

## Primordial non-Gaussianity and non-Gaussian covariance

Thomas Flöss<sup>\*</sup>

*Van Swinderen Institute for Particle Physics and Gravity, University of Groningen,  
Nijenborgh 4, 9747 AG Groningen, Netherlands  
and Kapteyn Astronomical Institute, University of Groningen, 9700 AV Groningen, Netherlands*

Matteo Biagetti

*Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy;  
International School for Advanced Studies (SISSA), 34136 Trieste, Italy;  
INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, 34143 Trieste, Italy  
and INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste, Italy*

P. Daniel Meerburg

*Van Swinderen Institute for Particle Physics and Gravity, University of Groningen,  
Nijenborgh 4, 9747 AG Groningen, Netherlands*

 (Received 21 July 2022; accepted 2 December 2022; published 30 January 2023)

In the pursuit of primordial non-Gaussianities, we hope to access smaller scales across larger comoving volumes. At low redshift, the search for primordial non-Gaussianities is hindered by gravitational collapse, with which we often associate a scale  $k_{\text{NL}}$ . Beyond these scales, it will be hard to reconstruct the modes sensitive to the primordial distribution. When forecasting future constraints on the amplitude of primordial non-Gaussianity,  $f_{\text{NL}}$ , off-diagonal components are usually neglected in the covariance because they are small compared to the diagonal components. We show that the induced non-Gaussian off-diagonal components in the covariance degrade forecast constraints on primordial non-Gaussianity, even when all modes are well within what is usually considered the linear regime. As a testing ground, we examine the effects of these off-diagonal components on the constraining power of the matter bispectrum on  $f_{\text{NL}}$  as a function of  $k_{\text{max}}$  and redshift, confirming our results against  $N$ -body simulations out to redshift  $z = 10$ . We then consider these effects on the hydrogen bispectrum as observed from a PUMA-like 21-cm intensity mapping survey at redshifts  $2 < z < 6$  and show that not including off-diagonal covariance overpredicts the constraining power on  $f_{\text{NL}}$  by up to a factor of 5. For future surveys targeting even higher redshifts, such as Cosmic Dawn and the Dark Ages, which are considered ultimate surveys for primordial non-Gaussianity, we predict that non-Gaussian covariance would severely limit prospects to constrain  $f_{\text{NL}}$  from the bispectrum.

DOI: [10.1103/PhysRevD.107.023528](https://doi.org/10.1103/PhysRevD.107.023528)

### I. INTRODUCTION

Over the past few decades, inflation has been established as the leading paradigm for describing the early Universe. It proposes a period of rapidly accelerated expansion during the first fraction of a second after the Universe came to be [1–3]. At the classical level, such an expansion can explain why the Universe looks nearly identical in every direction (i.e., it is homogeneous and isotropic), while at the quantum level it gives rise to the tiny density fluctuations that we observe in the cosmic microwave background radiation (CMB), which eventually grow into the large-scale structure (LSS) of the Universe. By precisely mapping the anisotropies in the CMB, we have determined the

fluctuations to be very close to Gaussian-distributed, which matches the predictions of even the simplest theories of inflation [4]. However, in order to sift through the vast landscape of consistent inflationary theories, we are required to look beyond such general predictions. One avenue to discriminate theories of inflation, is through the study of primordial non-Gaussianities (pnGs) (see Ref. [5] and references therein). Signatures of pnGs would appear as nonzero higher  $n$ -point functions of the initial conditions, where the three-point function, the so-called *bispectrum*, is generally the most sensitive. A measurement of pnGs can tell us a great deal about the dynamics driving the expansion (see Ref. [6] for a recent overview). To give an example, a detection of a local-type pnG in future experiments would exclude single-field models of inflation altogether [7,8], while generic values of  $f_{\text{NL}}^{\text{loc}}$  in multifield

<sup>\*</sup>t.s.floss@rug.nl

models are  $\gtrsim O(1)$  or higher. On the other hand, a detection of an equilateral-type pnG of order  $f_{\text{NL}}^{\text{eq}} \geq 1$  would imply strong coupling of the inflaton, since it is typically generated by self-interactions of the inflaton field. Furthermore, particles (fields) present in the primordial Universe leave their unique imprint in the distribution of fluctuations through pnGs, effectively making inflation a particle collider at the highest conceivable energy scale [9]. Hence, a detailed study of primordial non-Gaussianity is imperative in order to advance our understanding of the Universe as a whole.

While the most stringent constraints on pnGs are derived from measurements of the CMB bispectrum, future CMB experiments will be limited by its two-dimensional nature and its damping of primary fluctuations. In our search for signatures of pnGs, we are therefore required to look for alternative probes. Surveys of the large-scale structure of the Universe provide us with a huge observable volume, all the way into the cosmological Dark Ages, by mapping the distribution of galaxies and neutral hydrogen. While the anisotropies in the CMB are pristine (i.e., linearly related to the primordial fluctuations), the density field has since evolved. Gravity, being intrinsically nonlinear, breaks the linear relation between density fluctuations and primordial initial conditions, giving rise to a number of complications. First, even if the primordial fluctuations are purely Gaussian, the nonlinear gravitational evolution introduces *secondary* non-Gaussianities (snGs), typically many orders of magnitude stronger than any primordial signal. Thus, an accurate modeling of snGs is required in order to properly extract information about pnGs. Furthermore, snGs introduce non-Gaussian covariance in the measurements, reducing the amount of *unique* information present in the data. Although the impact of non-Gaussian covariance has been appreciated at low redshifts [10–19], its relevance for high-redshift surveys has typically been neglected [20–26]. As upcoming surveys aim to close in on important thresholds such as  $f_{\text{NL}}^{\text{loc}} \sim 1$ , it is important to assess the assumptions made in forecasting their performance. In this paper, we show that by not including non-Gaussian covariance in forecasts of the constraining power of the hydrogen bispectrum observed by a PUMA-like 21-cm intensity mapping experiment [23,27], one can underestimate the uncertainty in the linear regime by up to factors of  $\sim 5$  and  $\sim 2$  for local- and equilateral-type non-Gaussianity, respectively.

*Conventions and Notation:* We denote a spatial vector as  $\mathbf{k}_i$  and its magnitude as  $|\mathbf{k}_i| = k_i$ . Sums of momenta are written as  $\mathbf{k}_{1\dots n} = \sum_1^n \mathbf{k}_i$ ; e.g.,  $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_{12}$ . Momentum integrals are compactly written as  $\int \frac{d^3 \mathbf{k}_i}{(2\pi)^3} = \int_{\mathbf{k}_i}$ , and  $\delta_D$  denotes the Dirac delta function. In order to compare to simulations of the matter bispectrum, our cosmology equals the fiducial cosmology of the QUIJOTE suite [28], which closely resembles the 2018 Planck constraints [29]. For the analysis of the PUMA survey, we use the 2015 Planck constraints [30] to match previous forecasts.

## II. THEORETICAL FRAMEWORK AND SETUP

In order to estimate the signal-to-noise ratio for high-redshift survey observables, we need to introduce a few concepts. We are ultimately interested in constraining the early Universe through primordial non-Gaussianities; thus, we start off by defining correlations of primordial fluctuations—i.e., our signal of interest. Next, we introduce density perturbations, whose correlations at different positions in the sky are the building blocks of what we actually observe in high- (and low-) redshift surveys. Their dynamics driven by gravity determine the noise we need to overcome.

### A. Initial conditions

Quantum fluctuations during the inflationary epoch cause the expansion to end at slightly different times in different places, giving rise to tiny scalar density fluctuations  $\zeta$  that source linear perturbations in the matter density field. In this way, linear fluctuations of the density field trace the primordial initial conditions of the Universe. Even a small non-Gaussianity in the distribution of primordial fluctuations serves as an important way to discriminate between different models of inflation. Furthermore, it allows one to directly probe the particle content and interactions of the inflationary epoch [9,31]. Since such non-Gaussianities are constrained to be small by CMB observations [4], in this work we consider only the first non-Gaussian statistic, which is the bispectrum. Hence, we require only the first two statistical moments of the primordial density distributions. In Fourier space, these are the power spectrum  $P_\zeta(k)$  and bispectrum  $B_\zeta(k_1, k_2, k_3)$ , defined as

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\zeta(k_1), \quad (1)$$

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\zeta(k_1, k_2, k_3). \quad (2)$$

Different inflationary mechanisms give rise to distinct sizes and shapes of bispectra. It is customary to classify these bispectra into three main templates, the so-called local, equilateral, and orthogonal templates, whose expressions are given in Appendix A. The local shape typically arises in models of multifield inflation and peaks in squeezed triangle configurations  $k_1 \ll k_2 \sim k_3$ , while the equilateral shape peaks for equilateral configurations  $k_1 = k_2 = k_3$ . Finally, the orthogonal shape, along with the equilateral one, is a natural prediction of the effective field theory (EFT) of (single-field) inflation [32] and peaks for both equilateral and flattened configurations  $k_1 = k_2 + k_3$ .

### B. Matter field and correlators

The primordial initial conditions serve as the seed for the distribution of matter in the Universe. We can therefore study the initial conditions of the Universe by studying fluctuations of the matter density field,  $\rho$ , defined as

$\delta(t, \mathbf{x}) = \rho(t, \mathbf{x})/\bar{\rho}(t) - 1$ , with  $\bar{\rho}$  being the mean density in a volume. Similar to the primordial case, we define correlations of  $\delta(\mathbf{x})$  in Fourier space as

$$\langle \delta_{k_1} \delta_{k_2} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_\delta(k_1), \quad (3)$$

$$\langle \delta_{k_1} \delta_{k_2} \delta_{k_3} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \quad (4)$$

$$\langle \delta_{k_1} \delta_{k_2} \delta_{k_3} \delta_{k_4} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{1234}) T_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (5)$$

where we assume all fluctuations to be at equal times. In this work, we also need the four-point correlation function in Fourier space, known as the *trispectrum*, for the computation of the non-Gaussian covariance. Even in the absence of a primordial bispectrum, or higher-order primordial correlators, fluctuations in the matter field grow via gravitational instability and become nonlinear, thereby sourcing the matter bispectrum, trispectrum, and higher-order correlations. The dynamical equations for  $\delta$  describing this process can be solved perturbatively (see, e.g., Ref. [33] for a review). This allows one to compute correlators analytically up to a mildly nonlinear scale  $k_{\text{NL}}$ . One way to estimate this scale is by computing

$$k_{\text{NL}}(z) = \left[ \frac{1}{6\pi^2} \int_0^\infty dk P_\delta^L(k, z) \right]^{-1/2}, \quad (6)$$

where  $P_\delta^L$  is the linear matter power spectrum as defined in Eq. (B4). We use this scale to confine ourselves to the linear regime [34]. The gravitationally induced bispectrum and trispectrum in this framework are presented in Appendix B.

To complement the perturbative approach, we resort to  $N$ -body simulations of the Universe at large scales, solving the dynamical equations for  $\delta$  numerically (see Ref. [36] for a review). The advantage of  $N$ -body simulations is that they allow us to directly measure correlations of  $\delta$  even at nonlinear scales, and to test analytic predictions. The drawback is that they are computationally expensive to run. We make use of publicly available QUIJOTE simulations [28] for our estimates of signal-to-noise ratio at low redshift (i.e., up to  $z = 3$ ). For higher redshifts, as the nonlinear scale is pushed to very small scales, instead of fully solving dynamical equations we resort to Monofonic [37], which computes particle positions by solving third-order Lagrangian perturbation theory (3LPT) equations. Further details on how simulation data are used can be found in Appendix C.

### C. Fisher information and estimated uncertainty

In this section, we introduce the quantities we use to estimate the uncertainty on the amplitude of primordial non-Gaussianity,  $f_{\text{NL}}$ , from observations of the bispectrum.

#### 1. Fisher matrix

A common way to quantify the information content of an observable is through the Fisher matrix. It encodes both the amount of information available from a measurement to constrain a parameter and the correlation between different parameters. Given  $N$  measurements of an observable, which for us will be the matter or hydrogen bispectrum, and a set of parameters we want to constrain,  $\mathbf{p}$ , the Fisher matrix is defined as

$$F_{ab} = \sum_{TT'} \frac{\partial B_T}{\partial p_a} (C)_{TT'}^{-1} \frac{\partial B_{T'}}{\partial p_b}, \quad (7)$$

where  $T$ 's are triangle configurations in which the bispectra are measured or calculated,  $\mathbf{B}$  is the data vector of the bispectra, and  $C_{TT'}$  is the covariance of  $\mathbf{B}$ , defined as

$$C_{TT'} = \langle B_T B_{T'} \rangle - \langle B_T \rangle \langle B_{T'} \rangle. \quad (8)$$

The estimated uncertainty on a parameter  $p_a$  is then defined as

$$\sigma_{p_a} = (F^{-1})_{aa}^{1/2}, \quad (9)$$

where  $F^{-1}$  indicates the matrix inverse of  $F$ .

#### 2. $N$ -body measurements

When using numerical simulations, we measure the matter bispectrum on a finite-sized box with periodic boundary conditions, such that in this case  $\delta_{\mathbf{k}}$  is a discrete Fourier transform of the density contrast. The bispectrum estimator then is defined as [38]

$$\hat{B}(k_1, k_2, k_3) \equiv \frac{k_F^3}{N_{tr}} \sum_{\mathbf{q} \in k} \delta_K(\mathbf{q}_{123}) \delta_{\mathbf{q}_1} \delta_{\mathbf{q}_2} \delta_{\mathbf{q}_3}, \quad (10)$$

with  $k_F = 2\pi/L$  being the fundamental frequency in a cubic box of side  $L$ , and  $N_{tr}$  gives the number of ‘‘fundamental triangles’’ formed by the vectors  $\mathbf{q}_i$  satisfying the condition  $\mathbf{q}_{123} = 0$  that belong in the ‘‘triangle bin’’ defined by the triplet of bin centers  $(k_1, k_2, k_3)$  and bin width  $\Delta k$  [39].

The advantage of using  $N$ -body simulations is that the full covariance can be estimated numerically from a sample of simulations using Eq. (8), where now the average  $\langle \cdot \rangle$  is over different realizations of the same simulation.

It is also straightforward to compute the Gaussian contribution only—i.e., the case where different modes are uncorrelated. This contribution is given by the product of three power spectra [40],

$$C_{TT'}^G \simeq \frac{(2\pi)^3 k_F^3}{V_{123}} s_{123} \hat{P}(k_1) \hat{P}(k_2) \hat{P}(k_3) \delta_{TT'}, \quad (11)$$

where  $T, T'$  denote triangle bins,  $V_{123} \simeq 8\pi^2 k_1 k_2 k_3 \Delta k^3$  is the volume of the bin,  $s_{123} = 1, 2,$  and  $6$  for scalene, isosceles, and equilateral triangles, respectively, and  $\hat{P}(k_i)$ 's are power spectrum measurements.

### 3. Limit of infinitely thin bins

At high redshifts, the nonlinear scale  $k_{\text{NL}}$  is pushed to smaller scales. At fixed bin width  $\Delta k$ , this implies a wider range of scales explored, and consequently a larger data vector and covariance. In order to keep the calculations within reasonable computational cost, one solution is to widen the range of bins, and to sample wave numbers in log space. Alternatively, we choose to approach the limit of infinitely thin bins and promote the sums to integrals, such that the Fisher matrix becomes

$$F_{ab} = \int_{TT'} \frac{\partial B_T}{\partial p_a} C_{TT'}^{-1} \frac{\partial B_{T'}}{\partial p_b}, \quad (12)$$

where now the matter, or hydrogen, bispectrum is estimated using perturbation theory, as explained in the appendixes. Calculating Eq. (12) now implies knowledge of the dependence on triangle configurations  $T, T'$  of the inverted full covariance matrix, which is typically hard to compute. In Appendix D, we outline a strategy that is based on splitting the covariance into Gaussian and non-Gaussian contributions,  $C = C_G + C_{\text{nG}}$ , and expanding the inverse as a Neumann series. We then approximate this series such that the Fisher matrix in the limit of thin bins becomes

$$F_{ab} = \frac{(F_{ab}^G)^2}{F_{ab}^G + \delta F_{ab}^{\text{nG}}}, \quad (13)$$

where  $F_{ab}^G$  is the Fisher matrix computed using only the Gaussian covariance  $C_G$  to compute the inverse covariance, and  $\delta F_{ab}^{\text{nG}}$  is the non-Gaussian correction computed using as inverse the product of matrices  $-C_G^{-1} C_{\text{nG}} C_G^{-1}$ .

### 4. Model for non-Gaussian covariance

The goal of this paper is to compute how  $\sigma$  varies for  $f_{\text{NL}}$  whether we are considering only the Gaussian term  $C_G$  or a more complete modeling of the covariance including non-Gaussian terms. As explained above, when using  $N$ -body simulations, the Gaussian and the full covariance are computed numerically. In the case of thin bins, we need to introduce a model of the bispectrum covariance. Inserting Eq. (10) into Eq. (8), the computation involves the correlator of six fields in Fourier space, which can be combined in four different ways: the Gaussian term is the product of three power spectra (the PPP term), given by Eq. (11). Non-Gaussian terms are represented by either the product of two bispectra (the BB term), or the product of a power spectrum and a trispectrum (the PT term), or finally the connected six-point function, the so-called *pentaspectrum*.

The pentaspectrum is negligible in most practical cases (see Ref. [18] for a rough estimate). The key point of this paper is to account for the BB and PT terms in signal-to-noise estimates at high redshifts. The BB term, again assuming that correlators are slowly varying in the  $k$ -shells, can be written as

$$C_{\text{nG}}^{\text{BB}} \simeq B_T B_{T'} (\Sigma_{TT'}^{11} + 8 \text{ perm}), \quad (14)$$

where  $\Sigma_{TT'}^{ij}$  is a mode-counting factor that again depends on the shape of the triangles. The PT term is similarly written. We calculate these terms for the matter bispectrum predictions using Eqs. (B4), (B3), and (B5). For the hydrogen bispectrum, we use the following model for the covariance:

$$C \approx C_G + 2C_{\text{nG}}^{\text{BB}}, \quad (15)$$

where the PT term is approximated to be equal to the BB term, which is a good approximation for squeezed triangles for which the non-Gaussian terms are largest [18].

## III. CONSTRAINING $f_{\text{NL}}$ AT HIGH REDSHIFTS

The primary goal of this work is to show the importance of including non-Gaussian terms in the covariance when estimating the uncertainty to the primordial non-Gaussian amplitude  $f_{\text{NL}}$  in high-redshift surveys.

One could be tempted to neglect the non-Gaussian covariance at high redshifts on scales larger than  $k_{\text{NL}}$  at that redshift. In this linear regime, one might expect modes of different wavelength to be mostly uncorrelated, such that the covariance is diagonal and Gaussian terms dominate. As we show in what follows, this intuition fails: off-diagonal terms become important well within what is typically considered the linear regime based on Eq. (6).

### A. Uncertainty on $f_{\text{NL}}$ from the matter bispectrum

As a testing ground, we first consider the matter bispectrum in real space as our observable and compute the estimated uncertainty of the primordial non-Gaussian amplitude  $f_{\text{NL}}$  for the primordial bispectra of the local, equilateral, and orthogonal types as defined previously. In this setup,  $f_{\text{NL}}$  is the only parameter. When using finite-sized bins, we evaluate the derivative  $\partial \mathbf{B} / \partial f_{\text{NL}}$  averaging over the bins [see Eq. (C1) of Appendix C], while in the case of infinitely thin bins, the derivative is analytically computed directly from the templates in Eqs. (A1)–(A3).

Figure 1 shows the ratio of the estimated uncertainty computed with non-Gaussian over Gaussian covariance as a function of the maximum wave number  $k_{\text{max}}$ . The uncertainty computed using the infinitely thin bin approximation is shown with solid lines, while simulation measurements are shown with diamonds. Solid lines are computed up to the nonlinear scale  $k_{\text{NL}}$  at that redshift. The uncertainty on local-type non-Gaussianity is most

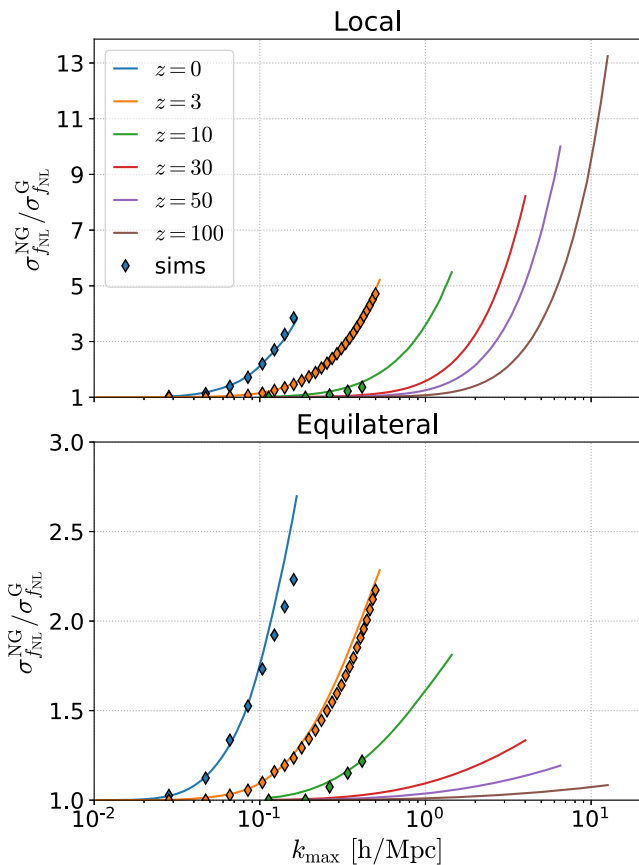


FIG. 1. Estimate of relative increase in error on  $f_{\text{NL}}$  due to non-Gaussian covariance as a function of  $k_{\text{max}}$ . The diamonds present results obtained using the QUIJOTE simulations ( $z = 0, 3$ ) or 3LPT ( $z = 10$ ). Note the different scales on the vertical axes. The local bispectrum is expected to be significantly affected when accounting for nondiagonal covariance even at very high redshifts. Solid lines are estimated up to the nonlinear scale  $k_{\text{NL}}$  at each redshift. For  $z = 0$  and  $3$ , the simulation results are also shown up to the nonlinear scale, while for  $z = 10$  they are shown up to the scale at which shot noise becomes a significant contribution to the covariance.

affected by the introduction of off-diagonal covariance, increasing by a factor of  $\sim 5$  at  $k_{\text{max}} \approx k_{\text{NL}}$  for redshifts lower than  $z = 10$  and even higher at higher redshifts. This is because the off-diagonal covariance is the largest for squeezed triangle configurations, where the local-type non-Gaussianity has most of its signal [18,41]. Equilateral non-Gaussianity is less affected, since most of its signal comes from equilateral triangle configurations whose covariance is large only when approaching nonlinear scales. Still, the loss is almost a factor of 2 at  $z \lesssim 10$ . For a discussion and the results of the orthogonal bispectrum, we refer to Appendix C.

It is important to note that these results do not imply that the uncertainty does not improve overall, since we are still able to access more modes as we increase  $k_{\text{max}}$ . Rather, our analysis shows that off-diagonal non-Gaussian covariance reduces the amount of information gained by probing

smaller scales, or in other words, the signal-to-noise saturates. We show a clear representation of this fact in Fig. 2, where we directly plot the uncertainty of  $f_{\text{NL}}$ , including non-Gaussian terms, as a function of  $k_{\text{max}}$  at different redshifts for a fictitious matter field survey.

## B. Uncertainty on $f_{\text{NL}}$ from the hydrogen bispectrum

To make contact with actual future observations, we consider a realistic PUMA-like intensity mapping survey setup. PUMA is a proposed 21-cm intensity mapping experiment aimed at measuring the distribution of neutral hydrogen through the 21-cm hyperfine transition between redshifts 2 and 6. One of the key science drivers for PUMA is to provide better constraints on primordial non-Gaussianity with respect to the CMB [27] (see also Fig. 5 in Ref. [6] for a comparison to other future surveys).

As compared to the simplified scenario considered in Fig. 1, the calculation of the estimated uncertainty in this case involves several complications. First of all, neutral

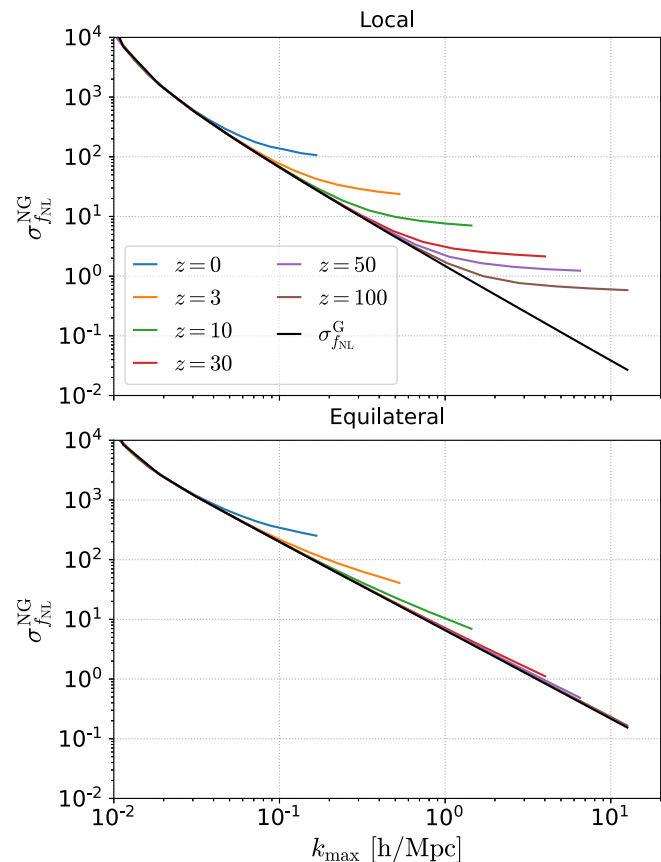


FIG. 2. The estimated uncertainty on  $f_{\text{NL}}$  as a function of  $k_{\text{max}}$  in the matter field at different redshifts, when including non-Gaussian covariance. The volume of the survey is taken to be  $1 \text{ (Gpc/h)}^3$ . Each redshift is shown up to the corresponding nonlinear scale  $k_{\text{NL}}$ . The solid black line shows the (redshift-independent) uncertainty assuming only Gaussian covariance.

hydrogen is a biased tracer of the matter field. This introduces additional nonlinearities, and we need to define a set of nuisance (bias) parameters that are fixed through observations (see Ref. [43] for a review). Second, we need to compute correlators in redshift space, taking into account the survey geometry and foregrounds. Lastly, the presence of primordial non-Gaussianity introduces additional bias parameters. This last effect famously appears already at the power spectrum level for the local template, known as scale-dependent bias (see Refs. [44–46] and Ref. [47] for a recent review). For this reasons, forecasts of  $\sigma(f_{\text{NL}})$  depend sensitively on many assumptions, and would need to include the tracer power spectrum in order to be realistic. Here we limit ourselves to calculate the uncertainty using the tracer bispectrum only, rather than performing a full forecast, since our goal is to show the loss of constraining power due to the inclusion of non-Gaussian covariance on the bispectrum [48]. For our computations, we follow the setup presented in Ref. [23]. The 21-cm intensity mapping is complicated by foregrounds, especially on large scales in the line-of-sight direction. Therefore, the largest scale is effectively set by a foreground cut in the line-of-sight direction ( $k_{\parallel, \text{min}} = 0.01 \text{ h/Mpc}$ ). The analysis is limited to linear scales by choosing  $k_{\text{max}}(z)$  to be fractions of  $k_{\text{NL}}(z)$ , namely  $0.5k_{\text{NL}}(z)$  and  $0.75k_{\text{NL}}(z)$ , where the nonlinear scale is given in Eq. (6). We also include the foreground wedge, further reducing the amount of modes. Besides  $f_{\text{NL}}$ , the hydrogen bispectrum is a function of seven parameters: three bias parameters, two shot-noise parameters, the dimensionless linear growth rate  $f$ , and the velocity dispersion  $\sigma_v$ . We compute the fiducial value of these parameters as a function of redshift following [23,50], and the expression for the hydrogen bispectrum is found in Eq. (E4) within Appendix E. We also calculate the hydrogen power spectrum, given in Eq. (E3), as we use it to compute the Gaussian covariance. To compute the non-Gaussian covariance, we use the model of Eq. (15). We then proceed in computing the Fisher matrix, which is estimated in the thin bins form of Eq. (13). We marginalize over all seven nuisance parameters entering the bispectrum. For more details, we refer to Appendix E.

Figure 3 shows the ratio of the estimated uncertainty computed using a non-Gaussian covariance over a Gaussian approximation for the local- and equilateral-type non-Gaussianities as a function of redshift for a PUMA-like experiment. We compute the uncertainty for two different values of  $k_{\text{max}}$ , corresponding to  $0.5k_{\text{NL}}$  (dashed lines) and  $0.75k_{\text{NL}}$  (solid lines), as we expect Eq. (6) to be less accurate for tracers. Our results show that even for a more conservative choice of  $k_{\text{max}} = 0.5k_{\text{NL}}$ , the effect is significant. The increase in uncertainty ranges from a factor of 2 to a factor of 5 for local-type non-Gaussianity. We therefore conclude that previous forecasts on constraining  $f_{\text{NL}}$  at high redshift are too optimistic [20,22–25], and non-Gaussian covariance will have to be considered in order

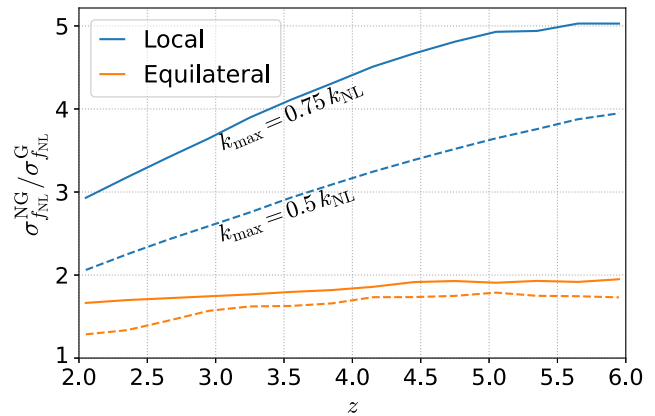


FIG. 3. Estimate of relative increase in error on the non-Gaussian amplitude  $f_{\text{NL}}$  due to non-Gaussian covariance of the hydrogen bispectrum, as a function of redshift for a PUMA-like experiment when marginalizing over the seven additional parameters of the hydrogen bispectrum. We show the results for  $k_{\text{max}} = 0.75k_{\text{NL}}$  (solid lines) and  $k_{\text{max}} = 0.5k_{\text{NL}}$  (dashed lines).

to produce more realistic forecasts. A similar estimation for a generic biased tracer was performed in Ref. [51] up to  $z = 10$  and shows qualitative agreement with Fig. 3.

#### IV. DISCUSSION AND CONCLUSIONS

We studied the impact of non-Gaussian terms in the covariance on measurements of cosmological correlators. Specifically, our aim was to quantify the effect on the estimated uncertainty of the primordial non-Gaussian amplitude  $f_{\text{NL}}$  when using the bispectrum at high redshift as an observable. Because off-diagonal components are small compared to the diagonal, most studies have typically neglected this covariance. We showed that, when looking at the information content, there is a significant impact on the constraining power on primordial non-Gaussianity due to this non-Gaussian mode coupling, even at high redshifts and well below the nonlinear scale as defined in Eq. (6).

We have first computed the effect of non-Gaussian covariance on  $\sigma_{f_{\text{NL}}}$  using the matter bispectrum in real space, and then performed a more realistic estimation using the hydrogen bispectrum as measured from a PUMA-like experiment. This proposed 21-cm intensity mapping experiment has the potential to constrain primordial bispectra to reach beyond constraints set by the CMB. However, our analysis shows that not accounting for the full covariance can overestimate the constraining power of the hydrogen bispectrum measured by PUMA up to a factor of 5 for local-type non-Gaussianity and 2 for equilateral-type. For local-type non-Gaussianity, the primary observable is actually the tracer power spectrum, thanks to the so-called scale-dependent bias, which we do not include in our analysis. Nevertheless, our results imply that combining it with the bispectrum does not help as much as it is expected to considering a Gaussian covariance only. Moreover, they motivate including the

bispectrum–power spectrum cross covariance in the joint analysis, which is also a non-Gaussian contribution [18]. Overall, our result suggests we should reconsider some of the existing forecasts and make sure the projected numbers are not overly optimistic for future high-redshift surveys such as PUMA, MegaMapper [52], and the Maunakea Spectroscopic Explorer [53].

Future constraints on primordial non-Gaussianities will depend on our ability to extract information from large-scale structures. Intuitively, the main obstacle to constrain primordial spectra is set by the nonlinear scale, which estimates when loop corrections become important. Here we show that for the Fisher information on  $f_{\text{NL}}$ , it is important to account for non-Gaussian bispectrum covariance, even for modes that are still considered linear. The results are comparable to what was found for measurements of the CMB bispectrum, where lensing-induced off-diagonal covariance is the main limitation as we start to measure smaller scales and increase the number of accessible modes.

For the CMB, it was shown that the lensing-induced covariance can be accounted for by delensing the data before applying the standard estimators [42]. The analogy here would be to “degravitate” the data, a technique that is well established in studies of the baryon acoustic oscillations in galaxy surveys [54]. In fact, this approach has been shown to improve constraints on primordial non-Gaussianity from the bispectrum [55]. It might be possible to explore this option at high redshifts, where the physics is still perturbatively tractable. At lower redshifts, however, it likely suggests that existing estimators are suboptimal, or that we have adopted inefficient summary statistics that need to be revisited. Similar conclusions were drawn in Ref. [56]. Applying reconstruction methods [57] or using simulations (e.g., through simulation-based inference [58]) [59–63], both active fields of investigation, will certainly help to establish to what degree we have to modify our analysis tools in search of signs of primordial non-Gaussianity.

The code used to produce the results in this work is publicly available [64].

### ACKNOWLEDGMENTS

The authors would like to thank Emanuele Castorina, Will Coulton, Simon Foreman, Dionysios Karagiannis, Emiliano Sefusatti, and Anže Slosar for useful discussions and comments on a draft. Recently, we became aware of related work by Coulton *et al.* [56], and we thank the authors for sharing their draft ahead of submission. We would also like to thank Francisco Villaescusa-Navarro and the whole QUIJOTE team for making the simulation suite available. We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high-performance computing cluster. T. F. is supported by the Fundamentals of the Universe research program within the University of Groningen. M. B.

acknowledges support from the Netherlands Organization for Scientific Research (NWO), which is funded by the Dutch Ministry of Education, Culture and Science (OCW) under VENI Grant No. 016.Veni.192.210. P. D. M. acknowledges support from the Netherlands Organization for Scientific Research (NWO) under a VIDI Grant (Dossier No. 639.042.730).

### APPENDIX A: PRIMORDIAL BISPECTRA

The local, equilateral, and orthogonal bispectrum templates are given by

$$B_{\zeta}^{\text{loc}}(k_1, k_2, k_3) = \frac{6}{5} f_{\text{NL}}^{\text{loc}} (P_{\zeta,1} P_{\zeta,2} + P_{\zeta,1} P_{\zeta,3} + P_{\zeta,2} P_{\zeta,3}), \quad (\text{A1})$$

$$B_{\zeta}^{\text{equil}}(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}}^{\text{equil}} (- (P_{\zeta,1} P_{\zeta,2} + 2 \text{ perms}) - 2 P_{\zeta,1}^{2/3} P_{\zeta,2}^{2/3} P_{\zeta,3}^{2/3} + (P_{\zeta,1}^{1/3} P_{\zeta,2}^{2/3} P_{\zeta,3} + 5 \text{ perms})), \quad (\text{A2})$$

$$B_{\zeta}^{\text{ortho}}(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}}^{\text{ortho}} (-3 (P_{\zeta,1} P_{\zeta,2} + 2 \text{ perms}) - 8 P_{\zeta,1}^{2/3} P_{\zeta,2}^{2/3} P_{\zeta,3}^{2/3} + 3 (P_{\zeta,1}^{1/3} P_{\zeta,2}^{2/3} P_{\zeta,3} + 5 \text{ perms})), \quad (\text{A3})$$

where we have introduced the shorthand notation  $P_{\zeta,1} = P_{\zeta}(k_1)$ . Note that although Eq. (2) demonstrates that the primordial bispectrum can in principle depend on the full three-momenta, for the above primordial shapes there is no angular dependence, and they only depend on the magnitudes of the triangle’s momenta (i.e., the shape of the triangle).

### APPENDIX B: STANDARD PERTURBATION THEORY AT TREE LEVEL

Within the perturbative regime, gravitational interactions can be treated within the framework of standard perturbation theory (SPT). For an extensive review, we refer to, e.g., Ref. [33]. Here, we will present only the results relevant to our work. Since we require the gravitational trispectrum, we expand the density field to third order:

$$\delta_{\mathbf{k}}(z) = \delta_{\mathbf{k}}^{(1)}(z) + \delta_{\mathbf{k}}^{(2)}(z) + \delta_{\mathbf{k}}^{(3)}, \quad (\text{B1})$$

where the superscript denotes the order of the perturbation. Solving the evolution equations order by order in perturbations, one finds (dropping the explicit time dependence)

$$\begin{aligned}\delta_{\mathbf{k}}^{(2)} &= \int_{\mathbf{q}} F_2(\mathbf{q}, \mathbf{k} - \mathbf{q}) \delta_{\mathbf{q}}^{(1)} \delta_{\mathbf{k}-\mathbf{q}}^{(1)}, \\ \delta_{\mathbf{k}}^{(3)} &= \int_{\mathbf{q}_1, \mathbf{q}_2} F_3(\mathbf{q}_1, \mathbf{q}_2, \mathbf{k} - \mathbf{q}_{12}) \delta_{\mathbf{q}_1}^{(1)} \delta_{\mathbf{q}_2}^{(1)} \delta_{\mathbf{k}-\mathbf{q}_{12}}^{(1)}.\end{aligned}\quad (\text{B2})$$

These higher-order perturbations will induce the gravitational correlations. At tree level, then, the bispectrum of the density field due to gravitational interactions is found to be

$$\begin{aligned}B_{\delta}^{\text{snG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \rangle + 2 \text{ perms} \\ &= 2F_2(\mathbf{k}_1, \mathbf{k}_2) P_{\delta}^L(k_1) P_{\delta}^L(k_2) + 2 \text{ perms},\end{aligned}\quad (\text{B3})$$

where the linear power spectrum is defined as

$$\begin{aligned}\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \rangle &= (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_{\delta}^L(k_1) \\ &= (2\pi)^3 \delta_D(\mathbf{k}_{12}) \mathcal{M}(k_1, z) P_{\zeta}(k_1).\end{aligned}\quad (\text{B4})$$

Here,  $\mathcal{M}$  is the linear transfer function. The tree-level trispectrum consists of two contributions:

$$\begin{aligned}T_{\delta}^{\text{snG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= T_{\delta}^{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\ &\quad + T_{\delta}^{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4),\end{aligned}\quad (\text{B5})$$

with the two contributions given by

$$\begin{aligned}T_{\delta}^{1113}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(1)} \delta_{\mathbf{k}_4}^{(3)} \rangle + 3 \text{ perms} \\ &= 6F_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \prod_{i=1}^3 P_{\delta}^L(k_i) \\ &\quad + 3 \text{ perms}, \\ T_{\delta}^{1122}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \delta_{\mathbf{k}_4}^{(2)} \rangle + 5 \text{ perms} \\ &\quad + 4[F_2(-\mathbf{k}_1, \mathbf{k}_{13}) P_{\delta}^L(k_{13}) \\ &\quad + F_2(-\mathbf{k}_1, \mathbf{k}_{14}) P_{\delta}^L(k_{14})] P_{\delta}^L(k_1) P_{\delta}^L(k_2) \\ &\quad + 5 \text{ perms}.\end{aligned}\quad (\text{B6})$$

The kernels  $F_2$  and  $F_3$  determine how modes of different wavelength are coupled by gravity. For their explicit form, see, e.g., Ref. [33]. Beyond tree level, the higher-order perturbations will also induce a correction to the power spectrum, known as the one-loop power spectrum:

$$\langle \delta_{\mathbf{k}_1}^{(2)} \delta_{\mathbf{k}_2}^{(2)} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{12}) P_{\delta}^{1\text{-loop}}(k_1). \quad (\text{B7})$$

## APPENDIX C: DETAILS ON SIMULATIONS

In order to verify our computations of non-Gaussian covariance, we compare our results to simulations. For redshifts  $z = 0$  and 3, we use the QUIJOTE simulation suite [28].

QUIJOTE consists of 15000  $N$ -body simulations using a fiducial cosmology, enough to obtain an accurate covariance matrix up to the scales of interest in this work. The simulations consist of  $512^3$  particles in a box with sides 1000 Mpc/h, setting the fundamental mode to  $k_F = (2\pi/1000) \approx 0.0063$  h/Mpc. To construct the covariance matrix, we use the power spectrum and bispectrum measurements as provided in the suite, which use a binning of  $\Delta k = 3k_F$ ,  $k_{\min} = \frac{3}{2}k_F$ , and an interpolation grid of size  $360^3$ . For redshift  $z = 10$ , we use  $\sim 4000$  realizations of initial conditions (ICs) generated using third-order Lagrangian perturbation theory (3LPT) with the `MONOFONIC` code [37]. Since at higher redshifts the power spectrum and bispectrum are smaller, shot noise becomes increasingly dominant. The  $z = 10$  realizations are therefore generated with  $512^3$  particles in a box of 250 Mpc/h, setting the fundamental mode to  $k_F = (2\pi/250) \approx 0.025$  h/Mpc. Measurements of the power spectrum and bispectrum are done using the codes `PyLIANS` [65] and `PySpectrum` [66], respectively, with the same settings as used for the QUIJOTE measurements. Having measured the power spectrum and bispectrum from many realizations of the simulation, we obtain the covariance matrix through Eq. (8). When inverting the covariance matrix from simulations, we include the Hartlap factor to unbiased the numerical matrix [67].

### 1. Evaluation of bin-averaged primordial bispectra

In order to compute the Fisher information for primordial bispectra from the simulations, we need the bin-averaged derivatives of the theoretical bispectrum with respect to  $f_{\text{NL}}$  that appear in Eq. (7):

$$\begin{aligned}\frac{\partial \hat{B}(k_1, k_2, k_3)}{\partial f_{\text{NL}}} &= \frac{1}{V_{123}} \int_{k_1} d^3 \mathbf{q}_1 \int_{k_2} d^3 \mathbf{q}_2 \int_{k_3} d^3 \mathbf{q}_3 \\ \frac{\partial B(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)}{\partial f_{\text{NL}}} &= \frac{1}{V_{123}} \int_{k_1} d^3 \mathbf{q}_1 \int_{k_2} d^3 \mathbf{q}_2 \int_{k_3} d^3 \mathbf{q}_3 \\ &\quad \times \left( \prod_{i=1}^3 \mathcal{M}(q_i, z) \right) B_{\zeta}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)|_{f_{\text{NL}}=1},\end{aligned}\quad (\text{C1})$$

where the hat denotes a bin-averaged quantity, and the volume of the bin is given by

$$V_{123} = \int_{k_1} d^3 \mathbf{q}_1 \int_{k_2} d^3 \mathbf{q}_2 \int_{k_3} d^3 \mathbf{q}_3, \quad (\text{C2})$$

and the integrals denote a binning similar to that of the simulation measurements—i.e., over spherical shells with centers  $k_i$  and width  $[k_i - 3k_F/2, k_i + 3k_F/2]$ .

### 2. Orthogonal shape

When computing the bin-averaged orthogonal shape bispectrum, one realizes that it becomes negative for certain



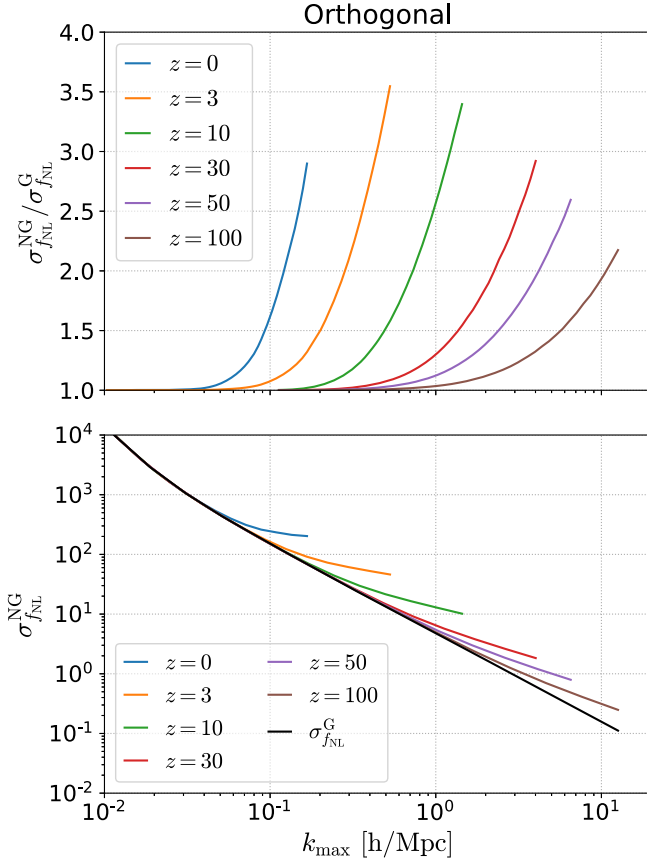


FIG. 4. Similar plots to Figs. 1 and 2, but now for orthogonal non-Gaussianity.

triangle configurations. For the coarse binning ( $\Delta k = 3k_F$ ) of the simulation data we use, the bin-averaged bispectrum in Eq. (C1) suffers from cancellations within the bin, removing part of the signal. Since remeasuring the bispectrum in QUIJOTE simulations with finer binning goes beyond the scope of this paper, we decided to omit these results. Nonetheless, our predictions in the thin bin approximation are presented in Fig. 4 and for  $z = 0$  agree well with the simulation results presented in Ref. [56].

#### APPENDIX D: DETAILS ON THE FISHER MATRIX IN THE THIN BINS LIMIT

To produce our theoretical predictions of the loss of constraining power due to non-Gaussian covariance, we take the continuous limit of Eq. (7) while approximating the inverse covariance matrix.

As explained in the main text, this allows us to probe a wide range of scales at low computational cost. The main complication for Eq. (12) is to explicitly compute the inverse covariance as a function of triangle configurations. To this end, we expand the inverse using a Neumann series:

$$C^{-1} = \sum_{n=0}^{\infty} (-C_G^{-1} C_{\text{nG}})^n C_G^{-1}, \quad (\text{D1})$$

such that we never have to invert the non-Gaussian covariance matrix that contains off-diagonal terms. Hence, the Fisher matrix becomes the sum of infinitely many terms:

$$F_{ab} = \sum_{n=0}^{\infty} \sum_{TT'} \frac{\partial B_T}{\partial p_a} ((-C_G^{-1} C_{\text{nG}})^n C_G^{-1})_{TT'} \frac{\partial B_T}{\partial p_b}. \quad (\text{D2})$$

Since the terms in this sum are increasingly complicated to compute (in the continuous limit, every matrix multiplication becomes an integral over triangle configurations), we choose to approximate the expansion using

$$\begin{aligned} F_{ab} &= \sum_{TT'} \frac{\partial B_T}{\partial p_a} (C_G^{-1})_{TT'} \frac{\partial B_T}{\partial p_b} \\ &+ \sum_{n=1}^{\infty} \frac{(\sum_{TT'} \frac{\partial B_T}{\partial p_a} (-C_G^{-1} C_{\text{nG}} C_G^{-1})_{TT'} \frac{\partial B_T}{\partial p_b})^n}{(\sum_{TT'} \frac{\partial B_T}{\partial p_a} (C_G^{-1})_{TT'} \frac{\partial B_T}{\partial p_b})^{n-1}} \\ &= F_{ab}^{\text{G}} + \sum_{n=1}^{\infty} (-1)^n \frac{(\delta F_{ab}^{\text{nG}})^n}{(F_{ab}^{\text{G}})^{n-1}}, \end{aligned} \quad (\text{D3})$$

which can be recognised as the expansion of Eq. (13). The approximation of Eq. (D3) seems to work reasonably well in the range of  $k_{\text{max}}$  which we compare with simulation results (see Fig. 1 in the main text). On the other hand, our key result is that these non-Gaussian terms are actually important; therefore, we expect our approximation to break down. This motivates further work in defining a proper estimator for primordial non-Gaussianity in the presence of non-Gaussian covariance terms.

#### APPENDIX E: DETAILS ON THE PUMA ANALYSIS

We perform the analysis of the impact of non-Gaussian covariance on the PUMA survey along the lines of Ref. [23]. We use a binning of  $\Delta z = 0.1$  between  $2 < z < 6$ . The largest available scale is set by the volume  $V_s(z)$  of the redshift bin through  $k_{\text{min}}(z) = k_F(z) = 2\pi/L(z)$ , where  $L(z) = V_s(z)^{1/3}$ ,  $V_s(z) = \frac{4\pi}{3}(r(z+\Delta z)^3 - r(z-\Delta z)^3)$  and  $r(z)$  is the comoving distance to redshift  $z$  in units of Mpc/h.

##### 1. Foregrounds

The largest scale is set by a foreground cut in the line-of-sight direction ( $k_{\parallel, \text{min}} = 0.01$  h/Mpc), removing much of the dependence on the choice of redshift binning. We limit ourselves to linear scales by choosing  $k_{\text{max}}(z)$  to be  $0.5k_{\text{NL}}(z)$  or  $0.75k_{\text{NL}}(z)$ . We model the foreground wedge by excluding all modes for which

$$k_{\parallel} < \frac{r(z)H(z)}{c(1+z)} \sin(0.66N_w \theta_{\text{FOV}}(z)) \times k_{\perp}. \quad (\text{E1})$$

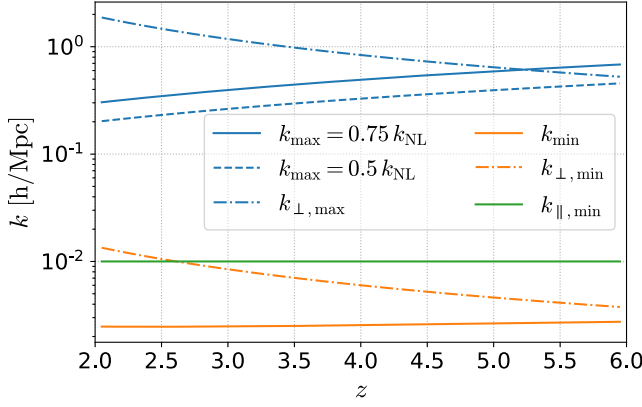


FIG. 5. Smallest and largest accessible scales as a function of redshift for the PUMA survey. The smallest overall scale  $k_{\max}$  is determined by the nonlinear scale. The largest overall scale  $k_{\min}$  is set by the volume of the redshift bin and hence depends on the choice of binning. The smallest and largest scales in the perpendicular direction,  $k_{\perp,\max}$  and  $k_{\perp,\min}$ , are set by properties of the experiment. Finally, we apply a foreground cut  $k_{\parallel,\min} = 0.01$  h/Mpc in the line-of-sight direction that effectively replaces  $k_{\min}$ , removing most dependence on the choice of binning.

Here  $H(z)$  is the Hubble parameter,  $\theta_{\text{FOV}}(z) = \lambda_{21}(z)/D_{\text{eff}}$ ,  $\lambda_{21}(z)$  is the redshifted 21-cm wavelength in meters, and  $D_{\text{eff}} = (\sqrt{0.7} \times 6 \text{ m})$  is the effective dish size of the PUMA setup.  $N_w$  determines the severity of the foreground wedge. We apply a pessimistic wedge cut of  $N_w = 3$  and a foreground cut  $k_{\parallel,\min} = 0.01$  h/Mpc, in order to show that

the loss of constraining power due to non-Gaussian covariance persists in such a setup. Finally, the largest and smallest accessible scales in the perpendicular direction are set by

$$k_{\perp,\max}(z) = \frac{2\pi D_{\max}}{\lambda_{21}(z)r(z)}, \quad k_{\perp,\min}(z) = \frac{2\pi}{r(z)\theta_{\text{FOV}}(z)}, \quad (\text{E2})$$

where  $D_{\max} = 700$  m is the largest baseline of the PUMA setup. We show all the relevant scales together in Fig. 5.

## 2. The hydrogen power spectrum and bispectrum

The calculation of hydrogen correlation functions in redshift space is significantly more involved than the matter field case from the modeling point of view. Besides nonlinearities in the matter field, we need to account for the biased relation between the hydrogen and matter distributions, redshift space distortions, and stochasticity introduced by the discreteness effects and Poisson noise. A complete explanation of these modeling efforts can be found in Ref. [23]. Here we quote the hydrogen power spectrum and bispectrum for reference, defined as

$$P_{\text{HI}}(z, \mathbf{k}) = P_N(z, \mathbf{k}) + T_b(z)^2 D_{\text{FOG}}^P(z, \mathbf{k}) \times [Z_1(z, \mathbf{k})^2 P_{\delta}^L(z, \mathbf{k}) + P_{\epsilon}(z)], \quad (\text{E3})$$

and

$$B_{\text{HI}}(z, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = T_b(z)^3 (D_{\text{FOG}}^B(z, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)) \left[ \prod_{i=1}^3 Z_1(z, \mathbf{k}_i) B_{\delta}^{\text{pnG}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + 2Z_1(z, \mathbf{k}_1)Z_1(z, \mathbf{k}_2)Z_2(z, \mathbf{k}_1, \mathbf{k}_2)P_{\delta}^L(z, \mathbf{k}_1)P_{\delta}^L(z, \mathbf{k}_2) + 2 \text{ perm} \right] + P_{\epsilon\epsilon\delta}(z) \left[ \sum_{i=1}^3 Z_1(z, \mathbf{k}_i)P_{\delta}^L(z, \mathbf{k}_i) \right] + B_{\epsilon}(z), \quad (\text{E4})$$

where  $P_N$  is the instrumental noise,  $T_b(z)$  is the brightness temperature of the 21-cm signal at a given redshift,  $P_{\epsilon}$ ,  $P_{\epsilon\epsilon\delta}$ ,  $B_{\epsilon}$  are stochastic noise contributions,  $Z_1$ ,  $Z_2$  are the first- and second-order redshift space kernels,  $B_{\delta}^{\text{pnG}}$  is the primordial contribution to the matter bispectrum as used in Eq. (C1), and  $D_{\text{FOG}}^B$  models the finger-of-God (FOG) dumping effect. For the explicit expressions of these quantities, we refer to Ref. [23]. The redshift space kernels contain bias parameters  $\{b_1, b_2, b_{s^2}, b_{\Psi}, b_{\Psi\delta}\}$ , as well as the linear growth rate  $f$  due to redshift space distortions (RSDs). The scale-dependent biases  $\{b_{\Psi}, b_{\Psi\delta}\}$  can be modeled in terms of  $\{b_1, b_2, f_{\text{NL}}\}$  (though see Ref. [68] for a study of this approximation). This means that

primordial non-Gaussianity enters not only through  $B_{\delta}^{\text{pnG}}$ , but also through the terms involving  $Z_1$ ,  $Z_2$  that contain the scale-dependent biases. Finally, the FOG factor is modeled using the velocity dispersion  $\sigma_v$ . In this work, we are only interested in the signal-to-noise ratio for  $f_{\text{NL}}$  coming from the hydrogen bispectrum; hence, the total number of parameters including the stochastic noise contributions equals 8:

$$\mathbf{p} = \{f_{\text{NL}}, b_1, b_2, b_{s^2}, f, \sigma_v, P_{\epsilon\epsilon\delta}, B_{\epsilon}\}. \quad (\text{E5})$$

We calculate the Fisher matrix of the eight parameters that enter the hydrogen bispectrum, with and without

non-Gaussian covariance at each redshift bin using the weighted estimator approach described above [40]. Contrary to Ref. [23], we do not account for theoretical errors on the bias parameters in our analysis, which adds additional covariance (including off-diagonal) to account for uncertainties in the bias model along the lines of Ref. [69]. Once we have the Gaussian and non-Gaussian Fisher matrices, we marginalize over the seven nuisance

parameters by inverting the Fisher matrix at every redshift. The marginalized uncertainty for  $f_{\text{NL}}$  is then given by

$$\sigma_{f_{\text{NL}}}(z) = (F^{-1}(z))_{f_{\text{NL}}f_{\text{NL}}}^{1/2}. \quad (\text{E6})$$

The ratio of the estimated uncertainty including non-Gaussian covariance over Gaussian covariance is shown in Fig. 3.

- 
- [1] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).  
 [2] A. D. Linde, *Phys. Lett.* **108B**, 389 (1982).  
 [3] A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).  
 [4] Y. Akrami *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A10 (2020).  
 [5] P. D. Meerburg *et al.*, [arXiv:1903.04409](https://arxiv.org/abs/1903.04409).  
 [6] A. Achúcarro *et al.*, [arXiv:2203.08128](https://arxiv.org/abs/2203.08128).  
 [7] J. M. Maldacena, *J. High Energy Phys.* **05** (2003) 013.  
 [8] P. Creminelli and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **10** (2004) 006.  
 [9] N. Arkani-Hamed and J. Maldacena, [arXiv:1503.08043](https://arxiv.org/abs/1503.08043).  
 [10] R. Takahashi, N. Yoshida, M. Takada, T. Matsubara, N. Sugiyama, I. Kayo, T. Nishimichi, S. Saito, and A. Taruya, *Astrophys. J.* **726**, 7 (2011).  
 [11] K. C. Chan and L. Blot, *Phys. Rev. D* **96**, 023528 (2017).  
 [12] K. C. Chan, A. Moradinezhad Dizgah, and J. Noreña, *Phys. Rev. D* **97**, 043532 (2018).  
 [13] D. Wadekar and R. Scoccimarro, *Phys. Rev. D* **102**, 123517 (2020).  
 [14] A. Barreira, *J. Cosmol. Astropart. Phys.* **03** (2019) 008.  
 [15] D. Gualdi and L. Verde, *J. Cosmol. Astropart. Phys.* **06** (2020) 041.  
 [16] A. Oddo, F. Rizzo, E. Sefusatti, C. Porciani, and P. Monaco, *J. Cosmol. Astropart. Phys.* **11** (2021) 038.  
 [17] A. Barreira, *J. Cosmol. Astropart. Phys.* **01** (2022) 033.  
 [18] M. Biagetti, L. Castiblanco, J. Noreña, and E. Sefusatti, *J. Cosmol. Astropart. Phys.* **09** (2022) 009.  
 [19] F. Rizzo, C. Moretti, K. Pardede, A. Eggemeier, A. Oddo, E. Sefusatti, C. Porciani, and P. Monaco, [arXiv:2204.13628](https://arxiv.org/abs/2204.13628).  
 [20] J. B. Muñoz, Y. Ali-Haïmoud, and M. Kamionkowski, *Phys. Rev. D* **92**, 083508 (2015).  
 [21] X. Chen, P. D. Meerburg, and M. Münchmeyer, *J. Cosmol. Astropart. Phys.* **09** (2016) 023.  
 [22] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz, and X. Chen, *J. Cosmol. Astropart. Phys.* **03** (2017) 050.  
 [23] D. Karagiannis, A. Slosar, and M. Liguori, *J. Cosmol. Astropart. Phys.* **11** (2020) 052.  
 [24] T. Flöss, T. de Wild, P. D. Meerburg, and L. V. E. Koopmans, *J. Cosmol. Astropart. Phys.* **06** (2022) 020.  
 [25] D. Yamauchi, *Prog. Theor. Exp. Phys.* **2022**, 073E02 (2022).  
 [26] D. Karagiannis, R. Maartens, and L. Randrianjanahary, *J. Cosmol. Astropart. Phys.* **11** (2022) 003.  
 [27] R. Ansari *et al.* (Cosmic Visions 21 cm Collaboration), [arXiv:1810.09572](https://arxiv.org/abs/1810.09572).  
 [28] F. Villaescusa-Navarro *et al.*, *Astrophys. J. Suppl. Ser.* **250**, 2 (2020).  
 [29] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020); **652**, C4(E) (2021).  
 [30] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016).  
 [31] H. Lee, D. Baumann, and G. L. Pimentel, *J. High Energy Phys.* **12** (2016) 040.  
 [32] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, *J. High Energy Phys.* **03** (2008) 014.  
 [33] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, *Phys. Rep.* **367**, 1 (2002).  
 [34] Other definitions have been considered in the literature—e.g., Ref. [35] studies the nonlinear scale for the bispectrum specifically. The precise definition of  $k_{\text{NL}}$  does not qualitatively change the results of this paper.  
 [35] J. Tomlinson and D. Jeong, [arXiv:2204.00668](https://arxiv.org/abs/2204.00668).  
 [36] R. E. Angulo and O. Hahn, *Living Rev. Comput. Astrophys.* **8**, 1 (2022).  
 [37] M. Michaux, O. Hahn, C. Rampf, and R. E. Angulo, *Mon. Not. R. Astron. Soc.* **500**, 663 (2020).  
 [38] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, *Astrophys. J.* **496**, 586 (1998).  
 [39] We also measure the power spectrum, since, as we show below, it enters in the calculation of the covariance. The estimator of the power spectrum is
- $$\hat{P}(k) \equiv \frac{k_F^3}{N_k} \sum_{\mathbf{q} \in k} \delta_{\mathbf{q}} \delta_{-\mathbf{q}}, \quad (\text{16})$$
- where  $N_k$  gives the number of modes in each  $k$  bin.  
 [40] The approximate equality indicates the thin shell approximation.  
 [41] This is very similar to how lensing-induced covariance mostly affects measurements of the local bispectrum in the CMB [42].  
 [42] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, *Phys. Rev. D* **101**, 123504 (2020).  
 [43] V. Desjacques, D. Jeong, and F. Schmidt, *Phys. Rep.* **733**, 1 (2018).  
 [44] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, *Phys. Rev. D* **77**, 123514 (2008).  
 [45] S. Matarrese and L. Verde, *Astrophys. J. Lett.* **677**, L77 (2008).

- [46] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, *J. Cosmol. Astropart. Phys.* **08** (2008) 031.
- [47] M. Biagetti, *Galaxies* **7**, 71 (2019).
- [48] We have confirmed that our forecasts, using very similar assumptions about the PUMA survey, result in forecasts on  $\sigma(f_{\text{NL}})$  that are consistent with those presented in Refs. [23,49] when neglecting non-Gaussian covariance.
- [49] N. Sailer, E. Castorina, S. Ferraro, and M. White, *J. Cosmol. Astropart. Phys.* **12** (2021) 049.
- [50] E. Castorina and F. Villaescusa-Navarro, *Mon. Not. R. Astron. Soc.* **471**, 1788 (2017).
- [51] R. de Putter, [arXiv:1802.06762](https://arxiv.org/abs/1802.06762).
- [52] D. J. Schlegel *et al.*, [arXiv:1907.11171](https://arxiv.org/abs/1907.11171).
- [53] C. Babusiaux *et al.* (MSE Science Team), [arXiv:1904.04907](https://arxiv.org/abs/1904.04907).
- [54] D. J. Eisenstein, H.-j. Seo, E. Sirko, and D. Spergel, *Astrophys. J.* **664**, 675 (2007).
- [55] M. Shirasaki, N. S. Sugiyama, R. Takahashi, and F.-S. Kitaura, *Phys. Rev. D* **103**, 023506 (2021).
- [56] W. R. Coulton, F. Villaescusa-Navarro, D. Jamieson, M. Baldi, G. Jung, D. Karagiannis, M. Liguori, L. Verde, and B. D. Wandelt, [arXiv:2206.01619](https://arxiv.org/abs/2206.01619).
- [57] F. Leclercq and B. Wandelt, *IAU Symp.* **306**, 1 (2014).
- [58] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055 (2020).
- [59] J. Alsing, B. Wandelt, and S. Feeney, *Mon. Not. R. Astron. Soc.* **477**, 2874 (2018).
- [60] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 5093 (2019).
- [61] N. Jeffrey and B. D. Wandelt, in *Proceedings of the 34th Conference on Neural Information Processing Systems* (2020), [arXiv:2011.05991](https://arxiv.org/abs/2011.05991).
- [62] B. K. Miller, A. Cole, G. Louppe, and C. Weniger, [arXiv:2011.13951](https://arxiv.org/abs/2011.13951).
- [63] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, *J. Cosmol. Astropart. Phys.* **09** (2022) 004.
- [64] <https://github.com/tsfloss/pyNG>.
- [65] <https://github.com/franciscovillaescusa/Pylians>.
- [66] <https://github.com/changhoonhahn/pySpectrum>.
- [67] J. Hartlap, P. Simon, and P. Schneider, *Astron. Astrophys.* **464**, 399 (2007).
- [68] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques, and F. Schmidt, *Mon. Not. R. Astron. Soc.* **468**, 3277 (2017).
- [69] T. Baldauf, M. Mirbabayi, M. Simonović, and M. Zaldarriaga, [arXiv:1602.00674](https://arxiv.org/abs/1602.00674).