

Unitary null energy condition violation in $P(X)$ cosmologiesClaudia de Rham^{1,2,*} and Scott Melville^{1,†}¹*Theoretical Physics, Blackett Laboratory, Imperial College, London SW7 2AZ, United Kingdom*²*CERCA, Department of Physics, Case Western Reserve University,**10900 Euclid Avenue, Cleveland, Ohio 44106, USA*

(Received 10 March 2017; published 19 June 2017)

A non singular cosmological bounce in the Einstein frame can only take place if the null energy condition (NEC) is violated. We explore situations where a single scalar field drives the NEC violation and derive the constraints imposed by demanding tree level unitarity on a cosmological background. We then focus on the explicit constraints that arise in $P(X)$ theories and show that constraints from perturbative unitarity make it impossible for the NEC violation to occur within the region of validity of the effective field theory without also involving irrelevant operators that arise at a higher scale that would enter from integrating out more massive degrees of freedom. Within the context of $P(X)$ theories we show that including such operators allows for a bounce that does not manifestly violate tree level unitarity, but at the price of either imposing a shift symmetry or involving technically unnatural small operator coefficients within the low-energy effective field theory.

DOI: [10.1103/PhysRevD.95.123523](https://doi.org/10.1103/PhysRevD.95.123523)**I. INTRODUCTION**

Describing the earliest moments of our Universe remains one of the greatest challenges of physics. The singularity theorem states that if the Universe is described by general relativity (GR) with a Friedmann-Lemaître-Robertson-Walker (FLRW) metric and matter that respects the null energy condition, then extrapolating backwards in time, our present understanding must break down and “new physics” has to become important. This can happen in one of two ways: (i) Either the Hubble parameter reaches Planckian values at very early times, and the effects of quantum gravity become important (“big bang singularity”), (ii) or the null energy condition (NEC) is violated and the Universe underwent a cosmological bounce. While theories of quantum gravity are still in development, scenario (ii) may be amenable to current QFT techniques. A great deal of recent work has focused on constructing an early Universe cosmology that takes advantage of (ii) in order to have a nonsingular bounce, some of which can even provide an alternative to inflation (see for instance [1–17]).

In addition to modeling the very early Universe, violating the NEC in a stable, unitary way can be useful in other contexts [18]. It has been proposed as a mechanism for quintessence [19], solving the cosmological constant problem [6,20,21], and is also required in theories with traversable wormholes [22–24].

Whether NEC violation can ever be stable (free of ghosts and gradient instabilities) has been the subject of much discussion [25–28]. In [29], it was for instance shown that within the context of “kinetic gravity braiding” (see also

[30]) one can reach a NEC violating phase while remaining free of ghosts and gradient instabilities. In the context of the most general single scalar degree of freedom coupled to gravity (so-called “Horndeski gravity” [31] or “generalized Galileons” [32]), it is known that there is always a ghost or gradient instability in any bouncing solution as was first pointed out in [33] and further in [34–37]. However this instability can be made to occur long before or after the NEC violation as was already constructed in the context of G bounces in [33] and more recently in a cubic Galileon bounce in [38], or in some cases removed by imposing particular asymptotic conditions [39].

Within the context of single scalar field $P(\Phi, X)$ theories [denoted as simply $P(X)$ in this work unless we wish to emphasize the distinction], it is well known that the bounce, or the onset of the NEC violation, is necessarily linked with a vanishing speed of sound, and potentially classical instabilities. A way out is to include higher derivative operators in the effective field theory (EFT), which may naturally capture the effect of high-energy degrees of freedom without needing to commit to a particular UV completion of the scalar effective field theory considered. Such additional higher derivative operators were previously used in an attempt to regulate the sound speed in the simplest scalar field $P(X)$ theories [6], for example, by adding higher spatial derivatives and arranging for the instability to be much slower than the Hubble rate [28].

In addition to potential classical stability issues linked with a vanishing speed of sound in a relativistic field theory, we emphasize here that there are strong coupling issues associated with it that signal the breakdown of the quantum effective field theory. This can be seen intuitively by noting that the non-Gaussianities typically scale as $f_{\text{NL}} \sim c_s^{-2}$. A vanishing of the sound speed therefore directly implies that

*c.de-rham@imperial.ac.uk

†s.melville16@imperial.ac.uk

the effect from the higher order operators dominates and hence there is a breakdown of perturbative unitarity. To make this statement more concrete and quantify it, we consider the scattering amplitude of high frequency fluctuations on a given bouncing background and determine the energy scale at which perturbative unitarity is violated. Strong coupling effects from cubic operators in $P(X)$ cosmological bounces were previously discussed in [40]. Here we explore all the operators that enter the EFT (cubic and beyond) and fully investigate the effect that a small sound speed has on the validity of the EFT. In particular, we show that a vanishing sound speed inexorably leads to a vanishing of the strong coupling scale, which is unacceptable, and which agrees with the well-known results of [41,42]. Our conclusion on the fate of the strong coupling issue in any *pure*¹ $P(X)$ cosmological bounce [including a $P(\Phi, X)$ bounce] therefore departs from some other previous analysis, but agrees with [41,42].

However, by including irrelevant operators that enter from high-energy effects, we show that it is possible to restore perturbative unitarity throughout the bounce in $P(X)$ theories. This was for instance proposed in [9,28] to avoid any gradient instabilities. We show explicitly here how these operators affect the strong coupling scale. In particular, we show that for these high-energy irrelevant operators to “save” perturbative unitarity their associated scale needs to be chosen carefully: It must be low enough to restore unitarity, but high enough so as to decouple the low-energy effective field theory from specific high-energy completions.

In what follows we focus our attention on the requirements set by perturbative unitarity (and particularly tree level unitarity) within the vicinity of a NEC violating region, so as to determine whether a classical NEC violation can be trusted. We provide an example where this NEC violation leads to a cosmological bounce, but do not incorporate this bounce within a full cosmological scenario. The model considered here does not attempt to circumvent the no-go mentioned in [34–37]. Rather, the approach of this manuscript is that our $P(X)$ theory ought to successfully capture the duration of the bounce, but that this low-energy effective field theory will likely break down (and new physics ought to be included) if followed sufficiently far in the past (well before the bounce or the start of the NEC violating region).

With this approach in mind, we begin with a general discussion of EFT consistency conditions in Sec. II, which makes a few assumptions on the details of the scalar field theory considered. This involves considering a precisely defined *decoupling limit* of the gravity/scalar effective field theory, and establishing the constraints arising from the requirement of a (at least perturbatively) unitary S

matrix. We then turn our attention in Sec. III to the high-energy effects that can help regulate the classical gradient instabilities as well as the violation of unitarity at the onset of a NEC violation. Keeping track of those effects, we derive the full bound, bearing in mind that the high-energy effects have to occur at a sufficiently low scale to restore unitarity, but at a sufficiently large scale to decouple the high-energy states from the low-energy effective field theory. This provides a restricted window of possibility. After having derived the formal requirements set by perturbative unitarity, we focus our attention in Sec. IV to $P(X)$ models that violate the NEC. We show that a level of tuning is required for those models that may call into question their naturalness, but show that, in principle, a violation of the NEC could occur while maintaining classical stability and perturbative unitarity once a particular (albeit technically unnatural) tuning is chosen. We illustrate this result by providing an explicit covariant example of a $P(\Phi, X)$ model that can allow for a bouncing solution while preserving perturbative unitarity. Our results are then summarized in Sec. V. In Appendix A we analyze the well-known ghost-condensate model and show that NEC violation in this model cannot preserve unitarity unless the irrelevant operators coming from high-energy effects take a particular form that breaks the shift symmetry.

II. SINGLE SCALAR EFT

To set the stage, we start by considering the theory of a single scalar field Φ coupled to a metric $g_{\mu\nu}$, with no mention of any additional high derivative operators. The consistency of this EFT requires absence of gradient and ghost instabilities as well as unitarity (in the sense that any n -point scattering amplitude should satisfy the optical theorem).

A. Decoupling limit

Throughout this manuscript we focus on a *decoupling limit* of the full theory, which is designed to focus on the leading interactions that determine the strong coupling scale of the theory, i.e., the scale at which perturbative unitarity is broken. The existence of this decoupling limit comes from the assumption of a hierarchy of scales $\Lambda \ll M_{\text{Pl}}$, where Λ is the typical interaction scale for the scalar field. Any stability and unitarity bounds determined in that limit represent necessary conditions that must be satisfied by the full theory. In other words, it is sufficient to use the decoupling limit to infer the strong coupling physics, even if not the full cosmological predictions (e.g., power spectrum, bispectrum, etc.). Crucially, this limit continues to allow for bouncing solutions that temporarily violate the null energy condition.

To derive the decoupling limit, we begin with an action of the form

¹Here by pure we mean without the inclusion of high-energy irrelevant operators.

$$S[g_{\mu\nu}, \Phi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{Pl}}^2}{2} R + \mathcal{L}(g, \Phi) \right). \quad (2.1)$$

We then expand this action around the cosmological background describing the bouncing solution ($\bar{\gamma}_{\mu\nu}$ and $\bar{\phi}$), defining the perturbations in an *arbitrary gauge* ($h_{\mu\nu}$ and φ) via

$$g_{\mu\nu} = \bar{\gamma}_{\mu\nu} + h_{\mu\nu} \quad \text{and} \quad \Phi = \bar{\phi} + \varphi. \quad (2.2)$$

The physical action for the perturbations is (schematically)

$$\begin{aligned} S[h_{\mu\nu}, \varphi] = & \int d^4x \sqrt{-\bar{\gamma}} (-M_{\text{Pl}}^2 \bar{\mathcal{B}}(\bar{\gamma})(\partial h)^2 - \bar{R}(\bar{\gamma}) h^2 \\ & + M_{\text{Pl}}^2 f_1(\bar{\gamma}) h(\partial h)^2 + \dots - Z(\bar{\gamma}, \bar{\phi})(\partial \varphi)^2 \\ & - m^2(\bar{\gamma}, \bar{\phi}) \varphi^2 + f_2(\bar{\gamma}, \bar{\phi}) \varphi(\partial \varphi)^2 + f_3(\bar{\gamma}, \bar{\phi}) \varphi^3 \\ & + \dots + f_4(\bar{\gamma}, \bar{\phi}) h \partial^2 \varphi + f_5(\bar{\gamma}, \bar{\phi}) h \varphi \\ & + f_6(\bar{\gamma}, \bar{\phi}) h^2 \partial^2 \varphi + \dots), \end{aligned} \quad (2.3)$$

where the last line represents all the potential mixing between h and φ . All the functions $\bar{\mathcal{B}}$, \bar{R} , Z , and f_i depend on the background (and of course carry indices); however, their exact expressions are irrelevant for the rest of the scaling argument. The scale of the background naturally enters all these functions—for instance, $\bar{R}(\bar{\gamma})$ is related to the background curvature. In principle, the linear kinetic mixing between h and φ could be taken care of by performing an appropriate field redefinition and absorbing the function f_3/M_{Pl}^2 in the expression of Z . In practice however such a shift is irrelevant in the decoupling limit we consider below as it is suppressed by the Planck scale.

Our working assumption is that the scalar field contains interactions at the scale Λ , such that $\mathcal{L}(g, \Phi) \sim \Lambda^4$, and that there is a large hierarchy $\Lambda \ll M_{\text{Pl}}$. If this is the case then we typically expect $\dot{H} \sim \Lambda^4/M_{\text{Pl}}^2$. The leading interactions that determine the strong coupling physics are determined principally by the scalar field, with those that arise from the mixing with gravity being M_{Pl} suppressed. As such we may take a decoupling limit $M_{\text{Pl}} \rightarrow \infty$, keeping Λ fixed. In the explicit bouncing solutions we construct in Sec. IV C, the time scale for the null energy violation is set by the scale of the scalar interactions $\Delta t \sim \Lambda^{-1}$. This in turn implies an even stronger suppression for the Hubble rate $H \sim \Lambda^3/M_{\text{Pl}}^2$, since $H \sim \dot{H} \Lambda^{-1}$. This is borne out by the explicit solutions of Sec. IV C.

The first step in dealing with this effective field theory of $h_{\mu\nu}$ and φ is to properly canonically normalize the fields. In this case, the appropriate canonical normalization of the field is²

²In practice the tensors $\bar{\mathcal{B}}$ and Z are usually not conformal with respect to the background metric and the rescaling should be taken with slightly more care as is performed in Sec. II B; however, those subtleties do not affect the essence of the decoupling limit.

$$\tilde{h}_{\mu\nu} \sim \frac{1}{\sqrt{\bar{\mathcal{B}} M_{\text{Pl}}}} h_{\mu\nu} \quad \text{and} \quad \tilde{\varphi} \sim \frac{1}{\sqrt{Z}} \varphi, \quad (2.4)$$

and the action is (again symbolically)

$$\begin{aligned} S[\tilde{h}_{\mu\nu}, \tilde{\varphi}] = & \int d^4x \sqrt{-\bar{\gamma}} \left(-(\partial \tilde{h})^2 - \frac{\bar{R}}{\bar{\mathcal{B}}} \tilde{h}^2 - (\partial \tilde{\varphi})^2 - \frac{m^2}{Z} \tilde{\varphi}^2 \right. \\ & + \frac{f_2}{Z^{3/2}} \tilde{\varphi}(\partial \tilde{\varphi})^2 + \frac{f_3}{Z^{3/2}} \tilde{\varphi}^3 + \dots \\ & + \frac{1}{M_{\text{Pl}}} \frac{f_1}{\bar{\mathcal{B}}^{3/2}} \tilde{h}(\partial \tilde{h})^2 + \frac{f_4}{M_{\text{Pl}} \sqrt{\bar{\mathcal{B}} Z}} \tilde{h} \partial^2 \tilde{\varphi} \\ & \left. + \frac{f_5}{M_{\text{Pl}} \sqrt{\bar{\mathcal{B}} Z}} \tilde{h} \tilde{\varphi} + \frac{f_6}{M_{\text{Pl}}^2 \bar{\mathcal{B}} \sqrt{Z}} \tilde{h}^2 \partial^2 \tilde{\varphi} + \dots \right). \end{aligned} \quad (2.5)$$

Then taking a limit where $M_{\text{Pl}} \rightarrow \infty$, while maintaining the scales that enter the scalar field background fixed, we see that all of the whole second line becomes unimportant, and the scalar field fluctuations $\tilde{\varphi}$ entirely decouple from the metric fluctuations (which become a trivial free theory). A significant virtue of this decoupling limit is that the gauge degrees of freedom remain in h and decouple. This procedure is thus insensitive to any gauge issues. For this reason it is unnecessary to work with the comoving curvature perturbation ζ , or similar gauge invariant variables (this point is discussed in more detail at the end of this subsection).

The relevant effective action in this limit is hence solely that of the scalar field fluctuations

$$\begin{aligned} S_{\text{dec}} = & \int d^4x \sqrt{-\bar{\gamma}} \left(-(\partial \tilde{\varphi})^2 - \frac{m^2}{Z} \tilde{\varphi}^2 + \frac{f_2}{Z^{3/2}} \tilde{\varphi}(\partial \tilde{\varphi})^2 \right. \\ & \left. + \frac{f_3}{Z^{3/2}} \tilde{\varphi}^3 + \dots \right), \end{aligned} \quad (2.6)$$

where the ellipses carry operators to all orders in φ (and potentially $\partial \varphi$ and even $\partial^2 \varphi$ and higher order in derivatives). Although gravity has decoupled, this is not the same as the scalar theory on Minkowski spacetime. Information about the background is carried through the process, and we are effectively looking at a scalar field on curved background. In the explicit solutions given in Sec. IV C, $\phi(t)$ remains finite in the limit $M_{\text{Pl}} \rightarrow \infty$, with a time dependence at the scale $\Delta t \sim 1/\Lambda$. In other words the scalar field is varying at a time scale $1/\Lambda$ that is much faster than the naive background scale $1/H$; similarly $\dot{H}/H \sim \Lambda \gg H$.

Before proceeding, it is worth noting that a decoupling limit of this form is not always appropriate; for example, in slow roll inflation, the flatness of the potential means that the interactions coming from the mixing with gravity are actually the dominant ones, and it is incorrect to perform the above limit. Another example would be if the interactions with the metric were made artificially large, for instance in the following example,

$$S = \int d^4x \sqrt{-g} \left(M_{\text{Pl}}^2 R + \frac{M_{\text{Pl}}}{\Lambda} R \Phi^2 - \frac{1}{2} (\partial\Phi)^2 + \frac{1}{\Lambda^4} \Phi^4 \right), \quad (2.7)$$

where $\Lambda \ll M_{\text{Pl}}$. In this case the mixing term between the metric and the scalar field fluctuations ought to be taken into account for background configurations with $\phi \sim \Lambda$. Of course in this case, we can simply remedy this issue by first going to the Einstein frame, which automatically accounts for the scalar/gravity mixing. When dealing with generic scalar-tensor theories there is not necessarily a covariant definition of the Einstein frame; however, the appropriate procedure is simply to first go to the relevant Einstein frame at the perturbed level about the cosmological background and then perform the appropriate decoupling limit. In the present case, the decoupling limit we have defined is justified by the existence of a large class of explicit solutions, which we discuss in Sec. IV C, which have the property that $H \sim \Lambda^3/M_{\text{Pl}}^2$ and $\dot{H} \sim \Lambda^4/M_{\text{Pl}}^2$.

We emphasize that our concern here is about the consistency of the effective action, and the scale at which perturbative unitarity breaks down. This is a very different question than that of say, the precise form of the power spectrum or bispectrum. For precise questions of this sort, performing a decoupling limit and focusing on the scalar field effective action would not provide an accurate answer; however, for the question of perturbative unitarity we address in this manuscript, focusing on the scalar field decoupling limit on the appropriate background gives necessary conditions that must be satisfied.

1. Gauge issues

Before analyzing this decoupling limit we briefly comment on the gauge issues that have been highlighted in the literature (see for instance Ref. [40] for a recent discussion). In particular, if we first chose a gauge where the scalar part of h_{ij} vanishes we would find that the constraints for the shift and the lapse would impose them to scale as $(M_{\text{Pl}}^2 H)^{-1}$. The resulting Lagrangian (before taking any decoupling limit) would then involve inverse powers of $M_{\text{Pl}}^2 H$. Since in the explicit solution we provide in Sec. IV C $H \sim \Lambda^3/M_{\text{Pl}}^2$, such a behavior, if generic, would invalidate the decoupling limit arguments. Moreover there would seem to be an ‘‘apparent’’ singularity at the bounce itself (i.e., when $H = 0$). This should come as no surprise since it is impossible to fix the gauge $h_{ij}^{\text{scalar}} = 0$ when $H = 0$, so the previous apparent singularity is simply a gauge artefact and so is the scaling found in that gauge.

Switching to comoving gauge however does not help with the previous issue either since in that gauge the kinetic coefficient of the curvature perturbation does itself then scale as H^{-1} , which makes it impossible to properly normalize the field. Rather than focusing on any of these two gauges (or any standard ‘‘local’’ gauges), the problem at

hand can be entirely dealt with by going for instance to de Donder gauge or any other gauge of that form (or at leading order by going to harmonic gauge as performed in [43]). The appropriate way to perform this is to consider the action (2.1) and add the appropriate Fadeev-Popov gauge fixing terms so that they combine with the Einstein curvature term to give kinetic terms for the metric fluctuation that take the remarkably simple form $M_{\text{Pl}}^2 h^{\mu\nu} \bar{\square} (h_{\mu\nu} - \frac{1}{2} h \bar{\gamma}_{\mu\nu})$ in addition to higher order fluctuations (and where $\bar{\square}$ is the d’Alembertian with respect to the background metric $\bar{\gamma}_{\mu\nu}$). In this language there is no constraint to solve for since the gauge fixing terms are precisely there to break gauge invariance. The absence of constraints ensures that at no point would one need to perform an inversion of the Hubble parameter and the appropriate canonical normalization follows the same behavior as in (2.4) (where \mathcal{B} is manifestly finite and is trivial in the flat space limit). The breaking of gauge invariance from the gauge fixing terms comes at the price of including other spurious degrees of freedom but it is well understood how to introduce the Fadeev-Popov ghosts to deal with those and they only contribute to loops. All the tree level amplitudes computed with these de Donder gauge fixing terms are therefore the same as that of the original theory. In this formulation the decoupling limit can therefore be taken precisely as discussed previously. For the questions we are interested in (namely the size of the strong coupling scale, whether or not the theory preserves tree level unitarity, etc.), we can therefore safely perform this decoupling limit and work with the low-energy effective field theory for the scalar field φ on the cosmological background.

In the case of the specific example that is provided in Sec. IV C, we can compute (at least to a given order) the corrections that arise beyond the decoupling limit. By performing a complementary analysis to that described above, and keeping track of the corrections that arise due to the mixing with the metric, we have checked that we obtain a correction to the effective mass for the scalar field that is suppressed by six orders of magnitude, which is precisely what one would have expected in our example.

B. Scalar on FLRW

Since the main interest of this manuscript is cosmology (and the potential description of cosmological bounces), the following analysis takes place on a flat FLRW background with scale factor $a(t)$. The effective metric for the scalar field fluctuations then takes the form

$$Z^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi = -A(t) \dot{\varphi}^2 + \frac{B(t)}{a^2} (\partial_i \varphi)^2, \quad (2.8)$$

where A and B depend on the background behavior (here and in what follows dots represent the physical time derivatives and ∂_i designate spatial gradients). The

corresponding sound speed is $c_s^2 = \frac{B}{A}$. The absence of a ghost (in the scalar sector) then implies $A > 0$, while the absence of gradient instabilities implies $B > 0$.

For the rest of the analysis we assume that we are within the WKB regime, meaning that we consider frequencies well above the scale set by the background time variation, E_{back} , with

$$E_{\text{back}} = \text{Max}(H, \sqrt{\dot{H}}, m, \dot{\phi}/\phi, \ddot{\phi}/\dot{\phi}, \dots). \quad (2.9)$$

Indeed to probe unitarity violation we are interested in analyzing the interactions of modes close to the strong coupling scale, which should be much larger than the background scale for the effective field theory to make sense. In the explicit models given later, the scale of variation of the background solution is $1/\Lambda$, i.e., $E_{\text{back}} \sim \Lambda$, and so to ensure the validity of the WKB regime we require $\omega \gg \Lambda$.

1. Canonical normalization

Before going through the formal unitarity bounds arising from the optical theorem, it is useful to first estimate the strong coupling scale by simply analyzing the scale of the operators present in the scalar field theory (on the FLRW background). Since Lorentz invariance is *spontaneously* broken by the FLRW background, the effective metric $Z^{\mu\nu}$ is not conformally flat, so before canonically normalizing the field, it is useful to perform the following coordinate rescaling,³

$$\tilde{t} = \int c_s(t) dt \quad \text{and} \quad \tilde{x}^i = x^i, \quad (2.10)$$

so that in this system of coordinates the effective metric is conformally flat,

$$\begin{aligned} & \int dt d^3x a^3 \left(-\frac{1}{2} Z^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi \right) \\ &= \int d\tilde{t} d^3\tilde{x} a^3 \frac{B}{2c_s} \left(\left(\frac{\partial \varphi}{\partial \tilde{t}} \right)^2 - \frac{1}{a^2} \left(\frac{\partial \varphi}{\partial \tilde{x}^i} \right)^2 \right), \end{aligned} \quad (2.11)$$

and we can simply canonically normalize the field by setting

$$\varphi = \sqrt{\frac{c_s}{B}} \tilde{\varphi} = (AB)^{-1/4} \tilde{\varphi} \quad (2.12)$$

³As we see, as soon as B , or the speed of sound, approaches small enough (positive) values, the whole EFT runs out of control and the classical background is not to be trusted; therefore, there is no sense in which B or the speed of sound actually ever vanishes, and even less became negative in the first place. The rescaling performed in this section is thus well defined.

(the derivatives of AB are then simply absorbed into the mass term).

2. Irrelevant operators

Now consider an EFT on the FLRW background that contains an irrelevant operator of the form

$$S_{NML} = \int dt d^3x \frac{a^{3-2L}}{\Lambda_{NML}^{N+2M+4L-4}} \varphi^N \dot{\varphi}^M (\partial_i \varphi)^{2L}, \quad (2.13)$$

where N , M , and L are arbitrary positive integers with $N + 2M + 4L > 4$ and the scale Λ_{NML} depends on the background and the particular theory one is dealing with. Then in terms of the rescaled coordinates and the properly canonically normalized field, this interaction is

$$\begin{aligned} S_{NML} &= \int d\tilde{t} d^3\tilde{x} \frac{a^{3-2L}}{\tilde{\mu}_{NML}^{N+2M+4L-4}} (\tilde{\partial}_i \tilde{\varphi})^{2L} \\ &\times \sum_{j=0}^M \left(\frac{\partial_t (AB)}{c_s AB} \right)^j \tilde{\varphi}^{N+j} (\partial_i \tilde{\varphi})^{M-j}, \end{aligned} \quad (2.14)$$

where we have ignored signs and other order one and combinatory numbers. The scale $\tilde{\mu}$ is given by

$$\tilde{\mu}_{NML} = A^{\frac{N+3M+2L-2}{4(N+2M+4L-4)}} B^{\frac{N-M+2L+2}{4(N+2M+4L-4)}} \Lambda_{NML}. \quad (2.15)$$

Now when rescaling back to the original coordinates, as an energy scale we have

$$\mu_{NML} = \frac{\partial \tilde{t}}{\partial t} \tilde{\mu}_{NML} = c_s \tilde{\mu}_{NML}, \quad (2.16)$$

and so the scale that enters these interactions is

$$\mu_{NML} = A^{n_A} B^{n_B} \Lambda_{NML}, \quad (2.17)$$

with

$$\begin{aligned} n_A &= -\frac{N+M+6L-6}{4(N+2M+4L-4)} \quad \text{and} \\ n_B &= \frac{3N+3M+10L-6}{4(N+2M+4L-4)}. \end{aligned} \quad (2.18)$$

We therefore see that in a situation where A becomes parametrically large or B is parametrically small, for a fixed scale Λ_{NML} there are typically operators that enter at a parametrically small scale $\mu_{NML} \ll \Lambda_{NML}$ and therefore spoil the validity of the effective field theory at a low scale.

For these arguments to be valid one should have $\frac{\partial_t (AB)}{AB} \sim E_{\text{back}} \lesssim \text{Min}(\mu_{NML})$ [where $\text{Min}(\mu_{NML})$ is the lowest of all the possible scales derived in (2.17) for any non-negative integers N, M, L with $N + 2M + 4L > 4$]. This allows us to ignore the contributions from $\frac{\partial_t (AB)}{c_s AB}$ in (2.14). If

$\frac{\partial_t(AB)}{AB} \sim E_{\text{back}}$ was larger than any of those scales μ_{NML} , then by definition the background would be varying faster than the time scale set by the strong coupling scale of the effective theory and it would no longer be possible to trust the validity of the background solution.

3. Marginal and relevant operators

For the relevant $\Lambda_{300}/3!\varphi^3$ and marginal $\lambda/4!\varphi^4$ operators, the previous rescaling can also be performed and leads to

$$\mu_{300} = \frac{\Lambda_{300}}{(AB)^{3/4}}, \quad (2.19)$$

$$\mu_{400} = \frac{\lambda}{(AB^3)^{1/2}}, \quad (2.20)$$

where of course μ_{400} does not represent an actual scale but rather a dimensionless coupling constant. Remaining in the perturbative regime requires the dimensionless coupling constant to be $\mu_{400} \ll \mathcal{O}(16\pi)$ and the scale of the marginal operator to be $\mu_{300} \lesssim E_{\text{back}}$. However since these operators are renormalizable, we can deal with them in the strong coupling regime and we therefore do not necessarily need to impose that these couplings are small to preserve unitarity.

C. Optical theorem

The previous section provided a generic scaling argument to determine the typical interaction scale of an operator on a FLRW background. We now provide more substance to this argument by showing how it relates to the optical theorem by computing a precise scattering amplitude.

Going back to the effective metric (2.8) in its original formulation on FLRW (before any canonical normalization), in the WKB regime the quantized modes can be decomposed in the following way:

$$\begin{aligned} \hat{\varphi}(t, x^i) = & \int \frac{d^3 k_i}{(2\pi)^3 a^3} \frac{1}{\sqrt{\mathcal{N}(k)}} \left(\hat{a}^\dagger(k_i) e^{i(k \int_{a(t)}^{c_s(t)} dt - k_i x^i)} \right. \\ & \left. + \hat{a}(k_i) e^{-i(k \int_{a(t)}^{c_s(t)} dt - k_i x^i)} \right). \end{aligned} \quad (2.21)$$

To derive the normalization $\mathcal{N}(k)$ we go back to the well-known definition of the Klein-Gordon norm along a three-dimensional surface Σ with unit normal vector n^μ and induced metric $\gamma_{\mu\nu}$,

$$|\hat{\varphi}|^2 = -i \int_{\Sigma} d^3 x \sqrt{\gamma} n^\mu (\hat{\varphi} \overleftrightarrow{\partial}_\mu \hat{\varphi}^\dagger). \quad (2.22)$$

Choosing a spacelike surface $t = \text{const}$, the unit normal vector is $n^\mu = \sqrt{A} \delta_0^\mu$ and the induced volume element is $\sqrt{\gamma} = B^{-3/2}$, leading to the Klein-Gordon norm,

$$|\hat{\varphi}|^2 = -i \frac{\sqrt{A}}{B^{3/2}} \int d^3 x a^3 (\hat{\varphi} \overleftrightarrow{\partial}_t \hat{\varphi}^\dagger). \quad (2.23)$$

We can therefore infer the field normalization,

$$\mathcal{N}(k) = \frac{\partial F}{\partial \omega} = 2A\omega = 2(AB)^{1/2} \frac{k}{a}, \quad (2.24)$$

where we evaluated the mode on shell, $\omega = c_s k/a$, and where the function F determines the dispersion relation,

$$F = A\omega^2 - B \frac{k^2}{a^2} = 0. \quad (2.25)$$

In other words the speed of sound is given by $c_s^2 = B/A$. The operators \hat{a} and \hat{a}^\dagger that appear in (2.21) are the creation and annihilation operators for one-particle states of definite momentum that obey the usual commutation relations.

In the high-energy WKB regime, it is meaningful to talk about an approximate S-matrix for scattering of quanta, provided it is defined over a time scale shorter than the background variation $\Delta t \ll 1/E_{\text{back}}$. The constraints from unitarity on this approximate S-matrix are many; however, it is useful to focus on the special case of n to n scattering, between states of equal momenta, $|i\rangle = |f\rangle = |k_1 \dots k_n\rangle$. This gives a unitarity bound for every integer $n \geq 2$,

$$\begin{aligned} 2\text{Im}(\langle i|T|i\rangle) &= \sum_N \langle i|T^\dagger|q_1 \dots q_N\rangle \langle q_1 \dots q_N|T|i\rangle \\ &\geq |\langle q_1 \dots q_n|T|i\rangle|^2, \end{aligned} \quad (2.26)$$

so that the scattering amplitudes \mathcal{A}_{2n} (with momentum conserving delta function removed) should satisfy the following relation:

$$\begin{aligned} 2\text{Im}\mathcal{A}_{2n}(k_j; k_j) &\geq \int \frac{d^3 q_1}{(2\pi)^3 \mathcal{N}(q_1)} \dots \frac{d^3 q_n}{(2\pi)^3 \mathcal{N}(q_n)} |\mathcal{A}_{2n}(k_j; q_j)|^2 \\ &\times \left[(2\pi)^4 \delta\left(\frac{c_s^2}{a^2} \left(\sum k_j^2 - \sum q_j^2\right)\right) \delta^{(3)}\left(\sum k_j - \sum q_j\right) \right]. \end{aligned} \quad (2.27)$$

The scattering amplitudes may be computed using standard Feynman diagrams using the WKB form for the propagator for scattering at time intervals shorter than $\Delta t \ll 1/E_{\text{back}}$.

1. Four-point function

In the case of $n = 2$, the two-body phase space factor on the right-hand side of (2.27) can be evaluated simply in terms of a center of mass energy, \sqrt{s} , and scattering angle, θ , giving the well-known optical theorem for the 2 to 2 scattering amplitude $\mathcal{A}_4(s, \theta)$. For our dispersion relation

$\omega^2 = c_s^2 k^2 / a^2$, a partial wave expansion yields the following bound for every $\ell \geq 0$,

$$|\mathcal{A}_{4,\ell}(s)| \leq 8\pi^2 \frac{\omega}{k/a} AB = 8\pi^2 (AB^3)^{1/2}, \quad (2.28)$$

with

$$\mathcal{A}_{4,\ell}(s) = \int_{-1}^1 d\cos\theta P_\ell(\cos\theta) \mathcal{A}_4(s, \theta). \quad (2.29)$$

Now consider a cubic operator of the form given in (2.13) with $N + M + 2L = 3$ (keeping in mind we are still dealing with an irrelevant operator, $N + 2M + 4L > 4$). Taking into account the proper normalization of the propagator, such a cubic operator leads to a contribution to the four-point function scaling as (at fixed angle)

$$\mathcal{A}_4^{(\text{cubic})} \sim \frac{k^{4L-2} \omega^{2M}}{B \Lambda_{NML}^{2(N+2M+4L-4)}} \sim \frac{A^{2L-1}}{B^{2L}} \left(\frac{\sqrt{s}}{\Lambda_{NML}} \right)^{2(M+2L-1)} \quad (2.30)$$

(keeping in mind that $N = 3 - M - 2L$). We have ignored any combinatory factor in this estimation since we are mainly after a scaling argument. However once a particular amplitude is diagnosed as being potentially problematic the proper factors are included. Now applying the bound (2.28), we determine that perturbative unitarity gets broken at the scale

$$\sqrt{s} \sim A^{\frac{3-4L}{4(M+2L-1)}} B^{\frac{3+4L}{4(M+2L-1)}} \Lambda_{NML}, \quad (2.31)$$

which is precisely the scale μ_{NML} inferred from the simple scaling argument in (2.17) when $N = 3 - M - 2L$. When dealing with the contribution to the $2-2$ scattering amplitude from the cubic vertex φ^3 , we see that the amplitude (2.30) is dominated instead by IR effects and one should go beyond the tree level amplitude when $\Lambda_{300} \gtrsim (AB)^{3/4} E_{\text{back}}$, as deduced in (2.19).

Similarly, we can consider a quartic irrelevant operator of the form given in (2.13) with $N + M + 2L = 4$ ($N + 2M + 4L > 4$), which leads to a contribution to the four-point function going as (at fixed angle)

$$\mathcal{A}_4^{(\text{quartic})} \sim \frac{k^{2L} \omega^M}{\Lambda_{NML}^{N+2M+4L-4}} \sim \frac{A^L}{B^L} \left(\frac{\sqrt{s}}{\Lambda_{NML}} \right)^{M+2L}. \quad (2.32)$$

From the perturbative unitarity bound (2.28) we see that such a quartic operator is responsible for breaking perturbative unitarity at the scale

$$\sqrt{s} \sim \Lambda_{NML} A^{\frac{1-2L}{2M+4L}} B^{\frac{3-2L}{2M+4L}}, \quad (2.33)$$

which is again precisely the scale μ_{NML} derived in (2.17) for a quartic operator $N = 4 - M - 2L$. When dealing with the marginal operator $\lambda \varphi^4 / 4!$, its contribution to the

previous amplitude simply goes as λ and remaining in the perturbative regime then requires $\lambda \ll (AB^3)^{1/2}$, which is once again precisely the requirement deduced previously in (2.20), although breaking this bound does not yet imply breaking unitarity since that operator is renormalizable.

One obvious worry is that cancellations could occur for instance between different cubic operators or between the contributions from $\mathcal{A}_4^{(\text{cubic})}$ and $\mathcal{A}_4^{(\text{quartic})}$. If that were the case, it would simply imply that a field redefinition could be performed to remove such operators (or part of them). For the simple scalar field effective theory we are dealing with this is fortunately relatively straightforward to monitor and provided we are not dealing with an unnecessarily complicated formulation of the effective theory the strong coupling scale is indeed given by the smallest of the scales μ_{NML} provided in (2.17). A more complete discussion of this aspect is given in Sec. IID.

2. 2n-point function

To complete the perturbative unitarity requirement as directly imposed from the optical theorem, we provide here the bounds from higher n -point functions. In that case the n -body phase space factor is a complicated integral over several momenta and scattering angles, but those do not affect the overall scaling of the bound and we obtain the following perturbative unitarity bound for the $2n$ -point functions,

$$|\mathcal{A}_{2n}| \lesssim A^{(3-n)/2} B^{3(n-1)/2} s^{2-n}, \quad (2.34)$$

where we have ignored order one numerical factors. We can again compare this bound with the strong coupling scale we infer from an operator (2.13) with $N + M + 2L = 2n$ and we see that we infer precisely the same scaling for the strong coupling scale in terms of A and B as was found in (2.17).

D. Field redefinitions and redundant operators

The previous arguments to deduce the strong coupling scale and the breaking of perturbative unitarity implicitly assumed that any operator of the form (2.13) leads to a contribution to the tree level scattering amplitude. There are of course cases where this assumption fails.

- (i) First, it may be possible that the contribution of an operator to a scattering amplitude accidentally canceled at tree level. If the cancellation only occurs at tree level but the operator still contributes to loops with an order of magnitude comparable to what is estimated in for instance (2.32) or (2.30), then the argument remains unaffected as the loops still lead to a violation of perturbative unitarity (even though this may occur at higher order in the loop expansion).
- (ii) Second, the contribution of an operator to *any* scattering amplitude may exactly cancel (to all orders in loops). This simply means that this operator is actually not present and in a single scalar

field theory (where there is no gauge issue) this can only happen if that operator is simply removable by integrations by parts. This case is of course trivial.

- (iii) Finally the particular contribution of an operator may not cancel by itself but may be canceled by the contributions of other operators. If this cancellation does not occur for scattering amplitudes and to all orders in loops, then the previous arguments on the breaking of perturbative unitarity are effectively unaffected. On the other hand, if this set of operators was simply removable by an appropriate field redefinition without leading to other new higher order operators, then in practice these operators are redundant and superficial to the description of any physics. In practice when analyzing a scalar field effective theory we ought to take care of all such redundancies in the first place.

Besides these previous special cases, the arguments presented here are robust in diagnosing the potentially dangerous operators and applicable to any scalar field effective field theory that carries operators of the form (2.13). Once a diagnosis has been established one can directly compute the amplitude with the appropriate combinatorial factors to fully determine the fate of unitarity.

The main motivation of this work is to apply these bounds to $P(X)$ models that can, in principle, allow for a violation of the NEC and potentially allow for a cosmological bounce. Within the context of pure $P(X)$ theories (coupled) to gravity, it is well known that the speed of sound turns negative at the onset of the NEC violation and thus leads to gradient instabilities; this agrees with previous literature such as [41,42]. This can potentially be remedied by considering new irrelevant operators that arise at a higher scale and regulate this instability. For instance considering the irrelevant operator $(\square\Phi)^2/\Lambda_c^2$ it can allow for a stable bounce [9,28]. In what follows we therefore explore the effects of such an irrelevant operator on the unitarity bound considered previously, before focusing on $P(X)$ models endowed with such an operator.

III. INCLUSION OF HIGH-ENERGY EFFECTS

We now consider the inclusion of high-energy effects that naturally enter any EFT. For concreteness, we can imagine that the next massive particle enters with a mass Λ_c , and integrating out such a massive particle leads to an operator for the form⁴ $(\square\Phi)^2/\Lambda_c^2$ in the low-energy effective action for Φ . To see this more precisely, we could

⁴Even though the operator considered here carries more derivatives, the would-be Ostrogradsky ghost associated with it is not present since its mass lies above the cutoff of the EFT, by construction. In what follows we take great care in making sure that in the regime of validity of the cosmological bounce, the would-be ghost is not present; see Refs. [44–46] for a discussion of this point.

for instance consider the two-scalar field example coupled to gravity,

$$S[g_{\mu\nu}, \Phi, \chi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{Pl}}^2}{2} R + \mathcal{L}(g, \Phi) - \frac{1}{2} (\partial\chi)^2 - \frac{\Lambda_c^2}{2} \chi^2 + \frac{1}{2} \chi \square\Phi \right), \quad (3.1)$$

where we assume no ghosts or other types of pathologies in $\mathcal{L}(g, \Phi)$. Note that at this level the model is entirely free of any type of ghost. If the field χ is sufficiently massive [i.e., the scale Λ_c is sufficiently large as compared with the other scales in $\mathcal{L}(g, \Phi)$], the field χ is frozen and its dynamics decouple, and we can integrate it out. At leading order,

$$\chi = -\frac{\square\Phi}{\Lambda_c^2} = \frac{\square\Phi}{\Lambda_c^2} + \dots, \quad (3.2)$$

and we are left with irrelevant operators in the low-energy effective field theory for Φ ,

$$S[g_{\mu\nu}, \Phi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{Pl}}^2}{2} R + \mathcal{L}(g, \Phi) + \frac{1}{2\Lambda_c^2} (\square\Phi)^2 + \dots \right), \quad (3.3)$$

where the ellipses designate higher order corrections in \square/Λ_c^2 . As mentioned in footnote 4, the irrelevant operator we have included $(\square\Phi)^2$ seems to carry higher derivatives and one could be worried about the associated Ostrogradsky ghost; however, in this context, when reaching the scale Λ_c , it is no longer appropriate to integrate out the dynamics of χ and one should go back to (3.1) for the appropriate description, which is clearly free of ghosts.

In general we could expect the scale Λ_c to also be background dependent (e.g., through Φ), and hence to carry a time dependence within this cosmological setup, but within the WKB approximation (keeping in mind that we are interested in energy scales much smaller than Λ_c) we ignore any possible time dependence of Λ_c for the rest of the manuscript.

Since $(\square\Phi)^2/\Lambda_c^2$ is a higher derivative operator, it leads to inevitable unitarity violation at the scale Λ_c (or possibly a background redressed version Λ'_c). Consequently this sets an upper bound on the cutoff of the effective theory $\Lambda_{\text{cutoff}} \leq \Lambda'_c$. This cutoff scale is not necessarily the same as the strong coupling scale Λ_s , which signals the breaking of perturbative unitarity of the theory (see Refs. [47,48] for a clear distinction between these two scales). New physics may not necessarily be required at the scale of the breaking of perturbative unitarity; however, the classical background can no longer be trusted in that case. So if a classical violation of the NEC occurred at a scale E_{back} beyond the strong coupling scale, there would be no reason to believe

the NEC violation actually took place. Putting this together, then overall consistency of our EFT description requires $E_{\text{back}} \ll \Lambda_s \leq \Lambda_{\text{cutoff}} \leq \Lambda'_c$.

A. Dispersion relation

At first sight, the inclusion of the operator $(\square\Phi)^2/\Lambda_c^2$, suppressed by the large scale Λ_c , would appear to have a negligible effect on the theory. Furthermore, if it does have an effect, one would worry that it would be necessary to include the infinite number of higher order operators that enter at the scale Λ_c , such as $(\Phi\square^{n+2}\Phi)/\Lambda_c^{2+2n}$. The reason that this is not the case is that in the absence of this operator, for the NEC violating solutions we consider, c_s passes through 0 and temporarily becomes negative. The inclusion of the operator $(\square\Phi)^2/\Lambda_c^2$ then creates a large correction (relative to 0), which must be included, while at the same time higher order operators like $(\Phi\square^{n+2}\Phi)/\Lambda_c^{2+2n}$ are negligible. To see how this works, we note that on including the irrelevant operator $(\square\Phi)^2/\Lambda_c^2$ in the scalar field effective theory on the curved background, the quadratic action for φ is then

$$S^{(2)} = \int d^4x \frac{a^3}{2} \left(A\dot{\varphi}^2 - B\frac{(\partial_i\varphi)^2}{a^2} - m^2\varphi^2 + \frac{1}{\Lambda_c^2} \left(\ddot{\varphi} - \frac{\partial_i^2\varphi}{a^2} \right)^2 \right), \quad (3.4)$$

where A , B , and m are determined by the background behavior and we still use the notation $c_s^2 = B/A$. The expression for the function F that provides the equation for the dispersion relation in (2.25) is therefore now promoted to

$$F = A\omega^2 - B\frac{k^2}{a^2} - \frac{1}{\Lambda_c^2} \left(\omega^2 - \frac{k^2}{a^2} \right)^2 = 0, \quad (3.5)$$

where we work with modes well above the background and hence the mass, so the mass term is safely ignored in the previous expression. The previous equation can be solved as

$$\omega^2(k) = c_s^2 \frac{k^2}{a^2} + \frac{(A-B)^2}{A^3} \frac{k^4}{a^4\Lambda_c^2} + \mathcal{O}(k^6/\Lambda_c^4). \quad (3.6)$$

In general the first term dominates in the naive region of validity of the EFT $k/a \ll \Lambda_c$. However, for the bouncing solutions c_s^2 could become very small, potentially passing through 0 and becoming negative, and the second term then regulates any instability that may occur, while the higher order corrections $\mathcal{O}(k^6/\Lambda_c^4)$ remain small in comparison. Interestingly in our explicit solutions, we find that the inclusion of the operator $(\square\Phi)^2/\Lambda_c^2$ also modifies the background solution in such a way that c_s^2 remains positive

throughout the bounce. Again this is achieved while terms like $(\Phi\square^{n+2}\Phi)/\Lambda_c^{2+2n}$ remain negligible.

In addition to the dispersion relation (3.6), there is a second solution, which is the Ostrogradsky ghost mode

$$\omega_{\text{ghost}}^2 \sim A\Lambda_c^2 \left(1 + \mathcal{O}\left(\frac{k^2}{\Lambda_c^2}\right) \right). \quad (3.7)$$

These nonunitary degrees of freedom must be excluded from the EFT, and since they enter at an energy scale $\sqrt{A}\Lambda_c$, this sets the maximal value of the energy cutoff of the theory on this background,

$$\Lambda_{\text{cutoff}} \leq \Lambda'_c = \sqrt{A}\Lambda_c. \quad (3.8)$$

In addition to Λ_{cutoff} , it is useful in what follows to introduce the scale μ_c , which is the energy scale at which the dispersion relation transitions from the relativistic form $\omega = c_s k$ to the nonrelativistic form dominated by the operator $(\square\Phi)^2/\Lambda_c^2$,

$$\mu_c = c_s^2 \frac{A^{3/2}}{(A-B)} \Lambda_c. \quad (3.9)$$

Because of the background redressing, μ_c may differ significantly from the scale Λ_c and from the cutoff of the theory. For instance if we were in a situation where $B \ll 1 \ll A$, then $\mu_c \sim B/\sqrt{A}\Lambda_c \ll \Lambda_c \ll \Lambda'_c$. In practice, the situations where the irrelevant operator has an effect on the dispersion relation arise because the speed of sound $c_s^2 = B/A$ is small, which implies that $B \ll A$. In what follows we slightly simplify the notation by making that assumption and hence write

$$\mu_c \sim c_s^2 \sqrt{A}\Lambda_c = \frac{B}{\sqrt{A}} \Lambda_c. \quad (3.10)$$

From the quadratic action (3.4) [or the dispersion relation (3.6)], it is clear that when considering modes $k/a \gg \mu_c/c_s$, or equivalently when looking at energy scales $\omega \gg \mu_c$, the irrelevant operator takes over from the standard gradient term. There are now therefore two different regimes to consider when checking perturbative unitarity of the theory.

B. Intermediate energy modes with $E_{\text{back}}^2 \ll s \ll \mu_c^2$

For modes with energy well below the scale μ_c (yet well above any of the background scales) the high-energy effects are irrelevant and we can perform the analysis in the same way as we did in the previous section in the absence of the irrelevant operator $(\square\Phi)^2/\Lambda_c^2$. We can therefore infer the following.

- (i) If any of the energy scales μ_{NML} derived in (2.17) happened to be smaller than the energy scale μ_c , then the smallest of those scales would be the maximal

value of the strong coupling scale and perturbative unitarity would break down at that scale.

- (ii) If on the other hand *all* the energy scales μ_{NML} derived in (2.17) are larger than μ_c then higher energy effects (which are still below the cutoff) ought to be taken into account to properly diagnose the scale of perturbative unitarity breaking and the scales computed in (2.17) should then be ignored.

C. High-energy modes with $\mu_c^2 \ll s \ll \Lambda_{\text{cutoff}}^2$

For the high-energy modes, $\mu_c^2 \ll s \ll \Lambda_{\text{cutoff}}^2$, as we have seen already at the quadratic level, the irrelevant operator $k^4 \varphi^2 / \Lambda_c^2$ takes over from the standard gradient term. At those energy scales, the second contribution in (3.6) dominates the dispersion relation, which hence becomes

$$\omega \sim \frac{1}{A^{1/2} \Lambda_c} \frac{k^2}{a^2} \sim \frac{c_s^2 k^2}{\mu_c a^2}. \quad (3.11)$$

1. Scaling

The scaling performed in Sec. II B should hence be reinvestigated. Considering the operator $(\partial_i^2 \varphi)^2 / \Lambda_c^2$ rather than the subdominant gradient term $B(\partial_i \varphi)^2$, the appropriate rescaling then corresponds to substituting B by $(k/a\Lambda_c)^2 \sim (As/\Lambda_c^2)^{1/2}$. This is a slight abuse of notation; however, when working in the WKB regime, such a rescaling can be justified. This means that an operator of the form (2.13) (with $N + 2M + 4L > 4$) preserves perturbative unitarity so long as the energy \sqrt{s} satisfies

$$\mu_c \lesssim \sqrt{s} \ll A^{n_A} \left(\frac{As}{\Lambda_c^2} \right)^{n_B/2} \Lambda_{NML}, \quad (3.12)$$

where the powers n_A and n_B are given in (2.18). If $n_B < 1$, i.e., for $N + 5M + 6L > 10$ this implies that unitarity is broken above the energy

$$\sqrt{s} \sim A^{\frac{N+M-2L+6}{2(N+5M+6L-10)}} \left(\frac{\Lambda_{NML}}{\Lambda_c} \right)^{\frac{4(N+2M+4L-4)}{N+5M+6L-10}} \Lambda_c$$

for $N + 5M + 6L > 10$. (3.13)

On the other hand for $n_B \geq 1$, the bound (3.12) is always satisfied so long as it is satisfied at $s = \mu_c^2$, i.e., so long as

$$\Lambda_{NML} \gg A^{\frac{2N+3M+6L-4}{2(N+2M+4L-4)}} B^{\frac{7N+11M+11L-22}{4(N+2M+4L-4)}} \Lambda_c$$

for $N + 5M + 6L \leq 10$, (3.14)

(and for $N + 2M + 4L > 4$). For the relevant and marginal operators φ^3 and φ^4 , this distinction between intermediate and low-energy modes is unimportant and we can follow the discussion of Eqs. (2.19) and (2.20).

2. Scattering amplitude

The previous argument required a scaling that itself depended on momenta. For a more rigorous treatment, one can instead first directly compute the bounds imposed by the optical theorem in the presence of the irrelevant operator and then estimate the $2n$ -point functions without the need of ever performing the rescaling of the coordinates mentioned previously. Just as in Sec. II C, both methods give the same result, but we sketch the direct bounds from the optical theorem (without the need of any rescaling) to solidify the argument and the results.

When including the irrelevant operator, first notice that the correct normalization of the field in (2.21) is still given by $\mathcal{N} = \frac{\partial F}{\partial \omega}$ as in the first equality of Eq. (2.24) but where the function F is now given by (3.5) and where the dispersion relation $\omega = \frac{1}{A^{1/2} \Lambda_c} \frac{k^2}{a^2}$, hence leading to

$$\mathcal{N} = \frac{\partial F}{\partial \omega} = \frac{2}{A^{5/2} \Lambda_c} \frac{k^2}{a^2}. \quad (3.15)$$

This change of normalization \mathcal{N} affects the momenta integrals in (2.27), which in turn affects the bounds provided by the optical theorem. Taking these effects into account, the perturbative unitarity bound from the $2n$ -point functions (2.34) now becomes

$$|\mathcal{A}_{2n}| \lesssim A^{(n+3)/4} \left(\frac{\sqrt{s}}{\Lambda_c} \right)^{\frac{3(n-1)}{2}} s^{2-n}, \quad (3.16)$$

again up to numerical factors that have been ignored.

Now if we consider the effect of an operator given in (2.13), with $N + M + 2L = 2n$ and $N + 2M + 4L > 4$, it leads to a contribution to the $2n$ -point function going as

$$\mathcal{A}_{NML} \sim \frac{k^{2L} \omega^M}{\Lambda_{NML}^{N+2M+4L-4}} \sim A^{L/2} \frac{\Lambda_c^L s^{(M+L)/2}}{\Lambda_{NML}^{M+2L+2n-4}}. \quad (3.17)$$

Of course this is only one contribution to that amplitude, and this does not account for the very special case where this contribution happens to be precisely canceled from another operator. This means that these two (or more) operators are removable by field redefinition. This requires a very precise tuning between these operators and simply means that the original formulation of the theory was unnecessarily complicated. We deal with this here by defining the relevant scale Λ_{NML} *after* all the appropriate and relevant field redefinitions have been performed. See Sec. II D for more discussion on this point. We have also ignored any combinatorial factors that ought to be included when computing the amplitude, just like we have ignored for now any factorial that enters in the definition of the operator (2.13) but those can easily be accounted for once we have diagnosed the potentially dangerous operators and amplitudes.

Then from Eq. (3.16), the contribution to amplitude (3.17) diagnoses a violation of tree level unitarity at the energy scale

$$\sqrt{s} \sim A^{\frac{3-2L+2n}{2(-5+2L+2M+n)}} \left(\frac{\Lambda_{NML}}{\Lambda_c} \right)^{\frac{2(-4+2L+M+2n)}{-5+2L+2M+n}} \Lambda_c, \quad (3.18)$$

for $N + 2M + 2L > 5$ (or $N + 5M + 6L > 10$), which is precisely the scale inferred from the simple scaling argument in (3.13) when using $N + M + 2L = 2n$. For operators with $N + 5M + 6L \leq 10$ on the other hand, they respect unitarity so long as they satisfy the same bound as that derived in (3.14).

3. Strong coupling scale

To summarize, we have computed in two different and yet complementary ways the strong coupling scale associated with a class of scalar EFTs on a cosmological background. We consider any scalar field EFT that takes the following form on a cosmological background,

$$S = \int d^4x a^3 \left(\frac{A}{2} \dot{\varphi}^2 - \frac{B}{2} \frac{(\partial_i \varphi)^2}{a^2} - \frac{1}{2} m^2 \varphi^2 + \sum_{\substack{N, M, L \geq 0 \\ N+M+2L \geq 3}} \frac{\varphi^N \dot{\varphi}^M (\partial_i \varphi)^{2L}}{\Lambda_{NML}^{N+2M+4L-4} a^{2L}} + \frac{(\square \varphi)^2}{2\Lambda_c^2} \right), \quad (3.19)$$

where A , B , m , and the scales Λ_{NML} are all functions of time and depend on the precise theory one is dealing with and on the particular cosmological background considered. The scales Λ_{NML} are the relevant scales of the theory on the cosmological background after all the appropriate and relevant field redefinitions have been performed and typically $\Lambda_{NML} \ll A\Lambda_c$. We denote by E_{back} the scale of the background.

Then for that theory (modulo the subtleties related to field redefinitions and redundant operators discussed in Sec. IV D) perturbative unitarity requires the following conditions on the scale of the different irrelevant operators.

- (i) For any irrelevant operator ($N + 2M + 4L > 4$),

$$A^{\frac{N+M+6L-6}{4(N+2M+4L-4)}} B^{\frac{3N+3M+10L-6}{4(N+2M+4L-4)}} \Lambda_{NML} \gg \mu_c \sim \frac{B}{\sqrt{A}} \Lambda_c. \quad (3.20)$$

- (ii) For all irrelevant operators with $N + 5M + 6L > 10$,

$$A^{\frac{N+M-2L+6}{2(N+5M+6L-10)}} \left(\frac{\Lambda_{NML}}{\Lambda_c} \right)^{\frac{4(N+2M+4L-4)}{N+5M+6L-10}} \Lambda_c \gg E_{\text{back}}. \quad (3.21)$$

- (iii) For all the irrelevant operators with $N + 2M + 4L - 4 > 0$ and $N + 5M + 6L \leq 10$,

$$A^{\frac{2N+3M+6L-4}{2(N+2M+4L-4)}} B^{\frac{7N+11M+11L-22}{4(N+2M+4L-4)}} \Lambda_{NML} \gg \Lambda_c. \quad (3.22)$$

The strong coupling scale is designated by

$$\Lambda_s = \text{Min}_{N, M, L} \left(A^{\frac{N+M-2L+6}{2(N+5M+6L-10)}} \left(\frac{\Lambda_{NML}}{\Lambda_c} \right)^{\frac{4(N+2M+4L-4)}{N+5M+6L-10}} \right) \Lambda_c, \quad (3.23)$$

where the minimum is taken over all the N , M , L that satisfy $N + 5M + 6L > 10$ and where we have implicitly assumed that $E_{\text{back}} \gtrsim \mu_c \sim B/\sqrt{A}\Lambda_c$ and the second condition (3.20) is satisfied. If $E_{\text{back}} \lesssim \mu_c$, the last two requirements are irrelevant and the right-hand side of (3.20) is simply E_{back} instead of μ_c and the strong coupling scale of interest is instead given by the left-hand side of (3.20). Once again, these conditions should be taken as a diagnosis. If any of those was broken one could go back to computing the explicit contributions to the scattering amplitudes, which would account for the proper numerical factors and is independent of field redefinitions.

When it comes to the relevant and marginal operators, since those operators are renormalizable, one can, in principle, deal with those in the strong coupling regime. Remaining in the perturbative regime requires

$$\begin{aligned} \mu_{300} &= \frac{\Lambda_{300}}{(AB)^{3/4}} \lesssim E_{\text{back}} \quad \text{and} \\ \mu_{400} &= \frac{\lambda}{(AB^3)^{1/2}} \lesssim \mathcal{O}(16\pi) \end{aligned} \quad (3.24)$$

(where λ is the coupling constant that enters for the φ^4 operator); however, we emphasize that breaking this bound (3.24) does not directly imply breaking of unitarity but rather that loops from these operators ought to be considered.

In the rest of this manuscript, we now use these bounds to infer if and how a NEC violation can occur in $P(\Phi, X)$ theories of gravity and whether a bounce is possible without violation of unitarity.

IV. VIOLATING THE NEC IN $P(\Phi, X)$ THEORY

We focus the discussion of perturbative unitarity on $P(\Phi, X)$ theories near a NEC violating region. In that case the scalar field Lagrangian in (2.1) takes the form

$$S[g_{\mu\nu}, \Phi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{Pl}}^2}{2} R + P(\Phi, X) \right), \quad (4.1)$$

with

$$X = -\frac{1}{2} g^{\mu\nu} \partial_\mu \Phi \partial_\nu \Phi, \quad (4.2)$$

and as before, we are interested in the effective scalar field theory for φ on the FLRW cosmological background where we have set $\Phi = \phi + \varphi(t, x^i)$ and the background satisfies the appropriate equations of motion. In particular, the background energy density is given by

$$\rho = -\bar{P} + \bar{P}_{,X}\dot{\phi}^2, \quad (4.3)$$

where “bar” quantities are related to the background, i.e., $\bar{X} = \frac{1}{2}\dot{\phi}^2$ and $\bar{P} = P(\phi, \bar{X})$.

The kinetic coefficients and the mass that determine the quadratic action for φ are given by

$$\begin{aligned} A &= 2\bar{X}\bar{P}_{,XX} + \bar{P}_{,X}, & B &= \bar{P}_{,X} \quad \text{and} \\ m^2 &= \partial_t[\bar{P}_{,\Phi X}\dot{\phi}] - \bar{P}_{,\Phi\Phi}, \end{aligned} \quad (4.4)$$

and stability of the theory requires these three functions be positive. Note that when accounting for the kinetic factors A and B , the effective mass scale “perceived” by the properly normalized scalar field is given by

$$m_{\text{eff}}^2 = A^{-1}m^2 = A^{-1}(\partial_t[\bar{P}_{,\Phi X}\dot{\phi}] - \bar{P}_{,\Phi\Phi}). \quad (4.5)$$

The operators S_{NML} (2.13) of the effective theory on the cosmological background can be derived for any given $P(\Phi, X)$ model, and the scales Λ_{NML} are given by

$$\Lambda_{NML}^{-(N+2M+4L-4)} = \sum_{j=0}^{M/2} \frac{(-1)^L \dot{\phi}^{M-2j}}{2^{j+L} N! L! j! (M-2j)!} \partial_\Phi^N \partial_X^{M+L-j} \bar{P}. \quad (4.6)$$

Without having more insight on the precise form of the function P and whether or not it truncates at any order in Φ or X , there is little more one can say about those scales. In what follows we hence take into account the fact that the function P allows for a NEC violation, and even consider the case where P is such that a cosmological bounce occurs. We first focus on the case without any higher order effects before including their effects.

A. Ignoring high-energy effects

When considering a pure $P(\Phi, X)$ model, without the inclusion of higher energy effects as mimicked by the operator $(\square\Phi)/\Lambda_c^2$ in Sec. III, the background equations of motion are simply

$$3M_{\text{pl}}^2 H^2 = \rho = -\bar{P} + 2\bar{P}_{,X}\bar{X}, \quad (4.7)$$

$$M_{\text{pl}}^2 \dot{H} = -\frac{1}{2}(p + \rho) = -\bar{X}\bar{P}_{,X}. \quad (4.8)$$

It is well known that in a $P(\Phi, X)$ theory, at the onset of a NEC violation⁵ $\bar{P}_{,X} = B = 0$. Even more worrisome than the standard stability issues caused by the gradient terms vanishing or becoming negative is that we can directly see that the onset of the NEC violation in any pure $P(\Phi, X)$ theory is inexorably linked with a violation of unitarity, which makes it impossible to trust that the NEC violation actually occurred in the first place.

Indeed, if $B \rightarrow 0$ the strong coupling scale derived in (2.17) [or in (3.23)] vanishes (hence signaling violation of perturbative unitarity at an arbitrarily small scale) unless *all* the operators with $3N + 3M + 10L > 6$ vanished (or could all have been simultaneously removed by a field redefinition), which means there can be no interactions at all, which of course means there could not have been any operator to set a NEC violation in the first place. So without including higher energy effects, even putting aside any stability issue that may occur at the classical level, there can be no NEC violation within the regime of validity of a pure $P(\Phi, X)$ model. In what follows we therefore include the high-energy effects that are naturally expected to be present in any effective field theory. Notice that we still work within the low-energy effective field theory and therefore should not be sensitive about the exact details of the UV physics or its exact realization, but their effect plays the role of a regulator for the low-energy effective theory that is crucial for the onset of a NEC violation in $P(\Phi, X)$ types of theories.

B. Including high-energy effects

When including the operator $(\square\Phi)^2/\Lambda_c^2$, i.e., when considering the following action,

$$S[g_{\mu\nu}, \Phi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{pl}}^2}{2} R + P(\Phi, X) + \frac{1}{2\Lambda_c^2} (\square\Phi)^2 \right), \quad (4.9)$$

the background equations of motion are slightly modified to

$$\begin{aligned} 3M_{\text{pl}}^2 H^2 &= -\bar{P} + 2\bar{P}_{,X}\bar{X} \\ &+ \frac{1}{2\Lambda_c^2} [-\ddot{\phi}^2 + 2\dot{\phi}\ddot{\phi} + 6\dot{H}\dot{\phi} - 9H^2\dot{\phi}^2], \end{aligned} \quad (4.10)$$

⁵From the Raychaudhuri Eq. (4.8) we could try to onset the NEC by setting $\dot{\phi} = 0$ rather than $\bar{P}_{,X} = B = 0$. However, setting $\dot{\phi} = 0$ also implies $\dot{H} = 0$ at that time, which means that we are not within the NEC violating region unless one also has $\ddot{H} = 0$ at that time, which itself also implies $\bar{P}_{,X} = 0$ (or $\ddot{\phi} = 0$, in which case the same story continues). So to summarize, starting outside the NEC region $\dot{H} < 0$, one can never enter within the NEC region $\dot{H} > 0$ without passing through a point for which $\bar{P}_{,X} = 0$ in a pure $P(\Phi, X)$ theory. We can therefore assume that $\dot{\phi} \neq 0$ at the onset of the NEC violation without loss of generality.

$$M_{\text{Pl}}^2 \dot{H} = -\bar{X}\bar{P}_{,X} + \frac{1}{\Lambda_c^2} [\dot{\phi} \ddot{\phi} + 3\dot{H}\dot{\phi}^2 + 3H\dot{\phi}\ddot{\phi}]. \quad (4.11)$$

The effect of Λ_c on the Friedman equation is irrelevant when working within the regime of validity of the effective field theory, but it allows for a violation of the NEC at finite positive B ,

$$B = \bar{P}_{,X} = \frac{2}{\Lambda_c^2 \dot{\phi}} (\ddot{\phi} + 3H\ddot{\phi}) > 0 \quad \text{when } \dot{H} = 0. \quad (4.12)$$

As mentioned in (2.9), the scale of the background, E_{back} , is given by

$$E_{\text{back}} = \text{Max}(H, \sqrt{\dot{H}}, m, \dot{\phi}/\phi, \ddot{\phi}/\dot{\phi}, \dots) \gtrsim \sqrt{B}\Lambda_c, \quad (4.13)$$

where the last inequality is valid when $\dot{H} \sim 0$. Preserving perturbative unitarity requires the background scale to be smaller than the strong coupling scale derived in (3.23),

$$E_{\text{back}} \ll \Lambda_s, \quad (4.14)$$

which requires

$$\sqrt{B}\Lambda_c \ll \Lambda_s. \quad (4.15)$$

Some of the literature does also require the background energy density ρ to be much smaller than the strong coupling scale of the perturbed effective field theory $\rho \ll \Lambda_s^4$. This is certainly a safe requirement to impose as one would not *expect* the EFT to be stable under quantum corrections otherwise. From a pure unitarity-preserving perspective, however, we are not forced to impose that condition.

Since we can assume $\dot{\phi} \neq 0$, and since the derivatives of the background should be small compared with the scale of high-energy physics Λ_c , we can immediately infer the hierarchy

$$B \ll A \quad \text{and} \quad B \ll 1 \quad \text{when } \dot{H} = 0, \quad (4.16)$$

which implies that the speed of sound ought to be small in this model for a NEC violation to occur. A small sound speed may not necessarily be problematic at the classical level, especially not when the high-energy effects regulate any gradient instability that could occur, but as we see it greatly affects unitarity.

To establish whether or not perturbative unitarity can be preserved in the vicinity of a NEC violating region we start by looking at the effect of the intermediate modes as described in Sec. II B. We therefore consider modes with energy scale $E_{\text{back}}^2 \ll s \ll \mu_c^2$. In that regime, if any of the scales derived in (2.17) happened to be smaller than $\mu_c \sim BA^{-1/2}\Lambda_c \ll \Lambda_c$, with now the Λ_{NML} given by (4.6)

(and keeping in mind that $B \ll A \sim 2\bar{X}\bar{P}_{,XX}$), then that scale would set the strong coupling scale.

1. Breaking the shift symmetry

Let us start by considering a $P(\Phi, X)$ theory that does not necessarily preserve the shift symmetry (i.e., has explicit Φ dependence). We also start by making the very natural assumption that the fundamental theory does not carry any hierarchy of scales by which we mean that one can formulate the function $P(\Phi, X)$ in terms of just one scale $\Lambda (\ll \Lambda_c \ll M_{\text{Pl}})$ and order one dimensionless coefficients $c_{\ell,m}$,

$$P(\Phi, X) = \Lambda^4 \sum_{\ell,n} \frac{c_{\ell,n}}{\Lambda^{\ell+4n}} \Phi^\ell X^n. \quad (4.17)$$

We then have $\partial_\Phi^4 \bar{P} \sim 1$ (or $\partial_\Phi^4 \bar{P} \gg 1$ if the background involves $\phi \gg \Lambda$ or $\bar{X} \gg \Lambda^4$, but we are not able to have $\partial_\Phi^4 \bar{P} \ll 1$ unless a very specifically tuned cancellation occurs precisely at the onset of the NEC violation). Similarly we would never expect to have $\Lambda^{N+4M-4} \partial_\Phi^N \partial_X^M \bar{P} \ll 1$ unless a very particular tuning was set to occur precisely at the onset of the NEC violation, or unless the shift symmetry or another precise type of symmetry was present. So in a typical theory (4.17) with no hierarchy of scales, we expect $\Lambda_{NML} \sim \Lambda$ for all $N + 2M + 4L \neq 4$.

With this assumption in mind, we can consider the quintic operator

$$\mathcal{L}_{500} = \frac{\varphi^5}{\Lambda_{500}} \quad \text{with} \quad \Lambda_{500} = 5!(\partial_\Phi^5 \bar{P})^{-1} \sim 5!\Lambda. \quad (4.18)$$

As we have seen, to avoid any breaking of unitarity this operator should satisfy the requirements (3.20) and (3.22), leading to

$$\Lambda \gtrsim (A^3 B^5)^{-1/4} \Lambda_c, \quad (4.19)$$

which further requires $A \gtrsim B^{-5/3} (\Lambda_c/\Lambda)^{4/3} \gg 1$, which is only possible if some of the coefficients present in (4.17) are much larger than unity. Once the door is opened for such a special tuning [i.e., when some coefficients $c_{\ell,m}$ in (4.17) are allowed to be parametrically much larger than others], any vertex can, in principle, dominate the scattering amplitudes and lead to much stronger bounds than would be inferred from the other vertices.

2. Preserving the shift symmetry

An obvious way to evade the previous argument is to keep the shift symmetry and hence avoid any operator that depends explicitly on φ . We can start by assuming as we did earlier that the covariant theory contains no large hierarchies,

$$P(X) = \Lambda^4 \sum_{n \geq 1} \frac{c_n}{\Lambda^{4n}} X^n, \quad (4.20)$$

with all the c_n of order 1, with the possibility that some of them may vanish and that the sum may truncate at order \bar{N} . In that case one should have $\Lambda^{4(n-1)} \partial_X^n \bar{P} \sim 1$ (or $\gg 1$) for any $2 \leq n \leq \bar{N}$ unless there is particular artificial tuning (or hierarchy of scales). In this case the most stringent bounds come from the cubic and quartic operators $\dot{\varphi}(\partial_i \varphi)^2$ and $(\partial_i \varphi)^4$, which impose the requirement (3.20) $\Lambda \gtrsim c_s^{1/4} \Lambda_c$. Since $c_s \ll 1$ at the onset of the NEC violation in these types of theories, this requirement is not *a priori* unreasonable, but it does constrain the theory. Once this constraint is satisfied, none of the other operators of the theory break perturbative unitarity so long as no artificially large coefficient is included in $P(X)$. Expressed as a constraint on $P(X)$ we see that perturbative unitarity requires

$$1 \ll \Lambda_c^4 \bar{P}_{,XX} \ll c_s^{-1} \quad (4.21)$$

(where the lower bound is coming from the requirement that $\Lambda \ll \Lambda_c$). The pure ghost condensate [6,9,28,40] lies in that category of models and is discussed in more detail in Appendix A. For that model we see that the unitarity bounds put severe constraints on the high-energy operators, which end up needing to break the shift symmetry.

The previous unitarity bound was derived by estimating the contribution of the cubic and quartic vertices to the $2-2$ scattering amplitudes. Having identified the potentially most dangerous vertices, we can go ahead and compute their actual contributions to the tree level scattering amplitudes to ensure that no “accidental cancellations” occur. The direct calculation of the $2-2$ tree scattering amplitude taking into account both the cubic and quartic vertices is in complete agreement with the estimations and provides the following upper bound: $\Lambda_c^4 \bar{P}_{,XX} \ll 192\pi^2/(103c_s)$.

3. Summary

To summarize, a NEC violation in shift-symmetric $P(X)$ theory is, in principle, possible so long as higher energy effects enter at a sufficiently low energy scale to regulate the scattering amplitudes, but still at sufficiently high energies so as not to entirely spoil the low-energy EFT. With these conditions in mind we obtain a limited, but not necessarily empty, window of possibility.

In practice, however, when it comes to obtaining explicit bouncing solutions, breaking the shift symmetry may make the bounce “easier” to model. In the absence of a protecting shift symmetry, the unitarity bounds are tighter and require an additional level of tuning of the model. These additional tunings imply that the $2-2$ scattering is no longer necessarily the dominant scattering amplitude and all processes should be examined with care to determine whether unitarity is preserved. One should also ensure

that the effective mass and couplings of the marginal and relevant operators are sufficiently small. This can be done explicitly, and to illustrate the process we now provide an explicit model that allows for a stable cosmological bounce that preserves unitarity at the price of introducing an unnaturally small parameter.

C. Explicit model

We now present an explicit model of the form

$$S[g_{\mu\nu}, \Phi] = \int d^4x \sqrt{-g} \left(\frac{M_{\text{Pl}}^2}{2} R + P(\Phi, X) + \frac{1}{2\Lambda_c^2} (\square\Phi)^2 \right), \quad (4.22)$$

with

$$P(\Phi, X) = -\Lambda^4 V(\Phi) + p(\Phi)X + \frac{q(\Phi)}{\Lambda^4} X^2, \quad (4.23)$$

so the background equations of motion are given by

$$3M_{\text{Pl}}^2 H^2 = \Lambda^4 V + p\bar{X} + \frac{3}{\Lambda^4} q\bar{X}^2 + \frac{1}{2\Lambda_c^2} [-\ddot{\phi}^2 + 2\dot{\phi}\ddot{\phi} + 6\dot{H}\dot{\phi} - 9H^2\dot{\phi}^2], \quad (4.24)$$

$$M_{\text{Pl}}^2 \dot{H} = -p\bar{X} - 2\frac{q}{\Lambda^4} \bar{X}^2 + \frac{1}{\Lambda_c^2} [\dot{\phi}\ddot{\phi} + 3\dot{H}\dot{\phi}^2 + 3H\dot{\phi}\ddot{\phi}]. \quad (4.25)$$

We can then explicitly check that the following profile,

$$\dot{\phi} = \Lambda\phi, \quad H = \frac{\Lambda^3}{M_{\text{Pl}}^2} h(\phi), \quad (4.26)$$

is an exact solution of the background equations of motion if we choose the following potential and functions of Φ ,

$$V(\Phi) = -\frac{1}{4} q(\Phi) \frac{\Phi^4}{\Lambda^4} + \left(3 \frac{\Lambda^2}{M_{\text{Pl}}^2} h^2(\Phi) + \Phi h'(\Phi) \right) - \frac{\Phi^2}{2\Lambda_c^2} \left(1 + 3 \frac{\Lambda^2}{M_{\text{Pl}}^2} h(\Phi) \right)^2, \quad (4.27)$$

$$p(\Phi) = -q(\Phi) \frac{\Phi^2}{\Lambda^2} - 2 \frac{\Lambda^2}{\Phi} h'(\Phi) + \frac{2\Lambda^2}{\Lambda_c^2} \left[1 + 3 \frac{\Lambda^2}{M_{\text{Pl}}^2} h(\Phi) + 3 \frac{\Lambda^2}{M_{\text{Pl}}^2} \Phi h'(\Phi) \right], \quad (4.28)$$

and for any pair of free functions $q(\Phi)$ and $h(\Phi)$. This freedom ensures that we can choose a profile that undergoes a cosmological bounce while preserving unitarity.

Focusing on the decoupling limit about the FLRW background as discussed in Sec. II A, the coefficients of the kinetic matrix are then given by

$$A = 2q(\phi) \frac{\phi^2}{\Lambda^2} - 2 \frac{\Lambda^2}{\phi} h'(\phi) + 2 \frac{\Lambda^2}{\Lambda_c^2} \left[1 + 3 \frac{\Lambda^2}{M_{\text{Pl}}^2} (h(\phi) + \phi h'(\phi)) \right], \quad (4.29)$$

$$B = -2 \frac{\Lambda^2 h'(\phi)}{\phi} + 2 \frac{\Lambda^2}{\Lambda_c^2} \left[1 + 3 \frac{\Lambda^2}{M_{\text{Pl}}^2} (h(\phi) + \phi h'(\phi)) \right], \quad (4.30)$$

and so the sound speed is given by

$$c_s^2 = \frac{\phi h'(\phi) - \frac{\phi^2}{\Lambda_c^2}}{\phi h'(\phi) - q \frac{\phi^4}{\Lambda^4} - \frac{\phi^2}{\Lambda_c^2}} + \mathcal{O}\left(\frac{\Lambda}{M_{\text{Pl}}}\right). \quad (4.31)$$

For example, if we choose

$$\phi(t) = \phi_0 e^{\Lambda t}, \quad (4.32)$$

$$h(\Phi) = -\frac{\Phi}{\phi_0} \frac{1 - \Phi^2/\phi_0^2}{1 + \Phi^4/\phi_0^4} \frac{\Phi^{10}/\phi_0^{10}}{1 + \Phi^{20}/\phi_0^{20}}, \quad (4.33)$$

then the Hubble parameter goes through a bounce at $t = 0$,

$$H(t) = \frac{\Lambda^3}{2M_{\text{Pl}}^2} \frac{\sinh(\Lambda t)}{\cosh(2\Lambda t) \cosh(10\Lambda t)}, \quad (4.34)$$

which is a smooth bounce that violates the NEC for a time $\Delta t \sim \Lambda^{-1}$ and produces a Hubble rate on the order of $H_m \sim \Lambda^3/M_{\text{Pl}}^2$ on exiting the NEC violating region. This is shown in Fig. 1. To ensure that the perturbed field on that background is well behaved, we can for instance choose

$$q(\Phi) = \frac{\Lambda^2}{M_{\text{Pl}}^2} \left(1 + q_1 \frac{\phi_0^4/\Phi^4}{(1 + \Phi/\phi_0)} \right). \quad (4.35)$$

As discussed in Sec. IV B, without a shift symmetry the model has to involve a hierarchy of scales and the ratio Λ/M_{Pl} is chosen to that effect (hence H_m , q , etc. are parametrically smaller than naively expected). This model gives the desired bouncing solution for $\phi(t)$, $H(t)$, when $\Lambda_c \rightarrow \infty$; however, in that limit fluctuations on this background have negative sound speed (there is a gradient instability), and it badly violates unitarity. To remedy this, we switch on the high-energy effects (i.e., bring Λ_c to a finite value) and include the irrelevant operator $(\square\Phi)^2$ at that scale. As derived previously, this scale should be in the appropriate range to save unitarity, but without spoiling the low-energy effective theory.

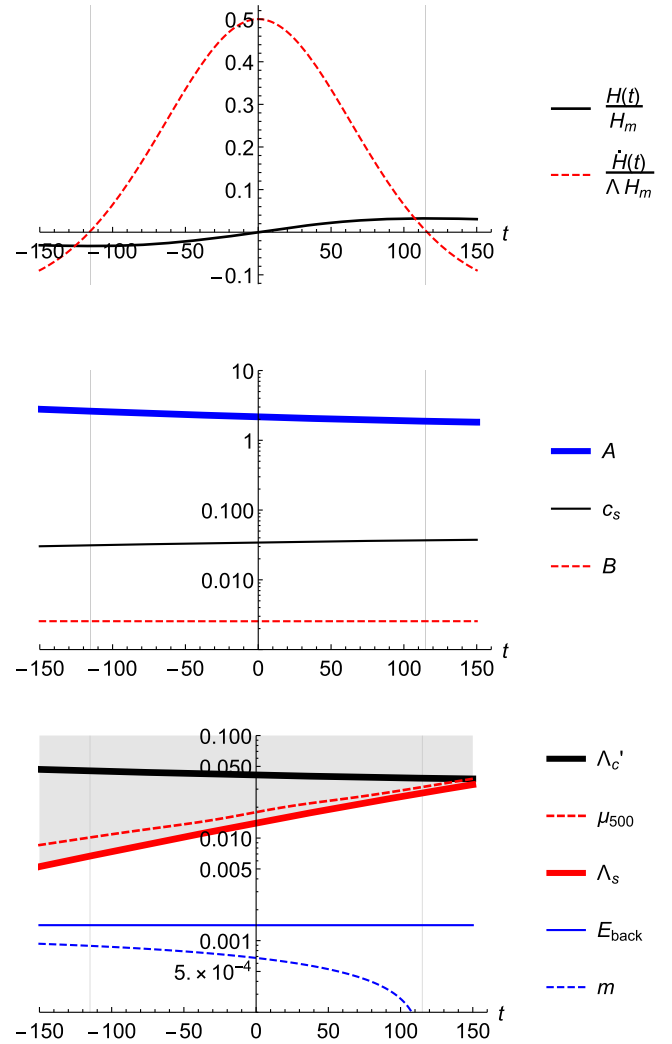


FIG. 1. The behavior of the specific example provided in (4.34). The time is given in Planck-scale units. The Hubble parameter is measured in terms of the scale $H_m = \Lambda^3/M_{\text{Pl}}^2$. The sound speed is manifestly positive throughout the NEC violating region. On the lower plot, the relevant scales of the system are represented relative to M_{Pl} . Λ_c' represents the scale at which the higher energy effects enter, ρ is the energy density of the background, m is the mass of the perturbed scalar field, and E_{back} is the background scale and is manifestly smaller than the scale Λ_s at which perturbative unitarity breaks down. That scale is dominated by the cubic and quartic operators $\dot{\phi}(\partial_i\phi)^2$ and $(\partial_i\phi)^4$. For comparison μ_{500} represents the scale at which the operator ϕ^5 breaks tree level unitarity. The example (4.34) is specifically engineered so that the contribution of that operator and any other operator is subdominant.

For concreteness we plot the behavior of the background and the strong coupling scale for the following specific choices of parameters:

$$\begin{aligned} \Lambda &= 10^{-3} M_{\text{Pl}}, & \phi_0 &= 0.5 M_{\text{Pl}}, \\ q_1 &= 6.4 & \text{and} & \quad \Lambda_c = 0.028 M_{\text{Pl}}. \end{aligned} \quad (4.36)$$

For this choice of parameters, we now have a finite positive sound speed and satisfy tree level unitarity throughout the bounce. The strong coupling scale is indeed set by the cubic and quartic operator $\dot{\varphi}(\partial_i\varphi)^2$ and $(\partial_i\varphi)^4$, while the higher order effective interaction scales are all much larger than the strong coupling scale Λ_s . This example has indeed been specifically engineered so as to suppress the effect of any other operator.

1. Strong coupling scale

An explicit calculation shows that within the NEC violating region, we have $A \sim 2$, $B \sim 2 \times 10^{-3}$ and the strong coupling scale deduced from the 2–2 scattering amplitude (including all the operators that would affect that amplitude) is $\Lambda_s \sim 10\Lambda$. Then we can check that the mass of the fluctuations on that background is indeed small,

$$m_{\text{eff}}^2 \sim \Lambda^2 \sim 10^{-2}\Lambda_s^2, \quad (4.37)$$

and the scale associated with the relevant operator φ^3 is sufficiently small, $\mu_{300} \sim \Lambda$, as is the dimensionless coupling constant in front of the marginal operator φ^4 , $\mu_{400} \sim \mathcal{O}(1)$. As for all the infinite number of irrelevant operators, their respective scales should be at least Λ_s . A direct calculation shows that within the region of interest (i.e., within the NEC violating region), all the requirements derived from tree level unitarity in (3.24) and (3.22) are indeed satisfied. For illustration purposes, we represent the scale μ_{500} we would have naively derived from where the quintic operator φ^5 breaks unitarity in Fig. 1 and it indeed lies above Λ_s throughout the bounce. The same remains true for all the other operators [other than $\dot{\varphi}(\partial_i\varphi)^2$ and $(\partial_i\varphi)^4$].

We can also directly see that the derivatives of the background remain small; obviously $\phi^{(n+1)}/\phi^{(n)} \sim \Lambda \ll \Lambda_s$ (the hierarchy involved in the example is not an important one, but it simply serves as an illustration of the principle). Moreover, we can check that the variation of the mass and the coefficients of the kinetic matrix are small, $\dot{B}/B \sim 10^{-6}M_{\text{Pl}}$, and $\dot{A}/A \sim 1.5 \times 10^{-3}M_{\text{Pl}}$, which is actually what sets the scale of the background, $E_{\text{back}} \sim 1.5\Lambda$.

We have therefore shown how, for couplings given by (4.28), and solution (4.34), while if one took $\Lambda_c \rightarrow \infty$ the bounce would have a classical gradient instability and would badly violate perturbative unitarity, this can be remedied by including high-energy effects that are present at sufficiently high energy (finite Λ_c) without severely affecting the predictions of the low-energy EFT. For a bounce to occur without violating unitarity in a $P(\Phi, X)$ theory, the parameters have to be carefully tuned and a hierarchy of scale has to be introduced already in the $P(\Phi, X)$ model.

V. SUMMARY

While solutions that violate the NEC condition are relatively easy to find classically, they may not be trusted if they are derived beyond the regime of validity of their effective field theory. In this work, we have derived the conditions set by tree level unitarity on NEC violating effective field theories on a cosmological background. In $P(\Phi, X)$ theories minimally coupled to gravity, without including any higher energy effects, it is impossible to describe a NEC violation (much less a complete bounce): Any classical solutions automatically severely violate unitarity and cannot be trusted. A natural resolution is to include the high-energy effects, irrelevant operators suppressed by a higher energy scale that are naturally expected to be present. We have shown that these can regulate not only the classical instabilities that arise in these classical NEC violating solutions, but can also regulate scattering amplitudes, hence providing a much better handle on unitarity. We have derived the precise requirements set by tree level unitarity in $P(\Phi, X)$ models with additional irrelevant operators, and shown that while the theory should be very carefully tuned, in principle, there is an open window of possibility for a stable cosmological bounce that preserves unitarity. To further illustrate the constraints set by unitarity and level of tuning required, we have presented an explicit $P(\Phi, X)$ that generates a stable cosmological bounce and preserves unitarity within the region of the bounce, albeit at the price of introducing an unnaturally small parameter into, and hence finely tuning, the EFT.

The analysis and example provided here was not aimed at providing a full cosmological framework and we have only focused on the possibility of violating the NEC while avoiding any instability and remaining within the regime of validity of the effective field theory. We have not addressed the question of particle production throughout the bounce, which is beyond the scope of this work; however, given that there is no instability and the scales are under control, there is *a priori* no reason to expect a large particle production at this point. Moreover, the high level of tuning required to obtain a stable and unitary bounce certainly raises the question of whether such a model would ever accommodate the precise cosmology we observe today and be able to reproduce the precise value of the spectral index (without an excess of non-Gaussianities, [49]) and the tensor to scalar ratio to be embedded in a consistent and viable model for the cosmological history of our Universe.

ACKNOWLEDGMENTS

We thank Anna Ijjas, Jean-Luc Lehners, Paul Steinhardt, and Andrew Tolley for very useful discussions and comments on the manuscript. C. d. R. is supported by a Royal Society Wolfson Merit Award. S. M. is supported by the von Clemm Fellowship and the Imperial College President's Scholarship.

APPENDIX: BOUNCING WITH A PURE GHOST CONDENSATE

In this appendix we review the violation of the NEC in a pure ghost-condensate model, [6,9,28,40]. Starting with the particular $P(X)$,

$$P(X) = -pX + \frac{q}{\Lambda^4} X^2, \quad (\text{A1})$$

a violation of the NEC is possible if $p > 0$. In that case the standard vacuum $\langle \phi \rangle = 0$ carries a ghost, but no ghost is present in the ghost-condensate phase ($\langle \dot{\phi} \rangle \neq 0$) where the quadratic terms X^2 become relevant. Just like any other $P(\Phi, X)$ theory, this model is unstable and breaks unitarity even before entering the NEC violating region, unless higher energy effects are considered as discussed in Sec. II B and so we have these high-energy effects entering at Λ_c in mind throughout this appendix.

The expression for the kinetic coefficients is

$$A = -p + \frac{6qX}{\Lambda^4} \quad \text{and} \quad B = -p + \frac{2qX}{\Lambda^4}, \quad (\text{A2})$$

so interestingly the variations of A and B are linked, $\dot{A} = 3\dot{B}$, and we have

$$X = \frac{\Lambda^4}{4q} (A - B). \quad (\text{A3})$$

As a consequence, we therefore have

$$\frac{\dot{X}}{X} = \frac{\dot{A} - \dot{B}}{A - B} = \frac{2\dot{B}}{A - B} \quad \text{and similarly} \quad \frac{\ddot{X}}{X} = \frac{2\ddot{B}}{A - B}. \quad (\text{A4})$$

Now from the Raychaudhuri equation, within the NEC violating region $\dot{H} > 0$, we have

$$0 < B < \mathcal{O}\left(\frac{1}{\Lambda_c^2} \frac{\ddot{X}}{X}\right), \quad (\text{A5})$$

where the exact expression on the right-hand side depends on the very precise operators that enter at Λ_c (and could

also for instance involve terms of the form $H\dot{X}/\Lambda_c^2 X$), but this analysis is independent of the precise form (as it should be). We merely use the fact that they involve higher derivatives as is required if those terms are to cure the instabilities associated with pure $P(X)$ bounce. Then from (A4) within the NEC violating region we ought to have

$$A - B < \mathcal{O}\left(\frac{1}{\Lambda_c^2} \frac{\ddot{B}}{B}\right). \quad (\text{A6})$$

Requiring that the background does not vary faster than $\Lambda_c' = \sqrt{A}\Lambda_c$ sets $\ddot{B}/B \ll A\Lambda_c^2$, which therefore implies that

$$1 - c_s^2 \ll 1, \quad (\text{A7})$$

i.e., the speed of sound should be very close to luminal. Now for a $P(X)$ model with close to luminal speed of sound the unitarity bound (4.21) cannot be satisfied.

One may be worried that the bound (4.21) is not technically valid if the speed of sound is not small since in going from (3.9) to (3.10) we have assumed $A \gg B$. It is straightforward to rederive the bound (4.21) when this assumption is relaxed and we then find that unitarity imposes

$$1 \ll \Lambda_c^4 \bar{P}_{,XX} \ll \frac{(1 - c_s^2)^4}{c_s}, \quad (\text{A8})$$

which is even more impossible to satisfy when $1 - c_s^2 \ll 1$. So we can conclude that a pure ghost-condensate model of the form (A1) can never give rise to a unitarity NEC violation (let alone a bounce) even when introducing higher derivative terms at a higher energy scale. The only way to avoid this argument in the ghost-condensate model is if the higher energy effects also involve operators that are not higher derivatives, i.e., involve terms that break the shift symmetry at high energy. Instead of trying to maintain a low-energy EFT that preserves the shift symmetry and only breaks that symmetry softly at high energy, in this manuscript we consider instead an explicit model that directly breaks the shift symmetry at low energy as is described in Sec. IV C.

-
- [1] J. D. Barrow, Nonsingular scalar-tensor cosmologies, *Phys. Rev. D* **48**, 3592 (1993).
 [2] M. Gasperini and G. Veneziano, Pre-big bang in string cosmology, *Astropart. Phys.* **1**, 317 (1993).
 [3] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok, The ekpyrotic universe: Colliding branes and the origin of the hot big bang, *Phys. Rev. D* **64**, 123522 (2001).

- [4] P. J. Steinhardt and N. Turok, Cosmic evolution in a cyclic universe, *Phys. Rev. D* **65**, 126003 (2002).
 [5] J. Khoury, B. A. Ovrut, N. Seiberg, P. J. Steinhardt, and N. Turok, From big crunch to big bang, *Phys. Rev. D* **65**, 086007 (2002).
 [6] N. Arkani-Hamed, H.-C. Cheng, M. A. Luty, and S. Mukohyama, Ghost condensation and a consistent infrared

- modification of gravity, *J. High Energy Phys.* **05** (2004) 074.
- [7] D. Battefeld and P. Peter, A critical review of classical bouncing cosmologies, *Phys. Rep.* **571**, 1 (2015).
- [8] P. Peter, E. J. C. Pinho, and N. Pinto-Neto, A Noninflationary model with scale invariant cosmological perturbations, *Phys. Rev. D* **75**, 023516 (2007).
- [9] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, New ekpyrotic cosmology, *Phys. Rev. D* **76**, 123503 (2007).
- [10] P. Peter and N. Pinto-Neto, Cosmology without inflation, *Phys. Rev. D* **78**, 063506 (2008).
- [11] M. Novello and S. E. P. Bergliaffa, Bouncing cosmologies, *Phys. Rep.* **463**, 127 (2008).
- [12] J.-L. Lehners, Ekpyrotic Non-Gaussianity—a review, *Adv. Astron.* **2010**, 1 (2010).
- [13] A. Ijjas and P. J. Steinhardt, The anamorphic universe, *J. Cosmol. Astropart. Phys.* **10** (2015) 001.
- [14] R. Brandenberger and P. Peter, Bouncing cosmologies: Progress and problems, [arXiv:1603.05834](https://arxiv.org/abs/1603.05834).
- [15] P. Peter and S. D. P. Viteni, The simplest possible bouncing quantum cosmological model, *Mod. Phys. Lett. A* **31**, 1640006 (2016).
- [16] Y. Cai, Y. Wan, H.-G. Li, T. Qiu, and Y.-S. Piao, The effective field theory of nonsingular cosmology, *J. High Energy Phys.* **01** (2017) 090.
- [17] Y. Cai, H.-G. Li, T. Qiu, and Y.-S. Piao, The effective field theory of nonsingular cosmology: II, [arXiv:1701.04330](https://arxiv.org/abs/1701.04330).
- [18] M. Visser and C. Barcelo, in *Proceedings, 3rd International Conference on Particle Physics and the Early Universe (COSMO 1999), Trieste, Italy, 1999* (World Scientific, Singapore, 2000), pp. 98–112.
- [19] A. E. Schulz and M. J. White, The tensor to scalar ratio of phantom dark energy models, *Phys. Rev. D* **64**, 043514 (2001).
- [20] L. F. Abbott, A mechanism for reducing the value of the cosmological constant, *Phys. Lett.* **150B**, 427 (1985).
- [21] L. Alberte, P. Creminelli, A. Khmelnitsky, D. Pirtskhalava, and E. Trincherini, Relaxing the cosmological constant: A proof of concept, *J. High Energy Phys.* **12** (2016) 022.
- [22] D. Hochberg and M. Visser, Geometric structure of the generic static traversable wormhole throat, *Phys. Rev. D* **56**, 4745 (1997).
- [23] D. Hochberg and M. Visser, Dynamic wormholes, anti-trapped surfaces, and energy conditions, *Phys. Rev. D* **58**, 044021 (1998).
- [24] D. Hochberg and M. Visser, The Null Energy Condition in Dynamic Wormholes, *Phys. Rev. Lett.* **81**, 746 (1998).
- [25] B. Holdom, Accelerated expansion, and the Goldstone ghost, *J. High Energy Phys.* **07** (2004) 063.
- [26] J. M. Cline, S. Jeon, and G. D. Moore, The phantom menaced: Constraints on low-energy effective ghosts, *Phys. Rev. D* **70**, 043543 (2004).
- [27] S. Dubovsky, T. Gregoire, A. Nicolis, and R. Rattazzi, Null energy condition and superluminal propagation, *J. High Energy Phys.* **03** (2006) 025.
- [28] P. Creminelli, M. A. Luty, A. Nicolis, and L. Senatore, Starting the Universe: Stable Violation of the null energy condition and nonstandard Cosmologies, *J. High Energy Phys.* **12** (2006) 080.
- [29] C. Deffayet, O. Pujolas, I. Sawicki, and A. Vikman, Imperfect dark energy from kinetic gravity braiding, *J. Cosmol. Astropart. Phys.* **10** (2010) 026.
- [30] O. Pujolas, I. Sawicki, and A. Vikman, The imperfect fluid behind kinetic gravity braiding, *J. High Energy Phys.* **11** (2011) 156.
- [31] G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, *Int. J. Theor. Phys.* **10**, 363 (1974).
- [32] C. Deffayet, S. Deser, and G. Esposito-Farese, Generalized Galileons: All scalar models whose curved background extensions maintain second-order field equations and stress tensors, *Phys. Rev. D* **80**, 064015 (2009).
- [33] D. A. Easson, I. Sawicki, and A. Vikman, G-bounce, *J. Cosmol. Astropart. Phys.* **11** (2011) 021.
- [34] V. A. Rubakov, Consistent NEC violation: Towards creating a universe in the laboratory, *Phys. Rev. D* **88**, 044015 (2013).
- [35] M. Libanov, S. Mironov, and V. Rubakov, Generalized Galileons: Instabilities of bouncing and Genesis cosmologies and modified Genesis, *J. Cosmol. Astropart. Phys.* **08** (2016) 037.
- [36] T. Kobayashi, Generic instabilities of nonsingular cosmologies in Horndeski theory: A no-go theorem, *Phys. Rev. D* **94**, 043511 (2016).
- [37] S. Akama and T. Kobayashi, Generalized multi-Galileons, covariantized new terms, and the no-go theorem for nonsingular cosmologies, *Phys. Rev. D* **95**, 064011 (2017).
- [38] A. Ijjas and P. J. Steinhardt, Classically Stable Nonsingular Cosmological Bounces, *Phys. Rev. Lett.* **117**, 121304 (2016).
- [39] A. Ijjas and P. J. Steinhardt, Fully stable cosmological solutions with a nonsingular classical bounce, *Phys. Lett. B* **764**, 289 (2017).
- [40] M. Koehn, J.-L. Lehners, and B. Ovrut, Nonsingular bouncing cosmology: Consistency of the effective description, *Phys. Rev. D* **93**, 103501 (2016).
- [41] A. Vikman, Can dark energy evolve to the phantom?, *Phys. Rev. D* **71**, 023515 (2005).
- [42] D. A. Easson and A. Vikman, The phantom of the new oscillatory cosmological phase, [arXiv:1607.00996](https://arxiv.org/abs/1607.00996).
- [43] L. Battarra, M. Koehn, J.-L. Lehners, and B. A. Ovrut, Cosmological perturbations through a nonsingular ghost-condensate/Galileon bounce, *J. Cosmol. Astropart. Phys.* **07** (2014) 007.
- [44] C. P. Burgess, Introduction to effective field theory, *Annu. Rev. Nucl. Part. Sci.* **57**, 329 (2007).
- [45] P. Creminelli, G. D’Amico, M. Musso, J. Norena, and E. Trincherini, Galilean symmetry in the effective theory of inflation: New shapes of non-Gaussianity, *J. Cosmol. Astropart. Phys.* **02** (2011) 006.
- [46] C. de Rham, L. Heisenberg, and R. H. Ribeiro, Ghosts and matter couplings in massive gravity, bigravity and multi-gravity, *Phys. Rev. D* **90**, 124042 (2014).
- [47] U. Aydemir, M. M. Anber, and J. F. Donoghue, Self-healing of unitarity in effective field theories and the onset of new physics, *Phys. Rev. D* **86**, 014025 (2012).
- [48] C. de Rham and R. H. Ribeiro, Riding on irrelevant operators, *J. Cosmol. Astropart. Phys.* **11** (2014) 016.
- [49] X. Gao, M. Lilley, and P. Peter, Non-Gaussianity excess problem in classical bouncing cosmologies, *Phys. Rev. D* **91**, 023516 (2015).