MODELING THE LARGE-SCALE BIAS OF NEUTRAL HYDROGEN

Felipe A. Marín\textsuperscript{1,2}, Nickolay Y. Gnedin\textsuperscript{1,2,3}, Hee-Jong Seo\textsuperscript{3}, and Alberto Vallinotto\textsuperscript{3}

\textsuperscript{1} Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA; fmarin@oddjob.uchicago.edu
\textsuperscript{2} Kavli Institute for Cosmological Physics and Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637, USA
\textsuperscript{3} Center for Particle Astrophysics, Fermi National Accelerator Laboratory, P.O. Box 500 Kirk Rd. & Pine St., Batavia, IL 60510-0500, USA; gnedin@fnal.gov, sheejong@fnal.gov, avalli@fnal.gov

Received 2009 October 30; accepted 2010 May 15; published 2010 July 9

ABSTRACT

We present new analytical estimates of the large-scale bias of neutral hydrogen (\textsc{H}i). We use a simple, non-parametric model which monotonically relates the total mass of a halo \(M_{\text{tot}}\) with its \textsc{H}i mass \(M_{\text{H}i}\) at zero redshift; for earlier times we assume limiting models for the \(\Omega_{\text{H}i}\) evolution consistent with the data presently available, as well as two main scenarios for the evolution of our \(M_{\text{H}i} - M_{\text{tot}}\) relation. We find that both the linear and the first nonlinear bias terms exhibit a strong evolution with redshift, regardless of the specific limiting model assumed for the \textsc{H}i density over time. These analytical predictions are then shown to be consistent with measurements performed on the Millennium Simulation. Additionally, we show that this strong bias evolution does not sensibly affect the measurement of the \textsc{H}i power spectrum.

Key words: diffuse radiation – large-scale structure of universe

Online-only material: color figures

1. INTRODUCTION

The 21 cm emission line of neutral hydrogen has been the workhorse of astronomy for over 50 years. Its use for galactic astronomy cannot be overestimated, and it now promises to revolutionize cosmology as well (cf. Furlanetto et al. 2006; Barkana & Loeb 2007; Pritchard & Loeb 2008; Mao et al. 2008; Visbal et al. 2009, and references therein).

Compared to present probes of large-scale structure, the 21 cm intensity mapping technique (Chang et al. 2008a; Wyithe et al. 2008) offers three advantages. First, it allows us to reconstruct the spatial distribution of \textsc{H}i in the universe over volumes much larger than the ones currently probed by galaxy redshift surveys. Since the precision with which the power spectrum of density fluctuations can be measured is approximately proportional to the number of independent Fourier modes that fit into the survey volume, a larger survey volume leads to higher precision in the measured power spectrum and hence cosmological parameters (Loeb & Wyithe 2008). Second, 21 cm intensity mapping allows us to probe a very wide range of redshifts and therefore dark energy properties (see, for instance, Hu & White 1996; Eisenstein et al. 1998; Eisenstein 2002; Blake & Glazebrook 2003; Linder 2003; Hu & Haiman 2003; Seo & Eisenstein 2003). The distinct oscillatory feature of the BAO allows us to isolate and measure the BAO information independent of the overall broadband shape of the \textsc{H}i power spectrum. Therefore, in principle, BAO measurements do not require a knowledge of the \textsc{H}i bias, i.e., the relation between the neutral hydrogen and the total matter clustering, which affects the overall shape of the \textsc{H}i power spectrum.

However, understanding the \textsc{H}i bias is important for the 21 cm BAO surveys in the following ways. First, the prediction of the signal to noise of the BAO measurements is sensitive to the knowledge of the clustering bias of \textsc{H}i (Chang et al. 2008b; Seo et al. 2010). Second, understanding \textsc{H}i bias will help us to extract cosmological information from the broadband shape of the power spectrum in addition to the BAO. Third, the nonlinear effects on the BAO, such as the degradation and the shift of the feature, may differ for different biased tracers (Padmanabhan & White 2009; K. Metha et al. 2010, in preparation). Knowing the properties of \textsc{H}i bias will allow us to approximately estimate the degree of the expected nonlinear effect on the BAO in the \textsc{H}i distribution.

In this work, we focus on two critical aspects of the large-scale \textsc{H}i bias: its dependence on the relation between \textsc{H}i mass and halo mass and its evolution with redshift. To study the first aspect, we use a non-parametric model where we assume a one-to-one correspondence between a \textsc{H}i mass present in a dark matter halo \(M_{\text{H}i}\) and the total mass \(M_{\text{tot}}\) of this halo. Using this \(M_{\text{H}i} - M_{\text{tot}}\) relation, we estimate the large-scale bias of the \textsc{H}i using a model containing elements of the halo occupation distribution (HOD) formalism and the \textsc{H}i mass function measured at \(z = 0\) (Zwaan et al. 2005). We then use this relation to paint \textsc{H}i on Millennium Simulation halos (Springel et al. 2005) and to measure the power spectrum of neutral hydrogen in order to test our model assumptions.

With respect to the redshift evolution of the bias, it is absolutely critical to know how the \textsc{H}i mass function evolves with time. However, to the best of our knowledge, little is known about the redshift evolution of \textsc{H}i both on the theoretical and the
observational sides (but see Wyithe et al. 2009). A detailed study of the bias evolution requires two extra pieces of information which are currently poorly constrained by observational data: the evolution of the total H\textsubscript{i} mass in the universe and the evolution of the H\textsubscript{i} mass function. For both of these, we empirically assume limiting evolution models consistent with the few available data that should bracket the actual evolution. We then proceed to predict the bias evolution in these scenarios, being aware of the fact that the actual evolution will lie somewhere in between these limiting cases.

The paper is organized as follows. In Section 2 we introduce the basics of our model, we derive the relation \( M_{\text{H\textsubscript{i}}} - M_{\text{tot}} \) at \( z = 0 \), we introduce the limiting cases for its evolution, and we define the bias. In Section 3, we present our main results: analytic predictions for the evolution of H\textsubscript{i} bias, measurement of its scale dependence on the Millennium Simulation, and how it would affect measurements of the power spectrum. Finally, we conclude in Section 4. To compare analytical and numerical results, throughout the paper we use a flat ΛCDM cosmology with \( \Omega_m = 0.25 \), \( \sigma_8 = 0.9 \), Hubble parameter \( h = 0.73 \), and initial power spectrum index of \( n_s = 1 \), consistent with the cosmological parameters used in the Millennium Simulation.

2. BACKGROUND

2.1. Basic Formalism

The HOD model (cf. Cooray & Sheth 2002; Berlind & Weinberg 2002, and references therein) provides a parameterized prescription for the spatial distribution of objects that can be found inside dark matter halos. Originally designed to model the distribution of galaxies, it can also be used to model the spatial distributions of other astronomical objects, such as the neutral hydrogen clouds. The advantage in using this approach is that the problem of deriving the spatial distribution naturally falls into two separate parts: the first part is concerned with the distribution of halos in the universe, and the second part involves how objects (i.e., galaxies, hydrogen clouds, etc.) populate these halos. In this work, we use some of the elements of this HOD formalism to attempt to estimate the large-scale bias, i.e., in a quasi-linear regime, where the estimations of the 21 cm velocity distribution of the H\textsubscript{i} inside the halos are needed; these are out of the scope of this paper (see, for instance, Bagla & Khandai 2010; Obreschkow et al. 2009; Wyithe et al. 2009 for treatments that include small-scale clustering).

In the following subsection, we explore different models for the neutral hydrogen mass function and the relations between the neutral hydrogen \( M_{\text{H\textsubscript{i}}} \) and the total mass of a halo \( M_{\text{tot}} \).

2.2. Neutral Hydrogen Mass Function

The neutral hydrogen mass function of galaxies in the local universe has been measured by the HIPASS survey (Zwaan et al. 2005). However, the masses of the halos that host galaxies with a given H\textsubscript{i} mass are not generally known. Thus, we need to construct a plausible model for the relation between the total mass of a halo \( M_{\text{tot}} \) and its neutral hydrogen mass \( M_{\text{H\textsubscript{i}}} \).

While such a relation does not have to be a simple function, models that assume a one-to-one correspondence between the galaxy luminosity and the halo mass (Colin et al. 1999; Kravtsov & Klypin 1999; Kravtsov et al. 2004) provide remarkably good fits not only to the galaxy correlation functions (Conroy et al. 2006; Marín et al. 2008) but also to a wide variety of observational tests (Nagai & Kravtsov 2005; Gnedin 2008; Volonteri & Gnedin 2009; Guo et al. 2010; Conroy & Wechsler 2009; Tinker & Conroy 2009). Therefore, it is tempting to assume such a relation between the H\textsubscript{i} mass and total halo mass as well, that is, to assign one HIPASS source to one dark matter halo.

There is one complication with such an approach—not all galaxies contain H\textsubscript{i}. Thus, we need to account for the fraction of galaxies that do. Unfortunately, there are no reliable observational measurements of the fraction of H\textsubscript{i}-rich galaxies as a function of galaxy mass or luminosity. Instead, we adopt a simple ansatz that all blue galaxies contain H\textsubscript{i} and all red galaxies do not. Then, we can use the abundance matching technique if we match the abundances of HIPASS and blue galaxies.

Specifically, the relation between the total mass of a dark matter halo \( M_{\text{tot}} \) and its H\textsubscript{i} mass \( M_{\text{H\textsubscript{i}}} \) can be obtained by matching the two cumulative mass functions,

\[
 n_{\text{H\textsubscript{i}}} (>M_{\text{H\textsubscript{i}}}) = n_B (>M_{\text{tot}}),
\]
where \( n_{H^1} \) is the observed cumulative mass function (Zwaan et al. 2005), and

\[
n_B(>M_{tot}) = \int_{M_{tot}}^{\infty} dM f_B(M)n_B(M) \quad \text{(5)}
\]

In the last equation, \( f_B \) is the fraction of blue galaxies as a function of halo mass.

That fraction can be computed from the measured Sloan Digital Sky Survey (SDSS) red and blue galaxy luminosity functions (Montero-Dorta & Prada 2009) using the same abundance-matching technique. Namely, using the total SDSS luminosity function from Montero-Dorta & Prada (2009) we first match halo masses to \( r \)-band magnitudes as

\[
n_{gal}(<m_r) = \int_{M_{tot}}^{\infty} dM n_B(M) \quad \text{(6)}
\]

such a matching gives us a relation between a galaxy magnitude \( m_r \) and the halo mass \( M_{tot} \). Then, for each \( m_r \) (and, hence, \( M_{tot} \)) we compute the fraction of blue galaxies from the measured blue galaxy luminosity function (Montero-Dorta & Prada 2009). The so-measured blue galaxy fraction is shown in Figure 1 together with an estimate of the blue Brightest Cluster Galaxy (BCG) fraction from the maxBCG galaxy cluster catalog (E. Rykoff 2010, private communication). Estimates of the blue BCG fraction in X-ray selected clusters usually give substantially larger numbers (Bildfell et al. 2008); however, because X-ray emission is likely to correlate with the existence of a cooling flow (and, hence, a blue BCG), the latter estimates should be taken as upper limits (E. Rykoff 2010, private communication).

The resulting matching between the \( H^1 \) mass of a galaxy and its total mass from Equation (4) is shown in Figure 2. The relation between \( M_{H^1} \) and \( M_{tot} \) from the central panel allows us to compute the clustering properties of \( H^1 \) from the known clustering properties of the dark matter halos.

Unfortunately, little is known about the \( H^1 \) mass function beyond \( z \approx 0.1 \). To obtain the \( M_{H^1} \rightarrow M_{tot} \) matching at higher redshifts, we therefore must extrapolate the observed \( H^1 \) mass function to earlier times. This extrapolation is, of course, not unique and it requires two ingredients: the evolution of the total \( H^1 \) mass and how the \( H^1 \) mass is distributed among halos. To make the matter even worse, the observational constraints on the total \( H^1 \) mass density in the universe \( \rho_{H^1} \) (or, equivalently, the cosmological density parameter \( \Omega_{H^1} \)) are imprecise. Figure 3 shows a representative sample of these constraints for a range of redshifts (Zwaan et al. 2005; Prochaska et al. 2005; Rao et al. 2006). While at \( z \approx 0 \) and \( z \gtrsim 2 \) the measurements are reasonably robust, at intermediate redshifts the error bars are large and the behavior is non-monotonic. Therefore, to explore a wide enough range of possible evolutionary histories of \( \Omega_{H^1} \), we choose three representative models.

1. Model A provides a reasonable fit to all data points except for a single \( z \approx 2.3 \) point from Prochaska et al. (2005).
2. Model B follows the suggestion of Prochaska & Wolfe (2009), who argue that \( \Omega_{H^1} \) does not evolve between \( z = 0 \) and \( z \approx 2 \). Since it provides a plausible lower limit to the redshift evolution of \( \Omega_{H^1} \), we use this model even if it ignores the measurements of Rao et al. (2006).
3. In the same spirit, Model C serves as an upper limit to the possible \( \Omega_{H^1}(z) \). Formally, it is inconsistent with the HIPASS measurement, but we assume that \( \Omega_{H^1} \) in Model C increases rapidly from the HIPASS value of \( 3.8 \times 10^{-4} \) to about \( 10^{-3} \) within a redshift interval \( \Delta z \ll 1 \).

To fully specify the \( H^1 \) mass function, we then need to address how the total \( H^1 \) mass is distributed among halos as a function of redshift. If the shape of the \( H^1 \) mass function is approximately preserved at \( z > 0 \), then

\[
n_{H^1}(>M_{H^1}, z) = q_N(z)n_0(>q_M(z)M_{H^1}) \quad \text{(7)}
\]

where \( n_0(>M_{H^1}) \) is the observed cumulative HIPASS mass function (Zwaan et al. 2005) at \( z = 0 \). The two quantities \( q_N \) and \( q_M \) are not independent; for a given evolutionary history of neutral hydrogen contents of the universe \( \rho_{H^1}(z) \equiv \Omega_{H^1}(z)\rho_{\text{crit}} \).
and $q_M$ vertically for clarity—all models adopt $\Omega_{H_1} = 10^{-3}$ at $z > 3$. Black symbols with error bars show the observational measurements from Zwaan et al. (2005) (triangle), Rao et al. (2006) (circles), and Prochaska et al. (2005) (squares). (A color version of this figure is available in the online journal.)

$q_N$ and $q_M$ are related by the constraint

$$q_N(z) = \int_0^\infty \rho_{H_1}(M_{H_1}, z) dM_{H_1} = \int_0^\infty n_{H_1}(M_{H_1}, z) dM_{H_1} = \frac{q_N(z)}{q_M(z)}.$$  

(8)

Obviously, $q_N(0) = q_M(0) = 1$.

It is not practical to explore all possible evolutionary histories for $n_{H_1}(M_{H_1}, z)$; therefore, we restrict this work to only two scenarios.

**Pure number density evolution.** The pure number density evolution (hereafter, PNE) scenario assumes that the $H_1$ masses of individual halos do not change at high redshift ($q_M = 1$ for all $z$), but only the number density of halos of a given mass changes to reflect the change in the total neutral hydrogen content of the universe, $q_N = \Omega_{H_1}(z)/\Omega_{H_1}(0)$.

**Pure mass evolution.** The pure mass evolution scenario (hereafter, PME—a direct analog of a “pure luminosity evolution” in optical surveys) adopts an opposite extreme where the number density of halos remains the same ($q_N = 1$ for all $z$), but, instead, their masses change, $q_M = \frac{\Omega_{H_1}(0)}{\Omega_{H_1}(z)}$.

The justification for such a restriction is as follows. If the number density of galaxies with a given $H_1$ mass increases monotonically with increasing redshift, then $q_N(z)$ is a monotonically increasing function of $z$ (which is expected, since galaxies merge but do not break up). If, for a given halo, its $H_1$ also increases monotonically with $z$ (which is expected, if $H_1$ gets consumed by star formation and new accretion never exceeds the consumption), then $q_M(z)$ is also a monotonically increasing function of $z$. Since $q_N(0) = q_M(0) = 1$, our two scenarios bracket all possible evolutionary paths for $q_N$ and $q_M$.

Obviously, these assumptions do not hold all the time—for example, for some galaxies accretion of fresh gas can exceed the consumption by star formation at some times. However, over long timescales and over the whole galaxy population, it appears plausible that our two scenarios bracket the majority of evolutionary histories for $n_{H_1}(M_{H_1}, z)$.

One more ingredient of our model can evolve with redshift and that is the fraction of blue galaxies $f_B(M_{tot})$. The measured fraction of blue galaxies at $z = 0$ from Figure 1 is exceptionally well fitted (within the thickness of the line in Figure 1) by the following simple formula:

$$f_B(M_{tot}) = \left(1 + \frac{M_{tot}}{M_1}\right)^{-0.25} \exp\left(-\frac{M_{tot}}{M_2}\right)^{0.65}.$$  

(9)

with $M_1 = 2.5 \times 10^{11} M_\odot$ and $M_2 = 3.0 \times 10^{14} M_\odot$. It is tempting to associate $M_1$ with the characteristic mass of dwarf ellipticals and $M_2$ with the characteristic mass of a galaxy cluster. Then the behavior of $f_B$ with $M$ becomes well justified physically: for $M_{tot} \lesssim M_1$ a small fraction of all halos host dwarf and (at larger masses) regular elliptical galaxies, and $f_B$ becomes a gradually decreasing function of mass. But for $M_{tot} \lesssim M_2$ halos turn into galaxy groups and clusters, and the fraction of halos dominated by blue galaxies plummets.

With this plausible (although by no means proven) interpretation, we can come up with reasonable evolutionary scenarios for $f_B$ as a function of $z$. Specifically, we adopt the parametric form (Equation (9)) but allow $M_1$ and $M_2$ to evolve with redshift as

$$M_1(z) = 2.5 \times 10^{11}(1 + z)^\alpha M_\odot,$$

$$M_2(z) = 3.0 \times 10^{14}(1 + z)^\beta M_\odot.$$  

(10)

It seems plausible to assume that the fraction of blue galaxies at a given halo mass is higher at high redshift, i.e., $\alpha > 0$ and $\beta < 0$. To investigate the dependence of our results on these two parameters, we consider below three representative cases with $(\alpha, \beta) = (0, 0), (0, 1), \text{and} (1, 1)$. This set of values is, of course, not exclusive, but it is sufficient to evaluate the robustness of our results to these parameters.

In this work, we therefore consider a range of possibilities—our models A, B, and C for $\Omega_{H_1}(z)$ and two evolutionary scenarios (PNE and PME)—as sampling a plausible range of the redshift dependence of the $H_1$ mass function. It seems reasonable to speculate that the actual evolution should lie somewhere in between these different scenarios.

### 2.3. Large-scale Bias of $H_1$

On large scales, and not considering redshift distortions, the $H_1$ bias parameters provide the relations between the $H_1$ and dark matter correlation functions (Pan & Szapudi 2005)

$$\xi_{H_1} \approx b_{1,H_1}^i \xi_{dm}^i,$$  

(11)

$$\xi_{H_1} \approx b_{3,H_1}^3 \xi_{dm}^3 + b_{2,H_1}^2 b_{1,H_1}=
Bias coefficients of neutral hydrogen clouds as a function of redshift. Left: linear bias $b_1$ vs. redshift in the pure mass evolution scenario for all models (thick black solid line), and the pure number evolution scenario for model A (thin red solid line), model B (dashed magenta line), and model C (blue dotted line). Right: first nonlinear bias term $b_2$ vs. redshift for the same models described in the left plot. All coefficients are calculated with $M_{\text{min}} = 10^4 \, h^{-1} M_\odot$.

(A color version of this figure is available in the online journal.)

\begin{align}
  b^2(M) &= \frac{8}{21} (\epsilon_1 + E_1) + \epsilon_2 + E_2, \\
  \epsilon_1 &= \frac{a v^2 - 1}{\delta_c}, \\
  \epsilon_2 &= \frac{a v^2(a v^2 - 3)}{\delta_c^2}, \\
  E_1 &= \frac{2p}{\delta_c(1 + (a v^2)^\rho)}, \\
  E_2 &= \frac{1 + 2p}{\delta_c} + 2\epsilon_1,
\end{align}

where $i = 1, 2$ represent the linear and first nonlinear terms, where their redshift dependence is encompassed in $v = \delta_c/\sigma(M, z)$, which depends on the redshift evolution of the linear power spectrum.

Using the $M_{\text{HI}} - M_{\text{tot}}$ relation obtained in the previous section, the $\text{HI}$ bias from all halos with a mass greater than $M_{\text{min}}$ is given by the following integral:

\begin{equation}
  b_{\text{HI}} = \frac{1}{\rho_{\text{HI}}} \int_{M_{\text{min}}}^\infty dM n_b(M) b^1(M) \langle M_{\text{HI}}(M) \rangle,
\end{equation}

where $\rho_{\text{HI}}$ is the density of neutral hydrogen:

\begin{equation}
  \rho_{\text{HI}} = \int_{M_{\text{min}}}^\infty dM n_b(M) \langle M_{\text{HI}}(M) \rangle.
\end{equation}

In what follows, we present the analytic results for the large-scale bias and its evolution in redshift.

3. RESULTS

3.1. Analytic Results

We calculate the bias parameters as a function of redshift for all models described in Section 2.2. In Figure 4, we show the results for the models where the blue galaxy fraction does not evolve with redshift, i.e., in Equation (10), $\alpha = \beta = 0$. In the left panel, we present the linear ($b_1$) bias redshift evolution; in the right panel, the nonlinear ($b_2$) bias evolution, with $M_{\text{min}} = 10^4 \, h^{-1} M_\odot$, which, given the fact that it is not expected that low mass halos (below $10^7 \, h^{-1} M_\odot$) contain significant amounts of H$\text{I}$, is set to the value mentioned above as a conservative limit. The black thick line shows the evolution of bias in the PME scenario; as for the PNE scenario, the red thin solid line shows the results for model A, the magenta dashed line shows the results for model B, and the blue dotted line shows the results for model C.

The first feature that arises from these analytical results is the strong dependence of the bias terms on the redshift for all the models considered. Given that $\Omega_{\text{HI}}$ does not evolve significantly with redshift, the reason for this behavior lies in the fact that preserving the matching relations (Equation (4)) at higher redshift has the consequence that the most massive halos, which are less numerous at high redshift (i.e., are more biased with respect to the dark matter distribution), carry more H$\text{I}$ mass than the halos with the same $M_{\text{tot}}$ at $z = 0$. Therefore, it is expected for both $b_1$ and $b_2$ to show a strong redshift dependence.

For the PNE scenario, we can see differences in both bias parameters, though they are relatively small. Between models A and B, the differences are increasing with redshift, with a peak around $z \sim 2$—at the moment where $\Omega_{\text{HI}}(z)$ of model B is the lowest compared to the rest of the models, since the fraction of neutral hydrogen in that model is low, in the PNE scenario this implies fewer halos, and therefore the bias is expected to be larger. In model C, the difference in the bias evolution compared to model A appears only at low redshifts, since it starts with a higher $\Omega_{\text{HI}}$, and therefore has a smaller bias on redshifts close to $z = 0$. The differences for the nonlinear bias term $b_2$ are qualitatively similar but much milder.

As opposed to the PNE scenario, there are no differences between the model biases in the PME scenario; this is not surprising since, by definition, in that scenario only the masses of halos change but not their total number density. Hence, irrespective of the $\Omega_{\text{HI}}(z)$ evolution, the behavior of the bias here depends only on the cosmological evolution. When we compare both scenarios, the differences in the bias parameter $b_1$, albeit small in general, grow steadily with redshift reaching 10% level between $z = 1$ and 4; note here that there are differences between the PME scenario and model B in the PNE scenario only after $z \sim 2$. For $b_2$ between $z = 0$ and 1 there are almost
no differences between the models, but after that the differences grow until reaching almost a factor of 2 by $z = 4$. In both cases the PNE scenario has a larger bias; this difference is due to the fact that more halos are needed in the PNE scenario to account for the change of neutral hydrogen in the universe, which leads to a lower halo mass for a given H\textsc{i} mass and hence a smaller bias than in the PME scenario.

As for the differences between the different blue fraction evolution models and the case where we simply assume all galaxies contain H\textsc{i}, we show in Figure 5 the quantity

$$\Delta b_i = b_i - b_{i,\text{FID}},$$

where $i = 1, 2,$ and the fiducial (FID) model corresponds to model A in the PNE scenario, with $(\alpha, \beta) = (0, 0)$. The black dotted line represents the $f_b(M, z) = 1$, i.e., all galaxies carry H\textsc{i}; the blue short-dashed line represents the evolution model with $(\alpha, \beta) = (0, 1)$, and the long-dashed line shows the results for the blue fraction evolution model with $(\alpha, \beta) = (1, 1)$, both in model A of the PNE scenario. For other models and scenarios the results are qualitatively similar: between the different blue fraction evolution models both the linear and nonlinear bias do not differ significantly. But when we compare our FID model to that where all galaxies carry neutral hydrogen, i.e., $f_b = 1$, then the differences are more significant, considering a realistic blue galaxy fraction lowers the H\textsc{i} bias, since the neutral hydrogen avoids accumulating in the bigger halos and concentrates more in mid-size and small mass halos.

The weak dependence of the H\textsc{i} bias on the model adopted for the neutral hydrogen evolution means that the intensity mapping observations will be able to provide only a limited amount of information on the global evolution of the neutral hydrogen abundance in the universe. On the other hand, from a cosmologist’s point of view, this is good news: a detailed knowledge of the H\textsc{i} evolution may not be needed for studying the dark energy evolution in the radio intensity mapping experiments. Note that even though the differences in $b_2$ between the PME and PNE models at high redshift are larger than those in $b_1$, measurements on higher-order correlations have larger uncertainties which would make the $b_2$ estimations less reliable. Nevertheless, if good signal-to-noise measurements can be done, this parameter can play a relevant role in determining the right evolution model.

### 3.2. Linear Bias Calculated from Simulations

In order to have a consistency check of our analytical calculations, we turn to estimate the linear H\textsc{i} bias from the Millennium Simulation (Springel et al. 2005). The N-body simulation was carried out for 2160$^3$ particles with mass $m_p = 8.6 \times 10^8$ $h^{-1}$ $M_\odot$ in a box with $L_{\text{box}} = 500$ $h^{-1}$ Mpc on the side, with the same cosmological parameters we use for our analytical estimates.

Following Equation (11), we can estimate the linear bias in the simulation by

$$b_{H\textsc{i},\text{FID}}(k) = \sqrt{\frac{P_{H\textsc{i}}(k)}{P_{\text{dm}}(k)}},$$

where $P_{\text{dm}}(k)$ is the real-space dark matter power spectrum and $P_{H\textsc{i}}(k)$ is the real-space H\textsc{i} power spectrum. In addition to validating our analytical method, this measurement should allow us to estimate the scales below which nonlinear effects become relevant.

We use the $M_{H\textsc{i}}-\text{M_{tot}}$ relation obtained in Section 2.2 to assign the neutral hydrogen mass to dark matter halos and calculate the H\textsc{i} power spectrum $P_{H\textsc{i}}$. Note that due to the mass resolution of the simulation, our dark matter halo catalog does not well resolve the halos below a certain mass threshold, and those halos may contribute substantially to the value of the H\textsc{i} bias. Therefore, consistency requires that the bias measured in the Millennium Simulation be compared to the value obtained analytically through Equation (19) assuming the same mass threshold $M_{\text{min}}$. For this reason, we compute analytical estimates and measured values of the bias obtained by assuming $M_{\text{min}} \sim 10^{11} h^{-1} M_\odot$, which should correspond to halos with $N_p \sim 100$ particles. Another aspect when we are “painting” H\textsc{i} in the simulation is its distribution inside the halos. In our case, we place all hydrogen at the center of the halo, i.e., we are not assuming diffuse H\textsc{i} emission inside. While this might not model accurately the small-scale clustering of the H\textsc{i}, what matters in the large-scale clustering are the details of the $M_{H\textsc{i}}-\text{M_{tot}}$ relation; the spatial profile of the neutral hydrogen within a halo will have a very limited effect on the measured bias and its scale dependence on large scales we present in this paper (Cooray & Sheth 2002; Schulz & White 2006).
should agree with differences to the one measured in simulations. The figure also shows analytical predictions for the whole set of halos ($M_{\text{min}} = 10^{11} h^{-1} M_\odot$), while solid lines show the analytical calculations for the set of halos resolved in the Millennium Simulation ($M_{\text{min}} = 10^{15} h^{-1} M_\odot$).

(A color version of this figure is available in the online journal.)

Figure 6 shows, in symbols, the measured bias $b_{\text{HI}}^{\text{MS}}(k)$ for a series of redshifts from $z = 0$ to $z = 3$ for model A in the PNE scenario. The solid lines represent $b_{\text{HI}}$ from our analytical modeling, with a minimum mass cut $M_{\text{min}} = 10^{11} h^{-1} M_\odot$; the dashed lines are the results of $b_{\text{HI}1}$ with $M_{\text{min}} = 10^4 h^{-1} M_\odot$.

The bias measured from the simulation using Equation (22) should agree with $b_{\text{HI}1}$ obtained from Equation (19) on large scales; on smaller scales, a contribution from the clustering within a halo results in a scale dependence in bias (Seljak 2000) which we refer to as a nonlinear bias effect. From the figure, we indeed observe a strong redshift evolution of large-scale bias in agreement with the behavior found in our analytical calculations; the agreement is of the order of 5%. The small differences are due mainly to the fact that our analytical halo mass function, though a good approximation, indeed has small differences to the one measured in simulations. The figure also shows that the linear bias assumptions break down for $k > 0.15 h \, \text{Mpc}^{-1}$, but the effective nonlinear scale (where the linear bias deviates strongly from a constant value) appears close to $10^{11} h^{-1} M_\odot$.

As stated earlier, the difference between the two sets of horizontal lines in Figure 6 shows the contribution of halos with $M_{\text{total}} < M_{\text{min}} = 10^{11} h^{-1} M_\odot$. This contribution is substantial. It is important to note that in the case of the 21 cm intensity mapping, we suffer no lower detection limit in neutral hydrogen mass and therefore halo mass. In that case, the bias derived with $M_{\text{min}} = 0$ is the appropriate one. As a much higher mass resolution would be required to reduce $M_{\text{min}}$ much below $10^{11} h^{-1} M_\odot$, the Millennium Simulation—or another simulation with similar resolution—cannot be used to provide an accurate estimate for the H I bias for this kind of survey but only as a useful tool to validate the semi-analytical estimates obtained in the previous section.

3.3. Impact of Bias Evolution on the Power Spectrum

Having shown how the H I bias evolves with redshift, we turn to assess to what extent this may or may not have an effect on the measurement of the power spectrum. At tree level in perturbation theory, the density fluctuations in the neutral hydrogen $\delta_{\text{HI}}$ and in the dark matter $\delta_{\text{dm}} \equiv \delta$ are related through

$$\delta_{\text{HI}}(\vec{x}, \chi) = b(\chi) \delta(\vec{x}, \chi) = b(\chi) D_s(\chi) \delta(\vec{x}, 0) \equiv E(\chi) \delta(\vec{x}, 0),$$

where $D_s(\chi)$ is the linear growth factor, and we exchanged redshift $z$ for comoving distance $\chi$. When observations are performed, only the product $E(\chi) \delta$ can be measured, as the redshift evolution of the bias contributes a component that combines with the growth of structure. To understand to what extent this has an influence on the measurement of the power spectrum, we start by Taylor expanding $E(\chi) = \sum E_0^n(\chi) n^n$, where all derivatives $E_0^n = d^n E/d\chi^n$ are evaluated at the observer’s position. Next, we Fourier transform $\delta_{\text{HI}}(\vec{x})$ to obtain

$$\delta_{\text{HI}}(\vec{k}, \chi) = \int d^3 \vec{x} e^{-i \vec{k} \cdot \vec{x}} \delta_{\text{HI}}(\vec{x}, \chi) \equiv \int d^3 \vec{x} e^{-i \vec{k} \cdot \vec{x}} \left[ \sum_{n=0}^{\infty} \frac{E_0^n(\chi)}{n!} \delta(\vec{x}, 0) \right],$$

where we used the fact that $\chi^n = n! \delta_{k_0} e^{-i \vec{k} \cdot \vec{x}}$. Next, we note that $\delta(\vec{k}, 0)$ is a Gaussian random variable with dispersion $\delta_{\text{HI}}(\vec{k}, 0)$. We can then rewrite it as

$$\delta(\vec{k}, 0) = \sqrt{P(\vec{k}, 0) / \delta_{\text{HI}}(\vec{k}, 0)},$$

where $\delta_{\text{HI}}(\vec{k}, 0)$ is a Gaussian random variable with $\langle \delta_{\text{HI}}(\vec{k}, 0) \rangle = 0$ and $\langle \delta_{\text{HI}}^2(\vec{k}, 0) \rangle = (2\pi)^3 \delta_{\text{HI}}^2(\vec{k} - \vec{k}')$. Note that $\vec{k}$ for $\lambda_{\text{HI}}$ is nothing but a label and that all the dependence of $\delta(\vec{k}, 0)$ on the wavevector is really encoded in the prefactor. Defining the HI power spectrum by

$$P_{\text{HI}}(k, \chi) = P_{\text{HI}}(k, \chi) \equiv P_{\text{HI}}(k, \chi) (2\pi)^3 \delta_{\text{HI}}^2(\vec{k} - \vec{k}')$$

we can then express $P_{\text{HI}}$ as

$$P_{\text{HI}}(k, \chi) \approx [L(k) + \mu^2 H(k)] P(k, 0),$$

where $\mu = \cos(\theta)$ is the cosine of the angle the wavevector $\vec{k}$ makes with the line of sight and $L(k)$ and $H(k)$ are the first two terms that encode the $\chi$ dependence of $\delta_{\text{HI}}$. We calculate the first few terms of $L$ and $H$ using the fact that $k_0 = k \cos(\theta) = \mu k$ and $k_{\perp} = k \sin(\theta) = k \sqrt{1 - \mu^2}$. They turn out to be

$$L \approx E_0^2 - E_0^0 \frac{P'}{2k P},$$

$$H \approx E_0^0 \left[ \frac{P^2}{4P'^2} - \frac{P''}{2P} + \frac{P'}{2k P} \right] + E_0^2 \frac{P'^2}{4P^2},$$

where $P'$ and $P''$ denote first and second derivatives of $P$ with respect to $k$, while $E_0^0$ and $E_0^2$ denote derivatives of $E(\chi)$ with respect to the comoving distance evaluated at the observer position. In general, the comoving distance dependence generates in the power spectrum a directional dependence that is reminiscent of redshift space distortions. Despite the fact that the nature of the effect considered here is completely different from the effect of redshift space distortions (which is due to peculiar velocities), the two effects may be in practice difficult to separate from one another, unless we have a sufficiently small sample variance on the clustering measurement. However, we point out that the $n$th derivative with respect to the comoving...
distance scales as $H_0^n$ and therefore $E'_{0} \sim 10^{-4}$ and $E''_{0} \sim 10^{-8}$. As such, all but the first term appearing in Equations (26) and (27) are of order $10^{-8}$ and only the $E^3_0$ term appearing in $L$ will give a non-negligible contribution. Also, it is possible to note that the above treatment is valid even in the case of a perfect unbiased dark matter tracer.

In Figure 7, we show the functions $L(k)$ and $|H(k)|$. Consistent with the above estimates, $H$ turns out to be extremely small. On the other hand, $L$ is dominated by the term $E^3_0$ and therefore does not show any appreciable $k$ dependence, except for extremely large scales $k \sim 10^{-3}$, where it deviates from $E^3_0$ by about 0.5%. These two facts together allow us to conclude that the measurement of the 21 cm power spectrum through intensity mapping should not be sensibly affected by the redshift dependence of the bias.

### 4. SUMMARY AND CONCLUSIONS

In this paper, we have estimated analytically and in the Millennium Simulation the large-scale bias of neutral hydrogen from a non-parametric model which establish a one-to-one relation between the total mass and the H I mass of galactic halos. Although only an approximation, this relation is expected to hold for a large range of masses.

The bias parameters have been calculated using a model which uses elements of the HOD formalism, in a way that can be considered complementary to the galaxy-HOD model of H I galaxies carried out by Wyithe et al. (2009); we do not deal in our case with the individual galaxies inside dark matter halos but with the total neutral hydrogen inside them. In terms of large-scale clustering, these approaches are completely equivalent.

Using this approach together with different models for the redshift evolution of the H I content in the universe and the relation between neutral hydrogen and total matter contents of galactic halos, we analyze the evolution of bias with redshift.

We find that both linear and nonlinear bias increase significantly with redshift. This contrasts with the finding by Wyithe et al. (2008), based on an extrapolation to lower redshifts from a reionization model of Wyithe & Loeb (2008). Since Wyithe et al. (2008) do not account for clustering of galactic halos that host all neutral hydrogen after reionization, they underestimate the H I bias at $z \lesssim 4$. On the other hand, our results are in agreement with the high value of the linear large-scale bias at redshift $z \sim 3$ measured in the model by Bagla & Khandai (2010) where they populate H I in high-resolution $N$-body simulations.

We also find that, given the limitations of our model at high masses, the estimation of the bias does not change significantly when the very high mass halos are neglected.

Our main conclusion is that, despite a wide range of possible models for the evolution of H I mass function that we consider, in all models the bias evolution is similar—the neutral hydrogen distribution is mildly antibiased at $z = 0$ but becomes strongly biased ($b_{HI} \sim 2$) by $z \sim 4$. This result is encouraging for the planned radio intensity mapping experiments, since large bias implies a stronger 21 cm signal. Nevertheless, this strong redshift evolution does not significantly compromise the measurements of the neutral hydrogen power spectrum along the radial direction.

We thank Gerard Lemson for his support and patience in providing the Millennium Simulation halo catalogs. We also thank Andrey Kravtsov for his comments and suggestions.

This work was supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through grants NSF PHY-0114422 and NSF PHY-0551142 and an endowment from the Kavli Foundation and its founder Fred Kavli. A.V. and H.S. are supported by the DOE at Fermilab. The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory.

### REFERENCES

- Ansari, R., Le Goff, J., Magneville, C., Moniez, M., Palanque-Delabrouille, N., Rich, J., Ruhlmann-Kleider, V., & Yèche, C. 2008, arXiv:0807.3614
- Bagla, J. S., & Khandai, N. 2010, MNRAS, in press (arXiv:0908.3796)
- Barkana, R., & Loeb, A. 2007, Prog. Phys., 70, 627
- Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
- Bildfell, C., Hoekstra, H., Babul, A., & Mahdavi, A. 2008, MNRAS, 389, 1637
- Blake, C., & Glazebrook, K. 2003, ApJ, 594, 665
- Chang, T., Pen, U., Peterson, J. B., & McDonald, P. 2008a, Phys. Rev. Lett., 100, 091303
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008b, Phys. Rev. Lett., 100, 091303
- Colín, P., Klypin, A. A., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 523, 32
- Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
- Eisenstein, D. 2002, in ASP Conf. Ser. 280, Next Generation Wide-Field Multi-Object Spectroscopy, ed. M. J. L. Brown & A. Dey (San Francisco, CA: ASP), 35
- Eisenstein, D. J., Hu, W., & Tegmark, M. 1998, ApJ, 504, L57
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Phys. Rep., 433, 181
- Gnedin, N. Y. 2008, ApJ, 673, L1
- Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, MNRAS, 404, 1111
- Hu, W., & Haiman, Z. 2003, Phys. Rev. D, 68, 063004
- Hu, W., & White, M. 1996, ApJ, 471, 30
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B., & Primack, J. R. 2004, ApJ, 609, 35
- Kravtsov, A. V., & Klypin, A. A. 1999, ApJ, 520, 437
- Linder, E. V. 2003, Phys. Rev. D, 68, 083504
- Loeb, A., & Wyithe, S. 2008, Phys. Rev. Lett., 100, 161301
- Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, Phys. Rev. D, 78, 023529
Marín, F. A., Wechsler, R. H., Frieman, J. A., & Nichol, R. C. 2008, ApJ, 672, 849
Montero-Dorta, A. D., & Prada, F. 2009, MNRAS, 399, 1106
Nagai, D., & Kravtsov, A. V. 2005, ApJ, 618, 557
Obreschkow, D., Croton, D., De Lucia, G., Khochfar, S., & Rawlings, S. 2009, ApJ, 698, 1467
Padmanabhan, N., & White, M. 2009, Phys. Rev. D, 80, 063508
Pan, J., & Szapudi, I. 2005, MNRAS, 362, 1363
Pritchard, J. R., & Loeb, A. 2008, Phys. Rev. D, 78, 103511
Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123
Prochaska, J. X., & Wolfe, A. M. 2009, ApJ, 696, 1543
Rao, S. M., Turnshek, D. A., & Nestor, D. B. 2006, ApJ, 636, 610
Schulz, A. E., & White, M. 2006, Astropart. Phys., 25, 172
Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20
Seljak, U. 2000, MNRAS, 318, 203
Seo, H., Dodelson, S., Marriner, J., McGinnis, D., Stebbins, A., Stoughton, C., & Vallinotto, A. 2010, ApJ, submitted (arXiv:0910.5007)
Seo, H., & Eisenstein, D. J. 2003, ApJ, 598, 720
Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1
Sheth, R. K., & Tormen, G. 2002, MNRAS, 329, 61
Springel, V., et al. 2005, Nature, 435, 629
Tinker, J. L., & Conroy, C. 2009, ApJ, 691, 633
Vishal, E., Loeb, A., & Wyithe, S. 2009, J. Cosmol. Astropart. Phys., JCAP10(2009)030
Volonteri, M., & Gnedin, N. 2009, ApJ, 703, 2113
Wyithe, J. S. B., Loeb, A., & Geil, P. M. 2008, MNRAS, 383, 1195
Wyithe, S., Brown, M. J. I., Zwaan, M. A., & Meyer, M. J. 2009, arXiv:0908.2854
Wyithe, S., & Loeb, A. 2008, MNRAS, 383, 606
Zwaan, M. A., Meyer, M. J., Staveley-Smith, L., & Webster, R. L. 2005, MNRAS, 359, L30