

RECEIVED: October 3, 2024  
REVISED: January 24, 2025  
ACCEPTED: February 24, 2025  
PUBLISHED: April 1, 2025

## Fast simulation of cosmological neutral hydrogen based on the halo model

Pascal Hitz ,<sup>a,\*</sup> Pascale Berner ,<sup>a</sup> Devin Crichton ,<sup>a</sup> John Hennig ,<sup>a,b</sup> and Alexandre Refregier ,<sup>a</sup>

<sup>a</sup> Institute for Particle Physics and Astrophysics, ETH Zurich,  
8092 Zurich, Switzerland

<sup>b</sup> Scientific IT Services, ETH Zurich,  
8092 Zurich, Switzerland

E-mail: [hitzpa@phys.ethz.ch](mailto:hitzpa@phys.ethz.ch), [bernerp@phys.ethz.ch](mailto:bernerp@phys.ethz.ch), [dcrichton@phys.ethz.ch](mailto:dcrichton@phys.ethz.ch),  
[john.hennig@id.ethz.ch](mailto:john.hennig@id.ethz.ch), [alexandre.refregier@phys.ethz.ch](mailto:alexandre.refregier@phys.ethz.ch)

**ABSTRACT:** Cosmological neutral hydrogen (HI) surveys provide a promising tomographic probe of the post-reionization era and of the standard model of cosmology. Simulations of this signal are crucial for maximizing the utility of these surveys. We present a fast method for simulating the cosmological distribution of HI based on a halo model approach. Employing the approximate PINOCCHIO code, we generate the past light cone of dark matter halos. Subsequently, the halos are populated with HI according to a HI-halo mass relation. The nature of 21 cm intensity mapping demands large-volume simulations with a high halo mass resolution. To fulfill both requirements, we simulate a past light cone for declinations between  $-15^\circ$  and  $-35^\circ$  in the frequency range from 700 to 800 MHz, matching HIRAX, the Hydrogen Intensity and Real-time Analysis eXperiment. We run PINOCCHIO for a  $1 \text{ h}^{-3} \text{Gpc}^3$  box with  $6700^3$  simulation particles. With this configuration, halos with masses as low as  $M_{\min} = 4.3 \times 10^9 M_\odot$  are simulated, resulting in the recovery of more than 97% of the expected HI density. From the dark matter and HI past light cone, maps with a width of 5 MHz are created. To validate the simulation, we have implemented and present here an analytical dark matter and HI halo model in PyCosmo, a Python package tailored for theoretical cosmological predictions. We perform extensive comparisons between analytical predictions and the simulation for the mass function, mass density, power spectrum, and angular power spectrum for dark matter and HI. We find close agreement in the mass function and mass densities, with discrepancies within a few percent. For the three-dimensional power spectra and angular power spectra, we observe an agreement better than 10%. This approach is broadly applicable for forecasting and forward-modeling of dedicated intensity mapping

\*Corresponding author.



© 2025 The Author(s). Published by IOP Publishing Ltd on behalf of SISSA Medialab. Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

experiments, such as HIRAX, as well as cosmological surveys with the SKA. Additionally, it is particularly well suited to be extended for cross-correlation studies. The `PyCosmo` halo model extension and simulation datasets are made publicly available.

KEYWORDS: cosmological simulations, dark matter simulations, power spectrum, dark matter theory

ARXIV EPRINT: [2410.01694](https://arxiv.org/abs/2410.01694)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Analytical methods</b>	<b>3</b>
2.1	Dark matter halo model	4
2.2	HI halo model	8
<b>3</b>	<b>Simulation methods</b>	<b>11</b>
3.1	Simulation of dark matter and HI halos	11
3.2	Map generation	13
<b>4</b>	<b>Results</b>	<b>14</b>
4.1	Mass function	15
4.2	Density	17
4.3	Power spectrum	18
4.4	Angular power spectrum	21
<b>5</b>	<b>Conclusions</b>	<b>23</b>
<b>A</b>	<b>Poisson noise contribution to the angular power spectrum</b>	<b>25</b>

---

**1** **Introduction**

The measurement of the recent accelerated expansion of the universe confronts us with a fundamental problem for our understanding of its nature. Current cosmological models invoke an unknown dark energy, which is now the dominant energy component in the universe. To shed light on dark energy, its characteristic properties need to be empirically determined, in particular the equation of state. In order to achieve this, the expansion history of the universe must be constrained in tomographic analyses, starting from the epoch when dark energy began to influence the cosmic acceleration, around a redshift of  $z \sim 2$ . Baryonic Acoustic Oscillations (BAO) provide a unique standard ruler which can be used to measure the expansion history. Measuring the observed scale of the BAO in the matter power spectrum in the radial and transverse directions allows to measure the Hubble parameter and the comoving angular-diameter distance. These in turn provide information about the expansion history [1, 2].

The BAO have been well-studied in galaxy surveys before [3–7], but there exists another promising technique with several advantages: 21 cm intensity mapping [8–11]. After the end of reionization, neutral hydrogen (HI) is expected to be a good tracer of the underlying matter distribution [12]. Even in its ground state HI can be observed through the emission of radiation due to the hyperfine transition between the state in which the spins of the electron and proton are parallel aligned to the state in which they are anti-aligned. This transition produces a photon of about 21 cm wavelength corresponding to a rest frequency

of  $f_{21} = 1420.406$  MHz. The fact that this is a forbidden transition is compensated by the vast amount of HI and is an advantage because the probability that the photon is absorbed again after emission is very small. Since the expansion of the universe causes the emitted photons to be redshifted, the spectral line of this transition is observed at a lower frequency  $f_{\text{obs}}$ . The relation between the observed frequency and redshift is given by:

$$z = \frac{f_{21}}{f_{\text{obs}}} - 1. \quad (1.1)$$

Consequently, individual galaxies do not have to be resolved to infer their redshift, as is the case with galaxy surveys. Instead, mapping the emission of unresolved HI as a function of frequency immediately provides three-dimensional information about its distribution. While post-reionization 21 cm intensity mapping can target the same science as galaxy surveys, it has the advantage of observing much larger volumes, i.e. wide fields in a broad redshift range, with a lower angular resolution sufficient for cosmological scales, in a shorter time. It therefore has the potential to constrain cosmological parameters to very high precision.

There are several instruments aiming to perform 21 cm intensity mapping both in single-dish mode (e.g. SKAO-MID [13, 14], MeerKAT [15], BINGO [16], FAST [17], GBT [18], and Parkes [19]) and in interferometric mode (e.g. CHIME [20], CHORD [21], HERA [22], PUMA [23], Tianlai [24], and HIRAX [25, 26]). HIRAX, which stands for the Hydrogen Intensity and Real-time Analysis eXperiment, is a radio interferometer currently in development. It will be located in the Karoo desert in South Africa covering the southern sky from the celestial equator to a declination of  $-60^\circ$ . Eventually, HIRAX will consist of 1024 six meter dishes arranged in a redundant configuration and operating in drift-scan mode, i.e. scanning the sky for constant but adjustable declinations through the natural rotation of the Earth. HIRAX will be sensitive to the frequency range of 400 to 800 MHz and therefore probe in the redshift range from 0.78 to 2.55. Although HIRAX is still under construction and therefore observational data is not yet available, simulations of the 21 cm signal in the form of tomographic maps can be used to test the instrument and observation modeling as well as the analysis pipeline. This work focuses on simulating post-reionization HI signal maps and verifying their quality.

On one end, statistical simulations, such as lognormal simulations [27–30], provide a computationally efficient option to generate many realisations of a known statistical distribution. However, their accuracy is limited by our understanding of the input distribution, which is often what we seek to improve through simulations. On the other end, large  $N$ -body or hydrodynamical simulations can model complex non-linear processes that are analytically intractable, but their high computational costs prevent us from running a large number of realisations. Lagrangian Perturbation Theory (LPT) based methods aim to bridge the gap between the two approaches, providing accurate realisations of the matter distribution up to mildly non-linear scales while accepting some inaccuracies on very small scales. One approach is to use LPT to compute the evolution of the dark matter density field and subsequently applying a global bias model to link HI to this on the field level [30]. However, in the post-reionization era, most HI is expected to reside within dense regions, i.e. in dark matter halos, where it can self-shield from ionizing radiation. Although intensity mapping instruments typically have low angular resolution, they have a high frequency resolution in

the radial direction. Consequently, shot noise, arising from the discrete nature of the HI distribution, becomes increasingly significant. To capture this more realistic description of HI and to account for the discreteness of the sources, we employ the `PINOCCHIO` code [31–36]. This code uses LPT to efficiently simulate the past light cone of dark matter halos in an approximate but quicker way than full  $N$ -body simulations. The HI is then subsequently assigned to these halos using a HI-halo mass relation that has been fitted to observations. Due to its approximate nature, the small-scale clustering and halo positions of `PINOCCHIO` are probably not accurate enough for future simulation-based inference on real observational data. However, to generate forecasts for our intended applications, this level of accuracy is sufficient.

This method has been previously applied in cosmological HI signal simulations for the SKAO HI intensity blind foreground subtraction challenge described in [37]. There, a full-sky past light cone was generated and a HI-halo mass relation was used, which was calibrated using a semi-analytical model [38]. In this work, we limit the construction of the light cone to the region of the sky covered by the HIRAX telescope, allowing us to simulate a larger number of particles with `PINOCCHIO`. This is essential for accurately capturing most of the cosmological HI.

In future work, these maps will serve to test the HIRAX pipeline used to reconstruct the HI emission signal from our simulated signal. Furthermore, the fact that the simulations are based on consistent halo catalogues opens up the possibility for upcoming cross-correlation studies. This paper is dedicated to the development and running of a HI simulation pipeline and the validation of the simulated signal by comparison with analytical predictions. For the latter, we have implemented an extension including a theoretical dark matter and HI halo model in `PyCosmo`<sup>1</sup> [39–41], a Python-based framework for theoretical cosmological predictions. In particular, predictions of the mass function, mass densities and non-linear power spectra are compared to the simulations. As part of this work, we present the halo model and the specific implementations that are made publicly available in `PyCosmo` 2.2.0.

This paper is organised as follows: In section 2, all details of the implemented dark matter and HI halo model are described. Section 3 proceeds with the procedure of simulating the past light cone of dark matter and HI halos and the map generation step. Section 4 presents the resulting dark matter and HI temperature maps and compares the simulated mass function, mass densities, power spectra and angular power spectra to predictions within the framework of the dark matter and HI halo model. Finally, section 5 serves as a summary of this paper and provides an outlook to future work.

Throughout this paper, a flat  $\Lambda$ CDM cosmology with the cosmological parameters from the Planck 2018 results [42] is assumed:  $\Omega_m = 0.3153$ ,  $\Omega_\Lambda = 0.6847$ ,  $\Omega_b = 0.0493$ ,  $h = 0.6736$ ,  $\sigma_8 = 0.8111$ ,  $n_s = 0.9649$ , and  $Y_p = 0.24$ .

## 2 Analytical methods

Our simulations are based on the distribution of dark matter halos. Therefore, we make use of the halo model formalism [43–45] to make theoretical predictions for comparison with the simulated data. The fundamental assumption of this formalism is that all matter in

---

<sup>1</sup>PyCosmo webpage: <https://cosmology.ethz.ch/research/software-lab/PyCosmo.html>.

the universe is arranged in halos of different sizes and masses. Given that HI is expected to trace these halos, we can apply a halo model analogous to that used for dark matter to HI as well. At the heart of every halo model are three ingredients: the halo mass function, the halo bias, and the halo profile. Once these are defined, the formalism can be used to calculate the mean mass densities, the non-linear three-dimensional power spectrum and, consequently, the angular power spectrum.

Given the various implementations and conventions, and for the convenience of the reader, we give a detailed review of the halo model in the following sections as we have implemented it in `PyCosmo`.

## 2.1 Dark matter halo model

### 2.1.1 Ingredients

The mass function, which gives the number of dark matter halos of mass  $M$  per unit volume and unit mass at a given redshift  $z$ , is given as:

$$\frac{dn(M, z)}{dM} = \frac{\bar{\rho}_{m,0}}{M^2} \nu f(\nu) \left| \frac{d \ln \nu}{d \ln M} \right|. \quad (2.1)$$

The peak height is defined as  $\nu(M, z) = \delta_c(\Omega_m(z))/\sigma(M, z)$ , with  $\delta_c$  the critical overdensity needed for spherical collapse of a halo and  $\sigma^2(M, z)$  being the mass variance of the density field smoothed on the scale  $R = (3M/(4\pi\bar{\rho}_{m,0}))^{1/3}$ . In the spherical collapse model regions with mean overdensity larger than the critical overdensity are assumed to have collapsed into a virialized object. This criterion can either be evaluated at the present time or at the time of collapse. In the first case,  $\delta_c$  must be linearly extrapolated to today's value by dividing by the linear growth factor. In the later case, the density field must be considered at this redshift and the mass variance is linearly extrapolated to the specific redshift. We decide to use the second convention and make the mass variance redshift dependent by multiplying it by the linear growth factor. For the multiplicity function  $f(\nu)$ , we provide four different options in `PyCosmo`: Press & Schechter [46], Sheth & Tormen [47], Tinker et al. [48], and Watson et al. [49]. For the mass function of Tinker et al., the critical overdensity is set to  $\delta_c = \frac{3}{20}(12\pi)^{2/3} \approx 1.686$ , the value found in spherical collapse in a flat matter dominated universe  $\Omega_m = 1$ , and we have implemented the mass function with the parameters found for  $\Delta_v = 200$  in [48].<sup>2</sup> For the other mass functions, we allow cosmology dependent expressions [50, 51]. For a flat  $\Lambda$ CDM model,  $\Omega_\Lambda \neq 0$  and  $\Omega_\Lambda + \Omega_m = 1$ :

$$\delta_c(\Omega_m(z)) = \frac{3}{20}(12\pi)^{2/3}\Omega_m(z)^{0.0055}, \quad (2.2)$$

and for  $\Omega_\Lambda = 0$  and  $\Omega_m \neq 1$ :

$$\delta_c(\Omega_m(z)) = \frac{3}{20}(12\pi)^{2/3}\Omega_m(z)^{0.0185}. \quad (2.3)$$

Ultimately, we analyse a simulation that is in comoving coordinates and so we use today's mass density  $\bar{\rho}_{m,0}$  in equation (2.1). This means that the halo mass function, as it is defined

---

<sup>2</sup>See equation (2.7) for the formal definition of  $\Delta_v$ .

here, captures only the growth of structure due to gravitational attraction as a function of time, but it factors out the expansion of the universe.<sup>3</sup>

Using the halo mass function, the mean dark matter density can be calculated as an integral over the dark matter mass. To predict the density in the simulations at a specific redshift, the lower and upper bound of the integral are given by the minimal and maximal simulated halo mass at that time  $\mathbf{M}_{\text{cut}}(z) = [M_{\text{min}}(z), M_{\text{max}}(z)]$ :

$$\bar{\rho}_{\text{m},0}(z; \mathbf{M}_{\text{cut}}(z)) = \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} M \frac{dn(M, z)}{dM} dM. \quad (2.4)$$

This expression is again in comoving coordinates and the redshift dependency only describes the hierarchical growth of halos, but it does not include the expansion of the universe.

The fundamental assumption of the halo model, that all dark matter is arranged in halos,  $\bar{\rho}_{\text{m},0} = \bar{\rho}_{\text{m},0}(z; M_{\text{min}} = 0, M_{\text{max}} = \infty)$ , results in a normalisation condition for the mass function or multiplicity function, respectively:

$$1 \equiv \frac{1}{\bar{\rho}_{\text{m},0}} \int_0^{\infty} M \frac{dn(M, z)}{dM} dM = \int_0^{\infty} f(\nu) d\nu. \quad (2.5)$$

In PyCosmo, we also provide four different linear halo biases to describe how halos cluster on large scales relative to the underlying matter field: Mo and White [52], Sheth & Tormen [47], Sheth, Mo & Tormen [53], and Tinker et al. [48]. As for the mass function of Tinker et al., we have also implemented the bias parameters for  $\Delta_v = 200$  found in [48].

The spatial distribution of dark matter within individual halos is modeled using a Navarro-Frenk-White (NFW) profile [54]:

$$\rho(r, M) = \frac{\rho_s}{\left(\frac{r}{r_s}\right) \left(1 + \frac{r}{r_s}\right)^2}. \quad (2.6)$$

The profile is parameterised by the scale radius  $r_s$  and the normalisation  $\rho_s$ . The scale radius is usually defined by the halo mass dependent concentration parameter, which is the ratio between the virial radius of a halo with mass  $M$ ,  $r_{\text{vir}}(M)$ , and the scale radius:  $c(M) = r_{\text{vir}}(M)/r_s$ . For the concentration parameter, we make use of the fitting function from [55] and the virial radius is given as:

$$r_{\text{vir}}(M) = \left( \frac{3M}{4\pi\bar{\rho}_{\text{m},0}\Delta_v} \right)^{1/3}. \quad (2.7)$$

As for the critical overdensity for spherical collapse, we also allow for a cosmology dependent expression for  $\Delta_v$  [56].<sup>4</sup> For a flat  $\Lambda$ CDM cosmology,  $\Omega_\Lambda \neq 0$  and  $\Omega_\Lambda + \Omega_m = 1$ :

$$\Delta_v = \frac{18\pi^2 + 82d(z) - 39d(z)^2}{\Omega_m(z)}, \quad (2.8)$$

<sup>3</sup>If the expansion of the universe should be included, one has to use  $\bar{\rho}_m(z) = \bar{\rho}_{m,0}(1+z)^3$  in the definition of the halo mass function.

<sup>4</sup>Since  $r_{\text{vir}}$  in equation (2.7) is defined with respect to the matter density rather than the critical density, we have an additional  $1/\Omega_m(z)$  compared to [56].

and for  $\Omega_\Lambda = 0$  and  $\Omega_m \neq 1$ :

$$\Delta_v = \frac{18\pi^2 + 60d(z) - 32d(z)^2}{\Omega_m(z)}, \quad (2.9)$$

with  $d(z) = \Omega_m(z) - 1$ . Both expressions simplify to  $\Delta_v = 18\pi^2 \approx 178$  for  $\Omega_m = 1$ . The normalisation  $\rho_s$  results from the requirement that integrating the profile over a sphere with a radius equal to the virial radius returns the total halo mass  $M$ . This leads to the following expression:

$$\rho_s = \frac{\Delta_v \bar{\rho}_{m,0}}{3} \frac{c(M)^3}{\ln(1 + c(M)) - c(M)/(1 + c(M))}. \quad (2.10)$$

### 2.1.2 Dark matter power spectrum

After defining the mass function, the halo bias, and the halo profile, the non-linear dark matter power spectrum can be calculated. Therefore, we adopt the convention given by the Fourier transform of the dark matter overdensity field,  $\delta_m(\vec{x}, z) = (\rho_m(\vec{x}, z) - \bar{\rho}_m(z))/\bar{\rho}_m(z)$ :<sup>5</sup>

$$\langle \tilde{\delta}_m(\vec{k}, z) \tilde{\delta}_m^*(\vec{k}', z) \rangle = (2\pi)^3 \delta_D^{(3)}(\vec{k} - \vec{k}') P_{\text{DM}}(k, z). \quad (2.11)$$

In the halo model formalism, the power spectrum can be divided into a one- and two-halo term,  $P_{\text{DM},1h}$  and  $P_{\text{DM},2h}$ , respectively:

$$P_{\text{DM}}(k, z; \mathbf{M}_{\text{cut}}(z)) = P_{\text{DM},1h}(k, z; \mathbf{M}_{\text{cut}}(z)) + P_{\text{DM},2h}(k, z; \mathbf{M}_{\text{cut}}(z)). \quad (2.12)$$

While the one-halo term describes the contributions that come from correlations within a single halo, the two-halo term describes the correlations between different halos. As with the mean dark matter density, the two terms can be written as integrals over the halo mass  $M$ . To enable a direct comparison with the simulations, the lower and upper limits are again given by the smallest and largest simulated halos,  $M_{\text{min}}(z)$  and  $M_{\text{max}}(z)$ :

$$P_{\text{DM},1h}(k, z) = \frac{1}{\bar{\rho}_{m,0}^2(z; \mathbf{M}_{\text{cut}}(z))} \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} \frac{dn(M, z)}{dM} M^2 |\tilde{u}(k|M)|^2 dM \quad (2.13)$$

$$P_{\text{DM},2h}(k, z) = \frac{P_{\text{lin}}(k, z)}{\bar{\rho}_{m,0}^2(z; \mathbf{M}_{\text{cut}}(z))} \left[ \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} \frac{dn(M, z)}{dM} M b(M, z) \tilde{u}(k|M) dM \right]^2. \quad (2.14)$$

These expressions include the linear matter power spectrum  $P_{\text{lin}}(k, z)$ , for which we use the fit of Eisenstein and Hu [57], and the Fourier transform of the normalised spherically symmetric halo profile truncated at the virial radius  $\tilde{u}(k|M)$ . If we use the NFW profile from equation (2.6), we get [45]:

$$\begin{aligned} \tilde{u}(k|M) &= \frac{4\pi}{M} \int_0^{r_{\text{vir}}(M)} \rho(r, M) \frac{\sin(kr)}{kr} r^2 dr \\ &= \frac{4\pi \rho_s r_s^3}{M} \left( \sin(kr_s) [\text{Si}([1 + c(M)]kr_s) - \text{Si}(kr_s)] - \frac{\sin(c(M)kr_s)}{(1 + c(M))kr_s} \right. \\ &\quad \left. + \cos(kr_s) [\text{Ci}([1 + c(M)]kr_s) - \text{Ci}(kr_s)] \right). \end{aligned} \quad (2.15)$$

---

<sup>5</sup> $\delta_D^{(3)}$  denotes the three-dimensional Dirac delta function.

In this expression, we make use of the sine and cosine integrals:

$$\text{Ci}(x) = - \int_x^\infty \frac{\cos t}{t} dt \quad \text{and} \quad \text{Si}(x) = \int_0^x \frac{\sin t}{t} dt. \quad (2.16)$$

However, since in our simulations the halos are considered as point sources and their profile is therefore a Dirac delta function,  $\tilde{u}(k|M) = 1$  in our case. It is worth noting that regardless of whether it is assumed that the halos have a profile or not, on very large scales (i.e. for small  $k$ )  $\tilde{u}(k|M) \rightarrow 1$ . If it is also considered that on large scales the two-halo term has to approach  $P_{\text{lin}}(k, z)$ , we find a second normalisation condition similar to equation (2.5):

$$1 \equiv \frac{1}{\bar{\rho}_{\text{m},0}} \int_0^\infty \frac{dn(M, z)}{dM} M b(M, z) dM = \int_0^\infty f(\nu) b(\nu) d\nu. \quad (2.17)$$

### 2.1.3 Dark matter angular power spectrum

The final products of our simulation pipeline are sky maps of the projected dark matter distribution with a given width in the radial direction. The key statistic that we are interested in is thus the angular power spectrum. It is a three-dimensional integral in Fourier space and along the radial comoving distance  $\chi(z)$  of the power spectrum. For a flat cosmology, it takes the following form:

$$C_{\ell, \text{DM}} = \frac{2}{\pi} \int k^2 dk \int_0^\infty d\chi W(\chi) j_\ell(k\chi) \int_0^\infty d\chi' W(\chi') j_\ell(k\chi') \times \sqrt{P_{\text{DM}}(k, z(\chi); \mathbf{M}_{\text{cut}}(z(\chi))) \cdot P_{\text{DM}}(k, z(\chi'); \mathbf{M}_{\text{cut}}(z(\chi')))}. \quad (2.18)$$

The power spectrum is weighted by window functions  $W(\chi(z))$  and spherical Bessel functions  $j_\ell(k\chi)$ . It should be noted that for non-flat cosmologies, the spherical Bessel functions are replaced by hyper-spherical Bessel functions. In our case,  $W(\chi(z))$  is a top hat function in the redshift interval of a map and is therefore only non-zero in a very limited redshift or  $\chi(z)$  range. Furthermore, the function is normalised as:

$$\int_0^\infty W(\chi) d\chi = 1. \quad (2.19)$$

Since equation (2.18) involves three nested integrals, it is common to use the Limber approximation [58] to accelerate the computation:

$$\begin{aligned} C_{\ell, \text{DM}} &\simeq \int d\chi \frac{W^2(\chi)}{r^2(\chi)} P_{\text{DM}}\left(k = \frac{\ell + 1/2}{r(\chi)}, z(\chi); \mathbf{M}_{\text{cut}}(z(\chi))\right) \\ &= \int dz \frac{c}{H(z)} \frac{W^2(z)}{r^2(z)} P_{\text{DM}}\left(k = \frac{\ell + 1/2}{r(z)}, z; \mathbf{M}_{\text{cut}}(z)\right), \end{aligned} \quad (2.20)$$

with the angular diameter distance  $r(\chi)$  and the Hubble parameter  $H(z)$ . The Limber approximation simplifies further in the case of a very narrow window function. Under the assumption that the power spectrum and the angular diameter distance remain constant

over the limited integration range, they can be evaluated at a fixed  $\chi^*$  and factored out of the integral. The normalisation of the window function then ensures that no integration is required:

$$C_{\ell, \text{DM}} \simeq \frac{W(\chi^*)}{r^2(\chi^*)} P_{\text{DM}} \left( k = \frac{\ell + 1/2}{r(\chi^*)}, z(\chi^*); \mathbf{M}_{\text{cut}}(z(\chi^*)) \right). \quad (2.21)$$

When  $r(\chi)$  is used in the expression of the Limber approximation, it is generally valid for flat and non-flat cosmologies. For a flat cosmology, it is straight forward to derive equation (2.20) from equation (2.18) [59]: By using that the product of two spherical Bessel functions for high  $\ell$  has a sharp peak around  $k\chi \approx k\chi' \approx \sqrt{\ell(\ell+1)} \approx \ell + 1/2$ ,  $P_{\text{DM}}$  can be taken out of the integration over  $k$  as long as it is approximately constant over the peak. Subsequently using the orthogonal property of spherical Bessel functions<sup>6</sup> and remembering that  $r(\chi) = \chi$  for the flat case finally results in equation (2.20). The result for non-flat cosmologies can be recovered by generalising the calculations found in e.g. [60, 61] for the flat case by using  $r(\chi)$  in the definition of the comoving position  $\vec{x} \approx (r(\chi)\theta_1, r(\chi)\theta_2, \chi)$  instead of just  $\chi$ .

However, the Limber approximation tends to diverge for narrow window functions, as is the case for our maps. We will see this in our results for the angular power spectrum (see section 4.4) when we compare it to the spectrum measured from our maps.

Since performing the complete integration in equation (2.18) is numerically challenging, particularly for very narrow redshift bins, we opted to use a Monte-Carlo integration. In future work, we aim to explore more advanced methods, e.g. [62, 63], to perform the integration. Note that this computation is not implemented in `PyCosmo`, but we make use of its halo model power spectrum to perform the computation. However, the Limber approximation is fully integrated into the package and functional.

## 2.2 HI halo model

Similarly to the dark matter halo model, one can also describe a HI halo model, which only needs minor changes compared to the former. Since in the post-reionization era HI is assumed to reside mainly in dark matter halos, we model the HI mass distribution by assigning HI mass to the dark matter halos through a  $M_{\text{HI}}-M$  relation. Two of the before mentioned ingredients for a complete halo model can be inherited from the dark matter halo model: the halo mass function and the halo bias. We only introduce a new HI halo profile which replaces the dark matter halo profile from the last section.

### 2.2.1 HI-halo mass relation

To assign the right amount of HI to the dark matter halos, we use a HI-halo mass relation that has been constrained by HI data in [64]. As constraints, they utilized data from the small-scale clustering, column density and mass function of HI galaxies at low redshift, intensity mapping measurements at intermediate redshifts and UV/optical observations of Damped Lyman Alpha systems at higher redshifts. The amount of HI mass in a dark matter halo of mass  $M$  as a function of redshift is given as:

$$M_{\text{HI}}(M, z) = \alpha f_{\text{H,c}} M \left( \frac{M}{10^{11} h^{-1} \text{M}_\odot} \right)^\beta \exp \left[ - \left( \frac{v_{\text{c},0}}{v_{\text{c}}(M, z)} \right)^3 \right]. \quad (2.22)$$

---

<sup>6</sup>  $\frac{2}{\pi} \int k^2 dk j_\ell(k\chi) j_\ell(k\chi') = \delta_{\text{D}}^{(1)}(\chi - \chi')/\chi^2$ , where  $\delta_{\text{D}}^{(1)}$  denotes the one-dimensional Dirac delta function.

The overall normalisation  $\alpha$  stands for the mass fraction of HI in halos relative to the cosmic fraction  $f_{\text{H},c} = (1 - Y_p)\Omega_b/\Omega_m$ . The logarithmic slope of the relation is described by  $\beta$ . The minimum virial velocity,  $v_{c,0}$ , corresponds to a minimum virial halo mass of halos that are still able to host HI. In [64], they find the best-fit values  $\alpha = 0.09$ ,  $\beta = -0.58$ , and  $\log v_{c,0} = 1.56$  for these parameters. Although they assumed a slightly different cosmology,<sup>7</sup> we also use these values in this work. We use this relation with the best-fit parameter values in both the theoretical analysis and the simulation although they could be varied to account for uncertainties on their values. For the virial velocity as a function of halo mass, we use [65]:

$$v_c(M, z) = 96.6 \text{ km s}^{-1} \left( \frac{\Delta_v \Omega_m h^2}{24.4} \right)^{1/6} \left( \frac{1+z}{3.3} \right)^{1/2} \left( \frac{M}{10^{11} \text{ M}_\odot} \right)^{1/3}. \quad (2.23)$$

The relationship between the HI mass and the halo mass describes an exponential cut-off at low halo masses, which is due to the fact that the halos are too small to self-shield HI from ionization and thus host HI. The redshift dependence of the virial velocity leads to less and less HI being hosted in small halos as time evolves.

With this prescription for the assignment of HI to the halos, the mean comoving HI density at a given redshift for given minimum and maximum dark matter halo masses can again be written as an integral over the dark matter mass. The halo mass function only has to be weighted with the HI mass instead of dark matter mass:

$$\bar{\rho}_{\text{HI},0}(z; \mathbf{M}_{\text{cut}}(z)) = \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} M_{\text{HI}}(M, z) \frac{dn(M, z)}{dM} dM. \quad (2.24)$$

### 2.2.2 HI halo profile

To characterise the HI halo profile, we also follow [64]. In observations and simulations [66, 67] it has been found that the HI distribution can be modelled by a spherically symmetric exponential profile in the radial direction:

$$\rho(r, M) = \rho_0 \exp(-r/r_{s,\text{HI}}). \quad (2.25)$$

Here  $r_{s,\text{HI}} = r_{\text{vir}}(M)/c_{\text{HI}}(M, z)$  is now the scale radius of HI defined through the concentration parameter  $c_{\text{HI}}(M, z)$ :

$$c_{\text{HI}}(M, z) = c_{\text{HI},0} \left( \frac{M}{10^{11} \text{ M}_\odot} \right)^{-0.109} \frac{4}{(1+z)^\gamma}, \quad (2.26)$$

with the best-fit values found in [64] for the two parameters:  $c_{\text{HI},0} = 28.65$  and  $\gamma = 1.45$ . The normalisation constant of the profile is approximated to be  $\rho_0 = M_{\text{HI}}(M, z)/(8\pi r_s^3)$  such that the HI mass inside the virial radius integrates to  $M_{\text{HI}}(M, z)$ .

### 2.2.3 HI power spectrum

Analogous to the dark matter power spectrum, we adopt the convention for the non-linear HI power spectrum given by the Fourier transform of the HI mass overdensity field,  $\delta_{\text{HI}}(\vec{x}, z) =$

---

<sup>7</sup> $\Omega_m = 0.281$ ,  $\Omega_\Lambda = 0.719$ ,  $h = 0.71$ ,  $\Omega_b = 0.0462$ ,  $\sigma_8 = 0.8$ ,  $n_s = 0.96$ , and  $Y_p = 0.24$ .

$(\rho_{\text{HI}}(\vec{x}, z) - \bar{\rho}_{\text{HI}}(z))/\bar{\rho}_{\text{HI}}(z)$ :

$$\langle \tilde{\delta}_{\text{HI}}(\vec{k}, z) \tilde{\delta}_{\text{HI}}^*(\vec{k}', z) \rangle = (2\pi)^3 \delta_{\text{D}}^{(3)}(\vec{k} - \vec{k}') P_{\text{HI}}(k, z). \quad (2.27)$$

It can be expressed again as the sum of a one- and two-halo term, which are given as:

$$P_{\text{HI},1\text{h}}(k, z) = \frac{1}{\bar{\rho}_{\text{HI},0}^2(z; \mathbf{M}_{\text{cut}}(z))} \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} \frac{dn(M, z)}{dM} M_{\text{HI}}^2(M, z) |\tilde{u}_{\text{HI}}(k|M)|^2 dM \quad (2.28)$$

$$P_{\text{HI},2\text{h}}(k, z) = \frac{P_{\text{lin}}(k, z)}{\bar{\rho}_{\text{HI},0}^2(z; \mathbf{M}_{\text{cut}}(z))} \left[ \int_{M_{\text{min}}(z)}^{M_{\text{max}}(z)} \frac{dn(M, z)}{dM} M_{\text{HI}}(M, z) b(M, z) \tilde{u}_{\text{HI}}(k|M) dM \right]^2. \quad (2.29)$$

The Fourier transform of the normalised HI halo profile truncated at the virial radius is given by:

$$\tilde{u}_{\text{HI}}(k|M) = \frac{4\pi}{M_{\text{HI}}(M, z)} \int_0^{r_{\text{vir}}(M)} \rho(r, M) \frac{\sin(kr)}{kr} r^2 dr. \quad (2.30)$$

Substituting the exponential profile (2.25) into this expression, we obtain:

$$\tilde{u}_{\text{HI}}(k|M) = \frac{4\pi \rho_0 r_{\text{s,HI}}^3 u_1(k|M)}{M_{\text{HI}}(M, z)}, \quad (2.31)$$

with

$$u_1(k|M) = \frac{2}{(1 + k^2 r_{\text{s,HI}}^2)^2}. \quad (2.32)$$

In comparison to our halo-based HI simulations, we again consider point sources and therefore use  $\tilde{u}_{\text{HI}}(k|M) = 1$ .

#### 2.2.4 HI angular power spectrum

The full expression of the HI angular power spectrum and its Limber approximation are again obtained by integrating the three-dimensional HI power spectrum. The full expression for a flat cosmology is given by:

$$C_{\ell,\text{HI}} = \frac{2}{\pi} \int k^2 dk \int_0^\infty d\chi W(\chi) j_\ell(k\chi) \int_0^\infty d\chi' W(\chi') j_\ell(k\chi') \times \sqrt{P_{\text{HI}}(k, z(\chi); \mathbf{M}_{\text{cut}}(z(\chi))) \cdot P_{\text{HI}}(k, z(\chi'); \mathbf{M}_{\text{cut}}(z(\chi')))}. \quad (2.33)$$

And the Limber approximation takes the form:

$$\begin{aligned} C_{\ell,\text{HI}} &\simeq \int d\chi \frac{W^2(\chi)}{r^2(\chi)} P_{\text{HI}}\left(k = \frac{\ell + 1/2}{r(\chi)}, z(\chi); \mathbf{M}_{\text{cut}}(z(\chi))\right) \\ &= \int dz \frac{c}{H(z)} \frac{W^2(z)}{r^2(z)} P_{\text{HI}}\left(k = \frac{\ell + 1/2}{r(z)}, z; \mathbf{M}_{\text{cut}}(z)\right), \end{aligned} \quad (2.34)$$

which further simplifies again when assuming a very narrow window function:

$$C_{\ell, \text{HI}} \simeq \frac{W(\chi^*)}{r^2(\chi^*)} P_{\text{HI}} \left( k = \frac{\ell + 1/2}{r(\chi^*)}, z(\chi^*); \mathbf{M}_{\text{cut}}(z(\chi^*)) \right). \quad (2.35)$$

As already for dark matter, equation (2.33) is performed with a Monte-Carlo integration and is not implemented in `PyCosmo`, but the Limber approximation is.

### 3 Simulation methods

Having discussed the theoretical halo model in detail, we turn our attention in the following sections to the procedure for generating 21 cm emission maps based on a halo model approach. First, we describe the simulation of the dark matter halo distribution, i.e. the simulation of a past light cone halo catalogue, which is subsequently populated with HI using equation (2.22). We also explain how maps are generated from the past light cone.

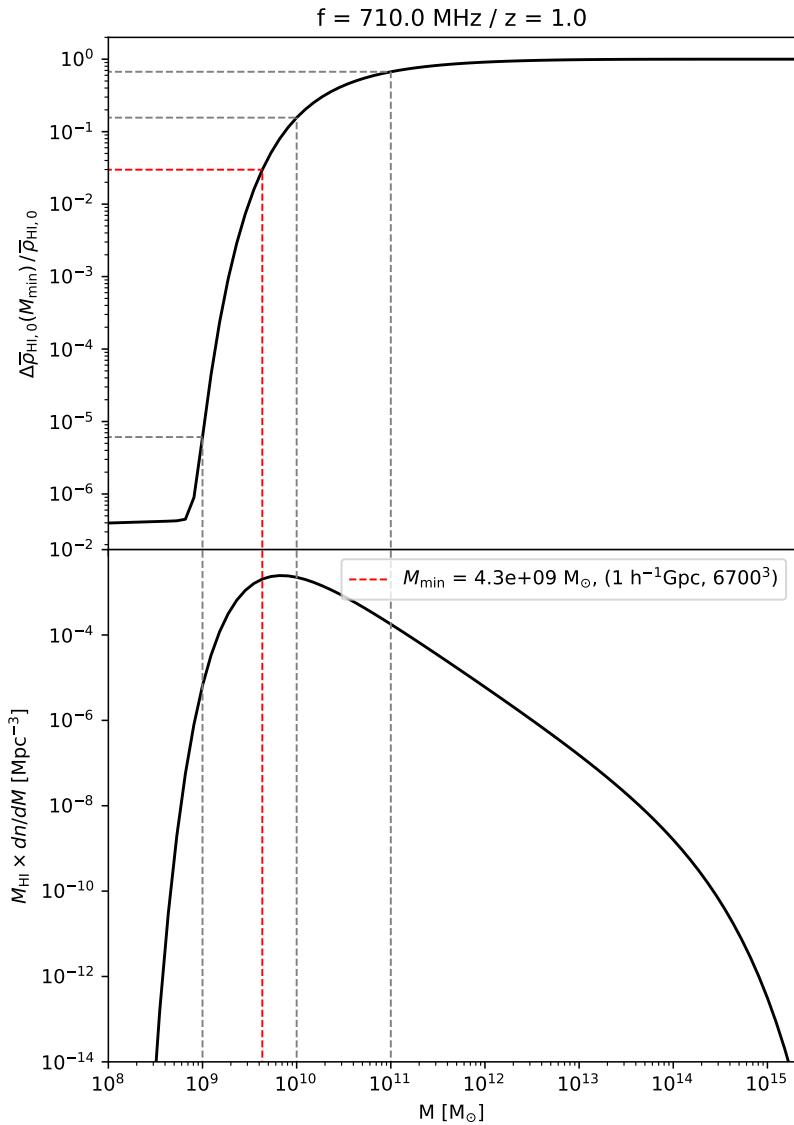
#### 3.1 Simulation of dark matter and HI halos

As already mentioned in section 1, we try to seek an intermediate solution between realistic but computationally intensive  $N$ -body simulations and computationally efficient but simplified statistical simulations. Therefore, we use the `PINOCCHIO` code<sup>8</sup> (PINpointing Orbit Crossing Collapsed HIerarchical Objects) [31–36]. It has been shown to be much faster while still recovering clustering properties at the 10 %–level when compared to  $N$ -body. The massive speed-up is due to the usage of Lagrangian Perturbation Theory together with the Extended Press & Schechter formalism [68] for non-spherical collapse to calculate the masses, positions, velocities, and merger histories of dark matter halos. Furthermore, `PINOCCHIO` generates the past light cone halo catalogue with continuous time sampling on-the-fly.

The goal of intensity mapping is to observe large cosmological volumes, e.g. a comoving volume of approximately  $90 \text{ h}^{-3} \text{Gpc}^3$  with HIRAX, at low angular resolution rather than resolving individual objects. Therefore, our interest lies in the simulation of such vast volumes. However, although there is more HI in large dark matter halos, as we have discussed in section 2.2.1, there are many more small halos than large ones, making their contribution to the total HI content not negligible. Thus, not only large volumes are required, but also a sufficiently good mass resolution, which would be hard to achieve with conventional full  $N$ -body simulations. Figure 1 quantifies how much HI is lost at redshift  $z = 1$  for a given mass resolution. The upper panel shows the relative difference in the mean HI density, as described in equation (2.24), for different lower mass cuts, compared to the total density when no cut is applied. The bottom panel shows the integrand of  $\bar{\rho}_{\text{HI},0}$ .<sup>9</sup> The decrease of the integrand at the high mass end results from the rarity of high mass halos described by the mass function. The exponential decrease at the low mass end results from the inability of these halos to self-shield HI from ionization. However, this exponential decrease facilitates to simulate a significant part of HI with a limited resolution in the first place: As long as halo masses just below the peak of the integrand are simulated, most of the HI is captured. As described in section 2.2.1, this cut-off shifts to lower masses with increasing redshift, which requires a better mass resolution of the simulation.

<sup>8</sup>In this work, we use version 4.1.3 available on <https://github.com/pigimonaco/Pinocchio>.

<sup>9</sup>For this figure the mass function of Watson is assumed.



**Figure 1. Upper panel:** Loss in the mean HI density as a function of the minimum simulated dark matter halo mass  $M_{\min}$  relative to the total HI density. In the calculation of the density the mass function of Watson [49] is assumed. The red dashed line indicates the loss for the simulated minimum halo mass achieved for our simulation with a box size of  $1 \text{ h}^{-1} \text{Gpc}$  and  $6700^3$  simulation particles per box. The grey dashed lines are shown for better visualisation. **Lower panel:** Integrand of the mean HI density as given in equation (2.24).

Since, in future work, we are interested in applying our generated 21 cm maps to the instrument simulation and analysis pipeline of HIRAX, we choose our light cone geometry according to the survey configuration. Given that memory requirements are a limiting factor for the simulation, a good balance must be achieved between the volume of the past light cone and the mass resolution. HIRAX will operate in drift-scan mode, observing strips of constant declination. Therefore, instead of running a full-sky simulation, we can confine the simulation to the volume covered by HIRAX, allowing computational resources to be allocated toward maximizing the mass resolution. We decide to simulate the past light cone to cover declinations between  $-15^\circ$  and  $-35^\circ$  and a frequency range from 700 to 800 MHz, that means for redshifts from 0.77 to 1.03. Pushing `PINOCCHIO` to its limits for this light cone geometry, we manage to run a simulation for a box size of  $1 \text{ h}^{-1}\text{Gpc}$  and  $6700^3$  simulation particles per box, choosing that the smallest identified halos consist of at least ten particles. To cover the cosmological volume, 40 boxes are replicated and stacked together with periodic boundary conditions. Although the full light cone has a width of  $0.46 \text{ h}^{-1}\text{Gpc}$ , which is about half the box size, the geometry of the light cone requires two box replications in the radial direction. In the transverse direction, the circumference of the light cone is between 10.4 and  $14.2 \text{ h}^{-1}\text{Gpc}$  at  $z = 1.03$  and between 8.3 and  $11.4 \text{ h}^{-1}\text{Gpc}$  at  $z = 0.77$ . This requires approximately between 20 and 24 box replications, depending on the distance from the central observer. Box replications will introduce artefacts in the clustering signal at scales corresponding to the box size and larger. With the given configuration, the smallest simulated halos have masses of  $4.3 \times 10^9 M_\odot$ , which is indicated by the red dashed line in figure 1. At  $z = 1$ , the simulation is therefore missing around 3% of the total HI. The missing fraction shrinks to around 2% at  $z = 0.77$ .

The displacements of the particles are calculated with third-order LPT, while the grouping of the collapsed particles into halos, a step called fragmentation, is performed with second-order LPT. For fragmentation, the simulation box is divided into sub-volumes. Each sub-volume is assigned to a separate MPI task. Since the different tasks do not communicate with each other, the calculations need to be extended to a boundary layer that overlaps between the sub-volumes. The thickness of the boundary layer is defined as the Lagrangian radius of the largest halo expected at the last redshift multiplied by a boundary layer factor set in the `PINOCCHIO` parameter file. As recommended in the documentation,<sup>10</sup> we set this parameter to 1, which leads to a significant overhead in memory requirements. The simulation is run on 2400 nodes with 12 cores each on the supercomputer Piz Daint (Cray XC40/50) [69], hosted by the Swiss National Supercomputing Centre (CSCS). It requires a total of  $\sim 150 \text{ TB}$  of memory and  $\sim 40'000$  CPU hours.

### 3.2 Map generation

Since the great advantage of intensity mapping is the direct mapping between the frequency of the observation and redshift, we generate thin maps from the past light cone by projecting equally spaced bins in frequency to resolve the structure along the line of sight. To generate pixelated maps on a sphere, the software package `HEALPix`<sup>11</sup> [70], in particular the Python

<sup>10</sup><https://github.com/pigimonaco/Pinocchio/blob/master/DOCUMENTATION>.

<sup>11</sup><http://healpix.sourceforge.net>.

package `healpy` [71], is used. The *NSIDE* parameter, which defines the number of pixels on the map, is set to 512. This corresponds to an angular resolution of about  $6.9'$ , which is just below the maximum resolution of  $12'$  of the initially planned HIRAX 256-element array [26]. The frequency resolution of HIRAX will be 0.39 MHz [26]. However, as illustrated in appendix A, we found that for maps with such a narrow bandwidth, Poisson noise (see also section 4.4) dominates the cosmological signal for a wide range of scales in the angular power spectrum due to the small number of discrete halos per pixel. Since we are mainly interested here in the simulated cosmological signal and whether it matches theoretical expectations, we choose a coarser bandwidth of 5 MHz, resulting in 20 separate maps. In this work, we deliberately use cosmological redshifts rather than observed redshifts, as the latter would introduce additional complexities in comparing our results to theoretical predictions. We plan to incorporate observed redshifts and their implications in future work. A thorough analysis of PINOCCHIO’s accuracy in simulating the velocity field will be essential for that. Nevertheless, prior studies have demonstrated good agreement with an *N*-body simulation in the redshift space power spectrum when PINOCCHIO employs second-order LPT for halo construction and third-order LPT for halo displacements [35]. To create dark matter- or HI-mass maps, the pixels are weighted by the masses of the halos falling into them. However, the key observable of a radio interferometer is the brightness temperature. To convert the HI mass maps into HI brightness temperature maps, we use [72]:

$$T_b(\vec{x}, z) = \frac{3hc^3 A_{10}}{32\pi k_B f_{21}^2 m_p} \frac{(1+z)^2}{H(z)} \rho_{\text{HI}}(\vec{x}, z), \quad (3.1)$$

where  $h$  is the Planck constant,  $c$  the speed of light,  $A_{10} \approx 2.869 \times 10^{-15} \text{ s}^{-1}$  [73] the Einstein coefficient for spontaneous emission,  $k_B$  the Boltzmann constant, and  $m_p$  the proton mass. To calculate  $\rho_{\text{HI}}$ , the HI mass maps are divided by the volume of a voxel.<sup>12</sup> The redshift dependent factor  $(1+z)^2/H(z)$  is calculated by averaging the redshifts of all halos that fall within a pixel.<sup>13</sup>

It is important to note that when comparing the power spectra of the simulation with the theoretical predictions outlined in sections 2.2.3 and 2.2.4, converting the HI mass into brightness temperature is not necessary. As long as overdensities are calculated initially, the spectra in equations (2.27) and (2.33) are computed. This is because the brightness temperature is proportional to the HI mass density:

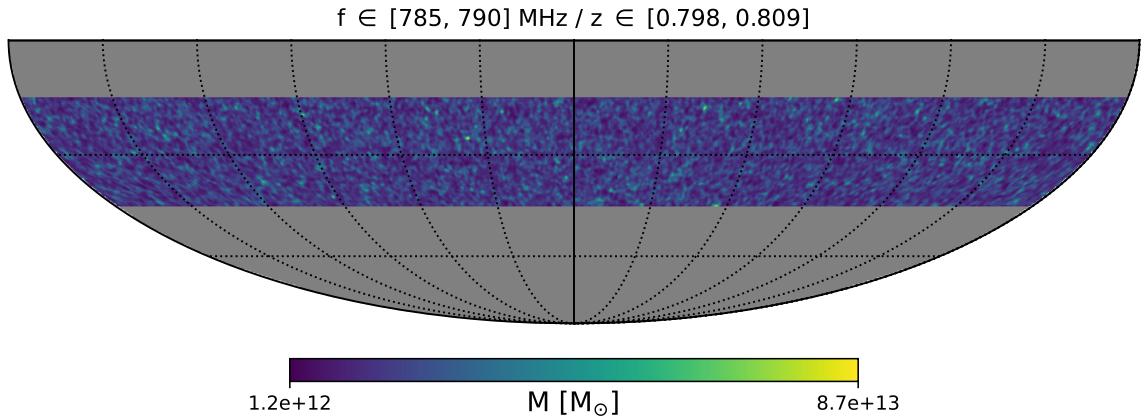
$$\delta T_b(\vec{x}, z) = \frac{T_b(\vec{x}, z) - \bar{T}_b(z)}{\bar{T}_b(z)} = \frac{\rho_{\text{HI}}(\vec{x}, z) - \bar{\rho}_{\text{HI}}(z)}{\bar{\rho}_{\text{HI}}(z)} = \delta_{\text{HI}}(\vec{x}, z). \quad (3.2)$$

## 4 Results

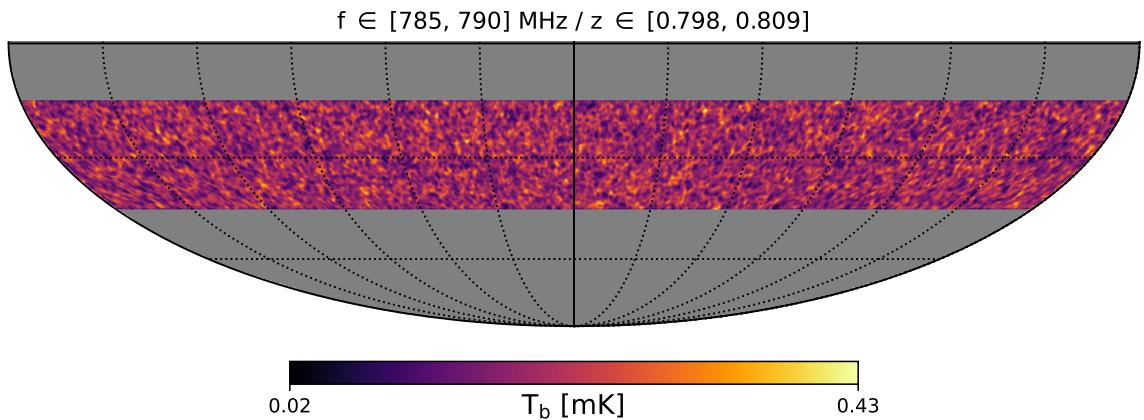
In the following sections, we will compare the statistics of the simulations with the theoretical predictions based on the halo model. This serves as a check to ensure that the simulated signal agrees with theoretical expectations before applying the simulations to instrument-related analyses. The compared statistics are the mass function, the dark matter and HI mass

<sup>12</sup>The volume is calculated by dividing the volume of a 5 MHz thick shell by the number of pixels.

<sup>13</sup>We have found that the central redshift of each bin would also be sufficient for this expression, since the bins are small. However, this would not work for wider bins in redshift.



**Figure 2.** Dark matter map for the frequency bin from 785 to 790 MHz, smoothed with a Gaussian kernel with  $\sigma = 0.005 \text{ rad} \approx 17.2'$ , generated from a halo light cone produced with PINOCCHIO.



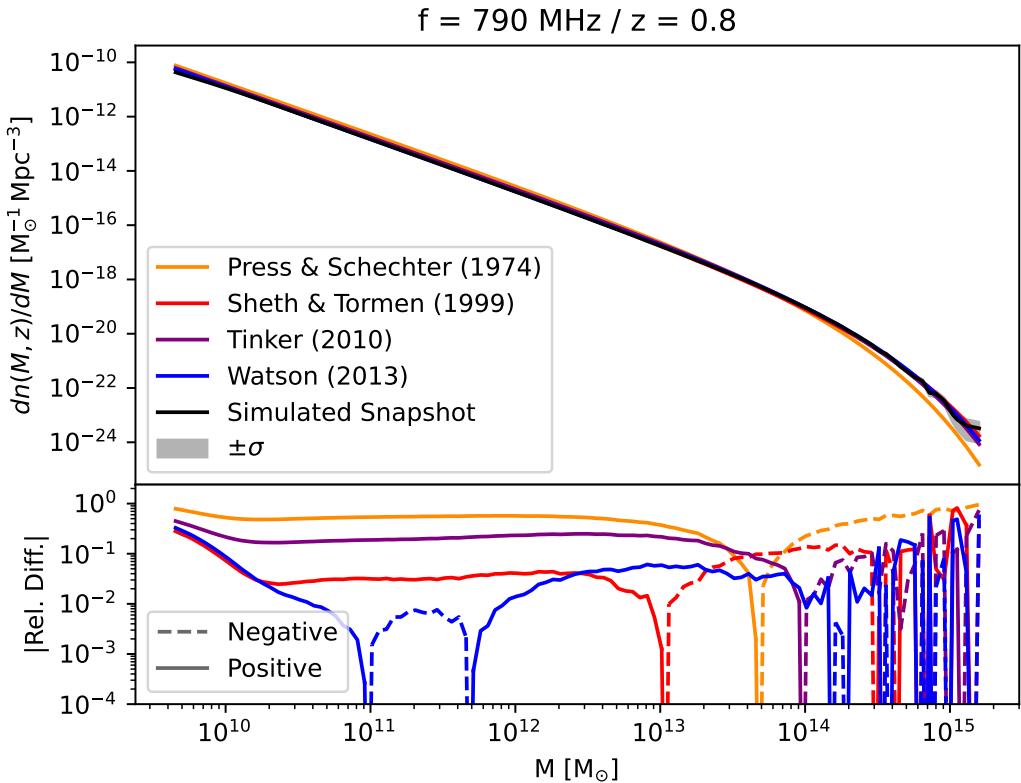
**Figure 3.** HI brightness temperature map for the frequency bin from 785 to 790 MHz, smoothed with a Gaussian kernel with  $\sigma = 0.005 \text{ rad} \approx 17.2'$ .

densities as a function of redshift, the three-dimensional dark matter and HI power spectra, and the corresponding angular power spectra.

Figures 2 and 3 show examples of a dark matter mass map and a HI brightness temperature map, respectively. The maps are for the frequency bin from 785 to 790 MHz corresponding to redshifts from 0.798 to 0.809. For better visualisation, we present versions smoothed with a Gaussian kernel with  $\sigma = 0.005 \text{ rad} \approx 17.2'$ . Unless a quantity is shown as a function of redshift, only the results for the 785 to 790 MHz bin or the snapshot at  $z = 0.8$  are shown in the following sections, as the results for the other frequency bins are similar.

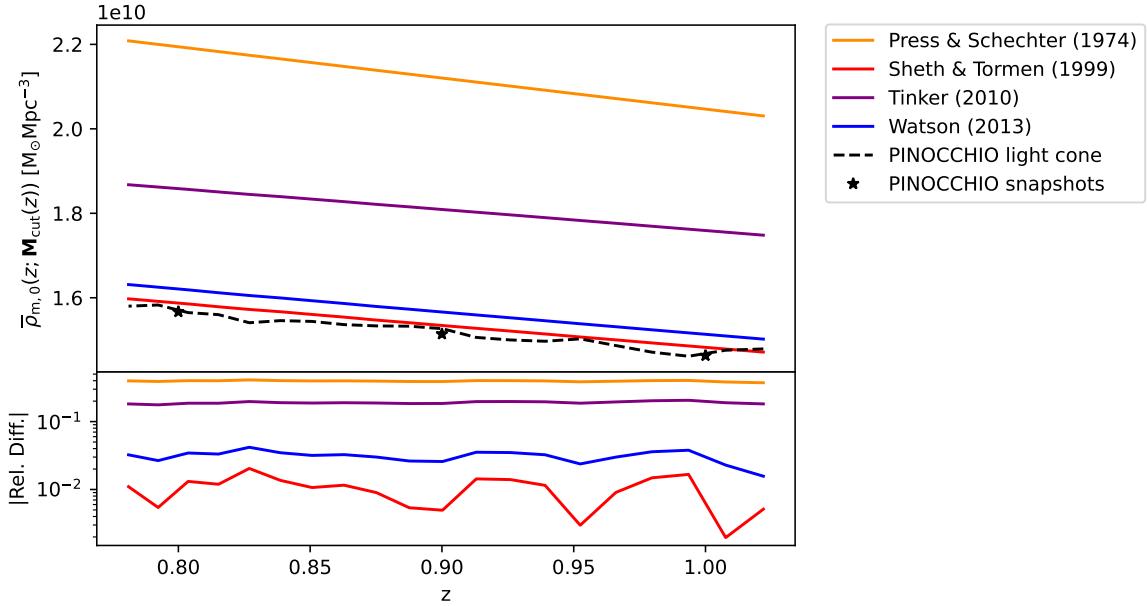
#### 4.1 Mass function

PINOCCHIO offers the option to output snapshots at different redshifts, i.e. dark matter halo catalogues of a single simulation box with mass, position, and velocity for each halo. It also directly provides the corresponding mass functions of these snapshots. We write out snapshots at  $z = 0.8, 0.9, 1$ .



**Figure 4.** Comparison between mass functions implemented in `PyCosmo` (orange, red, purple, blue) [46–49] and the mass function measured on a snapshot of the simulation (black) at frequency 790 MHz or redshift 0.8, respectively (upper panel). The grey shaded area indicates the Poisson error. The lower panel shows the absolute relative difference, with solid (dashed) lines indicating positive (negative) differences.

Figure 4 compares the measured mass function at  $z = 0.8$  with the four options of the mass function implemented in `PyCosmo`. The lower part shows the absolute relative difference between the implementations and the measured function, distinguishing between positive (solid line) and negative differences (dashed line). We find that the mass functions of Watson and Sheth & Tormen agree well with the simulation for intermediate halo masses, with agreement at the level of a few percent. At the low mass end, the simulation lacks halos. This is most likely due to our low choice of at least ten particles per halo. It is possible that the simulation may not be able to identify such small halos accurately. At the high mass end, the measured mass function is dominated by Poisson noise, as there are only a limited number of large mass halos in the snapshot. The grey shaded area indicates the Poisson noise of the snapshot. Overall, these observations are consistent with the mass function findings reported in [35], where they find that `PINOCCHIO` reproduces the analytic mass function of Watson to within 5%.



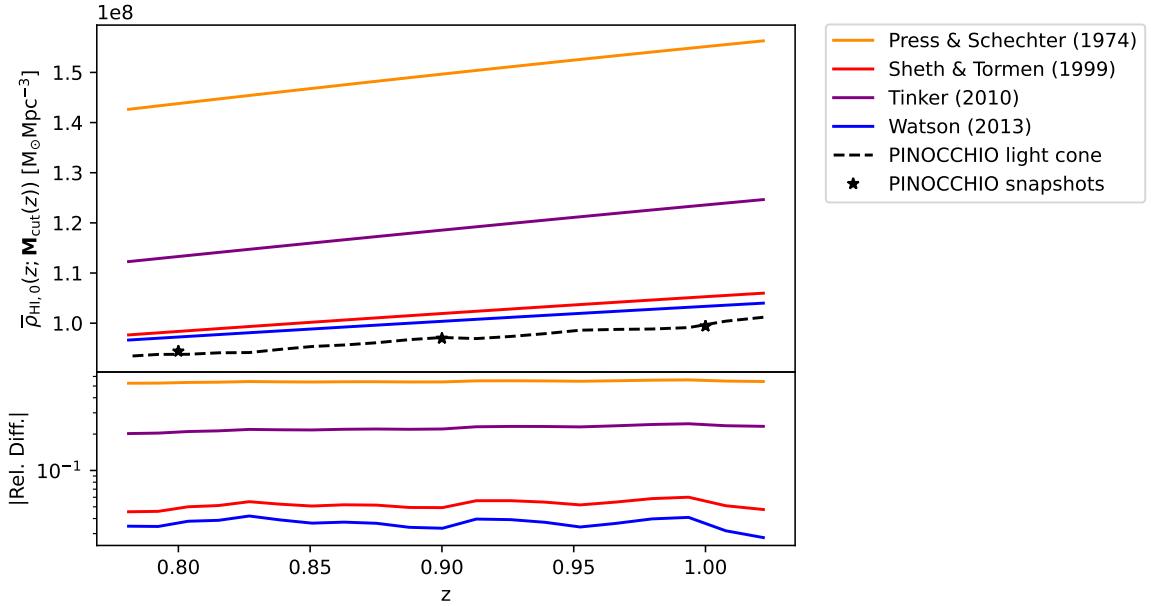
**Figure 5.** Comparison of the mean dark matter density over the simulated redshift range. The predicted densities for the different mass functions in `PyCosmo` (orange, red, purple, blue) [46–49] are compared with the density computed from the snapshots of the simulation at  $z = 0.8, 0.9, 1$  (black stars), and with the density computed from the maps we created (black dashed line). In the lower panel, we show the absolute relative difference between the different lines to the density measured on the light cone.

## 4.2 Density

In a next step, the mass functions weighted by the mass can be integrated over the simulated mass range, as described in equations (2.4) and (2.24), to predict the mean dark matter and HI mass densities for the different frequency bins. Figures 5 and 6 compare the mass densities calculated with the different mass functions in `PyCosmo` to the simulated densities. We check the snapshots and the maps that are generated based on the past light cone. The density in a snapshot is calculated by integrating the measured mass function. The density in a map is calculated by summing the total mass in a map and dividing it by the volume projected onto the map.

As mentioned in section 2.1.1, the densities considered here only capture the hierarchical growth of the halos, but neglect the expansion of the universe. This explains the increasing dark matter density with decreasing redshift. In the case of HI, it must also be taken into account that in the HI model used here (see equation (2.22)) the HI fraction in low-mass halos, with virial velocities below the cut-off velocity  $v_{c,0}$ , decreases as time evolves. Therefore, the HI mass density decreases with decreasing redshift.

There is good agreement between the densities calculated from the snapshots and the maps, i.e. the past light cone. The lower panels show the relative difference between the `PyCosmo` predictions and the densities in the maps. As expected from the comparisons in figure 4, the mass functions from Watson and Sheth & Tormen also give the best agreement for the densities compared to the simulation. For dark matter, the Sheth & Tormen mass



**Figure 6.** Same comparison as in figure 5, but for the HI mass density, equation (2.24), instead of dark matter density.

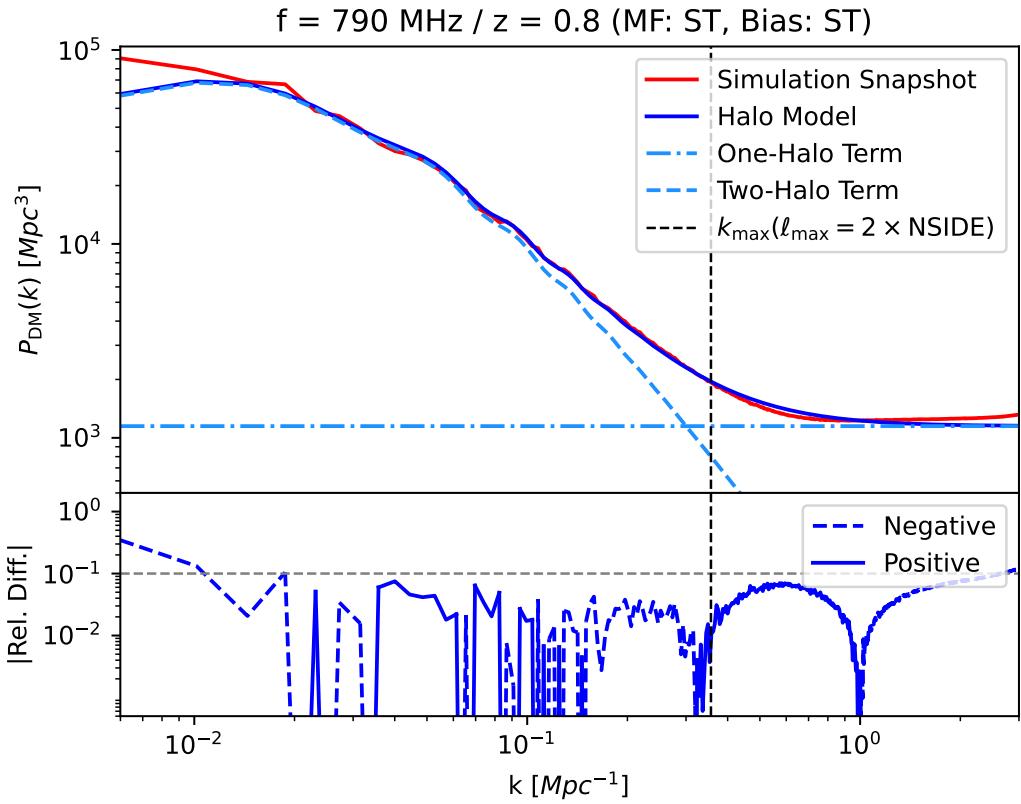
function agrees slightly better with a 0.2 – 2 % difference than the Watson mass function with 1.5 – 4.2 %. Similar results are found in [35], where the number density of halos more massive than  $9.9 \times 10^{12} M_{\odot}$  and  $4.9 \times 10^{13} M_{\odot}$  in the past light cone is compared to the number density in snapshots and a prediction based on the Watson mass function. They find very good agreement between the snapshots and the light cone and again within 5 % when compared to Watson. For HI, the Watson mass function with a relative difference of 2.8 – 4.2 % achieves a better agreement than the Sheth & Tormen mass function, for which we find an agreement of 4.5 – 6 %. This is due to the different mass weighting when considering dark matter halos or HI halos.

### 4.3 Power spectrum

In figures 7 and 8, the three-dimensional power spectra for dark matter and HI calculated from the snapshot are compared with the halo model predictions. To compute the three-dimensional power spectra of the snapshot the Python libraries from `Pylians` [74] have been used. The snapshot provided by `PINOCCHIO` can be used directly to compute the dark matter power spectrum. For HI, however, the dark matter halo mass must first be converted into HI mass using equation (2.22) again. Both figures explicitly show the contributions of the one- and two-halo terms for the theoretical predictions with the halo model. The black dashed line indicates  $k_{\text{max}} = (\ell_{\text{max}} + 1/2)/r(\chi(z))$ ,<sup>14</sup> which corresponds to the maximum multipole moment for which we calculate the angular power spectrum later in section 4.4.

In the previous two sections, we have seen that the mass functions of Watson or Sheth & Tormen agree best with the simulation. When the measured mass function of the snapshot is

<sup>14</sup>Note that this correspondence between  $k$  and  $\ell$  is taken from the argument of the power spectrum in the expression of the Limber approximation.

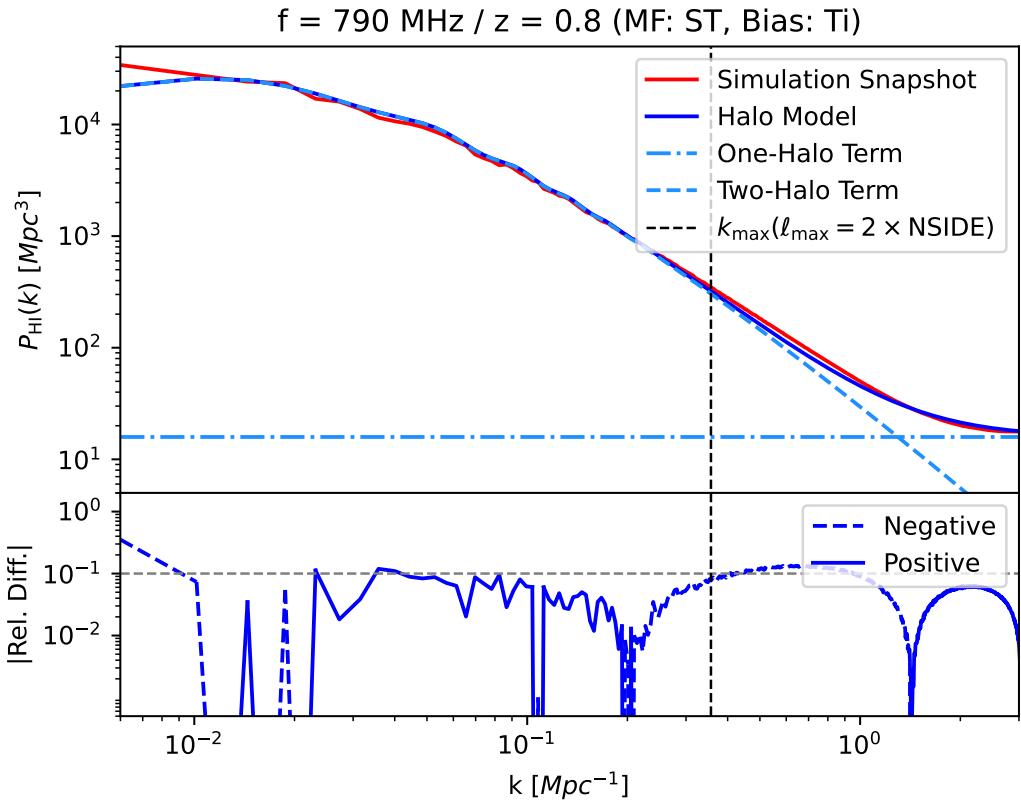


**Figure 7.** Dark matter power spectrum of a simulation snapshot (red) at 790 MHz, or redshift 0.8, compared to the halo model prediction (blue). In addition, the contributions of the one-halo term (dashed-dotted light blue line) and two-halo term (dashed light blue line) are shown. We assume the mass function and linear halo bias of Sheth & Tormen [47] in the halo model calculations. The vertical dashed black line indicates  $k_{\max}$ , which corresponds to the maximum multipole  $\ell_{\max}$  that is considered in the angular power spectrum in section 4.4. In the bottom panel, we show the absolute relative difference between the halo model and the simulation. The horizontal dashed grey line indicates the 10 % relative difference.

used to calculate the power spectrum and the linear halo bias is varied, it is found that for the three snapshots at different redshifts, the bias from the Sheth & Tormen mass function best matches the power spectrum of the snapshot for dark matter and the bias from the Tinker mass function for HI. Looking at all the combinations of mass functions and linear halo biases in PyCosmo, it is found that for dark matter, the combination of the Sheth & Tormen mass function and linear halo bias best fits the power spectrum of the snapshot. For HI, it is the combination of the Sheth & Tormen mass function and the linear halo bias from the Tinker mass function.<sup>15</sup>

As expected, the two-halo term is dominant on large scales, i.e. for small  $k$ 's, describing the correlation between different halos, while the one-halo term is dominant on small scales, i.e. for large  $k$ 's, describing the correlation within a single halo. The latter contributes as an

<sup>15</sup>For HI, the combination of the bias derived from the Tinker mass function with the mass functions from either Watson or Tinker also agrees well with the measured power spectrum.



**Figure 8.** Same comparison as in figure 7, but for the HI power spectrum instead of the dark matter power spectrum. Here, we assume the Sheth & Tormen mass function [47] and the linear halo bias from Tinker [48].

overall constant shot-noise because the halos are assumed to be point sources with no radial profile, as previously described in sections 2.1.2 and 2.2.3, in agreement with the simulations.

We again plot the absolute relative difference between the halo model prediction and the spectrum of the snapshot and find good agreement within a few percent over a broad range of scales for both dark matter and HI. For dark matter this level of agreement is consistent with similar results reported in [35] where the power spectrum of PINOCCHIO is compared to an  $N$ -body simulation within the range of  $2 \times 10^{-2} \text{ Mpc}^{-1} < k < 3 \times 10^{-1} \text{ Mpc}^{-1}$ . The agreement we find here is particularly remarkable given the simplicity of the halo model, especially for HI, since the mass functions and linear halo biases were originally calibrated on dark matter halo simulations. That could potentially explain why the agreement is slightly worse for HI compared to dark matter. The discrepancy between the simulation and the halo model observed for  $k < 10^{-2} \text{ Mpc}^{-1}$  arises from the limited box size of our simulation, i.e.  $1 \text{ h}^{-1} \text{Gpc}$ . This box size corresponds to  $k \approx 4 \times 10^{-3} \text{ Mpc}^{-1}$ , and as a result, modes on scales comparable to that are sampled less often which results in weak statistical power.

We are aware of the optimized variant of the halo model presented in [75, 76], where physically motivated free parameters have been introduced and determined by fitting to simulated power spectra. While these parameters are well-motivated for dark matter, their validity for HI is not guaranteed. To enable a clear comparison between the dark matter

and HI results in this work, we refrained from introducing the adapted dark matter halo model. However, the model described in [75, 76] is implemented in `PyCosmo`, though not made use of in this work.

#### 4.4 Angular power spectrum

Finally, the angular power spectra of the maps are calculated. For comparison with the halo model expressions described in sections 2.1.3 and 2.2.4, we calculate overdensity maps by subtracting and dividing by the mean pixel mass for the dark matter maps and the mean pixel brightness temperature for the HI maps, respectively. The *anafast* function of `HEALPix` is then used to calculate the angular power spectrum. As recommended in the documentation,<sup>16</sup> the angular power spectrum is calculated up to a multipole moment of  $\ell_{\max} = 2 \times NSIDE = 1024$  for  $NSIDE = 512$ . For a clean comparison between the halo model and the maps, one has to correct the theoretical predictions for the partial sky coverage and the pixel window function.<sup>17</sup> For the former, it was verified that dividing the angular power spectrum expression from the halo model by the sky coverage is sufficient in the multipole range of interest, with no further deconvolution of the sky mask required. For the latter, it is essential to subtract the Poisson shot-noise of the simulation before applying the correction. The pixel window function represents the smoothing of the signal from individual sources across the pixel area. However, the Poisson noise arises due to the discrete nature of the halos and is therefore independent of pixelization. Applying the pixel window function to the Poisson noise would result in an overestimation of the signal measured from the maps. By adapting the expression from [77] for mass overdensity maps, the Poisson contribution to the angular power spectrum can be calculated as:

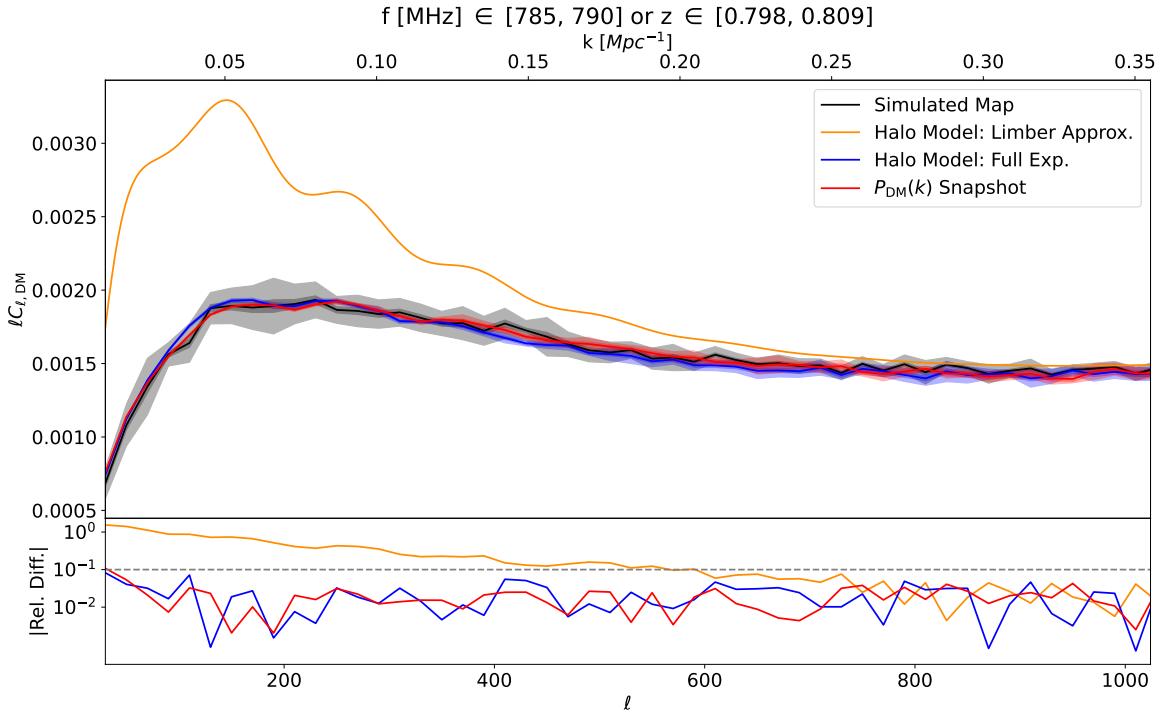
$$C_{\ell}^{SN} = \Omega_{\text{pix}} \frac{\langle m^2 \rangle}{\bar{N} \bar{m}^2}, \quad (4.1)$$

where  $\Omega_{\text{pix}}$  is the pixel area in steradians,  $\langle m^2 \rangle$  the average of the squared halo masses in the map,  $\bar{N}$  the average number of halos per pixel, and  $\bar{m}$  the mean halo mass in the map. After applying the pixel window function correction, the Poisson noise is reintroduced to the signal. Figures 9 and 10 show the angular power spectrum for dark matter and HI calculated from the maps (black). For better visualisation, the  $C_{\ell}$ 's are averaged over 20 neighbouring  $\ell$ 's. The dark grey band shows the error of the mean value and the light grey band the standard deviation calculated on the bins.

The  $C_{\ell}$ 's of the maps are compared to the halo model. As for the power spectra in the last section, the combinations of the Sheth & Tormen mass function and linear halo bias for the dark matter maps and the Tinker bias for HI have been used. In both cases, it is apparent that the Limber approximation (orange), i.e. equations (2.20) and (2.34), strongly overestimates the spectra due to the thinness of the maps for which the approximation diverges. The full expressions for the spectra (blue), equations (2.18) and (2.33), agree much better. As for the power spectra, we again find very good agreement below 10% over all shown scales. For the angular power spectra shown, we integrate over the  $k$ -range that is given in figures 7 and 8, i.e.

<sup>16</sup>[https://healpix.sourceforge.io/html/fac\\_anafast.htm](https://healpix.sourceforge.io/html/fac_anafast.htm).

<sup>17</sup>[https://healpix.sourceforge.io/html/intro\\_Pixel\\_window\\_functions.htm](https://healpix.sourceforge.io/html/intro_Pixel_window_functions.htm).



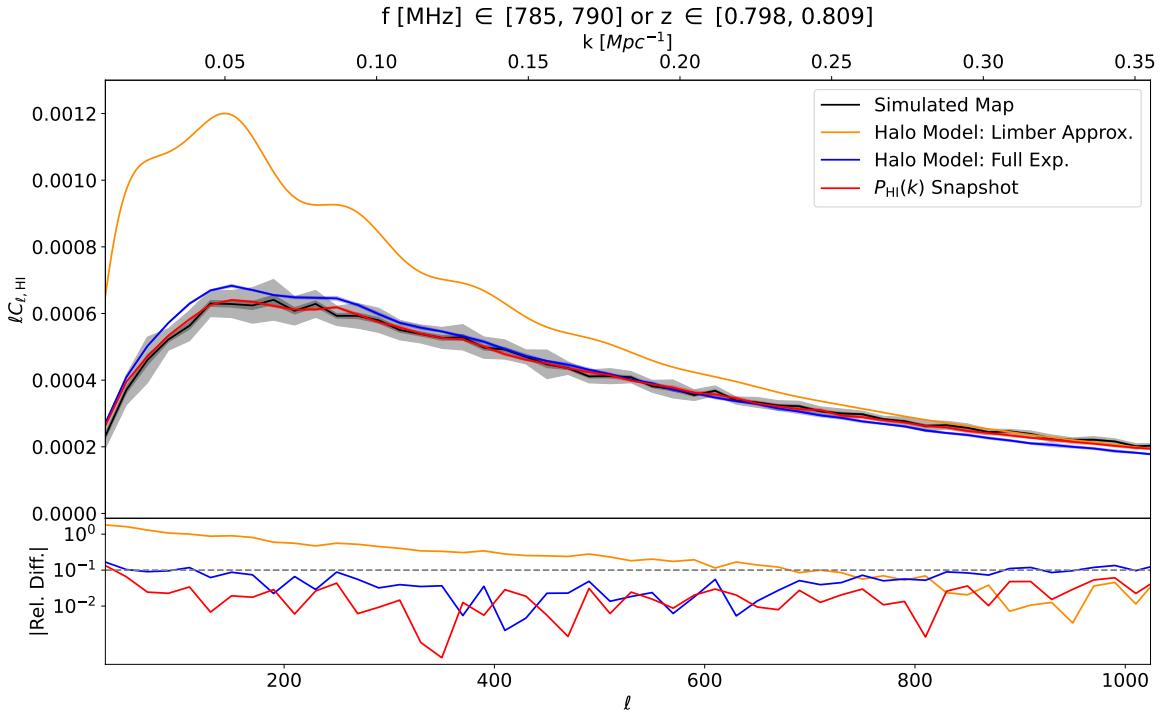
**Figure 9.** Angular Power Spectrum for the dark matter overdensity map for the frequency bin from 785 to 790 MHz (black). We show the mean  $C_{\ell}$ 's averaged over 20 adjacent  $\ell$ 's. The dark grey band indicates the error on the mean and the light grey band the standard deviation. The spectrum of the map is compared with the halo model predictions: The orange line shows the Limber approximation. The blue and red lines show the spectra resulting from the integration without the approximation for the  $P(k)$  based on the halo model or the snapshot, respectively (see equation (2.18)). The three-dimensional integration is performed with a Monte-Carlo integration. The blue and red shaded bands indicate the error of the integration. The bottom panel shows the absolute relative difference and the grey dashed line indicates 10 %.

$\sim 6 \times 10^{-3} - 3 \text{ Mpc}^{-1}$ . It has been further checked that this  $k$ -range is sufficiently large for the integration to converge. The shaded areas give the errors of the Monte-Carlo integrations.

We also calculate the  $C_{\ell}$ 's with equations (2.18) and (2.33) by using the measured  $P(k)$  of the snapshots as discussed in the last section instead of the halo model. To facilitate the comparison between the power spectrum and the angular power spectrum, we highlight  $k_{\text{max}}$  in figure 7 and 8, which corresponds to the maximum multipole moment shown in figure 9 and 10.<sup>18</sup> As already with the power spectra for dark matter, the angular power spectra match very well too. For HI, the agreement with the angular power spectra of the maps for  $\ell < 350$  and  $\ell > 700$  gets better when the snapshot power spectrum is used. Therefore, the differences seen in the  $P_{\text{HI}}(k)$  transfer to the angular power spectrum and explain the discrepancies that we observe there.

As discussed in section 3.1, scales comparable to or larger than the simulated box size are affected by box replications. The box size corresponds to  $\ell \approx 10$ , but as highlighted in

<sup>18</sup>The same relation between  $k$  and  $\ell$ , given in section 4.3, is used to give the corresponding  $k$ -values on the upper horizontal axis for the angular power spectra.



**Figure 10.** Same comparison as in figure 9 but for the HI angular power spectrum.

section 4.3, already scales  $k < 10^{-2} \text{ Mpc}^{-1}$  or  $\ell < 30$  are impacted by the limited box size. While a disagreement becomes apparent at these scales, the plots do not emphasize this region.

## 5 Conclusions

Post-reionization 21 cm intensity mapping is a promising, powerful and efficient cosmological probe. By inherently encoding the position along the line of sight within the redshifted signal from unresolved neutral hydrogen, this method eliminates the need for spectroscopic follow-up of individual galaxies to determine their redshifts. This holds the potential to cover large volumes in a short time, enabling highly accurate constraints on cosmological parameters. The increasing recognition of the importance of this probe is evident in the growing number of radio interferometric projects that have been realised or are being planned. Simulations of the 21 cm signal are invaluable for testing observation modeling and analysis pipelines, as well as for conducting forecast studies of future surveys and parameter constraints in the context of simulation-based inference.

In this paper, we developed a method to generate realistic post-reionization 21 cm emission maps. Utilizing the approximate but efficient PINOCCHIO code, we simulated the past light cone of dark matter halos and populated them with HI according to an HI-halo mass relation. This approach enabled the creation of large-volume simulations with high halo mass resolution, essential for 21 cm intensity mapping. We focused on simulating a past light cone specified for the HIRAX radio interferometer, i.e. covering declinations between  $-15^\circ$  and  $-35^\circ$  in the frequency range from 700 to 800 MHz, i.e. a redshift range of  $0.77 < z < 1.03$ . We ran PINOCCHIO for a box size of  $1 \text{ h}^{-1} \text{ Gpc}$  including  $6700^3$  simulation particles. It needs 40

box replications which are stacked together with harmonic boundary conditions to cover the entire light cone. We simulate dark matter halos down to masses of  $4.3 \times 10^9 M_\odot$ , achieving an HI loss of only 2 – 3 % within the simulated redshift range. Subsequently, we generated dark matter and HI signal maps by dividing the light cone into 20 equally spaced frequency channels with a width of 5 MHz to account for the radial resolution of the signal.

To validate the quality of our simulation and the generated maps, we implemented a dark matter halo model and an adapted HI halo model in `PyCosmo`. Based on the basic assumption that all dark matter resides in halos of any size and assuming a halo mass function, a linear halo bias, and a halo profile, the models provide predictions for the mass densities, the non-linear power spectra, and consequently the angular power spectra. The implementations are flexible enough to account for the simulated mass resolution, allowing direct comparison with the simulations. We provide four different halo mass functions [46–49] and four linear halo biases [47, 48, 52, 53] in `PyCosmo`. We offer NFW profiles for dark matter and exponential profiles for HI halos. However, the simulation treats halos as point sources without substructure. Therefore, we compared the models to the simulation using Dirac delta profiles. Currently, `PyCosmo` only includes the angular power spectrum assuming the Limber approximation. We found that this approximation does not apply very well to thin maps. Hence, we considered the full calculation without the approximation. But because it consists of an expensive three-dimensional integration including spherical Bessel functions, we perform it outside of `PyCosmo` with a Monte-Carlo integration method.

Our findings show that the models and the simulation agree on the mass function and the mass densities within a few percent. We also find good agreement below 10 % for the power spectrum and angular power spectrum of dark matter and HI. This agreement, consistent with the expectations of a simple halo model and the approximate `PINOCCHIO` code, gives us confidence in the quality and understanding of the simulated signal.

To avoid complicating the analysis, we have made the comparison between the simulation and theory in real space, considering the actual redshift of the halos. However, redshift measurements are always performed in redshift space, introducing redshift space distortions due to the peculiar velocities of the halos. Fortunately, `PINOCCHIO` provides the velocities and the observed redshifts in the halo catalogue, making it straightforward to include them from the simulation perspective. On the theory side, distortion terms must be included in the power spectrum, as for example described in [78]. Dedicated work on these implementations has already started and will continue henceforth.

In future work, we plan to apply the maps to the instrument simulation and analysis pipeline of the HIRAX instrument, which is based on the *m-mode* formalism described in [79, 80]. Our goal is to incorporate instrumental noise along with galactic and extra-galactic foregrounds that contaminate the cosmological signal. The latter poses probably the foremost challenge for 21 cm cosmology. We aim to test the pipeline’s ability to remove the foregrounds and retrieve the underlying 21 cm signal. Additionally, we will investigate the pipeline’s sensitivity to variations in cosmological parameters and astrophysics, such as alternative HI-halo mass relations, within our simulations.

Cross-correlating the 21 cm intensity mapping signal with other cosmological probes, as demonstrated in previous studies [81–86], allow for detection of the HI signal with

reduced systematics. Our approach facilitates future cross-correlation studies between 21 cm intensity mapping and other probes by providing consistent halo catalogues that can be used to model other observables. One such probe could be CMB lensing. In [87], they illustrate that HI intensity mapping suffers from long-wavelength line-of-sight mode loss due to galactic foreground subtraction, which significantly reduces the constraining power of the cross-correlation power spectrum with CMB lensing. Nonetheless, a proposed cross-bispectrum estimator, which cross-correlates the HI power spectrum with the CMB lensing field, enhances constraints on cosmological parameters. Forecasts combining hypothetical HIRAX HI intensity mapping measurements with CMB lensing measurements from ground based CMB experiments show that this estimator can yield competitive constraints on cosmological parameters, particularly on the dark energy equation of state parameters, compared to next-generation galaxy redshift surveys. Reinforcing these results with realistic simulations would confirm that, despite the significant challenge posed by foregrounds, 21 cm intensity mapping remains a promising cosmological probe.

`PyCosmo` 2.2.0 with the halo model extension is available at <https://cosmology.ethz.ch/research/software-lab/PyCosmo.html> and can be installed from the Python Package Index at <https://pypi.org>. The simulation dataset is available at <https://cosmology.ethz.ch/research/software-lab/cosmological-neutral-hydrogen-simulation.html>. The dataset includes `PINOCCHIO`'s dark matter halo catalogues of the snapshots at  $z = 0.8, 0.9, 1$  and the past light cone saved as text files as well as the dark matter mass and HI brightness temperature maps for the 20 frequency bins stored as HDF5 files.

## Acknowledgments

We thank Marta Spinelli, Sara Aliqolizadehsafari, Jennifer Studer, and Luis Machado Poletti Valle for helpful discussions and ongoing work to apply our simulations to instruments other than HIRAX and therefore help spreading awareness. We thank Hamsa Padmanabhan for insightful discussions on the halo model. We thank Uwe Schmitt for his continuous software support throughout the project, especially with respect to `PyCosmo`. We thank Pierluigi Monaco for his support in the use of `PINOCCHIO`. We thank Andrina Nicola for her great help in interpreting the angular power spectrum.

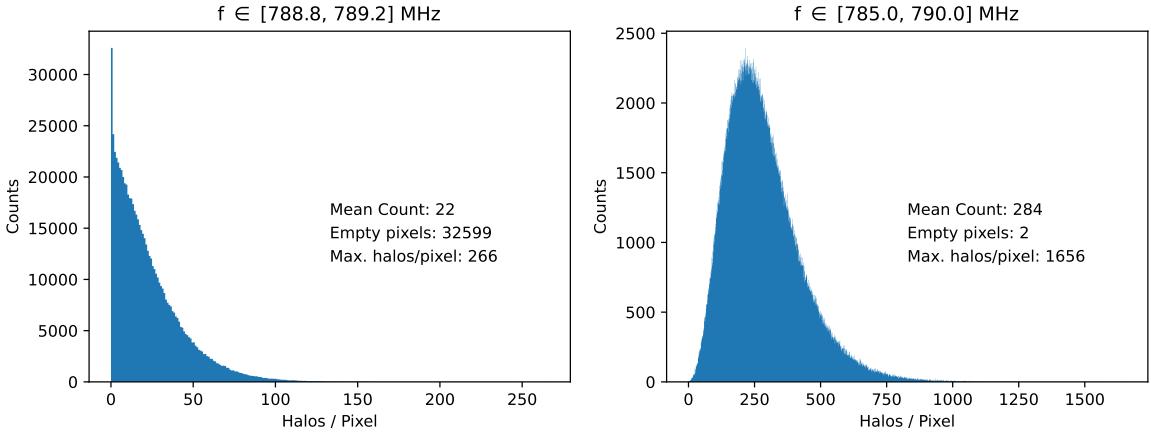
We acknowledge access to Piz Daint at the Swiss National Supercomputing Centre, Switzerland, under the SKA's share with the project IDs go06, sk05, and sk19.

This work was supported in part by grant 200021\_192243 from the Swiss National Science Foundation, and by SERI as part of the SKACH consortium.

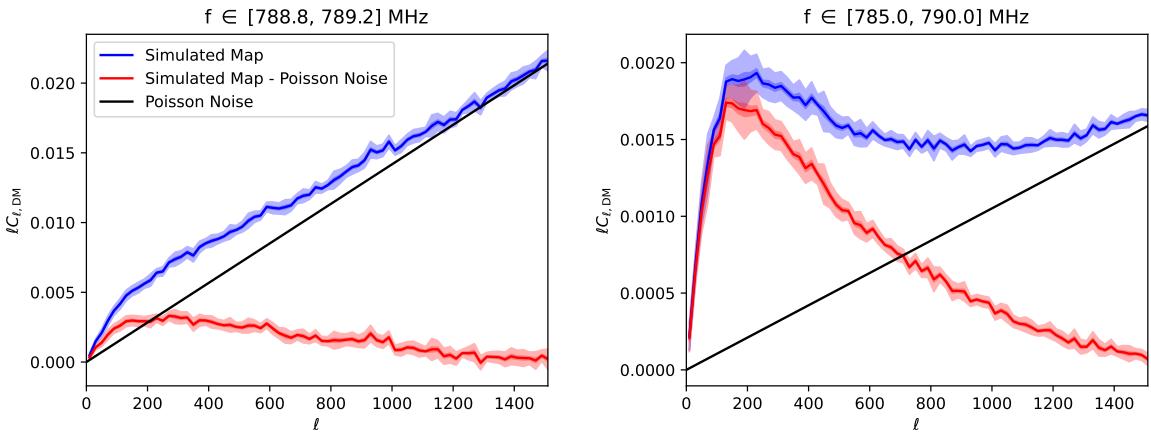
This work made use of the `numpy` [88], `scipy` [89], `matplotlib` [90], `h5py` [91], and `Pylians` [74] software packages. Some of the results in this paper have been derived using the `healpy` [71] and `HEALPix` [70] package.

## A Poisson noise contribution to the angular power spectrum

Figure 11 shows the halo counts per pixel for a bandwidth of 0.4 MHz (left) and 5 MHz (right). The former shows a more Poisson-like distribution with a much lower mean count and many more empty pixels compared to the latter, which shows a more lognormal distribution.



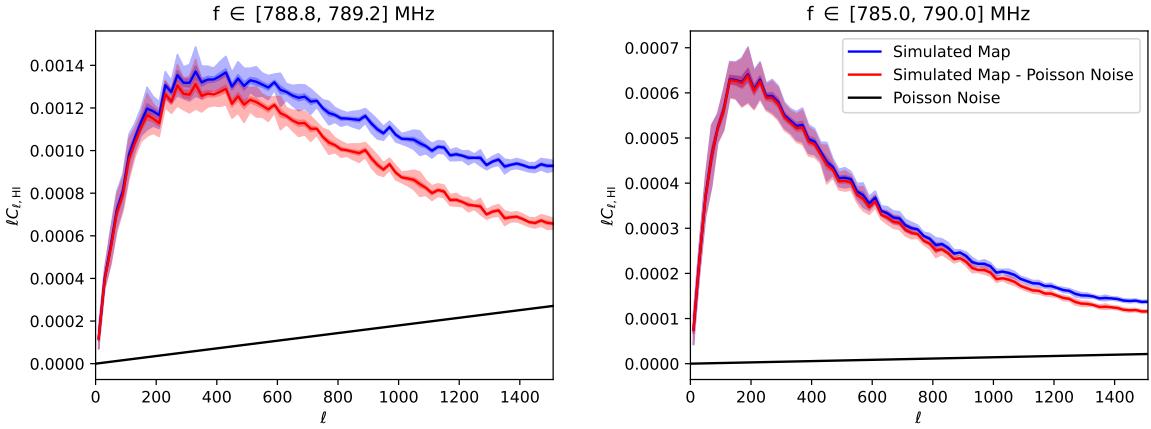
**Figure 11.** Number of halos per pixel for maps with bandwidths of 0.4 MHz (left) and 5 MHz (right).



**Figure 12.** Poisson noise contribution to the angular power spectrum for dark matter maps with bandwidths of 0.4 MHz (left) and 5 MHz (right). For the simulated maps, we show the mean  $C_{\ell}$  values averaged over 20 adjacent  $\ell$ 's. The dark blue/red band represents the error on the mean, while the light blue/red band indicates the standard deviation.

Figures 12 and 13 show the contributions of Poisson noise to the angular power spectra for bandwidths of 0.4 MHz and 5 MHz for dark matter and HI brightness temperature maps, respectively. For maps weighted by the dark matter halo masses, it can be clearly seen that for a bandwidth of 0.4 MHz the Poisson noise (black) dominates over the cosmological signal (red) for  $\ell > 200$ , while this happens only for  $\ell > 700$  for a bandwidth of 5 MHz.

For HI, the contribution of Poisson noise is also larger for thinner maps, but overall less significant than for dark matter. This is due to the different mass weighting of the halos: Since the logarithmic slope in equation (2.22) is  $\beta = -0.58$ ,  $M_{\text{HI}}(M, z) \propto M^{0.42}$ . This means that when weighting the halo mass function by the HI mass, the smaller and more abundant halos are weighted more than the larger and rarer halos. The Poisson noise is inversely proportional to the number of halos (see equation (4.1)) and hence small halos have a lower Poisson noise than large halos. This results in a lower Poisson noise for HI maps than dark matter maps.



**Figure 13.** Poisson noise contribution to the angular power spectrum for HI brightness temperature maps with bandwidths of 0.4 MHz (left) and 5 MHz (right). As in figure 12, the dark blue/red band represents the error on the mean  $C_\ell$  values averaged over 20 adjacent  $\ell$ 's and the light blue/red band the standard deviation.

## References

- [1] D.J. Eisenstein, W. Hu and M. Tegmark, *Cosmic Complementarity:  $H_0$  and  $\Omega_m$  from Combining CMB Experiments and Redshift Surveys*, *Astrophys. J. Lett.* **504** (1998) L57 [[astro-ph/9805239](#)] [[INSPIRE](#)].
- [2] H.-J. Seo and D.J. Eisenstein, *Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys*, *Astrophys. J.* **598** (2003) 720 [[astro-ph/0307460](#)] [[INSPIRE](#)].
- [3] DESI collaboration, *DESI 2024 III: Baryon Acoustic Oscillations from Galaxies and Quasars*, [arXiv:2404.03000](#) [[INSPIRE](#)].
- [4] DESI collaboration, *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest*, *JCAP* **01** (2025) 124 [[arXiv:2404.03001](#)] [[INSPIRE](#)].
- [5] DESI collaboration, *DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations*, *JCAP* **02** (2025) 021 [[arXiv:2404.03002](#)] [[INSPIRE](#)].
- [6] EBOSS collaboration, *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, *Phys. Rev. D* **103** (2021) 083533 [[arXiv:2007.08991](#)] [[INSPIRE](#)].
- [7] BOSS collaboration, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **470** (2017) 2617 [[arXiv:1607.03155](#)] [[INSPIRE](#)].
- [8] R.A. Battye, R.D. Davies and J. Weller, *Neutral hydrogen surveys for high redshift galaxy clusters and proto-clusters*, *Mon. Not. Roy. Astron. Soc.* **355** (2004) 1339 [[astro-ph/0401340](#)] [[INSPIRE](#)].
- [9] J.B. Peterson, K. Bandura and U.L. Pen, *The Hubble Sphere Hydrogen Survey*, in the proceedings of the *41st Rencontres de Moriond: Workshop on Cosmology: Contents and Structures of the Universe*, La Thuile, Italy, March 18–25 (2006) [[astro-ph/0606104](#)] [[INSPIRE](#)].
- [10] T.-C. Chang, U.-L. Pen, J.B. Peterson and P. McDonald, *Baryon Acoustic Oscillation Intensity Mapping of Dark Energy*, *Phys. Rev. Lett.* **100** (2008) 091303 [[arXiv:0709.3672](#)] [[INSPIRE](#)].

[11] S. Wyithe and A. Loeb, *Fluctuations in 21cm Emission After Reionization*, *Mon. Not. Roy. Astron. Soc.* **383** (2008) 606 [[arXiv:0708.3392](https://arxiv.org/abs/0708.3392)] [[INSPIRE](#)].

[12] A. Liu and J.R. Shaw, *Data Analysis for Precision 21cm Cosmology*, *Publ. Astron. Soc. Pac.* **132** (2020) 062001 [[arXiv:1907.08211](https://arxiv.org/abs/1907.08211)] [[INSPIRE](#)].

[13] M.G. Santos et al., *Cosmology from a SKA HI intensity mapping survey*, *PoS AASKA14* (2015) 019 [[arXiv:1501.03989](https://arxiv.org/abs/1501.03989)] [[INSPIRE](#)].

[14] SKA collaboration, *Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical specifications and performance forecasts*, *Publ. Astron. Soc. Austral.* **37** (2020) e007 [[arXiv:1811.02743](https://arxiv.org/abs/1811.02743)] [[INSPIRE](#)].

[15] MEERKCLASS collaboration, *A Large Sky Survey with MeerKAT*, in the proceedings of the *MeerKAT Science: On the Pathway to the SKA*, Stellenbosch, South Africa, May 25–27 (2016) [[DOI:10.22323/1.277.0032](https://doi.org/10.22323/1.277.0032)] [[arXiv:1709.06099](https://arxiv.org/abs/1709.06099)] [[INSPIRE](#)].

[16] BINGO collaboration, *The BINGO telescope: a new instrument exploring the new 21 cm cosmology window*, *J. Phys. Conf. Ser.* **1269** (2019) 012002 [[arXiv:1803.01644](https://arxiv.org/abs/1803.01644)] [[INSPIRE](#)].

[17] R. Nan et al., *The Five-Hundred-Meter Aperture Spherical Radio Telescope (FAST) Project*, *Int. J. Mod. Phys. D* **20** (2011) 989 [[arXiv:1105.3794](https://arxiv.org/abs/1105.3794)] [[INSPIRE](#)].

[18] R.M. Prestage et al., *The Green Bank Telescope*, *Proc. IEEE* **97** (2009) 1382.

[19] L. Staveley-Smith et al., *The Parkes 21 cm Multibeam Receiver*, *Publ. Astron. Soc. Austral.* **13** (1996) 243.

[20] K. Bandura et al., *Canadian Hydrogen Intensity Mapping Experiment (CHIME) Pathfinder*, *Proc. SPIE Int. Soc. Opt. Eng.* **9145** (2014) 22 [[arXiv:1406.2288](https://arxiv.org/abs/1406.2288)] [[INSPIRE](#)].

[21] K. Vanderlinde et al., *LRP 2020 Whitepaper: The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD)*, [arXiv:1911.01777](https://arxiv.org/abs/1911.01777) [[DOI:10.5281/zenodo.3765414](https://doi.org/10.5281/zenodo.3765414)] [[INSPIRE](#)].

[22] D.R. DeBoer et al., *Hydrogen Epoch of Reionization Array (HERA)*, *Publ. Astron. Soc. Pac.* **129** (2017) 045001 [[arXiv:1606.07473](https://arxiv.org/abs/1606.07473)] [[INSPIRE](#)].

[23] PUMA collaboration, *Packed Ultra-wideband Mapping Array (PUMA): A Radio Telescope for Cosmology and Transients*, *Bull. Am. Astron. Soc.* **51** (2019) 53 [[arXiv:1907.12559](https://arxiv.org/abs/1907.12559)] [[INSPIRE](#)].

[24] Y. Xu, X. Wang and X. Chen, *Forecasts on the Dark Energy and Primordial Non-Gaussianity Observations with the Tianlai Cylinder Array*, *Astrophys. J.* **798** (2015) 40 [[arXiv:1410.7794](https://arxiv.org/abs/1410.7794)] [[INSPIRE](#)].

[25] L.B. Newburgh et al., *HIRAX: A Probe of Dark Energy and Radio Transients*, *Proc. SPIE Int. Soc. Opt. Eng.* **9906** (2016) 99065X [[arXiv:1607.02059](https://arxiv.org/abs/1607.02059)] [[INSPIRE](#)].

[26] D. Crichton et al., *Hydrogen Intensity and Real-Time Analysis Experiment: 256-element array status and overview*, *J. Astron. Telesc. Instrum. Syst.* **8** (2022) 011019 [[arXiv:2109.13755](https://arxiv.org/abs/2109.13755)] [[INSPIRE](#)].

[27] P. Coles and B. Jones, *A lognormal model for the cosmological mass distribution*, *Mon. Not. Roy. Astron. Soc.* **248** (1991) 1 [[INSPIRE](#)].

[28] H.S. Xavier, F.B. Abdalla and B. Joachimi, *Improving lognormal models for cosmological fields*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 3693 [[arXiv:1602.08503](https://arxiv.org/abs/1602.08503)] [[INSPIRE](#)].

[29] J.R. Shaw, R. Nitsche, S. Foreman and A. Kefala, *radiocosmology/cora: 20.5.0 (v20.5.0)*, [DOI:10.5281/zenodo.5949195](https://doi.org/10.5281/zenodo.5949195).

- [30] C. Ramírez-Pérez, J. Sánchez, D. Alonso and A. Font-Ribera, *CoLoRe: fast cosmological realisations over large volumes with multiple tracers*, *JCAP* **05** (2022) 002 [[arXiv:2111.05069](https://arxiv.org/abs/2111.05069)] [[INSPIRE](#)].
- [31] P. Monaco, T. Theuns and G. Taffoni, *The PINOCCHIO algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field*, *Mon. Not. Roy. Astron. Soc.* **331** (2002) 587 [[astro-ph/0109323](https://arxiv.org/abs/astro-ph/0109323)] [[INSPIRE](#)].
- [32] P. Monaco et al., *Predicting the number, spatial distribution and merging history of dark matter haloes*, *Astrophys. J.* **564** (2002) 8 [[astro-ph/0109322](https://arxiv.org/abs/astro-ph/0109322)] [[INSPIRE](#)].
- [33] G. Taffoni, P. Monaco and T. Theuns, *PINOCCHIO and the hierarchical build-up of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **333** (2002) 623 [[astro-ph/0109324](https://arxiv.org/abs/astro-ph/0109324)] [[INSPIRE](#)].
- [34] P. Monaco et al., *An accurate tool for the fast generation of dark matter halo catalogues*, *Mon. Not. Roy. Astron. Soc.* **433** (2013) 2389 [[arXiv:1305.1505](https://arxiv.org/abs/1305.1505)] [[INSPIRE](#)].
- [35] E. Munari et al., *Improving fast generation of halo catalogs with higher-order Lagrangian perturbation theory*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 4658 [[arXiv:1605.04788](https://arxiv.org/abs/1605.04788)] [[INSPIRE](#)].
- [36] L.A. Rizzo et al., *Simulating cosmologies beyond  $\Lambda$ CDM with PINOCCHIO*, *JCAP* **01** (2017) 008 [[arXiv:1610.07624](https://arxiv.org/abs/1610.07624)] [[INSPIRE](#)].
- [37] M. Spinelli et al., *SKAO HI intensity mapping: blind foreground subtraction challenge*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 2048 [[arXiv:2107.10814](https://arxiv.org/abs/2107.10814)] [[INSPIRE](#)].
- [38] M. Spinelli et al., *The atomic hydrogen content of the post-reionization Universe*, *Mon. Not. Roy. Astron. Soc.* **493** (2020) 5434 [[arXiv:1909.02242](https://arxiv.org/abs/1909.02242)] [[INSPIRE](#)].
- [39] A. Refregier, L. Gamper, A. Amara and L. Heisenberg, *PyCosmo: An Integrated Cosmological Boltzmann Solver*, *Astron. Comput.* **25** (2018) 38 [[arXiv:1708.05177](https://arxiv.org/abs/1708.05177)] [[INSPIRE](#)].
- [40] F. Tarsitano et al., *Predicting cosmological observables with PyCosmo*, *Astron. Comput.* **36** (2021) 100484 [[arXiv:2005.00543](https://arxiv.org/abs/2005.00543)] [[INSPIRE](#)].
- [41] B. Moser et al., *Symbolic implementation of extensions of the PyCosmo Boltzmann solver*, *Astron. Comput.* **40** (2022) 100603 [[arXiv:2112.08395](https://arxiv.org/abs/2112.08395)] [[INSPIRE](#)].
- [42] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](https://arxiv.org/abs/1807.06209)] [[INSPIRE](#)].
- [43] J.A. Peacock and R.E. Smith, *Halo occupation numbers and galaxy bias*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 1144 [[astro-ph/0005010](https://arxiv.org/abs/astro-ph/0005010)] [[INSPIRE](#)].
- [44] U. Seljak, *Analytic model for galaxy and dark matter clustering*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 203 [[astro-ph/0001493](https://arxiv.org/abs/astro-ph/0001493)] [[INSPIRE](#)].
- [45] A. Cooray and R.K. Sheth, *Halo Models of Large Scale Structure*, *Phys. Rept.* **372** (2002) 1 [[astro-ph/0206508](https://arxiv.org/abs/astro-ph/0206508)] [[INSPIRE](#)].
- [46] W.H. Press and P. Schechter, *Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation*, *Astrophys. J.* **187** (1974) 425 [[INSPIRE](#)].
- [47] R.K. Sheth and G. Tormen, *Large scale bias and the peak background split*, *Mon. Not. Roy. Astron. Soc.* **308** (1999) 119 [[astro-ph/9901122](https://arxiv.org/abs/astro-ph/9901122)] [[INSPIRE](#)].
- [48] J.L. Tinker et al., *The Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests*, *Astrophys. J.* **724** (2010) 878 [[arXiv:1001.3162](https://arxiv.org/abs/1001.3162)] [[INSPIRE](#)].

[49] W.A. Watson et al., *The halo mass function through the cosmic ages*, *Mon. Not. Roy. Astron. Soc.* **433** (2013) 1230 [[arXiv:1212.0095](https://arxiv.org/abs/1212.0095)] [[INSPIRE](#)].

[50] T.T. Nakamura and Y. Suto, *Strong gravitational lensing and velocity function as tools to probe cosmological parameters: Current constraints and future predictions*, *Prog. Theor. Phys.* **97** (1997) 49 [[astro-ph/9612074](https://arxiv.org/abs/astro-ph/9612074)] [[INSPIRE](#)].

[51] H. Mo, F. van den Bosch and S. White, *Galaxy Formation and Evolution*, Cambridge University Press (2010) [[DOI:10.1017/cbo9780511807244](https://doi.org/10.1017/cbo9780511807244)].

[52] H.J. Mo and S.D.M. White, *An analytic model for the spatial clustering of dark matter halos*, *Mon. Not. Roy. Astron. Soc.* **282** (1996) 347 [[astro-ph/9512127](https://arxiv.org/abs/astro-ph/9512127)] [[INSPIRE](#)].

[53] R.K. Sheth, H.J. Mo and G. Tormen, *Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **323** (2001) 1 [[astro-ph/9907024](https://arxiv.org/abs/astro-ph/9907024)] [[INSPIRE](#)].

[54] J.F. Navarro, C.S. Frenk and S.D.M. White, *A universal density profile from hierarchical clustering*, *Astrophys. J.* **490** (1997) 493 [[astro-ph/9611107](https://arxiv.org/abs/astro-ph/9611107)] [[INSPIRE](#)].

[55] J.S. Bullock et al., *Profiles of dark haloes. Evolution, scatter, and environment*, *Mon. Not. Roy. Astron. Soc.* **321** (2001) 559 [[astro-ph/9908159](https://arxiv.org/abs/astro-ph/9908159)] [[INSPIRE](#)].

[56] G.L. Bryan and M.L. Norman, *Statistical properties of x-ray clusters: Analytic and numerical comparisons*, *Astrophys. J.* **495** (1998) 80 [[astro-ph/9710107](https://arxiv.org/abs/astro-ph/9710107)] [[INSPIRE](#)].

[57] D.J. Eisenstein and W. Hu, *Power spectra for cold dark matter and its variants*, *Astrophys. J.* **511** (1997) 5 [[astro-ph/9710252](https://arxiv.org/abs/astro-ph/9710252)] [[INSPIRE](#)].

[58] D.N. Limber, *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II*, *Astrophys. J.* **119** (1954) 655 [[INSPIRE](#)].

[59] S. Dodelson and F. Schmidt, *Modern Cosmology Second Edition*, Academic Press (2020).

[60] S. Dodelson, *Modern Cosmology*, Academic Press (2003).

[61] N. Kaiser, *Weak gravitational lensing of distant galaxies*, *Astrophys. J.* **388** (1992) 272 [[INSPIRE](#)].

[62] J.-E. Campagne, J. Neveu and S. Plaszczynski, *Angpow: a software for the fast computation of accurate tomographic power spectra*, *Astron. Astrophys.* **602** (2017) A72 [[arXiv:1701.03592](https://arxiv.org/abs/1701.03592)] [[INSPIRE](#)].

[63] M. Simonović et al., *Cosmological perturbation theory using the FFTLog: formalism and connection to QFT loop integrals*, *JCAP* **04** (2018) 030 [[arXiv:1708.08130](https://arxiv.org/abs/1708.08130)] [[INSPIRE](#)].

[64] H. Padmanabhan, A. Refregier and A. Amara, *A halo model for cosmological neutral hydrogen: abundances and clustering*, *Mon. Not. Roy. Astron. Soc.* **469** (2017) 2323 [[arXiv:1611.06235](https://arxiv.org/abs/1611.06235)] [[INSPIRE](#)].

[65] L.A. Barnes and M.G. Haehnelt, *The bias of DLAs at  $z \sim 2.3$ : evidence for very strong stellar feedback in shallow potential wells*, *Mon. Not. Roy. Astron. Soc.* **440** (2014) 2313 [[arXiv:1403.1873](https://arxiv.org/abs/1403.1873)] [[INSPIRE](#)].

[66] D. Obreschkow, H.-R. Klöckner, I. Heywood, F. Levrier and S. Rawlings, *A Virtual Sky with Extragalactic HI and CO Lines for the Square Kilometre Array and the Atacama Large Millimeter/Submillimeter Array*, *Astrophys. J.* **703** (2009) 1890 [[arXiv:0908.0983](https://arxiv.org/abs/0908.0983)].

[67] J. Wang et al., *An observational and theoretical view of the radial distribution of HI gas in galaxies*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 2159 [[arXiv:1401.8164](https://arxiv.org/abs/1401.8164)] [[INSPIRE](#)].

[68] J.R. Bond, S. Cole, G. Efstathiou and N. Kaiser, *Excursion set mass functions for hierarchical Gaussian fluctuations*, *Astrophys. J.* **379** (1991) 440 [[INSPIRE](#)].

[69] <https://www.csics.ch/computers/piz-daint>.

[70] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [[astro-ph/0409513](#)] [[INSPIRE](#)].

[71] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Softw.* **4** (2019) 1298 [[INSPIRE](#)].

[72] P. Bull, P.G. Ferreira, P. Patel and M.G. Santos, *Late-time cosmology with 21cm intensity mapping experiments*, *Astrophys. J.* **803** (2015) 21 [[arXiv:1405.1452](#)] [[INSPIRE](#)].

[73] T. L. Wilson, K. Rohlfs and Susanne Hüttemeister, *Tools of Radio Astronomy*, Springer-Verlag, Berlin Heidelberg Germany (2009), pg. 331.

[74] F. Villaescusa-Navarro, *Pylians: Python libraries for the analysis of numerical simulations*, [ascl.soft11008V](#).

[75] A. Mead et al., *An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 1958 [[arXiv:1505.07833](#)] [[INSPIRE](#)].

[76] A. Mead et al., *Accurate halo-model matter power spectra with dark energy, massive neutrinos and modified gravitational forces*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 1468 [[arXiv:1602.02154](#)] [[INSPIRE](#)].

[77] LSST collaboration, *Tomographic galaxy clustering with the Subaru Hyper Suprime-Cam first year public data release*, *JCAP* **03** (2020) 044 [[arXiv:1912.08209](#)] [[INSPIRE](#)].

[78] H. Padmanabhan, R. Maartens, O. Umeh and S. Camera, *The HI intensity mapping power spectrum: insights from recent measurements*, [arXiv:2305.09720](#) [[INSPIRE](#)].

[79] J.R. Shaw et al., *All-Sky Interferometry with Spherical Harmonic Transit Telescopes*, *Astrophys. J.* **781** (2014) 57 [[arXiv:1302.0327](#)] [[INSPIRE](#)].

[80] J.R. Shaw et al., *Coaxing cosmic 21 cm fluctuations from the polarized sky using m-mode analysis*, *Phys. Rev. D* **91** (2015) 083514 [[arXiv:1401.2095](#)] [[INSPIRE](#)].

[81] MEERKCLASS collaboration, *MeerKCLASS L-band deep-field intensity maps: entering the HI dominated regime*, *Mon. Not. Roy. Astron. Soc.* **537** (2025) 3632 [[arXiv:2407.21626](#)] [[INSPIRE](#)].

[82] CHIME collaboration, *Detection of Cosmological 21 cm Emission with the Canadian Hydrogen Intensity Mapping Experiment*, *Astrophys. J.* **947** (2023) 16 [[arXiv:2202.01242](#)] [[INSPIRE](#)].

[83] S. Cunningham et al., *HI intensity mapping with MeerKAT: power spectrum detection in cross-correlation with WiggleZ galaxies*, *Mon. Not. Roy. Astron. Soc.* **518** (2022) 6262 [[arXiv:2206.01579](#)] [[INSPIRE](#)].

[84] EBOSS collaboration, *HI constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3495 [[arXiv:2102.04946](#)] [[INSPIRE](#)].

[85] K.W. Masui et al., *Measurement of 21 cm brightness fluctuations at  $z \sim 0.8$  in cross-correlation*, *Astrophys. J. Lett.* **763** (2013) L20 [[arXiv:1208.0331](#)] [[INSPIRE](#)].

[86] T.-C. Chang, U.-L. Pen, K. Bandura and J.B. Peterson, *An intensity map of hydrogen 21-cm emission at redshift  $z \approx 0.8$* , *Nature* **466** (2010) 463 [[arXiv:1007.3709](#)] [[INSPIRE](#)].

- [87] K. Moodley, W. Naidoo, H. Prince and A. Penin, *A cross-bispectrum estimator for CMB-HI intensity mapping correlations*, [arXiv:2311.05904](https://arxiv.org/abs/2311.05904) [INSPIRE].
- [88] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](https://arxiv.org/abs/2006.10256)] [INSPIRE].
- [89] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261 [[arXiv:1907.10121](https://arxiv.org/abs/1907.10121)] [INSPIRE].
- [90] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [INSPIRE].
- [91] A. Collette, *Python and HDF5*, O'Reilly (2013).