

# Possible Signatures of Inflationary Particle Content: Spin-2 Fields

Matteo Biagetti <sup>a</sup>, Emanuela Dimastrogiovanni <sup>b</sup>, Matteo Fasiello <sup>c</sup>

<sup>a</sup> Institute of Physics, Universiteit van Amsterdam, Science Park, 1098XH Amsterdam, The Netherlands

<sup>b</sup> CERCA & Department of Physics, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>c</sup> Stanford Institute for Theoretical Physics and Department of Physics, Stanford University, Stanford, CA 94306

**Abstract.** We study the imprints of a massive spin-2 field on inflationary observables, and in particular on the breaking of consistency relations. In this setup, the minimal inflationary field content interacts with the massive spin-2 field through dRGT interactions, thus guaranteeing the absence of Boulware-Deser ghostly degrees of freedom. The unitarity requirement on spinning particles, known as Higuchi bound, plays a crucial role for the size of the observable signal.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The framework</b>	<b>3</b>
2.1	The Higuchi bound and its observational consequences	6
2.2	Strong coupling scales	7
2.3	Coupling to matter	7
<b>3</b>	<b>Non-standard tensor-scalar-scalar consistency relations</b>	<b>8</b>
3.1	Background	8
3.2	Linear perturbations	9
3.3	Tensor-scalar-scalar bispectrum	10
3.4	All of the freedom in the unitarity bound	13
3.5	New cubic interactions	13
<b>4</b>	<b>Observational consequences</b>	<b>15</b>
<b>5</b>	<b>Conclusions</b>	<b>17</b>
<b>A</b>	<b>Consistency relations: general considerations</b>	<b>18</b>
<b>B</b>	<b>Alternative route to CRs breaking</b>	<b>20</b>

---

## 1 Introduction

The main reason for investigating non-minimal particle content during inflation is the energy scale at which inflationary dynamics may take place. With high-scale inflation set at about  $10^{14}$  GeV, there is no other comparable window on such high energy processes. The characterization of the signatures corresponding to different particle species is indispensable for a precise observation-to-theory mapping in case of detection. Cosmological probes can indeed inform our model building and even UV completion efforts. This line of reasoning is at the heart of recent literature on the subject [1–5].

Additional degrees of freedom (dof) during inflation may take many shapes (mass, spin). Perhaps the most immediate indication of their impact on observables stems from the value of their mass as compared to the Hubble scale  $H$ . The latter is indeed a crucial, threshold, value. The reasons are manifold. First of all, the energy scale of inflation is itself generally of order  $H$ ; it naturally follows that too large a mass,  $m \gg H$ , would justify the integrating out of any dof leaving very little in terms of imprints (although see e.g. [6–8] for important caveats). A mass of order Hubble can instead support distinct signatures and has been the subject of many studies, especially in the scalar case [3, 9–15]. This range of values for the mass is also natural from the point of view of supersymmetric theories: if such symmetry is not broken at a

higher scale, the inflationary vacuum energy will eventually break it, thereby justifying the expectation that some of the fields populating the supersymmetric multiplets will have masses of order Hubble. The behaviour of inflationary correlation functions also dictates that particles whose Compton wavelength is much smaller than the horizon<sup>1</sup> have negligible effects on late-time observables, identifying again  $1/H$  as a lower limit. Finally, for particles of spin  $s = 2$  or higher, unitarity imposes lower bounds on the mass in de Sitter

$$m^2 > s(s - 1)H^2, \quad (1.1)$$

which further underscores the frontier nature of the  $m \sim H$  mass range. The unitarity or Higuchi [16–18] constraint is less stringent away from de Sitter [19, 20] but, as we shall see, it remains a powerful bound.

The most efficient way to probe both the field content and inner structure of the inflationary Lagrangian is to consider  $N$ -point correlation functions, non-Gaussianities ( $N \geq 3$ ). The Planck satellite mission has set precise bounds on non-Gaussianities, starting with the three-point function of curvature fluctuations [21] in the key momenta configurations such as local, equilateral, orthogonal. Within observationally viable values, there is still plenty of room to seek imprints of dynamics beyond the single-field inflation scenario. Of particular interest are the squeezed limit of the three-point function and its higher order generalizations. This configuration is particularly sensitive to the presence of additional dof in light of consistency relations (CRs). The latter are relations between the squeezed limit of an  $N + 1$ -point function and its  $N$ -point counterpart.

CRs stem from residual gauge diffeomorphisms of the (inflationary) theory [22–26]. One may derive non-trivial CRs when the soft mode characterizing any squeezed limit transforms non-linearly under the residual diff. In the minimal inflationary setup the curvature perturbation  $\zeta$  and tensor mode  $\gamma$  transform non linearly respectively under dilatation and anisotropic rescaling, thereby generating the well-known scalar and tensor CRs. In each of these two cases, the leading effect of a long scalar (tensor) mode on  $N$  short modes corresponds to the action of a residual gauge-symmetry on the  $N$ -point function and can therefore be gauged away.

In this work we focus on the case of a three-point function with a long tensor mode and two scalar short modes. In this case, the CR in single-field inflation reads

$$\lim_{\vec{k}_1 \rightarrow 0} \left\langle \gamma_{\vec{k}_1}^\lambda \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \right\rangle' \simeq \frac{3}{2} \left\langle \gamma_{\vec{k}_1}^\lambda \gamma_{-\vec{k}_1}^\lambda \right\rangle' \epsilon_{ij}^\lambda \hat{k}_{3i} \hat{k}_{3j} \left\langle \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \right\rangle', \quad (1.2)$$

where  $\gamma$  and  $\zeta$  are the tensor and scalar perturbations respectively, we defined the polarization tensors  $\epsilon_{ij}^\lambda$  as  $\gamma_{ij} = \sum_\lambda \epsilon_{ij}^\lambda \gamma^\lambda$  and the prime indicates that we left implicit the momentum-conserving delta.

Whenever additional fields appear on the scene, CRs can be broken<sup>2</sup> if, for ex-

<sup>1</sup>The equations of motion, for example in quasi-de Sitter, show that as soon as  $m \sim H$  the mass of a generic particle becomes relevant to its dynamics and leads to decaying solutions.

<sup>2</sup>For our purposes it suffice to have non-standard consistency relations in the sense that the variation of the power spectrum is not related to the squeezed limit of one three-point function but rather to a weighted linear combinations of at least two such observables.

ample, there exists one more independent mode that transforms non-linearly under the diffeomorphism. It follows that the squeezed bispectrum may no longer be gauged away and an enhanced signal is made possible in this configuration. The power of CRs breaking as a probe of new physics is not limited to inflationary setups and has been widely employed in other contexts such as large scale structure [27–36]. One should further note that the squeezed one is the best suited of configurations to reconstruct the inflationary potential as it is less sensitive to possible variations of the statistics of the fluctuations in sub-volumes [37].

In this manuscript we will report on work where the inflationary dynamics is enriched by a massive spin-2 field non-minimally coupled to the usual massless graviton via ghost-free interactions [38–40]. As a result, the standard tensor CR is modified. We quantify the consequences of this “breaking” on the squeezed limit of the tensor-scalar-scalar bispectrum signal and show how the constraints originating from a (softened) Higuchi bound propagate all the way to observables. Despite a weakened unitarity constraint away from dS, we find that the amplitude of the signal produced in this setup is too weak for detection by upcoming large scale structure surveys.

This work is organized as follows. In *Section 2* we introduce the model under study comprising the inflaton field minimally coupled to a spin-2 field interacting with another (massive) spin-2 field through dRGT-interactions. We discuss at length the features of the model that will be most relevant in the inflationary context. *Section 3* is devoted to outlining how consistency relations play out in our specific setup and calculating the resulting tensor-scalar-scalar bispectrum. In *Section 4* we discuss the observational consequences of these results. We conclude with *Section 5*, where we comment on our findings and suggest possible directions for further work on the subject. In *Appendix A* we provide a general overview of CRs and their breaking. In *Appendix B* we detail on a route to modified CRs which is alternative, up to horizon scales, to the one in *Section 3*, and include a diagrammatic proof of CRs breaking.

## 2 The framework

It is well-known that additional degrees of freedom populating the inflationary Lagrangian have the potential to modify CRs in favour of a enhanced squeezed bispectrum signal (see *Appendix A* for a general discussion). There already exist several studies on the subject in the literature. We focus our analysis here on the case where the minimal inflationary particle content is enriched by a spin-2 field. Our discussion ought to start from the realization due to [54] that a theory of interacting<sup>3</sup> spin-2 particles admits at most one massless spin-2 field. As we are working from the perspective of adding particle content to the minimal (GR + scalar field) inflationary scenario, we are then immediately led to introducing a new, massive, graviton.

The theory of massive gravity has received enormous attention in recent years. This is due to the combination of experimental results [55–57] confirming the current

---

<sup>3</sup>Note that there is a further implicit restriction: even indirect, e.g. mediated by other fields, interactions are forbidden.

acceleration of the universe and the recent non-linear formulation [38, 39] of a ghost-free massive gravity theory known as dRGT. An infrared modification of gravity is indeed among the most studied mechanisms, along with dark energy, to explain late time acceleration. The interest for a theory of massive gravity is of course much wider than pertains its use for late-time cosmology. Within string theory for example, open strings have spin-2 excitations whose lowest energy state is massive at tree level. The formulation of [39] in particular, has found applications as varied as e.g. its use as a framework for translational symmetry breaking and dissipation of momentum in holography [58, 59]. Our interest is focussed on the inflationary context: here a consistent massive spin-2 field next to GR and a scalar inflaton field takes the form of a theory known as bigravity. This is an extension of dRGT theory that contains the same ghost-free structure. In bigravity each of the two metrics,  $g$  and  $f$ , has its own Einstein-Hilbert term and they interact via the dRGT potential. The action reads

$$S = \int d^4x \left[ M_P^2 \sqrt{-g} R[g] + \sqrt{-g} P_g(X, \varphi) + 2\sqrt{-g} m^2 M^2 V + M_f^2 \sqrt{-f} R[f] \right], \quad (2.1)$$

where a few details are in order:

- The interaction potential  $V$  is defined as

$$V = \sum_{n=0}^4 \beta_n \mathcal{E}_n(\sqrt{g^{-1}f}), \quad (2.2)$$

where the  $\beta_n$  are free parameters and the polynomials  $\mathcal{E}_n(\mathbb{X})$  take the form

$$\begin{aligned} \mathcal{E}_0(\mathbb{X}) &= 1, & \mathcal{E}_1(\mathbb{X}) &= \text{Tr}(\mathbb{X}) \equiv [\mathbb{X}], & \mathcal{E}_2(\mathbb{X}) &= \frac{1}{2} ([\mathbb{X}]^2 - [\mathbb{X}^2]), \\ \mathcal{E}_3(\mathbb{X}) &= \frac{1}{3!} ([\mathbb{X}]^3 - 3[\mathbb{X}^2][\mathbb{X}] + 2[\mathbb{X}^3]), \\ \mathcal{E}_4(\mathbb{X}) &= \frac{1}{4!} ([\mathbb{X}]^4 - 6[\mathbb{X}^2][\mathbb{X}]^2 + 8[\mathbb{X}^3][\mathbb{X}] + 3[\mathbb{X}^2]^2 - 6[\mathbb{X}^4]). \end{aligned} \quad (2.3)$$

- Due to the properties of the  $\mathcal{E}_n$  polynomials, if it were not for the coupling to matter (here restricted to the metric “g”) the action would be symmetric under the exchange  $g \leftrightarrow f$   $M_g \equiv M_P \leftrightarrow M_f$ ,  $\beta_n \leftrightarrow \beta_{4-n}$ .
- The mass  $M$  has already been symmetrized via

$$M^2 = \frac{M_P^2 M_f^2}{M_P^2 + M_f^2}. \quad (2.4)$$

and we have introduced  $\kappa \equiv \frac{M_f^2}{M_P^2}$ .

- As a consequence of the above points, it is not strictly true that  $g$  is the massless spin-2 field and  $f$  is the massive one. In fact, the mass eigenstate of the theory are in general time-dependent. We can on the other hand work in a low energy configuration [62] where this is approximately true.

- The  $P_g(X, \varphi)$  stands for a generic matter Lagrangian and  $X = -\frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi$ . We will assume the potential within  $P_g(X, \varphi)$  is driving the background dynamics of the inflaton field  $\varphi$ , minimally coupled to  $g$  only. We refer the reader to *Section 2.3* for the more general case.

The theory in Eq. (2.1) has 8 propagating degrees of freedom: 5 (2T+2V+1S) from a healthy spin-2, 2 more from a massless spin-2 (2T) and a scalar. In what follows we will focus our attention on the  $2 \times 2T$  degrees of freedom plus one scalar. Let us elaborate on the reasoning behind this choice. Non-linear massive (bi)gravity is endowed with a very efficient screening mechanism known as Vainshtein screening [60, 61]. In specific configurations the coupling to matter of the helicity-0 and helicity-1 modes is highly suppressed. This is, for example, what allows massive gravity to pass a host of cosmological tests in setups (e.g. solar system scales) where observations are in exquisite agreement with general relativity: the additional degrees of freedom are screened. We are working in a rather different, inflationary, background here; we will nevertheless assume that the non-linearities in the helicity-0 and helicity-1 part of the theory will generate sufficient screening so that we can focus the analysis on the remaining dofs.

There exist in the literature several important studies [63–67] on the perturbation theory of all the degrees of freedom (including matter) populating bigravity during the various cosmological epochs. Requiring that an efficient screening is in place in an inflationary background is tantamount to requiring that non-linearities in the theory (these studies usually stop at quadratic order in perturbation theory) will downplay any effect due to the vector and helicity-0 modes. In support of this assumption is also the intuition that Vainshtein screening is typically more efficient in homogeneous and isotropic backgrounds. On the other hand, we stress here that studies of perturbations limited to quadratic order already indicate [66] that all bigravity degrees of freedom are well-behaved during inflation.

The Vainshtein mechanism relies on the fact that whenever non-linearities are important the canonical normalization of e.g. the helicity-zero mode is redressed and so will be also the coupling with matter. The helicity-0 (usually indicated by  $\pi$ ) coupling to matter is dictated by the structure in Eq. (2.1) and includes <sup>4</sup> the term  $\pi T_\mu^\mu$ . Whenever non-linearities in  $\pi$  modify the normalization of the kinetic term, the corresponding coupling to matter will be

$$\pi_c T_\mu^\mu = \frac{1}{Z(\pi_0, \beta_n)} \pi T_\mu^\mu, \quad (2.5)$$

where  $Z$  is a function of background quantities and the non-linearities are weighted by the  $\beta_n$ s. As soon as  $\pi$  self-interactions are important,  $Z \gg 1$  and screening takes place. It is important to note at this stage that the same does not apply to the tensor dofs introduced by making the metric  $f$  dynamical. Even though  $f$  does not directly couple to matter, it does so via  $g$  and so the coupling does not undergo the same kind of suppression as the helicity-0/1 modes. On this basis we can go on and prioritize

---

<sup>4</sup>Also terms such as  $\frac{1}{\Lambda^3}\partial_\mu\pi\partial_\nu\pi T^{\mu\nu}$  are allowed, see [68].

the observational effects of the tensor degrees of freedom contained in  $f$ . What must not be omitted in our analysis are the consequences that purely formal consistency criteria on the helicity-0 and vector sector have on the observables made up by the remaining dofs. The most important case in point is the helicity-0 unitarity bound, which we address next. The reader familiar with the matter and more interested in the calculation for the bispectrum signal may want to skip the content of the rest of *Section 2* and go directly to *Section 3*.

## 2.1 The Higuchi bound and its observational consequences

The unitarity bound on the helicity-0 mode corresponds to the condition on the coefficient of the kinetic term in the quadratic theory to be positive<sup>5</sup>. This stems from requiring that the Hamiltonian is bounded from below so that the dynamical system is stable. Naturally, the latter also implies that the coefficients of the mass and spatial gradient terms, if present, need also be positive. It turns out that the helicity-zero Higuchi bound is the most restrictive on the parameter space of bigravity theory, so that it is justified to study it more in detail. It is well-known that unitary representations of the de Sitter group for spin-2 fields are massless, partially massless, and massive, with

$$m^2 = 0 \text{ (GR)}, \quad m^2 = 2H^2 \text{ (partially massless)}, \quad m^2 > 2H^2 \text{ (massive)}. \quad (2.6)$$

It has been shown [20] that the existence of helicity-1 and helicity-0 interactions in the partially massless case implies the absence of the intriguing conjectured partially massless symmetry for both massive gravity and bigravity, at least as long as the kinetic term is of the standard form. In the massive representation, our stepping away from de Sitter and, most importantly, our dealing with (i) a fully non-linear theory of (ii) bigravity rather than massive gravity, relaxes [20] to some extent the bound on the now dressed mass  $\tilde{m}$ :

$$\tilde{m}^2 \left( 1 + \frac{M_P^2 H_f^2}{M_f^2 H^2} \right) \geq 2H^2 \quad (2.7)$$

where we have defined  $\tilde{m}^2$  as

$$\tilde{m}^2 \equiv \frac{m^2}{2} \frac{H}{H_f} \frac{M^2}{M_P^2} \left( \beta_1 + 2\beta_2 \frac{H}{H_f} + \beta_3 \frac{H^2}{H_f^2} \right), \quad (2.8)$$

and the quantities  $H$  and  $H_f$  are the Hubble scales associated to the FLRW solution for the metrics  $g$  and  $f$  respectively<sup>6</sup>. These notions have immediate consequences also for the tensor sector and, as we shall see, for the observability of the  $\langle \gamma\zeta\zeta \rangle$  bispectrum in the squeezed limit. In fact, we anticipate here (see Eqs.(3.5)-(3.10) in the text)

<sup>5</sup>See [69] for a derivation of the Higuchi bound from the perspective of the AdS/CFT correspondence and in particular as a consequence of reflection positivity in radially quantized CFT<sub>3</sub>.

<sup>6</sup>Note that the absence of  $\beta_0, \beta_4$  is due to the fact that these quantities are typically fixed by the tadpole cancellation requirement.



that the mass in the (suitably diagonalized) tensor equations of motion in the low  $k$  limit is the very same LHS of Eq. (2.7) [18]. This implies that the helicity-0 mode unitarity bound forces an effective mass on the tensor sector, and in particular one of at least order Hubble: the corresponding tensor wavefunction will then inevitably have a component that decays outside the horizon, to the detriment of the observed signal. As for possible tachyonic and gradient instabilities, we will work in a regime where any such possibility is above the reach of the effective theory: the “low energy” regime of [62].

## 2.2 Strong coupling scales

For the sake of providing a more complete picture, we find it necessary to briefly expound on the strong coupling scale (scs) of the bigravity theory at hand. Naturally, it is the presence of the  $g - f$  interaction term that dictates the lowest scales in the theory. In particular, depending on the hierarchy between the two masses  $M_g$  and  $M_f$ , the lowest “naive” strong coupling scale (in the helicity-0 sector) can be as low as:

$$\Lambda_3^{g,f} = (m^2 M_{g,f})^{1/3} . \quad (2.9)$$

A few comments are in order.

Taking at face value the above quantities may lead to some worries about a scs perilously close to the energy we are working at. Indeed, although the Higuchi bound forces the dressed mass to be larger than the Hubble scale  $H$ , one typically satisfies this condition by requiring  $H_f/M_f \gg H/M_g$ , and allowing  $m^2 \lesssim H^2$ . This will be also the case for our parameter space of choice as keeping the ratio  $M_f^2/M_g^2 \cdot H^2/H_f^2$  small will allow approximately constant mass eigenstates in the tensor sector [62]. However, the scs is naive in the sense that the background via the Vainshtein mechanism changes the canonical scale and so the scale at which the fluctuations are strongly coupled. Instead of the naive scale  $\Lambda$ , one should rather think of a larger quantity as the strong coupling scale,  $\Lambda^* = \sqrt{Z}\Lambda$ , where  $Z$  is the same quantity discussed below equation (2.5). The efficient screening we require goes then in the same direction of a larger strong coupling scale.

## 2.3 Coupling to matter

In writing Eq. (2.1) we have assumed the existence of a unique matter sector that is minimally coupled to the metric  $g$  only; it is in fact this coupling (together with the hierarchy between  $M_g$  and  $M_f$ ) that breaks the otherwise symmetric role played by the two metrics in our setup. One might ask: is this the only possibility? Would any other coupling lead to a ghostly (classical or quantum) degree of freedom? How about coupling different matter sectors to different metrics? Naturally, coupling matter just to the  $f$  sector is just as acceptable but the  $g \leftrightarrow f$  exchange would be mirrored by a swapping of the observables we are going to calculate, and deliver no detectable advantage. Coupling different matter sectors to different metric fields is also not immediately helpful because communication between different matter sectors will be mediated and



therefore suppressed. On the other hand, what might prove to be more intriguing for us is the coupling via the effective composite metric [70]:

$$g_{\mu\nu}^{\text{eff}} = \alpha^2 g_{\mu\nu} + 2\alpha\beta g_{\mu\alpha} \left( \sqrt{g^{-1}f} \right)_\nu^\alpha + \beta^2 f_{\mu\nu} , \quad (2.10)$$

where  $\alpha$  and  $\beta$  are two arbitrary real dimensionless parameters. In this case the matter Lagrangian in Eq. (2.1) would be replaced, in a simple example, by:

$$\mathcal{L}_m = \sqrt{-g_{\text{eff}}} \left[ g_{\text{eff}}^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi + V(\varphi) \right]. \quad (2.11)$$

This specific form for the coupling preserves the ghost-free nature of the theory both at the classical and quantum level [70]. From the EFT perspective any ghost that is above the cutoff scale of the theory is completely harmless and it is in this sense that one can freely use the coupling via composite metric in Eq. (2.10).

### 3 Non-standard tensor-scalar-scalar consistency relations

Equipped with sufficient information on our bigravity model, we now go on to tackle the observational signal generated by the tensor-scalar-scalar three-point function  $\langle \gamma_g \zeta \zeta \rangle$ , where  $\gamma_g$  is the  $g$ -metric tensor fluctuation and  $\zeta$  is a repository of scalar fluctuations. As mentioned in *Section 1* and detailed in *Appendix A*, non standard CRs are possible whenever additional field content has non-linear transformation under a gauge diff. The extra field content of choice here includes the  $f$ -metric new tensor dof. In order to most clearly show how consistency relations are modified in our setup we provide in this manuscript two different routes to non standard CRs. We present the first one below, in *Section 3.2*, and leave the second one to *Appendix B*. The latter relies on an approximation that does not hold outside the horizon and cannot therefore be trusted all the way to late-time observables. Nevertheless, we find the presentation in *Appendix B* very instructive and choose to include it in the manuscript. In *Section 3.2* we will solve the full<sup>7</sup> coupled system of equations for the traceless transverse part of  $h_f, h_g$ . We will then calculate the tree level  $\langle \gamma_g \zeta \zeta \rangle$  diagram and show that its late time limit is different than that of the standard single-field model as it contains a term that depends on the mass  $m$ . The late-time power spectra  $P_{\gamma_g}$ <sup>8</sup> and  $P_\zeta$  cannot capture this mass dependence at tree level, signaling that CRs are indeed modified in our setup.

#### 3.1 Background

Following for example [18, 20, 62], it is straightforward to derive background equations and study the quadratic perturbation theory around FLRW solutions for the metric  $f$  and  $g$ . We denote

$$\bar{g}_{\mu\nu} dx^\mu dx^\nu = -N^2 dt^2 + a^2 \delta_{ij} dx^i dx^j, \quad \bar{f}_{\mu\nu} dx^\mu dx^\nu = -\tilde{N}^2 dt^2 + b^2 \delta_{ij} dx^i dx^j. \quad (3.1)$$

<sup>7</sup>This is in contradistinction with the procedure outlined in *Appendix B* where an approximation of the type  $\delta\mathcal{L}^{(2)} \ll \mathcal{L}_0^{(2)}$  on parts of the quadratic Lagrangian is necessary.

<sup>8</sup>We do not include  $P_{\gamma_f}$  here because  $\gamma_f$  does not couple directly with matter. However, the conclusion would not change as at tree level and in the late time limit  $P_{\gamma_f}$  too is unaware of the mass  $m^2$ .

Here  $N$  and  $\tilde{N}$  are background lapse functions and  $a$  and  $b$  are the scale factors for the corresponding metrics  $g$  and  $f$ . The Friedmann equations read:

$$3M_P^2 H^2 = \rho(\varphi) + 3m^2 M^2 \sum_{n=0}^3 \frac{\beta_n}{(3-n)!n!} \left(\frac{b}{a}\right)^n, \quad (3.2)$$

$$3M_f^2 H_f^2 = 3m^2 M^2 \sum_{n=0}^3 \frac{\beta_{n+1}}{(3-n)!n!} \left(\frac{b}{a}\right)^{n-3}, \quad (3.3)$$

where  $\rho(\varphi)$  in Eq. (3.2) and in particular the potential  $V(\varphi)$  in it, is the leading term, driving inflation, on the RHS. We notice in passing that we will be working in the so-called healthy branch (see e.g. [62]) of bigravity solutions in order to avoid strong coupling issues. The healthy branch is defined by the condition  $H/H_f = b/a \equiv \xi$ .<sup>9</sup> Upon using

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + a^2 h_{g\mu\nu}, \quad f_{\mu\nu} = \bar{f}_{\mu\nu} + b^2 h_{f\mu\nu}, \quad (3.4)$$

one obtains from the quadratic action the equations of motion for  $h_f, h_g$ .

### 3.2 Linear perturbations

Let us start with the equations of motion for  $\gamma_g^{TT}$  and  $\gamma_f^{TT}$ , the traceless transverse components of, respectively,  $h_g$  and  $h_f$ <sup>10</sup>. Tensor perturbations of a given  $k$ -mode are separable into two independent helicity modes with the same mode-function:  $\gamma_g, \gamma_f$ . Combining Eq. (2.1) and (3.4) one obtains[18]:

$$\gamma_g'' - \frac{2}{\tau} \gamma_g' + k^2 \gamma_g + \frac{\mathcal{M}^2}{H^2 \tau^2} \left( \frac{r_{fg}^2}{1 + r_{fg}^2} \right) (\gamma_g - \gamma_f) = 0, \quad (3.5)$$

$$\gamma_f'' - \frac{2}{\tau} \gamma_f' + k^2 \gamma_f + \frac{\mathcal{M}^2}{H^2 \tau^2} \frac{1}{(1 + r_{fg}^2)} (\gamma_f - \gamma_g) = 0, \quad (3.6)$$

where  $\mathcal{M}^2 = \tilde{m}^2 [1 + 1/r_{fg}^2]$  and we introduced the parameter

$$r_{fg} \equiv \frac{M_f H}{M_P H_f}. \quad (3.7)$$

In writing Eqs. (3.5)-(3.6) we have also assumed a slowly varying  $\xi = b/a$ . It turns out to be convenient at this stage to change basis, from  $\gamma_{g,f}$  to  $\gamma_{+,-}$ , where

<sup>9</sup>To conform with standard notation we have defined the ratio of the two scale factors as  $\xi \equiv b/a$ . This is also the symbol generally used for gauge transformations. Which one is meant should be clear from the context.

<sup>10</sup>As anticipated, henceforth we shall focus on the tensor sectors perturbations. For a systematic study of quadratic perturbations in bigravity, we refer the reader to, for example, the work in [18].

$$\gamma_{\pm} = \gamma_g \pm r_{fg}^2 \gamma_f, \quad (3.8)$$

so that the equations of motion now read

$$\gamma_+'' - \frac{2}{\tau} \gamma_+' + k^2 \gamma_+ = 0, \quad (3.9)$$

$$\gamma_-'' - \frac{2}{\tau} \gamma_-' + \left( k^2 + \frac{\mathcal{M}^2}{H^2 \tau^2} \right) \gamma_- - \frac{\mathcal{M}^2}{H^2 \tau^2} \left( \frac{1 - r_{fg}^2}{1 + r_{fg}^2} \right) \gamma_+ = 0. \quad (3.10)$$

The solutions, once written back in the  $\gamma_g, \gamma_f$  basis are

$$\begin{aligned} \gamma_g &= \left( \frac{1 + r_{fg}}{1 + r_{fg}^2} \right) \frac{H}{M_P} \frac{\sqrt{2} (i - k\tau)}{k^{3/2}} e^{-ik\tau} \\ &\quad + \left( \frac{-1 + r_{fg}}{1 + r_{fg}^2} \right) r_{fg} \frac{H}{M_P} e^{i\nu\pi} (1 + i) \sqrt{\frac{\pi}{2}} (-\tau)^{3/2} H_\nu^{(1)}(-k\tau), \end{aligned} \quad (3.11)$$

$$\begin{aligned} \gamma_f &= \left( \frac{1 + r_{fg}}{1 + r_{fg}^2} \right) \frac{H}{M_P} \frac{\sqrt{2} (i - k\tau)}{k^{3/2}} e^{-ik\tau} \\ &\quad + \left( \frac{1 - r_{fg}}{1 + r_{fg}^2} \right) \frac{1}{r_{fg}} \frac{H}{M_P} e^{i\nu\pi} (1 + i) \sqrt{\frac{\pi}{2}} (-\tau)^{3/2} H_\nu^{(1)}(-k\tau), \end{aligned} \quad (3.12)$$

where we have defined  $\nu \equiv \sqrt{9/4 - \mathcal{M}^2/H^2}$ .

Note that the solution for the mode-functions in Eqs. (3.11)-(3.12) were found by solving Eqs.(3.9)-(3.10) first and requiring that the mode functions asymptote to Bunch-Davies vacua deep inside the horizon:

$$\gamma_g(k, \tau) \xrightarrow{|k\tau| \gg 1} \frac{2}{a M_P} \frac{e^{-ik\tau}}{\sqrt{2k}}, \quad \gamma_f(k, \tau) \xrightarrow{|k\tau| \gg 1} \frac{2}{b M_f} \frac{e^{-ik\tau}}{\sqrt{2k}}. \quad (3.13)$$

It is worth mentioning the special case  $r_{fg}^2 = 1$ . In this particular configuration, despite the presence of the mass term in the equations of motion, the two solutions for  $\gamma_g, \gamma_f$  are independent and identical to two copies of the usual massless solution for  $\gamma_g$  in standard single-field inflation. One concludes that in this case, at the level of the quadratic action for the traceless transverse tensor components, the theory behaves as two copies of GR and one has to go to cubic order to start probing the massive gravity interactions.

### 3.3 Tensor-scalar-scalar bispectrum

Equipped with the solution to the quadratic theory, we can now consider the effect of the  $m^2$  driven  $g - f$  coupling on standard observables such as the tensor-scalar-scalar bispectrum. We shall be considering the contribution due to the  $\gamma_g^{ij} \partial_i \zeta \partial_j \zeta$  interaction.

Note that we define the gauge-invariant comoving curvature perturbation  $\zeta$  in the standard way,  $\zeta = \psi_g + \mathcal{H}\delta\rho/\bar{\rho}'$ , where  $\psi_g$  is the gravitational potential identified with the diagonal part of the spatial components of the metric  $g$  (see also [63, 66]). The important question is if this quantity is conserved on scales larger than the Hubble radius. Since it is only the metric  $g$  that interacts with matter, in light of the conservation equation and of our assumption that the helicity-0 and helicity-1 modes are screened, one immediately finds that  $\zeta' = \psi_g' = 0$  outside the horizon, and we conclude that one can adopt for  $\zeta$  the usual expression. This is consistent with e.g. the analysis in [66]. Since the  $\gamma_g^{ij}\partial_i\zeta\partial_j\zeta$  interaction is already present in single-field inflation, we can already anticipate that the imprint of the massive gravity potential is all stored in the wavefunctions and, in particular, their  $\mathcal{M}$ -dependent parts. The formal expression for the bispectrum in the in-in formalism is

$$\left\langle \gamma_g^\lambda(t, \vec{k}_1)\zeta(t, \vec{k}_2)\zeta(t, \vec{k}_3) \right\rangle = -i \int_{t_0}^t dt' \left\langle \left[ \gamma_g^\lambda(t, \vec{k}_1)\zeta(t, \vec{k}_2)\zeta(t, \vec{k}_3), H_{\text{int}}(t') \right] \right\rangle \quad (3.14)$$

where

$$H_{\text{int}}(t') = -\mathcal{L}_{\text{int}}(t') = -M_P^2\epsilon \int d^3x a(t') \gamma_{g,ij}(\vec{x}, t') \partial_i\zeta(\vec{x}, t') \partial_j\zeta(\vec{x}, t'). \quad (3.15)$$

It is convenient to split the result of the bispectrum calculation in two pieces. One, which we indicate as  $\mathcal{B}_{\mathfrak{m}}$ , will be due to the “massless”<sup>11</sup> component of  $\gamma_g$  in the integral of Eq. (3.15); the other one, corresponding to the remaining component of the wavefunction, will be indicated by  $\mathcal{B}_m$ . One obtains:

$$\left\langle \gamma_g^\lambda(t, \vec{k}_1)\zeta(t, \vec{k}_2)\zeta(t, \vec{k}_3) \right\rangle = (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) e_{ij}^\lambda(\hat{k}_1) \left( -\hat{k}_{2i}\hat{k}_{3j} - \hat{k}_{3i}\hat{k}_{2j} \right) [\mathcal{B}_{\mathfrak{m}} + \mathcal{B}_m]. \quad (3.16)$$

It is straightforward (and identical to the single-field case) to derive the quantity in  $\mathcal{B}_{\mathfrak{m}}$ . The integral necessary to obtain an explicit expression for  $\mathcal{B}_m$  can be performed analytically by fixing each time the value of the mass  $\mathcal{M}$ . Let us fix it to  $\mathcal{M} = \sqrt{2}H$ , a value that saturates the Higuchi bound. The explicit expression for the bispectrum now reads:

$$\mathcal{B}_{\mathfrak{m}}(k_1, k_2, k_3) = \frac{H^4}{4\epsilon M_P^4} \left( \frac{1 + r_{fg}}{1 + r_{fg}^2} \right)^2 \mathcal{I}_{\mathfrak{m}}(k_1, k_2, k_3), \quad (3.17)$$

$$\mathcal{B}_m(k_1, k_2, k_3) = \frac{H^4}{4\epsilon M_P^4} \frac{1 - r_{fg}^2}{(1 + r_{fg}^2)^2} r_{fg} \mathcal{I}_m(k_1, k_2, k_3), \quad (3.18)$$

---

<sup>11</sup>The massless component of  $\gamma_g$  is the first term in Eq. (3.11), which is unaware of  $\nu = \nu(\mathcal{M})$  and is therefore termed massless.

where

$$\mathcal{I}_{\mathfrak{m}}(k_1, k_2, k_3) \equiv \frac{\sum_i k_i^3 + 2 \sum_{i \neq j} k_i k_j^2 + 2 \Pi_i k_i}{k_1^3 k_2^2 k_3^2 k_t^2}, \quad (3.19)$$

$$\mathcal{I}_{\mathfrak{m}}(k_1, k_2, k_3) \equiv \frac{1}{\Pi_i k_i^2} \left( \frac{k_1(k_2 + k_3) + 3k_2 k_3 + k_2^2 + k_3^2}{k_t^2} - \gamma - \ln[-k_t \tau_0] \right), \quad (3.20)$$

with  $k_t \equiv k_1 + k_2 + k_3$  and where  $\gamma$  is the Euler's constant. The logarithm in the last expression is handled in the standard fashion [86–88]. The presence of a log term in the bispectrum signals a non-trivial interaction between the hard  $\zeta$  modes and the soft graviton outside the Hubble radius and accounts for the fact that the massive graviton mode, despite its decay on large scales, sources the massless graviton perturbation from the moment it leaves the Hubble radius till the end of inflation. Such an effect is not found in the standard single-field scenario.

We can now take the squeezed limit ( $k_1 \ll k_2 \simeq k_3$ ) of the expressions in (3.17)-(3.18), to get

$$\mathcal{B}_{\mathfrak{m}}(k_1, k_2, k_3) \xrightarrow{k_1 \ll k_2 \simeq k_3} \frac{3}{4} P_{\gamma_g}^\lambda(k_1) P_\zeta(k_3), \quad (3.21)$$

$$\mathcal{B}_{\mathfrak{m}}(k_1, k_2, k_3) \xrightarrow{k_1 \ll k_2 \simeq k_3} \frac{1}{8} P_{\gamma_g}^\lambda(k_1) P_\zeta(k_3) \frac{k_1}{k_3} \left( \frac{1 - r_{fg}}{1 + r_{fg}} \right) r_{fg} (5 - 4\gamma - 4 \ln[-2k_3 \tau_0]), \quad (3.22)$$

where we have defined the tensor and scalar power spectra at late times<sup>12</sup>

$$P_{\gamma_g}^\lambda(k_1) = \lim_{\tau \rightarrow 0} [\gamma_g(k_1, \tau) \gamma_g^*(k_1, \tau)] \simeq \left( \frac{1 + r_{fg}}{1 + r_{fg}^2} \right)^2 \frac{2 H^2}{M_P^2} \frac{1}{k_1^3}, \quad (3.23)$$

$$P_\zeta(k_3) = \lim_{\tau \rightarrow 0} [\zeta(k_3, \tau) \zeta^*(k_3, \tau)] \simeq \frac{H^2}{4 \epsilon M_P^2} \frac{1}{k_3^3} \quad (3.24)$$

A tensor-scalar-scalar bispectrum that satisfies the consistency condition would, in the squeezed limit, behave according to

$$\langle \gamma_g^\lambda(\vec{k}_1) \zeta(\vec{k}_2) \zeta(\vec{k}_3) \rangle \xrightarrow{k_1 \ll k_2 \simeq k_3} -(2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) \frac{1}{2} e_{ij}^\lambda(\hat{k}_1) \hat{k}_{3i} \hat{k}_{3j} P_{\gamma_g}^\lambda(k_1) P_\zeta(k_3) \frac{\partial \ln P_\zeta(k_3)}{\partial \ln k_3}. \quad (3.25)$$

As expected, the contribution from  $\mathcal{B}_{\mathfrak{m}}$  in (3.21) does indeed obey such rule. The squeezed limit contribution from  $\mathcal{B}_{\mathfrak{m}}$  in (3.22), however, does not: its expression depends on  $\mathcal{M}$  whilst both the tree level scalar and tensor power spectrum do not. The  $\mathcal{M}$  dependence in Eq. (3.22) is implicit in that the exponent of the very last term,  $k_1/k_3$ , can be thought of as proportional to  $3/2 - \sqrt{9/4 - \mathcal{M}^2/H^2}$ . We stress that this

<sup>12</sup>Note that the second term in the mode-function for  $\gamma_g$ , Eq. (3.11), generates terms scaling as positive powers of  $k\tau$  for  $k\tau \rightarrow 0$ , therefore they do not contribute to the power spectrum at late times.

is specific to the case  $\mathcal{M} = \sqrt{2}H$  and in contradistinction to the case where a similar (except for an angular dependence) suppression arises from a different diffeomorphism altogether in standard single-field inflation where CRs are conserved [89]. Note also that we have not included the contribution from  $\gamma_f$  to the late time tensor power spectrum. This is because this contribution is suppressed in light of  $\gamma_f$  not coupling directly to matter. Nevertheless, our conclusions would not change as the late-time limit of  $\gamma_f$  is unaware of  $\mathcal{M}$  and therefore cannot cancel off the  $\mathcal{B}_m$  bispectrum contribution in the squeezed limit. As we shall see more at length in *Section 4*, the small ratio  $k_1/k_3$  in Eq. (3.22) and the proportionality to  $r_{fg}$  play a key role in determining the detectability of the observational signal. As we have remarked above, a smaller  $\mathcal{M}$  would lead to a smaller exponent for  $k_1/k_3$  and thus a larger signal. We shall now elaborate on different strategies, within the same model, to enhance the signal.

### 3.4 All of the freedom in the unitarity bound

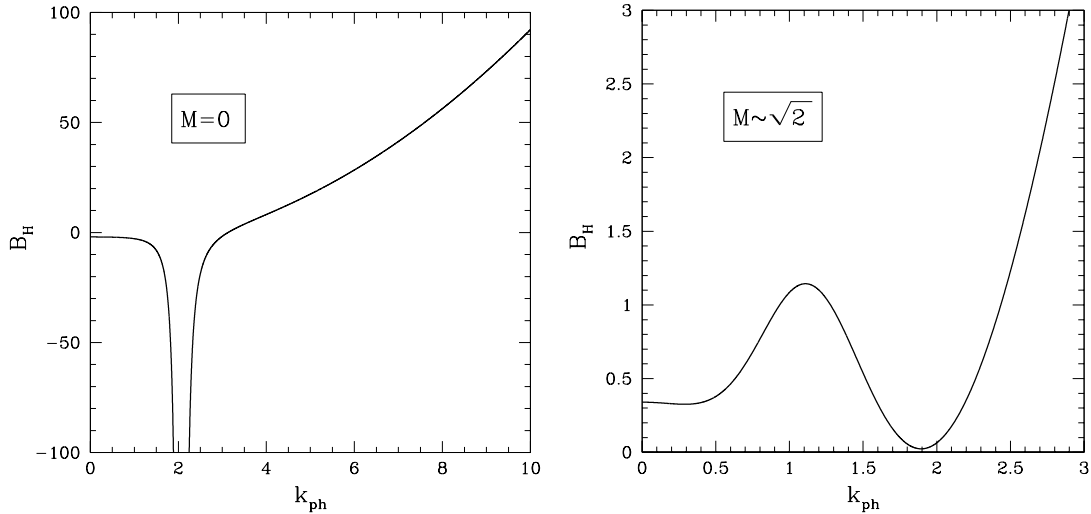
The Higuchi bound we inherit from the helicity-0 mode is strictly the one in Eq. (2.7) only in the low  $k$  limit, at or outside the horizon, in inflationary terms. From Ref. [18] we know that the scalar sector of the perturbations develops an instability (see Fig. 2) unless

$$\frac{4/3 k^6 + 6k^4 \mathcal{H}^2}{9a^2 \mathcal{H}^2 \mathcal{M}^2 + k^4 - 18\mathcal{H}^4} - \frac{12k^8 \mathcal{H}^2}{(9a^2 \mathcal{H}^2 \mathcal{M}^2 + k^4 - 18\mathcal{H}^4)^2} + a^2 \mathcal{M}^2 - \frac{k^2}{3} - 2\mathcal{H}^2 \geq 0, \quad (3.26)$$

where  $\mathcal{H} = aH$ . The above implies that for,  $k^2 > 8a^2 H^2$ , there is effectively no bound on  $\mathcal{M}$ . In other words, up to what corresponds approximately to the horizon for the long mode,  $\mathcal{M}$  can be much smaller than  $H^2$  so that the corresponding exponent of the ratio  $k_1/k_3$  would be much smaller and the corresponding signal in Eq. (3.22) enhanced. For this to be true we would need a time-dependent  $\mathcal{M}$ , something not allowed in de Sitter but possible in quasi-dS. The crucial point however is that, upon studying the  $k$  dependence of Eq. (3.26), one can see immediately that the inequality abruptly requires  $\mathcal{M} > 2H^2$  before we cross the  $k^2 = a^2 H^2$  threshold, with  $k$  being the long mode. This implies that a small and time-dependent  $\mathcal{M}$  would be in place only when short modes are still well inside their horizon and the long mode is about to exit. It is well-known that in this configuration the contribution from the integral in Eq. (3.15) is in fact suppressed. We conclude then that a suitably adjusted time-dependent  $\mathcal{M}$  can do little to improve the signal.

### 3.5 New cubic interactions

As stated, the  $k_1/k_3$  suppression in Eq. (3.22) is due to the non-zero  $\mathcal{M}$ -dependent term in the tensor wavefunction. Of course, simply considering the wavefunction massless component leads to the standard CR. The next logical step is to consider the massless component of the wavefunction and switch on new cubic interactions originating from the massive gravity potential  $\mathcal{E}(\sqrt{g^{-1}}f)$ . In light of the fact that these interactions diagram(s) will be inserted within the  $\langle \gamma \zeta \zeta \rangle$  three-point function, it is clear that we will be dealing with one-loop diagrams. Loops would typically lead to a  $(H/M_P)^2$



**Figure 1.** Left: the LHS of Eq. (3.26) as a function of the physical  $k$  for  $\mathcal{M} = 0$ . One can notice that a  $k_{\text{ph}} \geq \sqrt{8}H$  is enough for a positive  $B_H$ , regardless of the value of  $\mathcal{M}$ . Right: the LHS of Eq. (3.26) as a function of the physical  $k$  for  $\mathcal{M} \sim \sqrt{2}$ ; for small  $k$  the value of  $\mathcal{M}$  needs to surpass this threshold in order for the Higuchi bound to be satisfied.

suppression but in bigravity we have a little more freedom in the form of the  $M_P/M_f$  hierarchy.

To ensure non-standard CRs, we also need to verify that the resulting squeezed bispectrum cannot be expressed as in Eq. (3.25) by considering the corresponding possible one-loop contributions in  $P_\gamma$  or  $P_\zeta$ . The diagrams below should be considered as indicative of the typical behaviour in our setup. In view of the general result, it is not necessary to calculate the exact numerical value of the interactions and it shall suffice to give an estimate. We will also temporarily relax the assumption that matter couples only to the metric  $g$ , we do so relying on couplings such as those in Eq. (2.10). We resort here to direct coupling of  $f$  to matter because it allows us to construct diagrams involving the new spin-2 degrees of freedom, have  $\gamma_g$  and two  $\zeta$ s as external fields, and stay at one loop order. Two-loops would be necessary without direct coupling between  $f$  and matter. One possible diagram is shown in Fig. 2.



**Figure 2.** Diagrammatic representation of a typical 1-loop interaction: – the left vertex should be thought of as originating from a “new”  $m^2 M^2$ -type interaction; – the right vertex stems from standard-type direct minimal coupling of  $f$  with matter. The black (blue) wiggly line represents  $\gamma_g$  ( $\gamma_f$ ), the solid ones stand for  $\zeta$ .

The first vertex from the left should be thought of as originating from the new  $m^2 M^2$ -proportional interactions whilst the second vertex is “standard” whenever matter is



coupled directly to  $f$  as well. A quick estimate of the amplitude of this contribution returns

$$\mathcal{B}_m \sim \frac{H^4}{\epsilon_f M_P^4} \cdot \frac{H^2}{M_P^2} \quad (3.27)$$

where we have imposed a  $H \sim H_f$ ,  $M_f \ll M_P$  hierarchy. This result is to be compared with Eq. (3.22). As expected, employing the massless components of the wavefunctions guarantees the absence of the  $k_1/k_3$  factor. However, the price to pay is a Planck suppression squared. One can check that, regardless of the hierarchy, a suppression at least as strong as the typical loop scaling remains. Notice also that the outcome does not qualitatively change for the special configuration  $r_{fg} = 1$  discussed in Section 3.2.

## 4 Observational consequences

If a coupling exists between a long-wavelength tensor and two short-wavelength scalar fluctuations, different scalar Fourier modes will appear to be correlated with one another. These *off-diagonal* correlations can be employed to estimate the amplitude of the tensor fluctuation. Note that such an observable is only sensitive to sub-horizon tensor modes, as it is not possible to resolve scalar modes that differ by an amount smaller than  $H_0^{-1}$ . An optimal unbiased estimator for the tensor modes amplitude built from off-diagonal correlations was proposed in [72, 73], it reads:

$$\hat{A}_\gamma = \sigma_\gamma^2 \sum_{\vec{K}, \lambda} \frac{(P_f^\lambda(K))^2}{2(P_n^\lambda(K))^2} \left( \frac{|\hat{\gamma}^\lambda(\vec{K})|^2}{V} - P_n^\lambda(K) \right), \quad (4.1)$$

where  $V \equiv (2\pi/k_{\min})^3$  is the survey volume,  $P_f^\lambda \equiv P_\gamma^\lambda(K)/A_\gamma$  is a fiducial power spectrum for the tensor modes,  $\hat{\gamma}^\lambda(\vec{K})$  is the optimal estimator for a Fourier mode amplitude defined as

$$\hat{\gamma}^\lambda(\vec{K}) \equiv P_n^\lambda(K) \sum_{\vec{k}} \frac{B^\lambda(\vec{K}, \vec{k}, \vec{K} - \vec{k})/P_\gamma(K)}{2VP_{\text{tot}}(k)P_{\text{tot}}(|\vec{K} - \vec{k}|)} \delta(\vec{k})\delta(\vec{K} - \vec{k}). \quad (4.2)$$

In Eq. (4.2)  $P_n^\lambda$  stands for the variance

$$P_n^\lambda(K) \equiv \left[ \sum_{\vec{k}} \frac{|B^\lambda(\vec{K}, \vec{k}, \vec{K} - \vec{k})/P_\gamma(K)|^2}{2VP_{\text{tot}}(k)P_{\text{tot}}(|\vec{K} - \vec{k}|)} \right]^{-1}. \quad (4.3)$$

The quantity  $P_{\text{tot}}$  is the total scalar power spectrum (signal+noise) and  $B^\lambda$  is the effective (i.e. after subtraction of the consistency-conditions-preserving part) tensor-scalar-scalar bispectrum. The variance associated with (4.1) is given by

$$\sigma_\gamma^{-2} = \frac{1}{2} \sum_{\vec{K}, \lambda} [K^3 P_n^\lambda(K)]^{-2}, \quad (4.4)$$

and can be used to quantify, for a given survey size and for given model parameters, the smallest detectable tensor-mode amplitude.

Using the expression we derived in Eq. (3.22) for the non-trivial contribution to the bispectrum in the squeezed limit, one finds

$$(P_n^\lambda(K))^{-1} \approx \frac{\beta^2}{\pi^2 15} K^2 k_{\max}, \quad (4.5)$$

where  $k_{\max}$  is the maximum wave-number accessed with a given survey and the bispectrum amplitude  $\beta$  in the squeezed limit has been defined, from Eq. (3.22), via  $\|\mathcal{B}_m\| \equiv \beta P_\gamma^\lambda(K) P_\zeta(k) \frac{K}{k}$ , hence  $\beta = \mathcal{O}(1) \times [r_{fg}|(1-r_{fg})|]/(1+r_{fg})$ . The variance for the tensor power spectrum amplitude is then found to be

$$\sigma_\gamma^{-1} = \frac{2\beta^2}{15\pi\sqrt{\pi}} \left( \frac{k_{\max}}{k_{\min}} \right)^{3/2}. \quad (4.6)$$

From Eq. (4.6) it is clear that, for an amplitude  $\mathcal{A}_\gamma$  close to the current observational bounds (i.e.  $10^{-10}$ ) to be within reach of upcoming 21-cm surveys, one would need  $\beta \gtrsim \mathcal{O}(10^3)$ , assuming  $k_{\max}/k_{\min} \sim 5000$  (see e.g. [46]). In our configuration  $r_{fg} \ll 1$ , therefore given the above definition for  $\beta$  we conclude that the amplitude of the tensor-scalar-scalar correlation predicted by this model is too weak for detection by upcoming large scale structure surveys. Coupling  $f$  directly to matter as well, via the composite metric, does allow for  $r_{fg} \gg 1$ . However the consistency/stability requirements comprising background equations, the Higuchi bound, and the fact that it is the matter sector that ought to drive inflation, will also change the definition of  $\beta$  so that a high  $r_{fg}$  will not enhance the signal in the way the above expression of  $\beta$  suggests.

A non-trivial squeezed-limit tensor-scalar-scalar bispectrum introduces also a local quadrupolar anisotropy in the scalar power spectrum which, unlike the off-diagonal correlation, is sensitive to super-horizon tensor modes [74]

$$P_\zeta(\vec{k})|_{\gamma^\lambda(\vec{K})} = P_\zeta(k) \left[ 1 + \mathcal{Q}_{ij}^\lambda(\vec{K}) \hat{k}_i \hat{k}_j \right], \quad (4.7)$$

where

$$\mathcal{Q}_{ij}^\lambda(\vec{K}) \equiv \frac{\mathcal{B}(K, k, |\vec{K} - \vec{k}|)}{P_\gamma(K) P_\zeta(k)} \gamma_{ij}^\lambda(\vec{K}). \quad (4.8)$$

The variance of the quadrupole reads

$$\overline{\mathcal{Q}^2} \equiv \frac{8\pi}{15} \langle \mathcal{Q}_{ij}^\lambda \mathcal{Q}^{\lambda,ij} \rangle = \frac{16}{15\pi} \int_{k_{\min}}^{k_{\max}} K^2 dK \left[ \frac{\mathcal{B}(K, k, |\vec{K} - \vec{k}|)}{P_\gamma(K) P_\zeta(k)} \right]^2 P_\gamma(K). \quad (4.9)$$

where  $k_{\min}$  is the longest-wavelength tensor mode produced during inflation and  $k_{\min}$  is the smallest observable wavenumber. Replacing the results for our model, Eqs. (3.22), (3.23) and (3.24), one arrives at

$$\overline{\mathcal{Q}^2} \sim \frac{1}{15\pi} \left( \frac{H}{M_P} \right)^2 \frac{r_{fg}^2 (1-r_{fg})^2}{(1+r_{fg}^2)^2} \left( \frac{k_{\min}}{k} \right)^2 \lesssim \left( \frac{H}{M_P} \right)^2 \frac{r_{fg}^2 (1-r_{fg})^2}{(1+r_{fg}^2)^2}. \quad (4.10)$$

Given the current limits on the quadrupolar anisotropy,  $\mathcal{Q} \lesssim 10^{-2}$  [75–82], Eq. (4.10) shows that the model allows for ample room for those to be satisfied.

## 5 Conclusions

Surveying the possible signatures of the inflationary particle content is crucial to make the best possible use of this cosmological window on high energy physics. The squeezed limit (and generalizations thereof) of  $N \geq 3$  correlation functions is particularly sensitive a probe to the presence of “new” physics beyond the minimal single-field inflationary scenario. Indeed, the relation between the soft limit of  $N + 1$ -point correlators and their  $N$ -point counterpart may be modified as a result of the additional dynamics and reveal information about the mass and spin [1, 2, 72] of the extra degrees of freedom. In this work we have studied the case of the minimal inflationary scenario, GR + scalar, coupled through so-called dRGT-interactions to a (massive) spin-2 field. This is a non-trivial yet economic choice going up the particle spin ladder as we capitalize on the vast existing literature on massive spin-2 fields in the context of late-time cosmic acceleration. We employ a model which is ghost-free at the fully non-linear level and show how consistency relations are modified in this setup.

The property that is most consequential for the purposes of detection turns out to be the fact that massive unitary representations of the dS group come with a lower bound on the effective mass of the graviton  $m^2 \geq 2H^2$ . Particles whose Compton wavelength is smaller than  $1/H$  will typically decay outside the horizon. This reflects on the impact that modified consistency relations have on observables in our setup: the contribution of each mode comes only from an integrated-over-time effect. The resulting signal for the tensor-scalar-scalar correlation is too weak for detection by upcoming large scale structure surveys. One can exploit the fact that the unitarity condition is softened in FLRW backgrounds and matter may be coupled to both spin-2 fields. However, demanding consistency of the overall setup leads to similar conclusions.

In order to rescue the massive tensor from decay, the most natural mechanism is that of a non-minimal coupling to the matter sector [84], where special care needs to be exerted so as not to excite ghostly degrees of freedom both at the linear and fully non-linear level. Many of these features characterize also the intriguing case of higher spin fields which will be the subject of future work.

## Acknowledgments

We are delighted to thank Alex Kehagias and Toni Riotto for collaboration at the early stages of this work. We thank especially T. Riotto without whose input this paper would not have been possible. M.B. thanks Daniel Baumann, Garrett Goon, Hayden Lee and Guilherme Pimentel for helpful discussions and comments and Luigi Pilo for useful correspondence. E.D. and M.F. are grateful to Kurt Hinterbichler for fruitful conversation and especially to Claudia de Rham and Andrew J. Tolley for illuminating discussions and comments. M.B. acknowledges support from Delta ITP consortium, a program of the Netherlands Organisation for Scientific Research (NWO)

that is funded by the Dutch Ministry of Education, Culture and Science (OCW). E.D. acknowledges support by DOE grant DE-SC0009946. M.F. is supported in part by NSF PHY-1068380.

## A Consistency relations: general considerations

In Appendix A and B we will adopt the notation of [24, 25]. The key requirement for consistency relations to be in place between an  $N + 1$ -function with at least one long mode and its  $N$ -point counterpart is the existence of gauge diffeomorphisms (diff) in the action (or the equations of motion) of the physical system at hand. In cosmological setups (e.g. spatially-flat FLRW background) it is often said that certain gauges, for example unitary gauge, completely fix gauge invariance. This is indeed true for diffeomorphisms that vanish at spatial infinity. Crucially, the residual gauge symmetries that CRs rely upon do not vanish at infinity [41]. It can be shown [25] that these transformations can be smoothly extended to configurations that do fall off at infinity and therefore corresponds to the action of adiabatic modes. The key identification is between these gauge transformations and the soft limit of transformations that fall off at infinity and are therefore physical. In the inflationary context, depending on the specific gauge diff in question, the transformation of a given observable  $\mathcal{O}$  can be identified with the action of a long scalar or tensor mode on the same quantity  $\mathcal{O}$ .

For the CR to be non-trivial it is essential that the long mode transforms non linearly under the symmetry. All CRs originate from the invariance of the action under specific residual space diffs:

$$x^i \rightarrow x^i + \bar{\xi}^i(x) \tag{A.1}$$

where the  $\bar{\phantom{\xi}}$  on top of  $\xi$  indicates that the gauge mode  $\xi$  must satisfy equation of motions similar to that of physical modes. Let us consider the example of dilations, which are defined as

$$\mathbf{x} \mapsto (1 + \lambda)\mathbf{x}, \tag{A.2}$$

$$\zeta \mapsto \zeta + \lambda(1 + \mathbf{x} \cdot \partial_{\mathbf{x}}\zeta). \tag{A.3}$$

This symmetry of the action is associated with a conserved current  $J_d^\mu$  and a corresponding charge  $Q_d = \int d^3x J_d^0$ , such that locally we can write the transformation of the curvature fluctuation under dilations as

$$\delta_d \zeta = i[Q_d, \zeta] = -1 - \mathbf{x} \cdot \partial_{\mathbf{x}}\zeta. \tag{A.4}$$

A similar relation holds for the tensor fluctuation  $\gamma$  under anisotropic spatial rescaling,

$$\bar{\xi}_{\text{ar}}^i = S_{il} x^l \quad \text{with} \quad S_{il} = S_{li}; \quad S_{ii} = 0 = \hat{q}^i S_{il}(\vec{q}). \tag{A.5}$$

Let us very schematically track the action of a gauge diff  $s$  on a generic observable  $\langle \mathcal{O} \rangle = \langle \zeta_{1..l}, \gamma_{l+1..m}, ..\sigma_N \rangle$ , which we take to be the  $N$ -point function made up by

any combinations of fields populating the Lagrangian and consider  $\sigma$  as a placeholder for any type of particle content. The transformation  $s$  acts according to

$$\delta_s \langle \mathcal{O} \rangle \Big|_{\text{connected}} \propto \mathcal{D} \cdot \langle \mathcal{O} \rangle , \quad (\text{A.6})$$

where we have been deliberately agnostic about the rules of the  $s$  transformation except for indicating by  $\mathcal{D} \cdot$  a generic differential operator. Crucially, only the linear component of  $\delta_s$  acting on the fields is relevant for connected diagrams<sup>13</sup>. The very same action of the  $s$  diff can be expressed in an alternative but equivalent fashion and it is the equivalence of the two that sits at the heart of consistency relations. The effect of the  $\delta_s$  transformation is associated to the conserved current, and consequently to its corresponding charge so that  $\delta_s \langle \mathcal{O} \rangle \sim \langle Q_s | \mathcal{O} \rangle$ . At this stage we are certainly at liberty to introduce the identity operator in between

$$\langle Q_s | \mathcal{O} \rangle = \sum_n \langle Q_s | n \rangle \langle n | \mathcal{O} \rangle , \quad (\text{A.7})$$

where the  $|n\rangle$  represent mutually orthogonal independent states. Setting up this basis requires no work in the minimal inflationary scenario as naturally  $\langle \zeta \gamma \rangle = 0$ . In the case of anisotropic rescaling, the tensor field transformation has both a linear and a non-linear component while the scalar mode only the former. It follows that  $Q_{\text{ar}}$  acts non-trivially only on  $\gamma$  :

$$\langle Q_{\text{ar}} | \gamma \rangle \sim \langle \gamma \rangle + c_{\text{number}} , \quad (\text{A.8})$$

the first term on the RHS being zero<sup>14</sup>. With no additional field content present, it follows that:

$$\mathcal{D} \cdot \langle \mathcal{O} \rangle \propto \langle Q_{\text{ar}} | \gamma \rangle \langle \gamma | \mathcal{O} \rangle \quad (\text{A.9})$$

or, rather less schematically, the CR reads:

$$-\frac{1}{2} \epsilon_{ij}^\lambda(\hat{k}_1) \sum_{a=2}^3 \left( k_a^i \frac{\partial}{\partial k_a^j} \langle \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle'_c \right) = \lim_{\mathbf{k}_1 \rightarrow 0} \frac{\langle \gamma_{\mathbf{k}_1}^\lambda \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle'_c}{P_\gamma(\mathbf{k}_1)} \quad (\text{A.10})$$

where “ $c$ ” stands for the connected part of the diagrams and the prime symbol is a reminder that the momentum-conserving delta has been removed. Albeit admittedly very schematic, the above account is sufficient to see that one obvious route to non-standard CRs is the presence of an additional field whose component *orthogonal* to the

<sup>13</sup>Note here that the action of the linear component of the  $\delta_s$  transformation cannot always, strictly speaking, be put in the clean form of the RHS of Eq. (A.6), see e.g. [25]. For example, the linear transformation of the curvature fluctuation  $\zeta$  under a generic diff comprises a term proportional to the tensor mode  $\gamma_{ij}$ . However, this technical consideration is in no way relevant to the discussion in this section and the simplified form is indeed in place for the specific case of anisotropic rescaling acting on two hard scalar modes.

<sup>14</sup>The ket  $|\gamma\rangle$  describes the action of the operator  $\hat{\gamma}_{ij}$  on the Bunch-Davies vacuum of the theory. For a more general and detailed treatment see, for example, the work in [25].

$\{|\gamma\rangle, |\zeta\rangle\}$  basis transforms non-linearly under anisotropic rescaling and whose  $N + 1$ -point function with  $\mathcal{O}$  is non-trivial<sup>15</sup>:

$$\mathcal{D} \cdot \langle \mathcal{O} \rangle \propto \langle Q_{\text{ar}} | \gamma \rangle \langle \gamma | \mathcal{O} \rangle + \langle Q_{\text{ar}} | \sigma^\perp \rangle \langle \sigma^\perp | \mathcal{O} \rangle, \quad (\text{A.11})$$

where  $\sigma$  stands here for a generic new field. This will be precisely the case of the additional (traceless transverse component of) tensor field that will populate our non-minimal inflationary setup. Of course, there are other ways to non-standard or even broken consistency relations: one option is to have a different symmetry breaking pattern or no residual diffs altogether [42–50]; another possibility is to have excited initial states [51]. The reader might wonder how would our brief discussion change in the latter case. In the above we have been rather casual about defining the free fields eigenstates  $|\zeta\rangle, |\gamma\rangle$  and the implicit vacuum wavefunction; the definition is modified in the case of non Bunch-Davies initial conditions (see [25] for a more thorough discussion on this and e.g. [52, 53] for examples of excited initial states wavefunctionals).

## B Alternative route to CRs breaking

In this section we shall work under the approximation that the quadratic Lagrangian (Hamiltonian) of the model, Eq. (2.1), can be written down as  $\mathcal{L}_{\text{full}}^{(2)} = \mathcal{L}_0^{(2)} + \delta\mathcal{L}_{f-g}^{(2)}$  where the last piece, even though quadratic, is handled just like an interaction in the in-in formalism, with its own interaction vertex. This treatment is allowed so long as  $\delta\mathcal{L}_{f-g}^{(2)} \ll \mathcal{L}_0^{(2)}$  and we refer the reader to [9] for a well-studied example within the quasi-single-field inflation paradigm. This approximation is the basis for an “alternative route” to CRs in our setup which, we stress it here, will be mostly qualitative.

We anticipate that treating the  $f - g$  coupling in the quadratic Lagrangian as an interaction is an approximation that is no longer valid when the long mode crosses the horizon. In the approach outlined in Section 3, one solves for the *fully* coupled equations in the quadratic Lagrangian so that all of the information on the (quadratic)  $f - g$  coupling is stored in the mode-functions.

The quadratic action for the traceless transverse part of the tensors  $h_f, h_g$  is schematically as follows:

$$\mathcal{L}^{(2)} \sim M_P^2 (\partial\gamma_g)^2 + M_f^2 (\partial\gamma_f)^2 + m^2 M^2 \Gamma(H/H_f) (\gamma_g^2 + \gamma_g \gamma_f + \gamma_f^2) \quad (\text{B.1})$$

where  $\Gamma$  is a specific function of the  $\beta_n$ s (one can think of them as order one) and the  $H/H_f$  ratio. The latter is not too relevant here as the  $M_f/M_P$  ratio will play the main role. We want to show that the  $\gamma_g - \gamma_f$  mixing can be treated as a perturbation on top of the remaining free Lagrangian. To this end, we recall that the usual normalization factor for the wavefunctions is  $\gamma_g \propto \frac{H}{M_P}$  and  $\gamma_f \propto \frac{H_f}{M_f}$ , so that the Einstein-Hilbert contributions to the quadratic Lagrangian go respectively like  $H^4, H_f^4$  and we take the former as being of the same order or smaller than the latter. It follows that the

---

<sup>15</sup>What we mean by non-trivial here is rather more than requiring a non-zero  $N + 1$ -point function, as will be clarified below.

condition for each term proportional to  $M^2$  in Eq. (B.1) to be smaller than the E-H terms is, in order:

$$\frac{m^2}{H^2} \frac{M^2}{M_P^2} \Gamma \ll 1, \quad \frac{m^2}{H^2} \frac{H_f}{H} \frac{M^2}{M_f M_P} \Gamma \ll 1, \quad \frac{m^2}{H^2} \frac{H_f^2}{H^2} \frac{M^2}{M_f^2} \Gamma \ll 1, \quad (\text{B.2})$$

the most relevant being of course the second inequality because it stems from requiring that the  $\gamma_g - \gamma_f$  mixing term can be treated as an interaction. It is important to check at this stage the compatibility of the second inequality<sup>16</sup> with the Higuchi bound as well as the Friedman equations. For convenience, we report the Higuchi bound as in Eq. (2.7) and assume the term regulated by  $\beta_1$  to be the leading one. The bound to be compared with the conditions in Eq. (B.2) reads:

$$\frac{1}{4} \frac{m^2}{H^2} \frac{H_f}{H} \frac{M^2}{M_f^2} \geq 1. \quad (\text{B.3})$$

Upon inspection, it is immediate that the first two requirements in Eq. (B.2) are satisfied if the hierarchy  $M_f \ll M_P$  is in place<sup>17</sup>; we shall assume it for the remaining parts of the manuscript<sup>18</sup>. This in particular implies that the  $\gamma_g - \gamma_f$  mixing can be treated as an interaction (see Fig. 3) without conflicting with the all-important unitarity bound. For consistency, we shall also require that the leading part in the RHS of Eq. (3.2) is also due to the inflationary potential and not the  $m^2$ -proportional part. To that end, and again focussing the attention on the  $\propto \beta_1$  contribution, we ought to require that:

$$\frac{m^2}{H^2} \frac{H}{H_f} \frac{M^2}{M_P^2} \ll 1, \quad (\text{B.4})$$

which is compatible with all of the above. Finally, Eq. (3.3) requires

$$\frac{m^2}{H^2} \frac{H_f}{H} \frac{M^2}{M_f^2} \sim 1, \quad (\text{B.5})$$

also consistent. Reassured by these checks, let us turn to consistency relations in this approximation scheme.

We want to provide evidence that tensor CRs in our setup take a non standard form. We have reviewed in *Section A* the origin of the CRs relevant for observables such as the tensor-scalar-scalar three-point function. We are interested in the action of anisotropic

---

<sup>16</sup>As we shall see, the first one is a byproduct of the second one in our parameter space and we will not need the third inequality to be satisfied at all.

<sup>17</sup>We must stress here that, at this stage, this is not a necessary assumption. However, the combination of Higuchi bound, the requirement on the Friedman equation to be dominated by matter driving inflation, and having a slowly varying  $\xi$ , leads to  $M_f \ll M_P$ . See [62, 71] for further details.

<sup>18</sup>The third requirement in Eq. (B.2) is not necessarily satisfied: the ratio can be order one and translates into a massive term for  $\gamma_f$  that must be contained within  $\mathcal{L}_0^{(2)}$  and is not a part of  $\delta\mathcal{L}^{(2)}$ . The same ratio appears e.g. in Eq. (B.5).





**Figure 3.** Diagrammatic representation of the quadratic  $\gamma_g$ - $\gamma_f$  interaction.

rescaling on long tensor modes defined in Eq. (A.5). Recall now from Eq. (A.11) that a non-standard CR demands at least two terms in the RHS of

$$\langle Q_{\text{ar}}|\zeta\zeta\rangle = \sum_n \langle Q_{\text{ar}}|n\rangle \langle n|\zeta\zeta\rangle. \quad (\text{B.6})$$

If at least two fields (or rather two states)  $\{|1\rangle \equiv |\gamma_g\rangle, |2\rangle, \dots\}$ , once orthonormalized, transform non-linearly under  $Q_{\text{ar}}$  then we will have a non-standard CR:

$$\langle Q_{\text{ar}}|\gamma_g\rangle \langle \gamma_g|\zeta\zeta\rangle + \langle Q_{\text{ar}}|2\rangle \langle 2|\zeta\zeta\rangle + \dots \sim \langle [Q_{\text{ar}}, \zeta\zeta]\rangle \sim \mathcal{D}\langle \zeta\zeta\rangle \quad (\text{B.7})$$

In standard single-field inflation the only field non-linearly transforming under an anisotropic rescaling is  $\gamma_g$ , hence the standard CR. In the setup of Eq. (2.1), can  $\gamma_f$ , the traceless and transverse part of the fluctuation of the  $f$  metric, suitably orthogonalized, serve as a non zero contribution from the  $|2\rangle$  “state”? To answer this question in the affirmative we ought to show that

$$0 \neq \langle Q_{\text{ar}}|\gamma_f^\perp\rangle \langle \gamma_f^\perp|\zeta\zeta\rangle \neq \langle Q_{\text{ar}}|\gamma_g\rangle \langle \gamma_g|\zeta\zeta\rangle, \quad (\text{B.8})$$

the last condition guaranteeing that the action of the two long modes  $\gamma_f$  and  $\gamma_g$  is not identical. Note also that we have orthogonalized  $\gamma_f$  w.r.t.  $\gamma_g$  in the standard way:

$$|\gamma_f^\perp\rangle = N \left[ |\gamma_f\rangle - |\gamma_g\rangle \langle \gamma_g|\gamma_f\rangle \right], \quad (\text{B.9})$$

with  $N$  a normalization factor. The RHS of Eq. (B.7) can now be written more explicitly as

$$\begin{aligned} \sum_n \langle Q_{\text{ar}}|n\rangle \langle n|\zeta\zeta\rangle &= \langle Q_{\text{ar}}|\gamma_g\rangle \langle \gamma_g|\zeta\zeta\rangle + \langle Q_{\text{ar}}|\gamma_f^\perp\rangle \langle \gamma_f^\perp|\zeta\zeta\rangle = \\ &c_g \langle \gamma_g|\zeta\zeta\rangle + \langle Q_{\text{ar}}|N \left[ |\gamma_f\rangle - |\gamma_g\rangle \langle \gamma_g|\gamma_f\rangle \right] \cdot N^* \left[ \langle \gamma_f| - \langle \gamma_f|\gamma_g\rangle \langle \gamma_g| \right] |\zeta\zeta\rangle = \\ &c_g \langle \gamma_g|\zeta\zeta\rangle + |N|^2 \left[ (c_f - c_g I_{gf}) \cdot (\langle \gamma_f|\zeta\zeta\rangle - I_{gf}^* \langle \gamma_g|\zeta\zeta\rangle) \right], \end{aligned} \quad (\text{B.10})$$

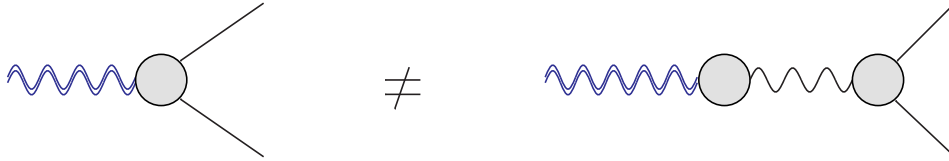
where we have assumed for simplicity that  $|\gamma\rangle$  has norm one. We have also defined the quantities  $c_g \equiv \langle Q_{\text{ar}}|\gamma_g\rangle$ ,  $c_f \equiv \langle Q_{\text{ar}}|\gamma_f\rangle$ , and  $I_{gf} \equiv \langle \gamma_g|\gamma_f\rangle$ . Note that the latter stems precisely from the interaction in Fig. 3. Non-standard CRs now correspond to both factors between square brackets in the last line of Eq. (B.10) being non-zero. In order to show that is indeed the case, let us write down the gauge transformation properties

of  $\gamma_f, \gamma_g$ . These, in our language, are stored into  $c_f, c_g$ . For the gauge parameter,  $\xi^\mu$ , the transformation law reads:

$$\begin{aligned}\delta\gamma_{g\mu\nu} &= a^{-2} (\xi^\alpha \partial_\alpha \bar{g}_{\mu\nu} + \bar{g}_{\alpha\nu} \partial_\mu \xi^\alpha + \bar{g}_{\mu\alpha} \partial_\nu \xi^\alpha) , \\ \delta\gamma_{f\mu\nu} &= b^{-2} (\xi^\alpha \partial_\alpha \bar{f}_{\mu\nu} + \bar{f}_{\alpha\nu} \partial_\mu \xi^\alpha + \bar{f}_{\mu\alpha} \partial_\nu \xi^\alpha) .\end{aligned}\tag{B.11}$$

The reasoning proceeds as follows. The gauge transformation of  $\gamma_f, \gamma_g$  are the same except for a factor of  $b/a$ , the ratio of the two metrics scale factors, a quantity also equal to  $H/H_f$  in the branch of the theory we are considering. From here, one should in principle solve the constraint equations and derive the transformation law specifically for the symmetric traceless transverse part of  $\gamma_f, \gamma_g$ . On the other hand, what matters for us is that the following term from Eq. (B.10) is non zero:  $c_f - c_g I_{gf} \neq 0$ . It is immediate to see, for example from the function  $\Gamma$  in Eq. (B.1), that the quadratic interaction term  $I_{gf}$  contains different powers of  $b/a = H/H_f$  and is also a function of  $\beta_1, \beta_2, \beta_3$  so that  $c_f - c_g I_{gf}$  does not in general vanish.

What remains to be proven now is that  $\langle \gamma_f | \zeta \zeta \rangle - I_{gf}^* \langle \gamma_g | \zeta \zeta \rangle \neq 0$  which is, written differently,  $\langle \gamma_f | \zeta \zeta \rangle \neq \langle \gamma_f | \gamma_g \rangle \langle \gamma_g | \zeta \zeta \rangle$ . Perhaps the best way to prove this last part is diagrammatically, as shown in Fig. 4, where the grey colored circles stand for generic interactions. It will be enough to find an internal<sup>19</sup> diagram corresponding to  $\langle \gamma_f | \zeta \zeta \rangle$  that cannot be expressed as  $\langle \gamma_f | \gamma_g \rangle \langle \gamma_g | \zeta \zeta \rangle$ . Let us give a few examples of interactions



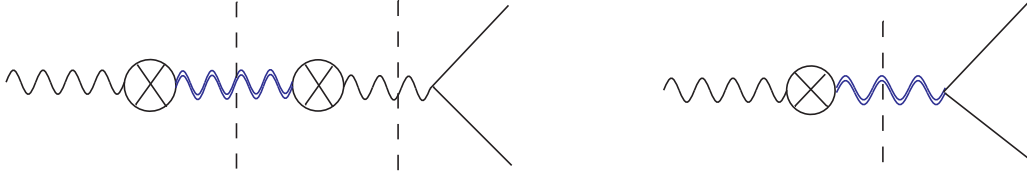
**Figure 4.** Diagrammatic representation non standard CRs condition.

that do and do not lead to a modified CR. On the left side of Fig. 5 one can see a diagram that can be drawn as in the RHS of Fig. 4 with dashed vertical lines just to the left of where the gray vertices would go.

On the other hand, the diagram on the right side of Fig. 5 cannot be put in that form and leads therefore to a non-standard consistency relation.

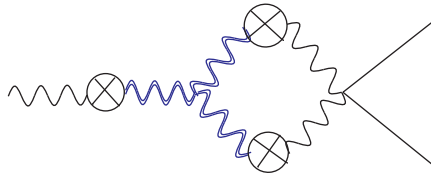
However, it must be noted that this requires a direct coupling between  $\gamma_f$  and  $\zeta$ . The diagram in Fig. 6 is instead a loop effect and delivers a non-standard CR regardless of the coupling of  $\gamma_f$ . We consider this last one as the clearest example of a modified consistency relation among those we present in this *Section B*. As anticipated, we will not discuss in detail the value of the bispectrum generated by these diagrams, we postpone a quantitative analysis of modified CRs to *Section 3.2*. One reason, as we shall see shortly, is that the approximation scheme justifying the use of  $\delta\mathcal{L}^{(2)}$ , or in other words the diagram in Fig. 3, breaks down at the horizon. Nevertheless, we find it worthwhile to present a diagrammatic proof of CRs “breaking” in this setup. The existence of a quadratic interaction lends itself nicely to such proof, it shows very

<sup>19</sup>In this approximation  $\gamma_f$  is massive and will therefore decay at late times.



**Figure 5.** Left: contribution to  $\langle \gamma_g \zeta \zeta \rangle$  consisting of two “quadratic”  $\gamma_g - \gamma_f$  vertices and the usual tree-level tensor-scalar-scalar interaction. Right: A  $\gamma_g - \gamma_f$  vertex and a three-vertex which is there only if  $\gamma_f$  couples directly with  $\zeta$ .

clearly why CRs are indeed broken, and is very reminiscent of the one that applies to, among others, quasi-single-field inflation models.



**Figure 6.** A loop-diagram contribution to  $\langle \gamma_g \zeta \zeta \rangle$ . This diagram will be “Planck” suppressed by the  $H/M_p$  and  $H_f/M_f$  in the wavefunctions normalizations.

Let us briefly show how the approximation breaks down at the horizon. The reason is quite simple, the quadratic interaction term is proportional to  $m^2 a(\tau)^2 h^2$ . In a quasi-de Sitter phase it is true that  $a^2 \simeq -\frac{1}{H^2 \tau^2}$ . At late times (starting at horizon exit) this term becomes of the same order and eventually much larger than the Einstein-Hilbert terms we have placed in  $\mathcal{L}_0^{(2)}$ , thereby invalidating the approximation. Considering that most of the contribution to cosmological correlation functions is known to originate at and after horizon exit it is advisable to seek another scheme whose validity extends further than this “alternative route”. We provided such a treatment in [Section 3.2](#).

## References

- [1] A. Kehagias and A. Riotto, *Fortsch. Phys.* **63**, 531 (2015) [[arXiv:1501.03515](#)].
- [2] N. Arkani-Hamed and J. Maldacena, [[arXiv:1503.08043](#)].
- [3] E. Dimastrogiovanni, M. Fasiello and M. Kamionkowski, *JCAP* **1602**, 017 (2016) [[arXiv:1504.05993](#)].
- [4] H. Lee, D. Baumann and G. L. Pimentel, *JHEP* **1612**, 040 (2016) [[arXiv:1607.03735](#)].
- [5] D. Baumann and L. McAllister, [[arXiv:1404.2601](#)].
- [6] A. Achúcarro, J. O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, *JHEP* **1205** (2012) 066 [[arXiv:1201.6342](#)].
- [7] C. P. Burgess, M. W. Horbatsch and S. P. Patil, *JHEP* **1301** (2013) 133 [[arXiv:1209.5701](#)].

- [8] E. Silverstein, [[arXiv:1706.02790](#)].
- [9] X. Chen and Y. Wang, Phys. Rev. D **81**, 063511 (2010) [[arXiv:0909.0496](#)].
- [10] X. Chen and Y. Wang, JCAP **1004**, 027 (2010) [[arXiv:0911.3380](#)].
- [11] N. Craig and D. Green, JHEP **1407**, 102 (2014) [[arXiv:1403.7193](#)].
- [12] T. Noumi, M. Yamaguchi and D. Yokoyama, JHEP **1306**, 051 (2013) [[arXiv:1211.1624](#)].
- [13] E. Sefusatti, J. R. Fergusson, X. Chen and E. P. S. Shellard, JCAP **1208**, 033 (2012) [[arXiv:1204.6318](#)].
- [14] L. McAllister, S. Renaux-Petel and G. Xu, JCAP **1210**, 046 (2012) [[arXiv:1207.0317](#)].
- [15] D. Baumann and D. Green, Phys. Rev. D **85**, 103520 (2012) [[arXiv:1109.0292](#)].
- [16] A. Higuchi, Nucl. Phys. B **282**, 397 (1987).
- [17] D. Blas, Int. J. Theor. Phys. **46**, 2258 (2007).
- [18] D. Comelli, M. Crisostomi and L. Pilo, JHEP **1206**, 085 (2012) [[arXiv:1202.1986](#)].
- [19] M. Fasiello and A. J. Tolley, JCAP **1211**, 035 (2012) [[arXiv:1206.3852](#)].
- [20] M. Fasiello and A. J. Tolley, JCAP **1312**, 002 (2013) [[arXiv:1308.1647](#)].
- [21] P. A. R. Ade *et al.* [Planck Collaboration], Astron. Astrophys. **594**, A17 (2016) [[arXiv:1502.01592](#)].
- [22] J. M. Maldacena, JHEP **0305**, 013 (2003) [[arXiv:0210603](#)].
- [23] P. Creminelli and M. Zaldarriaga, JCAP **0410**, 006 (2004) [[arXiv:0407059](#)].
- [24] V. Assassi, D. Baumann and D. Green, JCAP **1211**, 047 (2012) [[arXiv:1204.4207](#)].
- [25] K. Hinterbichler, L. Hui and J. Khoury, JCAP **1401**, 039 (2014) [[arXiv:1304.5527](#)].
- [26] T. Tanaka and Y. Urakawa, [[arXiv:1707.05485](#)].
- [27] A. Kehagias and A. Riotto, Nucl. Phys. B **873**, 514 (2013) [[arXiv:1302.0130](#)].
- [28] M. Peloso and M. Pietroni, JCAP **1305**, 031 (2013) [[arXiv:1302.0223](#)].
- [29] A. Kehagias, J. Noreña, H. Perrier and A. Riotto, Nucl. Phys. B **883**, 83 (2014) [[arXiv:1311.0786](#)].
- [30] P. Creminelli, J. Noreña, M. Simonović and F. Vernizzi, JCAP **1312**, 025 (2013) [[arXiv:1309.3557](#)].
- [31] A. Kehagias, H. Perrier and A. Riotto, Mod. Phys. Lett. A **29** (2014) 1450152 [[arXiv:1311.5524](#)].
- [32] P. Creminelli, J. Gleyzes, L. Hui, M. Simonović and F. Vernizzi, JCAP **1406**, 009 (2014) [[arXiv:1312.6074](#)].
- [33] B. Horn, L. Hui and X. Xiao, JCAP **1409**, no. 09, 044 (2014) [[arXiv:1406.0842](#)].
- [34] A. Kehagias, A. M. Dizgah, J. Noreña, H. Perrier and A. Riotto, JCAP **1508** (2015) no.08, 018 [[arXiv:1503.04467](#)].
- [35] E. Di Dio, H. Perrier, R. Durrer, G. Marozzi, A. M. Dizgah, J. Noreña and A. Riotto, JCAP **1703**, no. 03, 006 (2017) [[arXiv:1611.03720](#)].

- [36] M. Fasiello and Z. Vlah, [[arXiv:1611.00542](#)].
- [37] B. Bonga, S. Brahma, A. S. Deutsch and S. Shandera, JCAP **1605**, no. 05, 018 (2016) [[arXiv:1512.05365](#)].
- [38] C. de Rham and G. Gabadadze, Phys. Rev. D **82**, 044020 (2010) [[arXiv:1007.0443](#)].
- [39] C. de Rham, G. Gabadadze and A. J. Tolley, Phys. Rev. Lett. **106**, 231101 (2011) [[arXiv:1011.1232](#)].
- [40] S. F. Hassan, R. A. Rosen and A. Schmidt-May, JHEP **1202**, 026 (2012) [[arXiv:1109.3230](#)].
- [41] S. Weinberg, Phys. Rev. D **67**, 123504 (2003) [[arXiv:0302326](#)].
- [42] S. Endlich, A. Nicolis and J. Wang, JCAP **1310**, 011 (2013) [[arXiv:1210.0569](#)].
- [43] S. Endlich and A. Nicolis, [[arXiv:1303.3289](#)].
- [44] N. Bartolo, S. Matarrese, M. Peloso and A. Ricciardone, JCAP **1308**, 022 (2013) [[arXiv:1306.4160](#)].
- [45] S. Endlich, B. Horn, A. Nicolis and J. Wang, Phys. Rev. D **90**, no. 6, 063506 (2014) [[arXiv:1307.8114](#)].
- [46] E. Dimastrogiovanni, M. Fasiello, D. Jeong and M. Kamionkowski, JCAP **1412**, 050 (2014) [[arXiv:1407.8204](#)].
- [47] M. Akhshik, JCAP **1505**, no. 05, 043 (2015) [[arXiv:1409.3004](#)].
- [48] D. Cannone, J. O. Gong and G. Tasinato, JCAP **1508**, no. 08, 003 (2015) [[arXiv:1505.05773](#)].
- [49] N. Bartolo, D. Cannone, A. Ricciardone and G. Tasinato, JCAP **1603**, no. 03, 044 (2016) [[arXiv:1511.07414](#)].
- [50] A. Ricciardone and G. Tasinato, [[arXiv:1611.04516](#)].
- [51] S. Brahma, E. Nelson and S. Shandera, Phys. Rev. D **89**, no. 2, 023507 (2014) [[arXiv:1310.0471](#)].
- [52] R. Flauger, D. Green and R. A. Porto, JCAP **1308**, 032 (2013) [[arXiv:1303.1430](#)].
- [53] A. Aravind, D. Lorshbough and S. Paban, JHEP **1307**, 076 (2013) [[arXiv:1303.1440](#)].
- [54] N. Boulanger, T. Damour, L. Gualtieri and M. Henneaux, Nucl. Phys. B **597**, 127 (2001) [[arXiv:0007220](#)].
- [55] A. G. Riess *et al.* [Supernova Search Team], Astron. J. **116**, 1009 (1998) [[arXiv:9805201](#)].
- [56] B. P. Schmidt *et al.* [Supernova Search Team], Astrophys. J. **507** (1998) 46 [[arXiv:9805200](#)].
- [57] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], Bull. Am. Astron. Soc. **29** (1997) 1351 [[astro-ph/9812473](#)]. [[arXiv:9805200](#)].
- [58] D. Vegh, [[arXiv:1301.0537](#)].
- [59] M. Blake, D. Tong and D. Vegh, Phys. Rev. Lett. **112**, no. 7, 071602 (2014) [[arXiv:1310.3832](#)].

- [60] A. I. Vainshtein, Phys. Lett. **39B**, 393 (1972).
- [61] E. Babichev and C. Deffayet, Class. Quant. Grav. **30**, 184001 (2013) [[arXiv:1304.7240](#)].
- [62] A. De Felice, A. E. Gumrukcuoglu, S. Mukohyama, N. Tanahashi and T. Tanaka, JCAP **1406**, 037 (2014) [[arXiv:1404.0008](#)].
- [63] M. Lagos and P. G. Ferreira, JCAP **1412**, 026 (2014) [[arXiv:1410.0207](#)].
- [64] M. Johnson and A. Terrana, Phys. Rev. D **92**, no. 4, 044001 (2015) [[arXiv:1503.05560](#)].
- [65] Y. Akrami, S. F. Hassan, F. Knig, A. Schmidt-May and A. R. Solomon, Phys. Lett. B **748**, 37 (2015) [[arXiv:1503.07521](#)].
- [66] G. Cusin, R. Durrer, P. Guarato and M. Motta, JCAP **1509**, no. 09, 043 (2015) [[arXiv:1505.01091](#)].
- [67] P. Brax, A. C. Davis and J. Noller, [[arXiv:1703.08016](#)].
- [68] C. de Rham, Living Rev. Rel. **17**, 7 (2014) [[arXiv:1401.4173](#)].
- [69] A. Kehagias and A. Riotto, Fortsch. Phys. **65**, no. 5, 1700023 (2017) [[arXiv:1701.05462](#)].
- [70] C. de Rham, L. Heisenberg and R. H. Ribeiro, Class. Quant. Grav. **32**, 035022 (2015) [[arXiv:1408.1678](#)].
- [71] M. Fasiello and R. H. Ribeiro, JCAP **1507**, no. 07, 027 (2015) [[arXiv:1505.00404](#)].
- [72] D. Jeong and M. Kamionkowski, Phys. Rev. Lett. **108**, 251301 (2012) [[arXiv:1203.0302](#)].
- [73] L. Dai, D. Jeong and M. Kamionkowski, Phys. Rev. D **87**, no. 10, 103006 (2013) [[arXiv:1302.1868](#)].
- [74] L. Dai, D. Jeong and M. Kamionkowski, Phys. Rev. D **88**, no. 4, 043507 (2013) [[arXiv:1306.3985](#)].
- [75] A. R. Pullen and M. Kamionkowski, Phys. Rev. D **76**, 103529 (2007) [[arXiv:0709.1144](#)].
- [76] S. Ando and M. Kamionkowski, Phys. Rev. Lett. **100**, 071301 (2008) [[arXiv:0711.0779](#)].
- [77] N. E. Groeneboom and H. K. Eriksen, Astrophys. J. **690**, 1807 (2009) [[arXiv:0807.2242](#)].
- [78] D. Hanson and A. Lewis, Phys. Rev. D **80**, 063004 (2009) [[arXiv:0908.0963](#)].
- [79] C. L. Bennett *et al.*, Astrophys. J. Suppl. **192**, 17 (2011) [[arXiv:1001.4758](#)].
- [80] A. R. Pullen and C. M. Hirata, JCAP **1005**, 027 (2010) [[arXiv:1003.0673](#)].
- [81] P. A. R. Ade *et al.* [Planck Collaboration], Astron. Astrophys. **571**, A23 (2014) [[arXiv:1303.5083](#)].
- [82] J. Kim and E. Komatsu, Phys. Rev. D **88**, 101301 (2013) [[arXiv:1310.1605](#)].
- [83] N. Arkani-Hamed, H. Georgi and M. D. Schwartz, Annals Phys. **305**, 96 (2003) [[arXiv:0210184](#)].
- [84] A. Kehagias and A. Riotto, accepted for publication in JCAP [[arXiv:1705.05834](#)].

- [85] I. L. Buchbinder, V. A. Krykhtin and V. D. Pershin, Phys. Lett. B **466**, 216 (1999) [[arXiv:9908028](#)].
- [86] A. Riotto and M. S. Sloth, JCAP **0804**, 030 (2008) [[arXiv:0801.1845](#)].
- [87] C. P. Burgess, L. Leblond, R. Holman and S. Shandera, JCAP **1003**, 033 (2010) [[arXiv:0912.1608](#)].
- [88] D. Seery, Class. Quant. Grav. **27**, 124005 (2010) [[arXiv:1005.1649](#)].
- [89] P. Creminelli, J. Norena and M. Simonovic, JCAP **1207**, 052 (2012) [[arXiv:1203.4595](#)].