Cosmological constant problem on the horizon

Hassan Firouzjahi *

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

(Received 7 June 2022; accepted 27 September 2022; published 12 October 2022)

We revisit the quantum cosmological constant problem and highlight the important roles played by the de Sitter (dS) horizon of zero-point energy. We argue that fields which are light enough to have a dS horizon of zero-point energy comparable to the Friedmann-Lemaître-Robertson-Walker (FLRW) Hubble radius are the main contributors to dark energy. On the other hand, the zero-point energy of heavy fields develop nonlinearities on sub-Hubble scales and cannot contribute to dark energy. We speculate that our proposal may provide a resolution for both the old and new cosmological constant problems by noting that there exists a field, the (lightest) neutrino, which happens to have a mass comparable to the present background photon temperature. The proposal predicts multiple transient periods of dark energy in the early and late expansion history of the Universe, yielding a higher value of the current Hubble expansion rate which can resolve the H_0 tension problem.

DOI: 10.1103/PhysRevD.106.083510

I. INTRODUCTION

The lambda–cold dark matter (ΛCDM) model has emerged as a successful phenomenological model, explaining the dynamics of the evolution of the cosmos with a handful of free parameters [1,2]. Among the key unknown ingredients are dark matter and dark energy. There are hopes that the former may be addressed one way or another by new physics beyond the Standard Model (SM) of particle physics, while the latter has proven more elusive.

There have been numerous attempts to address the nature of dark energy; for a review, see Refs. [3–5] and references therein. One simple and natural possibility is that the dark energy is simply a cosmological constant term, like what Einstein initially introduced to construct a seemingly static Universe. As is well known, the cosmological constant is associated with big problems in theoretical physics; for a classic review, see Ref. [6]. As we shall review below, it is expected that the cosmological constant receives contributions from the zero-point energy of all fields in the SM. It is usually argued that the zero-point energy density is of the order M^4, in which M is the UV cutoff of the theory. The old cosmological constant problem is why it is not as large as expected from the typical energy scale of high-energy physics, like $(\mathrm{TeV})^4$ or M_P^4, in which $M_P \sim 10^{18}$ GeV is the reduced Planck mass related to the Newton constant G via $M_P^2 = 1/8\pi G$. The new cosmological constant problem is why it is comparable to the energy density of matter at the current epoch in cosmic history, entering into the dynamics of the expansion of the Universe at redshift values around $z_\Lambda \sim 0.3$.

A number of approaches were proposed to address the cosmological constant problem(s). One approach is to employ the fundamental symmetries, like supersymmetry. In supersymmetric theories, there is a one-to-one relation between the number of bosons and the number of fermions which keeps the value of the cosmological constant zero. Of course, nature is not seen to be supersymmetric on energy scales below TeV, so supersymmetry cannot explain the smallness of dark energy. The other approach is based on ideas like quintessence [7,8], in which the dark energy is dynamically evolving to its attractor value, which is zero. However, these kinds of models cannot address the severe fine-tunings involved, as one has to tune the model parameters carefully in order to obtain a tiny value of dark energy at the current epoch. Another class of solutions are based on the self-tuning idea in the context of extra dimensions [9–13]. These solutions are not convincing either, as they either run into singularities or require the fine-tuning of model parameters. A solution of a different kind has been proposed based on the anthropic principle, in which the cosmological constant should choose a range of values that allow gravitationally bound objects like galaxies to form to host the star formation required for the existence of intelligent observers [6,14].

In this paper, we revisit the quantum cosmological constant problem. We emphasize the important roles played by the scale of the de Sitter (dS) horizon associated with the zero-point energy. In particular, we compare the scale of its Hubble radius with the Hubble radius of the Friedmann-Lemaître-Robertson-Walker (FLRW) universe and argue

*firouz@ipm.ir

that the conventional treatment of the quantum cosmological constant problem misses this important effect.

The rest of the paper is organized as follows: In Sec. II, we review the cosmological constant problem and gather some basic formulas. In Sec. III, we highlight the important roles that the horizon scale of the patch of zero-point energy plays, while in Sec. IV, we show why, unlike in conventional treatment, the heavy fields cannot contribute to the cosmological constant. In Sec. V, we calculate the zero-point energy in a dS background, while in Sec. VI, we present various cosmological implications of our proposal, followed by the summary and discussions in Sec. VII.

II. THE COSMOLOGICAL CONSTANT PROBLEM

In this section, we briefly review the cosmological constant problem. For a more extensive review, see Refs. [6,15].

Numerous observations confirm the detection of dark energy at a level comparable to the matter-energy density today [2,16,17]. Therefore, the cosmological constant, if not exactly zero by some symmetry considerations, is as small as $\rho_v \sim (10^{-3}\ \mathrm{eV})^4$. Now the big trouble with the cosmological constant is what mechanism keeps it small enough to be consistent with cosmic evolution.

In its simplest form, the cosmological constant is associated with the problem that the zero-point energy of the fields has quartic divergence with the UV energy scale. To see it more specifically, consider a real scalar field ϕ with the mass m described by the following simple Lagrangian in flat spacetime:

$$S = \int d^4x \left[-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{m^2}{2}\phi^2\right]. \tag{2.1}$$

The energy momentum tensor $T^{\mu\nu}$ is given by

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \eta_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + \frac{m^2}{2}\phi^2\right]. \tag{2.2}$$

In particular, the energy density ρ is given by

$$\rho = T_{00} = \mathcal{H} = \frac{\dot{\phi}^2}{2} + \frac{1}{2}\delta^{ij}\partial_i\phi\partial_j\phi + \frac{m^2}{2}\phi^2, \tag{2.3}$$

in which $\mathcal{H}$ is the Hamiltonian density.

We expand the quantum field in terms of the creation and the annihilation operators in Fourier space,

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int \frac{d^3\mathbf{k}}{\sqrt{2\omega(k)}}[e^{ik\cdot x}a_\mathbf{k} + e^{-ik\cdot x}a^\dagger_\mathbf{k}], \tag{2.4}$$

in which $k^\mu = (\omega(k), \mathbf{k})$, $\omega(k) \equiv \sqrt{\mathbf{k}^2 + m^2}$ and $[a_\mathbf{k}, a^\dagger_\mathbf{q}] = \delta^3(\mathbf{k} - \mathbf{q})$.

The effective energy density contributing to the Einstein field equation is $\langle 0|\rho|0\rangle \equiv \langle\rho\rangle$, in which $\langle 0|X|0\rangle$ means the quantum expectation value of the quantum operator X with respect to the vacuum. Using the form of the operator ρ from Eq. (2.3), we obtain

$$\langle\rho\rangle = \frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\omega(k). \tag{2.5}$$

Performing the same calculation for the pressure p, we obtain [15]

$$\langle p\rangle = \left\langle \frac{1}{3}\perp^{\mu\nu}T_{\mu\nu}\right\rangle = \frac{1}{6}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{\mathbf{k}^2}{\omega(k)}, \tag{2.6}$$

in which $\perp^{\mu\nu} \equiv \eta^{\mu\nu} + u^\mu u^\nu$ is a projection operator and $u^\mu = (1, \mathbf{0})$ is the comoving four-velocity.

To perform the above integrals, one usually imposes a hard UV momentum cutoff $k \le M$, and the integral in Eq. (2.5) is estimated as

$$\langle\rho\rangle \simeq \frac{M^4}{16\pi^2}. \tag{2.7}$$

This is the usual representation of the old cosmological constant problem. If one assumes $M \sim M_P$, then the above estimation yields a contribution to $\langle\rho_v\rangle$ larger than the observed value by some factor of 10^{120}. Of course, if one chooses a smaller UV energy scale, say the scale of SM particle physics $M \sim 10^2$ GeV, then the discrepancy becomes less severe, but still the theoretical estimation is larger by some 10^{56} orders of magnitude from the observed value.

The above analysis can be repeated for other fields, bosonic or fermionic, with the important distinction that for fermionic fields one obtains a negative contribution in $\langle\rho\rangle$. This is because the fermionic fields, unlike the bosonic fields, anticommute with each other.

However, as pointed out in Refs. [15,18–21], the hard momentum cutoff implemented above is problematic, as it does not respect the Lorentz invariance of the setup. In other words, the hard cutoff only respects the $O(3)$ invariance over the three-dimensional momentum space while breaking the four-dimensional $O(1,3)$ Lorentz invariance. To see the fatal problem with this naive hard momentum cutoff regularization, note that to leading order we obtain $\langle p\rangle \simeq M^4/(3\times 16\pi^2)$, which means that $\langle p\rangle/\langle\rho\rangle \simeq 1/3$. On the other hand, the Lorentz invariance of the vacuum requires that

$$\langle T_{\mu\nu}\rangle = -\rho_{\mathrm{vac}}\eta_{\mu\nu}, \tag{2.8}$$

which immediately enforces $\langle p\rangle = -\langle\rho\rangle$. To bypass this problem, one has to employ a regularization scheme which respects the underlying Lorentz invariance. A good candidate is the dimensional regularization approach, which keeps the four-dimensional Lorentz invariance intact.

By performing the integral using the dimensional regularization approach (or any Lorentz-invariant scheme) for the massive real scalar field, one obtains [15,18–21]

$$\langle\rho\rangle = -\langle p\rangle = \frac{m^4}{64\pi^2}\ln\left(\frac{m^2}{\mu^2}\right), \tag{2.9}$$

in which μ is the renormalization scale.

There are a few comments in order concerning Eq. (2.9). First, the requirement of Lorentz invariance $\langle p\rangle = -\langle\rho\rangle$ is explicit. Second, the massless fields such as gravitons, photons, or gluons do not contribute to the energy density of the vacuum. This may be understood by noting that for massless particles, the equation of state is simply $p = \rho/3$, so the requirement of Lorentz invariance $\langle p\rangle = -\langle\rho\rangle$ can be met only if $\langle\rho\rangle = 0$. Third, depending on the renormalization scale, the vacuum energy density can be either positive (dS spacetime) or negative (AdS spacetime). Again, this is unlike the conclusion based on a hard momentum cutoff prescription in which the bosonic (fermionic) fields always yield positive (negative) contributions to $\langle\rho\rangle$. Finally, there is no quartic dependence on the cutoff scale. This can reduce the naive fine-tuning of order 10^{-120} by many orders of magnitudes. As a rough estimate, for the heavy SM fields with mass on the order of 10^2 GeV, we obtain $|\langle\rho\rangle| \sim 10^{44}$ eV^4, which is much smaller than the naive estimation $\langle\rho\rangle \sim M_P^4$. However, it is still vastly larger than the observed value of the vacuum energy density. In addition, note that the contribution from the (lightest) neutrino with $m_\nu \sim 10^{-2}$ eV is $\langle\rho\rangle \sim m_\nu^4 \sim (10^{-2}\ \text{eV})^4$, which is at the same order as the observed value of dark energy. As we shall see, this is not an accident and will be part of our solution for the cosmological constant problem.

Combining the contributions from all fields, bosonic and fermionic, the total contribution in vacuum zero-point energy is obtained to be [15,19]

$$\langle\rho\rangle = \sum_i n_i \frac{m_i^4}{64\pi^2}\ln\left(\frac{m_i^2}{\mu^2}\right), \tag{2.10}$$

in which n_i represents the degree of freedom (polarizations) of each field. For example, for a real scalar field $n = 1$, and for a massive vector field $n = 3$, while for a Dirac fermion $n = -4$. Note that while a fermionic field has an opposite sign contribution in $\langle\rho\rangle$ compared to a bosonic field, it is the combination $n_i \ln(\frac{m_i^2}{\mu^2})$ which determines the sign of the contribution of the corresponding field in $\langle\rho\rangle$. For example, if we take μ to be on the order of the electroweak symmetry breaking scale, then all fermionic (bosnic) fields in the SM spectrum contribute positively (negatively) in $\langle\rho\rangle$.

As discussed above, we comment that while a hard momentum cutoff breaks the underlying Lorentz invariance and yields to a number of problems, such as quartic and quadratic divergences in $\langle\rho\rangle$, with careful considerations one can still employ a hard momentum cutoff. For this purpose, one has to implement noncovariant counterterms in the Lagrangian to cancel the corresponding power-law divergences in $\langle\rho\rangle$. More specifically, expanding $\langle\rho\rangle$ in terms of the momentum cutoff scale M, the leading divergences are M^4 and M^2, followed by the logarithmic contribution as in Eq. (2.9). The quartic divergence has the equation of state of radiation, while the quadratic divergence has that of a spatial curvature. As emphasized in Ref. [15], it is only the logarithmic contribution which has the equation of state of the vacuum, which is what is captured by the dimensional regularization scheme.

III. THE QUESTION OF THE DS HORIZON

To calculate the vacuum energy density, we have assumed a flat background yielding to Eq. (2.9). This seems reasonable, as the energy density is a local quantity which is not sensitive to the large-scale properties of the cosmological background. Indeed, the combination of Lorentz invariance and the equivalence principle guarantee that the flat spacetime approximation to calculate the vacuum energy density as given in Eq. (2.9) is valid. However, there is a subtle issue that was not taken into account when applying the estimated zero-point energy to cosmology. In the presence of gravity—i.e., when the Newton constant G is turned on—associated with each positive vacuum energy is a horizon radius which controls the causal properties of the corresponding dS spacetime. However, in the usual treatment of the cosmological constant problem, it is assumed that the *entire* observable Universe with the current Hubble radius H_0^{-1} is *simultaneously* endowed with a vacuum energy density ρ_v given by Eq. (2.9). This is equivalent to assuming that the quantum zero-point energy fills the entire observable Universe in a single patch of size H_0^{-1}. This is too much to ask.

The correct way of thinking about this problem is to look at the Hubble radius associated with Eq. (2.10) for each field. Let us denote $\rho_v(m)$ and $H_{(m)}$, respectively, as the vacuum energy density and the Hubble rate of dS spacetime created from the zero-point energy of each field with the mass m. Then, we have $3M_P^2 H_{(m)}^2 = \rho_v(m) \sim m^4$, yielding

$$H_{(m)} \simeq \frac{m^2}{M_P}. \tag{3.1}$$

This is one important relation to keep in mind. For example, for an electron field with $m_e \sim$ MeV, we obtain $H_{(m_e)} \sim 10^{-15}$ eV, yielding the Hubble radius $H_{(m_e)}^{-1} \sim 10^9 m$. Logically, this is the largest dS patch which the zero-point energy associated with the electron field can cover coherently. On the other hand, the current Hubble radius is $H_0^{-1} \sim \mathrm{Gpc} \sim 10^{26} m$. From the above simple estimation, we conclude that the observable FLRW Universe encompasses as many as $(H_0^{-1}/H_{(m_e)}^{-1})^3 \sim 10^{51}$ independent patches with the size $H_{(m_e)}^{-1}$. Demanding that as many as 10^{51} dS patches of the electron zero-point energy coherently cover the entire observable Universe is too strong a condition to be realistic. Considering fields heavier than the electron, the situation becomes even worse—i.e,, the ratio $H_0^{-1}/H_{(m)}^{-1}$ becomes much larger.

The picture which emerges from the above discussions is as follows: For heavy fields, the Hubble radius associated with a dS spacetime generated from the zero-point energy is much smaller than the Hubble radius of the observable Universe, $H_{(m)}^{-1} \ll H_0^{-1}$. Therefore, one needs as many as

$$N_{\text{patches}} \sim \left(\frac{H_{(m)}}{H_0}\right)^3 \sim \left(\frac{m}{10^{-2}\ \mathrm{eV}}\right)^6 \tag{3.2}$$

tiny dS patches to cover the current observable Universe. As these tiny patches are created quantum mechanically, they are uncorrelated, so they cannot provide a coherent energy density to be the origin of the observed cosmological constant (or dark energy). This also provides an interesting resolution for the cosmological constant problem: considering the light fields. Specifically, very light fields have much smaller zero-point energy, but at the same time they have a very large Hubble radius, which can encompass the entire observable Universe in a single dS patch. From Eq. (3.2), we see that the (lightest) neutrino with the mass $m_\nu \sim 10^{-2}$ eV has just the right scale to address the cosmological constant problem in which $H_0^{-1} \sim H_{(m_\nu)}^{-1}$. In this view, the entire observable Universe currently is within a single dS patch created by the zero-point energy of the light neutrino field with the vacuum energy density $(10^{-3}\ \mathrm{eV})^4$; see the left panel of Fig. 1 for a schematic view. Correspondingly, the observed vacuum energy density today with the fractional energy density [2] $\Omega_\Lambda^{(0)} \sim 0.7$ is sourced by the zero-point energy of the lightest neutrino.

In the case of heavy fields, as the huge number of dS patches in Eq. (3.2) are uncorrelated, one expects the full volume between these regions to be highly inhomogeneous; for a schematic view, see the right panel of Fig. 1. Consequently, one expects the variance in the energy density to be large, yielding to a large density contrast, $\frac{\delta\rho_v(m)}{\rho_v(m)} \sim 1$. In the next section, we calculate the variance and the density contrast of the vacuum zero-point energy which confirm the above conclusion. In addition, we calculate the correlation length associated with the zero-point energy and show that the correlation length is on the order of Compton radius m^{-1}.

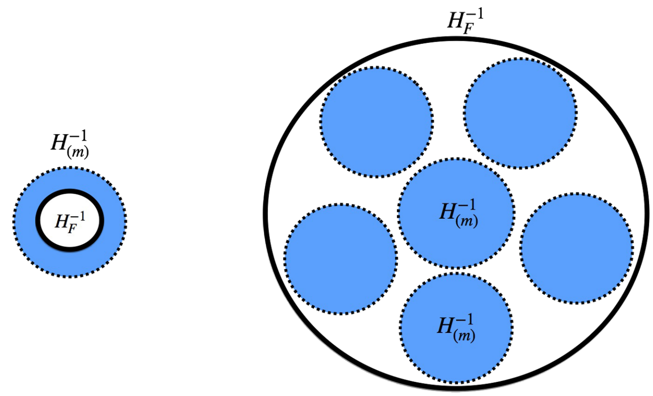

FIG. 1. A schematic view comparing the FLRW Hubble radius H_F^{-1} with the dS horizon of size $H_{(m)}^{-1}$ associated with the zero-point energy of the field with the mass m. Left: this is the case when the field is light, so the FLRW horizon is within a single dS patch of zero-point energy. Right: this represents the case when the field is heavy, so its dS horizon is much smaller than the FLRW horizon, such that a FLRW Hubble patch encompasses many small patches of zero-point energy.

However, before calculating the variance of the vacuum energy density, let us check the accumulated energy associated with the modes which are outside the correlation length—i.e., the modes which have crossed the quantum Compton radius with $k \leq m$. Denoting this accumulated energy by $\delta\rho_C(m)$, we obtain

$$\delta\rho_C(m) = \frac{1}{2}\int_0^m \frac{d^3\mathbf{k}}{(2\pi)^3}\sqrt{k^2+m^2} = \frac{m^4}{32\pi^2}[3\sqrt{2} - \ln(1+\sqrt{2})] \sim \frac{6m^4}{64\pi^2}. \tag{3.3}$$

Now, we can compare the above value of the accumulated energy density of the "long modes" with the supposedly background vacuum energy density $\rho_v(m)$ given in Eq. (2.9). Assuming that there is no exponential hierarchy between m and the renormalization scale μ, we obtain

$$\frac{\delta\rho_C(m)}{\rho_v(m)} \sim 1. \tag{3.4}$$

We conclude that the accumulated energy density of the modes which are outside the correlation length m^{-1} is comparable to the background vacuum energy density. This rings the bell that the background space can be highly inhomogeneous, and the nearby dS patches cannot provide a coherent background for the large-scale Universe.

IV. DENSITY CONTRAST AND CORRELATION LENGTH

The quantum average of the vacuum zero-point energy density $\langle\rho_v\rangle = \langle\mathcal{H}\rangle$ for the real scalar field is given in

Eq. (2.9). Now, for the moment, suppose we neglect the contribution from the FLRW matter and radiation energy density and assume that ρ_v is the dominant energy density in cosmic expansion. In order for $\langle\rho_v\rangle$ to be viewed as a viable background energy density for the cosmic evolution, we have to make sure that the variance in energy density is small, yielding to a small density contrast $\delta\rho_v/\langle\rho_v\rangle \ll 1$. Otherwise, a region with large density contrast $\delta\rho_v/\langle\rho_v\rangle \sim 1$ develops inhomogeneities and may even collapse into black holes.

In this section, we calculate explicitly the variance $\delta\rho_v^2 \equiv \langle\rho_v^2\rangle - \langle\rho_v\rangle^2$ and the density contrast $\delta\rho_v/\langle\rho_v\rangle$. First, we present the analysis for the case of a real scale field, and then we go to the case of a Dirac fermion field.

A. Real scalar field

First, we consider the case of a real scalar field with the action (2.1). Let us denote the three different contributions to energy density of the scalar field in Eq. (2.3) by ρ_1, ρ_2, and ρ_3 as follows:

$$\rho_1 \equiv \frac{\dot{\phi}^2}{2}, \qquad \rho_2 \equiv \frac{1}{2}\delta^{ij}\partial_i\phi\partial_j\phi, \qquad \rho_3 \equiv \frac{m^2}{2}\phi^2. \tag{4.1}$$

To calculate $\langle\rho_i\rangle$, we go to Fourier space and expand the field in terms of the mode functions as in Eq. (2.4). Performing the momentum integrals using the dimensional regularization scheme to regularize the UV divergences (for a sample analysis of dimensional regularization, see Sec. V), we obtain

$$\langle\rho_1\rangle = \frac{1}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\frac{\omega(q)}{2} = \frac{m^4}{128\pi^2}\ln\left(\frac{m^2}{\mu^2}\right), \tag{4.2}$$

$$\langle\rho_2\rangle = \frac{1}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\frac{\mathbf{q}^2}{2\omega(q)} = \frac{-3m^4}{128\pi^2}\ln\left(\frac{m^2}{\mu^2}\right), \tag{4.3}$$

$$\langle\rho_3\rangle = \frac{1}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\frac{m^2}{2\omega(q)} = \frac{4m^4}{128\pi^2}\ln\left(\frac{m^2}{\mu^2}\right). \tag{4.4}$$

Curiously, note that while all three operators ρ_1, ρ_2, and ρ_3 are positive definite and Hermitian operators, the expectation value of ρ_2 has the opposite sign compared to those of ρ_1 and ρ_2. Furthermore, noting that $\langle\rho\rangle = \langle\rho_1\rangle + \langle\rho_2\rangle + \langle\rho_3\rangle$, the following relations hold, which will be useful later on:

$$\langle\rho\rangle = 2\langle\rho_1\rangle = -\frac{2}{3}\langle\rho_2\rangle = \frac{1}{2}\langle\rho_3\rangle. \tag{4.5}$$

Now, to calculate $\delta\rho^2$, we have

$$\begin{aligned}\delta\rho^2 = {}& \langle\rho_1^2\rangle + \langle\rho_2^2\rangle + \langle\rho_3^2\rangle + \langle\rho_1\rho_2\rangle + \langle\rho_2\rho_1\rangle \\ &+ \langle\rho_1\rho_3\rangle + \langle\rho_3\rho_1\rangle + \langle\rho_2\rho_3\rangle + \langle\rho_3\rho_2\rangle \\ &- (\langle\rho_1\rangle^2 + \langle\rho_2\rangle^2 + \langle\rho_3\rangle^2) \\ &- 2\langle\rho_1\rangle\langle\rho_2\rangle - 2\langle\rho_1\rangle\langle\rho_3\rangle - 2\langle\rho_2\rangle\langle\rho_3\rangle. \end{aligned} \tag{4.6}$$

Note that since $\dot{\phi}$ and ϕ do not commute, we have considered all possible orderings of the operators. However, as we shall see, these orderings do not matter.

First, we note that ϕ is a free Gaussian field, so using the standard Wick theorem (or by direct calculations), one can show that

$$\langle\rho_i^2\rangle = 3\langle\rho_i\rangle^2, \quad i = 1, 3. \tag{4.7}$$

However, the calculation of $\langle\rho_2^2\rangle$ is somewhat tricky. Performing various contractions, we obtain

$$\begin{aligned}\langle\rho_2^2\rangle &= \langle\rho_2\rangle^2 + (2)\frac{1}{4}\int\frac{d^3\mathbf{q}_1}{(2\pi)^3 2\omega(q_1)}\frac{d^3\mathbf{q}_2}{(2\pi)^3 2\omega(q_2)}(\mathbf{q}_1\cdot\mathbf{q}_2)^2 \\ &= \langle\rho_2\rangle^2 + \left(2\times\frac{1}{3}\right)\langle\rho_2\rangle^2, \end{aligned} \tag{4.8}$$

in which the factor $1/3$ originates from an integral of the form $\int_{-1}^{+1} d(\cos\theta)\cos^2\theta$.

For the mix correlations, we calculate one case for illustration, and the rest would be similar. For this purpose, let us calculate $\langle\rho_1\rho_2\rangle$. To simplify the analysis, note that due to translation invariance we can simply set $\mathbf{x} = 0$, obtaining

$$\begin{aligned}\langle\rho_1\rho_2\rangle &= \frac{i^4}{4}\int\frac{d^3\mathbf{q}_i}{(2\pi)^{\frac{3}{2}}}\frac{\omega(q_1)\omega(q_2)\mathbf{q}_3\cdot\mathbf{q}_4}{2\omega(q_i)}\langle 0|(-a_{\mathbf{q}_1} + a^\dagger_{\mathbf{q}_1})(-a_{\mathbf{q}_2} + a^\dagger_{\mathbf{q}_2})(a_{\mathbf{q}_3} - a^\dagger_{\mathbf{q}_3})(a_{\mathbf{q}_4} - a^\dagger_{\mathbf{q}_4})|0\rangle \\ &= \frac{1}{4}\int\frac{d^3\mathbf{q}_1}{(2\pi)^3 2\omega(q_1)}\frac{d^3\mathbf{q}_2}{(2\pi)^3 2\omega(q_2)}[2\omega(q_1)\omega(q_2)\mathbf{q}_1\cdot\mathbf{q}_2 + \omega(q_1)^2\mathbf{q}_2^2]. \end{aligned} \tag{4.9}$$

On the other hand, due to rotation invariance, the integral over $\mathbf{q}_1\cdot\mathbf{q}_2$ vanishes, and we obtain

$$\langle\rho_1\rho_2\rangle = \frac{1}{4}\int\frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{d^3\mathbf{q}_2}{(2\pi)^3}\frac{\omega(q_1)}{2}\frac{\mathbf{q}_2^2}{2\omega(q_2)} = \langle\rho_1\rangle\langle\rho_2\rangle. \tag{4.10}$$

Similarly, we obtain

$$\langle\rho_2\rho_3\rangle = \langle\rho_2\rangle\langle\rho_3\rangle. \tag{4.11}$$

Since $\dot{\phi}$ and ϕ do not commute, the calculation for $\langle\rho_1\rho_3\rangle$ is a bit different, yielding

$$\langle\rho_1\rho_3\rangle = \langle\rho_1\rangle\langle\rho_3\rangle - \frac{m^2}{2}\left(\int\frac{d^3\mathbf{q}_1}{(2\pi)^3}\right)^2. \qquad (4.12)$$

Regularizing the last term above which is UV divergent by absorbing it into counterterms, we obtain

$$\langle\rho_1\rho_3\rangle = \langle\rho_1\rangle\langle\rho_3\rangle = \langle\rho_3\rho_1\rangle. \qquad (4.13)$$

Combining all, we obtain

$$\delta\rho^2 = 2(\langle\rho_1\rangle^2 + \langle\rho_2\rangle^2 + \langle\rho_3\rangle^2) - \frac{4}{3}\langle\rho_2\rangle^2 = 10\langle\rho\rangle^2, \qquad (4.14)$$

where to obtain the final result, the relations in Eq. (4.5) have been used.

From the above expression, the density contrast (taking both signs after the square root) is obtained to be

$$\frac{\delta\rho_v}{\langle\rho_v\rangle} = \pm\sqrt{10}. \qquad (4.15)$$

This is an interesting result. As argued before, the energy density of the background constructed from the zero-point energy is very inhomogeneous. This is in agreement with the intuition that a large number of incoherent patches of zero-point energy cannot cover the entire observable Universe. Also, note that both signs are allowed, so we can have either overdensity or underdensity. Specifically, the quantity $\langle\rho\rangle + \delta\rho$ can take either sign, so some regions of spacetime can be dS type, and some other parts AdS (anti–de Sitter) type. The parts of the spacetime which are AdS type may collapse due to instabilities.

The above conclusion that $\langle\rho\rangle + \delta\rho$ can be negative is based on the hidden assumption that the statistical distribution of ρ is mostly symmetric around its average value $\langle\rho\rangle$, so a large variance can yield to a negative local value of $\langle\rho\rangle + \delta\rho$. This is certainly the case for a Gaussian distribution, while for other distribution, like Poisson distribution, it may not be the case. Therefore, the nature of the distribution of ρ is an important question. In our setup with a large uncorrelated dS patches of heavy fields created quantum mechanically from the vacuum, it may be reasonable to assume that ρ follows a normal (i.e., Gaussian) distribution. But it would be interesting to calculate the distribution of ρ from first principles. In the following analysis, we assume a Gaussian-type distribution such that $\langle\rho\rangle + \delta\rho$ can become negative for a large variance.

In the above analysis, we have neglected the conventional matter and radiation energy density of the FLRW Universe, ρ_F. Now, let us look at the total energy density from the combination of the FLRW matter-radiation ρ_F and the vacuum energy density, defined via $\rho_T = \rho_F + \rho_v$ in which ρ_v is the vacuum energy density given in Eq. (2.9). Note that ρ_F is independent of the zero-point energy, so its quantum expectation is trivial—i.e., $\langle\rho_F\rangle = \rho_F$. Correspondingly, $\langle\rho_T\rangle = \rho_F + \langle\rho_v\rangle$ and $\delta\rho = \delta\rho_v$ in which $\delta\rho_v$ is as given in Eq. (4.14). Therefore, the total density contrast is given by

$$\frac{\delta\rho_T}{\langle\rho_T\rangle} = \frac{\delta\rho_v}{\rho_F + \langle\rho_v\rangle} = \pm\sqrt{10}\frac{\langle\rho_v\rangle}{\rho_F + \langle\rho_v\rangle}. \qquad (4.16)$$

Demanding that the amplitude of the total density contrast be small, $|\frac{\delta\rho_T}{\langle\rho_T\rangle}| < 1$, we obtain

$$\langle\rho_v\rangle < (\sqrt{10} - 1)^{-1}\rho_F \sim \frac{\rho_F}{2}. \qquad (4.17)$$

This means that in order for the background consisting of matter-radiation plus zero-point energy to be stable for cosmological expansion, one requires that the vacuum energy be less than the FLRW matter-radiation energy density.

Motivated by the above analysis, it would be instructive to calculate the fractional contrast in $\mathcal{K} \equiv \rho + 3p$, as the quantity $\mathcal{K}$ (the strong energy condition with $\mathcal{K} \geq 0$) plays key roles in gravitational collapse and black hole formations in the Hawking-Penrose theorem. One can check that the vacuum pressure is given by

$$p_v = \rho_1 - \frac{\rho_2}{3} - \rho_3, \qquad (4.18)$$

so we have

$$\mathcal{K} = 4\rho_1 - 2\rho_3. \qquad (4.19)$$

Now, repeating the same steps as in the case of ρ, one can show that $\langle\mathcal{K}\rangle = -2\langle\rho_v\rangle$ and $\langle\mathcal{K}^2\rangle = 44\langle\rho_v\rangle^2$. Combining these two results, we obtain

$$\delta\mathcal{K}^2 \equiv \langle\mathcal{K}^2\rangle - \langle\mathcal{K}\rangle^2 = 40\langle\rho_v\rangle^2, \qquad (4.20)$$

and

$$\frac{\delta\mathcal{K}}{\langle\mathcal{K}\rangle} = \pm\sqrt{10}. \qquad (4.21)$$

This suggests that the local curvature of the spacetime fluctuates rapidly, in which some regions are expected to become AdS type and may collapse to form black holes.

Having calculated the variance and the density contrast, it is also instructive to calculate the correlation length of the energy density. The correlation length is related to the connected part of the correlation function as follows: The two-point correlation function is given by $\langle\rho(\mathbf{x})\rho(\mathbf{y})\rangle$. Due to translation invariance, the correlation function depends only on $\mathbf{x} - \mathbf{y}$, so we simply set $\mathbf{y} = \mathbf{0}$. The connected part of the two-point correlation function is given by

$$\langle\rho(\mathbf{x})\rho(\mathbf{0})\rangle_c \equiv \langle\rho(\mathbf{x})\rho(\mathbf{0})\rangle - \langle\rho^2\rangle. \tag{4.22}$$

Performing the analysis similar to the case of variance, one can calculate $\langle\rho(\mathbf{x})\rho(\mathbf{0})\rangle_c$. The analysis is somewhat involved. However, what we are interested in is the correlation length ξ, which controls the exponential falloff of the connected correlation function at a large distance:

$$\langle\rho(\mathbf{x})\rho(\mathbf{0})\rangle_c \to e^{-\frac{r}{\xi}}, \qquad (r\to\infty). \tag{4.23}$$

To estimate ξ, let us look at $\langle\rho_3(\mathbf{x})\rho_3(\mathbf{0})\rangle_c$, which is easier. Upon performing the various contractions, we obtain

$$\langle\rho_3(\mathbf{x})\rho_3(\mathbf{0})\rangle_c = \frac{m^4}{2}\left(\int\frac{d^3\mathbf{q}}{(2\pi)^3 2\omega(q)}e^{-i\mathbf{q}\cdot\mathbf{x}}\right)^2. \tag{4.24}$$

The above integral is well known in QFT, which yields [22]

$$\langle\rho_3(\mathbf{x})\rho_3(\mathbf{0})\rangle_c = \frac{m^8}{32\pi^4}\left(\frac{K_1(mr)}{mr}\right)^2, \tag{4.25}$$

in which $K_1(x)$ is the modified Bessel function. Using the asymptotic behaviour $K_1(x)\sim e^{-x}$, we obtain $\langle\rho_3(\mathbf{x})\rho_3(\mathbf{0})\rangle_c \sim e^{-2mr}$, yielding the correlation length $\xi = 1/(2m)$. This is somewhat expected, as m is the only mass scale relevant for this free QFT. Calculating the two-point correlation functions for other components of the energy density yields the same exponential behavior for large r.

Now having the correlation length at hand, we can compare the Hubble radius of zero-point energy $H_{(m)}^{-1}$ and ξ, obtaining

$$\frac{\xi}{H_{(m)}^{-1}} \sim \frac{H_{(m)}}{m} \sim \frac{m}{M_P} \ll 1. \tag{4.26}$$

This is in line with our intuitive argument in the previous section. The correlation length of the zero-point energy density is much smaller than the Hubble radius of each dS patch, so the dS patches are practically uncorrelated. As such, one cannot cover the entire FLRW horizon with a vast number of uncorrelated dS patches (like the right panel of Fig. 1) which are created quantum mechanically and yet expect them to behave as a uniform cosmological constant.

As the above discussions suggest, the dS spacetime created purely from the vacuum zero-point energy with $\delta\rho_v/\langle\rho_v\rangle \sim \pm 1$ is unstable under perturbations. Does this suggest that the locally formed AdS regions will collapse to black holes? We cannot be sure about the answer to this question (see Sec. VI for further discussions/speculations). But suppose that some regions may collapse to form black holes. It is therefore helpful to compare the Jeans length of the corresponding gravitational instability and the horizon size of the supposedly formed black holes. Assuming a dS patch of size $H_{(m)}^{-1}$ collapses into a black hole, the mass of the corresponding black hole is on the order of $M \sim \rho_v H_{(m)}^{-3} \sim M_P^2 H_{(m)}^{-1} \sim M_P^3/m^2$. This yields to a Schwarzschild radius $r_S \sim H_{(m)}^{-1}$. In other words, the Schwarzschild radius is nothing but the horizon radius of the dS patch. On the other hand, the Jeans length of perturbations is given by $\lambda_J \sim c_s/\sqrt{G\rho_v}$, in which c_s is the sound speed of perturbations. In our picture with $\delta\rho_v/\langle\rho_v\rangle \sim \pm 1$, parts of spacetime fragment to AdS type while other pars are dS type (though with a higher Hubble expansion rate than the background dS). Assuming that the sound speed of perturbations is not exponentially small, the Jeans length is on the order of $\lambda_J \sim \sqrt{M_P^2/\rho_v} \sim H_{(m)}^{-1} \sim r_S$. Interestingly, we conclude that the Jeans length of gravitational instability is on the same order as the black hole horizon. This may support the conclusion that the dS spacetime constructed from the vacuum zero-point energy is unstable to fragmentations and the formation of black holes of size $H_{(m)}^{-1}$. However, as just mentioned above, the situation is more complicated, and we leave it as an open question whether locally formed AdS regions from large quantum fluctuations can collapse to form black holes.

Before closing this subsection, there is an important comment in order. In our analysis, we have not taken into account the backreactions of strong inhomogeneities on the background geometry. As we have seen, for a pure vacuum-dominated energy density, we have $\delta\rho_v/\rho_v \sim 1$. Correspondingly, this induces strong inhomogeneities in geometry. At the start, assuming a uniformly distributed vacuum energy, we have assumed a homogenous and isotropic FLRW background determined by the induced Hubble expansion rate $H_{(m)}$. In the presence of strong inhomogeneities, this assumption is violated, and one has to consider a general inhomogeneous and anisotropic background to properly take into account the strong inhomogeneities from the vacuum fluctuations. This is a difficult and open question. On the other hand, as we have stressed above, to have a consistent background, we require a classical source of energy density, ρ_F, such that the total density contrast $\delta\rho_v/\rho_T$ remains under perturbative control. However, this assumption breaks down when ρ_F falls off as the Universe expands and when $\rho_F \sim \rho_v$. At this stage, the inhomogeneities from vacuum fluctuations cannot be neglected. This will bring up important questions, which we will come back to in Sec. VI.

B. Dirac fermion field

Here we present the analysis of variance and density contrast for a Dirac fermion field Ψ. In what follows, we follow the notations of Weinberg [22].

Expressing the mode function in Fourier space, we have

$$\Psi = \int\frac{d^3\mathbf{q}}{(2\pi)^{\frac{3}{2}}}\sum_\sigma[u(\mathbf{q},\sigma)e^{iq\cdot x}a(\mathbf{q},\sigma) + v(\mathbf{q},\sigma)e^{-iq\cdot x}b^\dagger(\mathbf{q},\sigma)], \tag{4.27}$$

in which $a(\mathbf{q},\sigma)$ and $b(\mathbf{q},\sigma)$ are the annihilation operators for the "particle" and "antiparticle," and $\sigma=\pm\frac{1}{2}$ are two spin polarizations, while $u(\mathbf{q},\sigma)$ and $v(\mathbf{q},\sigma)$ are the corresponding mode functions. As usual, $a(\mathbf{q},\sigma)$ and $b(\mathbf{q},\sigma)$ are anticommuting operators in which $\{a(\mathbf{q},\sigma), a^\dagger(\mathbf{q}',\sigma')\}=\delta_{\sigma\sigma'}\delta(\mathbf{q}-\mathbf{q}')$ and $\{b(\mathbf{q},\sigma), b^\dagger(\mathbf{q}',\sigma')\}=\delta_{\sigma\sigma'}\delta(\mathbf{q}-\mathbf{q}')$, while the rest of the anticommutators are zero.

The action of the field with the mass m is given by

$$S=-\int d^4x\bar{\Psi}(\gamma^\mu\partial_\mu+m)\Psi, \tag{4.28}$$

in which γ^μ are the Dirac matrices satisfying the anticommutation relation

$$\{\gamma^\mu,\gamma^\nu\}=2\eta^{\mu\nu}, \tag{4.29}$$

while $\bar{\Psi}\equiv\Psi^\dagger\beta$, in which the matrix β is given by $\beta\equiv i\gamma^0$.

The mode functions $u(\mathbf{q},\sigma)$ and $v(\mathbf{q},\sigma)$ satisfy the following relations:

$$\sum_\sigma u_\mu(\mathbf{q},\sigma)\bar{u}_\nu(\mathbf{q},\sigma)=\left(\frac{-i\gamma^\mu q_\mu+m}{2q^0}\right)_{\mu\nu}, \tag{4.30}$$

and

$$\sum_\sigma v_\mu(\mathbf{q},\sigma)\bar{v}_\nu(\mathbf{q},\sigma)=\left(\frac{-i\gamma^\mu q_\mu-m}{2q^0}\right)_{\mu\nu}, \tag{4.31}$$

in which $q^0=\sqrt{\mathbf{q}^2+m^2}$ is the energy of the particle. Furthermore, the following relation is useful as well:

$$\sum_\sigma \bar{v}(\mathbf{q},\sigma)\gamma^0 v(\mathbf{q},\sigma)=-2i. \tag{4.32}$$

Finally, the energy density $\rho=T_{00}$ is given by

$$\rho=-\frac{1}{2}\bar{\Psi}\gamma^0\partial_0\Psi+\frac{1}{2}\partial_0\bar{\Psi}\gamma^0\Psi. \tag{4.33}$$

It is convenient to define the above two terms via $\rho_1\equiv-\frac{1}{2}\bar{\Psi}\gamma^0\partial_0\Psi$ and $\rho_2\equiv\frac{1}{2}\partial_0\bar{\Psi}\gamma^0\Psi$ (these values of ρ_i should not be confused with ρ_i in the case of a real scalar field studied in the previous subsection). One can check that $\rho_2=\rho_1^\dagger$, so $\rho=\rho_1+\rho_1^\dagger$. This shows, as expected, that ρ is a Hermitian operator.

As a warmup, it is helpful to calculate $\langle\rho\rangle$ from Eq. (4.33). We obtain $\langle\rho\rangle=\langle\rho_1\rangle+\langle\rho_1\rangle^*$, so we only need to calculate $\langle\rho_1\rangle$. Using the decomposition of the field in terms of the mode functions in Eq. (4.27), we have

$$\begin{aligned}\langle\rho_1\rangle&=-\frac{1}{2}\langle 0|\bar{\Psi}\gamma^0\partial_0\Psi|0\rangle\\&=-\frac{i}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\sum_\sigma\bar{v}(\mathbf{q},\sigma)\gamma^0 v(\mathbf{q},\sigma)q^0\\&=-\int\frac{d^3\mathbf{q}}{(2\pi)^3}q^0.\end{aligned} \tag{4.34}$$

As a result,

$$\langle\rho\rangle=2\langle\rho_1\rangle=-4\int\frac{d^3\mathbf{q}}{(2\pi)^3}\frac{q^0}{2}=-4\left[\frac{m^4}{64\pi^2}\ln\left(\frac{m^2}{\mu^2}\right)\right]. \tag{4.35}$$

The above result confirms the factor 4, since we have four independent degrees of freedom in the Dirac field. In addition, compared to the case of a real scalar field given in Eq. (2.5), we have an overall minus sign which is the hallmark of the fermionic field.

To calculate $\langle\rho^2\rangle$, note that $\langle\rho_2\rangle=\langle\rho_1\rangle^*$, $\langle\rho_2^2\rangle=\langle\rho_1^2\rangle^*$, and $\langle\rho_1\rho_2\rangle=\langle\rho_1\rho_2\rangle^*=\langle\rho_2\rho_1\rangle$, which will be handy in the following analysis. Furthermore, in the analysis below, we show that $\langle\rho_1^2\rangle$ and $\langle\rho_1\rho_2\rangle$ are real, so using the above mentioned relations, we obtain

$$\langle\rho^2\rangle=2\langle\rho_1^2\rangle+2\langle\rho_1\rho_2\rangle. \tag{4.36}$$

We calculate each term in turn. Due to translation invariance, we set $x^\mu=0$ without loss of generality.

Starting with $\langle\rho_1^2\rangle$, we have

$$\langle\rho_1^2\rangle=\frac{1}{4}\langle 0|(\bar{\Psi}\gamma^0\partial_0\Psi)(\bar{\Psi}\gamma^0\partial_0\Psi)|0\rangle. \tag{4.37}$$

Using the decomposition in terms of mode functions from Eq. (4.27) and performing all the possible contractions involving the creation and annihilation operators, we end up with two types of contributions in $\langle\rho_1^2\rangle$. Denoting these contributions as term A and term B, we have $\langle\rho_1^2\rangle\equiv A+B$, in which

$$A\equiv\frac{1}{4}\sum_{\sigma_1\sigma_2}\int\frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{d^3\mathbf{q}_2}{(2\pi)^3}q_1^0q_2^0\bar{v}(\mathbf{q}_1,\sigma_1)\gamma^0u(\mathbf{q}_2,\sigma_2)\bar{u}(\mathbf{q}_2,\sigma_2)\gamma^0v(\mathbf{q}_1,\sigma_1) \tag{4.38}$$

and

$$B \equiv -\frac{1}{4}\sum_{\sigma_1\sigma_2}\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{d^3\mathbf{q}_2}{(2\pi)^3} q_1^0 q_2^0 \bar{v}(\mathbf{q}_1,\sigma_1)\gamma^0 v(\mathbf{q}_1,\sigma_1)\bar{v}(\mathbf{q}_2,\sigma_2)\gamma^0 v(\mathbf{q}_2,\sigma_2). \tag{4.39}$$

Using the relations (4.30) and (4.31), we obtain

$$A = \frac{1}{4}\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{d^3\mathbf{q}_2}{(2\pi)^3}(m^2 - q_1^0 q_2^0)$$
$$= \frac{m^2}{4}\left(\int \frac{d^3\mathbf{q}}{(2\pi)^3}\right)^2 - \frac{1}{4}\langle\rho_1\rangle^2, \tag{4.40}$$

where to obtain the last result, Eq. (4.35) for $\langle\rho_1\rangle$ has been used.

Similarly, using the relation (4.32), for the B term we obtain

$$B = \left(\int \frac{d^3\mathbf{q}}{(2\pi)^3} q^0\right)^2 = \langle\rho_1\rangle^2. \tag{4.41}$$

Combining the above two contributions, we obtain

$$\langle\rho_1^2\rangle = A + B = \frac{m^2}{4}\left(\int \frac{d^3\mathbf{q}}{(2\pi)^3}\right)^2 + \frac{3}{4}\langle\rho_1\rangle^2. \tag{4.42}$$

As promised, $\langle\rho_1^2\rangle$ is real, so we do not need to calculate $\langle\rho_2^2\rangle$, since $\langle\rho_2^2\rangle = \langle\rho_1^2\rangle^* = \langle\rho_1^2\rangle$. Furthermore, as in the case of a scalar field, the first term above is UV divergent and has to be absorbed into counterterms via regularization, so after regularization we obtain $\langle\rho_1^2\rangle \cong \frac{3}{4}\langle\rho_1\rangle^2$.

Now, we calculate $\langle\rho_1\rho_2\rangle$, which is

$$\langle\rho_1\rho_2\rangle = -\frac{1}{4}\langle 0|(\bar{\Psi}\gamma^0\partial_0\Psi)(\partial_0\bar{\Psi}\gamma^0\Psi)|0\rangle. \tag{4.43}$$

Like in the previous case, after performing all contractions involving the creation and the annihilation operators, we have two different contributions in $\langle\rho_1\rho_2\rangle$. Incidentally, one of these contributions is the B term above, while the other contribution is new, denoted by the C term, so $\langle\rho_1\rho_2\rangle = B + C$, in which

$$C \equiv -\frac{1}{4}\sum_{\sigma_1\sigma_2}\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{d^3\mathbf{q}_2}{(2\pi)^3}(q_2^0)^2 \bar{v}(\mathbf{q}_1,\sigma_1)$$
$$\times \gamma^0 u(\mathbf{q}_2,\sigma_2)\bar{u}(\mathbf{q}_2,\sigma_2)\gamma^0 v(\mathbf{q}_1,\sigma_1). \tag{4.44}$$

Note that while the C term looks very similar to the A term, they are different, as the latter has the product $q_1^0 q_2^0$, while the former has $(q_2^0)^2$ inside the double integrals.

Using the relations (4.30) and (4.31), we obtain

$$C = -\frac{m^2}{4}\left(\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\frac{1}{q_1^0}\right)\left(\int \frac{d^3\mathbf{q}_2}{(2\pi)^3} q_2^0\right)$$
$$+ \frac{1}{4}\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\int \frac{d^3\mathbf{q}_2}{(2\pi)^3}(q_2^0)^2. \tag{4.45}$$

Combining the two contributions B and C, we obtain

$$\langle\rho_1\rho_2\rangle = \langle\rho_1\rangle^2 + \frac{m^2}{4}\left(\int \frac{d^3\mathbf{q}}{(2\pi)^3}\frac{1}{q^0}\right)\langle\rho_1\rangle$$
$$+ \frac{1}{4}\int \frac{d^3\mathbf{q}_1}{(2\pi)^3}\int \frac{d^3\mathbf{q}_2}{(2\pi)^3}(q_2^0)^2. \tag{4.46}$$

The last term above is again UV divergent and has to be absorbed into counterterms, so we keep the first two terms above in the regularized $\langle\rho_1\rho_2\rangle$.

Having calculated $\langle\rho_1^2\rangle$ and $\langle\rho_1\rho_2\rangle$ and discarding the two UV divergent terms as discussed above, from Eq. (4.36) $\langle\rho^2\rangle$ is calculated to be

$$\langle\rho^2\rangle = \frac{7}{2}\langle\rho_1\rangle^2 + \frac{m^2}{2}\left(\int \frac{d^3\mathbf{q}}{(2\pi)^3}\frac{1}{q^0}\right)\langle\rho_1\rangle. \tag{4.47}$$

On the other hand, the integral in the second term above is already calculated in the case of the real scalar field in Eq. (4.4), which, combined with Eq. (4.35), can be expressed in terms of $\langle\rho_1\rangle$ as

$$\frac{m^2}{2}\left(\int \frac{d^3\mathbf{q}}{(2\pi)^3}\frac{1}{q^0}\right) = -2\langle\rho_1\rangle. \tag{4.48}$$

Now, combining all the terms, we obtain

$$\langle\rho^2\rangle = \frac{3}{2}\langle\rho_1\rangle^2 = \frac{3}{8}\langle\rho\rangle^2. \tag{4.49}$$

Finally, the variance $\delta\rho^2 = \langle\rho^2\rangle - \langle\rho\rangle^2$ is obtained to be

$$\delta\rho^2 = -\frac{10}{16}\langle\rho\rangle^2. \tag{4.50}$$

Surprisingly, the variance is negative. This may look absurd; however, we already encountered seemingly negative values for the expectation values of positive definite operators, such as the expectation value of $\langle\rho_2\rangle$ in Eq. (4.3) for the case of a real scalar field. This phenomenon may be attributed to regularization which absorbs positive-infinity terms, leaving the finite possible negative terms.

The amplitude of the zero-point density perturbation is obtained to be

$$\left|\frac{\delta\rho_v}{\rho_v}\right| = \frac{\sqrt{10}}{4}, \tag{4.51}$$

in which by $|\frac{\delta\rho_v}{\rho_v}|$ we mean the amplitude with the minus sign absorbed. Curiously, the factor $\sqrt{10}$ is common as in the case of a real scalar field; see Eq. (4.15). The factor $\frac{1}{4}$ may be attributed to the four degrees of freedom encoded in the Dirac fermion field. Furthermore, by calculating the variance of $\mathcal{K}$ (defined like in the case of a scalar field), we obtain $|\frac{\delta\mathcal{K}}{\langle\mathcal{K}\rangle}| = \frac{\sqrt{10}}{4}$.

As in the case of a scalar field, we see that the amplitude of density contrast associated with the zero-point energy of the fermionic field is on the order of unity. As before, we conclude that the vacuum zero-point energy by itself cannot be the dominant source of the cosmic energy density. One requires a dominant (classical) source of energy density like the matter-radiation energy density ρ_F with $\rho_F > \rho_v$ at each stage in cosmic expansion history to obtain a stable cosmological background.

Finally, calculating the correlation length ξ of the zero-point energy, one can check that, as in the case of a real scalar field, $\xi \sim m^{-1}$ as expected.

V. ZERO-POINT ENERGY IN DS BACKGROUND

The vacuum zero-point energy density [Eq. (2.9)] is calculated assuming a flat background. As discussed before, one expects this result to hold true for a curved background as well. This is because general relativity is a local theory, and the combination of the Lorentz invariance and the equivalence principle guarantees the validity of Eq. (2.9) even in a curved field space. Having said this, it is a nontrivial exercise to see how this works in practice. Specifically, the accumulated zero-point energy yielding to Eq. (2.9) generates a dS background. At the same time, one has to solve the mode function $\phi(x^\mu)$ in this dS spacetime and then construct $\langle T_{00}\rangle$. Technically speaking, it seems nontrivial how Eq. (2.9) emerges as the final result. This is more intriguing for a dS spacetime, noting that a dS spacetime is characterized by the size of its horizon, $H^{-1}_{(m)}$ in our notation, while Eq. (2.9) is indifferent to this length scale.

In this section, we solve the mode function for a real scalar field in a dS background and calculate the vacuum zero-point energy. Note the nontrivial step that we have to calculate the mode function in a dS space which itself is created from the vacuum zero-point energy. In other words, there is no background dS space other than what would be generated from the vacuum zero-point energy in the first place. This resembles a strong backreaction problem where we have to create a background out of quantum perturbations and at the same time solve the mode function in the geometry of the created background. Then we have to regularize the energy density to make sure the resultant value of $\langle\rho\rangle$ agrees with its original value in the supposedly created background. This seems like moving in a loop. For relevant works, see also Refs. [15,23–25] and Chapter 6 of Ref. [26], which study the analysis of zero-point energy in a *given* curved background. For earlier works concerning the renormalized stress energy tensor, see Refs. [27–30].

To treat the regularization properly, we employ dimensional regularization for a free scalar field in a curved field space with the spacetime dimension d, with the action

$$S = \int d^dx\sqrt{-g}\left[-\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{m^2}{2}\phi^2\right]. \tag{5.1}$$

The background metric is given by a FLRW-type metric in d dimensions,

$$ds^2 = -dt^2 + a(t)^2 d\mathbf{x}^2, \tag{5.2}$$

in which $a(t)$ is the scale factor.

Using the conformal time $d\tau = dt/a(t)$ and defining the canonically normalized field σ via $\sigma \equiv a^{\frac{d-2}{2}}\phi$, the action takes the following form:

$$S = \frac{1}{2}\int d^dx\left[\sigma'^2 + \left(\frac{(d-4)(d-2)}{4}\left(\frac{a'}{a}\right)^2 + \frac{d-2}{2}\frac{a''}{a} - m^2a^2\right)\sigma^2\right], \tag{5.3}$$

in which a prime denotes a derivative with respect to the conformal time.

Going to the Fourier space and expanding the field σ in terms of the creation and annihilation operators, we have

$$\sigma(x^\mu) = \int\frac{d^{d-1}\mathbf{k}}{(2\pi)^{(d-1)/2}}(\sigma_k(\tau)e^{i\mathbf{k}\cdot\mathbf{x}}a_{\mathbf{k}} + \sigma_k(\tau)^* e^{-i\mathbf{k}\cdot\mathbf{x}}a^\dagger_{\mathbf{k}}). \tag{5.4}$$

First, we have to solve the mode function $\sigma_k(\tau)$ in a dS spacetime. Considering a dS spacetime with the Hubble expansion rate H, we have $aH\tau = -1$, where we have taken $-\infty < \tau < 0$. Note that since there is no rolling scalar field (i.e., there is no inflaton field), the dS symmetry is exact and in particular $\dot{H} = 0$, yielding to $a' = Ha^2$ and $a'' = 2H^2a^3$. Then the equation of motion for the mode function $\sigma_{\mathbf{k}}(\tau)$ is given by

$$\sigma_k'' + \left[k^2 + \frac{m^2}{H^2\tau^2} - \frac{(d-2)^2}{2\tau^2}\right]\sigma_k = 0. \tag{5.5}$$

As a check, note that if we set $d = 4$, then the last term in the bracket above yields the well-known contribution $-2/\tau^2$, which is the hallmark of gravitational particle production in a dS background for the light scalar field perturbations.

If we expect Eq. (2.9) to hold true, then $H \sim m^2/M_P$, and therefore $H/m \sim m/M_P \ll 1$. Therefore, we can safely ignore the last term in Eq. (5.5) compared to the mass term. Imposing the Bunch-Davies (Minkowski) vacuum for the modes deep inside the dS horizon $k\tau \to -\infty$,

$$\sigma_k = \frac{1}{\sqrt{2k}} e^{-ik\tau}, \tag{5.6}$$

the solution of the mode function is given in terms of the Hankel function of the first type as follows:

$$\phi_k(t) = a^{\frac{2-d}{2}} \sigma_k = (-H\tau)^{\frac{d-1}{2}} \left(\frac{\pi}{4H}\right)^{\frac{1}{2}} e^{-\frac{\pi}{2}\nu} H_{i\nu}^{(1)}(-k\tau), \tag{5.7}$$

where

$$\nu \equiv \sqrt{\frac{m^2}{H^2} - 1} \simeq \frac{m}{H}. \tag{5.8}$$

Having calculated the mode function $\phi_k(\tau)$, we can calculate the vacuum energy density as follows:

$$\langle\rho\rangle = \frac{\mu^{4-d}}{2a^2} \int \frac{d^{d-1}\mathbf{k}}{(2\pi)^{d-1}} [|\phi_k'(\tau)|^2 + (k^2 + m^2 a^2)|\phi_k(\tau)|^2], \tag{5.9}$$

in which μ is a mass scale to properly take care of the energy dimension.

Defining the dimensionless variable $x \equiv -k\tau$, and separating the angular and the radial contributions of the integral, we obtain

$$\langle\rho\rangle = \frac{\pi\mu^{4-d}}{8(2\pi)^{d-1}} e^{-\pi\nu} H^d \left(\int d^{d-2}\Omega\right) \times \int_0^\infty x^d \left[\left|\frac{d}{dx} H_{i\nu}^{(1)}(x)\right|^2 + \left(1 + \frac{\nu^2}{x^2}\right)|H_{i\nu}^{(1)}(x)|^2\right]. \tag{5.10}$$

In deriving the above expression, we have neglected the time derivative of the prefactor $(-H\tau)^{(d-1)/2}$ in the mode function Eq. (5.7). One can check that this is a very good approximation in the limit of interest here where $\nu \gg 1$. The fractional errors induced in the final result are smaller by a factor $1/\nu^2 \ll 1$.

The integral over the azimuthal directions in Eq. (5.10) is given by [15]

$$\int d^{d-2}\Omega = \frac{2\pi^{(d-1)/2}}{\Gamma(\frac{d}{2} - \frac{1}{2})}, \tag{5.11}$$

where $\Gamma(x)$ is the gamma function.

The final step is to perform the following integral:

$$I \equiv \int_0^\infty x^d \left[\left|\frac{d}{dx} H_{i\nu}^{(1)}(x)\right|^2 + \left(1 + \frac{\nu^2}{x^2}\right)|H_{i\nu}^{(1)}(x)|^2\right]. \tag{5.12}$$

Fortunately, the above integral can be taken exactly, yielding[1]

$$I = \frac{(1 - d + 2i\nu)}{4\pi^{5/2}\sinh^2(\nu\pi)} \Gamma\left(-i\nu - \frac{1}{2} + \frac{d}{2}\right)\Gamma\left(i\nu + \frac{1}{2} + \frac{d}{2}\right) \times \Gamma\left(\frac{d}{2} - \frac{1}{2}\right)\Gamma\left(-\frac{d}{2}\right) \times \mathcal{C}, \tag{5.13}$$

where $\mathcal{C}$ is given by

$$\mathcal{C} \equiv 2\cosh(\nu\pi)\cos\left(\frac{\pi(2i\nu - d)}{2}\right)\cos\left(\frac{\pi(2i\nu + d)}{2}\right) - \cos\left(\frac{\pi d}{2}\right)\left[\cos\left(\frac{\pi(2i\nu + d)}{2}\right) + \cos\left(\frac{\pi(2i\nu - d)}{2}\right)\right]. \tag{5.14}$$

The singular limit of the integral I governing the UV divergence of the zero-point energy is controlled by the term $\Gamma(-\frac{d}{2})$ in Eq. (5.13).

To perform the dimensional regularization, we expand $d = 4 - \epsilon$ in $\langle\rho\rangle$, obtaining

$$\langle\rho\rangle = \frac{H^4}{2048\pi^2}(4\nu^2 + 1)(4\nu^2 + 9) \times \left[-\frac{4}{\epsilon} + 2\ln\left(\frac{H^2}{4\pi\mu^2}\right) + \Delta + \mathcal{O}(\epsilon)\right], \tag{5.15}$$

where

$$\Delta \equiv 2\Psi\left(\frac{5}{2} + i\nu\right) + \Psi\left(\frac{3}{2} - i\nu\right) + \frac{(8\gamma - 12)\nu^2 + 8i\nu + 18\gamma - 15}{(4\nu^2 + 9)}, \tag{5.16}$$

in which $\Psi(x)$ is the polygamma function and γ is the Euler-Mascheroni constant. One can check that for the large-ν limit (which is the case here), $\Delta \to 4\ln(\nu)$. Finally, note that the $\frac{1}{\epsilon}$ term in Eq. (5.15) originates from $\Gamma(-\frac{d}{2})$, which is divergent near $d = 4$.

[1]We use the Maple computational package to perform the integral.

Now, expanding $\langle\rho\rangle$ to leading order in ϵ and taking the limit $\nu \simeq m/H \gg 1$, we obtain

$$\langle\rho\rangle = -\frac{\nu^4 H^4}{64\pi^2}\left[\frac{2}{\epsilon} - \ln\left(\frac{H^2\nu^2}{4\pi\mu^2}\right) + \cdots\right], \qquad (5.17)$$

where the terms in $\cdots$ are on the order of ν^{-2} and smaller, which are subleading compared to the log term.

As expected, the divergent $1/\epsilon$ term in Eq. (5.17) represents the logarithmic divergence of $\langle\rho\rangle$ with the UV scales. After regularizing this divergent term and absorbing the factor 4π into our renormalization scale and noting that $\nu \simeq m/H$, we recover the desired formula

$$\langle\rho_v\rangle = \frac{m^4}{64\pi^2}\ln\left(\frac{m^2}{\mu^2}\right) + \mathcal{O}(m^2H^2), \qquad (5.18)$$

in agreement with Eq. (2.9) obtained in a flat background up to subleading corrections $\mathcal{O}(\nu^{-2})$. It is intriguing that the factor H drops out to leading order both in the prefactor and inside the log term in Eq. (5.17).

There are a few comments in order. First, while the leading order in Eq. (5.18) agrees with Eq. (2.9) in a flat background, we see that there are subleading corrections as well. These subleading corrections contain various even powers of H, such as m^2H^2 and H^4. This is because the dS background has a curvature radius of order H^{-2}, which should enter the analysis. This is not inconsistent with our starting argument based on the equivalence principle. More specifically, the equivalence principle can pin down the local term, which is independent of the curvature radius of the spacetime. However, the subleading terms containing powers of H involve the curvature radius of the spacetime, and as such are inaccessible via the equivalence principle. Second, for the massless term, the first two leading terms vanish, and we end up with the H^4 contribution in $\langle\rho_v\rangle$. This is again a unique feature of the curved background, which is not accessible for flat spacetime. Third, as shown in Ref. [23], for the massless field with a quartic coupling λ, the contribution $\langle\rho_v + P_v\rangle \propto \lambda H^4$, indicating the violation of the weak energy condition. This is again a nontrivial property of the vacuum stress energy tensor in a curved background.

To complete this discussion, we also calculate the density contrast $\delta\rho/\langle\rho\rangle$ in this background. The analysis is similar to that performed in Sec. IV A. As before, let us define the three contributions ρ_i to energy density as in Eq. (4.1), in which the specific forms of $\langle\rho_1\rangle$, $\langle\rho_2\rangle$, and $\langle\rho_3\rangle$ are given in Eq. (5.9). In addition, by performing the dimensional regularization for each component of $\langle\rho_i\rangle$, one confirms that the relations in Eq. (4.5) are valid here as well with subleading corrections $\mathcal{O}(\nu^{-2})$. Furthermore, by performing the same steps as in Sec. IV A, one can easily check that $\langle\rho_3^2\rangle$ and $\langle\rho_2^2\rangle$ still satisfy the relations (4.7) and (4.8), respectively. However, there are new contributions in $\langle\rho_1^2\rangle$ due to the fact that $\phi(x^\mu) = \sigma(x^\mu)/a(\tau)$ and $\dot{a} = aH$. One can check that the new contributions in $\langle\rho_1^2\rangle$ compared to Eq. (4.7) are suppressed by the factor $\frac{H^2}{m^2} \simeq \nu^{-2} \ll 1$, so $\langle\rho_1^2\rangle = \langle\rho_1\rangle^2(3 + \mathcal{O}(\nu^{-2}))$.

Correspondingly, combining all combinations in $\delta\rho$ as in Eq. (4.14), the density contrast with a dS background is obtained to be

$$\frac{\delta\rho_v}{\langle\rho_v\rangle} = \pm\sqrt{10} + \mathcal{O}(\nu^{-2}). \qquad (5.19)$$

It is both interesting and reassuring that the expressions for $\langle\rho_v\rangle$ and $\frac{\delta\rho_v}{\langle\rho_v\rangle}$ match the corresponding formula in the flat background up to subleading corrections $\mathcal{O}(\nu^{-2})$.

Similarly to the case of the flat background, one can also calculate the correlation length associated with the zero-point energy density fluctuations in the dS background. The analysis is somewhat complicated due to technicalities associated with the Hankel functions. Using the approximate relations of the Hankel functions, we have verified that the correlation length of the zero-point fluctuation in dS backgrounds is indeed $\xi \sim 1/m$ as with a flat background. This is consistent with our notion that the intrinsic properties of the zero-point energy density fluctuation such as its amplitude and the correlation length are local phenomena and are insensitive to the large-scale structure of spacetime.

Now equipped with the exact mode function in a dS background, we can calculate the accumulated energy density on the super-Compton scale $\delta\rho_C$ and compare the result with its flat counterpart Eq. (3.3). Simply setting $d = 4$ in Eq. (5.10) and $\nu \simeq m/H$, we obtain

$$\delta\rho_C \simeq \frac{H^4}{16\pi}\int_0^\nu x^4\left[\left|\frac{d}{dx}H_{i\nu}^{(1)}(x)\right|^2 + \left(1 + \frac{\nu^2}{x^2}\right)|H_{i\nu}^{(1)}(x)|^2\right],$$
$$\simeq \frac{H^4}{16\pi}\frac{\nu^4}{\pi} \simeq \frac{m^4}{16\pi^2}. \qquad (5.20)$$

Interestingly, we see that the above value of $\delta\rho_C$ is in good agreement with its flat counterpart given in Eq. (3.3).

As we have mentioned above, the correlation length of zero-point energy density is $\xi \sim 1/m$, so on length scales far beyond this correlation length, the distribution of the zero-point energy is expected to be uncorrelated. On the other hand, from Eqs. (5.18) and (5.20), we see that the accumulated energy density on super-Compton scales (i.e., for modes with wavelength larger than ξ) is comparable to the total zero-point energy, $\delta\rho_C/\langle\rho_v\rangle \sim 1$. This again confirms that the zero-point energy of the heavy fields cannot be used to cover the entire spacetime as a background energy density. In other words, one needs to impose $\langle\rho_v\rangle \ll \rho_T \simeq \rho_F$ in order to have a sensible cosmological background.

VI. COSMOLOGICAL IMPLICATIONS

The results from the previous three sections exclude the heavy fields with $\langle\rho_v\rangle \gg \rho_F$ from contributing to the observed dark energy today. Originally, we came to this conclusion based on the fact that one requires an enormous number $N_{\rm patches} \gg 1$ of incoherent small patches of size $H_{(m)}^{-1}$ to cover the FLRW horizon today. This conclusion was specifically confirmed in Sec. IV by looking at the density contrast $|\frac{\delta\rho_v}{\langle\rho_v\rangle}| \sim 1$. If the heavy fields do not contribute to the current dark energy (cosmological constant), then a natural question to ask now is what roles in cosmology their vacuum energy density plays.

To answer the above question, we again employ the picture based on the scale of dS horizons as presented in Fig. 1. We consider the early stage in cosmic expansion history when $\rho_F \gg \rho_v$, so $H_F^{-1} \ll H_{(m)}^{-1}$, as in the left panel of Fig. 1. In this situation, the observable Universe is inside the horizon of the zero-point energy, but at the same time, since $\rho_v \ll \rho_F$, the contribution of ρ_v in expansion rate is too small to be important. As time proceeds and ρ_F decreases further, we approach the epoch when $\rho_F \sim \rho_v$, and the zero-point energy becomes relevant; we notice the effects of the vacuum energy. As time goes by further, ρ_F falls off rapidly, so we end up with the situation where $\rho_v \gg \rho_F$, as in the right panel of Fig. 1. In this case, the regions filled with the zero-point energy develop strong inhomogeneities while falling into the FLRW Hubble horizon. The timescale for the dS patches to create inhomogeneities is about $2/H_{(m)}$, as it takes a period of about $1/H_{(m)}$ for each dS horizon to enter the FLRW horizon. Once the FLRW horizon grows large enough (or equivalently when $\rho_F \ll \rho_v$), more and more dS patches enter the FLRW Hubble radius. The variance condition (4.16) indicates that the mass inside the dS patches inside the FLRW Hubble radius may collapse to form black holes. However, as we mentioned at the end of Sec. IV A, the formation of black holes from the collapse of local AdS regions created from inhomogeneities is far from obvious, as the interplay between the turbulent dynamics of local AdS and dS inhomogeneities is complicated. But if the masses inside these patches collapse into black holes, their contribution to the background cosmological dynamics is expected to be in the form of matter. This indicates that the resulting mass may behave like dark matter or the seeds of dark matter.

Despite the above qualitative discussions, at this stage we cannot be more specific about the subsequent contribution of the zero-point energy of heavy fields after the time when $H_F^{-1} > H_{(m)}^{-1}$. Different scenarios are possible, which we study below using a phenomenological fluid description.

A. Phenomenological fluid description

As mentioned above, we do not know the detailed mechanism which governs the contribution of the zero-point energy of heavy fields after the time when $H_F^{-1} > H_{(m)}^{-1}$. Here, we consider a phenomenological approach and treat the effects of heavy fields like a fluid with an undetermined equation of state.

To be more specific, let us denote the energy density of massive fields after the time when ρ_F is diluted enough and $\rho_F < \rho_m$ by $\tilde{\rho}_m$. We also assume that each component of effective energy density evolves separately by its own energy conservation: $\dot{\rho}_F + 3H(\rho_F + p_F) = 0$ and $\dot{\tilde{\rho}}_m + 3H(\tilde{\rho}_m + \tilde{p}_m) = 0$, in which $\tilde{p}_m$ represents the effective pressure of the new matterlike energy component. Defining the equation of states via $p_F \equiv w_F\rho_F$ and $\tilde{p}_m \equiv w_m\tilde{\rho}_m$, the total energy density now is $\rho_T = \rho_F + \tilde{\rho}_m$, yielding

$$\rho_T(t) = \left[\rho_m\left(\frac{a(t)}{a(t_m)}\right)^{-3(1+w_m)} + \rho_F(t_m)\left(\frac{a(t)}{a(t_m)}\right)^{-3(1+w_F)}\right], \tag{6.1}$$

in which t_m is the time when $\rho_F(t_m) \sim \rho_m$—i.e., when the two Hubble radii H_F^{-1} and $H_{(m)}^{-1}$ nearly coincide. Considering the numerical uncertainties, we set $\tilde{\rho}_m(t_m) \equiv \kappa'\rho_F(t_m)$ in which κ' is a numerical constant which may depend on m. Then, the total energy density is given by

$$\rho_T(t) = \rho_F(t_m)\left(\frac{a(t)}{a(t_m)}\right)^{-3(1+w_F)}\left[1 + \kappa'\left(\frac{a(t)}{a(t_m)}\right)^{3(w_F-w_m)}\right]. \tag{6.2}$$

The above is a phenomenological description of the contribution of the heavy fields after the time when $\rho_F < \rho_m$. We have introduced the effective equation-of-state parameter ω_m to capture the uncertain behavior of the zero-point energy associated with the massive fields. If strong inhomogeneities with $\delta\rho_v/\rho_v \sim 1$ yield to black hole formation, then $\omega_m = 0$ and $\tilde{\rho}_m$ can play the roles of dark matter. Of course, this is good news for this proposal, in which case the nature of dark matter and dark energy is unified, with both being generated dynamically from the zero-point energy. However, the difficulty with $\omega_m \simeq 0$ is that at early times in cosmic history, the produced dark matter from the zero-point energy of electron and heavier fields rapidly dominates over the radiation energy density, long before the time of matter-radiation equality at the temperature $T_{\rm eq} \sim 3$ eV, altering the hot big bang cosmology in various unwanted ways. One way out may be to consider a situation in which the phenomenological parameter κ' in Eq. (6.2) is (exponentially) small, so it will take a long time for the seed of dark matter to take over the radiation energy density. On the other hand, if $\omega_m \simeq \frac{1}{3}$, then $\tilde{\rho}_m$ may behave like dark radiation. Considering the uncertainties involved in the process, ω_m may take different values for various fields, so $\tilde{\rho}_m$ may behave differently during cosmic history, and we may have dark radiation, dark matter, or stiff fluid with $\omega_m \lesssim 1$.

Another possibility is that the effective fluid associated with the zero-point energy is such that $\omega_m = \omega_F$, so $\tilde{\rho}_m(t) \propto \rho_F(t)$—i.e.,

$$\tilde{\rho}_{(m)}(t) = 3\kappa M_P^2 H_F^2 = \kappa \rho_F, \qquad (6.3)$$

in which κ is a numerical constant. This corresponds to a tracking scenario in which the energy density of the heavy fields at each stage in cosmic history tracks the background FLRW energy density. During the radiation-dominated era, their energy densities follow that of a radiation energy density, while during the matter-dominated era, they behave like matter. This is an interesting case, so below we concentrate on this scenario more closely.

The tracking scenario has interesting implications for the current energy density of dark matter. To see this, let us assume that $\rho_F = \rho_\gamma + \rho_b$, in which ρ_γ and ρ_b represent the energy density of radiation and baryons. Specifically, we assume that there is no conventional (say WIMP-like) dark matter in our setup. Let us define the time of matter-radiation equality at the redshift $z_{\rm eq}$ as when $\rho_b(z_{\rm eq}) = \rho_\gamma(z_{\rm eq})$. Now, observe that what we mean by dark matter in the ΛCDM model is the zero-point energy density given by the heavy field in Eq. (6.3). Specifically, the current would-be dark matter energy density is actually $\rho_{\rm DM}^{(0)} = \tilde{\rho}_m(t_0) = \kappa\rho_F^{(0)} = \kappa\rho_b^{(0)} + \kappa\rho_\gamma^{(0)}$. Taking into account that the energy densities of radiation and baryons fall off like a^{-4} and a^{-3}, respectively, we obtain

$$1 + z_{\rm eq} = \frac{1}{1+\kappa}\left(\frac{\Omega_M^{(0)}}{\Omega_\gamma^{(0)}} - \kappa\right), \qquad (6.4)$$

in which $\Omega_M^{(0)} \sim 0.31$ ($\Omega_\gamma^{(0)} \sim 10^{-4}$) is the fractional total matter (radiation) energy density today as inferred from the ΛCDM model. Now, defining $z_{\rm eq}^{\Lambda{\rm CDM}}$ as the time of matter-radiation equality in the ΛCDM model and assuming $\kappa \sim \mathcal{O}(1)$, we obtain

$$z_{\rm eq} \simeq \frac{z_{\rm eq}^{\Lambda{\rm CDM}}}{1+\kappa}. \qquad (6.5)$$

This means that in the current setup, the time of matter-radiation equality is shifted toward a later epoch in cosmic history. This is because we have assumed that the time of matter-radiation equality happens when $\rho_b(z_{\rm eq}) = \rho_\gamma(z_{\rm eq})$. Taking $\kappa \lesssim 1$, we may have $z_{\rm eq} \sim {\rm few} \times 1000$. However, note that shifting $z_{\rm eq}$ close to the time of the cosmic microwave background (CMB) last scattering can be dangerous for CMB observations. But we also note that shifting the time of matter-radiation equality towards a later epoch can play an important role in solving the H_0 tension problem.

The above conclusion about the roles of the heavy fields may seem in conflict with our starting point that the vacuum energy should be locally Lorentz invariant and $\langle p\rangle = -\langle\rho\rangle$, which is trivially violated for both radiation and matter. The answer is that the relation $\langle p\rangle = -\langle\rho\rangle$ is enforced locally, say deep inside each dS patch. However, what we have in Eqs. (6.3) or (6.2) is the collective energy density of a large number of patches of zero-point energy inside the FLRW horizon, which may not be in conflict with the local requirement $\langle p\rangle = -\langle\rho\rangle$ deep inside each dS patch.

B. Selection rules

Based on the results from the previous sections, we end up with an interesting "selection rule," which works as follows: At each stage in the cosmic epoch, only a field with $H_{(m)} \sim H_F$ and energy density $\langle\rho_v\rangle \sim \rho_F$ can be relevant as the source of dark energy. Fields which are much lighter ($H_{(m)} \ll H_F$) are irrelevant in cosmic expansion. This is simply because their contributions in $\langle\rho_v\rangle$ as given in Eq. (2.9) are exceedingly smaller than ρ_F, so as to be unnoticeable. Finally, heavy fields with $H_{(m)} \gg H_F$ cannot be the source of dark energy, while they may form the seeds of dark matter from the start or track the background FLRW energy density ρ_F and be the source of dark matter after the time of matter-radiation equality. Alternatively, they may behave like a stiff fluid with $\omega_m \lesssim 1$, in which case their contributions in background energy density are diluted faster than radiation after the time $t > H_{(m)}^{-1}$.

The above selection rule can easily address both the old and the new cosmological constant problems. Recall that the old cosmological constant problem is why ρ_v is not large—i.e., why it is as small as $(10^{-3}\ {\rm eV})^4$. The new cosmological constant problem is why the vacuum energy density becomes comparable to the matter-energy density at the current stage of the cosmic history. The resolution is that there is a field in the SM field content, the lightest neutrino,[2] with the vacuum zero-point energy $\rho_{(\nu)} \sim m_\nu^4$ which happens to have a mass at the same order as $\rho_{F0}^{1/4} \simeq \rho_c^{1/4}$, in which $\rho_c \sim (10^{-3}\ {\rm eV})^4$ is the critical energy density. The entire FLRW Universe is currently within a single patch of the lightest neutrino with the horizon radius $H_{(m_\nu)}^{-1}$. The current dark energy survives in the future for another period of roughly $1/H_{(m_\nu)} \sim 1/H_0 \sim 10^{10}$ years before multiple dS patches of the lightest neutrino enter the FLRW horizon. Of course, this conclusion about the future dynamics of the Universe is based on the assumption that

[2]While the differences in neutrino mass squared are known, the absolute masses and the mass of the lightest neutrino are not exactly known; see Ref. [31]. Here we adopt the simple assumption that the lightest neutrino has a mass on the order of 10^{-2} eV [32,33].

there is no light field in the beyond-SM sector (like a light axion) with a mass lighter than 10^{-3} eV to contribute to the future dark energy. We comment that the relations between the zero-point energy of neutrinos and the observed dark energy (cosmological constant) have been studied in different contexts in the past in Refs. [34–38].

An immediate corollary of our selection rule is that at early stages in cosmic history (i.e., for all times prior to the time of matter-radiation equality), only fields with masses on the order of the FLRW photon temperature could contribute to dark energy in that epoch. This is because the energy density of the FLRW background during the radiation-dominated era is related to the photon temperature via

$$\rho_F = \frac{\pi^2}{30} g_* T^4, \tag{6.6}$$

in which g_* is the effective relativistic degree of freedom at the temperature T. Now, comparing the above equation with Eq. (2.9) for the vacuum energy density, and noting that for the field of interest $H_{(m)} \sim H_F$, we conclude that

$$m \sim T. \tag{6.7}$$

If a field is much lighter than T, then its vacuum energy density $\propto m^4$ is much smaller than T^4 and cannot be important in cosmic energy density. On the other hand, if a field is much heavier than T, then based on the arguments mentioned in previous sections, the dS horizon radius associated with its vacuum energy density is much smaller than the Hubble radius H_F^{-1}, and they collapse into a FLRW horizon behaving like dark radiation, dark matter, or a stiff fluid. Note that while Eq. (6.7) works well for the time prior to matter-radiation equality, it also gives a reasonable estimation for the relation between the mass of the lightest neutrino and the energy scale of the current dark energy. To see this, note that $T_{\gamma_0} \sim 10^{-4}$ eV which is only about 2 orders of magnitude below the expected mass of the lightest neutrino. Of course, this is easily understandable, since after the time of matter-radiation equality, the energy density of radiation falls off by an additional factor of $1/a$ compared to the matter-energy density. This is why we had to use the critical mass density ρ_c instead of $T_{\gamma 0}^4$ to correctly estimate the magnitude of the current dark energy density.

Since only fields with $m \neq 0$ contribute to the vacuum energy, we conclude that before the electroweak symmetry breaking at the temperature $T_{\mathrm{EW}} \sim 160$ GeV, all SM fields were massless, and they did not contribute to dark energy. Immediately after the electroweak phase transition, the relevant fields are the top quark $m_t \simeq 170$ GeV, the Higgs field $m_H \sim 125$ GeV, and the three gauge bosons $m_Z \simeq 91$ GeV, $m_{W^\pm} \simeq 80$ GeV. The top quark and the Higgs field are somewhat heavy, so they are at the threshold of being able to contribute to dark energy at that time, while the three gauge bosons have good chances to play important roles as the source of dark energy. Of course, as the temperature of the Universe falls below the mass of vector bosons, their contribution in dark energy expires, and they may contribute to dark radiation, dark matter, or behaving like a stiff fluid.

This story is repeated when the background temperature approaches the mass of other fundamental particles such as $m_\tau \simeq 1.7$ GeV, $m_\mu \simeq 105$ MeV, and $m_e \simeq 0.5$ MeV. Especially important is the time when the temperature is around the MeV scale when big bang nucleosynthesis (BBN) is at work. Changing the Hubble expansion rate of the Universe at this epoch, either by a contribution from the dark energy of the electron field or its contributions in the form of dark radiation, can affect the dynamics of BBN. This issue requires careful investigation, which can provide a nontrivial check for the consistency of the whole picture presented here. Another important time to look for is when the heavier neutrino fields (i.e., not the lightest neutrino) with mass $m \sim 0.1$ eV become relevant to contribute to dark energy. This occurs around $T \sim 10^{-2}$ eV and redshifts of $z \sim 10^2$.

Based on the above discussions, we have transient periods of dark energy for a few e-folds any time in cosmic history when the condition $m \sim T$ is met, during which the Hubble expansion rate stays nearly constant and then falls off as in conventional big bang cosmology. These can happen in cosmic history both after the surface of last scattering (for heavier neutrinos) at redshift $z \sim 10^2$, and before the surface of last scattering (for e, μ, and τ fields) at much higher redshifts. The curious conclusion is that since the energy density is nearly constant for a few e-folds, the value of the Hubble expansion rate today (and at the time of last scattering) would be larger than what the standard ΛCDM model predicts. Multiple transient periods of dark energy both at early and intermediate cosmic expansion history are a nontrivial prediction of the model that have a good potential to solve the H_0 tension problem [39–43]. In this view, our proposal can incorporate the early dark energy (EDE) mechanism [44,45] which is proposed to solve the H_0 tension problem.

VII. SUMMARY AND DISCUSSIONS

In this work, we have revisited the quantum cosmological constant problem. As already known in the literature, the conventional approach of imposing a hard UV momentum cutoff violates the underlying Lorentz invariance of the vacuum. Employing a regularization scheme which respects the underlying Lorentz invariance of the theory, such as the dimensional regularization method, one obtains that the energy density of each field scales like $\rho_v \sim m^4$. Furthermore, the condition of Lorentz invariance of the vacuum, $\langle p \rangle = -\langle \rho \rangle$, is manifest. Some noticeable conclusions are that the massless fields such as gravitons, photons, and gluons do not contribute to the vacuum energy density. Furthermore, $\langle \rho \rangle$ runs logarithmically with the renormalization scale, μ and depending on the ratio m/μ, the value of the cosmological constant can be either positive (dS) or negative (AdS).

We have highlighted that the dS horizon associated with the zero-point energy plays important roles. For a field with mass m, the associated dS space has a horizon radius $H_{(m)}^{-1} \sim M_P/m^2$, which was not taken into account in previous studies of the cosmological constant problems. In the conventional approach, it is naively assumed that the entire observable FLRW patch is covered by the vacuum energy density $\langle \rho \rangle \sim m^4$ irrespective of the vast hierarchy between the two Hubble radii H_F^{-1} and $H_{(m)}^{-1}$. For the heavy fields with the large hierarchy $H_{(m)} \gg H_F$, this requires as many as $(\frac{H_{(m)}}{H_F})^3 \gg 1$ independent patches of zero-point energy to fill the FLRW patch. However, since the patches of zero-point energy are created quantum mechanically, they are uncorrelated. Therefore, one expects that a space filled with as many as $(\frac{H_{(m)}}{H_F})^3$ uncorrelated patches of zero-point energy will be highly chaotic and may even collapse to form black holes. To support this picture, we have calculated the variance of the zero-point energy for both the real scalar field and the Dirac fermion field. In both cases, we have found that $|\delta\rho_v/\langle \rho_v \rangle| \sim 1$, supporting the above intuitive picture. Furthermore, the correlation length of the fluctuations of the zero-point energy is calculated to be on the order of $\xi \sim m^{-1}$. This results in the vast hierarchy $\xi \ll H_{(m)}^{-1} \ll H_F^{-1}$. It is shown that the accumulated energy density on super-Compton scales—i.e., scales which are far beyond the correlation length—is comparable to the would-be background vacuum energy density. This conclusion, in conjunction with the above hierarchies among the three scales ξ, $H_{(m)}^{-1}$, and H_F^{-1}, is another sign that one cannot use the vacuum energy density $\langle \rho_v \rangle$ as the dominant background energy density. To have a viable cosmological background, we have to impose $\langle \rho_v \rangle \lesssim \rho_F$ at each stage in cosmic expansion history in which ρ_F is the nonvacuum energy density (i.e., the classical energy density) of the background FLRW cosmology.

We have established a "selection rule" stating that at each stage in cosmic expansion history, only fields with a mass m satisfying the condition $H_{(m)} \sim H_F$ can contribute to dark energy in that epoch. Specifically, only fields with mass on the order of the background photon temperature, $m \sim T$, are relevant as the source for dark energy. Lighter fields carry too little energy ($m^4 \ll T^4$) to be important in the cosmic energy budget. On the other hand, heavy fields with $H_{(m)} \gg H_F$ (or $m \gg T$) cannot be the source of dark energy, while they can be the seeds of dark matter, dark radiation, or a stiff fluid with equation of state $\frac{1}{3} < \omega_m \leq 1$.

We have speculated that both the old and new cosmological constant problems can be addressed readily. Specifically, these puzzles are solved by noting that there is a field in the SM spectrum, the lightest neutrino, which happens to have a mass on the order of $\rho_c^{1/4} \sim 10^{-2}$ eV. Another important implication is that at various stages in cosmic history, the energy density experiences loitering stages of dark energy in which the Hubble expansion rate stays nearly constant for about one or two e-folds in the expansion rate, and then the energy density falls off as in the ΛCDM model. As a consequence, the Hubble expansion rate at the time of CMB last scattering is expected to be higher than what is inferred from ΛCDM model. Currently, we are at the last transient stage associated with the lightest neutrino. However, prior to this stage there were two transient phases of dark energy happening at around the mass scale of the heavier neutrinos at $T \sim 10^{-2}$ eV, corresponding to the redshift $z \sim 10^2$. This period will be some time after the time of CMB decoupling. Multiple transient phases of dark energy, both before and after the surface of last scattering, yield to higher values of the current Hubble expansion rate compared to what is inferred from the ΛCDM model. In this view, our proposal can incorporate the EDE proposal and has the potential to solve the H_0 tension problem. Another interesting prediction of the model (as mentioned above) is that there may be no dark matter. All matter is in the form of known baryonic matter, while the role of dark matter is played by the heavy fields which may track the background FLRW energy density. As such, the time of matter-radiation equality is shifted towards a later time in cosmic history.

There are a number of open questions in this study. The immediate one is how the collapse for the patches of zero-point energy associated with heavy fields can happen. We provided only a rough phenomenological picture of this mechanism. This is an important question which controls whether or not we have the tracking energy or a fluid with the equation of state $\omega_m \sim 0$. Another open question is the natural value of the renormalization scale μ. Note that the vacuum energy runs logarithmically with μ. Furthermore, depending on the statistics of the field (i.e., being a boson or a fermion) and on the ratio m/μ, the vacuum energy can be either positive (dS) or negative (AdS). To solve the cosmological constant problems, we have implicitly assumed that the resulting vacuum energy is positive, so taking the neutrinos as the relevant fields for this purpose, we require $\mu >$ eV. If we take μ to be the scale of electroweak symmetry breaking, say $\mu \gtrsim 10^2$ GeV, then all fermionic (bosonic) fields add positive (negative) contributions to the vacuum energy. If the energy density is AdS type, then there may be a falloff period (instead of a loitering phase) in the Hubble expansion rate at the corresponding stage in cosmic history. This will modify the inferred value of the Hubble expansion rate at the time of CMB last scattering. Therefore, the dS and AdS contributions for all SM degrees of freedom from the mass scale 10^2 GeV after the electroweak symmetry breaking down to the neutrino scale should be carefully included to see how dark energy and dark matter behave at each stage in cosmic history.

Finally, it is important to examine the implications of the scenario for other aspects of early-Universe cosmology. Specifically, the interplay between this proposal and cosmic inflation is an open and interesting question. However, we comment that our proposal is not in conflict with an early stage of inflation. More specifically, during the slow-roll inflation driven by an inflaton field of mass m and the potential V, the dominant energy is given by the inflaton classical potential V_c, while the vacuum zero-point energy is on the order of m^4. Since for a rolling inflaton $V_c \simeq 3M_P^2H^2 \gg m^4$, the induced energy density from the zero-point fluctuations during inflation is negligible. Therefore, the quantum cosmological constant is not an issue during inflation, and a large enough classical vacuum energy during inflation is stable against quantum zero-point corrections.

ACKNOWLEDGMENTS

We would like to thank Xingang Chen, Jerome Martin, Shinji Mukohyama, Mohammad Hossein Namjoo, and Misao Sasaki for insightful comments about the draft. We also thank Hossein Moshafi for the preliminary investigations of the predictions of the proposal for the H_0 tension problem and Omid Sameie for useful discussions.

Note added.—We have recently become aware[3] of a series of papers in Refs. [46–49] that question the assumption of the homogeneity of the spacetime in the presence of the vacuum zero-point energy. It is concluded, among other things, that a uniform cosmological constant cannot cover the large scale spacetime, and the local spacetime is very inhomogeneous as in Wheeler's spacetime foam. Using the semiclassical GR, they have studied the cosmological implications of their proposal such as in resolving the old cosmological constant problem. Our analysis in Sec. IV for the fluctuations in the zero-point energy density and large density contrast is conceptually in line with their investigation.

[3]We are grateful to Jerome Martin for bringing to our attention the papers by Unruh and his collaborators [46–49].

[1] S. Weinberg, *Cosmology* (Oxford University Press, New York, 2008).
[2] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **641**, A6 (2020); **652**, C4(E) (2021).
[3] V. Sahni and A. A. Starobinsky, Int. J. Mod. Phys. D **09**, 373 (2000).
[4] P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75**, 559 (2003).
[5] E. J. Copeland, M. Sami, and S. Tsujikawa, Int. J. Mod. Phys. D **15**, 1753 (2006).
[6] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
[7] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, Phys. Rev. Lett. **85**, 4438 (2000).
[8] I. Zlatev, L. M. Wang, and P. J. Steinhardt, Phys. Rev. Lett. **82**, 896 (1999).
[9] N. Arkani-Hamed, S. Dimopoulos, N. Kaloper, and R. Sundrum, Phys. Lett. B **480**, 193 (2000).
[10] S. Kachru, M. B. Schulz, and E. Silverstein, Phys. Rev. D **62**, 045021 (2000).
[11] C. Csaki, J. Erlich, and C. Grojean, Nucl. Phys. **B604**, 312 (2001).
[12] J. M. Cline and H. Firouzjahi, Phys. Lett. B **514**, 205 (2001).
[13] J. M. Cline and H. Firouzjahi, Phys. Rev. D **65**, 043501 (2002).
[14] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).
[15] J. Martin, C.R. Phys. **13**, 566 (2012).
[16] S. Perlmutter *et al.* (Supernova Cosmology Project Collaboration), Astrophys. J. **517**, 565 (1999).
[17] A. G. Riess *et al.* (Supernova Search Team Collaboration), Astron. J. **116**, 1009 (1998).
[18] E. K. Akhmedov, arXiv:hep-th/0204048.
[19] J. F. Koksma and T. Prokopec, arXiv:1105.6296.
[20] G. Ossola and A. Sirlin, Eur. Phys. J. C **31**, 165 (2003).
[21] M. Visser, Particles **1**, 138 (2018).
[22] S. Weinberg, *The Quantum Theory of Fields* Vol. 1: Foundations (Cambridge University Press, 1995), 10.1017/CBO9781139644167.
[23] V. K. Onemli and R. P. Woodard, Classical Quantum Gravity **19**, 4607 (2002).
[24] C. Moreno-Pulido and J. Sola, Eur. Phys. J. C **80**, 692 (2020).
[25] C. Moreno-Pulido and J. S. Peracaula, arXiv:2110.08070.
[26] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1982).
[27] J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 3224 (1976).
[28] T. S. Bunch and P. C. W. Davies, Proc. R. Soc. A **360**, 117 (1978).
[29] T. S. Bunch and P. C. W. Davies, J. Phys. A **11**, 1315 (1978).
[30] P. C. W. Davies, S. A. Fulling, S. M. Christensen, and T. S. Bunch, Ann. Phys. (N.Y.) **109**, 108 (1977).
[31] P. A. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).
[32] A. Loureiro *et al.*, Phys. Rev. Lett. **123**, 081301 (2019).

[33] P. Stöcker *et al.*, Phys. Rev. D **103**, 123508 (2021).
[34] I. L. Shapiro and J. Sola, J. High Energy Phys. 02 (2002) 006.
[35] I. L. Shapiro and J. Sola, Phys. Lett. B **475**, 236 (2000).
[36] D. Glavan and T. Prokopec, arXiv:1504.00842.
[37] L. E. Ibanez, V. Martin-Lozano, and I. Valenzuela, J. High Energy Phys. 11 (2017) 066.
[38] M. Blasone, A. Capolupo, S. Capozziello, S. Carloni, and G. Vitiello, Braz. J. Phys. **35**, 455 (2005).
[39] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, Astrophys. J. **876**, 85 (2019).
[40] A. G. Riess *et al.*, Astrophys. J. Lett. **934**, L7 (2022).
[41] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, Classical Quantum Gravity **38**, 153001 (2021).
[42] M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, and G. Lambiase, Astrophys. J. **912**, 150 (2021).
[43] M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, G. Lambiase, M. Bogdan, and S. Ugale, Galaxies **10**, 24 (2022).
[44] V. Poulin, T. L. Smith, T. Karwal, and M. Kamionkowski, Phys. Rev. Lett. **122**, 221301 (2019).
[45] T. L. Smith, V. Poulin, and M. A. Amin, Phys. Rev. D **101**, 063523 (2020).
[46] Q. Wang, Z. Zhu, and W. G. Unruh, Phys. Rev. D **95**, 103504 (2017).
[47] S. S. Cree, T. M. Davis, T. C. Ralph, Q. Wang, Z. Zhu, and W. G. Unruh, Phys. Rev. D **98**, 063506 (2018).
[48] Q. Wang and W. G. Unruh, Phys. Rev. D **102**, 023537 (2020).
[49] Q. Wang, Phys. Rev. Lett. **125**, 051301 (2020).