de Sitter space decay and cosmological constant relaxation in unimodular gravity with charged membranes

Nemanja Kaloper[*]

QMAP, Department of Physics and Astronomy, University of California Davis, California 95616, USA

General covariant unimodular gravity frameworks, based on the Henneaux-Teitelboim formulation, are, in disguise, precisely 4-form field theories corrected with higher dimension operators. In the presence of charged tensional membranes, any de Sitter space in all such theories is unstable and decays. If the fluxes sourced by membranes are mutually incommensurate, de Sitter geometries comprise a very refined discretuum of states. Whenever the 4-form sector is dominated by terms linear in flux the almost-Minkowski space is the unique long-time attractor. As a result, a tiny cosmological constant is natural in all such frameworks, without appealing to anthropic reasoning.

I. SYNOPSIS

In this paper we demonstrate that general covariant unimodular gravity frameworks, which naturally generalize the Henneaux-Teitelboim formulation [1], are nothing other than 4-form field theories corrected with higher (and lower) dimension operators, after they are recast in canonically dual variables, analogous to the exchange of coordinates and momenta in classical mechanics. Specifically, the general 4-form theories take exactly the form of generalized unimodular gravity when rewritten in terms of magnetic 4-form variables, as opposed to the more commonly encountered electric variables.

When the theory is completed with the inclusion of charged tensional membranes, entailing local gauge symmetry, any de Sitter space in all such theories is unstable and decays: de Sitter geometries comprise a refined discretuum of unstable states, which decay by nucleation of bubbles bounded by the membranes [2,3]. In the presence of multiple 4-forms and their associated membrane towers, with at least some of the fluxes sourced by charges being incommensurate, the de Sitter discretuum is extremely finely grained, with cosmological constant values arbitrarily close to zero. Whenever the energy of the 4-form sector is dominated by leading linear terms in 4-form fluxes, an almost-Minkowski space is the unique long-time attractor, in the statistical sense: almost-flat regions greatly outnumber more curved ones [4–6].

As a result, a tiny cosmological constant becomes natural in all such frameworks, without appealing to anthropic reasoning. We discuss how this can be employed to address the cosmological constant problem in our Universe, and various aspects of cosmology.

In the interest of clarity, let us stress again that our main new results here are

(1) establishing the 1-to-1 correspondence between generic 4-form theories and a dynamical generalization of unimodular gravity, which includes charged membranes which source the fluxes that screen the cosmological constant; in this vein, we find that the standard covariantly formulated unimodular gravity [1] is interpreted as a Routhian transform of the 4-form theories; this changes the perspective on the theory and should affect how the theory is UV-completed and quantized;

(2) demonstrating that for generic 4-form fluxes, the presence of at least two systems of membranes that source incommensurate contributions to the cosmological constant such that the terms linear in fluxes dominate over higher powers, the attractor mechanism discovered and elaborated in [4–6] remains fully operational;

(3) confirming that quantum dynamics statistically biases the distribution of vacua, dynamically exponentially favoring $\Lambda \to 0^+$; other values of Λ are possible, but dynamically suppressed,[1] since the linear flux terms, when dominant, allow only processes which have rates $\approx \exp(-24\pi^2 M_{\rm Pl}^4/\Lambda)$; this is

---

[*]kaloper@physics.ucdavis.edu

---

[1]Although [4–6] focus only on linear flux terms, it was noted that more general examples will behave in the same manner, which we show in detail here.

a loophole around the more traditional approaches to flux generated landscapes that have nearly uniform distribution of vacua.

II. WHAT IS GENERAL RELATIVITY?

In 1915, Einstein [7] and Hilbert [8] laid foundations of general relativity (GR). The former wrote the field equations governing geometry sourced by stress energy of dynamical matter. The latter formulated the simplest action principle that the field equations can be derived from. Enforcing locality and causality by imposing on the theory to involve no more than two derivatives, and encoding local gauge invariance via Bianchi's identities, dimensional analysis requires that gravity's coupling to matter is a dimensional constant, identified with Newton's constant $G_N = 1/8\pi M_{\rm Pl}^2$, where $M_{\rm Pl}$ is Planck mass. Another dimensional constant Λ arises representing the energy of the vacuum of the theory, which not only does not need to be zero, but in quantum theory most likely cannot be, thanks to the equivalence principle and quantum uncertainty, as originally noted by Bronstein [9] and Pauli [10], and formulated in more modern terms by Zeldovich [11]. Since the engineering dimension of Λ is four, it is often viewed as a quartic power of a UV cutoff scale (or some symmetry breaking scale, such as shorthand for supersymmetry (SUSY) and/or conformal symmetry). Yet, since it sources gravity via Einstein's equations, its effect on the geometry is via inducing a vacuum curvature $\mathcal{H}^2 = \Lambda/3M_{\rm Pl}^2$. Thus inverting this, Λ can be viewed as a square of an IR cutoff set by curvature, since at longer distances the bending of space cannot be ignored.

Since both of these terms are constants it is tempting—and common—to set them to their observed values from the get-go, and treat them as given dimensional parameters of the theory, in both the field equations and the action. Yet there is absolutely no *à priori* argument to select their numerical values from first principles. On the contrary, the hugely discrepant scales controlling these numbers, and additionally their disparity with any of the known scales in the matter sector, are commonly recognized as the gauge hierarchy and cosmological constant problems, respectively (see, e.g., [12,13]). However, a careful examination of quantum field theory (QFT) coupled to semiclassical gravity, specifically renormalization and UV sensitivity of theory's operator expansion, clearly shows that in this limit there is no reason to expect these numbers to be correlated [14]. Both are UV sensitive, and, after renormalization, they both depend on independent finite parts of the bare counterterms added to cancel divergences. This, along with the naïve idea of naturalness in effective field theory, suggests that the cosmological constant problem, which can be loosely phrased as

$$\left(\frac{\Lambda}{M_{\rm Pl}^4}\right)\Bigg|_{\rm observed} \ll 1 \quad \text{versus}$$

$$\left(\frac{\Lambda}{M_{\rm Pl}^4}\right)\Bigg|_{\rm calculated} \gg \left(\frac{\Lambda}{M_{\rm Pl}^4}\right)\Bigg|_{\rm observed},$$

which is a "radiative instability": quantum corrections want to drive Λ up. Yet, without a UV completion including gravity, this issue is moot, since in generic cases renormalization introduces a separate counterterm for each UV sensitive quantity with totally arbitrary finite parts.[2] Thus in a QFT, $\left(\frac{\Lambda}{M_{\rm Pl}^4}\right)|_{\rm calculated}$ can be anything. A generic theory does not fail to fix the counterterms, it simply is not supposed to. That does not mean that there is no cosmological constant problem. We will discuss this subtle issue in more detail in what follows.

From the point of view of "canonical" GR based on Einstein-Hilbert theory [7,8], this is particularly jarring. After all, the canonical action does not even seem to have the degrees of freedom that could accommodate variation of Λ, and the many attempts to append them to the theory in the effective field theory framework have run into the wall of venerable Weinberg's no-go theorem [13] (see also [14]). Yet, only a few years after the paper on formulating GR, Einstein proposed its, perhaps most minimal, generalization [15], which today is commonly dubbed unimodular gravity (see, e.g., [1,16–23]). In a nutshell, in this approach the cosmological parameter Λ is viewed as yet another free parameter of the theory, which solves a field equation, albeit trivially: $\partial_\mu \Lambda = 0$, and hence it is an arbitrary constant. This follows from gauge invariance of the theory: the local contributions to the metric determinant $\det(g_{\mu\nu})$ are pure gauge, since all local propagating metric modes cancel out of it. It can be set to any value by a diffeomorphism, under which $\det(g_{\mu\nu}) \to \det(g'_{\mu\nu}) = \det(g_{\mu\nu})[\det(\partial x/\partial x')]^2$. In particular one could choose the gauge $\det(g_{\mu\nu}) = 1$. This gauge fixes the diffeomorphism transformations to only those with $|\det(\partial x/\partial x')| = 1$ (hence the moniker "unimodular"[3]). Since it is now fixed, one drops the field equation obtained by $\det(g_{\mu\nu})$ variation. This amounts to retaining only the traceless subset of Einstein's equations. However, Bianchi identities, which follow from local gauge invariance, recover the trace, albeit under a derivative [13]. The effective theory is split into $9 + 1$ field equations,

---

[2]Symmetries can correlate them, but an example of such symmetries that fits our Universe has not been found to date.

[3]This notion of unimodularity is somewhat of a distraction, since gauge fixing can be altered at will; the one global gauge invariant degree of freedom in $\int d^4x \sqrt{g}$ is completely unaffected by it, as we discuss in what follows.

$$M_{\text{Pl}}^2 \left(R^\mu_{\ \nu} - \frac{1}{4} R \delta^\mu_{\ \nu} \right) = -\left(T^\mu_{\ \nu} - \frac{1}{4} T \delta^\mu_{\ \nu} \right),$$

$$\partial_\mu (M_{\text{Pl}}^2 R - T) = 0. \tag{1}$$

Integrating the last equation, and plugging the result back into the traceless field equations, yields the "rearranged" full set of 10 covariant equations valid in any gauge

$$M_{\text{Pl}}^2 \left(R^\mu_{\ \nu} - \frac{1}{2} R \delta^\mu_{\ \nu} \right) = -T^\mu_{\ \nu} - \Lambda \delta^\mu_{\ \nu}, \tag{2}$$

where Λ is the integration constant arising from $\partial_\mu (M_{\text{Pl}}^2 R - T) = 0$ (normalized to dim. 4). Thus the only distinction between canonical GR with fixed G_N and Λ, and unimodular GR is the complete arbitrariness of Λ in the latter interpretation, which is actually more aligned with the description of renormalized QFT dwelling on the background spacetime. The integration constant Λ is precisely the required QFT counterterm that subtracts the UV sensitive contribution from QFT [14]. While this feature motivated some claims that as a result cosmological constant problem is solved in unimodular GR, this is manifestly not true [13]. The physical cosmological constant that bends the spacetime vacuum $|0\rangle$ is[4]

$$\Lambda_{\text{phys}} = \Lambda + \frac{1}{4} \langle 0 | T^\mu_{\ \mu} | 0 \rangle, \tag{3}$$

and so Λ can indeed be chosen such that Λ_{phys} is finite, but its value, without additional ingredients (see, e.g., [14,24,25]), remains indeterminate (see also Sec. VII of [13]).

This one parameter freedom arises because $\det(g_{\mu\nu})$ does depend on a single global gauge invariant degree of freedom: the spacetime volume $\Omega_4 = \int d^4x \sqrt{|\det(g_{\mu\nu})|}$ is gauge invariant (although often ill defined due to infinities) [14,22,24,25]. Nevertheless, its Legendre dual is precisely Λ, which can be treated as a regulated, gauge invariant, global degree of freedom, of so-generalized GR.

Formulating the correct gauge invariant action principle for unimodular gravity, however, took a while. Gauge-fixing $g = \det(g_{\mu\nu}) = 1$ via a Lagrange multiplier in Einstein-Hilbert action,

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\text{Pl}}^2}{2} R - \mathcal{L}_{\text{QFT}} \right) - \Lambda (\sqrt{g} - 1) \right\}, \tag{4}$$

and treating Λ as a global variable (set to a constant to satisfy Bianchi identities) reproduces (1) at the classical

level. However, the very last term violates gauge symmetry (i.e., diffeomorphism invariance) and so, by itself, it could compromise the gauge symmetry that enforces $\Lambda = \text{const}$ in full quantum theory. However the manifestly covariant formulation given by Henneaux and Teitelboim in [1] precludes any such concerns. With our normalizations, the covariant unimodular action of Henneaux and Teitelboim is[5]

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\text{Pl}}^2}{2} R - \mathcal{L}_{\text{QFT}} - \Lambda \right) \right.$$
$$\left. + \frac{1}{3M_{\text{Pl}}^2} \epsilon^{\mu\nu\lambda\sigma} (\partial_\mu \Lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}. \tag{5}$$

Here $\mathcal{A}_{\nu\lambda\sigma}$ is an auxiliary 3-form gauge potential that serves as a Lagrange multiplier. Variation of this action with respect to the metric yields the standard GR equations with a "free function" Λ, which is constrained to a constant by the field equation $\partial_\mu \Lambda = 0$ arising from $\mathcal{A}_{\nu\lambda\sigma}$ variation. Finally, variation with respect to Λ itself yields a "spectator" equation $\epsilon^{\mu\nu\lambda\sigma} \partial_\mu \mathcal{A}_{\nu\lambda\sigma} = -3M_{\text{Pl}}^2 \sqrt{g}$. If we introduce a "spectator" 4-form $\mathcal{F}_{\mu\nu\lambda\sigma} = 4\partial_{[\mu} \mathcal{A}_{\nu\lambda\sigma]}$ (where $[\dots]$ denotes antisymmetrization), then we can invert this equation to $\mathcal{F}_{\mu\nu\lambda\sigma} = \frac{M_{\text{Pl}}^2}{2} \sqrt{g} \epsilon_{\mu\nu\lambda\sigma}$.

Note, that this 4-form flux does not gravitate directly since it is completely decoupled from the metric. Yet, it points to an interesting link between unimodular gravity and theories with 4-forms, already noted in Sec. VII of [13], but otherwise being an apparently as yet untold story. We will unveil this link in detail in the next section, after we give a straightforward generalization of the Henneaux-Teitelboim action (5). But first, let us remark that it has been proven that the action (5), which is clearly classically identical to conventional GR in every respect except that Λ is an arbitrary integration constant, preserves this identity in quantum theory, both in the semiclassical limit [21] and in the loop expansion as a QFT [22,23].

To conclude this section, we first slightly rewrite (5) by redefining $\Lambda = M_{\text{Pl}}^2 \lambda$; the action becomes

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\text{Pl}}^2}{2} R - M_{\text{Pl}}^2 \lambda - \mathcal{L}_{\text{QFT}} \right) \right.$$
$$\left. + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} (\partial_\mu \lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}. \tag{6}$$

One can readily verify that this action is identical to a subset of actions recently discussed in [4–6], in connection to the cosmological constant relaxation toward zero (neglecting for the moment the boundary terms in the actions in [4–6]).

---

[4]The vacuum expectation $\langle 0 | T^\mu_{\ \mu} | 0 \rangle$ projects $T^\mu_{\ \mu}$ to only its constant part.

[5]In [1], the authors mostly use the vector Hodge dual to the 3-form potential, $\mathcal{A}_{\mu\nu\lambda} = \sqrt{g} \epsilon_{\mu\nu\lambda\sigma} \mathcal{T}^\sigma$, but they do identify the dual relationship explicitly in their Eqs. (23)–(26).

Now, viewing λ as the independent variable, it immediately leaps to the eye that the "bulk" term $M_{\mathrm{Pl}}^2 \lambda$ could be viewed as a truncation of some more general "Lagrangian" $\mathcal{L}(\lambda)$ to the leading term in Taylor expansion. Indeed, if one is to compute the cosmological constant contributions from $\mathcal{L}_{\mathrm{QFT}}$ using some regulator μ, one will find that the loop expansion yields $\quad \Lambda_{\mathrm{QFT}} = c_0 \mu^4 + c_1 \mu^2 m^2 + c_2 m^4 + c_3 m^6/\mu^2 + \cdots$, up to logarithms of the regulator μ, where m is the mass describing local QFT degrees of freedom. In this instance, one can take $\mu = M_{\mathrm{Pl}}$ in order to include graviton loops as well as those from the QFT.

It is thus only natural to generalize (5) to

$$S = \int d^4 x \left\{ \sqrt{g} \left(\frac{M_{\mathrm{Pl}}^2}{2} R - \mathcal{L}(\lambda) - \mathcal{L}_{\mathrm{QFT}} \right) + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} (\partial_\mu \lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}, \quad (7)$$

where[6]

$$\mathcal{L}(\lambda) = M_{\mathrm{Pl}}^4 \left(c_1 \frac{\lambda}{M_{\mathrm{Pl}}^2} + 2c_2 \frac{\lambda^2}{M_{\mathrm{Pl}}^4} + 4c_3 \frac{\lambda^3}{M_{\mathrm{Pl}}^6} + \ldots \right). \quad (8)$$

At the classical level and without charges, however, this generalization is trivial, since it is merely replacing the integration constant $M_{\mathrm{Pl}}^2 \lambda$ by a more general function $\mathcal{L}(\lambda)$ in the gravitational field equations, which can be undone with a field redefinition.[7] However, when we transition to the dual variables and add charged discrete degrees of freedom that can change λ, the real purpose of this generalization will become manifest. We now turn to these issues.

III. UNIMODULAR GRAVITY IN DUAL VARIABLES

To illustrate the dualization procedure, let us consider a simple harmonic oscillator, with the Hamiltonian $H = p^2/2 + q^2/2$. Clearly, p and q are canonical variables, whose Hamilton's equations are $\dot{q} = \partial_p H = p$ and $\dot{p} = -\partial_q H = -q$. The equivalent second order equation is $\ddot{q} + q = 0$. However, consider now the change of variables $P = q, Q = -p$. In terms of those variables, the Hamiltonian is $H = P^2/2 + Q^2/2$, and the equations of motion retain the same structure as before: $\dot{Q} = \partial_P H$, $\dot{P} = -\partial_Q H$. Indeed, as is well known, the symplectic transformations of generalized coordinates are dynamical symmetries and, in our example, comprise a canonical transformation of the theory. Either set of variables is as

good a description, since no information about the dynamics is neither "lost" nor "found." The only question is, which of the variables is more convenient to address a specific application of the theory.

The same is true for other dynamical structures. Given a dynamical system with its set of generalized coordinates and momenta, we are free to rename and reshuffle variables at will provided that we preserve the canonical structure of the theory—i.e., that the transformations are canonical in the sense of Hamiltonian mechanics. We are particularly interested in applying such transformations to the theories involving 4-forms, which have a long history in the pursuit of mechanisms to address the cosmological constant problem [2,3,26–40]. We look below at a few examples and eventually at the general case.

A. Quadratic 4-form

A most common example encountered in the past works is a theory given by the bulk action

$$S = \int d^4 x \sqrt{g} \left(\frac{M_{\mathrm{Pl}}^2}{2} R - \mathcal{L}_{\mathrm{QFT}} - \frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 \right), \quad (9)$$

where $\mathcal{F}_{\mu\nu\lambda\sigma} = 4\partial_{[\mu} \mathcal{A}_{\nu\lambda\sigma]}$ for some 3-form $\mathcal{A}$, ignoring for the moment the membranes charged under $\mathcal{A}$. Rather than discussing the field equations, we cut to the chase and dualize this theory using the standard techniques explained in, e.g., [40–43]. That trick will make the connection of (9) to unimodular gravity quickly and clearly. The idea is to reformulate (9) as a first-order theory, by adding the Lagrangian constraint $\frac{1}{12} \lambda \epsilon^{\mu\nu\lambda\sigma} (\mathcal{F}_{\mu\nu\lambda\sigma} - 4\partial_\mu \mathcal{A}_{\nu\lambda\sigma})$, and then integrating out $\mathcal{F}$. This is most simply illustrated with the path integral

$$Z = \int \ldots [\mathcal{D}\mathcal{A}][\mathcal{D}\mathcal{F}][\mathcal{D}\lambda] e^{iS + 2i \int \lambda (\mathcal{F} - d\mathcal{A})} \ldots, \quad (10)$$

where the total action is (with a boundary term to be added below)

$$S_{\mathrm{total}} = S + 2 \int \lambda (\mathcal{F} - d\mathcal{A}),$$

$$= \int d^4 x \left\{ \sqrt{g} \left(\frac{M_{\mathrm{Pl}}^2}{2} R - \mathcal{L}_{\mathrm{QFT}} - \frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 \right) + \frac{1}{12} \lambda \epsilon^{\mu\nu\lambda\sigma} (\mathcal{F}_{\mu\nu\lambda\sigma} - 4\partial_\mu \mathcal{A}_{\nu\lambda\sigma}) \right\}. \quad (11)$$

Defining a new independent degree of freedom

$$\tilde{\mathcal{F}}_{\mu\nu\lambda\sigma} = \mathcal{F}_{\mu\nu\lambda\sigma} - 2\lambda \sqrt{g} \epsilon_{\mu\nu\lambda\sigma}, \quad (12)$$

and recalling that the translation of variables as in (12) do not change the path integral measure since the functional Jacobian is unity, we rewrite the action (*sans* the subscript),

---

[6]We ignore $c_0 M_{\mathrm{Pl}}^4$ in the expansion since that term can be absorbed away by a finite renormalization of Λ_{QFT}. Normalizing powers of 2 are for later convenience.

[7]To be discussed in more detail shortly.

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\rm Pl}^2}{2} R - \mathcal{L}_{\rm QFT} - 2\lambda^2 - \frac{1}{4} \tilde{\mathcal{F}}_{\mu\nu\lambda\sigma}^2 \right) \right. $$
$$\left. - \frac{1}{3} \lambda \epsilon^{\mu\nu\lambda\sigma} \partial_\mu \mathcal{A}_{\nu\lambda\sigma} \right\}. \qquad (13)$$

Since $\tilde{\mathcal{F}}$ does not appear anywhere else, the integration over it yields a factorizable Gaussian normalization factor,

$$Z = \int \dots [\mathcal{D}\tilde{\mathcal{F}}] \dots e^{\dots + i \int d^4x \sqrt{g} \left(-\frac{1}{48} \tilde{\mathcal{F}}_{\mu\nu\lambda\sigma}^2 \right)} \dots, \qquad (14)$$

which can be dropped. Then adding the boundary term for the 4-form sector, required by intrinsic consistency of the variational principle [32,41], which is just $\int \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu (\lambda \mathcal{A}_{\nu\lambda\sigma})$, our action, in terms of the new dual variable λ, becomes

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\rm Pl}^2}{2} R - \mathcal{L}_{\rm QFT} - 2\lambda^2 \right) + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu (\lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}. \qquad (15)$$

Note that the signs came up as they do due to $\epsilon_{\mu\nu\lambda\sigma} \epsilon^{\mu\nu\lambda\sigma} = -4!$ in Lorentzian signature.

This is precisely the action (7) with $\mathcal{L}(\lambda) = 2\lambda^2$. The actual normalizing factor of λ^2 does not matter when we ignore the membranes charged under $\mathcal{A}$, since it can be set to an arbitrary value by rescaling λ and $\mathcal{A}$ (with charges present, this involves a finite rescaling of charges as well). However, the important point is that the bulk 4-form theory is nothing other than a variant of unimodular gravity without charges. It has been noticed by Weinberg that 4-form flux energy behaves like a unimodular gravity cosmological counterterm (see Sec. VII of [10]), but in fact we will argue here that this connection is more than just an analogy. The precise map linking these pictures is that a theory with 4-forms and tensional membranes becomes a unimodular gravity in the decoupling limit, when the membrane tension is taken to infinity.

B. Linear + quadratic

It is obvious that these statements are true for more general examples of $\mathcal{L}(\lambda)$, or in other words for actions that include a variety of powers of $F_{\mu\nu\lambda\sigma}$. The dualization procedure is more complicated, but it amounts to performing a dual transformation $\mathcal{F} \leftrightarrow {}^*\lambda$, and replacing $\hat{\mathcal{L}}(\mathcal{F})$ by its Legendre transform $\mathcal{L}(\lambda)$ [35,40,44]. As an illustration, let us first dualize

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\rm Pl}^2}{2} R - \mathcal{L}_{\rm QFT} - \frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 \right) \right. $$
$$\left. - \frac{\alpha}{24} \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma} \right\}. \qquad (16)$$

Such a combination of 4-form terms can be found in, e.g., [26]. Here, α is a fixed 4-form theory coupling parameter that can be induced by nontrivial CP-breaking effects [26]. Following the same steps as outlined above (and adding the boundary term $\int \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu (\lambda \mathcal{A}_{\nu\lambda\sigma})$ to (16))

$$S = \int \left\{ \sqrt{g} \left(\frac{M_{\rm Pl}^2}{2} R - \mathcal{L}_{\rm QFT} - \frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 \right) - \frac{\alpha}{24} \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma} \right. $$
$$\left. + \frac{1}{12} \lambda \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma} + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu (\lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}. \qquad (17)$$

Defining a new variable $\tilde{\mathcal{F}}_{\mu\nu\lambda\sigma} = \mathcal{F}_{\mu\nu\lambda\sigma} - (2\lambda - \alpha)\sqrt{g}\epsilon_{\mu\nu\lambda\sigma}$ and integrating out $\tilde{\mathcal{F}}$, after a straightforward algebra we find

$$S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\rm Pl}^2}{2} R - \mathcal{L}_{\rm QFT} - 2 \left(\lambda - \frac{\alpha}{2} \right)^2 \right) \right. $$
$$\left. + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu (\lambda) \mathcal{A}_{\nu\lambda\sigma} \right\}. \qquad (18)$$

Note that $\mathcal{L}(\lambda) = 2(\lambda - \frac{\alpha}{2})^2$ is precisely the Legendre transform of $\hat{\mathcal{L}}(\mathcal{F}) = \frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 + \frac{\alpha}{24\sqrt{g}} \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma}$. Indeed, defining the variable[8] $\mathcal{F} = 2(\frac{1}{4!\sqrt{g}} \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma})$ and using the identity $\mathcal{F}_{\mu\nu\lambda\sigma}^2 = -3!\mathcal{F}^2$, it follows that λ and $\mathcal{F}$ are related by

$$\lambda \equiv \frac{\partial \hat{\mathcal{L}}(\mathcal{F})}{\partial \mathcal{F}} = -\frac{\mathcal{F}}{4} + \frac{1}{2}\alpha, \qquad (19)$$

and solving it to eliminate $\mathcal{F} = 2(\alpha - 2\lambda)$, we find the Legendre transform of $\mathcal{L}(\mathcal{F})$ to be

$$\mathcal{L}(\lambda) = \hat{\mathcal{L}}(\mathcal{F}) - \lambda\mathcal{F} \equiv 2 \left(\lambda - \frac{\alpha}{2} \right)^2. \qquad (20)$$

C. General unimodular case: Multiple powers of the 4-form

The same procedure works for a general $\hat{\mathcal{L}}(\mathcal{F})$. In Eq. (17), all we do is replace $\frac{1}{48} \mathcal{F}_{\mu\nu\lambda\sigma}^2 + \frac{\alpha}{24\sqrt{g}} \epsilon^{\mu\nu\lambda\sigma} \mathcal{F}_{\mu\nu\lambda\sigma}$ with $\hat{\mathcal{L}}(\mathcal{F})$, and define the variable $\tilde{\mathcal{F}}_{\mu\nu\lambda\sigma} = \mathcal{F}_{\mu\nu\lambda\sigma} - 2\lambda\sqrt{g}\epsilon_{\mu\nu\lambda\sigma}$ which is guaranteed to gather all the $\mathcal{F}$-dependent pieces (and produce the leftover $\propto \lambda$ terms), which can then be integrated out in the path integral. What remains is the dual action for the new variable λ, which will take the form of (18), with $2(\lambda - \frac{\alpha}{2})^2$ replaced by the

---

[8]This shows that the odd powers of $\mathcal{F}$ in the action may be viewed as integrals over an alternative measure $\mathcal{F} = \frac{1}{4!} \mathcal{F}_{\mu\nu\lambda\sigma} dx^\mu dx^\nu dx^\lambda dx^\sigma$, as opposed to $\sqrt{g}d^4x$, as noted in [4,5] and also in, e.g., [17,45–47].

Legendre transform of $\hat{\mathcal{L}}(\mathcal{F})$: defining the variable $\mathcal{F} = 2(\frac{1}{4!\sqrt{g}}\epsilon^{\mu\nu\lambda\sigma}\mathcal{F}_{\mu\nu\lambda\sigma})$, one can show that

$$\lambda \equiv \frac{\partial \hat{\mathcal{L}}(\mathcal{F})}{\partial \mathcal{F}}, \qquad \mathcal{L}(\lambda) = \hat{\mathcal{L}}(\mathcal{F}) - \lambda \mathcal{F}. \qquad (21)$$

Thus the magnetic dual representation of any 4-form theory is precisely the general form of unimodular gravity action introduced in the previous section. We repeat it here for clarity:

$$S = \int d^4x \left\{ \sqrt{g}\left(\frac{M_{\rm Pl}^2}{2}R - \mathcal{L}_{\rm QFT} - \mathcal{L}(\lambda)\right) \right.$$
$$\left. + \frac{1}{3}\epsilon^{\mu\nu\lambda\sigma}\partial_\mu(\lambda)\mathcal{A}_{\nu\lambda\sigma} \right\}. \qquad (22)$$

Obviously, this procedure is in general invertible, and so starting with the theory (22), we can rewrite it as a theory of a 4-form

$$S = \int \left\{ \sqrt{g}\left(\frac{M_{\rm Pl}^2}{2}R - \mathcal{L}_{\rm QFT} - \hat{\mathcal{L}}(\mathcal{F})\right) + \frac{1}{12}\lambda\epsilon^{\mu\nu\lambda\sigma}(\mathcal{F}_{\mu\nu\lambda\sigma} \right.$$
$$\left. - 4\partial_\mu\mathcal{A}_{\nu\lambda\sigma}) + \frac{1}{3}\epsilon^{\mu\nu\lambda\sigma}\partial_\mu(\lambda\mathcal{A}_{\nu\lambda\sigma}) \right\}. \qquad (23)$$

We stress again, that despite appearances (22) and (23) represent the same theory. Note also that in all cases of interest to us we have assumed that the bulk action only depended on $\mathcal{F} = d\mathcal{A}$, which means that $\mathcal{A}$ is the field theory analogue of cyclic coordinates in mechanics. This means that our procedure of trading the "conserved" quantity $\mathcal{F}$ for its "integral of motion" λ is really the procedure of replacing the Lagrangian $\hat{\mathcal{L}}(\mathcal{F})$ by the Routhian $\mathcal{L}(\lambda)$.

D. Linear limit

Before proceeding with adding the charged membranes to (22), (23), let us briefly discuss the special limit when $\mathcal{L}$ is a linear function of λ—i.e., precisely the Henneaux-Teitelboim example, which was augmented with charged tensional membranes in [6]. It is immediately clear that the Legendre transformation (21) breaks down in this case: if, e.g., $\hat{\mathcal{L}}(\mathcal{F}) = c\mathcal{F}$, $\lambda = c$ and $\mathcal{L}(\lambda) \equiv 0$. In turn, this is not surprising at all since, as is well known, Legendre transformation establishes a relation between a family of tangents to a curve that is the envelope of this family of straight lines. When the curve is a straight line, it has a fixed tangent—i.e., itself—which is its own envelope. This shows that the linear example is a degenerate limit of the general case. To see it, we can start with a general $\hat{\mathcal{L}}(\mathcal{F})$ and truncate it for simplicity to only quadratic terms,

$$\hat{\mathcal{L}}(\mathcal{F}) = \frac{\alpha}{2}\mathcal{F} - \frac{c}{8}\mathcal{F}^2 + \dots. \qquad (24)$$

The Legendre transform rules (21) yield

$$\lambda = \frac{\alpha}{2} - \frac{c}{4}\mathcal{F} + \dots, \qquad \mathcal{L}(\lambda) = \frac{2}{c}\left(\lambda - \frac{\alpha}{2}\right)^2 + \dots \qquad (25)$$

In the first equation of (25), the $(\dots)$ terms depend on higher powers of c, while in the second, the $(\dots)$ terms depend on higher powers of $1/c$. Clearly, the limit $c \to 0$, which reduces (24) to a linear term only is singular.

Yet note that without charges a field redefinition exists (as remarked earlier),

$$\tilde{\lambda} = \mathcal{L}(\lambda), \qquad \tilde{\mathcal{A}}_{\nu\lambda\sigma} = \frac{1}{\partial_\lambda\mathcal{L}(\lambda)}\mathcal{A}_{\nu\lambda\sigma}, \qquad (26)$$

which completely removes the nonlinearities, since the functional Jacobian of this redefinition is unity,

$$\mathcal{J} = \det\frac{\partial(\tilde{\lambda}, \tilde{\mathcal{A}}_{\nu\lambda\sigma})}{\partial(\lambda, \mathcal{A}_{\nu\lambda\sigma})} = \det\begin{pmatrix} \partial_\lambda\mathcal{L}(\lambda) & 0 \\ -\frac{\mathcal{A}_{\nu\lambda\sigma}}{\partial_\lambda^2\mathcal{L}(\lambda)} & \frac{1}{\partial_\lambda\mathcal{L}(\lambda)} \end{pmatrix} = 1. \qquad (27)$$

Although this shows that the singularity $c \to 0$ can be sidestepped, it also points that the procedure is noncommutative since, if we took the limit $c \to 0$ before dualizing, it would be unclear how to trade variables. This is because the steps involve formally divergent contributions to the cosmological constant which should be regulated and subtracted away. This problem was circumvented in [4–6] by promoting $M_{\rm Pl}^2$ into a flux of a second 4-form and dualizing the theory using bilinear terms.

Note however that when charged membranes are included, they obstruct the field redefinition (26), because the couplings change.[9] Hence with membranes included, purely linear theory and higher power theories are physically distinct. In any case, in what follows we will mostly work with the quadratic truncation of $\mathcal{L}$, which is practically indistinguishable for our purposes whenever the linear term dominates.

IV. CHARGING UP

So far we have been working with unimodular gravity/ 4-form theory without charged membranes. However, a theory of 4-forms is a gauge theory, describing dynamics of membranes charged under $\mathcal{A}$. What is more, the "braneless" theory of 4-forms, in addition to usual local gauge symmetries $\mathcal{A} \to \mathcal{A} + d\omega$ also has generalized global 3-form symmetry, associated with the "current conservation" $d^*\mathcal{F} = 0$ [48]. The lore that QG does not permit

---

[9]Analogously to the change of basis from interaction to propagation eigenstates in the theory of flavor oscillations.

global symmetries [49,50] (see also [51] for a review) is simply incorporated by breaking the generalized global symmetry by adding objects charged under $\mathcal{A}$—the membranes, precisely. For our purposes, in the semiclassical limit, this means that we should enhance (22), (23) with the inclusion of membrane contributions and boundary terms required to properly provide junction conditions across the membrane walls. Working with the λ variables for convenience, the action becomes

$$
S = \int d^4x \left\{ \sqrt{g} \left(\frac{M_{\mathrm{Pl}}^2}{2} R - \mathcal{L}_{\mathrm{QFT}} - \mathcal{L}(\lambda) \right) \right.
$$
$$
\left. + \frac{1}{3} \epsilon^{\mu\nu\lambda\sigma} \partial_\mu(\lambda) \mathcal{A}_{\nu\lambda\sigma} \right\} - \int d^3\xi \sqrt{\gamma} M_{\mathrm{Pl}}^2 [K]
$$
$$
- \mathcal{T}_{\mathcal{A}} \int d^3\xi \sqrt{\gamma}_{\mathcal{A}} - \mathcal{Q}_{\mathcal{A}} \int \mathcal{A}, \tag{28}
$$

where $\mathcal{T}_{\mathcal{A}}$ and $\mathcal{Q}_{\mathcal{A}}$ are the membrane tension and charge, respectively, the term $\propto K$ is the Israel-Gibbons-Hawking term for gravity which encodes boundary conditions across membrane walls, and $[\ldots]$ is the jump across a membrane. The coordinates ξ are coordinates along membrane world volume, embedding it in spacetime. The charge terms are

$$
\int \mathcal{A} = \frac{1}{6} \int d^3\xi \mathcal{A}_{\mu\nu\lambda} \frac{\partial x^\mu}{\partial \xi^\alpha} \frac{\partial x^\nu}{\partial \xi^\beta} \frac{\partial x^\lambda}{\partial \xi^\gamma} \epsilon^{\alpha\beta\gamma}. \tag{29}
$$

As is customary, $\mathcal{T}_{\mathcal{A}} > 0$ to avoid problems with negative energies.

Classically, membranes can be added to the spacetime as sources of the gauge fields. Their number is fixed, and the location can change by their interactions among themselves and also with other sources of gravitational fields. Quantum mechanically, however, the membranes can nucleate in background fields [2,3]. Such processes change the distribution of sources and their number, and lead to nontrivial transitions in the background geometry. The classical "superselection sectors" (with a fixed number of membranes) now all mix up.

This induces the evolution in the space of geometries due to the variation of λ. In particular, such transitions change the local value of the cosmological "constant," which unlike in chargeless unimodular gravity is not a constant anymore. It can change discretely by quantum creation of membranes. Thus, if we start with a given de Sitter space with an initial value of cosmological constant, it will evolve by membrane production into a geometry that is locally de Sitter, but where the cosmological constant varies from bubble to bubble. This will look like the original "old inflation" of Guth [52], but with a huge number of discretely separated false vacua.

We have initiated a study of those phenomena in [4–6] for the special case of linear $\mathcal{L}(\lambda) = M_{\mathrm{Pl}}^2 \lambda$. Here, we will redo this analysis for the general case, and extract the

salient features of such "unimodular," or pancosmic, landscapes.

V. DE SITTER INSTABILITY

The quantum membrane discharge in the semiclassical limit can be described by the action (28) in Euclidean time. This action controls the nucleation processes and sets their rates, $\Gamma \sim e^{-S_E}$ [53–55]. Here we follow the same steps as in [4–6]: first, we Wick rotate the action using $t = -ix_E^0$, which gives $-i \int d^4x \sqrt{g} \mathcal{L}_{\mathrm{QFT}} = -\int d^4x_E \sqrt{g} \mathcal{L}_{\mathrm{QFT}}^E$. Next, with the convention $\mathcal{A}_{0jk} = \mathcal{A}_{0jk}^E$, $\mathcal{A}_{jkl} = \mathcal{A}_{jkl}^E$ we get $\mathcal{F}_{\mu\nu\lambda\sigma} = \mathcal{F}_{\mu\nu\lambda\sigma}^E$. In addition $\epsilon_{0ijk} = \epsilon_{0ijk}^E$ and $\epsilon^{0ijk} = -\epsilon_E^{0ijk}$. The tension and charge terms transform to $-i\mathcal{T}_{\mathcal{A}} \int d^3\xi \sqrt{\gamma} = -\mathcal{T}_{\mathcal{A}} \int d^3\xi_E \sqrt{\gamma}$ and $i\mathcal{Q}_{\mathcal{A}} \int \mathcal{A}_i = -\mathcal{Q}_{\mathcal{A}} \int \mathcal{A}_i$. The scalars do not change (but if they include time derivatives, those terms change accordingly). We define Euclidean action by $iS = -S_E$. Our Euclidean action is

$$
S_E = \int d^4x_E \left\{ \sqrt{g} \left(-\frac{M_{\mathrm{Pl}}^2}{2} R_E + \mathcal{L}(\lambda) + \Lambda_{\mathrm{QFT}} \right) \right.
$$
$$
\left. + \frac{1}{3} \epsilon_E^{\mu\nu\lambda\sigma} \partial_\mu(\lambda) \mathcal{A}_{\nu\lambda\sigma}^E + \int d^3\xi \sqrt{\gamma} M_{\mathrm{Pl}}^2 [K_E] \right.
$$
$$
\left. + \mathcal{T}_{\mathcal{A}} \int d^3\xi_E \sqrt{\gamma}_{\mathcal{A}} - \frac{\mathcal{Q}_{\mathcal{A}}}{6} \int d^3\xi_E \mathcal{A}_{\mu\nu\lambda}^E \frac{\partial x^\mu}{\partial \xi^\alpha} \frac{\partial x^\nu}{\partial \xi^\beta} \frac{\partial x^\lambda}{\partial \xi^\gamma} \epsilon_E^{\alpha\beta\gamma} \right\}. \tag{30}
$$

From here on we drop the index E.

We then restrict our attention to "vacuum evolution": we only consider transitions between locally maximally symmetric backgrounds. Those have local $O(4)$ symmetry and so dominate in semiclassical limit since they have minimal Euclidean action [53–55]. Therefore we set $\langle \mathcal{L}_{\mathrm{QFT}}^E \rangle = \Lambda_{\mathrm{QFT}}$, with Λ_{QFT} the regulated matter sector vacuum energy to an arbitrary loop order. We can further imagine that the divergent parts in the limit where regulator decouples are subtracted away by the counterterm $\mathcal{L}(\lambda)$, whose finite part is still completely arbitrary. From the QFT/gravity couplings, we infer, as before [4–6], that $\Lambda_{\mathrm{QFT}} = \mathcal{M}_{\mathrm{UV}}^4 + \ldots \equiv M_{\mathrm{Pl}}^2 \lambda_{\mathrm{QFT}}$, where $\mathcal{M}_{\mathrm{UV}}^4$ is the QFT UV cutoff and ellipsis denote subleading terms [56,57]. This means that we can write the total cosmological constant in any patch as

$$
\Lambda_{\mathrm{total}} = \Lambda_{\mathrm{QFT}} + \mathcal{L}(\lambda) = M_{\mathrm{Pl}}^2 \lambda_{\mathrm{QFT}} + \mathcal{L}(\lambda), \tag{31}
$$

where λ can vary from patch to patch across membrane walls.

The transitions induced by nucleations of a single membrane, which are dominant channels here, can be approximated by geometries with $\Lambda_{\mathrm{total\,out/in}}$ glued together along a membrane, where the subscripts out/in denote parent and offspring geometries (exterior and

interior of a membrane, respectively). Both of the out/in geometries are described with the metrics

$$ds_E^2 = dr^2 + a^2(r)d\Omega_3, \tag{32}$$

where $d\Omega_3$ is the line element on a unit S^3. The Euclidean scale factor a is the solution of the Euclidean "Friedmann equation,"

$$3M_{\text{Pl}}^2\left(\left(\frac{a'}{a}\right)^2 - \frac{1}{a^2}\right) = -\Lambda_{\text{total}}, \tag{33}$$

which follows because the bulk metric-dependent part of (30) is structurally the same as in standard general relativity. The prime designates an r derivative. From here on, we will drop the subscript "total."

To construct these geometries, we need to assemble together two patches, each with a local metrics given by (32) but with different Λ, and then use the junction conditions to connect the patches. The boundary conditions induced on a membrane follow from (30) by varying with respect to $\mathcal{A}$. Similarly, the boundary conditions on the metric follow from Israel-Gibbon-Hawking junction conditions. Summarizing [4–6],

$$a_{\text{out}} = a_{\text{in}}, \qquad \lambda_{\text{out}} - \lambda_{\text{in}} = \frac{\mathcal{Q}_\mathcal{A}}{2},$$
$$M_{\text{Pl}}^2\left(\frac{a'_{\text{out}}}{a} - \frac{a'_{\text{in}}}{a}\right) = -\frac{\mathcal{T}_\mathcal{A}}{2}, \tag{34}$$

in the coordinate system where the outward membrane normal vector is oriented in the direction of the radial coordinate; r measures the distance in this direction. The discontinuities in λ and a' follow since a membrane is a Dirac δ-function source of charge and tension.

The next step is to solve (33) for $a' = \zeta_j\sqrt{1 - \frac{\Lambda a^2}{3M_{\text{Pl}}^2}}$, with $\zeta_j = \pm 1$ designating the two possible branches of the square root, and rewrite the discontinuous junction conditions in a more convenient form, as in [4–6]. Using $a'^2_{\text{out}} - a'^2_{\text{in}} = -a^2(\Lambda_{\text{out}} - \Lambda_{\text{in}})/3M_{\text{Pl}}^2$, which follows from (33), and the last equation of (34), we can extract an equation for $a'_{\text{out}} + a'_{\text{in}}$. Then, adding and subtracting those two equations, we solve for $a'_{\text{out/in}}$. Finally, we solve for $a'_{\text{out/in}}$ from (33) to obtain

$$\zeta_{\text{out}}\sqrt{1 - \frac{\Lambda_{\text{out}}a^2}{3M_{\text{Pl}}^2}} = -\frac{\mathcal{T}_\mathcal{A}}{4M_{\text{Pl}}^2}\left(1 - \frac{4M_{\text{Pl}}^2}{3\mathcal{T}_\mathcal{A}^2}(\Lambda_{\text{out}} - \Lambda_{\text{in}})\right)a,$$

$$\zeta_{\text{in}}\sqrt{1 - \frac{\Lambda_{\text{in}}a^2}{3M_{\text{Pl}}^2}} = \frac{\mathcal{T}_\mathcal{A}}{4M_{\text{Pl}}^2}\left(1 + \frac{4M_{\text{Pl}}^2}{3\mathcal{T}_\mathcal{A}^2}(\Lambda_{\text{out}} - \Lambda_{\text{in}})\right)a. \tag{35}$$

Here $\zeta_i = \pm$ can be viewed as discrete conserved charges. They pick one of two possible branches of the square root

of $(\frac{a'}{a})^2 - \frac{1}{a^2} = -\Lambda/3M_{\text{Pl}}^2$. The configurations that result from gluing together sections of exterior and interior metrics (32) are therefore counted by the variations of the sign of Λ and the branches of solutions ($\zeta_j = \pm 1$) of Euclidean Friedmann equation (33). The taxonomy of allowed solutions is presented in detail in [4–6], and follows the steps described by [2,3]. Here we focus on the novel ingredients that follow from the generalized theory (28).

The main new ingredient is that from (31), $\Lambda_{\text{out}} - \Lambda_{\text{in}} = \mathcal{L}(\lambda_{\text{out}}) - \mathcal{L}(\lambda_{\text{in}})$. Combining this with our last remaining junction condition from (34), $\lambda_{\text{out}} - \lambda_{\text{in}} = \frac{\mathcal{Q}_\mathcal{A}}{2}$, yields

$$\Lambda_{\text{out}} - \Lambda_{\text{in}} = \mathcal{L}(\lambda_{\text{out}}) - \mathcal{L}\left(\lambda_{\text{out}} - \frac{\mathcal{Q}_\mathcal{A}}{2}\right). \tag{36}$$

This equation has important consequences. To see this, first off, we can check that, in the linear limit, $\mathcal{L} = M_{\text{Pl}}^2\lambda$, this yields $\Lambda_{\text{out}} - \Lambda_{\text{in}} = \frac{M_{\text{Pl}}^2\mathcal{Q}_\mathcal{A}}{2}$. Plugging this into Eqs. (35) precisely reproduces the formulas found in [4–6], since the right-hand side (rhs) of (35) becomes $\mp\frac{\mathcal{T}_\mathcal{A}}{4M_{\text{Pl}}^2}(1 \mp q)$, where $q = \frac{2M_{\text{Pl}}^4\mathcal{Q}_\mathcal{A}}{3\mathcal{T}_\mathcal{A}^2}$.

When $\mathcal{L}$ is quadratic, $\mathcal{L} = 2\lambda^2$, we obtain[10] $\Lambda_{\text{out}} - \Lambda_{\text{in}} \simeq 4\lambda\Delta\lambda = 2\lambda\mathcal{Q}_\mathcal{A}$, and so the terms controlling the boundary conditions (35) on the rhs are $\simeq \mp\frac{\mathcal{T}_\mathcal{A}}{4M_{\text{Pl}}^2}(1 \mp \frac{8M_{\text{Pl}}^2\lambda\mathcal{Q}_\mathcal{A}}{3\mathcal{T}_\mathcal{A}^2})$. This is precisely the behavior found in [2,3], and later also encountered in, e.g., [36]. The implications for the dynamics are crucial, since these terms control the selection rules that allow or prohibit a specific type of instanton that mediates transitions of initial de Sitter.

In the linear case, the point is that for $|q| < 1$ the only allowed instantons are those where the rhs has a fixed sign: $(-)$ in the first and $(+)$ in the second equation. Since $\mathcal{T}_\mathcal{A} > 0$, this uniquely fixes the signs of $\zeta_{\text{out/in}}$: $\zeta_{\text{out}} = -1, \zeta_{\text{in}} = +1$. Thus only instantons that have $(\zeta_{\text{out}}, \zeta_{\text{in}}) = (-, +)$ are allowed. Specifically, for the transitions of de Sitter to de Sitter, completely independently of the initial and final value of cosmological constant, there is a unique allowed process, described by the instanton in Fig. 1; all other possibilities are prohibited.[11]

In contrast, in the quadratic case, the boundary condition depends on $(1 \mp \frac{8M_{\text{Pl}}^2\lambda\mathcal{Q}_\mathcal{A}}{3\mathcal{T}_\mathcal{A}^2})$ and therefore on the parent value of the 4-form flux $\lambda = -(\frac{1}{48\sqrt{g}}\epsilon^{\mu\nu\lambda\sigma}\mathcal{F}_{\mu\nu\lambda\sigma})$. This means that, for a fixed $\mathcal{Q}_\mathcal{A}$, when the flux $\mathcal{F}$ is small the signs on the rhs will again uniquely fix $\zeta_{\text{out}}, \zeta_{\text{in}}$, allowing only the instanton of Fig. 1. However, for large flux, which is generically

---

[10]For a small $\mathcal{Q}_\mathcal{A}$, the difference between $\lambda_{\text{out/in}}$ is much smaller than either one, and we can ignore the subscript here.

[11]There is only one more process, whereby $dS \to AdS$ [4,5].

required to screen a large QFT contribution to cosmological constant Λ_{QFT} [36], the situation is reversed: the instanton of Fig. 1 is disallowed, since the large $\mathcal{F}$ limit corresponds to $|q| > 1$. Instead, only $(\zeta_{out}, \zeta_{in}) = (+, +)$ or $(-, -)$ can occur. The dominant instanton in this case is given in Fig. 2. The reason this instanton dominates follows from the membrane nucleation rates, $\Gamma \sim e^{-S_{bounce}}$ [53–55]. To compute them, we need the bounce action, which is defined by $S(\text{bounce}) = S(\text{instanton}) - S(\text{parent})$. A precise calculation then yields [2–5]

FIG. 1. A $|q| < 1$ instanton mediating $dS \to dS$.

$$S(\text{bounce}, a) = 2\pi^2 a^3 \mathcal{T}_\mathcal{A} - 2\pi^2 \left(\Lambda_{in} \int_{\text{North Pole}}^a da \left(\frac{a^3}{a'}\right)_{in} + 3M_{Pl}^2 a^2 \zeta_{in} \sqrt{1 - \frac{\Lambda_{in} a^2}{3M_{Pl}^2}} \right)$$
$$+ 2\pi^2 \left(\Lambda_{out} \int_{\text{North Pole}}^a da \left(\frac{a^3}{a'}\right)_{out} + 3M_{Pl}^2 a^2 \zeta_{out} \sqrt{1 - \frac{\Lambda_{out} a^2}{3M_{Pl}^2}} \right). \tag{37}$$

In this formula, the "outside" contributions to the integrals coming from the parent geometry completely cancel between the instanton and the "parent reference" actions. The only contributions from the parent come from the complement of the outside geometry of the instantons [2,3,53]. In the $(+, +)$ instanton, this means that the parent contributions to the bounce action are small, comparable in size to the "offspring" contribution. As a result, the integral is dominated by the area surrounding the membrane, since the contributions from the polar regions are small. This is in contrast to the $(-, +)$ instanton of Fig. 1, where the complement of the parent section of the instanton geometry is most of the initial de Sitter, and so the bounce action will be dominated by the Euclidean–de Sitter parent area—i.e., precisely the Gibbons-Hawking action, as we will show in detail below.

Whenever the $(+, +)$ instanton dominates, as Λ becomes smaller, the action for the $(+, +)$ instantons converges to (see [2–5,53–55] and many other papers)

$$S_{bounce} \simeq \frac{27\pi^2}{2} \frac{\mathcal{T}_\mathcal{A}^4}{(\Delta\Lambda)^3}. \tag{38}$$

Hence, the decay rate saturates as the cosmological constant decreases and, as a result, while any initial de Sitter space will steadily evolve towards lower cosmological constant values (easily reaching $\Lambda < 0$), all values of Λ will be approximately equally likely, and, below some critical value, very long lived. It thus seems that as long as these instantons dominate, the only way to argue that the terminal cosmological constant induced by transitions is small and nonnegative is to invoke anthropics. This is the reason we think the alternative framework where the only allowed instanton is $(-, +)$ of (1) is preferred.
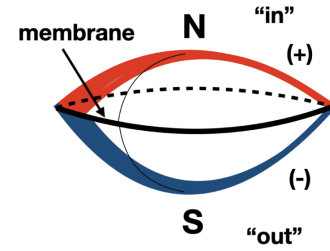
This conclusion remains valid for more general cases, when $\mathcal{L}$ includes more powers of λ as in the Taylor series (8), $\mathcal{L}(\lambda) = c_1 M_{Pl}^2 \lambda + 2c_2 \lambda^2 + 4c_3 \frac{\lambda^3}{M_{Pl}^2} + \cdots$. The equations look more involved, but the physical outcome is similar to the two simple limits above: generically, for small fluxes the instanton (1) is the only one allowed, whereas for large fluxes the instanton (2) dominates. This means, that in the absence of the linear term, all theories with higher powers of 4-form flux behave like pure $\mathcal{F}^2$. However, when the linear term is present, it dominates for all sub-Planckian fluxes $\mathcal{F}/M_{Pl}^2 < 1$ [assuming $c_1 \sim \mathcal{O}(1)$]. The simplest case $\mathcal{L} = c_1 M_{Pl}^2 \lambda + 2c_2 M_{Pl}^2 \lambda^2$, which we elaborated upon above, serves as a clear example. Since $\Lambda_{out} - \Lambda_{in} = \frac{c_1}{2} M_{Pl}^4 \mathcal{Q}_\mathcal{A} + 2c_2 M_{Pl}^2 \lambda \mathcal{Q}_\mathcal{A} + \mathcal{O}(\mathcal{Q}_\mathcal{A}^2)$, the junction conditions (35) to leading order become
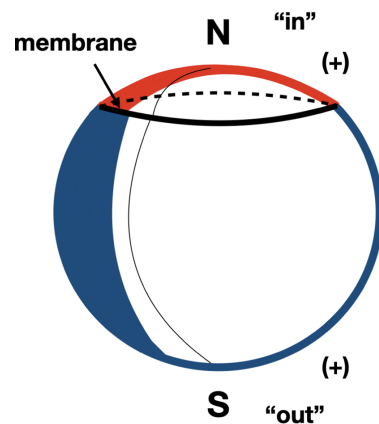

FIG. 2. A large flux instanton mediating $dS \to dS$.

$$\zeta_{\rm out}\sqrt{1-\frac{\Lambda_{\rm out}a^2}{3M_{\rm Pl}^2}}=-\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\left(1-\frac{2M_{\rm Pl}^4\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\right.$$

$$\left.\times\left(c_1+4c_2\frac{\lambda}{M_{\rm Pl}^2}\right)\right)a,$$

$$\zeta_{\rm in}\sqrt{1-\frac{\Lambda_{\rm in}a^2}{3M_{\rm Pl}^2}}=\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\left(1+\frac{2M_{\rm Pl}^4\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\right.$$

$$\left.\times\left(c_1+4c_2\frac{\lambda}{M_{\rm Pl}^2}\right)\right)a. \tag{39}$$

Hence whenever $c_2\sim c_1$ and the flux is sub-Planckian, $\lambda<M_{\rm Pl}^2$, the flux contribution to the junction conditions (39) is subleading. Ergo, the large sub-Planckian fluxes might only influence the nonperturbative physics selection rules initially, at scales close to $M_{\rm Pl}$, but their influence quickly wanes. In natural theories where it is not prohibited, the linear term does all the driving.

This persists for general $\mathcal{L}$. After a straightforward manipulation, one finds to linear order in $\Delta\lambda$

$$\zeta_{\rm out}\sqrt{1-\frac{\Lambda_{\rm out}a^2}{3M_{\rm Pl}^2}}=-\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\left(1-\frac{2\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda/M_{\rm Pl}^2)}\right)a,$$

$$\zeta_{\rm in}\sqrt{1-\frac{\Lambda_{\rm in}a^2}{3M_{\rm Pl}^2}}=\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\left(1+\frac{2\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda/M_{\rm Pl}^2)}\right)a. \tag{40}$$

This means that as long as

$$|Q|<1, \qquad Q=\frac{2\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda/M_{\rm Pl}^2)}, \tag{41}$$

the only instantons that can facilitate $dS\to dS$ transitions (in the limit $\Delta\lambda/\lambda\ll1$) are those of Fig. 1. This equation generalizes $|q|<1$ and is always satisfied when the linear term dominates[12] $\mathcal{L}$.

From here on, for simplicity we will assume that (41) holds and approximate $\mathcal{L}$ by the linear term. To proceed, we can solve Eqs. (35) for a^2:

$$\frac{1}{a^2}=\frac{\Lambda_{\rm out}}{3M_{\rm Pl}^2}+\left(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\right)^2(1-Q)^2$$

$$=\frac{\Lambda_{\rm in}}{3M_{\rm Pl}^2}+\left(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2}\right)^2(1+Q)^2. \tag{42}$$

As noted in [4–6], using this we discern two regimes of membrane nucleations. When a^2 is comparable to de Sitter

---

[12]One may ask how this compares to the "weak gravity conjecture." Transliterating the inequalities of [58], Weak Gravity Conjecture (WGC) requires $\mathcal{Q}_{\mathcal{A}}/\mathcal{T}_{\mathcal{A}}>1/M_{\rm Pl}$. Combining this and (41) yields $\frac{1}{M_{\rm Pl}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda)}<\frac{3}{2}\frac{\mathcal{T}_{\mathcal{A}}}{M_{\rm Pl}^3}$. Thus for sub-Planckian tension membranes this will be satisfied if $\mathcal{L}=cM_{\rm Pl}^2\lambda$ with $c<\frac{3}{2}\frac{\mathcal{T}_{\mathcal{A}}}{M_{\rm Pl}^3}<\frac{3}{2}$. This might be realized by, e.g., strong dynamics effects and SUSY breaking below the Planck scale [26]. Suffice it to say for now that the bound does not appear prohibitive.

---

radii, from Eq. (42), $a^2\sim(1-\frac{\Lambda_j a^2}{3M_{\rm Pl}^2})^{1/2}\ll1$ and so the bounce action is approximately

$$S_{\rm bounce}\simeq-\frac{12\pi^2M_{\rm Pl}^4\Delta\Lambda}{\Lambda_{\rm out}\Lambda_{\rm in}}. \tag{43}$$

In this regime the transitions are fast because $S_{\rm bounce}<0$. The reverse processes increasing Λ are suppressed because their bounce action is the sign reversed (43) and so they are rarer. Every once in a while de Sitter space is given a push up. But mostly it decays to flatter space, decreasing Λ. This lasts as long as $\Lambda_{\rm out}\gg3M_{\rm Pl}^2(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2})^2$.

Once Λ decreases to $\Lambda\lesssim3M_{\rm Pl}^2(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2})^2$, the discharge nucleations proceed via production of small bubbles, which have the bounce action [5]

$$S_{\rm bounce}\simeq\frac{24\pi^2M_{\rm Pl}^4}{\Lambda_{\rm out}}\left(1-\frac{8}{3}\frac{M_{\rm Pl}^2\Lambda_{\rm out}}{\mathcal{T}_{\mathcal{A}}^2}\right), \tag{44}$$

and $S_{\rm bounce}>0$ because $\Lambda<3M_{\rm Pl}^2(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2})^2$. This action is remarkable: as $\Lambda_{\rm out}\to0$ the action diverges. Perhaps this is not entirely surprising: as Λ decreases the geometric entropy $\mathcal{S}_{GH}\simeq\frac{24\pi^2M_{\rm Pl}^4}{\Lambda_{\rm out}}$ grows, and the ensuing "chaos" takes over. The decay towards Minkowski then simply looks like the enforcement of the second law of thermodynamics. As a result the bubbling rate $\Gamma\sim e^{-S_{\rm bounce}}$ has an essential singularity at $\Lambda_{\rm out}\to0$, where the rate goes to zero [4–6]. Hence when $|Q|<1$, small cosmological constants become very long lived, and the closer the geometry gets to a locally Minkowski space, the more stable it becomes to discharges. If it ends up with zero cosmological constant, then further discharge stops.

Hence we see that as a result of augmenting unimodular gravity with charged membranes, de Sitter spaces of the theory are immediately rendered unstable. Membranes will nucleate quantum mechanically, with the dominant trend being the discharge of the cosmological constant in their interior. When the membrane charge and tension in any generalized unimodular gravity satisfy $\left|\frac{2\mathcal{Q}_{\mathcal{A}}}{3\mathcal{T}_{\mathcal{A}}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda/M_{\rm Pl}^2)}\right|<1$ of Eq. (41), which can be realized, for example, when $\mathcal{L}$ is dominated by a term liner in λ for $|\lambda|<M_{\rm Pl}^2$, the discharge can only occur via a single channel controlled by the instanton of Fig. 1. In this case, when the background cosmological constant decreases below $\Lambda_{\rm critical}\simeq3M_{\rm Pl}^2(\frac{\mathcal{T}_{\mathcal{A}}}{4M_{\rm Pl}^2})^2$, the subsequent transitions are slow, since the bounce action is approximately given by the parent geometry horizon area (44) [59,60]. The deluge of bubbles bounded by membranes eventually comes down to a trickle, clearly favoring Minkowski space as an asymptotic attractor. In the next section we discuss how close to Minkowski this evolution gets, and how it can provide a means to address the cosmological constant problem.

VI. RELAXATION OF THE COSMOLOGICAL CONSTANT

The discussion above shows that

(1) de Sitter space is unstable in braney unimodular gravity,

(2) the evolutionary trend is towards decreasing Λ, and

(3) when the flux sector is such that $|\frac{2\mathcal{Q}_\mathcal{A}}{3\mathcal{T}_\mathcal{A}^2}\frac{\partial\mathcal{L}(\lambda)}{\partial(\lambda/M_{\rm Pl}^2)}| < 1$, which is most simply realized when $\mathcal{L}(\lambda)$ is dominated by the linear term, as in [4–6], the Minkowski space $\Lambda/M_{\rm Pl}^4 \to 0^+$ is the unique attractor.

These evolutionary trends go in the right direction. Nevertheless, this mechanism of cosmological constant relaxation requires additional ingredients if it is to be employed to address the cosmological constant problem.

The first issue is the question of how close does the evolution get to Minkowski. On the one hand, the observations imply that the "terminal" achievable value of Λ should be in the range $\Lambda/M_{\rm Pl}^4 \lesssim 10^{-120}$ or so. On the other hand, the evolution of the fluxes by the nucleation of membranes shows that variable λ is quantized, changing in units of $\mathcal{Q}_\mathcal{A}/2$ [61,62]

$$\lambda = N\frac{\mathcal{Q}_\mathcal{A}}{2}. \tag{45}$$

Here, as we include general operators $\mathcal{F}^n, n = 2, 3, \ldots$ and not only the linear terms, we have set the possible "background value" λ_0 to zero imposing the Dirac quantization condition [61,62] amounting to the addition of magnetic monopole sources. Previously, working with strictly linear theory (in the action) in fluxes [4–6], we kept λ_0 explicitly, but since we could absorb it into the cosmological constant counterterm, it was of no consequence. This means that the total cosmological constant in the theory is

$$\Lambda_{\rm total} = M_{\rm Pl}^2\lambda_{\rm QFT} + \mathcal{L}\left(N\frac{\mathcal{Q}_\mathcal{A}}{2}\right)$$
$$= M_{\rm Pl}^2\left(\lambda_{\rm QFT} + Nc\frac{\mathcal{Q}_\mathcal{A}}{2} + \ldots\right), \tag{46}$$

where $\lambda_{\rm QFT}$ includes all the contributions to $\Lambda_{\rm total}$ which are not sourced by membranes. The last equality follows since we are assuming that the linear term dominates in $\mathcal{L}$. Ergo, since the membrane nucleations can only change $\Lambda_{\rm total}$ in the units of $\mathcal{Q}_\mathcal{A}/2$, it means that a sequence of bubbles can evolve an initial $\Lambda_{\rm total}$ to a terminal value $\Lambda/M_{\rm Pl}^4 \lesssim 10^{-120}$ only if either (1) the initial value is tuned to be $10^{-120}M_{\rm Pl}^4 + N_{\rm initial}cM_{\rm Pl}^2\frac{\mathcal{Q}_\mathcal{A}}{2}$ or (2) $cM_{\rm Pl}^2\frac{\mathcal{Q}_\mathcal{A}}{2} < 10^{-120}M_{\rm Pl}^4$ [2–6]. The first option is a clear fine-tuning, and the second is not only theoretically dubitable, but cosmologically problematic since it leads to the empty universe problem, to be discussed in the next section.

To get around the fine tuning problem, in [4–6] a second system of forms and membranes was introduced, which
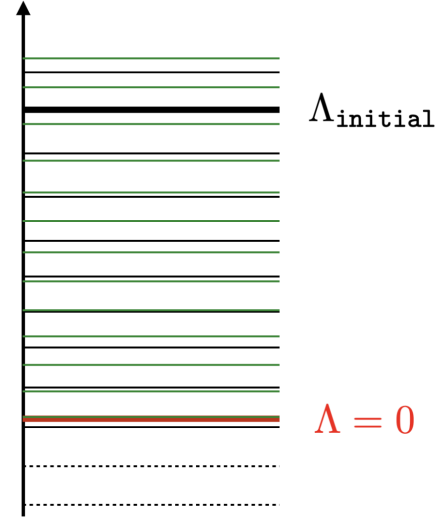

FIG. 3. A discharge cascade in the cosmological constant discretuum. Starting with some initial state given by a black line, the evolution will reduce Λ to a level where the black and green lines are degenerate, and subsequent evolution can proceed on the green ladder with the terminal step coming arbitrarily close to zero.

looked exactly the same as the theory of $\mathcal{A}$ and its membranes. Since the second form sector is completely degenerate with the first one, the total cosmological constant is now[13]

$$\Lambda_{\rm total} = M_{\rm Pl}^2(\lambda_{\rm QFT} + c\lambda + \hat{c}\,\hat{\lambda} + \ldots)$$
$$= M_{\rm Pl}^2\left(\lambda_{\rm QFT} + \frac{\mathcal{Q}_\mathcal{A}}{2}(cN + \hat{c}\,\hat{N}\,\omega) + \ldots\right). \tag{47}$$

Here $\omega = \hat{\mathcal{Q}}_{\hat{\mathcal{A}}}/\mathcal{Q}_\mathcal{A}$ is the ratio of membrane charges. In [4–6] we worked in the linear limit with $c = \hat{c} = 1$, having renormalized the charges $\mathcal{Q}_\mathcal{A}, \hat{\mathcal{Q}}_{\hat{\mathcal{A}}}$. This means that, to get arbitrarily close to $\Lambda \to 0^+$, the charge ratio ω needs to be an irrational number [49,63]. If so, there exist integers $N, \hat{N}$ such that $N + \hat{N}\omega$ is arbitrarily close to $-\frac{2\lambda_{\rm QFT}^2}{\mathcal{Q}_\mathcal{A}}$. The set of values of $\lambda_{\rm QFT} + c\lambda + \hat{c}\,\hat{\lambda}$ is discrete but it is dense in a set of reals, with values arbitrarily close to any real number including zero [49,63]. Moreover, there exist many sequences of discharging membranes, which will arrive to $N, \hat{N}$ at which point the cosmological constant is arbitrarily close to zero, and the underlying nearly flat space is very long lived, due to the pole of the bounce action, Eq. (44). We illustrate this in Fig. 3.

To reach this conclusion, what is really required is that the ratio of the energy densities that individual fluxes contribute to $\Lambda_{\rm total}$ should be irrational. If charges are mutually irrational, and $\mathcal{L}(\lambda)$ is dominated by linear terms this follows, but in general other options might occur.

---

[13]In [4–6] we had $c = \hat{c} = 1$. Here we keep them arbitrary, in principle.

The respective flux energies might be mutually irrational due to compactification effects, if dual 4-forms arise from dimensional reduction, and/or due to strong coupling effects in some gauge theory, or thanks to higher powers in $\mathcal{L}$; e.g., $\sqrt{1+x^2} = \sum_k \frac{\Gamma(3/2)}{k!\Gamma(3/2-k)} x^2$ will map integers into irrational numbers. We will not delve here into the specific details of how irrational variations can arise, deferring that for later work. For the rest of the current paper, we will restrict to linear terms, and treat $c\mathcal{Q}_\mathcal{A}, \hat{c}\hat{\mathcal{Q}}_{\hat{\mathcal{A}}}$ as mutually incommensurate, for simplicity's sake.

In this case, as Fig. 3 shows, given that the bounce action (44) has a pole at vanishing Λ, the quantum attractor of the evolution, as we will now explain [4–6], is

$$\frac{\Lambda}{M_{\mathrm{Pl}}^4} \to 0^+. \qquad (48)$$

Recall the previous approaches with at least quadratic flux contributions to the total cosmological constant [2,3,30–33,36]. For a natural cosmological constant $|\Lambda_{\mathrm{QFT}}| \sim M_{\mathrm{Pl}}^4$, by Eq. (31), this means that $\mathcal{L} \sim \mathcal{F}^2 \sim M_{\mathrm{Pl}}^4$ and $\mathcal{F} \sim M_{\mathrm{Pl}}^2$. In the pure quadratic theory, this forces $\lambda \sim M_{\mathrm{Pl}}^2$. Hence the variable Q of Eq. (41), which controls which instanton the transition goes by, is $|Q| \sim \frac{M_{\mathrm{Pl}}\lambda}{\mathcal{T}_\mathcal{A}} | \frac{M_{\mathrm{Pl}}\mathcal{Q}_\mathcal{A}}{\mathcal{T}_\mathcal{A}} | \gg 1$ for typical cases where membrane charges and tensions are sub-Planckian.[14] Thus in this case the only transitions which occur are those mediated by the instanton of Fig. 2, whose bounce action saturates at (38). The transition rate from one value of Λ to another is independent of Λ, and so all values will be approximately equally likely. As a result, with a uniform distribution of initial values, the resulting dynamical distribution of values of Λ will remain uniform, since "anything goes" at the same rate when the instantons of Fig. 2 are allowed. To favor a small positive cosmological constant, one then needs to invoke the anthropic principle [65].

The addition of the linear term in $\mathcal{F}$ changes this picture dramatically [4–6]. When cosmological constant is $\Lambda_{\mathrm{QFT}} < M_{\mathrm{Pl}}{}^4$, linear term will dominate when $|Q| < 1$ and the only processes which allow subsequent discharges of Λ are those mediated by the instanton of Fig. 1. Hence when the cosmological constant reduces to below $\Lambda_{\mathrm{critical}} \simeq 3M_{\mathrm{Pl}}^2 (\frac{\mathcal{T}_\mathcal{A}}{4M_{\mathrm{Pl}}^2})^2$, their bounce action is approximated by the exponent of the Gibbons-Hawking action (44), and the further decay progressively slows down; the evolution

relaxes Λ to 0^+ by quantum Brownian drift, and it exponentially slows down as $\Lambda \to 0^+$. This means, as long as $\Lambda > 0$ and there are smaller values, eventually regions of smaller Λ will be created inside the parent region. The most stable universes are those with the smallest value of the positive cosmological constant.

Further, as long as the cosmological constant is not exactly zero, one further decay to $\Lambda < 0$ or an up jump to a large positive Λ are possible. In the former case, the evolution completely stops. If there is reheating and normal matter in the Universe, then this region will collapse into a black hole surrounded by the parent geometry. It is a terminal sink. On the other hand, if there is an up jump, then the sequence of discharges will repeat, seeking to find an ever smaller Λ once again. This gives the universe with a larger Λ yet another chance to discharge it by going first up and then discharging toward $\Lambda \to 0^+$. The process may be extremely slow, but that does not really matter as long as it can occur. As a result, the smaller values of Λ will also be statistically more likely.[15] The posterior distribution of Λ including the tunneling dynamics effects will not be uniform even if the prior (number frequency) distribution is flat. This dynamics favors smallest values.

Of course, to make sure that arbitrarily small values of cosmological constant can be approached, as we noted above there must be at least two sets of membranes, with incommensurate fluxes. In this case the configurations with tiny positive cosmological constant will accumulate, overwhelming other outcomes. Thus finding a tiny positive Λ is natural without anthropics. Rather interestingly, this can be achieved already with the simple linear + quadratic actions mentioned in [26]. Still we stress that this assumes no phase transitions at late times, in the matter sector after inflation, and it does not immediately lead to a prediction of current cosmic acceleration.

So we see that the likeliest values of the cosmological constant are small as opposed to large, addressing the "why is the cosmological constant not huge?" part of the problem. However we would be too hasty to declare victory at this point. Besides the obvious point that observations suggest that $\frac{\Lambda}{M_{\mathrm{Pl}}^4} \sim 10^{-120}$ instead of zero, in our analysis of the evolution we have treated Λ_{QFT} as a fixed number, ignoring any possible phase transitions in a late universe. This is an issue, since while the early evolution favors the tiniest values of Λ_{total}, the membrane nucleation processes that lead to it slow down tremendously when the local value of Λ becomes smaller than $\Lambda_{\mathrm{critical}} \simeq 3M_{\mathrm{Pl}}^2 (\frac{\mathcal{T}_\mathcal{A}}{4M_{\mathrm{Pl}}^2})^2$. Indeed,

---

[14]In the examples of [36], $\mathcal{Q}_\mathcal{A} \sim M_{11}^3/M_{\mathrm{Pl}}$ and $\mathcal{T}_\mathcal{A} \sim M_{11}^3$, where $M_{11} \ll M_{\mathrm{Pl}}$ is the 11D Planck scale, and so $|Q| \sim \frac{M_{\mathrm{Pl}}\lambda}{M_{11}^3} \gg 1$, as claimed. Note also that this yields $\frac{\mathcal{Q}_\mathcal{A}}{\mathcal{T}_\mathcal{A}} \sim \frac{1}{M_{\mathrm{Pl}}}$ and so the setup of [36] can marginally satisfy the weak gravity conjecture, as does the present work (see footnote 12). The recent paper [64], while generally confirming our observations from [5,6] and the present work, claims the opposite.

[15]If the bounce action is sufficiently big, a universe with a positive cosmological constant larger than observed may still be very long lived. But that is not the issue: with evolution driven by the instantons of Fig. 1, it will not be typical. There will be many more longer lived universes that have a smaller cosmological constant, and where subsequent cosmology could produce structures regardless of anthropics.
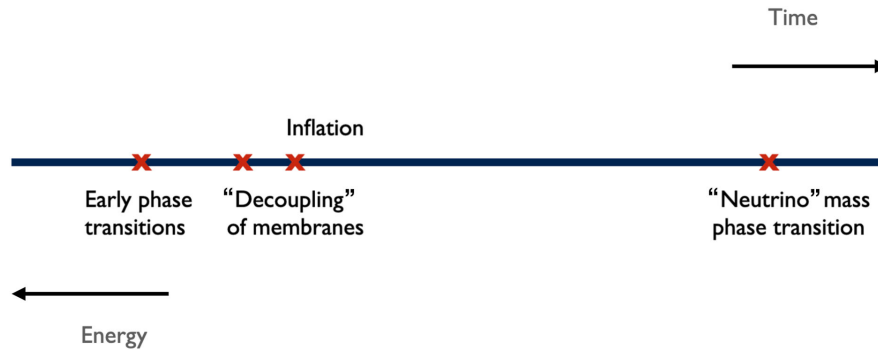
FIG. 4.   A toy model universe with the cosmological constant problem completely solved.

from a practical point of view, we need to worry about a framework where the early evolution sets Λ_{total} to almost zero, and then at a much later stage a phase transition in the matter sector occurs, changing the QFT vacuum energy dramatically, and by a large value. A region of the universe where this occurs would typically develop a large negative energy density, which would force it to stop expanding and recollapse, presumably at least forming a large black hole. Note, that we can solve this problem immediately by going back to anthropics to fix the very late value of the cosmological constant [66].

On the other hand, this also points towards a toy model universe where the cosmological constant could be completely solved using our mechanism, although it is not our Universe. Imagine a universe where all but one matter sector phase transitions occur very early at very high energy scales that are above the critical energy density controlling membrane decoupling and the scale of inflation. Quantum dynamics would favor vacua with nearly vanishing energy density, and after membranes stop being nucleated these regions could inflate in slow roll regime, reheat, and undergo radiation and matter evolution. If then a new phase transition occurs at a very low scale, it would typically induce a net negative cosmological constant. However if the theory also contains an ultralight field (an "axion"), which develops a potential, its potential energy due to misalignment could compensate the negative vacuum energy and make the expansion rate of this region start to accelerate while the axion is away from its minimum (see Fig. 4) [67]. This looks like an approximation to our Universe, except for the fact that it does not contain electroweak and QCD phase transition.

The discussion above however points to an important difference in how the real cosmological constant problem manifests in our framework, and also other frameworks that strive to address it using dynamical decay—i.e., adjustment. Some of these points have been already raised in [65]. Here we will revisit them and elaborate some aspects further. First, note that the oft-encountered phrasing of the cosmological constant problem in terms of the UV sensitivity of Λ_{total} is in fact misleading: the fact that the cosmological constant might get a large contribution

from a very heavy field is only an issue if such a field exists. On the other hand, there are many fields in our Universe whose mass scales are much greater than the observed scale of vacuum energy. What this shows is that the cosmological constant problem really is not a UV problem ("why is it not huge?") or an IR problem ("why is it tiny but not zero?") *per se*—it is really an all scales problem, or at least, all scales where we find massive fields (which means at least most of the stuff coming from the Standard Model, and possibly from dark matter sectors). In [65], Polchinski argued compellingly that to get all those contributions to conspire to cancel each other down to $10^{-120}M_{\text{Pl}}^4$ in local QFT seems well nigh impossible.

Despite some brave forays in the literature exploring alternatives to local QFT, we share the view of [65], which adheres to more conventional physics, whereby the cosmological constant decays via a de Sitter instability is a more palatable avenue. In this approach, regardless of the diversity of contributions from various scales, since they all add to the cosmological constant in the same way thanks to equivalence principle, they all get screened by bubble nucleation by which de Sitter decays.

VII. COSMIC CONNECTIONS

Our mechanism is a cosmological constant adjustment mechanism, albeit not one mediated by a smooth field, but by a discretely varying flux of a set of 4-forms. It shares some features with the adjustment mechanism proposed by Abbott in a very insightful paper [68]. Abbott proposed a field-theoretic adjustment mechanism using a scalar with a linear potential which is degenerate with the cosmological constant so that the scalar was screening away the cosmological term. Near the zero value of the cosmological constant, Abbott proposed that strong-coupling nonperturbative corrections arise, like for an axion. They catch the scalar and stop further rolling (see Fig. 5). However since the adjustment is continuous, the scalar must dominate the Universe at least until the cosmological constant were nearly zero. In such a Universe, inflation does not end until the Hubble parameter decreases to $\simeq 10^{-34}$ eV, and so the
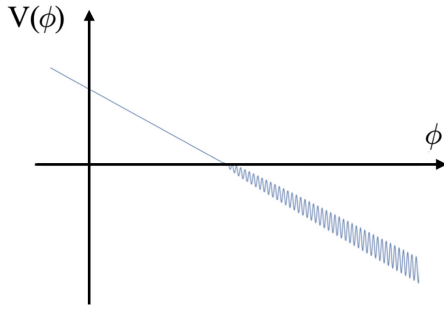
FIG. 5. A schematic of Abbott's adjustment mechanism.

universe never reheats.[16] In contrast, in our case this empty universe problem is averted since the relaxation of Λ involves jumps by large charges $\mathcal{Q}_j$: the cosmological constant jumps by a large step $\propto M_{\rm Pl}^2 \mathcal{Q}_j$. The tiny terminal Λ arises from the misalignment of the large fluxes sourced by charges. As a result, the cosmological constant does not always dominate, but just early on [4–6,36]. Before the universe jumps to tiny Λ, it had a large cosmological constant, and so immediately after its decay it can inflate and reheat using a standard slow roll stage [69]. The discrete adjustment also bypasses Weinberg's no-go theorem [13].

In striking contrast to what we would expect if an adjustment were based on a smooth variation of a field, the evolution does not completely wipe out the future of the information about the ancestry of the final near-Minkowski space. If the adjustment were via a smooth field, the universe everywhere would transition to the final near Minkowski at roughly the same time (as Abbott explained that would have to be about now). Instead, discrete adjustment happens locally. In each terminal bubble the interior geometry will preferentially be nearly flat. However those regions are separated from each other by highly curved spacetime, with large cosmological constant values, which are still bubbling.

This image of global structure of spacetime is just like the multiverse of eternal inflation (for reviews, see, e.g., [70,71]). After a moment's thought, this is not surprising in the least. In a setup where the net cosmological constant includes a contribution from a background flux, and where the 4-form gauge fields are sourced by charged tensional membranes, the discharge processes analogous to Schwinger particle production are inevitable [2,3]. Since they change the cosmological constant locally, inside a membrane, given some "initial" parent geometry (by this we mean any patch without bubbles at an arbitrarily selected moment of observation) it will start to convert into the bubbles of other vacua. However, since the parent is de Sitter, the descendant bubbles will not percolate, just

---

[16]This also happens in the adjustment mechanism employing a single 4-form when the value of charge $\mathcal{Q}_{\mathcal{A}}$ is selected to be tiny, as noted in [2,3].

like in the original "old inflation" of Guth [52]. In most of the bubbles, the cosmological constant will be smaller, but in some it will be (much) larger than the parent's [72,73]. In those regions inflation will restart itself. In our case, since we look at models with two systems of charged membranes with mutually irrational fluxes, there will be a huge number of different vacua, since the values of the cosmological constant span a discretuum like in [36], instead of only a few types usually encountered in simple models of false vacuum inflation.

Amusingly, our adjustment mechanism could address a critique directed at inflation in [74]. The authors of this work open up with the assertion that "inflationary theory has for some time had two skeletons in its cupboard." They elaborate that one problem is the selection of the initial conditions for the inflaton sector, required to start inflation, and the other is the cosmological constant, which they argue is troublesome since inflation happens because it is supported by cosmological constant which in turn must be canceled. The authors infer that both of "these issues point to severe fine-tunings". In our picture, both of these problems could be addressed with the help of eternal inflation.

Indeed, as long as there is even a single region of the universe with $\Lambda > 0$, it can happen that an empty universe with a nearly vanishing Λ "restarts" itself by a rare quantum jump which increases the cosmological constant [75,76]. In subsequent evolution back to $\Lambda \to 0$, the repeated process of membrane discharges reducing Λ back toward zero can scan for an inflationary stage [76]. This suggests that the natural "ground state" of the theory is near Minkowski, which is a dynamical attractor of the evolution, similar to the proposal of [77,78] (motivated by the ideas about reformulating inflationary initial conditions selection). If we follow this path, inflation seems *a priori* rare. However repeated "recycles" allow for filling up the phase space, and so even a "rare" event will be found eventually. As Guth points in [79], "once inflation starts, it generically continues forever, creating an infinite number of 'pocket' universes." The only difference here is that this eternal self-reproduction does not require a scalar field and its large quantum fluctuations. The large discrete variations of the cosmological constant come from the quantized flux which sources vacuum energy.

Guth also stresses the need for the beginning of inflation, due to the past incompleteness theorem [80]. However, as [70,78] note, it is possible that some quantum creation event may have initiated the whole process. We will set this issue aside for now, and only note that it should be possible to embed a standard late cosmology in the present framework.

Finally, one can consider specific predictions and implications for observations [81], among which might be a past record of colliding with other bubbleworlds [82–84], applications to particle physics hierarchies [85,86], and maybe

even late time variations of cosmological parameters, as in, e.g., [87,88]. These questions warrant further scrutiny.

VIII. SUMMARY

An established view of general relativity is that it is a theory of gravity with fixed values of G_N and Λ. This description is incredibly successful for explaining a huge host of observed phenomena, and passes a number of tests and bounds with flying colors. However from the theoretical point of view, this description runs into a quandary when we ask questions about the reasons why the values of G_N and Λ are what they are.

Here, and also earlier, in [4–6], we have considered an approach that "liberates" these gravitational parameters and allows them to be treated as global gauge invariant degrees of freedom. Focusing on the cosmological constant, this leads to a generalized unimodular gravity theory, which is locally completely indistinguishable from GR, classically and as a quantum field theory [21–23]. We showed that this theory is a theory of 4-forms in disguise, and to make sure that the gauge symmetry is properly encoded, we added a system of charged tensional membranes for each form present. It then follows that cosmological constant is unstable to quantum-mechanical, nonperturbative, discharge of membranes. This is because the 4-form fluxes are degenerate with the cosmological

constant by covariance and construction. Cosmological constant decays toward $\Lambda/M_{\rm Pl}^4 \to 0$, which is a dynamical attractor when the theory is dominated by terms linear in 4-form flux, at least in the statistical sense, once charge to tension ratio is fixed to (41). This is an avatar of Coleman and De Luccia's "gravitational stabilization" of flat space to nonperturbative instabilities. This addresses a part of the cosmological constant problem in our Universe: it shows why the cosmological constant is not huge without using anthropics.

The postinflationary phase transitions in the matter sector—the electroweak and the QCD ones—obscure the full solution at the moment. It would be interesting to explore this issue further. It is also of interest to determine precisely how to connect our mechanism with slow roll inflation, and how to accommodate current observed cosmic acceleration (by a very late phase transition?). We plan to return to these questions elsewhere.

ACKNOWLEDGMENTS

[1] M. Henneaux and C. Teitelboim, Phys. Lett. B **222**, 195 (1989).

[2] J. D. Brown and C. Teitelboim, Phys. Lett. B **195**, 177 (1987).

[3] J. D. Brown and C. Teitelboim, Nucl. Phys. **B297**, 787 (1988).

[4] N. Kaloper, Phys. Rev. D **106**, 065009 (2022).

[5] N. Kaloper, Phys. Rev. D **106**, 044023 (2022).

[6] N. Kaloper and A. Westphal, Phys. Rev. D **106**, L101701 (2022).

[7] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1915**, 844 (1915).

[8] D. Hilbert, Gott. Nachr. **27**, 395 (1915).

[9] G. E. Volovik, arXiv:0705.0991.

[10] S. E. Rugh and H. Zinkernagel, Stud. Hist. Phil. Mod. Phys. **33**, 663 (2002).

[11] Y. B. Zeldovich, JETP Lett. **6**, 316 (1967); Sov. Phys. Usp. **11**, 381 (1968).

[12] F. Wilczek, Phys. Rep. **104**, 143 (1984).

[13] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).

[14] N. Kaloper and A. Padilla, Phys. Rev. D **90**, 084023 (2014).

[15] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1919**, 349 (1919).

[16] J. L. Anderson and D. Finkelstein, Am. J. Phys. **39**, 901 (1971).

[17] W. G. Unruh, Phys. Rev. D **40**, 1048 (1989).

[18] W. Buchmuller and N. Dragon, Phys. Lett. B **207**, 292 (1988).

[19] W. Buchmuller and N. Dragon, Phys. Lett. B **223**, 313 (1989).

[20] Y. J. Ng and H. van Dam, J. Math. Phys. (N.Y.) **32**, 1337 (1991).

[21] B. Fiol and J. Garriga, J. Cosmol. Astropart. Phys. 08 (2010) 015.

[22] G. P. de Brito, O. Melichev, R. Percacci, and A. D. Pereira, J. High Energy Phys. 12 (2021) 090.

[23] T. Kugo, R. Nakayama, and N. Ohta, Phys. Rev. D **104**, 126021 (2021).

[24] N. Kaloper and A. Padilla, Phys. Rev. Lett. **112**, 091304 (2014).

[25] N. Kaloper, A. Padilla, D. Stefanyszyn, and G. Zahariade, Phys. Rev. Lett. **116**, 051302 (2016).

[26] A. Aurilia, H. Nicolai, and P. K. Townsend, Nucl. Phys. **B176**, 509 (1980).

[27] M. J. Duff and P. van Nieuwenhuizen, Phys. Lett. **94B**, 179 (1980).

[28] S. W. Hawking, The cosmological constant and the weak anthropic principle, Contribution to *Nuffield Workshop on Quantum Structure of Space and Time* (Cambridge University Press, Cambridge, 1982), pp. 423–432, Print-82-0177 (CAMBRIDGE).

[29] E. Baum, Phys. Lett. **133B,** 185 (1983).

[30] S. W. Hawking, Phys. Lett. **134B,** 403 (1984).

[31] M. J. Duff, Phys. Lett. B **226,** 36 (1989).

[32] M. J. Duncan and L. G. Jensen, Nucl. Phys. **B336,** 100 (1990).

[33] M. J. Duncan and L. G. Jensen, Nucl. Phys. **B361,** 695 (1991).

[34] A. Aurilia and E. Spallucci, Classical Quantum Gravity **10,** 1217 (1993).

[35] F. R. Klinkhamer and G. E. Volovik, Phys. Rev. D **77,** 085015 (2008).

[36] R. Bousso and J. Polchinski, J. High Energy Phys. 06 (2000) 006.

[37] J. L. Feng, J. March-Russell, S. Sethi, and F. Wilczek, Nucl. Phys. **B602,** 307 (2001).

[38] G. R. Dvali and A. Vilenkin, Phys. Rev. D **64,** 063509 (2001).

[39] G. Dvali and A. Vilenkin, Phys. Rev. D **70,** 063501 (2004).

[40] G. Dvali, arXiv:hep-th/0507215.

[41] N. Kaloper and L. Sorbo, Phys. Rev. D **79,** 043528 (2009).

[42] N. Kaloper and L. Sorbo, Phys. Rev. Lett. **102,** 121301 (2009).

[43] N. Kaloper, A. Lawrence, and L. Sorbo, J. Cosmol. Astropart. Phys. 03 (2011) 023.

[44] N. Kaloper, M. König, A. Lawrence, and J. H. C. Scargill, J. Cosmol. Astropart. Phys. 03 (2021) 024.

[45] E. I. Guendelman and A. B. Kaganovich, Phys. Rev. D **53,** 7020 (1996).

[46] F. Gronwald, U. Muench, A. Macias, and F. W. Hehl, Phys. Rev. D **58,** 084021 (1998).

[47] F. Wilczek, Phys. Rev. Lett. **80,** 4851 (1998).

[48] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, J. High Energy Phys. 02 (2015) 172.

[49] T. Banks, M. Dine, and N. Seiberg, Phys. Lett. B **273,** 105 (1991).

[50] T. Banks and N. Seiberg, Phys. Rev. D **83,** 084019 (2011).

[51] M. Reece, arXiv:2304.08512.

[52] A. H. Guth, Phys. Rev. D **23,** 347 (1981).

[53] S. R. Coleman and F. De Luccia, Phys. Rev. D **21,** 3305 (1980).

[54] S. R. Coleman, Phys. Rev. D **15,** 2929 (1977); **16,** 1248(E) (1977).

[55] C. G. Callan, Jr. and S. R. Coleman, Phys. Rev. D **16,** 1762 (1977).

[56] F. Englert, E. Gunzig, C. Truffin, and P. Windey, Phys. Lett. **57B,** 73 (1975).

[57] N. Arkani-Hamed, S. Dimopoulos, N. Kaloper, and R. Sundrum, Phys. Lett. B **480,** 193 (2000).

[58] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, J. High Energy Phys. 06 (2007) 060.

[59] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15,** 2752 (1977).

[60] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15,** 2738 (1977).

[61] C. Teitelboim, Phys. Lett. **167B,** 63 (1986).

[62] C. Teitelboim, Phys. Lett. **167B,** 69 (1986).

[63] I. Niven, *Numbers: Rational and Irrational* (Mathematical Association of America: New Mathematical Library, Washington DC, 1961).

[64] Y. Liu, A. Padilla, and F. G. Pedro, arXiv:2303.17723v1.

[65] J. Polchinski, arXiv:hep-th/0603249.

[66] (work in progress).

[67] R. Barbieri, L. J. Hall, S. J. Oliver, and A. Strumia, Phys. Lett. B **625,** 189 (2005).

[68] L. F. Abbott, Phys. Lett. **150B,** 427 (1985).

[69] (work in progress).

[70] A. H. Guth, Phys. Rep. **333,** 555 (2000).

[71] A. Linde, Rep. Prog. Phys. **80,** 022001 (2017).

[72] P. J. Steinhardt, in *Proceedings of Nuffield Workshop on Quantum Structure of Space and Time* (Cambridge University Press, Cambridge, 1983), pp. 423–432.

[73] A. D. Linde, Mod. Phys. Lett. A **01,** 81 (1986).

[74] N. Turok and S. W. Hawking, Phys. Lett. B **432,** 271 (1998).

[75] J. Garriga and A. Vilenkin, Phys. Rev. D **57,** 2230 (1998).

[76] J. Garriga and A. Vilenkin, Phys. Rev. D **64,** 023517 (2001).

[77] S. M. Carroll and J. Chen, arXiv:hep-th/0410270.

[78] S. M. Carroll and J. Chen, Gen. Relativ. Gravit. **37,** 1671 (2005).

[79] A. H. Guth, arXiv:astro-ph/0404546.

[80] A. Borde, A. H. Guth, and A. Vilenkin, Phys. Rev. Lett. **90,** 151301 (2003).

[81] B. Freivogel, M. Kleban, M. Rodriguez Martinez, and L. Susskind, J. High Energy Phys. 03 (2006) 039.

[82] A. Aguirre, M. C. Johnson, and A. Shomer, Phys. Rev. D **76,** 063509 (2007).

[83] S. Chang, M. Kleban, and T. S. Levi, J. Cosmol. Astropart. Phys. 04 (2008) 034.

[84] S. M. Feeney, M. C. Johnson, D. J. Mortlock, and H. V. Peiris, Phys. Rev. Lett. **107,** 071301 (2011).

[85] G. F. Giudice, A. Kehagias, and A. Riotto, J. High Energy Phys. 10 (2019) 199.

[86] N. Kaloper and A. Westphal, Phys. Lett. B **808,** 135616 (2020).

[87] P. H. Coleman and L. Pietronero, Phys. Rep. **213,** 311 (1992).

[88] V. Marra and L. Perivolaropoulos, Phys. Rev. D **104,** L021303 (2021).