

$$\xi = \frac{DD(r)}{RR(r)} - 1 \tag{12}$$

where $DD(r)$ is the number of pairs of galaxies that can be found at a certain distance in the data sample and $RR(r)$ is the number of pairs of galaxies that can be found at a certain distance in the random sample. In our case we are not only interested in the number of galaxies and their positions but also on their HI masses. Thus we have to assign a weighting scheme to the correlation function for each galaxy. These weights will correspond with the HI mass for each.

The calculation of the 2PCF for such large samples is very computationally demanding, thus, routines to improve performance have been created and are publicly available. Such is the case of Corrfunc [Sinha and Garrison, 2020], a parallelized implementation which we have used for the calculation of the 2PCFs in this work and which, in this case, uses the Natural Estimator [Sinha and Garrison, 2020]. Since our 2PCF was going to be weighted by the density, our first approach was to use the density we had found as a function of position. As explained before, the density was actually calculated in small bins of sides of 5Mpc/h. This way, for the 2PCF we tried using this density as the weight and the center of the bins as the positions. However, this did not work as expected since Corrfunc is not designed to work with equally-spaced positions and thus the output was incorrect.

Since the previous method did not work, we decided to follow the more traditional approach and use the actual positions of the cold gas to calculate the 2PCF. As weights we used the HI masses we calculated for each of the models. However, our simulation has approximately 60M galaxies which is an excessive number for a calculation of these characteristics. Therefore, we decided to select random portions from the whole galaxy sample. Specifically, we used the following sample sizes: 0.5M, 10M, 30M and 50M.

The results are shown in Fig.3 where we have plotted the correlation function for distances between r = 5Mpc/h and r = 150Mpc/h. We must note however that

the correlation function $\xi(r)$ plotted has suffered some transformations. Firstly, it is multiplied by the distance squared which is common practice to bring out the maximum at around 100Mpc/h which, as we explain later, corresponds with the Baryonic Acoustic Oscillations peak. Secondly, it has been multiplied by a factor which converts it to temperature variation. The reason behind this is that surveys actually measure the intensity of the 21cm line which is closely related to the temperature of the gas and to the mass of HI [Dickey, 1990]. The conversion factor is indicated in [Battye et al., 2013] and is:

$$\bar{T} = 44\mu\mathrm{K}\left(\frac{\Omega_{\mathrm{HI}}(z)h}{2.45\times10^{-4}}\right)\left(\frac{(1+z)^2 H_0}{H(z)}\right) = \begin{cases} 0.517\,\mathrm{mK} \\ 0.260\,\mathrm{mK} \end{cases}, \tag{13}$$

where z is the redshift, Ω_{HI} is the average density normalized to the critical density, $H(z)$ is the Hubble constant at redshift z = 1.372 and H_0 is the Hubble constant in the present universe. The two values found correspond respectively to model 1 and model 2. The conversion coefficients differ because of the distinct average HI density parameter values for each model.

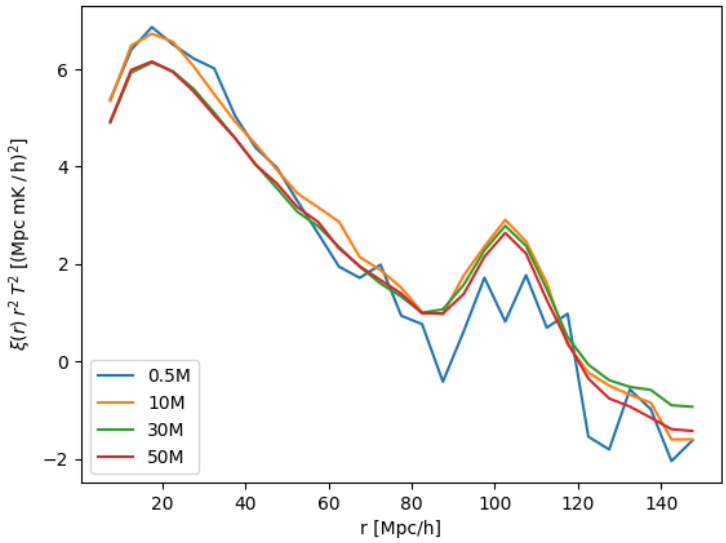

Figure 3: Representation of the two point correlation function multiplied by the temperature conversion factor squared and the distance squared, for the HI mass distribution calculated using model 1. This is done for different sample sizes: 0.5M, 10M, 30M and 50M

Finally, we go on to examine Fig.3 where the two point correlation function is represented. In this case the data for model 1 was used together with different sample sizes ranging 0.5M to 50M data points. As can be seen, smaller samples result in great fluctuations due to the presence of the Poisson noise inherent of a random distribution. However, as sample size increases, we observe the stabilization of the plot with the results provided by 30M and 50M being highly similar. As for the shape of the plot, we observe how the clustering decreases with length, i.e., there is less HI mass as the distance increases. Being gravity a force which decreases with distance, this is no

surprise . However, what does catch our attention is the presence of a maximum at around 100Mpc/h. At these lengths there is an increase in clustering which is due to the Baryonic Acoustic Oscillations (BAO). During the Recombination epoch, the Universe is finally cool enough to become neutral. This resulted in the drastic drop of the sound speed which freezes the sound waves in. The imprint of this feature in the matter distribution is what we call the BAO whose characteristic length is 100Mpc/h.

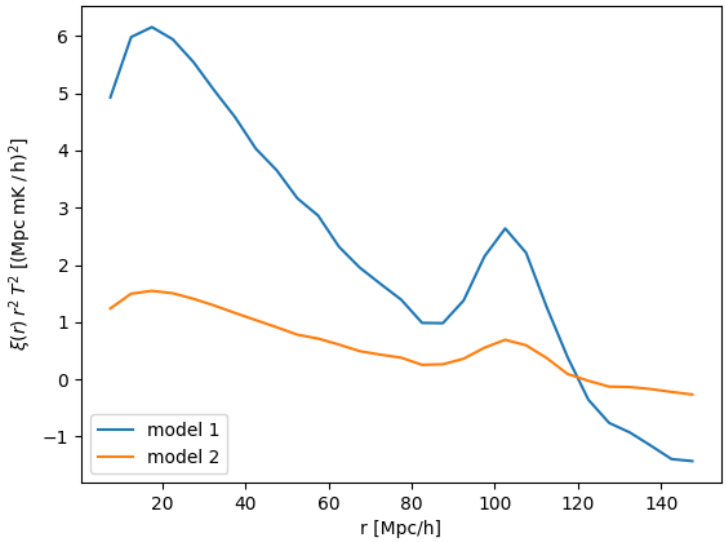

Figure 4: Representation of the 2PCF multiplied by the temperature conversion factor squared and the distance squared for model 1 and model 2 in the range $r = 5 - 150$ Mpc/h. For both models we consider a galaxy sample size of 50M.

In Fig. 4 we compare both models by plotting the correlation function for each and using in both the same sample size of 50M galaxies.

We can observe that both models predict a decrease in clustering with increasing distance as well as a peak at around $r = 100$Mpc/h. However, they clearly present differences with the steepness of the decrease being much more pronounced for model 1. Similarly, the height of the peak at $r = 100$Mpc/h is also more pronounced for model 1. The main reason for these differences is the temperature coefficient whose value depends on the average normalized density. This density is lower for model 2 thus enhancing the differences between the models.

Model 2 predicts a higher clustering for distances greater than $\approx 120 Mpc/h$ where the correlation function becomes negative. This means that there is less clustering than there would be in a random Poisson distribution. Once again the difference between the models is enhanced by the temperature coefficient.

In order to study the small scale predictions of the models we have also calculated the correlation function at smaller scales. The results ($\xi(r)$) for model 1 with different sample sizes are plotted in Fig.**??** in logarithmic scales. It is important to note that for small scales it was actually possible to plot the whole data set of points which allows the evaluation of the adequacy of our estimation. Since the results obtained with the bigger sample sizes and the whole set of data are extremely similar, we can confirm its validity.

We now observe the logarithmic decrease in clustering as distance increases. This decrease can also be seen in Fig.5 where models 1 and 2 are compared, again in logarithmic scale with the correlation function calculated from the whole data set. In this case, and contrary to what happened at bigger scales, model 2 predicts higher clustering. This difference is due to the use of the temperature normalization in the previous Figure. As can be observed, this difference in clustering is higher at the smallest scales, with model 1 predicting only 60% of the clustering that model 2 predicts. As the scale increases, the difference in clustering is reduced and model 1 stabilizes with a prediction of 90% of the clustering predicted by model 2.

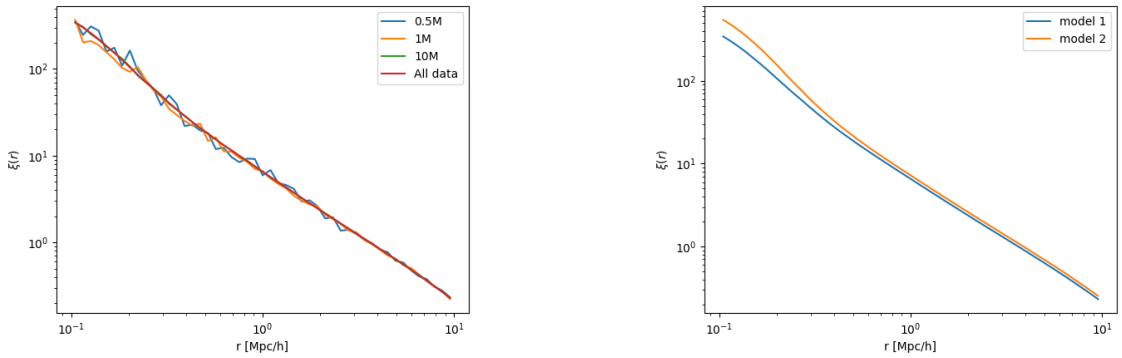

Figure 5: Two point correlation function in real space. *Left:* Small scales for model 1 using different sample sizes: 0.5M, 1M, 10M and the whole data set. *Right:* Comparison between the two models for HI using the whole data set.

5 Conclusions

In this work we have studied the distribution of neutral hydrogen in the Universe at a redshift $z = 1.372$. For this, we have studied how HI mass is distributed as a function of host halo mass by examining four theoretical prescriptions proposed in the literature, specifically [Padmanabhan and Kulkarni, 2017], [Bagla et al., 2010], [Baugh et al., 2018] and [Spinelli et al., 2020]. While all the models are different, they all predict low HI mass for low halo masses due to supernovae feedback. As host halo mass increases, all show a decrease in HI mass due to the onset of AGN feedback. However, at the highest host halo masses, the predictions differ more notably due to the consideration, or lack thereof, of subhaloes. When considered, high HI mass is predicted due to the domination of satellite galaxies and when not considered, low HI mass is predicted due to the efficacy of AGN feedback.

Connecting halo masses directly with cold gas might be too big of a leap, thus, we have obtained its mass, among other parameters, from SAGE semi-analytical models. These parameters have been employed to compare the HI mass distribution predicted by two models, model 1 and model 2, that were suggested in the literature in Zoldan et al. [2016]. After the comparison of the HI mass and halo mass relation of both two models and of the four theoretical prescriptions, we proceeded to study these two models' predictions about HI distribution and its large scale and small scale structure.

This was achieved by computing the HI density contrast across the simulation cube and identifying the differences predicted by the models. Positions of high density peaks are very similar between the two models. Nevertheless, density differences are enhanced precisely in those places. Further information about the clustering was achieved via the calculation and study of a weighted two point correlation function both for large and small scales. For larger scales we have obtained a difference between the models that was largely due to the different temperature coefficient values. As for the small scale, we observed that model 2 predicts higher clustering.

6 Acknowledgements

Quiero agradecerle a mi tutor Bernhard toda la dedicación e interés sin el cual este trabajo no habría salido adelante. Me ha encantado compartir este proyecto contigo, pues no podría haber pedido un mejor tutor.

Por supuesto también agradecer a Violeta su ayuda y su guía que definitivamente han ayudado a hacer de este trabajo el que es.

A mis padres, por siempre apoyarme en mis estudios. No lo habría conseguido sin vosotros.

A Mª Angeles, que no sólo es mi profesora de violín, sino mi familia después de tantos años juntas. No sería quien soy sin la música y el violín, pero sobre todo, sin el criterio que me has ayudado a desarrollar. Este trabajo no habría sido posible sin la música como la entiendo en mi vida y entiendo la música como lo hago gracias a ti.

A mis personas cercanas por estar ahí, por hacerme reír, por verme llorar y por dejarme desahogar mi frustación. A Dunaeb, A Sophia, a todes les Antarianes, a Diana, a Ruth, a Melia, a Aleph, a Delfa...sois muchos para nombraros pero a todos os agradezco estar ahí.

Bibliography

P. A. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, and et al. planck2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571, 2014. doi: 10.1051/0004-6361/201321591.

S. Avila, B. Vos-Ginés, S. Cunnington, A. R. H. Stevens, G. Yepes, A. Knebe, and C.-H. Chuang. Hi im correlation function from UNIT simulations: BAO and observationally induced anisotropy. *Monthly Notices of the Royal Astronomical Society*, 510(1):292–308, nov 2021. doi: 10.1093/mnras/stab3406. URL https://doi.org/10.1093%2Fmnras%2Fstab3406.

D. J. Bacon, R. A. Battye, P. Bull, S. Camera, P. G. Ferreira, I. Harrison, D. Parkinson, A. Pourtsidou, M. G. Santos, L. Wolz, and et al. Cosmology with phase 1 of the square kilometre array red book 2018: Technical specifications and performance forecasts. *Publications of the Astronomical Society of Australia*, 37:e007, 2020. doi: 10.1017/pasa.2019.51.

J. Bagla and S. Ray. Performance characteristics of treepm codes. *New Astronomy*, 8 (7):665–677, 2003.

J. S. Bagla. TreePM: A code for cosmological n-body simulations. *Journal of Astrophysics and Astronomy*, 23(3-4):185–196, dec 2002. doi: 10.1007/bf02702282. URL https://doi.org/10.1007%2Fbf02702282.

J. S. Bagla, N. Khandai, and K. K. Datta. H i as a probe of the large-scale structure in the post-reionization universe. *Monthly Notices of the Royal Astronomical Society*, 407(1):567–580, 08 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16933.x. URL https://doi.org/10.1111/j.1365-2966.2010.16933.x.

R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou. H i intensity mapping: a single dish approach. *Monthly Notices of the Royal Astronomical Society*, 434(2):1239–1256, jul 2013. doi: 10.1093/mnras/stt1082. URL https://doi.org/10.1093%2Fmnras%2Fstt1082.

C. M. Baugh, C. G. Lacey, C. S. Frenk, G. L. Granato, L. Silva, A. Bressan, A. J. Benson, and S. Cole. Can the faint submillimetre galaxies be explained in the λ cold dark matter model? *Monthly Notices of the Royal Astronomical Society*, 356(3): 1191–1200, 2005. doi: 10.1111/j.1365-2966.2004.08553.x.

C. M. Baugh, V. Gonzalez-Perez, C. D. P. Lagos, C. G. Lacey, J. C. Helly, A. Jenkins, C. S. Frenk, A. J. Benson, R. G. Bower, and S. Cole. Galaxy formation in the Planck Millennium: the atomic hydrogen content of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 483(4):4922–4937, 12 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty3427. URL https://doi.org/10.1093/mnras/sty3427.

P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. THE ROCKSTAR PHASE-SPACE TEMPORAL HALO FINDER AND THE VELOCITY OFFSETS OF CLUSTER CORES. *The Astrophysical Journal*, 762(2):109, dec 2012. doi: 10.1088/0004-637x/762/2/109. URL https://doi.org/10.1088%2F0004-637x%2F762%2F2%2F109.

A. Bera, N. Kanekar, J. N. Chengalur, and J. S. Bagla. The hi mass function of star-forming galaxies at z $\approx$ 0.35. *The Astrophysical Journal Letters*, 940(1), 2022. doi: 10.3847/2041-8213/ac9d32.

S. Bharadwaj, B. B. Nath, and S. K. Sethi. Using HI to probe large scale structures at z$\sim$3. *Journal of Astrophysics and Astronomy*, 22(1):21–34, mar 2001. doi: 10.1007/bf02933588. URL `https://doi.org/10.1007%2Fbf02933588`.

L. Blitz and E. Rosolowsky. The role of pressure in gmc formation ii: The h_2-pressure relation. *The Astrophysical Journal*, 650(2):933–944, oct 2006. doi: 10.1086/505417. URL `https://doi.org/10.1086%2F505417`.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *mnras*, 370(2):645–655, Aug. 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

G. Chuang, Chia-Hsunand Yepes, F.-S. Kitaura, M. Pellejero-Ibanez, S. Rodríguez-Torres, Y. Feng, R. B. Metcalf, R. H. Wechsler, C. Zhao, C.-H. To, S. Alam, A. Banerjee, J. DeRose, C. Giocoli, A. Knebe, and G. Reyes. UNIT project: Universe N-body simulations for the Investigation of Theoretical models from galaxy surveys. *MNRAS*, 487(1):48–59, July 2019. doi: 10.1093/mnras/stz1233.

S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. Hierarchical galaxy formation. *mnras*, 319(1):168–204, Nov. 2000. doi: 10.1046/j.1365-8711.2000.03879.x.

D. J. Croton, A. R. H. Stevens, C. Tonini, T. Garel, M. Bernyk, A. Bibiano, L. Hodkinson, S. J. Mutch, G. B. Poole, and G. M. Shattow. SEMI-ANALYTIC GALAXY EVOLUTION (SAGE): MODEL CALIBRATION AND BASIC RESULTS. *The Astrophysical Journal Supplement Series*, 222(2):22, feb 2016. doi: 10.3847/0067-0049/222/2/22. URL `https://doi.org/10.3847%2F0067-0049%2F222%2F2%2F22`.

G. De Lucia, L. Tornatore, C. S. Frenk, A. Helmi, J. F. Navarro, and S. D. M. White. Elemental abundances in Milky Way-like galaxies from a hierarchical galaxy formation model. *mnras*, 445(1):970–987, Nov. 2014. doi: 10.1093/mnras/stu1752.

J. M. Dickey. *Measuring Atomic Hydrogen Masses Using the 21-cm Line*, pages 473–482. Springer Netherlands, Dordrecht, 1990. ISBN 978-94-009-0595-5. doi: 10.1007/978-94-009-0595-5_19. URL `https://doi.org/10.1007/978-94-009-0595-5_19`.

J. Fu, G. Kauffmann, M.-l. Huang, R. M. Yates, S. Moran, T. M. Heckman, R. Davé, Q. Guo, and B. M. Henriques. Star formation and metallicity gradients in semi-analytic models of disc galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 434(2):1531–1548, 2013. doi: 10.1093/mnras/stt1117.

V. Gonzalez-Perez, C. G. Lacey, C. M. Baugh, C. D. Lagos, J. Helly, D. J. Campbell, and P. D. Mitchell. How sensitive are predicted galaxy luminosities to the choice of stellar population synthesis model? *Monthly Notices of the Royal Astronomical Society*, 439(1):264–283, 2014a. doi: 10.1093/mnras/stt2410.

V. Gonzalez-Perez, C. G. Lacey, C. M. Baugh, C. D. P. Lagos, J. Helly, D. J. R. Campbell, and P. D. Mitchell. How sensitive are predicted galaxy luminosities to the choice of stellar population synthesis model? *Monthly Notices of the Royal Astronomical Society*, 439(1):264–283, 02 2014b. ISSN 0035-8711. doi: 10.1093/mnras/stt2410. URL https://doi.org/10.1093/mnras/stt2410.

M. Hirschmann, G. De Lucia, and F. Fontanot. Galaxy assembly, stellar feedback and metal enrichment: the view from the GAEA model. *mnras*, 461(2):1760–1785, Sept. 2016. doi: 10.1093/mnras/stw1318.

D. Keres, M. S. Yun, and J. S. Young. CO Luminosity Functions for Far-Infrared- and B-Band-selected Galaxies and the First Estimate for Ω_{HI}+H_2. *Astrophysical Journal*, 582(2):659–667, Jan. 2003. doi: 10.1086/344820.

N. Khandai and J. Bagla. A modified treepm code. *Research in Astronomy and Astrophysics*, 9(8):861, 2009.

C. G. Lacey, C. M. Baugh, C. S. Frenk, A. J. Benson, R. G. Bower, S. Cole, V. Gonzalez-Perez, J. C. Helly, C. D. Lagos, P. D. Mitchell, and et al. A unified multiwavelength model of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 462 (4):3854–3911, 2016. doi: 10.1093/mnras/stw1888.

C. d. Lagos, C. M. Baugh, C. G. Lacey, A. J. Benson, H.-S. Kim, and C. Power. Cosmic evolution of the atomic and molecular gas contents of galaxies. *Monthly Notices of the Royal Astronomical Society*, 418(3):1649–1667, 2011a. doi: 10.1111/j.1365-2966.2011.19583.x.

C. d. Lagos, C. G. Lacey, C. M. Baugh, R. G. Bower, and A. J. Benson. On the impact of empirical and theoretical star formation laws on galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 416(2):1566–1584, 2011b. doi: 10.1111/j.1365-2966.2011.19160.x.

M. D. Lallo. Experience with the hubble space telescope: 20 years of an archetype. *Optical Engineering*, 51(1):011011, feb 2012. doi: 10.1117/1.oe.51.1.011011. URL https://doi.org/10.1117%2F1.oe.51.1.011011.

B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the relationship between stellar mass and halo mass at low and high redshift. *The Astrophysical Journal*, 710(2):903–923, 2010. doi: 10.1088/0004-637x/710/2/903.

B. P. Moster, T. Naab, and S. D. M. White. Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 428(4):3121–3138, dec 2012. doi: 10.1093/mnras/sts261. URL https://doi.org/10.1093%2Fmnras%2Fsts261.

J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *APJ*, 462:563, May 1996. doi: 10.1086/177173.

D. Obreschkow and S. Rawlings. THE COSMIC DECLINE IN THE h_2 /h i RATIO IN GALAXIES. *The Astrophysical Journal*, 696(2):L129–L132, apr 2009. doi: 10.1088/0004-637x/696/2/l129. URL https://doi.org/10.1088%2F0004-637x%2F696%2F2%2Fl129.

H. Padmanabhan and G. Kulkarni. Constraints on the evolution of the relationship between h i mass and halo mass in the last 12≈ gyr. *Monthly Notices of the Royal Astronomical Society*, 470(1):340–349, may 2017. doi: 10.1093/mnras/stx1178. URL https://doi.org/10.1093%2Fmnras%2Fstx1178.

P. J. E. Peebles and M. G. Hauser. Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. *Astrophysical Journal Supplement*, 28:19, Nov. 1974. doi: 10.1086/190308.

Planck Collaboration. Planck 2015 results. xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, sep 2016. doi: 10.1051/0004-6361/201525830. URL https://doi.org/10.1051%2F0004-6361%2F201525830.

J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, Aug. 2012. doi: 10.1088/0034-4885/75/8/086901.

D. Rouan. *VLT*, pages 1750–1753. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-11274-4. doi: 10.1007/978-3-642-11274-4_1668. URL https://doi.org/10.1007/978-3-642-11274-4_1668.

M. G. Santos, M. Cluver, M. Hilton, M. Jarvis, G. I. G. Jozsa, L. Leeuw, O. Smirnov, R. Taylor, F. Abdalla, J. Afonso, D. Alonso, D. Bacon, B. A. Bassett, G. Bernardi, P. Bull, S. Camera, H. C. Chiang, S. Colafrancesco, P. G. Ferreira, J. Fonseca, K. van der Heyden, I. Heywood, K. Knowles, M. Lochner, Y.-Z. Ma, R. Maartens, S. Makhathini, K. Moodley, A. Pourtsidou, M. Prescott, J. Sievers, K. Spekkens, M. Vaccari, A. Weltman, I. Whittam, A. Witzemann, L. Wolz, and J. T. L. Zwart. Meerklass: Meerkat large area synoptic survey, 2017.

M. Sinha and L. H. Garrison. CORRFUNC - a suite of blazing fast correlation functions on the CPU. *Monthly Notices of the Royal Astronomical Society*, 491(2):3022–3041, Jan 2020. doi: 10.1093/mnras/stz3157.

M. Spinelli, A. Zoldan, G. D. Lucia, L. Xie, and M. Viel. The atomic hydrogen content of the post-reionization universe. *Monthly Notices of the Royal Astronomical Society*, 493(4):5434–5455, mar 2020. doi: 10.1093/mnras/staa604. URL https://doi.org/10.1093%2Fmnras%2Fstaa604.

K. Storey-Fisher and D. W. Hogg. Two-point statistics without bins: A continuous-function generalization of the correlation function estimator for large-scale structure. *The Astrophysical Journal*, 909(2):220, mar 2021. doi: 10.3847/1538-4357/abdc21. URL https://doi.org/10.3847%2F1538-4357%2Fabdc21.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson,

C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17: 261–272, 2020. doi: 10.1038/s41592-019-0686-2.

A. K. Weigel, K. Schawinski, and C. Bruderer. Stellar mass functions: Methods, systematics and results for the local universe. *Monthly Notices of the Royal Astronomical Society*, 459(2):2150–2187, 2016. doi: 10.1093/mnras/stw756.

A. Zoldan, G. D. Lucia, L. Xie, F. Fontanot, and M. Hirschmann. H i-selected galaxies in hierarchical models of galaxy formation and evolution. *Monthly Notices of the Royal Astronomical Society*, 465(2):2236–2253, nov 2016. doi: 10.1093/mnras/stw2901. URL `https://doi.org/10.1093%2Fmnras%2Fstw2901`.

M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. The HIPASS catalogue: Ω_{HI} and environmental effects on the HI mass function of galaxies. *Monthly Notices Royal Astronomical Society*, 359(1):L30–L34, May 2005. doi: 10.1111/j.1745-3933.2005.00029.x.