Loop corrections in gravitational wave spectrum in single field inflation

Hassan Firouzjahi [*]

School of Astronomy, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5531, Tehran, Iran

(Received 4 May 2023; accepted 7 August 2023; published 28 August 2023)

We study the one-loop corrections in the power spectrum of long gravitational waves induced from small scale modes in the models of single field inflation undergoing a phase of ultra-slow-roll (USR). We show that the spectrum of long tensor perturbations are largely unaffected by the loop corrections from the short scalar modes. In particular, the spectrum of long tensor perturbations is insensitive to the sharpness of the transition from the USR phase to the final slow-roll phase. This is in contrast to the case of scalar power spectrum in which the loop corrections can be large for a sharp transition while it is slow-roll suppressed in a mild transition. We study the tensor-scalar-scalar bispectrum in the squeezed limit and demonstrate that the Maldacena consistency condition does hold.

DOI: 10.1103/PhysRevD.108.043532

I. INTRODUCTION

Recently the question of one-loop corrections in the power spectrum of large CMB scale scalar perturbations from the small scale modes in the setup of single field inflation undergoing a phase of ultra-slow-roll (USR) was debated extensively [1–10] (for a related earlier work see [11]). This is particularly an important question since the models of single field inflation with an intermediate USR phase have been employed extensively in recent years as a viable mechanism to generate primordial black holes (PBHs) which may comprise all or parts of cold dark matter [12–14] (for a review see [15,16]). More specifically, to have a successful mechanism of PBH formation, one requires the amplitude of curvature perturbations to be enhanced by a factor of 10^7 or so in the allowed small scales compared to the large CMB scales. It turns out that an intermediate phase of USR inflation can provide this enhancement naturally.

The USR setup is a phase of inflation in which the potential is very flat [17–19]. Consequently, the inflaton velocity falls off exponentially and the curvature perturbations grow on superhorizon scales [20]. As the curvature perturbations grow on superhorizon scales, it provides a nontrivial example for the violation of the celebrated Maldacena consistency condition [21,22] for the non-Gaussianity of single field inflation [20,23–30]. More specifically, it was shown in [20] that the amplitude of local-type non-Gaussianity in the USR model is $f_{NL} = \frac{5}{2}$. This question was further investigated in [31] in which it was demonstrated that the final amplitude of f_{NL} crucially depends on the sharpness of the transition from the USR phase to the final slow-roll (SR) phase. In particular, in an extreme sharp transition from the USR phase to the SR phase, as assumed in [20], f_{NL} reaches its maximum value $\frac{5}{2}$. However, if the transition is mild, then the curvature perturbations evolve after the USR phase until it reaches its final attractor value. Correspondingly, much of the amplitude of f_{NL} is washed out, and it ends up at a value of the order of the slow-roll parameters though the Maldacena consistency condition is still violated. The lesson is that the sharpness of the transition from the USR phase to the final SR phase plays important roles to read off the amplitude of cosmological observables at the end of inflation.

Originally, it was argued in [1] (see also [2]) that the one-loop corrections from small USR modes can significantly affect the large CMB scale modes. Therefore, it was argued that to keep these loop corrections under perturbative control, the model loses its applicability to generate the desired PBH abundance. This conclusion was criticized in [3,4] where it was advocated that this conclusion is model-dependent and the dangerous one-loop corrections can be harmless in a smooth transition. This question was further investigated in [8] in a consistent manner where the effects of both cubic and quartic Hamiltonians were taken into account. While the analysis in [8] supported the conclusion of [1] for the setup with a sharp transition, it was argued that the situation can be very different in a mild transition. Finally, this question was further studied in [10] where, using δN formalism, it was shown that for a mild transition the one-loop corrections are suppressed by the slow-roll parameters and the setup can still be viable for PBH formation, in agreement with [3,4]. The conclusion from these works, as in the old story of f_{NL} alluded to before, is that the amplitude of one-loop corrections

[*]firouz@ipm.ir

crucially depends on the sharpness of the transition from the USR phase to the final SR phase. For a physical smooth transition, the dangerous one-loop corrections are washed out during the subsequent evolutions of the modes after the USR phase.

With the above discussions in mind, in this work we extend the motivation of [1] and calculate the one-loop correction from small USR modes on large CMB scale gravitational waves (GWs) perturbations.[1] On the physical ground, similar to the reasonings of [3,4], it is expected that the tensor perturbations will be less sensitive to the USR phase transition. This is because the amplitude of GWs are determined by the Hubble scale, H, during inflation. As the value of H is not much modified during the USR transition, then the background for GWs propagation is not much modified either. Add to it the important effect that the tensor perturbations are frozen on superhorizon scales at the linear level in perturbation theory [37–40]. However, the lesson of large loop corrections in a sharp transition for the case of a scalar power spectrum sets a nontrivial example to examine more directly the validity of the above physical expectations for the long GWs. This is the goal of this work.

II. THE SETUP

Here we briefly review our setup and present the formulas which will be required for our subsequent analysis.

We consider a three-phase model of inflation in which a USR phase is sandwiched between two phases of SR inflation (SR → USR → SR). The early SR phase is when the large CMB scale mode leaves the horizon. The USR phase is extended in the interval $t_i \leq t \leq t_e$ in which the potential is flat $V(\phi) = V_0$. The background equations during the USR phase are

$$\ddot{\phi}(t) + 3H\dot{\phi}(t) = 0, \qquad 3M_P^2 H^2 \simeq V_0, \tag{1}$$

where M_P is the reduced Planck mass and H is the Hubble expansion rate during inflation. During the USR phase H is very nearly constant while $\dot{\phi} \propto \frac{1}{a^3}$.

The two slow-roll parameters related to H are given as follows:

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{\dot{\phi}^2}{2M_P^2 H^2}, \qquad \eta \equiv \frac{\dot{\epsilon}}{H\epsilon}. \tag{2}$$

Since ϵ falls off like a^{-6} during the USR setup, we see that $\eta \simeq -6$ which is the hallmark of the USR inflation [17]. Going to conformal time $d\tau = dt/a(t)$ with $aH\tau \simeq -1$, the evolution of ϵ is given by

$$\epsilon(\tau) = \epsilon_i \left(\frac{\tau}{\tau_i}\right)^6, \tag{3}$$

in which ϵ_i is the value of ϵ at the start of the USR phase. Correspondingly, at the end of the USR phase $\epsilon_e = \epsilon_i (\frac{\tau_e}{\tau_i})^6$. Using the number of e-folds, $dN = Hdt$, the duration of the USR phase is denoted by $\Delta N \equiv N(\tau_e) - N(\tau_i)$ so $\epsilon_e = e^{-6\Delta N}\epsilon_i$.

As shown in [8], a crucial role is played by the sharpness of the transition from the USR phase to the final SR phase. To take this into account, following [31], we define the parameter associated with the sharpness of the transition, h, as follows:

$$h \equiv \frac{6\sqrt{2\epsilon_V}}{\dot{\phi}(t_e)} = -6\sqrt{\frac{\epsilon_V}{\epsilon_e}}. \tag{4}$$

Here, ϵ_V represents the slow-roll parameter at the final SR phase when the system reaches its attractor regime. Since we assume (without lack of generality) that ϕ is decreasing during the USR phase, then $\dot{\phi} < 0$ so $h < 0$.

As shown in [31] near the transition we can approximate η as

$$\eta = -6 - h\theta(\tau - \tau_e), \qquad \tau_e^- < \tau < \tau_e^+. \tag{5}$$

In particular, for the derivative of η, we have

$$\frac{d\eta}{d\tau} = -h\delta(\tau - \tau_e), \qquad \tau_e^- < \tau < \tau_e^+. \tag{6}$$

In the following analysis we consider two cases of sharp transition: a "natural" sharp transition in which η drops to zero immediately after the transition corresponding to $h = -6$. In this situation ϵ after the transition is frozen to its value at the end of USR given by ϵ_e. This limit was studied in [1,2]. The other case is an "extreme" sharp transition where $|h| \gg 1$. In this situation, ϵ after the transition evolves toward the end of inflation (or when the evolution in the final stage has reached its attractor phase) so $\epsilon_V = \epsilon_e (\frac{h}{6})^2$.

As $\epsilon(\tau)$ falls off exponentially during the USR phase, the comoving curvature perturbation $\mathcal{R}(\tau)$ grows exponentially during the USR phase, $\mathcal{R}(\tau) \propto a(\tau)^3 \propto \tau^{-3}$. After the USR period, the curvature perturbation may evolve during the final USR phase until it reaches its final attractor value to be measured at the end of inflation. To read off the final value of $\mathcal{R}$, we have to track it from the first phase of inflation toward the USR phase and then eventually into the final SR phase. This is achieved by requiring that both $\mathcal{R}(\tau)$ and $\mathcal{R}'(\tau)$ be continuous across the transitions SR → USR → SR.

Starting with a Bunch-Davies initial condition in the first SR phase, the mode function in the Fourier space is given by

$$\mathcal{R}_k^{(1)} = \frac{H}{M_P\sqrt{4\epsilon_i k^3}}(1 + ik\tau)e^{-ik\tau} \qquad (\tau < \tau_i), \tag{7}$$

[1]For earlier works concerning the loop corrections in the tensor power spectrum during inflation see [32–36].

where ϵ_i is the value of the slow-roll parameter at the start of inflation when the CMB scale mode leaves the horizon. The superscript (1) indicates the first SR phase. During the USR phase, the mode function is given formally by the superposition of the positive and negative frequency modes,

$$\mathcal{R}_k^{(2)} = \frac{H}{M_P\sqrt{4\epsilon_i k^3}}\left(\frac{\tau_i}{\tau}\right)^3\left[\alpha_k^{(2)}(1+ik\tau)e^{-ik\tau} + \beta_k^{(2)}(1-ik\tau)e^{ik\tau}\right], \tag{8}$$

with the coefficients $\alpha_k^{(2)}$ and $\beta_k^{(2)}$, after imposing the matching condition at $\tau = \tau_i$, and they are obtained to be

$$\alpha_k^{(2)} = 1 + \frac{3i}{2k^3\tau_i^3}(1+k^2\tau_i^2), \qquad \beta_k^{(2)} = -\frac{3i}{2k^3\tau_i^3}(1+ik\tau_i)^2 e^{-2ik\tau_i}. \tag{9}$$

Finally, imposing the matching conditions at τ_e, the mode function in the final SR phase, denoted by the superscript (3), is obtained to be

$$\mathcal{R}_k^{(3)} = \frac{H}{M_P\sqrt{4\epsilon(\tau)k^3}}\left[\alpha_k^{(3)}(1+ik\tau)e^{-ik\tau} + \beta_k^{(3)}(1-ik\tau)e^{ik\tau}\right], \tag{10}$$

with the coefficients $\alpha_k^{(3)}$ and $\beta_k^{(3)}$ given by

$$\alpha_k^{(3)} = \frac{1}{8k^6\tau_i^3\tau_e^3}\left[3h(1-ik\tau_e)^2(1+ik\tau_i)^2 e^{2ik(\tau_e-\tau_i)} + (-2ik^3\tau_i^3 + 3k^2\tau_i^2 + 3)(4ik^3\tau_e^3 - hk^2\tau_e^2 - h)\right]$$

and

$$\beta_k^{(3)} = \frac{-1}{8k^6\tau_i^3\tau_e^3}\left[3(1+ik\tau_i)^2(h + hk^2\tau_e^2 + 4ik^3\tau_e^3)e^{-2ik\tau_i} - h(1+ik\tau_e)^2(3 + 3k^2\tau_i^2 - 2ik^3\tau_i^3)e^{-2ik\tau_e}\right].$$

Finally, the power spectrum of curvature perturbations at the end of inflation $\tau = \tau_0 \to 0$ for the mode in the interval $k_i < k < k_e$ which leaves the horizon during the USR phase given by

$$P_{\mathcal{R}}(\tau_0, k) = \left(\frac{h-6}{h}\right)^2 \frac{H^2}{4M_P^2\epsilon_e k^3} = \left(\frac{h-6}{h}\right)^2 P_{\mathcal{R}}(\tau_e, k) \qquad (k_i < k < k_e). \tag{11}$$

Curiously, we see that the power spectrum is scaled with a factor $(\frac{h-6}{h})^2$ compared to its value at the end of the USR phase. In the limit of extreme sharp transition, $h \to -\infty$, we see that $P_{\mathcal{R}}(\tau_0, k) \simeq P_{\mathcal{R}}(\tau_e, k)$. This is expected, since in this limit the mode function is frozen immediately after the USR phase and does not experience evolution after the USR phase. On the other hand, for the case of a natural sharp transition with $h = -6$, we see that $P_{\mathcal{R}}(\tau_0, k) \simeq 4P_{\mathcal{R}}(\tau_e, k)$ so the power spectrum actually becomes larger toward the end of inflation. This is because the mode function is still evolving after the USR phase until it reaches its final attractor value. We comment that there are subleading corrections of order $O(\frac{k^2}{k_e^2})$ in Eq. (11) which we have neglected.

On the other hand, the modes which leave the horizon during the first SR phase are frozen during the intermediate USR phase. Correspondingly, for these modes (at the tree level) we have

$$P_{\mathcal{R}}(\tau_0, k) = \frac{H^2}{4M_P^2\epsilon_i k^3} \qquad (k < k_i). \tag{12}$$

III. CUBIC AND QUARTIC HAMILTONIANS

Our goal is to calculate the one-loop corrections in the tensor power spectrum induced by the scalar perturbations which experience a growth during the USR phase. For this purpose, we need to calculate the cubic and quartic interaction Hamiltonians. Schematically, the cubic Hamiltonian represents an interaction of the type $\gamma\mathcal{R}^2$ while the quartic Hamiltonian is in the form $\gamma^2\mathcal{R}^2$. A schematic view of the corresponding one-loop diagrams associated with these interactions are presented in Fig. 1. The left panel in Fig. 1 represents the contribution of the cubic Hamiltonian involving a nested in-in integral while the right panel represents the contribution of the quartic Hamiltonian involving a single in-in integral.

We consider the tensor perturbations of the Friedmann-Lemaitre-Robertson-Walker (FLRW) background as follows:

$$ds^2 = -dt^2 + g_{ij}dx^i dx^j, \qquad g_{ij} \equiv a(t)^2\hat{h}_{ij}, \tag{13}$$

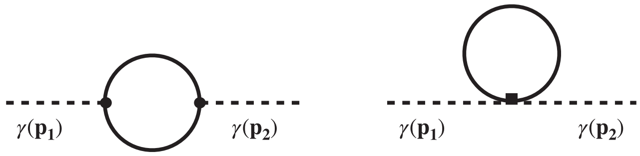

FIG. 1. The Feynman diagrams for the one-loop correction in the tensor power spectrum. The dotted line represents the tensor perturbations while the solid line in the loop represents the scalar perturbations. The left and right panels represent the contribution of the cubic and quartic Hamiltonians, respectively.

in which $\hat{h}_{ij}$ is expanded in terms of the tensor perturbations γ_{ij} as follows [21]:

$$\hat{h}_{ij} = \delta_{ij} + \gamma_{ij} + \frac{1}{2}\gamma_{i\ell}\gamma_{\ell j} + \cdots. \tag{14}$$

The tensor perturbations are transverse and traceless, $\gamma^i_i = \partial_i\gamma_{ij} = 0$ in which the indices are raised via δ^{ij}. With this construction, there is no contribution of γ_{ij} in $\sqrt{-g}$.

The total action is $S_{\rm total} = S_{\rm matter} + S_{\rm EH}$ in which $S_{\rm matter}$ is the matter part of the action while $S_{\rm EH}$ represents the usual Einstein-Hilbert action. To calculate the leading interaction Hamiltonian, we use the effective field theory (EFT) of inflation [41,42]. In a near de Sitter (dS) spacetime with a background inflaton field $\phi(t)$, the four-dimensional diffeomorphism invariance is spontaneously broken to a three-dimensional spatial diffeomorphism invariance. Starting with the unitary (or comoving) gauge where the perturbations of inflaton are turned off, one is allowed to write down all terms in the action which are consistent with the remaining three-dimensional diffeomorphsim invariance. Upon doing so, the background inflation dynamics is controlled via the known Hubble expansion rate $H(t)$ and its derivative $\dot{H}(t)$. After writing the full action consistent with the three-dimensional diffeomorphsim invariance, one restores the full four-dimensional diffeomorphsim invariance by introducing scalar field fluctuations, $\pi(x^\mu)$, which is the Goldstone boson associated with the breaking of the time diffeomorphsim invariance. One big advantage of the EFT approach is when one works in the decoupling limit where the gravitational backreactions are neglected. In this limit one neglects the slow-roll suppressed interactions in cubic and quartic actions while keeping only the leading terms which can yield large non-Gaussianities. In our study concerning the USR setup, these are the interactions which induce large corrections in one-loop integrals. For earlier work employing the EFT approach for the bispectrum analysis in a general nonattractor setup (including the USR setup) see [43]. The EFT approach was employed in [8] to study the one-loop corrections in the scalar power spectrum.

Assuming we have a canonical scalar field with a sound speed $c_s = 1$, the matter part of the action consistent with the FLRW inflationary background is given by [41]

$$S_{\rm matter} = \int d^4x\sqrt{-g}\left[-M_P^2\dot{H}(t+\pi)\left(\frac{1}{N^2}(1+\dot{\pi}-N^i\partial_i\pi)^2 - g^{ij}\partial_i\pi\partial_j\pi\right) - M_P^2\left(3H^2(t+\pi)+\dot{H}(t+\pi)\right)\right], \tag{15}$$

in which N and N^i are the lapse and shift function in the standard Arnowitt-Deser-Misner (ADM) formalism. In the decoupling limit where the gravitational backreactions are neglected we set $N=1$, $N^i=0$, and $\sqrt{-g}=a^3$. Our goal is to read off the interaction between π and γ_{ij}. Since γ_{ij} does not contribute into $\sqrt{-g}$, the coupling between π and γ_{ij} to leading order comes via the interaction $g^{ij}\partial_i\pi\partial_j\pi$. On the other hand, to quadratic order, we have

$$g^{ij} = a^{-2}\left(\delta_{ij} - \gamma_{ij} + \frac{1}{2}\gamma_{i\ell}\gamma_{\ell j}\right), \tag{16}$$

where in the right-hand side above, we raise and lower the indices via δ_{ij}. Correspondingly, the interaction between π and γ_{ij} to quartic order has the following terms:

$$g^{ij}\partial_i\pi\partial_j\pi \to -\gamma_{ij}\partial_i\pi\partial_j\pi + \frac{1}{2}\gamma_{i\ell}\gamma_{\ell j}\partial_i\pi\partial_j\pi. \tag{17}$$

On the other hand, expanding $\dot{H}(t+\pi)$ to first order in π we have

$$\begin{aligned}\dot{H}(t+\pi) &= \dot{H} + \ddot{H}\pi + \cdots,\\ &\simeq -\epsilon H^2 - \epsilon\eta H^3\pi.\end{aligned} \tag{18}$$

It is important to note that in the USR setup $\eta \simeq -6$, we cannot discard the last term above.

Plugging Eqs. (17) and (18) into the action (15) the cubic action is obtained to be [44]

$$S_{\gamma\pi^2} = M_P^2H^2\int d\tau d^3x\epsilon a^2\gamma_{ij}\partial_i\pi\partial_j\pi, \tag{19}$$

while the quartic action is given by

$$S_{\gamma^2\pi^2} = M_P^2H^2\int d\tau d^3x\epsilon a^2\left[-\frac{1}{2}\gamma_{i\ell}\gamma_{\ell j}\partial_i\pi\partial_j\pi + \eta\pi\gamma_{ij}\partial_i\pi\partial_j\pi\right]. \tag{20}$$

Correspondingly, the cubic and quartic interaction Hamiltonians are

$$\mathbf{H_3} = -M_P^2 H^2 \int d^3x \epsilon a^2 \gamma_{ij} \partial_i \pi \partial_j \pi \tag{21}$$

and

$$\mathbf{H_4} = M_P^2 H^2 \int d^3x \epsilon a^2 \left[\frac{1}{2}\gamma_{i\ell}\gamma_{\ell j}\partial_i \pi \partial_j \pi - \eta \gamma_{ij} \pi \partial_i \pi \partial_j \pi\right]. \tag{22}$$

As we see, the quartic Hamiltonian has two terms. One can easily check that the second term above, containing $\gamma_{ij}\pi\partial_i\pi\partial_j\pi$, does not contribute to the graviton power spectrum at the one-loop level while it contributes to the graviton power spectrum at the two-loop level. Therefore, in the following analysis where we study the one-loop correction in the graviton power spectrum, we neglect the effects of the second term in $\mathbf{H_4}$.

From the above interaction Hamiltonians we see that both $\mathbf{H_3}$ and $\mathbf{H_4}$ contain spatial derivatives of the scalar perturbations. This is required because the tensor perturbations carry the indices i, j so they should be contracted with the spatial derivatives of the scalar perturbations. Consequently, one expects that the induced loop corrections in the tensor power spectrum will be suppressed compared to the case of the scalar power spectrum. However, the amplitude of one-loop corrections in the tensor spectrum has yet to be calculated.

Finally, note that curvature perturbations $\mathcal{R}$ are related to π via [8]

$$\mathcal{R} = -H\pi + O(\pi^2), \tag{23}$$

in which the higher order terms contain the derivatives of π or H [45,46]. However, we calculate the two-point correlation functions at the end of inflation $\tau = \tau_0 \to 0$ where it is assumed that the system is in the slow-roll regime and the perturbations are frozen on superhorizon scales. In this case, the higher order corrections in Eq. (23) are suppressed, and we can simply use the linear relation between $\mathcal{R}$ and π in the following in-in integrals [8].

Going to Fourier space, the tensor perturbations are expended as follows:

$$\gamma_{ij}(x) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \sum_{s=\pm} \epsilon_{ij}^s(\mathbf{k}) \gamma_{\mathbf{k}}^s e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{24}$$

in which $s = \pm$ are two polarizations of the tensor perturbation. The polarization tensor is transverse and traceless, $\epsilon_{ii} = k^i \epsilon_{ij} = 0$, and satisfies

$$\epsilon_{ij}^{s*}(\mathbf{k}) = \epsilon_{ij}^s(-\mathbf{k}), \qquad \epsilon_{ij}^s(\mathbf{k})\epsilon_{ij}^{s'*}(\mathbf{k}) = 2\delta_{ss'}. \tag{25}$$

As an example of the polarization tensor, taking $\hat{\mathbf{k}}$ along the third direction, we choose [37]

$$\epsilon_{11}(\hat{z}, \pm 2) = -\epsilon_{22}(\hat{z}, \pm 2) = \mp i\epsilon_{12}(\hat{z}, \pm 2) = \mp i\epsilon_{21}(\hat{z}, \pm 2) = \frac{1}{\sqrt{2}}, \quad \epsilon_{i3} = \epsilon_{3i} = 0. \tag{26}$$

To quantize the free tensor perturbation, as usual we expand the Einstein-Hilbert action to quadratic order in γ_{ij} obtaining [38]

$$S_{\gamma^2} = \frac{M_P^2}{8} \int d\tau d^3x a^2 [(\gamma'_{ij})^2 - (\nabla\gamma_{ij})^2]. \tag{27}$$

Expanding the quantum operators in terms of the corresponding creation and annihilation operators as

$$\gamma_{\mathbf{k}}^s = b_{\mathbf{k}}^s \gamma_k(\tau) + b_{-\mathbf{k}}^{s\dagger} \gamma_k(\tau)^*, \tag{28}$$

with the usual commutation relation $[b_{\mathbf{k}}^{s_1}, b_{\mathbf{k}'}^{s_2}] = \delta^{s_1 s_2}\delta^3(\mathbf{k} - \mathbf{k}')$, the mode function is given by

$$\gamma_k(\tau) = \frac{H\sqrt{2}}{M_P k^{\frac{3}{2}}}(1 + ik\tau)e^{-ik\tau}. \tag{29}$$

Correspondingly, the two-point correlation is given by

$$\langle \gamma_{\mathbf{k}}^s \gamma_{\mathbf{k}'}^{s'} \rangle = \frac{\delta^{ss'}}{2} P_\gamma(k) = \frac{2H^2}{k^3 M_P^2}\delta^{ss'}, \tag{30}$$

with the dimensionless tensor power spectrum given by

$$\mathcal{P}_\gamma = \frac{k^3}{2\pi^2} P_\gamma(k) = \frac{2H^2}{\pi^2 M_P^2}. \tag{31}$$

To calculate the loop corrections, we employ the standard in-in formalism [47] in which the expectation value of the operator $\hat{O}$ at the end of inflation τ_0 is given by the Dyson series,

$$\langle \hat{O}(\tau_0) \rangle = \left\langle \left[\bar{\mathrm{T}} \exp\left(i\int_{-\infty}^{\tau_0} d\tau' H_{\mathrm{in}}(\tau')\right)\right] \times \hat{O}(\tau_0) \left[\mathrm{T}\exp\left(-i\int_{-\infty}^{\tau_0} d\tau' H_{\mathrm{in}}(\tau')\right)\right] \right\rangle, \tag{32}$$

in which T and $\bar{\mathrm{T}}$ represent the time ordering and anti-time ordering, respectively, while $H_{\mathrm{in}}(t)$ collectively represents the interaction Hamiltonian. In our case at hand $H_{\mathrm{in}}(\tau) = \mathbf{H}_3 + \mathbf{H}_4$.

IV. TENSOR-SCALAR-SCALAR CONSISTENCY CONDITION

While our main goal is to calculate the one-loop corrections in the tensor power spectrum, as a prelude here we study the bispectrum of $\langle \gamma^\lambda_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle$ in the squeezed limit $k_1 \ll k_2 \simeq k_3$. This is mainly to check that our EFT approach with the interaction Hamiltonians given above are trusted for the one-loop corrections in the tensor power spectrum. While this analysis is interesting and new (in the current SR → USR → SR setup), the reader who is only interested in loop corrections can skip directly to the next section.

To calculate $\langle \gamma^\lambda_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle$ in the squeezed limit we assume that the tensor perturbation has left the horizon during the first SR phase while the scalar perturbations have left the horizon during the intermediate USR phase. As such, the hierarchy $k_1 \to 0$ and $k_2 \simeq k_3$ is assumed. On the physical ground, as the tensor mode is frozen on the superhorizon scale, we expect a consistency condition similar to that of Maldacena [21] for the tensor-scalar-scalar to hold.

To calculate $\langle \gamma^\lambda_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle$ at the tree level, we only need the cubic interaction Hamiltonian $\mathbf{H_3}$. Plugging $\mathbf{H_3}$ from Eq. (21) in the in-in integral (32), we have

$$\langle \gamma^\lambda_{\mathbf{k}_1}(\tau_0)\mathcal{R}_{\mathbf{k}_2}(\tau_0)\mathcal{R}_{\mathbf{k}_3}(\tau_0)\rangle = -2\mathrm{Im}\int_{-\infty}^{\tau_0} d\tau \langle \mathbf{H_3}(\tau)\gamma^\lambda_{\mathbf{k}_1}(\tau_0)\mathcal{R}_{\mathbf{k}_2}(\tau_0)\mathcal{R}_{\mathbf{k}_3}(\tau_0)\rangle. \quad (33)$$

Using the linear relation $\mathcal{R} = -H\pi$, and noting that $\mathbf{k}_2 \simeq -\mathbf{k}_3$, we obtain

$$\langle \gamma^\lambda_{\mathbf{k}_1}(\tau_0)\mathcal{R}_{\mathbf{k}_2}(\tau_0)\mathcal{R}_{\mathbf{k}_3}(\tau_0)\rangle' = -4M_P^2\epsilon^\lambda_{ij}(\mathbf{k}_1)\hat{\mathbf{k}}_{2i}\hat{\mathbf{k}}_{2j}\mathcal{I}, \quad (34)$$

in which here and below a prime over $\langle\cdots\rangle$ means we have pulled out the overall factor $(2\pi)^3\delta^3(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)$. The factor $\mathcal{I}$ is calculated via the in-in integral as follows:

$$\mathcal{I} \equiv k_2^2\int_{-\infty}^{\tau_0} d\tau\epsilon(\tau)a^2\mathrm{Im}\left[\gamma^*_{k_1}(\tau_0)\mathcal{R}^*_{k_2}(\tau_0)^2\gamma_{k_1}(\tau)\mathcal{R}_{k_2}(\tau)^2\right]. \quad (35)$$

As the scalar modes leave the horizon during the USR period, there are two contributions in the above integral, from the USR period $\tau_i < \tau < \tau_e$ and after the USR period $\tau_e < \tau < \tau_0$. Performing the integral over the USR period and neglecting the contribution of a rapidly oscillating term in the form of $\cos(2k_2\tau_i)$, we obtain

$$\mathcal{I}(\tau_i < \tau < \tau_e) = -\frac{3}{4}\left(\frac{h-6}{h}\right)^2\frac{H^4}{k_1^3k_2^3M_P^4\epsilon_e} + \mathcal{O}\left(\frac{k_2^2}{k_e^2}\right). \quad (36)$$

On the other hand, calculating $\mathcal{I}$ for the period $\tau_e < \tau < \tau_0$ we obtain

$$\mathcal{I}(\tau_e < \tau < \tau_0) = -\frac{(6-h)(h-10)}{10h^2}\frac{H^4}{k_1^3k_2^3M_P^4\epsilon_e}\times\left(\frac{k_2^2}{k_e^2}\right). \quad (37)$$

For the modes which $k_2 \ll k_e$, we may neglect the contribution $\mathcal{I}(\tau_e < \tau < \tau_0)$ and to leading order

$$\langle \gamma^\lambda_{\mathbf{k}_1}(\tau_0)\mathcal{R}_{\mathbf{k}_2}(\tau_0)\mathcal{R}_{\mathbf{k}_3}(\tau_0)\rangle' = \frac{3}{4}\epsilon^\lambda_{ij}(\mathbf{k}_1)\hat{\mathbf{k}}_{2i}\hat{\mathbf{k}}_{2j}P_\mathcal{R}(k_2,\tau_0)P_\gamma(k_1,\tau_0), \quad (38)$$

in which $P_\mathcal{R}(k_2,\tau_0)$ and $P_\gamma(k_1,\tau_0)$ are the scalar and tensor power spectrum as given in Eqs. (11) and (30).

The above result is obtained employing a direct in-in calculation. However, as the tensor mode is frozen on superhorizon scales and is not affected by the USR phase, we expect a consistency condition similar to [21] to hold. Below we demonstrate that this is indeed the case.

As $k_1 \to 0$, one can assume that the long tensor mode only modifies the background for the short scalar modes [21] in the form of a quadrupolar anisotropy by changing $k_2^2 \to k_2^2 - \gamma_{ij}k_2^ik_2^j$. Following the logic of [21] we can write

$$\langle \gamma^\lambda_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle' \simeq -\langle \gamma^\lambda_{\mathbf{k}_1}\gamma^\lambda_{\mathbf{k}_1}\rangle\epsilon^\lambda_{ij}(\mathbf{k}_1)\hat{\mathbf{k}}_{2i}\hat{\mathbf{k}}_{2j}\frac{\partial}{\partial k_2^2}\langle\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle. \quad (39)$$

Using the specific form of the scalar power spectrum given in Eq. (11) we have

$$\frac{\partial}{\partial k_2^2}P_\mathcal{R}(k_2) = -\frac{3}{2k_2^2}P_\mathcal{R}(k_2), \quad (40)$$

and consequently, plugging this into Eq. (39), we obtain

$$\langle \gamma^\lambda_{\mathbf{k}_1}(\tau_0)\mathcal{R}_{\mathbf{k}_2}(\tau_0)\mathcal{R}_{\mathbf{k}_3}(\tau_0)\rangle' = \frac{3}{4}\epsilon^\lambda_{ij}(\mathbf{k}_1)\hat{\mathbf{k}}_{2i}\hat{\mathbf{k}}_{2j}P_\mathcal{R}(k_2,\tau_0)P_\gamma(k_1,\tau_0), \quad (41)$$

in exact agreement with Eq. (38).

As explained above, one expects the above consistency condition to hold. This is because the tensor perturbation has left the horizon during the early SR phase which is frozen afterwards and is largely unaffected by the USR phase. Consequently, it can only modify the background for the short scalar modes, which leave the horizon much later in the USR phase, in a form of quadrupolar anisotropy.

The above analysis confirms the applicability of our EFT approach. In addition, as the consistency condition is unaffected, the above results imply that the loop corrections from the short scalar perturbations will be minimal on long tensor perturbations which have left the horizon much earlier. We study this issue more directly in the next section.

V. LOOP CORRECTIONS IN TENSOR POWER SPECTRUM

Now we study the one-loop corrections in long CMB scale gravitational power spectrum $\langle \gamma^{s_1}(\mathbf{p}_1)\gamma^{s_2}(\mathbf{p}_2)\rangle$ induced from the short scalar modes which leave the horizon during the intermediate USR phase. In our convention the CMB scale tensor modes have momentum $\mathbf{p}_1$ and $\mathbf{p}_2$ while that of short scalar perturbations running in the loop is $\mathbf{q}$.

For a consistent one-loop corrections, we have to calculate the contributions of both Feynman diagrams shown in Fig. 1. We start with the right panel which is easier, containing a four vertex involving one in-in integral over the quartic Hamiltonian $\mathbf{H_4}$.

A. Loop corrections from quartic Hamiltonian

With the quartic Hamiltonian given in Eq. (22) the one-loop correction from the right panel of Fig. 1 is given by

$$\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{\mathbf{H_4}} = -2\mathrm{Im}\int_{-\infty}^{\tau_0} d\tau \langle \mathbf{H_4}(\tau)\gamma_{s_1}(\mathbf{p}_1,\tau_0)\gamma_{s_2}(\mathbf{p}_1,\tau_0)\rangle, \quad (42)$$

yielding

$$\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H_4}} = -2M_P^2\mathrm{Im}\left[\epsilon^{s_1}_{i\ell}(-\mathbf{p}_1)\epsilon^{s_2}_{\ell j}(\mathbf{p}_1)\int \frac{d^3\mathbf{q}}{(2\pi)^3} q_i q_j \mathcal{I}_4(q)\right], \quad (43)$$

in which the factor $\mathcal{I}_4(q)$ associated with the quartic Hamiltonian in-in integral will be given shortly below.

Using the isotropy of the background, the integral $\int d^3\mathbf{q} q_i q_j \mathcal{I}_4(q)$ is nonzero only $i=j$ so one can replace this momentum integral by $\frac{1}{3}\delta_{ij}\int d^3\mathbf{q} q^2 \mathcal{I}_4(q)$. Now using the properties of the polarization tensor given in Eq. (25) we obtain

$$\epsilon^{s_1}_{i\ell}(-\mathbf{p}_1)\epsilon^{s_2}_{\ell j}(\mathbf{p}_1)\int \frac{d^3\mathbf{q}}{(2\pi)^3} q_i q_j \mathcal{I}_4(q) = \frac{2}{3}\delta_{s_1 s_2}\int \frac{d^3\mathbf{q}}{(2\pi)^3} q^2 \mathcal{I}_4(q). \quad (44)$$

Combining all together, we obtain

$$\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H_4}} = -\frac{4\delta^{s_1 s_2}}{3} M_P^2 \int \frac{d^3\mathbf{q}}{(2\pi)^3} q^2 \mathrm{Im}\mathcal{I}_4(q), \quad (45)$$

in which the factor $\mathcal{I}_4(q)$ is given by

$$\mathcal{I}_4(q) \equiv \int_{-\infty}^{\tau_0} d\tau \epsilon(\tau) a^2 \left[\gamma^*(p_1,\tau_0)^2\gamma(p_1,\tau)^2\right]|\mathcal{R}(q,\tau)|^2. \quad (46)$$

In performing the time integral above we should only consider the contribution of the superhorizon modes, so the actual time interval in Eq. (46) should be $-\frac{1}{q} < \tau < \tau_0$. This guarantees that we do not count the contributions of the modes which are subhorizon (i.e. not yet classical) in the time integral in Eq. (46). On the other hand, the modes which are subhorizon during the USR phase are quantum mechanical in nature so their contributions should be collected via a UV renormalization scheme. While renormalization is an important issue to read off the final physical quantity, here we are mainly interested in the effects of superhorizon modes to obtain a rough estimate for the magnitude of the loop corrections. In addition, as $q\tau \to 0$, the integral in Eq. (46) receives its contribution from its lower end. In particular, the contribution from the period after the USR phase $\tau_e < \tau < \tau_0$ is subleading.

In the limit that $p \to 0$, we have

$$\mathrm{Im}\left[\gamma^*(p_1,\tau_0)^2\gamma(p_1,\tau)^2\right] \simeq -\frac{8}{3}\frac{H^4\tau^3}{M_P^4 p^3}. \quad (47)$$

Furthermore, on the superhorizon in which $q\tau \to 0$, we have $\epsilon(\tau)|\mathcal{R}(q,\tau)|^2 \simeq \frac{H^2}{4q^3 M_P^2}$, yielding

$$\mathrm{Im}\mathcal{I}_4(q) \simeq -\frac{2H^4}{3M_P^6 q^3 p^3}\int_{-\frac{1}{q}}^{\tau_0} d\tau\tau \simeq \frac{H^4}{3M_P^6 q^5 p^3}. \quad (48)$$

Plugging the above result in Eq. (45) and integrating over the USR modes $q_i < q < q_e$, we obtain

$$\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H_4}} \simeq -\frac{4\delta^{s_1 s_2}}{9}\frac{H^4}{M_P^4 p^3}\frac{\Delta N}{2\pi^2}, \quad (49)$$

in which $\Delta N = \ln(\frac{\tau_i}{\tau_e})$ is the duration of the USR phase.

It is convenient to express the loop correction in terms of the dimensionless power spectrum $\mathcal{P}_\gamma$ defined in Eq. (31). Using the result from Eq. (49), for the one-loop correction in the tensor power spectrum from the quartic Hamiltonian $\mathbf{H_4}$ we obtain

$$\mathcal{P}^{(\mathrm{loop})}_\gamma\Big|_{\mathbf{H_4}} \simeq -\frac{\Delta N}{36}\mathcal{P}_\gamma^2. \quad (50)$$

Since we calculate the loop corrections induced from the scalar perturbations on the tensor power spectrum, then one expects the loop correction to scale like $\mathcal{P}_\gamma\mathcal{P}_\mathcal{R}$. However, from Eq. (50) we see that the loop correction actually scales like $\mathcal{P}_\gamma^2$. The reason is that the interaction vertices in $\mathbf{H_3}$ and $\mathbf{H_4}$ contain the factor ϵ so the combination $\epsilon\mathcal{R}^2$ appears inside the in-in integral as in Eq. (46). Since $\epsilon\mathcal{R}^2 \sim \mathcal{P}_\gamma$, then the final result for the loop correction is given as $\mathcal{P}_\gamma^2$ instead of $\mathcal{P}_\gamma\mathcal{P}_\mathcal{R}$.

B. Loop corrections from cubic Hamiltonian

Now we calculate the loop corrections from the cubic Hamiltonian corresponding to the left panel of Fig. 1. It involves a nested integral containing the product of two three-vertices. More schematically, expanding the Dyson series to second order in $\mathbf{H}_3$ we have

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{\mathbf{H}_3} = \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(2,0)} + \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(1,1)} + \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(0,2)} \tag{51}$$

in which

$$\begin{aligned}\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(2,0)} &= -\int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_1} d\tau_2 \langle\mathbf{H}_3(\tau_1)\mathbf{H}_3(\tau_2)\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle \\ &= \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle^\dagger_{(0,2)}\end{aligned} \tag{52}$$

and

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(1,1)} = \int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_0} d\tau_2 \langle\mathbf{H}_3(\tau_1)\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\mathbf{H}_3(\tau_2)\rangle. \tag{53}$$

We leave the details of the in-in analysis to the Appendix. After a long calculation, one obtains

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H}_3} = -8M_P^4\delta^{s_1s_2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}|\epsilon^{s_1}_{ij}(\mathbf{p})q_iq_j|^2\int_{-\infty}^{\tau_0}d\tau_1\int_{-\infty}^{\tau_1}d\tau_2\,\mathrm{Im}[X^*(\tau_1)\delta(\tau_2)], \tag{54}$$

in which

$$X(\tau) \equiv \epsilon a^2\gamma(p,\tau)\gamma^*(p,\tau_0)\mathcal{R}(q,\tau)^2 \tag{55}$$

and

$$\delta(\tau) \equiv 2\epsilon a^2\mathcal{R}(q,\tau)^2\mathrm{Im}[\gamma(p,\tau)\gamma^*(p,\tau_0)]. \tag{56}$$

Using the orthogonality properties of the polarization tensor one can show that

$$\int d\Omega|\epsilon^{s_1}_{ij}(\mathbf{p})\widehat{q}_i\,\widehat{q}_j|^2 = \frac{16\pi}{15}, \tag{57}$$

in which $d\Omega$ represents the angular parts of $d^3\mathbf{q}$.

Combining all contributions, we obtain (see the Appendix for further details)

$$\mathcal{P}_\gamma^{(\mathrm{loop})}|_{\mathbf{H_3}} \simeq (c_1e^{\Delta N} + c_2\Delta N)\mathcal{P}_\gamma^2, \tag{58}$$

in which $c_1 \simeq -0.003$ and $c_2 \simeq 0.005$. As in the quartic case the loop correction scales like $\mathcal{P}_\gamma^2$ instead of $\mathcal{P}_\mathcal{R}\mathcal{P}_\gamma$.

Unlike the correction from the quartic case we see a mild dependence on $e^{\Delta N}$. However, this does not cause much harm. Specifically, for the typical USR setup employed for the PBH formation, one has $\Delta N \sim \mathcal{O}(1)$ so the contribution $c_1e^{\Delta N} < 1$. For example, for $\Delta N = 5$, we obtain $c_1e^{\Delta N} \sim 0.4$. However, note that with $\Delta N = 5$ the loop corrections in the scalar sector already become very large if the transition is sharp [1,8].

Now combining the results from the cubic and quartic interactions, Eqs. (50) and (58), the total one-loop correction is obtained to be

$$\mathcal{P}_\gamma^{(\mathrm{loop})} \simeq (c_1e^{\Delta N} + c_3\Delta N)\mathcal{P}_\gamma^2 \tag{59}$$

in which $c_3 \simeq -0.02$.

From the above result we see that the loop corrections in the tensor power spectrum induced from the USR modes are quite insensitive to the sharpness of the transition from the USR phase to the SR phase. Indeed, we do not see any explicit dependence to the sharpness parameter h in Eq. (59). This is unlike the loop corrections induced on long scalar perturbations in which the loop corrections increase linearly with h [8] for $|h| \gg 1$ in which $\mathcal{P}_\mathcal{R}^{(\mathrm{loop})} \sim h\mathcal{P}_\mathcal{R}^{\mathrm{CMB}}\mathcal{P}_\mathcal{R}^{\mathrm{short}} \sim h(\mathcal{P}_\mathcal{R}^{\mathrm{CMB}})^2e^{6\Delta N}$. The dependence on the duration of the USR phase via the exponential factor $e^{6\Delta N}$ is the hallmark of USR loop corrections in the scalar power spectrum which can invalidate the perturbative treatment.

In addition, we see that the induced loop corrections in GWs are quite small in all practical setups. More specifically we obtain $\frac{\mathcal{P}_\gamma^{(\mathrm{loop})}}{\mathcal{P}_\gamma} \sim 10^{-3}e^{\Delta N}\mathcal{P}_\gamma$. Assuming $\mathcal{P}_\gamma \lesssim 10^{-10}$

from the CMB observations, we need $\Delta N \sim 30$ in order for the ratio $\frac{\mathcal{P}_\gamma^{(\mathrm{loop})}}{\mathcal{P}_\gamma}$ to approach unity. However, this does not happen, because by that time the scalar power spectrum $\mathcal{P}_\mathcal{R}$ has increased by the gigantic amount $e^{6\Delta N} \sim e^{180}$, invalidating the perturbative approach completely. One may wonder if considering a mild transition can change the above conclusion. More specifically, in a mild transition one expects that the loop corrections in the scalar power spectrum becomes suppressed (by slow-roll parameters) so one may have more freedom in increasing ΔN. However, we note that the loop correction in the scalar power spectrum scales like $e^{6\Delta N} h$ [8]. So the sensitivity of h is not strong enough to relax the value of ΔN considerably. For example, changing h by a factor of 100 from $h = -6$ to a mild transition with $h = -0.06$ will increase ΔN by the amount $\frac{1}{6}\ln(100) \simeq 0.8$ which cannot change our conclusion above that the loop correction in the tensor spectrum is harmless.

The conclusion is that the long CMB scale gravitational waves are practically unaffected by the short scalar perturbations which leave the horizon during the USR phase. This conclusion is largely independent of the mechanism of the transition from the USR phase to the final SR phase.

VI. SUMMARY AND DISCUSSIONS

In this work we have studied the one-loop correction in the power spectrum of long gravitational waves from small scale modes which leave the horizon during the intermediate USR phase. This study is motivated by similar recent studies performed for loop corrections in the scalar power spectrum.

As one might have guessed, the results are quite different from what were obtained for the case of scalar power spectrum. We have shown that the long tensor power spectrum is largely unaffected by the loop corrections from small USR modes. In particular, the one-loop corrections are quite insensitive to the sharpness of the transition. This might have been expected from the physical ground that the tensor perturbations only probe the Hubble expansion rate of the corresponding inflationary background and are insensitive to slow-roll parameters. Having said this, it is still a good cross-check to verify the validity of this physical expectation since a similar intuition, suggesting that the scalar power spectrum should be unaffected by intermediate short modes, proved to fail for the case of a sharp transition [1,2,8]. While our analysis was focused to the particular setup of SR → USR → SR, this conclusion may be general. As long as there are no dramatic changes in the background Hubble expansion rate, then independent of the nature of transitions in slow-roll parameters, the superhorizon tensor modes are unaffected by the short scalar modes which may experience rapid growth. It would be useful to verify this conjecture in its generality.

In addition, we have shown that the Maldacena consistency condition for the tensor-scalar-scalar bispectrum in the squeezed limit does hold. The fact that the long tensor mode is frozen on the superhorizon scale is the key reason for the validity of this consistency condition. The long tensor perturbations only induce small anisotropies on the background for the short modes yielding to the expected tensor-scalar-scalar consistency condition.

We comment that the loop corrections on the tensor power spectrum calculated here should not be confused with the induced gravitational waves from second order scalar perturbations which have been actively investigated recently (for a review see [48]), and for works studying secondary GWs induced in models with a non-Gaussian feature or a USR setup see [49–52]. While these two questions are related but the induced GWs from large second order scalar perturbations are mostly concerned with small scale GWs, the modes near the peak of scalar perturbations reenter the horizon during the radiation dominated era. Here, on the other hand, we look at the enhancement of the GW spectrum at the CMB scales.

ACKNOWLEDGMENTS

I am grateful to Mohammad Ali Gorji and Antonio Riotto for useful discussions and for comments on the draft. I thank Amin Nassiri-Rad for checking the in-in analysis in Sec. V. This research is partially supported by the "Saramadan" Federation of Iran.

APPENDIX: IN-IN ANALYSIS FOR CUBIC HAMILTONIAN

In this appendix we present the details of the in-in integral for the cubic Hamiltonians $\mathbf{H}_3$.

As discussed before, the loop interaction from the cubic Hamiltonian is given by

$$\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{\mathbf{H}_3} = \langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(2,0)} + \langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(1,1)} + \langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(0,2)} \tag{A1}$$

with

$$\begin{aligned}\langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(2,0)} &= -\int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_1} d\tau_2 \langle \mathbf{H}_3(\tau_1)\mathbf{H}_3(\tau_2)\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle \\ &= \langle \gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle^\dagger_{(0,2)}\end{aligned} \tag{A2}$$

and

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(1,1)} = \int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_0} d\tau_2 \langle \mathbf{H}_3(\tau_1)\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\mathbf{H}_3(\tau_2)\rangle. \tag{A3}$$

Let us start with $\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(1,1)}$. Using the Hamiltonian (21), performing all contractions, and employing the properties of the polarization tensor given in Eq. (25) one obtains

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(1,1)} = 4M_P^4\delta^{s_1s_2}\int \frac{d^3\mathbf{q}}{(2\pi)^3}|\epsilon^{s_1}_{ij}(\mathbf{p})q_iq_j|^2 \left|\int_{-\infty}^{\tau_0} d\tau X(\tau)\right|^2, \tag{A4}$$

in which

$$X(\tau) \equiv \epsilon a^2\gamma(p,\tau)\gamma^*(p,\tau_0)\mathcal{R}(q,\tau)^2. \tag{A5}$$

Similarly, for $\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(2,0)}$ we obtain

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(2,0)} = -4M_P^4\delta^{s_1s_2}\int \frac{d^3\mathbf{q}}{(2\pi)^3}|\epsilon^{s_1}_{ij}(\mathbf{p})q_iq_j|^2 \int_{-\infty}^{\tau_0} d\tau_1 X(\tau_1)\int_{-\infty}^{\tau_1} d\tau_2 Z(\tau_2), \tag{A6}$$

in which

$$Z(\tau) \equiv \epsilon a^2\gamma(p,\tau)\gamma^*(p,\tau_0)\mathcal{R}^*(q,\tau)^2. \tag{A7}$$

Noting that $\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle_{(2,0)} = \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle^\dagger_{(0,2)}$, we obtain

$$\begin{aligned}&\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(2,0)} + \langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(0,2)}\\ &\quad = -4M_P^4\delta^{s_1s_2}\int \frac{d^3\mathbf{q}}{(2\pi)^3}|\epsilon^{s_1}_{ij}(\mathbf{p})q_iq_j|^2 \int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_1} d\tau_2\Big[X(\tau_1)Z(\tau_2) + X^*(\tau_1)Z^*(\tau_2)\Big].\end{aligned} \tag{A8}$$

To proceed further, let us define

$$Z(\tau) \equiv X^*(\tau) + i\delta(\tau)^*, \tag{A9}$$

in which the new variable δ, from Eqs. (A5) and (A7), is obtained to be

$$\delta(\tau) = 2\epsilon a^2\mathcal{R}(q,\tau)^2\mathrm{Im}[\gamma(p,\tau)\gamma^*(p,\tau_0)]. \tag{A10}$$

With the above relation between $X(\tau)$ and $Z(\tau)$, one can show that the nested time integrals in Eq. (A8) are rearranged in the following form:

$$\int_{-\infty}^{\tau_0} d\tau_1 \int_{-\infty}^{\tau_1} d\tau_2[X(\tau_1)Z(\tau_2) + X^*(\tau_1)Z^*(\tau_2)] = \int_{-\infty}^{\tau_0} d\tau|X(\tau)|^2 - 2\int_{-\infty}^{\tau_0} d\tau_1\int_{-\infty}^{\tau_1} d\tau_2\mathrm{Im}[X(\tau_1)\delta^*(\tau_2)]. \tag{A11}$$

We see that the first integral in Eq. (A11) cancels the contribution of $\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{(1,1)}$ in Eq. (A4) so at the end we are left with

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H}_3} = -8M_P^4\delta^{s_1s_2}\int \frac{d^3\mathbf{q}}{(2\pi)^3}|\epsilon^{s_1}_{ij}(\mathbf{p})q_iq_j|^2 \int_{-\infty}^{\tau_0} d\tau_1\int_{-\infty}^{\tau_1} d\tau_2\mathrm{Im}[X^*(\tau_1)\delta(\tau_2)]. \tag{A12}$$

To go further, we need to calculate the contribution of the polarization tensor in the above integral. With the specific form of the polarization tensor given in Eq. (26), one can show that

$$\epsilon^{\pm}_{ij}(\mathbf{p})\widehat{q}_i\,\widehat{q}_j = \frac{1}{\sqrt{2}}\sin^2(\theta)e^{\pm 2i\phi}, \tag{A13}$$

in which the orientation of the unit vector $\hat{q}$ in a coordinate where $\hat{\mathbf{p}}$ is along the third axis is specified by the angles (ϕ, θ) in which $\hat{q} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. Consequently, one can easily check that

$$\int d\Omega(\epsilon^{s_1}_{ij}(\mathbf{p})\widehat{q}_i\,\widehat{q}_j)(\epsilon^{s_2*}_{mn}(\mathbf{p})\widehat{q}_m\,\widehat{q}_n) = \frac{16\pi}{15}\delta^{s_1 s_2}. \tag{A14}$$

Plugging the above result into Eq. (A12) we obtain

$$\langle\gamma^{s_1}_{\mathbf{p}_1}(\tau_0)\gamma^{s_2}_{\mathbf{p}_2}(\tau_0)\rangle'_{\mathbf{H}_3} = -\frac{16}{15\pi^2}M_P^4\delta^{s_1 s_2}\int dq q^6\int_{-\infty}^{\tau_0}d\tau_1\int_{-\infty}^{\tau_1}d\tau_2 \mathrm{Im}[X^*(\tau_1)\delta(\tau_2)]. \tag{A15}$$

In performing the above nested integral, it is useful to note that

$$\mathrm{Im}[\gamma(p,\tau)\gamma^*(p,\tau_0)] = -\frac{2H^2}{3M_P^2}\tau^3 \tag{A16}$$

and

$$\gamma(p,\tau)\gamma^*(p,\tau_0) = \frac{2H^2}{M_P^2 p^3} + \mathcal{O}(p^{-1}). \tag{A17}$$

There is an important comment in order. We emphasize that we integrate over the modes which become superhorizontal during the USR phase, so the time integrals in Eq. (A15) are actually restricted to $-\frac{1}{q} < \tau_2 < \tau_1 < \tau_e$. This is to make sure that we only count the modes which become classical during the USR phase. The modes which are subhorizontal during the USR phase are not classical, and their effects may be collected under a UV renormalization scheme which is not our question of interest here. With the same logic, for the integral over the momentum q we integrate over the modes $q_i < q < q_e$ which become superhorizontal during the USR phase.

Using the relations (A16) and (A17) for $\delta(\tau_2)$ and $X(\tau_1)$ in the nested integral (A15) we obtain Eq. (58) in the main text. We comment that the main contribution in the time integral in Eq. (A15) comes for the USR period, $\tau_i < \tau_{1,2} < \tau_e$, while the contribution from the final SR phase, $\tau_e < \tau_{1,2} < \tau_0$, is subleading. To perform the analysis of the nested integral in Eq. (A15) we use the Maple computational software.

[1] J. Kristiano and J. Yokoyama, arXiv:2211.03395.
[2] J. Kristiano and J. Yokoyama, arXiv:2303.00341.
[3] A. Riotto, arXiv:2303.01727.
[4] A. Riotto, arXiv:2301.00599.
[5] S. Choudhury, M. R. Gangopadhyay, and M. Sami, arXiv:2301.10000.
[6] S. Choudhury, S. Panda, and M. Sami, arXiv:2302.05655.
[7] S. Choudhury, S. Panda, and M. Sami, arXiv:2303.06066.
[8] H. Firouzjahi, arXiv:2303.12025.
[9] H. Motohashi and Y. Tada, arXiv:2303.16035.
[10] H. Firouzjahi and A. Riotto, arXiv:2304.07801.
[11] S. L. Cheng, D. S. Lee, and K. W. Ng, Phys. Lett. B **827**, 136956 (2022).
[12] P. Ivanov, P. Naselsky, and I. Novikov, Phys. Rev. D **50**, 7173 (1994).
[13] J. Garcia-Bellido and E. Ruiz Morales, Phys. Dark Universe **18**, 47 (2017).
[14] M. Biagetti, G. Franciolini, A. Kehagias, and A. Riotto, J. Cosmol. Astropart. Phys. 07 (2018) 032.
[15] O. Özsoy and G. Tasinato, Universe **9**, 203 (2023).
[16] C. T. Byrnes and P. S. Cole, arXiv:2112.05716.
[17] W. H. Kinney, Phys. Rev. D **72**, 023515 (2005).
[18] M. J. P. Morse and W. H. Kinney, Phys. Rev. D **97**, 123519 (2018).
[19] W. C. Lin, M. J. P. Morse, and W. H. Kinney, J. Cosmol. Astropart. Phys. 09 (2019) 063.
[20] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, Europhys. Lett. **101**, 39001 (2013).
[21] J. M. Maldacena, J. High Energy Phys. 05 (2003) 013.
[22] P. Creminelli and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 10 (2004) 006.
[23] J. Martin, H. Motohashi, and T. Suyama, Phys. Rev. D **87**, 023514 (2013).
[24] X. Chen, H. Firouzjahi, M. H. Namjoo, and M. Sasaki, Europhys. Lett. **102**, 59001 (2013).
[25] X. Chen, H. Firouzjahi, E. Komatsu, M. H. Namjoo, and M. Sasaki, J. Cosmol. Astropart. Phys. 12 (2013) 039.
[26] M. Akhshik, H. Firouzjahi, and S. Jazayeri, J. Cosmol. Astropart. Phys. 12 (2015) 027.
[27] S. Mooij and G. A. Palma, J. Cosmol. Astropart. Phys. 11 (2015) 025.
[28] R. Bravo, S. Mooij, G. A. Palma, and B. Pradenas, J. Cosmol. Astropart. Phys. 05 (2018) 024.
[29] B. Finelli, G. Goon, E. Pajer, and L. Santoni, Phys. Rev. D **97**, 063531 (2018).

[30] S. Pi and M. Sasaki, Phys. Rev. Lett. **131**, 011002 (2023).
[31] Y. F. Cai, X. Chen, M. H. Namjoo, M. Sasaki, D. G. Wang, and Z. Wang, J. Cosmol. Astropart. Phys. 05 (2018) 012.
[32] A. Ota, M. Sasaki, and Y. Wang, arXiv:2211.12766.
[33] C. Chen, A. Ota, H. Y. Zhu, and Y. Zhu, Phys. Rev. D **107**, 083518 (2023).
[34] A. Ota, M. Sasaki, and Y. Wang, arXiv:2209.02272.
[35] D. S. Meng, C. Yuan, and Q. G. Huang, Phys. Rev. D **106**, 063508 (2022).
[36] S. Brahma, A. Berera, and J. Calderón-Figueroa, J. High Energy Phys. 08 (2022) 225.
[37] S. Weinberg, *Cosmology* (Oxford University Press, New York, 2008).
[38] D. Baumann, *Cosmology* (Cambridge University Press, Cambridge, England, 2022).
[39] H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. **78**, 1 (1984).
[40] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rep. **215**, 203 (1992).
[41] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, J. High Energy Phys. 03 (2008) 014.
[42] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, J. Cosmol. Astropart. Phys. 02 (2008) 021.
[43] M. Akhshik, H. Firouzjahi, and S. Jazayeri, J. Cosmol. Astropart. Phys. 07 (2015) 048.
[44] T. Noumi and M. Yamaguchi, arXiv:1403.6065.
[45] P. R. Jarnhus and M. S. Sloth, J. Cosmol. Astropart. Phys. 02 (2008) 013.
[46] F. Arroja and K. Koyama, Phys. Rev. D **77**, 083517 (2008).
[47] S. Weinberg, Phys. Rev. D **72**, 043514 (2005).
[48] G. Domènech, Universe **7**, 398 (2021).
[49] R. G. Cai, S. Pi, and M. Sasaki, Phys. Rev. Lett. **122**, 201101 (2019).
[50] J. Liu, Z. K. Guo, and R. G. Cai, Phys. Rev. D **101**, 083535 (2020).
[51] A. Talebian, S. A. Hosseini Mansoori, and H. Firouzjahi, Astrophys. J. **948**, 48 (2023).
[52] H. V. Ragavendra, Phys. Rev. D **105**, 063533 (2022).