DESY-23-143

Induced gravitational waves and baryon asymmetry fluctuations from primordial black hole formation

Chiara Altavista[⋆] and Julián Rey[†]

[⋆]*Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185, Roma, Italy*
[†]*Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

Abstract

We consider black hole formation due to the gravitational collapse produced by large density fluctuations during an epoch of reheating with a stiff equation of state and calculate the induced gravitational wave spectrum. By considering the existing bounds on the total energy density of gravitational waves today, we find constraints on the parameter space of this scenario. We also calculate the lepton asymmetry generated by metric perturbations via the chiral gravitational anomaly present in the Standard Model and find that, once the electroweak sphaleron processes have taken place, the large spectrum of scalar perturbations responsible for black hole formation induces a peak in the baryon asymmetry fluctuations on small scales.

Contents

⋆ altavista.1844243@studenti.uniroma1.it
† julian.rey@desy.de

Introduction

The baryon-to-photon ratio observed in CMB experiments [1] as well as in measurements of the abundance of light elements produced during Big Bang nucleosynthesis (see e.g. [2]), leads to the requirement of some mechanism, known as baryogenesis, to produce an asymmetry between matter and antimatter in the early Universe. One possibility that has been extensively considered is that of leptogenesis [3], which consists in generating a lepton asymmetry at early stages which is later converted to a baryon asymmetry via sphaleron transitions, non-perturbative processes that take place in the electroweak sector of the Standard Model. Typical models require extending the particle content of the theory by adding, for instance, right-handed neutrinos [4]. An alternative possibility is gravitational leptogenesis, which consists in exploiting the chiral gravitational anomaly [5, 6] present in the Standard Model (that is, in the presence of only left-handed neutrinos[1]),

$$\nabla_\mu J^\mu_{\rm L} = -\frac{N_{\rm L-R}}{24(4\pi)^2} R\tilde{R}, \tag{1}$$

where $R\tilde{R} = (1/2)\epsilon^{\mu\nu}{}_{\rho\sigma} R_{\mu\nu\alpha\beta} R^{\rho\sigma\alpha\beta}$ denotes the contraction between the Riemann tensor and its dual, and the factor $N_{\rm L-R} = 3$ arises from the difference between the number of left- and right-handed neutrinos in the theory.[2] By expanding the right-hand side of the above equation to quadratic order in perturbations, one obtains a term proportional to the product $h_{ij}\phi$, as well as terms quadratic in h_{ij}, where ϕ denotes the Newtonian potential and h_{ij} the transverse, traceless tensor perturbation of the metric. From the structure of the resulting terms one can check that, in order for the mechanism to work, a chiral gravitational wave background is required, as one would expect from the Sakharov conditions [9]. Such a spectrum can be generated, for instance, during inflation, in models in which the inflaton φ contains a CP-odd component and couples to gravity through a term of the form $f(\varphi)R\tilde{R}$ [7].[3]

It was recently suggested in [11] that, due to the fact that inflation is a stochastic process, there is actually no need to invoke the presence of these couplings to produce a lepton asymmetry via the above mechanism. Indeed, due to the dependence of the anomaly on the stochastic variables ϕ and h_{ij}, we expect a non-vanishing variance for the lepton number density $\langle |n_{\rm L}|^2 \rangle$ to be generated in different patches as inflation progresses. The average of the root-mean-square variance over some particular region therefore quantifies the expected asymmetry. Since these fluctuations are generated during inflation (out of thermal equilibrium) and the anomaly violates C, CP, and L (with the electroweak sphalerons providing the required B violation later on), the Sakharov conditions are fulfilled. As shown in [11], however, when this average is computed over the entire observable Universe today, the resulting asymmetry turns out to be extremely suppressed, rendering the proposal incapable of producing baryogenesis. Although unable to explain the observed background value of the baryon asymmetry, the presence of the chiral gravitational anomaly would nevertheless

[1] Even if right-handed neutrinos are added to the Standard Model, this anomaly is still present once they are integrated out, at sufficiently low energies [7].

[2] In general, all chiral fermions in the theory contribute to $N_{\rm L-R}$, with the contribution of each particle weighted by its corresponding B − L factor (see e.g. [8]), but since the Standard Model contains an equal amount of left- and right-handed quarks, their contribution vanishes and B drops out. The same argument applies to charged leptons.

[3] See [10] and the references therein for a discussion about the caveats of this mechanism, in particular the presence of ghost modes, and some proposed solutions.

lead to unavoidable fluctuations in the baryon number density on sufficiently large scales,[4] which could potentially be used as an observable to probe different models of inflation. We anticipate, however, that in the particular scenario discussed here the fluctuations are much smaller than the observed background value, so measuring them would likely require some additional enhancement mechanism.

The calculation presented in [11] makes use of the $h_{ij}\phi$ term that arises after expanding eq. (1) in perturbations, so that the size of the fluctuations in the baryon asymmetry is proportional to the amplitude of both the tensor and scalar power spectrum, and becomes negligible if the tensor-to-scalar ratio r is sufficiently small. A natural question is therefore whether an enhancement in either of these quantities could significantly increase the size of these fluctuations. Such an enhancement can be obtained, for instance, in single-field models of inflation that aim to generate a significant population of primordial black holes by introducing an inflection point in the potential (see e.g. [12–14] for particular implementations of this mechanism). In this class of models, the presence of the inflection point leads to a phase of ultra-slow-roll that enhances the curvature power spectrum, producing large energy overdensities once the perturbations re-enter the horizon after the end of inflation, which in turn induce gravitational collapse, generating the black holes. These black holes could account for the entirety of the observed dark matter provided their masses lie in the range

$$10^{-16} M_\odot \lesssim M_{\rm PBH} \lesssim 10^{-11} M_\odot, \tag{2}$$

where the upper bound comes from microlensing observations [15] and the lower one from their Hawking evaporation [16–19]. Such a large scalar power spectrum would also source gravitational waves at second order in perturbations after the end of inflation [20–22], leading to a peaked gravitational wave signal which would also contribute to the variance of the lepton number density and which is completely independent of the primordial tensor spectrum arising from inflation. The spectrum of gravitational waves induced by large scalar perturbations during a radiation era has been extensively studied in the literature [21–24]. However, since the resulting signal depends on the evolution of the scalar perturbations after they re-enter the horizon, we expect the result to change if the background fluid has a different equation of state. The case in which the Universe is dominated by non-relativistic matter before transitioning to the radiation era was studied in [24–26], and the scenario with a general background was considered in [27], where an enhancement of the spectrum for background fluids with stiff equations of state (that is, for $w \lesssim 1$, with $p = w\rho$) was reported.

This paper has two objectives. The first is to connect the enhancement of the induced gravitational wave spectrum for stiff background equations of state to the abundance of primordial black holes.[5] We consider a scenario in which primordial black holes form during a reheating stage[6] with an equation of state parameter $1/3 < w < 1$ before the Universe transitions to the radiation era at some temperature T_r. We find that, depending on the specific values of the equation of

[4] As explained in [11], the baryon asymmetry survives only on large scales due to the matter-antimatter annihilation processes taking place on small patches. We discuss this in detail in Section 5.

[5] See also [28, 29] for recent work in similar directions.

[6] Throughout the paper we refer to this early stiff epoch as reheating, since we assume it takes place immediately after inflation ends. However, we remark that our results do not depend on the origin of this epoch, which we remain agnostic about. In particular, it does not necessarily have to come from the dynamics of the inflaton.

state, the transition temperature, and the scale at which the peak in the scalar power spectrum is located (therefore, on the mass and abundance of the black holes that form), the enhancement of the signal can be large enough to violate the existing bounds on the total gravitational wave energy density derived from CMB observations and the abundance of light elements produced during Big Bang nucleosynthesis [30, 31], effectively reducing the parameter space of this scenario. We perform the calculation of the gravitational wave spectrum for two cases, one in which the transition between epochs is instantaneous, and one in which the stiff fluid gradually decays into radiation, and show that the resulting constraints on the parameter space for PBH formation depend only mildly on the smoothness of the transition. We improve upon the results of [27] by implementing a matching procedure for the transfer function of the scalar perturbations and the Green's function of the tensor modes in the sudden transition case akin to the one presented in [24] for the case of matter-domination, effectively taking into account the full time evolution of both quantities. We also perform a fully numerical calculation of the scalar transfer function and the tensor Green's functions in the gradual transition case, in contrast to [26]. The second objective is to calculate the baryon asymmetry fluctuations induced by the chiral gravitational anomaly in eq. (1) for the same scenario.[7] We extend the results of [11] in two significant ways. The first is that we consider a peaked scalar power spectrum responsible for PBH formation, as opposed to a scale-invariant one, thereby enhancing the asymmetry reported there. The second one is that we also consider the purely scalar contribution to the asymmetry due to induced gravitational waves, effectively removing one of the essential ingredients in [11], namely, the need to have a non-vanishing gravitational wave background generated during inflation. We once again remark that the mechanism studied here is unable to produced the observed baryon asymmetry of the Universe, but it does allow us to predict a spectrum of fluctuations in this quantity which would be present in any model of PBH formation from single-field inflation using an inflection point in the potential, assuming only the matter content of the Standard Model. Moreover, the machinery presented here could be used to compute these fluctuations in other models of gravitational leptogenesis, such as the one in [7].

The paper is structured as follows. In Section 1 we expand eq. (1) in perturbations and derive the corresponding expressions for the lepton number density at each order. In Section 2 we derive the expressions for the black hole mass and abundance assuming that they form during a reheating stage with a stiff equation of state, and discuss the relevant constraints on the parameter space. In Section 3 we derive the induced gravitational wave spectrum when the transition between the reheating stage and the radiation era is instantaneous, and determine the constraint on the PBH masses due to the aforementioned bounds on the gravitational wave energy density. In Section 4 we repeat the calculation by assuming that the stiff fluid gradually decays into radiation. Finally, in Section 5 we determine the baryon asymmetry produced by the chiral gravitational anomaly in this scenario.

1. Gravitational lepton anomaly

In this section we expand eq. (1) to third order in perturbations and find expressions for the lepton number density $n_{\rm L}$ at each order. Throughout the paper we denote second-order perturbations of

[7] The mechanism studied here is different from leptogenesis via PBH evaporation, see e.g. [32].

the metric with bold symbols. The perturbed FLRW metric is, in conformal time $d\eta = dt/a$,

$$ds^2 = a^2\Big\{-(1+2\phi+\Phi)d\eta^2 + \Big[(1-2\phi-\Psi)\delta_{ij} + \frac{1}{2}(\partial_i \boldsymbol{E}_j + \partial_j \boldsymbol{E}_i) + h_{ij} + \frac{1}{2}\boldsymbol{h}_{ij}\Big]dx^i dx^j\Big\}, \quad (3)$$

where we have fixed the Newtonian gauge, so that $E = B = \boldsymbol{E} = \boldsymbol{B} = 0$, we have assumed that the first-order vector perturbations vanish, $E_i = B_i = 0$, and we have used a second-order vector gauge-transformation to set $\boldsymbol{B}_i = 0$. The second-order vector perturbation $\boldsymbol{E}_i$ cannot be set to zero, since it is sourced by terms quadratic in first-order scalar perturbations by virtue of Einstein's equations. We have additionally assumed that no anisotropic stress is present, so that the two first-order Newtonian potentials are equal to each other, $\phi = \psi$. The difference between the two Newtonian potentials at second order $\Phi - \Psi$ does not vanish in the absence of anisotropic stress, since it is sourced by terms quadratic in first-order scalars.

The quantity $R\tilde{R}$ which appears on the right-hand side of eq. (1) can be easily checked to vanish at the background level. This quantity can be calculated at each order in perturbations by direct expansion using the above metric. However, the result can also be obtained by noting that every term in the expansion must contain four derivatives (with two coming from each factor of R) and the indices can only be saturated in a limited number of ways. For instance, this quantity vanishes at leading order[8] (LO), since the only perturbations available are ϕ and h_{ij}, and every possible contraction (for instance, $\epsilon_{ijk}\partial_i\partial_j\partial_k\phi$) vanishes due to the antisymmetry of ϵ_{ijk}. At the next-to-leading order (NLO) we have only one possible term mixing scalar and tensor perturbations,

$$R\tilde{R} \supset \epsilon_{ijk}\partial_\ell\partial_i\phi\partial_j h'_{k\ell}, \quad (4)$$

where primes denote derivatives with respect to conformal time ($' = d/d\eta$), as well as several possible terms mixing two tensor perturbations, such as

$$R\tilde{R} \supset \epsilon_{ijk}\partial_m h'_{i\ell}\partial_m\partial_k h_{j\ell}. \quad (5)$$

We generically expect these terms to be suppressed with respect to the scalar-tensor ones (since the stochastic gravitational wave background produced during inflation is much smaller than the scalar power spectrum) and thus we do not consider them.

At NNLO, we focus on the situation in which the first-order tensor perturbation h_{ij} is negligible and the only relevant contribution comes from the scalar-induced $\boldsymbol{h}_{ij}$, which is the case in models of PBH formation from single-field inflation. In this case, all terms containing only scalar modes (such as $\epsilon_{ijk}\partial_i\phi\partial_j\phi\partial_k\phi'$ or $\epsilon_{ijk}\partial_i\partial_j\phi\partial_k\Phi'$) can be easily seen to vanish due to the antisymmetry of ϵ_{ijk}. Similarly, the scalar-vector terms vanish due to the fact that terms such as $\epsilon_{ijk}\partial_i\phi'\partial_j\boldsymbol{E}'_k$ are always accompanied by $\epsilon_{ijk}\partial_i\phi'\partial_k\boldsymbol{E}'_j$, since only the symmetric combination $\partial_i\boldsymbol{E}_j + \partial_j\boldsymbol{E}_i$ appears in the metric. Thus, the only relevant term at this order is the scalar-tensor one

$$R\tilde{R} \supset \epsilon_{ijk}\partial_\ell\partial_i\phi\partial_j\boldsymbol{h}'_{k\ell}, \quad (6)$$

[8] We use the terminology LO, NLO, and NNLO for the expansion in powers of the metric perturbations. Note, however, that this does not imply anything about the relative amplitude of each term. As we will see, the dominant term changes depending on the scale considered.

since terms quadratic in $\boldsymbol{h}_{ij}$ are of higher order, and we neglect terms mixing h_{ij} and $\boldsymbol{h}_{ij}$, which are subdominant with respect to the above contribution.

We conclude that, at NLO, the only relevant term is the one in eq. (4), assuming that the first-order gravitational wave background produced during inflation is suppressed with respect to the scalar one at the scales of interest. The numerical prefactor can be obtained by explicit calculation, and the result is

$$R\tilde{R} = -\frac{8}{a^4}\epsilon_{ijk}\partial_\ell\partial_i\phi\partial_j h'_{k\ell}. \tag{7}$$

On the other hand, at NNLO only the term in eq. (6) contributes, assuming vanishing first-order tensor modes. The prefactor can be obtained by simply substituting $h_{ij} \to \frac{1}{2}\boldsymbol{h}_{ij}$ in eq. (7),

$$R\tilde{R} = -\frac{4}{a^4}\epsilon_{ijk}\partial_\ell\partial_i\phi\partial_j \boldsymbol{h}'_{k\ell}. \tag{8}$$

We remark that the NNLO terminology in no way implies that the term in eq. (8) is smaller than the one in eq. (7). The reason is that, since $\boldsymbol{h}_{ij}$ is sourced by scalar perturbations, its amplitude will depend on their initial conditions and time evolution, which can be very different from those of the unsourced h_{ij}.

The left-hand side of eq. (1) can be expanded by writing $J^\mu = (a^{-1}n_{\rm L}, \boldsymbol{0})$ and using the following identity,

$$\nabla_\mu J^\mu = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}J^\mu). \tag{9}$$

We therefore obtain, for the NLO term,

$$a^3 n_{\rm L} = \frac{1}{16\pi^2}\int \epsilon_{ijk}\partial_j\partial_\ell\phi\partial_k h'_{i\ell}d\eta. \tag{10}$$

We can expand the perturbations in Fourier modes

$$h_{ij}(\boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}e^{i\boldsymbol{k}\cdot\boldsymbol{x}}(h^+_k e^+_{ij} + h^\times_k e^\times_{ij}), \tag{11}$$

$$\phi(\boldsymbol{x}) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\phi_k, \tag{12}$$

where e^s_{ij} denotes the two transverse, traceless polarization tensors, defined via

$$e^+_{ij} = \frac{1}{\sqrt{2}}(v_i v_j - \bar{v}_i\bar{v}_j), \quad e^\times_{ij} = \frac{1}{\sqrt{2}}(v_i\bar{v}_j + \bar{v}_i v_j), \tag{13}$$

where $\boldsymbol{v}$ and $\bar{\boldsymbol{v}}$ are two unit vectors satisfying $\boldsymbol{k}\cdot\boldsymbol{v} = \boldsymbol{k}\cdot\bar{\boldsymbol{v}} = \boldsymbol{v}\cdot\bar{\boldsymbol{v}} = 0$. We can therefore regard e^s_{ij} as a $\boldsymbol{k}$-dependent quantity. We then find

$$a^3 n_{\rm L} = \frac{1}{16\pi^2}\epsilon_{ijk}\sum_s\int d\eta\int\frac{d^3\boldsymbol{p}}{(2\pi)^3}\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}e^{i(\boldsymbol{p}+\boldsymbol{q})\cdot\boldsymbol{x}}(iq_j)(iq_\ell)\phi_q(ip_k)h^{s'}_p e^s_{i\ell}(\boldsymbol{p}). \tag{14}$$

The mean value of this quantity clearly vanishes, as can be easily checked by expanding ϕ_k and h^s_k in terms of creation and annihilation operators. The variance, however, does not, and will be computed in Section 5.

To find the expression for the NNLO term in Fourier space it is necessary to solve the equation of motion for the second-order tensor modes induced by scalar fluctuations, given, in momentum space and in the absence of anisotropic stress, by (see e.g. [22–25])

$$\boldsymbol{h}_k^{s\prime\prime} + 2\mathcal{H}\boldsymbol{h}_k^{s\prime} + k^2\boldsymbol{h}_k^s = S_k^s, \tag{15}$$

where $\mathcal{H} = a'/a$ denotes the conformal Hubble factor and the source term is, in the Newtonian gauge,

$$S_k^s = \int \frac{d^3p}{(2\pi)^3}\Big[\boldsymbol{p}\cdot\boldsymbol{e}^s(\boldsymbol{k})\cdot\boldsymbol{p}\Big]\left[8\phi_p\phi_{|\boldsymbol{k}-\boldsymbol{p}|} + \frac{16}{3(1+p/\rho)}\left(\phi_p + \frac{\phi_p'}{\mathcal{H}}\right)\left(\phi_{|\boldsymbol{k}-\boldsymbol{p}|} + \frac{\phi_{|\boldsymbol{k}-\boldsymbol{p}|}'}{\mathcal{H}}\right)\right], \tag{16}$$

where p and ρ are the background pressure and energy density, and we have used the boldface $\boldsymbol{e}^s(\boldsymbol{k})$ to denote the matrix $e_{ij}^s(\boldsymbol{k})$. Explicitly, we have $\boldsymbol{p}\cdot\boldsymbol{e}^s(\boldsymbol{k})\cdot\boldsymbol{p} \equiv p_i e_{ij}^s(\boldsymbol{k}) p_j$. The solution to eq. (15) is given by

$$\boldsymbol{h}_k^s(\eta) = \int_0^\eta G_k(\eta,\eta')S_k^s(\eta')d\eta', \tag{17}$$

where $G_k(\eta,\eta')$ is the Green's function of the differential equation, and we have assumed that inflation ends at $\eta = 0$ and the gravitational waves induced up to that point are negligible. This solution can be rewritten as

$$\boldsymbol{h}_k^s(\eta) = \frac{1}{k^2}\int \frac{d^3\boldsymbol{p}}{(2\pi)^3}\Big[\boldsymbol{p}\cdot\boldsymbol{e}^s(\boldsymbol{k})\cdot\boldsymbol{p}\Big]\mathcal{R}_p\mathcal{R}_{|\boldsymbol{k}-\boldsymbol{p}|}I_k(\eta,p,|\boldsymbol{k}-\boldsymbol{p}|), \tag{18}$$

where

$$I_k(\eta,p,|\boldsymbol{k}-\boldsymbol{p}|) = \left(\frac{3+3w}{5+3w}\right)^2\int_0^\eta kG_k(\eta,\eta')Q(\eta',p,|\boldsymbol{k}-\boldsymbol{p}|)kd\eta', \tag{19}$$

with

$$\begin{aligned} Q(\eta,p,|\boldsymbol{k}-\boldsymbol{p}|) &= 8T_\phi(p\eta)T_\phi(|\boldsymbol{k}-\boldsymbol{p}|\eta) \\ &+ \frac{16}{3(1+p/\rho)}\left(T_\phi(p\eta) + \frac{T_\phi'(p\eta)}{\mathcal{H}}\right)\left(T_\phi(|\boldsymbol{k}-\boldsymbol{p}|\eta) + \frac{T_\phi'(|\boldsymbol{k}-\boldsymbol{p}|\eta)}{\mathcal{H}}\right). \end{aligned} \tag{20}$$

Here we have used the definition of the scalar transfer function $\phi_k(\eta) \equiv T_\phi(k\eta)\phi_k(0)$, obtained by solving the equation of motion for ϕ_k, as well as the fact that the initial value of the Newtonian potential $\phi_k(0)$ is related to the frozen curvature perturbation $\mathcal{R}_k$ on superhorizon scales via [33]

$$\phi_k(0) = \frac{3+3w}{5+3w}\mathcal{R}_k, \tag{21}$$

where w is the equation of state of the Universe at the time at which the initial conditions are imposed (that is, shortly after the end of inflation). In Section 3 we will give explicit expressions for the Green's function $G_k(\eta,\eta')$ and the transfer function.

At NNLO, the lepton number density is given by

$$a^3 n_{\rm L} = \frac{1}{32\pi^2}\int \epsilon_{ijk}\partial_j\partial_\ell\phi\partial_k\boldsymbol{h}_{i\ell}'d\eta, \tag{22}$$

where we reiterate that we have neglected all terms involving the first-order tensor modes h_{ij}. Using eq. (18) we can write this quantity as

$$a^3 n_{\rm L} = \frac{1}{32\pi^2}\left(\frac{3+3w}{5+3w}\right)\epsilon_{ijk}\sum_s \int d\eta \int \frac{d^3\boldsymbol{p}}{(2\pi)^3}\int \frac{d^3\boldsymbol{q}}{(2\pi)^3} e^{i(\boldsymbol{p}+\boldsymbol{q})\cdot\boldsymbol{x}}(iq_j)(iq_\ell)(ip_k)e^s_{i\ell}(\boldsymbol{p})$$
$$T_\phi(q\eta)\frac{1}{p^2}\int \frac{d^3\boldsymbol{k}}{(2\pi)^3}\Big[\boldsymbol{k}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{k}\Big] I'_p(\eta,k,|\boldsymbol{p}-\boldsymbol{k}|)\mathcal{R}_{\boldsymbol{q}}\mathcal{R}_{\boldsymbol{k}}\mathcal{R}_{\boldsymbol{p}-\boldsymbol{k}}. \tag{23}$$

As we will see in Section 5, the mean value of this quantity once again vanishes, but the variance does not. In general, $n_{\rm L}$ is given by the sum of both the NLO and NNLO contributions.

2. Black hole masses and abundance

Black holes form when sufficiently large perturbations re-enter the horizon after the end of inflation. Their abundance can be described using the Press-Schechter formalism, which states that collapse occurs only when the density contrast $\delta = \delta\rho/\rho$ is above a critical value δ_c,

$$\beta = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{\delta_c}^{\infty} e^{-\delta^2/(2\sigma^2)}d\delta, \tag{24}$$

where $\beta = \rho_{\rm PBH}/(\gamma\rho)$ is the ratio of the energy density in a Hubble patch that ends up in the form of PBHs to the total energy density at the time of collapse, with γ an $\mathcal{O}(1)$ factor encoding the efficiency of the collapse.[9] In the above expression we have assumed the probability distribution of the fluctuations to be Gaussian and the smoothed variance of the density contrast σ can be related to the dimensionless primordial spectrum of curvature perturbations $\mathcal{P}_\mathcal{R}$ by means of the gradient expansion [35]

$$\sigma^2 = \frac{4(1+w)^2}{(5+3w)^2}\int \frac{dq}{q}\left(\frac{q}{k}\right)^4 \mathcal{P}_\mathcal{R}(q)W^2(q/k), \tag{25}$$

where $W(x) = e^{-x^2/2}\sqrt{2/\pi}$ is a window function, which we take as a Gaussian. The dependence of the critical value δ_c in eq. (24) on the equation of state parameter w was analytically estimated in [36] to be

$$\delta_c = \frac{3(1+w)}{5+3w}\sin^2\left(\frac{\pi\sqrt{w}}{1+3w}\right). \tag{26}$$

Note that this formula is not valid for $w \ll 1/3$, since in this case the non-sphericity and angular momentum of the collapsing cloud, neglected in [36], become relevant, see [37, 38].

We remark that recent studies have focused on using the peak theory formalism instead of Press-Schechter to calculate the PBH abundance, see e.g. [39]. Similarly, the fact that in models of PBH formation the probability distribution deviates significantly from the Gaussian approximation is well-understood [40]. However, as noted in [39, 40], the change in abundance induced by taking into account these modifications can always be compensated by multiplying the power spectrum by an $\mathcal{O}(1)$ factor, since eq. (24) depends exponentially on $\mathcal{P}_\mathcal{R}$. Since the induced gravitational wave spectrum depends quadratically on $\mathcal{P}_\mathcal{R}$, these corrections will not heavily impact the results

[9] We set $\gamma = 0.2$ throughout this work [34], but we remark that this number could depend on the equation of state parameter w at the time at which the black holes form.

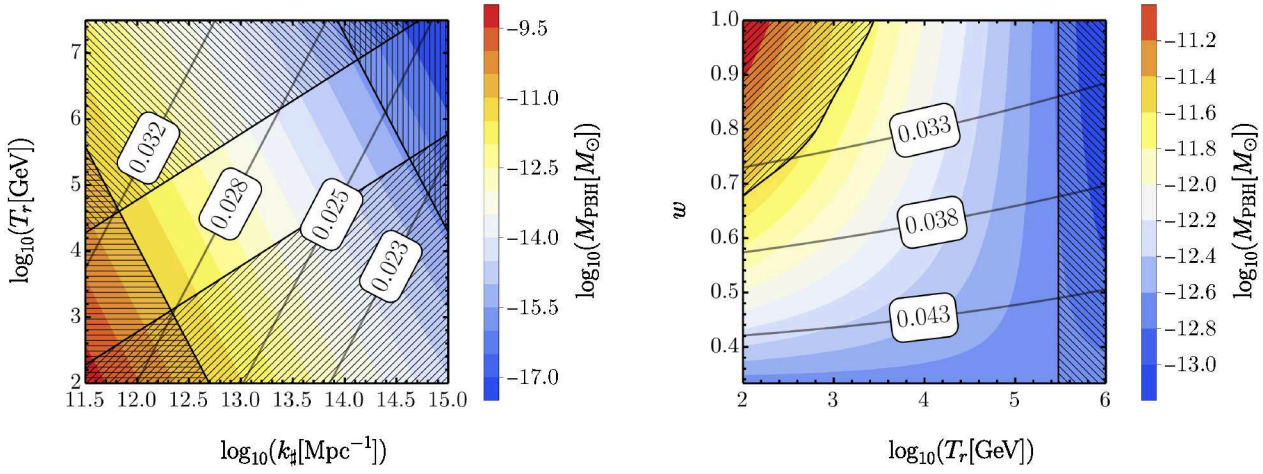

Figure 1: *Left panel: contours depicting the black hole masses as a function of the reheating temperature T_r and the scale $k_\sharp$ at which the peak in the power spectrum is located for $w = 1$, together with the amplitude of the power spectrum $\mathcal{A}_\sharp$ necessary to obtain $f_{\mathrm{PBH}} = 1$ (labeled solid lines). The horizontal and vertical shaded regions represent the constraints on the PBH masses in eq. (2). The region with bottom-right to top-left shading represents the re-entry constraint in eq. (32), and the region with bottom-left to top-right shading represents the GW constraint in eq. (54). Right panel: same as left panel, but leaving w as a free parameter, for $k_\sharp = 5 \times 10^{12}\,\mathrm{Mpc}^{-1}$.*

presented in this work. The same argument applies to the threshold δ_c, which strictly speaking should be calculated by finding the local maxima of the compaction function, a quantity that measures how close the overdensity is to fulfilling the hoop conjecture, see e.g. [41]. The specific value of δ_c will not be particularly important for our arguments, since the only issue relevant for us is the value of $\mathcal{P}_\mathcal{R}$ necessary to obtain $\Omega^0_{\mathrm{PBH}} = \Omega^0_{\mathrm{DM}}$ (that is, for PBHs to form all of the dark matter today), which is always of order $\mathcal{P}_\mathcal{R} \sim 10^{-2}$. We take eq. (26) simply as a convenient benchmark.

In what follows we consider a situation in which the Universe goes through a long phase of reheating after inflation with an arbitrary equation of state $w > 1/3$ during which PBHs form, and then transitions instantaneously into a radiation-dominated era at temperature T_r. By following a procedure analogous to the one presented in [42] for the case $w = 0$, we can write the present fraction of dark matter in the form of PBHs $f_{\mathrm{PBH}} \equiv \Omega^0_{\mathrm{PBH}}/\Omega^0_{\mathrm{DM}}$ as

$$f_{\mathrm{PBH}} = \gamma\beta \frac{\Omega^0_\gamma}{\Omega^0_{\mathrm{DM}}} \left[\left(\frac{g_{\star s}(T_r)}{g_{\star s}(T_0)}\right)^{1/3} \frac{T_r}{T_0}\right]^{\frac{1+9w}{1+3w}} \left(\frac{M_p^2 k^2}{T_r^4} \frac{90}{\pi^2 g_\star(T_r)}\right)^{\frac{3w}{1+3w}}, \tag{27}$$

where T_0 is the temperature of the Universe today, Ω_γ the energy density in radiation, $g_\star$ denotes the effective number of relativistic degrees of freedom, and $g_{\star s}$ the effective number of degrees of freedom in entropy.[10] In deriving the above equation we have assumed that collapse occurs immediately after perturbations with wavenumber k re-enter the horizon, at a time t_k defined by $k = a(t_k)H(t_k)$. We assume entropy is conserved only between the transition time t_r and today, but not necessarily between the time of collapse and t_r. The mass of the PBHs is proportional to

[10] We take $g_\star(T_r) = g_{\star s}(T_r) = 106.75$, $g_\star(T_0) = 3.36$, and $g_{\star s}(T_0) = 3.94$ throughout the paper [33].

the mass contained in a Hubble patch at the time of formation, $M_{\rm PBH} = 4\pi\gamma M_p^2/H$. By following a similar procedure to the one we used for $f_{\rm PBH}$, we can write

$$M_{\rm PBH} = 4\pi\gamma M_p^2 \left(\frac{\pi^2}{90} g_\star(T_r) \frac{T_r^4}{M_p^2}\right)^{\frac{1}{1+3w}} \left(\frac{1}{k^3} \frac{g_{\star s}(T_0) T_0^3}{g_{\star s}(T_r) T_r^3}\right)^{\frac{1+w}{1+3w}}. \tag{28}$$

As a cross-check, note that, for $w = 1/3$, eqs. (27) and (28) become independent of T_r and, in particular, we recover the well-known scaling relation $M_{\rm PBH} \propto k^{-2}$ [13].

It is clear from eqs. (24) and (25) that to produce a large PBH population one requires a spectrum of curvature perturbations $\mathcal{P}_\mathcal{R}$ which is peaked at a particular scale k so as not to overproduce black holes with masses outside of the unconstrained range in eq. (2). Such a power spectrum can be obtained from single-field models of inflation in which the potential has an approximate inflection point leading to a short phase of ultra-slow-roll, see e.g. [13, 14]. In what follows we will consider two possibilities, a flat scale-invariant spectrum with an amplitude $\mathcal{A}_\flat$ that matches the value measured on CMB scales of $\mathcal{O}(10^{-9})$,

$$\mathcal{P}_\mathcal{R}(k) = \mathcal{A}_\flat, \tag{29}$$

and a sharp Dirac delta spectrum with an amplitude $\mathcal{A}_\sharp$ of $\mathcal{O}(10^{-2})$ in order to produce an $\mathcal{O}(1)$ fraction of PBHs as dark matter,

$$\mathcal{P}_\mathcal{R}(k) = k_\sharp \mathcal{A}_\sharp \delta(k - k_\sharp), \tag{30}$$

where we have assumed that the peak occurs at a scale $k_\sharp$. This form of the power spectrum will allow us to derive several results analytically later on. Putting everything together and considering the sharp power spectrum, we find the following expression for β at the peak of the distribution, which is located at $k = k_\sharp/\sqrt{2}$ (as can be checked by setting $d\sigma/dk = 0$),

$$\beta = \frac{1}{2}\text{erfc}\left[\frac{3e}{8}\sin^2\left(\frac{\pi\sqrt{w}}{1+3w}\right)\sqrt{\frac{\pi}{\mathcal{A}_\sharp}}\right]. \tag{31}$$

Note that the total PBH abundance is obtained by integrating $f_{\rm PBH}$ over k, but since we are considering essentially monochromatic mass distributions, we can approximate the abundance by the value of $f_{\rm PBH}$ at its peak, $f_{\rm PBH}(k = k_\sharp/\sqrt{2})$.

The black hole mass and abundance are plotted in Fig. 1. As anticipated, the power spectrum required to obtain $f_{\rm PBH} = 1$ is of order $\mathcal{P}_\mathcal{R} \sim 10^{-2}$ over the entire parameter space, so the w-dependent threshold δ_c does not have a particularly strong effect. In the Figures we also show the GW constraint derived in the next Section, in eq. (54), for a fixed value $\mathcal{A}_\sharp = 0.05$. Choosing the value of $\mathcal{A}_\sharp$ necessary to obtain $f_{\rm PBH} = 1$ at each point in parameter space would only marginally change this constraint. We have also included the constraint imposed by the requirement that collapse occurs in the reheating stage. That is, the mode with comoving wavenumber $k_\sharp$ that induces collapse must re-enter the horizon at most at the time of the transition,

$$k_\sharp > a_r H_r = \frac{T_0}{T_r}\left(\frac{g_{\star s}(T_0)}{g_{\star s}(T_r)}\right)^{1/3}\left(\frac{\pi^2 g_\star(T_r) T_r^4}{90 M_p^2}\right)^{1/2}. \tag{32}$$

We remark that this constraint does not impede the formation of black holes, but if it is not satisfied then collapse occurs during the radiation era, and one should use the expressions for the mass and abundance for $w = 1/3$ instead.

We conclude that a population of PBHs with unconstrained masses which is large enough to explain the observed dark matter abundance can only form during an early stiff era (with $w = 1$) in a narrow region of parameter space, for $10^3\,\mathrm{GeV} < T_r < 10^7\,\mathrm{GeV}$, depending on the location of the peak in the power spectrum. We therefore find that the new GW constraint derived in this work (see the following Section) significantly restricts the parameter space for this mechanism.

3. Induced gravitational waves

In this Section we solve the equation of motion (15) for the tensor modes induced by scalar perturbations at second order. The solution, shown in eq. (18) can be written in terms of the Green's function $G_k(\eta, \eta')$ of the differential equation and the scalar transfer function $T_\phi(k\eta)$, which we now determine.

The Green's function can be obtained by solving the homogeneous equation

$$\boldsymbol{h}_k^{s\prime\prime} + 2\mathcal{H}\boldsymbol{h}_k^{s\prime} + k^2\boldsymbol{h}_k^s = 0. \tag{33}$$

Since we assume that the transition between the reheating stage with equation of state parameter w and the radiation epoch occurs instantaneously at η_r, the evolution of $\mathcal{H}$ is given by

$$\mathcal{H} = \begin{cases} \dfrac{2}{(1+3w)\eta}, & \text{for } \eta < \eta_r, \\ \dfrac{1}{\eta - \eta_w}, & \text{for } \eta_r < \eta, \end{cases} \tag{34}$$

where we have chosen the integration constant $\eta_w \equiv (1-3w)\eta_r/2$ to make the function continuous at η_r. We remark that throughout the paper w always denotes the equation of state parameter during the stage of reheating after inflation. We never use w to denote the equation of state parameter during the radiation era, and instead write explicitly $p/\rho = 1/3$ wherever necessary.

The general solution to eq. (33) is

$$h_k^s(\eta) = \begin{cases} (k\eta)^{m_\ell}\Big[A_{h\ell}J_{m_\ell}(k\eta) + B_{h\ell}Y_{m_\ell}(k\eta)\Big], & \text{for } \eta < \eta_r, \\ \dfrac{1}{k\eta - k\eta_w}\Big[A_{hr}\sin(k\eta - k\eta_w) + B_{hr}\cos(k\eta - k\eta_w)\Big], & \text{for } \eta_r < \eta, \end{cases} \tag{35}$$

where the ℓ subscript denotes quantities in the reheating stage, and r refers to quantities in the radiation era. We have also defined

$$m_\ell \equiv \frac{3(w-1)}{2(1+3w)}. \tag{36}$$

The Green's function is given by

$$G_k(\eta, \eta') = \frac{h_1(\eta)h_2(\eta') - h_1(\eta')h_2(\eta)}{h_1'(\eta')h_2(\eta') - h_1(\eta')h_2'(\eta')}, \tag{37}$$

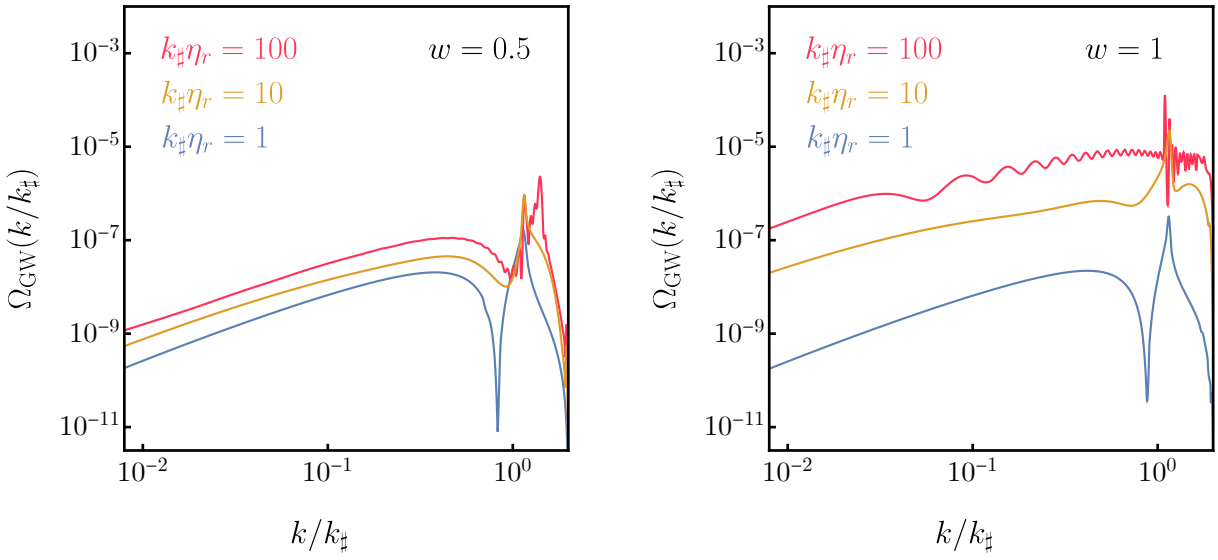

Figure 2: *Left panel: gravitational wave energy density as a function of $k/k_\sharp$ for $w = 0.5$ varying the dimensionless parameter $k_\sharp\eta_r$. The energy density grows as this parameter increases. Right panel: same as left panel, for $w = 1$.*

where h_1 and h_2 are any two linearly independent solutions to eq. (33) and we have suppressed the k and s indices in h for readability. Since this expression is independent of which pair of solutions is chosen, we can simply fix the integration constants in one of the regions by hand, and obtain the constants in the other epoch by imposing continuity of the solutions and their derivatives at η_r. We therefore set $A^1_{hr} = 1$ and $B^1_{hr} = 0$ for the first solution, together with $A^2_{hr} = 0$ and $B^2_{hr} = 1$ for the second one.

For the calculation of the lepton number density we need the tensor transfer function, defined by

$$h^s_k(\eta) = T_h(k\eta) h^s_k(0). \tag{38}$$

We remind the reader that initial conditions are imposed at the end of inflation, which we choose as $\eta = 0$. Assuming that tensor modes are initially frozen outside the horizon (so that we can impose the initial conditions $T_h(0) = 0$ and $T'_h(0) = 1$ for superhorizon modes after inflation ends), we find the following expression for the tensor transfer function

$$T_h(k\eta) = \begin{cases} \dfrac{\Gamma(1-m_\ell)}{2^{m_\ell}} (k\eta)^{m_\ell} J_{-m_\ell}(k\eta), & \text{for } \eta < \eta_r, \\ \dfrac{1}{k\eta - k\eta_w}\Big[A_{hr}\sin(k\eta - k\eta_w) + B_{hr}\cos(k\eta - k\eta_w)\Big], & \text{for } \eta_r < \eta. \end{cases} \tag{39}$$

We remark that the transfer function is unrelated to the Green's function, and in this case the constants during the reheating stage are fixed by the initial conditions mentioned above, whereas the constants A_{hr} and B_{hr} must be determined by imposing continuity of the solutions and their derivatives at η_r.

The equation of motion for the Newtonian potential is

$$\phi_k'' + 3\left(1+\frac{p}{\rho}\right)\mathcal{H}\phi_k' + \frac{p}{\rho}k^2\phi_k = 0, \tag{40}$$

where we have assumed that the background is dominated by a perfect fluid with $\delta p = (p/\rho)\delta\rho$. We remind the reader that we are working in the Newtonian gauge and in the absence of anisotropic stress. Assuming once again that the modes are initially frozen outside the horizon, we obtain the following expression for the transfer function, valid for $w \neq 0$,

$$T_\phi(k\eta) = \begin{cases} \dfrac{\Gamma(n_\ell+1)2^{n_\ell}}{w^{n_\ell/2}}(k\eta)^{-n_\ell}J_{n_\ell}\left(k\eta\sqrt{w}\right), & \text{for } \eta < \eta_r, \\ \left(\dfrac{\sqrt{3}}{k\eta-k\eta_w}\right)^2\left\{A_{\phi r}\left[\sin\left(\dfrac{k\eta-k\eta_w}{\sqrt{3}}\right)+\left(\dfrac{\sqrt{3}}{k\eta-k\eta_w}\right)\cos\left(\dfrac{k\eta-k\eta_w}{\sqrt{3}}\right)\right]+\right. & \\ \left. +B_{\phi r}\left[\cos\left(\dfrac{k\eta-k\eta_w}{\sqrt{3}}\right)-\left(\dfrac{\sqrt{3}}{k\eta-k\eta_w}\right)\sin\left(\dfrac{k\eta-k\eta_w}{\sqrt{3}}\right)\right]\right\}, & \text{for } \eta_r < \eta, \end{cases} \tag{41}$$

where

$$n_\ell = \frac{5+3w}{2(1+3w)} = 1 - m_\ell \tag{42}$$

and the constants $A_{\phi r}$ and $B_{\phi r}$ can once again be found by matching the solutions and their derivatives at η_r.

With these ingredients in hand we can calculate the I_k function in eq. (19). The GW energy density is given by [43]

$$\Omega_{\rm GW}(\eta,k) = \frac{1}{24}\left(\frac{k}{\mathcal{H}}\right)^2\langle\mathcal{P}_{\boldsymbol{h}}(\eta,k)\rangle, \tag{43}$$

where the brackets $\langle\cdots\rangle$ denote a time average, which must be taken due to the stochasticity of the signal, and $\mathcal{P}_{\boldsymbol{h}}$ denotes the power spectrum of $\boldsymbol{h}_{ij}$. Starting from eq. (18), we find the following expression for the power spectrum (see e.g. [23, 24])

$$\mathcal{P}_{\boldsymbol{h}}(\eta,k) = \int_0^\infty dy\int_{|1-y|}^{1+y}dz\left[\frac{4y^2-(1+y^2-z^2)^2}{8yz}\right]^2\mathcal{P}_{\mathcal{R}}(ky)\mathcal{P}_{\mathcal{R}}(kz)I_k^2(\eta,ky,kz). \tag{44}$$

Since we are interested in measuring this quantity at late times, we can take $\eta\to\infty$ in eq. (19). After pulling the solutions $h_1(\eta)$ and $h_2(\eta)$ in the Green's function out of the integral and splitting the limits of integration into the two contributions $(0,\eta_r)$ and (η_r,∞), we obtain the following expression, valid late into the RD era,

$$I_k(\eta>\eta_r) = \frac{\sin(k\eta-k\eta_w)}{k\eta-k\eta_w}(J_\ell^s+J_r^s) - \frac{\cos(k\eta-k\eta_w)}{k\eta-k\eta_w}(J_\ell^c+J_r^c), \tag{45}$$

where

$$J_\ell^s = \left(\frac{3+3w}{5+3w}\right)^2\left(\frac{\pi/2}{A_{\ell h}^2B_{\ell h}^1-A_{\ell h}^1B_{\ell h}^2}\right)\int_0^{\eta_r}(k\eta')^{-2m_\ell+1}h_2^\ell(\eta')Q_\ell(\eta',ky,kz)kd\eta', \tag{46}$$

$$J_\ell^c = \left(\frac{3+3w}{5+3w}\right)^2\left(\frac{\pi/2}{A_{\ell h}^2B_{\ell h}^1-A_{\ell h}^1B_{\ell h}^2}\right)\int_0^{\eta_r}(k\eta')^{-2m_\ell+1}h_1^\ell(\eta')Q_\ell(\eta',ky,kz)kd\eta', \tag{47}$$

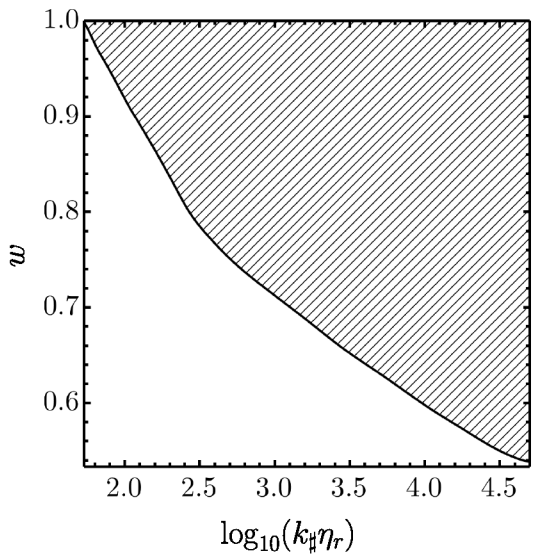

Figure 3: *Constraint in eq. (54) as a function of the two parameters w and $k_\sharp\eta_r$. The top-right region is forbidden due to the fact that the redshift of the tensor modes during an epoch with a stiff equation of state makes the energy density grow, as illustrated in Fig. 2.*

and

$$J_r^s = \left(\frac{3+3w}{5+3w}\right)^2 \int_{\eta_r}^{\infty} (k\eta' - k\eta_w)\cos(k\eta' - k\eta_w) Q_r(\eta', ky, kz) k d\eta', \tag{48}$$

$$J_r^c = \left(\frac{3+3w}{5+3w}\right)^2 \int_{\eta_r}^{\infty} (k\eta' - k\eta_w)\sin(k\eta' - k\eta_w) Q_r(\eta', ky, kz) k d\eta'. \tag{49}$$

Analytical expressions can be found for the last two integrals, whereas the first two must in general be performed numerically.

Squaring I_k and taking the average as required by eq. (43), we find

$$\frac{k^2}{\mathcal{H}^2}\langle I_k^2\rangle = \frac{1}{2}\left[(J_\ell^s + J_r^s)^2 + (J_\ell^c + J_r^c)^2\right] \equiv \frac{1}{2}J^2(y, z, k\eta_r), \tag{50}$$

where we have used $\langle\sin^2(k\eta)\rangle = \langle\cos^2(k\eta)\rangle = 1/2$ and $\langle\sin(k\eta)\cos(k\eta)\rangle = 0$ and highlighted the dependence of J on the dimensionless parameter $k\eta_r$. We can evaluate the integrals over y and z in eq. (44) by using the sharp power spectrum in eq. (30) necessary to obtain a significant PBH population. The result is

$$\Omega_{\rm GW}(k/k_\sharp) = \mathcal{A}_\sharp^2 \frac{\Omega_\gamma(T_0)}{3072}\frac{g_\star(T)}{g_\star(T_0)}\left(\frac{g_{\star s}(T_0)}{g_{\star s}(T)}\right)^{4/3}\frac{k^2}{k_\sharp^2}\left(4\frac{k_\sharp^2}{k^2} - 1\right)^2 J^2\left(\frac{k_\sharp}{k}, \frac{k_\sharp}{k}, k\eta_r\right)\Theta_{\rm H}\left(2\frac{k_\sharp}{k} - 1\right), \tag{51}$$

where $\Theta_{\rm H}$ denotes the Heaviside function, and we have used the fact that these gravitational waves behave as radiation at late times to evolve the result from some late time deep into the radiation era until today. This expression only depends on the two dimensionless quantities $k_\sharp/k$ and $k_\sharp\eta_r$.

The transition time is, in terms of temperature,

$$\eta_r = \frac{1}{T_r}\left(\frac{2}{1+3w}\right)\sqrt{\frac{90}{\pi^2 g_\star(T_r)}}\left(\frac{g_{\star s}(T_r)}{g_{\star s}(T_0)}\right)^{1/3}\frac{M_p}{T_0}. \tag{52}$$

In Fig. 2 we show the GW energy density from eq. (51) for different choices of the two parameters w and $k_\sharp \eta_r$. We find that the GW abundance grows for stiff equations of state, as reported in [27], due to the fact that the tensor modes redshift more slowly than the background. We also find that the energy density grows as $k_\sharp \eta_r$ increases. The reason for this is that the largest contribution to the momentum integral in eq. (44) comes from scales around the narrow peak in the scalar power spectrum, and therefore we should only expect the signal to be affected if the relevant modes re-enter the horizon before the transition to radiation has taken place, so that $k_\sharp > \mathcal{H}_r \propto \eta_r^{-1}$. For $k_\sharp \eta_r \ll 1$, the relevant modes enter during the radiation era and the effect of the stiff epoch on the signal is washed out, as can be clearly seen in the Figure.

The total energy density can be found by integrating

$$\Omega_{\rm GW} = \int d\log k\, \Omega_{\rm GW}(k). \tag{53}$$

As we anticipated earlier, there is a bound on this quantity arising from CMB observations and the abundance of light elements produced during Big Bang nucleosynthesis [30, 31],

$$\Omega_{\rm GW} h^2 < 1.8 \times 10^{-6}. \tag{54}$$

Since the signal grows as both w and $k_\sharp \eta_r$ increase, this bound is eventually violated. In Fig. 3 we perform a numerical scan over the parameter space, showing the forbidden region.[11] By using eq. (52), this bound can be written in terms of the transition temperature, and translated to a constraint on the PBH mass and abundance, shown in Fig. 1 for $w = 1$ (left panel) and for $k_\sharp = 5 \times 10^{12}\,{\rm Mpc}^{-1}$ (right panel). These constraints are one of the main results of this paper. As anticipated at the end of the previous Section, this bound significantly limits the available parameter space for PBH formation during a stiff epoch. The constraint disappears completely for the standard radiation scenario with $w = 1/3$, where the bound is satisfied. In these Figures we have taken $\mathcal{A}_\sharp = 5 \times 10^{-2}$ (thereby assuming that $f_{\rm PBH} \simeq 1$) and neglected the mild dependence of this number on w depicted in Fig. 1.

Before moving on, we remark that, although we have restricted our attention to stiff epochs due to the fact that PBH formation during a matter-dominated era is not particularly well understood at the time of writing (since the effect of angular momentum and non-sphericity of the collapsing cloud becomes important [37, 38]), the calculation of the GW spectrum presented in this Section is also valid for the case $w = 0$, with the only modification being the scalar transfer function in eq. (41), which takes the simpler form $T_\phi(k\eta) = 1$ for $\eta < \eta_r$.

[11] The small kink towards the center of the plot is a numerical artifact due to the precision of our scan and has no physical significance.

4. Gradual decay into radiation

For the calculation in the previous Section, we assumed that the transition between the reheating stage and the radiation era occurred instantaneously. We now study the case in which the initial fluid decays gradually, thereby making the transition between epochs smooth. This scenario was studied in [26] for the particular case $w = 0$ and for a scale-invariant scalar spectrum. Here we extend these results for a generic fluid with constant equation of state parameter w and using a peaked scalar spectrum relevant for PBH formation. It was reported in [26] that, for a scale-invariant scalar spectrum, the induced gravitational wave spectrum in the smooth transition case is suppressed with respect to its sudden counterpart. As we will see, although the same conclusion remains true for a peaked spectrum, the suppression is very mild and therefore does not strongly affect the bounds derived in the previous Section.

In what follows we denote quantities related to the radiation fluid by a subscript γ, and quantities related to the decaying fluid by a subscript w. We model the transition by considering a rate of energy transfer of the form $\Gamma\rho_w$ between the two fluids, with Γ constant, such that the total stress-energy tensor $T_{\mu\nu} = T_{\mu\nu}{}^{(\gamma)} + T_{\mu\nu}{}^{(w)}$ is conserved, but the separate components are not, so

$$\nabla_\nu T^{\mu\nu(\gamma)} = Q^\mu = -\nabla_\nu T^{\mu\nu(w)} \tag{55}$$

for some vector Q^μ [44]. The background equations then become, using the number of e-folds $dN \equiv \mathcal{H}d\eta$ as the time variable,

$$3M_p^2 H^2 = \rho_\gamma + \rho_w, \tag{56}$$

$$\frac{d\rho_w}{dN} + 3(1+w)\rho_w = -\frac{\Gamma}{H}\rho_w, \tag{57}$$

$$\frac{d\rho_\gamma}{dN} + 4\rho_\gamma = \frac{\Gamma}{H}\rho_w. \tag{58}$$

At the perturbative level, we work in Newtonian gauge and assuming no anisotropic stress. We use the following dimensionless set of variables

$$\delta_\gamma \equiv \frac{\delta\rho_\gamma}{\rho_\gamma}, \qquad \delta_w \equiv \frac{\delta\rho_w}{\rho_w}, \qquad \theta_\gamma \equiv H\frac{\delta q_\gamma}{\rho_\gamma}, \qquad \theta_w \equiv H\frac{\delta q_w}{\rho_w}, \tag{59}$$

where $\delta q_b = a(\rho_b + p_b)\delta v_b$, and δv_b denotes the velocity perturbation for each fluid component.[12] The Einstein equation that determines the evolution of the Newtonian potential ϕ is

$$\frac{d\phi}{dN} + \left(1 + \frac{1}{3}\frac{k^2}{\mathcal{H}^2}\right)\phi + \frac{1}{2}\left(\frac{\rho_w}{\rho}\delta_w + \frac{\rho_\gamma}{\rho}\delta_\gamma\right) = 0. \tag{60}$$

[12] Note that θ is sometimes used instead to denote the divergence of the velocity perturbation (e.g. in [26]), so one must be careful when comparing different references.

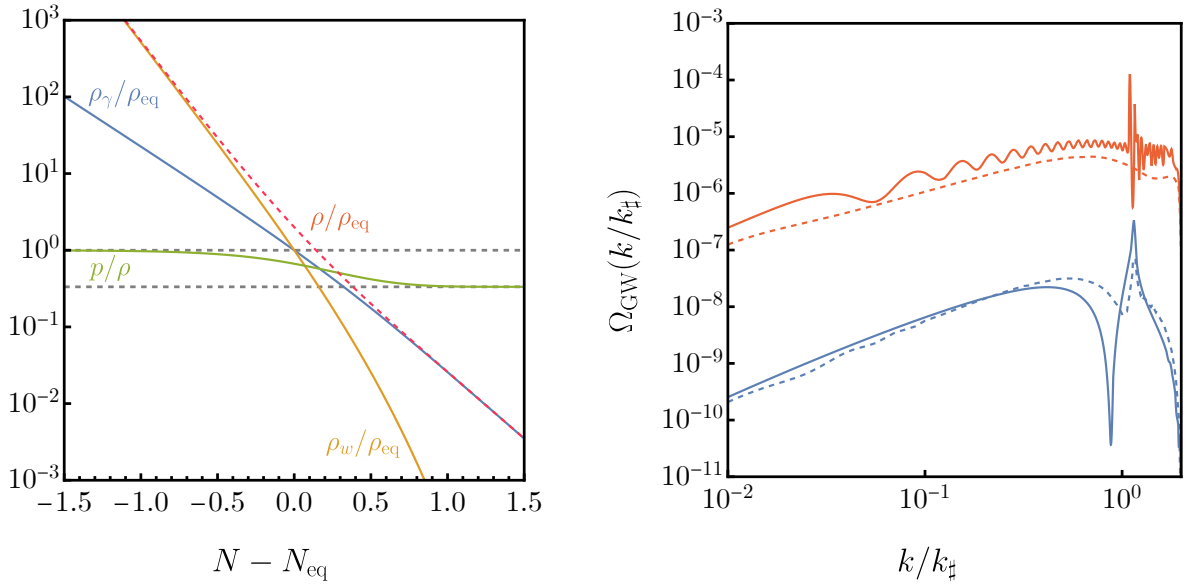

Figure 4: *Left panel: evolution of the background energy density of each fluid, as well as the total energy density ρ and the equation of state p/ρ, for the case $w = 1$. By normalizing all quantities with respect to their value at the time $N_{\rm eq}$ defined by $\rho_\gamma = \rho_w$, the plot becomes independent of Γ. Right panel: calculation of the GW energy density for the sudden transition case (solid) and the gradual case (dashed) for $w = 1$ with $k_\sharp \eta_r = 1$ (blue) and $k_\sharp \eta_r = 100$ (red). The time η_r in the gradual transition scenario is defined through $k_{\rm eq}$ in eq. (69). The integrated energy density is very similar for both cases, leaving the bounds of the previous Section essentially unaltered.*

The continuity and Euler equations for the fluid are [44]

$$\frac{d\delta_w}{dN} + \frac{\Gamma}{H}\phi - 3(1+w)\frac{d\phi}{dN} - \frac{k^2}{\mathcal{H}^2}\theta_w = 0, \tag{61}$$

$$\frac{d\delta_\gamma}{dN} + \frac{\Gamma}{H}\frac{\rho_m}{\rho_\gamma}(\delta_\gamma - \delta_w - \phi) - 4\frac{d\phi}{dN} - \frac{k^2}{\mathcal{H}^2}\theta_\gamma = 0, \tag{62}$$

$$\frac{d\theta_w}{dN} + \left(\frac{3}{2}\frac{\rho+p}{\rho} - \frac{w}{1+w}\frac{\Gamma}{H} - 3w\right)\theta_w + (1+w)\phi + w\delta_w = 0, \tag{63}$$

$$\frac{d\theta_\gamma}{dN} + \left(\frac{\Gamma}{H}\frac{\rho_w}{\rho_\gamma} + \frac{3p+\rho}{2\rho}\right)\theta_\gamma + \frac{4}{3}\phi + \frac{1}{3}\delta_\gamma - \frac{\Gamma}{H}\frac{\rho_w}{\rho_\gamma}\frac{1}{1+w}\theta_w = 0, \tag{64}$$

where we have introduced the additional assumption that $\delta p_w = w\delta\rho_w$, and similarly for the radiation component. Together, these five equations form a complete system that can be solved numerically.

To solve the above system of equations, we choose adiabatic initial conditions [33]. On super-horizon scales $k \ll \mathcal{H}$ and at sufficiently early times so that $\Gamma \ll H$ and $\rho_\gamma \ll \rho_w$, the equation for the Newtonian potential ϕ becomes

$$\frac{d\phi}{dN} + \phi + \frac{1}{2}\delta_w = 0. \tag{65}$$

Since there is essentially only one fluid present at this time, and we have assumed $\delta p_w = w\delta\rho_w$,

the Newtonian potential can be shown to satisfy eq. (40) and is therefore frozen on superhorizon scales. Thus, on these scales, we have from the above equation the relation

$$\delta_w = -2\phi. \tag{66}$$

The adiabatic initial conditions for the energy density perturbations are [33]

$$\frac{\delta_\gamma}{4/3} = \frac{\delta_w}{1+w}. \tag{67}$$

Finally, if we assume that θ_γ and θ_w are also frozen on superhorizon scales, we obtain, from the last two equations,

$$\theta_\gamma = -\frac{8\phi}{9(1+w)}, \qquad \theta_w = -\frac{2}{3}\phi. \tag{68}$$

By using the relation (21), we can therefore fix all initial conditions in terms of $\mathcal{R}$.

The evolution of the background energy density of each fluid is shown in Fig. 4 for $w = 1$. Note that, as explained in [26], the evolution of the background quantities (and in fact, that of the perturbations) is completely independent of the choice of Γ once they are normalized by their values at the time $N_{\rm eq}$ defined by $\rho_\gamma(N_{\rm eq}) = \rho_w(N_{\rm eq}) \equiv \rho_{\rm eq}$, as can be checked from eqs. (56-58). The left panel of Fig. 5 shows the time evolution of the perturbations for particular values of the parameters. We define the transition time η_r in this case via

$$\eta_r \equiv \frac{2}{(1+3w)k_{\rm eq}}, \tag{69}$$

where $k_{\rm eq}$ is the mode that re-enters the horizon at $N_{\rm eq}$. This is the relation between $k_{\rm eq}$ and η_r in the sudden transition case of the previous Section, so this definition allows us to make a fair comparison between both scenarios.

For the calculation of the gravitational wave spectrum we need both the scalar transfer function T_ϕ, obtained by solving the system in eqs. (60-64) with initial condition $T_\phi = 1$ (the condition $\frac{d}{dN}T_\phi = 0$ is guaranteed by the choice of adiabatic initial conditions for the other perturbations), and the Green's function for the tensor modes, which we now turn our attention to. The integral in eq. (19) can be rewritten in terms of the number of e-folds as

$$I_k = \left(\frac{3+3w}{5+3w}\right)^2 \int_{-\infty}^{\infty} \frac{k^2}{\mathcal{H}^2} \frac{h_1(N)h_2(N') - h_2(N)h_1(N')}{h_2(N')\frac{d}{dN}h_1(N') - h_1(N')\frac{d}{dN}h_2(N')} Q(N', p, |\boldsymbol{k}-\boldsymbol{p}|)dN'. \tag{70}$$

The equation of motion for the tensor modes is, in terms of the number of e-folds,

$$\frac{d^2 h_k^s}{dN^2} + \frac{3}{2}\left(1 - \frac{p}{\rho}\right)\frac{dh_k^s}{dN} + \frac{k^2}{\mathcal{H}^2}h_k^s = 0. \tag{71}$$

To calculate the GW spectrum today, we must evaluate eq. (70) at late times, deep into the radiation era. At this stage the background pressure is $p = \rho/3$, and the Hubble scales as $\mathcal{H} = \mathcal{H}_c(a_c/a)$, where the c subscript denotes some late time $N_c \gg N_{\rm eq}$. The product $a_c\mathcal{H}_c$ approaches a constant value, which can be found numerically, after the transition. Using this scaling in the above equation,

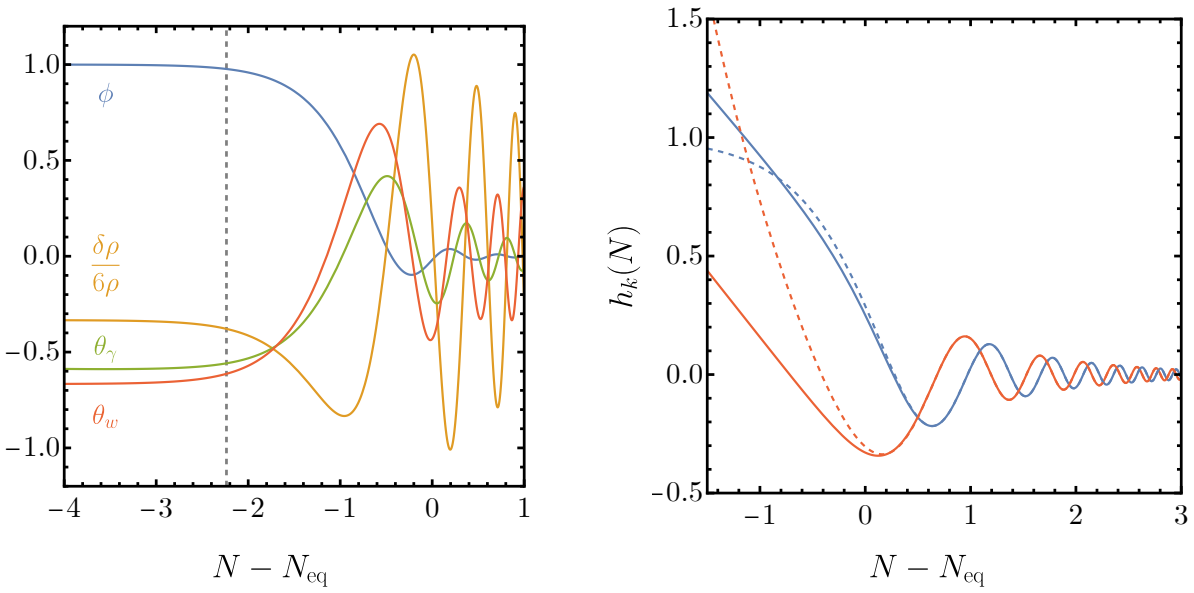

Figure 5: *Left panel: evolution of the perturbations with adiabatic initial conditions for* $w = 0.5$ *and* $k_\sharp \eta_r = 10$*, with* η_r *defined via eq. (69). The vertical dashed line denotes the horizon crossing time for this mode. Right panel: evolution of the two independent solutions to the homogeneous equation for the tensor modes used to construct the Green's function (solid lines) for* $w = 1$*,* $k_\sharp \eta_r = 1$*, and* $k/k_\sharp = 1$*. At late times the two solutions approach the asymptotic values in eq. (72) (dashed lines).*

we find the following two independent solutions at late times

$$h_1^{\rm late} = \frac{a_c}{a}\frac{\mathcal{H}_c}{k}\sin\left(\frac{a}{a_c}\frac{k}{\mathcal{H}_c}\right), \qquad h_2^{\rm late} = \frac{a_c}{a}\frac{\mathcal{H}_c}{k}\cos\left(\frac{a}{a_c}\frac{k}{\mathcal{H}_c}\right). \tag{72}$$

By using these two solutions we can take the time-average of I_k^2 as we did in the previous Section, and we obtain, after performing the momentum integral over the Dirac delta power spectrum,

$$\frac{k^2}{\mathcal{H}^2}\langle I_k^2\rangle = \frac{1}{2}(J_s^2 + J_c^2) = \frac{1}{2}J^2\left(\frac{k}{k_{\rm eq}}, \frac{k_\sharp}{k_{\rm eq}}\right), \tag{73}$$

where

$$J_s = \left(\frac{3+3w}{5+3w}\right)^2 \int_{-\infty}^{\infty} \frac{(k/k_{\rm eq})^2}{(\mathcal{H}/k_{\rm eq})^2} G_2(N', k/k_{\rm eq}) Q(N', k_\sharp/k_{\rm eq}) dN', \tag{74}$$

$$J_c = \left(\frac{3+3w}{5+3w}\right)^2 \int_{-\infty}^{\infty} \frac{(k/k_{\rm eq})^2}{(\mathcal{H}/k_{\rm eq})^2} G_1(N', k/k_{\rm eq}) Q(N', k_\sharp/k_{\rm eq}) dN', \tag{75}$$

with

$$G_1(N', k/k_{\rm eq}) \equiv \frac{h_1(N')}{h_2(N')\frac{d}{dN}h_1(N') - h_1(N')\frac{d}{dN}h_2(N')}, \tag{76}$$

$$G_2(N', k/k_{\rm eq}) \equiv \frac{h_2(N')}{h_2(N')\frac{d}{dN}h_1(N') - h_1(N')\frac{d}{dN}h_2(N')}. \tag{77}$$

Once again we see that the GW energy depends only on $k/k_{\rm eq} = (k/k_\sharp)(k_\sharp/k_{\rm eq})$, as well as the two parameters $k_\sharp/k_{\rm eq}$ and w. The dependence on the parameter $k_{\rm eq}$ therefore replaces η_r from the previous section, with the relation between both given by eq. (69).

We remark that, although the Green's function does not depend on which two linearly independent solutions are chosen to construct it, the fact that we perform the time average over I_k^2 by using $\langle \sin^2(k\eta)\rangle = \langle \cos^2(k\eta)\rangle = 1/2$ means we need to project the two solutions that behave as eq. (72) at late times. This can be accomplished by imposing the boundary conditions $h_{1,2}(N_{\rm late}) = h_{1,2}^{\rm late}(N_{\rm late})$ at some late time $N_{\rm late} \gg N_{\rm eq}$, sufficiently deep into the radiation era (numerically, we find that a few e-folds after the transition suffice). The corresponding solutions are shown in the right panel of Fig. 5 together with their late time limits for $w = 1$, $k_\sharp \eta_r = 1$ and $k/k_\sharp = 1$.

We show the resulting GW energy density for $w = 1$ in the right panel of Fig. 4, for two illustrative examples with $k_\sharp \eta_r = 1$ and $k_\sharp \eta_r = 100$. We find that, although the signal is very mildly suppressed with respect to the sudden transition case of the previous Section, the difference between both is essentially negligible for the purpose of estimating the curve in Fig. 3, so we conclude that the results from the previous Section are robust. We nevertheless remark that this conclusion might change if the transition is modelled differently.

Having established that the gradual transition scenario is not significantly different from the sudden case, we go back to using the analytical formulas of the latter in the following Section.

5. Smoothed lepton number density

In this Section we estimate the size of the baryon asymmetry fluctuations induced by the gravitational chiral anomaly. The quantity of interest is the variance of the lepton number density[13] $n_{\rm L}$ computed in eqs. (14) and (23), averaged over a region of size r_σ,

$$\left\langle |n_{\rm L}|^2 \right\rangle_{r_\sigma} \equiv \left\langle \left| \int d^3\boldsymbol{r} W_{r_\sigma}(r) n_{\rm L}(\boldsymbol{x}+\boldsymbol{r}) \right|^2 \right\rangle, \tag{78}$$

where $W_{r_\sigma}(r)$ is a Window function that decays smoothly on scales $r \gg r_\sigma$, which we take to be a Gaussian for concreteness,

$$W_{r_\sigma}(r) = \left(\frac{1}{r_\sigma \sqrt{\pi}}\right)^3 e^{-r^2/r_\sigma^2}. \tag{79}$$

Let us clarify why we focus on the variance of this quantity. We would naively expect the baryon asymmetry to be determined by the mean value of $n_{\rm L}$. However, in the absence of a chiral gravitational wave background (generated for instance, via some inflationary coupling of the form $f(\varphi)R\tilde{R}$ such as the one considered in [7], where φ is the inflaton field), which would make the terms quadratic in h_{ij} of the form shown in eq. (5) nonzero, the mean value $\langle n_{\rm L}\rangle$ vanishes. This can be easily seen at NLO, since, schematically, taking the mean value of eq. (14) yields $\langle \phi h\rangle \sim \langle \phi\rangle\langle h\rangle = 0$. The fact that the mean value also vanishes at NNLO is less obvious. The dimensionless bispectrum

[13] Throughout this Section we often abuse language and refer to this quantity simply as the baryon asymmetry. In the leptogenesis scenario, the relation between the baryon and lepton number densities is, assuming the Standard Model matter content, $n_{\rm L} = (79/28)n_{\rm B}$ [45]. The measured value for this quantity is $n_{\rm L}/s \simeq 2 \times 10^{-10}$ [1, 2].

$\mathcal{B}_\mathcal{R}$ of curvature perturbations is defined by

$$\left\langle \mathcal{R}_{\boldsymbol{q}} \mathcal{R}_{\boldsymbol{k}} \mathcal{R}_{\boldsymbol{p}-\boldsymbol{k}} \right\rangle \equiv (2\pi)^3 \delta^3_{\boldsymbol{p}+\boldsymbol{q}} \frac{(2\pi^2)^2}{q^2 k^2 |\boldsymbol{p}-\boldsymbol{k}|^2} \mathcal{B}_\mathcal{R}(q, k, |\boldsymbol{p}-\boldsymbol{k}|), \tag{80}$$

where we have introduced the shorthand $\delta^3_{\boldsymbol{p}} \equiv \delta^3(\boldsymbol{p})$ for later convenience. This quantity vanishes if the fluctuations are Gaussian and is therefore slow-roll suppressed in conventional inflationary scenarios, but this could change depending on the dynamics of the inflaton (particularly in models of PBH production featuring a potential with a near-inflection point, in which the slow-roll approximation breaks down [40]). After some manipulation, we obtain, from eq. (23),

$$\begin{aligned} a^3 \left\langle |n_\mathrm{L}| \right\rangle_{r_\sigma} = \frac{\pi^2}{8} \left(\frac{3+3w}{5+3w} \right) \sum_s \int d\eta \int \frac{d^3\boldsymbol{p}}{(2\pi)^3} \epsilon_{ijk} (p_j p_\ell p_k) e^s_{i\ell}(\boldsymbol{p}) T_\phi(p\eta) \\ \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \Big[\boldsymbol{k} \cdot \boldsymbol{e}^s(\boldsymbol{p}) \cdot \boldsymbol{k} \Big] I'_p(\eta, k, |\boldsymbol{p}-\boldsymbol{k}|) \frac{\mathcal{B}_\mathcal{R}(p, k, |\boldsymbol{p}-\boldsymbol{k}|)}{p^4 k^2 |\boldsymbol{p}-\boldsymbol{k}|^2}, \end{aligned} \tag{81}$$

where we have used the Dirac delta function in the definition of the bispectrum to perform one of the momentum integrals, as well as the normalization of the window function

$$\int d^3\boldsymbol{r} W_{r_\sigma}(r) = 1. \tag{82}$$

The right-hand side of eq. (81) clearly vanishes, since $p_j p_k$ is symmetric, but ϵ_{ijk} is not. In what follows we assume that the curvature perturbation $\mathcal{R}$ follows a Gaussian distribution.

Instead of focusing on the mean value of this quantity, we assume that the baryon asymmetry of the Universe is generated via some other mechanism which we remain agnostic about, and shift our attention to the variance of n_L.[14] As noted in [11], since inflation is a stochastic process, we expect the lepton number density to deviate from its mean value in different patches. This deviation will generically have a magnitude of order $\sim \sqrt{\langle |n_\mathrm{L}|^2 \rangle_{r_\sigma}}$ in a region of size r_σ, leading to fluctuations in the baryon asymmetry. Before moving on with the calculation however, let us note that, as discussed in [11], this asymmetry does not survive at late times on arbitrarily small patches, due to annihilation processes. Let us consider two neighboring patches, one with a matter excess, and another one with an antimatter excess. If particles are able to freely travel from one patch to the other, annihilation will take place, leading to a smaller asymmetry overall. On sufficiently large patches, however, the asymmetry always survives at late times. To see why, note that the maximum distance a particle in the radiation bath can travel between collisions is λ/a, where λ is its mean free path. In a time Δt, the particle undergoes $N = \Delta t/\lambda$ collisions. The average displacement in a time Δt for a random walk is therefore $\Delta r = (\lambda/a)\sqrt{N} = (\lambda/a)\sqrt{\Delta t/\lambda}$. Integrating this displacement yields the Silk length [46], which sets the limit below which annihilation can take place.[15] On the other hand, at late times, after the electroweak sphaleron processes have taken place, the asymmetry is carried by quarks, which are relativistic. Once the temperature is low enough, quarks are confined into non-relativistic baryons and their mean free path drops significantly. The

[14] It is conceivable that the mechanism responsible for generating the baryon asymmetry will also lead to a spectrum of fluctuations which could be potentially larger than the contributions discussed here. The relative importance of each term is of course model-dependent and can only be assessed once a particular mechanism is fixed.

[15] This is analogous to photon diffusion in the CMB [33].

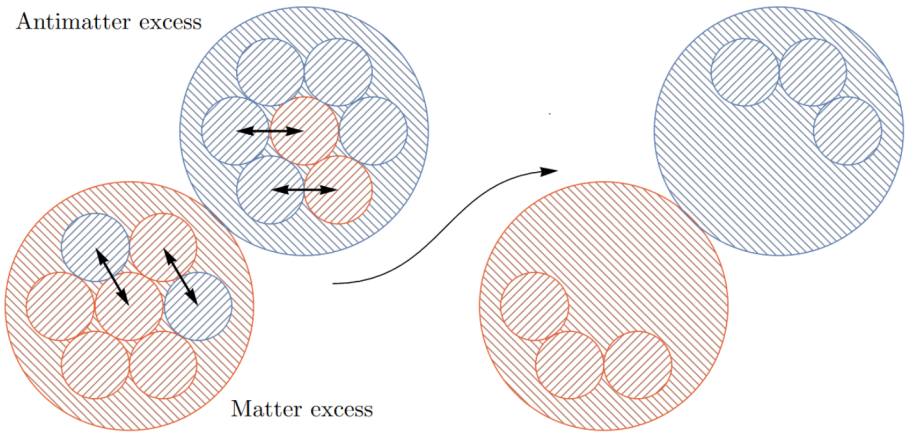

Figure 6: *Schematic depiction of the late-time behaviour of the baryon asymmetry fluctuations. Small neighbouring patches with matter and antimatter excesses annihilate, homogenizing the distribution on small scales. Larger regions remain unaffected, since quarks cannot travel from one patch to the other if the distance separating them is larger than the Silk length (see text). This length scale sets the range of validity for the calculations in this Section.*

Silk length is therefore approximately given by

$$r_{\rm S}^2 \simeq \int_0^{t_{\rm QCD}} dt\, \frac{\lambda_{\rm Q}}{a^2}, \tag{83}$$

where $\lambda_{\rm Q} = 1/\Gamma_{\rm Q}$ denotes the mean free path of quarks in the plasma, and $\Gamma_{\rm Q}$ is the interaction rate, which can be estimated as $\Gamma_{\rm Q} \simeq T$ [47]. For simplicity, let us neglect the effect of the equation of state on the evolution of a and assume that the Universe is always radiation-dominated. We also neglect the change in the relativistic degrees of freedom in entropy $g_{\star s}$ between the end of inflation and the QCD scale. We then have $T \simeq T_0/a$ and $H = H_0 a^{-2}$, which leads to

$$r_{\rm S}^2 \simeq \int_0^{a_{\rm QCD}} da\, \frac{\Gamma_{\rm Q}^{-1}}{a^3 H} \simeq \frac{1}{H_0 T_{\rm QCD}} \simeq \left(10^{-17}{\rm Mpc}\right)^2. \tag{84}$$

We remark that this is only an order of magnitude estimate,[16] and an accurate description of the damping dynamics requires solving the Boltzmann equation. We conclude that the asymmetry is preserved on scales larger than the Silk length of quarks at the confinement scale, $r_\sigma \gtrsim r_{\rm S} \simeq 10^{-17}$Mpc. This is the domain of validity of the calculations in this Section. Below this scale, annihilation dynamics become important and we expect the amplitude of the baryon asymmetry fluctuations to drop significantly.

In what follows we will compute the fluctuations in three different cases, 1. for the NLO term in eq. (14) with a flat scalar spectrum (29) and with the tensor spectrum given by the relation $\mathcal{P}_h = r\mathcal{A}_\flat$, where $r \lesssim 10^{-2}$ [48, 49] is the tensor-to-scalar ratio, 2. the same NLO term with a sharp

[16] A similar value of $r_{\rm S} \simeq 10^{-16}$Mpc was assumed in [11]. However, the specific number is not particularly important for our purposes since, as we will see, the peak in the spectrum of fluctuations generated by PBHs lies far above this scale.

scalar spectrum (30), but with a flat tensor spectrum given by the same relation as in the first case, and 3. the NNLO term in eq. (23) for a peaked scalar spectrum, and with the tensor modes induced by scalar perturbations. As we will see, each term dominates at different scales, and the total asymmetry fluctuations will in general be given by the sum of the three contributions, plus the subdominant mixed terms that we have neglected as per the arguments of Section 1. We also neglect the term involving the flat part of the scalar spectrum together with the induced GW piece, since we expect it to be suppressed with respect to the NNLO term mentioned above. Before moving on with the technical details of the calculation, let us anticipate that, since the variance of $n_{\rm L}$ is proportional to both the tensor and scalar power spectrum, from dimensional analysis we can guess the result in all three cases to be $\langle |n_{\rm L}|^2 \rangle_{r_\sigma}^{1/2} \sim \sqrt{\mathcal{P}_\mathcal{R} \mathcal{P}_h}/(a r_\sigma)^3$, since r_σ is the only dimensionful parameter. Dividing by the entropy density $s \sim g_{\star s}(T) T^3$ and using entropy conservation, we find

$$\frac{n_{\rm L}}{s} \sim \frac{\sqrt{\mathcal{P}_\mathcal{R} \mathcal{P}_h}}{g_{\star s}(T_0)(r_\sigma T_0)^3}, \tag{85}$$

up to an overall numerical factor. The NNLO term will be highly peaked around $r_\sigma \sim k_\sharp^{-1}$, whereas the NLO term due to the flat part of the power spectrum will instead grow simply as r_σ^{-3} and reach its highest value at the Silk scale $r_{\rm S}$, as per the previous discussion.

Case 1: NLO, flat scalar spectrum (NLO♭)

Putting together eqs. (14) and (78), we find, after some manipulation,

$$\begin{aligned}\langle a^6 |n_{\rm L}|^2 \rangle_{r_\sigma} = \left(\frac{1}{16\pi^2}\right)^2 \sum_{st} \int d\eta \int d\hat{\eta} \int \frac{d^3\boldsymbol{p}}{(2\pi)^3} \int \frac{d^3\boldsymbol{q}}{(2\pi)^3} \int \frac{d^3\hat{\boldsymbol{p}}}{(2\pi)^3} \int \frac{d^3\hat{\boldsymbol{q}}}{(2\pi)^3} \int d^3\boldsymbol{r} \int d^3\hat{\boldsymbol{r}} \\ \left(\frac{3+3w}{5+3w}\right)^2 T_\phi(q\eta) T_\phi(\hat{q}\hat{\eta}) T_h'(p\eta) T_h'(\hat{p}\hat{\eta}) \langle \mathcal{R}_{\boldsymbol{q}} \mathcal{R}^\dagger_{\hat{\boldsymbol{q}}} \rangle \langle h^s_{\boldsymbol{p}}(0) h^{t\dagger}_{\hat{\boldsymbol{p}}}(0) \rangle \\ W_{r_\sigma}(r) W_{r_\sigma}(\hat{r}) e^{i(\boldsymbol{p}+\boldsymbol{q})\cdot(\boldsymbol{x}+\boldsymbol{r})} e^{-i(\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})\cdot(\boldsymbol{x}+\hat{\boldsymbol{r}})} \left[\epsilon_{ijk}\epsilon_{abc} q_j q_\ell p_k e^s_{i\ell}(\boldsymbol{p}) \hat{q}_b \hat{q}_d \hat{p}_c e^t_{ad}(\hat{\boldsymbol{p}})\right].\end{aligned} \tag{86}$$

Straightforward evaluation of the correlation functions yields

$$\langle \mathcal{R}_{\boldsymbol{k}} \mathcal{R}^\dagger_{\boldsymbol{p}} \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_\mathcal{R}(k) \delta^3_{\boldsymbol{k}-\boldsymbol{p}} (2\pi)^3, \tag{87}$$

$$\langle h^s_{\boldsymbol{k}}(0) h^{t\dagger}_{\boldsymbol{p}}(0) \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_h(k) \delta^3_{\boldsymbol{k}-\boldsymbol{p}} (2\pi)^3 \delta^{st}, \tag{88}$$

where we have introduced the shorthand $\delta^3(\boldsymbol{k}) \equiv \delta^3_{\boldsymbol{k}}$ and assumed that $h^+_k = h^\times_k = h_k$. Using the Dirac delta functions to perform two of the momentum integrals, we find

$$\begin{aligned}\langle a^6 |n_{\rm L}|^2 \rangle_{r_\sigma} = \frac{1}{64}\left(\frac{3+3w}{5+3w}\right)^2 \int d\eta \int d\hat{\eta} \int \frac{d^3\boldsymbol{p}}{(2\pi)^3} \int \frac{d^3\boldsymbol{q}}{(2\pi)^3} \frac{\mathcal{P}_\mathcal{R}(q)}{q^3} \frac{\mathcal{P}_h(p)}{p^3} \left|\widehat{W}_{r_\sigma}(|\boldsymbol{p}+\boldsymbol{q}|)\right|^2 \\ T_\phi(q\eta) T_\phi(q\hat{\eta}) T_h'(p\eta) T_h'(p\hat{\eta}) \left[\sum_s \epsilon_{ijk}\epsilon_{abc} q_j q_\ell p_k e^s_{i\ell}(\boldsymbol{p}) q_b q_d p_c e^s_{ad}(\boldsymbol{p})\right],\end{aligned} \tag{89}$$

where we have introduced the Fourier transform of the window function,

$$\widehat{W}_{r_\sigma}(k) \equiv \int d^3\boldsymbol{r}\, W_{r_\sigma}(r) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{90}$$

By expanding the Levi-Civita symbols in terms of Kronecker deltas, we find the following expression for the term in brackets

$$\sum_s \epsilon_{ijk}\epsilon_{abc} q_j q_\ell p_k e^s_{i\ell}(\boldsymbol{p}) q_b q_d p_c e^s_{ad}(\boldsymbol{p}) = \Big[q^2p^2 - (\boldsymbol{q}\cdot\boldsymbol{p})^2\Big]\Big[\boldsymbol{q}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{q}\Big] - p^2\Big[\boldsymbol{q}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{q}\Big]^2. \quad (91)$$

If we choose the coordinate system in such a way that the $\boldsymbol{z}$ axis is aligned with $\boldsymbol{p}$, then the vectors in the definition of the polarization tensors in eq. (13) are simply $\boldsymbol{v} = \boldsymbol{x}$ and $\bar{\boldsymbol{v}} = \boldsymbol{y}$. Using θ to denote the angle between $\boldsymbol{q}$ and the $\boldsymbol{z}$ axis, and φ for the corresponding azimuthal angle, we obtain, after some straightforward algebra,

$$\sum_s \epsilon_{ijk}\epsilon_{abc} q_j q_\ell p_k e^s_{i\ell}(\boldsymbol{p}) q_b q_d p_c e^s_{ad}(\boldsymbol{p}) = \frac{1}{2}p^2q^4\sin^4\theta. \quad (92)$$

The window function is obtained by Fourier-transforming eq. (79),

$$\widehat{W}_{r_\sigma}(|\boldsymbol{p}+\boldsymbol{q}|) = \exp\left[-\frac{r_\sigma^2}{4}(p^2+q^2+2pq\cos\theta)\right]. \quad (93)$$

After switching to spherical coordinates, all of the angular integrals can be performed explicitly, and we arrive at the expression

$$\begin{aligned}\left\langle a^6|n_{\rm L}|^2\right\rangle_{r_\sigma} = \frac{1}{64\pi^4}\left(\frac{3+3w}{5+3w}\right)^2 \int dp \int dq\; pq^3\mathcal{P}_{\mathcal{R}}(q)\mathcal{P}_h(p)\left|\int d\eta\; T_\phi(q\eta)T_h'(p\eta)\right|^2 \\ \exp\left[-\frac{r_\sigma^2}{2}(p^2+q^2)\right]\left[-3\frac{\cosh(pqr_\sigma^2)}{(pqr_\sigma^2)^4} + (3+p^2q^2r_\sigma^4)\frac{\sinh(pqr_\sigma^2)}{(pqr_\sigma^2)^5}\right]. \quad (94)\end{aligned}$$

As explained earlier, the observable quantity of interest is the root mean square of the lepton number density. In order to compare with the baryon asymmetry after the electroweak sphaleron processes have taken place, we need to divide by the entropy density

$$s = \frac{2\pi^2}{45}g_{\star s}(T)T^3. \quad (95)$$

The main contribution to the time integrals in eq. (94) occurs during the reheating stage, since the transfer functions for both tensors and scalars decay quickly during the radiation era. Let us suppose that the integrals freeze at some time t_f shortly after reheating ends. Then, all of the time dependence of the lepton number density is in the a^3 factor, and we can relate the value of $n_{\rm L}/s$ to its value today by using entropy conservation,

$$g_{\star s}(T_f)a_f^3T_f^3 = g_{\star s}(T_0)a_0^3T_0^3. \quad (96)$$

Moreover, the upper limit in the time integrals can then be taken as $\eta \to \infty$.

So far, we have not made any explicit choice for the power spectra. Let us set

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{A}_\flat, \qquad \mathcal{P}_h(k) = r\mathcal{A}_\flat. \quad (97)$$

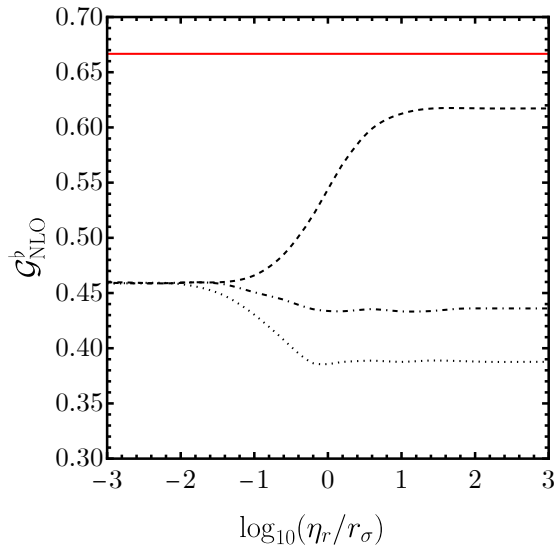

Figure 7: *Numerical calculation of the coefficient $\mathcal{G}^{\flat}_{\rm NLO}$ in eq. (99) as a function of the parameter η_r/r_σ for the different equations of state $w = 0$ (black, dashed), $w = 0.5$ (black, dot-dashed), and $w = 1$ (black, dotted), together with the analytical estimate in eq. (104) (red, solid).*

We then obtain

$$\left.\frac{n_{\rm L}}{s}\right|_{\rm NLO\flat} = \frac{45}{16\pi^4}\left(\frac{3+3w}{5+3w}\right)\frac{\sqrt{r\mathcal{A}_\flat^2\mathcal{G}^{\flat}_{\rm NLO}}}{g_{\star s}(T_0)(T_0 r_\sigma)^3}, \tag{98}$$

with

$$\mathcal{G}^{\flat}_{\rm NLO} = \int_0^\infty dp_\sigma \int_0^\infty dq_\sigma \left| \int_0^\infty dx\; T_\phi\left(\frac{q_\sigma}{p_\sigma}x, q_\sigma\frac{\eta_r}{r_\sigma}\right)\frac{d}{dx}T_h\left(x, p_\sigma\frac{\eta_r}{r_\sigma}\right)\right|^2 \exp\left[-\frac{1}{2}(p_\sigma^2+q_\sigma^2)\right]\left[-3\frac{\cosh(p_\sigma q_\sigma)}{p_\sigma^3 q_\sigma} + (3+p_\sigma^2 q_\sigma^2)\frac{\sinh(p_\sigma q_\sigma)}{p_\sigma^4 q_\sigma^2}\right], \tag{99}$$

where we have introduced the dimensionless variables

$$x = p\eta, \qquad p_\sigma = r_\sigma p, \qquad q_\sigma = r_\sigma q, \tag{100}$$

as well as the following notation for the transfer functions

$$T_\phi(k\eta) \equiv T_\phi(k\eta, k\eta_r), \qquad T_h(k\eta) \equiv T_h(k\eta, k\eta_r), \tag{101}$$

which simply makes explicit the dependence on η_r. There is, of course, also an implicit dependence on the equation of state w during the reheating stage. In fact, $\mathcal{G}^{\flat}_{\rm NLO}$ is a dimensionless numerical factor that depends only on the parameters w and η_r/r_σ, which can be found from eq. (52),

$$\frac{\eta_r}{r_\sigma} \simeq \left(\frac{1}{1+3w}\right)\left(\frac{r_\sigma^{-1}}{10^{14}{\rm Mpc}^{-1}}\right)\left(\frac{10^7{\rm GeV}}{T_r}\right). \tag{102}$$

The dependence on both of these quantities is quite mild, as illustrated by the numerical results presented in Fig. 7.

It is instructive to obtain an analytical result for $\mathcal{G}^{\flat}_{\rm NLO}$ by introducing some approximations, following the procedure in [11]. The first is that since the time integrals freeze around the end of reheating, we can take the upper limit as η_r instead of ∞, which allows us to use the expressions for the transfer functions valid during the reheating stage. Moreover, since the dependence on the equation of state w is quite mild, we can simply set $w = 0$, so that $T_\phi(k\eta) = 1$. The time integral can then be performed immediately, yielding

$$\int_0^{p\eta_r} dx\; T_\phi\left(\frac{q_\sigma}{p_\sigma}x\right)\frac{d}{dx}T_h(x) = T_h\left(p_\sigma\frac{\eta_r}{r_\sigma}\right) - 1, \tag{103}$$

where we have suppressed the second arguments in the transfer functions, since we are using the expressions during the reheating stage, and no matching of coefficients is involved. We can finally make the assumption that $\eta_r \gg r_\sigma$, so that $T_h(p_\sigma\eta_r/r_\sigma) \to 0$. The remaining integrals over p_σ and q_σ can then be performed analytically, and we find

$$\mathcal{G}^{\flat}_{\rm NLO} = \frac{2}{3}. \tag{104}$$

This result is very close to those obtained by numerically calculating the integral in eq. (99), as shown in Fig. 7, even for $\eta_r \lesssim r_\sigma$ and $w \neq 0$.

Our result differs from the one in [11] in two ways. The first is that we obtain a different expression for the integral in eq. (99). The momentum integrals in [11] diverge in both the IR and UV, so the authors have calculated it by imposing cutoffs on both limits. Our integral, in contrast, converges without the need for cutoffs, so we obtain a finite result for all parameter choices. The second difference is that the evolution of the transfer functions during the radiation era was not considered in [11], and instead an analytical estimate similar to the one performed above was presented. We have taken the effect of this evolution into account in our calculation, and computed the integrals in eq. (99) numerically by varying the two parameters η_r/r_σ and w, confirming that the dependence of the result on these quantities is very mild.

We conclude that, apart from the mild dependence of $\mathcal{G}^{\flat}_{\rm NLO}$ on r_σ, the resulting $n_{\rm L}/s$ scales as r_σ^{-3}. As discussed earlier, this result is only valid up to the Silk scale $r_{\rm S}$, where the annihilation damping becomes relevant. The maximum value it reaches is therefore, taking $r = 0.01$ [48, 49] and $\mathcal{G}^{\flat}_{\rm NLO} = 2/3$ for simplicity,

$$\left.\frac{n_{\rm L}}{s}\right|_{\rm NLO\flat} \simeq \frac{45}{16\pi^4}\left(\frac{3+3w}{5+3w}\right)\frac{\sqrt{r\mathcal{A}_\flat^2(2/3)}}{g_{\star s}(T_0)(T_0 r_{\rm S})^3} \simeq 10^{-38}. \tag{105}$$

The asymmetry over the entire Hubble patch today is much smaller,

$$\left.\frac{n_{\rm L}}{s}\right|_{\rm NLO\flat} \simeq \frac{45}{16\pi^4}\left(\frac{3+3w}{5+3w}\right)\frac{\sqrt{r\mathcal{A}_\flat^2(2/3)}}{g_{\star s}(T_0)}\left(\frac{H_0}{T_0}\right)^3 \simeq 10^{-100}, \tag{106}$$

so we confirm, as claimed in [11], that these fluctuations alone cannot explain the observed value

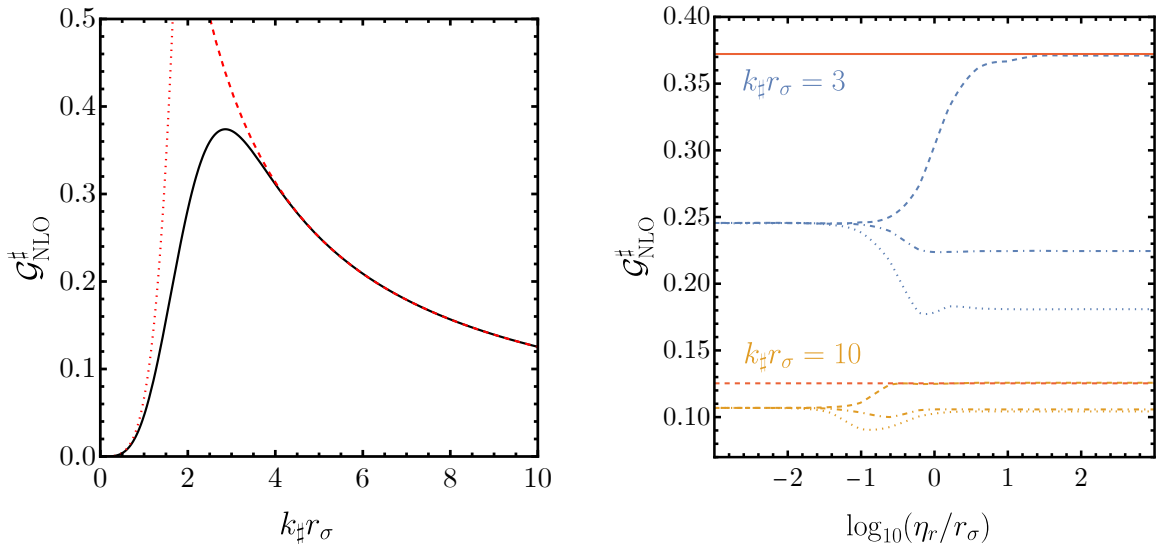

Figure 8: *Left panel: approximate expression for* $\mathcal{G}^\sharp_{\rm NLO}$ *in eq. (110) (solid black), together with the asymptotic limits in (111) (dashed and dotted red). Right panel: numerical calculation of the full integral in eq. (109) for* $k_\sharp r_\sigma = 3$ *(blue) and* $k_\sharp r_\sigma = 10$ *(orange) for* $w = 0$ *(dashed),* $w = 0.5$ *(dot-dashed), and* $w = 1$ *(dotted), together with the approximate expression depicted in the left panel for each case (solid red and dashed red).*

of $n_{\rm L}/s \simeq 10^{-10}$. The same conclusion holds for the two following cases, but, as we will see momentarily, the resulting spectrum of fluctuations has a rich structure that could, in principle, allow us to probe different inflationary models if it were observable on small scales.

Case 2: NLO, sharp scalar spectrum (NLO♯)

We now turn our attention to the cases in which the power spectrum is sharply peaked at some particular scale (not considered in [11]), and compute the corresponding enhancement to the baryon asymmetry fluctuations. The calculation in this case is exactly the same as the previous one up to the choice of the power spectra in eq. (97). We now set

$$\mathcal{P}_\mathcal{R}(k) = \mathcal{A}_\sharp k_\sharp \delta(k - k_\sharp), \qquad \mathcal{P}_h(k) = r\mathcal{A}_\flat, \tag{107}$$

and obtain

$$\left.\frac{n_{\rm L}}{s}\right|_{\rm NLO\sharp} = \frac{45}{16\pi^4}\left(\frac{3+3w}{5+3w}\right)\frac{\sqrt{r\mathcal{A}_\flat \mathcal{A}_\sharp \mathcal{G}^\sharp_{\rm NLO}}}{g_{\star s}(T_0)(T_0 r_\sigma)^3}, \tag{108}$$

with

$$\mathcal{G}^\sharp_{\rm NLO} = \int_0^\infty dp_\sigma \left| \int_0^\infty dx\, T_\phi\left(\frac{k_\sharp r_\sigma}{p_\sigma}x, k_\sharp r_\sigma \frac{\eta_r}{r_\sigma}\right)\frac{d}{dx}T_h\left(x, p_\sigma\frac{\eta_r}{r_\sigma}\right)\right|^2 \\ \exp\left[-\frac{1}{2}(p_\sigma^2 + k_\sharp^2 r_\sigma^2)\right]\left[-3\frac{\cosh(p_\sigma k_\sharp r_\sigma)}{p_\sigma^3} + (3 + p_\sigma^2 k_\sharp^2 r_\sigma^2)\frac{\sinh(p_\sigma k_\sharp r_\sigma)}{p_\sigma^4 k_\sharp r_\sigma}\right]. \tag{109}$$

Now the dimensionless factor $\mathcal{G}^\sharp_{\rm NLO}$ depends not only on w and η_r/r_σ, but also on the parameter

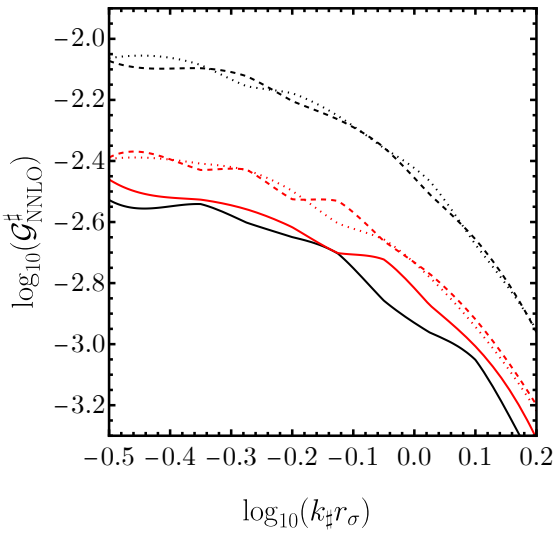

Figure 9: *Numerical calculation of the function* $\mathcal{G}^{\sharp}_{\rm NNLO}$ *for* $w = 1$ *(black) and* $w = 0.5$ *(red) for* $k_\sharp \eta_r = 1$, 10, *and* 100 *(solid, dashed, and dotted lines, respectively).*

$k_\sharp r_\sigma$. It is once again instructive to obtain an analytical result using the approximations of the previous case. Setting $w = 0$ and following the same procedure, we obtain

$$\mathcal{G}^{\sharp}_{\rm NLO} = \frac{1}{k_\sharp r_\sigma}\sqrt{\frac{\pi}{2}}\,\mathrm{erf}\left(\frac{k_\sharp r_\sigma}{\sqrt{2}}\right) - e^{-k_\sharp^2 r_\sigma^2/2}\left(1 + \frac{1}{3}k_\sharp^2 r_\sigma^2\right). \tag{110}$$

This function is highly peaked around $k_\sharp r_\sigma \simeq 3$, for which we find $\mathcal{G}^{\sharp}_{\rm NLO} \simeq 0.38$. This quantity has the asymptotic limits

$$\mathcal{G}^{\sharp}_{\rm NLO} \simeq \begin{cases} \dfrac{1}{15}(k_\sharp r_\sigma)^4, & \text{for } k_\sharp r_\sigma \ll 1, \\ \dfrac{1}{k_\sharp r_\sigma}\sqrt{\dfrac{\pi}{2}}, & \text{for } k_\sharp r_\sigma \gg 1. \end{cases} \tag{111}$$

The function and its asymptotic limits are shown in the left panel of Fig. 8. The numerical results for the full integral in eq. (109) are shown in the right panel of the same Figure for $k_\sharp r_\sigma = 3$ and $k_\sharp r_\sigma = 10$ as a function of η_r/r_σ and w. We once again find that the dependence of the result on these two parameters is very mild, so only the scaling with $k_\sharp r_\sigma$ is relevant for our purposes. The dependence of the result on r_σ is shown in Fig. 10 for values of $k_\sharp$ of interest for PBH dark matter, together with the NLO♭ and NNLO♯ contributions, see the discussion below.

Case 3: NNLO, sharp scalar spectrum (NNLO♯)

We now turn our attention to the NNLO contribution to the baryon asymmetry fluctuations. The amplitude of this term is entirely determined by the initial conditions and evolution of the scalar perturbations, and is therefore present even if the tensor-to-scalar ratio r is vanishingly small. The calculation in this case is more involved than the previous ones, so we perform it in Appendix A.

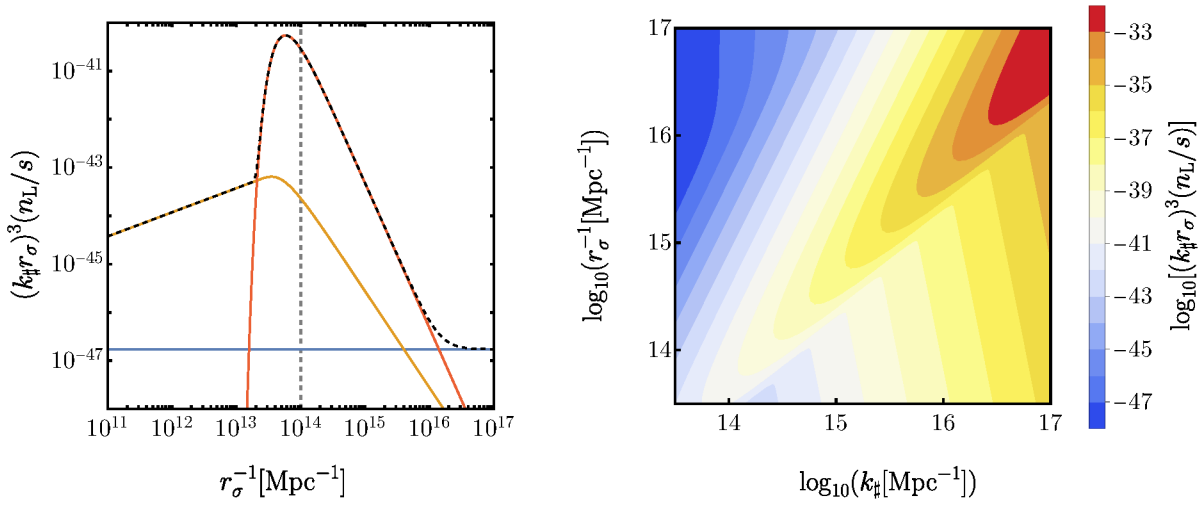

Figure 10: *Left panel: sum of the* NLO♭ *(solid, blue),* NLO♯ *(solid, orange), and* NNLO♯ *(solid, red) contributions to the baryon asymmetry fluctuations, represented by the dashed black line and scaled by the factor* $(k_\sharp r_\sigma)^3$*. The plot is cut at the Silk scale* r_S*, where the calculation stops being valid. We have taken* $k_\sharp = 10^{14}\text{Mpc}^{-1}$ *as a representative value for PBH dark matter (see Fig. 1), together with* $\mathcal{A}_\sharp = 5 \times 10^{-2}$*,* $r = 0.01$*,* $\mathcal{A}_\flat = 2 \times 10^{-9}$ *and* $w = 1$*. Right panel: sum of the three contributions to the baryon asymmetry for the same parameters as in the left panel, but varying* $k_\sharp$*.*

The resulting expression is

$$\left.\frac{n_L}{s}\right|_{\text{NNLO}\sharp} = \frac{45}{16\pi^4}\left(\frac{3+3w}{5+3w}\right)\frac{\sqrt{\mathcal{A}_\sharp^3 \mathcal{G}^\sharp_{\text{NNLO}}}}{g_{\star s}(T_0)}\left(\frac{k_\sharp}{T_0}\right)^3, \tag{112}$$

where $\mathcal{G}^\sharp_{\text{NNLO}}$ is computed in eq. (121). The coefficient $\mathcal{G}^\sharp_{\text{NNLO}}$ is, in this case, a function of w and $k_\sharp \eta_r$, as well as $k_\sharp r_\sigma$. The numerical calculation of this quantity is shown in Fig. 9 for different values of the parameters. We once again find that only the dependence on $k_\sharp r_\sigma$ is relevant, and for $w = 1$ and $k_\sharp \eta_r = 1$ the function can be approximated by $\mathcal{G}^\sharp_{\text{NNLO}} \simeq 10^{-2.8} \exp\left(-k_\sharp^2 r_\sigma^2\right)$. We adopt this expression for the following discussion.

The sum of the contributions in all three cases is shown on the left panel of Fig. 10 for the representative value $k_\sharp = 10^{14}\text{Mpc}^{-1}$, and on the right panel of the same Figure as a function of $k_\sharp$. We have rescaled n_L/s by the factor $(k_\sharp r_\sigma)^3$ to make the peaked structure of the spectrum more clear. This is the main result of this Section. We find that the NLO♯ contribution dominates for $r_\sigma^{-1} \ll k_\sharp$, and the NLO♭ contribution becomes relevant only for $r_\sigma^{-1} \gg k_\sharp$. For $r_\sigma^{-1} \simeq k_\sharp$, the distribution is heavily peaked due to the NNLO♯ term, which yields an enhancement over the NLO♭ result of $\mathcal{O}(10^6)$. As the right panel shows, the amplitude of the fluctuations at the peak of the distribution increases with $k_\sharp$. We note, however, that for $k_\sharp \simeq r_S^{-1}$, where the enhancement is largest, the PBH masses obtained are far too small, as can be checked from eq. (28), yielding a distribution of black holes that would have evaporated today [50] (though having a peak in the scalar spectrum at this scale is still possible, provided its amplitude is small enough so that PBHs are not overproduced). It is also clear from the Figure that the NLO♯ and NNLO♯ contributions

do not change the distribution significantly on large scales, so the estimate in eq. (106) of the asymmetry on a Hubble-sized patch today holds in the presence of the new terms.

Conclusions

We have calculated the expressions for the mass and abundance of PBHs produced during an early epoch of reheating with a stiff equation of state. We find that the parameters that determine the black hole distribution are the scale at which the peak in the power spectrum is located $k_\sharp$, the temperature at which the transition to radiation occurs T_r, and the equation of state w. There are three relevant constraints in this scenario. The first is that, in order to reproduce the observed dark matter abundance, the black hole masses must be in the range (2). The second is that for collapse to occur before the radiation era, the scale at which the peak in the spectrum is located must re-enter the horizon before the transition occurs, see eq. (32). Finally, since tensor modes are enhanced in the presence of a stiff epoch, we must also ask that GWs are not overproduced so that the bound on their energy density today, which arises from BBN and CMB observations, is not violated, see eq. (54). The allowed region of parameter space is shown in Fig. 1.

In order to calculate the induced GW signal we have implemented a matching procedure for both the Green's function of the tensor modes and the transfer function of the scalar perturbations, thereby extending the results of [27] by taking into account the full time evolution of these quantities. We confirm that, in the presence of a stiff epoch, the induced GW signal is enhanced. We have explicitly checked that the smoothness of the transition does not significantly change our results, and provided a procedure to compute the signal in these gradual decay scenarios exactly, without resorting to approximate analytical expressions of the Green's and transfer functions, extending the results of [25] to decaying fluids with stiff equations of state. As mentioned above, we have used these results to translate the bound on the GW abundance to a constraint on the PBH masses formed in this scenario. The results are shown in Figs. 1 and 3.

Finally, we have computed the chiral gravitational anomaly to third order in perturbations, and we find that the large scalar spectrum responsible for PBH formation induces a peak in the baryon asymmetry fluctuations on small scales. These results are shown in Fig. 10. We have shown that this spectrum is essentially independent of the cosmological history (i.e. the reheating scale and equation of state) and is therefore a generic prediction present in every model of PBH formation from collapse induced by large density perturbations, even in the standard scenario where the black holes form during radiation domination, assuming only the matter content of the Standard Model. These fluctuations could, in principle, be used as an observable to probe not only the existence of PBHs, but also different models of inflation. Assessing whether these fluctuations can be measured by any future experiments is beyond the scope of this paper, but we point out that the main obstacle to this end is the fact that they are much smaller than the observed background value, and therefore some sort of enhancement mechanism would likely be required in the particular scenario studied here. We remark, however, that the machinery we have developed could also be applied to other models of gravitational leptogenesis, such as the one in [7], in which the fluctuations might be less suppressed.

Acknowledgments

We are grateful to Antonio Maroto and Alfredo Miravet for discussions about their work and the mechanism studied here, to Mehmet Gümüş for an extended discussion about the diagrams contributing to the six-point function entering the NNLO calculation, to Thomas Konstandin and Guillermo Ballesteros for comments on the draft, and to Alexander Westphal, Hyungjin Kim, and Mathias Pierre for discussions in the early stages of this project. This work is supported by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy – EXC 2121 Quantum Universe – 390833306.

Appendices

A. Evaluation of $\mathcal{G}^{\sharp}_{\rm NNLO}$

In this Appendix we compute the quantity $\mathcal{G}^{\sharp}_{\rm NNLO}$ relevant for the calculation of the lepton number density variance in the third case of Section 5. By putting together eqs. (23) and (78) we obtain

$$\begin{aligned}\left\langle a^6|n_{\rm L}|^2\right\rangle_{r_\sigma} = &\left(\frac{1}{32\pi^2}\right)^2 \sum_{st}\int d\eta\int d\hat\eta\int\frac{d^3\boldsymbol{p}}{(2\pi)^3}\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\int\frac{d^3\hat{\boldsymbol{p}}}{(2\pi)^3}\int\frac{d^3\hat{\boldsymbol{q}}}{(2\pi)^3}\int\frac{d^3\hat{\boldsymbol{k}}}{(2\pi)^3}\\ &\left(\frac{3+3w}{5+3w}\right)^2 T_\phi(q\eta)T_\phi(\hat q\hat\eta)I'_p(\eta,k,|\boldsymbol{p}-\boldsymbol{k}|)I'_{\hat p}(\hat\eta,\hat k,|\hat{\boldsymbol{p}}-\hat{\boldsymbol{k}}|)\left\langle\mathcal{R}_{\boldsymbol{q}}\mathcal{R}_{\boldsymbol{k}}\mathcal{R}_{\boldsymbol{p}-\boldsymbol{k}}\mathcal{R}^\dagger_{\hat{\boldsymbol{q}}}\mathcal{R}^\dagger_{\hat{\boldsymbol{k}}}\mathcal{R}^\dagger_{\hat{\boldsymbol{p}}-\hat{\boldsymbol{k}}}\right\rangle\\ &\int d^3\boldsymbol{r}\int d^3\hat{\boldsymbol{r}}\; W_{r_\sigma}(r)W_{r_\sigma}(\hat r)e^{i(\boldsymbol{p}+\boldsymbol{q})\cdot(\boldsymbol{x}+\boldsymbol{r})}e^{-i(\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})\cdot(\boldsymbol{x}+\hat{\boldsymbol{r}})}\\ &\frac{1}{p^2\hat p^2}\Big[\boldsymbol{k}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{k}\Big]\Big[\hat{\boldsymbol{k}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{k}}\Big]\Big[\epsilon_{ijk}\epsilon_{abc}q_jq_\ell p_k e^s_{i\ell}(\boldsymbol{p})\hat q_b\hat q_d\hat p_c e^t_{ad}(\hat{\boldsymbol{p}})\Big].\end{aligned} \tag{113}$$

We can evaluate the six-point function by using Wick's theorem

$$\begin{aligned}\left[2(2\pi)^9\right]^{-1}\left\langle\mathcal{R}_{\boldsymbol{q}}\mathcal{R}_{\boldsymbol{k}}\mathcal{R}_{\boldsymbol{p}-\boldsymbol{k}}\mathcal{R}^\dagger_{\hat{\boldsymbol{q}}}\mathcal{R}^\dagger_{\hat{\boldsymbol{k}}}\mathcal{R}^\dagger_{\hat{\boldsymbol{p}}-\hat{\boldsymbol{k}}}\right\rangle = &|\mathcal{R}_q|^2|\mathcal{R}_{|\boldsymbol{p}-\boldsymbol{k}|}|^2|\mathcal{R}_{\hat q}|^2\delta^3_{\boldsymbol{q}+\boldsymbol{k}}\delta^3_{\boldsymbol{p}-\boldsymbol{k}-\hat{\boldsymbol{p}}+\hat{\boldsymbol{k}}}\delta^3_{\hat{\boldsymbol{q}}+\hat{\boldsymbol{k}}}\\ &+|\mathcal{R}_q|^2|\mathcal{R}_k|^2|\mathcal{R}_{\hat q}|^2\delta^3_{\boldsymbol{q}+\boldsymbol{p}-\boldsymbol{k}}\delta^3_{\boldsymbol{k}-\hat{\boldsymbol{k}}}\delta^3_{\hat{\boldsymbol{q}}+\hat{\boldsymbol{p}}-\hat{\boldsymbol{k}}}\\ &+|\mathcal{R}_q|^2|\mathcal{R}_k|^2|\mathcal{R}_{\hat q}|^2\delta^3_{\boldsymbol{q}-\hat{\boldsymbol{k}}}\delta^3_{\boldsymbol{k}-\hat{\boldsymbol{p}}+\hat{\boldsymbol{k}}}\delta^3_{\hat{\boldsymbol{q}}-\boldsymbol{p}+\boldsymbol{k}}\\ &+|\mathcal{R}_q|^2|\mathcal{R}_{|\boldsymbol{p}-\boldsymbol{k}|}|^2|\mathcal{R}_{\hat q}|^2\delta^3_{\boldsymbol{q}-\hat{\boldsymbol{p}}+\hat{\boldsymbol{k}}}\delta^3_{\boldsymbol{k}-\hat{\boldsymbol{q}}}\delta^3_{\boldsymbol{p}-\boldsymbol{k}-\hat{\boldsymbol{k}}},\end{aligned} \tag{114}$$

where we have used the symmetry of the integrand under $\boldsymbol{k}\to\boldsymbol{p}-\boldsymbol{k}$ and $\hat{\boldsymbol{k}}\to\hat{\boldsymbol{p}}-\hat{\boldsymbol{k}}$ to simplify the result.[17] We can use the Dirac delta functions to perform the momentum integrals over $\boldsymbol{k}$, $\hat{\boldsymbol{k}}$ and $\hat{\boldsymbol{q}}$. After using the symmetry of I_k under the exchange of the two momenta in the argument,

[17] We have also ignored unphysical contact terms containing factors of the form $\delta^3_{\boldsymbol{k}}$, and one term with $\delta^3_{\boldsymbol{q}-\hat{\boldsymbol{q}}}\delta^3_{\boldsymbol{p}-\hat{\boldsymbol{p}}}$, which is not proportional to an overall momentum-conserving delta function $\delta^3_{\boldsymbol{p}+\boldsymbol{q}-\hat{\boldsymbol{p}}-\hat{\boldsymbol{q}}}$.

we obtain

$$\begin{aligned}\left\langle a^6|n_{\rm L}|^2\right\rangle_{r_\sigma} = \frac{\pi^2}{32}\left(\frac{3+3w}{5+3w}\right)^2 \sum_{st}\int d\eta\int d\hat{\eta}\int\frac{d^3\boldsymbol{p}}{(2\pi)^3}\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\,\frac{1}{p^2k^2}\left|\widehat{W}_{r_\sigma}(|\boldsymbol{p}+\boldsymbol{q}|)\right|^2 \\ T_\phi(q\eta)T_\phi(|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|\hat{\eta})\Big[\epsilon_{ijk}q_jq_\ell p_k e^s_{i\ell}(\boldsymbol{p})\Big]\Big[\epsilon_{abc}(p_b+q_b-k_b)(p_d+q_d)k_c e^t_{ad}(\boldsymbol{k})\Big] \\ \frac{\mathcal{P}_{\mathcal{R}}(q)}{q^3}\frac{\mathcal{P}_{\mathcal{R}}(|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|)}{|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|^3}\Bigg\{\frac{\mathcal{P}_{\mathcal{R}}(|\boldsymbol{p}+\boldsymbol{q}|)}{|\boldsymbol{p}+\boldsymbol{q}|^3}\Big[\boldsymbol{q}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{q}\Big]\Big[(\boldsymbol{p}+\boldsymbol{q})\cdot\boldsymbol{e}^t(\boldsymbol{k})\cdot(\boldsymbol{p}+\boldsymbol{q})\Big] \\ I'_p(\eta,q,|\boldsymbol{p}+\boldsymbol{q}|)I'_k(\hat{\eta},|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|,|\boldsymbol{p}+\boldsymbol{q}|) \\ +\frac{\mathcal{P}_{\mathcal{R}}(|\boldsymbol{q}-\boldsymbol{k}|)}{|\boldsymbol{q}-\boldsymbol{k}|^3}\Big[(\boldsymbol{q}-\boldsymbol{k})\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot(\boldsymbol{q}-\boldsymbol{k})\Big]\Big[\boldsymbol{q}\cdot\boldsymbol{e}^t(\boldsymbol{k})\cdot\boldsymbol{q}\Big] \\ I'_p(\eta,|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|,|\boldsymbol{q}-\boldsymbol{k}|)I'_k(\hat{\eta},q,|\boldsymbol{q}-\boldsymbol{k}|)\Bigg\},\end{aligned} \tag{115}$$

where we have also renamed the remaining dummy variable $\hat{\boldsymbol{p}}\to\boldsymbol{k}$. By choosing the sharp power spectrum in eq. (30) and following the same procedure as for the two NLO cases in Section 5, we find eq. (112), with

$$\begin{aligned}\mathcal{G}^{\sharp}_{\rm NNLO} \equiv \frac{2\pi^6}{k_\sharp^3}\int\frac{d^3\boldsymbol{p}}{(2\pi)^3}\int\frac{d^3\boldsymbol{q}}{(2\pi)^3}\int\frac{d^3\boldsymbol{k}}{(2\pi)^3}\,\frac{\delta(q-k_\sharp)}{k_\sharp^3}\,\frac{\delta(|\boldsymbol{p}+\boldsymbol{q}|-k_\sharp)}{k_\sharp^3}\,\frac{\delta(|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|-k_\sharp)}{k_\sharp^3}\,\frac{1}{p^2k^2} \\ \int d\eta\, T_\phi(k_\sharp\eta)I'_p(\eta,k_\sharp,k_\sharp)\int d\hat{\eta}\, T_\phi(k_\sharp\hat{\eta})I'_k(\hat{\eta},k_\sharp,k_\sharp) \\ \sum_{st}\Big[\boldsymbol{q}\cdot\boldsymbol{e}^s(\boldsymbol{p})\cdot\boldsymbol{q}\Big]\Big[(\boldsymbol{p}+\boldsymbol{q})\cdot\boldsymbol{e}^t(\boldsymbol{k})\cdot(\boldsymbol{p}+\boldsymbol{q})\Big] \\ \Big[F_{st}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k})\left|\widehat{W}_{r_\sigma}(|\boldsymbol{p}+\boldsymbol{q}|)\right|^2+F_{ts}(-\boldsymbol{k},\boldsymbol{q},-\boldsymbol{p})\left|\widehat{W}_{r_\sigma}(|\boldsymbol{q}-\boldsymbol{k}|)\right|^2\Big],\end{aligned} \tag{116}$$

where

$$F_{st}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{k}) \equiv \Big[\epsilon_{ijk}q_jq_\ell p_k e^s_{i\ell}(\boldsymbol{p})\Big]\Big[\epsilon_{abc}(p_b+q_b-k_b)(p_d+q_d)k_c e^t_{ad}(\boldsymbol{k})\Big]. \tag{117}$$

To obtain this expression from eq. (115) we have renamed the dummy variables $\boldsymbol{k}\to-\boldsymbol{p}$ and $\boldsymbol{p}\to-\boldsymbol{k}$, as well as $s\leftrightarrow t$, in the second term inside the brackets.

Let us orient our coordinate system in such a way that $\boldsymbol{k}$ coincides with the $\boldsymbol{z}$ axis, and denote the angle between $\boldsymbol{p}$ and $\boldsymbol{q}$ by Θ, and the angle between $\boldsymbol{p}+\boldsymbol{q}$ and $\boldsymbol{k}$ by Φ. We can then use the following property of the Dirac delta function,

$$\delta[f(x)] = \sum_j\frac{\delta(x-x_j)}{|f'(x_j)|}, \tag{118}$$

where x_j are the roots of $f(x)$, to find

$$\delta(|\boldsymbol{p}+\boldsymbol{q}|-k_\sharp) = \frac{1}{|\cos\Theta|}\delta(p+2k_\sharp\cos\Theta), \tag{119}$$

$$\delta(|\boldsymbol{p}+\boldsymbol{q}-\boldsymbol{k}|-k_\sharp) = \frac{1}{|\cos\Phi|}\delta(k-2k_\sharp\cos\Phi). \tag{120}$$

We now switch to spherical coordinates in all of the momentum integrals in eq. (116). We denote the polar and azimuthal angles of $\boldsymbol{p}$ by θ_p and ϕ_p, respectively, and similarly for $\boldsymbol{q}$ and $\boldsymbol{k}$. Since $\boldsymbol{k}$

coincides with the $\boldsymbol{z}$ axis, we can immediately perform the integrals over θ_k and ϕ_k. We can also perform the integrals over the moduli by using the Dirac delta functions and including a Heaviside function for each to take into account the fact that the moduli must be positive. We find

$$\begin{aligned}\mathcal{G}^{\sharp}_{\rm NNLO} = \frac{1}{16\pi^2}\int d\theta_p \int d\phi_p \int d\theta_q \int d\phi_q \sin\theta_p \sin\theta_q \Theta_{\rm H}(-\cos\Theta)\Theta_{\rm H}(\cos\Phi) \\ \int dx\; T_\phi(x)\frac{d}{dx} I_{-2k_\sharp\cos\Theta}\left(\frac{x}{k_\sharp}, k_\sharp, k_\sharp\right)\int d\hat{x}\; T_\phi(\hat{x})\frac{d}{d\hat{x}} I_{2k_\sharp\cos\Phi}\left(\frac{\hat{x}}{k_\sharp}, k_\sharp, k_\sharp\right) \\ \sum_{st}\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{q}}\Big]\Big[(-2\cos\Theta\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(-2\cos\Theta\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})\Big] \\ \Big[\widehat{F}_{st}(\hat{\boldsymbol{p}},\hat{\boldsymbol{q}},\hat{\boldsymbol{k}})\big|\widehat{W}_{k_\sharp r_\sigma}(|\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}|)\big|^2 + \widehat{F}_{ts}(-\hat{\boldsymbol{k}},\hat{\boldsymbol{q}},-\hat{\boldsymbol{p}})\big|\widehat{W}_{k_\sharp r_\sigma}(|\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}}|)\big|^2\Big], \end{aligned} \tag{121}$$

where $x \equiv k_\sharp\eta$, the hatted vectors are normalized to unity (and bear no relation to the hatted dummy variables in eq. (113)), and

$$\begin{aligned}\widehat{F}_{st}(\hat{\boldsymbol{p}},\hat{\boldsymbol{q}},\hat{\boldsymbol{k}}) &\equiv \Big[\epsilon_{ijk}\hat{q}_j\hat{q}_\ell\hat{p}_k e^s_{i\ell}(\hat{\boldsymbol{p}})\Big]\Big[\epsilon_{abc}\Big(\hat{q}_b - 2\cos\Theta\hat{p}_b - 2\cos\Phi\hat{k}_b\Big)\Big(\hat{q}_d - 2\cos\Theta\hat{p}_d\Big)\hat{k}_c e^t_{ad}(\hat{\boldsymbol{k}})\Big], \\ \widehat{F}_{ts}(-\hat{\boldsymbol{k}},\hat{\boldsymbol{q}},-\hat{\boldsymbol{p}}) &\equiv \Big[\epsilon_{ijk}\hat{q}_j\hat{q}_\ell\hat{k}_k e^t_{i\ell}(\hat{\boldsymbol{k}})\Big]\Big[\epsilon_{abc}\Big(\hat{q}_b - 2\cos\Theta\hat{p}_b - 2\cos\Phi\hat{k}_b\Big)\Big(\hat{q}_d - 2\cos\Phi\hat{k}_d\Big)\hat{p}_c e^s_{ad}(\hat{\boldsymbol{p}})\Big].\end{aligned} \tag{122}$$

The window functions can be explicitly written as

$$\begin{aligned}\big|\widehat{W}_{k_\sharp r_\sigma}(|\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}|)\big|^2 &= e^{-k_\sharp^2 r_\sigma^2/2}, \\ \big|\widehat{W}_{k_\sharp r_\sigma}(|\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}}|)\big|^2 &= e^{-k_\sharp^2 r_\sigma^2/2}\exp\Big[2k_\sharp^2 r_\sigma^2(\cos\theta_q - \cos\Phi)\cos\Phi\Big].\end{aligned} \tag{123}$$

We can also write the Θ and Φ angles explicitly as follows,

$$\cos\Theta = \sin\theta_p\cos\phi_p\sin\theta_q\cos\phi_q + \sin\theta_p\sin\phi_p\sin\theta_q\sin\phi_q + \cos\theta_p\cos\theta_q, \tag{124}$$

$$\cos\Phi = \cos\theta_q - 2\cos\Theta\cos\theta_p. \tag{125}$$

We can write the expressions in eq. (122) explicitly by expanding the Levi-Civita symbols. After some straightforward algebra, we find

$$\begin{aligned}\widehat{F}_{st}(\hat{\boldsymbol{p}},\hat{\boldsymbol{q}},\hat{\boldsymbol{k}}) = &-\Big[\hat{\boldsymbol{q}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{k}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{p}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big] \\ &+\Big[\hat{\boldsymbol{p}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{k}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big] \\ &-\Big[\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{k}}\Big]\Big[(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big] \\ &+\Big[\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}\Big]\Big[(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{p}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big] \\ &-\Big[\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}\Big]\Big[\hat{\boldsymbol{p}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big] \\ &+\Big[\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{k}}\Big]\Big[\hat{\boldsymbol{q}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\Big],\end{aligned} \tag{126}$$

and, similarly,

$$\begin{aligned}\widehat{F}_{ts}(-\hat{\boldsymbol{k}},\hat{\boldsymbol{q}},-\hat{\boldsymbol{p}}) = &-\Big[\hat{\boldsymbol{q}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{p}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{k}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\\
&+\Big[\hat{\boldsymbol{k}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{p}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\\
&-\Big[\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{k}}\Big]\Big[(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\\
&+\Big[\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{p}}\Big]\Big[(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}})\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\hat{\boldsymbol{q}}\Big]\Big[\hat{\boldsymbol{k}}\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\\
&-\Big[\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{p}}\Big]\Big[\hat{\boldsymbol{k}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\\
&+\Big[\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{k}}\Big]\Big[\hat{\boldsymbol{q}}\cdot(\hat{\boldsymbol{q}}-2\cos\Theta\hat{\boldsymbol{p}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]\Big[\hat{\boldsymbol{q}}\cdot\boldsymbol{e}^t(\hat{\boldsymbol{k}})\cdot\boldsymbol{e}^s(\hat{\boldsymbol{p}})\cdot(\hat{\boldsymbol{q}}-2\cos\Phi\hat{\boldsymbol{k}})\Big]. \quad (127)\end{aligned}$$

The $\boldsymbol{e}^s$ matrices are in general given by

$$e^t_{ij}(\hat{\boldsymbol{k}}) = \frac{1}{\sqrt{2}}\delta^{t+}(e_ie_j-\bar{e}_i\bar{e}_j)+\frac{1}{\sqrt{2}}\delta^{t\times}(e_i\bar{e}_j+\bar{e}_ie_j), \quad (128)$$

where e_i and $\bar{e}_i$ are orthonormal vectors orthogonal to $\hat{\boldsymbol{k}}$. Since $\boldsymbol{k}$ coincides with the $\boldsymbol{z}$ axis, we can choose $e_i = x_i$ and $\bar{e}_i = y_i$ for $\boldsymbol{e}^t(\hat{\boldsymbol{k}})$. For $\boldsymbol{e}^s(\hat{\boldsymbol{p}})$ we can instead choose $e_i = v_i$ and $\bar{e}_i = w_i$, where

$$v_i = \cos\theta_p\cos\phi_p x_i + \cos\theta_p\sin\phi_p y_i - \sin\theta_p z_i, \quad (129)$$

$$w_i = -\sin\phi_p x_i + \cos\phi_p y_i. \quad (130)$$

These vectors can be easily checked to be orthogonal to p by using

$$p_i = p\sin\theta_p\cos\phi_p x_i + p\sin\theta_p\sin\phi_p y_i + p\cos\theta_p z_i. \quad (131)$$

With these matrices in hand, the dot products in eqs. (126, 127) can be written in terms of θ_p, θ_q, ϕ_p, and ϕ_q.

Let us turn our attention to the time integrals in eq. (121),

$$\mathcal{I} \equiv \int dx\, T_\phi(x)\frac{d}{dx}I_q\left(\frac{x}{k_\sharp},k_\sharp,k_\sharp\right). \quad (132)$$

The I_q function is given by eq. (19), so that

$$\frac{d}{dx}I_q\left(\frac{x}{k_\sharp},k_\sharp,k_\sharp\right) = \left(\frac{3+3w}{5+3w}\right)^2\int_0^{x(q/k_\sharp)}\frac{d}{dx}qG_q\left(\frac{x}{k_\sharp},\frac{y}{q}\right)Q\left(\frac{y}{q},k_\sharp,k_\sharp\right)dy, \quad (133)$$

with $y = q\eta'$, together with

$$Q\left(\frac{y}{q},k_\sharp,k_\sharp\right) = 8T_\phi\left(\frac{k_\sharp}{q}y,k_\sharp\eta_r\right)^2+\frac{16}{3(1+p/\rho)}\left[T_\phi\left(\frac{k_\sharp}{q}y,k_\sharp\eta_r\right)+\frac{q}{\mathcal{H}}\frac{d}{dy}T_\phi\left(\frac{k_\sharp}{q}y,k_\sharp\eta_r\right)\right]^2, \quad (134)$$

and

$$qG_q\left(\frac{x}{k_\sharp},\frac{y}{q}\right) = h_1\left(\frac{q}{k_\sharp}x,\frac{q}{k_\sharp}k_\sharp\eta_r\right)G_2\left(y,\frac{q}{k_\sharp}k_\sharp\eta_r\right)-h_2\left(\frac{q}{k_\sharp}x,\frac{q}{k_\sharp}k_\sharp\eta_r\right)G_1\left(y,\frac{q}{k_\sharp}k_\sharp\eta_r\right). \quad (135)$$

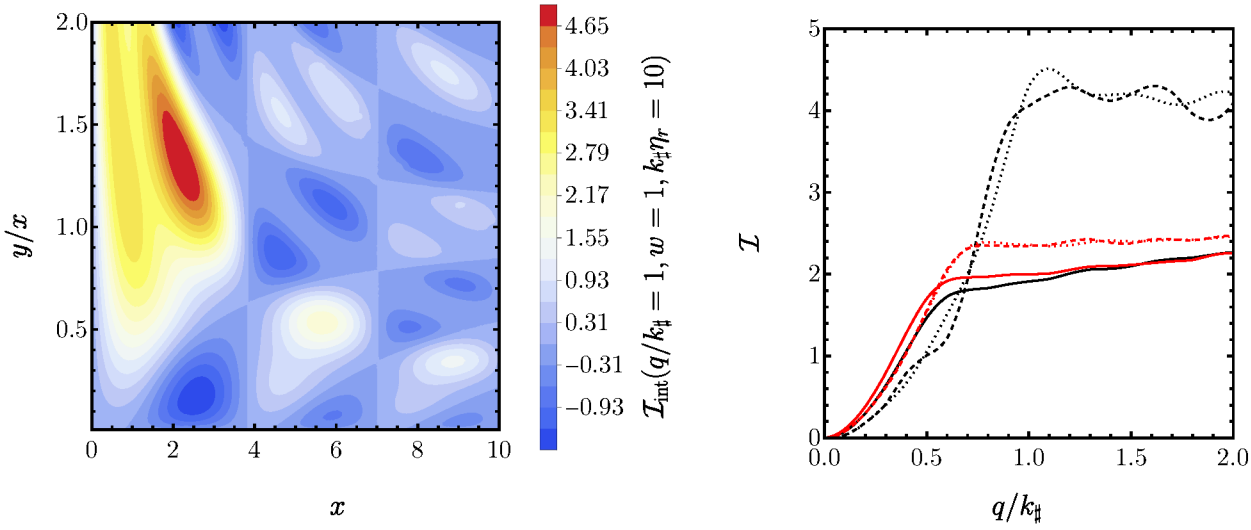

Figure 11: *Left panel: integrand of $\mathcal{I}$ ($\mathcal{I}_{\rm int}$) for the specific parameters shown in the label. The integral converges quickly for $x \gtrsim 4$. Right panel: integral $\mathcal{I}$ as a function of $q/k_\sharp$ for $w = 1$ (black) and $w = 0.5$ (red) for $k_\sharp\eta_r = 1$, 10, and 100 (solid, dashed, and dotted lines, respectively).*

In these equations we have highlighted the dependence of the transfer function T_ϕ on the parameter $k_\sharp\eta_r$. Similarly, the factor $q/\mathcal{H}$ in eq. (134) is a function of y and $q\eta_r = (q/k_\sharp)k_\sharp\eta_r$. For the Green's function, we think of the independent solutions $h_{1,2}$ and the functions $G_{1,2}$ defined in eqs. (76, 77) as functions of $(q\eta, q\eta_r)$. We therefore find that I_q is a function of $(x, q/k_\sharp, k_\sharp\eta_r, w)$.

Computing the 8-dimensional integral in eq. (121) numerically is difficult, even with Monte Carlo methods. The strategy we adopt is to first calculate the integral $\mathcal{I}$ for different choices of the parameters w and $k_\sharp\eta_r$ as a function of $q/k_\sharp$. The results are shown on the right panel of Fig. 11. Since $q/k_\sharp$ lies between 0 and 2 in eq. (121), we restrict our calculation to this range. The integrand of $\mathcal{I}$ is shown on the left panel of the same Figure. With these integrals in hand, we can then calculate the remaining four integrals over the angles in eq. (121) numerically by varying the remaining parameter, $k_\sharp r_\sigma$. The results are shown in Fig. 9.

References

[1] N. Aghanim et al. **Planck 2018 results. VI. Cosmological parameters**. *Astron. Astrophys.* 641 (2020). arXiv: 1807.06209 [astro-ph.CO].

[2] V. Mossa et al. **The baryon density of the Universe from an improved rate of deuterium burning**. *Nature* 587.7833 (2020).

[3] Sacha Davidson, Enrico Nardi, and Yosef Nir. **Leptogenesis**. *Phys. Rept.* 466 (2008). arXiv: 0802.2962 [hep-ph].

[4] M. Fukugita and T. Yanagida. **Baryogenesis Without Grand Unification**. *Phys. Lett. B* 174 (1986).

[5] T. Kimura. **Divergence of axial-vector current in the gravitational field**. *Prog. Theor. Phys.* 42 (1969).

[6] Luis Alvarez-Gaume and Edward Witten. **Gravitational Anomalies**. *Nucl. Phys. B* 234 (1984). Ed. by A. Salam and E. Sezgin.

[7] Stephon Haigh-Solom Alexander, Michael E. Peskin, and Mohammad M. Sheikh-Jabbari. **Leptogenesis from gravity waves in models of inflation**. *Phys. Rev. Lett.* 96 (2006). arXiv: hep-th/0403069.

[8] Peter Adshead, Andrew J. Long, and Evangelos I. Sfakianakis. **Gravitational Leptogenesis, Reheating, and Models of Neutrino Mass**. *Phys. Rev. D* 97.4 (2018). arXiv: 1711.04800 [hep-ph].

[9] A. D. Sakharov. **Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe**. *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967).

[10] Kohei Kamada, Jun'ya Kume, Yusuke Yamada, and Jun'ichi Yokoyama. **Gravitational leptogenesis with kination and gravitational reheating**. *JCAP* 01 (2020). arXiv: 1911.02657 [hep-ph].

[11] Antonio L. Maroto and Alfredo D. Miravet. **Gravitational leptogenesis from metric perturbations**. *Phys. Rev. D* 107.4 (2023). arXiv: 2207.00465 [astro-ph.CO].

[12] Juan Garcia-Bellido and Ester Ruiz Morales. **Primordial black holes from single field models of inflation**. *Phys. Dark Univ.* 18 (2017). arXiv: 1702.03901 [astro-ph.CO].

[13] Guillermo Ballesteros and Marco Taoso. **Primordial black hole dark matter from single field inflation**. *Phys. Rev. D* 97.2 (2018). arXiv: 1709.05565 [hep-ph].

[14] Guillermo Ballesteros, Julián Rey, Marco Taoso, and Alfredo Urbano. **Primordial black holes as dark matter and gravitational waves from single-field polynomial inflation**. *JCAP* 07 (2020). arXiv: 2001.08220 [astro-ph.CO].

[15] Hiroko Niikura et al. **Microlensing constraints on primordial black holes with Subaru/HSC Andromeda observations**. *Nature Astron.* 3.6 (2019). arXiv: 1701.02151 [astro-ph.CO].

[16] B. J. Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun'ichi Yokoyama. **Constraints on primordial black holes from the Galactic gamma-ray background**. *Phys. Rev. D* 94.4 (2016). arXiv: 1604.05349 [astro-ph.CO].

[17] Alexandre Arbey, Jérémy Auffinger, and Joseph Silk. **Constraining primordial black hole masses with the isotropic gamma ray background**. *Phys. Rev. D* 101.2 (2020). arXiv: 1906.04750 [astro-ph.CO].

[18] Ranjan Laha, Julian B. Muñoz, and Tracy R. Slatyer. **INTEGRAL constraints on primordial black holes and particle dark matter**. *Phys. Rev. D* 101.12 (2020). arXiv: 2004.00627 [astro-ph.CO].

[19] J. Berteaud, F. Calore, J. Iguaz, P. D. Serpico, and T. Siegert. **Strong constraints on primordial black hole dark matter from 16 years of INTEGRAL/SPI observations**. *Phys. Rev. D* 106.2 (2022). arXiv: 2202.07483 [astro-ph.HE].

[20] Kenji Tomita. **Non-Linear Theory of Gravitational Instability in the Expanding Universe**. *Progress of Theoretical Physics* 37.5 (May 1967).

[21] Kishore N. Ananda, Chris Clarkson, and David Wands. **The Cosmological gravitational wave background from primordial density perturbations**. *Phys. Rev. D* 75 (2007). arXiv: gr-qc/0612013.

[22] Daniel Baumann, Paul J. Steinhardt, Keitaro Takahashi, and Kiyotomo Ichiki. **Gravitational Wave Spectrum Induced by Primordial Scalar Perturbations**. *Phys. Rev. D* 76 (2007). arXiv: hep-th/0703290.

[23] José Ramón Espinosa, Davide Racco, and Antonio Riotto. **A Cosmological Signature of the SM Higgs Instability: Gravitational Waves**. *JCAP* 09 (2018). arXiv: 1804.07732 [hep-ph].

[24] Kazunori Kohri and Takahiro Terada. **Semianalytic calculation of gravitational wave spectrum nonlinearly induced from primordial curvature perturbations**. *Phys. Rev. D* 97.12 (2018). arXiv: 1804.08577 [gr-qc].

[25] Keisuke Inomata, Kazunori Kohri, Tomohiro Nakama, and Takahiro Terada. **Enhancement of Gravitational Waves Induced by Scalar Perturbations due to a Sudden Transition from an Early Matter Era to the Radiation Era**. *Phys. Rev. D* 100.4 (2019). arXiv: 1904.12879 [astro-ph.CO].

[26] Keisuke Inomata, Kazunori Kohri, Tomohiro Nakama, and Takahiro Terada. **Gravitational Waves Induced by Scalar Perturbations during a Gradual Transition from an Early Matter Era to the Radiation Era**. *JCAP* 10 (2019). arXiv: 1904.12878 [astro-ph.CO].

[27] Guillem Domènech. **Induced gravitational waves in a general cosmological background**. *Int. J. Mod. Phys. D* 29.03 (2020). arXiv: 1912.05583 [gr-qc].

[28] Lang Liu, Zu-Cheng Chen, and Qing-Guo Huang. **Probing the equation of state of the early Universe with pulsar timing arrays** (July 2023). arXiv: 2307.14911 [astro-ph.CO].

[29] Tadeo D. Gomez-Aguilar, Luis E. Padilla, Encieh Erfani, and Juan Carlos Hidalgo. **Constraints on primordial black holes for nonstandard cosmologies** (Aug. 2023). arXiv: 2308.04642 [astro-ph.CO].

[30] Thomas J. Clarke, Edmund J. Copeland, and Adam Moss. **Constraints on primordial gravitational waves from the Cosmic Microwave Background**. *JCAP* 10 (2020). arXiv: 2004.11396 [astro-ph.CO].

[31] Luca Pagano, Laura Salvati, and Alessandro Melchiorri. **New constraints on primordial gravitational waves from Planck 2015**. *Phys. Lett. B* 760 (2016). arXiv: 1508.02393 [astro-ph.CO].

[32] Daniel Baumann, Paul J. Steinhardt, and Neil Turok. **Primordial Black Hole Baryogenesis** (Mar. 2007). arXiv: hep-th/0703250.

[33] Daniel Baumann. **Cosmology**. Cambridge University Press, 2022.

[34] Bernard J. Carr. **The Primordial black hole mass spectrum**. *Astrophys. J.* 201 (1975).

[35] Tomohiro Harada, Chul-Moon Yoo, Tomohiro Nakama, and Yasutaka Koga. **Cosmological long-wavelength solutions and primordial black hole formation**. *Phys. Rev. D* 91.8 (2015). arXiv: 1503.03934 [gr-qc].

[36] Tomohiro Harada, Chul-Moon Yoo, and Kazunori Kohri. **Threshold of primordial black hole formation**. *Phys. Rev. D* 88.8 (2013). [Erratum: Phys.Rev.D 89, 029903 (2014)]. arXiv: 1309.4201 [astro-ph.CO].

[37] Tomohiro Harada, Chul-Moon Yoo, Kazunori Kohri, Ken-ichi Nakao, and Sanjay Jhingan. **Primordial black hole formation in the matter-dominated phase of the Universe**. *Astrophys. J.* 833.1 (2016). arXiv: 1609.01588 [astro-ph.CO].

[38] Tomohiro Harada, Chul-Moon Yoo, Kazunori Kohri, and Ken-Ichi Nakao. **Spins of primordial black holes formed in the matter-dominated phase of the Universe**. *Phys. Rev. D* 96.8 (2017). [Erratum: Phys.Rev.D 99, 069904 (2019)]. arXiv: 1707.03595 [gr-qc].

[39] Flavio Riccardi, Marco Taoso, and Alfredo Urbano. **Solving peak theory in the presence of local non-gaussianities**. *JCAP* 08 (2021). arXiv: 2102.04084 [astro-ph.CO].

[40] Marco Taoso and Alfredo Urbano. **Non-gaussianities for primordial black hole formation**. *JCAP* 08 (2021). arXiv: 2102.03610 [astro-ph.CO].

[41] Cristiano Germani and Ravi K. Sheth. **The statistics of primordial black holes in a radiation dominated Universe – a review and new results** (Aug. 2023). arXiv: 2308.02971 [astro-ph.CO].

[42] Guillermo Ballesteros, Julián Rey, and Fabrizio Rompineve. **Detuning primordial black hole dark matter with early matter domination and axion monodromy**. *JCAP* 06 (2020). arXiv: 1912.01638 [astro-ph.CO].

[43] Michele Maggiore. **Gravitational Waves. Vol. 1: Theory and Experiments**. Oxford Master Series in Physics. Oxford University Press, 2007.

[44] Karim A. Malik, David Wands, and Carlo Ungarelli. **Large scale curvature and entropy perturbations for multiple interacting fluids**. *Phys. Rev. D* 67 (2003). arXiv: astro-ph/0211602.

[45] Steven Weinberg. **Cosmology**. 2008.

[46] Edward W. Kolb and Michael S. Turner. **The Early Universe**. Vol. 69. 1990.

[47] M. H. Thoma. **Parton interaction rates in the quark - gluon plasma**. *Phys. Rev. D* 49 (1994). arXiv: hep-ph/9308257.

[48] P. A. R. Ade et al. **Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season**. *Phys. Rev. Lett.* 127.15 (2021). arXiv: 2110.00483 [astro-ph.CO].

[49] M. Tristram et al. **Improved limits on the tensor-to-scalar ratio using BICEP and Planck data**. *Phys. Rev. D* 105.8 (2022). arXiv: 2112.07961 [astro-ph.CO].

[50] Bernard Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun'ichi Yokoyama. **Constraints on primordial black holes**. *Rept. Prog. Phys.* 84.11 (2021). arXiv: 2002.12778 [astro-ph.CO].