Consistency conditions and primordial black holes in single field inflation

Ogan Özsoy$^{\heartsuit}$, Gianmassimo Tasinato$^{\spadesuit}$

$\spadesuit$ *Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom*
$\heartsuit$ *CEICO, Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 21, Prague.*

Abstract

We discuss new consistency relations for single field models of inflation capable of generating primordial black holes (PBH), and their observational implications for CMB μ-space distortions. These inflationary models include a short period of non-attractor evolution: the scale-dependent profile of curvature perturbation is characterised by a pronounced dip, followed by a rapid growth leading to a peak responsible for PBH formation. We investigate the squeezed and the collapsed limits of three and four point functions of curvature perturbation around the dip, showing that they satisfy consistency relations connecting their values to the total amplification of the curvature spectrum, and to the duration of the non-attractor era. Moreover, the corresponding non-Gaussian parameters are scale-dependent in proximity of the dip, with features that again depend on the amplification of the spectrum. For typical PBH scenarios requiring an order $\mathcal{O}(10^7)$ enhancement of the spectrum from large towards small scales, we generally find values $f_{\rm NL}^{\rm sq} = \mathcal{O}(10)$ and $\tau_{\rm NL}^{\rm col} = \mathcal{O}(10^3)$ in a range of scales that can be probed by CMB μ-space distortions. Using these consistency relations, we carefully analyse how the scale-dependence of non-Gaussian parameters leads to characteristic features in $\langle \mu T \rangle$ and $\langle \mu\mu \rangle$ correlators, providing distinctive probes of inflationary PBH scenarios that can be tested using well-understood CMB physics.

Contents

1 Introduction

Primordial black holes (PBH) might constitute a fraction of the dark matter of our universe being formed by the gravitational collapse of overdense regions in the early stages of our universe evolution [1–3]. Such overdense regions can be produced by an amplification of the spectrum of primordial curvature fluctuations from inflation [4, 5]. We refer the reader to the recent reviews [6–9] for constraints on PBH populations and details on their formation mechanisms.

Within the framework of single-field inflation, the enhancement of the spectrum of fluctuations can occur during a brief phase of non-attractor evolution, for example associated with an inflection point in the inflaton potential (for explicit constructions, see [10–17]). In this case, the would-be decaying mode of curvature fluctuations is actually active at superhorizon scales, where it plays an important role in determining the amplitude of the curvature spectrum. In particular, it causes an enhancement of its amplitude by several orders of magnitude from large towards small scales – see Fig 1 for a representative example. Despite the large variety of single-field scenarios, there are universal features that are common to all models with a short-phase of non–slow-roll evolution. First, the spectrum as a function of momentum does not grow faster than k^4 in its way towards the peak [18], see for example Fig 1. Second, the growth of the spectrum is preceded by a pronounced dip, whose position is controlled by the total enhancement of the spectrum [19]. In [20], building on the original idea of [21], we showed that large squeezed non-Gaussianity in proximity of the dip of the spectrum induces sizeable cross-correlations between CMB μ-distortions and temperature fluctuations, which can be used as a test of PBH scenarios based on single-field inflation. Interestingly, this proposal is based only on well understood perturbation theory at large cosmological scales, and does not need to consider complex non-linear phenomena occurring at PBH formation.

In this work we make further steps in exploring this subject, by determining new properties for the statistics of curvature fluctuations around the dip of the spectrum, and their consequences for CMB μ-distortion anisotropies. We first show that appropriate limits of n-point correlators of curvature fluctuations satisfy new consistency relations, connecting their maximal values to the total amplification of the spectrum. For example, calling Π_T the total amplification of the spectrum from large to small scales – and having in mind a typical values a $\Pi_T \simeq 10^7$ for generating PBH – we focus on the parameters $f_{\rm NL}^{\rm sqz}$ and $\tau_{\rm NL}^{\rm coll}$ controlling respectively the squeezed bispectrum and collapsed trispectrum. We find that around the dip these quantities satisfy universal consistency relations, and have maximal values that scale as $f_{\rm NL}^{\rm sqz} \simeq \Pi_T^{1/4} \simeq \mathcal{O}(10)$, and $\tau_{\rm NL}^{\rm coll} \simeq \Pi_T^{1/2} \simeq \mathcal{O}(10^3)$ with respect to Π_T, up to overall order one coefficients determined by the duration of the non-attractor epoch. Moreover, such squeezed and collapsed limits are scale-dependent (see Fig 3), and their features are again controlled by the total amplification of the spectrum. We prove our consistency relations for non-Gaussian parameters in two ways. In section 2 we make use of a heuristic approach based on [19], which allows one to get a simple, intuitive understanding of the physics behind the single-field system we consider. Then, in section 3, we re-derive the consistency conditions using the rigorous gradient expansion approach first introduced [22], so to place the results of section 2 in firmer footings. In section 4, we study the implications of our findings for CMB μ-distortions. In this respect, we go beyond the analysis carried in [20] by developing the following points: First, we show how the information provided by the consistency relations allows us to carry on a more detailed analysis of $\langle \mu T \rangle$ correlators, whose quantitative and qualitative features depend on the properties of the scale-dependent squeezed bispectrum. Then, we study for the first time the implications for the $\langle \mu\mu \rangle$ self-correlator of a scale-dependent collapsed trispectrum around the dip. Finally, at the light of the results above, we discuss improved estimates for the detectability of non-Gaussian consistency relations with PIXIE or PRISM-like experiment, and physical implications for PBH populations. We conclude our work with a Discussion in Section 5, followed by technical appendixes.

2 Consistency relations at the dip: a heuristic approach

We begin by developing a heuristic, intuitive approach for characterising the statistics of curvature perturbations. We consider single-field inflationary scenarios that include a short phase of non-slow-roll evolution, which can lead to PBH production. We characterize the scale-dependence of the curvature perturbation spectrum, that in turn implies new consistency relations for squeezed limits of n-point functions. The results of these heuristic considerations will be supported by a more rigorous analysis in section 3. Phenomenological implications are then developed in section 4, where we apply our findings to CMB μ-distortions as a probe of early universe scenarios leading to PBH formation.
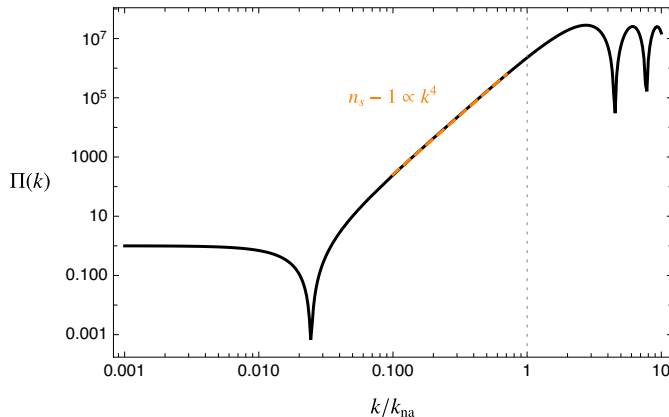

Figure 1: *A representative plot of the spectrum of fluctuations in inflationary scenarios including short phase of non-attractor expansion. Horizontal axis: $k/k_{\rm na}$, with $k_{\rm na}$ the scale of horizon-crossing at the onset of the non-slow-roll epoch. Vertical axis: ratio $\Pi(k)$ of the spectrum evaluated at scale k, versus its large-scale value at $k = 0$. The orange dashed line has a profile proportional to k^4.*

A short phase of non-attractor evolution during inflation is a common ingredient of inflationary scenarios leading to primordial black hole production [10, 12–17, 23–29]. During the non-attractor epoch, the slow-roll conditions are violated, and the would-be decaying mode influences the dynamics of curvature fluctuations, enhancing by several orders of magnitude the amplitude of the corresponding spectrum. Fig 1 shows a representative plot for the enhancement of the spectrum of curvature fluctuations caused by a non-attractor phase. We notice interesting features that are universal in single-field models of inflation:

- As first shown in [18], the spectrum towards the peak grows as function of momentum k with a power not larger [1] than k^4.

- The growth of the spectrum towards the peak is preceded by a deep dip, whose properties depend on the total amplification of the spectrum, and on the duration of the non-attractor era.

In this section, using methods based on [19], we discuss universal properties of curvature fluctuations at the dip of the spectrum, in the limit of short duration of the non-attractor era. The spectral index around the dip, as well as the squeezed limits of the bispectrum and trispectrum in its proximity, satisfy consistency relations that connect their values with the total size of amplification of the spectrum during the non-attractor regime. In the next section, using a gradient formalism, we prove our results more rigorously, and extend them to include a sizeable duration of the non-slow-roll era during inflation.

---

[1]Some exceptions to this conclusion can be made through a prolonged non-attractor era with $\eta_c = -1$ [30] (See also [31]) which can result with a growth rate of $k^5(\ln k)^2 > k^4$ or through multiple non-attractor phases that exhibit an almost instantaneous transition [19, 32].

In the limit of short duration of non-attractor regime, the equations controlling the curvature fluctuations and the corresponding spectrum at the dip position simplify considerably. We assume a conformally flat Friedmann-Lemaitre-Roberson-Walker (FLRW) background metric during inflation, with scale factor $a(\tau)$ and $-\infty < \tau < 0$ the conformal time. The system consists of three phases:

1. An initial prolonged phase of quasi de Sitter expansion, for $\tau < \tau_1$ (we neglect any slow-roll corrections in this epoch, and in epoch 3 below).

2. A short phase of non-attractor expansion for $\tau_1 \leq \tau \leq \tau_2$, whose duration is parameterized by $\Delta\tau_A = \tau_2 - \tau_1$.

3. A final phase of inflationary quasi de Sitter expansion, lasting $\Delta\tau_B = -\tau_2$ for $\tau_2 \leq \tau \leq 0$. We assume $\Delta\tau_A \ll \Delta\tau_B$, and this relation quantifies the limit of short-duration of the non-attractor process.

The succession of these three phases requires appropriate matchings at times $\tau_{1,2}$: this fact has phenomenological implications that we will discuss in due time.

The quadratic action for a Fourier mode of comoving curvature perturbation $\mathcal{R}_k$ is given by

$$S_k^{(2)} = \frac{1}{2} \int d\tau \, z^2(\tau) \left[\mathcal{R}_k'^2(\tau) - k^2 \mathcal{R}_k^2(\tau) \right], \tag{2.1}$$

with $z(\tau) = a\dot{\phi}/H$ a function of time only, dubbed *pump field*. During the quasi-de Sitter phases of inflationary evolution, $\tau < \tau_1$ and $\tau > \tau_2$, the pump field is proportional to the scale factor $a(\tau)$, with (nearly) constant coefficient depending on the inflationary model. During the phase of non-attractor, instead, the time profile of the pump field changes considerably, potentially inducing amplifications of the spectrum of curvature fluctuations.

We define

$$\beta_h \equiv \frac{\Delta\tau_A}{\Delta\tau_B} \left[2 + \left(\frac{d \ln z^2(\tau)}{d \ln \tau} \right)_{\tau=\tau_1} \right], \tag{2.2}$$

as a (possibly) large parameter associated with the variations in the pump field during the non-slow-roll phase $\tau_1 \leq \tau \leq \tau_2$. In the limit of pure de Sitter expansion in the intervals $\tau < \tau_1$ and $\tau > \tau_2$, and considering small values for the ratio $\Delta\tau_A/\Delta\tau_B$, the work [19] determined the following analytical expression for the mode function $\mathcal{R}_k(\tau)$ during the final phase of evolution $\tau > \tau_2$:

$$\mathcal{R}_k(\tau) = \frac{\mathcal{R}_0}{\sqrt{2\,k^3}} \left[\mathcal{C}_1(k)\,e^{-ik\tau}\,(1+ik\tau) + \mathcal{C}_2(k)\,e^{ik\tau}\,(1-ik\tau) \right], \qquad \tau \geq \tau_2\,. \tag{2.3}$$

The previous expression contains an overall constant normalization $\mathcal{R}_0$ common to all modes k, whose value depends on the inflationary model and is not relevant for our arguments. The scale-dependent functions $\mathcal{C}_{1,2}(k)$ read as

$$\mathcal{C}_1(k) = 1 + \frac{\beta_h}{8k^2 \Delta\tau_A \, \Delta\tau_B} \left(1 - e^{2ik\Delta\tau_A} - 2ik\Delta\tau_A \left(1 + 2ik\Delta\tau_B \right) \right), \tag{2.4}$$

$$\mathcal{C}_2(k) = -\frac{\beta_h \, e^{-2ik\tau_2}}{8k^2 \Delta\tau_A \, \Delta\tau_B} \left(1 - e^{2ik\Delta\tau_A} \left(1 - 2ik\Delta\tau_B \right) - 2ik \left(\Delta\tau_A + \Delta\tau_B \right) \right). \tag{2.5}$$

When $\beta_h = 0$, the configuration reduces to the usual solution for fluctuations during a de Sitter phase of expansion. Notice that the solution (2.3), (2.4), (2.5) satisfies the Bunch-Davies condition at early times. We refer the reader to [19] for a complete discussion regarding the formula (2.3).

Equation (2.3) allows us to compute the spectrum of curvature fluctuations at the end of inflation $\tau = 0$, which corresponds to the end of the second slow-roll epoch following the intermediate non-attractor phase. In our context, we are interested to study the enhancement of the curvature spectrum from large

$k \to 0$ towards small scales. Denoting with a prime $\langle \dots \rangle'$ the n-point correlators without the corresponding momentum-conserving δ-functions, we consider the ratio

$$\Pi(k) \equiv \frac{\langle \mathcal{R}_k^2(\tau = 0) \rangle'}{\langle \mathcal{R}_{k=0}^2(\tau = 0) \rangle'} \tag{2.6}$$

of correlation functions evaluated at the end of inflation. The resulting function $\Pi(k)$ quantifies the amplification of the spectrum from large to small scales, and depends on the scale k of horizon exit for modes produced at early times during inflation. In our case it reads:

$$\Pi(k) = |\mathcal{C}_1(k) + \mathcal{C}_2(k)|^2. \tag{2.7}$$

The maximal value of $\Pi(k)$ at small scales, which we define as Π_T can be analytically determined from the expressions (2.4), (2.5) as a simple function of β_h:

$$\Pi_T \equiv \Pi(k \to \infty) = \left(1 + \frac{\beta_h}{2} \right)^2. \tag{2.8}$$

In the limit of large β_h, the value of Π_T can be large, enhancing the spectrum to the values needed for producing PBHs. For example, a β_h of orders of a few thousands can produce an enhancement $\Pi_T \simeq 10^7$, the typical enhancement of the curvature spectrum required for PBH formation mechanisms.

The position of the dip of the power spectrum. In the present instance, we are interested in using the previous formulas for examining in details the properties of the dip feature we notice in Fig 1, which corresponds to a universal property of single-field models with a short non-attractor phase. In particular we are interested to its position, and the scale-dependence of the spectrum in its proximity. We take the simplifying limit of very small duration of non-attractor phase, $\Delta\tau_A/\Delta\tau_B \to 0$. We then identify the momentum scale k_{na}:

$$k_{\mathrm{na}} = 1/\Delta\tau_B = -1/\tau_2 \tag{2.9}$$

with the scale at which modes start leaving the horizon during the non-attractor era (while for larger scales $k \le k_{\mathrm{na}}$ modes leave the horizon during the initial de Sitter phase). We introduce the dimensionless quantity

$$\kappa = k/k_{\mathrm{na}}. \tag{2.10}$$

to describe the quantity $\Pi(\kappa)$ that characterize the scale dependence of the power spectrum as

$$\Pi(\kappa) = 1 - 2\beta_h \kappa j_1(\kappa) \cos\kappa + \beta_h^2 \kappa^2 j_1^2(\kappa), \tag{2.11}$$

with $j_1(\kappa)$ the spherical Bessel function:

$$j_1(\kappa) = \frac{1}{\kappa} \left(\frac{\sin\kappa}{\kappa} - \cos\kappa \right). \tag{2.12}$$

Notice that the spectrum $\Pi(\kappa)$ in eq (2.11) depends on a single extra parameter β_h, which is the only free parameter available describing the impact of the very short phase of non-attractor dynamics. As shown in Figure 2, (2.11) describes very well the global features we mentioned above including the intermediate dip feature and the expected k^4 growth that follows the dip towards small scales.

In the limit of large values of β_h (and $\kappa \le 1$), it is straightforward to compute the position of the dip, using $1/\beta_h$ as expansion parameter. For this purpose, we introduce a convenient variable x through the
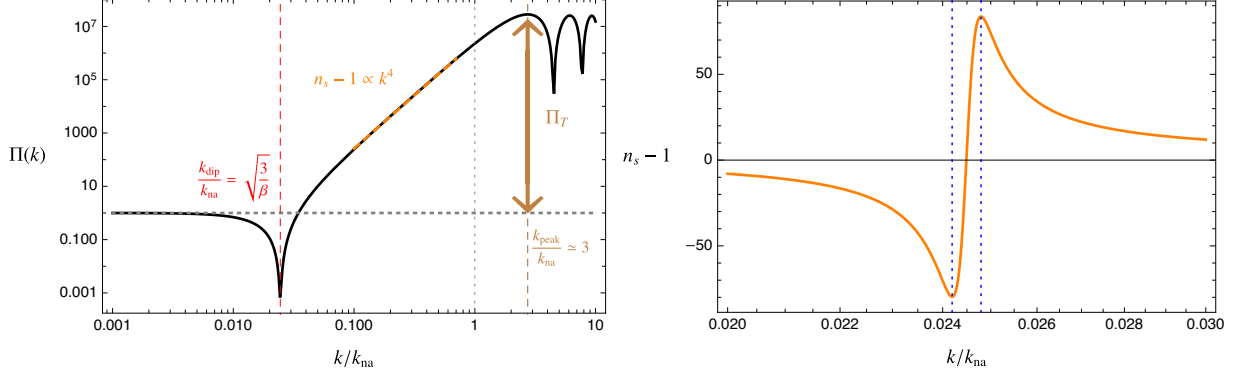
Figure 2: *Left: Scale dependence of the Power spectrum and its important global features. Right: Scale dependence of the spectral index $n_s - 1$ around the dip feature. In the right panel, the location of the minimum/maximum of the spectral index $n_s - 1$ (see e.g. eq. (2.20)) are indicated with blue dashed lines. In both plots we use $\beta_h = 5 \times 10^3$.*

rescaling

$$\kappa \equiv \sqrt{\frac{3}{\beta_h}} x, \tag{2.13}$$

to express $\Pi(\kappa)$ of eq (2.11) as (see *e.g.* eq. (B.10)),

$$\Pi(x) = \left(1 - x^2\right)^2 + \frac{3x^4 \left(6 - x^2\right)}{5\beta_h} + \mathcal{O}\left(\frac{1}{\beta_h^2}\right). \tag{2.14}$$

To leading order in the large β_h limit, $x = 1$ therefore locates the position of the universal dip feature [19]. In terms of the original variable $\kappa = k/k_{\mathrm{na}}$, the location and depth of the dip feature are then given by

$$\frac{k_{\mathrm{dip}}}{k_{\mathrm{na}}} = \sqrt{\frac{3}{\beta_h}} = \sqrt{\frac{3}{2}} \Pi_T^{-1/4} \quad \longrightarrow \quad \Pi\left(\frac{k_{\mathrm{dip}}}{k_{\mathrm{na}}}\right) = \frac{3}{2} \Pi_T^{-1/2}, \tag{2.15}$$

where we used (2.8) to relate the position of the dip and its depth to the total enhancement in the spectrum. These simple interesting relations connect the position of the dip with an inverse power of the total enhancement Π_T of the spectrum. In the limit of large Π_T, the position of the dip in momentum space is parametrically well smaller than the scale at which modes start to leave the horizon during the non-attractor era k_{na}, *i.e.* $\kappa_{\mathrm{dip}} = k_{\mathrm{dip}}/k_{\mathrm{na}} \ll 1$. *The position of dip scale in momentum space therefore corresponds to modes that leave the horizon during the initial attractor slow-roll regime.* For example, for a typical PBH forming scenario with $\Pi_T \sim 10^7$, we find $\kappa_{\mathrm{dip}} \sim 10^{-2}$.

We will reconsider and prove relations similar to (2.15) with more rigorous methods in section 3 to further relate these features to the additional properties of the system such as the duration of the non-attractor era and the value of the slow-roll parameter during this epoch.

The spectral index. We continue our discussion by studying the behaviour of the spectral tilt

$$n_s - 1 = \frac{\mathrm{d} \ln \Pi(\kappa)}{\mathrm{d} \ln \kappa}, \tag{2.16}$$

around the position of the dip. Using the simplified profile of eq (2.14) and (2.13) in (2.16), we present the behaviour of the spectral index around the dip feature in the right panel of Fig 2. The plot shows that the spectral index can acquire values of order $\mathcal{O}(10)$ around the dip. We can utilize the same formulas to analytically determine the maximum and the minimum acquired by $n_s - 1$, focusing on the zeros of the

6

running defined as

$$\alpha_s \equiv \frac{\mathrm{d}n_s}{\mathrm{d}\ln k}.\tag{2.17}$$

With this aim, we notice from the right panel of Figure 2 that extremal points of the spectral tilt are located in the close vicinity of the dip feature. We therefore we make the convenient substitution in terms of an expansion in the small quantity $1/\beta_h$,

$$\kappa = \sqrt{\frac{3}{\beta_h}}\left(1 + \frac{x_1}{\sqrt{\beta_h}} + \frac{x_2}{\beta_h}\right),\tag{2.18}$$

for two yet to be determined quantities $x_{1,2}$. Using (2.14) and (2.13), we find that α_s (2.17) vanishes with the following choices for the parameters $x_{1,2}$:

$$x_1 = \pm\frac{\sqrt{3}}{2}, \qquad x_2 = -\frac{9}{40}.\tag{2.19}$$

Then the position of the extrema of the spectral tilt are located in

$$\frac{k_{\mathrm{min/max}}}{k_{\mathrm{na}}} = \sqrt{\frac{3}{\beta_h}}\left(1 \mp \frac{\sqrt{3}}{2\sqrt{\beta_h}} - \frac{9}{40\beta_h}\right),\tag{2.20}$$

with corresponding value of the minimum and maximum value of the $n_s - 1$:

$$(n_s - 1)_{\mathrm{min/max}} = 2 \mp \frac{2\sqrt{\beta_h}}{\sqrt{3}} = 2 \mp \sqrt{\frac{8}{3}}\,\Pi_T^{1/4},\tag{2.21}$$

in good agreement with the profile shown in the right panel of Figure 2. These formulas provide new consistency relations for the extremal values of the spectral index around the dip in single-field inflationary models which include a short non-slow-roll phase. Interestingly, these relations all involve combinations of the quantity $\Pi_T^{1/4}$, which controls the position of the dip and the properties of the slope of the spectrum in its proximity.

Implications for non-Gaussianity

We find that the location of the dip feature occurs at momentum scales much smaller than the scale at which inflationary modes leave the horizon during the non-slow-roll phase. In this regime, since modes that leave the horizon around the dip scale are associated with the initial single-field slow-roll epoch, we can reasonably expect the validity of Maldacena consistency relations for squeezed and collapsed limits of n-point correlation functions. Namely, defining $P_1 \equiv \langle \mathcal{R}_{k_1} \mathcal{R}_{k_1} \rangle'$, $k_{12} = |\vec{k}_1 + \vec{k}_2|$, and

$$f_{\mathrm{NL}}^{\mathrm{sq}} = \frac{5}{12}\lim_{k_1 \to 0}\frac{\langle \mathcal{R}_{k_1}\mathcal{R}_{k_2}\mathcal{R}_{k_3}\rangle'}{P_1\,P_2}, \qquad \tau_{\mathrm{NL}}^{\mathrm{col}} = \frac{1}{4}\lim_{k_{12}\to 0}\frac{\langle \mathcal{R}_{k_1}\mathcal{R}_{k_2}\mathcal{R}_{k_3}\mathcal{R}_{k_4}\rangle'}{P_1\,P_3\,P_{12}},\tag{2.22}$$

we expect the following relations to hold [33–38],

$$f_{\mathrm{NL}}^{\mathrm{sq}}(k) = \frac{5}{12}\left(1 - n_s(k)\right), \qquad \tau_{\mathrm{NL}}^{\mathrm{col}}(k) = \left(\frac{6}{5}f_{\mathrm{NL}}^{\mathrm{sqz}}(k)\right)^2,\tag{2.23}$$

respectively for the squeezed bispectrum and collapsed bispectrum. We represent the scale dependence of these non-linearity parameters in Figure 3. Following our discussions above, we find that for large values of the enhancement factor Π_T the maximal/minimal value for $f_{\mathrm{NL}}^{\mathrm{sq}}$ and the extremal values of $\tau_{\mathrm{NL}}^{\mathrm{col}}$ around the
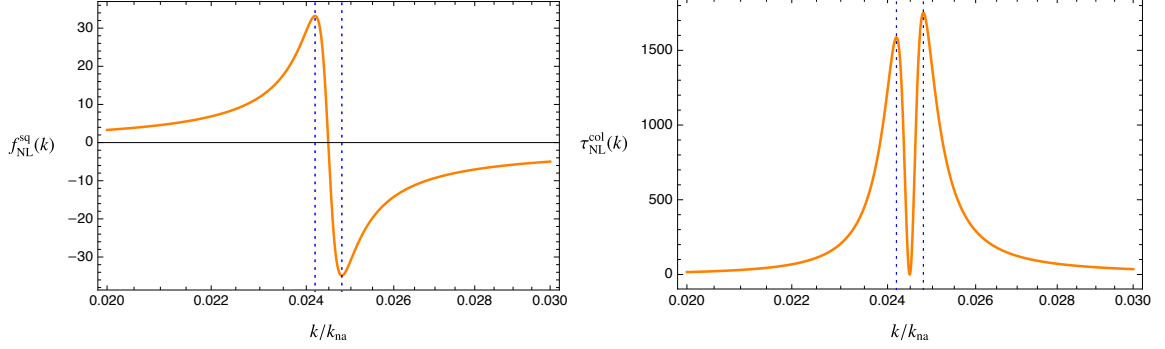
Figure 3: *Scale dependence of $f_{\rm NL}^{\rm sqz}$ (Left) and $\tau_{\rm NL}^{\rm coll}$ (Right) around the dip scale $k_{\rm dip}$ assuming the consistency relations (2.23) in the squeezed and collapsed limits respectively. In both plots we take $\beta_h = 5 \times 10^3$ corresponding to a total enhancement of $\Pi_T \simeq 10^7$ as shown in Figure 2.*

dip are given by

$$(f_{\rm NL}^{\rm sq})_{\rm max/min} = \frac{5}{12}\left(\pm\sqrt{\frac{8}{3}}\,\Pi_T^{1/4} - 1\right), \qquad (\tau_{\rm NL}^{\rm col})_{\rm ext} = \frac{3}{5}\left(\pm\sqrt{\frac{8}{3}}\,\Pi_T^{1/4} - 1\right)^2. \qquad (2.24)$$

For a total enhancement of $\Pi_T \sim 10^7$, these results yield to $|(f_{\rm NL}^{\rm sq})_{\rm max}| \sim 38$ and $(\tau_{\rm NL}^{\rm col})_{\rm ext} \sim 2 \times 10^3$.

We find it interesting that the maximal values of non-Gaussianity parameters are controlled by the total enhancement of the curvature spectrum, Π_T, the only free parameter that enters in our discussion. As we will learn in the next section, more refined proofs of the new consistency relations lead to the same order of magnitude for the amplitudes of $f_{\rm NL}$, $\tau_{\rm NL}$, up to order-one corrections related to the details of the non-attractor phase. Moreover, we will confirm that the quantities $f_{\rm NL}$ and $\tau_{\rm NL}$ are characterized by the specific scale dependence shown in Figure 3, whose features are also controlled by Π_T.

Before discussing a more rigorous derivation of these results, we comment on some of their physical implications, once we assume that $\Pi_T \sim \mathcal{O}(10^7)$ as usually required by PBH formation mechanisms:

- The position of the dip roughly satisfies $k_{\rm dip} \simeq 10^{-2} k_{\rm peak}$, see the left panel of Figure 2. It occurs at relatively large scales where modes still leave the horizon during the initial attractor phase.

- The consistency relations (2.23) and (2.24) dictate that in any single-field model of inflation with a non-attractor phase we should expect amplitudes of order $f_{\rm NL}^{\rm sq} \sim \mathcal{O}(10)$ and $\tau_{\rm NL}^{\rm col} \sim \mathcal{O}(10^3)$ for the non-linearity parameters around the dip.

- Besides their large magnitude, the parameters $f_{\rm NL}^{\rm sq}$ and $\tau_{\rm NL}^{\rm col}$ acquire a pronounced, specific scale dependence around the dip feature – see Figure 3. As we will show in section 4, the amplitude and the characteristic scale dependence of these parameters play an important role for studying observable quantities –such as the statistics of CMB μ-distortions– at scales much larger than the peak scale in the power spectrum. This is a distinctive feature of these scenarios producing PBH, that may allow us to differentiate non-attractor single-field inflation from other models that exhibit large non-Gaussianity at μ-distortion scales.

- Since the scenarios we consider undergo phases of non-attractor evolution that need dedicated matching conditions to connect to distinct slow-roll phases, the squeezed limits of non-Gaussian parameters correspond to *physical* quantities, and can not be removed by coordinate transformations as in [39]. Non-Gaussianity in explicit models undergoing non-attractor evolution are studied for example in [40–45] and the subtlety with matching conditions clearly discussed in the recent work [46].

3 The gradient expansion approach to consistency relations

The aim of this technical section is to use the *gradient expansion formalism*[2] – first introduced in [22] – for reproducing and place in firm footing the heuristic perspective to consistency relations we discussed in the previous section.

In the gradient expansion framework, iterative solutions of the comoving curvature perturbation are generated at any desired order in a small-k expansion, in terms of analytic functions controlled by the background evolution. For example, the amplitude of curvature fluctuations at a fiducial late time $\tau = \tau_*$ (*e.g.* at the reheating surface) is mapped to its value at an initial time around horizon exit $\tau = \tau_k$ in terms of a complex-valued k-dependent coefficient:

$$\mathcal{R}_k(\tau_*) = \alpha_k \mathcal{R}_k(\tau_k) = \left(\alpha_k^R + i\alpha_k^I\right)\mathcal{R}_k(\tau_k), \tag{3.1}$$

where in the last equality we split the enhancement factor into its real and imaginary parts:

$$\alpha_k \equiv \alpha_k^R + i\alpha_k^I. \tag{3.2}$$

Once expanded up to second, k^2-order in the gradient expansion, the real and imaginary parts of this coefficient are given by

$$\alpha_k^R = 1 + D(\tau_k)\, v_{\mathcal{R}}^R - F(\tau_k)\, k^2, \tag{3.3}$$

$$\alpha_k^I = D(\tau_k)\, v_{\mathcal{R}}^I. \tag{3.4}$$

The quantity $v_{\mathcal{R}}^R$ and $v_{\mathcal{I}}^R$ denotes the real and imaginary part of the k-dependent fractional velocity of the curvature perturbation evaluated at $\tau = \tau_k$ defined as

$$v_{\mathcal{R}}(\tau_k) = \left.\frac{\mathcal{R}_k'}{3\mathcal{H}_k \mathcal{R}_k}\right|_{\tau=\tau_k}. \tag{3.5}$$

The full k dependence of the expressions (3.3) and (3.4) on super-horizon scales is encoded in the quantities $v_{\mathcal{R}}^R$, $v_{\mathcal{R}}^I$ (see Appendix A for details) and in the functions $D(\tau_k), F(\tau_k)$, given by the following nested integrals of the pump field appearing in eq (2.1) (see [22, 31] for further details):

$$D(\tau) = 3\mathcal{H}_k \int_\tau^{\tau_*} d\tau'\, \frac{z^2(\tau_k)}{z^2(\tau')}, \tag{3.6}$$

$$F(\tau) = \int_\tau^{\tau_*} \frac{d\tau'}{z^2(\tau')} \int_{\tau_k}^{\tau'} d\tau''\, z^2(\tau''). \tag{3.7}$$

Expressions (3.6) and (3.7) indicate that if the pump field increases with time – as in standard slow-roll inflation, where $z \propto a(\tau)$ – the functions D, F rapidly decrease to zero after horizon crossing (*i.e.* $\alpha_k \to 1$). In this case, the curvature perturbation in (3.1) settles to a constant shortly after horizon exit ($\mathcal{R}_k(\tau_*) \simeq \mathcal{R}_k(\tau_k)$). On the contrary, in inflationary models containing phases of non-attractor evolution, $z(\tau)$ transiently decreases and the functions D, F can grow, amplifying the spectrum of curvature perturbation (*i.e.* $|\alpha_k| \gg 1$ in eq (3.1)) at super-horizon scales: see Appendix B.

The curvature perturbation power spectrum. We define the late-time power spectrum evaluated at $\tau = \tau_*$ as

$$\langle \mathcal{R}_k(\tau_*)\mathcal{R}_{k'}(\tau_*)\rangle = (2\pi)^3\, P_{\mathcal{R}}(\tau_*, k)\, \delta\left(\vec{k} + \vec{k}'\right). \tag{3.8}$$

---

[2]For a detailed account on the spectral profile of the scalar power spectrum that leads to PBH formation and its applications in the context of induced gravitational waves, we refer the reader to [31] where further developments of the gradient expansion formalism is discussed.

Using eq. (3.1), we can then relate the power spectrum at late times to the power spectrum evaluated at horizon crossing via

$$P_{\mathcal{R}}(\tau_*, k) \equiv \frac{2\pi^2}{k^3} \mathcal{P}_{\mathcal{R}}(\tau_*, k) = \frac{2\pi^2}{k^3} \left[|\alpha_k|^2 \mathcal{P}_{\mathcal{R}}(\tau_k) \right] \equiv |\alpha_k|^2 P_{\mathcal{R}}(\tau_k), \tag{3.9}$$

where $\mathcal{P}_{\mathcal{R}}(\tau_k) \equiv k^3 P_{\mathcal{R}}(\tau_k, k)/2\pi^2$, and $|\alpha_k|^2 = (\alpha_k^R)^2 + (\alpha_k^I)^2$.

A non-linear expression for $\mathcal{R}$. We assume that $\mathcal{R}_k(\tau_k)$ is a Gaussian random variable: nevertheless, the superhorizon evolution typically introduces non-linearities. In fact, we can go beyond the linear theory used for eq. (3.1) to compute the bispectrum of the late time curvature perturbation $\mathcal{R}_k(\tau_*)$. For the purpose of deriving an analytic expression for the bispectrum and to study the corresponding consistency relations, we adopt the following non-linear expression for $\mathcal{R}_k(\tau_*)$, first derived in [47, 48]

$$\mathcal{R}_k(\tau_*) = \alpha_k \mathcal{R}_k(\tau_k) + \frac{F(\tau_k)}{2} \left\{ \int \frac{\mathrm{d}^3 k'}{(2\pi)^3} \left[4k'^2 - \vec{k}'.(\vec{k} - \vec{k}') \right] \mathcal{R}_{k'}(\tau_{k'}) \mathcal{R}_{|\vec{k}-\vec{k}'|}(\tau_{|\vec{k}-\vec{k}'|}) \right\}. \tag{3.10}$$

where the last term represents the non-linear contribution parametrized by the convolution of the Gaussian variable $\mathcal{R}_k(\tau_k)$.

Bispectrum. We define the late-time bispectrum of curvature perturbations as

$$\langle \mathcal{R}_{k_1}(\tau_*) \mathcal{R}_{k_2}(\tau_*) \mathcal{R}_{k_3}(\tau_*) \rangle = (2\pi)^3 B_{\mathcal{R}}(k_1, k_2, k_3) \delta\left(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 \right), \tag{3.11}$$

and using (3.10), at leading order in the non-linear term, the bispectrum is given by [47, 48]

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{(2\pi^2)^2}{2(k_1 k_2 k_3)^3} \left[\mathrm{Re}[\alpha_{k_1}^* \alpha_{k_2}] F(\tau_{k_3}) \left\{ 5\left(k_1^2 + k_2^2\right) - k_3^2 \right\} k_3^3 \mathcal{P}_{\mathcal{R}}(\tau_{k_1}) \mathcal{P}_{\mathcal{R}}(\tau_{k_2}) + \text{ perms} \right], \tag{3.12}$$

where the permutations are cyclic among the three external momenta $\{\vec{k}_1, \vec{k}_2, \vec{k}_3\}$. Correspondingly, we define the scale-dependent non-linearity parameter f_{NL} as

$$f_{\mathrm{NL}}(k_1, k_2, k_3) = \frac{5}{6} \frac{B_{\mathcal{R}}(k_1, k_2, k_3)}{[P_{\mathcal{R}}(\tau_*, k_1) P_{\mathcal{R}}(\tau_*, k_2) + \text{ perms}]}, \tag{3.13}$$

Using (3.9) and (3.12) in (3.13), we then obtain the scale-dependent f_{NL} as

$$f_{\mathrm{NL}}(k_1, k_2, k_3) = \frac{5}{12} \frac{\left(\mathrm{Re}[\alpha_{k_1}^* \alpha_{k_2}] F(\tau_{k_3}) \left\{ 5\left(k_1^2 + k_2^2\right) - k_3^2 \right\} k_3^3 \mathcal{P}_{\mathcal{R}}(\tau_{k_1}) \mathcal{P}_{\mathcal{R}}(\tau_{k_2}) + \text{ perms} \right)}{\left[|\alpha_{k_1} \alpha_{k_2}|^2 \mathcal{P}_{\mathcal{R}}(\tau_{k_1}) \mathcal{P}_{\mathcal{R}}(\tau_{k_2}) k_3^3 + \text{ perms} \right]}. \tag{3.14}$$

Trispectrum. We define the connected part of the late time trispectrum of curvature perturbation as

$$\langle \mathcal{R}_{k_1} \mathcal{R}_{k_2} \mathcal{R}_{k_3} \mathcal{R}_{k_4} \rangle = \langle \mathcal{R}_{k_1}^{\mathrm{G}}(\tau_*) \mathcal{R}_{k_2}^{\mathrm{G}}(\tau_*) \mathcal{R}_{k_3}^{\mathrm{G}}(\tau_*) \mathcal{R}_{k_4}^{\mathrm{G}}(\tau_*) \rangle + (2\pi)^3 \delta\left(\vec{K}_{\mathrm{tot}} \right) T_{\mathcal{R}}(k_1, k_2, k_3, k_4), \tag{3.15}$$

where the first term is the Gaussian part of the trispectrum (see appendix D) and $\vec{K}_{\mathrm{tot}} \equiv \sum_{i=1}^4 \vec{k}_i$.

Utilizing again the non-linear expression (3.10), the leading order scale-dependent trispectrum is found to be

$$T_{\mathcal{R}}(k_1, k_2, k_3, k_4) = \frac{\mathrm{Re}[\alpha_{k_1}^* \alpha_{k_2}] F(\tau_{k_3}) F(\tau_{k_4})}{4} \left\{ 4\left(k_1^2 + k_{13}^2\right) + 2\vec{k}_1.\vec{k}_{13} \right\} \left\{ 4\left(k_2^2 + k_{13}^2\right) - 2\vec{k}_2.\vec{k}_{13} \right\}$$
$$\times P_{\mathcal{R}}(\tau_{k_1}) P_{\mathcal{R}}(\tau_{k_2}) P_{\mathcal{R}}(\tau_{k_{13}}) + 11 \text{ perms}, \tag{3.16}$$

where we introduce $\vec{k}_{ij} = \vec{k}_i + \vec{k}_j$, $k_{ij} = |\vec{k}_i + \vec{k}_j|$ for $i \neq j$ and perms denote distinct permutations among the external momenta $\{\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4\}$. We define the corresponding scale dependent non-linearity parameter τ_{NL} as

$$\tau_{\mathrm{NL}}(k_1, k_2, k_3, k_4) = \frac{T_{\mathcal{R}}(k_1, k_2, k_3, k_4)}{P_{\mathcal{R}}(\tau_*, k_1) P_{\mathcal{R}}(\tau_*, k_2) P_{\mathcal{R}}(\tau_*, k_{13}) + 11 \text{ perms}}. \tag{3.17}$$

Eqs (3.14), (3.16) and (3.17) suggest that the size and the scale-dependence of the f_{NL} and τ_{NL} parameters depend on the function $F(\tau_k)$ and the quantity α_k, whose behavior depends on the pump field profile. Given these preliminary results, we now study the scale dependence of the power spectrum and bispectrum on a representative set-up capable of producing a large PBH population during inflation, to further support the findings of section 2.

3.1 Spectral profile of the scalar power spectrum and its properties

To study the spectral shape and enhancement in the power spectrum, we consider a typical two phase inflationary scenario that instantly connects an initial slow-roll era with $\eta_{\mathrm{sr}} = 0$, to a slow-roll violating, non-attractor phase with $\eta_{\mathrm{c}} \leq -6$, within the time range $\tau_0 \leq \tau \leq \tau_f$. Here η denotes the second slow-roll parameter, $H \eta \equiv d \ln \epsilon / dt$. The first slow-roll parameter is given by $H \epsilon = -d \ln H / dt$. The pump field $z(\tau)$ appearing in the quadratic action (2.1) is assumed to have a profile (we take $\tau < 0$):

$$z(\tau) = \begin{cases} z_0 (\tau/\tau_0)^{-1} & \tau/\tau_0 \geq 1, \\ z_0 (\tau/\tau_0)^{-(\eta_{\mathrm{c}}+2)/2} & \tau_f/\tau_0 \leq \tau/\tau_0 \leq 1, \end{cases} \tag{3.18}$$

describing collectively the initial slow-roll and the slow-roll violating phases. Although the set-up is qualitatively similar to the one discussed in the previous section, in this case we consider a *finite duration* for the non-attractor era denoted by $\Delta N = \ln(\tau_0/\tau_f)$ in e-fold numbers. We relate the quantity z_0 with a constant slow-roll parameter ϵ_{sr} via $z_0 = -a(\tau_0)\sqrt{2\epsilon_{\mathrm{sr}}} M_{\mathrm{pl}}$. For simplicity we parametrize the scale factor as in de Sitter space: $a = -1/(H\tau)$ with a constant Hubble rate H during inflation, which defines comoving horizon $\tau_0 = (a_0 H)^{-1} = \mathcal{H}_0^{-1}$ at the time of the transition to the non-attractor era.

We proceed with determining spectral profile of the power spectrum by evaluating the dimensionless power spectrum (3.9) at the end of the non-attractor era:

$$\mathcal{P}_{\mathcal{R}}(\tau_f, k) \equiv |\alpha_k|^2 \mathcal{P}_{\mathcal{R}}(\tau_k) = \left[(\alpha_k^R)^2 + (\alpha_k^I)^2 \right] \mathcal{P}_{\mathcal{R}}(\tau_k). \tag{3.19}$$

Using the pump field profile (3.18), we can derive analytic formulas for α_k^R (3.3) and α_k^I (3.4) (see Appendix A and B) to characterize the shape of the power spectrum. The resulting profile of the late time power spectrum is shown in Figure 4 for representative scenarios leading to PBH formation given the pump field profile of (3.18). For understanding the physical implications of our findings within the gradient expansion formalism, it is convenient to introduce a fixed quantity

$$c_k \equiv -k\tau_k \leq 1, \tag{3.20}$$

which determines the size of a mode k with respect to the horizon $(aH)^{-1}$ at time $\tau = \tau_k$, corresponding to the horizon crossing epoch. We stress that by virtue of the relation (3.20) and the super-horizon gradient formalism, all modes we consider in Figure 4 (and in general in this work) are outside the horizon at the initial time τ_k. We then distinguish modes whose momenta lie in the following ranges:

i) modes that become super-horizon during the initial slow-roll era, *i.e.* modes satisfying $\tau_k/\tau_0 > 1$ or equivalently $k/\mathcal{H}_0 < c_k \leq 1$, and

ii) modes that leave the horizon during the non-attractor $\eta_c \leq -6$ phase, $c_k < k/\mathcal{H}_0$.
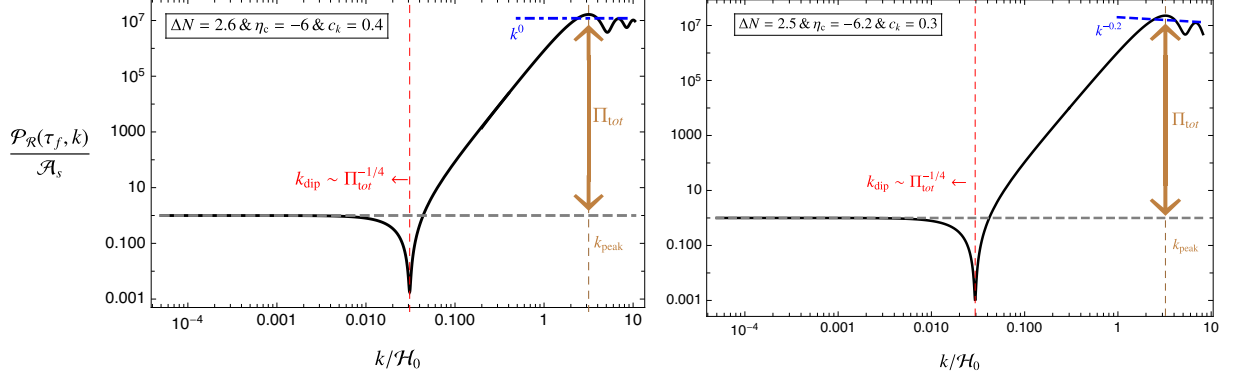
Figure 4: *Scale dependence of the power spectrum for an inflationary scenario that includes a transient ultra slow-roll or constant roll phase. The location of the dip feature (red dashed line) and its relation to the global enhancement $\Pi_{\rm tot}$ in the power spectrum is shown. The plots are obtained using the gradient expansion formalism, and normalized with respect to power spectrum at the largest scales, i.e. $\mathcal{A}_s = H^2/(8\pi^2 \epsilon_{\rm sr} M_{\rm pl}^2) \simeq 2.1 \times 10^{-9}$.*

Focusing on these regimes separately, we discuss below the spectral behavior of $\mathcal{P}_{\mathcal{R}}$ and its global features.

Total enhancement in the power spectrum. Figure 4 indicates that independently from the choice of model parameters, the power spectrum reaches its peak at $k_{\rm peak} \simeq 3\mathcal{H}_0$ (See also [49] for an analysis of the ultra–slow-roll case). To obtain the parametric dependence of the total enhancement on the model parameters, we evaluate the power spectrum at this scale to determine its height with respect to the largest scales. For this purpose, we first focus on $\mathcal{P}_{\mathcal{R}}(\tau_k)$ as in (A.7) in the non-attractor phase ($c_k < k/\mathcal{H}_0$), which reads

$$\frac{\mathcal{P}_{\mathcal{R}}(\tau_k)}{\mathcal{A}_s} = \frac{c_k^{2\nu}\pi^2}{4}\left(\frac{k}{\mathcal{H}_0}\right)^{-2\nu+2}\left[f_3^2 + 2\left(\frac{k}{\mathcal{H}_0}\right)f_3 f_4 + \left(\frac{k}{\mathcal{H}_0}\right)^2(f_3^2 + f_4^2)\right]_{\tau=\tau_k}, \qquad (3.21)$$

where $\nu = (3 + \eta_{\rm c})/2$ and $\mathcal{A}_s = H^2/8\pi^2 \epsilon_{\rm sr} M_{\rm pl}^2$ is the normalization of the power spectrum at very large scales, $k \to 0$.

In appendix A we show that the functions f_3 and f_4 that appear in (3.21) are given by (J, Y being Bessel functions of first and second kind)

$$f_3\left(-k\tau, \frac{k}{\mathcal{H}_0}, \nu\right) = J_{\nu-1}\left(\frac{k}{\mathcal{H}_0}\right)Y_\nu(-k\tau) - Y_{\nu-1}\left(\frac{k}{\mathcal{H}_0}\right)J_\nu(-k\tau), \qquad (3.22)$$

$$f_4\left(-k\tau, \frac{k}{\mathcal{H}_0}, \nu\right) = J_\nu\left(\frac{k}{\mathcal{H}_0}\right)Y_\nu(-k\tau) - Y_\nu\left(\frac{k}{\mathcal{H}_0}\right)J_\nu(-k\tau). \qquad (3.23)$$

Using (3.22) and (3.23), it is straightforward to realise that the expression in the square brackets of (3.21) is dominated by the last term around the peak scale. Also, around the scale of the peak, the modulus square of the enhancement factor are well approximated by the following expression

$$|\alpha_k|^2 \simeq (\alpha_k^R)^2 \simeq \left(1 + D(\tau_k)v_{\mathcal{R}}^R\right)^2 \simeq D(\tau_k)^2, \qquad (3.24)$$

where in the last step we take $v_{\mathcal{R}}^R \to 1$ which can be verified explicitly applying eq (A.4) at scales around $k_{\rm peak}$. Combining (3.24) and (3.21) evaluated at the peak scale, the total enhancement of the power spectrum from large scales to small scales can be computed by making use of the expression in (3.19). This

gives,

$$\Pi_{\text{tot}} \equiv \frac{\mathcal{P}_{\mathcal{R}}(\tau_f, k_{\text{peak}})}{\mathcal{A}_s} \simeq C(c_k, \nu) \, e^{-2(3+\eta_c)\Delta N}, \tag{3.25}$$

where $C(c_k, \nu)$ is

$$C(c_k, \nu) = \frac{\pi^2 \, 3^{2\nu+6}}{16 \, \nu^2 \, c_k^{2\nu}} \left[f_3^2(c_k, 3, \nu) + f_4^2(c_k, 3, \nu) \right]. \tag{3.26}$$

Eq (3.25) indicates that the total enhancement in the scalar power spectrum is exponentially sensitive to slow-roll parameter $\eta_c \leq -6$ and in particular to the duration ΔN of the non-attractor phase. Using this expression, one can confirm that for typical parameter choices that leads to a $\Pi_{\text{tot}} \simeq 10^7$ enhancement required for PBH formation, one obtains $C \simeq \mathcal{O}(1)$.

The dip and its properties. For modes that exit the horizon during the initial slow-roll era, there is a pronounced dip in the power spectrum occurring far away from the peak scale, $k_{\text{dip}} \ll \mathcal{H}_0 < k_{\text{peak}}$. It is worth mentioning that the dip appears due to competing contributions in the power spectrum that are weighted by opposite signs. To determine its location, we focus our attention on the scale dependence of the power spectrum (3.19) for scales satisfying $k/\mathcal{H}_0 < c_k \leq 1$. As we show in Appendix B.1, in this regime the enhancement factor α_k of eq (3.2) is dominated by its real part whose scale dependence can be accurately captured by eq. (B.3):

$$\alpha_k^R \simeq \alpha_{(0)}^R \left[1 - \beta \left(\frac{k}{\mathcal{H}_0} \right)^2 \right] + \mathcal{O}\left(\frac{k^3}{\mathcal{H}_0^3} \right) \tag{3.27}$$

where $\alpha_{(0)}^R$ is an order quantity and $\beta \gg 1$ is an exponentially large number parametrized in terms of the duration ΔN of the slow-roll violating phase and the value of η_c in this era. (See eq. (B.4).) In Appendix A – see in particular eq (A.3) – we show that the quantity $\mathcal{P}_{\mathcal{R}}(\tau_k)$ is nearly scale-independent for modes that leave the horizon in the initial slow-roll era. This implies that the scale dependence of the late time power spectrum (3.19) around the dip is completely dictated by α_k^R. Therefore the zero of (3.27) provides an accurate description for the location of the dip in the power spectrum, which reads as

$$\frac{k_{\text{dip}}}{\mathcal{H}_0} = \frac{1}{\sqrt{\beta}} \equiv \left[-\frac{\eta_c \, e^{-(3+\eta_c)\Delta N}}{\alpha_{(0)}^R(c_k) \, (\eta_c + 1)(\eta_c + 3)} \right]^{-1/2}. \tag{3.28}$$

Notice that by virtue of the above definition (3.28), the exponentially large number β plays the same role of the parameter β_h introduced in the heuristic approach we discussed in the previous section. A clear advantage of the gradient approach is the fact that it makes apparent why such a large number appears in PBH forming inflationary scenarios by relating β to the duration ΔN and η_c of the slow-roll violating phase as $\beta \propto e^{-(3+\eta_c)\Delta N} \gg 1$.

The presence of such a pronounced dip in the spectrum is a universal feature, being virtually present in all single field models based on non-attractor evolution that are aiming to generate a sizeable peak in the power spectrum for producing PBH, say of order $\Pi_{\text{tot}} \simeq 10^7$ (see e.g. [10, 12, 14–17, 23–26]). Using (3.28) we can universally relate the location of the dip feature to the total enhancement in the power spectrum in (3.25), as first found in [19], which reads as

$$\frac{k_{\text{dip}}}{\mathcal{H}_0} \simeq \left[-\frac{\sqrt{C(c_k, \nu)} \, \alpha_{(0)}^R(c_k) \, (\eta_c + 1)(\eta_c + 3)}{\eta_c} \right]^{1/2} \Pi_{\text{tot}}^{-1/4} \approx 10^{-2}. \tag{3.29}$$

A close examination of Figure 4 confirms these arguments and shows that (3.29) is a robust relation, valid in all single field inflationary scenarios that can produce a pronounced peak in the power spectrum. Note also that – considering the relation between the peak scale and $\mathcal{H}_0$ we mentioned before – we can connect

the peak scale to the dip scale as $k_{\rm dip} \simeq 10^{-2}\, k_{\rm peak}$. Relation (3.29) is in agreement with the results of section 2, and includes an overall, order-one factor depending on parameters controlling the duration of the dip and the properties of the system.

3.2 Bispectrum in the squeezed limit: a consistency relation around the dip

Let us concentrate on modes that exit during the initial slow-roll era, $k/\mathcal{H}_0 < c_k$, to investigate the scale-dependence of the bispectrum. We anticipate that the bispectrum exhibits features and be amplified around the dip scale $k_{\rm dip}$ in the power spectrum since non-linearities are usually enhanced at the location of rapid changes in the power spectrum [50]. More importantly, we show that scale dependence of the squeezed bispectrum closely follows the prediction of Maldacena consistency condition: *i.e.* $f_{\rm NL}(k) = 5(1-n_s(k))/12$ [33], proving the heuristic results of section 2.

Consistency condition. We begin by the squeezed limit of the scale dependent non-linearity parameter[3]. For scales that exit the horizon during the initial slow-roll era, noting the scale invariance of $\mathcal{P}_\mathcal{R}(\tau_k)$ factors, we take the squeezed limit $\vec{k}_3 \to 0$ ($k_1 \simeq k_2 \equiv q$) of (3.14) which yields [20,47]

$$f_{\rm NL}(q,q,k_3 \to 0) \equiv f_{\rm NL}^{\rm sq} = \frac{5}{12}\left(\frac{4\,{\rm Re}[\alpha_q^* \alpha_{k_3}]\,F(\tau_{k_q})q^2}{|\alpha_q \alpha_{k_3}|^2}\right) \simeq \frac{5}{12}\left(\frac{4\,\alpha_q^R\,F(\tau_{k_q})q^2}{|\alpha_q|^2}\right), \qquad (3.30)$$

where in the last equality we make the approximation $\alpha_{k_3} \to 1$ in the $k_3 \to 0$ limit, and we use the fact that the enhancement factor is dominated by its real part α_k^R in the initial slow-roll phase. To prove that the consistency condition holds, we now show that the term inside the brackets in (3.30) is equivalent to $(1-n_s)$. For this purpose, approximating $\alpha_k \simeq \alpha_k^R$ as before, we utilize (3.19) and the definition $n_s - 1 \equiv {\rm d}\ln\mathcal{P}_\mathcal{R}(\tau_f, k)/{\rm d}\ln k$ to write

$$(1-n_s) \equiv -\frac{{\rm d}\ln\mathcal{P}_\mathcal{R}(\tau_f, q)}{{\rm d}\ln q} = -\frac{2\alpha_q^R\left(\frac{q}{\mathcal{H}_0}\times\alpha_q^{R'}\right)}{|\alpha_q|^2} \approx \frac{4\alpha_q^R F(\tau_{k_q})q^2}{|\alpha_q|^2}, \qquad (3.31)$$

where the prime denotes derivative with respect to the normalized wave-number $k/\mathcal{H}_0$. Notice that in the last equality of (3.31) we take $-(q/\mathcal{H}_0)\alpha_q^{R'} \simeq 2F(\tau_q)q^2$. Employing further simplifications for α_k^R and its derivative (see eqs. (B.8) and (B.9)), we prove in appendix B.1 that this relation holds to a very good approximation. This fact confirms that the consistency condition always holds in the initial slow-roll stage of inflationary scenarios that can generate PBH populations. To be more explicit, in Figure 5 we show the behavior of squeezed limit $f_{\rm NL}$ superimposed with $5(1-n_s)/12$ around the dip scale for a representative choice of parameters that can generate a $\Pi_{\rm tot} \simeq 10^7$ enhancement in the power spectrum as required for PBH production. The figure shows clearly that the scale dependence of the non-linearity parameter $f_{\rm NL}$ follows Maldacena consistency condition.

Maximal value of $f_{\rm NL}^{\rm sq}$ around the dip. Another conclusion we can guess from Figure 5 is that the maximal amplitude of the squeezed limit non-linearity parameter becomes $|f_{\rm NL}| \simeq \mathcal{O}(10)$ around $k_{\rm dip}$ which is consistent with the large values obtained by the spectral tilt $|n_s - 1|$ around the dip region: see Figure 4, and recall the results of section 2. In what follows, we will show that $|f_{\rm NL}| \simeq \mathcal{O}(10)$ universally holds around the dip scale $k_{\rm dip}$ for any inflationary scenario that support a $\Pi_{\rm tot} \simeq 10^7$ enhancement in the power spectrum.

We proceed as in section 2: we focus on the maximum and the minimum values obtained by the $n_s - 1$ and specifically the zeros of its running defined in (2.17). On the other hand, as Figure 5 suggests, the spectral index reaches its maximal values very close to the dip scale $k_{\rm dip}$. Therefore we find it convenient

---

[3]A scale-dependent $f_{\rm NL}$ may arise in a various other inflationary contexts, see e.g. the early works [51–53].
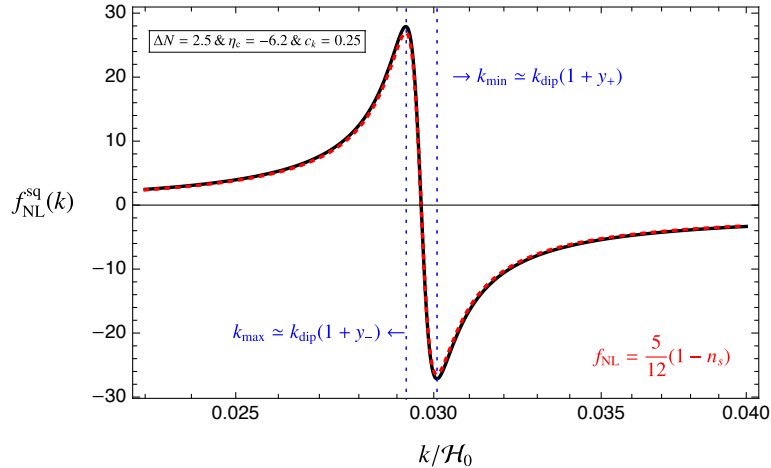
Figure 5: *Scale dependence of the squeezed configuration non-linearity parameter $f_{\rm NL}$ around $k_{\rm dip}$ for a transient constant-roll model that generates a $\Pi_{\rm tot} \simeq 10^7$ enhancement in the power spectrum. The accuracy of the consistency relation $f_{\rm NL} = 5(1 - n_s)/12$ in capturing the behavior of the $f_{\rm NL}$ in the squeezed limit is shown by red dotted curve.*

to define a new variable,

$$\frac{k}{\mathcal{H}_0} \equiv \frac{x}{\sqrt{\beta}}, \tag{3.32}$$

where $\beta = \beta(\eta_c, \Delta N, c_k) \gg 1$ is defined as in (3.28) (See also eq. (B.4)). In the parametrization of eq. (3.32), the dip feature in the power spectrum corresponds to $x = 1$ which in turn overlaps with the zero of the spectral index, since the latter is proportional to the real part of the enhancement factor as can be verified from eq. (3.31). Furthermore, we expect the maximal values of $n_s - 1$ to be very close to the dip feature. We then define $x = 1 + y$ where $y \ll 1$ and expand the resulting expression using α_s (2.17) up to quadratic order in y. In this way we obtain a simple quadratic equation for y, and its roots provide us the zeros of the running α_s and hence the location of the maximum and minimum of the spectral index. In particular, in terms of the model parameters, we find the following solutions,

$$y_{\pm} \simeq \pm \frac{\bar{D}}{2\,\alpha_{(0)}^R(c_k)\sqrt{\beta}} + \frac{3\bar{D}^2}{4\,[\alpha_{(0)}^R(c_k)]^2\beta}, \tag{3.33}$$

where we define

$$\bar{D} = \frac{\alpha_{(0)}^R(c_k)(\eta_c + 1)}{(1 + c_k^2)\eta_c} - \alpha_{(0)}^I(c_k)\,\sqrt{\beta}. \tag{3.34}$$

We note that although the latter expression appear to be of the order of $\sqrt{\beta} \gg 1$, it is genuinely an order one number due to the small factor $\alpha_{(0)}^I \propto c_k^3 \ll 1$ (See *e.g.* (B.4)). In fact, for any scenario that leads to a $\Pi_{\rm tot} \simeq 10^7$ enhancement in the power spectrum, the second term in (3.34) is comparable to the first term in absolute magnitude and therefore it is a order-one number as one can verify explicitly from (B.4). Using (3.33), the location of the maximum and the minimum of the spectral index is therefore given by

$$\frac{k_{\pm}}{\mathcal{H}_0} \simeq \frac{k_{\rm dip}}{\mathcal{H}_0}\,(1 \pm y_{\pm}) = \frac{1}{\sqrt{\beta}}\,(1 \pm y_{\pm}). \tag{3.35}$$

The accuracy of these formulas in locating the max/min values of $n_s - 1$ is shown in Figure 5. Using these results, we can then determine the maximal values obtained by the non-linearity parameter in the squeezed limit. Utilizing the consistency condition, we plug (3.35) in (3.31) and at leading order in the

15

large parameter β, we found

$$(f_{\text{NL}}^{\text{sq}})_{\text{max/min}} = \frac{5}{12}(1 - n_s) \simeq -\frac{10}{3}\frac{\alpha_{(0)}^R(c_k)^2}{\bar{D}^2}\beta\, y_{\mp} \simeq \pm\frac{5}{3}\frac{\alpha_{(0)}^R(c_k)}{\bar{D}}\sqrt{\beta}. \tag{3.36}$$

Using (3.25), we can then re-write the max/min value of the non-linearity parameter in terms of the total enhancement in the power spectrum as

$$(f_{\text{NL}}^{\text{sq}})_{\text{max/min}} \simeq \pm\frac{5}{3}\left[-\frac{\alpha_{(0)}^R(c_k)\,\eta_{\text{c}}}{\bar{D}^2\,\sqrt{C(c_k,\nu)}\,(\eta_{\text{c}} + 1)(\eta_{\text{c}} + 3)}\right]^{1/2}\Pi_{\text{tot}}^{1/4}. \tag{3.37}$$

For typical scenarios with $\Pi_{\text{tot}} \simeq 10^7$, the expression above evaluates to $(f_{\text{NL}}^{\text{sq}})_{max/min} \simeq \mathcal{O}(10)$ as can be checked explicitly. This results confirms and supports the findings of section 2.

3.3   Trispectrum in the collapsed limit and its consistency relation

We now focus on the scale dependence of the trispectrum around the dip feature, *i.e.* for $k/\mathcal{H}_0 < c_k$. As we show below, the resulting scale dependent trispectrum closely tracks a specific single-field relation [35–38]

$$\tau_{\text{NL}}^{\text{col}} = \frac{36}{25}\left(f_{\text{NL}}^{\text{sq}}\right)^2. \tag{3.38}$$

in the collapsed limit $|\vec{k}_{12}| \to 0$.

Consistency relation. To prove tha relation (3.38) holds for scales that exit the horizon in the initial slow-roll era, we take the collapsed limit $\vec{k}_{12} \to 0$ ($\vec{k}_{34} \to 0$) of eq. (3.17) assuming a symmetric folded kite configuration for the external momenta, *i.e.* $k_1 \simeq k_2 \simeq k_3 \simeq k_4 = q$ with $k_{12}, k_{34} \to 0$. In this limit, noting the expressions (3.16) and (3.9), we find that τ_{NL} (3.17) reduces to

$$\lim_{k_{12}\to 0}\tau_{\text{NL}} \equiv \tau_{\text{NL}}^{\text{col}}(q, k_{12}) = 4\frac{|\alpha_q|^2\left(F(\tau_q)\,q^2\right)^2}{|\alpha_q|^4\,|\alpha_{k_{12}}|^2} \simeq 4\frac{(\alpha_q^R)^2\left(F(\tau_q)\,q^2\right)^2}{|\alpha_q|^4}, \tag{3.39}$$

where in the last step we approximately take $\alpha_{k_{12}} \simeq 1$ in the $k_{12} \to 0$ limit and assume $\alpha_q \simeq \alpha_q^R$ during the initial slow-roll stage. Using the squeezed limit expression for $f_{\text{NL}}^{\text{sq}}$ (3.30), it simply follows from (3.39) that relation (3.38) holds in a non-trivial scale dependent manner during the initial slow-roll stage, *i.e.* around the dip feature in the power spectrum. To illustrate these points concretely, we show in Figure 6 the scale dependence of the $\tau_{\text{NL}}^{\text{col}}$ around the dip scale k_{dip}. We clearly see that $\tau_{\text{NL}}^{\text{col}}$ satisfies the relation (3.38) and therefore it becomes maximal at wave numbers where $|f_{\text{NL}}^{\text{sq}}|$ does: *i.e.* at $k_{\pm}$ as given by eq (3.35). Using (3.37) and (3.38), the maximal value that $\tau_{\text{NL}}^{\text{col}}$ can acquire is found to be

$$\left(\tau_{\text{NL}}^{\text{col}}\right)_{\text{max}} = -\frac{4\,\alpha_{(0)}^R(c_k)\,\eta_{\text{c}}}{\bar{D}^2\,\sqrt{C(c_k,\nu)}\,(\eta_{\text{c}} + 1)(\eta_{\text{c}} + 3)}\Pi_{\text{tot}}^{1/2}. \tag{3.40}$$

Therefore for any slow-roll violating transient phase that can generate a $\Pi_{\text{tot}} \simeq 10^7$ enhancement in the power spectrum, the non-linearity parameter becomes $\tau_{\text{NL}}^{\text{col}} \simeq \mathcal{O}(10^3)$ around the dip feature, in agreement with the heuristic results of section 2.

Before we conclude this section, we stress that our findings on the scale dependence of the bispectrum in Section 3.2 agree well both qualitatively and quantitatively with the previous studies focusing on the same issue using numerical techniques (See *e.g.* [54–56]). As we mentioned earlier in [20], the gradient expansion formalism has the advantage of analytic control that allows us to transparently capture the features of n-point scalar correlation functions using a few parameters such as the duration of the transient
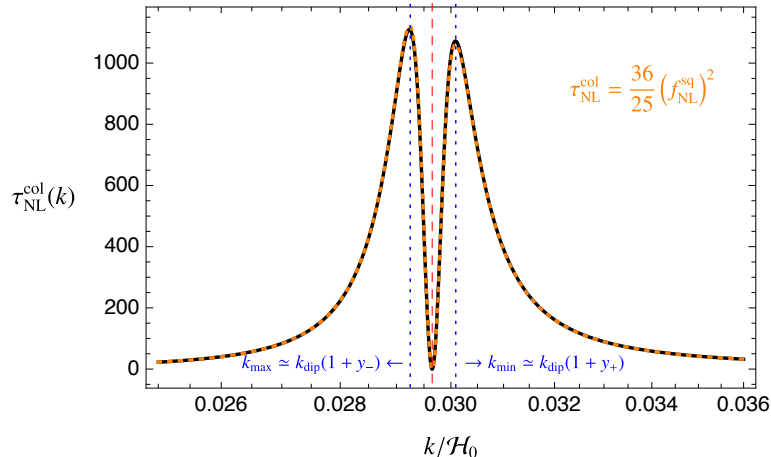
Figure 6: *Scale dependence of τ_{NL} around k_{dip} in the collapsed limit for the transient constant-roll scenario with the same model parameters provided in Figure 5. $\tau_{\mathrm{NL}}^{\mathrm{col}}$ as predicted by the consistency relation (3.38) is shown by orange dotted curve. $\tau_{\mathrm{NL}}^{\mathrm{col}}$ becomes maximal at the same wave numbers (blue dotted vertical lines) where f_{NL} in the squeezed configuration does.*

non-attractor era ΔN and the slow-roll parameter η_{c} in this phase. More importantly, using the gradient formalism we proved analytically in Section 3.2 and 3.3 that consistency conditions[4] $f_{\mathrm{NL}}^{\mathrm{sq}} = 5(1 - n_s)/12$ and $\tau_{\mathrm{NL}}^{\mathrm{col}} = 36 \left(f_{\mathrm{NL}}^{\mathrm{sq}}\right)^2 /25$ (3.38) hold for modes that leave the horizon during the initial slow-roll era.

4   CMB μ-distortions and non-Gaussianity around the dip

In the previous section we studied consistency conditions for non-Gaussian parameters around the dip of the spectrum. If the dip feature occurs at relatively large scales, say $10\,\mathrm{Mpc}^{-1} \leq k_{\mathrm{dip}} \leq 10^4\,\mathrm{Mpc}^{-1}$, the properties of the resulting curvature spectrum can be probed through CMB $\mu-$distortions, using well controlled CMB physics in the linear regime [58, 59] (see [60, 61] for a review).

This possibility was suggested in [20], building on the ideas first developed in [21]. In this section we further develop and extend these arguments. We stress that the range of scales we identified for k_{dip} above is well motivated for PBH populations with astrophysical masses. For example, given that $k_{\mathrm{dip}}/k_{\mathrm{peak}} \sim 10^{-2}$, with a dip scale located at $k_{\mathrm{dip}} \simeq 10^3 - 10^4\,\mathrm{Mpc}^{-1}$ relates to a peak scale $k_{\mathrm{peak}} \simeq 10^5 - 10^6\,\mathrm{Mpc}^{-1}$, that correspond to the formation of PBHs in the mass range $M_{\mathrm{PBH}} \simeq 1 - 100\,M_{\odot}$ (see e.g. [8, 62]).

CMB μ-distortions can impose constraints on PBH formation mechanisms [18, 63]; we wish to emphasize that, in case of detection, the statistics of μ-distortions at large scales provide information on the physics sourcing PBHs at much smaller, non-linear scales. This is possible thanks to the coupling between large and small scales through the squeezed and collapsed limits of non-Gaussianity.

Our starting point are the consistency relations of sections 2, 3 which imply that for an $\mathcal{O}(10^7)$ enhancement in the spectrum, we obtain $f_{\mathrm{NL}}^{\mathrm{sqz}} \sim \mathcal{O}(10)$, $\tau_{\mathrm{NL}}^{\mathrm{coll}} \sim \mathcal{O}(10^3)$ around the dip feature for any model of single-field inflation including a short period of non-attractor evolution. Additionally, we find that the squeezed non-Gaussian parameters has features specific of these scenarios – see Fig 3.

As first proposed in [21] (and further explored in [64–70]) non-Gaussianity at μ-distortion scales have distinctive consequences for $\langle \mu T \rangle$ correlators [5] among CMB μ-distortions and temperature fluctuations –

---

[4]See also [57] and [56] for an investigation on the Maldacena's consistency condition in slow-roll violating inflationary scenarios.

[5]See also [70, 71] for the influence of primordial non-Gaussianities on μE and μB [69] cross correlations.

sensitive to the squeezed limit of the bispectrum – and for $\langle \mu\mu \rangle$ self correlations – sensitive to the collapsed limit of the trispectrum. In this section we go beyond work in [20] by developing the following points

- First, we show how the information provided by the consistency relations allows us to carry on a more detailed analysis of $\langle \mu T \rangle$ correlators, whose quantitative and qualitative features depend on the properties of the squeezed bispectrum. See Section 4.1.

- Then, we study for the first time the implications for the $\langle \mu\mu \rangle$ selfcorrelator of a scale-dependent collapsed trispectrum around the dip. See Section 4.2.

- Finally, at the light of the results above, we discuss improved estimates for the detectability of non-Gaussian consistency relations with PIXIE or PRISM-like experiments, and physical implications for PBH populations. See Section 4.3.

Before covering the points above, we present some preliminary formulas [6] that we are going to use extensively. We begin by relating the initial curvature perturbation $\mathcal{R}_k$ to the harmonic coefficients of the CMB temperature anisotropies $\Theta(\hat{n}) = \sum_{lm} a_{lm}^T Y_{lm}(\hat{n})$ and CMB distortion anisotropies $\mu(\hat{n}) = \sum_{lm} a_{lm}^\mu Y_{lm}(\hat{n})$. The coefficients $a_{lm}^{T,\mu}$ associated with $\mathcal{R}_k$ are given by [21,64],

$$a_{lm}^T = \frac{12\pi}{5}(-i)^l \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \, \mathcal{R}_k \, \Delta_l(k) \, Y_{lm}^*(\hat{k}), \tag{4.1}$$

$$a_{lm}^\mu \simeq 18.4\pi(-i)^l \int \frac{\mathrm{d}^3 k_1 \, \mathrm{d}^3 k_2}{(2\pi)^6} Y_{lm}^* \left(\hat{k}_+ \right) \mathcal{R}_{k_1} \mathcal{R}_{k_2} \, W\left(\frac{k_+}{k_s} \right) j_l\left(k_+ \chi_* \right)$$
$$\times \langle \cos\left(c_s k_1 \tau \right) \cos\left(c_s k_2 \tau \right) \rangle_p \left[e^{-\left(k_1^2 + k_2^2 \right)/k_D^2} \right]_{z_f}^{z_i}, \tag{4.2}$$

where p denotes time averaging over the period acoustic oscillations; $\vec{k}_+ \equiv \vec{k}_1 + \vec{k}_2$; c_s is the sound speed of the radiation perturbations; $\chi_* = \tau_0 - \tau_* \simeq 14\,\mathrm{Gpc}$ is the comoving distance between the last scattering surface and today; and $\Delta_l(k)$ is the transfer function during radiation dominated universe. Moreover, in (4.2), $W(k) = 3k^{-3}[\sin(k) - k\cos(k)]$ is a top-hat filter function in Fourier space that smears the dissipated energy over a volume of radius $k_s^{-1} \gtrsim k_D(z_f)^{-1}$ where k_D is the diffusion damping scale during radiation domination. It depends on redshift as $k_D(z) \simeq [(1+z)/10^5]^{3/2} \, 130\,\mathrm{Mpc}^{-1}$ and the range of z associated with the μ distortions is given by

$$z_f \equiv 5 \times 10^4 < z < 2 \times 10^6 \equiv z_i. \tag{4.3}$$

We then define angular correlators involving anisotropies labeled by $\{i, j\}$ as

$$\left\langle \left(a_{lm}^i \right)^* a_{l'm'}^j \right\rangle = \delta_{ll'} \delta_{mm'} C_l^{ij}, \qquad i = \mu, T. \tag{4.4}$$

In what comes next, we make use of the definition (4.4) to study angular correlations $\langle \mu T \rangle$ and $\langle \mu\mu \rangle$, and to relate them to scale dependent bispectrum and trispectrum present around the dip scale of the PBH forming inflationary scenarios.

4.1   Phenomenology of the scale dependent squeezed bispectrum: $C_l^{\mu T}$

We start discussing cross correlations between μ distortions and temperature anisotropies Θ. As we will see, the specific scale-dependent profile of the squeezed bispectrum can considerably enhance such cross correlations in comparison with more standard non-Gaussian models.

---

[6]We refer the reader to [60], or section 3.1 of [20] for a mini-introduction of CMB μ-distortions and additional motivations for the formulas that follow.

Using the definition (3.11) of the bispectrum together with (4.1) and (4.2), the cross correlations $C_l^{\mu T}$ for a squeezed bispectrum ($\vec{k}_+ = \vec{k}_1 + \vec{k}_2 \to 0$) can be expressed as [21, 64],

$$C_l^{\mu T} \simeq \frac{27.6}{20\pi^3} \int \mathrm{d}\ln k_+ \, \Delta_l(k_+) \, j_l\left(k_+\chi_*\right) \int \mathrm{d}\ln q \left[k_+^3 q^3 B_\mathcal{R}\left(q, q, k_+ \to 0\right) \right] \left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}, \tag{4.5}$$

where we take the filter function $W \to 1$ in the squeezed limit $k_+ \ll k_s \simeq k_d(z_f)$, and we relabel $|\vec{k}_1| \simeq |-\vec{k}_2| \equiv q$. Using the general definition (3.13) of the non-linearity parameter, the squeezed limit bispectrum inside the $\mathrm{d}\ln q$ integral in (4.5) can be expressed as

$$B_\mathcal{R}\left(q, q, k_+ \to 0\right) \simeq \frac{12}{5} \underbrace{f_{\mathrm{NL}}^{\mathrm{sq}}(q) E(q)}_{\equiv f_{\mathrm{NL}}^{\mathrm{eff}}(q)} \, P_\mathcal{R}(\tau_q) P_\mathcal{R}(\tau_{k_+}). \tag{4.6}$$

In (4.6), we defined an effective $f_{\mathrm{NL}}^{\mathrm{eff}}(q)$ by combining the squeezed limit $f_{\mathrm{NL}}^{\mathrm{sq}}$ with a scale-dependent enhancement factor

$$E(k) = \{\Pi(k), |\alpha_k|^2\}, \tag{4.7}$$

which can assume two values, depending on whether we consider the heuristic formulas of section 2 for the consistency relations, or the gradient expansion formulas of section 3. For the rest of this section, we consider in parallel the two cases, normalizing E assuming that $E(k_+ \to 0) \to 1$ as we did in (4.6).

Noticing that $P_\mathcal{R}(\tau_p) = 2\pi^2 \mathcal{P}_\mathcal{R}(\tau_p)/p^3$ and $\mathcal{P}_\mathcal{R}(\tau_p) \simeq 2.1 \times 10^{-9}$ for mode exit during the initial slow-roll era (see e.g. (A.3)), we insert (4.6) in (4.5) to express the angular cross correlation as

$$C_l^{\mu T} \simeq 2.7 \times 10^{-17} \frac{2\pi}{l(l+1)} \, b_{(\mathrm{pbh})}(l), \tag{4.8}$$

where we define

$$b_{(\mathrm{pbh})}(l) \equiv \frac{6l(l+1)}{\ln\left(\frac{k_D(z_i)}{k_D(z_f)}\right)} \int \mathrm{d}\ln k_+ \, \Delta_l(k_+) \, j_l\left(k_+\chi_*\right) \int \mathrm{d}\ln q \, f_{\mathrm{NL}}^{\mathrm{eff}}(q) \left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}, \tag{4.9}$$

as the key quantity that parameterizes the multipole l dependence and the size of the angular correlator $\langle \mu T \rangle$ of eq (4.8). Notice that this quantity depends on the transfer function $\Delta_l(k)$ which we decompose as [66]

$$\Delta_l(k) = \rho(l)\Delta_l^{\mathrm{sw}}(k). \tag{4.10}$$

$\Delta_l^{\mathrm{sw}}(k)$ is the transfer function in the large-scale Sachs-Wolfe (SW) limit

$$\Delta_l^{\mathrm{sw}}(k) = \frac{1}{3} j_l(k\chi_*), \tag{4.11}$$

while

$$\rho(l) \simeq 1.08 \left[1 - 0.022l - 1.72 \times 10^{-4}l^2 + 2 \times 10^{-6}l^3 - 4.56 \times 10^{-9}l^4\right], \tag{4.12}$$

is an analytic fit that includes high-l corrections to the SW approximation [66]. Inserting (4.10) in (4.9) and using (4.11), we can analytically carry the integral over long momenta k_+ in (4.9), and factorize the result as

$$b_{(\mathrm{pbh})}(l) = \rho(l) \, b_{(\mathrm{pbh})}^{\mathrm{sw}}, \tag{4.13}$$

with

$$b_{(\mathrm{pbh})}^{\mathrm{sw}} = \left[\ln\left(\frac{k_D(z_i)}{k_D(z_f)}\right)\right]^{-1} \int \mathrm{d}\ln q \, f_{\mathrm{NL}}^{\mathrm{eff}}(q) \left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}. \tag{4.14}$$
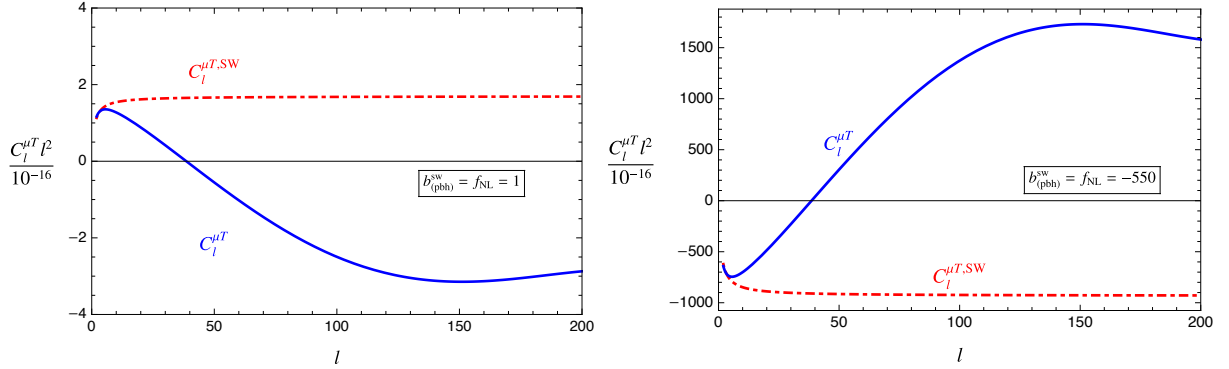
Figure 7: $C_l^{\mu T}$ (see eqs. (4.15) and (4.16)) for a scale independent purely local type bispectrum with $f_{\rm NL} = 1$ (Left) and $f_{\rm NL} = -550$ (Right).

The function $b_{\rm (pbh)}^{\rm sw}$ is normalized in such a way that for a purely local spectrum with constant non-linearity parameter, $f_{\rm NL}^{\rm eff}(q) = f_{\rm NL}$, we get $b_{\rm (pbh)}^{\rm sw} = f_{\rm NL}$. In this case, we reproduce the standard results for $C_l^{\mu T}$ in the SW limit [21, 64] as we show in the left panel of Figure 7. However, a non-trivial scale dependent of $f_{\rm NL}^{\rm eff}$ in (4.14) can significantly impact the amplitude $C_l^{\mu T}$ (4.8). We will concretely see an example of this fact in the PBH forming inflationary scenarios we are analyzing.

To summarize the formulas so far, for inflationary models that can produce a sizeable peak in the power spectrum, the amplitude and scale dependence (l) of the angular correlator $C_l^{\mu T}$ can be determined by

$$C_l^{\mu T} \simeq 2.7 \times 10^{-17} \frac{2\pi}{l(l+1)} \rho(l)\, b_{\rm (pbh)}^{\rm sw} \equiv \rho(l)\, C_l^{\mu T, {\rm sw}}, \tag{4.15}$$

where we defined the angular $\langle \mu T \rangle$ cross correlation in the Sachs-Wolfe limit $l \to 0$ as

$$C_l^{\mu T, {\rm sw}} = 2.7 \times 10^{-17} \frac{2\pi}{l(l+1)}\, b_{\rm (pbh)}^{\rm sw}. \tag{4.16}$$

Consistency relations and the function $b_{\rm (pbh)}^{\rm sw}$. The integral of eq (4.14) which provides the quantity $b_{\rm (pbh)}^{\rm sw}$ depends on the scale-dependent non-linearity parameter $f_{\rm NL}^{\rm eff}(q)$ of eq (4.6). We now make use of the information given by the consistency relation for the squeezed bispectrum of sections 2 and 3 ($f_{\rm NL}^{\rm sq}(q) = 5(1 - n_s(q))/12$) to characterise the size of $b_{\rm (pbh)}^{\rm sw}$ and its dependence on the dip position $k_{\rm dip}$. We show that for interesting values of $k_{\rm dip}$ this quantity is large and negative, and its value depends on the total amplification of the spectrum, as well as on properties of the non-attractor phase.

In Appendix C, we derive a power-law expression for $f_{\rm NL}^{\rm eff}(q)$: see eq (C.2). Using it, we can express $b_{\rm (pbh)}^{\rm sw}$ in terms of the following polynomial

$$b_{\rm (pbh)}^{\rm sw} = \frac{5}{3 \ln \left[k_d\left(z_i\right)/k_d\left(z_f\right) \right]} \sum_{n=2}^{n_f} \tilde{c}_E^{(n)} \frac{\Gamma(n/2)}{2^{(n+2)/2}} \left(\frac{k_D(z)}{k_{\rm dip}} \right)^n \Bigg|_{z_f}^{z_i}, \tag{4.17}$$

where the coefficients $\tilde{c}_E$, with $E = \{\Pi, |\alpha|^2\}$, include as above both the heuristic and gradient expansion approaches. They are given in (C.3) and (C.4). In the same way, $n_f = \{5, 6\}$ depending on which approach one adopts.
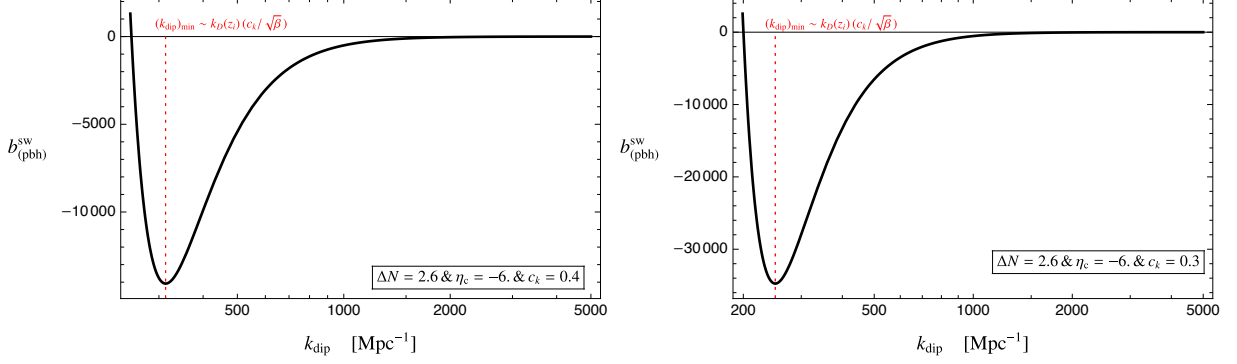
20

Figure 8: *The quantity $b_{(pbh)}^{sw}$ (4.17) vs k_{dip} within the gradient expansion formalism for inflationary scenario that contains a transient slow-roll violating phase that is characterized by the parameter choices: $\{\Delta N = 2.6, \eta_c = -6, c_k = 0.4\}$ (Left) and $\{\Delta N = 2.6, \eta_c = -6, c_k = 0.3\}$ (Right).*

This formula depends on k_{dip}, which we take within the μ-distortion band of eq (4.3) as

$$k_D(z_f) \simeq 46\,\mathrm{Mpc}^{-1} < k_{dip} < k_D(z_i) \simeq 11600\,\mathrm{Mpc}^{-1}. \tag{4.18}$$

The reason for this choice is two folds. First of all, if we take $k_{dip} \lesssim k_D(z_f)$ then the peak of the power spectrum would lie within the range of scales associated with μ-distortions, where we have stringent constraints on the peak amplitude of the power spectrum from COBE and FIRAS which limits $\langle \mu \rangle \lesssim 10^{-5}$ [72, 73]. Instead, when choosing $k_{dip} > k_D(z_f)$, the interesting effects of the scale dependent bispectrum around the dip scale are no longer present, as can be realized from (4.17) which behaves as $b_{(pbh)}^{sw} \to 0$ in the $k_D(z_i)/k_{dip} \to 0$ limit. Another concrete consequence of the choice of scales (4.18) is that the amplitude of $b_{(pbh)}^{sw}$ is controlled by the upper limit shown in eq (4.17), namely by powers of $k_D(z_i)/k_{dip}$. For the derivations we present below, we will repeatedly make us such simplification.

In the regime of interest (4.18), we find that $b_{(pbh)}^{sw}$ acquires a minimum at a critical value of the ratio $k_{dip}/k_D(z_i)$, whose location is

$$\left(\frac{k_{dip}}{k_D(z_i)} \right)^{n_f - 4}_{min} = -\frac{n_f\,\Gamma(n_f/2)}{2^{n_f/2}} \frac{\tilde{c}_E^{(n_f)}}{\tilde{c}_E^{(4)}}, \tag{4.19}$$

At its minimum (4.19), the amplitude of $b_{(pbh)}^{sw}$ (4.17) is negative and its final value is set by a competition between the n_f and $n = 4$ terms in the sum (4.17). Using (4.19) in (4.17), we can obtain this negative value at the minimum as

$$b_{(pbh)}^{sw}\Big|_{min} \simeq -\frac{5}{3\ln\left[k_D(z_i)/k_D(z_f)\right]} \frac{(n_f - 4)}{8n_f} |\tilde{c}_E^{(4)}| \left(\frac{k_D(z_i)}{k_{dip}} \right)^4_{min}, \tag{4.20}$$

where we use the fact $\tilde{c}_E^{(4)} < 0$ for both the approaches we are focusing in. We can be more explicit in discussing separately the two cases we are considering:

- *Heuristic approach:* Using the formulas (4.19) and (4.20), we can relate the location of the minimum of $b_{(pbh)}^{sw}$, as well as its amplitude at that position to the free parameter β_h, by setting $n_f = 6$, and making use of the coefficients we provide in (C.4).
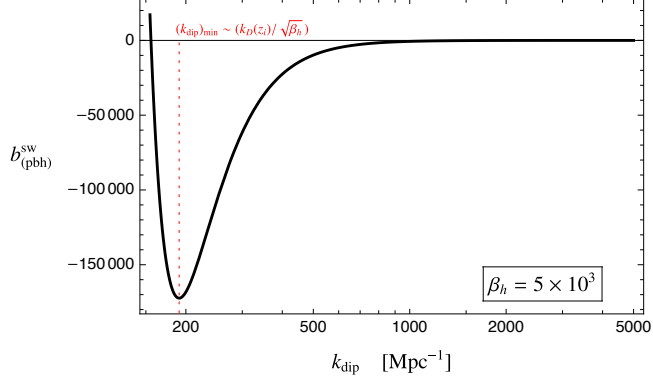
21

Figure 9: *The quantity $b_{(\text{pbh})}^{\text{sw}}$ (4.17) vs k_{dip} in the heuristic approach of section 2 for an inflationary scenario that can generate a $\Pi_T \simeq 10^7$ growth in the power spectrum.*

At leading order in the large parameter $\beta_h \gg 1$, this procedure gives

$$\left(\frac{k_D(z_i)}{k_{\text{dip}}}\right)^2_{\text{min}} \simeq 0.74\beta_h \simeq 1.5\,\Pi_T^{1/2} \quad \longrightarrow \quad b_{(\text{pbh})}^{\text{sw}}\Big|_{\text{min}} \simeq -7\times 10^{-3}\,\beta_h^2 \simeq -2.8\times 10^{-2}\,\Pi_T, \quad (4.21)$$

where we use $\beta_h \simeq 2\Pi_T^{1/2}$ from (2.8). For inflationary scenarios with a total of $\Pi_T \simeq 10^7$ enhancement, these results predict $k_D(z_i) \simeq 69\,k_{\text{dip,min}}$ and $b_{(\text{pbh})}^{\text{sw}}\big|_{\text{min}} \simeq -\mathcal{O}(10^5)$. Importantly, the size of $|b_{(\text{pbh})}^{\text{sw}}|$ is much larger than the maximal values of $|f_{\text{NL}}^{\text{sqz}}|$ around the dip position, as we analyzed in section 2 in eq (2.24). This is due to the fact that $b_{(\text{pbh})}^{\text{sw}}$ involves an integration over momenta that – picking up contributions from the scale-dependent squeezed bispectrum – considerably enhances its value with respect to the constant $f_{\text{NL}}^{\text{sq}}$ case (see comment after eq (4.14)). This fact improves the chances of detection.

- *Gradient expansion formalism:* In this approach, we use (C.3) together with (4.19) and (4.20) to determine the location of the minimum and the resulting $b_{(\text{pbh})}^{\text{sw}}\big|_{\text{min}}$ in terms of model parameters as

$$\left(\frac{k_D(z_i)}{k_{\text{dip}}}\right)_{\text{min}} \simeq \frac{0.34}{c_k}\frac{\eta_c(1+c_k^2)}{(\eta_c+1)}\sqrt{\beta} \quad \longrightarrow \quad b_{(\text{pbh})}^{\text{sw}}\Big|_{\text{min}} \simeq -7.5\times 10^{-3}\left(\frac{0.34}{c_k}\right)^4\frac{\eta_c^4(1+c_k^2)^4}{(\eta_c+1)^4}\beta^2,$$
$$(4.22)$$

where dependence of $\beta \gg 1$ on the properties of the non-attractor era is given by (B.4). For typical parameter choices that leads to a $\Pi_{\text{tot}} \simeq 10^7$ enhancement in the power spectrum (see Figure 4), we have $\beta \simeq 10^3$ and (4.22) predict $k_D(z_i) \simeq (40-50)\,k_{\text{dip,min}}$ and $b_{(\text{pbh})}^{\text{sw}}\big|_{\text{min}} \simeq -\mathcal{O}(10^4)$.

Comparing the result obtained from the heuristic approach and the gradient expansion formalism, we notice that although they agree for the location of the minimum $k_{\text{dip,min}}$, the amplitude of $b_{(\text{pbh})}^{\text{sw}}$ at the minimum differs by an order of magnitude. However, expression (4.22) indicates that for smaller choices of the model parameter c_k the overall amplitude $|b_{(\text{pbh})}^{\text{sw}}|$ and the location of the minimum of $b_{(\text{pbh})}^{\text{sw}}$ tend to agree better with the heuristic approach, as in this case $k_{\text{dip,min}}$ shifts to slightly larger scales, causing the overall amplitude $|b_{(\text{pbh})}^{\text{sw}}|_{\text{min}}$ to become larger. We confirm these findings in Figures 8 and 9, where we present the dependence of $b_{(\text{pbh})}^{\text{sw}}$ on the dip scale k_{dip} in both the heuristic approach and gradient expansion formalism for representative parameter choices that can generate substantial growth in the power spectrum, as required for PBH formation.

The multipole dependence of $C_l^{\mu T}$. We collect these results to consider the scale dependence (l) of angular cross correlation $C_l^{\mu T}$. For concreteness, we set $k_{\text{dip}} = 10^3\,\text{Mpc}^{-1}$ to first determine the amplitude
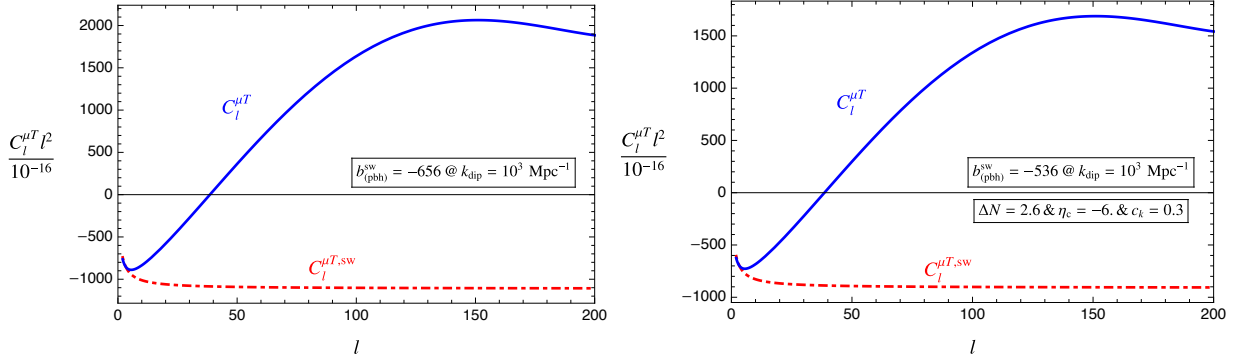
Figure 10: $C_l^{\mu T}$ for inflationary models that can generate a $\Pi_{\text{tot}} \simeq 10^7$ enhancement in the power spectrum. The left panel represents the results obtained in the heuristic approach with $\beta_h = 5 \times 10^3$. In the right panel, we show the corresponding $C_l^{\mu T}$ using the formulas in the gradient formalism where the parameter choices that characterize the slow-roll violating phase is indicated in the boxes.

of $b_{(\text{pbh})}^{\text{sw}}$ from (4.17). We then represent in Figure 10 the multipole dependence of the quantity $C_l^{\mu T}$ using (4.12), (4.15) and (4.16). We notice that, in addition to the enhancement of the amplitude of the angular correlator $C_l^{\mu T} \propto b_{(\text{pbh})}^{\text{sw}}$ for $|b_{(\text{pbh})}^{\text{sw}}| \gg 1$, its scale dependence from large (small l) to small scales (large l). This is because $b_{(\text{pbh})}^{\text{sw}} < 0$ for the phenomenologically interesting k_{dip} values we are focusing defined in (4.18). As explained in detail in [20], the origin of this behavior can be traced back to the change of sign of the scale dependent $f_{\text{NL}}^{\text{eff}}$ in (4.14) which occurs at $q = q_{\text{dip}}$ followed by its growth in the negative direction. This result implies that for inflationary models that is capable to generate PBH populations, μ distortions become anti-correlated with temperature anisotropies at large scales. We note that a similar profile for $C_l^{\mu T}$ can be generated by a standard local type non-Gaussianity that exhibit a large and negative scale-independent non-linearity parameter $f_{\text{NL}} = -\mathcal{O}(100)$ as we show in the right panel of Figure 7.

4.2 Phenomenology of scale dependent trispectrum: $C_l^{\mu\mu}$

We now investigate non-Gaussian[7] contribution to $\langle \mu\mu \rangle$ self correlation induced by the scale dependent trispectrum in PBH forming inflationary models. With this aim, we focus our attention to the collapsed limit $k_+ \to 0$ of the curvature perturbation 4-pt function. In this limit, the non-Gaussian contribution (D.10) to $\langle \mu\mu \rangle$ reads as

$$C_{l,\text{NG}}^{\mu\mu} \simeq \frac{2.65}{\pi^5} \int d\ln k_+ \, j_l^2 \left(k_+ \chi_* \right) \int d\ln q \, d\ln p \, k_+^3 q^3 p^3 \lim_{k_+ \to 0} T_{\mathcal{R}}\left(q, p\right) \left[e^{-2q^2/k_D^2(z)} \right]_{z_f}^{z_i} \left[e^{-2p^2/k_D^2(z)} \right]_{z_f}^{z_i},$$
(4.23)

where in the collapsed limit we use $\vec{k}_1 \to -\vec{k}_2 \Rightarrow k_1 = k_2 \equiv q$ and $\vec{k}_3 \to -\vec{k}_4 \Rightarrow k_3 = k_4 \equiv p$. The expression (4.23) simplifies further if we consider symmetric folded kite configuration $q = p$. To see this fact explicitly, we first focus on the trispectrum in this configuration which reads as

$$\lim_{k_+ \to 0} T_{\mathcal{R}}\left(q, k_+\right) \simeq 4 \underbrace{\tau_{\text{NL}}^{\text{col}}(q) E(q)^2}_{\equiv \left[\frac{6}{5} f_{\text{NL}}^{\text{eff}}(q)\right]^2} P_{\mathcal{R}}(\tau_q) P_{\mathcal{R}}(\tau_q) P_{\mathcal{R}}(\tau_{k_+}),$$
(4.24)

---

[7]Disconnected part of the 4-pt function also leads to a Gaussian contribution for $\langle \mu\mu \rangle$ with $l \neq 0$. We analyze this contribution in Appendix D and find that it can be neglected compared to the $C_{l,\text{NG}}^{\mu\mu}$ we investigate in this section.
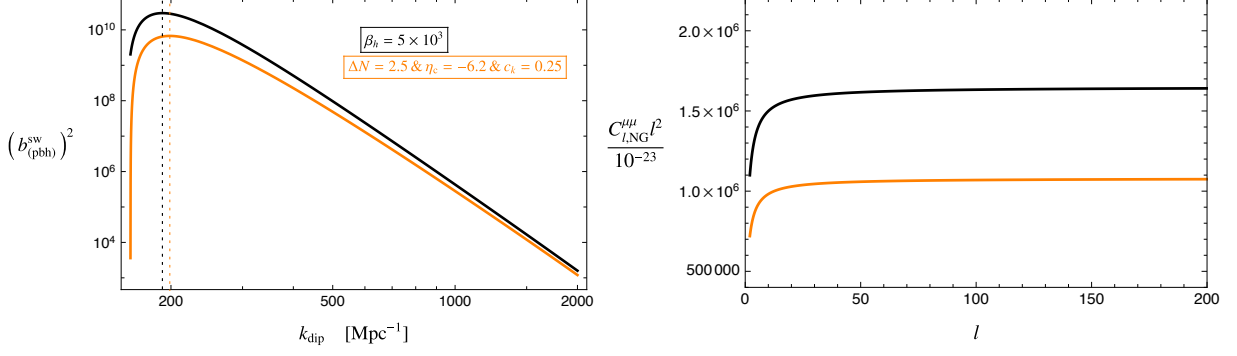
Figure 11: *Left: Logarithmic plot of $b_{(\text{pbh})}^{\text{sw}\ 2}$ as a function of the location of the dip feature for inflationary models that can generate a peak required for PBH formation. Right: Scale invariance of $C_l^{\mu\mu}$ for the parameter choices shown in the left panel. The location of the dip feature in the power spectrum is taken as $k_{\text{dip}} = 10^3\,\text{Mpc}^{-1}$ for both curves, corresponding to $b_{(\text{pbh})}^{\text{sw}\ 2} \simeq \{4.3 \times 10^5, 2.8 \times 10^5\}$ for the black and orange curves respectively.*

where we make use of the consistency relation (3.38) and the definition (4.6) of $f_{\text{NL}}^{\text{eff}}$. The terms under the braces in (4.24) can be identified as the effective τ_{NL}, namely $\tau_{\text{NL}}^{\text{eff}}(q) \equiv 36 f_{\text{NL}}^{\text{eff}}(q)^2/25$, which represents a scale-dependent generalization of the standard local type trispectrum with a constant τ_{NL} [74,75]. Inserting (4.24) into (4.23) and focusing on $q = p$ configuration, we can describe the non-Gaussian $\langle\mu\mu\rangle$ selfcorrelation as

$$C_{l,\text{NG}}^{\mu\mu} \simeq 6.1 \times 10^{-24}\, \frac{2\pi}{l(l+1)} \left(b_{(\text{pbh})}^{\text{sw}}\right)^2, \tag{4.25}$$

where $b_{(\text{pbh})}^{\text{sw}}$ is defined as in (4.14). This result implies that, in single-field PBH inflationary scenarios, $\langle\mu\mu\rangle$ is scale invariant, *i.e.* $l(l+1)C_l^{\mu\mu} = $ constant similarly to a purely local form trispectrum [21]. However, its amplitude can be enhanced by a factor of $b_{(\text{pbh})}^{\text{sw}\ 2} \gg 1$ compared to the latter. In particular, considering a $|b_{(\text{pbh})}^{\text{sw}}| \simeq \mathcal{O}(10^2)$ which arises in scenarios with a dip feature located at $k_{\text{dip}} \approx 10^3\,\text{Mpc}^{-1}$, we can obtain a total enhancement of $\mathcal{O}(10^5)$. Another point worth stressing is the fact that $C_{l,\text{NG}}^{\mu\mu} > 0$ for all multipoles, as it should be clear from (4.25). We illustrate these points in Figure 11 where we show $b_{(\text{pbh})}^{\text{sw}\ 2}$ and $C_l^{\mu\mu}$ for representative scenarios that can produce a large peak in the power spectrum. The left panel of the Figure informs us that $b_{(\text{pbh})}^{\text{sw}\ 2}$ is maximal at the same location (4.19) where $b_{(\text{pbh})}^{\text{sw}}$ has a minimum.

4.3 Prospects of detectability of μ distortion anisotropies

We now develop Fisher forecasts to estimate the detectability of the signals we investigated in Sections 4.1 and 4.2. We use a 1×1 Fisher information matrix for the parameter $b_{(\text{pbh})}^{\text{sw}} \equiv b$ [76],

$$F = \sum_l \sum_{X,Y} \frac{\partial C_l^{\mu X, S}}{\partial b} \text{Cov}^{-1}(C_l^{\mu X} C_l^{\mu Y}) \frac{\partial C_l^{\mu Y, S}}{\partial b}, \tag{4.26}$$

where the sum runs over $X, Y = \{T, \mu\}$; S refers to the theoretical predictions we derived in eqs. (4.15) and (4.25) and Cov^{-1} is the inverse of the covariance matrix, defined as [76]

$$\text{Cov}(C_l^{\mu X} C_l^{\mu Y}) = \frac{1}{2l+1} \left[(C_l^{\mu X, S} + C_l^{\mu X, N})(C_l^{\mu Y, S} + C_l^{\mu Y, N}) + (C_l^{\mu\mu, S} + C_l^{\mu\mu, N})(C_l^{XY, S} + C_l^{XY, N}) \right], \tag{4.27}$$

where N represents the experimental noise. Using (4.26) and (4.27), the signal to noise ratio (SNR) is given by $(S/N)^2 = b^2 F$. In the derivation of the components of the covariance matrix, some simplifications
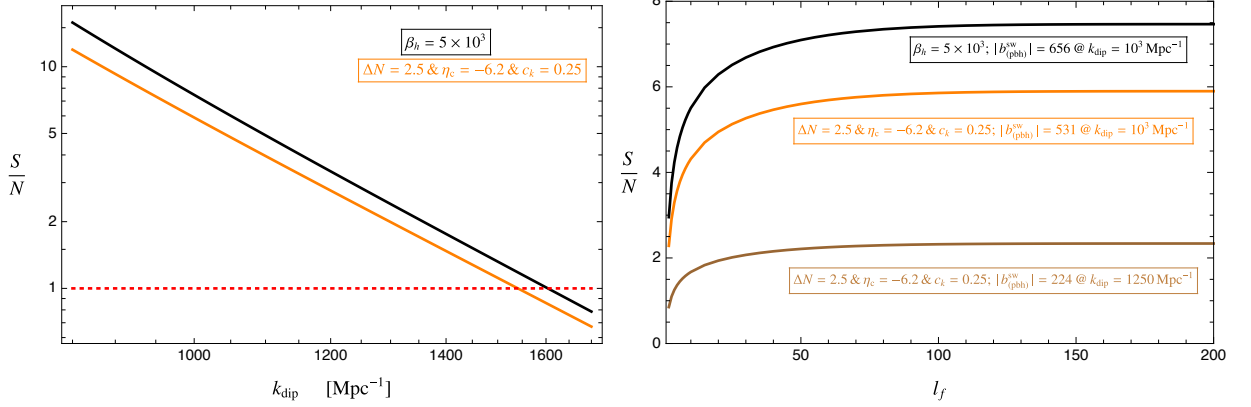
Figure 12: *The cumulative signal to noise ratio (4.30) as a function of the dip location in the power spectrum for $k_{dip} \gtrsim 830 \, \mathrm{Mpc}^{-1}$ (Left). SNR as a function of maximum multipole l_f at fixed k_{dip} and hence $|b^{sw}_{(pbh)}|$ (Right).*

can be made. First of all, note that for scales we are interested in $2 < l < \mathcal{O}(100)$, the experimental noise for TT correlator can be neglected $C_l^{TT,S} \gg C^{TT,N}$, with

$$C_l^{TT,S} = \frac{36\pi}{25} \int d \ln k \, \mathcal{P}_\mathcal{R}(\tau_f, k) \Delta_l^2(k). \tag{4.28}$$

Similarly, μ and T instrumental noises are uncorrelated and therefore we can set $C_l^{\mu T,N} = 0$ in (4.27). On the other hand, as can be confirmed from (4.25) and the discussion it follows, $\mu\mu$ correlation is dominated by the instrumental noise, *i.e.* $C_l^{\mu\mu,S} \ll C_l^{\mu\mu,N}$, *i.e.* for phenomenologically interesting values we are focusing where $(b^{sw}_{(pbh)})^2 \lesssim 10^5$. For a PIXIE like experiment [77], this noise can be modeled as $C_l^{\mu\mu,N} \simeq 4\pi \, \mu_{min}^2 \, e^{l^2/84^2}$ [21] where μ_{min} denotes the minimum detectable μ distortion signal. Finally, for the $X = Y = T$ component of the covariance matrix another simplification arise by noticing that $(C_l^{\mu T})^2 \ll C_l^{\mu\mu,N} C_l^{TT}$. This relation can be confirmed noticing the SW limit of (4.28): $C_l^{TT,sw} = 2\pi \mathcal{A}_s/(25l(l+1))$ with $\mathcal{A}_s = 2.1 \times 10^{-9}$, (4.15) and the relation for $C_l^{\mu\mu,N}$ above. In light of these arguments, SNR yields as

$$\left(\frac{S}{N}\right)^2 = \sum_{l=2}^{l_f} (2l+1) \left[\frac{\left(C_l^{\mu T,S}\right)^2}{C_l^{TT,S} C_l^{\mu\mu,N}} + 2 \frac{C_l^{\mu\mu,S}}{C_l^{\mu\mu,N}} + 2 \left(\frac{C_l^{\mu\mu,S}}{C_l^{\mu\mu,N}}\right)^2 \right], \tag{4.29}$$

where we carry the sum up to $l_f = 200$. To accurately estimate the first term in (4.29), we require the knowledge of C_l^{TT} in (4.28) by taking into account the transfer function Δ_l. In this respect, [64, 66] found that the contribution from the first term in (4.29) corresponds to 40% of the result obtained by adopting the SW limit, *i.e.* by taking $\rho(l) \to 1$ in (4.15), and adopting $C_l^{TT,sw} = 2\pi \mathcal{A}_s/(25l(l+1))$ [8]. At the same time, the second and third term in eq (4.29) can be evaluated directly from our results of section (4.2). We obtain

$$\frac{S}{N} \simeq 0.99 \times 10^{-3} \left(\frac{10^{-8}}{\mu_{min}}\right) |b^{sw}_{(pbh)}| + 2.15 \times 10^{-8} \left(\frac{10^{-8}}{\mu_{min}}\right)^2 |b^{sw}_{(pbh)}|^2, \tag{4.30}$$

where we normalized the minimum detectable distortion to $\mu_{min} \approx 10^{-8}$, as relevant for a PIXIE-like experiment [78]. In (4.30), the first contribution corresponds to the sum of the first two terms in (4.29)

---

[8]It should be noted that we ignore the enhancement of the power spectrum in (4.28) to arrive this expression. This is justified because the smallest scales we are interested in corresponds to $l_f = 200$, while for the scenarios we are focusing in this work, the enhancement in the power spectrum occurs for scales corresponding to $l \gg l_f$.

which are of the same order of magnitude for any values of b and $\mu_{\min}$. In fact, the second term is weighted by the cross component of the covariance matrix, and it contributes twice as large as the μT (the first) term in (4.29). At the same time, the last term represents the contribution to SNR from $\mu\mu$ alone: it is generically sub-dominant compared to the first two terms in (4.29), but it nevertheless provides a non-negligible contribution. These results imply that $\langle\mu\mu\rangle$ self correlations improve the prospects for the observability of distortion anisotropies.

Using (4.30), we argue that an inflationary scenario with $|b^{\mathrm{sw}}_{(\mathrm{pbh})}| \gtrsim 989$ is detectable at 1σ level for a PIXIE like experiment. We note from (4.17) that such values of $|b^{\mathrm{sw}}_{(\mathrm{pbh})}|$ can be obtained for PBH forming scenarios where k_{dip} lies close its smallest allowed values ($k_{\mathrm{dip}} \gtrsim 830\,\mathrm{Mpc}^{-1}$) dictated by the limits on average distortion $\mu \lesssim 10^{-5}$ (See figure 14). Moreover, for an experimental design comparable to PRISM [79] with $\mu_{\min} = 10^{-9}$, this situation can be improved since a smaller value $|b^{\mathrm{sw}}_{(\mathrm{pbh})}| \gtrsim 99$ is required for the detectability. Focusing our attention to the latter, in the left panel of Figure 12 we present the SNR (4.30) in terms of the location of the dip scales k_{dip} allowed by the $\mu \lesssim 10^{-5}$ limit. As can be also inferred from (4.17), for $k_{\mathrm{dip}} \gtrsim 1600\,\mathrm{Mpc}^{-1}$ SNR drops below unity. However it satisfies $S/N > 1$ for the allowed region of $830\,\mathrm{Mpc}^{-1} \lesssim k_{\mathrm{dip}} \lesssim 1600\,\mathrm{Mpc}^{-1}$. In the right panel, we show cumulative SNR as a function of l_f for three different scenarios defined in this range of scales together with their corresponding $|b^{\mathrm{sw}}_{(\mathrm{pbh})}|$. (Recall that l_f corresponds to the upper limit of the sum in eq (4.29).) This implies that our SNR estimate does not improve significantly and saturates for $l_f \gtrsim 100$.

μ-distortion anisotropies as a probe of PBH populations. We conclude our analysis with some implications of our findings for PBH populations. In inflationary scenarios where the curvature power spectrum has a pronounced peak of order $\Pi_{\mathrm{tot}} \simeq 10^7$ located at wave-number $k_{\mathrm{peak}} \simeq 100 k_{\mathrm{dip}} \gg k_{\mathrm{CMB}} = 0.05\,\mathrm{Mpc}^{-1}$, PBHs may have formed during the radiation dominated era upon horizon re-entry of modes whose wave-number is comparable with the peak scale [1, 2]. Assuming that the power spectrum is sufficiently peaked, we can relate the mass of the PBH today to the location of the dip scale in the power spectrum as [17, 63],

$$M_{\mathrm{pbh},0} = \mathcal{A}\,\mathcal{M}\,M_{\mathrm{pbh,f}} \simeq \mathcal{A}\,\mathcal{M}\left(\frac{\gamma}{0.2}\right)\left(\frac{k_{\mathrm{dip}}}{10^4\,\mathrm{Mpc}^{-1}}\right)^{-2} 2.4 M_\odot, \tag{4.31}$$

where $\mathcal{A}$ and $\mathcal{M}$ indicate the amount mass gain that can arise due to accretion and merger effects for the corresponding PBH seed mass $M_{\mathrm{pbh,f}}$ at the time of formation. γ is the ratio of the PBH mass to the mass within the causal horizon and can take values between $\gamma = 0.2$ [3, 80] and $\gamma = 0.8$ [81] depending on the assumptions about PBH formation in the radiation dominated era.

Considering the phenomenologically interesting values of dip scales $830\,\mathrm{Mpc}^{-1} \lesssim k_{\mathrm{dip}} \lesssim 1600\,\mathrm{Mpc}^{-1}$ we identified above, eq (4.31) implies that PBHs with $M_{\mathrm{pbh},0} = 10 - 100 M_\odot$ can be probed by μ distortion anisotropies assuming negligible accretion and merger coefficients: $\mathcal{A}, \mathcal{M} \to 1$. Taking into account these effects within the ranges $10^5 \geq \mathcal{A} \geq 1$, $10^5 \geq \mathcal{M} \geq 1$ [63, 82], a detection of μ distortion anisotropies through $\langle\mu T\rangle$ and $\langle\mu\mu\rangle$ angular correlators can be therefore considered as a useful tool to distinguish astrophysical vs primordial the origin of super massive black holes (SMBH) with masses $M_{\mathrm{pbh},0} = 10^6 - 10^9\,M_\odot$ today.

5 Discussion

In single field-inflationary models that are capable of generating PBH populations, the power spectrum of curvature perturbation has interesting universal features such as the presence of a pronounced dip. Focusing on the heuristic approach introduced in [19] and gradient expansion formalism [22] (see sections 2 and 3), we explicitly demonstrated that the position of the dip in momentum space is uniquely determined by the global enhancement $\Pi \simeq 10^7$ of the power spectrum via $k_{\mathrm{dip}} \simeq \Pi^{-1/4} k_{\mathrm{peak}} \ll k_{\mathrm{peak}}$, implying its occurrence on scales much larger than the peak scale associated with PBH formation. More importantly,

in sections 3.2 and 3.3 we analyzed consistency relations for n-point correlators ($n = 3, 4$) of curvature perturbation in the vicinity of the dip feature. We found that non-Gaussianity parameters satisfy the conditions $f_{\rm NL}^{\rm sq} = 5(1 - n_s)/12$ and $\tau_{\rm NL}^{\rm col} = (6 f_{\rm NL}^{\rm sq}/5)^2$ in a non-trivial scale dependent manner, allowing us to derive a new set of consistency conditions in terms of the the global enhancement $\Pi \simeq 10^7$ in the power spectrum and relate their scale dependence to its slope. In scenarios where the dip feature lies within the scale range (4.18) where μ-distortions are generated, the characteristic scale dependence of such n-point correlators offers us a unique chance to probe the underlying PBH formation mechanism at relatively large scales through the CMB spectral μ-distortion anisotropies.

In fact, developing upon the ideas first presented in [20], we explored the implications of the consistency conditions for the bispectrum and trispectrum on the cross correlation between spectral distortions and temperature anisotropies $\langle \mu T \rangle$ and distortion self correlations $\langle \mu \mu \rangle$. In this context, in section 4.1, we studied $\langle \mu T \rangle$ angular correlator induced by the squeezed limit bispectrum derived from the consistency relation $f_{\rm NL}^{\rm sq} = 5(1 - n_s)/12$. We found that the pronounced characteristic scale dependence of the bispectrum can alter the amplitude and overall multipole dependence of $\langle \mu T \rangle$ significantly with respect to more standard cases (See section 4.1). These results confirm the findings obtained earlier in [20] and put them in a firmer footing through the use of consistency condition. In Section 4.2, utilizing the consistency relation $\tau_{\rm NL}^{\rm col} = (6 f_{\rm NL}^{\rm sq}/5)^2$, we studied for the first time the influence of the enhanced collapsed limit trispectrum present around the dip scale and found that it induces sizeable distortion self-correlations $\langle \mu \mu \rangle$ that are scale invariant: $l(l+1)C_l^{\mu\mu} = \text{constant}$. Including the information that we can gain from the μ self-correlations, in section 4.3, we then showed that the prospects of detectability of μ distortion anisotropies are enhanced compared to considering μT correlations alone [20]. In particular, we found that for phenomenologically allowed and interesting values of dip location in the range $830\,\text{Mpc}^{-1} \lesssim k_{\rm dip} \lesssim 1600\,\text{Mpc}^{-1}$, μ-distortion anisotropies – induced by non-Gaussian consistency relations – should be observable for a PIXIE or PRISM like experimental design.

Considering the relation between the dip and peak scale in the power spectrum $k_{\rm peak} \simeq 100 k_{\rm dip}$, spectral distortion anisotropies we derived in this work can be utilized to identify the formation mechanism of BHs with masses $M_{\rm pbh,0} \simeq 10 - 100\,M_\odot$ today, and/or SMBHs with $M_{\rm pbh,0} \simeq 10^6 - 10^9\,M_\odot$ taking into account strong accretion and merger effects (see Section 4.3). Furthermore, the properties we studied in this work can also be also considered as a useful tool for discriminating inflationary models of PBH formation as in some scenarios based on particle production [83–86], the dip feature is not present and therefore unlikely to produce interesting μ-distortion anisotropies at large scales.

This work can be extended in a few directions. First of all, our analysis on the squeezed limit bispectrum and the collapsed limit trispectrum does not include finite but sub-leading corrections of order $\mathcal{O}(k_+/q)$ in terms of the soft momenta k_+. It would be interesting to include such corrections to to investigate the impact of *not so squeezed and collapsed limit* n-point correlators of curvature perturbation on the μT and $\mu\mu$ angular correlators. Finally, it would be interesting to extend our analysis to derive predictions on the μ distortion anisotropies for scenarios including multiple scalar fields [87–89]. Such an analysis would be helpful in comparing the general single field predictions we derived in this work and guide us towards a better understanding for the formation mechanism of BHs with astrophysically relevent masses.

Acknowledgments

A The power spectrum $\mathcal{P}_\mathcal{R}(\tau_k)$ and fractional velocity $v_\mathcal{R}$

For the initial slow-roll phase, the standard solution for the curvature perturbation that reduces to the standard Bunch Davies vacuum can be written as [31],

$$\mathcal{R}_k^{\mathrm{sr}} = \frac{iH}{M_{\mathrm{pl}}} \frac{e^{-ik\tau}}{\sqrt{4\epsilon_{\mathrm{sr}}k^3}} \left(1 + ik\tau\right), \qquad \tau_k/\tau_0 > 1, \tag{A.1}$$

with $\epsilon_{\mathrm{sr}} \ll 1$ is the slow-roll parameter which we assume to be constant adopting $\eta = \dot{\epsilon}/\epsilon H \to 0$. Using the solution (A.1), the real and imaginary part of $v_\mathcal{R}$ (3.5) can be derived as

$$v_\mathcal{R}^R(c_k) = -\frac{c_k^2}{3(1 + c_k^2)}, \quad v_\mathcal{R}^I(c_k) = -\frac{c_k^3}{3(1 + c_k^2)}, \qquad \tau_k/\tau_0 > 1, \tag{A.2}$$

where c_k is defined as in (3.20). Notice that the imaginary part of $v_\mathcal{R}$ in (A.2) includes an extra factor of c_k compared to the real part. We note that unless $c_k = 1$, this translates into an extra suppression for the imaginary part of the fractional velocity and hence the imaginary part of α_k (3.4) as we will show explicitly below. On the other hand, using (A.1), the power spectrum evaluated at around horizon crossing is given by

$$\mathcal{P}_\mathcal{R}(\tau_k) = \frac{k^3}{2\pi^2} |\mathcal{R}_k(\tau_k)|^2 = \frac{H^2}{8\pi^2 \epsilon_{\mathrm{sr}} M_{\mathrm{pl}}^2} \left(1 + c_k^2\right), \qquad \tau_k/\tau_0 > 1. \tag{A.3}$$

Next, we need to determine $\mathcal{R}(\tau_k)$, $v_\mathcal{R}$ and $\mathcal{P}_\mathcal{R}(\tau_k)$ in the non-attractor era, i.e. for $\tau_k/\tau_0 < 1$. This was done in [20] using a matching procedure for $\mathcal{R}_k$ and its derivative at the transition time $\tau = \tau_0$. The resulting fractional velocity in the non-attractor era, i.e. for $\tau_k/\tau_0 < 1$ is given by [20]

$$v_\mathcal{R}^R(\tau) = -\frac{y}{3} \left[\frac{f_1 f_3 - y_0 (f_1 f_4 + f_2 f_3) + y_0^2 (f_1 f_3 + f_2 f_4)}{f_3^2 - 2y_0 f_3 f_4 + y_0^2 (f_3^2 + f_4^2)} \right], \tag{A.4}$$

$$v_\mathcal{R}^I(\tau) = -\frac{y}{3} \left[\frac{y_0^2 (f_1 f_4 - f_2 f_3)}{f_3^2 - 2y_0 f_3 f_4 + y_0^2 (f_3^2 + f_4^2)} \right], \tag{A.5}$$

where we defined $y \equiv -k\tau$ and the functions $f_n = f_n(y, y_0, \nu)$, for $n = 1, 2, 3, 4$ in terms of the Bessel function of the first and second kind as

$$\begin{aligned} f_1(y, y_0, \nu) &= J_{\nu-1}(y_0)Y_{\nu-1}(y) - Y_{\nu-1}(y_0)J_{\nu-1}(y), \\ f_2(y, y_0, \nu) &= J_\nu(y_0)Y_{\nu-1}(y) - Y_\nu(y_0)J_{\nu-1}(y), \end{aligned} \tag{A.6}$$

satisfying the following relations $f_4 = f_1(y, y_0, \nu+1)$, $f_3 = -f_2(y_0, y, \nu)$ with $\nu = (3 + \eta_{\mathrm{c}})/2$. The continuity of the real and the imaginary part of the fractional velocity in passing from slow-roll to non-attractor era can be confirmed explicitly from the eqs in (A.4) and (A.5) at $\tau_k = \tau_0$. Finally, the power spectrum evaluated at $\tau = \tau_k$ for modes that leave the horizon during the non-attractor phase ($\tau_k/\tau_0 < 1$) is given by [20]

$$\mathcal{P}_\mathcal{R}(\tau_k) = \frac{H^2}{8\pi^2 \epsilon_{\mathrm{sr}} M_{\mathrm{pl}}^2} \frac{c_k^{2\nu} \pi^2}{4} \left(\frac{k}{\mathcal{H}_0}\right)^{2-2\nu} \left[f_3^2 - 2y_0 f_3 f_4 + y_0^2(f_3^2 + f_4^2) \right]_{\tau=\tau_k}. \tag{A.7}$$

Similarly, the continuity of the power spectrum can be confirmed explicitly by evaluating (A.7) at the transition point $\tau_k = \tau_0$ at which it reduces to (A.3).

B   The functions $D(\tau_k)$, $F_k(\tau_k)$ and the enhancement factor α_k

For the two phase background model parametrized by the pump field profile in eq. (3.18), scale dependent functions $D(\tau_k)$, $F(\tau_k)$ (see eqs. (3.6) and (3.7)) are calculated in [31] for modes that exit during the initial slow-roll era ($k/\mathcal{H}_0 < c_k$) and in [20] for mode exit during the non-attractor era ($e^{\Delta N} > k/\mathcal{H}_0 > c_k$). In particular, for mode exit during the initial slow-roll stage, these functions are found to be

$$D(\tau_k) = 1 - \frac{3}{(\eta_c + 3)c_k^3} \left[e^{-(\eta_c+3)\Delta N} + \frac{\eta_c}{3} \right] \left(\frac{k}{\mathcal{H}_0} \right)^3, \tag{B.1}$$

$$F(\tau_k)\, k^2 = \frac{c_k^2}{6} - \left[\frac{\eta_c\, e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)(\eta_c+1)} + \frac{\eta_c}{2(\eta_c+3)} \right] \left(\frac{k}{\mathcal{H}_0} \right)^2 + \left[\frac{e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)c_k} + \frac{\eta_c}{3(\eta_c+3)c_k} \right] \left(\frac{k}{\mathcal{H}_0} \right)^3,$$

where $\tau_f/\tau_0 \equiv x_f = e^{-\Delta N}$ with ΔN denoting the duration of the non-attractor era. On the other hand, for modes that exit during the non-attractor era, $e^{\Delta N} > k/\mathcal{H}_0 > c_k$ they are given by

$$D(\tau_k) = -\frac{3}{\eta_c + 3} - \frac{3\, e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)c_k^{\eta_c+3}} \left(\frac{k}{\mathcal{H}_0} \right)^{(\eta_c+3)},$$

$$F(\tau_k)\, k^2 = \frac{e^{-(\eta_c+3)\Delta N}}{(\eta_c+1)(\eta_c+3)c_k^{\eta_c+1}} \left(\frac{k}{\mathcal{H}_0} \right)^{(\eta_c+3)} + \frac{c_k^2}{2(\eta_c+3)} - \frac{e^{-2\Delta N}}{2\,(\eta_c+1)} \left(\frac{k}{\mathcal{H}_0} \right)^2. \tag{B.2}$$

Together with the fractional velocities $v_{\mathcal{R}}^R$, $v_{\mathcal{R}}^I$ we found in (A.2), (A.4) and (A.5), one can make use of the formulas (B.1) and (B.2) to describe full spectral behavior of the enhancement factor α_k from large to small scales in a continuous way by utilizing the formulas (3.3) and (3.4).

B.1 α_k for mode exit during the initial slow-roll era.

In this subsection, using the formulas of the previous section, we provide an expression for α_k for $k/\mathcal{H}_0 < c_k \leq 1$ that we utilize in the main text extensively.

We begin by noticing that for inflationary scenarios that contains a transient $\Delta N \sim \mathcal{O}(1)$ slow-roll violating $\eta_c \leq -6$ phase, the terms in the square brackets of (B.1) are dominated by the exponential factors so we can further simplify the quantities $D(\tau_k)$ and $F(\tau_k)k^2$. Then noting the definitions (3.3),(3.4) and (A.2), the real and imaginary part of the enhancement factor can be parametrized as

$$\alpha_k^R \simeq \alpha_{(0)}^R \left[1 - \beta \left(\frac{k}{\mathcal{H}_0} \right)^2 + \frac{c_k(\eta_c+1)}{(1+c_k^2)\eta_c} \beta \left(\frac{k}{\mathcal{H}_0} \right)^3 \right], \tag{B.3}$$

$$\alpha_k^I \simeq \alpha_{(0)}^I - \alpha_{(0)}^R \frac{(\eta_c+1)}{(1+c_k^2)\eta_c} \beta \left(\frac{k}{\mathcal{H}_0} \right)^3,$$

where we defined an exponentially large number $\beta = \beta(\eta_c\, \Delta N, c_k)$, $\alpha_{(0)}^R$ and $\alpha_{(0)}^I$ as

$$\alpha_{(0)}^R = 1 - \frac{c_k^2}{6} - \frac{c_k^2}{3(1+c_k^2)}, \quad \alpha_{(0)}^I = -\frac{c_k^3}{3(1+c_k^2)}, \quad \beta = -\frac{\eta_c\, e^{-(3+\eta_c)\Delta N}}{\alpha_{(0)}^R (\eta_c+1)(\eta_c+3)}. \tag{B.4}$$

Note that $\alpha_{(0)}^R$ and $\alpha_{(0)}^I$ parametrize the initial values of the real and imaginary part of α_k in the large scale limit $k \to 0$, respectively. The parameter $\beta \gg 1$ we introduced is instead characterize the amplification that higher order terms in the gradient expansion of α_k obtains for a non-attractor era $\eta_c \leq -6$ with a duration of ΔN e-folds. As it is clear from (B.4), for $\Delta N \to 0$, $\beta \to \mathcal{O}(1)$, higher order terms in the gradient expansion becomes insignificant for providing an enhancement of the power spectrum at small
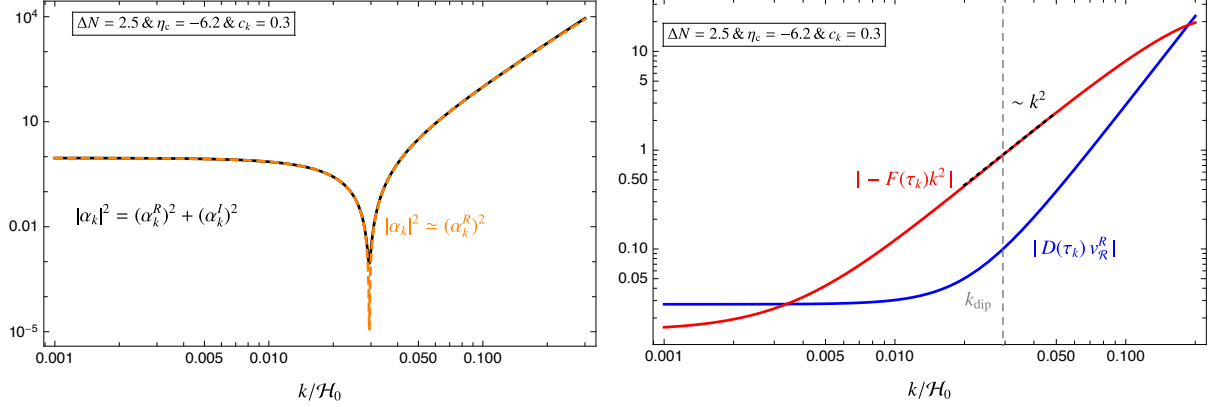
Figure 13: *The exact behavior of $|\alpha_k|^2$ (black solid) and the approximation $|\alpha_k|^2 \simeq (\alpha_k^R)^2$ (B.5) (orange dashed) from large to small scales associated with the initial slow-roll era, $k/\mathcal{H}_0 < c_k$ (Left). Comparison of the individual terms in $\alpha_k^R = 1 + D(\tau_k)v_{\mathcal{R}}^R - F(\tau_k)k^2$ (Right).*

scales $k/\mathcal{H}_0 \to 1$.

Finally, for the purpose of calculating scale dependent $f_{\mathrm{NL}}^{\mathrm{sq}}$ and $\tau_{\mathrm{NL}}^{\mathrm{col}}$ using the consistency conditions we derived in Section 3.2 and 3.3, we derive a simplified expression for the enhancement factor $|\alpha_k|^2$ in the gradient expansion formalism. As we discuss explicitly below, for modes associated with the initial era we can neglect the imaginary part of the enhancement factor $\alpha_k^I \ll \alpha_k^R$ and hence $|\alpha_k^R|^2 \simeq (\alpha_k^R)^2$. Furthermore, using (B.3), truncating the resulting $(\alpha_k^R)^2$ to fifth order in k provides a very accurate description to the exact result:

$$|\alpha_k|^2 \simeq (\alpha_{(0)}^R)^2 \left[1 - 2\beta \left(\frac{k}{\mathcal{H}_0} \right)^2 + \frac{2c_k\,(\eta_{\mathrm{c}}+1)}{(1+c_k^2)\,\eta_{\mathrm{c}}} \beta \left(\frac{k}{\mathcal{H}_0} \right)^3 + \beta^2 \left(\frac{k}{\mathcal{H}_0} \right)^4 - \frac{2c_k\,(\eta_{\mathrm{c}}+1)}{(1+c_k^2)\,\eta_{\mathrm{c}}} \beta^2 \left(\frac{k}{\mathcal{H}_0} \right)^5 \right],$$

$$\simeq (\alpha_k^R)^2. \tag{B.5}$$

Using the relation between scales (3.28), we can rewrite (B.5) in a compact way as

$$|\alpha_k|^2 \simeq (\alpha_k^R)^2 \simeq \sum_{n=0}^{5} c_{|\alpha|^2}^{(n)} \left(\frac{k}{k_{\mathrm{dip}}} \right)^n, \tag{B.6}$$

where in terms of the parametrization (B.4) above, the coefficients of the sum are given by

$$c_{|\alpha|^2}^{(0)} = (\alpha_{(0)}^R)^2, \quad c_{|\alpha|^2}^{(1)} = 0, \quad c_{|\alpha|^2}^{(2)} = -2c_{|\alpha|^2}^{(4)} = -2\,(\alpha_{(0)}^R)^2, \quad c_{|\alpha|^2}^{(3)} = -c_{|\alpha|^2}^{(5)} = \frac{2\,(\alpha_{(0)}^R)^2\,c_k\,(\eta_{\mathrm{c}}+1)}{(1+c_k^2)\,\eta_{\mathrm{c}}\sqrt{\beta}}. \tag{B.7}$$

Comparison between α_k^R and α_k^I. Another important point we will use repeatedly in the main text is the fact that imaginary part of α_k is subdominant compared to its real counterpart for modes associated with the initial slow-roll era. To see this, we can recall (3.4) and notice that $\alpha_k^I = D(\tau_k)v_{\mathcal{R}}^I \ll D(\tau_k)v_{\mathcal{R}}^R \subset \alpha_k^R$ which holds as long as we assume $c_k \ll 1$ implying $v_{\mathcal{R}}^R \gg v_{\mathcal{R}}^I$ as can be inferred from (A.2). We would like to remind that the choice $-k\tau_k = c_k \ll 1$ is a natural one considering that in the gradient expansion formalism $c_k \ll 1$ simply implies that all the k modes we focus are outside the horizon at the initial time τ_k. To explicitly check the statements above, we compare the exact $|\alpha_k|^2$ quantity with the approximate $|\alpha_k|^2 \simeq (\alpha_k^R)^2$ relation (B.5) in Figure 13. We observe that two expressions match really well for all scales except at the position of the dip k_{dip} where the approximate relation $|\alpha_k|^2 \simeq (\alpha_k^R)^2$ generically leads to a

more pronounced dip feature compared to the full expression $|\alpha_k|^2 = (\alpha_k^R)^2 + (\alpha_k^I)^2$.

In this work, except for the part where we study the global shape of the power spectrum (see Section 3.1), we will utilize the approximation $\alpha_k \simeq \alpha_k^R$ especially when we derive the scale dependence of bispectrum and trispectrum around the dip feature $k_{\rm dip}$ (See *e.g.* Section 3.2 and 3.3). We note however that whenever α_k appears in the denominator of an expression, we kept its full expression $\alpha_k = \alpha_k^R + i\alpha_k^I$ to avoid unphysical divergences that might appear around especially the dip scale.

α_k^R **around the dip scale.** In addition to the approximation we mentioned above, one can make further simplifications for the real part of the enhancement factor around the location of the dip feature of the power spectrum. In particular, an analysis of the individual terms that constitutes α_k^R (3.3) for $k/\mathcal{H}_0 < c_k$ reveals that $D(\tau_k)v_{\mathcal{R}}^R \ll F(\tau_k)k^2$ as can be also observed from the right panel of Figure 13. Therefore around the $k \sim k_{\rm dip}$, we can simply approximate α_k^R as

$$\alpha_k^R \simeq 1 - F(\tau_k)k^2. \tag{B.8}$$

The relation above automatically implies that $(\alpha_k^R)' = -(F(\tau_k)k^2)'$ where prime denotes a derivative with respect to the normalized wave-number $k/\mathcal{H}_0$. Recall that for the discussion that follows equation (3.31) in section 3.2, we require $-(k/\mathcal{H}_0)(\alpha_k^R)' = -(F(\tau_k)k^2)' \simeq 2F(\tau_k)k^2$ to prove that Maldacena's consistency condition $f_{\rm NL} = 5(1 - n_s)/12$ is satisfied around the dip feature. To show this explicitly, notice from the Figure 13 that $F(\tau_k)k^2$ in (B.1) can be truncated to second order in the k expansion around the dip feature. Keeping this in mind, using (B.1) up to second order in k, we can therefore write

$$-\frac{k}{\mathcal{H}_0}(\alpha_k^R)' = \frac{k}{\mathcal{H}_0}(F(\tau_k)k^2)' \simeq 2\left(F(\tau_k)k^2 - \frac{c_k^2}{6}\right) \simeq 2F(\tau_k)k^2, \tag{B.9}$$

where in the last equality we ignored the term proportional to $c_k^2 \ll 1$. Indeed, we utilize the approximation (B.9) (together with (B.8)) in the derivation of the general, scale dependent consistency condition between bispectrum and power spectrum (See section 3.2).

B.2 $\Pi(k)$ for mode exit during the initial slow-roll era

We now provide a simplified power law expression for the scale dependence of the enhancement factor $\Pi(\kappa)$ of Section 2 for modes that exit the horizon in the initial slow-roll stage $\kappa = k/k_{\rm na} < 1$. For this purpose, we find it sufficient to adopt a small κ expansion of the expression (2.11) up to k^6 order which yields

$$\Pi(k) \simeq 1 - \frac{2\beta_h}{3}\left(\frac{k}{k_{\rm na}}\right)^2 + \left[\frac{2\beta_h}{5} + \frac{\beta_h^2}{9}\right]\left(\frac{k}{k_{\rm na}}\right)^4 - \frac{\beta_h^2}{45}\left(\frac{k}{k_{\rm na}}\right)^6. \tag{B.10}$$

By explicitly comparing with (2.11), we found that (B.10) reproduces the exact scale dependence of the power spectrum very accurately for modes associated with the initial slow-roll era, $k/k_{\rm na} < 1$. Using the relation (2.15) between the scales, we can rewrite (B.10) in terms of the dip scale $k_{\rm dip}$ as

$$\Pi(k) \simeq \sum_{n=0}^{6} c_\Pi^{(n)}\left(\frac{k}{k_{\rm dip}}\right)^n, \tag{B.11}$$

where in terms of the only free parameter β_h of the heuristic approach the coefficients c_Π are given by

$$c_\Pi^{(0)} = 1, \quad c_\Pi^{(1)} = 0, \quad c_\Pi^{(2)} = -2, \quad c_\Pi^{(4)} = \left[1 + \frac{18}{5\beta_h}\right], \quad c_\Pi^{(6)} = -\frac{3}{5\beta_h}, \quad c_\Pi^{(3)} = c_\Pi^{(5)} = 0. \tag{B.12}$$

C   Squeezed limit $f_{\rm NL}^{\rm eff}$ and average μ distortions

Building upon our results in the previous appendix, we now would like to identify the effective non-linearity parameter defined in *e.g.* (4.6) for both approaches (heuristic vs gradient expansion) we focus in this work. Using the consistency condition $f_{\rm NL}^{\rm sq} = 5(1 - n_s)/12$ and noting the definition (4.6), we can describe the squeezed limit $(k_+ \to 0)$ $f_{\rm NL}^{\rm eff}$ in terms of small momenta q as

$$f_{\rm NL}^{\rm eff}(q) \simeq -\frac{5}{12} q \frac{{\rm d}E(q)}{{\rm d}q}, \tag{C.1}$$

where we denote the enhancement factors collectively as $E(q) = \{\Pi(q), |\alpha_q|^2\}$ for heuristic and gradient approaches respectively. Using the expressions (B.6) and (B.11) we can then describe scale dependent $f_{\rm NL}^{\rm eff}$ in terms of simple power law expansion. For both approaches we undertake, this power law expression can be expressed collectively as

$$f_{\rm NL}^{\rm eff}(q) = \frac{5}{3} \sum_{n=2}^{n_f} \tilde{c}_{\rm E}^{(n)} \left(\frac{q}{q_{\rm dip}}\right)^n, \tag{C.2}$$

where $n_f = \{5, 6\}$ for the gradient/heuristic approach respectively. In terms of the model parameters that characterize the both approaches, the coefficients of this expansion are given by

$$\tilde{c}_{|\alpha|^2}^{(2)} = -\tilde{c}_{|\alpha|^2}^{(4)} = (\alpha_{(0)}^R)^2, \;\; \tilde{c}_{|\alpha|^2}^{(3)} = -\frac{5}{3}\tilde{c}_{|\alpha|^2}^{(5)} = -\frac{3}{2}\frac{(\alpha_{(0)}^R)^2 c_k (\eta_{\rm c} + 1)}{(1 + c_k^2) \eta_{\rm c} \sqrt{\beta}}, \tag{C.3}$$

in the gradient expansion formalism and

$$\tilde{c}_\Pi^{(2)} = 1, \;\; \tilde{c}_\Pi^{(3)} = \tilde{c}_\Pi^{(5)} = 0, \;\; \tilde{c}_\Pi^{(4)} = -\left[1 + \frac{18}{5\beta_h}\right], \;\; \tilde{c}_\Pi^{(6)} = \frac{9}{10\beta_h}, \tag{C.4}$$

in the heuristic approach of Section 2. In Section 4.1 and 4.2, we utilize the formulas (C.2), (C.3) and (C.4) to calculate the amplitude of $\langle \mu T \rangle$ and $\langle \mu\mu \rangle$ angular correlators as a function of the location of the dip scale $q_{\rm dip}$ in the power spectrum.

Average $\langle \mu \rangle$. In the following discussion, we provide $\langle \mu \rangle$ induced due to the scale dependence of the power spectrum in PBH forming single-field inflationary models. For concreteness, we will utilize gradient expansion formalism to describe average $\langle \mu \rangle$ in the sky as a function of the scale $k_{\rm dip}$ beyond which most of the enhancement in the scalar power spectrum occurs. To compute average distortions we start with the following expression[9] [21],

$$\langle \mu \rangle \simeq 2.3 \int {\rm d} \ln k \, \mathcal{P}_{\mathcal{R}}(\tau_f, k) \left[e^{-2k^2/k_D^2(z)}\right]_{z_f}^{z_i}, \tag{C.5}$$

where $\mathcal{P}_{\mathcal{R}}(\tau_f, k) = |\alpha_k|^2 \mathcal{P}_{\mathcal{R}}(\tau_k)$ and assuming the scale invariance, $\mathcal{P}_{\mathcal{R}}(\tau_k) \simeq \mathcal{A}_s \simeq 2.1 \times 10^{-9}$. Using (B.6) and (B.7) in (C.5), we then estimate the amplitude of $\langle \mu \rangle$ in terms of the location of the dip scale $k_{\rm dip}$. For this purpose, we take the integral in (C.5) over the scales associated μ era, in particular from $k_{\rm in}/{\rm Mpc}^{-1} = 1$ to $k_{\rm fin}/{\rm Mpc}^{-1} = k_D(z_i) = 11600$. Results obtained in this way are presented in Figure 14 for two representative set of parameter choices that leads to $\Pi_{\rm tot} \simeq 10^7$ growth in the power spectrum. In this plot, the gray dotted lines indicate the threshold value of $k_{\rm dip}$ allowed by the current constraints $\langle \mu \rangle \lesssim 10^{-5}$ such that the right hand side of these vertical lines belongs to the allowed choices of the dip scale. In light of these results, we find it convenient to make the choice $k_{\rm dip} = 10^3 \, {\rm Mpc}^{-1}$ to derive predictions for the $\langle \mu T \rangle$ and $\langle \mu\mu \rangle$ correlators (see Sections 4.1 and 4.2).

We find it worth stressing that our estimates here can be regarded as rough indicator of average distortion

---

[9]More refined expressions that relates $\langle \mu \rangle$ to the primordial scalar power spectrum can be found in the earlier works [90, 91].
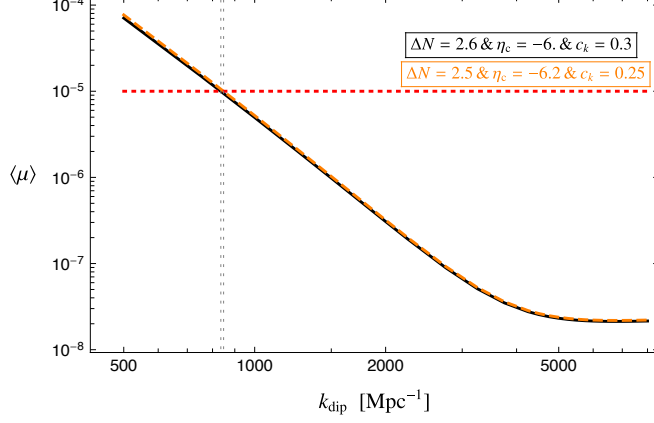
Figure 14: *Average μ distortion with respect to the location of the dip scale in the power spectrum within PBH forming inflationary scenarios.*

in PBH forming single field inflationary models. This is because in realistic models (see *e.g.* [17, 92]), the spectral shape and amplitude of the power spectrum prior to the dip scale can differ substantially compared to the simple assumption $\mathcal{P}_{\mathcal{R}}(\tau_k) \simeq \mathcal{A}_s \simeq 2.1 \times 10^{-9}$ we are undertaking here. As a mild indicator of such scenarios, by introducing a red tilt to $\mathcal{P}_{\mathcal{R}}(\tau_k)$ for scales prior to $k_{\rm dip}$ one can confirm that the overall amplitude of $\langle \mu \rangle$ becomes smaller than the estimates we present here.

D   Gaussian and non-Gaussian contributions to $C_l^{\mu\mu}$

In this appendix, we derive expressions for the Gaussian and non-Gaussian contribution to the $C_l^{\mu\mu}$ that arise through the connected and disconnected part of the trispectrum respectively. To calculate the distortion angular auto-correlations, we require

$$
\begin{aligned}
\left\langle a_{lm}^{\mu*} a_{l'm'}^{\mu} \right\rangle = (18.4\pi)^2 i^l (-i)^{l'} \int \frac{\mathrm{d}^3 k_1 \mathrm{d}^3 k_2 \mathrm{d}^3 k_3 \mathrm{d}^3 k_4}{(2\pi)^{12}} \, Y_{lm}(\hat{k}_+) \, Y_{l'm'}^*(\hat{k}'_+) \, W\left(\frac{k_+}{k_s}\right) W\left(\frac{k'_+}{k_s}\right) \\
\times \, j_l\left(k_+ \chi_*\right) j_{l'}\left(k'_+ \chi_*\right) \langle \cos\left(c_s k_1 \tau\right) \cos\left(c_s k_2 \tau\right) \rangle_p \, \langle \cos\left(c_s k_3 \tau\right) \cos\left(c_s k_4 \tau\right) \rangle_p \\
\times \left[e^{-(k_1^2 + k_2^2)/k_D^2(z)}\right]_f^i \left[e^{-(k_3^2 + k_4^2)/k_D^2(z)}\right]_f^i \langle \mathcal{R}_{k_1}(\tau_f) \mathcal{R}_{k_2}(\tau_f) \mathcal{R}_{k_3}^*(\tau_f) \mathcal{R}_{k_4}^*(\tau_f) \rangle, \quad (D.1)
\end{aligned}
$$

where we used (4.2) with (4.4) and defined $\vec{k}_+ = \vec{k}_1 + \vec{k}_2$, $\vec{k}'_+ = \vec{k}_3 + \vec{k}_4$. Assuming $\mathcal{R}_k$ is Gaussian at leading order (see *e.g.* (3.10)), we can generically split the 4-pt correlator in (D.1) into its connected (non-Gaussian) and disconnected (Gaussian) parts as

$$
\langle \mathcal{R}_{k_1} \mathcal{R}_{k_2} \mathcal{R}_{k_3}^* \mathcal{R}_{k_4}^* \rangle = \langle \mathcal{R}_{k_1}^{\rm G}(\tau_f) \mathcal{R}_{k_2}^{\rm G}(\tau_f) \mathcal{R}_{k_3}^{\rm G*}(\tau_f) \mathcal{R}_{k_4}^{\rm G*}(\tau_f) \rangle + \langle \mathcal{R}_{k_1}(\tau_f) \mathcal{R}_{k_2}(\tau_f) \mathcal{R}_{k_3}^*(\tau_f) \mathcal{R}_{k_4}^*(\tau_f) \rangle_{\rm c}. \quad (D.2)
$$

Taking into account all possible pair contractions and using (3.8), the disconnected part of the 4-pt correlator (D.2) is given by

$$
\begin{aligned}
\langle \mathcal{R}_{k_1}^{\rm G} \mathcal{R}_{k_2}^{\rm G} \mathcal{R}_{k_3}^{\rm G*} \mathcal{R}_{k_4}^{\rm G*} \rangle = (2\pi)^6 \Bigg\{ & \delta(\vec{k}_+) \, \delta(\vec{k}'_+) \, P_{\mathcal{R}}(\tau_f, k_1) \, P_{\mathcal{R}}(\tau_f, k_3) \\
& + P_{\mathcal{R}}(\tau_f, k_1) P_{\mathcal{R}}(\tau_f, k_2) \left[\delta(\vec{k}_1 - \vec{k}_4)\delta(\vec{k}_2 - \vec{k}_3) + \delta(\vec{k}_1 - \vec{k}_3)\delta(\vec{k}_2 - \vec{k}_4) \right] \Bigg\}.
\end{aligned}
\quad (\text{D.3})
$$

33

On the other hand, the connected part of (D.2) can be defined as

$$\langle \mathcal{R}_{k_1}(\tau_f)\mathcal{R}_{k_2}(\tau_f)\mathcal{R}^*_{k_3}(\tau_f)\mathcal{R}^*_{k_4}(\tau_f)\rangle_{\rm c} = (2\pi)^3\delta\left(\vec{k}_+ - \vec{k}'_+\right)T_{\mathcal{R}}\left(k_1,k_2,k_3,k_4\right), \qquad (\text{D.4})$$

where $T_{\mathcal{R}}$ is the trispectrum. Noting (D.3) and (D.4), we can then make use of (D.2) in (D.1) to separate the angular distortion auto-correlator into Gaussian and non-Gaussian part as $C_l^{\mu\mu} = C_{l,\rm G}^{\mu\mu} + C_{l,\rm NG}^{\mu\mu}$.

Gaussian $C_l^{\mu\mu}$. Inserting (D.3) in (D.1), notice that the contribution associated with the first term in (D.3) vanishes unless the index of the spherical Bessel functions are zero, *i.e.* $l = l' = 0$ ($m = m' = 0$). Therefore, after carrying the integrals over $\rm d^3k_2$ and $\rm d^3k_4$, the contribution of the first term is given by

$$\langle a^{\mu,*}_{lm}a^{\mu}_{l'm'}\rangle_{\rm G,1} = 4\pi\,\delta_{ll'}\,\delta_{mm'}\delta_{l0}\,(2.3)^2\int\frac{\rm d^3k_1 d^3k_3}{(2\pi)^6}P_{\mathcal{R}}(\tau_f,k_1)P_{\mathcal{R}}(\tau_f,k_3)\,e^{-2k_1^2/k_D^2(z)}\Big|_{z_f}^{z_i}\,e^{-2k_1^2/k_D^2(z)}\Big|_{z_f}^{z_i},$$

$$= 4\pi\,\delta_{ll'}\,\delta_{mm'}\,\delta_{l0}\,\langle\mu\rangle^2, \qquad (\text{D.5})$$

where the average distortion is given by (C.5). (D.5) describes the average $\langle\mu\mu\rangle$ in the sky (*i.e.* a monopole $l = 0$) and should be subtracted from the total $\langle\mu\mu\rangle$. Here our focus is on the non-trivial anisotropies with $l \neq 0$ that originates from the second term in (D.3). Taking the integrals over $\rm d^3k_3$ and $\rm d^3k_4$ in (D.1), this contribution to the Gaussian $C_{l,\rm G}^{\mu\mu}$ reads as

$$\langle a^{\mu,*}_{lm}a^{\mu}_{l'm'}\rangle_{\rm G,2} = (18.4\pi)^2 i^l(-i)^{l'}\int\frac{\rm d^3k_1 d^3k_2}{(2\pi)^6}\,Y_{lm}(\hat{k}_+)Y^*_{l'm'}(\hat{k}_+)\,W\left(\frac{k_+}{k_s}\right)^2 j_l\left(k_+\chi_*\right)j_{l'}\left(k_+\chi_*\right)$$

$$\times\,\langle\cos\left(c_s k_1\tau\right)\cos\left(c_s k_2\tau\right)\rangle_p^2\,\left(\left[e^{-(k_1^2+k_2^2)/k_D^2(z)}\right]_{z_f}^{z_i}\right)^2$$

$$\times\left[\frac{2\pi^2}{k_1^3}\mathcal{P}_{\mathcal{R}}(\tau_f,k_1)\right]\left[\frac{2\pi^2}{k_2^3}\mathcal{P}_{\mathcal{R}}(\tau_f,k_2)\right]. \qquad (\text{D.6})$$

We then make the transformation $\rm d^3k_2 \to d^3k_+$ and carry out the integral over directions $\rm d^2\hat{k}_+$. This gives

$$\langle a^{\mu,*}_{lm}a^{\mu}_{l'm'}\rangle_{\rm G,2} = (4.6)^2\delta_{ll'}\delta_{mm'}\int {\rm d}k_+\,k_+^2\,j_l(k_+\chi_*)^2\,W\left(\frac{k_+}{k_s}\right)^2\int\frac{\rm d^3k_1}{k_1^3|\vec{k}_+ - \vec{k}_1|^3} \qquad (\text{D.7})$$

$$\times\,\langle\cos\left(c_s k_1\tau\right)\cos\left(c_s|\vec{k}_+ - \vec{k}_1|\tau\right)\rangle_p^2\,\mathcal{P}_{\mathcal{R}}(\tau_f,k_1)\,\mathcal{P}_{\mathcal{R}}(\tau_f,|\vec{k}_+ - \vec{k}_1|)$$

$$\times\left(\left[e^{-(k_1^2+|\vec{k}_+-\vec{k}_1|^2)/k_D^2(z)}\right]_{z_f}^{z_i}\right)^2.$$

Furthermore, making the replacement $\vec{k}_3 \leftrightarrow \vec{k}_4$ in the last term of (D.3) and noticing the fact the rest of the integrand in (D.1) is symmetric under $\vec{k}_3 \leftrightarrow \vec{k}_4$, we have $\langle a^{\mu,*}_{lm}a^{\mu}_{l'm'}\rangle_{\rm G,2} = \langle a^{\mu,*}_{lm}a^{\mu}_{l'm'}\rangle_{\rm G,3}$. Therefore, the non-trivial part ($l \neq 0$) part of the Gaussian contribution to $C_l^{\mu\mu}$ is given by

$$C_{l,\rm G}^{\mu\mu} = 2\,(4.6)^2\int {\rm d}k_+\,k_+^2\,j_l(k_+\chi_*)^2\,W\left(\frac{k_+}{k_s}\right)^2\int\frac{\rm d^3k_1}{k_1^3|\vec{k}_+ - \vec{k}_1|^3}\mathcal{P}_{\mathcal{R}}(\tau_f,k_1)\,\mathcal{P}_{\mathcal{R}}(\tau_f,|\vec{k}_+ - \vec{k}_1|)$$

$$\times\,\langle\cos\left(c_s k_1\tau\right)\cos\left(c_s|\vec{k}_+ - \vec{k}_1|\tau\right)\rangle_p^2\left(\left[e^{-(k_1^2+|\vec{k}_+-\vec{k}_1|^2)/k_D^2(z)}\right]_{z_f}^{z_i}\right)^2. \qquad (\text{D.8})$$

non-Gaussian $C_l^{\mu\mu}$. To derive the non-Gaussian contribution to $\langle\mu\mu\rangle$, we insert (D.4) in (D.1). In particular, transforming $\rm d^3k_2 \to d^3k_+$ and $\rm d^3k_4 \to d^3k'_+$ we first take the integral over $\rm d^3k'_+$ using (D.4)

$(\vec{k}_+ = \vec{k}'_+)$ to obtain

$$
\left\langle a_{lm}^{\mu^*} a_{l'm'}^{\mu} \right\rangle_{\mathrm{NG}} = (18.4\pi)^2 i^l (-i)^{l'} \int \frac{\mathrm{d}^3 k_+ \mathrm{d}^3 k_1 \mathrm{d}^3 k_3}{(2\pi)^9} \, Y_{lm}(\hat{k}_+) \, Y_{l'm'}^*(\hat{k}_+) \, W\left(\frac{k_+}{k_s}\right)^2 \, j_l\left(k_+\chi_*\right) j_{l'}\left(k_+\chi_*\right)
$$
$$
\times \langle \cos\left(c_s k_1 \tau\right) \cos\left(c_s |\vec{k}_+ - \vec{k}_1| \tau\right) \rangle_p \, \langle \cos\left(c_s k_3 \tau\right) \cos\left(c_s |\vec{k}_+ - \vec{k}_3| \tau\right) \rangle_p
$$
$$
\times \left[e^{-(k_1^2 + |\vec{k}_+ - \vec{k}_1|^2)/k_D^2(z)} \right]_{z_f}^{z_i} \left[e^{-(k_3^2 + |\vec{k}_+ - \vec{k}_3|^2)/k_D^2(z)} \right]_{z_f}^{z_i} T_{\mathcal{R}}\left(k_1, |\vec{k}_+ - \vec{k}_1|, k_3, |\vec{k}_+ - \vec{k}_3|\right).
$$
$$(D.9)$$

In (D.9), the presence of integration over $\mathrm{d}^2 \hat{k}_1$ and $\mathrm{d}^2 \hat{k}_3$ ensures the independence of the integrand on $\hat{k}_+$ except in the arguments of Y_{lm}'s [66]. We can therefore take the integral over $\mathrm{d}^2 \hat{k}_+$ using $\int \mathrm{d}^2 \hat{k}_+ Y_{lm}(\hat{k}_+) Y_{l'm'}^*(\hat{k}_+) = \delta_{ll'} \delta_{mm'}$ to yield

$$
C_{l,\mathrm{NG}}^{\mu\mu} \simeq \frac{(2.3)^2}{8\pi^7} \int \mathrm{d}k_+ \, k_+^2 \, j_l^2\left(k_+\chi_*\right) W\left(\frac{k_+}{k_s}\right)^2 \int \mathrm{d}^3 k_1 \mathrm{d}^3 k_3 \, T_{\mathcal{R}}\left(k_1, |\vec{k}_+ - \vec{k}_1|, k_3, |\vec{k}_+ - \vec{k}_3|\right)
$$
$$
\times \langle \cos\left(c_s k_1 \tau\right) \cos\left(c_s |\vec{k}_+ - \vec{k}_1| \tau\right) \rangle_p \, \langle \cos\left(c_s k_3 \tau\right) \cos\left(c_s |\vec{k}_+ - \vec{k}_3| \tau\right) \rangle_p
$$
$$
\times \left[e^{-(k_1^2 + |\vec{k}_+ - \vec{k}_1|^2)/k_D^2(z)} \right]_{z_f}^{z_i} \left[e^{-(k_3^2 + |\vec{k}_+ - \vec{k}_3|^2)/k_D^2(z)} \right]_{z_f}^{z_i},
$$
$$(D.10)$$

where we extracted $C_l^{\mu\mu}$ from (D.10) using the definition (4.4).

D.1 $C_{l,\mathrm{G}}^{\mu\mu}$ in the collapsed $k_+ \to 0$ limit.

We now provide an estimate for the Gaussian $\mu\mu$ correlator (D.8) in the collapsed limit. Taking $k_+ \to 0$ in (D.8) we have

$$
C_{l,\mathrm{G}}^{\mu\mu} \simeq 5.9 \times 10^{-16} \int \mathrm{d}k_+ \, k_+^2 \, j_l(k_+\chi_*)^2 \int \frac{\mathrm{d}k_1}{k_1^4} E(k_1)^2 \left(\left[e^{-2k_1^2/k_D^2(z)} \right]_{z_f}^{z_i} \right)^2,
$$
$$(D.11)$$

where we used $\mathcal{P}_{\mathcal{R}}(\tau_f, k_1) \simeq E(k_1)\mathcal{A}_s$ with $\mathcal{A}_s = 2.1 \times 10^{-9}$ and $E(k_1) = \{\Pi(k_1), |\alpha_{k_1}|^2\}$ for heuristic and gradient approaches respectively. Using the dimensionless variables $\tilde{k} = k_1/k_{\mathrm{dip}}$ and $x = k_+\chi_*$, we can rewrite (D.11) as

$$
C_{l,\mathrm{G}}^{\mu\mu} \simeq 5.9 \times 10^{-16} \, \mathcal{I}\left[k_{\mathrm{dip}}\right] \int \mathrm{d}x \, x^2 \, j_l(x)^2,
$$
$$(D.12)$$

with the integral that depends on the location of k_{dip} (and model parameters) defined as

$$
\mathcal{I}\left[k_{\mathrm{dip}}\right] = \left(\frac{k_*}{k_{\mathrm{dip}}}\right)^3 \int \frac{\mathrm{d}\tilde{k}}{\tilde{k}^4} E(\tilde{k})^2 \left(\left[e^{-2\tilde{k}^2/\tilde{k}_D^2(z)} \right]_{z_f}^{z_i} \right)^2,
$$
$$(D.13)$$

where we defined $k_* = \chi_*^{-1} \simeq 1/(14\,\mathrm{Gpc}) \simeq 7.1 \times 10^{-5}\,\mathrm{Mpc}^{-1}$ and $\tilde{k}_D(z) = k_D(z)/k_{\mathrm{dip}}$. Notice that $(k_*/k_{\mathrm{dip}})^3$ factor in (D.13) introduces a large suppression factor to the amplitude of $\mathcal{I}$ for phenomenologically interesting values around $k_{\mathrm{dip}} \simeq 10^3$. To illustrate this, in Figure 15, we plot $\mathcal{I}$ as a function of k_{dip} for representative parameter choices within both heuristic and gradient approach. On the other hand, for large enough multipoles l, $j_l^2(x)$ is highly peaked around $x \approx l$ and acts like a delta function in the last integral in (D.12) and therefore we can approximate $\int \mathrm{d}x \, x^2 j_l^2(x) \approx l^2$. As a result, we anticipate that the Gaussian contribution can be approximated as for $l > 0$ as

$$
C_{l,\mathrm{G}}^{\mu\mu} \approx 10^{-32} \, l^2.
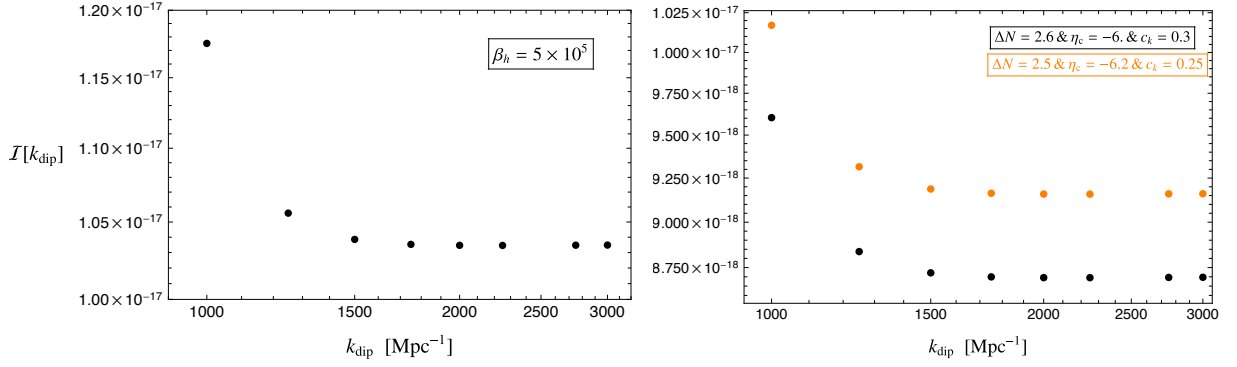$$
$$(D.14)$$

Figure 15: *The integral $\mathcal{I}$ (D.13) for a grid of phenomenologically interesting k_{dip} values using heuristic approach (Left) and the gradient expansion formalism (Right).*

For scales we can probe with μ distortions, we have $l_{\mathrm{max}} = 200$, this result above (D.14) gives $(C_{l,\mathrm{G}}^{\mu\mu})_{\mathrm{max}} \approx 10^{-28}$ and therefore we can safely conclude that in the collapsed limit, $C_{l,\mathrm{G}}^{\mu\mu} \ll C_{l,\mathrm{NG}}^{\mu\mu}$ holds for scales we would like to probe μ distortion anisotropies (See section 4.2).

References

[1] S. Hawking, "Gravitationally collapsed objects of very low mass," *Mon. Not. Roy. Astron. Soc.* **152** (1971) 75.

[2] B. J. Carr and S. W. Hawking, "Black holes in the early Universe," *Mon. Not. Roy. Astron. Soc.* **168** (1974) 399–415.

[3] B. J. Carr, "The Primordial black hole mass spectrum," *Astrophys. J.* **201** (1975) 1–19.

[4] P. Ivanov, P. Naselsky, and I. Novikov, "Inflation and primordial black holes as dark matter," *Phys. Rev.* **D50** (1994) 7173–7178.

[5] J. Garcia-Bellido, A. D. Linde, and D. Wands, "Density perturbations and black hole formation in hybrid inflation," *Phys. Rev.* **D54** (1996) 6040–6058, arXiv:astro-ph/9605094 [astro-ph].

[6] B. Carr, F. Kuhnel, and M. Sandstad, "Primordial Black Holes as Dark Matter," *Phys. Rev.* **D94** no. 8, (2016) 083504, arXiv:1607.06077 [astro-ph.CO].

[7] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, "Primordial black holes—perspectives in gravitational wave astronomy," *Class. Quant. Grav.* **35** no. 6, (2018) 063001, arXiv:1801.05235 [astro-ph.CO].

[8] B. Carr and F. Kuhnel, "Primordial Black Holes as Dark Matter: Recent Developments," *Ann. Rev. Nucl. Part. Sci.* **70** (2020) 355–394, arXiv:2006.02838 [astro-ph.CO].

[9] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, "Constraints on Primordial Black Holes," arXiv:2002.12778 [astro-ph.CO].

[10] J. Garcia-Bellido and E. Ruiz Morales, "Primordial black holes from single field models of inflation," *Phys. Dark Univ.* **18** (2017) 47–54, arXiv:1702.03901 [astro-ph.CO].

[11] C. Germani and T. Prokopec, "On primordial black holes from an inflection point," *Phys. Dark Univ.* **18** (2017) 6–10, arXiv:1706.04226 [astro-ph.CO].

[12] J. M. Ezquiaga, J. Garcia-Bellido, and E. Ruiz Morales, "Primordial Black Hole production in Critical Higgs Inflation," *Phys. Lett.* **B776** (2018) 345–349, arXiv:1705.04861 [astro-ph.CO].

[13] H. Motohashi and W. Hu, "Primordial Black Holes and Slow-Roll Violation," *Phys. Rev.* **D96** no. 6, (2017) 063503, arXiv:1706.06784 [astro-ph.CO].

[14] G. Ballesteros and M. Taoso, "Primordial black hole dark matter from single field inflation," *Phys. Rev.* **D97** no. 2, (2018) 023501, arXiv:1709.05565 [hep-ph].

[15] M. P. Hertzberg and M. Yamada, "Primordial Black Holes from Polynomial Potentials in Single Field Inflation," *Phys. Rev.* **D97** no. 8, (2018) 083509, arXiv:1712.09750 [astro-ph.CO].

[16] M. Cicoli, V. A. Diaz, and F. G. Pedro, "Primordial Black Holes from String Inflation," *JCAP* **1806** (2018) 034, arXiv:1803.02837 [hep-th].

[17] O. Ozsoy, S. Parameswaran, G. Tasinato, and I. Zavala, "Mechanisms for Primordial Black Hole Production in String Theory," *JCAP* **1807** (2018) 005, arXiv:1803.07626 [hep-th].

[18] C. T. Byrnes, P. S. Cole, and S. P. Patil, "Steepest growth of the power spectrum and primordial black holes," *JCAP* **1906** no. 06, (2019) 028, arXiv:1811.11158 [astro-ph.CO].

[19] G. Tasinato, "An analytic approach to non-slow-roll inflation," *Phys. Rev. D* **103** no. 2, (2021) 023535, arXiv:2012.02518 [hep-th].

[20] O. Özsoy and G. Tasinato, "CMB μT cross-correlations as a probe of PBH scenarios," arXiv:2104.12792 [astro-ph.CO].

[21] E. Pajer and M. Zaldarriaga, "A New Window on Primordial non-Gaussianity," *Phys. Rev. Lett.* **109** (2012) 021302, arXiv:1201.5375 [astro-ph.CO].

[22] S. M. Leach, M. Sasaki, D. Wands, and A. R. Liddle, "Enhancement of superhorizon scale inflationary curvature perturbations," *Phys. Rev.* **D64** (2001) 023512, arXiv:astro-ph/0101406 [astro-ph].

[23] R. Mahbub, "Primordial black hole formation in inflationary α-attractor models," *Phys. Rev. D* **101** no. 2, (2020) 023533, arXiv:1910.10602 [astro-ph.CO].

[24] G. Ballesteros, J. Rey, M. Taoso, and A. Urbano, "Primordial black holes as dark matter and gravitational waves from single-field polynomial inflation," *JCAP* **07** (2020) 025, arXiv:2001.08220 [astro-ph.CO].

[25] J. Liu, Z.-K. Guo, and R.-G. Cai, "Analytical approximation of the scalar spectrum in the ultraslow-roll inflationary models," *Phys. Rev. D* **101** no. 8, (2020) 083535, arXiv:2003.02075 [astro-ph.CO].

[26] K. Kefala, G. P. Kodaxis, I. D. Stamou, and N. Tetradis, "Features of the inflaton potential and the power spectrum of cosmological perturbations," arXiv:2010.12483 [astro-ph.CO].

[27] K.-W. Ng and Y.-P. Wu, "Constant-rate inflation: primordial black holes from conformal weight transitions," arXiv:2102.05620 [astro-ph.CO].

[28] K. Inomata, E. Mcdonough, and W. Hu, "Primordial Black Holes Arise When The Inflaton Falls," arXiv:2104.03972 [astro-ph.CO].

[29] K. Inomata, E. McDonough, and W. Hu, "Amplification of Primordial Perturbations from the Rise or Fall of the Inflaton," `arXiv:2110.14641 [astro-ph.CO]`.

[30] P. Carrilho, K. A. Malik, and D. J. Mulryne, "Dissecting the growth of the power spectrum for primordial black holes," *Phys. Rev. D* **100** no. 10, (2019) 103529, `arXiv:1907.05237 [astro-ph.CO]`.

[31] O. Özsoy and G. Tasinato, "On the slope of the curvature power spectrum in non-attractor inflation," *JCAP* **04** (2020) 048, `arXiv:1912.01061 [astro-ph.CO]`.

[32] M. W. Davies, P. Carrilho, and D. J. Mulryne, "Non-Gaussianity in inflationary scenarios for primordial black holes," `arXiv:2110.08189 [astro-ph.CO]`.

[33] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *JHEP* **05** (2003) 013, `arXiv:astro-ph/0210603`.

[34] P. Creminelli and M. Zaldarriaga, "Single field consistency relation for the 3-point function," *JCAP* **0410** (2004) 006, `arXiv:astro-ph/0407059 [astro-ph]`.

[35] C. T. Byrnes, M. Sasaki, and D. Wands, "The primordial trispectrum from inflation," *Phys. Rev. D* **74** (2006) 123519, `arXiv:astro-ph/0611075`.

[36] T. Suyama and M. Yamaguchi, "Non-Gaussianity in the modulated reheating scenario," *Phys. Rev. D* **77** (2008) 023505, `arXiv:0709.2545 [astro-ph]`.

[37] K. M. Smith, M. LoVerde, and M. Zaldarriaga, "A universal bound on N-point correlations from inflation," *Phys. Rev. Lett.* **107** (2011) 191301, `arXiv:1108.1805 [astro-ph.CO]`.

[38] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," *JCAP* **11** (2012) 047, `arXiv:1204.4207 [hep-th]`.

[39] E. Pajer, F. Schmidt, and M. Zaldarriaga, "The Observed Squeezed Limit of Cosmological Three-Point Functions," *Phys. Rev. D* **88** no. 8, (2013) 083502, `arXiv:1305.0824 [astro-ph.CO]`.

[40] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, "Violation of non-Gaussianity consistency relation in a single field inflationary model," *EPL* **101** no. 3, (2013) 39001, `arXiv:1210.3692 [astro-ph.CO]`.

[41] J. Martin, H. Motohashi, and T. Suyama, "Ultra Slow-Roll Inflation and the non-Gaussianity Consistency Relation," *Phys. Rev. D* **87** no. 2, (2013) 023514, `arXiv:1211.0083 [astro-ph.CO]`.

[42] X. Chen, H. Firouzjahi, M. H. Namjoo, and M. Sasaki, "A Single Field Inflation Model with Large Local Non-Gaussianity," *EPL* **102** no. 5, (2013) 59001, `arXiv:1301.5699 [hep-th]`.

[43] X. Chen, H. Firouzjahi, E. Komatsu, M. H. Namjoo, and M. Sasaki, "In-in and δN calculations of the bispectrum from non-attractor single-field inflation," *JCAP* **12** (2013) 039, `arXiv:1308.5341 [astro-ph.CO]`.

[44] B. Finelli, G. Goon, E. Pajer, and L. Santoni, "Soft Theorems For Shift-Symmetric Cosmologies," *Phys. Rev. D* **97** no. 6, (2018) 063531, `arXiv:1711.03737 [hep-th]`.

[45] Y.-F. Cai, X. Chen, M. H. Namjoo, M. Sasaki, D.-G. Wang, and Z. Wang, "Revisiting non-Gaussianity from non-attractor inflation models," *JCAP* **05** (2018) 012, `arXiv:1712.09998 [astro-ph.CO]`.

[46] T. Suyama, Y. Tada, and M. Yamaguchi, "Revisiting non-Gaussianity in non-attractor inflation models in the light of the cosmological soft theorem," *PTEP* **2021** no. 7, (2021) 073E02, `arXiv:2101.10682 [hep-th]`.

[47] Y.-i. Takamizu, S. Mukohyama, M. Sasaki, and Y. Tanaka, "Non-Gaussianity of superhorizon curvature perturbations beyond δ N formalism," *JCAP* **1006** (2010) 019, `arXiv:1004.1870 [astro-ph.CO]`.

[48] Y.-i. Takamizu, "Application of beyond δN formalism: Varying sound speed," *Phys. Rev. D* **89** no. 4, (2014) 043528, `arXiv:1311.6986 [astro-ph.CO]`.

[49] M. Biagetti, G. Franciolini, A. Kehagias, and A. Riotto, "Primordial Black Holes from Inflation and Quantum Diffusion," *JCAP* **1807** no. 07, (2018) 032, `arXiv:1804.07124 [astro-ph.CO]`.

[50] J. Chluba, J. Hamann, and S. P. Patil, "Features and New Physical Scales in Primordial Observables: Theory and Observation," *Int. J. Mod. Phys.* **D24** no. 10, (2015) 1530023, `arXiv:1505.01834 [astro-ph.CO]`.

[51] X. Chen, "Running non-Gaussianities in DBI inflation," *Phys. Rev.* **D72** (2005) 123518, `arXiv:astro-ph/0507053 [astro-ph]`.

[52] C. T. Byrnes, S. Nurmi, G. Tasinato, and D. Wands, "Scale dependence of local fNL," *JCAP* **1002** (2010) 034, `arXiv:0911.2780 [astro-ph.CO]`.

[53] C. T. Byrnes, M. Gerstenlauer, S. Nurmi, G. Tasinato, and D. Wands, "Scale-dependent non-Gaussianity probes inflationary physics," *JCAP* **1010** (2010) 004, `arXiv:1007.4277 [astro-ph.CO]`.

[54] S. Passaglia, W. Hu, and H. Motohashi, "Primordial black holes and local non-Gaussianity in canonical inflation," *Phys. Rev. D* **99** no. 4, (2019) 043536, `arXiv:1812.08243 [astro-ph.CO]`.

[55] Q. Gao, "Primordial black holes and secondary gravitational waves from chaotic inflation," `arXiv:2102.07369 [gr-qc]`.

[56] H. V. Ragavendra, P. Saha, L. Sriramkumar, and J. Silk, "Primordial black holes and secondary gravitational waves from ultraslow roll and punctuated inflation," *Phys. Rev. D* **103** no. 8, (2021) 083510, `arXiv:2008.12202 [astro-ph.CO]`.

[57] V. Sreenath, D. K. Hazra, and L. Sriramkumar, "On the scalar consistency relation away from slow roll," *JCAP* **02** (2015) 029, `arXiv:1410.0252 [astro-ph.CO]`.

[58] W. Hu, D. Scott, and J. Silk, "Power spectrum constraints from spectral distortions in the cosmic microwave background," *Astrophys. J. Lett.* **430** (1994) L5–L8, `arXiv:astro-ph/9402045`.

[59] R. Khatri, R. A. Sunyaev, and J. Chluba, "Does Bose-Einstein condensation of CMB photons cancel \mu distortions created by dissipation of sound waves in the early Universe?," *Astron. Astrophys.* **540** (2012) A124, `arXiv:1110.0475 [astro-ph.CO]`.

[60] J. Chluba and R. A. Sunyaev, "The evolution of CMB spectral distortions in the early Universe," *Mon. Not. Roy. Astron. Soc.* **419** (2012) 1294–1314, `arXiv:1109.6552 [astro-ph.CO]`.

[61] J. Chluba *et al.*, "Spectral Distortions of the CMB as a Probe of Inflation, Recombination, Structure Formation and Particle Physics: Astro2020 Science White Paper," *Bull. Am. Astron. Soc.* **51** no. 3, (2019) 184, `arXiv:1903.04218 [astro-ph.CO]`.

[62] J. García-Bellido, "Massive Primordial Black Holes as Dark Matter and their detection with Gravitational Waves," *J. Phys. Conf. Ser.* **840** no. 1, (2017) 012032, `arXiv:1702.08275 [astro-ph.CO]`.

[63] C. Ünal, E. D. Kovetz, and S. P. Patil, "Multimessenger probes of inflationary fluctuations and primordial black holes," *Phys. Rev. D* **103** no. 6, (2021) 063519, `arXiv:2008.11184 [astro-ph.CO]`.

[64] J. Ganc and E. Komatsu, "Scale-dependent bias of galaxies and mu-type distortion of the cosmic microwave background spectrum from single-field inflation with a modified initial state," *Phys. Rev. D* **86** (2012) 023518, `arXiv:1204.4241 [astro-ph.CO]`.

[65] R. Emami, E. Dimastrogiovanni, J. Chluba, and M. Kamionkowski, "Probing the scale dependence of non-Gaussianity with spectral distortions of the cosmic microwave background," *Phys. Rev. D* **91** no. 12, (2015) 123531, `arXiv:1504.00675 [astro-ph.CO]`.

[66] J. Chluba, E. Dimastrogiovanni, M. A. Amin, and M. Kamionkowski, "Evolution of CMB spectral distortion anisotropies and tests of primordial non-Gaussianity," *Mon. Not. Roy. Astron. Soc.* **466** no. 2, (2017) 2390–2401, `arXiv:1610.08711 [astro-ph.CO]`.

[67] E. Dimastrogiovanni and R. Emami, "Correlating CMB Spectral Distortions with Temperature: what do we learn on Inflation?," *JCAP* **12** (2016) 015, `arXiv:1606.04286 [astro-ph.CO]`.

[68] A. Ravenni, M. Liguori, N. Bartolo, and M. Shiraishi, "Primordial non-Gaussianity with μ-type and y-type spectral distortions: exploiting Cosmic Microwave Background polarization and dealing with secondary sources," *JCAP* **09** (2017) 042, `arXiv:1707.04759 [astro-ph.CO]`.

[69] G. Orlando, P. D. Meerburg, and S. P. Patil, "Primordial tensor bispectra in μ-CMB cross-correlations," `arXiv:2109.01095 [astro-ph.CO]`.

[70] M. Remazeilles, A. Ravenni, and J. Chluba, "Leverage on small-scale primordial non-Gaussianity through cross-correlations between CMB E-mode and μ-distortion anisotropies," `arXiv:2110.14664 [astro-ph.CO]`.

[71] A. Ota, "Cosmological constraints from μE cross-correlations," *Phys. Rev. D* **94** no. 10, (2016) 103520, `arXiv:1607.00212 [astro-ph.CO]`.

[72] J. C. Mather *et al.*, "Measurement of the Cosmic Microwave Background spectrum by the COBE FIRAS instrument," *Astrophys. J.* **420** (1994) 439–444.

[73] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, "The Cosmic Microwave Background spectrum from the full COBE FIRAS data set," *Astrophys. J.* **473** (1996) 576, `arXiv:astro-ph/9605054`.

[74] T. Okamoto and W. Hu, "The angular trispectra of CMB temperature and polarization," *Phys. Rev. D* **66** (2002) 063008, `arXiv:astro-ph/0206155`.

[75] L. Boubekeur and D. H. Lyth, "Detecting a small perturbation through its non-Gaussianity," *Phys. Rev. D* **73** (2006) 021301, `arXiv:astro-ph/0504046`.

[76] M. Zaldarriaga and U. Seljak, "An all sky analysis of polarization in the microwave background," *Phys. Rev. D* **55** (1997) 1830–1840, `arXiv:astro-ph/9609170`.

[77] A. Kogut *et al.*, "The Primordial Inflation Explorer (PIXIE): A Nulling Polarimeter for Cosmic Microwave Background Observations," *JCAP* **07** (2011) 025, `arXiv:1105.2044 [astro-ph.CO]`.

[78] J. Chluba and D. Jeong, "Teasing bits of information out of the CMB energy spectrum," *Mon. Not. Roy. Astron. Soc.* **438** no. 3, (2014) 2065–2082, `arXiv:1306.5751 [astro-ph.CO]`.

[79] **PRISM** Collaboration, P. André *et al.*, "PRISM (Polarized Radiation Imaging and Spectroscopy Mission): An Extended White Paper," *JCAP* **02** (2014) 006, `arXiv:1310.1554 [astro-ph.CO]`.

[80] B. J. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, "New cosmological constraints on primordial black holes," *Phys. Rev. D* **81** (2010) 104019, `arXiv:0912.5297 [astro-ph.CO]`.

[81] C. Germani and I. Musco, "Abundance of Primordial Black Holes Depends on the Shape of the Inflationary Power Spectrum," *Phys. Rev. Lett.* **122** no. 14, (2019) 141302, `arXiv:1805.04087 [astro-ph.CO]`.

[82] J. Garcia-Bellido, M. Peloso, and C. Unal, "Gravitational Wave signatures of inflationary models from Primordial Black Hole Dark Matter," *JCAP* **09** (2017) 013, `arXiv:1707.02441 [astro-ph.CO]`.

[83] E. Bugaev and P. Klimai, "Axion inflation with gauge field production and primordial black holes," *Phys. Rev. D* **90** no. 10, (2014) 103501, `arXiv:1312.7435 [astro-ph.CO]`.

[84] J. Garcia-Bellido, M. Peloso, and C. Unal, "Gravitational waves at interferometer scales and primordial black holes in axion inflation," *JCAP* **1612** no. 12, (2016) 031, `arXiv:1610.03763 [astro-ph.CO]`.

[85] V. Domcke, F. Muia, M. Pieroni, and L. T. Witkowski, "PBH dark matter from axion inflation," *JCAP* **07** (2017) 048, `arXiv:1704.03464 [astro-ph.CO]`.

[86] O. Özsoy, "Synthetic Gravitational Waves from a Rolling Axion Monodromy," *JCAP* **04** (2021) 040, `arXiv:2005.10280 [astro-ph.CO]`.

[87] G. A. Palma, S. Sypsas, and C. Zenteno, "Seeding primordial black holes in multifield inflation," *Phys. Rev. Lett.* **125** no. 12, (2020) 121301, `arXiv:2004.06106 [astro-ph.CO]`.

[88] J. Fumagalli, S. Renaux-Petel, J. W. Ronayne, and L. T. Witkowski, "Turning in the landscape: a new mechanism for generating Primordial Black Holes," `arXiv:2004.08369 [hep-th]`.

[89] M. Braglia, X. Chen, and D. Kumar Hazra, "Probing Primordial Features with the Stochastic Gravitational Wave Background," *JCAP* **03** (2021) 005, `arXiv:2012.05821 [astro-ph.CO]`.

[90] J. Chluba, A. L. Erickcek, and I. Ben-Dayan, "Probing the inflaton: Small-scale power spectrum constraints from measurements of the CMB energy spectrum," *Astrophys. J.* **758** (2012) 76, `arXiv:1203.2681 [astro-ph.CO]`.

[91] J. Chluba and D. Grin, "CMB spectral distortions from small-scale isocurvature fluctuations," *Mon. Not. Roy. Astron. Soc.* **434** (2013) 1619–1635, `arXiv:1304.4596 [astro-ph.CO]`.

[92] O. Özsoy and Z. Lalak, "Primordial black holes as dark matter and gravitational waves from bumpy axion inflation," *JCAP* **01** (2021) 040, `arXiv:2008.07549 [astro-ph.CO]`.