Modeling HI at the field level

Andrej Obuljen[1,*] Marko Simonović[2] Aurel Schneider[1] and Robert Feldmann[1]

[1]*Institute for Computational Science, University of Zurich,*
Winterthurerstrasse 190, 8057 Zurich, Switzerland
[2]*Theoretical Physics Department, CERN, 1 Esplanade des Particules, Geneva 23, CH-1211, Switzerland*

We use an analytical forward model based on perturbation theory to predict the neutral hydrogen (HI) overdensity maps at low redshifts. We investigate its performance by comparing it directly at the field level to the simulated HI from the IllustrisTNG magneto-hydrodynamical simulation TNG300-1 ($L = 205\ h^{-1}$ Mpc), in both real and redshift space. We demonstrate that HI is a biased tracer of the underlying matter field and find that the cubic bias model describes the simulated HI power spectrum to within 1% up to $k = 0.4(0.3)\ h\,\mathrm{Mpc}^{-1}$ in real (redshift) space at redshifts $z = 0, 1$. Looking at counts in cells, we find an excellent agreement between the theory and simulations for cells as small as $5\ h^{-1}$ Mpc. These results are in line with expectations from perturbation theory, and they imply that a perturbative description of the HI field is sufficiently accurate given the characteristics of upcoming 21 cm intensity mapping surveys. Additionally, we study the statistical properties of the model error—the difference between the truth and the model. We show that on large scales this error is nearly Gaussian and that it has a flat power spectrum, with amplitude significantly lower than the standard noise inferred from the HI power spectrum. We explain the origin of this discrepancy, discuss its implications for the HI power spectrum Fisher matrix forecasts, and argue that it motivates the HI field-level cosmological inference. On small scales in redshift space, we use the difference between the model and the truth as a proxy for the Fingers-of-God effect. This allows us to estimate the nonlinear velocity dispersion of HI and show that it is smaller than for the typical spectroscopic galaxy samples at the same redshift. Finally, we provide a simple prescription based on the perturbative forward model which can be used to efficiently generate accurate HI mock data, in real and redshift space.

I. INTRODUCTION

The main goal of current galaxy surveys is to measure cosmological parameters using galaxy clustering as a probe of the underlying matter density field over large volumes and large redshift ranges. Spectroscopic galaxy redshift surveys such as DESI [1] and Euclid [2] will measure approximately 30 million galaxy positions, an order-of-magnitude better than the largest currently available spectroscopic sample of BOSS [3,4].

Another way to probe the distribution of matter is to use neutral hydrogen (HI), relying on the technique of 21 cm intensity mapping (IM) [5–9]. In this approach the aim is to collect an integrated HI signal from the hyperfine spin-flip transition of atomic hydrogen at 21 cm across different redshifts. A number of upcoming 21 cm surveys such as CHIME [10], HIRAX [11], SKA [12], Tianlai [13], and proposed PUMA [14] will use the 21 cm line to probe the large-scale clustering of HI across the wide range of redshifts in the post-reionization ($z < 6$) universe, hopefully

providing some of the tightest constraints on the ΛCDM cosmological model and its extensions.

One of the keys in achieving these ambitious goals is the simple, reliable, and accurate theoretical description of the nonlinear density field of HI in terms of the initial conditions. The clustering of HI is somewhat different with respect to the standard galaxy clustering due to the particular form of the HI-halo connection. On average, the total HI mass in halos increases with the halo mass as a power law with an exponential cutoff at small halo masses. Together with the shape of the halo mass function, this results in an HI signal originating mainly from small and intermediate halo masses $M_h \sim 10^{10-12}\ [h^{-1} M_\odot]$. This has two important consequences. First, as these halos are very abundant, the expected HI shot noise is low [15,16]. Second, the fraction of HI in central versus satellite galaxies significantly depends on the halo mass [16,17]. For this reason it is hard to estimate whether the Fingers-of-God (FoG) effects in HI are larger or smaller compared to the FoG in typical galaxy samples, even though the velocity dispersion of HI is smaller than that of matter [16]. Finally, it is worth noting that, in contrast to the discrete nature of

*andrej.obuljen@uzh.ch

galaxy catalogs, the observed HI is smoothed by the angular and/or frequency resolution and therefore represented by a continuous field.

However, despite these peculiarities, the HI is still expected to be a biased tracer of the underlying matter field. Since the nonlinear evolution of HI is expected to be local in space [18] and to obey the equivalence principle, the same bias expansion formalism developed for galaxies (see [19] for a review) is valid for the HI as well. This approach is inherently perturbative. Therefore, one can use analytical techniques to find a relatively simple relation between initial conditions and the final HI field, relying only on the basic features of gravitational dynamics and symmetries of the system. The major advantage of perturbation theory is its universality and robustness. It is valid for *any* (local) baryonic physics, up to only a handful of free nuisance parameters which capture all the intricacies of the galaxy formation process, as formalized in the effective field theory approach to galaxy clustering [20–26].

Nevertheless, it is useful to carefully test these predictions against reality, since the relevant scales that determine the range of validity of perturbation theory can depend on the type of tracer considered. The main goal of this paper is to provide such careful analysis for the case of HI at low redshifts and in the context of planned IM surveys. To that end, we use the perturbative forward modeling to predict the HI maps and compare it to the state-of-the-art IllustrisTNG magneto-hydrodynamical simulations which have been instrumental in studying the clustering of HI in the late universe [27–31]. Such field level comparison has a number of advantages over the standard fit to the power spectrum or other n-point functions. First, choosing the same initial conditions for theory and simulations, one does not pay the price of cosmic variance. Second, in order to match the two realizations, amplitudes and phases of all Fourier modes have to be correct. This is a much more stringent test than looking at a particular summary statistics which unavoidably involves averaging over Fourier modes. Finally, at the field level one can also study observables such as counts in cells, which are not simply related to the n-point functions. For all these reasons, the field level methods have been exploited to study the performance of perturbation theory for dark matter [32,33] as well as biased tracers, in real and redshift space [34,35]. In the context of HI, this approach has already been used in several works [17,36,37], mainly at higher redshifts ($z > 2$). While based on the same ideas as in these earlier papers, our implementation of the field level analysis is slightly different, and we will comment on these differences in what follows. Other approaches, based on machine learning techniques, have also been used to predict HI at the field level [38,39].

Another important advantage of the field level comparison is that it provides a clear path to study the difference between the theory and simulations. On large scales, where the unaccounted nonlinearities are very

small, this difference is a true stochastic noise of galaxy formation, which depends on the number density and the type of tracers being observed. On small scales, this difference is dominated by the nonlinear evolution and can be used as a proxy of the truly nonlinear effects that cannot be captured by perturbation theory. We will show how both of these regimes can be exploited in order to gain a deeper understanding of clustering properties of HI. Those include measuring the true amplitude of the shot noise and estimating the HI nonlinear velocity dispersion.

Finally, perturbation theory can provide a good starting point to describe the HI density field even in the nonlinear regime. This is achieved by projecting the HI map to a few perturbative templates, allowing the coefficients in this projection, the so-called transfer functions, to be free functions of scale. These transfer functions are smooth and can be fitted by low-degree polynomials. As long as this simple expansion provides a decent phenomenological description of the nonlinear field, the transfer functions can be used to generate realistic HI mock data. We exploit this fact and provide a publicly available code `Hi-Fi mocks`[1] that can be used by the community to efficiently generate maps of HI whose properties resemble the one from the IllustrisTNG simulations but with arbitrary volumes, different realizations of the initial conditions, or even slightly different cosmologies.

The outline of the paper is the following. In Sec. II we describe the theoretical model that we use to make HI maps, given some cosmology and a realization of the initial conditions. In Sec. III we describe the IllustrisTNG simulations that we use as a reference point of comparison. The results in real and redshift space for the best-fit realization, the power spectrum analysis, and counts in cells are presented in Sec. IV. In Sec. V we discuss in detail the HI noise properties, including the amplitude of the noise on large scales and nonlinear velocity dispersion. In Sec. VI we show how to create fast and accurate HI mock data based on perturbative templates, and in Sec. VII we conclude and lay out directions for future research.

II. THEORETICAL MODEL

In this section we briefly describe the perturbative model that we use to predict HI maps at a given redshift, for a given cosmology and initial conditions. We use the exact same formalism of [34] in real space and of [35] in redshift space, and we refer the reader to these papers for more details. Before we provide the main ingredients of this approach, let us point out that it is designed to reproduce the correct one-loop power spectrum and the tree-level bispectrum, including all relevant dark matter nonlinearities, effective field theory counterterms, bias parameters, and infrared resummation. Therefore, the one-loop power

---

[1]https://github.com/andrejobuljen/Hi-Fi_mocks.

spectrum of this perturbative forward model is identical to the one computed by the nonlinear extensions of the Boltzmann codes, such as CLASS-PT [40], PyBird [41], or velocileptors [42].

The main difference in making a perturbative prediction at the field level compared to the n-point functions is the treatment of large displacements. While by the equivalence principle the large displacements do not affect the smooth part of the correlation functions [43–46] and impact only the features such as BAO [22,47–49], their effect at the field level is much more dramatic. Inspired by the Lagrangian perturbation theory which treats the large displacements nonperturbatively, in this paper we use the shifted operators of [34,35] as the building blocks for describing the HI realizations. These shifted operators, which we denote with the tilde signs, have the following form (in redshift space):

$$\tilde{\mathcal{O}}(\boldsymbol{k}, \hat{\boldsymbol{n}}) \equiv \int d^3\boldsymbol{q} \, \mathcal{O}(\boldsymbol{q}) e^{-i\boldsymbol{k}\cdot(\boldsymbol{q}+\boldsymbol{\psi}_1(\boldsymbol{q})+f\hat{\boldsymbol{n}}(\boldsymbol{\psi}_1(\boldsymbol{q})\cdot\hat{\boldsymbol{n}}))}, \quad (1)$$

where q is the Lagrangian coordinate in the initial conditions, $\boldsymbol{\psi}_1$ is the Zel'dovich displacement field, $\hat{\boldsymbol{n}}$ is the line-of-sight direction, and f is the logarithmic growth function. These operators are the building blocks in calculating the Eulerian density field using Lagrangian perturbation theory [50–53]. Their statistics can be calculated straightforwardly [42], and the results can be shown to be equivalent to the IR-resummed Eulerian counterparts [34,42]. Note that, for simplicity, we suppress the explicit time dependence, but all quantities in this expression have to be evaluated as functions of time. For instance, in order to obtain the Zel'dovich displacement field, we use the initial density field δ_1 rescaled linearly to a given output redshift ($z = 0, 1$ in our case):

$$\boldsymbol{\psi}_1(\boldsymbol{k}, z) = \frac{i\boldsymbol{k}}{k^2} \delta_1(\boldsymbol{k}, z). \quad (2)$$

Clearly, in order to obtain the shifted operators in real space, one can set $f = 0$ in Eq. (1) and drop the dependence on the line of sight.

Let us now turn to the question of which operators we use in our forward model. We focus on the leading orders in perturbation theory, needed to ensure that the one-loop power spectrum and the tree-level bispectrum are predicted correctly. This implies that we have to keep all relevant terms up to cubic order in small density fluctuations. However, there is a simple trick which allows us to absorb the contribution of the cubic operators into the scale-dependent coefficients of the shifted linear field. To see this, let us decompose a generic cubic operator in the following way:

$$\tilde{\mathcal{O}}_3 = \frac{\langle\tilde{\mathcal{O}}_3\tilde{\delta}_1\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle}\tilde{\delta}_1 + \left(\tilde{\mathcal{O}}_3 - \frac{\langle\tilde{\mathcal{O}}_3\tilde{\delta}_1\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle}\tilde{\delta}_1\right). \quad (3)$$

We can define the second term in the brackets to be orthogonal to $\tilde{\delta}_1$,

$$\tilde{\mathcal{O}}_3^{\perp} \equiv \tilde{\mathcal{O}}_3 - \frac{\langle\tilde{\mathcal{O}}_3\tilde{\delta}_1\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle}\tilde{\delta}_1, \quad (4)$$

since, by definition, $\langle\tilde{\mathcal{O}}_3^{\perp}\tilde{\delta}_1\rangle = 0$. This means that these orthogonal contributions do not contribute to the one-loop power spectrum and can be neglected at this order in perturbation theory. The only impact of cubic operators is through the scale-dependent coefficient multiplying $\tilde{\delta}_1$, which we refer to as the transfer functions

$$\beta_3(\boldsymbol{k}, \hat{\boldsymbol{n}}) \equiv \frac{\langle\tilde{\mathcal{O}}_3\tilde{\delta}_1\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle}. \quad (5)$$

These functions can be computed analytically, without the need to perform the forward modeling for cubic operators on the grid. In conclusion, provided that the proper transfer functions are used, the only operators that have to be generated at the field level are $\mathcal{O} = \{1, \delta_1, \delta_2 \equiv (\delta_1^2 - \sigma_1^2), \mathcal{G}_2\}$, where $\sigma_1^2 \equiv \langle\delta_1^2\rangle$ is the rms fluctuation of the linear density field, while

$$\mathcal{G}_2 \equiv \left[\frac{\partial_i\partial_j}{\nabla^2}\delta_1\right]^2 - \delta_1^2 \quad (6)$$

is the second order bias operator related to the tidal field. In some analyses we will also explicitly include $\mathcal{O} = \delta_1^3$. While this operator is not needed for the model for the reasons we have just explained, it is the only cubic term relevant for studying the properties of the noise, given that it has a flat autospectrum on large scales.

In order to simplify calculations in practice, it is convenient to use the basis of operators that are all orthogonal to each other, such that $\langle\tilde{\mathcal{O}}_a^{\perp}\tilde{\mathcal{O}}_b^{\perp}\rangle = 0$, for any a and b. This is achieved by a simple linear transformation, which does not change the statistical properties of the map. We will use such an orthogonal basis in all our analyses.

A. Real-space model

Let us focus on real space first. The full HI field can be written as a sum of two contributions,

$$\delta_{HI}(\boldsymbol{k}) = \delta_{HI}^{model}(\boldsymbol{k}) + \epsilon_{HI}(\boldsymbol{k}), \quad (7)$$

where δ_{HI}^{model} is the theoretical model and ϵ_{HI} is the model error which we assume to be uncorrelated with δ_{HI}^{model}. For example, on large scales, we expect ϵ_{HI} to be dominated by the stochastic noise due to discreteness of galaxies as a tracer. We follow Ref. [34] and use the following model for the HI field in terms of shifted and orthogonalized operators $\tilde{\mathcal{O}}^{\perp}(\boldsymbol{k})$:

$$\delta_{\text{HI}}^{\text{model}}(\boldsymbol{k}) = \beta_1(k)\tilde{\delta}_1(\boldsymbol{k}) + \beta_2(k)\tilde{\delta}_2^{\perp}(\boldsymbol{k})$$
$$+ \beta_{\mathcal{G}_2}(k)\tilde{\mathcal{G}}_2^{\perp}(\boldsymbol{k}) + \beta_3(k)\tilde{\delta}_3^{\perp}(\boldsymbol{k}), \qquad (8)$$

where $\beta_i(k)$ are transfer functions. At next-to-leading order in perturbation theory these functions have the following form [34] (see also Appendix A):

$$\beta_1(k) = b_1 + c_s^2 k^2 + b_2 \frac{\langle \tilde{\delta}_1 \tilde{\delta}_2 \rangle}{\langle \tilde{\delta}_1 \tilde{\delta}_1 \rangle} + b_{\mathcal{G}_2} \frac{\langle \tilde{\delta}_1 \tilde{\delta}_2 \rangle}{\langle \tilde{\delta}_1 \tilde{\delta}_1 \rangle}$$
$$+ b_{\Gamma_3} \frac{\langle \tilde{\delta}_1 \tilde{\Gamma}_3 \rangle}{\langle \tilde{\delta}_1 \tilde{\delta}_1 \rangle} - b_1 \frac{\langle \tilde{\delta}_1 \tilde{\mathcal{S}}_3 \rangle}{\langle \tilde{\delta}_1 \tilde{\delta}_1 \rangle}, \qquad (9)$$

$$\beta_2(k) = b_2, \qquad \beta_{\mathcal{G}_2}(k) = b_{\mathcal{G}_2}, \qquad \beta_3(k) = b_3, \quad (10)$$

where $b_1, c_s^2, b_2, b_{\mathcal{G}_2}, b_{\Gamma_3}, b_3$ are constant free nuisance parameters. Note that the parameter c_s^2 is the sum of the dark matter one-loop counterterm and the so-called nonlocal bias. The shape of cubic operators Γ_3 and $\mathcal{S}_3$ can be found in Appendix A. Let us stress that $\mathcal{S}_3$ comes from the second order displacement acting on the halo density field in Lagrangian coordinates, and therefore, by the equivalence principle, it is not multiplied by a new free parameter.

We would like to emphasize once again that the exact form of the transfer function in Eq. (9) is needed in order to have the correct one-loop power spectrum. Sometimes in the literature the loop corrections are neglected, and $\beta_1(k)$ is approximated as a constant [17]. While this may be a decent approximation at high redshifts ($z > 2$), the loop corrections to $\beta_1(k)$ in the late universe can be significant. Furthermore, the $\tilde{\mathcal{S}}_3$ operator (usually neglected in the literature) is very important in order to get reliable estimates of bias parameters, as an important additional cross-check that our theoretical description makes sense. We will come back to this point in Sec. IV when we fit the transfer functions in real space using perturbative templates.

Finally, let us comment on evaluating β_1 analytically. The IllustrisTNG simulation that we use as a reference point for comparison has a box size of $L = 205 \ h^{-1}$ Mpc. Given the periodic boundary conditions, this implies that the linear power spectrum has an infrared cutoff at $k_{\text{IR}} = 2\pi/L = 0.031 \ h \, \text{Mpc}^{-1}$. In all theoretical calculations, the same cutoff has to be implemented. Note that this implies that the typical displacements are significantly smaller than in the real Universe. Cutting the power spectrum below k_{IR}, we find that the velocity dispersion at redshift zero is $\sigma_v = 4.9 \ h^{-1}$ Mpc, roughly 20% lower than σ_v in ΛCDM and in excellent agreement with measurements from Illustris. Keeping this in mind, all correlators in Eq. (9) can still be evaluated following the standard methods for calculating the one-loop power spectrum in Lagrangian perturbation theory [42], using the modified linear power spectrum with the appropriate infrared cutoff.

B. Redshift-space model

Let us next turn to redshift space. Following [35], we can write the perturbative model for the HI field as

$$\delta_{\text{HI}}^{s,\text{model}}(\boldsymbol{k}, \hat{\boldsymbol{n}}) = \delta_Z^s(\boldsymbol{k}, \hat{\boldsymbol{n}}) - \frac{3}{7} f \tilde{\mathcal{G}}_2^{\parallel}(\boldsymbol{k}, \hat{\boldsymbol{n}}) + \beta_1(k, \mu) \tilde{\delta}_1(\boldsymbol{k}, \hat{\boldsymbol{n}})$$
$$+ \beta_2(k, \mu) \tilde{\delta}_2^{\perp}(\boldsymbol{k}, \hat{\boldsymbol{n}}) + \beta_{\mathcal{G}_2}(k, \mu) \tilde{\mathcal{G}}_2^{\perp}(\boldsymbol{k}, \hat{\boldsymbol{n}})$$
$$+ \beta_3(k, \mu) \tilde{\delta}_3^{\perp}(\boldsymbol{k}, \hat{\boldsymbol{n}}), \qquad (11)$$

where $\beta_i(k, \mu)$ are transfer functions which now depend on both k and $\mu \equiv \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}$. Here we define

$$\mathcal{G}_2^{\parallel}(\boldsymbol{q}) \equiv \hat{n}_i \hat{n}_j \frac{\partial_i \partial_j}{\nabla^2} \mathcal{G}_2(\boldsymbol{q}), \qquad (12)$$

which can be used to compute the shifted operator $\tilde{\mathcal{G}}_2^{\parallel}(\boldsymbol{k}, \hat{\boldsymbol{n}})$ using Eq. (1). Note that in redshift space the model error depends on both $\boldsymbol{k}$ and $\hat{\boldsymbol{n}}$ such that the full HI field can be written as

$$\delta_{\text{HI}}^s(\boldsymbol{k}, \hat{\boldsymbol{n}}) = \delta_{\text{HI}}^{s,\text{model}}(\boldsymbol{k}, \hat{\boldsymbol{n}}) + \epsilon_{\text{HI}}^s(\boldsymbol{k}, \hat{\boldsymbol{n}}). \qquad (13)$$

This model requires some explanations. While having a similar structure as in real space, Eq. (11) has new ingredients. In particular, the first line contains the Zel'dovich density in redshift space and the shifted $\mathcal{G}_2^{\parallel}$ field. Note that both of these terms come without free coefficients. While these two terms can, in principle, be absorbed in other shifted operators (the same way as in real space δ_Z is absorbed in $\tilde{\delta}_1$, $\tilde{\mathcal{G}}_2$, and $\tilde{\mathcal{G}}_3$; see [34]), we decide to keep them explicitly. In this way it is guaranteed that β_1, β_2, $\beta_{\mathcal{G}_2}$, and β_3 have the same limit when $k \to 0$ *independently* of μ. This is important because it simplifies the analysis of the transfer functions in the low-k limit and allows all μ bins to be combined. Such a combination increases the signal-to-noise ratio for the amplitude of each β in redshift space, which in turn is important for comparison to real-space results. This is another important difference in our work compared to previous literature, where the transfer functions are treated as constants independent of μ even without keeping $\tilde{\mathcal{G}}_2^{\parallel}$ in the model explicitly.

One important consequence of explicitly keeping δ_Z^s and $\tilde{\mathcal{G}}_2^{\parallel}$ in the model is that the low-k limits of the transfer functions in real and redshift space are not necessarily the same. One can show that when $k \to 0$, the theoretical expectation is

$$\beta_1^{\text{real}} = \beta_1^{\text{rsd}} + 1, \qquad (14)$$

$$\beta_2^{\text{real}} = \beta_2^{\text{rsd}}, \qquad (15)$$

$$\beta_{\mathcal{G}_2}^{\text{real}} = \beta_{\mathcal{G}_2}^{\text{rsd}} + \frac{2}{7}, \quad (16)$$

$$\beta_3^{\text{real}} = \beta_3^{\text{rsd}}. \quad (17)$$

We will test these relations once we fit the transfer functions in real and redshift space (see Sec. VI for comparison of this theoretical prediction and the best-fit values of transfer functions from simulations).

In principle, one can compute the one-loop corrections for $\beta_1(k, \mu)$ in a similar way as in Eq. (9), and we give the equations in Appendix B. However, with the precision available from a single box of Illustris simulation, testing the redshift-space distortions (RSD) model by fitting this transfer function would not be very informative. For this reason we will provide only a phenomenological fit to $\beta_1(k, \mu)$ in this paper and leave a more detailed analysis for when larger simulation volumes are available. All other transfer functions are expected to be constant on large scales, at leading order in perturbation theory.

C. Numerical implementation

We first generate a Gaussian linear density field using the same initial conditions (IC) as were used for the full hydrodynamical simulation (IllustrisTNG). We use a regular 256^3 grid at the initial redshift $z_{\text{ini}} = 127$, using the linear power spectrum and random seed of TNG. We use the same ΛCDM cosmology as TNG: $\Omega_\Lambda = 0.6911$, $\Omega_M = 0.3089$, $\Omega_b = 0.0486$, $\sigma_8 = 0.8159$, $n_s = 0.9667$, and $h = 0.6774$, in agreement with the results from Planck [54].

We generate a catalog of particles positioned on a regular 256^3 grid within the simulation box ($L = 205\ h\,\text{Mpc}^{-1}$), which results in a particle separation $\Delta q = 0.8\ h\,\text{Mpc}^{-1}$. At each particle's Lagrangian position q we compute the following: the displacement field $\psi_1(q)$, $\delta_1(q)$, $\delta_1^2(q)$, and $\mathcal{G}_2(q)$. We then displace the particles by the displacement field $\psi_1(q)$ and place the displaced particles on the same regular grid using the cloud-in-cell (CIC) mass assigning scheme [55]. To obtain the shifted overdensity fields $\tilde{\delta}_1$, $\tilde{\delta}_2$, $\tilde{\mathcal{G}}_2$, and $\tilde{\delta}_3$, we additionally weight displaced particles by the appropriate weights $\delta_1(q)$, $\delta_1^2(q)$, $\mathcal{G}_2(q)$, and $\delta_1^3(q)$, respectively. We apply the same smoothing to δ_1 when computing the shifted cubic term $\tilde{\delta}_3$ [34].

In redshift space, we follow a similar procedure, except that the displaced particles are additionally shifted along the line of sight. We choose the simulation z axis to be the line of sight ($\hat{z}$), and for this choice the total displacement along the line of sight is then $(1 + f)\psi_{zz}$, where f is the logarithmic growth rate [35]. Furthermore, we compute the $\mathcal{G}_2^{\parallel}$ term in Fourier space in the initial conditions as $\mathcal{G}_2^{\parallel}(k, \hat{n}) = \mu^2 \mathcal{G}_2(k)$, before shifting it the usual way to obtain $\tilde{\mathcal{G}}_2^{\parallel}$.

III. SIMULATED HI FIELD

In this section we describe the simulations used to obtain the evolved HI field. We use the IllustrisTNG simulations [27–31]. We focus on the largest simulation box with the highest resolution, TNG300-1, with a box size $L = 205\ h\,\text{Mpc}^{-1}$. We note that the main results in Ref. [16] were obtained from a smaller simulation, TNG100-1 ($L = 75\ h\,\text{Mpc}^{-1}$), with higher mass resolution, which allowed one to probe the HI distribution within smaller halos. However, for our work we need larger scales for two reasons. One reason is that the Zel'dovich displacement field receives a significant contribution from modes comparable to or larger than the size of the smaller box; thus, a larger box is needed to properly account for the large modes. Another reason is that we want to measure the performance of the model on large, BAO scales $\approx 100\ h^{-1}\,\text{Mpc}$ relevant for future 21 cm IM surveys, which again requires a larger box. Furthermore, we shall focus on outputs at $z = 0$ and $z = 1$ as these are the key redshifts for several future surveys.

We compute the HI masses in postprocessing following the approach from Ref. [16]. We then assign the particles weighted by their HI masses on a regular 256^3 grid using the CIC mass assignment scheme [55]. We correct for the CIC window to obtain the final HI overdensity field in real space. Similarly, in order to obtain the HI overdensity field in redshift space, we apply the RSD shift along the line of sight prior to mesh assignment. We caution the reader that the HI simulation results are not fully converged against resolution (see Appendix A in [16]).

In order to compare to the standard HI shot noise estimates, we make use of halo catalogs with total HI masses obtained from the same simulation TNG300-1. The halos were identified using the Friends-of-Friends (FOF) finder, and the total HI mass per halo is computed by summing the HI masses of all the particles within each FOF halo [16].

IV. RESULTS

In this section we test how well the perturbative forward model matches the simulations, and we present some results. Our main analysis is performed at the field level, where we fit the entire perturbative realization to simulations without restricting ourselves to any summary statistics. The key object in this approach is the mean-square error/residual power spectrum, in general given by

$$P_{\text{err}}(k, \hat{n}) \equiv \langle |\delta_{\text{HI}}^{\text{truth}}(k, \hat{n}) - \delta_{\text{HI}}^{\text{model}}(k, \hat{n})|^2 \rangle. \quad (18)$$

This definition implies that the error contains everything that is not included in the model; therefore, it does not correlate with $\delta_{\text{HI}}^{\text{model}}$. Consequently,

$$P_{\text{HI}}^{\text{truth}}(k, \hat{n}) = P_{\text{HI}}^{\text{model}}(k, \hat{n}) + P_{\text{err}}(k, \hat{n}). \quad (19)$$

In order to obtain the best-fit model, we minimize P_{err} in each (k, μ) bin. Since the model is linear in transfer functions, this minimization is equivalent to a linear regression in every bin. For the orthogonal basis of operators, the estimate of transfer functions is straightforward,

$$\beta_i(k, \mu) = \langle \mathcal{O}_i^\perp(\mathbf{k}, \hat{\mathbf{n}}) \delta_{\text{HI}}^{\text{truth}}(\mathbf{k}, \hat{\mathbf{n}})^* \rangle / \langle |\mathcal{O}_i^\perp(\mathbf{k}, \hat{\mathbf{n}})|^2 \rangle. \quad (20)$$

Note that in real space the P_{err} and the transfer functions β_i are functions of the wave-number amplitude k only.

Before diving into the details, let us comment on general expectations for the behavior of the transfer functions. On large scales, where the model is known to be correct, transfer functions are simply numbers, and they do not depend on scale. On the other hand, deeply in the nonlinear regime where the model is completely wrong, transfer functions approach zero. This can be seen directly from Eq. (20), if one assumes that the truth does not correlate with the model on very small scales. In the intermediate regime, the transfer functions can be modeled by perturbation theory, and we will test these predictions in this section.

Let us also comment on the expected size of the error P_{err}. Even if the model is perfect, we still expect the nonzero error for realistic biased tracers, which comes from the sampling noise due to the finite number of objects in simulations or galaxy catalogs. Therefore, on large scales, we expect the amplitude of the noise to have a flat power spectrum, with the amplitude approximately equal to $1/\bar{n}$, where $\bar{n}$ is the mean number density of galaxies. On the other hand, approaching the nonlinear regime, the error power spectrum can become dominated by the true non-linearities not captured by the model. These include higher order terms in perturbation theory as well as nonperturbative contributions such as 1-halo terms. We will exploit this fact in redshift space to estimate the size of the nonlinear velocity dispersion for HI.

A. Results in real space

Let us begin presenting our results starting from real space. Minimizing the error power spectrum as explained above, we measure the transfer functions in all k bins. In Fig. 1 we show a 2D slice of the simulated HI overdensity field in comparison to the best-fit cubic model and the residual at redshifts $z = 0$ and $z = 1$. We can see a general agreement between the simulation and best-fit model. Notable differences in the residuals are present at nonlinear scales and match the positions of large overdensities. These
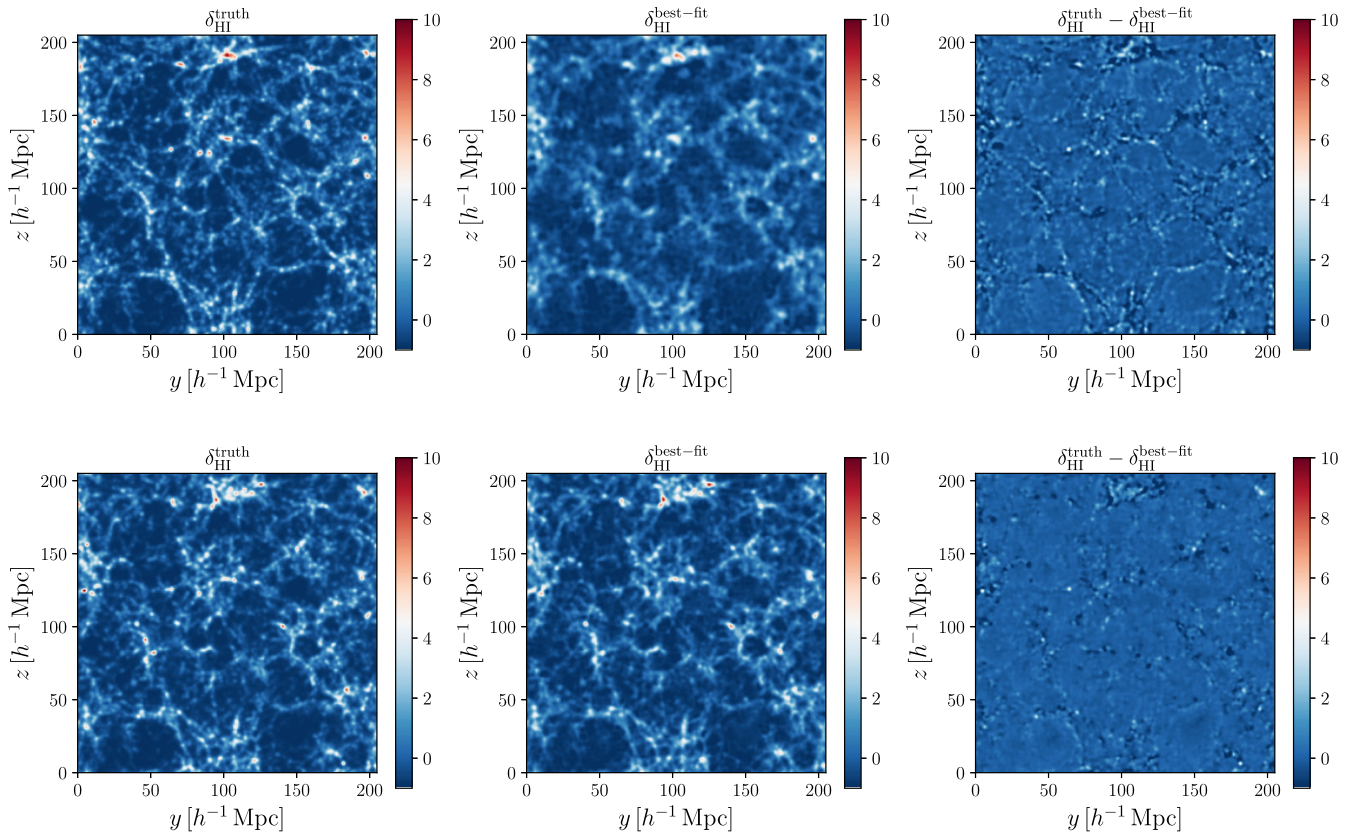

FIG. 1. Real-space slices in the y-z plane of the simulated HI overdensity field (left), the best-fit cubic bias model (middle), and the residuals (right), at $z = 0$ (top) and $z = 1$ (bottom). All density fields are smoothed with a $R = 1\ h^{-1}$ Mpc 3D Gaussian filter, while the depth of each slice is $20\ h^{-1}$ Mpc.
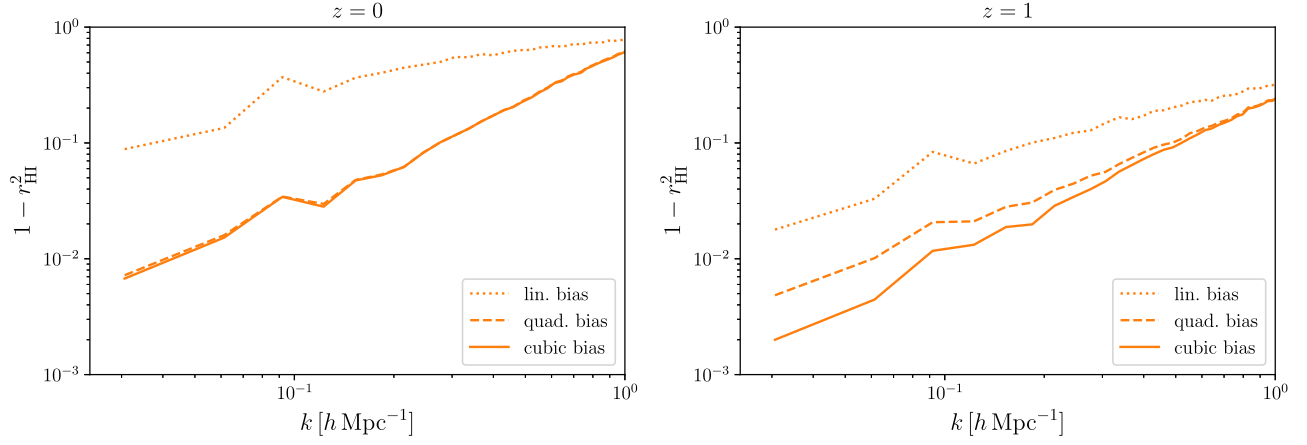
FIG. 2. Relative error of the best-fit model $P_{\rm err}$ compared to the true HI power spectrum $P_{\rm truth}$. This coincides with $1 - r_{\rm HI}^2$, where $r_{\rm HI}$ is the cross-correlation coefficient between simulations and the model. We show this quantity for different bias models: linear, quadratic, and cubic, at $z = 0$ (left) and $z = 1$ (right). We find that using quadratic or cubic models significantly decreases the relative error compared to the linear bias model.

differences are slightly more pronounced at later times, as expected from the nonlinear evolution. We note that a part of the difference is due to the accumulated projected error, given the thickness of the slice, which is $20\ h\,{\rm Mpc}^{-1}$.

A more quantitative way to compare the two maps is to calculate the cross-correlation coefficient between the simulated HI field and the best-fit model:

$$r_{\rm HI} \equiv \frac{\langle \delta_{\rm HI}^{\rm truth} \delta_{\rm HI}^{\rm best\text{-}fit*} \rangle}{(\langle |\delta_{\rm HI}^{\rm truth}|^2 \rangle \langle |\delta_{\rm HI}^{\rm best\text{-}fit}|^2 \rangle)^{1/2}}. \tag{21}$$

We show the quantity $1 - r_{\rm HI}^2$ in Fig. 2, which can be equal to $P_{\rm err}/P_{\rm truth}$ using Eq. (19) (see Appendix B in Ref. [34]). It is clear that for both $z = 0$ and $z = 1$ the quadratic and cubic bias models have much better cross-correlation coefficients than the linear bias model, even on the largest

scales where the linear theory is supposed to work well. We will come back to this peculiar feature of HI clustering and discuss it in detail in Sec. V. It is important to note that the major source of disagreement between the simulations and the model on large scales is the sampling noise and not the failure of perturbation theory to predict the correct realization of the maps.

In order to see this more explicitly, we measure the best-fit power spectrum as well as the model error power spectrum. The results are shown in Fig. 3. As previously mentioned, the model error power spectrum is expected to be flat if all the relevant higher order terms are properly accounted for. Indeed, we find that using either the quadratic or the cubic model results in almost scale-independent $P_{\rm err}$. On large scales we also find an excellent agreement between the power spectra in simulations and for the best-fit cubic bias model. Note that the sum of the model and error power
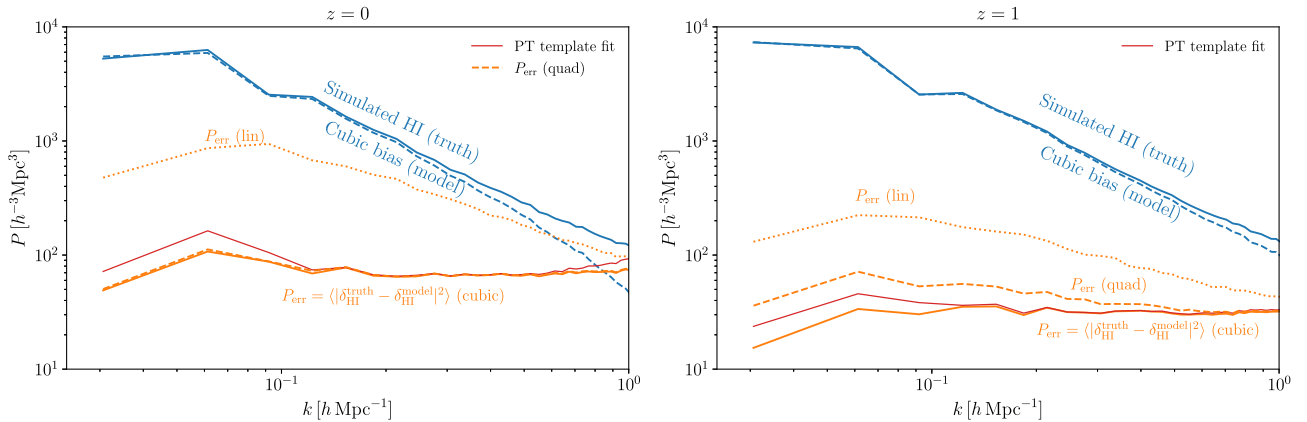

FIG. 3. Real-space power spectrum: simulated (blue solid line), best-fit cubic model (blue dashed line), and error power spectra $P_{\rm err}$ at $z = 0$ (left) and $z = 1$ (right). We show $P_{\rm err}$ for the different bias models used: cubic (solid orange line), quadratic (dashed orange), and linear (dotted orange line). We also show $P_{\rm err}$ when using the best-fit PT template based transfer functions for the cubic bias model (thin red line).
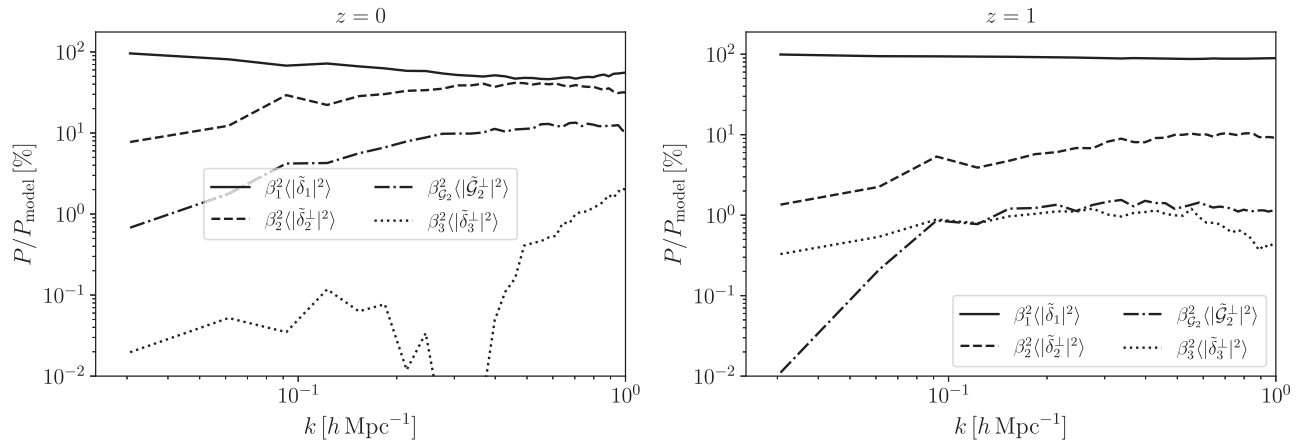
FIG. 4. Relative contribution of different terms of the best-fit power spectrum using the cubic bias model in real space. Different panels correspond to different redshifts at $z = 0$ (left) and $z = 1$ (right).

spectrum is equal to the simulated power spectrum by construction.

It is instructive to see how different nonlinear terms in the model contribute to the model power spectrum. In Fig. 4 we show the relative contributions to the best-fit power spectrum of each of the operators. We find that at both redshifts the linear term ($\tilde{\delta}_1$) dominates the signal. At $z = 0$ the quadratic term ($\tilde{\delta}_2$) contributes ∼10% or more on small scales, while the tidal field ($\tilde{\mathcal{G}}_2$) is relevant at the level of 1%–10% on all scales. The picture changes at $z = 1$, where the quadratic term contributes 1%–10% on all scales, while all higher order terms are below the subpercent level. It is important to stress that the $\langle|\tilde{\delta}_2^\perp|^2\rangle$ contains a large contribution which has a flat power spectrum. Indeed, this is the reason why including the quadratic operators significantly reduces the error power spectrum compared to the simple linear bias model. Therefore, the fact that $\langle|\tilde{\delta}_2^\perp|^2\rangle$ is larger than other nonlinear terms and even comparable to $\langle|\tilde{\delta}_1|^2\rangle$ on small scales for $z = 0$ should not be seen as a failure of perturbation theory but rather as a success in describing the true dynamics of HI clustering at the field level.

B. Best-fit transfer functions

So far we have compared the best-fit perturbative forward model and simulations and found excellent agreement between the two. However, the transfer functions that we used to minimize the mean-squared error have been completely free in each k bin. One may wonder if such large freedom is justified. In other words, if the transfer functions are very scale dependent or cannot be predicted in perturbation theory, this would undermine the success of the perturbative forward modeling.

In this section we address this issue by comparing the measured best-fit transfer functions to perturbation theory predictions. As we have already discussed, below the nonlinear scale the transfer functions should be described

by a handful of free (bias) parameters. We use the model from Sec. II A to test this prediction and measure the bias parameters. In this model only the first transfer function β_1 is scale dependent, while others are treated as constants. As this model is valid only in the perturbative regime, for the purposes of comparison we include only scales below $k < 0.4 \, h \, \text{Mpc}^{-1}$. We fit jointly all transfer functions $\beta_i(k)$ weighting each k bin by k to account for the different number of modes. The results are shown in Fig. 5, where dashed lines represent the best-fit curves, while the best-fit values of the parameters are provided in Table I. We also perform a convergence test with respect to our grid size choice and show that our results are converged in Appendix C.

As we can see, the analytical prediction provides a decent fit for the transfer functions on scales $k < 0.4 \, h \, \text{Mpc}^{-1}$. Note that the points in the first few bins have a large scatter, as a consequence of the numerical noise, given that we use only a single realization in simulations. This scatter could, in principle, be reduced by averaging over several realizations of the simulated HI field or using simulations with much larger volumes. It is also instructive to look at the values of the best-fit parameters. For consistency of the theory, the bias parameters measured in transfer functions should have the same values as those inferred using other methods, such as separate universe simulations [56–58]. While direct comparison with the biases for dark matter halos is not fully appropriate for HI, we can see from Table I that the values of b_1, b_2, and b_3 are close to the expectation for halos with $M_h \sim 10^{11} \, [h^{-1} M_\odot]$, which dominantly contribute to the HI signal. This is yet another confirmation that our perturbative framework is consistent.

Having obtained the best-fit PT template model for transfer functions, we are able to construct an approximate HI field using shifted operators and the best-fit $\beta_i(k)$ model. We test the performance of such an approximate model, hereafter the PT template model, by measuring
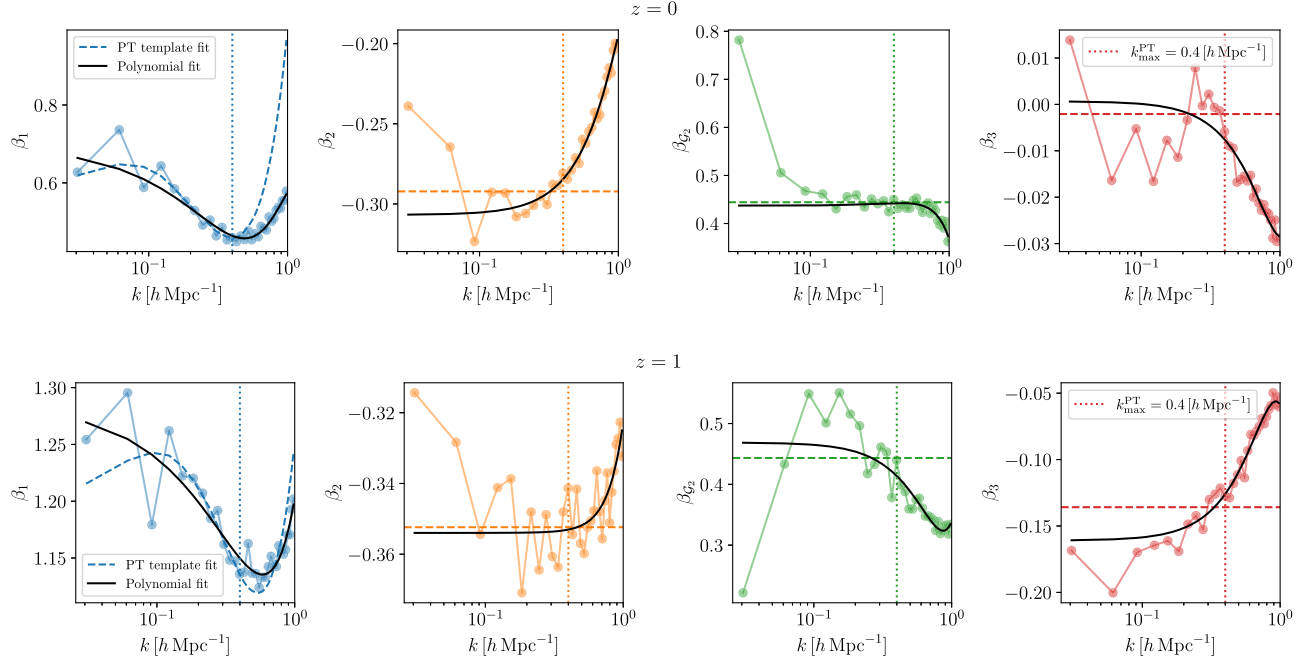
FIG. 5.   Transfer functions $\beta_i(k)$ in real space for the cubic bias model at $z = 0$ (top) and $z = 1$ (bottom) (lines with different colors). We show the best-fit PT prediction (dashed lines) obtained using scales below $k_{max} = 0.4\ h\,\mathrm{Mpc}^{-1}$ (vertical dotted lines). We also show the fits obtained using the following polynomials: $\beta_1(k) = a_0 + a_1 k + a_2 k^2 + a_4 k^4$ and $\beta_i(k) = a_0 + a_2 k^2 + a_4 k^4$ for higher order transfer functions (black solid lines) (see Table III).

P_{err}. The results are shown in Fig. 3 by a thin red line. We find that using only six free parameters gives results similar to the case of free transfer functions in each k bin, albeit with somewhat larger errors on small scales. Note that differences in the error power spectrum on large scales are a consequence of the numerical noise and the fact that we fit the noisy data with a smooth function, unlike for the free transfer functions case where all bins are independent.

Finally, in Fig. 6 we show the fractional deviation of P_{err} from a constant when using the PT template model. In comparison with the free transfer functions, the results show similar behavior albeit with a larger scatter. As we already explained, this scatter is a result of the numerical noise. The fractional deviation of P_{err} from a constant can also be seen as the relative error on the overall power spectrum model

$$\frac{P_{err} - \mathrm{const}}{P_{HI}^{truth}} = \frac{P_{HI}^{truth} - (P_{HI}^{model} + \mathrm{const})}{P_{HI}^{truth}}. \quad (22)$$

The value of the constant is chosen as an average plateau of $P_{err}(k)$ on $k = 0.1$–$0.3\ h\,\mathrm{Mpc}^{-1}$. We can see that the perturbation theory model describes the power spectrum well for $k < 0.4\ h\,\mathrm{Mpc}^{-1}$, with subpercent relative error. Due to the noise in the measurements coming from a single realization, it is hard to estimate the true amplitude of the error as a function of scale. We leave this for future work when more simulation boxes or larger volumes will be available. Let us only note that on general grounds we expect the error power spectrum to be scale dependent around the nonlinear scale, with relative corrections of the order of $(k/k_{NL})^2$. Including such terms in the error model would further improve the agreement with the simulated HI power spectrum.

C. The 21 cm IM surveys

It is useful to contrast the performance of the perturbative model that we tested in the previous section to the expected noise levels from future 21 cm IM surveys. These surveys are typically split into two main categories: single dish and interferometric experiments. The main difference between these is the range of scales they probe, with the former mainly probing large scales due to the size of the beam, especially at higher redshifts, while the latter are more sensitive to smaller scales. As we are interested in testing

TABLE I.   Best-fit transfer function parameters from the PT model at different redshifts obtained using $k_{max} = 0.4\ h\,\mathrm{Mpc}^{-1}$.

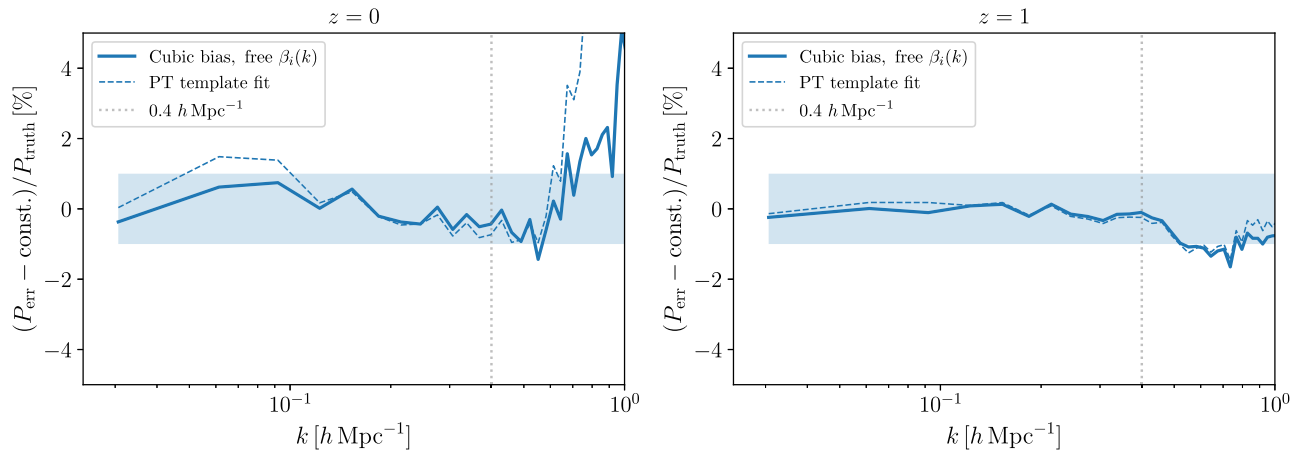| z	0	1
b_1	0.600	1.204
c_s^2	0.316	0.338
b_2	-0.292	-0.352
$b_{\mathcal{G}_2}$	0.444	0.443
b_{Γ_3}	-0.826	-1.361
b_3	-0.002	-0.136

FIG. 6.   Fractional deviation of the error power spectrum P_{err} from a constant using free transfer functions $\beta_i(k)$ (solid lines) and a PT template fit with six free parameters (dashed lines), at $z = 0$ (left) and $z = 1$ (right). Note that this quantity is identical to the relative error of the perturbative model compared to the simulated HI power spectrum. The blue shaded region shows $\pm 1\%$ of $P_{\mathrm{HI}}^{\mathrm{truth}}$.

the performance in the nonlinear regime, we focus on the interferometric approach. We follow the experimental setup similar to those of HIRAX [11] and Stage-II survey such as PUMA [14]. These surveys consist of square, compact arrays of, respectively, 32×32 and 256×256 dishes with a diameter of $D_{\mathrm{dish}} = 6$ m.

In this paper we consider an idealized setup, in which we neglect the impact of foreground cleaning [59–62] and the foreground wedge [63–66], and assume an idealized measurement. Even though this is not realistic, our main purpose is to demonstrate that even with such optimistic assumptions, the standard one-loop perturbative model is sufficiently accurate for upcoming 21 cm IM surveys. Having this in mind, the total error in the power spectrum measurements depends only on the cosmic variance and thermal noise. We first focus on instrumental thermal noise. Temperature fluctuations from the instrument and its surroundings, as well as the sky temperature fluctuations, add power to the visibility measurements. The power spectrum of the observed 21 cm signal contains the thermal noise power spectrum

$$P_{21}(k, \mu) = P_{\mathrm{HI}}(k, \mu) + P_{\mathrm{thermal}}(k, \mu). \qquad (23)$$

We estimate it following the approach in Refs. [9,67,68],

$$P_{\mathrm{thermal}} = \frac{T_{\mathrm{sys}}^2(z) X^2(z) Y(z) \lambda_{21}^4(z) S_{21}}{A_{\mathrm{eff}}^2 \mathrm{FOV}(z) t_0 n_{\mathrm{pol}} n(\mathbf{u}, z)}. \qquad (24)$$

In the expression above the system temperature T_{sys} is the sum of amplifier noise, background sky, and ground temperatures, i.e., $T_{\mathrm{sys}} = T_{\mathrm{amp}} + T_{\mathrm{sky}} + T_{\mathrm{gnd}}$. We assume the following [14]: $T_{\mathrm{amp}} = 62$ K, $T_{\mathrm{gnd}} = 33$ K and

$$T_{\mathrm{sky}} = 2.7 + 25 \left(\frac{\nu_{21}(z)}{400 \text{ MHz}} \right)^{-2.75} \text{ K}, \qquad (25)$$

where $\nu_{21}(z) = \nu_0/(1 + z)$ is the observed frequency of the HI line emitted at redshift z, with $\nu_0 = 1420$ MHz being the rest-frame frequency of the 21 cm line. Similarly, $\lambda_{21}(z) = \lambda_0(1 + z)$ is the observed wavelength of the HI emission from redshift z, with $\lambda_0 = 21$ cm. In addition, X is the comoving radial distance, and $Y(z) = \frac{c(1+z)^2}{\nu_0 H(z)}$. These terms are used to convert the angular and frequency space to the physical space. The total survey area S_{21} is assumed to be 15000 and 20000 deg^2 for HIRAX and Stage-II surveys, respectively. Furthermore, we assume an effective area per beam given by $A_{\mathrm{eff}} = \pi(D_{\mathrm{dish}}/2)^2$, with each dish having an isotropic primary beam with the field of view $\mathrm{FOV} = (1.22\lambda_{21}(z)/D_{\mathrm{dish}})^2$. For the number density of baselines $n(\mathbf{u}, z)$ we use an approximation from Ref. [14] with the parameters of a square array, where $\mathbf{u} = \mathbf{L}/\lambda_{21}$ is the baseline separation vector between two antennae in units of the observed HI wavelength. Finally, we assume $n_{\mathrm{pol}} = 2$ as the number of polarizations per antenna and an observing time of $t_0 = 1$ yr.

The power spectrum of the thermal noise from Eq. (24) is scale dependent through an uneven distribution of baselines which measure fluctuations on different spatial scales. This makes P_{thermal} dependent on both k and μ. In particular, the lack of baselines with large separations results in high thermal noise power on small scales. In this way the thermal noise we include accounts for the availability of scales each survey will be able to probe.

The second component of the error in our idealized setup is the cosmic variance. For concreteness we imagine a slice of a survey centered at $z = 1$ with the survey area S_{21}. The Gaussian contribution to the covariance matrix is given by

$$C_{k,k'} = \frac{(2\pi)^3}{V_{\text{survey}}} \frac{P_{21}^2}{2\pi k^2 \Delta k}, \qquad (26)$$

where V_{survey} is the comoving volume of the slice we consider and P_{HI} is the HI power spectrum. We choose a redshift bin width of $\Delta z = 0.2$ and $\Delta k \approx 0.03 \ h\,\text{Mpc}^{-1}$, the same k binning that we have used throughout the paper.

The expected uncertainty from the combined cosmic variance and thermal noise for the two different setups that we consider is shown in the top panel of Fig. 7, with the same k binning as we used for P_{HI}. To obtain only the k-dependent part of the thermal noise power spectrum, we compute the monopole part of P_{thermal} in each k bin. We can see that the theoretical uncertainties are within the observational noise in the perturbative regime. For a HIRAX-like survey this is true even on smaller scales. A more quantitative way to test the significance of the deviation of theory from simulations is to calculate the signal to noise for this
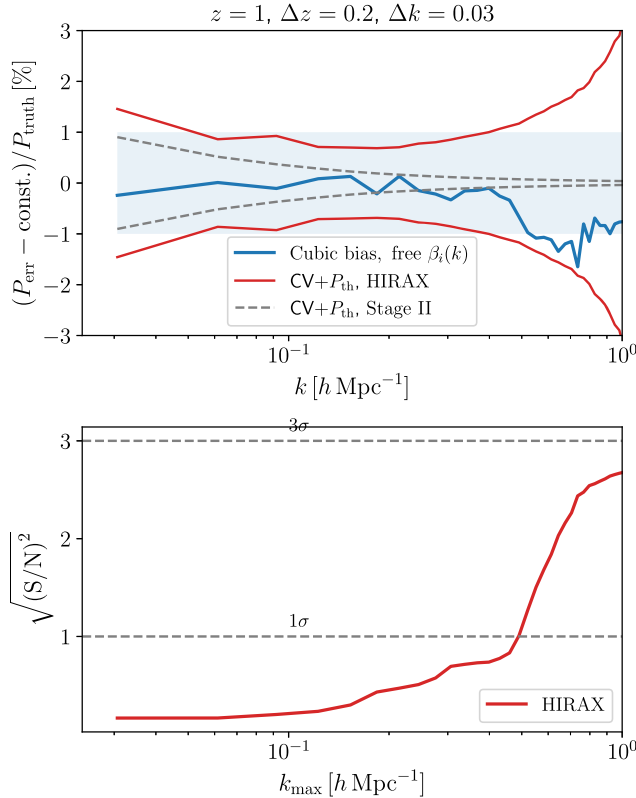

FIG. 7. Top panel: similar to Fig. 6 at $z = 1$ but comparing to the expected performance of different 21 cm IM surveys. The blue line is the relative error of the best-fit cubic bias model with free transfer functions. The solid red and dashed gray lines show the expected uncertainty after including the thermal noise and the cosmic variance for a HIRAX-like experiment and the Stage-II experiment, respectively. The blue shaded region represents $\pm 1\%$ of $P_{\text{HI}}^{\text{truth}}$. Bottom panel: cumulative signal-to-noise ratio squared $(\text{S/N})^2$ for detecting deviation of the best-fit cubic bias model from the true HI power spectrum as a function of k_{max} for a HIRAX-like survey.

difference given the error bars. We plot such cumulative signal to noise for a HIRAX-like survey in the bottom panel of Fig. 7. The perturbation theory model is clearly consistent with the data for $k < 0.4 \ h\,\text{Mpc}^{-1}$, and on smaller scales the significance of the difference compared to the truth never exceeds the 3σ level due to increasing thermal noise. In a realistic data analysis, where cosmological parameters are varied as well, we expect this difference to be even smaller. Therefore, this result can also be interpreted as a hint that the current level of theoretical modeling discussed in this paper is sufficient to analyze the data from the upcoming HI surveys such as HIRAX.

We would like to emphasize that these results should be taken with a grain of salt. Given a relatively small volume of the IllustrisTNG box and a single realization, the numerical noise in the transfer functions and the best-fit model, clearly visible in Fig. 7, can impact the signal-to-noise estimate. For this reason we do not explicitly show the cumulative signal-to-noise estimate for the Stage-II surveys that have much smaller error bars. Also, note that our choice of Δz in this exercise is somewhat arbitrary. In the real data analysis the choice of Δz will be driven by the compromise between increasing the cosmological volume in each z bin and probing the redshift evolution across the bin. While specific choices in the future are likely to be different, we follow what is done in the galaxy power spectrum analysis of the most recent galaxy surveys such as BOSS [3].

D. One-point probability distributions

So far we have mainly focused on the power spectrum as the standard observable used in cosmological data analyses. However, as we pointed out in the Introduction, one of the advantages of the field level method is that it allows us to predict other, very different, summary statistics. As an illustration we will focus on the one-point probability distribution function (PDF). The one-point PDF of HI from TNG100-1 has also been studied in Ref. [69]. Here we test our perturbative prediction for the one-point PDF by comparing the histograms of smoothed density fields from simulations and the best-fit cubic bias for various smoothing scales. The results are shown in Fig. 8.

For large smoothing scales the agreement between the two is rather good, particularly at $z = 1$, even for relatively large overdensities. Focusing on smaller smoothing scales, we find that the best-fit model tends to overpredict the number of pixels near the mean density ($\delta \sim 0$) and underpredict the number of highly overdense pixels. This is very much in line with what is expected from perturbation theory, as it is unable to account for the compactness of highly nonlinear objects, such as halos. In conclusion, this is yet more evidence that the perturbative description of the HI field works as expected, predicting the one-point PDF correctly, including the proper shape of the distribution even in the tails, for smoothing scales as small as $R = 5 \ h^{-1}\,\text{Mpc}$.
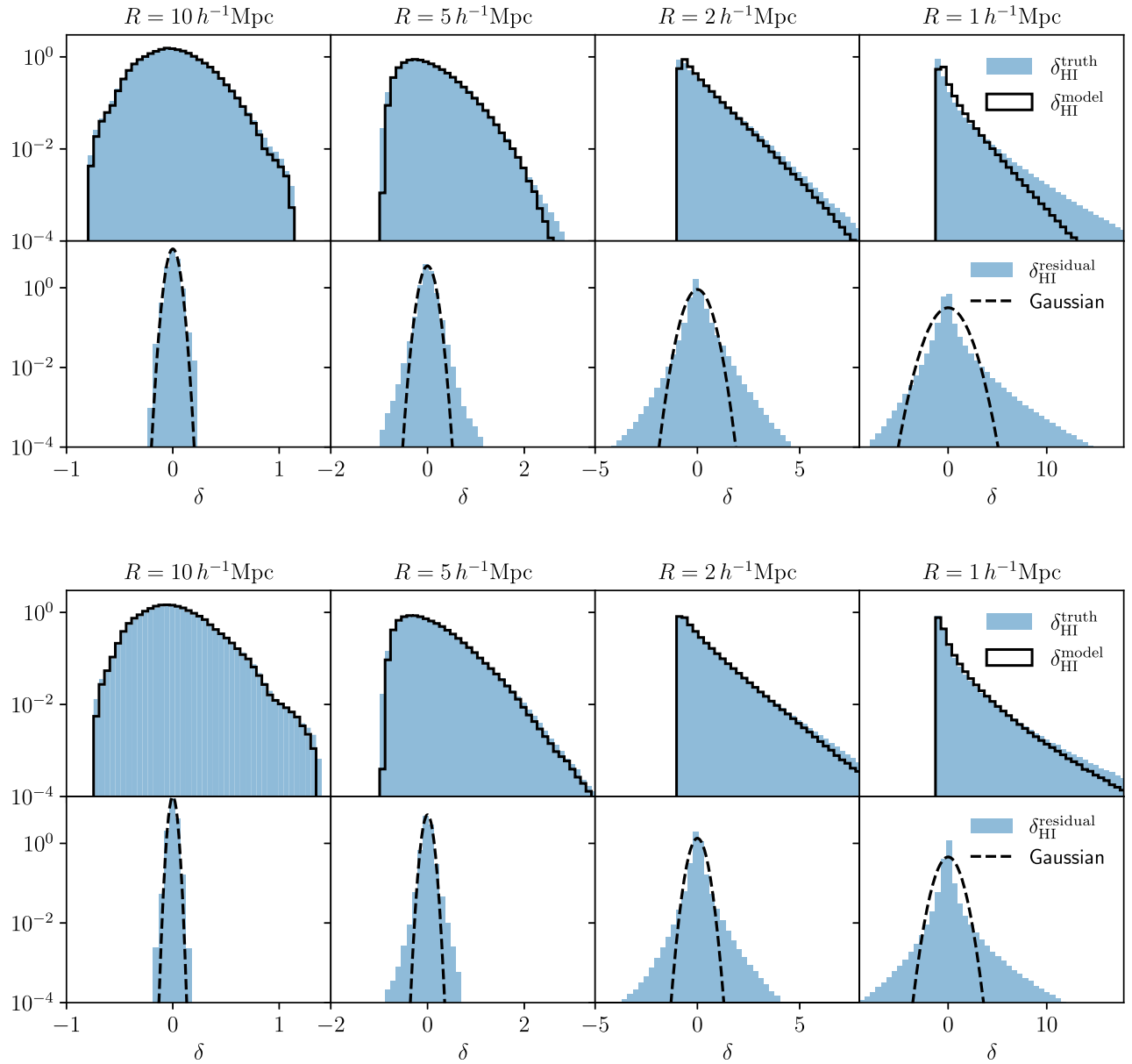
FIG. 8. Histogram of the true, best-fit, and residual HI overdensity fields using the cubic bias model at two redshifts: $z = 0$ (top panel) and $z = 1$ (bottom panel). Different panel columns correspond to applying a 3D Gaussian filter to residuals with varying smoothing scales R. Top rows in each panel show the simulated (blue) and the best-fit model (black lines) HI overdensity distribution. We find that the (cubic) bias model provides a good description of the simulated HI one-point PDFs, especially on larger smoothing scales. Bottom rows show the best-fit field level residuals (blue) compared to a Gaussian distribution with matching variance and zero mean. Using larger smoothing scales R, the residuals (or model errors) are nearly Gaussian, with larger deviations present in a small fraction of pixels.

E. Redshift space

Let us finally turn to redshift space. We can, in principle, repeat every analysis we did for real space, and this would lead to similar results. In this section we briefly present the most important tests of the perturbative forward model—comparison of theory and simulations at the level of the map and power spectrum.

In Fig. 9 we show a 2D slice of the simulated HI overdensity field in redshift space in comparison to the best-fit cubic model and the residuals at redshifts $z = 0$ and $z = 1$. As in the case of real space, we see a general agreement between the simulation and the best-fit model. Notable differences in the residuals are again more pronounced on small scales, and they match the locations of large overdensities.

In order to compare the anisotropic redshift-space power spectrum to the theoretical prediction, we measure it in the three wide μ bins centered around $\mu = 0.17$, $\mu = 0.50$, and
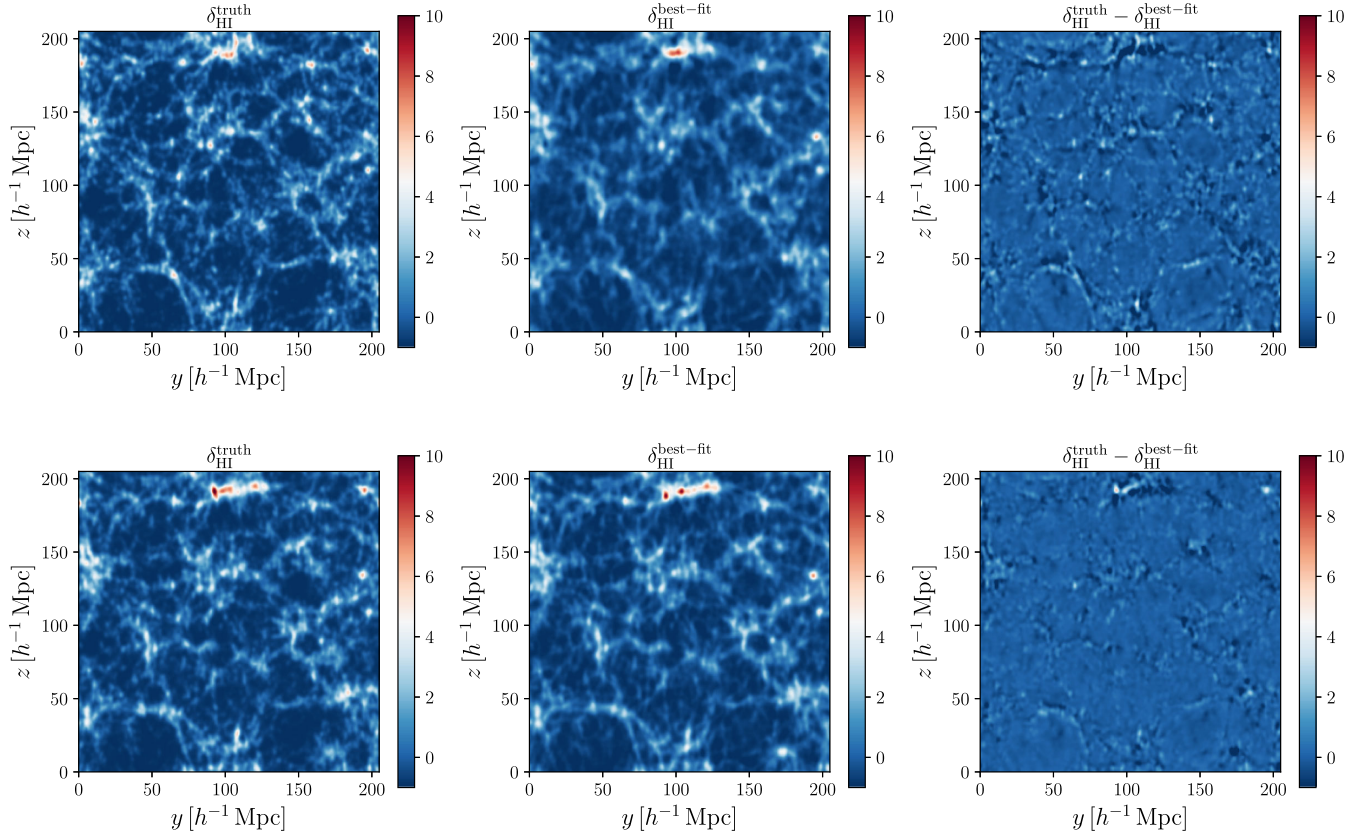
FIG. 9. Similar to Fig. 1, but now in redshift space: real-space slices in the y-z plane of the simulated HI overdensity field (left), the best-fit cubic bias model (middle), and the residuals (right), at $z = 0$ (top) and $z = 1$ (bottom). All density fields are smoothed with a $R = 1\ h^{-1}$ Mpc 3D Gaussian filter, while the depth of each slice is $20\ h^{-1}$ Mpc. The line of sight is along the $\hat{z}$ direction.

$\mu = 0.83$. The results are shown in Fig. 10. Even though the measurements have a large scatter, we find that perturbation theory predicts the measured power spectrum well on large scales. Looking at the error, we can see that it is still flat in k on large scales, with similar amplitude (in all bins) as in real space, as expected. On small scales the deviation from a flat power spectrum is more prominent than in real space. This

is a consequence of the nonlinear velocity, which can produce large distortions in redshift space. We will comment more on this scale dependence in Sec. V and use it to estimate the nonlinear HI velocity dispersion.

In Fig. 11 we show the redshift-space transfer functions $\beta_i(k, \mu)$. We find that these transfer functions are rather smooth, both in k and μ. On large scales, β_1 and $\beta_{\mathcal{G}_2}$ have
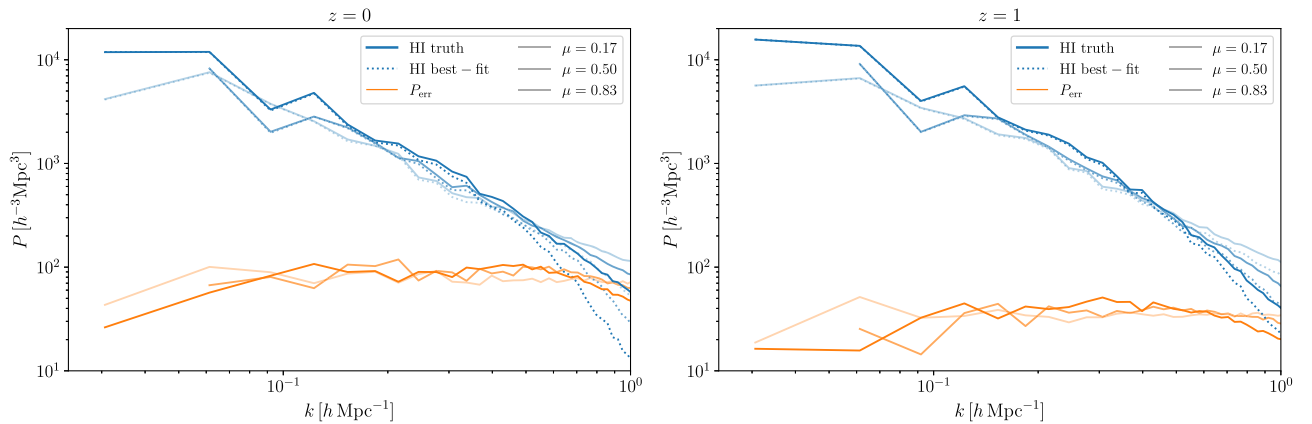

FIG. 10. Redshift-space power spectrum: simulated, best-fit model, and error power spectrum $P_{\mathrm{err}}(k, \mu)$ at $z = 0$ (left panel) and $z = 1$ (right panel). The power spectra are measured in three wide μ bins (different shading).
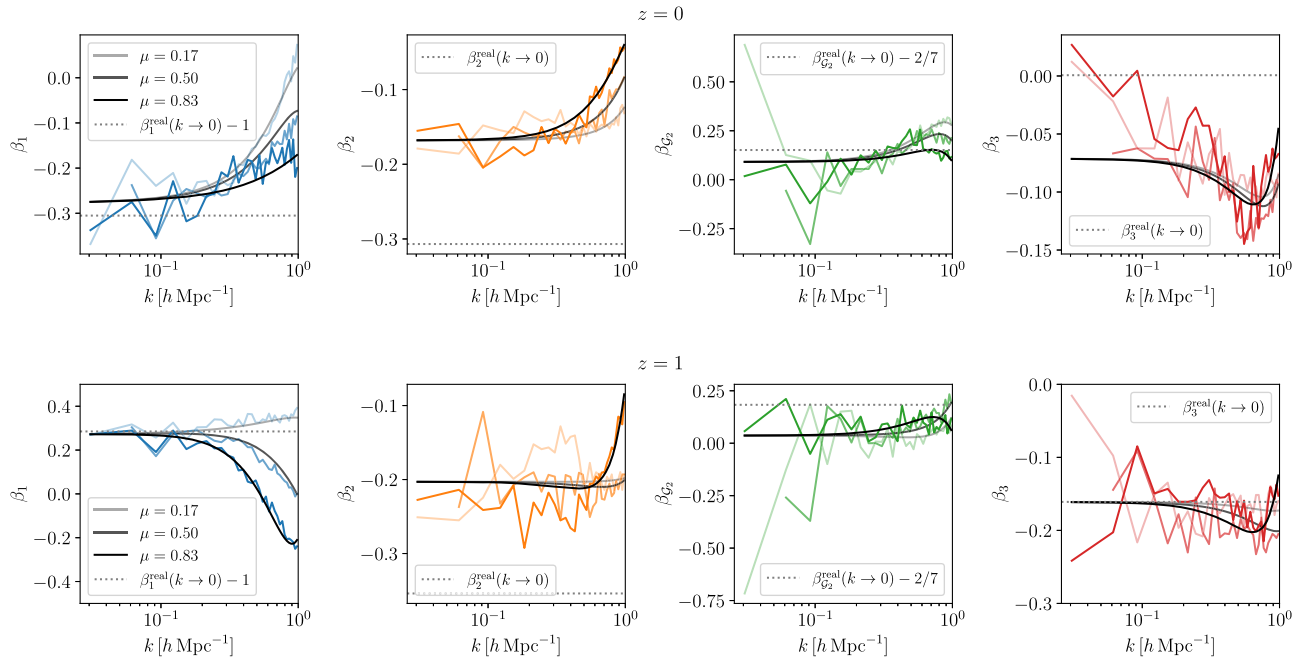
FIG. 11. Redshift-space transfer functions $\beta_i(k, \mu)$ for the cubic bias model at $z = 0$ (top) and $z = 1$ (bottom) (lines with different colors). We show the best-fit transfer functions (black solid lines) using the following polynomials: $\beta_1(k, \mu) = a_0 + a_1 k + a_2 k^2 + a_4 k^4 + a_{22}(k\mu)^2 + a_{44}(k\mu)^4$ and $\beta_i(k, \mu) = a_0 + a_2 k^2 + a_4 k^4 + a_{22}(k\mu)^2 + a_{44}(k\mu)^4$ for higher order transfer functions (see Table III). Different shading corresponds to different μ bins in all panels. For comparison we also show the low-k values using polynomial fits in real space (a_0 values from Table III) (dotted lines).

the same value in each μ bin, as a consequence of our choice to keep δ_Z^s and $\tilde{\mathcal{G}}_2^\parallel$ explicitly in the model for the HI field in redshift space. As expected, β_2, $\beta_{\mathcal{G}_2}$, and β_3 are rather flat on large scales, and they can be approximated as being constant. The first transfer function β_1 is more scale dependent, as expected, but the measurements in different μ bins on large scales are very noisy, not allowing for a robust test of the perturbative model in redshift space. For this reason, we will use a simple phenomenological fit for β_1 to account for the k and μ dependence. We will come back to

this and discuss details in Sec. VI. The bottom line is that, in the same way as in real space, the transfer function behavior on large scales is compatible with perturbation theory expectations.

Similarly to what we did for real space, in Fig. 12 we show the relative error of the model after subtracting a constant P_{err} piece. The value of the constant is chosen as an average plateau of $P_{\text{err}}(k, \mu)$ on $k = 0.1$–$0.3 \, h \, \text{Mpc}^{-1}$ scales in each μ bin. As before, the fractional deviation of $P_{\text{err}}(k, \mu)$ from the constant allows us to infer at which
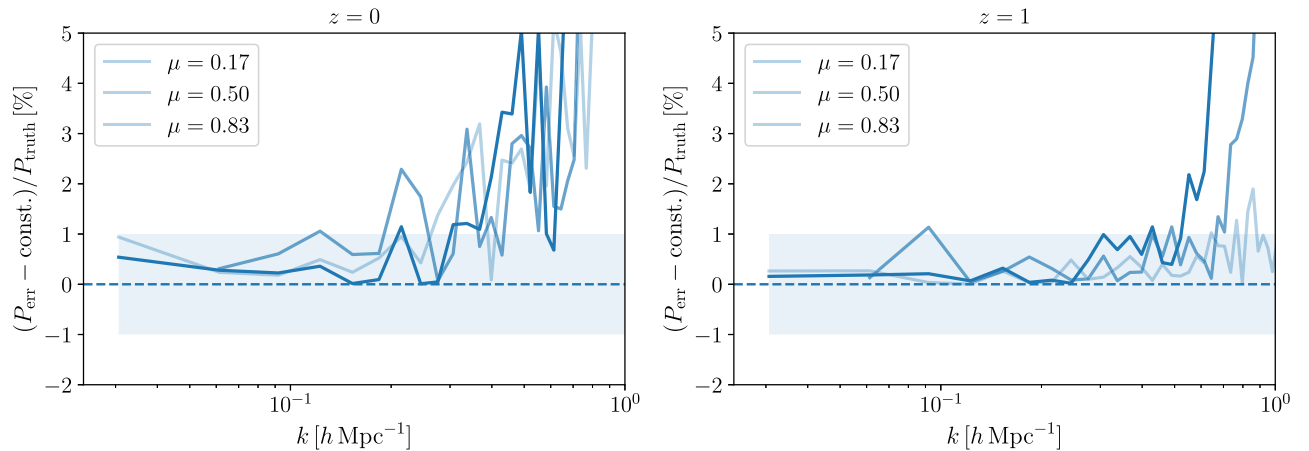

FIG. 12. Fractional deviation from the constant of the error power spectrum P_{err} using the cubic bias model in redshift space with free transfer functions $\beta_i(k, \mu)$ at $z = 0$ (left) and $z = 1$ (right). The blue shaded region represents $\pm 1\%$ of $P_{\text{HI}}^{\text{truth}}$.
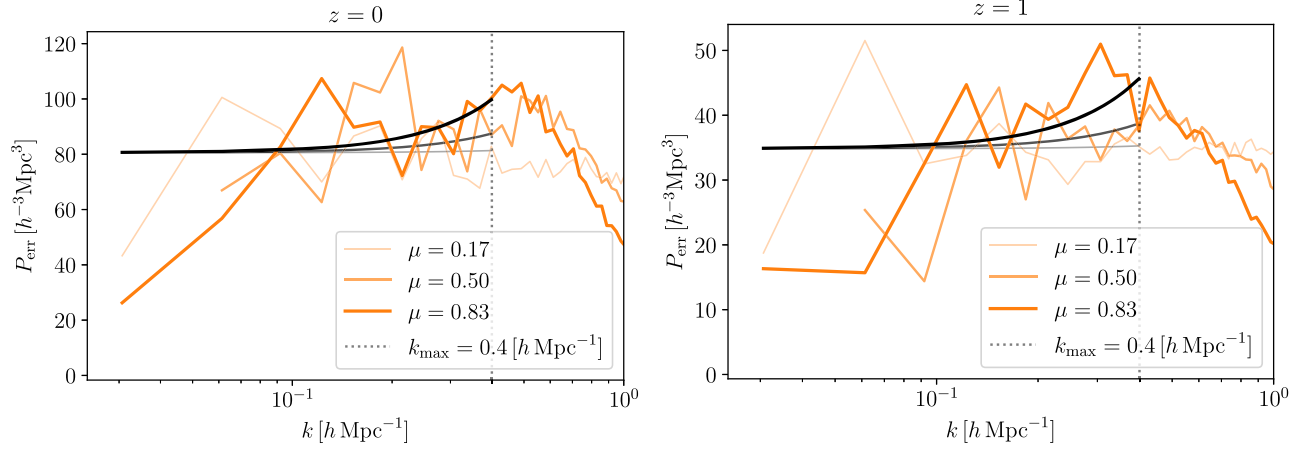
FIG. 13. Redshift-space error power spectrum $P_{\mathrm{err}}(k, \mu)$ at $z = 0$ (left panel) and $z = 1$ (right panel). The black solid lines show the best-fit model for P_{err} from Eq. (28) using $k_{\mathrm{max}} = 0.4\ h\,\mathrm{Mpc}^{-1}$ (vertical dotted line). All power spectra are measured in three wide μ bins (orange lines), and different shadings and line widths correspond to different μ bins.

scales we expect our PT model to break down. We can see that the cubic model is able to accurately model the true HI power spectrum within 1% up to $k_{\mathrm{max}} \sim 0.3\ \&\ 0.4\ h\,\mathrm{Mpc}^{-1}$ at redshifts $z = 0$ and 1, respectively. We can also see that the model is accurate up to smaller scales (higher ks) for lower μ bins, as these modes are less affected by the nonlinear RSDs.

V. HI NOISE PROPERTIES

We have argued in the Introduction that one of the main advantages comparing the theory and simulations at the field level is that such a procedure allows us to clearly isolate the model error. Understanding the properties of this error is equally important as having a good model, both for standard analyses based on n-point functions and for the more ambitious methods such as field-level forward modeling. In this section we study the properties of the error in detail and comment on its peculiarities in the case of the HI field.

A. Non-Gaussianity of the error

We first look at the one-point PDF of the model error. As we have seen in Sec. IV D, the cubic bias model provides a good description of the simulated HI field. However, we can also study the properties of the model error distribution. To obtain the model error distributions we compute the histogram of the residuals ($\delta_{\mathrm{HI}}^{\mathrm{truth}} - \delta_{\mathrm{HI}}^{\mathrm{model}}$) for different 3D Gaussian smoothing scales, and we show the results in the bottom panels of Fig. 8. We estimate the level of non-Gaussianity by comparing the model errors to the Gaussian distribution with matching variance and zero mean. We find that using smaller smoothing scales, the model error distribution deviates from the Gaussian distribution, especially in the tails. This means that assuming a Gaussian likelihood in the data analysis at the field level is not

adequate. However, these deviations become smaller, and the model errors become nearly Gaussian when larger smoothing scales ($R = 5$ or 10 $h^{-1}\,\mathrm{Mpc}$) are used. Some discrepancies in the tails of the distributions are still present, although the number of discrepant pixels is relatively small.

B. Amplitude of the noise on large scales

In the previous section we found the noise power spectrum to be flat in k. Here we want to discuss its amplitude in detail. Naively, one may think that the only fluctuations on large scales are the linear modes (multiplied by the linear bias) and random fluctuations due to the discreteness of the bias tracer. It is then natural to define the HI stochasticity as the power spectrum of the following difference: $\delta_{\mathrm{HI}}^{\mathrm{truth}} - b_1 \delta_{\mathrm{m}}$. Using the best-fit value for b_1, this can be expressed in terms of the HI-matter cross-correlation coefficient $r_{\mathrm{HI,m}}$ [16,70]:

$$P_{\mathrm{HI}} - \frac{P_{\mathrm{HI,m}}^2}{P_m} = P_{\mathrm{HI}}(1 - r_{\mathrm{HI,m}}^2), \qquad (27)$$

where $P_{\mathrm{HI,m}}$ and P_m are the HI-matter cross and matter power spectra, respectively. Note that the stochasticity is equivalent to P_{err} for the linear bias model.

However, the measured amplitude of the stochasticity is significantly larger than the naive sampling noise, given by $1/\bar{n}$, where $\bar{n}$ is the number density of M_{HI}-weighted halos [16]. We call this quantity M_{HI}-weighted sampling noise and compute it following Eq. (26) from Ref. [16]. Furthermore, we find good agreement between M_{HI}-weighted sampling noise and the amplitude of the high-k plateau of the M_{HI}-weighted halo power spectrum, similar to what was found in the case of TNG100-1 [16]. The discrepancy between stochasticity and sampling noise is well known for small-mass halos [16,34,71,72], and we
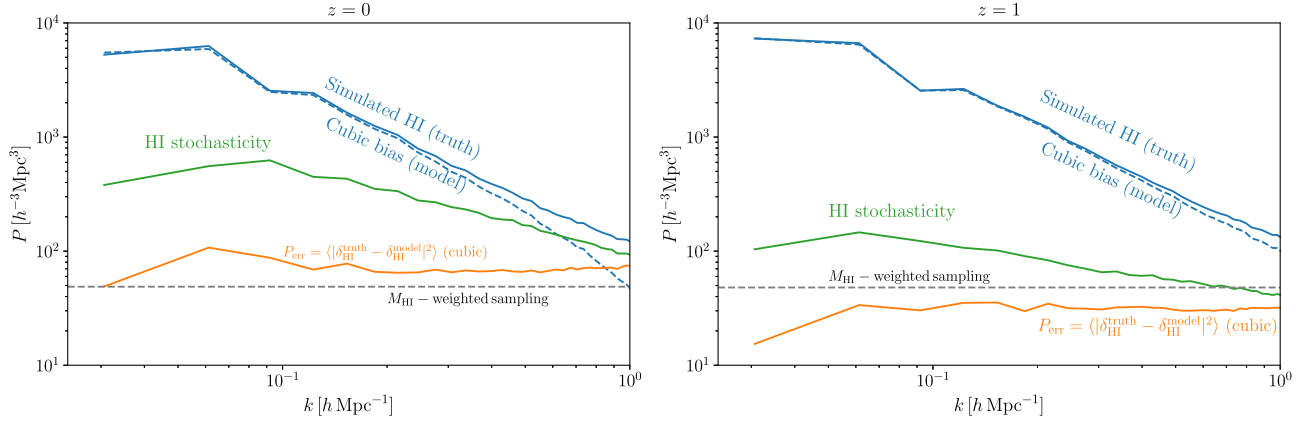
FIG. 14. HI noise power spectrum estimates: P_{err} using the cubic bias model (orange), HI stochasticity $P_{\mathrm{HI}}(1 - r_{\mathrm{HI,m}}^2)$ (green), and the M_{HI}-weighted sampling noise (gray dashed). For comparison we show the simulated and best-fit P_{HI} (blue lines) models. We find P_{err} is almost an order of magnitude smaller than HI stochasticity on the largest scales at both redshifts, while at $z = 1$, P_{err} is smaller than the usually assumed HI noise power spectrum, M_{HI}-weighted sampling noise.

show the measurements for Illustris in Fig. 14. This discrepancy between the M_{HI}-weighted sampling noise and stochasticity clearly indicates that something is inconsistent in the simple picture we have just described.

The resolution of this problem was given in [34]. Through the nonlinear bias operators, such as δ_2, the galaxy density field on very large scales receives contributions from small but perturbative modes. These long-wavelength fluctuations have a flat power spectrum which is indistinguishable from the sampling noise, but it can have much larger amplitudes for very dense tracers.

To explicitly test this hypothesis, we can compare the error power spectrum in the model, which includes quadratic biases to the M_{HI}-weighted sampling noise. This is shown in Fig. 14. In the full quadratic and cubic model, P_{err} is indeed comparable to the Poisson noise. This means that at the map level, perturbation theory correctly predicts the modes which lead to large stochasticity.

This has several important implications. First, the nonlinear bias is needed in order to have the correct map, not only on small scales in order to improve the model but also on large scales in order to consistently predict all long-wavelength modes. Second, when calculating the covariance matrix for the power spectrum, the total noiselike contribution is much larger than the M_{HI}-weighted sampling noise, and indeed, it is given by the stochasticity. While in simulations we were exploiting the fact that we knew the initial conditions, which allowed us to disentangle the contributions to the noise from the nonlinear bias and Poisson sampling, in a real analysis the initial conditions are unknown; as a consequence, the two contributions to the noise power spectrum are indistinguishable. This is an important observation since it implies that the realistic values for the amplitude of the noise in the HI power spectrum Fisher forecasts are given by stochasticity, which can be an order of magnitude larger than what is usually assumed [14,67], making the forecasts overly optimistic.

Finally, one can ask if an alternative analysis can be done such that the "true" noise, close to the M_{HI}-weighted sampling noise, is the only price to pay in the total error budget. After all, since the long-wavelength fluctuations induced by the nonlinear bias are predictable, should they be used as a signal rather than treated as the noise? While this cannot be done in the power spectrum analysis, it is possible to achieve it by including higher order n-point functions or with the full forward modeling at the field level. As explained in [73], this situation is similar to the BAO reconstruction, where the information at the map level is used to sharpen the BAO peak, which can be measured in the two-point function only with limited precision.

C. Nonlinear velocity dispersion estimates

As we already emphasized, the error on small scales can be dominated by the nonlinearities that are not taken into account in the model. The dominant nonlinearities in redshift space are Fingers of God induced by the nonlinear velocity dispersion $\sigma_{v,\mathrm{NL}}$. Their effect on the noise is to introduce a characteristic scale and angular dependence proportional, at leading order, to $\frac{1}{2}\sigma_{v,\mathrm{NL}}^2 f\mu^2 k^2$. We can use this fact to estimate the nonlinear velocity dispersion from the error power spectrum. We use the following model:

$$P_{\mathrm{err}}(k,\mu) = c_{\epsilon,1}\left(1 + \frac{1}{2}\sigma_{v,\mathrm{NL}}^2 f\mu^2 k^2\right), \quad (28)$$

where $c_{\epsilon,1}$ and $\sigma_{v,\mathrm{NL}}$ are free parameters. We obtain the best-fit values by fitting $P_{\mathrm{err}}(k,\mu)$ and weighting each k-bin by k. We show the best-fit model of P_{err} in Fig. 13, while the best-fit values at both redshifts are presented in Table II.

We find that the constant part of P_{err} is in agreement between real and redshift space. Our estimate of $\sigma_{v,\mathrm{NL}}$ shows that the HI nonlinear velocity dispersion is smaller than for the typical galaxy samples at similar redshifts [35,74], in line

TABLE II. Best-fit parameters of the $P_{\mathrm{err}}(k,\mu)$ model at different redshifts obtained using $k_{\max} = 0.4 \; h\,\mathrm{Mpc}^{-1}$.

z	0	1
$c_{\varepsilon,1} \; [h^{-3}\,\mathrm{Mpc}^3]$	80.6	34.8
$\sigma_{v,\mathrm{NL}} \; [h^{-1}\,\mathrm{Mpc}]$	2.87	2.54

with the conclusions of previous studies [16]. However, this measurement has to be taken with a grain of salt, given the large numerical noise in the measured $P_{\mathrm{err}}(k,\mu)$ on large scales.

VI. HI MOCK DATA

In all previous sections we focused on modeling based on perturbation theory. However, inspired by the perturbative templates (shifted operators), we can try to see how well such a simple description works even on scales where perturbation theory formally does not apply. In other words, we can think of shifted operators as a phenomenological basis in which we can decompose the nonlinear field, with a set of smooth transfer functions. Even though unjustified in the perturbative framework, such an approach can be useful as a simple tool to easily generate realistic HI fields with HI clustering properties similar to the ones of TNG simulation, even on small scales beyond k_{NL}. In this section we show how to do this.

For producing mocks, we will focus on scales up to $k_{\max} = 1 \; h\,\mathrm{Mpc}^{-1}$. The two main observations to keep in mind are that the measured transfer functions are smooth and that they do not vanish at small scales, indicating that the HI field still correlates with the perturbative templates. Given the smoothness of the transfer functions, we can describe them using simple, low-order polynomials in k and μ. In the case of real space we use the following model:

$$\beta_1^{\mathrm{real}}(k) = a_0 + a_1 k + a_2 k^2 + a_4 k^4,$$
$$\beta_{i\neq 1}^{\mathrm{real}}(k) = a_0 + a_2 k^2 + a_4 k^4. \tag{29}$$

Note that in β_1^{real} we also include the term linear in k. This is needed since several terms in Eq. (9) have very mild scale dependence around $k \sim 0.1 \; h\,\mathrm{Mpc}^{-1}$, given the form of the loop corrections and the ΛCDM linear power spectrum. In the case of redshift space we include μ-dependent terms equally for all transfer functions,

$$\beta_i^{\mathrm{rsd}}(k,\mu) = \beta_i^{\mathrm{real}}(k) + a_{22}(k\mu)^2 + a_{44}(k\mu)^4. \tag{30}$$

The polynomial coefficients a_i and a_{ii} are free parameters for each transfer function β_i, in both real and redshift-space models. The total number of free parameters in the cubic bias model is thus 13 ($4 + 3 \times 3$) for real and 26 ($6 + 4 \times 5$) for redshift space, at a single redshift.

We obtain the values of these parameters by doing the least-square fits to the measured transfer functions, weighting each k bin by k. The polynomial fits for real and redshift space are shown in Figs. 6 (black solid lines) and 11 (dashed lines), respectively. All the best-fit parameter values are shown in Table III. We find that this simple model provides a good description of the measured transfer functions at both considered redshifts and for our choice of $k_{\max} = 1 \; h\,\mathrm{Mpc}^{-1}$.

Before we move on, let us comment on the inferred values of transfer function parameters in the low-k limit. As we argued in Sec. II B, we expect certain relations between real- and redshift-space transfer functions on large scales. This expectation turns out to be correct. As we can see in Table III, the following approximate relations hold

$$\beta_1: \quad a_0^{\mathrm{real}} \approx a_0^{\mathrm{rsd}} + 1, \tag{31}$$

$$\beta_2: \quad a_0^{\mathrm{real}} \approx a_0^{\mathrm{rsd}}, \tag{32}$$

$$\beta_{\mathcal{G}_2}: \quad a_0^{\mathrm{real}} \approx a_0^{\mathrm{rsd}} + \frac{2}{7}, \tag{33}$$

$$\beta_3: \quad a_0^{\mathrm{real}} \approx a_0^{\mathrm{rsd}}, \tag{34}$$

TABLE III. Best-fit parameters of the transfer functions when using the polynomial model in real and redshift space $\beta_i(k,\mu)$ at different redshifts obtained using $k_{\max} = 1 \; h\,\mathrm{Mpc}^{-1}$ (see Figs. 5 and 11).

		Redshift space		Real space	
z		0	1	0	1
β_1	a_0	−0.28	0.27	0.69	1.28
	a_1	0.09	0.09	−1.04	−0.52
	a_2	0.40	0.08	1.19	0.46
	a_4	−0.17	−0.04	−0.28	−0.02
	a_{22}	−0.53	−2.11	⋯	⋯
	a_{24}	0.33	1.78	⋯	⋯
β_2	a_0	−0.17	−0.20	−0.31	−0.35
	a_2	−0.02	−0.002	0.14	−0.001
	a_4	0.02	0.01	−0.03	0.03
	a_{22}	0.22	−0.13	⋯	⋯
	a_{44}	−0.13	0.45	⋯	⋯
$\beta_{\mathcal{G}_2}$	a_0	0.09	0.04	0.44	0.47
	a_2	0.54	−0.09	0.05	−0.38
	a_4	−0.34	0.16	−0.12	0.26
	a_{22}	−0.42	0.62	⋯	⋯
	a_{44}	−0.20	−1.02	⋯	⋯
β_3	a_0	−0.07	−0.16	0.001	−0.16
	a_2	−0.12	−0.02	−0.06	0.24
	a_4	0.11	0.02	0.03	−0.14
	a_{22}	−0.10	−0.26	⋯	⋯
	a_{44}	0.23	0.48	⋯	⋯

both at $z = 0$ and $z = 1$. Even though the agreement is not perfect, up to $\Delta a_0 \sim 0.1$, the real-space and redshift-space results are compatible, given the large numerical noise at low k in the measured transfer function. This is yet another confirmation of the consistency of the theory.

So far we have only considered the part of the HI signal that comes from the model prediction and correlates with the IC. However, we also need to include the stochastic part in order to generate realistic HI mock data. For this purpose, we generate 3D fields with the power spectrum matching the inferred P_{err}. In real space we generate a noise realization with a flat power spectrum amplitude of $P_{\mathrm{err}} = (68.53, 33.00) \; [h^{-3} \, \mathrm{Mpc}^3]$ at $z = (0, 1)$, respectively. These amplitudes correspond to the P_{err} plateau between scales $k = 0.1$–$0.3 \; h \, \mathrm{Mpc}^{-1}$. In redshift space we generate a noise realization with the power spectrum constructed with a polynomial fit to the inferred $P_{\mathrm{err}}(k, \mu)$ (see Fig. 13). For the polynomial model we use the following ansatz,

$$P_{\mathrm{err}}(k, \mu) = a_0 + a_2 k^2 + a_3 k^3 + a_4 k^4 + \sum_{i=2}^{4} a_{ii}(k\mu)^i, \quad (35)$$

and perform a weighted least-squares fit up to $k_{\mathrm{max}} = 1 \; h \, \mathrm{Mpc}^{-1}$, weighting each k bin by k as before. The best-fit parameter values are shown in Table IV. Note that this is different from the model that we used in Eq. (28), where the goal was to estimate HI velocity dispersion and where we used only the scales in the perturbative regime. With the polynomial model that we use here, the goal is to generate the most realistic noise in redshift space and be able to use it even on smaller scales. Finally, having generated the noise fields, we add them to the signal fields to obtain the final HI mock data in either real or redshift space.

We can now use these fits to generate HI mock data. Since we measure the transfer functions without cosmic variance (using the same IC for the model and the simulated field), the transfer functions and their smooth polynomial fits can be applied to realizations with different IC as the dependence on IC is in the shifted fields. For any given realization of IC, we can then generate orthogonalized shifted fields and simply multiply them by the best-fit transfer function polynomials in order to obtain the HI signal map [cf. Eqs. (8) and (11)]. The code we use in this paper, `Hi-Fi mocks`, performs this procedure exactly.

TABLE IV. Best-fit parameters using the polynomial fit from Eq. (35) to model $P_{\mathrm{err}}(k, \mu)$ at different redshifts.

z	a_0	a_2	a_3	a_4	a_{22}	a_{33}	a_{44}
0	77.3	99.1	-232.3	111.8	463.6	-1282.1	861.9
1	36.6	16.4	-50.8	29.1	106.4	-317.6	206.7

This procedure has several free parameters. One is the grid size used to generate shifted fields, and the other is the box size. In our fiducial analysis we use $N_{\mathrm{mesh}} = 256^3$, which results in ≈ 16 million particles. Given the size of the TNG300-1 simulation ($L = 205 \; h^{-1} \, \mathrm{Mpc}$), this configuration makes generating HI mock data achievable even on personal computers in a short amount of time, and the full HI mock data can be obtained on a modern laptop in a few minutes. We note that for larger box sizes, a larger grid size is needed in order to accurately probe smaller scales. This makes computation times longer and the memory load larger.

We can contrast this procedure to the more standard way of generating HI mock data using N-body simulations and techniques such as halo occupation distribution, subhalo abundance matching, and semianalytical models. This is usually done by populating dark matter halos with HI mass after running the full N-body simulation. Approximate approaches to the full N-body simulations, such as COLA [75] or FastPM [76], are able to accelerate this step. However the main bottleneck is when running halo finder on top of simulation outputs, which takes a comparable computational time. This step is not differentiable, which makes using the forward model in this setup unfeasible. Approaches based on machine learning techniques have been used to generate HI fields in cosmological boxes from dark matter fields [38]; however, these require training and testing, and at the moment, we lack the interpretability. On the other hand, our method is computationally cheap as it does not require running N-body simulations or halo finders; in addition, it requires no training, and in the low-k limit, it is based on perturbation theory. Furthermore, it can be easily calibrated to other hydrodynamical simulations or simulations with different baryonic feedback prescriptions.

VII. CONCLUSIONS

In this paper we tested a perturbative forward model against the simulated HI at the field level. We found excellent agreement on large scales, with subpercent precision for $k < 0.3 \; h \, \mathrm{Mpc}^{-1}$ in real and redshift space, at redshifts of $z = 0$ and $z = 1$. This is confirmed by looking at different statistics, such as cross-correlation between the maps, the HI power spectrum, and one-point probability distribution functions of the HI field. We also studied the properties of the HI noise. We confirmed the well-known fact that the stochasticity for HI can be much larger than the M_{HI}-weighted sampling noise, and we explained the origin of this discrepancy. We argued that this has an important implication for the HI power spectrum Fisher matrix forecasts and that it motivates the field level inference. Finally, using perturbation theory as an inspiration for the basis of templates at the field level, we found a phenomenological fit for the HI maps and

provided a simple and efficient code to generate mock HI data for arbitrary initial conditions and volumes. Our results can also be seen as an important validation step for the HI modeling; they show that the currently existing pipelines for galaxy clustering analysis based on perturbation theory can, in principle, be applied to the neural hydrogen as well.

There are several interesting directions to explore in the future, and we mention some of them here. First, we can use our `Hi-Fi mocks` code to generate realistic synthetic data in large volume boxes. The same code can be used to produce many realizations needed to estimate the covariance matrix. The currently existing pipelines for the power spectrum analysis can then be applied to these synthetic data. The resulting errors on the cosmological parameters would be the most realistic estimates of the constraining power of the HI power spectrum, even though they are in a very idealistic setup. A step towards a more realistic analysis requires inclusion of foregrounds, wedge, and thermal noise, all of which can be added to the perturbative forward model. We leave this for future work.

Another direction is to explore the alternative summary statistics in more detail. We have already presented the results for one-point PDFs. Using predictions for the maps, one can also study to what extent perturbation theory predicts the size function of voids, the nearest-neighbor distribution, etc.

A natural question to ask is whether the perturbative forward model can be used to reconstruct the initial conditions. The simplest way would be to use perturbative inversion at the field level. This is along the lines of advanced reconstruction algorithms based on perturbative expansion, such as [77]. Alternatively, one can use full forward modeling, where the posterior for cosmological parameters is obtained by marginalizing over all amplitudes and phases of the initial conditions. The perturbative approach is useful in this context because it provides a simple, differentiable, forward model. In the future we plan to use our model for full field level inference.

It is worth mentioning that our analysis is based on a particular set of subgrid models used in the IllustrisTNG simulation. It is known that other hydrodynamical simulations show different baryonic effects [78] on the matter field. Therefore, our results may change if other simulation or different astrophysical models are used. However, the effect of baryons is typically important on smaller scales, outside the validity of perturbation theory. Therefore, while the particular values of transfer functions will change when using different simulations, we do not expect this to limit the method or change our main results. In the future, we plan to repeat our analysis for simulations with larger volumes, more redshifts, or different baryonic physics. This would lead to less noisy estimates of the transfer functions and provide an estimate of the impact of different galaxy formation scenarios.

APPENDIX A: TRANSFER FUNCTIONS IN REAL SPACE

In this appendix we derive the form of the real-space transfer functions used in the main text. The starting point for our forward model is

$$\delta_{HI}(\boldsymbol{k}) = \int d^3\boldsymbol{q}(1 + \delta_{HI}^{L}(\boldsymbol{q}))e^{-i\boldsymbol{k}\cdot(\boldsymbol{q}+\boldsymbol{\psi}(\boldsymbol{q}))}, \quad (A1)$$

where the HI field in Lagrangian coordinates $\boldsymbol{q}$ is expressed as

$$\begin{aligned}
\delta_{HI}^{L}(\boldsymbol{q}) &= b_1^{L}\delta_1(\boldsymbol{q}) + b_2^{L}[\delta_2(\boldsymbol{q}) - \sigma_1^2] + b_{\mathcal{G}_2}^{L}\mathcal{G}_2(\boldsymbol{q}) + b_3^{L}\delta_3(\boldsymbol{q}) \\
&+ b_{\mathcal{G}_2\delta}^{L}[\mathcal{G}_2\delta](\boldsymbol{q}) + b_{\mathcal{G}_3}^{L}\mathcal{G}_3(\boldsymbol{q}) + b_{\Gamma_3}^{L}\Gamma_3(\boldsymbol{q}) \\
&+ b_{\nabla^2}^{L}\nabla^2\delta_1(\boldsymbol{q}).
\end{aligned} \quad (A2)$$

The first line contains the linear and quadratic operators defined in the main text. Assuming the following form of bias operators,

$$\begin{aligned}
\mathcal{O}(\boldsymbol{k}) &= \int_{\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3}(2\pi)^3\delta^D(\boldsymbol{k} - \boldsymbol{p}_1 - \boldsymbol{p}_2 - \boldsymbol{p}_3) \\
&\times F_{\mathcal{O}}^s(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)\delta_1(\boldsymbol{p}_1)\delta_1(\boldsymbol{p}_2)\delta_1(\boldsymbol{p}_3), \quad (A3)
\end{aligned}$$

where $F_{\mathcal{O}}^s$ is a symmetrized kernel, the nontrivial cubic terms in the bias expansion have the following kernels:

$$F_{\mathcal{G}_3} = \frac{3}{2}\frac{(\boldsymbol{p}_1\cdot\boldsymbol{p}_2)^2}{p_1^2 p_2^2} - \frac{(\boldsymbol{p}_1\cdot\boldsymbol{p}_2)(\boldsymbol{p}_1\cdot\boldsymbol{p}_3)(\boldsymbol{p}_2\cdot\boldsymbol{p}_3)}{p_1^2 p_2^2 p_3^2} - \frac{1}{2}, \quad (A4)$$

$$F_{\Gamma_3} = \frac{4}{7}\left(1 - \frac{(\boldsymbol{p}_1\cdot\boldsymbol{p}_2)^2}{p_1^2 p_2^2}\right)\left(\frac{((\boldsymbol{p}_1+\boldsymbol{p}_2)\cdot\boldsymbol{p}_3)^2}{|\boldsymbol{p}_1+\boldsymbol{p}_2|^2 p_3^2} - 1\right). \quad (A5)$$

Finally, the last term in Eq. (A2) is a higher derivative bias. Even though it is linear in δ_1, this term is suppressed by the derivatives. Note that it has the same form as the leading order counterterm for the dark matter field.

The nonlinear displacement $\boldsymbol{\psi}(\boldsymbol{q})$ can be expanded in perturbation theory, too. The leading term is Zel'dovich displacement $\boldsymbol{\psi}_1(\boldsymbol{q})$. Being the largest contribution, it should be kept in the exponent, while higher order terms can be expanded. These higher order terms will mainly lead to operators which have the same form as in the bias expansion (but with fixed constants) and therefore can be absorbed in the bias coefficients. One exception is the second order shift acting on the HI density field, producing, at cubic order,

$$\mathcal{S}_3(\boldsymbol{q}) \equiv \boldsymbol{\psi}_2(\boldsymbol{q})\cdot\nabla\delta_1(\boldsymbol{q}). \quad (A6)$$

Such a term cannot be written as a linear combination of bias operators and has the following kernel:

$$F_{\mathcal{S}_3} = -\frac{3}{14}\left(1 - \frac{(\boldsymbol{p}_1\cdot\boldsymbol{p}_2)^2}{p_1^2 p_2^2}\right)\frac{(\boldsymbol{p}_1+\boldsymbol{p}_2)\cdot\boldsymbol{p}_3}{|\boldsymbol{p}_1+\boldsymbol{p}_2|^2}. \quad (A7)$$

The uniform density displaced by Zel'dovich displacement in Eq. (A1) is just the Zel'dovich field, which can be written in perturbation theory as

$$\delta_Z(\boldsymbol{k}) = \int d^3q\left(\delta_1 + \frac{1}{2}\mathcal{G}_2 - \frac{1}{3}\mathcal{G}_3\right)e^{-i\boldsymbol{k}\cdot(\boldsymbol{q}+\boldsymbol{\psi}_1)}, \quad (A8)$$

where all fields in the integrand depend on $\boldsymbol{q}$.

In conclusion, the real-space model can be expressed in terms of the shifted bias operators and $\mathcal{S}_3$. The transfer function in Eq. (9) is then given by the sum of the linear bias, the higher-derivative bias (and dark matter counterterm), and a set of all quadratic and cubic operators that correlate with $\tilde{\delta}_1$. Note that those do not include $\tilde{\mathcal{G}}_3$, $\tilde{\mathcal{G}}_2\delta$, and $\tilde{\delta}_3$ since their correlation with $\tilde{\delta}_1$ either vanishes or renormalizes the linear bias.

APPENDIX B: TRANSFER FUNCTIONS IN REDSHIFT SPACE

In this appendix we review the forward model in redshift space and give formulas for the transfer functions. In redshift space, the displacement in the exponent

contains an additional term, with projected velocity along the line of sight,

$$\delta_{\mathrm{HI}}^s(\boldsymbol{k},\mu) = \int d^3q(1+\delta_{\mathrm{HI}}^{\mathrm{L}}(\boldsymbol{q}))e^{-i\boldsymbol{k}\cdot\left(\boldsymbol{q}+\boldsymbol{\psi}(\boldsymbol{q})+\frac{\hat{z}\cdot\boldsymbol{\psi}(\boldsymbol{q})}{H}\hat{z}\right)}. \quad (B1)$$

Using the same bias model in Lagrangian space as before and expanding all nonlinear displacement terms in the exponent, we get the following expression [35]:

$$\delta_{\mathrm{HI}}^s(\boldsymbol{k},\hat{\boldsymbol{n}}) - \delta_Z^s(\boldsymbol{k},\boldsymbol{n}) + \frac{3}{7}f\tilde{\mathcal{G}}_2^{\|}(\boldsymbol{k},\boldsymbol{n})$$

$$= \int d^3q\Bigg[\delta_{\mathrm{HI}}^{\mathrm{L}} - \frac{3}{14}\mathcal{G}_2 - \frac{3}{14}(1+b_1^{\mathrm{L}})\delta_1\mathcal{G}_2 + \frac{1}{6}\Gamma_3 + \frac{1}{9}\mathcal{G}_3$$

$$- \frac{3}{7}fb_1^{\mathrm{L}}\delta_1\mathcal{G}_2^{\|} - \frac{5}{8}f\Gamma_3^{\|} + \frac{1}{3}f\mathcal{G}_3^{\|} - \frac{9}{14}f\mathcal{K}_3 - \frac{3}{14}f^2\delta_1\mathcal{G}_2^{\|}$$

$$- R_{ij}^{[2]}\psi_2^i\partial_j((1+b_1^L)\delta_1 + f\delta_1^{\|})\Bigg]e^{-i\boldsymbol{k}\cdot(\boldsymbol{q}+R^{[1]}\boldsymbol{\psi}_1)}, \quad (B2)$$

where the operators projected along the line of sight are defined in the following way:

$$\mathcal{O}^{\|}(\boldsymbol{q},z) \equiv \hat{z}^i\hat{z}^j\frac{\partial_i\partial_j}{\nabla^2}\mathcal{O}(\boldsymbol{q}), \quad (B3)$$

and the matrix $R^{[n]}$ is given by

$$R_{ij}^{[n]}(\hat{\boldsymbol{z}}) \equiv \delta_{ij} + nf\hat{z}_i\hat{z}_j. \quad (B4)$$

We also define

$$\mathcal{K}_3(\boldsymbol{q},z) \equiv \hat{z}_i\hat{z}_j\frac{\partial_i\partial_m}{\nabla^2}\delta_1(\boldsymbol{q})\frac{\partial_m\partial_j}{\nabla^2}\mathcal{G}_2(\boldsymbol{q}). \quad (B5)$$

The first line in Eq. (B2) contains the terms that we model in the main text as a linear combination of shifted operators. The second line is the same as in real space. The third line contains new cubic operators with projection along the line of sight. Finally, the last term is the anisotropic second order shift which acts on the linear field. Note that for the new cubic operators in redshift space, there are no new free parameters. This is consistent with the fact that velocities induced by gravity do not depend on the type of tracer.

Using Eq. (B2) it is easy to read off the form of the transfer function $\beta_1^{\mathrm{rsd}}(k,\mu)$ in redshift space. It has the following form:

$$\beta_1^{\mathrm{rsd}}(k,\mu) = \beta_1^{\mathrm{real}}(k) + \sum_i\frac{\langle\tilde{\delta}_1\tilde{\mathcal{O}}_i^{\|}\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle} - b_1\frac{\langle\tilde{\delta}_1\tilde{\mathcal{S}}_3^{\mathrm{rsd}}\rangle}{\langle\tilde{\delta}_1\tilde{\delta}_1\rangle}, \quad (B6)$$

where $\tilde{\mathcal{O}}_i^{\|}$ runs over all new cubic operators with projections along the line of sight, including the appropriate coefficients as given in Eq. (B2), and $\tilde{\mathcal{S}}_3^{\mathrm{rsd}}$ is the operator in the last line of Eq. (B2).

APPENDIX C: CONVERGENCE TEST WITH RESPECT TO GRID SIZE

In this appendix we demonstrate the convergence with respect to the grid size of our results on the scales we focus on. Throughout this work we use $N_{\text{mesh}} = 256$ grid cells per side (or equivalently the cell size of $0.8\ h^{-1}\ \text{Mpc}$) to analyze simulated and predicted HI fields. This choice determines the validity of our analysis. In order to test whether our results have converged, we repeat our analysis using different numbers of grid cells per side. In particular, we use lower ($N_{\text{mesh}} = 128$) and higher ($N_{\text{mesh}} = 512$) resolution and compare the resulting transfer function in Fig. 15, focusing on redshift $z = 1$. As expected, we find significant differences on very small scales and also when using a lower resolution grid. However, for the fiducial grid size and on scales $k < 1\ h\,\text{Mpc}^{-1}$ that we focus on in this work, we find our results have converged.
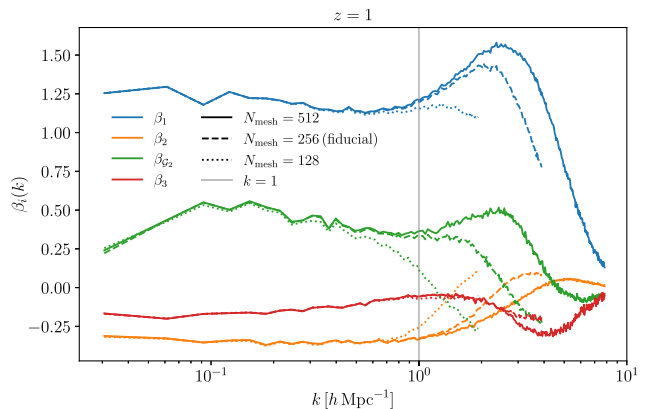

FIG. 15. Best-fit transfer functions $\beta_i(k)$ obtained from the simulated HI field at $z = 1$ for different choices of grid size N_{mesh}. We find that on relevant scales for this work, $k < 1\ h\,\text{Mpc}^{-1}$, our results converge with respect to the grid size for all transfer functions.

[1] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen *et al.* (DESI Collaboration), The DESI experiment part I: Science, targeting, and survey design, arXiv:1611.00036.

[2] L. Amendola, S. Appleby, A. Avgoustidis, D. Bacon, T. Baker, M. Baldi *et al.*, Cosmology and fundamental physics with the Euclid satellite, Living Rev. Relativity **21**, 2 (2018).

[3] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek *et al.*, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).

[4] S. Alam, M. Aubert, S. Avila, C. Balland, J. E. Bautista, M. A. Bershady *et al.*, 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**, 083533 (2021).

[5] S. Bharadwaj and S. K. Sethi, HI fluctuations at large redshifts: I–Visibility correlation, J. Astrophys. Astron. **22**, 293 (2001).

[6] S. Bharadwaj, B. B. Nath, and S. K. Sethi, Using HI to probe large scale structures at z ∼ 3, J. Astrophys. Astron. **22**, 21 (2001).

[7] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, Baryon acoustic oscillation intensity mapping of dark energy, Phys. Rev. Lett. **100**, 091303 (2008).

[8] J. B. Peterson, R. Aleksan, R. Ansari, K. Bandura, D. Bond, J. Bunton *et al.*, 21-cm intensity mapping, in astro2010: The Astronomy and Astrophysics Decadal Survey (2009), Vol. **2010**, p. 234, arXiv:0902.3091.

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

[10] M. Amiri, K. Bandura, A. Boskovic, T. Chen, J.-F. Cliche *et al.* (CHIME Collaboration), An overview of CHIME, the Canadian Hydrogen Intensity Mapping Experiment, Astrophys. J. Suppl. Ser. **261**, 29 (2022).

[11] L. B. Newburgh, K. Bandura, M. A. Bucher, T.-C. Chang, H. C. Chiang, J. F. Cliche *et al.*, HIRAX: A probe of dark energy and radio transients, in Ground-based and Airborne Telescopes VI Vol. 9906 (2016), (Proc. SPIE) p. 99065X, 10.1117/12.2234286.

[12] F. B. Abdalla, P. Bull, S. Camera, A. Benoit-Lévy, B. Joachimi, D. Kirk *et al.*, Cosmology from HI galaxy surveys with the SKA, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 17 (2015), arXiv:1501.04035.

[13] X. Chen, The Tianlai project: A 21 cm cosmology experiment, Int. J. Mod. Phys. Conf. Ser. **12**, 256 (2012).

[14] R. Ansari, E. J. Arena, K. Bandura, P. Bull, E. Castorina *et al.* (Cosmic Visions 21 cm Collaboration), Inflation and early dark energy with a stage II hydrogen intensity mapping experiment, arXiv:1810.09572.

[15] E. Castorina and F. Villaescusa-Navarro, On the spatial distribution of neutral hydrogen in the Universe: Bias and shot-noise of the H I power spectrum, Mon. Not. R. Astron. Soc. **471**, 1788 (2017).

[16] F. Villaescusa-Navarro, S. Genel, E. Castorina, A. Obuljen, D. N. Spergel, L. Hernquist *et al.*, Ingredients for 21 cm intensity mapping, Astrophys. J. **866**, 135 (2018).

[17] C. Modi, M. White, A. Slosar, and E. Castorina, Reconstructing large-scale structure with neutral hydrogen surveys, J. Cosmol. Astropart. Phys. 11 (2019) 023.

[18] C. Modi, E. Castorina, Y. Feng, and M. White, Intensity mapping with neutral hydrogen and the Hidden Valley simulations, J. Cosmol. Astropart. Phys. 09 (2019) 024.

[19] V. Desjacques, D. Jeong, and F. Schmidt, Large-scale galaxy bias, Phys. Rep. **733**, 1 (2018).

[20] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, Cosmological non-linearities as an effective fluid, J. Cosmol. Astropart. Phys. 07 (2012) 051.

[21] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, The effective field theory of cosmological large scale structures, J. High Energy Phys. 09 (2012) 082.

[22] L. Senatore and M. Zaldarriaga, The IR-resummed effective field theory of large scale structures, J. Cosmol. Astropart. Phys. 02 (2015) 013.

[23] L. Senatore, Bias in the effective field theory of large scale structures, J. Cosmol. Astropart. Phys. 11 (2015) 007.

[24] S. Foreman, H. Perrier, and L. Senatore, Precision comparison of the power spectrum in the EFTofLSS with simulations, J. Cosmol. Astropart. Phys. 05 (2016) 027.

[25] A. Perko, L. Senatore, E. Jennings, and R. H. Wechsler, Biased tracers in redshift space in the EFT of large-scale structure, arXiv:1610.09321.

[26] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, Snowmass white paper: Effective field theories in cosmology, Phys. Dark Universe **40**, 101193 (2023).

[27] D. Nelson, A. Pillepich, V. Springel, R. Weinberger, L. Hernquist, R. Pakmor et al., First results from the IllustrisTNG simulations: The galaxy colour bimodality, Mon. Not. R. Astron. Soc. **475**, 624 (2018).

[28] F. Marinacci, M. Vogelsberger, R. Pakmor, P. Torrey, V. Springel, L. Hernquist et al., First results from the IllustrisTNG simulations: Radio haloes and magnetic fields, Mon. Not. R. Astron. Soc. **480**, 5113 (2018).

[29] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist et al., First results from the IllustrisTNG simulations: Matter and galaxy clustering, Mon. Not. R. Astron. Soc. **475**, 676 (2018).

[30] A. Pillepich, D. Nelson, L. Hernquist, V. Springel, R. Pakmor, P. Torrey et al., First results from the IllustrisTNG simulations: The stellar mass content of groups and clusters of galaxies, Mon. Not. R. Astron. Soc. **475**, 648 (2018).

[31] J. P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger et al., First results from the IllustrisTNG simulations: A tale of two elements—chemical evolution of magnesium and europium, Mon. Not. R. Astron. Soc. **477**, 1206 (2018).

[32] T. Baldauf, E. Schaan, and M. Zaldarriaga, On the reach of perturbative descriptions for dark matter displacement fields, J. Cosmol. Astropart. Phys. 03 (2016) 017.

[33] T. Baldauf, E. Schaan, and M. Zaldarriaga, On the reach of perturbative methods for dark matter density fields, J. Cosmol. Astropart. Phys. 03 (2016) 007.

[34] M. Schmittfull, M. Simonović, V. Assassi, and M. Zaldarriaga, Modeling biased tracers at the field level, Phys. Rev. D **100**, 043514 (2019).

[35] M. Schmittfull, M. Simonović, M. M. Ivanov, O. H. E. Philcox, and M. Zaldarriaga, Modeling galaxies in redshift space at the field level, J. Cosmol. Astropart. Phys. 05 (2021) 059.

[36] M. McQuinn and A. D'Aloisio, The observable 21 cm signal from reionization may be perturbative, J. Cosmol. Astropart. Phys. 10 (2018) 016.

[37] W. Qin, K. Schutz, A. Smith, E. Garaldi, R. Kannan, T. R. Slatyer, and M. Vogelsberger, An effective bias expansion for 21 cm cosmology in redshift space, Phys. Rev. D **106**, 123506 (2022).

[38] D. Wadekar, F. Villaescusa-Navarro, S. Ho, and L. Perreault-Levasseur, HInet: Generating neutral hydrogen from dark matter with neural networks, Astrophys. J. **916**, 42 (2021).

[39] M. Bernardini, R. Feldmann, D. Anglés-Alcázar, M. Boylan-Kolchin, J. Bullock, L. Mayer et al., From EMBER to FIRE: Predicting high resolution baryon fields from dark matter simulations with deep learning, Mon. Not. R. Astron. Soc. **509**, 1323 (2022).

[40] A. Chudaykin, M. M. Ivanov, O. H. E. Philcox, and M. Simonović, Nonlinear perturbation theory extension of the Boltzmann code CLASS, Phys. Rev. D **102**, 063533 (2020).

[41] G. D'Amico, L. Senatore, and P. Zhang, Limits on wCDM from the EFTofLSS with the PyBird code, J. Cosmol. Astropart. Phys. 01 (2021) 006.

[42] S.-F. Chen, Z. Vlah, and M. White, Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory, J. Cosmol. Astropart. Phys. 07 (2020) 062.

[43] M. Peloso and M. Pietroni, Galilean invariance and the consistency relation for the nonlinear squeezed bispectrum of large scale structure, J. Cosmol. Astropart. Phys. 05 (2013) 031.

[44] A. Kehagias and A. Riotto, Symmetries and consistency relations in the large scale structure of the universe, Nucl. Phys. **B873**, 514 (2013).

[45] P. Creminelli, J. Noreña, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure, J. Cosmol. Astropart. Phys. 12 (2013) 025.

[46] P. Creminelli, J. Gleyzes, M. Simonović, and F. Vernizzi, Single-field consistency relations of large scale structure. Part II: Resummation and redshift space, J. Cosmol. Astropart. Phys. 02 (2014) 051.

[47] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, Equivalence principle and the baryon acoustic peak, Phys. Rev. D **92**, 043514 (2015).

[48] Z. Vlah, U. Seljak, M. Y. Chu, and Y. Feng, Perturbation theory, effective field theory, and oscillations in the power spectrum, J. Cosmol. Astropart. Phys. 03 (2016) 057.

[49] D. Blas, M. Garny, M. M. Ivanov, and S. Sibiryakov, Time-sliced perturbation theory II: Baryon acoustic oscillations and infrared resummation, J. Cosmol. Astropart. Phys. 07 (2016) 028.

[50] T. Matsubara, Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture, Phys. Rev. D **78**, 083519 (2008).

[51] J. Carlson, B. Reid, and M. White, Convolution Lagrangian perturbation theory for biased tracers, Mon. Not. R. Astron. Soc. **429**, 1674 (2013).

[52] Z. Vlah, E. Castorina, and M. White, The Gaussian streaming model and convolution Lagrangian effective field theory, J. Cosmol. Astropart. Phys. 12 (2016) 007.

[53] S.-F. Chen, Z. Vlah, E. Castorina, and M. White, Redshift-space distortions in Lagrangian perturbation theory, J. Cosmol. Astropart. Phys. 03 (2021) 100.

[54] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al. (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594**, A13 (2016).

[55] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (CRC Press, 1988), 10.1201/9780367806934.

[56] T. Lazeyras, C. Wagner, T. Baldauf, and F. Schmidt, Precision measurement of the local bias of dark matter halos, J. Cosmol. Astropart. Phys. 02 (2016) 018.

[57] M. M. Abidi and T. Baldauf, Cubic halo bias in Eulerian and Lagrangian space, J. Cosmol. Astropart. Phys. 07 (2018) 029.

[58] A. Barreira, G. Cabass, F. Schmidt, A. Pillepich, and D. Nelson, Galaxy bias and primordial non-Gaussianity: Insights from galaxy formation simulations with IllustrisTNG, J. Cosmol. Astropart. Phys. 12 (2020) 013.

[59] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell, All-sky interferometry with spherical harmonic transit telescopes, Astrophys. J. **781**, 57 (2014).

[60] J. R. Shaw, K. Sigurdson, M. Sitwell, A. Stebbins, and U.-L. Pen, Coaxing cosmic 21 cm fluctuations from the polarized sky using m-mode analysis, Phys. Rev. D **91**, 083514 (2015).

[61] R. Byrne, M. F. Morales, B. Hazelton, W. Li, N. Barry, A. P. Beardsley, R. Joseph, J. Pober, I. Sullivan, and C. Trott, Fundamental limitations on the calibration of redundant 21 cm cosmology instruments and implications for HERA and the SKA, Astrophys. J. **875**, 70 (2019).

[62] M. Spinelli, I. P. Carucci, S. Cunnington, S. E. Harper, M. O. Irfan, J. Fonseca et al., SKAO H I intensity mapping: Blind foreground subtraction challenge, Mon. Not. R. Astron. Soc. **509**, 2048 (2022).

[63] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs, and D. F. Moore, A per-baseline, delay-spectrum technique for accessing the 21 cm cosmic reionization signature, Astrophys. J. **756**, 165 (2012).

[64] J. C. Pober, The impact of foregrounds on redshift space distortion measurements with the highly redshifted 21-cm line, Mon. Not. R. Astron. Soc. **447**, 1705 (2015).

[65] A. Liu, A. R. Parsons, and C. M. Trott, Epoch of reionization window. I. Mathematical formalism, Phys. Rev. D **90**, 023018 (2014).

[66] A. Liu, A. R. Parsons, and C. M. Trott, Epoch of reionization window. II. Statistical methods for foreground wedge reduction, Phys. Rev. D **90**, 023019 (2014).

[67] A. Obuljen, E. Castorina, F. Villaescusa-Navarro, and M. Viel, High-redshift post-reionization cosmology with 21 cm intensity mapping, J. Cosmol. Astropart. Phys. 05 (2018) 004.

[68] S.-F. Chen, E. Castorina, M. White, and A. Slosar, Synergies between radio, optical and microwave observations at high redshift, J. Cosmol. Astropart. Phys. 07 (2019) 023.

[69] O. Leicht, C. Uhlemann, F. Villaescusa-Navarro, S. Codis, L. Hernquist, and S. Genel, Extreme spheres: Counts-in-cells for 21 cm intensity mapping, Mon. Not. R. Astron. Soc. **484**, 269 (2019).

[70] U. Seljak, N. Hamaus, and V. Desjacques, How to suppress the shot noise in galaxy surveys, Phys. Rev. Lett. **103**, 091303 (2009).

[71] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith, and T. Baldauf, Minimizing the stochasticity of halos in large-scale structure surveys, Phys. Rev. D **82**, 043515 (2010).

[72] C. Modi, E. Castorina, and U. Seljak, Halo bias in Lagrangian space: Estimators and theoretical predictions, Mon. Not. R. Astron. Soc. **472**, 3959 (2017).

[73] G. Cabass, M. Simonović, and M. Zaldarriaga, Cosmological information in perturbative forward modeling, arXiv:2307.04706.

[74] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological parameters from the BOSS galaxy power spectrum, J. Cosmol. Astropart. Phys. 05 (2020) 042.

[75] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein, Solving large scale structure in ten easy steps with COLA, J. Cosmol. Astropart. Phys. 06 (2013) 036.

[76] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, FASTPM: A new scheme for fast simulations of dark matter and haloes, Mon. Not. R. Astron. Soc. **463**, 2273 (2016).

[77] M. Schmittfull, T. Baldauf, and M. Zaldarriaga, Iterative initial condition reconstruction, Phys. Rev. D **96**, 023505 (2017).

[78] N. E. Chisari, A. J. Mead, S. Joudaki, P. G. Ferreira, A. Schneider, J. Mohr et al., Modelling baryonic feedback for survey cosmology, Open J. Astrophys. **2**, 4 (2019).

[79] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, nbodykit: An open-source, massively parallel toolkit for large-scale structure, Astron. J. **156**, 160 (2018).

[80] F. Perez and B. E. Granger, IPython: A system for interactive scientific computing, Comput. Sci. Eng. **9**, 21 (2007).

[81] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[82] S. van der Walt, S. C. Colbert, and G. Varoquaux, The NumPy array: A structure for efficient numerical computation, Comput. Sci. Eng. **13**, 22 (2011).

[83] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., Array programming with NumPy, Nature (London) **585**, 357 (2020).

[84] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[85] www.s3it.uzh.ch.