

## Inflation and gauge field holonomy

Andrey O. Barvinsky<sup>1</sup> and Ariel R. Zhitnitsky<sup>2</sup>

<sup>1</sup>*Theory Department, Lebedev Physics Institute, Leninsky Prospekt 53, Moscow 119991, Russia*

<sup>2</sup>*Department of Physics & Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*



(Received 10 October 2017; published 8 August 2018)

We discuss a novel scenario for early cosmology, when the inflationary quasi-de Sitter phase dynamically originates from the initial quantum state represented by the microcanonical density matrix. This genuine quantum effect occurs as a result of the dynamics of the topologically nontrivial sectors in a (conjectured) strongly coupled QCD-like gauge theory in expanding universe. The crucial element of our proposal is the presence in our framework of a nontrivial  $\mathbb{S}^1$  which plays the dual role in construction: it defines the periodic gravitational instanton (on the gravity side) and it also defines a nontrivial gauge holonomy (on the gauge side) generating the vacuum energy. The effect is global in nature and cannot be formulated in terms of a gradient expansion in an effective local field theory. We also discuss a graceful exit from holonomy inflation due to the helical instability. The number of  $e$ -folds in the holonomy inflation framework is determined by the gauge coupling constant at the moment of inflation, and estimated as  $N_{\text{infl}} \sim \alpha^{-2}(H_0) \sim 10^2$ . We also comment on the relation of our framework with the no-boundary and tunneling cosmological proposals and their recent criticism.

DOI: [10.1103/PhysRevD.98.045008](https://doi.org/10.1103/PhysRevD.98.045008)

### I. INTRODUCTION

The inflationary scenario is widely recognized as one of the most successful candidates for the description of the early Universe leading to its observable large scale structure. The majority of effective and fundamental models of this scenario are based on the assumption that matter energy density driving the quasiexponential expansion of the Universe during inflation stage is generated by local field-theoretical degrees of freedom (d.o.f.), like a scalaron field in the Starobinsky  $R^2$ -gravity [1] or a scalar field inflaton  $\Phi(x)$  with its potential  $V(\Phi)$  in chaotic and other inflationary models [2]; see textbook [3] for a general overview.

However, it is also very possible that the generation of this type of uniformly distributed energy might not be associated with any local propagating particles. Instead, it might be related to some global characteristics (such as holonomy) or topological d.o.f. which cannot be expressed in terms of any local fields such as inflaton  $\Phi$ . Examples, in particular, include the global d.o.f. arising in the context of the recently suggested generalized unimodular gravity theory [4]. Another example is represented by a strongly

coupled QCD-like gauge theory when the vacuum energy is generated by some nontrivial topological features of the gauge systems [5–7].

Here we apply the ideas when the vacuum energy is induced by the topologically nontrivial holonomy [5–7] to the mechanism of inflation in the early quantum Universe driven by the thermal states [8,9]. This model, which incorporates the idea of the microcanonical density matrix as the initial quantum state of the Universe [10], is conceptually very attractive because of the minimum set of assumptions underlying it and, moreover, because of a mechanism restricting the cosmological ensemble to sub-Planckian energy domain and avoiding the infrared catastrophe inherent in the no-boundary wave function [11]. Furthermore, this thermally driven cosmology [8,9] can serve as initial conditions for the observationally consistent models of  $R^2$  and Higgs inflation; see original paper [12] and the recent development [13,14] based on induced gravity aspects of the theory.

As we argue below our construction, which can be viewed as a synthesis of two naively unrelated ideas [8–10] and [5–7] correspondingly, shows a number of very desirable and remarkable features. On the gravity side [8–10] the nontrivial element of the construction is represented by the Euclidean spacetime with a time compactified to a circle  $\mathbb{S}^1$ . On the gauge field theory side [5–7] the same  $\mathbb{S}^1$  plays a crucial role when the gauge configurations may assume a nontrivial holonomy along  $\mathbb{S}^1$ . Precisely the gauge configurations with the nontrivial holonomy along

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

$\mathbb{S}^1$  may serve as a source of vacuum energy density sustaining the inflationary scenario. Furthermore, as we argue below this construction provides a system with a sub-Planckian energy scale such that a number of well-known and undesirable properties which always accompany the conventional inflationary scenario when a system is formulated in terms of a local field  $\Phi(x)$ , does not even occur in our framework.

Our presentation is organized as follows. We begin in Sec. II with a brief overview of the first crucial element of the proposal: the thermally driven cosmology when the initial state is described by the microcanonical density matrix as originally discussed in [8–10]. In Sec. III we overview a second crucial element of our proposal related to a fundamentally new source of the vacuum energy as suggested in [5–7].

Then in Secs. IV and V we construct two different inflationary models based on similar building principles, but different field contexts. In both models the inflationary vacuum energy is generated by the holonomy of gauge fields. In the first model studied in Sec. IV one can carry out all the computations in theoretically controllable semiclassical approximation as a result of special selection of the matter context. The second model studied in Sec. V is much more attractive phenomenologically, though the semiclassical approximation cannot be justified in this case.

We discuss how the inflation ends in our scenario (the so-called reheating epoch) in Sec. VI. In particular, we demonstrate that the number of  $e$ -folds  $N_{\text{infl}}$  is always very large  $N_{\text{infl}} \sim \alpha^{-2}(H) \sim 100$  as a result of small gauge coupling constant  $\alpha(H) \sim 0.1$  at the Hubble scale  $H$ . We also compare our holonomy inflation with conventional description in terms of the local inflaton  $\Phi$  and potential  $V[\Phi]$  in Sec. VI F.

We conclude in Sec. VII with a formulation of the basic results and profound consequences of our proposal. We also describe in Sec. VII C how this new form of the topological vacuum energy can be tested in tabletop experiments in the physical Maxwell system. We also comment in Sec. VII D on differences of our framework with well-known no-boundary and tunneling proposals. Finally, we summarize a number of technical aspects relevant to our topological inflation scenario in Appendix A. In particular, we overview the nature of the contact term in gauge theories in Sec. A 1, the generation of the “nondispersive” vacuum energy due to the holonomy in Sec. A 2, and its role in cosmological context in Sec. III C.

## II. ORIGIN OF INFLATION IN THE THERMALLY DRIVEN COSMOLOGY

Our goal here is to overview the previous results [8–10] with emphasis on the periodic properties of  $\mathbb{S}^1$  where gravitational instantons are defined and serve as initial conditions for the cosmological evolution of the scale factor  $a(t)$ . Analytical continuation to the physical

Lorentzian spacetime demonstrates the de Sitter-like behavior with constant  $H$ . This is precisely the main goal of this section.

The model of quantum initial conditions in cosmology in the form of the microcanonical density matrix was suggested in [10], where its statistical sum was built as the Euclidean quantum gravity path integral,

$$Z = \int_{\text{Periodic}} D[g_{\mu\nu}, \Phi] e^{-S[g_{\mu\nu}, \Phi]}, \quad (1)$$

over the metric  $g_{\mu\nu}$  and matter fields  $\Phi$  which are periodic on the Euclidean spacetime with a time compactified to a circle  $\mathbb{S}^1$ . This statistical sum has a good predictive power in the Einstein theory with the primordial cosmological constant and the matter sector which mainly consists of a large number of quantum fields conformally coupled to gravity [8,9]. The dominant contribution of numerous conformal modes allows one to overstep the limits of the usual semiclassical expansion, because the integration over these modes gives the quantum effective action of the conformal fields  $\Gamma_{\text{CFT}}[g_{\mu\nu}]$  exactly calculable by the method of conformal anomaly. On the Friedmann-Robertson-Walker (FRW) background,

$$ds^2 = d\tau^2 + a^2(\tau)d^2\Omega^{(3)}, \quad (2)$$

with a periodic scale factor  $a(\tau)$ , the functions of the Euclidean time belonging to the circle  $\mathbb{S}^1$  [8], this action is calculable by using the local conformal transformation to the static Einstein universe and the well-known trace anomaly,

$$g_{\mu\nu} \frac{\delta \Gamma_{\text{CFT}}}{\delta g_{\mu\nu}} = \frac{1}{4(4\pi)^2} g^{1/2} (\alpha \square R + \beta E + \gamma C_{\mu\nu\alpha\beta}^2), \quad (3)$$

which is a linear combination of Gauss-Bonnet  $E = R_{\mu\nu\alpha\gamma}^2 - 4R_{\mu\nu}^2 + R^2$ , Weyl tensor squared  $C_{\mu\nu\alpha\beta}^2$  and  $\square R$  curvature invariants with spin dependent coefficients.<sup>1</sup> The resulting  $\Gamma_{\text{CFT}}[g_{\mu\nu}]$  turns out to be the sum of the anomaly contribution and the contribution of the static Einstein universe—the Casimir and free energy of conformal matter fields at the temperature determined by the circumference of the compactified time dimension  $\mathbb{S}^1$ . This is the main calculational advantage provided by the local Weyl invariance of  $\Phi$  conformally coupled to  $g_{\mu\nu}$ . Solutions of equations of motion for the full effective action—saddle

<sup>1</sup>With the nonvanishing background values of matter fields there are additional contributions to the conformal anomaly like the square of the relevant Yang-Mills strength  $F_{\mu\nu}^2$  or a conformal scalar field  $\phi^4$ , [15]. We disregard them, because in what follows their values are assumed to be either 0 or negligible compared to the contribution of the gravitational structures with large coefficients  $\alpha$ ,  $\beta$  and  $\gamma$ .

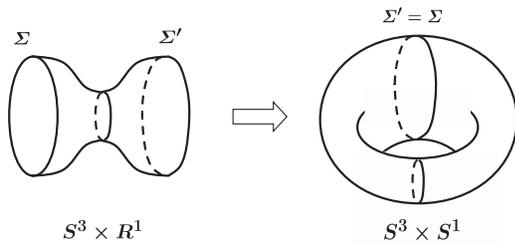


FIG. 1. Transition from the density matrix instanton to the periodic statistical sum instanton.

points of the microcanonical statistical sum (1)—are the periodic cosmological instantons of  $S^1 \times S^3$  topology (in what follows we assume spatially closed cosmology which explains spherical topology of its spatial sections). These statistical sum instantons follow by a usual tracing procedure from the two-boundary instantons of the relevant microcanonical density matrix, which is depicted on Fig. 1. In their turn, the density matrix instantons serve as initial conditions for the cosmological evolution  $a_L(t)$  in the physical Lorentzian spacetime. The latter follows from  $a(\tau)$  by the analytic continuation  $a_L(t) = a(\tau_* + it)$  at the point of the maximum value of the Euclidean scale factor  $a_+ = a(\tau_*)$ , as shown on Fig. 2.

This construction is described in [8–10] and we refer the readers to these original papers. The only comment we make here is that the starting point of the analysis [8–10] is, of course, the density matrix  $\rho(\phi, \phi')$  with two surfaces carrying its field arguments. These surfaces *semiclassically* are the boundaries of either Euclidean or Lorentzian spacetime, depending on the relevant size of the scale factor. The entire saddle-point solution for  $\rho(\phi, \phi')$  consists respectively of the Euclidean spacetime interpolating between them or of the Euclidean spacetime between  $\Sigma$  and  $\Sigma'$ , sandwiched between the two layers of the Lorentzian spacetime. These two layers interpolate from  $\Sigma$  to the unprimed argument of the density matrix and from  $\Sigma'$  to its primed argument and correspond in the density matrix to the chronological and antichronological evolution factors of the well-known Schwinger-Keldysh technique [16] for expectation values in thermal field theory. When calculating the trace in the statistical sum in view of unitarity these two factors cancel out, and the

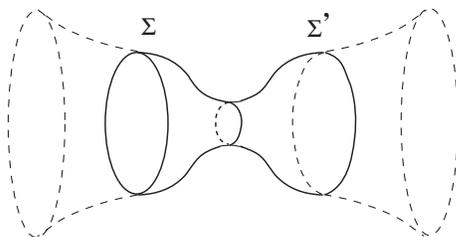


FIG. 2. Density matrix instanton. Dashed lines depict the Lorentzian Universe nucleating at minimal surfaces  $\Sigma$  and  $\Sigma'$ .

only contribution to the statistical sum remains from the Euclidean domain between the Euclidean-Lorentzian transition surfaces  $\Sigma$  and  $\Sigma'$ . These surfaces are uniquely determined from the condition of smooth periodicity in the Euclidean time on the compact  $S^1$ , or as two turning points of the Euclidean trajectory for  $a(\tau)$ .

The equations for these cosmological instantons have the form of the effective Friedmann equation in the Euclidean time  $\tau$  ( $\dot{a} = da/d\tau$ ),

$$-\frac{\dot{a}^2}{a^2} + \frac{1}{a^2} - B \left( \frac{\dot{a}^4}{2a^4} - \frac{\dot{a}^2}{a^4} \right) = \frac{\rho}{3M_P^2} + \frac{C}{a^4}, \quad (4)$$

$$C = \frac{B}{2} + \frac{R(\eta)}{3M_P^2}, \quad (5)$$

$$B = \frac{\beta}{8\pi^2 M_P^2}, \quad (6)$$

where  $M_P = 1/\sqrt{8\pi G}$  is the reduced Planck mass,  $\rho$  is the overall energy density of matter fields other than the conformal particles,  $\beta$  is the coefficient of the Gauss-Bonnet term in the total conformal anomaly of these particles and  $R(\eta)$  is their radiation energy density.<sup>2</sup> The latter is given by a boson or fermion sum over field modes with energies  $\omega$  on a unit 3-sphere at the comoving temperature  $1/\eta$ , the inverse of the instanton circumference  $S^1$  measured in units of the conformal time,

$$R(\eta) = \frac{1}{2\pi^2} \sum_{\omega} \frac{\omega}{e^{\omega\eta} \mp 1}, \quad \eta = \int_{S^1} \frac{d\tau}{a}. \quad (7)$$

Note that  $\gamma$  does not contribute to the above equations in view of conformal flatness of the FRW metric, while the coefficient  $\alpha$  can always be renormalized to 0 by a local  $R^2$  counterterm, changes in  $\alpha$  thus being equivalent to the inclusion of the nonminimally coupled scalaron of the Starobinsky model; see discussion in [17,18].

The integrodifferential equations (4) and (5) form a bootstrap—the amount of radiation constant  $C$  is determined from (5) by the underlying scale factor history  $a(\tau)$  which, in its turn, is generated by the backreaction of this

<sup>2</sup>It should be emphasized that nonconformal matter was not completely excluded in the original setup and the gravitational sector of the theory was not assumed to be Weyl invariant at all. In particular, the role of  $\rho$  could be played by a fundamental cosmological constant, its particular value being selected from the existence of the periodic Euclidean saddle-point solution, as it was in the simplest model of [9]. In more realistic models the role of  $\rho$  is played by the nonconformal inflaton field in the slow-roll regime or the scalaron field of the Starobinsky  $R^2$ -model [17,18]; see below. Moreover,  $\rho$  can contain ordinary particle matter of negligible amount in the early Universe, but quantum created during inflation in view of its nonconformal nature and, therefore, starting to dominate at later stages of the evolution; see footnote 4.

radiation on  $a(\tau)$  via the effective Friedmann equation (4). Their solutions represent the set of periodic  $\mathbb{S}^3 \times \mathbb{S}^1$  gravitational instantons<sup>3</sup> with the oscillating scale factor, *garlands* [8,10], that can be regarded as the thermal version of the Hartle-Hawking instantons [11]. When the matter density is constant or nearly constant and forms a ‘‘Hubble factor’’

$$H^2 \equiv \frac{\rho}{3M_P^2} \quad (8)$$

the scale factor oscillates  $m$  times ( $m = 1, 2, 3, \dots$ ) between the maximum and minimum values,  $a_- \leq a(\tau) \leq a_+$ , so that the full period of the conformal time (7) is the  $2m$ -multiple of the integral between the two neighboring turning points of  $a(\tau)$ ,  $\dot{a}(\tau_{\pm}) = 0$ . Similarly, the full period of the *proper* Euclidean time on these periodic  $m$ -fold *garland* instantons is given by the analogous integral,

$$\mathcal{T} = \oint_{\mathbb{S}^1} d\tau = 2m \int_{a_-}^{a_+} \frac{da}{\dot{a}}. \quad (9)$$

These garland-type instantons exist only in the limited range of  $H^2$  [8]. As shown in [8], periodic solutions should necessarily belong to the domain

$$B - B^2 H^2 \leq C \leq \frac{1}{4H^2}, \quad (10)$$

where they form a countable,  $m = 1, 2, \dots$ , sequence of one-parameter families interpolating between the lower and upper boundaries of this domain in the two-dimensional plane of  $H^2$  and  $C$ . This sequence with  $m \rightarrow \infty$  accumulates at the upper bound of  $H_{\max}^2 = 1/2B$  (and minimal value of  $C_{\min} = B/2$ ), which corresponds to the bound on the effective cosmological constant

$$\Lambda_{\max} = \frac{12\pi^2 M_P^2}{\beta}. \quad (11)$$

The lower bound  $H_{\min}^2$ , the lowest point of the  $m = 1$  family, can be obtained numerically for any field content of the model.

For solutions close to the upper boundary of the domain (10),  $C \simeq 1/4H^2$ , the scale factor oscillates with a very small amplitude, and one can write down the following approximation,

<sup>3</sup>We use the term gravitational instanton to avoid confusion with conventional instanton-type solutions which describe the interpolation between topologically distinct but physically identical winding sectors  $|k\rangle$  in gauge theories. The corresponding periodic instantons (the so-called calorons with nontrivial holonomy) are the subject of the Appendix A where we overview relevant for the present work results.

$$a^2 = \frac{1 + \sqrt{1 - 4CH^2} \cos(\Omega\tau)}{2H^2}, \quad (12)$$

$$\Omega = \frac{2H}{\sqrt{1 - 2BH^2}}, \quad (13)$$

which is valid for  $\sqrt{1 - 4CH^2} \ll 1 - 2BH^2$  [17,18]. The full period of the  $m$ -folded instanton is thus

$$\mathcal{T} = \frac{2\pi m}{\Omega} = \frac{\pi m}{H} \sqrt{1 - 2BH^2}. \quad (14)$$

Remarkably, the bootstrap equations (4) and (5) have explicit solution for large  $m$  and close to the upper boundary of the domain (10) [8]. In this limit the Hubble parameter is close to the upper bound of its range

$$H^2 \simeq \frac{1}{2B} \left( 1 - \frac{\ln^2 m^2}{2\pi^2 m^2} \right), \quad m \gg 1, \quad (15)$$

corresponding to the maximal value of the effective cosmological constant (11).

We make a few comments on the physical meaning of the topological parameter  $m$  which enters Eq. (9). This parameter looks very similar to the integer instanton number in gauge theories where the Euclidean path integral is defined as the sum over all topological sectors, so that it is tempting to consider summation over  $m$ . However,  $m$  is not an independent parameter of the cosmological instantons of the above type. Each instanton is parametrized by two dimensional parameters,  $H^2 = \Lambda/3$ , the cosmological constant or the energy scale of the model and  $M_P$ , the Planckian mass or gravitational coupling constant. The folding number  $m$  is in one-to-one correspondence with the energy scale  $H$  as in (15). Therefore, under a general assumption that at later times the cosmological models with different values of  $H$  decohere and become observable, one should not sum over different values of  $m$  in the contribution to the initial conditions for inflation with a given  $H$ . The concrete values of  $H$  and  $m$  should, thus, be selected by matching with observations.

Inflation stage in this model starts after the ‘‘nucleation’’ of the system from the gravitational instanton when the evolution in the Lorentzian time begins. The Lorentzian time history of the scale factor  $a_L(t)$  originates by the analytic continuation of the approximate solution (12) to  $\tau = 2m\pi/\Omega + it$ . This leads to the replacement of oscillatory behavior of  $\cos(\Omega\tau)$  by exponentially growing  $\cosh(\Omega t)$ , so that at later times nonlinear effects start dominating. When solved with respect to  $\dot{a}^2$  Eq. (4) takes in the Lorentzian spacetime,  $\dot{a}^2(\tau) = -\dot{a}_L^2(t)$ , the manifestly general relativistic form [cf. (26) and (27) below],

$$\frac{\dot{a}_L^2}{a_L^2} + \frac{1}{a_L^2} = \frac{\varepsilon}{3M_{\text{eff}}^2(\varepsilon)}, \quad \varepsilon = \rho + \frac{R}{a_L^4}, \quad (16)$$

$$M_{\text{eff}}^2(\varepsilon) = \frac{M_P^2}{2} \left( 1 + \sqrt{1 - \frac{\beta\varepsilon}{12\pi^2 M_P^4}} \right), \quad (17)$$

with the effective Planck mass  $M_{\text{eff}}(\varepsilon)$  depending on the full matter density  $\varepsilon$  which together with  $\rho$  includes the primordial radiation of the conformal cosmology.

As shown in [17,18], the above Euclidean-Lorentzian scenario remains valid also when the matter density  $\rho$  is represented by an appropriate potential of the slowly varying scalar field playing the role of the inflaton. The evolution consists in the fast quasiexponential expansion during which the primordial radiation gets diluted; the inflaton field and its density  $\rho$  slowly decay by a conventional exit scenario and go over into the quanta of conformally noninvariant fields produced from the vacuum.<sup>4</sup> They get thermalized and reheated to give a new postinflationary radiation with a sub-Planckian energy density,  $\varepsilon \rightarrow \varepsilon_{\text{rad}} \ll M_P^4/\beta$ . Therefore,  $M_{\text{eff}}$  tends to  $M_P$ , and one obtains a standard general relativistic inflationary scenario for which initial conditions were prepared by the garland instanton of the above type.

Interestingly, this model can serve as a source of quantum initial conditions for the Starobinsky  $R^2$ -inflation [1] and Higgs inflation theory [13,14], in which the effective  $H^2$  is generated respectively by the scalaron and Higgs field. In particular, the observable value of the cosmic microwave background (CMB) spectral tilt  $n_s \simeq 0.965$  in these models can be related to the exponentially high instanton folding number [17,18],

$$m \simeq \exp \frac{2\pi}{\sqrt{3(1-n_s)}} \sim 10^8, \quad (18)$$

whereas the needed inflation scale in these models  $H \sim 10^{-6}M_P$  determines the overall parameter  $\beta \sim 10^{13}$  generated by a hidden sector of conformal fields [18,19]. If this sector is built of higher spin conformal fields [19],<sup>5</sup> then the gravitational cutoff [20,21] of the model turns out to be several orders of magnitude higher than the inflation scale, which justifies the omission of the graviton loop contribution and the use of the above nonperturbative (trace anomaly based) method. This concludes our overview of

<sup>4</sup>A realistic model should contain a sector of nonconformal fields which can be negligible on top of conformal fields in the early Universe but eventually starts dominating in the course of cosmological expansion.

<sup>5</sup>Such a high value of  $\beta$  cannot be reached with low spin conformal fields having  $\beta = (1/180)(\mathbb{N}_0 + 11\mathbb{N}_{1/2} + 62\mathbb{N}_1)$ , unless the numbers  $\mathbb{N}_s$  of fields of spin  $s$  are tremendously high. On the contrary, this bound on  $\beta$  can be reached with a relatively low tower of higher spin conformal fields, because partial contribution of spin  $s$  to  $\beta$  grows as  $s^6$  and, moreover, this scaling guarantees that the theory with multiple quantum species remains deeply below its gravitational cutoff [20,21].

the previous results [8–10,17,18] which play an important role in constructions presented in the following sections.

### III. THE TOPOLOGY AS THE SOURCE OF THE VACUUM ENERGY

The goal here is to overview the basic ideas advocated in [5–7]. We explain a number of technical elements related to these ideas in Appendix A, while here we present the corresponding arguments using simple plain language and analogies; see next Sec. III A. The basic prescription of the vacuum energy which enters the Friedmann equations is explained in Sec. III B. In Sec. III C we list a number of key technical elements of the proposal relevant for cosmological applications.

#### A. Intuitive picture

The new paradigm advocated in [5] is based on a fundamentally novel view on the nature and origin of the inflaton field which is drastically different from the conventional viewpoint that the inflaton is a dynamical local field  $\Phi$ . In this new framework the inflation is a *genuine quantum effect* in which the role of the inflaton is played by an auxiliary topological field. A similar field, e.g., is known to emerge in the description of a topologically ordered condensed matter (CM) system realized in nature. This field does not propagate, and does not have a canonical kinetic term, as the sole role of the auxiliary field is to effectively describe the dynamics of the topological sectors of a gauge theory which are present in the system. The corresponding physics is fundamentally indescribable in terms of any local propagating fields [such as  $\Phi(x)$ ]. It might be instructive to get some intuitive picture for the vacuum energy in this framework formulated in terms of a CM analogy. Such an intuitive picture is quite helpful in getting a rough idea about the nature of the inflaton in the framework advocated in this work.

Imagine that we study the Aharonov-Casher effect. We insert an external charge into a superconductor when the electric field  $E$  is screened, i.e.,  $E \sim Q \exp(-r/\lambda)$  with  $\lambda$  being the penetration depth. Nevertheless, a neutral magnetic fluxon will still be sensitive to an inserted external charge  $Q$  at arbitrary large distances in spite of the screening of the physical field. This genuine quantum effect is purely topological and nonlocal in nature and can be explained in terms of the dynamics of the gauge sectors which are responsible for the long range dynamics. Imagine now that we study the same effect but in a time-dependent background. The corresponding topological sectors which saturate the vacuum energy are modified due to the external background. However, this modification cannot be described in terms of any local dynamical fields, as there are no propagating long range fields in the system since the physical electric field is screened. For this simplified example, the dynamics of the inflaton corresponds to the

effective description of the modification of topological sectors when the external background slowly changes. The effect is obviously nonlocal in nature as the Aharonov-Casher effect itself is a nonlocal phenomenon, and cannot be expressed in terms of  $F_{\mu\nu}$ .

One should emphasize that many crucial elements of this proposal have in fact been tested using the numerical lattice Monte Carlo simulations in strongly coupled QCD. Furthermore, this fundamentally new sort of energy can be in principle studied in tabletop experiments by measuring some specific corrections to the Casimir pressure in the Maxwell theory; see remarks and references in concluding Sec. VII C. In the next subsection we specifically list some important technical elements which are used in the construction.

### B. QCD holonomy mechanism of vacuum energy

Let us now get to the discussion of the nature of the effective cosmological constant or a Hubble factor (8) in the modified Euclidean Friedmann equation (4). Our interpretation in the present work is based on the prescription that the relevant energy is in fact the difference  $\Delta\rho \equiv \rho - \rho_{\text{flat}}$  between the energies of a system in a nontrivial background and flat spacetime geometry, similar to the well-known Casimir effect when the observed energy is a difference between the energy computed for a system with conducting boundaries and infinite Minkowski flat space. In this framework it is quite natural to define the “renormalized vacuum energy” to be 0 in flat spacetime vacuum wherein the Einstein equations are automatically satisfied as the Ricci tensor identically vanishes.

In the present context such a definition  $\Delta\rho \equiv \rho_{\text{FLRW}} - \rho_{\text{Mink}}$  for the vacuum energy for the first time was advocated in 1967 by Zeldovich [22] who argued that  $\rho_{\text{vac}} = \Delta\rho \sim Gm_p^6$  must be proportional to the gravitational constant with  $m_p$  being the proton’s mass. Later on such a definition for the relevant energy  $\Delta\rho \equiv \rho_{\text{FLRW}} - \rho_{\text{flat}}$  which enters the Einstein equations was advocated from different perspectives in a number of papers written by researchers from different fields, including particle physics, cosmology, and condensed matter physics; see e.g., relatively recent works [23–27], and review article [28] with a large number of the original references.

This subtraction prescription is consistent with conventional subtraction procedure of the divergent ultralocal bare cosmological constant because in the infinitely large flat spacetime the corresponding contribution is proportional to the  $\delta^4(x)$  function; see (A5). At the same time the nontrivial corrections to  $\Delta\rho$  are nonlocal functions of the geometry and cannot be renormalized by any UV counterterms.

Precisely this feature of nonlocality implies that the relevant energy  $\Delta\rho$  which enters the Friedmann equation, see (19) below, cannot be expressed in terms of a gradient expansion in any effective field theory. Additional arguments supporting the same claim on the impossibility to

formulate the relevant physics in terms of any local effective field, such as inflaton  $\Phi(x)$ , are presented in the following Sec. III C.

This prescription is also consistent with the renormalization group (RG) approach advocated in [28–30]. In fact, it is a direct consequence of the renormalization group approach when we fix a physical parameter at one point of normalization to predict its value at a different normalization point. In the present work with the geometry  $\mathbb{S}^3 \times \mathbb{S}^1$ , the proper length of the  $\mathbb{S}^1$ -period being  $\mathcal{T}$ , it implies that the vacuum energy in the Friedmann equation (4) is  $\rho \equiv \rho(\mathcal{T}^{-1}) - \rho(0)$ , where  $\rho(\mathcal{T}^{-1})$  is the energy of the gauge field holonomy on a compactified spacetime coordinate of length  $\mathcal{T}$ . It can be interpreted as the RG normalization point  $\mu \sim \mathcal{T}^{-1}$ , where  $\mathcal{T}$  is the size of the compactified Euclidean time dimension given by (9). As we already mentioned, this prescription is consistent with the Einstein equations when the vacuum energy approaches 0,  $\Delta\rho \rightarrow 0$ , for the flat spacetime which itself may be considered as a limiting case with  $\mathcal{T} \rightarrow \infty$ .

Finally, with the expression for the energy of the gauge field holonomy winding across the compactified coordinate of the length  $\mathcal{T}$  whose derivation we give in next Sec. III C, one has

$$\rho \equiv \rho_{\text{vac}}[\mathbb{S}^3 \times \mathbb{S}^1] - \rho_{\text{vac}}[\mathbb{R}^4] = \frac{\bar{c}_{\mathcal{T}} \Lambda_{\text{QCD}}^3}{\mathcal{T}}, \quad (19)$$

where  $\Lambda_{\text{QCD}}$  is the scale of the underlying QCD-like gauge field theory and  $\bar{c}_{\mathcal{T}}$  is some dimensionless  $\mathcal{O}(1)$  constant whose precise value is not important for our argumentation.

Our final comment in this subsection goes as follows. As we already mentioned the energy  $\Delta\rho$  can be interpreted as a running cosmological constant within the renormalization group approach advocated in [28–30] with the only difference being that odd powers of  $H$  are also included in the series as a result of the IR sensitivity and nonlocality (in contrast with conventional UV renormalization) as discussed in Appendix A. The linear correction (which is a particular example of the odd power of  $H$ ) to the vacuum energy can be interpreted in terminology [28–30] as the possibility of running cosmological constant at very low  $\mu \sim \mathcal{T}^{-1} \ll M_p$ . This running is originated from nonperturbative and nonlocal physics in QFT (through the nontrivial holonomy along  $\mathbb{S}^1$ ) and cannot be seen at any finite level in perturbation theory, as entire nondispersive vacuum energy cannot be generated in perturbation theory; see some technical comments on this matter in Appendix A 2.

As we see in the next subsection, the leading correction to the vacuum energy (19) is in fact proportional to  $H$ , and this linear in  $H$  correction in the effective Friedmann equation is saturated by the IR-sensitive topological configurations with nontrivial holonomy which cannot be expressed in terms of any local propagating d.o.f.

### C. Nondispersive vacuum energy. Cosmological context.

We define the nondispersive vacuum energy  $E_{\text{vac}}$  in gauge theory in conventional way in terms of the path integral; see Appendix A 1. Precisely this vacuum energy enters all the relevant correlation functions, including the topological susceptibility as defined by (A1).

1. From the arguments of Appendix A 1 one can infer that the  $\theta$ -dependent portion of the vacuum energy  $E_{\text{vac}}(\theta)$  cannot be identified with any propagating d.o.f. Furthermore, all effects are obviously nonanalytical in coupling constant  $\sim \exp(-1/g^2)$  and cannot be seen in perturbation theory. These arguments obviously suggest that there is no local effective field  $\Phi(x)$  (inflaton) which could describe these features of the vacuum energy in gauge theories. These arguments are obviously consistent with our discussions in previous Sec. III B.

2. One can view the relevant topological Euclidean configurations which satisfy the properties from item 1 above as the three-dimensional magnetic monopoles wrapping around  $S^1$  direction. These configurations are characterized by the nonvanishing holonomy (A6), which eventually generates the linear correction  $\sim 1/T$  to the vacuum energy density represented by Eqs. (20) and (24) below.

3. In the cosmological context such configurations are highly unusual objects: they obviously describe the nonlocal physics as the holonomy (A6) is a nonlocal object. Indeed, the holonomy defines the dynamics along the entire history of evolution of the system in the given confined phase: from the very beginning to the very end. There is no contradiction with causality in the system as there are no physical d.o.f. to propagate along this path; see item 1 above. Furthermore, this entire gauge configuration is a mere saddle point in Euclidean path integral computation which describes the instantaneous tunneling event, rather than propagation of a physical d.o.f. capable of carrying any information/signal.

4. The generation of the nondispersive energy  $E_{\text{vac}}$  is a highly nonlocal effect. In particular, formulas (20) and (24) below explicitly show that small variations of the background produce large linear correction  $\sim T^{-1}$  at small  $T^{-1} \rightarrow 0$  as a result of this nonlocality. Precisely this feature of nonlocality implies that the relevant energy  $\Delta\rho$  which enters the Friedmann equation (19) cannot be expressed in terms of a gradient expansion in any effective local field theory as emphasized in Sec. III B.

5. Our subtraction prescription as explained in Sec. III B is consistent with all fundamental principles of QFT. What is more important is that the correction to the energy  $\Delta\rho$  which enters the Friedmann equation (19) cannot be renormalized by any UV counterterms as it is generated by nonlocal configurations.

6. The basic assumption of this work is that the same pattern (as highlighted in items 1–5 above) holds for other

manifolds. In other words, we assume that the vacuum energy density for the  $S^3 \times S^1$  manifold receives a linear correction  $T^{-1}$  in comparison with flat  $\mathbb{R}^3 \times S^1$  geometry, similar to the computations in hyperbolic space  $S^1 \times \mathbb{H}^3$  where computations can be explicitly performed, as reviewed in Appendix A 2, i.e.,

$$\frac{E_{\text{vac}}[S^3 \times S^1]}{E_{\text{vac}}[\mathbb{R}^3 \times S^1]} \simeq \left(1 - \frac{c_T}{T\Lambda_{\text{QCD}}}\right), \quad (20)$$

where  $c_T$  is a coefficient of order one, similar to computations in Appendix A 2. Formula (20) plays the crucial role in our arguments in Secs. IV and V.

One can use conventional thermodynamical relation

$$dF = TdS - PdV, \quad P = -\left.\frac{\partial F}{\partial V}\right|_S \quad (21)$$

to convince oneself that the correction  $\sim T^{-1}$  does not modify the equation of state (EoS). In fact, it behaves exactly in the same way as the cosmological constant does, i.e.,

$$P = -\frac{\partial F}{\partial V} = +\frac{32\pi^2}{g^4}\Lambda_{\text{QCD}}^4 \left(1 - \frac{c_T}{T\Lambda_{\text{QCD}}}\right)$$

$$\rho = \frac{F}{V} = -\frac{32\pi^2}{g^4}\Lambda_{\text{QCD}}^4 \left(1 - \frac{c_T}{T\Lambda_{\text{QCD}}}\right), \quad (22)$$

where we use formula (A8) for  $F$  with correction factor (20). The correction  $\sim T^{-1}$  does not modify the equation of state  $w = -1$ , which is normally associated with the cosmological constant contribution,

$$w \equiv \frac{P}{\rho} = -1. \quad (23)$$

Finally, using (20) the vacuum energy for  $S^3 \times S^1$  manifold can be represented as follows,

$$E_{\text{vac}}[S^3 \times S^1] \simeq -\frac{32\pi^2}{g^4}\Lambda_{\text{QCD}}^4 \left(1 - \frac{c_T}{T\Lambda_{\text{QCD}}}\right)$$

$$\simeq -\frac{32\pi^2}{g^4}\Lambda_{\text{QCD}}^4 + \Lambda_{\text{QCD}}^3 \frac{\bar{c}_T}{T} + \mathcal{O}\left(\frac{1}{T^2}\right), \quad (24)$$

where we redefined  $\bar{c}_T \equiv \frac{32\pi^2}{g^4}c_T$  as the parameter  $c_T \sim 1$  is expected to be order of 1 (based on the previous experience) but is not yet known.

We conclude this section with a few important comments which are relevant for the physical interpretation of the obtained results.

- (i) All computations presented above, as usual, are performed in the Euclidean spacetime where the

relevant gauge configurations describing the tunneling processes are defined. Using this technique we computed the energy density  $\rho$  and the pressure  $P$  in the Euclidean space. As usual, we assume that there is analytical continuation to Lorentzian spacetime where the physical energy density has the same form. This is of course conventional procedure for the QCD practitioners who normally perform computations on the lattice using the Euclidean formulation, while the obtained results are expressed in physical terms in Minkowski spacetime. In our context it means that the parameters  $P$ ,  $\rho$  and EoS as given by (23) are interpreted as the corresponding parameters in physical Lorentzian spacetime.

- (ii) Therefore, the driving force for the de Sitter behavior in the Lorentzian space is not a local dynamical inflaton field  $\Phi(x)$  which never emerges in our framework. Rather the driving force in our scenario should be thought of as a Casimir-type vacuum energy which is generated by numerous tunneling transitions in a strongly coupled gauge theory determined by the dimensional parameter  $\Lambda_{\overline{\text{QCD}}}$ . Precisely this parameter replaces the dimensional parameters from inflaton potential  $V[\Phi(x)]$  which cosmology practitioners normally use in their studies.
- (iii) The equation of state (23) in Lorentzian spacetime obviously implies the de Sitter expansion. The corrections due to the radiation  $\rho_r$  and matter  $\rho_m$  can be easily incorporated into the Friedmann equation written in Lorentzian spacetime. The interaction of the system with standard model (SM) particles modifies the EoS (23). Precisely these modifications to EoS (23) are responsible for the end of inflation as described in Sec. VI.

#### IV. THE HOLONOMY INFLATION. MODEL 1

The origin of inflation in the model reviewed in previous Sec. II is based on two important ingredients—the vacuum energy (8) of a certain local nature and the hidden sector of conformal fields critically important for the contribution of the conformal anomaly and generation of the thermal radiation in effective Friedmann equation. The key technical element for the successful inflation is the presence of  $\mathbb{S}^1$  which emerges in the system as a result of thermal initial state formulated in terms of the density matrix. We keep this first ingredient of our construction from previous studies as is explained below in Sec. IVA.

A new idea which is advocated in the present work is that the second important ingredient of this framework, the vacuum energy, may be originated from some nontrivial nonlocal gauge configurations. This structure in our proposal is fundamentally different from all conventional

inflationary models because this source of the vacuum energy cannot be expressed in terms of any local d.o.f. such as scalar inflaton  $\Phi(x)$ .

In our construction this source of the vacuum energy is generated by the gauge configurations with nontrivial holonomy in the QCD-like field theory as explained in Sec. III B. This construction uses exactly the nontrivial topology  $\mathbb{S}^1 \times \mathbb{S}^3$  of the gravitational instanton considered above. In its turn, the origin of this topology, compactification of the Euclidean time on a circle  $\mathbb{S}^1$ , is entirely due to a subtle effect of conformal radiation, whereas the inflation compatible value of the vacuum energy is the effect of this holonomy in the QCD-like gauge theory with a sub-Planckian scale as explained in Sec. IV B.

We treat *model 1* considered in this section as a toy model where, one hand, one can demonstrate all the crucial elements of the construction. On the other hand, one can adjust parameters in a such a way that all computations are under complete theoretical control and the semiclassical approximation is justified. Unfortunately, this model is not very natural as it requires very large instanton folding number  $m$  and very large  $\beta$  to be consistent with observations.

In the next section V we consider *model 2*, which is naturally consistent with all presently available observations without special selection of the parameters  $\beta$  or  $m$ . However, we should relax some technical requirements for model 2 in which case the semiclassical approach is not formally justified.

##### A. The effect of the radiation generating $\mathbb{S}^1$

The effect of the radiation related to the difference  $C - B/2$  in (5) is indeed quite subtle because the radiation itself is strongly suppressed. For a high folding number  $m \gg 1$  according to equation (15) it is proportional to

$$C - \frac{B}{2} \simeq B \left( \frac{\ln m}{\pi m} \right)^2 \quad (25)$$

and very small for instanton solutions at the tip of the triangular domain (10) with  $H^2 \simeq 1/2B$  and  $C \simeq 1/4H^2$ . At the same time merely the existence of the radiation enforces us to consider the topology  $\mathbb{S}^3 \times \mathbb{S}^1$ . If one ignores the radiation then the topology  $\mathbb{S}^3 \times \mathbb{S}^1$  reduces to  $\mathbb{S}^4$ . This easily follows from the effective Friedmann equation (4) with  $\rho = 3M_p^2 H^2$  when it is cast, by solving it with respect to  $\dot{a}^2$ , into the form

$$-\frac{\dot{a}^2}{a^2} + \frac{1}{a^2} = \mathcal{H}^2(a), \quad (26)$$

$$\mathcal{H}^2(a) \equiv \frac{1}{B} \left( 1 - \sqrt{1 - 2BH^2 - \frac{2BR}{3M_p^2 a^4}} \right). \quad (27)$$

For  $R = 0$  it gives as a solution the Euclidean sphere,  $a(\tau) = \sin(\mathcal{H}\tau)/\mathcal{H}$ , of the radius  $1/\mathcal{H} = (1/2H^2) \times (1 - \sqrt{1 - 2BH^2})$ , whereas any however small amount of radiation would provide a bouncing of  $a$  back from some nonzero minimal value, otherwise  $a = 0$  occurring at the pole of  $\mathbb{S}^4$ .

But the contribution of such spherical (Hartle-Hawking) instantons to the path integral is completely suppressed as argued in [8,31]. Technically, this suppression occurs as a result of the conformal anomaly which changes the sign of the negative classical action on  $\mathbb{S}^4$  and, moreover, makes it divergent at the poles of the 4-sphere at  $a \rightarrow 0$ . Thus, it is entirely due to the radiation of conformal particles that the scale factor never shrinks to 0, which allows one to compactify time on a circle and get the  $\mathbb{S}^3 \times \mathbb{S}^1$  topology, which can bear a nontrivial gauge field holonomy.

### B. QCD holonomy and inflation scale

The prescription we are advocating in the present work essentially corresponds to the identification of the vacuum energy (19) with the energy density  $\rho$  in the Hubble factor  $H^2$  (8) of the effective Friedmann equation (4), i.e.,

$$H^2 = \frac{\lambda}{\mathcal{T}}, \quad \lambda \equiv \frac{\bar{c}_{\mathcal{T}} \Lambda_{\text{QCD}}^3}{3M_P^2}. \quad (28)$$

With the instanton period of  $m$ -folded garland (14), which is inverse proportional to  $H$ , this immediately gives

$$H\sqrt{1 - 2BH^2} = \frac{\lambda}{\pi m}. \quad (29)$$

This equation is always correct for any value of  $BH^2$ .

However, the bootstrap self-consistency solution always has the property that  $2BH^2 = \mathcal{O}(1)$  as shown in detail in previous papers on CFT driven cosmology. The corresponding results are discussed at the end of this subsection, while now we make a few comments related to the small value of parameter  $BH^2 \ll 1$ , which can be achieved in model 2, to be discussed in the next section V.

If we formally take  $BH^2 \ll 1$  in expression (29) the term  $BH^2 \ll 1$  can be numerically neglected in which case  $\lambda \sim H$  and Eq. (19) assumes the form

$$\rho(H) \equiv \rho_{\text{vac}}[\mathbb{S}^3 \times \mathbb{S}^1] - \rho_{\text{vac}}[\mathbb{R}^4] \simeq H \frac{\bar{c}_{\mathcal{T}} \Lambda_{\text{QCD}}^3}{\pi m}, \quad (30)$$

which explicitly shows the linear dependence of the vacuum energy on the Hubble constant,  $\rho(H) \sim H$ , as previously claimed. One can see from (20) and (24) that

the source of this linear correction to the vacuum energy is related to the term proportional to  $\mathcal{T}^{-1}$  which represents the inverse size of  $\mathbb{S}^1$  manifold for our geometry, and proportional to  $H$  in our framework. Needless to say, this linear (with respect to  $H$ ) correction is saturated by the IR topological configurations with nontrivial holonomy which cannot be expressed in terms of any local propagating fields as explained in Appendix A. Therefore, this term cannot be written in a conventional gradient expansion in an effective field theory as it represents a global rather than local characteristic of the system.

Our next step is to make these computations self-consistent satisfying the semiclassicality condition. Formally, this condition is expressed as the bootstrap equation with solution (15). The physical meaning of the enforcement of the bootstrap equation as explained in previous Sec. II and original papers [8–10] is that the temperature of the system (therefore the size of  $\mathbb{S}^1$ ) cannot be an arbitrary parameter. Instead, it must be determined by the system itself. In other words, the size of the manifold changes as a result of accounting for the feedback to adjust the changes of the vacuum energy. This formal enforcement obviously implies that all dimensional parameters must be of the order of  $M_P$  as the only scale of the problem. The deviation from the Planck scale may only occur if some very small or very large dimensionless parameters are present in the system. In our model 1 there are two free parameters,  $\beta$ , which effectively count the number of d.o.f., and the instanton number  $m$  which, in principle, assumes any value.

In this section, in model 1, we proceed with self-consistent computations. Therefore, we enforce the semiclassicality conditions. In this case, for large  $m$  and the value of  $H$  determined by the bootstrap solution (15) this equation gives the expression for the parameter  $\lambda$  which is equivalent to  $\Lambda_{\text{QCD}}$ , i.e.,

$$\lambda \simeq \frac{\ln m}{\sqrt{B}} = \sqrt{8\pi^2} \frac{M_P}{\sqrt{\beta}} \ln m, \quad (31)$$

$$\Lambda_{\text{QCD}} \simeq \left( \frac{6\sqrt{2}\pi \ln m}{\bar{c}_{\mathcal{T}}} \right)^{1/3} \frac{M_P}{\beta^{1/6}}. \quad (32)$$

As this model is considered to be a toy model, we can take  $\beta$  as a free parameter and consider  $\beta \gg 1$  such that

$$\frac{\Lambda_{\text{QCD}}}{M_P} \sim \frac{1}{\beta^{1/6}} \ll 1, \quad \frac{H}{M_P} \simeq \frac{2\pi}{\beta^{1/2}} \ll 1. \quad (33)$$

The key observation we make here is that both parameters,  $H$  and  $\Lambda_{\text{QCD}}$ , belong to the sub-Planckian scale according to (33), which justifies the use of the semiclassical expansion discarding a negligible contribution of

graviton loops.<sup>6</sup> Furthermore, there is a hierarchy of scales which parametrically holds for large  $\beta \gg 1$ ,

$$H \ll \Lambda_{\overline{\text{QCD}}} \ll M_P. \quad (34)$$

This hierarchy of scales once again demonstrates the self-consistency of the computations (on the gauge side) because the nondispersive vacuum energy (19) related to the holonomy is only generated in the confined phase of the gauge  $\overline{\text{QCD}}$  theory at temperature below  $\Lambda_{\overline{\text{QCD}}}$ , which is automatically satisfied as a result of the hierarchy (34).

The inflation scenario in the Lorentzian domain described in Sec. II [Eqs. (16) and (17) above] holds also in the gauge holonomy inflation model advocated in the present work. However the exit from inflation takes place via a decay of  $H$  due to *helical* instability to be discussed below. As mentioned above, if one attempts to match the parameters  $m$  and  $\beta$  with observational numbers, one should take extremely high values of these parameters. Indeed, the model becomes phenomenologically compatible with the CMB data within the Starobinsky  $R^2$  or Higgs inflation theory when the scale  $H \sim 10^{-6} M_P$ . It can be generated by the scenario [8] with  $\beta \simeq 10^{13}$ . In particular, it matches the observable value of the spectral tilt  $n_s \simeq 0.97$  when the number of instanton folds equals (18),  $m \sim 10^8$ . If we assume that instead of the  $R^2$ -mechanism or the Higgs potential the vacuum energy is entirely due to the  $\overline{\text{QCD}}$  holonomy mechanism of the above type, then from (32) it follows that

$$\Lambda_{\overline{\text{QCD}}} \sim 0.05 M_P, \quad H \sim 10^{-6} M_P. \quad (35)$$

The necessity to have a very high value of  $\beta$ , which now can only be generated by a large hidden sector of conformal higher spin fields [19], makes this model rather speculative even though it justifies semiclassical expansion below the gravitational cutoff of [20,21]. Therefore we consider the second, much more natural model without any hidden sectors filled by a large number of conformal fields.

## V. THE HOLONOMY INFLATION. MODEL 2

The starting point in this section is the same set of equations discussed in previous sections. However, in (4) we now ignore the higher derivative terms  $\sim B\dot{a}^2$  and  $\sim B\dot{a}^4$ .

<sup>6</sup>Sub-Planckian scale of the model does not imply, however, that the  $B$ -terms in (4) quadratic in curvature and generated by the conformal anomaly can be discarded. Effective action generating the conformal anomaly is nonlocal and strong in the infrared which is an artifact of the conformal invariance of the matter fields. In contrast to the quantum loops of the conformal non-invariant graviton, loop effects of conformal matter are not suppressed by inverse powers of the Planck mass and their gravitational effect is treated beyond perturbation theory.

It corresponds to disregarding the higher derivative terms in the effective action as the typical scales of the problem are much lower than the Planck scale  $M_P$ . The corresponding set of equations has been reviewed above, but now we consider the limit  $BH^2 \ll 1$  and for the convenience of the readers repeat some important formulas below.

### A. Overview of gravitational instanton solution

The scale factor  $a(\tau)$  oscillates between the maximum and minimum values  $a_{\pm}$  determined as follows:

$$a_{\pm}^2 = \frac{1}{2H^2} (1 \pm \sqrt{1 - 4CH^2}), \quad a_{\pm} \equiv a(\tau_{\pm}), \quad \frac{\Lambda}{3} \equiv H^2.$$

The solution for the scale factor  $a(\tau)$  is also known,

$$a^2(\tau) = \frac{1}{2H^2} (1 - \sqrt{1 - 4CH^2} \cos(2H\tau)). \quad (36)$$

Now we implement the ideas formulated in the previous subsection. To proceed with this task we identify the energy (24) with the vacuum energy entering the Friedmann equations as we discussed in the previous section, i.e.,

$$\rho \equiv \rho_{\text{vac}}[\mathbb{S}^3 \times \mathbb{S}^1] - \rho_{\text{vac}}[\mathbb{R}^4] = \frac{\bar{c}_T \Lambda_{\overline{\text{QCD}}}^3}{\mathcal{T}}. \quad (37)$$

The prescription we are advocating in the present work essentially corresponds to the identification of the vacuum energy (37) with the cosmological constant  $\Lambda/3$  entering the equation (8), i.e.,

$$3M_P^2 H^2 = \rho = \frac{\bar{c}_T \Lambda_{\overline{\text{QCD}}}^3}{\mathcal{T}}. \quad (38)$$

Up to this point Eq. (38) identically coincides with our analysis in Eqs. (19)–(28) from the previous section.

### B. Relaxing the semiclassical approximation

The new element for model 2 is as follows. We relax the bootstraplike equation and its solution (15) for this model. Essentially we unlink a few parameters which were previously tightly linked. In particular, the de Sitter temperature being expressed in terms of the size of  $\mathbb{S}^1$  is unambiguously fixed by the radiation parametrized by parameter  $\beta$ . This relation essentially fixes the size of  $\mathbb{S}^1$  which is generated by the radiation and determined by the backreaction of  $\mathbb{S}^1$  to the corresponding radiation. The size of  $\mathbb{S}^3$  is also not a free parameter in the semiclassical gravitational instanton solution. Essentially, by relaxing these links we assume that there could be another physics which determines the size of the gravitational instanton (or a complicated network of strongly interacting gravitational instantons). A self-consistent semiclassical approximation obviously cannot be justified when some parameters enter

from different physics. In Appendix A 3 we overview a well-known example in strongly coupled gauge theory where the holonomy (and corresponding size of the manifold) is not fixed by hand, but rather is determined dynamically by strong quantum fluctuations. We suspect that a similar physics may emerge here.

In any case, for model 2 we unlink the size of  $\mathbb{S}^1$  from the radiation and treat it as a free parameter. To simplify our formulas, we also assume the lowest possible instanton number  $m = 1$  in all expressions in this section, which should be contrasted with our studies in the previous section analyzing model 1 where the consistent description exists only for very large  $m \sim 10^8$ . This simplification does not modify our main results as the instanton number always accompanies dimensional parameter  $\Lambda_{\overline{\text{QCD}}}$  and dimensionless coefficient  $\bar{c}_T$ , which are not yet known and can always be redefined.<sup>7</sup>

With these preliminary remarks, and after substituting  $T = \pi/H$  (which is a good approximation in the regime  $CH^2 \ll 1$  we are interested in, see below), Eq. (38) can be rewritten in the following form:

$$3M_P^2 H^2 = \frac{\bar{c}_T \Lambda_{\overline{\text{QCD}}}^3}{\pi} H. \quad (39)$$

This equation is very important as it relates the Hubble constant  $H$  for our Euclidean geometry  $\mathbb{S}^3 \times \mathbb{S}^1$  with the vacuum energy generated by the gauge configurations with nontrivial holonomy,

$$H = \frac{\bar{c}_T \Lambda_{\overline{\text{QCD}}}^3}{3\pi M_P^2}, \quad \rho = \frac{\bar{c}_T^2 \Lambda_{\overline{\text{QCD}}}^6}{3\pi^2 M_P^2}. \quad (40)$$

A few comments are in order. First of all, the hierarchy of scales (33) and (34) characterizing model 1 from the previous section still holds in the present case,

$$H \ll \Lambda_{\overline{\text{QCD}}} \ll M_P. \quad (41)$$

However, in model 2 the hierarchy emerges not as a result of extremely large parameter  $\beta \gg 1$ , but rather, as a result of new scale of the problem,  $\Lambda_{\overline{\text{QCD}}}$ , which is a free dimensional parameter of the system generated by the dimensional transmutation in classically conformal field theory and plays the same role in  $\overline{\text{QCD}}$  as  $\Lambda_{\text{QCD}} \simeq 170$  MeV plays in QCD physics.

<sup>7</sup>It does not imply that the system suffers from some ambiguities. In fact, the coefficient  $\bar{c}_T$  can be in principle computed from the first principles, while the observation of the tensor fraction  $r$  would unambiguously fix the relation between  $H$  and  $\Lambda_{\overline{\text{QCD}}}$ ; see Sec. VI with details. If both these parameters are known, the instanton number  $m$  saturating the path integral can also be computed.

$$\frac{H}{M_P} \sim \frac{\bar{c}_T}{3\pi} \left( \frac{\Lambda_{\overline{\text{QCD}}}}{M_P} \right)^3 \ll 1, \quad \frac{\Lambda_{\overline{\text{QCD}}}}{M_P} \ll 1. \quad (42)$$

Parameter  $\Lambda_{\overline{\text{QCD}}}/M_P \ll 1$  plays the same role in model 2 as parameter  $\beta^{-1/6} \ll 1$  plays in model 1 as expressed by Eq. (33). The crucial difference, however, is that we unlink the size of  $\mathbb{S}^1$  from the radiation by treating  $\Lambda_{\overline{\text{QCD}}}$  as a free dimensional parameter which defines a new gauge theory coined as  $\overline{\text{QCD}}$ . It is assumed<sup>8</sup> at this point that the size  $\mathbb{S}^1$  where the holonomy is defined is determined by a different physics as discussed in Appendix A 3.

### C. Subtle effects of the radiation

Due to the hierarchy of scales mentioned in the previous subsection, one can explicitly check that the relevant parameter  $\epsilon \equiv 4CH^2$  entering Eq. (36) is very small,

$$\epsilon \equiv 4CH^2 \sim \left( \frac{\Lambda_{\overline{\text{QCD}}}}{M_P} \right)^6 \ll 1. \quad (43)$$

Indeed, as it follows from Eq. (36),  $\epsilon \leq 1$  because for larger  $\epsilon$  the turning points disappear and monotonically changing  $a(\tau)$  cannot form a periodic solution—the saddle point of the partition function path integral. Thus in view of (5) the amount of radiation  $R(\eta)$  is always bounded from above—though the Universe is born not in the vacuum state it is still essentially cold. The hottest possible Universe corresponding to a maximal value  $\epsilon = 1$  and minimal  $\eta = \pi\sqrt{2}$  has a moderate maximal value of  $R(\eta) = O(1)$ . Actual smallness of  $\epsilon$  assumed above follows from a sub-Planckian value of  $H \ll M_P$ , because Eq. (5) is then equivalent to  $\epsilon = (\beta + O(1))H^2/4\pi^2 M_P^2 \ll 1$ .

Thus one can simplify the formula (36) and present an approximate solution for  $a(\tau)$  in the following form,

$$a(\tau) \simeq \frac{1}{H} |\sin(H\tau)|, \quad (44)$$

which is valid everywhere except the points close to 0's of  $\sin(H\tau)$ . In the approximation (44) we neglected the terms  $\sim \epsilon$  in accordance with (43). In particular,  $a(\tau = 0)$  is in fact

<sup>8</sup>We make a short comment here why and how such unlink between these two parameters may occur. In weakly coupled semiclassical approximation in model 1 the two parameters (the intensity of radiation characterized by the size of  $\mathbb{S}^1$ , which in its turn depends on the anomaly parameter  $\beta$  in view of the bootstrap equation) are tightly linked. In strongly coupled gauge theory as reviewed in Appendix A 3 the holonomy and size of effective  $\mathbb{S}^1$  is determined dynamically. This is precisely the reason why these two parameters in strongly coupled regime are not linked. As reviewed in Appendix A 3 it is believed that in strongly coupled QCD the holonomy is also determined by the dynamics, the so-called “confining holonomy” when the instanton dissociates into  $N$  constituents. Such a phenomenon may only occur for topological configurations with nontrivial holonomy (A7). The known dependence of the vacuum energy on  $\theta$  as  $\cos(\frac{\theta}{N})$  is an explicit manifestation of the same nontrivial holonomy.

$\sim\sqrt{\epsilon}$  rather than 0, and the exact solution (36) is required for the computation of  $\eta$ ; see (46).

Now we consider only single-folded instantons and compute the full period of conformal time  $\eta$  which can be rewritten as follows,

$$\eta = \frac{1}{H} \left[ \int_0^{\pi/2} \frac{d\phi}{a(\tau)} + \int_{\pi/2}^{\pi} \frac{d\phi}{a(\tau)} \right], \quad \phi \equiv 2H\tau, \quad (45)$$

and reduced to incomplete elliptic integral. Within the  $\ln \epsilon$  accuracy it reads

$$\eta \simeq \frac{1}{\sqrt{2}} \ln \frac{1}{\epsilon}. \quad (46)$$

During this long evolution represented by conformal Euclidean time (46) the scale factor  $a(\tau)$  makes some drastic changes in size as one can see from the following estimation,

$$\frac{a_+}{a_-} \simeq \frac{1}{\sqrt{CH^2}} \sim \frac{1}{\sqrt{\epsilon}} \gg 1. \quad (47)$$

One should observe here that there is a qualitative difference with discussions of model 1 when the ratio (47) was always parametrically of order 1. In the present model 2 this ratio (47) could be parametrically very large which implies that the largest and smallest sizes in the garland construction could have parametrically different scales.<sup>9</sup>

We conclude with the following comment. Merely the existence of the radiation forces us to consider the topology  $\mathbb{S}^3 \times \mathbb{S}^1$ . If one ignores the radiation and the presence of  $\mathbb{S}^1$  then the system defined on  $\mathbb{S}^3 \times \mathbb{S}^1$  becomes defined on  $\mathbb{S}^4$ , in which case the corresponding contribution to the path integral is strongly suppressed as argued in [31]. Technically, this suppression occurs as a result of the conformal anomaly which changes the sign of the classical Euclidean action. In addition, the positive action which is generated due to the conformal anomaly is divergent at  $a \rightarrow 0$  for  $\mathbb{S}^4$ . This divergence leads to the infinitely strong suppression of these vacuum  $\mathbb{S}^4$  configurations; see [31] for comments and details. One should also add that in model 2 the relevant  $\mathbb{S}^1$  structure might be generated not only by radiation but also by the quantum interactions in strongly coupled gauge theories as argued in Appendix A 3 such that size of the  $\mathbb{S}^1$  is a free parameter of the model and it is

<sup>9</sup>Note that a large value of the ratio  $a_+/a_-$  does not essentially affect the thermal history of the inflation in the Lorentzian spacetime modulo the determination of its original energy scale, because the low temperature primordial radiation in Eq. (16) gets quickly diluted during inflationary expansion and does not contribute to the reheating at the exit from the inflationary scenario.

determined by the dimensional parameter  $\Lambda_{\overline{\text{QCD}}}$  of the strongly coupled  $\overline{\text{QCD}}$  gauge theory.

## VI. HOW THE HOLONOMY INFLATION ENDS

The main goal of this section is to argue that the holonomy inflation paradigm advocated in this work is consistent with all presently available observations. One should emphasize that a theory describing the end of inflation (similar to preheating and reheating stages in conventional inflationary scenario) in our framework is yet to be developed. The required technique which would answer the relevant questions is formulated in Sec. VI B by items 1–4. Therefore, this section should be treated as a description of a vision and foresight for a future development rather than a final formulation of the theory describing the end of inflation.

We focus on three items to demonstrate the consistency of the framework. First of all we argue that the EoS almost identically coincides with the EoS which is normally attributed to the cosmological constant. Secondly, we argue that the nondispersive vacuum energy which plays the key role in this framework is capable of transferring its energy to the real propagating gauge fields of the SM. Therefore, the topological inflation could end with a successful reheating epoch. Finally, we estimate the number of  $e$ -folds  $N_{\text{infl}}$  for this framework to show that it is perfectly consistent with presently available observations.

### A. Equation of state

We start with the following generic remark. Consider the holonomy which assumes a nontrivial value along  $\mathbb{S}^1$  directed in time direction as discussed in the previous section. In this case the Hubble constant and the energy density remain constant even after the nucleation from the gravitational instanton in spite of the fact that the topology of the manifold is not  $\mathbb{S}^3 \times \mathbb{S}^1$  anymore. Further to this point, the system is not described by the Euclidean metric after the nucleation, but rather assumes the conventional Lorentzian signature.

The corresponding Hubble constant  $H$  is unambiguously determined by the dimensional parameter  $\Lambda_{\overline{\text{QCD}}}$  of a strongly coupled gauge theory as Eq. (40) states. This solution after the nucleation corresponds to the inflationary (almost) de Sitter behavior such that the EoS and parameter  $a(t)$  assume the form

$$w \equiv \frac{P}{\rho} \simeq -1, \quad a(t) \sim \exp(Ht), \quad (48)$$

in accordance with Eqs. (22) and (23).

The inflationary regime described by (48) would be the final destination of our Universe if the interaction of the  $\overline{\text{QCD}}$  fields with SM particles were always switched off. One should emphasize that the driving force for this

inflationary de Sitter behavior (48) in the Lorentzian space is not a local inflaton field  $\Phi(x)$  which is not present in our system at all. Rather the driving force should be thought of as a Casimir-type vacuum energy which is generated by numerous tunneling transitions in a strongly coupled gauge theory determined by the dimensional parameter  $\Lambda_{\overline{\text{QCD}}}$ . Precisely this parameter replaces the dimensional parameters from inflaton potential  $V[\Phi(x)]$  which cosmology practitioners normally use in their studies.

When the coupling of the  $\overline{\text{QCD}}$  fields with SM particles is switched back on, the end of inflation is triggered precisely by this interaction which itself is unambiguously fixed by the triangle anomaly as we discuss below.

### B. Anomalous coupling of the nondispersive vacuum energy with gauge fields

Before we explain the structure of the relevant interaction we make a few comments in order to explain the physical nature of such unusual coupling between propagating and nonpropagating d.o.f. First of all, we have to recall that the physics responsible for the generating of the nondispersive vacuum energy (dubbed as a “strange energy” in [5–7]) which eventually leads to the de Sitter behavior (48) cannot be formulated in terms of any physical propagating d.o.f. as discussed in great detail in Sec. III C. Instead, the generation of this energy can be explained in terms of tunneling transitions between topologically distinct but physically identical  $|k\rangle$  states.

The corresponding technique to describe these tunneling transitions is normally formulated in terms of the Euclidean path integral and the corresponding field configurations interpolating between distinct topological sectors. In conventional QFT computations the corresponding procedure selects a specific superposition of the  $|k\rangle$  states which generates the  $|\theta\rangle$  state with energy  $E_{\text{vac}}(\theta)$ . In the context of inflation, when the background assumes a nontrivial geometry (instead of  $\mathbb{R}^4$  in the conventional case) the corresponding computations become profoundly more complicated, though the corresponding procedure is well defined.

1. One should describe the relevant Euclidean configurations satisfying the proper boundary conditions for a nontrivial geometry (similar to calorons with nontrivial holonomy, reviewed in Appendix A 2).

2. One should compute the corresponding path integral which includes all possible positions and orientations of the relevant gauge configurations.

3. The corresponding computations for the vacuum energy  $\rho$  and pressure  $P$  must be done with all fields which couple to  $\overline{\text{QCD}}$  gauge theory. Precisely this coupling is responsible for transferring the vacuum energy to SM particles.

4. As the last step, one should subtract the corresponding expression computed on  $\mathbb{R}^4$  as explained in Sec. III B. Precisely this remaining part of the vacuum energy is

interpreted as the relevant energy which enters the Friedmann equation, and which cannot be removed by any subtraction procedure and cannot be renormalized by any UV counterterms. The corresponding formulas for  $\rho, P$  depend, in general, on properties of the manifold and relevant coupling constants.

While these steps are well defined in principle, it is not feasible to perform the corresponding computations because even the first step in this direction, a finding of the relevant Euclidean configurations satisfying the proper boundary conditions for a nontrivial geometry, is yet unknown. Nevertheless, this procedure, in principle, shows that the de Sitter behavior (48) in this framework emerges without any local inflaton field  $\Phi(x)$  as explained in previous Sec. VI A because the physical force driving the inflation has a completely different nature in this proposal.

Fortunately, the key ingredients which are relevant for our future studies can be understood in an alternative way, in terms of the auxiliary topological nonpropagating fields  $b(x, H)$  which effectively describe the relevant IR physics representing the key elements of steps 1–4 highlighted above.

The corresponding formal technique is widely used in particle physics and CM communities. For the convenience of the readers we provide (within our cosmological context) the main ideas and results of this approach in Appendix B. In particular, this approach is extremely useful in the description of the topologically ordered phases when the IR physics is formulated in terms of the topological Chern-Simons (CS)-like Lagrangian. One should emphasize that the corresponding physics, such as the calculation of the braiding phases between quasiparticles, computation of the degeneracy etc., can be computed (and in fact originally was computed) without Chern-Simons Lagrangian and without auxiliary fields. Nevertheless, the discussion of the IR physics in terms of CS-like effective action proves to be very useful, beautiful and beneficial. In our case, unfortunately, we cannot proceed with explicit computations along the lines of 1–4 as explained above. Therefore, the alternative technique in terms of the auxiliary topological nonpropagating fields is the only remaining option in our case.

We refer to Appendix B where we overview the corresponding technique in the context of the inflationary cosmology. We also explain there the physical meaning of these auxiliary field  $b(x, H)$  which should be thought of as the source of the topological fluctuations, similar to the axion field; see below. Precisely this auxiliary nonpropagating field eventually generates the nondispersive energy (37) and consequently leads to the de Sitter behavior (48). This auxiliary field  $b(x, H)$  effectively describes (through the correlation functions) the modification of the tunneling rates between topological  $|k\rangle$  sectors as a result of external background field parametrized by  $H$ . In other words, a

profoundly complicated procedure of summation over all topological configurations interpolating between  $|k\rangle$ -sectors in the background parametrized by  $H$  as outlined above (steps 1–4) can be expressed in terms of the auxiliary field  $b(x, H)$  which, of course, remains the nonpropagating auxiliary field in background  $H$ .

The only information which is required for future analysis is that the relevant auxiliary field  $b(x, H)$ , saturating the nondispersive vacuum energy (37), couples to the SM particles precisely in the same way as the  $\theta$  parameter couples to the gauge fields. This claim is explained in Appendix B and is based on analysis of the exact anomalous Ward identities. In many respects the coupling of the  $b(x, H)$  field to the gauge fields is unambiguously determined similar to unique coupling of the  $\eta'$  field to the gluons, photons and gauge bosons.

As a consequence of this fundamental feature the topological auxiliary  $b(x, H)$  field is in fact an angular topological variable and it has the same  $2\pi$  periodic properties as the original  $\theta$  parameter. As it is known the  $\theta$  parameter can be promoted to the dynamical axion field  $\theta(x)$  by adding the canonical kinetic term  $[\partial_\mu \theta(x)]^2$  to the effective Lagrangian. The difference of the  $b(x, H)$  field with the dynamical axion  $\theta(x)$  field is that the auxiliary topological field  $b(x, H)$  does not have a conventional axion kinetic term.

For simplicity we also assume that  $\overline{\text{QCD}}$  has a single flavor  $N_f = 1$  quark which couples to the non-Abelian  $\overline{\text{QCD}}$  gauge gluons as well as to the  $E$  &  $W$  gauge fields, similar to conventional QCD quarks. This is precisely the coupling which provides the interaction between the (conjectured) high-energy  $\overline{\text{QCD}}$  and the low-energy  $E$  &  $W$  gauge fields. It is natural to assume that the mass of the corresponding  $\bar{\eta}'$  is of order  $m_{\bar{\eta}'} \sim \Lambda_{\overline{\text{QCD}}}$ , similar to the QCD case. Therefore, this heavy d.o.f. can be safely ignored in what follows. In other words, the desired coupling of  $b(x, H)$  auxiliary field with  $E$  &  $W$  gauge field is [5]

$$\mathcal{L}_{b\gamma\gamma}(x) = \frac{\alpha(H)}{8\pi} N Q^2 [\theta - b(x, H)] \cdot F_{\mu\nu} \tilde{F}^{\mu\nu}(x), \quad (49)$$

where  $\alpha(H)$  is the fine-structure constant measured during the period of inflation,  $Q$  is the electric charge of a  $\overline{\text{QCD}}$  quark,  $N$  is the number of colors of the strongly coupled  $\overline{\text{QCD}}$ , and  $F_{\mu\nu}$  is the usual electromagnetic field strength. As we already mentioned, the coupling (49) is unambiguously fixed because the auxiliary  $b(x)$  field always accompanies the so-called  $\theta$  parameter in the specific combination  $[\theta - b(x, H)]$  as explained in Appendix B, and describes the anomalous interaction of the topological auxiliary  $b(x, H)$  field with  $E$  &  $W$  photons. In formula (49) we also ignored the heavy  $\bar{\eta}'$  field which couples in the same way as auxiliary  $b(x, H)$  field, i.e.,  $[\theta - \bar{\eta}' - b(x, H)]$ . However,  $\bar{\eta}'$  field is very heavy as explained above, in

contrast with auxiliary field which generates a topologically protected pole as explained in Appendix B.

The coupling of the  $b(x, H)$  with other  $E$  &  $W$  gauge bosons can be unambiguously reconstructed as explained in [5], but we keep a single  $E$  &  $W$  field  $F_{\mu\nu}$  to simplify the notations and emphasize the crucial elements of the dynamics, related to the helical instability which triggers the end of inflation; see next Sec. VI C.

Based on coupling (49) we present our numerical estimates for number  $N_{\text{infl}}$  of  $e$ -folding in Sec. VI D. Finally, in Sec. VI E we interpret the obtained results and give an intuitive explanation of why and how the nondynamical auxiliary field  $b(x, H)$  can, nevertheless, produce the real physical propagating d.o.f. in a time-dependent background parametrized by  $H$ .

### C. The helical instability and the end of inflation

It has been known for quite some time that the structure of the interaction (49) in many respects has a unique and mathematically beautiful structure with a large number of very interesting features. The most profound property which is crucial for our present analysis of the inflationary Universe is the observation that the topological term (49) along with the conventional Maxwell term  $F_{\mu\nu}^2$  leads to an instability with respect to photon production if  $\dot{b}(x, H)$  does not vanish. This is the so-called helical instability and has been studied in condensed matter literature [32] as well as in particle physics literature including some cosmological applications [33].

In the context of our studies, the closest system where the helical instability develops is the system of heavy ion collisions [34] wherein  $\langle \dot{b}(x, H) \rangle$  can be identified<sup>10</sup> with the so-called axial chemical potential  $\mu_5$ . One can explicitly demonstrate that the interaction (49) leads to the exponential growth of the low-energy modes with

$$k \leq \frac{\alpha(H)\mu_5}{\pi}, \quad \mu_5 \equiv \langle \dot{b}(x, H) \rangle. \quad (50)$$

The growth (50) signals that the instability of the system with respect to production of the real photons develops [34]. It is also known that the fate of this instability is to reduce the axial chemical potential  $\mu_5$  which was the source of this instability. In the inflationary context the corresponding instability reduces  $H$  which plays the role of  $\mu_5$ ; see discussions below. One should also comment here that parameter  $\mu_5$  in heavy ion collisions is also not a dynamical field, but rather is an auxiliary fluctuating field which

<sup>10</sup>The simplest way to demonstrate the correctness of this identification is to perform the path integral  $U(1)_A$  chiral time-dependent transformation to rotate away the coupling (49). The corresponding interaction reappears in the form of a nonvanishing axial chemical potential  $\mu_5$ ; see Appendix B of Ref. [5] with details and references.

accounts for the dynamics of the topological sectors in QCD, similar to our case when  $\langle \dot{b}(x, H) \rangle$  describes the dynamics of the topological sectors in  $\overline{\text{QCD}}$ .

This short detour into the nature of helical instability as a result of interaction (49) has direct relevance to our studies because the auxiliary field  $b(x, H)$  entering Eq. (49) exhibits all the features of parameter  $\mu_5$  which was the crucial element in the analysis of the helical instability in heavy ion collisions. Indeed, both these auxiliary fields originated from the same physics and they both describe the dynamics of the topological sectors in strongly coupled gauge theories.

In terms of physics these nonpropagating fields effectively account for the long range variation of the tunneling processes as a result of some external influence of the backgrounds expressed in terms of  $H$  for inflation and in terms of  $\mu_5$  for heavy ion collisions respectively; see some additional comments on this analogy in Appendix B of Ref. [5].

The net result of the interaction (49) and instability (50) is that the holonomy inflation in this framework inevitably ends by transferring the nondispersive vacuum energy proportional to  $H$  as Eq. (30) states into the real propagating gauge fields. One can interpret this energy transfer as a backreaction to the auxiliary field  $b(x, H)$  as a result of adjustment of the system due to the interaction (49). How can this backreaction effect be in principle computed? The corresponding computations based on the first principles as listed in Sec. VI B by items 1–4 are not presently feasible as we already mentioned. Effective description in terms of the dynamics of the auxiliary field  $b(x, H)$  can be, in principle, carried out along the lines mentioned at the very end of Appendix B.

One may also wonder if the entire vacuum energy will be transferred to the radiation in the form of the SM gauge field, which is the key element for successful graceful exit from inflation. Our comment here is that the transfer of the vacuum energy in this framework is a continuous process, rather than a one-time event. This is obviously the same backreaction effect which is mentioned in the previous paragraph: the radiation decreases the magnitude of the vacuum energy. This process continues as long as the vacuum energy still remains a source of the radiation. This process lasts as long as  $\langle \dot{b}(x, H) \rangle \neq 0$ .

The physical picture of this energy transfer is as follows.<sup>11</sup> Nonvanishing value for  $\langle \dot{b}(x, H) \rangle \neq 0$  leads to the particle production. This radiation of particles obviously decreases the value of  $\langle \dot{b}(x, H) \rangle$  (and the corresponding vacuum energy) as the source of this radiation. In terms of real physical processes this energy transfer corresponds

to the modification of the tunneling transition rate with emission of the real particles in a nontrivial background which also varies. The radiation continues as long as the background deviates from the flat Minkowski spacetime.

The technical description of this energy transfer cannot be carried out in a conventional way, let us say, in terms of physical propagating d.o.f. For example, we cannot model these radiation processes by adding a kinetic term to  $b(x, H)$  field because the corresponding anomalous Ward identities cannot be satisfied with physical propagating d.o.f. as explained in Appendix B. We think that the holographic description mentioned in Appendix B offers a possible framework which potentially can accommodate the dynamics of the auxiliary  $b(x, H)$  field, strange features of the nondispersive vacuum energy and backreaction effects due to the coupling with the SM fields (49). At the present time we do not know yet how to formulate a proper computational framework to answer this question.

To conclude this subsection we comment that the energy transfer between nondynamical auxiliary fields and propagating dynamical fields can, in principle, be tested in a tabletop experiment based on the Maxwell system. We explain the relevant physics and also offer a possible design for a tabletop experiment in Sec. VI E where such an unusual effect can, in principle, be experimentally tested in simplified settings.

#### D. Estimates for the $e$ -folds

The number of  $e$ -folds in the holonomy inflation is determined by the time  $\tau_{\text{inst}}$  when the helical instability fully develops, which explains our subscript  $\tau_{\text{inst}}$ . This is exactly the time scale where a large portion of the energy density  $\rho$  from Eq. (30) which eventually generates the Hubble constant  $H$  according to (40) is transferred to SM light fields. The corresponding time scale for the heavy ion system is known [34] and it is given by  $\tau_{\text{inst}}^{-1} \sim \mu_5 \alpha^2$ . For our system  $\mu_5$  should be interpreted as  $\langle \dot{b}(x, H) \rangle \sim H$ , as the only relevant scale of the problem; see also a few additional arguments in Appendix B supporting this estimate. At the moment  $\tau_{\text{inst}}$  the de Sitter growth (48) cannot be maintained anymore as the source of this behavior  $\sim H$  is completely exhausted due to the transferring of its energy to the gauge fields of the SM.

Therefore we arrive at the following order of magnitude estimate for the number of  $e$ -folds  $N_{\text{infl}}$  in QCD inflationary paradigm,

$$\tau_{\text{inst}}^{-1} \sim H \alpha^2(H), \quad \Rightarrow \quad N_{\text{infl}} \sim \frac{1}{\alpha^2(H)} \sim 10^2, \quad (51)$$

where the number of  $e$ -folds  $N_{\text{Inf}}$  is, by definition, the coefficient in front of  $H^{-1}$  in the expression for the time scale  $\tau_{\text{inst}}$ . At this moment the energy density  $\rho$  from Eq. (30) ceases to exist as the dominant portion of the energy of the system.

<sup>11</sup>The intuitive picture presented below is based on our understanding of the fate of the helical instability in heavy ion collisions leading to reduction in the axial chemical potential  $\mu_5$  which itself is the source of this instability.

The key element of this holonomy inflationary scenario is that the number of  $e$ -folds  $N_{\text{infl}}$  when the de Sitter behavior (48) ends is determined in this framework by the gauge coupling constant  $\alpha(H)$  rather than by dynamics of *ad hoc* inflaton field  $\Phi$  governed by some *ad hoc* inflationary potential  $V(\Phi)$ .

In Sec. VI E we explain the concept of mechanism of the energy transfer at the end of inflation. It is very different from conventional mechanism when propagating inflaton  $\Phi$  couples with physical particles and transfers the energy. In Sec. VI F we compare our framework with the conventional inflationary scenario to show some similarities and differences between the two approaches.

### E. Interpretation

In this subsection we explain a fundamentally new type of particle production which is the key element in all our discussions in this section related to the question how the inflation ends in this framework due to the coupling (49) of the auxiliary field with real physical gauge fields from the SM.

The main point is that the driving force for inflation in this framework is the nondispersive vacuum energy which generates the EoS given by (48). Without anomalous coupling (49) it would be the final destination of the Universe. How does this coupling produce the particles? The main point is that the topological fluctuations with the typical scale  $\sim \Lambda_{\text{QCD}}$  which saturate the vacuum energy are slightly different in the presence of background with scale  $\sim H \ll \Lambda_{\text{QCD}}$ . This time-dependent background generates the particle production with the rate  $\sim H$  which is precisely the reason why inflation eventually ends in this framework on the time scale (51).

We test this mechanism of the particle production from nondispersive vacuum energy using the Maxwell theory as a playground. The corresponding Maxwell system can, in principle, be designed and fabricated with existing technology; see the relevant references in concluding Sec. VII C. Therefore, in principle, this novel phenomenon can be tested in a tabletop experiment in a lab.

The basic idea is that there is a new contribution to the Casimir pressure which emerges as a result of tunneling processes when the Maxwell system is formulated on a nontrivial manifold permitting the  $E$  &  $M$  configurations with nontrivial topology  $\pi_1[U(1)] = \mathbb{Z}$ . Precisely these tunneling transitions between physically identical but topologically distinct states play the same role in the Maxwell system as the topologically nontrivial configurations in QCD. The corresponding extra energy generated due to these transitions is the direct analog of the nondispersive contribution to the energy which is the key player of the present work as it explicitly enters (30), (28), and (38) in previous sections. This nondispersive energy in the Maxwell system is similar to our studies of the non-Abelian gauge theories reviewed in Appendix A; this extra

energy also cannot be formulated in terms of conventional propagating photons with two transverse polarizations.

If the same system is considered in the background of a small external time-dependent field, then real physical particles are emitted from the vacuum, similar to the dynamical Casimir effect (DCE) when photons are radiated from the vacuum due to time-dependent boundary conditions. Essentially, the reheating epoch as advocated in this section when the vacuum energy can radiate real particles in a time-dependent background is analogous to the DCE. The difference is that in conventional DCE the virtual particles from vacuum become real propagating particles in a time-dependent background and get emitted. In our case the  $E$  &  $M$  configurations which describe the interpolations between different topological sectors get excited in time-dependent background and emit real particles; see concluding Sec. VII C for references and details.

We hope this intuitive explanation provides the basic conceptual picture on how the particles can be produced from the vacuum, which represents the key element of the graceful exit from inflation.

### F. Relation to the conventional inflationary scenario

The goal of this section is to collect a number of comments made in different places in this work related to the (possible) connection between our framework and conventional description in terms of a scalar inflation  $\Phi(x)$  governed by a potential  $V[\Phi]$ . For obvious reasons this is not a one-to-one correspondence between drastically different descriptions. Nevertheless, these comments, hopefully, may generate some thoughts about the source of the vacuum energy in nature, and find a proper technical framework to describe it.

We start with a few generic remarks. The topological inflationary mechanism as formulated in this proposal is fundamentally nonlocal in nature and cannot be modeled by any local effective inflationary potential  $V(\Phi)$ . Furthermore, this mechanism is fundamentally “nondispersive” in nature and cannot be described in terms of any propagating physical d.o.f. such as inflaton  $\Phi(x)$  with canonical kinetic term  $(\partial_\mu \Phi(x))^2$ . Further to this point, we introduced the topological auxiliary fields  $a(x, H)$  and  $b(x, H)$  in Appendix B to describe the physics in terms of effective long range fields which, in principle, should describe the relevant IR physics. These fields are not propagating, in contrast with the inflaton  $\Phi(x)$  field. The physical meaning of these fields as explained in Appendix B is as follows: the  $\square a(x, H)$  describes the distribution of the topological density in the system, while  $b(x, H)$  acts as the axion field (without kinetic term) being the source of the topological density distribution.

These obvious differences between drastically different frameworks must obviously lead to distinct observational results. In particular, the conventional computations of the cosmological perturbations are based on treating the

inflaton  $\Phi(x)$  as the conventional scalar field with canonical kinetic term  $(\partial_\mu\Phi(x))^2$ . The corresponding results can be expressed in terms of the vacuum energy  $\rho$  and pressure  $P$  as it is formulated in [3]. However, mere existence of a local inflaton field  $\Phi(x)$  has been assumed in computations in [3], while the final results are presented in terms of energy-momentum tensor. Computations in our framework require a different technique, which is not yet developed as explained at the very beginning of this Sec. VI. Therefore, it is natural to expect that the outcome would be different even when the final results are expressed in terms of the energy-momentum tensor's parameters  $\rho$  and  $P$ . However, as the corresponding technical tools are not yet developed, it is very hard to quantify the corresponding differences.

In what follows we make a few comments on some similarities between these two distinct approaches. In particular, we identify (on an intuitive level) the topological auxiliary fields  $a(x, H)$  and  $b(x, H)$  with the inflaton  $\Phi(x)$  field in a sense that both fields eventually generate the de Sitter behavior, and both approaches lead to the inflationary EoS (48). The fundamental difference between the two is that the inflaton  $\Phi(t)$  satisfies the classical equation of motion and depends on time  $t$ , while  $a(x, H)$  and  $b(x, H)$  are truly quantum objects, such that all observables in principle must be expressed exclusively in terms of the correlation functions and expectation values when the time dependence enters the physics exclusively in terms of the Hubble parameter  $H$ .

Still, there are some hints which apparently suggest that some links between the two approaches may exist.

Indeed, let us introduce a few important parameters which are normally used in conventional inflationary analysis and compare them with our description. For this purpose we introduce conventional slow-roll parameters; see e.g., [35]

$$\epsilon = \frac{M_P^2}{2} \left( \frac{V_\phi}{V} \right)^2, \quad V_\phi \equiv \frac{\partial V[\phi]}{\partial \phi}. \quad (52)$$

For example, the computation of the number of  $e$ -foldings in conventional slow-roll approximation and estimates (51) in the holonomy inflationary scenario both produce numerically large magnitudes. In the conventional approach one can use the following relation [35]:

$$N_{\text{infl}} \simeq \frac{1}{M_P^2} \int_{\phi_{\text{end}}}^{\phi} d\phi \left( \frac{V}{V_\phi} \right). \quad (53)$$

The large numerical value for  $N_{\text{infl}} \gg 1$  in the conventional approach is due to the specific choice of the potential (52) when the integrand entering (53) is parametrically large and proportional to  $\epsilon^{-1}$ . It should be compared with the holonomy inflationary scenario when  $N_{\text{infl}} \gg 1$  is parametrically large due to the enhancement factor  $\alpha^{-2}$  as estimate (51) suggests.

We conclude this section with a few generic comments. First of all, while we identify (on the intuitive level) the auxiliary topological fields with inflaton, the  $a(x, H)$  and  $b(x, H)$  fields remain to be quantum (not classical) fluctuating fields, saturating the relevant correlation functions. We observed above that there are a number of instances when the holonomy inflationary scenario behaves very much in the same way as the conventional description represented by formulas (52) and (53) discussed above. Is it a coincidence or is there a deeper reason for these relations?

We formulate the same question in a different way: Is it possible to make any connection between conventional description in terms of auxiliary  $a(x, H)$  and  $b(x, H)$  fields and local inflaton  $\Phi(x)$  field which satisfies the classical equation of motion determined by the potential  $V[\Phi]$ ? We do not know how to do it. The main obstacle to make such a connection is related to the fact that the auxiliary topological fields, by construction (reviewed in Appendix B), saturate the topological susceptibility (and the corresponding vacuum energy) with the positive sign according to (A3) and (A5), generating the topologically protected pole (A4), while any conventional d.o.f. (including dynamical propagating inflaton) can only produce a negative sign according to (A2).

One possible path to overcome this obstacle is to define the auxiliary fields<sup>12</sup> using the holographic description along the lines suggested in [36]. In this case the axion field which is represented by our auxiliary field  $b(x, H)$  becomes the dynamical propagating field in the bulk of multidimensional space but acts as a conventional (non-dynamical) term on the boundary (representing our spacetime). This feature is precisely what is required for our auxiliary field  $b(x, H)$  defined on physical spacetime.

## VII. CONCLUDING COMMENTS

We conclude this work with formulation of our basic results in Sec. VII A. We next formulate the profound consequences of our framework in Sec. VII B. To convince the readers that we study a real physical effect, we suggest testing this new nondispersive type of vacuum energy in a laboratory using the physical Maxwell system as highlighted in Sec. VII C. Finally, we make a few comments on the relation of our approach with no-boundary and tunneling proposals in Sec. VII D.

### A. Basic results

The heart of the proposal suggested in the present work is a synthesis of two, naively unrelated, ideas.

<sup>12</sup>We recall the physical meaning of the auxiliary fields: the  $\square a(x, H)$  describes the distribution of the topological density in the system, while  $b(x, H)$  acts as the axion field (without kinetic term) being the source of the topological density distribution.

The first idea represents the self-consistent treatment of the problem formulated on the Euclidean  $\mathbb{S}^3 \times \mathbb{S}^1$  manifold through the bootstrap equation [8–10].

The second novel idea [5–7] is a proposal to treat the vacuum energy entering the Friedmann equation as a nondispersive vacuum energy which is always generated in non-Abelian gauge theories as a result of tunneling transitions between topologically nontrivial sectors in a system. This type of energy is very unusual in many respects. First of all, it is nonanalytical in coupling constant  $\sim \exp(-1/g^2)$  and cannot be seen in perturbation theory as reviewed in Appendix III C. Secondly, this vacuum energy is nonlocal in nature as it cannot be expressed in terms of any local operators in a gradient expansion in any effective field theory. Rather, it can be expressed in terms of the nonlocal holonomy, similar to Aharonov-Casher effect as mentioned in Sec. III.

We coin the marriage of these two sets of ideas as the holonomy inflation which has a number of very attractive and desirable features. First of all, there is the hierarchy of scales for both models given by Eqs. (34) and (41) correspondingly which indicates that the distances smaller than Planck scale  $M_P^{-1}$  never appear in our framework. Secondly, the equation of state (48) assumes its de Sitter behavior as a result of nucleation as Fig. 2 shows. Thirdly, the number of  $e$ -folds  $N_{\text{infl}}$  is naturally determined by the gauge coupling constant  $\alpha(H)$  as Eq. (51) suggests.

## B. Implications and future development

There are a few important and generic consequences of this framework.

1. The conventional scenarios of the eternal self-producing inflationary universes are always formulated in terms of a physical scalar dynamical inflaton field  $\Phi(x)$ . This problem with self-reproduction of the Universe does not even emerge in our framework as there are no any fundamental scalar dynamical fields in the system responsible for inflation. Instead, the de Sitter behavior in our framework is a pure quantum phenomenon, which is a consequence of the dynamics of the long ranged topological configurations with nontrivial holonomy, rather than a result of a physical fluctuating dynamical field. This type of energy manifests itself in terms of the “wrong” sign in the correlation function which cannot be formulated in terms of any local propagating d.o.f. as explained in Appendix A 1. Therefore, the problem with eternal inflation does not even occur in our framework.

2. There are many other problems in conventional formulation of the inflation in terms of scalar inflaton field  $\Phi(x)$ . For example, the initial value  $\Phi_{\text{in}} \gg M_{\text{PL}}$  for the inflaton is normally very large. This problem does not occur in our holonomy inflation scenario as the hierarchy of scales (34) always holds in our framework.

3. We should also mention that the energy described by a formula similar to Eqs. (19) and (30), which eventually

leads to the de Sitter behavior (36), has been previously postulated [37–39] as the driving force for the dark energy (admittedly, without much deep theoretical understanding behind the formula at that time). The model has been (successfully) confronted with observations<sup>13</sup>; see recent review papers [40,41], and many original references therein, where it has been claimed that this proposal is consistent with all presently available data.

Our comment here is that history of evolution of the Universe may repeat itself by realizing the de Sitter behavior twice in its history. The  $\overline{\text{QCD}}$ -dynamics is responsible for the holonomy inflation considered in present work, while the QCD dynamics is responsible for the dark energy in the present epoch. In this case the dark energy (DE) density is given by an expression similar to (30); i.e.,  $\rho_{\text{DE}} \sim H\Lambda_{\text{QCD}}^3 \sim (10^{-3} \text{ eV})^4$  is amazingly close to the observed value without any fine-tunings or adjustments of the parameters.

4. One should also mention that some recent lattice simulations [42] implicitly support our results. Indeed, the author of Ref. [42] studied the rate of particle production in the de Sitter background. The rate turns out to be linearly proportional to the Hubble constant  $\sim H$ , rather than naively expected  $H^2$ . It is fully consistent with our proposal.<sup>14</sup> We hope that some further lattice computations in time-dependent background can further elucidate the role of holonomy in generating the vacuum energy.

5. Finally, we make a comment about possible future development. As we already mentioned at the beginning of Sec. VI the relevant technique describing the end of inflation in our framework (including computations of the cosmological perturbations) is yet to be developed. We already mentioned in the text a number of technical challenging problems which need to be resolved, and do not repeat them here in the conclusion.

<sup>13</sup>We note that the structure of the relevant vacuum energy which enters the Friedman equations (19) and (30) is determined by the size of  $\mathbb{S}^1$  and behaves in all respects as the cosmological constant. Therefore, it is obviously consistent with presently available data as it does not modify the equation of state as explained in Appendix III C.

<sup>14</sup>Indeed, the rate of the particle production in quantum field theory in general is determined by the imaginary part of the stress tensor,  $\text{Im}[T_{\mu}^{\nu}]$ , while the vacuum energy is related to the real part of the stress tensor,  $\text{Re}[T_{\mu}^{\nu}]$ . Analyticity suggests that both components must have the same corrections on  $H$  at small  $H$ . Therefore, the lattice measurements [42] of the linear dependence on  $H$  of the particle production strongly suggest that the vacuum energy (which is determined by the real part of the same stress tensor) must also exhibit the same linear  $\sim H$  correction. The corresponding lattice computations of the  $\theta$ -dependent portion of the vacuum energy and topological susceptibility in time-dependent background are possible, in principle, but are technically much more involved than the analysis performed in Ref. [42].

### C. Possible tests of the cosmological ideas in a lab?

Our comment here is that we cannot experimentally test the first element of the proposal advocated in [8–10] in any simplified settings. However, we can test the second element of this proposal advocated in [5–7] in tabletop experiments. This subsection should convince the readers that we are dealing with new physical phenomena which can be realized in cosmology (which is the subject of the present work) as well as in the Maxwell  $U(1)$  gauge theory.

The basic idea goes as follows. The fundamentally new type of energy advocated in the present work can, in principle, be studied in a tabletop experiment by measuring some specific corrections to the Casimir vacuum energy in the Maxwell theory as suggested in [43–47]. This fundamentally new contribution to the Casimir pressure emerges as a result of tunneling processes, rather than due to the conventional fluctuations of the propagating photons with two physical transverse polarizations. Therefore, it was coined as the topological casimir effect. The extra energy computed in [43–47] is the direct analog of the QCD nondispersive vacuum energy (A10) and (24) which is the key player of the present work as it explicitly enters (37), (28), and (38) in the main text. In fact, an extra contribution to the Casimir pressure emerges in this system as a result of nontrivial holonomy similar to (A6) for the Maxwell field. The nontrivial holonomy in the  $E$  &  $M$  system is enforced by the nontrivial boundary conditions imposed in Refs. [43–47], and related to the nontrivial mapping  $\pi_1[U(1)] = \mathbb{Z}$  relevant for the Maxwell Abelian gauge theory. Furthermore, the reheating epoch when the physical particles can be emitted from the vacuum in a time-dependent background, similar to the dynamical Casimir effect, can also be tested in the Maxwell system as argued in [46].

A similar new type of energy can, in principle, also be studied in the superfluid He-II system which also shows a number of striking similarities with non-Abelian QCD as argued in [48]. For the superfluid He-II system the crucial role is played by the vortices which are classified by  $\pi_1[U(1)] = \mathbb{Z}$  similar to the Abelian quantum fluxes studied in the Maxwell system in [43–47].

### D. Cosmological density matrix vs no-boundary and tunneling states

We conclude this section with a few comments on status of the density matrix initial conditions in cosmology (which is the key element of the present work) as compared to the well-known no-boundary [11] and tunneling [49–51] proposals for the wave function of the Universe.

As is known, observer independent treatment of the no-boundary state leads to an insufficient amount of inflation. Phenomenologically, the volume weighting [52,53] or top-down approach [54] to the no-boundary state seems to resolve this issue but remains with the problem of

consistency of complex tunneling geometries and normalizability of the quantum ensemble in cosmology.

On the other hand, the tunneling state has a rather uncertain ground based on the hyperbolic rather than Schroedinger nature of the Wheeler-DeWitt equation. No-boundary wave function within the Euclidean path integral construction represents a special quasivacuum state. The tunneling state within the approach of path integration over Lorentzian geometries leads to nonnormalizable wave function with unstable quantum matter and gravity perturbations. This fact has been known since [50], long before the recent works [55–57] which extended this criticism also to the no-boundary wave function.

Diversity of the definitions of the no-boundary and tunneling states (defined either as propagators or solutions of the homogeneous Wheeler-DeWitt equation either in Euclidean or Lorentzian spacetime) as discussed in [55–57] actually indicates that neither of these states has a rigorous canonical quantization ground. However, the critical verdict of [55–57] invalidating both the no-boundary and tunneling states, though it requires deeper consideration, does not actually achieve its goal. This is because what is actually required is not the construction of the wave function itself, but rather scattering amplitudes, mean values and probabilities generated by it. The step from the wave function (or the density matrix) to these quantities is very nontrivial and requires additional integration over the end points of the path integral histories. This integration can also run along the complex contours of the steepest decent approximation; it can bear UV divergences and might lead to the effects invalidating the main conclusions of [55–57].<sup>15</sup>

This is exactly what is done in the microcanonical density matrix setup of [8–10]—we do not calculate the density matrix itself, but directly go over to its partition function dominated by the real valued periodic history in Euclidean spacetime. The starting point is the microcanonical density matrix of a spatially closed cosmology, which is defined as a projector on the space of solutions of the Wheeler-DeWitt equations—quantum Dirac constraints of the canonical quantization of gravity in physical Lorentzian spacetime [10]. The periodicity of the relevant saddle-point histories directly follows from the tracing procedure for the normalization of the density matrix (see Fig. 1), and their Euclideanization is the corollary of the fact that periodic solutions exist only in the imaginary (Euclidean) time, which is equivalent to the integration over the complex contour of the lapse ADM function [10].

Thus, our approach differs from the methods of [11,49] and [55–57] in two major points—first, the microcanonical

<sup>15</sup>This extra integration requires the selection of saddle points in the complex plane via the technique similar to that of [55–57] which can unpredictably alter the results of this work.

density matrix prescription for the initial state of the Universe rather than the pure state wave function and, secondly, the calculation of the physical quantity—partition function—rather than the wave function or the density matrix. Conceptual rigidity of this construction avoids ambiguities of the approach of [11,49,55–57] and unambiguously leads to  $\mathbb{S}^1$ -compactification of the Euclidean time bearing the holonomy of the gauge field—the corner stone of the strongly coupled nonperturbative QCD-like theory and its effect of generating the vacuum energy.

### ACKNOWLEDGMENTS

This project was initiated during the workshop “Quantum Vacuum and Gravitation: Testing General Relativity in Cosmology,” Mainz Institute for Theoretical Physics, Mainz. We are thankful to the organizers of this workshop for creating an amazing atmosphere and the opportunity to discuss some unorthodox ideas during the official talks and special evening sessions. The work of A. O. B. was supported by the Russian Foundation for Basic Research (RFBR) Grant No. 17-02-00651 and by the Yukawa Institute for Theoretical Physics. A. R. Z. was supported in part by the Natural Sciences and Engineering Research Council of Canada.

### APPENDIX A: THE NATURE OF THE NONDISPERSIVE VACUUM ENERGY

The main goal of this Appendix is to review a number of crucial elements relevant for our studies of the nondispersive vacuum energy and its cosmological significance. First, we start in Sec. A 1 with explanation of a highly nontrivial nature of this type of the vacuum energy in the Euclidean spacetime.

This type of the vacuum energy is well known to the QCD practitioners, while it is much less known in the general relativity and cosmology communities. We think this ignorance can be explained by the fact that this unusual type of the vacuum energy cannot be formulated in terms of conventional local propagating d.o.f. Precisely such a local formulation is a conventional framework for the cosmology community when the inflation or the dark energy is described in terms of a scalar field, such as inflaton  $\Phi(x)$  with specifically adjusted local potential  $V[\Phi]$ . On the other hand, this unusual type of energy has been known to the QCD community for quite some time. Furthermore, this unusual nondispersive nature of the vacuum energy has been supported by numerous lattice simulations; see A 1 with references and details.

We continue in Sec. A 2 by clarifying the crucial role of the holonomy (A6) in generating such a type of energy. We review a few known analytical calculations of this type of energy by emphasizing the role of the nonlocal holonomy (computed along  $\mathbb{S}^1$ ) which generates this

unusual energy. The  $\mathbb{S}^1$  in these computations represents an important portion of a larger Euclidean four-dimensional manifold  $\mathbb{S}^3 \times \mathbb{S}^1$ , which has been extensively employed in the main text of this work; see Secs. IV and V.

In Sec. A 3 we make a few historical remarks on fractionally charged topological objects as they are intimately related to nontrivial holonomy defined on  $\mathbb{S}^1$ .

### 1. The topological susceptibility and contact term in flat spacetime

We start our short overview on the nondispersive nature of the vacuum energy by reviewing a naively unrelated topic—the formulation and resolution of the so-called  $U(1)_A$  problem in strongly coupled QCD [58–60]. We introduce the topological susceptibility  $\chi$  which is ultimately related to the vacuum energy  $E_{\text{vac}}(\theta = 0)$  as follows,<sup>16</sup>

$$\chi = \left. \frac{\partial^2 E_{\text{vac}}(\theta)}{\partial \theta^2} \right|_{\theta=0} = \lim_{k \rightarrow 0} \int d^4x e^{ikx} \langle T\{q(x), q(0)\} \rangle, \quad (\text{A1})$$

where  $\theta$  parameter enters the Lagrangian along with topological density operator  $q(x) = \frac{1}{16\pi^2} \text{tr}[F_{\mu\nu} \tilde{F}^{\mu\nu}]$  and  $E_{\text{vac}}(\theta)$  is the vacuum energy density computed for the Euclidean infinitely large flat spacetime. This  $\theta$ -dependent portion of the vacuum energy (computed at  $\theta = 0$ ) has a number of unusual properties as we review below. The corresponding properties are easier to explain in terms of the correlation function (A1), rather than in terms of the vacuum energy  $E_{\text{vac}}(\theta = 0)$  itself. The relation between the two is given by Eq. (A1).

A few comments are in order. First of all, the topological susceptibility  $\chi$  does not vanish in spite of the fact that  $q(x) = \partial_\mu K^\mu(x)$  is total derivative. This feature is very different from any conventional correlation functions which normally must vanish at zero momentum  $\lim_{k \rightarrow 0}$  if the corresponding operator can be represented as total divergence.

Secondly, any physical  $|n\rangle$  state gives a *negative* contribution to this diagonal correlation function,

$$\begin{aligned} \chi_{\text{dispersive}} &\sim \lim_{k \rightarrow 0} \int d^4x e^{ikx} \langle T\{q(x), q(0)\} \rangle \\ &\sim \lim_{k \rightarrow 0} \sum_n \frac{\langle 0|q|n\rangle \langle n|q|0\rangle}{-k^2 - m_n^2} \simeq -\sum_n \frac{|c_n|^2}{m_n^2} \leq 0, \quad (\text{A2}) \end{aligned}$$

where  $m_n$  is the mass of a physical  $|n\rangle$  state,  $k \rightarrow 0$  is its momentum, and  $\langle 0|q|n\rangle = c_n$  is its coupling to topological density operator  $q(x)$ . At the same time the resolution of the  $U(1)_A$  problem requires a positive sign for the topological

<sup>16</sup>We use the Euclidean notations where path integral computations are normally performed.

susceptibility (A1); see the original reference [60] for a thorough discussion,

$$\chi_{\text{nondispersive}} = \lim_{k \rightarrow 0} \int d^4x e^{ikx} \langle T\{q(x), q(0)\} \rangle > 0. \quad (\text{A3})$$

Therefore, there must be a contact contribution to  $\chi$ , which is not related to any propagating physical d.o.f., and it must have the wrong sign. The wrong sign in this paper implies a sign which is opposite to any contributions related to the physical propagating d.o.f. (A2). The corresponding vacuum energy associated with nondispersive contribution to the topological susceptibility  $\chi$  as defined by (A1) can be coined as nondispersive vacuum energy  $E_{\text{vac}}(\theta = 0)$ . It is quite obvious that the nature of this energy is drastically different from any types of conventional energy because it *cannot* be formulated in terms of any conventional propagating d.o.f. according to (A2) and (A3). In the cosmological context relevant for the present work this type of energy in Refs. [5–7] was dubbed as the strange energy, while a scientific name would be the nondispersive vacuum energy  $E_{\text{vac}}(\theta = 0)$  generated by the contact term in the correlation function (A3). It should be contrasted with the “dispersive” energy which, by definition, is associated with some propagating d.o.f. and can be always restored from the absorptive portion of the correlation function through the dispersion relations according to (A2).

In the framework [58] the contact term with wrong sign has been simply postulated, while in Refs. [59,60] the Veneziano ghost (with a wrong kinetic term) had been introduced into the theory to saturate the required property (A3).

Our next comment is the observation that the contact term (A3) has the structure  $\chi \sim \int d^4x \delta^4(x)$ . The significance of this structure is that the gauge variant correlation function in momentum space

$$\lim_{k \rightarrow 0} \int d^4x e^{ikx} \langle K_\mu(x), K_\nu(0) \rangle \sim \frac{k_\mu k_\nu}{k^4} \quad (\text{A4})$$

develops a topologically protected “unphysical” pole which does not correspond to any propagating massless d.o.f., but nevertheless must be present in the system. Furthermore, the residue of this pole has the wrong sign which saturates the nondispersive term in gauge invariant correlation function (A3),

$$\langle q(x)q(0) \rangle \sim \langle \partial_\mu K^\mu(x), \partial_\nu K^\nu(0) \rangle \sim \delta^4(x). \quad (\text{A5})$$

We conclude this review-type subsection with the following remark. The entire framework, including the singular behavior of  $\langle q(x)q(0) \rangle$  with the wrong sign, has been well confirmed by numerous lattice simulations in strong coupling regime, and it is accepted by the community as a standard resolution of the  $U(1)_A$  problem. Furthermore, it has been argued long ago in Ref. [61] that the gauge

theories may exhibit the “secret long range forces” expressed in terms of the correlation function (A4) with topologically protected pole at  $k = 0$ .

Finally, in a weakly coupled gauge theory (the so-called “deformed QCD” model [62]) where all computations can be performed in a theoretically controllable way, one can explicitly test every single element of this entire framework, including the topologically protected pole (A4), the contact term with wrong sign, etc.; see Refs. [63–65] for the details. In particular, one can explicitly see that the Veneziano ghost is in fact an auxiliary topological field which saturates the vacuum energy and the topological susceptibility  $\chi$ . It does not violate unitarity, causality and any other fundamental principles of a quantum field theory. What is more important for the present studies is that one can explicitly see that the holonomy (A6) plays a crucial role in generating the strange vacuum energy defined in terms of the correlation function (A1).

While all these unusual features of the vacuum energy are well known and well supported by numerous lattice simulations in strongly coupled regime (see e.g., [63] for a large number of references on original lattice results), a precise quantitative understanding of these properties (on a level of analytical computational scheme) is still lacking. In the next subsection we review some known results on this matter specifically emphasizing on role of the holonomy (A6) in the analytical computations. Precisely a nontrivial holonomy (A6) plays a crucial role in generating the strange vacuum energy as we argue in next Sec. A2. This is the key technical element which pinpoints the source of this novel type of energy not expressible in terms of any local operators as the holonomy is obviously a nonlocal object.

## 2. The holonomy (A6) and generation of the nondispersive vacuum energy

Our goal here is to argue that the holonomy plays a key role in generation of the nondispersive vacuum energy in the system. We also compare the vacuum energy computed on different manifolds such as  $\mathbb{S}^1 \times \mathbb{R}^3$  versus  $\mathbb{S}^1 \times \mathbb{H}^3$  and  $\mathbb{S}^1 \times \mathbb{S}^3$ . Such studies play the crucial role in our analysis in the main text in Sec. II devoted to construction of the gravitation instanton formulated on  $\mathbb{S}^1 \times \mathbb{S}^3$ .

We start our analysis with  $\mathbb{S}^1 \times \mathbb{R}^3$  geometry. The key role in the discussions is played by the behavior of holonomy  $U(\mathbf{x}) \equiv \mathcal{P} \exp(i \int_0^T dx_4 A_4(x_4, \mathbf{x}))$  at spatial infinity, the Polyakov line,

$$L = \mathcal{P} \exp \left( i \int_0^T dx_4 A_4(x_4, |\mathbf{x}| \rightarrow \infty) \right). \quad (\text{A6})$$

The operator  $\text{Tr}L$  classifies the self-dual solutions which may contribute to the path integral at finite temperature  $T \equiv T^{-1}$ , including the low temperature limit  $T \rightarrow 0$ . There is a well-known generalization of the standard

self-dual instantons to nonzero temperature, which corresponds to the description on  $\mathbb{R}^3 \times \mathbb{S}^1$  geometry. These are so-called periodic instantons, or calorons [66], studied in detail in [67]. These calorons have trivial holonomy, which implies that the  $\text{Tr}L$  assumes values belonging to the group center  $\mathbb{Z}_N$  for the  $SU(N)$  gauge group.

A more general class of the self-dual solutions with nontrivial holonomy (A6), the so-called KvBLL calorons, was constructed much later in Refs. [68,69]. In this case the holonomy (A6) in general is not reduced to the group center  $\text{Tr}L \notin \mathbb{Z}_N$ . The fascinating feature of the KvBLL calorons is that they can be viewed as a set of  $N$  monopoles of  $N$  different types. Normally, one expects that monopoles come in  $N - 1$  different varieties carrying a unit magnetic charge from each of the  $U(1)$  factors of the  $U(1)^{N-1}$  gauge group left unbroken by vacuum expectation value due to nontrivial holonomy (A6). There is an additional, so-called Kaluza-Klein monopole which carries magnetic charges and instanton charge. All monopoles' charges are such that when a complete set of different types of monopoles is present, the magnetic charges exactly cancel, and the configuration of  $N$  different monopoles carries a unit instanton charge. In particular, for  $SU(2)$  gauge group the holonomy

$$\frac{1}{2}\text{Tr}L = \cos(\pi\nu) \quad (\text{A7})$$

belongs to the group center  $\frac{1}{2}\text{Tr}L = \pm 1$  when  $\nu$  assumes the integer values (trivial holonomy). The so-called ‘‘confining’’ value for the holonomy corresponds to  $\nu = 1/2$  when  $\text{Tr}L = 0$  vanishes.

It has been known since [67] that the gauge configurations with nontrivial holonomy are strongly suppressed in the partition function. Therefore, naively KvBLL calorons cannot produce a finite contribution to the partition function. However, this naive argument is based on consideration of the individual KvBLL caloron, or a finite number of them. If one considers a grand canonical assemble of these objects than their density is determined by the dynamics, and the old argument of Ref. [67] does not hold anymore. The corresponding objects in this case may in fact produce a finite contribution to the partition function. A self-consistent computation in a weak coupling regime supporting this picture has been carried out in the so-called deformed QCD model [62]. One can explicitly see how  $N$  different types of monopoles with nontrivial holonomy (A6) which carry fractional topological charge  $\pm 1/N$  produce confinement, generate the strange vacuum energy (A1) and associated with this energy the topological susceptibility (A5) with known, but highly unusual properties reviewed above in previous Sec. A 1; see [63–65] for the technical details on these computations.

In the strong coupling regime we are interested in, the corresponding analytical computations have never been

completed. There is a limited number of partial analytical and numerical results [70–72] on computations of moduli space and one loop determinant, controlling the dynamics and interaction properties of the constituents in a large ensemble of KvBLL calorons.

While complete analytical solution in strong coupling regime is still lacking, nevertheless there are a number of hints supporting the basic picture that the KvBLL configurations with nontrivial holonomy (A6) and representing  $N$  different types of monopoles with fractional topological charges  $\pm 1/N$  saturate the strange vacuum energy (A1) and associated with this energy the topological susceptibility (A5) in very much the same way as it happens in the deformed QCD model where all computations are performed in a theoretically controllable regime [62–64]. It is assumed in what follows that the topological susceptibility (A1) and associated with it the nondispersive vacuum energy  $E_{\text{vac}}(\theta)$  is indeed saturated by fractionally charged monopoles with  $Q = \pm 1/N$  which are constituents of KvBLL caloron with nontrivial holonomy (A6) and (A7).

The corresponding computations of the partition function and the free energy for the vacuum ground state for  $\mathbb{S}^1 \times \mathbb{R}^3$  geometry lead to the following result [70–72]:

$$\begin{aligned} \mathcal{Z} &\simeq \exp[4\pi f V], & f &= \frac{4\pi\Lambda_{\text{QCD}}^4}{g^4 T} \\ F_{\text{vac}} &= -T \ln \mathcal{Z} = -\frac{32\pi^2}{g^4} \Lambda_{\text{QCD}}^4 V, \end{aligned} \quad (\text{A8})$$

where  $V$  is the 3-volume of the system,  $g$  is the coupling constant of a non-Abelian gauge field, and the  $\Lambda_{\text{QCD}}$  is a single dimensional parameter of the system generated as a result of dimensional transmutation in classically conformal gauge theory, similar to conventional  $\Lambda_{\text{QCD}} \simeq 170$  MeV in QCD physics. Parameter  $f$  in (A8) can be interpreted as the monopoles' fugacity of the system, while the combination  $T F_{\text{vac}} \equiv E_{\text{vac}} V^{(4)} \equiv E_{\text{vac}} T V$  shows the extensive property when  $\ln \mathcal{Z}$  is proportional to the Euclidean 4-volume at large  $V^{(4)} \rightarrow \infty$ . In this framework  $E_{\text{vac}}$  has dimension 4 and represents the vacuum energy density of the system entering the fundamental formula (A1) and defining the nondispersive portion of the vacuum energy.

One can show that free energy (A8) as well as the topological susceptibility  $\chi$  demonstrate all the features of the strange energy briefly described in Sec. A 1 in a model-independent generic way, including the wrong sign for  $\chi$  which cannot be associated with any physical propagating d.o.f. The specific mechanism based on the KvBLL configurations reviewed above and describing the tunneling processes between the distinct topological sectors precisely generates all these required properties. In what follows we assume that the very same mechanism generates the nondispersive vacuum energy density  $E_{\text{vac}}$  for different

geometries, including  $\mathbb{S}^3 \times \mathbb{S}^1$  and  $\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1$  exactly in the same way as computed above for  $\mathbb{R}^3 \times \mathbb{S}^1$ .

With this assumption in hand the question we address below is as follows. How does the strange energy density  $E_{\text{vac}}$  vary if the geometry is slightly modified at large distances? The main motivation for this question is originated from our fundamental conjecture formulated in Sec. IV that the energy density which enters the Friedman equation represents in fact the difference  $\Delta E$  between the energy density computed in a nontrivial background by subtracting the ‘‘trivial’’ portion computed in the flat background similar to the Casimir-type computations.

Specifically, we want to know how the vacuum energy density depends on the geometry  $\mathbb{S}^3 \times \mathbb{S}^1$ . Precisely this information is required in our computations in Secs. IV and V. Unfortunately, there are a number of technical obstacles to carry out the computations similar to (A8) for  $\mathbb{S}^3 \times \mathbb{S}^1$  manifold. In particular, even the monopole solution (which is the crucial ingredient in this type of semiclassical computation) satisfying the appropriate boundary conditions on  $\mathbb{S}^3 \times \mathbb{S}^1$  is not exactly known. As a result of this deficiency, a semiclassical computation which would account for zero and nonzero mode contributions to the partition function [similar to formula (A8) derived for  $\mathbb{R}^3 \times \mathbb{S}^1$ ] is also not known.

Fortunately, the exact semiclassical computations are available for the hyperbolic space  $\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1$ . While this manifold is not exactly what we need for our analysis in Secs. IV and V, nevertheless, the corresponding computations give us a hint on possible corrections to the vacuum energy density  $E_{\text{vac}}$  due to a small dimensional parameter  $\sim \kappa$  which emerges in the  $\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1$  in comparison with computations (A8) corresponding to  $\mathbb{R}^3 \times \mathbb{S}^1$  geometry.

The main reason why the semiclassical computations can be carried out in hyperbolic space  $\mathbb{H}_\kappa^3$  with the constant negative curvature  $-\kappa^2$  is as follows. There is a conformal equivalence between  $(\mathbb{R}^4 - \mathbb{R}^2)$  and  $\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1$  where  $\mathbb{S}_{\kappa^{-1}}^1$  denotes the circle of radius  $\kappa^{-1}$ . As a result of this exact equivalence, the monopole’s solutions can be explicitly constructed in this case. The holonomy (A6) is computed along a closed loop  $\mathbb{S}_{\kappa^{-1}}^1$  and assumes a nontrivial value.

The key observation of this computation, see formula (A9) below, is that the topological configurations with nontrivial holonomy produce a finite contribution to the vacuum energy density with a small correction being linearly proportional to  $\kappa \rightarrow 0$ . This effect cannot be expressed in terms of any local operators such as curvature as  $|R| \sim \kappa^2$ . Rather, the leading correction  $\sim \kappa$  is generated due to topological vacuum configurations with nontrivial holonomy, not expressible in terms of any local observables. This is precisely the reason why the generic arguments [28–30] based on locality simply do not apply here.

Now we are ready to formulate the main result of the computations [7] relating the vacuum energy density  $E_{\text{vac}}$  computed on the original  $\mathbb{R}^3 \times \mathbb{S}^1$  manifold and on the hyperbolic space  $\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1$ . In formula (A9) presented below we assume that the sizes of  $\mathbb{S}^1$  from two different manifolds are identically the same; i.e., we identify  $\mathcal{T} = \kappa^{-1}$ . After this identification the only difference between two manifolds is the curvature of the hyperbolic space  $R[\mathbb{H}_\kappa^3] \sim \kappa^2$  at  $\kappa \rightarrow 0$ . Formula (A9) below suggests a linear dependence on  $\kappa$  at small  $\kappa$  which we interpret as a strong argument supporting our conjecture on linear dependence of nondispersive vacuum energy as a function of external parameter. Such linear scaling obviously implies that this background-dependent correction is not related to any local operators such as curvature, but rather is generated by nonlocal operator (A6) which is sensitive to the global characteristics of the background.

The relevant formula can be represented as follows [7]:

$$\frac{E_{\text{vac}}[\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1]}{E_{\text{vac}}[\mathbb{R}^3 \times \mathbb{S}^1]} \simeq \left( 1 - \frac{\nu(1-\nu)}{2} \cdot \frac{\kappa}{\Lambda_{\text{QCD}}} \right). \quad (\text{A9})$$

Using formula (A8) the same result can be written as follows:

$$\begin{aligned} E_{\text{vac}}[\mathbb{H}_\kappa^3 \times \mathbb{S}_{\kappa^{-1}}^1] &\simeq -\frac{32\pi^2}{g^4} \Lambda_{\text{QCD}}^4 \left( 1 - \frac{\nu(1-\nu)}{2} \cdot \frac{\kappa}{\Lambda_{\text{QCD}}} \right) \\ &\simeq -\frac{32\pi^2}{g^4} \Lambda_{\text{QCD}}^4 + \frac{32\pi^2}{g^4} \Lambda_{\text{QCD}}^3 \cdot \frac{\nu(1-\nu)}{2} \cdot \kappa. \end{aligned} \quad (\text{A10})$$

The key observation is that a small correction here is linear, rather than naively expected quadratic function at small  $\kappa \rightarrow 0$ . Furthermore, the correction  $\sim \kappa$  vanishes for configurations with trivial holonomy,  $\nu = 0, \nu = 1$ .

This observation unambiguously implies that the relevant Euclidean configurations which are capable of producing the linear correction (A10) must carry a nontrivial holonomy (A6); and therefore, they are nonlocal in nature. The computations [65] in weakly coupled deformed QCD model (where a configuration with nontrivial holonomy produces a linear correction) also support this claim.

### 3. Generation of the Holonomy in a strongly coupled gauge theory

The question we address in this Appendix can be formulated as follows. If we consider the thermodynamical limit in Eq. (A8) one can explicitly see that the combination  $\mathcal{T}F_{\text{vac}} \equiv E_{\text{vac}}V^{(4)} \equiv E_{\text{vac}}\mathcal{T}V$  shows the extensive property when  $\ln \mathcal{Z}$  is proportional to the Euclidean 4-volume at large  $V^{(4)} \rightarrow \infty$ . In this framework  $E_{\text{vac}}$  has dimension 4 and represents the vacuum energy density of the system

entering the fundamental formula (A1). This formula defines the nondispersive  $\theta$ -dependent portion of the vacuum energy which plays the crucial role in our analysis.

The key question we address now is as follows: if we start from description of the system on  $\mathbb{R}^4$  from the very beginning such that the semiclassical solutions (calorons with nontrivial holonomy) cannot be constructed on  $\mathbb{R}^4$ . How do we know anything about the holonomy defined on  $\mathbb{S}^1$  (and its direct consequence in form of the objects with fractional topological charges) if it was not a part of our construction to begin with? We should emphasize here that the configurations with fractional topological charges are a very strong signal that there is a nontrivial holonomy in the system as the only semiclassical solutions which can be defined on  $\mathbb{R}^4$  are integer value instantons.

We obviously do not know the answer on the hard question formulated above in strongly coupled four-dimensional QCD. However, there is a well-known example of the two-dimensional  $CP^{N-1}$  model which hints that such a kind of holonomy (and its manifestation in the form of the configurations with fractional topological charges) might be generated dynamically by strong quantum fluctuations such that the “effective calorons” with nontrivial holonomy do appear in the system, but they are strongly coupled quantum objects, rather than the semiclassical configurations defined on  $\mathbb{S}^1$ .

Historically, the configurations with fractional topological charges emerged in the two-dimensional  $CP^{N-1}$  model. These fractional objects have been coined as instanton quarks, also known as “fractional instantons” or “instanton partons.” Namely, using an exact accounting and resummation of the  $n$ -instanton solutions in two-dimensional  $CP^{N-1}$  models, the original statistical problem of a grand canonical instanton ensemble (with exclusively integer topological charges defined on  $\mathbb{R}^2$ ) was mapped onto a two-dimensional Coulomb gas system of pseudoparticles with fractional topological charges  $\sim 1/N$  [73,74]. This picture leads to the elegant explanation of the confinement phase and other important properties of the two-dimensional  $CP^{N-1}$  models [73,74]. The term instanton quarks was introduced to emphasize that there are precisely  $N$  constituents making an integer instanton, similar to  $N$  quarks making a baryon. These objects do not appear individually in path integral; instead, they appear as configurations consisting of  $N$  different objects with fractional charge  $1/N$  such that the total topological charge of each configuration is always integer. In this case  $2Nk$  zero modes for the  $k$  instanton solution are interpreted as  $2$  translation zero modes accompanied by every single instanton quark. While the instanton quarks emerge in the path integral coherently, these objects are highly delocalized: they may emerge on opposite sides of the spacetime or be close to each other with alike probabilities. A similar attempt in four-dimensional QCD was unfortunately unsuccessful due to a number of technical problems, which remain to be solved [75].

There is deep analogy with the deformed QCD model [62–64] where the size of  $\mathbb{S}^1$  is fixed for the semiclassical approximation to be justified. However, it is a common view in the QCD community that the physics in strongly coupled QCD is qualitatively the same as in the weakly coupled deformed QCD model with enforced semiclassicity by specifically chosen  $\mathbb{S}^1$  in which case the configurations with nontrivial holonomy (and fractionally charged monopoles) can be explicitly constructed on the semiclassical level. Furthermore, it is expected that even in a case when the corresponding  $\bar{\theta}$  parameter in strongly coupled  $\overline{\text{QCD}}$  does not vanish, the physics remains the same and the confinement in  $\overline{\text{QCD}}$  occurs as a result of condensation of the same fractionally charged monopoles as argued in [76].

The main lesson to be learned in the context of the present work is as follows. The configurations with fractional topological charges can serve as a trigger for a nontrivial holonomy because conventional semiclassical solutions defined on  $\mathbb{R}^4$  can carry only integer topological charges. The lesson from two-dimensional  $CP^{N-1}$  is as follows. The fractional topological charges are not present in the system when it is defined on  $\mathbb{R}^2$ . However, such objects do appear dynamically as a result of strong quantum fluctuations. In terms of effective semiclassical configurations these objects obviously require a nontrivial holonomy (and therefore, nontrivial  $\mathbb{S}^1$  where the holonomy is defined). However, this effective  $\mathbb{S}^1$  is not the original circle, but rather the effective one which emerges as a result of strong quantum fluctuations. This is precisely the motivation for our model 2 in Sec. V where we unlink the size of  $\mathbb{S}^1$  from matter context of the theory by relaxing the bootstrap equation. Unfortunately, we can only speculate on this matter at the present time without making any precise and solid claims.

## APPENDIX B: TOPOLOGICAL AUXILIARY FIELD AS A NONPROPAGATING AND NONDYNAMICAL INFLATON

The goal of this Appendix is to introduce the auxiliary field technique and demonstrate that the corresponding alternative computations reproduce the crucial elements of the vacuum energy and its unusual features listed in Sec. III C. Furthermore, this technique plays a crucial role in our studies on anomalous coupling with the SM fields described in Sec. VI B. Precisely this coupling is responsible for the successful *reheating* phase as advocated in Secs. VI C and VI D.

As we argue below we can identify (on an intuitive level) the corresponding auxiliary nondynamical, nonpropagating field with the *inflaton*, which is an emergent field in our framework: it only appears in the confined  $\overline{\text{QCD}}$  phase, while in the deconfined phase it did not exist in the system. It should be contrasted with conventional description in

terms of local dynamical field  $\Phi(x)$  which is always a part of the system, long before and long after the inflation.

We should emphasize that the reformulation of the same physics in terms of an auxiliary quantum field rather than in terms of explicit computation of the partition function by summing over all topological sectors is not a mandatory procedure, but a matter of convenience. Similarly, the description of a topologically ordered phase in condensed matter physics in terms of Chern-Simons effective Lagrangian is a matter of convenience rather than a necessity as emphasized in Sec. [VI B](#).

We demonstrate how this technique works in a simplified version of QCD, the so-called weakly coupled deformed QCD model [\[62\]](#) which preserves all relevant features of the strongly coupled QCD such as confinement, nontrivial  $\theta$  dependence, generation of the nondispersive vacuum energy, etc. At the same time, all computations can be performed under complete theoretical control. The computations of the nondispersive term by explicit summation over positions and orientations of the monopoles-instantons describing the tunneling transitions have been performed in [\[64\]](#). The corresponding results have been reproduced in [\[63\]](#) using the technique of the auxiliary topological fields.

One should also mention that the computations are performed in effectively three-dimensional weakly coupled gauge theory, rather than in strongly coupled four-dimensional QCD. Nevertheless, the emergent auxiliary field to be introduced below and identified with *inflaton* behaves, in all respects, as the four-dimensional Veneziano ghost [\[59,60\]](#) which was postulated long ago precisely with the purpose to describe these unusual features of the vacuum energy as reviewed in [Appendix A 1](#).

The basic idea to describe the relevant IR physics in terms of an auxiliary field is to insert the corresponding  $\delta$ -function into the path integral with a Lagrange multiplier and integrate out the fast d.o.f. while keeping the slow d.o.f. which are precisely the auxiliary fields. Here and in what follows we use notations from [\[63\]](#) where this technique was originally implemented to demonstrate that the famous Veneziano ghost is nothing but an auxiliary topological field. The  $\delta$ -function to be inserted into the path integral is defined as follows,

$$\begin{aligned} & \delta\left(q(\mathbf{x}) + \frac{1}{4\pi NL} [\vec{\nabla}^2 a(\mathbf{x})]\right) \\ & \sim \int \mathcal{D}[b] e^{i \int d^3x b(\mathbf{x}) \cdot (q(\mathbf{x}) + \frac{1}{4\pi NL} [\vec{\nabla}^2 a(\mathbf{x})])}, \end{aligned} \quad (\text{B1})$$

where  $q(\mathbf{x}) \sim \text{tr}[F_{\mu\nu} \tilde{F}^{\mu\nu}]$  in this formula is treated as the original expression for the topological density operator including the fast non-Abelian gluon d.o.f., while  $b(\mathbf{x})$ ,  $a(\mathbf{x})$  are treated as slow-varying external sources describing the large distance physics for a given monopole configuration. One can proceed now with conventional semiclassical computations by summation over all monopoles, their

positions and orientations to arrive to the following dual form for the effective action. The new additional topological term  $\sim b(\mathbf{x}) \vec{\nabla}^2 a(\mathbf{x})$  can be immediately recovered from [\(B1\)](#), while interaction of the  $b(\mathbf{x})$  field (playing the role of the Lagrange multiplier) coupled to topological density operator  $q(\mathbf{x})$  can be easily recovered as it has precisely the structure of the  $\theta$  term. This observation unambiguously implies that  $b(\mathbf{x})$  field enters the effective description in unique combination with  $\theta$  as follows,  $[\theta - b(\mathbf{x})]$ , as long as  $b(\mathbf{x})$  field can be treated as a slow d.o.f. In all respects it is similar to construction of the effective Lagrangian for the  $\eta'$  field which enters the action in unique combination with  $\theta$  as follows,  $[\theta - \eta']$ . The difference is, of course, that  $\eta'$  meson has a kinetic term as well, in contrast with  $b(\mathbf{x})$  field. Therefore, the final expression for the dual effective action which includes new auxiliary  $b(\mathbf{x})$ ,  $a(\mathbf{x})$  fields assumes the form [\[63\]](#)

$$\begin{aligned} \mathcal{Z}[\sigma, b, a] & \sim \int \mathcal{D}[b] \mathcal{D}[\sigma] \mathcal{D}[a] e^{-S_{\text{top}}[b, a] - S_{\text{dual}}[\sigma, b]}, \\ S_{\text{top}}[b, a] & = \frac{-i}{4\pi N} \int_{\mathbb{R}^3} d^3x b(\mathbf{x}) \vec{\nabla}^2 a(\mathbf{x}), \\ S_{\text{dual}}[\sigma, b] & = \int_{\mathbb{R}^3} d^3x \frac{1}{2L} \left(\frac{g}{2\pi}\right)^2 (\nabla\sigma)^2 \\ & \quad - \zeta \int_{\mathbb{R}^3} d^3x \sum_{a=1}^N \cos\left(\alpha_a \cdot \sigma + \frac{\theta - b(\mathbf{x})}{N}\right). \end{aligned} \quad (\text{B2})$$

In this formula parameter  $\zeta$  plays the role of the monopole's density in the system, such that the vacuum energy is explicitly proportional to  $\zeta$ ; see [\(B3\)](#) below. The dynamical  $\sigma$  fields effectively describe the monopole's ensemble. The most important element for our studies is the Lagrange multiplier  $b(\mathbf{x})$  field and topological  $a(\mathbf{x})$  field which are interpreted as the inflaton in what follows. Both fields are not dynamical, and not propagating d.o.f., by construction. We obviously do not introduce any new dynamical d.o.f. by inserting the  $\delta$  function [\(B1\)](#) and introducing the auxiliary topological fields  $b(\mathbf{x})$ ,  $a(\mathbf{x})$ . This is obviously an important remark when one tries to identify  $b(\mathbf{x})$ ,  $a(\mathbf{x})$  with inflaton.

One next step is to compute the vacuum energy and topological susceptibility within this framework to demonstrate that it satisfies all the features listed in [Sec. III C](#). The corresponding computations explicitly show that the physical meaning of the vacuum energy is the number of the tunneling events per unit volume per unit time. The corresponding formula can be represented in terms of the correlation function as follows,

$$\begin{aligned} E_{\text{vac}} & = -N^2 \lim_{k \rightarrow 0} \int d^4x e^{ikx} \langle q(\mathbf{x}), q(\mathbf{0}) \rangle \\ & = -\frac{N\zeta}{L} \int d^3x \delta^3(\mathbf{x}) = -\frac{N\zeta}{L}, \end{aligned} \quad (\text{B3})$$

where we represented  $q(\mathbf{x})$  in terms of the auxiliary field  $-\frac{1}{4\pi NL} [\nabla^2 a(\mathbf{x})]$  and performed the Gaussian path integral over  $\mathcal{D}[b]\mathcal{D}[\sigma]\mathcal{D}[a]$  fields entering (B2).

We obviously reproduce our previous result based on explicit computations of the monopoles [64]. Now it is formulated in terms of the long ranged auxiliary topological fields. The fluctuating  $b(\mathbf{x})$ ,  $a(\mathbf{x})$  fields simply reflect the long distance dynamics of the degenerate topological sectors which exist independently from our description in terms of  $b(\mathbf{x})$ ,  $a(\mathbf{x})$  fields. However, in previous computations [64] we had to sum over all monopoles, their positions, interactions and orientations. Now this problem is simplified as it is reduced to the computation of the correlation function constructed from the auxiliary fields governed by the action (B2).

We identify (intuitively) the corresponding auxiliary  $[a(\mathbf{x}), b(\mathbf{x})]$  fields which saturate this energy (B3) with *inflaton* in this model in a sense that both objects eventually lead to the de Sitter behavior. We emphasize again that the corresponding dynamics cannot be formulated in terms of a canonical scalar field  $\Phi$  with any local potential  $V(\Phi)$  as it is known that the dynamics governed by CS-like action is truly nonlocal. There are a large number of CM systems (realized in nature) where CS action plays a key role with explicit manifestation of the nonlocality in the system. It has been also argued that the deformed QCD model which is explored in this section also belongs to a topologically ordered phase with many features which normally accompany the topological phases [63].

What is the physical meaning of the auxiliary  $[a(\mathbf{x}), b(\mathbf{x})]$  fields which we identify with *inflaton*? What is the best way to visualize it on an intuitive level? From our construction one can easily see that both fields  $[a(\mathbf{x}), b(\mathbf{x})]$  do not carry a color index. However,  $a(\mathbf{x})$  field has nontrivial transformation properties under large gauge transformation. In fact our field  $\nabla_i a(\mathbf{x})$  transforms as the  $K_i(\mathbf{x})$  in the Veneziano construction (A4). One can support this identification by computing a gauge variant correlation function,

$$\lim_{k \rightarrow 0} \int d^4 x e^{ikx} \langle \nabla_i a(\mathbf{x}), \nabla_j a(\mathbf{0}) \rangle \sim \frac{k_i k_j}{k^4}. \quad (\text{B4})$$

The massless pole (B4) has precisely the same nature as the pole in the Veneziano construction (A4).

What is the physical meaning of  $b(\mathbf{x})$  field? This field can be thought of as an external axion  $\theta(x)$  field, without kinetic term, though.

Our comment here is that in spite of the gap  $\sim \zeta$  in the system, some correlation functions constructed from the topological auxiliary fields  $a(\mathbf{x})$ ,  $b(\mathbf{x})$  fields are still highly sensitive to the IR physics. Furthermore, while the behavior (B4) at small  $k$  can be considered to be very dangerous as it includes  $k^4$  in denominator (which is normally attributed to the negative norm states in QFT), the physics described here is perfectly unitary and causal as  $a(\mathbf{x})$ ,  $b(\mathbf{x})$  are in fact

auxiliary rather than propagating dynamical fields as all questions can be formulated and answered even without mentioning the auxiliary topological fields.

One should comment here that the results presented above are based on computations in a weakly coupled, effectively three-dimensional, gauge theory (where the system is under complete theoretical control), while we are interested in four-dimensional strongly coupled  $\overline{\text{QCD}}$  to study the inflationary phase. Nevertheless, the relation between the  $a(x)$  auxiliary field and four-dimensional  $K_\mu$  field still holds and assumes the form

$$K_\mu \sim \partial_\mu a(x), \quad q(x) \sim \partial_\mu K^\mu \sim \square a(x) \quad (\text{B5})$$

while  $b(x)$  field always enters the effective Lagrangian precisely in combination with the  $\theta$  term according to (B2). This observation allows us to exactly reconstruct the interaction with SM particles from the knowledge on their coupling to the  $\theta$  parameter as Eq. (49) states.

What are the typical fluctuation scales of the auxiliary quantum  $a(x)$  and  $b(x)$  fields? The answer is quite obvious: the typical fluctuations are of order  $\Lambda_{\overline{\text{QCD}}}$  as the UV fluctuations of order  $M_P$  are present in the original gluon fields, but not in the auxiliary  $a(x)$  and  $b(x)$  fields which effectively describe the long distance physics in Eq. (B2) where fast d.o.f. are integrated out.

What happens when the same system is defined on a nontrivial manifold characterized by some dimensional parameters such as  $T^{-1} \ll \Lambda_{\overline{\text{QCD}}}$ ? In this case the fields  $a(x, H)$  and  $b(x, H)$  continue to fluctuate with typical frequencies  $\Lambda_{\overline{\text{QCD}}}$ . However, the relevant correlation functions should demonstrate the emergence of the linear corrections with respect to these small parameters  $\sim T^{-1}$ . In particular, the correlation functions such as (B3) computed in terms of the auxiliary fields are of the order of  $\Lambda_{\overline{\text{QCD}}}^4$  with corrections of order  $(\Lambda_{\overline{\text{QCD}}} T)^{-1}$  in agreement with expression (20) in Sec. III C.

We also want to make a few comments on a typical scale of the expectation value of the field itself  $\langle b(x, H) \rangle$  because, e.g.,  $\langle \dot{b}(x, H) \rangle$  enters the estimate (51) for the  $e$ -folds. As we discussed in the previous paragraph, the typical expectation values for the auxiliary fields must be expressed in terms of  $\Lambda_{\overline{\text{QCD}}}$  according to their dimensionality because it reflects the typical topological density distribution in strongly coupled  $\overline{\text{QCD}}$ . According to our general prescription formulated in Secs. III B and III C we must subtract all the expectation values computed on  $\mathbb{R}^4$  corresponding to  $H = 0$ . This procedure unambiguously implies<sup>17</sup> that  $\langle \dot{b}(x, H) \rangle \sim H$  as it must vanish at  $H = 0$ .

<sup>17</sup>For this specific case the particle production obviously does not occur when  $H = 0$  as all topological transitions simply select a specific  $\theta$  sector, but do not generate the particle production. Therefore, the subtraction in this case is trivial.

One should comment here that  $b(x, H)$  is not a classical field which satisfies some equations, similar to conventional studies on inflation when the potential  $V[\Phi]$  completely determines the dynamics of the system. In our case one has to compute the relevant correlation functions and the expectation values to answer the questions about observables. We make a few additional comments in relation to the conventional approach in Sec. VI F.

We want to present one additional argument supporting the same claim on linear correction  $\sim \mathcal{T}^{-1}$  or  $\sim H$ . The behavior (B4) hints to the possibility of nonlocal effects (which indeed are known to be present in this system [63]). Such IR sensitivity suggests that the physics must be highly

sensitive to the properties of the manifolds and external background configurations. Precisely this sensitivity to large distances supports our analysis of Sec. III C where we argued that the corrections to the vacuum energy due to the finite manifold should be linear  $\mathcal{T}^{-1}$  rather than exponential as a conventional gapped theory would naively suggest.

Now we can infer the physical meaning of the auxiliary fields:  $a(x, H)$  describes the longitudinal portion of  $K_\mu$  field generating the topologically protected pole (B4); the  $\square a(x, H)$  describes the distribution of the topological density in the system; finally,  $b(x, H)$  acts as the axion field (without kinetic term) being the source of the topological density distribution.

- 
- [1] A. A. Starobinsky, *Phys. Lett.* **91B**, 99 (1980).  
 [2] A. Guth, *Phys. Rev. D* **23**, 347 (1981); A. Linde, *Phys. Lett.* **108B**, 389 (1982).  
 [3] V. Mukhanov, *Physical Foundation of Cosmology* (Cambridge University Press, Cambridge, 2005).  
 [4] A. O. Barvinsky and A. Yu. Kamenshchik, *Phys. Lett. B* **774**, 59 (2017).  
 [5] A. R. Zhitnitsky, *Phys. Rev. D* **89**, 063529 (2014).  
 [6] A. R. Zhitnitsky, *Phys. Rev. D* **90**, 043504 (2014).  
 [7] A. R. Zhitnitsky, *Phys. Rev. D* **92**, 043512 (2015).  
 [8] A. O. Barvinsky and A. Y. Kamenshchik, *J. Cosmol. Astropart. Phys.* **09** (2006) 014.  
 [9] A. O. Barvinsky and A. Y. Kamenshchik, *Phys. Rev. D* **74**, 121502 (2006).  
 [10] A. O. Barvinsky, *Phys. Rev. Lett.* **99**, 071301 (2007).  
 [11] J. B. Hartle and S. W. Hawking, *Phys. Rev. D* **28**, 2960 (1983); S. W. Hawking, *Nucl. Phys.* **B239**, 257 (1984).  
 [12] J. L. Cervantes-Cota and H. Dehnen, *Nucl. Phys.* **B442**, 391 (1995).  
 [13] F. Bezrukov and M. Shaposhnikov, *Phys. Lett. B* **659**, 703 (2008).  
 [14] A. O. Barvinsky, A. Yu. Kamenshchik, and A. A. Starobinsky, *J. Cosmol. Astropart. Phys.* **11** (2008) 021.  
 [15] M. J. Duff, *Nucl. Phys.* **B125**, 334 (1977).  
 [16] J. Schwinger, *J. Math. Phys.* **2**, 407 (1961); L. V. Keldysh, *Sov. Phys. JETP* **20**, 1018 (1965).  
 [17] A. O. Barvinsky, A. Yu. Kamenshchik, and D. V. Nesterov, *Eur. Phys. J. C* **75**, 584 (2015).  
 [18] A. O. Barvinsky, A. Yu. Kamenshchik, and D. V. Nesterov, *J. Cosmol. Astropart. Phys.* **01** (2016) 036.  
 [19] A. O. Barvinsky, *Phys. Rev. D* **93**, 103530 (2016).  
 [20] G. Veneziano, *J. High Energy Phys.* **06** (2002) 051.  
 [21] G. R. Dvali, G. Gabadadze, M. Kolanovic, and F. Nitti, *Phys. Rev. D* **65**, 024031 (2001).  
 [22] Y. B. Zeldovich, *Pis'ma Zh. Eksp. Teor. Fiz.* **6**, 883 (1967); [*JETP Lett.* **6**, 316 (1967)].  
 [23] J. Bjorken, [arXiv:hep-th/0111196](https://arxiv.org/abs/hep-th/0111196).  
 [24] R. Schutzhold, *Phys. Rev. Lett.* **89**, 081302 (2002).  
 [25] F. R. Klinkhamer and G. E. Volovik, *Phys. Rev. D* **77**, 085015 (2008).  
 [26] E. C. Thomas, F. R. Urban, and A. R. Zhitnitsky, *J. High Energy Phys.* **08** (2009) 043.  
 [27] M. Maggiore, *Phys. Rev. D* **83**, 063514 (2011).  
 [28] J. Sola, *J. Phys. Conf. Ser.* **453**, 012015 (2013).  
 [29] I. L. Shapiro and J. Sola, *Phys. Lett. B* **475**, 236 (2000).  
 [30] I. L. Shapiro and J. Sola, *J. High Energy Phys.* **02** (2002) 006.  
 [31] A. O. Barvinsky, *J. Cosmol. Astropart. Phys.* **09** (2012) 033.  
 [32] J. Frohlich and B. Pedrini, [arXiv:cond-mat/0201236](https://arxiv.org/abs/cond-mat/0201236).  
 [33] M. Joyce and M. E. Shaposhnikov, *Phys. Rev. Lett.* **79**, 1193 (1997).  
 [34] Y. Akamatsu and N. Yamamoto, *Phys. Rev. Lett.* **111**, 052002 (2013).  
 [35] A. D. Linde, *Lect. Notes Phys.* **738**, 1 (2008).  
 [36] A. R. Zhitnitsky, *Phys. Rev. D* **86**, 045026 (2012).  
 [37] F. R. Urban and A. R. Zhitnitsky, *Phys. Lett. B* **688**, 9 (2010).  
 [38] F. R. Urban and A. R. Zhitnitsky, *Nucl. Phys.* **B835**, 135 (2010).  
 [39] F. R. Urban and A. R. Zhitnitsky, *Phys. Rev. D* **80**, 063001 (2009).  
 [40] R.-G. Cai, Y. Gong, and B. Wang, *Universe* **2**, 21 (2014).  
 [41] R.-G. Cai, Z.-L. Tuo, Y.-B. Wu, and Y.-Y. Zhao, *Phys. Rev. D* **86**, 023511 (2012).  
 [42] A. Yamamoto, *Phys. Rev. D* **90**, 054510 (2014).  
 [43] C. Cao, M. van Caspel, and A. R. Zhitnitsky, *Phys. Rev. D* **87**, 105012 (2013).  
 [44] A. R. Zhitnitsky, *Phys. Rev. D* **91**, 105027 (2015).  
 [45] C. Cao, Y. Yao, and A. R. Zhitnitsky, *Phys. Rev. D* **93**, 065049 (2016).  
 [46] Y. Yao and A. R. Zhitnitsky, *Phys. Rev. D* **95**, 065018 (2017).  
 [47] C. Cao and A. Zhitnitsky, *Phys. Rev. D* **96**, 015013 (2017).  
 [48] A. Zhitnitsky, *Nucl. Phys.* **B916**, 510 (2017).  
 [49] A. Vilenkin, *Phys. Rev. D* **30**, 509 (1984); A. D. Linde, *J. Exp. Theor. Phys.* **60**, 211 (1984); V. A. Rubakov, *J. Exp.*

- Theor. Phys. Lett. **39**, 107 (1984); Ya. B. Zeldovich and A. A. Starobinsky, Sov. Astron. Lett. **10**, 135 (1984).
- [50] T. Vachaspati and A. Vilenkin, *Phys. Rev. D* **37**, 898 (1988).
- [51] A. Vilenkin, *Phys. Rev. D* **58**, 067301 (1998).
- [52] D. Page, *Phys. Rev. D* **56**, 2065 (1997).
- [53] S. W. Hawking, [arXiv:0710.2029](https://arxiv.org/abs/0710.2029).
- [54] J. B. Hartle and T. Hertog, *Phys. Rev. D* **80**, 063531 (2009).
- [55] J. Feldbrugge, J. L. Lehners, and N. Turok, *Phys. Rev. D* **95**, 103508 (2017).
- [56] J. Feldbrugge, J. L. Lehners, and N. Turok, *Phys. Rev. Lett.* **119**, 171301 (2017).
- [57] J. Feldbrugge, J. L. Lehners, and N. Turok, *Phys. Rev. D* **97**, 023509 (2018).
- [58] E. Witten, *Nucl. Phys.* **B156**, 269 (1979).
- [59] G. Veneziano, *Nucl. Phys.* **B159**, 213 (1979).
- [60] P. Di Vecchia and G. Veneziano, *Nucl. Phys.* **B171**, 253 (1980).
- [61] M. Luscher, *Phys. Lett. B* **78B**, 465 (1978).
- [62] M. Ünsal and L. G. Yaffe, *Phys. Rev. D* **78**, 065035 (2008).
- [63] A. R. Zhitnitsky, *Ann. Phys. (Amsterdam)* **336**, 462 (2013).
- [64] E. Thomas and A. R. Zhitnitsky, *Phys. Rev. D* **85**, 044039 (2012).
- [65] E. Thomas and A. R. Zhitnitsky, *Phys. Rev. D* **86**, 065029 (2012).
- [66] B. J. Harrington and H. K. Shepard, *Phys. Rev. D* **17**, 2122 (1978).
- [67] D. J. Gross, R. D. Pisarski, and L. G. Yaffe, *Rev. Mod. Phys.* **53**, 43 (1981).
- [68] T. C. Kraan and P. van Baal, *Nucl. Phys.* **B533**, 627 (1998).
- [69] K. M. Lee and C. H. Lu, *Phys. Rev. D* **58**, 025011 (1998).
- [70] D. Diakonov, N. Gromov, V. Petrov, and S. Slizovskiy, *Phys. Rev. D* **70**, 036003 (2004).
- [71] D. Diakonov and V. Petrov, *Phys. Rev. D* **76**, 056001 (2007).
- [72] Y. Liu, E. Shuryak, and I. Zahed, *Phys. Rev. D* **92**, 085006 (2015).
- [73] V. A. Fateev, I. V. Frolov, and A. S. Schwarz, *Nucl. Phys.* **B154**, 1 (1979).
- [74] B. Berg and M. Luscher, *Commun. Math. Phys.* **69**, 57 (1979).
- [75] A. A. Belavin, V. A. Fateev, A. S. Schwarz, and Y. S. Tyupkin, *Phys. Lett.* **83B**, 317 (1979).
- [76] M. M. Anber and A. R. Zhitnitsky, *Phys. Rev. D* **96**, 074022 (2017).