

PAPER • OPEN ACCESS

# Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for $f_{\text{NL}}$ constraints

To cite this article: Alexandre Barreira JCAP01(2022)033

View the [article online](#) for updates and enhancements.

You may also like

- [Can we actually constrain  \$f\_{\text{NL}}\$  using the scale-dependent bias effect? An illustration of the impact of galaxy bias uncertainties using the BOSS DR12 galaxy power spectrum](#)  
Alexandre Barreira
- [Learning to concentrate: multi-tracer forecasts on local primordial non-Gaussianity with machine-learned bias](#)  
James M. Sullivan, Tijan Prijon and Uroš Seljak
- [Local primordial non-Gaussian bias at the field level](#)  
James M. Sullivan and Shi-Fan Chen

# Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for $f_{\text{NL}}$ constraints

**Alexandre Barreira**

Excellence Cluster ORIGINS,  
Boltzmannstraße 2, 85748 Garching, Germany  
Ludwig-Maximilians-Universität,  
Schellingstraße 4, 80799 München, Germany  
E-mail: [alex.barreira@origins-cluster.de](mailto:alex.barreira@origins-cluster.de)

Received August 2, 2021

Accepted January 4, 2022

Published January 18, 2022

**Abstract.** We use hydrodynamical separate universe simulations with the IllustrisTNG model to predict the local primordial non-Gaussianity (PNG) bias parameters  $b_\phi$  and  $b_{\phi\delta}$ , which enter at leading order in the galaxy power spectrum and bispectrum. This is the first time that  $b_{\phi\delta}$  is measured from either gravity-only or galaxy formation simulations. For dark matter halos, the popular assumption of universality overpredicts the  $b_{\phi\delta}(b_1)$  relation in the range  $1 \lesssim b_1 \lesssim 3$  by up to  $\Delta b_{\phi\delta} \sim 3$  ( $b_1$  is the linear density bias). The adequacy of the universality relation is worse for the simulated galaxies, with the relations  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  being generically redshift-dependent and very sensitive to how galaxies are selected (we test total, stellar and black hole mass, black hole mass accretion rate and color). The uncertainties on  $b_\phi$  and  $b_{\phi\delta}$  have a direct, often overlooked impact on the constraints of the local PNG parameter  $f_{\text{NL}}$ , which we study and discuss. For a survey with  $V = 100 \text{Gpc}^3/h^3$  at  $z = 1$ , uncertainties  $\Delta b_\phi \lesssim 1$  and  $\Delta b_{\phi\delta} \lesssim 5$  around values close to the fiducial can yield relatively unbiased constraints on  $f_{\text{NL}}$  using power spectrum and bispectrum data. We also show why priors on galaxy bias are useful even in analyses that fit for products  $f_{\text{NL}}b_\phi$  and  $f_{\text{NL}}b_{\phi\delta}$ . The strategies we discuss to deal with galaxy bias uncertainties can be straightforwardly implemented in existing  $f_{\text{NL}}$  constraint analyses (we provide fits for some of the bias relations). Our results motivate more works with galaxy formation simulations to refine our understanding of  $b_\phi$  and  $b_{\phi\delta}$  towards improved constraints on  $f_{\text{NL}}$ .

**Keywords:** cosmological parameters from LSS, cosmological simulations, hydrodynamical simulations, redshift surveys

**ArXiv ePrint:** [2107.06887](https://arxiv.org/abs/2107.06887)



---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methodology</b>	<b>4</b>
2.1	Simulation data specifications	4
2.2	The linear density and local PNG bias parameters $b_1$ and $b_\phi$	5
2.3	The second-order local PNG bias parameter $b_{\phi\delta}$	6
<b>3</b>	<b>Galaxy bias results</b>	<b>8</b>
3.1	The $b_{\phi\delta}$ parameter in gravity-only simulations	8
3.2	The $b_\phi$ and $b_{\phi\delta}$ parameters in galaxy formation simulations	10
<b>4</b>	<b>Impact of galaxy bias uncertainties on <math>f_{\text{NL}}</math> constraints</b>	<b>14</b>
4.1	Forecast setup	14
4.2	Results from parametrization 1: direct priors on $b_\phi$ and $b_{\phi\delta}$	16
4.3	Results from parametrization 2: fit for products $f_{\text{NL}}b_\phi$ , $f_{\text{NL}}b_{\phi\delta}$	18
<b>5</b>	<b>Summary and conclusions</b>	<b>20</b>
<b>A</b>	<b>Validation of the method to estimate <math>b_{\phi\delta}</math> using the <math>b_2</math> parameter</b>	<b>22</b>
<b>B</b>	<b>Expressions of the tree-level galaxy power spectrum and bispectrum</b>	<b>24</b>

---

## 1 Introduction

One of the main open questions in cosmology today concerns the origin of the primordial density fluctuations generated in the early Universe during the epoch of inflation. The simplest theoretical explanation is that of single-field models, in which the density perturbations are due to quantum fluctuations of a single scalar degree of freedom that rolls slowly down its potential. These models crucially predict density fluctuations that are Gaussian distributed [1–6], and as a result, the detection of any deviation from Gaussianity would immediately rule them out and open the door for more elaborate, multifield models [7, 8]. *Local-type* primordial non-Gaussianity (PNG) is the most popular way to describe departures from perfectly Gaussian-distributed fluctuations. In this case, the primordial gravitational (Bardeen) potential  $\phi(\mathbf{x})$  is expanded as [9]

$$\phi(\mathbf{x}) = \phi_{\text{G}}(\mathbf{x}) + f_{\text{NL}} \left[ \phi_{\text{G}}(\mathbf{x})^2 - \langle \phi_{\text{G}}(\mathbf{x})^2 \rangle \right], \quad (1.1)$$

where  $\phi_{\text{G}}$  is a Gaussian distributed random field and  $\langle \dots \rangle$  indicates ensemble averaging. The parameter  $f_{\text{NL}}$  is a constant that quantifies the leading-order departure from Gaussianity (this expansion can continue to include third and higher powers of  $\phi_{\text{G}}$ ). Analyses of the three-point function of the cosmic microwave background (CMB) radiation measured by the Planck satellite currently constrain  $f_{\text{NL}} = -0.9 \pm 5.1$  ( $1\sigma$ ) [10], but next-generation large-scale structure surveys hope to improve upon this bound.

As galaxies form out of the primordial fluctuations, the statistics of their distribution can in principle be used to probe PNG. In cosmologies with local PNG, the deterministic galaxy density contrast  $\delta_g$  can be written to leading order (LO) as [11–13]

$$\delta_g(\mathbf{x}, z) \stackrel{\text{LO}}{=} b_1(z)\delta_m(\mathbf{x}, z) + b_\phi(z)f_{\text{NL}}\phi(\mathbf{q}), \quad (1.2)$$

where  $\delta_m$  is the total matter density contrast and  $\mathbf{q}$  is the initial Lagrangian coordinate associated with the final Eulerian coordinate  $\mathbf{x}$ . In keeping with the effective field theory approach to galaxy statistics and the galaxy bias expansion (see ref. [14] for a comprehensive review), the wavelength of the perturbations  $\delta_m(\mathbf{x}, z)$  and  $\phi(\mathbf{q})$  is implicitly assumed to be much larger than the spatial scales over which galaxy formation takes place. The coefficients  $b_1$  and  $b_\phi$  are called galaxy bias parameters and they describe, respectively, the linear *responses* of the galaxy number density to total mass and primordial potential perturbations with  $f_{\text{NL}} \neq 0$ . The bias parameters effectively encode all of the complicated dependence of galaxy formation on the large-scale environment, are functions not only of redshift but also of the properties of the galaxies considered, and are thus extremely challenging to predict theoretically. The bias parameters are therefore normally fitted alongside the cosmological ones in inference analyses using galaxy clustering data, but as degeneracies between the two sets of parameters arise, one often finds it necessary/beneficial to take theoretical priors on bias into account in order to tighten the constraints on cosmology. These degeneracies are especially critical in  $f_{\text{NL}}$  constraints. For example, ref. [15] showed that the galaxy power spectrum  $P_{gg}(k)$  (the Fourier transform of the two-point correlation function) in cosmologies with local PNG gets a contribution  $\propto b_\phi f_{\text{NL}}/k^2$ , that becomes important on large scales and that can be used to constrain  $f_{\text{NL}}$ . Given the perfect degeneracy between  $f_{\text{NL}}$  and  $b_\phi$ , however, it is simply impossible to use  $P_{gg}$  to constrain  $f_{\text{NL}}$ , unless some theoretical prior on  $b_\phi$  is assumed.<sup>1</sup>

In the large-scale structure literature, the most popular way to circumvent this problem in  $f_{\text{NL}}$  constraints is by relating  $b_\phi$  to  $b_1$  using the so-called *universality relation* [12, 15–17]

$$b_\phi = 2\delta_c(b_1 - 1), \quad (1.3)$$

which follows from assuming that the halo mass function is universal, and  $\delta_c = 1.686$  is the threshold overdensity for spherical collapse. In this way, the contribution scales as  $\propto (b_1 - 1) f_{\text{NL}}/k^2$ , and since  $b_1$  can in principle be constrained using the smaller-scale part of the power spectrum (where  $f_{\text{NL}}$  contributes weakly), it then becomes possible to constrain  $f_{\text{NL}}$ . The universality relation is adopted by almost all existing galaxy data constraints on  $f_{\text{NL}}$  [16, 18–25] (the current tightest bound is  $f_{\text{NL}} = -12 \pm 21$  ( $1\sigma$ ) [25]), as well as in forecast studies [7, 26–36] for next-generation surveys. Despite its widespread adoption, there is however no reason to expect the universality relation to hold for real-life galaxy samples, and in fact, studies using  $N$ -body simulations have been indicating this to be the case already. For example, gravity-only simulations have shown that the  $b_\phi(b_1)$  relation of dark matter halos underpredicts slightly the universality relation [37–41], ref. [42] showed using galaxy formation simulations that the  $b_\phi(b_1)$  relation of stellar-mass selected galaxies overpredicts the universality relation, and refs. [16, 43] discussed how the  $b_\phi(b_1)$  relation of recently-merged objects differs also from the universality prediction; we will return to these findings below when we reproduce some with our numerical results. The impact that uncertainties on the  $b_\phi(b_1)$  relation have on the resulting  $f_{\text{NL}}$  constraints is only now beginning to be explored [44, 45], but

<sup>1</sup>Formally, next-to-leading-order contributions to the galaxy power spectrum (e.g. 1-loop terms and beyond) can break the degeneracy, but they are unimportant given the current observationally allowed range for  $f_{\text{NL}}$ .

it is crucial that these investigations are made robust and mature since they will ultimately directly impact the final constraints on  $f_{\text{NL}}$ .

Beyond leading-order, the contributions at next-to-leading-order (NLO) are<sup>2</sup>

$$\delta_g(\mathbf{x}, z) \stackrel{\text{NLO}}{=} \frac{b_2(z)}{2} [\delta_m(\mathbf{x}, z)]^2 + b_{K^2}(z) [K_{ij}(\mathbf{x}, z)]^2 + b_{\phi\delta}(z) f_{\text{NL}} \phi(\mathbf{q}) \delta_m(\mathbf{x}, z), \quad (1.4)$$

where  $K_{ij} = (\partial_i \partial_j / \nabla^2 - \delta_{ij} / 3) \delta_m$  is a long-wavelength tidal field, and  $b_2$ ,  $b_{K^2}$  and  $b_{\phi\delta}$  are additional galaxy bias parameters. These terms contribute to the galaxy bispectrum  $B_{ggg}(k_1, k_2, k_3)$  (the Fourier transform of the three-point correlation function), which is also a very well-known probe of  $f_{\text{NL}}$  [46–50], and which may prove crucial to combine with the galaxy power spectrum if next-generation galaxy surveys are to improve over the current CMB constraints. For halos, the  $b_2(b_1)$  relation is very well understood [51] and works have been progressively improving our understanding of  $b_{K^2}(b_1)$  as well [52–59]; the corresponding relations for simulated galaxies have also begun to be recently studied [58]. In contrast, the  $b_{\phi\delta}$  parameter has never been the focus of dedicated simulation work. Assuming universality of the halo mass function, one finds [12, 14]

$$b_{\phi\delta} = b_\phi - b_1 + 1 + \delta_c [b_2 - (8/21)(b_1 - 1)], \quad (1.5)$$

but unlike the case for  $b_\phi$ , the performance of this relation has never been checked, even for the simpler case of dark matter halos in gravity-only simulations. There are currently no real-data constraints on  $f_{\text{NL}}$  using the galaxy bispectrum, but existing forecast studies effectively always assume the validity of this relation. Although it is formally possible to constrain  $f_{\text{NL}}$  using the bispectrum without any prior on  $b_{\phi\delta}$ , we will see below that adopting priors based on a  $b_{\phi\delta}(b_1)$  relation is necessary for competitive constraints.

This motivates the main goal of this paper, which is to use separate universe  $N$ -body simulations to obtain predictions for the bias parameters  $b_\phi$  and  $b_{\phi\delta}$ ; to the best of our knowledge, this is the first time that  $b_{\phi\delta}$  is estimated from dedicated simulation data. We will do so not only for the case of dark matter halos in gravity-only simulations, but also for simulated galaxies in hydrodynamical simulations with the IllustrisTNG galaxy formation model. One of our main new results is that the universality relation of eq. (1.5) is not a perfect description for either one of these large-scale structure tracers, with the size of the departures varying depending on the galaxy selection criteria adopted. Recently, ref. [58] showed that the  $b_2(b_1)$  and  $b_{K^2}(b_1)$  relations of halos are broadly preserved for galaxies selected by a variety of criteria (total mass, stellar mass, color and black hole mass accretion rate), which suggests that priors around these relations may be adopted relatively safely in real-data analyses. In contrast, our results below will show that the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations are markedly more sensitive to the selection criteria, which makes the design of theoretical priors more challenging. In a second part of this paper, we build upon the idealized forecast study of ref. [44] to study the impact that uncertainties on  $b_\phi$  and  $b_{\phi\delta}$  can have on  $f_{\text{NL}}$  constraints. We will see that these bias parameters can have a marked impact on the resulting  $f_{\text{NL}}$  bounds. One of our main takeaway messages is that more simulation work is needed to our current level.

The rest of this paper is organized as follows. In section 2 we describe the simulation data we use in this work, as well as our methods to estimate the galaxy bias parameters.

<sup>2</sup>There is an additional contribution  $\delta_g \supset b_{\phi^2} f_{\text{NL}}^2 \phi^2$  that is quadratic in  $f_{\text{NL}}$ , but which is unimportant given the current observational bounds, and so we skip writing it explicitly. Should this contribution have been important, then all our motivations to study  $b_{\phi\delta}$  in this paper would apply equally to  $b_{\phi^2}$ .

Section 3 contains our main numerical results on galaxy bias, where we show first the  $b_{\phi\delta}$  parameter measured in gravity-only simulations, and then compare the sensitivity of the  $b_{\phi}(b_1)$  and  $b_{\phi\delta}(b_1)$  relations to different galaxy selection criteria in hydrodynamical simulations. In section 4 we show and discuss the results of a simple forecast study for a fictitious survey aiming to illustrate the impact that bias uncertainties have on  $f_{\text{NL}}$  constraints. We summarize and conclude in section 5. In appendix A, we show estimates of the bias parameter  $b_2$  using the same method we use to estimate  $b_{\phi\delta}$ . In appendix B, we collect the theory model expressions for the galaxy power spectrum and bispectrum that we use in our forecast study.

## 2 Methodology

In this section we describe the estimation of the two leading-order local PNG galaxy bias parameters  $b_{\phi}$  and  $b_{\phi\delta}$ , as well as the linear bias parameter  $b_1$ , which recall, contribute to the bias expansion as

$$\delta_g(\mathbf{x}, z) \supset b_1(z)\delta_m(\mathbf{x}, z) + b_{\phi}(z)f_{\text{NL}}\phi(\mathbf{q}) + b_{\phi\delta}(z)f_{\text{NL}}\phi(\mathbf{q})\delta_m(\mathbf{x}, z). \quad (2.1)$$

As explained below, we estimate (i)  $b_1$  using the large-scale limit of the galaxy-matter cross-power spectrum; (ii)  $b_{\phi}$  as the response of the galaxy number density to long-wavelength primordial gravitational potentials with  $f_{\text{NL}}$ ; and (iii)  $b_{\phi\delta}$  using the response of  $b_1$  to long-wavelength primordial gravitational potentials with  $f_{\text{NL}}$ .

### 2.1 Simulation data specifications

The simulations we use in this work have been presented previously in ref. [42], and they were run using the moving-mesh gravity+hydrodynamical  $N$ -body code AREPO [60, 61] with the IllustrisTNG model of galaxy formation [62–64]. This model is an improved version of its precursor Illustris [65, 66], and it is characterized by *sub-grid* prescriptions for the physics of gas cooling, star formation, stellar feedback, chemical enrichment, and black hole growth/feedback, which were calibrated to broadly reproduce a number of observations such as the galaxy stellar mass function at low redshift, the star formation rate history, galaxy sizes and the gas mass fractions of galaxies and galaxy groups (see refs. [67–73] for the first results with IllustrisTNG). The initial conditions were generated at  $z_i = 127$  with N-GENIC code [74] using the Zel’dovich approximation, and a linear matter power spectrum calculated using the CAMB code [75]. The galaxy formation simulations were run on a cubic box with size  $L_{\text{box}} = 205 \text{ Mpc}/h$ , containing  $N_p = 1250^3$  dark matter mass elements and  $N_p = 1250^3$  initial gas elements. At this resolution, we have both full-physics hydrodynamical simulations, as well as gravity-only counterparts (we label these as “Gravity”); we follow the standard IllustrisTNG nomenclature and refer to this resolution as TNG300-2. In addition, we consider also a set of gravity-only simulations run also with AREPO with  $N_p = 1250^3$  dark matter mass elements, but on a bigger simulation box size  $L_{\text{box}} = 560 \text{ Mpc}/h \approx 800 \text{ Mpc}$ .

The cosmological parameters of our Fiducial cosmology are: mean baryon density today  $\Omega_{b0} = 0.0486$ , mean total matter density today  $\Omega_{m0} = 0.3089$ , mean dark energy density today  $\Omega_{\Lambda0} = 0.6911$ , dimensionless Hubble rate  $h = 0.6774$ , primordial scalar spectral index  $n_s = 0.967$ , and primordial scalar power spectrum amplitude  $\mathcal{A}_s = 2.068 \times 10^{-9}$  (at  $k_{\text{pivot}} = 0.05/\text{Mpc}$ , corresponding to  $\sigma_8(z=0) = 0.816$ ). In order to measure the galaxy bias parameters  $b_{\phi}$  and  $b_{\phi\delta}$  using the separate universe method (see below), we consider also two additional cosmologies, dubbed High $\mathcal{A}_s$  and Low $\mathcal{A}_s$ , which differ from the Fiducial only in the value of  $\mathcal{A}_s \rightarrow \mathcal{A}_s [1 + \delta\mathcal{A}_s]$ , where  $\delta\mathcal{A}_s = +0.05$  for High $\mathcal{A}_s$  and  $\delta\mathcal{A}_s = -0.05$  for Low $\mathcal{A}_s$ .

We will show measurements of the bias parameters for both halos and subhalos/galaxies. The halos are identified with a Friends-of-Friends (FoF) algorithm run on the dark matter elements with linking length  $b = 0.2$  times the mean interparticle distance. In turn, the subhalos correspond to the gravitationally bound structures found by the SUBFIND algorithm [76] inside each halo. In the hydrodynamical simulations, we refer to the subhalos that contain any mass in stars ( $M_* > 0$ ) as galaxies, and we do not explicitly distinguish between main and satellite galaxies. We also only consider objects with at least 100 member star particles to ensure we deal with objects that are sufficiently well resolved in the simulations. Throughout, we will show results for objects selected by their total mass  $M_t$ , stellar mass  $M_*$ , black hole mass  $M_{\text{BH}}$ , black hole mass accretion rate  $\dot{M}_{\text{BH}}$ , and (dust-uncorrected)  $g - r$  color. When quoting the value of these quantities for a given object (halo or subhalo), we always consider the summed contribution of all member particles to that quantity.

There are two points worth emphasizing about our numerical setup. One is that the TNG300-2 resolution is below that at which the IllustrisTNG model was calibrated at, and as a result, the predictions of our simulations are not in as good agreement with the above-mentioned observations. The galaxy bias predictions are however expected to be less affected by numerical resolution compared to quantities like the galaxy number density itself. In fact, ref. [42] has shown that the  $b_1$  and  $b_\phi$  predictions at TNG300-2 resolution are in very good agreement with those obtained at a higher resolution ( $N_p = 2 \times 1250^3$ ,  $L_{\text{box}} = 75\text{Mpc}/h$ ; called TNG100-1.5 there) that is closer to the nominal IllustrisTNG one ( $N_p = 2 \times 1820^3$ ,  $L_{\text{box}} = 75\text{Mpc}/h$ ). The second point is that (as in past works with separate universe simulations of galaxy formation [42, 77, 78]) we keep the parameters of the IllustrisTNG model fixed when we adjust the parameter  $\mathcal{A}_s$  in the High $\mathcal{A}_s$  and Low $\mathcal{A}_s$  cosmologies. This is the appropriate choice in order to interpret our galaxy bias measurements as predictions of the IllustrisTNG model, i.e., the galaxy bias parameters are the response of galaxy formation to long-wavelength perturbations, at fixed galaxy formation prescription (that of IllustrisTNG).

## 2.2 The linear density and local PNG bias parameters $b_1$ and $b_\phi$

We estimate the linear bias parameter  $b_1$  in the Fiducial simulations using the large-scale limit of the ratio of the galaxy-matter cross-power spectrum  $P_{gm}(k, z)$  and matter power spectrum  $P_{mm}(k, z)$

$$b_1 = \lim_{k \rightarrow 0} \frac{P_{gm}(k)}{P_{mm}(k)}, \quad (2.2)$$

which follows from eq. (2.1) for  $f_{\text{NL}} = 0$  (note we dropped the redshift  $z$  from the arguments to lighten the notation). For the relatively small volume of the TNG300-2 simulations, the scale-dependence of this ratio can still be nonnegligible on the largest scales probed. Thus, rather than simply fitting a constant to it, we account for the leading-order scale-dependence by fitting instead for  $b_1 + Ak^2$ , and take the constant coefficient as our estimate of the bias parameter. In practice, we use all modes with  $k < 0.15h/\text{Mpc}$ , and we have checked that the values of  $b_1$  estimated from the TNG300-2 box agree with those from the  $L_{\text{box}} \approx 800$  Mpc box that is less affected by this systematic. Our error bars on  $b_1$  are the error estimate from the least-squares fitting procedure.

On the other hand, we estimate the linear local PNG bias parameter  $b_\phi$  from the definition

$$b_\phi = \frac{1}{\bar{n}_g} \frac{\partial \bar{n}_g}{\partial (f_{\text{NL}} \phi)} \equiv \frac{4}{\bar{n}_g} \frac{\partial \bar{n}_g}{\partial \delta \mathcal{A}_s}, \quad (2.3)$$



where the first equality follows from eq. (2.1) and the second equality from the separate universe (or peak-background split) equivalence between (i) structure formation inside a long-wavelength potential perturbation  $f_{\text{NL}}\phi$  and (ii) structure formation without the perturbation, but with a modified value of the primordial scalar power spectrum amplitude  $\mathcal{A}_s$  [15, 16]. In particular, it is possible to show (see e.g. section 7.1.2. of ref. [14]) that if  $\phi_L$  is the amplitude of a long-wavelength potential perturbation, then galaxies forming inside it form as they would form in a cosmology with  $\mathcal{A}_s$  rescaled as  $\mathcal{A}_s \rightarrow \mathcal{A}_s [1 + \delta\mathcal{A}_s]$ , with  $\delta\mathcal{A}_s = 4f_{\text{NL}}\phi_L$  (hence the factor of 4 in eq. (2.3)). Concretely, we evaluate eq. (2.3) via finite-differencing using the results of our Fiducial, High $\mathcal{A}_s$  ( $\delta\mathcal{A}_s = +0.05$ ) and Low $\mathcal{A}_s$  ( $\delta\mathcal{A}_s = -0.05$ ) simulations as

$$b_\phi = \frac{b_\phi^{\text{High}\mathcal{A}_s} + b_\phi^{\text{Low}\mathcal{A}_s}}{2}, \quad (2.4)$$

with

$$\begin{aligned} b_\phi^{\text{High}\mathcal{A}_s} &= \frac{4}{+0.05} \left[ \frac{N_g^{\text{High}\mathcal{A}_s}}{N_g^{\text{Fiducial}}} - 1 \right], \\ b_\phi^{\text{Low}\mathcal{A}_s} &= \frac{4}{-0.05} \left[ \frac{N_g^{\text{Low}\mathcal{A}_s}}{N_g^{\text{Fiducial}}} - 1 \right], \end{aligned} \quad (2.5)$$

and where  $N_g$  represents the number of objects in some selection variable bin (total mass, stellar mass, black hole mass, etc.) and the superscripts indicate in which cosmology the number of galaxies is counted.<sup>3</sup> The choice of  $|\delta\mathcal{A}_s| = 0.05$  is motivated by the compromise between having a sizeable and measurable impact of the change in  $\mathcal{A}_s$  on the galaxy abundance, while keeping negligible higher-order corrections to the first-order finite-difference result. Further, having just a single realization of the initial conditions for each cosmology/resolution, it is not possible to estimate our measurement errors in a statistical ensemble sense. As a compromise, we take the difference between the values of  $b_\phi^{\text{High}\mathcal{A}_s}$  and  $b_\phi^{\text{Low}\mathcal{A}_s}$  (which should be the same up to numerical noise) as our estimate of the error on  $b_\phi$ ; see ref. [42] for a discussion of why this yields trustworthy error estimates.

Using the same simulations and methodology, refs. [42, 82] presented an indepth study of the total- and stellar-mass dependence of  $b_\phi$ . Here, we will reproduce some of these past results (while showing also additional ones for other galaxy selection criteria) to better compare with the results for the second-order  $b_{\phi\delta}$  parameter.

### 2.3 The second-order local PNG bias parameter $b_{\phi\delta}$

The same separate universe simulations can be used to estimate the bias parameter  $b_{\phi\delta}$  via the response of  $P_{gm}(k)$  to long-wavelength primordial potential perturbations  $f_{\text{NL}}\phi$ . Concretely, in all our cosmologies, the large-scale galaxy-matter cross-power spectrum is described by  $P_{gm} = b_1 P_{mm}$ . Defining its linear local PNG response function as  $R_{\phi,gm} = \partial \ln P_{gm} / \partial (f_{\text{NL}}\phi)$ , then it follows that (the power spectrum response functions can be defined in analogy to the galaxy bias parameters by treating the local power spectrum as a biased tracer [83])

$$R_{\phi,gm} = \frac{\partial \ln b_1}{\partial (f_{\text{NL}}\phi)} + R_{\phi,mm}, \quad (2.6)$$

<sup>3</sup>In reality, the  $N_g$  correspond to the number of galaxies in cosmologies with local PNG, i.e. the response to  $\delta\mathcal{A}_s$  should be evaluated using simulations with local PNG initial conditions, whereas we do so using Gaussian distributed initial conditions. We note, however, that for the currently allowed values of  $f_{\text{NL}}$ , as well as for the mass scales we consider in this paper, the impact of this approximation is negligible [79–81].



where  $R_{\phi,mm} = \partial \ln P_{mm} / \partial (f_{\text{NL}} \phi)$ . The first term on the right can be worked out by using that

$$b_1 = \frac{1}{\bar{n}_g} \frac{\partial \bar{n}_g}{\partial \delta_m}; \quad b_{\phi\delta} = \frac{1}{\bar{n}_g} \frac{\partial^2 \bar{n}_g}{\partial (f_{\text{NL}} \phi) \partial \delta_m}, \quad (2.7)$$

to yield

$$R_{\phi,gm} = \frac{b_{\phi\delta}}{b_1} - b_\phi + R_{\phi,mm}, \quad (2.8)$$

where we have used also eq. (2.3). The power spectrum response functions  $R_{\phi,gm}$  and  $R_{\phi,mm}$  can be measured straightforwardly using the separate universe simulations analogously to how  $b_\phi$  is estimated in eqs. (2.4) and (2.5). For example, using the Fiducial and High $\mathcal{A}_s$  cosmologies, we would estimate the response as  $R_{\phi,gm}^{\text{High}\mathcal{A}_s} = 4 \left[ P_{gm}^{\text{High}\mathcal{A}_s} / P_{gm}^{\text{Fiducial}} - 1 \right] / \delta \mathcal{A}_s$ . Thus, given the estimates of  $b_1$  and  $b_\phi$  described in the last subsection, as well as measurements of the power spectrum responses  $R_{\phi,gm}$  and  $R_{\phi,mm}$  on large scales, we can use eq. (2.8) to fit for  $b_{\phi\delta}$ . The linear power spectrum response functions effectively describe the mode-coupling structure of squeezed-limit bispectra [83–85], which are sensitive to second-order bias parameters to leading order, and is what allows us to fit for  $b_{\phi\delta}$  (as well as other second-order bias parameters; cf. appendix A). The parameter  $b_{\phi\delta}$  can also be estimated from the response of the galaxy power spectrum  $P_{gg} = b_1^2 P_{mm} + 1/\bar{n}_g$ , although in this case the contribution from shot noise lowers the signal-to-noise unnecessarily; further, it can be straightforwardly shown that the response of this galaxy power spectrum model to  $f_{\text{NL}} \phi$  agrees with the expressions shown in ref. [86] obtained with squeezed-limit bispectra.

The results on  $b_{\phi\delta}$  that we will show below were obtained, however, with a different (yet related) strategy. Rather than using the separate universe simulations to differentiate  $P_{gm}$  and  $P_{mm}$ , we first fit instead for  $b_1$  in the three cosmologies separately using eq. (2.2), and then we differentiate  $b_1$  using finite differences. With this estimate, together with the estimates of  $b_1$  and  $b_\phi$ ,  $b_{\phi\delta}$  is then simply given by

$$b_{\phi\delta} = \left[ \frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} + b_\phi \right] b_1. \quad (2.9)$$

We have explicitly checked that both strategies above give consistent results, but the latter yielded slightly better signal-to-noise estimates, which is why we adopt it as the default. Concretely, we evaluate the response of  $b_1$  as

$$\frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} = \frac{1}{2} \left[ \frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} \right]^{\text{High}\mathcal{A}_s} + \frac{1}{2} \left[ \frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} \right]^{\text{Low}\mathcal{A}_s}, \quad (2.10)$$

with

$$\begin{aligned} \left[ \frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} \right]^{\text{High}\mathcal{A}_s} &= \frac{4}{+0.05} \left[ \frac{b_1^{\text{High}\mathcal{A}_s}}{b_1^{\text{Fiducial}}} - 1 \right], \\ \left[ \frac{\partial \ln b_1}{\partial (f_{\text{NL}} \phi)} \right]^{\text{Low}\mathcal{A}_s} &= \frac{4}{-0.05} \left[ \frac{b_1^{\text{Low}\mathcal{A}_s}}{b_1^{\text{Fiducial}}} - 1 \right], \end{aligned} \quad (2.11)$$

and we estimate its error analogously to as for  $b_\phi$ . The final uncertainty on  $b_{\phi\delta}$  is worked out by standard propagation of uncertainty in eq. (2.9) assuming uncorrelated errors, which is conservative since  $b_\phi$ ,  $b_1$  and its response are measured from the same simulations and so their uncertainties are correlated. In appendix A we validate this strategy to estimate  $b_{\phi\delta}$  by

applying it (with the appropriate modifications described there) to estimate the second-order bias parameter  $b_2$ , whose values can be compared with known results in the literature.

We note before proceeding that the  $b_{\phi\delta}$  parameter can also be estimated by fitting perturbation theory models to the bispectrum measured from simulations with local PNG initial conditions [45]. This requires however very large simulation volumes (in fact still currently out of reach for self-consistent galaxy formation simulations) in order to measure the bispectrum precisely on large-scales where the effects of  $f_{\text{NL}}$  dominate. At fixed volume, separate universe simulations offer thus the ideal method to study the local PNG bias parameters. Within the separate universe approach, there is an alternative way to estimate  $b_{\phi\delta}$  using separate universe simulations that incorporate simultaneously the effects of total mass and primordial gravitational potential perturbations. This would require however additional simulations for higher amplitudes of the long-wavelength modes to be sensitive to the second-order terms in the galaxy bias expansion, similarly to how ref. [51] uses separate universe simulations to estimate higher-order bias parameters such as  $b_2$  and  $b_3$ .

### 3 Galaxy bias results

In this section we show and discuss our main numerical results on the galaxy bias parameters  $b_\phi$  and  $b_{\phi\delta}$ . We begin with the  $b_{\phi\delta}$  parameter measured for halos and subhalos in the gravity-only simulations, and then compare the results of the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations for a number of different galaxy samples in the IllustrisTNG simulations; we show results for galaxies selected by total mass  $M_t$ , stellar mass  $M_*$ , black hole mass  $M_{\text{BH}}$ , black hole mass accretion rate  $\dot{M}_{\text{BH}}$  and  $(g-r)$  color.

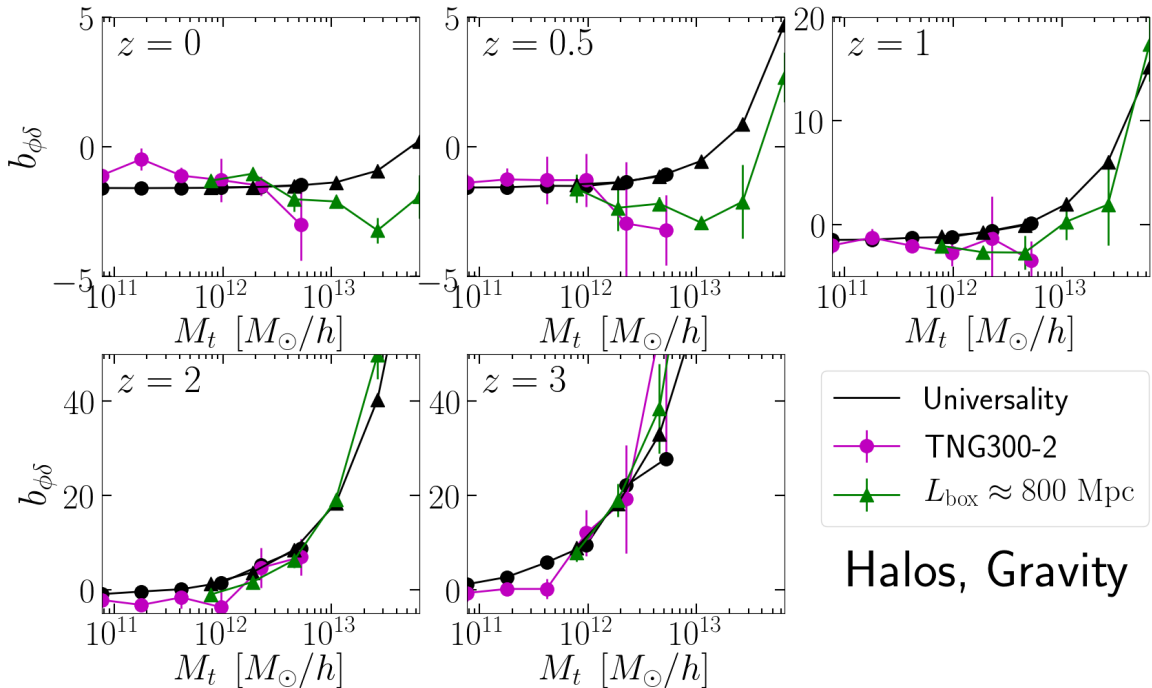
#### 3.1 The $b_{\phi\delta}$ parameter in gravity-only simulations

The total-mass dependence of  $b_{\phi\delta}$  is shown in figure 1 for dark matter halos at different redshifts, and for the gravity-only TNG300-2 and  $L_{\text{box}} \approx 800$  Mpc simulations, as labeled. The result is compared to the universality prediction of eq. (1.5) (shown in black), evaluated using the values of  $b_1$  estimated for the same mass bins, together with the universality relation for the linear local PNG parameter  $b_\phi = 2\delta_c(b_1 - 1)$  and  $b_2(b_1)$  given by the fit obtained for halos with separate universe simulations in ref. [51]:

$$b_2(b_1) = 0.412 - 2.143b_1 + 0.929b_1^2 + 0.008b_1^3. \quad (3.1)$$

The figure shows that the  $b_{\phi\delta}$  values measured in the simulations depart from the universality expectation, with the difference being both redshift- and mass-dependent. Concretely, for  $M_t \lesssim 10^{12}M_\odot/h$ , there is a slight trend for  $b_{\phi\delta}$  to overpredict (less negative) the universality prediction at  $z \leq 0.5$ , but to underpredict it at  $z \geq 2$  (more negative). On the other hand, for  $M_t \gtrsim 10^{12}M_\odot/h$ , the universality relation prediction is visibly above the measured  $b_{\phi\delta}$  at  $z \leq 1$ , but the two are consistent within the errors at  $z \geq 2$ . As one would expect for its smaller volume, the TNG300-2 results appear noisier compared to  $L_{\text{box}} \approx 800$  Mpc, but it is nonetheless possible to discern a good overall agreement between the two resolutions on the mass scales where they overlap.

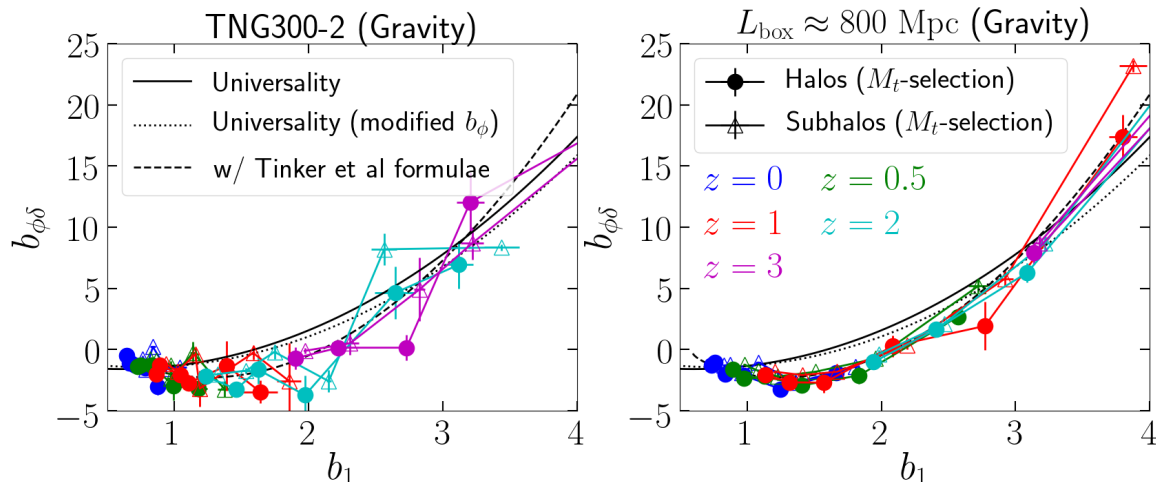
Figure 2 shows the  $b_{\phi\delta}(b_1)$  relation for total-mass selected halos and subhalos in the gravity-only TNG300-2 and  $L_{\text{box}} \approx 800$  Mpc simulations at different redshifts, as labeled. The relation is seen to depend only weakly on redshift; this can be better appreciated in the higher signal-to-noise results from the  $L_{\text{box}} \approx 800$  Mpc box, but the TNG300-2 results display



**Figure 1.** The total-mass dependence of  $b_{\phi\delta}$  for dark matter halos in the gravity-only simulations. The different panels are for different redshifts and the colored points with error bars show the results from the TNG300-2 and  $L_{\text{box}} \approx 800$  Mpc simulations, as labeled. The result in black shows the prediction of the universality relation of eq. (1.5),  $b_{\phi\delta} = b_{\phi} - b_1 + 1 + \delta_c [b_2 - (8/21)(b_1 - 1)]$ , using the values of  $b_1$  measured for the same halos, with  $b_{\phi} = 2\delta_c(b_1 - 1)$  and  $b_2(b_1)$  given by eq. (3.1).

a consistent picture. Further, in the range  $1 \lesssim b_1 \lesssim 3$ , the  $b_{\phi\delta}$  values of the simulations are systematically below (more negative/less positive) the universality prediction shown by the solid black line; the largest difference occurs at  $b_1 \approx 2$  and is  $\Delta b_{\phi\delta} \sim 3$ . For  $b_1 \lesssim 0.8$ , one can also discern a trend for the measured  $b_{\phi\delta}$  to overpredict the universality relation, which is a manifestation of the same trend shown in figure 1 at low redshift and lower masses. Note also that the results for halos and subhalos are consistent within the precision of our measurements, although some differences are in general to be expected (even if small) since the relation between  $b_{\phi\delta}(b_1)$  is nonlinear (see section 3.3 of ref. [58] for a discussion).

The dotted lines in figure 2 show the prediction of a variant of the universality relation of eq. (1.5), in which instead of using the universality relation for  $b_{\phi}$  (cf. eq. (1.3)), one replaces it by  $b_{\phi}(b_1) = 0.85 \times 2\delta_c(b_1 - 1)$ , which offers a more adequate approximation to the  $b_{\phi}(b_1)$  relation in gravity-only simulations [37–42] (see the dotted line in the top left panel of figure 3 below). This variant of the universality relation with a modified  $b_{\phi}(b_1)$  relation does get slightly closer to the simulation measurements over the range  $1 \lesssim b_1 \lesssim 3$ , but not sufficiently to bring the two results into agreement. In other words, the breakdown of the universality relation for the second-order parameter  $b_{\phi\delta}$  cannot be attributed solely to the breakdown of the universality relation for the linear parameter  $b_{\phi}$ . The dashed lines in figure 2 show the outcome of eq. (2.9) evaluated with the halo mass function and halo bias  $b_1$  formulae from Tinker et al. [87, 88]. Concretely, the  $\partial \ln b_1 / \partial (f_{\text{NL}} \phi) \equiv 4 \partial \ln b_1 / \partial (\delta \mathcal{A}_s)$  term is evaluated by finite-differencing the  $b_1$  formula w.r.t.  $\mathcal{A}_s$ , and  $b_{\phi}$  is obtained by finite-differencing the halo mass function also w.r.t.  $\mathcal{A}_s$  as in eq. (2.3) (this approach to  $b_{\phi}$  compares well with



**Figure 2.** The  $b_{\phi\delta}(b_1)$  relation for halos (filled circles) and subhalos (open triangles) in the gravity-only simulations. Each data point shows  $b_{\phi\delta}$  for the mass bins in figure 1, but now plotted against  $b_1$  in the same mass bin. The two panels are for the TNG300-2 and  $L_{\text{box}} \approx 800$  Mpc simulations and the different colors show the result at different redshifts, as labeled. In both panels, the solid black line shows the universality prediction of eq. (1.5),  $b_{\phi\delta} = b_\phi - b_1 + 1 + \delta_c[b_2 - (8/21)(b_1 - 1)]$ , with  $b_\phi = 2\delta_c(b_1 - 1)$  and  $b_2(b_1)$  given by eq. (3.1). The dotted line shows the same but with  $b_\phi = 0.85 \times 2\delta_c(b_1 - 1)$  to account for the departures from universality on this relation. The dashed line shows the outcome of eq. (2.9) obtained using the formulae from Tinker et al. [87, 88].

$b_\phi = 0.85 \times 2\delta_c(b_1 - 1)$ ). This semi-analytical calculation agrees extremely well with our numerical estimates, which represents a good cross-check of both results. Finally, we have also found the following best-fitting quadratic polynomial to the dark matter halo results (fitted up to  $b_1 < 4$  using all redshifts and both simulation boxes):

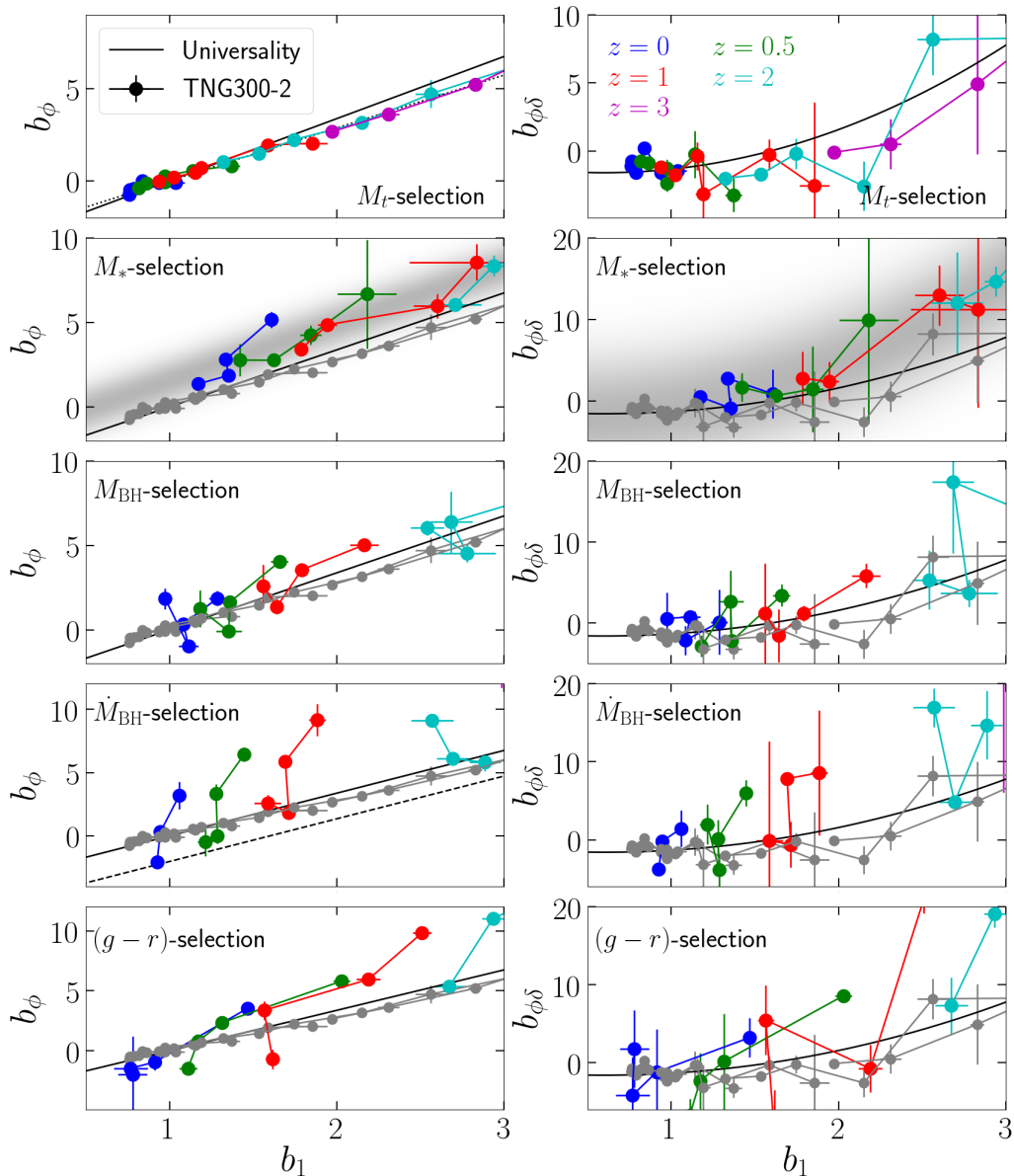
$$b_{\phi\delta}(b_1)^{\text{fit}} = 3.85 - 9.49b_1 + 3.44b_1^2, \quad (3.2)$$

but which we skip showing to avoid crowding the figure.

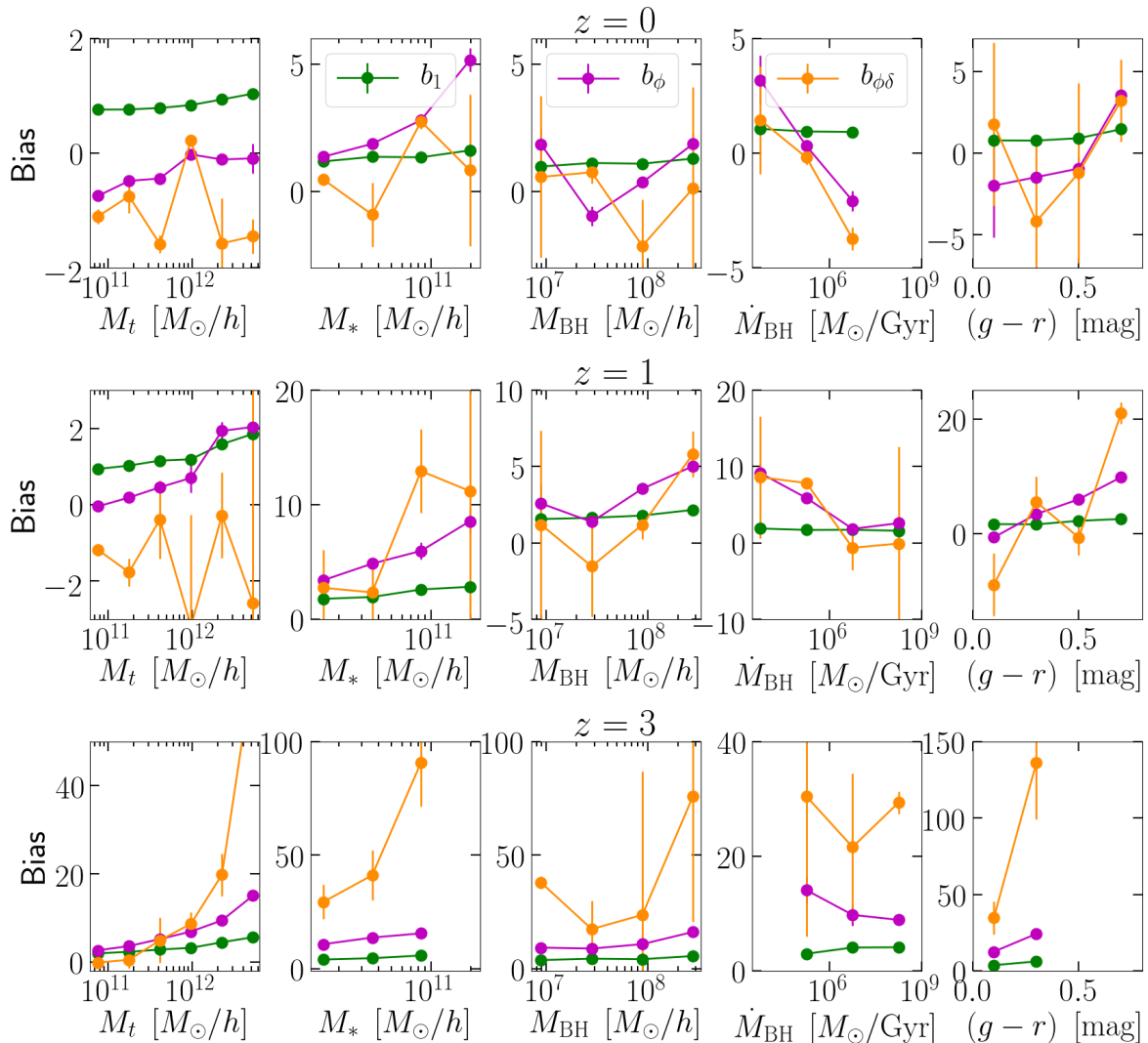
### 3.2 The $b_\phi$ and $b_{\phi\delta}$ parameters in galaxy formation simulations

We turn our attention now to the  $b_\phi$  and  $b_{\phi\delta}$  parameters measured from the hydrodynamical, full-physics IllustrisTNG runs with the TNG300-2 box; in this subsection we focus only on the results for galaxies (i.e., subhalos with mass in stars). Figure 3 shows the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations at different redshifts for galaxies selected by their total mass  $M_t$ , stellar mass  $M_*$ , black hole mass  $M_{\text{BH}}$ , black hole mass accretion rate  $\dot{M}_{\text{BH}}$ , and dust-uncorrected  $(g-r)$  color, as labeled. Concretely, we consider 6  $M_t$  bins log-spaced between  $[5 \times 10^{10}; 8 \times 10^{12}] M_\odot/h$ . For  $M_*$ ,  $M_{\text{BH}}$  and  $\dot{M}_{\text{BH}}$  we consider 4 bins log-spaced between  $[1 \times 10^{10}; 3 \times 10^{11}] M_\odot/h$ ,  $[5 \times 10^6; 5 \times 10^8] M_\odot/h$  and  $[1 \times 10^3; 1 \times 10^9] M_\odot/\text{Gyr}$ , respectively. And for  $(g-r)$  we consider 4 bins linearly spaced within  $[0, 0.8]$ . In figure 4, we show the dependence of the three galaxy bias parameters  $b_1$ ,  $b_\phi$  and  $b_{\phi\delta}$  on the galaxy selection variables themselves for  $z = 0$ ,  $z = 1$  and  $z = 3$ .

For the case of  $b_\phi$ , the total-mass and stellar-mass results in figure 3 (top two panels on the left) have been already discussed in detail in ref. [42] using the same simulations as in this paper. Namely, for  $M_t$ -selection, the figure shows the well-known result that the  $b_\phi(b_1)$  relation of the simulated objects is below the universality expectation for  $b_1 \gtrsim 1.5$ , and that



**Figure 3.** The  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations for IllustrisTNG galaxies. The left and right panels are for  $b_\phi$  and  $b_{\phi\delta}$ , the colors indicate the redshift, and from top to bottom, the panels show the result for galaxies selected by their total mass, stellar mass, black hole mass, black hole mass accretion rate and dust-uncorrected  $(g-r)$  color, as labeled (the  $M_t$  results for  $b_{\phi\delta}$  are shown also in figure 2, but are repeated here to ease comparisons). Each data point shows the value of  $b_\phi$  and  $b_{\phi\delta}$  for the galaxies in some property bin, plotted against the value of  $b_1$  for the same galaxies. The solid lines show the corresponding universality predictions (cf. eqs. (1.3) and (1.5)), and to ease comparisons, the grey points repeat the result of the top panels for total-mass selection. The  $b_\phi(b_1)$  panels for  $M_t$ - and  $\dot{M}_{BH}$ -selection show also the variants of the universality relation  $b_\phi(b_1) = 0.85 \times 2\delta_c(b_1 - 1)$  (dotted) and  $b_\phi(b_1) = 2\delta_c(b_1 - 1.6)$  (dashed), respectively. The grey color maps in the  $M_*$  panels indicate the amplitude of the Gaussian priors of eqs. (3.3) and (3.4). This figure displays results only on the observationally interesting range  $1 \lesssim b_1 \lesssim 3$ , but see figure 4 for bias parameter values that do not appear here.



**Figure 4.** The dependence of the bias parameters  $b_1$  (green),  $b_\phi$  (magenta) and  $b_{\phi\delta}$  (orange) on the galaxy total mass  $M_t$ , stellar mass  $M_*$ , black hole mass  $M_{\text{BH}}$ , black hole mass accretion rate  $\dot{M}_{\text{BH}}$  and dust-uncorrected  $(g-r)$  color, as labeled. From top to bottom, the panels show the result for  $z = 0$ ,  $z = 1$  and  $z = 3$ . This figure helps understand some of the trends seen in the main result of figure 3 on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations.

the modified relation  $b_\phi(b_1) = 0.85 \times 2\delta_c(b_1 - 1)$  (shown by the dotted line) provides a more adequate description. For  $M_*$ -selection, the  $b_\phi(b_1)$  relation of the galaxies systematically overpredicts the universality relation, with  $b_\phi(b_1) = 2\delta_c(b_1 - 0.55)$  being a more faithful approximation of the simulation measurements (center of the grey color map). The lower three panels on the left of figure 3 display the result for galaxies selected by  $M_{\text{BH}}$ ,  $\dot{M}_{\text{BH}}$  and  $(g-r)$ , and they illustrate the strong sensitivity of the  $b_\phi(b_1)$  relation to the galaxy selection criterion adopted, as well as the fact that the universality relation continues to be an inadequate description. We highlight for example the case of the  $\dot{M}_{\text{BH}}$ -selected galaxies, for which the  $b_\phi(b_1)$  relation becomes especially steep, and does not even admit a regular, redshift-independent function of  $b_1$ .



The dashed line in the  $\dot{M}_{\text{BH}}$  panel for  $b_\phi$  shows the prediction of the variant of the universality relation  $b_\phi(b_1) = 2\delta_c(b_1 - 1.6)$ , which was derived by ref. [16] for dark matter halos that had undergone a recent major merger. The authors further hypothesized it could be a better description of the  $b_\phi(b_1)$  relation of quasars compared to the universality relation, as recent mergers may correlate with strong active galactic nuclei (AGN) activity/luminosity. The reason why it is interesting to make this comparison is because AGN luminosity is also thought to be proportional to the accretion rate of the supermassive black hole. The result of figure 3 shows, however, that at least in the IllustrisTNG model, the relation  $b_\phi(b_1) = 2\delta_c(b_1 - 1.6)$  is not a good description of the objects selected by  $\dot{M}_{\text{BH}}$  (or by proxy, selected by their AGN luminosity). With just a single realization of the initial conditions of the simulations, and for a single galaxy formation model, our results do now allow us yet to conclude decisively on the  $b_\phi(b_1)$  relation of real-life quasars and AGN. However, in order to obtain competitive and unbiased constraints on  $f_{\text{NL}}$ , the  $\dot{M}_{\text{BH}}$  results depicted in figure 3 do strongly motivate more works with galaxy formation separate universe simulations to determine the precise bias relations for these objects. Note that the tightest constraints on  $f_{\text{NL}}$  using large-scale structure to date were obtained precisely with quasar samples from eBOSS DR14 [24] and DR16 [25], who assumed  $b_\phi(b_1) = 2\delta_c(b_1 - 1.6)$  in parts of their analysis.

For the case of the  $b_{\phi\delta}$  results on the right of figure 3, a first point to note concerns the lower signal-to-noise of the measurements compared to  $b_\phi$ . This is not surprising since  $b_{\phi\delta}$  is a second-order bias parameter, and our method to estimate it relies on estimating  $b_1$  using the large-scale limit of the galaxy-matter cross-power spectrum (cf. section 2.3), which can be somewhat uncertain on a  $L_{\text{box}} = 205 \text{ Mpc}/h$  box; recall from figure 2 how the signal-to-noise improves substantially from the TNG300-2 to the bigger  $L_{\text{box}} \approx 800 \text{ Mpc}$  box. There are nonetheless a few trends that one can discern. For example, similarly to the  $b_\phi(b_1)$  relation, there is also a visible trend for the  $b_{\phi\delta}$  values at fixed  $b_1$  to be higher for stellar-mass relative to total-mass selection, and the  $b_{\phi\delta}(b_1)$  relation displays similar *steep* variations for  $\dot{M}_{\text{BH}}$ -selected galaxies, i.e.,  $b_{\phi\delta}$  can vary rapidly in a narrow  $b_1$ -interval. Also similarly to  $b_\phi$ , the  $b_{\phi\delta}(b_1)$  relations do not show evidence of being robust to changes in the galaxy selection criterion, although the poorer signal-to-noise here makes it more difficult to draw decisive conclusions.

The details of the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations can be understood by inspecting the corresponding dependencies of the bias parameters on the galaxy selection variables in figure 4. For example, the shape of the dependence of the bias parameters can vary quite significantly from one galaxy property to another, which is the reason behind the strong sensitivity of the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations to the galaxy selection criteria seen in figure 3. It is interesting to contrast this result for  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$ , with the appreciably weaker sensitivity shown in ref. [58] for the  $b_2(b_1)$  and  $b_{K^2}(b_1)$  relations using also IllustrisTNG galaxies. As described in refs. [42, 58, 82], this can be explained using the halo model and halo occupation distribution formalisms, but we leave a more detailed investigation along these lines to future work.

Finally, the grey color maps in the stellar-mass panels in figure 3 describe the shape of Gaussian priors on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations that we will use in the next section to study the impact of galaxy bias uncertainties on  $f_{\text{NL}}$  constraints. Concretely, for  $b_\phi$  and  $b_{\phi\delta}$ , our assumed priors are, respectively,

$$\mathcal{P}(b_\phi|b_1) = \exp \left[ -\frac{1}{2} \frac{(b_\phi - \mu_{b_\phi}(b_1))^2}{\Delta b_\phi^2} \right]; \quad \mu_{b_\phi}(b_1) = 2\delta_c(b_1 - 0.55); \quad \Delta b_\phi = 1; \quad (3.3)$$



$$\mathcal{P}(b_{\phi\delta}|b_1) = \exp \left[ -\frac{1}{2} \frac{(b_{\phi\delta} - \mu_{b_{\phi\delta}}(b_1))^2}{\Delta b_{\phi\delta}^2} \right]; \quad \mu_{b_{\phi\delta}}(b_1) = -1.7 - 1.6b_1 + 2.4b_1^2; \quad \Delta b_{\phi\delta} = 5; \quad (3.4)$$

where, we stress, the mean relations are expected to describe only the results for  $M_*$ -selection, and the standard deviations are assumed  $b_1$ -independent for simplicity and chosen to roughly match the uncertainty in our numerical results with IllustrisTNG.

## 4 Impact of galaxy bias uncertainties on $f_{\text{NL}}$ constraints

In this section we show and discuss the results of an idealized forecast setup for a fictitious survey, to study the impact that uncertainties on  $b_\phi$  and  $b_{\phi\delta}$  have on  $f_{\text{NL}}$  constraints. This analysis effectively extends that of ref. [44] to include also uncertainties on the  $b_{\phi\delta}(b_1)$  relation.

### 4.1 Forecast setup

We work with a Gaussian likelihood function with a data vector consisting of hypothetical measurements of the real-space multitracer power spectrum  $P_{gg}$  [89, 90] and bispectrum  $B_{ggg}$

$$\mathbf{D} = \{ \hat{P}_{gg}^{\text{AA}}, \hat{P}_{gg}^{\text{AB}}, \hat{P}_{gg}^{\text{BB}}, \hat{B}_{ggg}^{\text{AAA}} \}, \quad (4.1)$$

where the superscripts A and B indicate two galaxy samples, e.g.,  $\hat{P}_{gg}^{\text{AB}}$  denotes the cross-power spectrum of the two samples. We consider only the bispectrum of sample A for simplicity; a more involved analysis could include also the bispectrum of sample B and the cross-bispectra  $B_{ggg}^{\text{AAB}}$  and  $B_{ggg}^{\text{ABB}}$ , but this additional complication is not critical to our discussion.

In our theory model, we evaluate the galaxy power spectrum and bispectrum at tree level using the following galaxy bias expansion

$$\begin{aligned} \delta_g(\mathbf{x}, z) = & b_1(z)\delta_m(\mathbf{x}, z) + \frac{1}{2}b_2(z)[\delta_m(\mathbf{x}, z)]^2 + b_{K^2}(z)[K_{ij}(\mathbf{x}, z)]^2 + \epsilon(\mathbf{x}) + \epsilon_\delta(\mathbf{x})\delta_m(\mathbf{x}, z) \\ & + f_{\text{NL}}[b_\phi(z)\phi(\mathbf{q}) + b_{\phi\delta}(z)\phi(\mathbf{q})\delta_m(\mathbf{x}, z) + \epsilon_\phi(\mathbf{x})\phi(\mathbf{q})], \end{aligned} \quad (4.2)$$

where in addition to the deterministic terms that appeared already before, we include now also the relevant stochastic contributions  $\epsilon, \epsilon_\delta, \epsilon_\phi$ . This expansion contains all terms that are needed to self-consistently derive the leading-order galaxy power spectrum and bispectrum [13]. We display the corresponding expressions for the power spectrum, bispectrum and our treatment of the covariance matrix in appendix B.

We consider a galaxy sample at redshift  $z = 1$  covering a volume of  $V = 100\text{Gpc}^3/h^3$ . For the multitracer part of the data vector, we split this galaxy sample into a low- and a high-stellar-mass subsamples, with  $M_* \in [5 \times 10^{10}; 2 \times 10^{11}] M_\odot/h$  for subsample A and  $M_* > 2 \times 10^{11} M_\odot/h$  for subsample B. For the stellar mass function of the IllustrisTNG simulations this corresponds to the number densities  $\bar{n}_g^{\text{A}} = 1.74 \times 10^{-3} [h^3/\text{Mpc}^3]$ ,  $\bar{n}_g^{\text{B}} = 1.07 \times 10^{-4} [h^3/\text{Mpc}^3]$ , and linear bias parameters  $b_1^{\text{A}} = 1.58$ ,  $b_1^{\text{B}} = 2.37$ . We evaluate  $b_2$  and  $b_{K^2}$  for these samples using  $b_2(b_1) = 0.30 - 0.79b_1 + 0.20b_1^2 + 0.12b_1^3$  and  $b_{K^2}(b_1) = 0.66 - 0.57b_1$ , which are fits for IllustrisTNG galaxies obtained by ref. [58]. The fiducial values of  $b_\phi$  and  $b_{\phi\delta}$  are given by the mean relations in eqs. (3.3) and (3.4). We assume Poissonian statistics for the fiducial power spectrum and bispectrum of the noise terms, i.e.,  $P_{\epsilon\epsilon} = 1/\bar{n}_g$ ,  $P_{\epsilon\epsilon_\delta} = b_1/(2\bar{n}_g)$

and  $B_{\epsilon\epsilon\epsilon} = 1/\bar{n}_g^2$  (but note that we sample these in our constraints too; see ref. [45] for the importance of  $k$ -dependent corrections to the shot noise in  $f_{\text{NL}}$  constraints).

We generate our data vector as a noiseless realization of our theory model for the bias parameters and noise terms listed above, and at a fiducial cosmology that has the same parameters as our Fiducial simulations (cf. section 2.1), except for  $f_{\text{NL}} = 5$ . The adoption of a noiseless data vector is naturally idealized, but which we note is the most adequate choice for our purpose here to isolate the impact of galaxy bias uncertainties; see also ref. [45], who takes halo power spectra and bispectra measurements from simulations as the data vector, and so the assessment of the true impact of bias uncertainties gets complicated by the limitations of the theory model to describe the simulation measurements. We consider 42  $k$ -bins between  $k_{\text{min}} = \pi/V_s^{1/3}$  and  $k_{\text{max}} = 0.2 h/\text{Mpc}$  (in intervals of  $k_{\text{min}}$  up to  $0.01h/\text{Mpc}$  and of  $10 \times k_{\text{min}}$  beyond that).

Similarly to ref. [44], we show results for two ways to deal with galaxy bias uncertainties:

- Parametrization 1: direct priors on  $b_\phi$  and  $b_{\phi\delta}$ .

In this case, we sample the following 13 dimensional parameter space:

$$\theta = \{f_{\text{NL}}, b_1^{\text{A}}, b_1^{\text{B}}, P_\epsilon^{\text{A}}, P_\epsilon^{\text{B}}, b_\phi^{\text{A}}, b_\phi^{\text{B}}, \mathcal{A}_s, b_2^{\text{A}}, b_{K^2}^{\text{A}}, b_{\phi\delta}^{\text{A}}, P_{\epsilon\epsilon\delta}^{\text{A}}, B_{\epsilon\epsilon\epsilon}^{\text{A}}\}. \quad (4.3)$$

With this parametrization we will study the impact that different priors on  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  have on  $f_{\text{NL}}$  constraints.

- Parametrization 2: fit for products  $f_{\text{NL}}b_\phi$ ,  $f_{\text{NL}}b_{\phi\delta}$ .

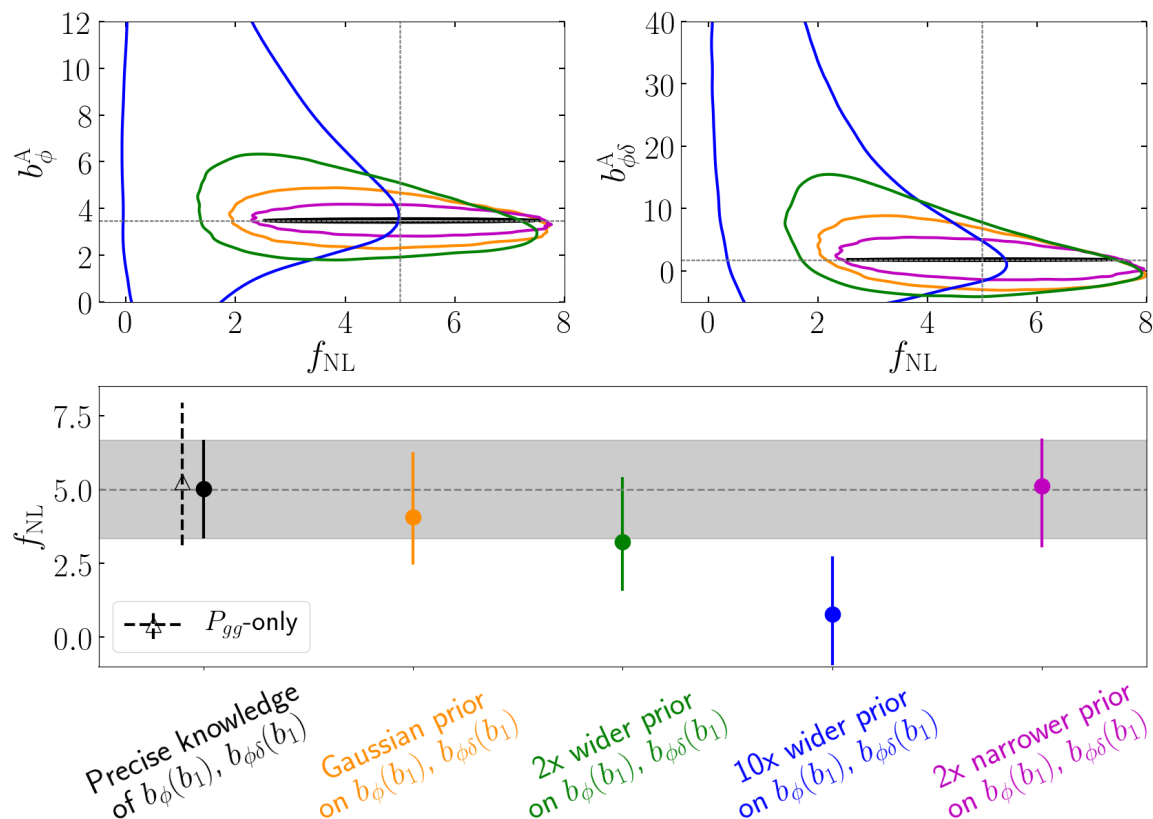
In this case, rather than making assumptions on the bias parameters, we fit directly for the parameter combinations  $f_{\text{NL}}b_\phi$  and  $f_{\text{NL}}b_{\phi\delta}$  as they enter the galaxy power spectrum and bispectrum (cf. appendix B). This approach makes it harder to constrain  $f_{\text{NL}}$  directly, but note it can still be powerful as it can provide evidence for  $f_{\text{NL}} \neq 0$  in a way that is independent of galaxy bias assumptions. This parametrization prevents however direct comparisons of the constraining power of galaxy and CMB data, as well as the combination of galaxy- and CMB-based bounds to obtain a tighter combined bound on  $f_{\text{NL}}$ . The parameter space is also 13 dimensional:

$$\theta = \{[f_{\text{NL}}b_\phi^{\text{A}}], [f_{\text{NL}}b_\phi^{\text{B}}], b_1^{\text{A}}, b_1^{\text{B}}, P_\epsilon^{\text{A}}, P_\epsilon^{\text{B}}, f_{\text{NL}}, \mathcal{A}_s, b_2^{\text{A}}, b_{K^2}^{\text{A}}, [f_{\text{NL}}b_{\phi\delta}^{\text{A}}], P_{\epsilon\epsilon\delta}^{\text{A}}, B_{\epsilon\epsilon\epsilon}^{\text{A}}\}. \quad (4.4)$$

In this part of the analysis, we will also consider cases in which we do not fit for  $[f_{\text{NL}}b_{\phi\delta}^{\text{A}}]$ , and replace it with priors on  $b_{\phi\delta}(b_1)$  to illustrate the breaking of parameter degeneracies.

The majority of our results are for the full data vector of eq. (4.1), but we shall also display results for power-spectrum-only analyses (shown in dashed-black and labeled as  $P_{gg}$ -only). In this case, in parametrizations 1 and 2 we need to consider only the first 7 and 6 parameters, respectively; we keep  $\mathcal{A}_s$  fixed at the fiducial value in these cases. Except for the  $b_\phi$  and  $b_{\phi\delta}$  parameters, we always assume wide uninformative linear priors when sampling the parameter space, which we do using the **EMCEE Python** implementation [91] of the affine-invariant Markov Chain Monte Carlo (MCMC) sampler in ref. [92]. We use 32 *walkers* with a chain convergence criteria that (i) the size of the chain must be 100 times the autocorrelation time and (ii) the latter having varied less than 1% since the last calculation point, which is every few thousand samples.

We stress that our idealized forecast setup serves primarily the purpose to provide a simple framework to visualize the impact that galaxy bias uncertainties can have on  $f_{\text{NL}}$  constraints, and that it is not intended to be representative of any specific current or future survey.

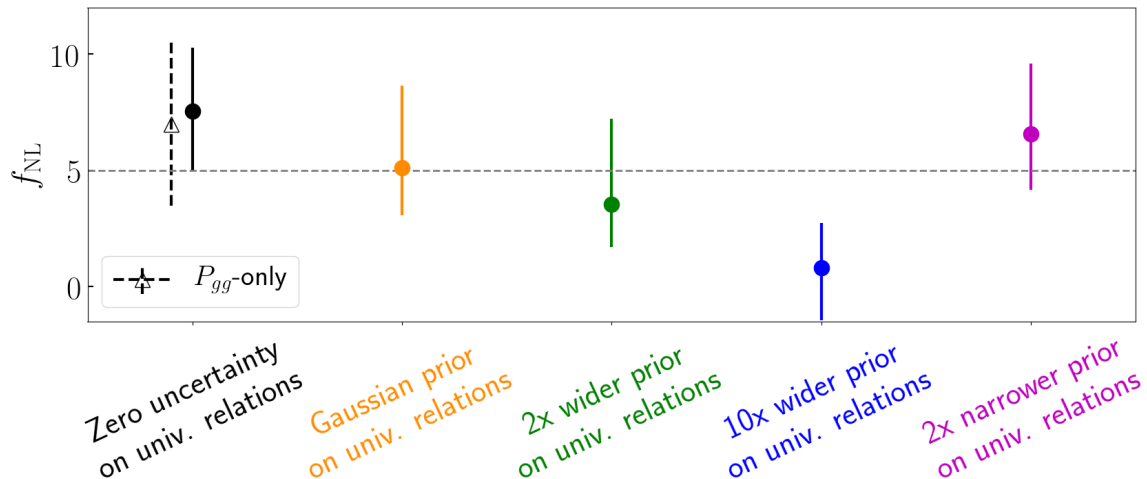


**Figure 5.** Constraint results from parametrization 1 with direct priors on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations. The upper panels show the two-dimensional marginalized  $1\sigma$  constraints on the  $b_\phi^A - f_{\text{NL}}$  and  $b_{\phi\delta}^A - f_{\text{NL}}$  planes, and the lower panel shows the marginalized  $1\sigma$  constraints on  $f_{\text{NL}}$ . The different colors are for different adopted priors on  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$ : the black lines show the result assuming perfect knowledge of the fiducial  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations, the results in orange are for the Gaussian priors of eqs. (3.3) and (3.4), and the remaining are for narrower and wider priors as labeled; all priors are centered around the fiducial bias relations. The grey band in the lower panel simply extends the solid black error bar to ease the comparison, and the dashed grey line marks the fiducial value of  $f_{\text{NL}} = 5$ . The dashed black error bar shows the result obtained with the power spectrum data alone to illustrate the gains from adding bispectrum information; all other results are for the combined power spectrum and bispectrum data.

## 4.2 Results from parametrization 1: direct priors on $b_\phi$ and $b_{\phi\delta}$

The marginalized  $1\sigma$  constraints on  $f_{\text{NL}}$  are shown in the lower panel of figure 5 for parametrization 1. The upper panels show the two-dimensional  $1\sigma$  marginalized constraints on the  $b_\phi^A - f_{\text{NL}}$  and  $b_{\phi\delta}^A - f_{\text{NL}}$  planes; we do not show the contours for all of the 13 parameters for brevity (the interested reader can find some of these triangle plots in ref. [44]). The different colors show the result for different assumed priors on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations, as labeled; the “Gaussian prior” result in orange corresponds to the use of eqs. (3.3) and (3.4) as priors.

Assuming perfect knowledge of the bias relations, our fictitious galaxy sample would be able to constrain  $f_{\text{NL}}$  with a  $1\sigma$  uncertainty of  $\sigma_{f_{\text{NL}}} = 1.7$  (solid black error bar). When we adopt our IllustrisTNG-inspired priors for the bias relations of  $M_*$ -selected galaxies, which recall are centered around the fiducial relations and assume an uncertainty on  $b_\phi(b_1)$  and



**Figure 6.** Same as the lower panel of figure 5, but for priors on  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  that are offset from the fiducial ones. The result in black assumes the universality relations (eqs. (1.3) and (1.5)) with zero uncertainty, the result in orange is for the Gaussian priors of eqs. (3.3) and (3.4) with the mean relations replaced by the universality relations, and the remaining data points are for narrower and wider priors as labeled.

$b_{\phi\delta}(b_1)$  of 1 and 5, respectively (cf. eqs. (3.3) and (3.4)), then the  $f_{\text{NL}}$  constraint becomes  $f_{\text{NL}} = 4.1^{+2.2}_{-1.6}(1\sigma)$  (orange error bar), i.e., the total  $1\sigma$  interval increases by about 10% and the mean distribution value gets shifted from the truth by  $\approx 0.5\sigma_{f_{\text{NL}}}$ . This shift is a manifestation of projection effects that can be understood as follows. In the parts of the parameter space that are close to  $f_{\text{NL}} = 0$  (which our chains still explore), the theory predictions become insensitive to both  $b_\phi$  and  $b_{\phi\delta}$ , which can take on any value allowed by the assumed priors. The wider the priors, the greater the volume of the parameter space that is near the  $f_{\text{NL}} = 0$  direction, and so marginalizing over the poorly constrained  $b_\phi$  and  $b_{\phi\delta}$  will progressively center the marginalized constraints around  $f_{\text{NL}} = 0$ . Indeed, the green and blue error bars show the result for  $2\times$  and  $10\times$  wider priors, which becomes more biased, as expected. On the other hand, halving the width of the priors (magenta), tightens the allowed range of  $b_\phi$  and  $b_{\phi\delta}$  sufficiently to the point where the constraints become effectively the same as the “perfect knowledge” case. A cautionary tale here is that contrary to what one might have naively expected, wide priors on the bias parameters are not conservative and should be interpreted carefully in light of projection effects like these.<sup>4</sup>

In real-life applications, however, the assumed priors will likely not be exactly centered around the bias of the observed galaxies, which introduces additional shifts in the best-fitting  $f_{\text{NL}}$  and its inferred error bar. This is illustrated in figure 6, which shows the same as the lower panel of figure 5, but for offset priors on the bias parameters centered around the universality relations, instead of the fiducial relations used to generate the data vector. Assuming zero uncertainty on the wrong bias relations (black error bar), our fictitious survey would constrain  $f_{\text{NL}} = 7.5 \pm 2.5 (1\sigma)$ , i.e., the inferred value would be  $1\sigma$  away from the fiducial  $f_{\text{NL}} = 5$

<sup>4</sup>We note also for completeness that the constraints on the  $b_\phi - f_{\text{NL}}$  and  $b_{\phi\delta} - f_{\text{NL}}$  planes are bimodal because the goodness-of-fit does not vary dramatically under a simultaneous change of sign of  $b_\phi$ ,  $b_{\phi\delta}$  and  $f_{\text{NL}}$ . This is not visible in the scale of the upper panels in figure 5 for the  $1\sigma$  contours, but see figures 2 and 3 in ref. [44] for a clear visualization of the bimodality of the constraints (the upper right panel of figure 7 in this paper also displays this bimodal nature).

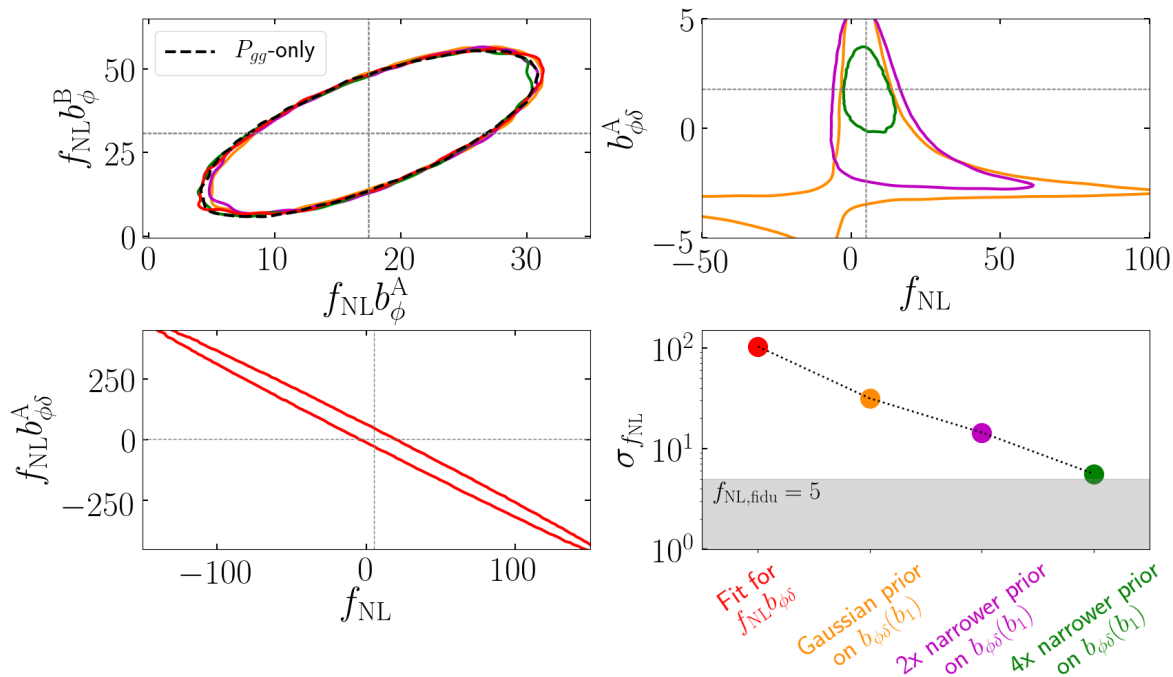
(the upward shift in the constraint and the increase of the size of the error bar compared to figure 5 is as expected since the universality relations underpredict the fiducial bias values). If one is interested only on the overall detection of  $f_{\text{NL}} \neq 0$ , then this represents a  $\approx 3\sigma$  detection, which is the same significance as the constraint  $f_{\text{NL}} = 5.0 \pm 1.7$  ( $1\sigma$ ) in figure 5 assuming perfect knowledge of the fiducial bias relations. Thus, if the true value is  $f_{\text{NL}} \neq 0$ , then from the point of view of *detection significance*, the assumptions on the bias parameters are not as critical. Note, however, that different assumptions on the bias relations still directly impact the inferred precision of the constraint  $\sigma_{f_{\text{NL}}}$ , which can lead to misinterpretations about the true constraining power of the data. For detection significance purposes, a much cleaner approach is thus that based on parametrization 2 discussed below. Further, as the width of the prior increases in figure 6, the constraint progressively approaches  $f_{\text{NL}} = 0$  due to the projection effects discussed above. Interestingly, the result in orange shows the case for widths of  $\Delta b_\phi = 1$  and  $\Delta b_{\phi\delta} = 5$ , for which the projection effects balance the shifts induced by centering the prior around the wrong  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations.

Although not shown, we have also repeated the analyses behind figures 5 and 6, but for a fiducial value of  $f_{\text{NL}} = 0$ , instead of  $f_{\text{NL}} = 5$ . In this case, for both the cases with priors centered around and offset from the fiducial bias relations, the constraints on  $f_{\text{NL}}$  are always unbiased, but too wide priors on  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  eventually also artificially shrink the marginalized error bar due to the projection effects. For the case of assuming perfect knowledge of the bias relations and zero uncertainty around the wrong ones, our fictitious survey would constrain  $|f_{\text{NL}}| < 1.2$  ( $1\sigma$ ) and  $|f_{\text{NL}}| < 2.0$  ( $1\sigma$ ), respectively, further illustrating how wrong assumptions about the bias relations can directly impact the apparent constraining power of the data.

We emphasize that the absolute values of the constraints in figures 5 and 6 correspond strictly to our fictitious galaxy survey, but the relative impact of the galaxy bias uncertainties on those constraints is a more trustworthy measure of what to expect for real surveys. Taken at face value, our results in figure 5 suggest that priors on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations centered around the truth with widths of order  $0.5 - 1$  and  $2.5 - 5$ , respectively (i.e., in between the results shown in orange and magenta in figure 5), may be necessary to guarantee unbiased constraints on  $f_{\text{NL}}$ . Within the IllustrisTNG model, figure 3 suggests that this target might be achievable for the case of  $M^*$ -selected galaxies with more simulations to beat the statistical errors. However, the figure shows that the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations for other selection criteria differ by more than these target widths, and it is still currently unclear how sensitive the bias relations are to changes in the star formation and baryonic feedback models in galaxy formation simulations. Further, recall also the lesson from figure 6 that considerations about the width of the priors on the bias relations cannot be disentangled from discussions about the assumed center values. This all strongly motivates more work to study the bias relations in galaxy formation models beyond IllustrisTNG, as well as for more refined criteria to select the objects, in particular, criteria that resemble as closely as possible those adopted for real galaxies. We highlight that progress along these lines may prove critical to the success of future galaxy surveys to improve over the CMB constraints on local PNG.

### 4.3 Results from parametrization 2: fit for products $f_{\text{NL}}b_\phi$ , $f_{\text{NL}}b_{\phi\delta}$

Our current limited understanding of the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations motivates the exploration of constraints on local PNG using parametrization 2, which constrains the products  $f_{\text{NL}}b_\phi$  and  $f_{\text{NL}}b_{\phi\delta}$ , and avoids the need to put priors on the local PNG bias parameters. For our forecast setup, the overall significance of the detection of local PNG is then determined by the



**Figure 7.** Constraint results from parametrization 2 that fits directly for products  $f_{\text{NL}}b_{\phi}$ ,  $f_{\text{NL}}b_{\phi\delta}$ . The upper panels and the lower left panel show two-dimensional marginalized  $1\sigma$  constraints, whereas the lower right panel shows the marginalized  $1\sigma$  errors on the parameter  $f_{\text{NL}}$  as a function of different assumed priors on  $b_{\phi\delta}$ , as labeled. The results in red correspond to fitting for  $f_{\text{NL}}b_{\phi\delta}$  and do not require priors on  $b_{\phi\delta}$ . All results are for the combined power spectrum and bispectrum data, except that in dashed black in the upper left panel which is only for the power spectrum data.

combined significance of the detection of nonzero values for the parameters  $f_{\text{NL}}b_{\phi}^{\text{A}}$ ,  $f_{\text{NL}}b_{\phi}^{\text{B}}$ ,  $f_{\text{NL}}$  and  $f_{\text{NL}}b_{\phi\delta}^{\text{A}}$ . The results are displayed in figure 7 in red color. For  $f_{\text{NL}}b_{\phi}^{\text{A}}$  and  $f_{\text{NL}}b_{\phi}^{\text{B}}$ , the top left panel shows that our fictitious survey would comfortably detect these parameters at the  $1\sigma$  level. However, as noted previously in ref. [44], these constraints are dominated by the power spectrum part of the data vector, without much contribution from the bispectrum. This can be seen by comparing the solid red contour in the top left panel, which is for the combined power spectrum and bispectrum data, with the dashed black contour for the  $P_{gg}$ -only analysis: the two are barely distinguishable, which shows that the bispectrum constrains the  $f_{\text{NL}}b_{\phi}^{\text{A}}$  and  $f_{\text{NL}}b_{\phi}^{\text{B}}$  parameters only weakly. In parametrization 2, the bispectrum thus contributes to the overall detection of local PNG only via a detection of the parameters  $f_{\text{NL}}$  or  $f_{\text{NL}}b_{\phi\delta}$ , but as the lower left panel of figure 7 shows, these are strongly degenerate and poorly constrained. Thus, while parametrization 2 can prove a powerful way to detect local PNG with power spectrum data without making assumptions on galaxy bias, the price to pay is that the bispectrum would not be able to further enhance the significance of the detection. This can be compared with the case of parametrization 1 in figure 5, in which the bispectrum helps visibly in improving the bounds on  $f_{\text{NL}}$ : for the “perfect knowledge” case, the uncertainty on  $f_{\text{NL}}$  shrinks by about 30% when the bispectrum information is considered (cf. the two error bars in black in the lower panel of figure 5).

As an attempt to “rescue” the constraining power of the bispectrum, we also show results where, in parametrization 2, we replace  $f_{\text{NL}}b_{\phi\delta}^{\text{A}}$  with  $b_{\phi\delta}^{\text{A}}$ , for which we assume Gaussian priors



again. The results are shown in orange, magenta and green in figure 7 for varying prior widths, as labeled. As expected, the upper left panel shows that tightening the prior on  $b_{\phi\delta}^A$  has no visible impact on the  $f_{\text{NL}}b_{\phi}^A$ ,  $f_{\text{NL}}b_{\phi}^B$  constraints, as they are dominated by the power spectrum data where  $b_{\phi\delta}$  does not contribute. The main effect is thus the progressive tightening of the  $f_{\text{NL}}$  bound for tighter  $b_{\phi\delta}$  priors, as shown on the right of figure 7. Specifically, when we fit for  $f_{\text{NL}}b_{\phi\delta}$ , we find  $\sigma_{f_{\text{NL}}} \approx 100$ , but for our default Gaussian prior on  $b_{\phi\delta}$ , a prior that is  $2\times$  narrower, and another prior that is  $4\times$  narrower, the uncertainties become, respectively,  $\sigma_{f_{\text{NL}}} \approx 30$ ,  $\sigma_{f_{\text{NL}}} \approx 15$  and  $\sigma_{f_{\text{NL}}} \approx 5$ . That is, it is only after the prior width on  $b_{\phi\delta}$  becomes of order unity (green;  $5/4 = 1.25 \sim 1$ ) that the constraints return  $\sigma_{f_{\text{NL}}}$  values comparable to our fiducial choice of  $f_{\text{NL}} = 5$ , or in other words, that the bispectrum begins to add to the overall significance of the detection of local PNG.

Before concluding, we clarify the origin of the peculiar shape of the orange contour in the upper right panel of figure 7, namely its sudden widening in the  $f_{\text{NL}}$  direction. It is possible to show that the  $\propto 2b_1^3 f_{\text{NL}}$  and  $\propto b_1^2 f_{\text{NL}} b_{\phi\delta}$  contributions to the galaxy bispectrum (cf. eq. (B.6)) have a similar scale-dependence on the range of scales that have the most constraining power (this is especially the small-scale squeezed bispectrum); this is in fact the reason behind the strong degeneracy between  $f_{\text{NL}}$  and  $f_{\text{NL}}b_{\phi\delta}$  discussed above. If the degeneracy was perfect and if  $b_{\phi\delta} = -2b_1$ , then the bispectrum would become independent of  $f_{\text{NL}}$  in parametrization 2, and unable to constrain it. In reality, the degeneracy is not perfect, but as the chains approach  $b_{\phi\delta}^A = -2b_1^A \approx -3.16$  (which is where the widening occurs), the sensitivity to  $f_{\text{NL}}$  still gets significantly reduced, and the contours get wider.

## 5 Summary and conclusions

Observational constraints on  $f_{\text{NL}}$  using galaxy power spectrum and bispectrum data are determined largely from contributions  $\propto b_{\phi} f_{\text{NL}}$  and  $\propto b_{\phi\delta} f_{\text{NL}}$ , and as a result, competitive constraints on  $f_{\text{NL}}$  using these data require priors on the bias parameters  $b_{\phi}$  and  $b_{\phi\delta}$  to be assumed. The most popular approach encountered in the literature (both in real-data constraints and forecasts) involves using the so-called *universality relations* of eqs. (1.3) and (1.5) (or certain variants thereof) to relate these two parameters to the parameter  $b_1$  that can be constrained using the parts of the data where  $f_{\text{NL}}$  contributes only weakly. The problem is that these relations are derived for dark matter halos assuming that their mass function is universal, and thus, there is no reason to expect them to hold for real-life tracers of the large-scale structure like galaxies.

In fact, prior to this work, it was already known that the universality relation does not describe perfectly the  $b_{\phi}(b_1)$  relation of halos in gravity-only simulations, nor the relation for stellar-mass selected galaxies in hydrodynamical simulations. In this paper, we tested, for the first time with dedicated measurements from separate universe  $N$ -body simulations, the validity of the universality relation for the  $b_{\phi\delta}$  parameter as well, which enters at leading order in the galaxy bispectrum. We carried out our analysis for both dark matter halos in gravity-only simulations, as well as simulated galaxies in hydrodynamical simulations with the IllustrisTNG galaxy formation model. We studied in particular the sensitivity of the  $b_{\phi}(b_1)$  and  $b_{\phi\delta}(b_1)$  relations to different ways of selecting the galaxy samples, including in terms of total mass, stellar mass, black hole mass, black hole mass accretion rate and color. Using an idealized forecast setup with galaxy power spectrum (multitracer) and bispectrum data for a fictitious survey with  $V = 100\text{Gpc}^3/h^3$  at  $z = 1$ , and a fiducial value of  $f_{\text{NL}} = 5$ , we then



explored ways for how to incorporate uncertainties on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations in  $f_{\text{NL}}$  constraint analyses.

Our main results can be summarized as follows:

- The  $b_{\phi\delta}(b_1)$  relation of dark matter halos is approximately redshift-independent, but it is not adequately described by the universality relation. The most notable difference is an overprediction by the latter in the range  $1 \lesssim b_1 \lesssim 3$ , by up to  $\Delta b_{\phi\delta} \approx 3$  (cf. figure 2).
- For the simulated galaxies, the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations are both very sensitive to the galaxy selection criteria, are poorly described by the corresponding universality relations, and do not generically admit a regular, redshift-independent function of  $b_1$  (cf. figure 3).
- Priors on  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  centered close to the fiducial with widths of order 1 and 5, respectively, proved sufficient to yield relatively unbiased  $f_{\text{NL}}$  constraints (cf. figures 5 and 6). This sets a rough target for how precisely these relations may need to be determined from simulations.
- Fitting for products of  $f_{\text{NL}}b_\phi$  and  $f_{\text{NL}}b_{\phi\delta}$  still allows to rule out  $f_{\text{NL}} = 0$  without assumptions on galaxy bias, but renders the bispectrum less useful in the constraints. For our fictitious survey analysis, the bispectrum only begins improving the overall detection of local PNG if priors on  $b_{\phi\delta}(b_1)$  of order  $\lesssim 1$  are assumed (cf. figure 7).

Overall, our current poor understanding of the  $b_\phi$  and  $b_{\phi\delta}$  parameters strongly motivates more works like this one to pin down the expected values of these parameters for real-life galaxy samples. For example, it is important to determine the extent to which the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations depend on the assumed galaxy formation physics. This can be done with simulations for variants of the IllustrisTNG model parameters, or for different self-consistent models of galaxy formation altogether. Given the sensitivity to the galaxy selection criteria that we encountered in this paper, it is also important that future works with galaxy formation simulations attempt also to mimic as closely as possible the selection strategy applied to the real-life galaxy samples considered. Further, beyond galaxies as tracers, it would be interesting to study also the  $b_\phi$  and  $b_{\phi\delta}$  parameters of the gas distribution [82], in particular, the distribution of neutral hydrogen mapped with 21cm line intensity mapping observations.

It is important to acknowledge, however, that the field of cosmological hydrodynamical simulations of galaxy formation has still many associated uncertainties, and that it can be nontrivial to work out the selection functions of real-life galaxies and apply them on simulated ones. This suggests that obtaining very tight theoretical priors on the  $b_\phi(b_1)$  and  $b_{\phi\delta}(b_1)$  relations can prove challenging, but this is work that needs to be carried out anyway to help shape our approach to  $f_{\text{NL}}$  constraints using large-scale structure data. For example, even if at the end of the day the conclusion is that our priors on these bias relations are not satisfactory, this will still be informative in that it will suggest abandoning approaches that focus on the numerical value of  $f_{\text{NL}}$  (like in parametrization 1), and limiting ourselves to analyses based on parametrization 2 more focused on simply detecting  $f_{\text{NL}} \neq 0$ .

We finish by highlighting also the importance for future studies on  $f_{\text{NL}}$  to begin taking galaxy bias uncertainties into account. The strategies that we described in section 4 are straightforward to implement in real-data constraint analyses like those of refs. [24, 25] using eBOSS survey data, as well as in the many forecast codes that exist in the literature for surveys like Euclid [93], SphereX [27] or SKA [94]. Doing so is important not only to guarantee

robust and trustworthy bounds on  $f_{\text{NL}}$ , but also to determine more precisely the theoretical precision requirements on the  $b_\phi$  and  $b_{\phi\delta}$  parameters for future galaxy surveys to reach their  $f_{\text{NL}}$  targets.

## Acknowledgments

We would like to thank Giovanni Cabass, Dragan Huterer, Eiichiro Komatsu, Titouan Lazeyras and Fabian Schmidt for very useful comments and discussions. The author acknowledges support from the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy — EXC-2094-390783311. The numerical analysis of the simulation data presented in this work was done on the Cobra supercomputer at the Max Planck Computing and Data Facility (MPCDF) in Garching near Munich.

## A Validation of the method to estimate $b_{\phi\delta}$ using the $b_2$ parameter

In this appendix we estimate the bias parameter  $b_2$  using a similar method to that used to estimate the  $b_{\phi\delta}$  parameter. This allows us to assess the performance of our method by comparing against known results for the  $b_2(b_1)$  relation of dark matter halos in the literature. Up to a subtle caveat that we comment on below, the method described here is in all similar to that used also by ref. [95].

Starting again from the galaxy-matter cross-power spectrum described by  $P_{gm} = b_1 P_{mm}$ , and defining its linear response to mass perturbations  $\delta_m$  as  $R_{1,gm} = \partial \ln P_{gm} / \partial \delta_m$ , we have

$$R_{1,gm}(k) = \frac{\partial \ln b_1}{\partial \delta_m} + R_{1,mm}(k) = \frac{b_2}{b_1} - b_1 + R_{1,mm}(k), \quad (\text{A.1})$$

where  $R_{1,mm} = \partial \ln P_{mm} / \partial \delta_m$  is the linear response of the matter power spectrum, and in the second equality we have used that  $b_2 = (\partial^2 \bar{n}_g / \partial \delta_m^2) / \bar{n}_g$ . Thus, given measurements of the power spectrum responses from separate universe simulations and estimates of the values of  $b_1$ , we can use the above equation to fit for  $b_2$ . Alternatively, and as we did in our main results, we can estimate  $b_2$  with higher signal-to-noise by computing the response of  $b_1$  measured in the separate universe simulations, i.e., using directly

$$b_2 = \left[ \frac{\partial \ln b_1}{\partial \delta_m} + b_1 \right] b_1, \quad (\text{A.2})$$

where

$$\frac{\partial \ln b_1}{\partial \delta_m} = \frac{1}{2} \left[ \frac{\partial \ln b_1}{\partial \delta_m} \right]^{\text{High}\delta_m} + \frac{1}{2} \left[ \frac{\partial \ln b_1}{\partial \delta_m} \right]^{\text{Low}\delta_m}, \quad (\text{A.3})$$

with

$$\begin{aligned} \left[ \frac{\partial \ln b_1}{\partial \delta_m} \right]^{\text{High}\delta_m} &= \frac{1}{\delta_L^{\text{High}\delta_m}(z)} \left[ \frac{b_1^{\text{High}\delta_m}}{b_1^{\text{Fiducial}}} - 1 \right], \\ \left[ \frac{\partial \ln b_1}{\partial \delta_m} \right]^{\text{Low}\delta_m} &= \frac{1}{\delta_L^{\text{Low}\delta_m}(z)} \left[ \frac{b_1^{\text{Low}\delta_m}}{b_1^{\text{Fiducial}}} - 1 \right]. \end{aligned} \quad (\text{A.4})$$

In these equations,  $\delta_L^{\text{High}\delta_m}(z)$  and  $\delta_L^{\text{Low}\delta_m}(z)$  are the amplitudes of the matter perturbation in the separate universe cosmologies, and the superscripts indicate in which cosmology  $b_1$

is evaluated using eq. (2.2). We evaluate the above equations using the separate universe simulations presented in refs. [42, 77], which correspond to the same Fiducial cosmology used in the main body of the paper, and two separate universe cosmologies, High $\delta_m$  and Low $\delta_m$ , characterized by  $\delta_L^{\text{High}\delta_m}(z=0) = +0.05$  and  $\delta_L^{\text{Low}\delta_m}(z=0) = -0.05$ , respectively.

The caveat that is important to mention here concerns the choice of the box size in these separate universe simulations. Contrary to the changes in  $\mathcal{A}_s$  that characterize the High $\mathcal{A}_s$  and Low $\mathcal{A}_s$  cosmologies in the main body of the paper, the values of  $\delta_L(z)$  modify the time-evolution of the scale-factor  $a(t)$ , and consequently, the relationship between comoving and physical distances. The separate universe simulations in refs. [42, 77] are for fixed comoving size of the boxes at all epochs in units of Mpc. This is important for how to interpret the corresponding response measurements since the finite-differences evaluated with these simulations are then being taken at fixed comoving volume, but strictly, the full responses in the equations above include the effects from changing the volume. For the case of the power spectrum responses, this means that the measured response corresponds actually to the so-called *growth-only* piece. Explicitly, the power spectrum response functions can be decomposed as [85, 96]

$$R_{1,mm}(k) = 1 - \frac{1}{3} \frac{d \ln P_{mm}(k)}{d \ln k} + G_{1,mm}(k), \quad (\text{A.5})$$

$$R_{1,gm}(k) = -\frac{1}{3} \frac{d \ln P_{gm}(k)}{d \ln k} + G_{1,gm}(k), \quad (\text{A.6})$$

where  $G_{1,mm}$  and  $G_{1,gm}$  are the growth-only responses that one measures with fixed-comoving-volume separate universe simulations, and the remaining terms account for so-called *reference density* and *dilation* effects, that describe respectively, the fact that in the separate universe cosmology the power spectrum is measured w.r.t. a modified mean density, and that the physical scales get modified (or *diluted*) by the modified scale factor; the missing +1 term in  $R_{1,gm}$  has to do with the fact that there is one less instance of  $\delta_m$  in  $P_{gm}$  and the galaxy overdensity is typically measured with respect to the observed local, mean galaxy density (and not the global one). Plugging these expressions in eq. (A.1) gives (noting that on large scales the two logarithmic derivative terms are the same)

$$G_{1,gm}(k) = \frac{b_2}{b_1} - b_1 + 1 + G_{1,mm}(k) \equiv \left. \frac{\partial \ln b_1}{\partial \delta_m} \right|_{\text{fixed vol.}} + G_{1,mm}(k), \quad (\text{A.7})$$

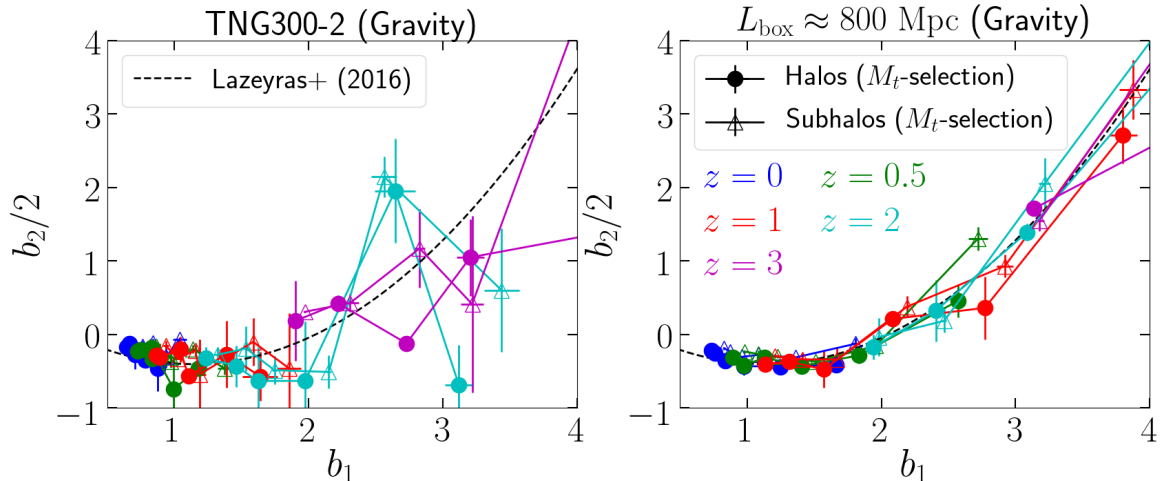
where the second equality defines the fixed-volume derivative of  $b_1$  that we actually evaluate with our separate universe simulations using eqs. (A.3) and (A.4), and which is related to the full, physical derivative as

$$\left. \frac{\partial \ln b_1}{\partial \delta_m} \right|_{\text{fixed vol.}} = \frac{\partial \ln b_1}{\partial \delta_m} + 1. \quad (\text{A.8})$$

We thus estimate  $b_2$  using

$$b_2 = \left[ \left. \frac{\partial \ln b_1}{\partial \delta_m} \right|_{\text{fixed vol.}} + b_1 - 1 \right] b_1. \quad (\text{A.9})$$

The result is in figure 8, which shows the same as figure 2, but for  $b_2$  instead of  $b_{\phi\delta}$ . Our estimated  $b_2(b_1)$  relation recovers well the expected result shown by the dashed line, obtained by ref. [51] for dark matter halos using separate universe simulations (cf. eq. (3.1)). Likewise for  $b_{\phi\delta}$ , the results have higher signal-to-noise in the  $L_{\text{box}} \approx 800$  Mpc, compared to the



**Figure 8.** The  $b_2(b_1)$  relation for halos (filled symbols) and subhalos (open symbols) in the gravity-only simulations. This is the same as figure 2 in the main text, but for  $b_2$ , instead of  $b_{\phi\delta}$ . In both panels, the dashed line shows the fit obtained by ref. [51] for dark matter halos using separate universe simulations (cf. eq. (3.1)).

TNG300-2 box. Overall, this successful recovery of the expected  $b_2(b_1)$  relation validates our estimates of  $b_{\phi\delta}(b_1)$  in the main body of the paper, which were obtained effectively in the same manner.

## B Expressions of the tree-level galaxy power spectrum and bispectrum

In section 4, we evaluate the galaxy power spectrum part of the data vector as (keeping only terms that are leading-order in  $f_{\text{NL}}$ )

$$P_{gg}^{\text{AA}}(k) = [b_1^{\text{A}}]^2 P_{mm}(k) + 2b_1^{\text{A}}b_{\phi}^{\text{A}}f_{\text{NL}}P_{m\phi}(k) + P_{\epsilon\epsilon}^{\text{AA}}, \quad (\text{B.1})$$

$$P_{gg}^{\text{AB}}(k) = b_1^{\text{A}}b_1^{\text{B}}P_{mm}(k) + [b_1^{\text{A}}b_{\phi}^{\text{B}} + b_1^{\text{B}}b_{\phi}^{\text{A}}]f_{\text{NL}}P_{m\phi}(k), \quad (\text{B.2})$$

$$P_{gg}^{\text{BB}}(k) = [b_1^{\text{B}}]^2 P_{mm}(k) + 2b_1^{\text{B}}b_{\phi}^{\text{B}}f_{\text{NL}}P_{m\phi}(k) + P_{\epsilon\epsilon}^{\text{BB}}, \quad (\text{B.3})$$

and the bispectrum part as

$$B_{ggg}(k_1, k_2, k_3) = B_{ggg}^{\text{G}}(k_1, k_2, k_3) + B_{ggg}^{\text{NG}}(k_1, k_2, k_3), \quad (\text{B.4})$$

with the  $f_{\text{NL}}$ -independent part given by

$$\begin{aligned} B_{ggg}^{\text{G}}(k_1, k_2, k_3) &= b_1^3 B_{mmm}(k_1, k_2, k_3) + [2b_1 P_{mm}(k_1) P_{\epsilon\epsilon\delta} + (2 \text{ perm.})] + B_{\epsilon\epsilon\epsilon} \\ &+ [b_1^2 b_2 P_{mm}(k_1) P_{mm}(k_2) + (2 \text{ perm.})] \\ &+ \left[ 2b_1^2 b_{K^2} \left( \mu_{12}^2 - \frac{1}{3} \right) P_{mm}(k_1) P_{mm}(k_2) + (2 \text{ perm.}) \right] \end{aligned} \quad (\text{B.5})$$

and the  $\propto f_{\text{NL}}$  part by

$$\begin{aligned}
 B_{ggg}^{\text{NG}}(k_1, k_2, k_3) = & \left[ 2b_1^3 f_{\text{NL}} \frac{P_{mm}(k_1)P_{mm}(k_2)}{\mathcal{M}(k_1)\mathcal{M}(k_2)} \mathcal{M}(k_3) + 2b_\phi f_{\text{NL}} \frac{P_{mm}(k_1)}{\mathcal{M}(k_1)} P_{\epsilon\epsilon\delta} \right. \\
 & + b_1^2 b_\phi f_{\text{NL}} P_{mm}(k_1) P_{mm}(k_2) \left( \mu_{12} \left( \frac{k_1}{k_2 \mathcal{M}(k_1)} + \frac{k_2}{k_1 \mathcal{M}(k_2)} \right) \right. \\
 & \left. \left. + 2F_2(k_1, k_2, \mu_{12}) \left( \frac{1}{\mathcal{M}(k_1)} + \frac{1}{\mathcal{M}(k_2)} \right) \right) \right. \\
 & \left. + \left( b_1^2 b_{\phi\delta} + b_1 b_2 b_\phi + 2b_1 b_{K^2} b_\phi \left( \mu_{12}^2 - \frac{1}{3} \right) \right) \right. \\
 & \left. \times f_{\text{NL}} P_{mm}(k_1) P_{mm}(k_2) \left( \frac{1}{\mathcal{M}(k_1)} + \frac{1}{\mathcal{M}(k_2)} \right) + (2 \text{ perm.}) \right]. \quad (\text{B.6})
 \end{aligned}$$

In the above expressions,  $\mathcal{M}(k) = (2/3)k^2 T_m(k) / (\Omega_{m0} H_0^2)$  with  $T_m$  the matter transfer function,  $B_{mmm}(k_1, k_2, k_3) = 2F_2(\mathbf{k}_1, \mathbf{k}_2) P_{mm}(k_1) P_{mm}(k_2) + (2 \text{ perm.})$  is the matter bispectrum,  $F_2(k_1, k_2, \mu_{12}) = 5/7 + (\mu_{12}/2) [k_1/k_2 + k_2/k_1] + (2/7)\mu_{12}^2$  is the second-order mode-coupling kernel in perturbation theory [97],  $\mu_{ab}$  is the cosine angle between the wavenumbers  $k_a$  and  $k_b$  in the sides of the bispectrum triangle, and  $P_{m\phi}(k) = P_{mm}(k)/\mathcal{M}(k)$  is the matter-potential cross-power spectrum. We calculate all of the spectra and transfer functions with the CAMB code [75]. Further,  $P_{\epsilon\epsilon}$ ,  $P_{\epsilon\epsilon\delta}$  and  $B_{\epsilon\epsilon\epsilon}$  are the power spectra and bispectra of the stochastic terms.

In our analysis, we consider  $k_{\text{max}} = 0.2h/\text{Mpc}$ , on which corrections to our tree-level theory model do become important, but since the data vector is obtained from the theory model this does not impact our main conclusions on the impact of galaxy bias uncertainties. For simplicity, we also skip modeling the effects of redshift space distortions [98], so-called projection/relativistic effects [28, 30, 35, 99–103], and observational systematics [104]. Any reduction in constraining power from taking these complications into account reduces the overall importance of the uncertainties on galaxy bias, but only in the sense that there are more sources of uncertainty worsening the constraints on  $f_{\text{NL}}$ . Our results in the main body of the paper thus assess how well we need to understand the  $b_\phi$  and  $b_{\phi\delta}$  parameters, assuming negligible sources of error from other aspects of observational constraints on local PNG.

The covariance matrix of the data vector can be written as

$$\mathbf{Cov} = \begin{pmatrix} \mathbf{Cov}^{PP} & \mathbf{Cov}^{BP} \\ \dots & \mathbf{Cov}^{BB} \end{pmatrix}, \quad (\text{B.7})$$

where  $\mathbf{Cov}^{PP}$ ,  $\mathbf{Cov}^{BP}$  and  $\mathbf{Cov}^{BB}$  indicate the covariance of the power spectrum part of the data vector, the cross-covariance of the bispectrum and power spectrum, and the covariance of the bispectrum part, respectively. For  $\mathbf{Cov}^{PP}$  we consider only the *Gaussian+shot noise* contribution, for  $\mathbf{Cov}^{BB}$  we consider in addition to the Gaussian contribution also contributions from leading-order non-Gaussian terms, and we set  $\mathbf{Cov}^{BP} = 0$  for simplicity. We do not repeat our covariance expressions here, but the interested reader can find them in ref. [44]; see also appendix B there for a look into the impact of different covariance compositions (including the cross-covariance term  $\mathbf{Cov}^{BP} = 0$ ) on the final  $f_{\text{NL}}$  bounds, which is another aspect of these analyses that is often overlooked too.

## References

- [1] J.M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, *JHEP* **05** (2003) 013 [[astro-ph/0210603](#)] [[INSPIRE](#)].
- [2] P. Creminelli and M. Zaldarriaga, *Single field consistency relation for the 3-point function*, *JCAP* **10** (2004) 006 [[astro-ph/0407059](#)] [[INSPIRE](#)].
- [3] P. Creminelli, G. D’Amico, M. Musso and J. Norena, *The (not so) squeezed limit of the primordial 3-point function*, *JCAP* **11** (2011) 038 [[arXiv:1106.1462](#)] [[INSPIRE](#)].
- [4] T. Tanaka and Y. Urakawa, *Dominance of gauge artifact in the consistency relation for the primordial bispectrum*, *JCAP* **05** (2011) 014 [[arXiv:1103.1251](#)] [[INSPIRE](#)].
- [5] E. Pajer, F. Schmidt and M. Zaldarriaga, *The Observed Squeezed Limit of Cosmological Three-Point Functions*, *Phys. Rev. D* **88** (2013) 083502 [[arXiv:1305.0824](#)] [[INSPIRE](#)].
- [6] R. de Putter, O. Doré and D. Green, *Is There Scale-Dependent Bias in Single-Field Inflation?*, *JCAP* **10** (2015) 024 [[arXiv:1504.05935](#)] [[INSPIRE](#)].
- [7] M. Alvarez et al., *Testing Inflation with Large Scale Structure: Connecting Hopes with Reality*, [arXiv:1412.4671](#) [[INSPIRE](#)].
- [8] M. Biagetti, *The Hunt for Primordial Interactions in the Large Scale Structures of the Universe*, *Galaxies* **7** (2019) 71 [[arXiv:1906.12244](#)] [[INSPIRE](#)].
- [9] E. Komatsu and D.N. Spergel, *Acoustic signatures in the primary microwave background bispectrum*, *Phys. Rev. D* **63** (2001) 063002 [[astro-ph/0005036](#)] [[INSPIRE](#)].
- [10] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [[arXiv:1905.05697](#)] [[INSPIRE](#)].
- [11] P. McDonald, *Primordial non-Gaussianity: large-scale structure signature in the perturbative bias model*, *Phys. Rev. D* **78** (2008) 123519 [[arXiv:0806.1061](#)] [[INSPIRE](#)].
- [12] T. Giannantonio and C. Porciani, *Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations*, *Phys. Rev. D* **81** (2010) 063530 [[arXiv:0911.0017](#)] [[INSPIRE](#)].
- [13] V. Assassi, D. Baumann and F. Schmidt, *Galaxy Bias and Primordial Non-Gaussianity*, *JCAP* **12** (2015) 043 [[arXiv:1510.03723](#)] [[INSPIRE](#)].
- [14] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1 [[arXiv:1611.09787](#)] [[INSPIRE](#)].
- [15] N. Dalal, O. Dore, D. Huterer and A. Shirokov, *The imprints of primordial non-Gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev. D* **77** (2008) 123514 [[arXiv:0710.4560](#)] [[INSPIRE](#)].
- [16] A. Slosar, C. Hirata, U. Seljak, S. Ho and N. Padmanabhan, *Constraints on local primordial non-Gaussianity from large scale structure*, *JCAP* **08** (2008) 031 [[arXiv:0805.3580](#)] [[INSPIRE](#)].
- [17] S. Matarrese and L. Verde, *The effect of primordial non-Gaussianity on halo bias*, *Astrophys. J. Lett.* **677** (2008) L77 [[arXiv:0801.4826](#)] [[INSPIRE](#)].
- [18] J.-Q. Xia, C. Baccigalupi, S. Matarrese, L. Verde and M. Viel, *Constraints on Primordial Non-Gaussianity from Large Scale Structure Probes*, *JCAP* **08** (2011) 033 [[arXiv:1104.5015](#)] [[INSPIRE](#)].
- [19] A.J. Ross et al., *The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 1116 [[arXiv:1208.1491](#)] [[INSPIRE](#)].



- [20] T. Giannantonio et al., *Improved Primordial Non-Gaussianity Constraints from Measurements of Galaxy Clustering and the Integrated Sachs-Wolfe Effect*, *Phys. Rev. D* **89** (2014) 023511 [[arXiv:1303.1349](#)] [[INSPIRE](#)].
- [21] B. Leistedt, H.V. Peiris and N. Roth, *Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars*, *Phys. Rev. Lett.* **113** (2014) 221301 [[arXiv:1405.4315](#)] [[INSPIRE](#)].
- [22] T. Giannantonio and W.J. Percival, *Using correlations between CMB lensing and large-scale structure to measure primordial non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) L16 [[arXiv:1312.5154](#)] [[INSPIRE](#)].
- [23] S. Ho et al., *Sloan Digital Sky Survey III photometric quasar clustering: probing the initial conditions of the Universe*, *JCAP* **05** (2015) 040 [[arXiv:1311.2597](#)] [[INSPIRE](#)].
- [24] E. Castorina et al., *Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space*, *JCAP* **09** (2019) 010 [[arXiv:1904.08859](#)] [[INSPIRE](#)].
- [25] E.-M. Mueller et al., *The clustering of galaxies in the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Primordial non-Gaussianity in Fourier Space*, [arXiv:2106.13725](#) [[INSPIRE](#)].
- [26] T. Giannantonio, C. Porciani, J. Carron, A. Amara and A. Pillepich, *Constraining primordial non-Gaussianity with future galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **422** (2012) 2854 [[arXiv:1109.0958](#)] [[INSPIRE](#)].
- [27] O. Doré et al., *Cosmology with the SPHEREX All-Sky Spectral Survey*, [arXiv:1412.4872](#) [[INSPIRE](#)].
- [28] D. Alonso, P. Bull, P.G. Ferreira, R. Maartens and M. Santos, *Ultra large-scale cosmology in next-generation experiments with single tracers*, *Astrophys. J.* **814** (2015) 145 [[arXiv:1505.07596](#)] [[INSPIRE](#)].
- [29] D. Alonso and P.G. Ferreira, *Constraining ultralarge-scale cosmology with multiple tracers in optical and radio surveys*, *Phys. Rev. D* **92** (2015) 063525 [[arXiv:1507.03550](#)] [[INSPIRE](#)].
- [30] A. Raccanelli, F. Montanari, D. Bertacca, O. Doré and R. Durrer, *Cosmological Measurements with General Relativistic Galaxy Correlations*, *JCAP* **05** (2016) 009 [[arXiv:1505.06179](#)] [[INSPIRE](#)].
- [31] R. de Putter and O. Doré, *Designing an Inflation Galaxy Survey: how to measure  $\sigma(f_{\text{NL}}) \sim 1$  using scale-dependent galaxy bias*, *Phys. Rev. D* **95** (2017) 123513 [[arXiv:1412.3854](#)] [[INSPIRE](#)].
- [32] D. Karagiannis, A. Lazanu, M. Liguori, A. Raccanelli, N. Bartolo and L. Verde, *Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 1341 [[arXiv:1801.09280](#)] [[INSPIRE](#)].
- [33] A. Moradinezhad Dizgah and G.K. Keating, *Line intensity mapping with [CII] and CO(1-0) as probes of primordial non-Gaussianity*, *Astrophys. J.* **872** (2019) 126 [[arXiv:1810.02850](#)] [[INSPIRE](#)].
- [34] M. Ballardini, W.L. Matthewson and R. Maartens, *Constraining primordial non-Gaussianity using two galaxy surveys and CMB lensing*, *Mon. Not. Roy. Astron. Soc.* **489** (2019) 1950 [[arXiv:1906.04730](#)] [[INSPIRE](#)].
- [35] R. Maartens, S. Jolicoeur, O. Umeh, E.M. De Weerd and C. Clarkson, *Local primordial non-Gaussianity in the relativistic galaxy bispectrum*, *JCAP* **04** (2021) 013 [[arXiv:2011.13660](#)] [[INSPIRE](#)].
- [36] N. Sailer, E. Castorina, S. Ferraro and M. White, *Cosmology at high redshift — a probe of fundamental physics*, *JCAP* **12** (2021) 049 [[arXiv:2106.09713](#)] [[INSPIRE](#)].



- [37] M. Grossi et al., *Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 321 [[arXiv:0902.2013](#)] [[INSPIRE](#)].
- [38] V. Desjacques, U. Seljak and I. Iliev, *Scale-dependent bias induced by local non-Gaussianity: A comparison to N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **396** (2009) 85 [[arXiv:0811.2748](#)] [[INSPIRE](#)].
- [39] A. Pillepich, C. Porciani and O. Hahn, *Universal halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions*, *Mon. Not. Roy. Astron. Soc.* **402** (2010) 191 [[arXiv:0811.4176](#)] [[INSPIRE](#)].
- [40] N. Hamaus, U. Seljak and V. Desjacques, *Optimal Constraints on Local Primordial Non-Gaussianity from the Two-Point Statistics of Large-Scale Structure*, *Phys. Rev. D* **84** (2011) 083509 [[arXiv:1104.2321](#)] [[INSPIRE](#)].
- [41] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques and F. Schmidt, *Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **468** (2017) 3277 [[arXiv:1611.04901](#)] [[INSPIRE](#)].
- [42] A. Barreira, G. Cabass, F. Schmidt, A. Pillepich and D. Nelson, *Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG*, *JCAP* **12** (2020) 013 [[arXiv:2006.09368](#)] [[INSPIRE](#)].
- [43] B.A. Reid, L. Verde, K. Dolag, S. Matarrese and L. Moscardini, *Non-Gaussian halo assembly bias*, *JCAP* **07** (2010) 013 [[arXiv:1004.1637](#)] [[INSPIRE](#)].
- [44] A. Barreira, *On the impact of galaxy bias uncertainties on primordial non-Gaussianity constraints*, *JCAP* **12** (2020) 031 [[arXiv:2009.06622](#)] [[INSPIRE](#)].
- [45] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques and J. Noreña, *Primordial Non-Gaussianity from Biased Tracers: Likelihood Analysis of Real-Space Power Spectrum and Bispectrum*, *JCAP* **05** (2021) 015 [[arXiv:2010.14523](#)] [[INSPIRE](#)].
- [46] R. Scoccimarro, E. Sefusatti and M. Zaldarriaga, *Probing primordial non-Gaussianity with large-scale structure*, *Phys. Rev. D* **69** (2004) 103513 [[astro-ph/0312286](#)] [[INSPIRE](#)].
- [47] E. Sefusatti and E. Komatsu, *The Bispectrum of Galaxies from High-Redshift Galaxy Surveys: Primordial Non-Gaussianity and Non-Linear Galaxy Bias*, *Phys. Rev. D* **76** (2007) 083004 [[arXiv:0705.0343](#)] [[INSPIRE](#)].
- [48] D. Jeong and E. Komatsu, *Primordial non-Gaussianity, scale-dependent bias, and the bispectrum of galaxies*, *Astrophys. J.* **703** (2009) 1230 [[arXiv:0904.0497](#)] [[INSPIRE](#)].
- [49] T. Baldauf, U. Seljak and L. Senatore, *Primordial non-Gaussianity in the Bispectrum of the Halo Density Field*, *JCAP* **04** (2011) 006 [[arXiv:1011.1513](#)] [[INSPIRE](#)].
- [50] E. Sefusatti, M. Crocce and V. Desjacques, *The Halo Bispectrum in N-body Simulations with non-Gaussian Initial Conditions*, *Mon. Not. Roy. Astron. Soc.* **425** (2012) 2903 [[arXiv:1111.6966](#)] [[INSPIRE](#)].
- [51] T. Lazeyras, C. Wagner, T. Baldauf and F. Schmidt, *Precision measurement of the local bias of dark matter halos*, *JCAP* **02** (2016) 018 [[arXiv:1511.01096](#)] [[INSPIRE](#)].
- [52] R.K. Sheth, K.C. Chan and R. Scoccimarro, *Nonlocal Lagrangian bias*, *Phys. Rev. D* **87** (2013) 083002 [[arXiv:1207.7117](#)] [[INSPIRE](#)].
- [53] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, *Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum*, *Phys. Rev. D* **86** (2012) 083540 [[arXiv:1201.4827](#)] [[INSPIRE](#)].
- [54] T. Lazeyras and F. Schmidt, *Beyond LIMD bias: a measurement of the complete set of third-order halo bias parameters*, *JCAP* **09** (2018) 008 [[arXiv:1712.07531](#)] [[INSPIRE](#)].

- [55] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura and P. McDonald, *Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum*, *Phys. Rev. D* **90** (2014) 123522 [[arXiv:1405.1447](#)] [[INSPIRE](#)].
- [56] M.M. Abidi and T. Baldauf, *Cubic Halo Bias in Eulerian and Lagrangian Space*, *JCAP* **07** (2018) 029 [[arXiv:1802.07622](#)] [[INSPIRE](#)].
- [57] A. Eggemeier, R. Scoccimarro, R.E. Smith, M. Crocce, A. Pezzotta and A.G. Sánchez, *Testing one-loop galaxy bias: Joint analysis of power spectrum and bispectrum*, *Phys. Rev. D* **103** (2021) 123550 [[arXiv:2102.06902](#)] [[INSPIRE](#)].
- [58] A. Barreira, T. Lazeyras and F. Schmidt, *Galaxy bias from forward models: linear and second-order bias of IllustrisTNG galaxies*, *JCAP* **08** (2021) 029 [[arXiv:2105.02876](#)] [[INSPIRE](#)].
- [59] T. Lazeyras, A. Barreira and F. Schmidt, *Assembly bias in quadratic bias parameters of dark matter halos from forward modeling*, *JCAP* **10** (2021) 063 [[arXiv:2106.14713](#)] [[INSPIRE](#)].
- [60] V. Springel, *E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh*, *Mon. Not. Roy. Astron. Soc.* **401** (2010) 791 [[arXiv:0901.4107](#)] [[INSPIRE](#)].
- [61] R. Pakmor et al., *Improving the convergence properties of the moving-mesh code AREPO*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 1134 [[arXiv:1503.00562](#)] [[INSPIRE](#)].
- [62] R. Weinberger et al., *Simulating galaxy formation with black hole driven thermal and kinetic feedback*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 3291 [[arXiv:1607.03486](#)].
- [63] A. Pillepich et al., *Simulating Galaxy Formation with the IllustrisTNG Model*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 4077 [[arXiv:1703.02970](#)] [[INSPIRE](#)].
- [64] D. Nelson et al., *The IllustrisTNG Simulations: Public Data Release*, [arXiv:1812.05609](#) [[INSPIRE](#)].
- [65] S. Genel et al., *Introducing the Illustris Project: the evolution of galaxy populations across cosmic time*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 175 [[arXiv:1405.3749](#)] [[INSPIRE](#)].
- [66] M. Vogelsberger et al., *Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 1518 [[arXiv:1405.2921](#)] [[INSPIRE](#)].
- [67] F. Marinacci et al., *First results from the IllustrisTNG simulations: radio haloes and magnetic fields*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 5113 [[arXiv:1707.03396](#)] [[INSPIRE](#)].
- [68] A. Pillepich et al., *First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 648 [[arXiv:1707.03406](#)] [[INSPIRE](#)].
- [69] J.P. Naiman et al., *First results from the IllustrisTNG simulations: a tale of two elements — chemical evolution of magnesium and europium*, *Mon. Not. Roy. Astron. Soc.* **477** (2018) 1206 [[arXiv:1707.03401](#)].
- [70] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [[arXiv:1707.03397](#)] [[INSPIRE](#)].
- [71] D. Nelson et al., *First results from the IllustrisTNG simulations: the galaxy colour bimodality*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 624 [[arXiv:1707.03395](#)] [[INSPIRE](#)].
- [72] D. Nelson et al., *First Results from the TNG50 Simulation: Galactic outflows driven by supernovae and black hole feedback*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 3234 [[arXiv:1902.05554](#)] [[INSPIRE](#)].
- [73] A. Pillepich et al., *First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 3196 [[arXiv:1902.05553](#)] [[INSPIRE](#)].

- [74] V. Springel, *N-GenIC: Cosmological structure initial conditions*, Astrophysics source code library (February 2015).
- [75] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [[astro-ph/9911177](#)] [[INSPIRE](#)].
- [76] V. Springel, S.D.M. White, G. Tormen and G. Kauffmann, *Populating a cluster of galaxies. 1. Results at  $z = 0$* , *Mon. Not. Roy. Astron. Soc.* **328** (2001) 726 [[astro-ph/0012055](#)] [[INSPIRE](#)].
- [77] A. Barreira et al., *Separate Universe Simulations with IllustrisTNG: baryonic effects on power spectrum responses and higher-order statistics*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 2079 [[arXiv:1904.02070](#)] [[INSPIRE](#)].
- [78] A. Barreira, G. Cabass, D. Nelson and F. Schmidt, *Baryon-CDM isocurvature galaxy bias with IllustrisTNG*, *JCAP* **02** (2020) 005 [[arXiv:1907.04317](#)] [[INSPIRE](#)].
- [79] S. Matarrese, L. Verde and R. Jimenez, *The abundance of high-redshift objects as a probe of non-Gaussian initial conditions*, *Astrophys. J.* **541** (2000) 10 [[astro-ph/0001366](#)] [[INSPIRE](#)].
- [80] M. LoVerde, A. Miller, S. Shandera and L. Verde, *Effects of Scale-Dependent Non-Gaussianity on Cosmological Structures*, *JCAP* **04** (2008) 014 [[arXiv:0711.4126](#)] [[INSPIRE](#)].
- [81] M. LoVerde and K.M. Smith, *The non-Gaussian halo mass function with  $f_{NL}$ ,  $g_{NL}$  and  $\tau_{NL}$* , *JCAP* **08** (2011) 003 [[arXiv:1102.1439](#)] [[INSPIRE](#)].
- [82] R. Voivodic and A. Barreira, *Responses of Halo Occupation Distributions: a new ingredient in the halo model & the impact on galaxy bias*, *JCAP* **05** (2021) 069 [[arXiv:2012.04637](#)] [[INSPIRE](#)].
- [83] A. Barreira and F. Schmidt, *Responses in Large-Scale Structure*, *JCAP* **06** (2017) 053 [[arXiv:1703.09212](#)] [[INSPIRE](#)].
- [84] C.-T. Chiang, C. Wagner, F. Schmidt and E. Komatsu, *Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum*, *JCAP* **05** (2014) 048 [[arXiv:1403.3411](#)] [[INSPIRE](#)].
- [85] C. Wagner, F. Schmidt, C.-T. Chiang and E. Komatsu, *The angle-averaged squeezed limit of nonlinear matter  $N$ -point functions*, *JCAP* **08** (2015) 042 [[arXiv:1503.03487](#)] [[INSPIRE](#)].
- [86] E. Castorina and A. Moradinezhad Dizgah, *Local Primordial Non-Gaussianities and Super-Sample Variance*, *JCAP* **10** (2020) 007 [[arXiv:2005.14677](#)] [[INSPIRE](#)].
- [87] J.L. Tinker et al., *Toward a halo mass function for precision cosmology: The limits of universality*, *Astrophys. J.* **688** (2008) 709 [[arXiv:0803.2706](#)] [[INSPIRE](#)].
- [88] J.L. Tinker et al., *The Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests*, *Astrophys. J.* **724** (2010) 878 [[arXiv:1001.3162](#)] [[INSPIRE](#)].
- [89] P. McDonald and U. Seljak, *How to measure redshift-space distortions without sample variance*, *JCAP* **10** (2009) 007 [[arXiv:0810.0323](#)] [[INSPIRE](#)].
- [90] U. Seljak, *Extracting primordial non-Gaussianity without cosmic variance*, *Phys. Rev. Lett.* **102** (2009) 021302 [[arXiv:0807.1770](#)] [[INSPIRE](#)].
- [91] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306 [[arXiv:1202.3665](#)] [[INSPIRE](#)].
- [92] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*, *Comm. App. Math. Comp. Sci.* **5** (2010) 65.
- [93] EUCLID collaboration, *Euclid Definition Study Report*, [arXiv:1110.3193](#) [[INSPIRE](#)].
- [94] A. Weltman et al., *Fundamental physics with the Square Kilometre Array*, *Publ. Astron. Soc. Austral.* **37** (2020) e002 [[arXiv:1810.02680](#)] [[INSPIRE](#)].

- [95] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Linear response to long wavelength fluctuations using curvature simulations*, *JCAP* **09** (2016) 007 [[arXiv:1511.01465](#)] [[INSPIRE](#)].
- [96] Y. Li, W. Hu and M. Takada, *Super-Sample Covariance in Simulations*, *Phys. Rev. D* **89** (2014) 083519 [[arXiv:1401.0385](#)] [[INSPIRE](#)].
- [97] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [[astro-ph/0112551](#)] [[INSPIRE](#)].
- [98] M. Tellarini, A.J. Ross, G. Tasinato and D. Wands, *Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions*, *JCAP* **06** (2016) 014 [[arXiv:1603.06814](#)] [[INSPIRE](#)].
- [99] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Galaxy Bias and non-Linear Structure Formation in General Relativity*, *JCAP* **10** (2011) 031 [[arXiv:1106.5507](#)] [[INSPIRE](#)].
- [100] J. Yoo, *General Relativistic Description of the Observed Galaxy Power Spectrum: Do We Understand What We Measure?*, *Phys. Rev. D* **82** (2010) 083508 [[arXiv:1009.3021](#)] [[INSPIRE](#)].
- [101] A. Challinor and A. Lewis, *The linear power spectrum of observed source number counts*, *Phys. Rev. D* **84** (2011) 043516 [[arXiv:1105.5292](#)] [[INSPIRE](#)].
- [102] D. Jeong, F. Schmidt and C.M. Hirata, *Large-scale clustering of galaxies in general relativity*, *Phys. Rev. D* **85** (2012) 023504 [[arXiv:1107.5427](#)] [[INSPIRE](#)].
- [103] M.S. Wang, F. Beutler and D. Bacon, *Impact of Relativistic Effects on the Primordial Non-Gaussianity Signature in the Large-Scale Clustering of Quasars*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 2598 [[arXiv:2007.01802](#)] [[INSPIRE](#)].
- [104] M. Rezaie et al., *Primordial non-Gaussianity from the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey I: Catalogue preparation and systematic mitigation*, *Mon. Not. Roy. Astron. Soc.* **506** (2021) 3439 [[arXiv:2106.13724](#)] [[INSPIRE](#)].