Background independent field quantization with sequences of gravity-coupled approximants

Maximilian Becker and Martin Reuter

Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, D-55099 Mainz, Germany

We outline, test, and apply a new scheme for nonperturbative analyses of quantized field systems in contact with dynamical gravity. While gravity is treated classically in the present paper, the approach lends itself for a generalization to full quantum gravity. We advocate the point of view that quantum field theories should be regularized by sequences of quasiphysical systems comprising a well-defined number of the field's degrees of freedom. In dependence on this number, each system backreacts autonomously and self-consistently on the gravitational field. In this approach, the limit which removes the regularization automatically generates the physically correct spacetime geometry, i.e., the metric the quantum states of the field prefer to "live" in. We apply the scheme to a Gaussian scalar field on maximally symmetric spacetimes, thereby confronting it with the standard approaches. As an application, the results are used to elucidate the cosmological constant problem allegedly arising from the vacuum fluctuations of quantum matter fields. An explicit calculation shows that the problem disappears if the pertinent continuum limit is performed in the improved way advocated here. A further application concerns the thermodynamics of de Sitter space where the approach offers a natural interpretation of the microstates that are counted by the Bekenstein-Hawking entropy.

I. INTRODUCTION

A way of regarding a quantum field theory is as a special quantum system whose degrees of freedom are parametrized by the points of a smooth manifold. In typical applications the latter is the theoretical model of choice for representing space or spacetime. Nevertheless, when it comes to computing concrete predictions one usually discovers that such a continuum of densely packed degrees of freedom comprises, in a sense, too many of them to allow for a straightforward (or any) interpretation like in elementary quantum mechanics. The generic symptom of this overabundance are the ultraviolet divergences which relativistic quantum field theories are notorious for. As a way out, a rich arsenal of tools for their *regularization* and subsequent *renormalization* have been devised. In many cases the first one of the two logically independent steps, regularization, is considered devoid of an immediate physical interpretation, being not more than a technical trick to render certain intermediate steps of the calculation mathematically meaningful. A classic among the many regularization schemes that are available today is the momentum space cutoff. It still possesses a certain physical flavor unlike, say, dimensional regularization or the zeta function technique. Its key ingredient is a regularization parameter which has the dimension of a mass typically, and which defines a *scale* therefore: Modes of the field having momenta below this mass scale are retained by the cutoff,

while the others are discarded. The point to be noticed here is that the specification of a cutoff scale in proper momentum units requires a metric on the base manifold. Clearly this is no cause for concern in the familiar quantum field theories formulated on a "prefabricated" and unchangeable Minkowski space. It is a concern, however, in quantum gravity: when the metric itself is among the fields to be quantized one easily runs into severe difficulties or unresolvable paradoxes if one tries to work with a regularization which is metric dependent in itself [1].

(1) As we are going to demonstrate in the present paper, similar remarks apply already one step before full-fledged quantum gravity, namely within the framework of quantum field theory in curved spacetimes, if one includes the backreaction of quantum matter fields on the classical metric. More generally, the purpose of this paper is as follows. We are going to investigate quantized matter fields on classical spacetime geometries whose metric is determined self-consistently by a "semiclassical" Einstein equation that involves a certain effective stress tensor due to the matter degrees of freedom. To tackle this problem, we propose an approach which goes beyond earlier investigations in that it respects three basic principles which we shall introduce, explain and motivate in detail. To a large extent those principles grew out of various general lessons that were learnt within full quantum gravity, but have a bearing on semiclassical gravity also [2,3]. They arose both in approaches to quantum gravity that build upon discrete

125001-1

structures at the fundamental level such as loop quantum gravity [3–5] or causal dynamical triangulations [6], as well as continuum approaches like asymptotic safety [1,7]. The approach we are advocating, while continuum-based, could in principle detect discreteness at the physical level, should it emerge in some theory. Furthermore, in a companion paper [8] we extend the framework presented here by also quantizing the gravitational field itself.

(2) The first and foremost among the three requirements is *background independence*. The gravitational interaction has the unique property of also being in charge of furnishing the stage all physics takes place on, namely the spacetime manifold. Background independence requires that the corresponding "furniture," the metric in particular, is obtained as the solution of some fundamental dynamical law rather than through an *ad hoc* selection "by hand" [9–11]. The second principle is an extension of background independence into the realm of the regularized theories, i.e., of the "approximants" we deal with as long as the regulator has not yet been removed. Contrary to the examples mentioned above, we insist that the regularization should yield the approximants which are, or come close to being, realizable physical systems in their own right. More precisely, we require two properties: First, those systems possess a well-defined number of degrees of freedom, and second, they are coupled to gravity and thereby respect background independence in the sense that their respective approximation of the spacetime metric emerges dynamically from a self-consistency condition. The third principle finally is of a more technical nature and describes how to set up the approximant systems pertaining to a given field theory. The requirement is to employ what we call *N-type cutoffs*. By definition, they amount to regularization schemes which do not involve the metric. Thus detaching cutoffs from scales allows us to take limits of the approximants that could not be considered within the standard approaches.

In Sec. II of this paper we describe these requirements and the new framework in more detail. In the subsequent sections we shall then present a first application, which is both instructive in its own right, and can shed new light on a particularly puzzling aspect of the cosmological constant problem [12–15], namely the gravitational field generated by the zero point oscillations of quantum fields. W. Pauli is credited for the first estimate of the influence quantum vacuum fluctuations should have on the curvature of spacetime [14,16]. Considering a free massless field on Minkowski space, with dispersion relation $\omega(\mathbf{p}) = |\mathbf{p}|$, he argued that the field is equivalent to a set of harmonic oscillators, and each of them should contribute its zero point energy density $\frac{1}{2}\omega$ to the energy of the vacuum state. Summing them up leads to an amount

$$\varrho_{\text{vac}} = \frac{1}{2} \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} |\mathbf{p}| \tag{1.1}$$

which is quartically ultraviolet (UV) divergent and needs regularization. Installing a momentum cutoff $|\mathbf{p}| \leq \mathscr{P}$, the result is $\varrho_{\text{vac}} = c\mathscr{P}^4$ with c a positive constant of order unity.[1] The argument then continues by giving a numerical value to $\mathscr{P}$, typically taken to be the energy scale up to which the matter field theory under consideration is believed to be valid. Only at this stage gravity comes into play. It is argued that like any other form of energy, ϱ_{vac} should contribute to the curvature of spacetime, and that this effect can be taken care of by adding the contribution $\Delta\Lambda = (8\pi G)\varrho_{\text{vac}} = 8\pi cG\mathscr{P}^4$ to the cosmological constant in Einstein's equation.

Here, then, comes the big disappointment. For every plausible scale $\mathscr{P}$, the curvature produced by $\Delta\Lambda$ is by far too large to be consistent with observation. Pauli himself identified $\mathscr{P}$ with typical energies in atomic physics; he had to conclude that the resulting curvature is so tremendous that, if correct, the Universe "would not even reach out to the moon" [14,16]. If we choose the Planck scale instead, $\mathscr{P} = m_{\text{Pl}}$, the calculation produces a curvature which is about 10^{120} times larger than the value from modern-day cosmological observations. According to a variant of this reasoning, Einstein's equation contains, besides $\Delta\Lambda$, also a bare cosmological constant, Λ_{b}, whose value is then tuned in dependence on $\mathscr{P}$ in such a way that the sum $\Lambda_{\text{obs}} = \Lambda_{\text{b}}(\mathscr{P}) + \Delta\Lambda(\mathscr{P})$ equals precisely the observed value. This version of the argument avoids making a false prediction (any prediction, in fact), but at the expense of an enormous naturalness problem. To achieve the desired value of Λ_{obs}, the bare quantity $\Lambda_{\text{b}}(\mathscr{P})$ must be fine-tuned with a precision of 120 digits, say.

One of the purposes of the present paper is to pinpoint precisely why the above reasoning must lead to a wrong answer. As it turns out employing the new and more powerful scheme no comparable "cosmological constant problem" arises.

(4) The rest of this paper is organized as follows. In Sec. II we outline our framework for an improved non-perturbative analysis of quantum fields in contact with dynamical gravity; in particular we explain and motivate the three basic requirements the corresponding calculations must meet. In the subsequent sections we apply those rules to a Gaussian scalar field self-consistently interacting with gravity: In Sec. III we derive a first type of approximant, in Sec. IV we explore the sequences of self-gravitating systems it gives rise to, and in Sec. V we analyze them from a path integral perspective, thereby also discovering a second type of natural approximants. While the primary application of our results is to the cosmological constant issue, Sec. VI is devoted to a brief discussion of the Bekenstein-Hawking entropy of de Sitter space, and the natural interpretation of its microstates we are led to.

---

[1] Its precise value depends on inessential implementational details.

The final Sec. VII contains a short summary and the conclusions.

II. THE FRAMEWORK: OUTLINE AND MOTIVATION

In this paper we advocate a scheme for the quantization of matter fields, and subsectors thereof, which are coupled to classical gravity.[2] This scheme satisfies three essential requirements; in the present section, we are going to explain and motivate them, and then in the rest of the paper we implement them in various sample calculations.

The three requirements are

> (**R1**) Background independence
>
> (**R2**) Gravity-coupled approximants
>
> (**R3**) N-type cutoffs

The requirements are not independent logically. In particular *(R2)* may be seen as an extension of *(R1)*, while *(R3)* is a tool for dealing with *(R2)*. We discuss them in turn now, focusing on the main aspects, and leaving aside inessential technical details or difficulties.

A. First requirement: Background independence

The desideratum of background independence is presumably the most powerful and far reaching concept that has been taken over from classical general relativity and integrated into the modern approaches to quantum gravity [3,17–19]. Depending on how they cope with this challenge, the various approaches can be grouped into two classes: those which, literally, do not use background structures like a metric, and those which do employ such fields, but at a certain point fix them self-consistently, namely by invoking the fundamental dynamical laws [9,11]. In this paper, we develop a continuum-based approach which follows the second strategy. It enforces background independence indirectly by invoking the background field technique in its general form [1,20].

B. Second requirement: Gravity-coupled approximants

The idea is to replace the notion of *regularization* by sequences of certain "quasiphysical" auxiliary systems describing matter; they are comprised of a well-defined, finite set of quantum degrees of freedom which couple to gravity. Referring to such systems as *approximants*, we denote them symbolically by App. The total configuration of an approximant is characterized by a quantum mechanical state $\Psi_\mathbf{f}$ of the matter system, having $\mathbf{f} < \infty$ degrees of freedom, together with a classical metric. Symbolically,

$\mathsf{App}(\mathbf{f}) \sim \Psi_\mathbf{f} \otimes$ metric. Approximants complying with the requirement *(R2)* are special in that their metric is fixed dynamically rather than by fiat. It arises as the gravitational response to the energy and momentum of the $\mathbf{f}$ matter degrees of freedom. They are allowed to backreact self-consistently on the geometry of the spacetime which they inhabit. Thus, by virtue of *(R2)*, background independence is manifest already at the regularized level. Regularized quantum field theories are represented by sequences of approximants, $\mathsf{App}(\mathbf{f})$, $\mathbf{f} \to \infty$. The removal of the regulator, corresponding in the standard case to, say, sending a lattice constant to zero, amounts to following a particular sequence for increasing $\mathbf{f}$. If the sequence has a limit, in an appropriate sense, we identify this limit with the quantum field theory (QFT) to be constructed, with the field in a particular state. Importantly, by this construction the (state of a) QFT arises always in combination with a self-consistently determined metric. Symbolically,

$$\mathsf{App}(\mathbf{f}) \overset{\mathbf{f} \to \infty}{\longrightarrow} \Psi_{\mathrm{QFT}} \otimes \text{self-consistent metric.} \quad (2.1)$$

Thus, loosely speaking, possible states of the QFT for the matter sector are already "born" in that particular spacetime which they like to live in. More precisely, the self-consistent metric in (2.1) is a solution to the semiclassical Einstein equation with an appropriate energy-momentum tensor $T_{\mu\nu}[\Psi_{\mathrm{QFT}}]$ on its right-hand side (rhs).

(A) Motivations for insisting on the requirement *(R2)* include

(A1) We want the approximants, in the best case, to constitute physically realizable systems in their own right, or at least come close to this ideal.[3] We expect that this enhances our chances to find sequences which converge to physically interesting limits. We believe that this property is particularly important if one is forced to resort to approximate calculations of some kind, as it is always the case in practice.

(A2) As for the assumption of *self-consistently* gravitating approximants subject to classical general relativity, this is a conceptually natural requirement if one regards matter QFTs in curved spacetime as an approximation to full-fledged quantum gravity that would additionally involve a quantized metric. In the present paper, our framework for matter QFTs is set up in such a way that it will generalize straightforwardly to full quantum gravity. The only difference is that for the time being quantum fluctuations of the metric, relative to the self-consistently adjusting background, are neglected.[4] Hence, for now, the metric is

---

[2]In a companion paper [8] we generalize the scheme by including a quantized gravitational field.

[3]As for being "physically realizable" or "quasiphysical" we are very liberal. What is important here is only that the drawbacks described below in connection with the counterexamples are avoided.

[4]See however [8] for the inclusion of such metric fluctuations.

obtained from the classical Einstein equation, with a quantum mechanical $T_{\mu\nu}[\Psi_{\mathbf{f}}]$, though.[5]

(A3) There is a further motivation for physical gravity-coupled approximants which goes beyond using the sequences of approximants merely as a tool for regularizing a QFT. Namely, our framework is open towards the possibility that experiment tells us that nature is actually better described by the approximant $\mathsf{App}(\mathbf{f}_{\mathrm{obs}})$ for some finite, observationally determined $\mathbf{f}_{\mathrm{obs}}$, rather than by $\mathsf{App}(\mathbf{f})$ in the limit $\mathbf{f} \to \infty$. In this hypothetical case, observational facts would suggest to abandon the original plan of removing the regulator fully. The quantization of a classical field would then result in a quantum system with, rigorously, only finitely many degrees of freedom. While this might sound like a quite exotic possibility, there is a natural and simple scenario which realizes it: Assume that the quantum gravitational dynamics is such that spacetime acquires physical, i.e., observable discreteness properties at some microscopic scale. It is a question of general significance then whether, and possibly how, this effect can be discovered by a continuum-based theoretical framework.[6] One possibility is the scenario above: If spacetime is roughly similar to a certain discrete point set with matter fields on it, it will, loosely speaking, have the appearance of an "incompletely quantized" classical field. Later on we shall encounter an explicit example where this is indeed what happens.

(B) Counter examples are perhaps the best way to characterize, and to further motivate, approximants that comply with *(R2)*:

(B1) Many regularization schemes that we employ routinely because they are convenient technically fail to generate quasiphysical approximants in the sense of *(R2)*. A typical example is *dimensional regularization*. Clearly it is impossible to interpret a regularized theory at a generic value of $\varepsilon \equiv 4 - d$ as a quantum system with a defined number $\mathbf{f}$ of degrees of freedom which an experimentalist could build. Similar remarks apply to other schemes based upon analytical continuation such as *zeta function regularization* [21,22] which, too, is unacceptable by the requirement *(R2)*.

(B2) Furthermore, *(R2)* rejects all schemes which are designed so as to make certain (usually power-law) divergences invisible, with the justification that they anyhow would be absorbed into bare parameters whose values do not matter for all practical purposes. Zeta function regularization is an example again. It is notorious for "identities" like

$$\sum_{n=1}^{\infty} n \overset{\text{zeta fct.}}{\equiv} -\frac{1}{12} \qquad (2.2)$$

which exhibit a finite part on top of a numerically much more important divergent one—which it suppresses however. By way of comparison, a scheme that does comply with *(R2)* is a cutoff regularization that would deal with the divergent sum $\sum_n n$ simply by stating that

$$\sum_{n=1}^{N} n = \frac{1}{2} N^2 \left\{ 1 + O\left(\frac{1}{N}\right) \right\}, \qquad (2.3)$$

without expressing any prejudice about the ultimate fate of the divergence which arises when $N \to \infty$ at a later stage. As a rule, acceptable regularization schemes must treat bare parameters as *potentially physical*. Hence contributions like the N^2-term above must be retained, and their physical impact along the sequence of approximants must be taken into account carefully.

(B3) Finally we turn to the perhaps most important issue, the question of why gravity should be included into the physical description of the approximants. Figure 1 contrasts field quantization via the sequences $\{\mathsf{App}(\mathbf{f})\}$ with the standard approach. The horizontal arrows of this highly schematic diagram represent taking the limit $\mathbf{f} \to \infty$, i.e., the transition from a finite system to a QFT, while the vertical arrows symbolize the inclusion of the gravitational backreaction into the matter system by solving a semiclassical Einstein equation. The upper left box in Fig. 1 stands for a generic regularized precursor of the QFT, which does not in general comply with *(R2)*. It consists of quasiphysical systems in states $\Psi_{\mathbf{f}} \otimes g_{\mu\nu}$, with $\mathbf{f} < \infty$, but the systems live in a spacetime with an arbitrary fixed metric $g_{\mu\nu}$, which is unrelated to $\mathbf{f}$ and the quantum states $\Psi_{\mathbf{f}}$ *a priori*. Now there exist two paths in the diagram which one can take in order to remove the regulator. Symbolically speaking they are, respectively,

$$(\mathbf{i}) \ \text{first down, then horizontal} \qquad (2.4)$$

$$(\mathbf{ii}) \ \text{first horizontal, then down.} \qquad (2.5)$$

The path (i) is the one advocated in the present paper: For each approximant separately we solve the respective Einstein equation. It contains $T_{\mu\nu}[\Psi_{\mathbf{f}}]$ which is self-consistently coupled to the Schrödinger equation for $\Psi_{\mathbf{f}}$; the latter involves the metric which is a solution to the former. In this manner we arrive at a sequence of configurations describing the total system, $\Psi_{\mathbf{f}}^{\mathrm{SC}} \otimes (g_{\mathbf{f}}^{\mathrm{SC}})_{\mu\nu}$, $\mathbf{f} = 0, 1, 2, \ldots$. Thereafter, as the second step, we let $\mathbf{f} \to \infty$ and (ideally) are led to a limiting state $\Psi_{\infty}^{\mathrm{SC}} \otimes (g_{\infty}^{\mathrm{SC}})_{\mu\nu}$ which can be interpreted as pertaining to a full-fledged QFT living on a spacetime with metric $(g_{\infty}^{\mathrm{SC}})_{\mu\nu}$. It is selected dynamically by self-consistency, in accord with background independence.

The path (ii) is the familiar one; it does not respect *(R2)* and the price one pays for that is the fatal cosmological constant problem in calculations like Pauli's. At first one
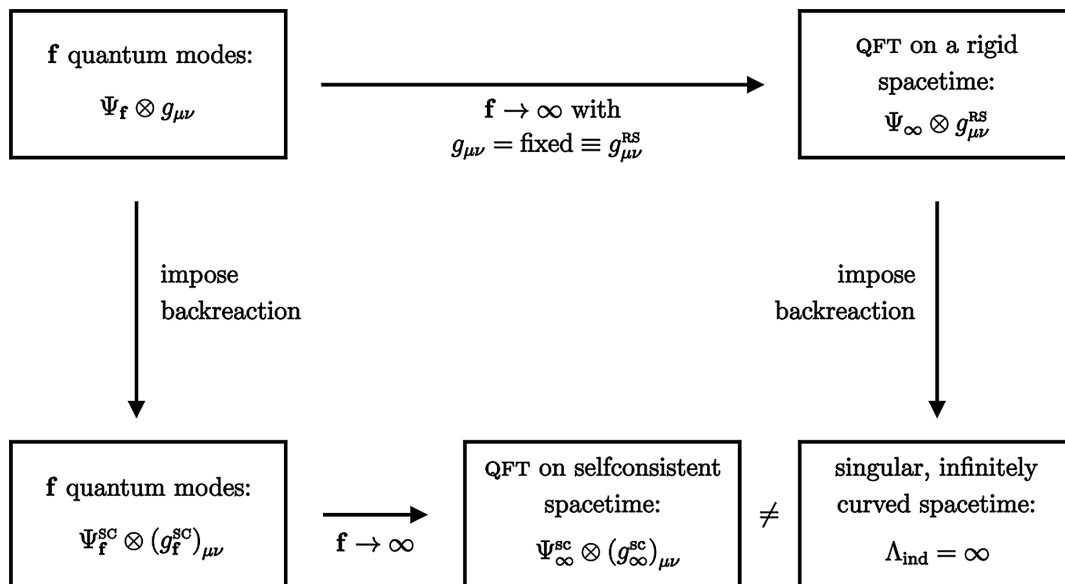
_____

[5]Often (but somewhat confusingly) called the "semiclassical Einstein equation".
[6]See Sec. 1.5 of [1] for a discussion of this point.

FIG. 1. Inclusion of the gravitational backreaction does not commute with the limit $\mathbf{f} \to \infty$.

takes the field theory limit leading to infinitely many degrees of freedom, and thereby one does not yet worry about gravity at all. The construction of the QFT is performed on a rigid spacetime (RS) whose metric $g_{\mu\nu}^{\mathrm{RS}}$ is chosen in a completely arbitrary way, often highly symmetric, typically flat, so as to ease the calculational difficulties. In this QFT, one then computes the vacuum energy density, or induced cosmological constant Λ_{ind}, and finds that it is formally infinite in absence of a cutoff.[7]

Only now, in the second stage one asks about the backreaction of the vacuum energy on the metric. The answer is very simple then: If $\Lambda_{\mathrm{ind}} = \infty$, the curvature is infinite and spacetime is degenerate. And even in an effective field theory setting, the predicted curvature is too large, by factors like 10^{120} or so. The lower right box in Fig. 1 indicates this singular, or at least phenomenologically unacceptable, state of the combined QFT-gravity system.[8]

The main goal of the present paper is to demonstrate by various sample calculations that the state obtained along track *(i)*, $\Psi_{\infty}^{\mathrm{SC}} \otimes (g_{\infty}^{\mathrm{SC}})_{\mu\nu}$, can be extremely different from the singular one of *(ii)*. In fact, choosing track *(i)* the kind of cosmological constant problem one encounters in the Pauli-type approaches does not occur. In essence, the requirement *(R2)* wants us to perform any QFT calculation in curved spacetime by proceeding in analogy with track *(i)* of this

example. This guarantees that background independence is respected not only by the final theory, but already at the level of its approximants. As a result, essential dynamical effects come within reach of practically feasible calculations.

C. Third requirement: *N*-type cutoffs

Given a classical field theory, the question is how to actually manufacture sequences of approximants by means of which we can hope to find interesting limits that would qualify as a QFT coupled to gravity. We propose to define the approximants by imposing a cutoff which is of "*N*-type," a concept which we outline next. It may be seen as a generalization of the "finite-mode regularization" that had been used before in a nongravitational context [23–25]. The expression *N*-type cutoff derives from the fact that in typical examples the regularization parameter is a positive integer, $N \in \mathbb{N}$, but other cases will occur as well. Using an *N*-type cutoff, the number of degrees of freedom $\mathbf{f} \equiv \mathbf{f}(N)$ becomes a function of the cutoff parameter N. In the case $N \in \mathbb{N}$ regularized QFTs are thus represented by sequences $\{\mathsf{App}(N), N = 0, 1, 2, \ldots\}$.

1. N cutoffs: Definition

Let us consider the general problem of giving a meaning to a formal functional integral of the type

$$Z = \int_{\mathcal{F}} \mathcal{D}(C) \mathrm{e}^{-S[C]}. \qquad (2.6)$$

The integration is over fields $C(x)$ that "live" on some differentiable manifold, $\mathcal{M}$, and belong to a certain space of functions, $\mathcal{F}$. We assume that the space $\mathcal{F}$ is the span of the basis

---

[7]Or it is finite, but still way too large if the QFT is regarded an *effective* one which is valid below a certain plausible UV scale only.

[8]Unless stated otherwise, bare parameters are always kept fixed when $N \to \infty$ in this paper. If one allows them to depend on N, the ($\Lambda_{\mathrm{ind}} = \infty$) problem can be traded for the naturalness problem of an infinite fine-tuning.

$$\mathscr{B} = \{w_\alpha(\cdot)|\alpha \in I\} \qquad (2.7)$$

where I is an appropriate index set. So we can expand the integration variable,

$$C(x) = \sum_{\alpha \in I} c_\alpha w_\alpha(x), \qquad (2.8)$$

and reexpress the measure $\mathcal{D}(C)$ in terms of the expansion coefficients c_α. In typical examples the result will be proportional to $\prod_{\alpha \in I} dc_\alpha$ then.

Nothing has been gained so far; (2.6) is still an ill-defined combination of infinitely many integrals. To regularize it, we introduce a one parameter family of subsets $\mathscr{B}_N \subset \mathscr{B}$ labeled by a dimensionless number $N \in \mathbb{N}$ (or $N \in [0, \infty) \equiv \mathbb{R}^+$).[9] This family of subsets is required to satisfy

$$\mathscr{B}_0 = \varnothing, \quad \mathscr{B}_\infty = \mathscr{B}, \quad \text{and} \quad N_2 > N_1 \Rightarrow \mathscr{B}_{N_2} \supset \mathscr{B}_{N_1}. \qquad (2.9)$$

Thus, for increasing values of N, a continually growing portion of the basis elements in $\mathscr{B}$ gets included into the set $\mathscr{B}_N$. Eligible sequences of subsets

$$\mathscr{B}_0 \subset \mathscr{B}_1 \subset \mathscr{B}_2 \subset \cdots \subset \mathscr{B}_N \subset \ldots \subset \mathscr{B} \qquad (2.10)$$

should be described entirely in terms of the indices in the set I. For each $N \in \mathbb{N}$ we introduce a subset of indices $I_N \subset I$ such that

$$\mathscr{B}_N = \{w_\alpha(\cdot)|\alpha \in I_N\}. \qquad (2.11)$$

Then the specification of a sequence like (2.10) boils down to a long "bit string" of yes-no decisions: For any given element of $\mathscr{B}$, say w_β, which is uniquely identified by its label $\beta \in I$, we must specify whether $\beta \in I_N$ or $\beta \notin I_N$, for all $N = 0, 1, 2, \ldots, \infty$. We stress that no length or momentum scales are involved in the specification of such sequences $\{\mathscr{B}_N\}_{N=0,\ldots,\infty}$. The rationale behind the $\mathscr{B}_N$s is that integration over their linear span, denoted by $\mathcal{F}_N \subset \mathcal{F}$, should lead to a well-defined regularized precursor of the functional integral, viz.

$$Z_N \equiv \int_{\mathcal{F}_N} \mathcal{D}(C) e^{-S[C]}. \qquad (2.12)$$

In the case $N \in \mathbb{N}$, (2.12) involves only finitely many integrals. (If $N \in \mathbb{R}^+$ we assume that $\mathcal{F}_N$ is defined

"sufficiently small" to make (2.12) mathematically meaningful.) We regard Z_N as a partition function which describes an approximant with a sufficiently small, or even finite number of degrees of freedom, $\mathbf{f} \equiv \mathbf{f}(N)$. They are realized by $\mathbf{f}$ selected modes of a field on $\mathscr{M}$. Replacing the full space of functions $\mathcal{F}$ successively by the chain of subspaces $\{\mathcal{F}_N\}_{N=0,\ldots,\infty}$ is what we will, very broadly, refer to as a regularization by means of a cutoff of the N-type.

2. N cutoffs: Properties

The most important property which an N cutoff must possess (while many others do not) is the following: Assigning a particular value to N does not imply a momentum or length scale separating the modes retained in $\mathscr{B}_N$ from those discarded.

(1) The benefit which we get from the complicated-looking definition above is that N cutoffs can be formulated without the need of a metric. In fact, $\mathscr{B}$ and $\mathscr{B}_N$ are certain bases of functions on a given differentiable manifold $\mathscr{M}$. Hence technically speaking they belong to the manifold's differentiable structure, and to define the cutoff no metric structure is required on $\mathscr{M}$. Above we therefore insisted that the $\mathscr{B}_N$s are defined in terms of binary decisions operating on the index set I. This highlights the importance of adequately labeling the basis functions: The enumeration of the $\mathscr{B}$ elements, i.e., the map $I \to \mathcal{F}$, $\alpha \mapsto w_\alpha(\cdot)$, the selection of the $\mathscr{B}_N$ elements and their enumeration by $I_N \to \mathcal{F}$, $\alpha \mapsto w_\alpha(\cdot)$, must not involve a metric in any way.

(2) Before discussing practical incarnations of N cutoffs let us clarify another point. It is essential for our purposes to keep the steps of *regularization* and *renormalization* strictly separated. By definition, N-type cutoffs are really not more than a regularization, and they must not implement any renormalization conditions implicitly.[10] At least for standard field theories, and in absence of gravity, there is a well-known procedure that can be followed when one tries to define a continuum limit[11] of regularized path integrals like (2.12). Namely, while the regulator is removed, one tries to change the bare parameters (masses, couplings, etc.) which are implicit in $S[C]$ in precisely such a manner that the Z_Ns do indeed converge to some limit. If this is (im-)possible, the theory in question is called (non-) renormalizable [28]. Clearly the same can be done on the basis of the above sequences of approximants. However, as it turns out it is also of interest to take the approximants at finite N seriously in their own right, and to compare them as physical systems. They are in different states Ψ_N, but refer to the same values of the bare parameters.

---

[9]Our general description focuses on $N \in \mathbb{N}$ mostly since our later examples will belong to this case. The most important features of an N cutoff carry over to the case $N \in \mathbb{R}^+$ straightforwardly, however. Note also that as such field systems having only a finite number of modes, $N \in \mathbb{N}$, are nothing exotic, of course. This is the usual situation in condensed matter physics [26], or lattice field theory [27], for example.

[10]Recall the counterexample of the zeta function regularization which unavoidably fixes finite parts in its own specific way.

[11]We follow the common practice and refer to the limit where the cutoff is removed generally as the "continuum limit" also when the regularization is not by a lattice.

The examples worked out in the present paper will be of this second kind.

3. Eigenbases of metric dependent operators

Next we discuss a special type of N cutoff. While particularly convenient from the technical point of view, superficially it might seem that they are inconsistent with *(R3)*. It is therefore important to see that this is not the case actually. Let us now explicitly assume that $\mathcal{M} \equiv (\mathcal{M}, g_{\mu\nu})$ is a Riemannian manifold, and let us allow the path integral to depend manifestly on the metric $g_{\mu\nu}$ as a background:

$$Z[g] = \int \mathcal{D}(C; g) \mathrm{e}^{-S[C;g]}. \tag{2.13}$$

Furthermore, let $\mathcal{K} \equiv \mathcal{K}[g]$ be a self-adjoint positive operator depending on the metric, the prime example being the (negative of the) Laplace-Beltrami operator, $\mathcal{K} = -\Box_g \equiv -g^{\mu\nu}D_\mu D_\nu$. We would like the basis $\mathcal{B} \equiv \mathcal{B}[g] = \{w_\alpha[g](\cdot)|\alpha \in I\}$ to be an eigenbasis of $\mathcal{K}[g]$ now, implying that in general the basis functions $w_\alpha(\cdot) \equiv w_\alpha[g](\cdot)$ and their eigenvalues $\lambda_\alpha[g]$ will have a parametric dependence on $g_{\mu\nu}$:

$$\mathcal{K}[g]w_\alpha[g](x) = \lambda_\alpha[g]w_\alpha[g](x), \quad \alpha \in I. \tag{2.14}$$

From here on we proceed in the usual way in order to install an N cutoff in (2.13). We declare, for all N, which indices $\alpha \in I$ are in I_N, so that setting

$$\mathcal{B}_N[g] = \{w_\alpha[g](\cdot)|\alpha \in I_N\} \tag{2.15}$$

yields a sequence $\{\mathcal{B}_N\}$ which obeys the general rules (2.15), and we can define the approximants by

$$Z_N[g] \equiv \int\limits_{\mathrm{span}\,\mathcal{B}_N[g]} \mathcal{D}(C; g)\mathrm{e}^{-S[C;g]}. \tag{2.16}$$

It is important to understand why this regularization still amounts to an N cutoff in accordance with *(R3)*: While the *elements* of the set $\mathcal{B}_N[g]$ do indeed depend on the metric, the crucial property is that the index set I_N is metric independent. This is why we repeatedly emphasized that (in the discrete case) the sets I_N should be the result of nothing but "binary decisions" applied to the elements of I; as such they do not define a proper length or mass scale.

To summarize: An N cutoff, even when applied to a formal functional integral with an explicit metric dependence, is such that its specification in terms of $\{I_N\}$ does not require, and does not involve, a metric. A simple example illustrates this point. Let $(\mathcal{M}, g)$ be the round 2-sphere with radius r. Then $\mathcal{K} = -\Box_g$ has the well-known spectrum $l(l+1)/r^2, l = 0, 1, 2, \ldots$, which is linked to the metric via the value of r. Now we can specify an N cutoff by decreeing that, for example, $\mathcal{B}_N$ contains all spherical harmonics with $l \leq N$ and none having $l > N$, i.e., $\mathcal{B}_N = \{Y_{lm}|(l, m) \in I_N\}$ where $I_N = \{(l, m)|l = 0, 1, 2, \ldots, N; m = -l, \ldots, +l\}$. The $\mathcal{B}_N$s are completely fixed upon specifying the rule according to which index pairs from the full $I \equiv I_\infty$ are allocated to I_N, and this rule has nothing to do with the continuous parameter r, i.e., with the metric which we put on the sphere.

4. Continuous spectra

In order to demonstrate that N cutoffs are not restricted to discrete spectra, we briefly consider the example of a foliated cosmological spacetime equipped with the correspondingly adapted spatially flat Robertson-Walker metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{2.17}$$

Here coordinate differences $(\Delta x^i) \equiv \Delta\mathbf{x}$, while often referred to as *comoving distances*, are not proper distances. Products $a(t)\Delta\mathbf{x}$ instead are "proper" with respect to the metric (2.17). Let us take $\mathcal{K}$ to be the spatial part of the corresponding Laplace-Beltrami operator,

$$\mathcal{K} = -\frac{1}{a(t)^2}\delta^{ij}\partial_i\partial_j. \tag{2.18}$$

Its eigenfunctions are plane waves clearly. A subtlety arises however when it comes to labeling them, since later on N cutoffs are defined in terms of the pertinent index set.

From this perspective, a "good" labeling amounts to writing

$$w_{\mathbf{q}}(\mathbf{x}) = \exp(i\mathbf{q} \cdot \mathbf{x}), \quad \mathbf{q} \in \mathbb{R}^3, \tag{2.19}$$

in terms of the *coordinate* momentum $\mathbf{q}$; like $\mathbf{x} \equiv (x^1, x^2, x^3)$, it is dimensionless, not proper, and in fact unrelated to any metric. It qualifies as a continuous version of the generic (multi-)indices α by means of which the $\mathcal{B}_N$ elements are selected. Hence a perfectly legitimate N cutoff would, for example, be specified by

$$\mathcal{B}_N = \{w_{\mathbf{q}}(\cdot)|\mathbf{q} \in I_N\} \quad \text{with} \quad I_N = \{\mathbf{q} \in \mathbb{R}^3||\mathbf{q}| \leq N\}. \tag{2.20}$$

In this case $N \in \mathbb{R}^+$, so the sequence of approximants is labeled in a continuous fashion now. However, there are also frequently used ways of labeling the eigenfunctions which are "bad" from the perspective of *(R3)*. In fact, the eigenvalue of (2.19) is given by

$$\lambda_{\mathbf{q}} = \left(\frac{\mathbf{q}}{a(t)}\right)^2 \equiv \mathbf{p}^2. \tag{2.21}$$

This motivates using the proper momentum $\mathbf{p} \equiv \mathbf{q}/a(t)$ in order to distinguish the eigenmodes of $\mathcal{K}$, rewriting (2.19) as

$$W_{\mathbf{p}}(\mathbf{x}) = \exp(ia(t)\mathbf{p} \cdot \mathbf{x}), \quad \mathbf{p} \in \mathbb{R}^3. \qquad (2.22)$$

When working with the mode functions (2.22) it would seem natural to impose a cutoff condition on the proper momentum, like $|\mathbf{p}| \leq \mathscr{P}$, say. But as we discuss next, this would violate *(R3)*.

5. N cutoffs vs $\mathscr{P}$ cutoffs

We close the discussion of the N-type cutoffs by exhibiting a class of counterexamples which we collectively refer to as cutoffs of "$\mathscr{P}$-type." They fail to satisfy the requirement *(R3)* because of a "mistake" that can be pinned down quite precisely. Later on we shall then see that this mistake has a significant impact on the cosmological constant issue. We return to the spectral problem (2.14) and, in order to simplify the argument, assume that all eigenvalues are nondegenerate. Hence, for $g_{\mu\nu}$ fixed, there exists a one-to-one map $\alpha \mapsto \lambda_\alpha[g]$ which relates eigenvalues λ and labels α. It is a common practice to solve this relationship for the label, obtaining $\alpha = \alpha[g](\lambda)$, and to use the eigenvalues in order to enumerate the eigenfunctions. The basis writes then

$$\mathscr{B}[g] = \{W_\lambda[g](\cdot)|\lambda \in \mathrm{spec}(\mathscr{K})\} \qquad (2.23)$$

with the reparametrized mode functions

$$W_\lambda[g](x) \equiv w_\alpha[g](x)|_{\alpha=\alpha[g](\lambda)}. \qquad (2.24)$$

Now we come to the delicate point: Being presented with the basis in the form of (2.23) it is tempting to construct regularizations by applying selection criteria to the new label λ, in the same way as with α above. Of course, the first example that comes to mind is a sequence $\{\mathscr{B}_{\mathscr{P}}\}_{\mathscr{P}\in\mathbb{R}^+}$ obtained by restricting the eigenvalues to lie below a fixed scale $\mathscr{P}^2$:

$$\mathscr{B}_{\mathscr{P}} = \{W_\lambda(\cdot)|\lambda \leq \mathscr{P}^2\}. \qquad (2.25)$$

Obviously the familiar momentum space cutoff that is used abundantly in field theory on flat space is precisely of this sort, with $\mathscr{P} \equiv \Lambda$ in the traditional notation. The background metric, $g_{\mu\nu} = \delta_{\mu\nu}$ or $\eta_{\mu\nu}$ usually, is fixed once and for all in this case. Nonetheless, the subsets $\mathscr{B}_{\mathscr{P}}$ in (2.25) do not define a cutoff in accord with *(R3)*, i.e., no N cutoff. The reason is obvious: Due to the substitution $\alpha \to \alpha[g](\lambda)$ the *enumeration* of the basis functions has become explicitly metric dependent. As we explained above, an "adiabatic" $g_{\mu\nu}$ dependence of $w_\alpha[g](x)$ is perfectly acceptable—as long as the labeling by the αs does not involve the metric. The new mode functions $W_\lambda[g](x)$ spoil this property. As a result, regularizations like (2.23) which are defined in terms of their "index" λ are not N-type cutoffs. Note also that λ and $\mathscr{P}$, unlike α and N, are not dimensionless: They have canonical mass dimensions $[\lambda] = 2$ and $[\mathscr{P}] = 1$, respectively.

An example of such a forbidden metric-related labeling is (2.22) in the cosmological example above. The proper momentum $\mathbf{p}$ has a magnitude determined by the eigenvalue, $|\mathbf{p}|^2 = \lambda$, and a direction specified by a unit vector $\mathbf{n} \equiv \mathbf{p}/|\mathbf{p}|$, which serves as a dimensionless degeneracy index here. Hence (2.22) is equivalent to writing

$$W_{\lambda,\mathbf{n}}(\mathbf{x}) = \exp(ia(t)\sqrt{\lambda}\mathbf{n} \cdot \mathbf{x}). \qquad (2.26)$$

This simple example also makes it clear that our argument generalizes trivially to spectra with degenerate eigenvalues if λ is combined with appropriate degeneracy indices. In the following we shall refer to regularizations of the form (2.25) collectively as cutoffs of the $\mathscr{P}$-type.

III. A FIRST TYPE OF APPROXIMANTS

In the rest of this paper we perform an explicit investigation which implements all three requirements *(R1)*, *(R2)* and *(R3)* simultaneously. As a model system, we consider scalar particles which couple to a classical gravitational field but do not interact among themselves. The present section covers the steps leading to the finite approximant systems.

A. The classical field

Our theoretical laboratory is a free scalar field $A(x)$ which lives on a classical d dimensional spacetime $\mathscr{M}$ that carries an externally prescribed Euclidean metric $g_{\mu\nu}$. We assume that $\mathscr{M}$ is compact and has no boundary, and that the dynamics of $A(x)$ is governed by the matter action

$$\begin{aligned} S[A;g] &= \frac{1}{2}\int_{\mathscr{M}} d^d x \sqrt{g}\{g^{\mu\nu}D_\mu A D_\nu A + M^2 A^2 + \xi R A^2\} \\ &= \frac{1}{2}\int_{\mathscr{M}} d^d x \sqrt{g} A \mathscr{K} A \end{aligned} \qquad (3.1)$$

with a self-adjoint kinetic operator ($\Box_g \equiv D^2 \equiv g^{\mu\nu}D_\mu D_\nu$)

$$\mathscr{K} = -\Box_g + M^2 + \xi R(x). \qquad (3.2)$$

Stationarity of $S[A;g]$ with respect to A implies the equation of motion

$$[-\Box_g + M^2 + \xi R(x)]A(x) = 0, \qquad (3.3)$$

while its functional derivative with respect to the metric gives rise to the Euclidean stress tensor:

$$T^{\mu\nu}[A;g](x) = -\frac{2}{\sqrt{g}}\frac{\delta}{\delta g_{\mu\nu}(x)}S[A;g]. \qquad (3.4)$$

For arbitrary parameters M and ξ we have

$$T^{\mu\nu}[A;g] = D^\mu A D^\nu A - \frac{1}{2} g^{\mu\nu}[g^{\rho\sigma} D_\rho A D_\sigma A + M^2 A^2 + \xi R A^2]$$
$$+ \xi[R^{\mu\nu} - D^\mu D^\nu + g^{\mu\nu} D^2]A^2. \qquad (3.5)$$

Evidently the Rosenfeld-type stress tensor (3.4) is symmetric. If evaluated on a solution to the equation of motion, $A \equiv A^{\mathrm{sol}}$, it is also well known to be conserved and, under certain conditions, traceless:

$$D_\mu T^{\mu\nu}[A^{\mathrm{sol}};g] = 0, \qquad (3.6)$$

$$g_{\mu\nu} T^{\mu\nu}[A^{\mathrm{sol}};g] = 0 \quad \text{if } \xi = \xi_c(d) \quad \text{and} \quad M = 0. \qquad (3.7)$$

Here we abbreviated $\xi_c(d) \equiv \frac{d-2}{4(d-1)}$.

B. The quantum system at finite N

Next, we employ the above field theory as a classical inspiration in setting up a quantum mechanical system with finitely many degrees of freedom. Concretely, we identify those degrees of freedom with the N lowest eigenvalues of the kinetic operator.

(1) The spectral problem. Given the metric $g_{\mu\nu}$, we construct the operator (3.2) and consider its eigenvalue problem on $\mathcal{M}$:

$$\mathcal{K} u_{n,m}(x) = \mathcal{F}_n u_{n,m}(x). \qquad (3.8)$$

The discrete eigenvalues $\mathcal{F}_n$ are enumerated by an integer $n = 0, 1, 2, 3, \ldots$ which labels them in ascending order: $\mathcal{F}_0 < \mathcal{F}_1 < \mathcal{F}_2 < \cdots$. Allowing for a D_n-fold degeneracy of $\mathcal{F}_n$, the eigenfunctions $u_{n,m}(x)$ carry an additional degeneracy index, or multi-index, m. By analogy with the generalized spherical harmonics [29,30] we may think of the indices n and m as a kind of angular momentum and magnetic quantum number, respectively.

The eigenfunctions $\{u_{n,m}|n \geq 0; m = 1,2,\ldots,D_n\} \equiv \mathcal{B}$ form a complete set of scalar functions on $\mathcal{M}$. They can be orthonormalized with respect to the inner product on $L^2(\mathcal{M})$ supplied by the metric,

$$\int d^d x \sqrt{g} u^*_{n,m}(x) u_{\bar{n},\bar{m}}(x) = \delta_{n\bar{n}} \delta_{m\bar{m}} \qquad (3.9)$$

so that the corresponding completeness relation reads

$$\sum_{n,m} u_{n,m}(x) u^*_{n,m}(y) = \frac{\delta(x-y)}{\sqrt{g(y)}}. \qquad (3.10)$$

(2) Definition of the approximants. Given the basis $\mathcal{B}$, we define the quantum mechanical system $\mathsf{App}(N)$ by truncating the set $\mathcal{B}$ at the level $n = N < \infty$, retaining only the eigenfunctions of $\mathcal{K}$ with eigenvalues $\mathcal{F}_n \leq \mathcal{F}_N$. Instead of arbitrary fields $A(x)$, we consider only those that can be expanded in the truncated basis $\mathcal{B}_N \equiv \{u_{n,m}|n = 0, 1, \ldots, N; m = 1, \ldots, D_n\}$, i.e.,

$$A(x) = \sum_{n=0}^{N} \sum_{m=1}^{D_n} \alpha_{n,m} u_{n,m}(x). \qquad (3.11)$$

The degrees of freedom of the quantum system are represented then by the coefficients $\{\alpha_{n,m}|0 \leq n \leq N, m = 1, \ldots, D_n\}$. Their total number equals

$$\sum_{n \leq N} D_n \equiv \mathbf{f}(N). \qquad (3.12)$$

In a path integral treatment (which will be the topic of Sec. V below) the functional integral over $A(x)$ reduces to an integration over the finitely many coefficients $\alpha_{n,m}$s then. In view of the truncated expansion (3.11) it is suggestive to visualize the spacetime of $\mathsf{App}(N)$ as a fuzzy sphere [31].

(3) The 2-point function. Ordinarily, the key ingredient of a free Euclidean field theory is the 2-point correlation function $\langle \hat{A}(x) \hat{A}(y) \rangle^g \equiv G(x,y)$.[12] By Wick's theorem it determines all higher n-point functions, and it satisfies

$$\mathcal{K} G(x,y) = \frac{\delta(x-y)}{\sqrt{g(x)}} \qquad (3.13)$$

with suitable boundary conditions being specified. In the case $N \to \infty$, the completeness relation (3.10) gives rise to a formal solution of (3.13), namely[13]

$$G(x,y) = \sum_{n=0}^{\infty} \sum_{m=1}^{D_n} \frac{u_{n,m}(x) u^*_{n,m}(y)}{\mathcal{F}_n}. \qquad (3.14)$$

For the time being, we define the quantum theory of the finite field system $\mathsf{App}(N)$ by the correspondingly truncated version of Eq. (3.13), namely

$$\mathcal{K} G_N(x,y) = \frac{1}{\sqrt{g(x)}} \delta_N(x,y) \qquad (3.15)$$

whereby now both the regularized correlator,

$$G_N(x,y) \equiv \sum_{n=0}^{N} \sum_{m=1}^{D_n} \frac{u_{n,m}(x) u^*_{n,m}(y)}{\mathcal{F}_n} \qquad (3.16)$$

and the modified delta function,

---

[12]Depending on the context the caret notation ($\hat{A}$, etc.) indicates operators or integration variables under a functional integral. Furthermore, the notation $\langle \cdots \rangle^g$ emphasizes that all expectation values must be regarded functionals of the metric on $\mathcal{M}$.

[13]Here and in the following it is understood that if $\mathcal{K}$ has a zero mode it is separated off in the usual way. We do not indicate this notationally.

$$\delta_N(x, y) \equiv \sqrt{g(x)} \sum_{n=0}^{N} \sum_{m=1}^{D_n} u_{n,m}(x) u_{n,m}^*(y), \qquad (3.17)$$

are constructed in terms of functions from the truncated set $\mathscr{B}_N$ only.

(4) Bilinear observables. In the following we are mostly interested in observables that are bilinear in $\hat{A}(x)$. With the regularized 2-point function $G_N(x, y) = \langle \hat{A}(x)\hat{A}(y)\rangle_N^g$ at hand it is in principle straightforward to calculate their expectation values in the state G_N corresponds to. Thanks to the UV cutoff, $G_N(x, y)$ is nonsingular in the limit $y \to x$, and so all those observables have well-defined expectation values. They can be expressed by finite sums over the functions $u_{n,m}$.

The local monomial $D_\mu \hat{A}(x) D^\mu \hat{A}(x)$, for instance, leads to

$$\langle (D_\mu \hat{A})^2(x)\rangle = \lim_{y \to x} \langle \partial_\mu^x \hat{A}(x) \partial_y^\mu \hat{A}(y)\rangle_N^g$$
$$= \lim_{y \to x} \partial_\mu^x \partial_y^\mu G_N(x, y)$$
$$= \lim_{y \to x} \partial_\mu^x \partial_y^\mu \sum_{n \leq N,m} \frac{1}{\mathscr{F}_n} u_{n,m}(x) u_{n,m}^*(y)$$
$$= \sum_{n \leq N,m} \frac{1}{\mathscr{F}_n} D_\mu u_{n,m}(x) D^\mu u_{n,m}^*(x) \qquad (3.18)$$

which is perfectly finite as long as $N < \infty$.

A particularly interesting integrated monomial is $\int d^d x \sqrt{g} \hat{A}(x) \mathscr{K} \hat{A}(x)$. Its expectation value counts the number of degrees of freedom which the quantum mechanical system App(N) possesses:

$$\left\langle \int d^d x \sqrt{g} \hat{A}(x) \mathscr{K} \hat{A}(x) \right\rangle_N^g$$
$$= \int d^d x \sqrt{g(x)} \lim_{y \to x} \mathscr{K}_x \langle \hat{A}(x)\hat{A}(y)\rangle_N^g$$
$$= \int d^d x \sqrt{g(x)} \lim_{y \to x} \sum_{n \leq N,m} \frac{1}{\mathscr{F}_n} \mathscr{K}_x u_{n,m}(x) u_{n,m}^*(y)$$
$$= \sum_{n \leq N,m} \int d^d x \sqrt{g(x)} u_{n,m}(x) u_{n,m}^*(x)$$
$$= \sum_{n \leq N} D_n$$
$$= \mathbf{f}(N). \qquad (3.19)$$

Here we also exploited the eigenvalue equation and the normalization condition satisfied by the mode functions.

(5) Trace of the stress tensor. The most important bilinear operator is the energy-momentum tensor of the field modes that inhabit $\mathcal{M}$. Thanks to the N cutoff, the operator $T^{\mu\nu}[\hat{A}; g]$ which we obtain from the classical

expression (3.5) by letting $A(x) \to \hat{A}(x)$ is not plagued by any operator product singularities.

Nevertheless, and this is important to be kept in mind, there is always an ambiguity with regard to the "correct" energy-momentum tensor of App(N). As always in quantum mechanics, the classical expression for an observable is at best an "inspiration" when guessing the quantum operator. After all, the two can differ by any number of $O(\hbar)$ terms that disappear in the classical limit.

Therefore, if we now *declare* that $T^{\mu\nu}[\hat{A}; g]$ as given by (3.5) with $A \to \hat{A}$ is the correct energy-momentum tensor of our quantum mechanical system, this amounts to a choice over and above the decision for an N cutoff.

We shall need in particular the operator which represents the trace of $T^{\mu\nu}$. It writes, without using the field equations,

$$T_\mu^\mu[\hat{A}; g] = \frac{1}{2}[2 - d + 4(d-1)\xi]D_\mu \hat{A} D^\mu \hat{A}$$
$$+ 2(d-1)\xi \hat{A} D_\mu D^\mu \hat{A}$$
$$+ \frac{1}{2}(2-d)\xi R \hat{A}^2 - \frac{1}{2} d M^2 \hat{A}^2. \qquad (3.20)$$

The expectation value of (3.20) is easily obtained by the same steps as above:

$$\langle T_\mu^\mu[\hat{A}; g](x)\rangle_N^g = \frac{1}{2} \sum_{n \leq N,m} \frac{1}{\mathscr{F}_n} \{[2 - d + 4(d-1)\xi]$$
$$\times D_\mu u_{n,m}^*(x) D^\mu u_{n,m}(x)$$
$$+ 4(d-1)\xi u_{n,m}^*(x) D_\mu D^\mu u_{n,m}(x)$$
$$+ [(2-d)\xi R(x) - dM^2]u_{n,m}^*(x) u_{n,m}(x)\}. \qquad (3.21)$$

Up to this point, $g_{\mu\nu}$ is an arbitrary externally prescribed metric. While it is usually difficult to solve the spectral problem of $\mathscr{K}$ in a concrete case, it is clear that in principle (3.21) and its untraced analogue provide us with well-defined finite expectation values for any choice of $g_{\mu\nu}$.

C. Backreaction on the metric

Next we promote the spacetime metric $g_{\mu\nu}$ to a dynamical, yet still classical, quantity which responds to the energy and momentum carried by the quantum fluctuations of the finite field system App(N). We assume this system to be in its ground state. Classically this means that $A = 0$ and hence $T_{\mu\nu} = 0$ everywhere on $\mathcal{M}$. Instead, quantum mechanically, the vacuum fluctuations of the $\mathbf{f}$ degrees of freedom contribute to the energy and momentum in the Universe which determine the metric on $\mathcal{M}$.

We assume that the metric is governed by the semi-classical Einstein equation

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda_{\mathrm{b}}g_{\mu\nu} = 8\pi G\langle T_{\mu\nu}[\hat{A};g]\rangle_N^g \qquad (3.22)$$

where Λ_{b} is a bare cosmological constant. Importantly, the rhs of the equation (3.22) involves the same metric $g_{\mu\nu}$ as its left-hand side (lhs), both explicitly via the operator $T_{\mu\nu}[\hat{A};g]$, and implicitly through the expectation value. This is what makes solutions to (3.22) *self-consistent*.

We denote such self-consistently determined metrics by $(g_N^{\mathrm{SC}})_{\mu\nu}$ in the following. The question we shall be particularly interested in concerns the dependence of the solutions on the parameter N, and thus on the number $\mathbf{f}(N)$ of degrees of freedom living on $\mathcal{M}$.

In full generality the semiclassical Einstein equation (3.22) represents an extremly hard problem; in principle the expectation value involved must be computed as an explicit functional of $g_{\mu\nu}$. It is given by sums over eigenfunctions like in (3.21). Evaluating them requires first of all solving the spectral problem of $\mathcal{K} \equiv \mathcal{K}[g]$ for "all" metrics $g_{\mu\nu}$.

(1) Maximally symmetric spacetimes. To make some progress here, we restrict the space of metrics to those of maximally symmetric Riemannian spaces of positive curvature, i.e., spheres $S^d(L)$. They come with only a single free parameter, namely the radius L, a Euclidean version of the Hubble length. We write the metric on $S^d(L)$ in the form

$$g_{\mu\nu}(x) = L^2\gamma_{\mu\nu}(x) \qquad (3.23)$$

where $\gamma_{\mu\nu}$ is the dimensionless metric on the unit d-sphere.[14] Thus, the determination of self-consistent background geometries of the type $\mathcal{M} = S^d(L)$ boils down to finding the N dependence of $L \equiv L^{\mathrm{SC}}(N)$.

The curvature scalar on spheres is a constant,

$$R = R(L) \equiv \frac{d(d-1)}{L^2}.$$

Therefore the eigenfunctions $u_{n,m}$ of $\mathcal{K} = -\Box_g + M^2 + \xi R$ coincide with those of the scalar Laplacian $\Box_g = D_\mu D^\mu$, and the eigenvalues of $\mathcal{K}$ are

$$\mathcal{F}_n = \mathcal{E}_n + M^2 + \xi R \qquad (3.24)$$

with $\{\mathcal{E}_n\}$ denoting the spectrum of $-\Box_g$:

$$-\Box_g u_{n,m}(x) = \mathcal{E}_n u_{n,m}(x). \qquad (3.25)$$

The eigenvalues $\mathcal{E}_n$ and their multiplicities D_n are well known [29,30]:

---

[14]Our conventions concerning the assignment of canonical mass dimensions are such that $[x^\mu] = [\partial_\mu] = [\mathrm{d}^d x] = [\delta(x)] = 0$, while $[g^{\mu\nu}] = -[g_{\mu\nu}] = [R] = [\Box_g] = [\mathcal{F}_n] = +2$, $[g] \equiv [\det(g_{\mu\nu})] = -2d$, $[\gamma_{\mu\nu}] = 0$ and $[A] = \frac{1}{2}(d-2)$, $[g^{1/4}\mathrm{d}A] = -1$, $[\mu] = -[L] = [\mathscr{P}] = +1$.

$$\mathcal{E}_n = \frac{n(n+d-1)}{L^2} \equiv \mathcal{E}_n(L), \quad n = 0, 1, 2, \ldots, \quad (3.26)$$

$$D_n = \frac{(2n+d-1)(n+d-2)!}{n!(d-1)!}. \qquad (3.27)$$

On the sphere, $\Box_g$ has a zero mode, the constant function appearing at $n = 0$. We exclude this mode from the degrees of freedom belonging to $\mathsf{App}(N)$. So, to be precise, their total number equals

$$\mathbf{f}(N) = \sum_{n=1}^{N} D_n. \qquad (3.28)$$

This, and all similar sums appearing below start at $n = 1$. However, for the present analysis the precise treatment of the low lying modes plays no role; the relevant regime will always be dominated by $n \gg 1$.

(2) The effective Einstein equation. Thanks to the maximum symmetry of the background geometry we have $\langle T_{\mu\nu}\rangle_N^g \propto g_{\mu\nu}$, and so it suffices to analyze the traced, and now, x-independent Einstein equation:

$$-\frac{1}{2}(d-2)R(L) + d\Lambda_{\mathrm{b}} = 8\pi G\langle T_\mu^\mu[\hat{A};g]\rangle_N^g. \qquad (3.29)$$

Moreover, no information is lost when we integrate (3.29) over spacetime:

$$\left\{-\frac{1}{2}(d-2)R(L) + d\Lambda_{\mathrm{b}}\right\}\mathrm{Vol}[S^d(L)]$$
$$= 8\pi G\int \mathrm{d}^d x\sqrt{g}\langle T_\mu^\mu[\hat{A};g]\rangle_N^g. \qquad (3.30)$$

The virtue of the latter integration is that, upon inserting our earlier result (3.21) into (3.30), it allows us to perform an integration by parts on the $D_\mu u^* D^\mu u$-terms, and then to simplify the entire sum by exploiting (3.25) and the orthonormality of the $u_{n,m}$s.

This brings us to the main result of this section, namely the following condition for self-consistency:

$$-\frac{1}{2}(d-2)R(L) + d\Lambda_{\mathrm{b}} = 8\pi G\frac{\Theta_N(L)}{\mathrm{Vol}[S^d(L)]}. \qquad (3.31)$$

Its main building block is the dimensionless and manifestly finite-mode sum representing the integrated trace of the stress tensor:

$$\Theta_N(L) \equiv \int \mathrm{d}^d x\sqrt{g}\langle T_\mu^\mu[\hat{A};g]\rangle_N^g$$
$$= -\sum_{n=1}^{N} D_n\left[\frac{1}{2}(d-2) + \frac{M^2}{\mathcal{E}_n(L) + \xi R(L) + M^2}\right]. \qquad (3.32)$$

Furthermore, the volume in Eq. (3.31) is given by

$$\text{Vol}[S^d(L)] = \frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)} L^d. \qquad (3.33)$$

Several remarks are in order at this point.

(3) Limiting cases. For vanishing and very large (infinite) mass we obtain, respectively,

$$\Theta_N(L)|_{M=0} = -\frac{1}{2}(d-2)\mathbf{f}(N) \qquad (3.34)$$

$$\Theta_N(L)|_{M\to\infty} = -\frac{1}{2}d\mathbf{f}(N). \qquad (3.35)$$

Note that the limit $M \to \infty$ is performed at fixed, finite N.

(4) Background dependent counterpart. It should be emphasized that the respective L dependencies on the lhs and rhs of the reduced Einstein equation (3.31) have a very different origin: The one on the lhs stems from the metric to be found, i.e., the one in the Einstein tensor. On the rhs instead, in $\Theta_N(L)/\text{Vol}[S^d(L)]$, the radius $L \equiv L_{\text{RS}}$ refers to a logically different metric of a certain rigid spacetime ("RS") in which energy and momentum of the vacuum fluctuations are computed. Hence, in a traditional *background dependent* calculation, Eq. (3.31) would appear replaced by

$$-\frac{1}{2}(d-2)R(L) + d\Lambda_{\text{b}} = 8\pi G \frac{\Theta_N(L_{\text{RS}})}{\text{Vol}[S^d(L_{\text{RS}})]}. \qquad (3.36)$$

Herein L_{RS} is an absolute constant, fixed by hand once and for all. In the literature a popular choice of such a rigid spacetime is flat space ($L_{\text{RS}} \to \infty$) since the corresponding mode functions are technically easiest to deal with.

However, in the rest of this paper we shall have many opportunities to see that solving (3.36) at fixed L_{RS} can result in sequences $L = L(N)$ which are very different from those derived in a background independent fashion, namely by first setting $L = L_{\text{RS}}$ and solving Einstein's equation thereafter. We believe that the L_{RS}-based sequences convey a physically wrong picture of what happens in the limit $N \to \infty$.

(5) Four dimensions. In the following sections we shall analyze the condition for self-consistent maximally symmetric backgrounds in detail and explore the corresponding sequences

$$N \mapsto L^{\text{SC}}(N; \xi, M, G, \Lambda_{\text{b}}) \qquad (3.37)$$

in dependence on the various bare parameters. We specialize for $d = 4$ dimensions then, so that the self-consistency condition becomes

$$-R(L) + 4\Lambda_{\text{b}} = \frac{3G}{\pi} \frac{\Theta_N(L)}{L^4} \qquad (3.38)$$

with $R(L) = 12/L^2$, and

$$-\Theta_N(L) = \sum_{n=1}^{N} D_n \left[1 + \frac{(ML)^2}{n(n+3) + 12\xi + (ML)^2} \right]. \qquad (3.39)$$

In 4 dimensions, $\mathcal{E}_n = n(n+3)/L^2$, and the multiplicities of the eigenvalues are governed by the cubic polynomial

$$D_n = \frac{1}{6}(2n+3)(n+2)(n+1). \qquad (3.40)$$

One easily proves by mathematical induction that the sum (3.28) evaluates to

$$\mathbf{f}(N) = \sum_{n=1}^{N} D_n = \frac{1}{12}[N^4 + 8N^3 + 23N^2 + 28N]. \qquad (3.41)$$

This is the total number of degrees of freedom inhabiting our spherical spacetimes in the four dimensional case.

(6) The fuzzy $\mathbf{S}^4$. On the 4-sphere, the eigenfunctions $u_{n,m} \equiv Y_{nl_1l_2m}$ are labeled by four integer quantum numbers. Besides $n = 0, 1, 2, \ldots$ which determines the eigenvalue, there is a triple of integers (l_1, l_2, m) satisfying $n \geq l_1 \geq l_2 \geq |m|$. They play the role of the degeneracy multi-index m now. The S^4 harmonics depend on four angular coordinates and have the general structure

$$Y_{nl_1l_2m}(\zeta, \eta, \vartheta, \varphi) \propto {}_4\bar{P}_n^{l_1}(\zeta){}_3\bar{P}_{l_1}^{l_2}(\eta){}_2\bar{P}_{l_2}^m(\vartheta)\frac{1}{\sqrt{2\pi}}e^{im\varphi} \qquad (3.42)$$

where the ${}_i\bar{P}_k^j$ denote generalized associated Legendre functions, see [32] for a detailed account.

Recalling the construction of the approximants, an interesting property is the "resolving power" of the basis functions (3.42) when one restricts $n \leq N$ in the series expansions. It is not difficult to show that functions $A(x) \equiv A(\zeta, \eta, \vartheta, \varphi)$ represented by such truncated series can display nontrivial structures down to angular separations of approximately [33]

$$\Delta\alpha \approx \frac{\pi}{N}. \qquad (3.43)$$

The minimum proper distance that can be resolved by the truncated basis is about $\Delta\ell \approx \pi L/N$ then. In this sense, the spacetimes of the approximants are reminiscent of "fuzzy spheres" [31].

IV. SEQUENCES OF SELF-GRAVITATING QUANTUM SYSTEMS

Next, we apply the above semiclassical Einstein equation in order to search for sequences of well-behaved self-gravitating quantum systems enumerated by the cutoff

quantum number N. The member labeled "N", $\mathsf{App}(N)$, consists of $\mathbf{f}(N)$ quantized field degrees of freedom. They inhabit a spacetime whose metric, $(g_N^{\mathrm{SC}})_{\mu\nu}$, they decide about in a self-determined, democratic way.

A. Massless scalar field

We begin by considering a massless scalar field in 4 dimensions. Setting $M = 0$ in (3.39) has the very welcome consequence that the sum that is to be evaluated boils down to nothing but the counting function (3.41):

$$-\Theta_N(L) = \sum_{n=1}^{N} D_n = \mathbf{f}(N) = \frac{N^4}{12}\left\{1 + O\left(\frac{1}{N}\right)\right\}. \quad (4.1)$$

We see that upon setting $M = 0$ the integrated trace Θ_N automatically becomes independent of L and ξ also. It is given by a pure number, namely the number of degrees of freedom the quantum system possesses. (The relative minus sign in $\Theta_N = -\mathbf{f}$ is a consequence of our Euclidean conventions for the stress tensor.) Hence the only L dependence on the rhs of the self-consistency condition (3.38) is due to the volume factor $\propto L^4$:

$$\frac{12}{L^2} - 4\Lambda_{\mathrm{b}} = \frac{3G}{\pi}\frac{\mathbf{f}(N)}{L^4}. \quad (4.2)$$

(1) *The classical initial point.* If we set $N = 0$ there are no quantum mechanical degrees of freedom, $\mathbf{f}(0) = 0$, and provided $\Lambda_{\mathrm{b}} > 0$, Eq. (4.2) yields $L = \sqrt{\Lambda_{\mathrm{b}}/3}$, i.e., the radius of the well-known S^4 solution to Einstein's equation *in vacuo*.

(2) *Exactness.* Incidentally, the modified Einstein equation (4.2) can be reexpressed succinctly in terms of the curvature scalar as

$$R - 4\Lambda_{\mathrm{b}} = \frac{G}{48\pi}\mathbf{f}(N)R^2. \quad (4.3)$$

The R^2 term on its rhs might be reminiscent of the derivative expansions that often are calculated on the basis of the asymptotic heat-kernel series. It must be stressed however that (4.3) actually enjoys a much more reliable status: For spherical spacetimes, the rhs of Eq. (4.3) is an exact, nonperturbative result and not merely a term in an asymptotic series. Its derivation does not involve any expansion in a small coupling or in the number of derivatives. For $M = 0$ and in 4 dimensions, curvature powers both higher and lower than R^2 are strictly absent.

B. The background dependent calculation

We begin the discussion of the self-consistency condition (4.2) by solving its background dependent counterpart. As we mentioned in Sec. III C, in its $\langle T_\mu^\mu\rangle$-term it has L replaced by a rigid scale L_{RS} independent of L, the radius

representing the dynamical metric in the symmetry reduced case. As a consequence, the $\langle T_\mu^\mu\rangle$-term can be treated as a contribution to the cosmological constant, yielding

$$\frac{3}{L^2} = \Lambda_{\mathrm{tot}}(N) \quad (4.4)$$

with the modified cosmological constant

$$\Lambda_{\mathrm{tot}}(N) \equiv \Lambda_{\mathrm{b}} + \frac{3G}{4\pi}\frac{\mathbf{f}(N)}{L_{\mathrm{RS}}^4}. \quad (4.5)$$

(1) Now we let $N \to \infty$ with L_{RS} and Λ_{b} fixed. Then the total cosmological constant behaves as $\Lambda_{\mathrm{tot}}(N) \propto N^4 \nearrow +\infty$ for sufficiently large N. This forces the dynamical radius to approach zero,

$$L(N) = \left(\frac{1}{3}\Lambda_{\mathrm{tot}}(N)\right)^{-1/2} \propto \frac{1}{N^2} \to 0, \quad (4.6)$$

so that the curvature scalar of the solution grows unboundedly, $R(N) = 12/L(N)^2 \propto N^2 \to \infty$.

What we encounter here is an incarnation of the cosmological constant problem as it arises from the Pauli-style calculations. They sum up the vacuum energies of a certain number of modes propagating *on a fixed spacetime*, and thereafter insert the resulting energy density *in one package* into Einstein's equation as part of the cosmological constant. Then either Λ_{tot} becomes unacceptably large for any physically plausible choice of the cutoff scale, or the bare parameter Λ_{b} must be fine-tuned with unnatural precision.

(2) In the background dependent calculation, one of the roles played by the rigid metric on $\mathcal{M} = S^4(L_{\mathrm{RS}})$ consists in relating the dimensionless cutoff N to a dimensionful one. The eigenvalue $\mathcal{E}_N$ of $-\square$, on this background geometry, is analogous to the dimensionful cutoff scale (traditionally denoted Λ^2) at which an ordinary momentum cutoff on flat space would become operative. In the case at hand, the pure number N gives rise to the UV cutoff scale $\mathscr{P}$ according to

$$\mathscr{P}^2 = \mathcal{E}_N(L_{\mathrm{RS}}) = \frac{N(N+3)}{L_{\mathrm{RS}}^2}. \quad (4.7)$$

Note that $\mathscr{P}$ and $\mathcal{E}_N$ have canonical mass dimensions $+1$ and $+2$, respectively.[15] Indeed, (4.7) is an example of a $\mathscr{P}$-type cutoff.

It sounds like a quite trivial remark—actually it is not, as we shall see—that $\mathscr{P}$ is a monotonically increasing function of N, and hence $N \to \infty$ is tantamount to $\mathscr{P} \to \infty$.

---

[15]In the QFT literature, $\mathscr{P}$ is denoted as Λ usually. We do not use this notation to avoid confusion with the cosmological constant.

(3) Note also that when reexpressed in terms of $\mathscr{P}$, the curvature scalar reads, for $N \gg 1$ and $\Lambda_{\mathrm{b}} = 0$, say

$$R = 4\Lambda_{\mathrm{tot}} = \frac{G}{4\pi}\mathscr{P}^4. \tag{4.8}$$

When presented in this fashion, the unphysical quantity L_{RS} drops out from the final result.[16] This may further contribute to the—false, as it turns out—impression that invoking a rigid auxiliary spacetime during the intermediate steps of the calculation is just a harmless technical convenience.

C. Self-consistent approximants: The case $\Lambda_{\mathrm{b}} = 0$

Let us now begin our search for sequences of approximants which satisfy the requirement of self-consistency. For various parameter choices of interest we determine their Hubble radii $L \equiv L^{\mathrm{SC}}(N)$, and more importantly, the pertinent scalar curvatures

$$R^{\mathrm{SC}}(N) = \frac{12}{L^{\mathrm{SC}}(N)^2}. \tag{4.9}$$

Sometimes it is suggestive to think of the quantum mechanical term in the effective Einstein equation, at the point of consistency, to contribute an additional piece to the cosmological constant; this shifts Λ_{b} to a certain $\Lambda^{\mathrm{SC}}(N)$ which satisfies, by definition,

$$R^{\mathrm{SC}}(N) = 4\Lambda^{\mathrm{SC}}(N). \tag{4.10}$$

For the specific example of Eq. (4.2), the modified cosmological constant reads

$$\Lambda^{\mathrm{SC}}(N) = \Lambda_{\mathrm{b}} + \frac{3G}{4\pi}\frac{\mathbf{f}(N)}{L^{\mathrm{SC}}(N)^4}. \tag{4.11}$$

Of course, the relation (4.11) can be employed only *after* having solved Einstein's equation: this is the very difference between the background dependent treatment and the background independent one.

(1) The $\Lambda_{\mathrm{b}} = 0$ solution. Letting $N = 1, 2, 3, \ldots$, we now populate spacetime with an increasing number of degrees of freedom and check if Eq. (4.2) admits self-consistent S^4 solutions.

In this subsection we focus on the particularly interesting case of a vanishing bare cosmological constant, $\Lambda_{\mathrm{b}} = 0$. After multiplication with L^4, assuming $L \neq 0, \infty$, the self-consistency condition becomes very simple therefore:

$$L^2 = \frac{G}{4\pi}\mathbf{f}(N). \tag{4.12}$$

Obviously, the quantum mechanical contribution to the Einstein equation has just the correct sign so that there does indeed exist a self-consistent solution $L \equiv L^{\mathrm{SC}}(N)$ for any number of degrees of freedom:

$$
\begin{aligned}
L^{\mathrm{SC}}(N) &= \left[\frac{G}{4\pi}\mathbf{f}(N)\right]^{1/2} \\
&= \left(\frac{G}{48\pi}\right)^{1/2} N^2\left\{1 + O\left(\frac{1}{N}\right)\right\}.
\end{aligned}
\tag{4.13}
$$

The scalar curvature of the spacetimes found are given by

$$
\begin{aligned}
R^{\mathrm{SC}}(N) &= \frac{48\pi}{G}\cdot\frac{1}{\mathbf{f}(N)} \\
&= \frac{576\pi}{G}\cdot\frac{1}{N^4}\left\{1 + O\left(\frac{1}{N}\right)\right\}.
\end{aligned}
\tag{4.14}
$$

This sequence of self-consistent spacetimes $\{\mathsf{App}(N)\}$ displays a number of highly surprising and unusual features:

(2) Inflating spheres. When additional scalar modes are added to the quantum system, i.e., N is increased, the radius $L^{\mathrm{SC}}(N)$ becomes *larger*, and the curvature $R^{\mathrm{SC}}(N)$ correspondingly *smaller*. In the limit $N \to \infty$, the radius of the sphere approaches infinity, and the self-consistent spacetime that supports those infinitely many field modes approaches *flat space.*

This behavior is the exact opposite of what we had found by means of the background dependent calculation: There, adding further modes led to a *smaller* radius, higher curvature, and a larger effective cosmological constant. And in the limit $N \to \infty$ the spacetime degenerated to a point even.

(3) The $\mathscr{P}$ cutoff. To understand the origin of the unexpected result in the background independent case, let us look at the dimensionful companion of the pure-number cutoff N. In absence of any absolute metric that could be employed to turn N into a dimensionful quantity, i.e., an *inverse proper length*, only the actually realized, dynamically determined self-consistent metric can be used for this purpose. Since this metric depends on N, the $(-\square)$ eigenvalue at the upper end of the sequence, i.e., $\mathcal{E}_N \equiv \mathscr{P}^2(N)$, acquires a second, indirect dependence on N now, namely via the radius:

$$\mathscr{P}(N)^2 = \frac{N(N+3)}{(L^{\mathrm{SC}}(N))^2} = \frac{4\pi}{G}\cdot\frac{N(N+3)}{\mathbf{f}(N)}. \tag{4.15}$$

Now, when N increases the novel factor $\mathbf{f}(N)$ in the denominator of (4.15) grows $\propto N^4$ for large N, and so it overrides the familiar $N(N+3)$ in the numerator. As a

consequence, the relationship between the dimensionless N cutoff and its dimensionful counterpart $\mathscr{P}(N)$ assumes a rather unusual and unexpected form in the background independent case; when $N \gg 1$,

$$\mathscr{P}(N) = \left(\frac{48\pi}{G}\right)^{1/2} \frac{1}{N} \left\{1 + O\left(\frac{1}{N}\right)\right\}. \quad (4.16)$$

We see that in the background independent setting the dimensionful cutoff $\mathscr{P}$ is a decreasing function of the dimensionless integer N.

This is in sharp contradistinction to the standard result $\mathscr{P} \approx N/L_{\mathrm{RS}}$ of (4.7) which one obtains in the background dependent case.

(i) Note also that the relation (4.16) that connects $\mathscr{P}$ and N depends *nonanalytically* on Newton's constant which hints at its nonperturbative character.

(ii) While at first sight the background independent relationship $\mathscr{P} \propto 1/N$ seems rather counterintuitive, it is nevertheless easy to understand its origin:

Each member in the sequence of self-gravitating systems $\{\mathsf{App}(N)\}_{N \in \mathbb{N}}$ comes with its own self-consistent, that is, dynamically determined metric $(g_N^{\mathrm{SC}})_{\mu\nu}$, here represented by the radius $L^{\mathrm{SC}}(N)$. And each member uses its own, individual metric in order to convert its number in the sequence, N, to a dimensionful cutoff scale which then enjoys the status of an inverse proper length with respect to this particular metric. It is clear therefore that if $(g_N^{\mathrm{SC}})_{\mu\nu}$ has a sufficiently strong N dependence, the emergent function $\mathscr{P}(N)$ no longer has any reason to depend on N monotonically.

In the case at hand we encounter the extreme situation where the N dependence of the metric is so strong that increasing N actually even lowers the mass scale of the UV cutoff, $\mathscr{P}(N)$.

(iii) It is also remarkable that upon using (4.16) to eliminate N in favor of $\mathscr{P}$, Eq. (4.14) assumes the form

$$R^{\mathrm{SC}}(N) = \frac{G}{4\pi} \mathscr{P}(N)^4. \quad (4.17)$$

This is exactly the same relationship between the curvature and $\mathscr{P}$ as in Eq. (4.8) which resulted from the background dependent calculation. An yet, there remains the crucial difference that now in (4.17) both $R^{\mathrm{SC}}(N) \to 0$ and $\mathscr{P}(N) \to 0$ when $N \to \infty$.

D. Self-consistent approximants: The case $\Lambda_{\mathbf{b}} \neq \mathbf{0}$

Next, we admit a nonvanishing bare cosmological constant Λ_{b}. For now we assume it independent of N. Provided $L \neq 0, \infty$, the self-consistency condition (4.2) is equivalent to a quadratic equation for L^2:
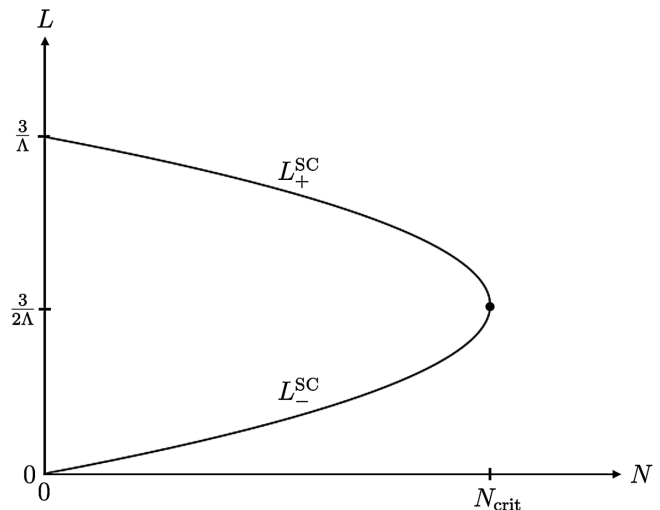

FIG. 2. The two branches of self-consistent solutions $L_{\pm}^{\mathrm{SC}}(N)$. The upper (lower) curve corresponds to the perturbative (nonperturbative) sequence. Both sequences terminate at N_{crit}. No S^4 solutions exist beyond this point.

$$\frac{2}{3}\Lambda_{\mathrm{b}}L^4 - 2L^2 + \frac{G}{2\pi}\mathbf{f}(N) = 0. \quad (4.18)$$

(1) Positive bare cosmological constant. If $\Lambda_{\mathrm{b}} > 0$ this equation admits the following two branches of solutions:

$$(L_{\pm}^{\mathrm{SC}}(N))^2 = \frac{3}{2\Lambda_{\mathrm{b}}}\left[1 \pm \sqrt{1 - \frac{G\Lambda_{\mathrm{b}}}{3\pi}\mathbf{f}(N)}\right]. \quad (4.19)$$

In Fig. 2 the radii $L_{\pm}^{\mathrm{SC}}(N)$ are plotted in dependence on N. The main properties of the two sequences are as follows.

(i) In the pure gravity limit $N = 0$, $\mathbf{f} = 0$ one obtains the radii

$$L_{+}^{\mathrm{SC}}(0) = \sqrt{3/\Lambda_{\mathrm{b}}}, \quad (4.20)$$

$$L_{-}^{\mathrm{SC}}(0) = 0, \quad (4.21)$$

so that the "plus" branch reproduces the result from general relativity, while the "minus" branch has a singular $N = 0$ limit, which corresponds to a degenerate geometry with infinite curvature.

(ii) Letting $N = 1, 2, 3, \ldots$, the radius $L_{+}^{\mathrm{SC}}(N)$ decreases, while $L_{-}^{\mathrm{SC}}(N)$ increases, describing a meaningful geometry now.

(iii) There exists a critical number N_{crit} at which L_{+}^{SC} and L_{-}^{SC} become equal, and beyond which there do not exist any S^4-type solutions to the self-consistency conditions. This critical number of degrees of freedom inhabiting the Universe is reached when $\frac{G\Lambda_{\mathrm{b}}}{3\pi}\mathbf{f}(N_{\mathrm{crit}}) = 1$. If $G\Lambda_{\mathrm{b}} \ll 1$ we may use the asymptotic form of $\mathbf{f}(N)$ and obtain

$$N_{\text{crit}} \approx \left(\frac{36\pi}{G\Lambda_{\text{b}}}\right)^{1/4}. \qquad (4.22)$$

(iv) Sending $\Lambda_{\text{b}} \searrow 0$ at fixed N, the point where the two branches meet moves out to infinity, both in the L and the N direction. In this limit the minus branch becomes

$$\lim_{\Lambda_{\text{b}} \searrow 0} L_-^{\text{SC}}(N) = \left[\frac{G}{4\pi}\mathbf{f}(N)\right]^{1/2}, \qquad (4.23)$$

and so it reproduces exactly the solution found in the previous subsection.

(v) To summarize: For nonzero, positive Λ_{b} there exist two N sequences, a "perturbative" one, $L_+^{\text{SC}}(N)$, and a "nonperturbative" one, $L_-^{\text{SC}}(N)$.

Along the *perturbative sequence*, the self-consistent radius $L_+^{\text{SC}}(N)$ starts out from its classical ($\hbar = 0$) value, decreases with the number of degrees of freedom, and causes the curvature $R_-^{\text{SC}}(N) = 12/L_-^{\text{SC}}(N)$ to increase correspondingly. As long as $N < N_{\text{crit}}$, this behavior is similar to the one we are familiar with from the Pauli-type calculations.

The *nonperturbative* $L_-^{\text{SC}}(N)$ *sequence* instead, while having a singular initial point at $N = 0$, describes a series of universes which become increasingly larger and more "inhabitable" when further degrees of freedom are added.

Both sequences terminate at the same critical number of modes living on the approximant's spacetime. Beyond N_{crit} self-consistent geometries, if they exist, are necessarily more complicated than round spheres. Qualitatively this picture is in accord with indirect evidence from the functional renormalization group indicating that vacuum fluctuations on a certain length scale generate curvature structures *on that particular scale* [34,35].[17]

(2) Negative bare cosmological constant. It is quite remarkable that even if $\Lambda_{\text{b}} < 0$ the condition (4.18) admits a (single) sequence of self-consistent spacetimes. As the classical Einstein equation possesses no S^4 solutions in this case, the sequence owes its existence entirely to the quantum effects:

$$L^{\text{SC}}(N)^2 = \frac{3}{2|\Lambda_{\text{b}}|}\left[\sqrt{1 + \frac{|\Lambda_{\text{b}}|G}{3\pi}\mathbf{f}(N)} - 1\right]. \qquad (4.24)$$

This sequence begins at $N = 1$, then the radius $L^{\text{SC}}(N)$ grows monotonically with N, the curvature decreases, and for $N \to \infty$ the self-consistent spacetime approaches flat space finally. Clearly this behavior is a perfect surprise from both the classical gravity and the background dependent QFT point of view.

---

[17]The picture is also in line with a purely classical proposal of "hiding" the cosmological constant at short scales [36]. See also [37,38] for related work in the context of QFT.

E. Massive and nonminimally coupled scalars

Up to now we assumed a vanishing bare mass, and as a consequence the integrated trace of the stress tensor was essentially the mode counting function, $\Theta_N = -\mathbf{f}(N)$. In the general case with $M \neq 0$ a partial fraction decomposition of (3.39) with (3.40) expresses Θ_N in the following suggestive form:

$$-\Theta_N(L) = \mathbf{f}(N) + \frac{1}{6}(ML)^2\left[\sum_{n=1}^{N}(2n+3)\right.$$

$$+ 2(2-z)\sum_{n=1}^{N}\frac{n}{n(n+3)+z}$$

$$\left. + 3(2-z)\sum_{n=1}^{N}\frac{1}{n(n+3)+z}\right]. \qquad (4.25)$$

Here we introduced the abbreviation $z \equiv (ML)^2 + 12\xi$.

The leading large-N behavior of the terms in the first, second and third line of (4.25) is, respectively, $\propto N^4$, N^2 and $\log(N)$; the last term, containing a convergent sum, becomes N independent asymptotically. Hence a nonzero M has comparatively little impact on Θ_N. When $N \gg 1$, it does not modify the dominant N^4 and N^3 terms coming from $\mathbf{f}(N)$, and shows up at the next-to-next-to leading order only, being proportional to $(ML)^2\sum_{n=1}^{N}(2n+3) \propto (ML)^2N^2$. The first term in which ξ makes its appearance is even more suppressed.

Moreover, taking the limit $ML \to \infty$ directly in Eq. (3.39) at fixed N, we see that even within the whole span of masses between $M = 0$ and $M \to \infty$ the trace $\Theta_N(L)$ changes only by a modest factor of 2:

$$\Theta_N(L)|_{M \to \infty} = 2\Theta_N(L)|_{M=0}. \qquad (4.26)$$

Because of their weak and qualitatively unimportant role when $N \gg 1$ we shall not discuss the case of a nonzero mass and nonminimal coupling any further.

V. THE PATH INTEGRAL ROUTE

In this section we consider the coupling of the scalar field degrees of freedom to gravity from an effective action and functional integral point of view. At first sight it might seem that not much new can be learned in this way, dealing with a simple Gaussian theory after all. However, the specific needs of the present investigation bring certain aspects to light which are important now, while they often can be brushed away in typical background dependent calculations.

For example, generically a scalar plus gravity system is described by an effective action $\Gamma[A; g]$ depending on arbitrary fields $A(x)$ and $g_{\mu\nu}(x)$. However, if one is interested in the "particle physics" of A on a rigid, say flat, spacetime only, there is no need to know the A-independent terms of Γ,

i.e., the invariants constructed from $g_{\mu\nu}$ alone. For us those terms are crucial though.

So the plan of this section is, first, to pinpoint precisely which parts of the functional integral for Γ do, or do not depend on the metric, and second, install a regulator that complies with the general requirements of an N cutoff.

The relevant object is the induced gravity action $\Gamma_{\text{scal}}[g]$ which is generated by the vacuum fluctuations of a scalar field. In the case at hand the latter is governed by the classical action $S[A;g] \equiv \frac{1}{2}\int_{\mathcal{M}} d^d x \sqrt{g} A \mathcal{K} A$ with $\mathcal{K}$ as in (3.2). The functional $\Gamma_{\text{scal}}[g]$ augments the Einstein-Hilbert action of pure classical gravity

$$S_{\text{EH}}[g] = \frac{1}{16\pi G}\int d^d x \sqrt{g}(-R + 2\Lambda_{\text{b}}) \qquad (5.1)$$

to the effective gravitational action

$$\Gamma[g] = S_{\text{EH}}[g] + \Gamma_{\text{scal}}[g]. \qquad (5.2)$$

The functional $\Gamma[g]$ is the restriction of the full Legendre effective action $\Gamma[A;g]$ to a vanishing scalar field, $\Gamma[g] \equiv \Gamma[A;g]|_{A=0}$. As A has no self-interactions, $\Gamma[A;g] = S[A;g] + \Gamma_{\text{scal}}[g] + S_{\text{EH}}[g]$, and therefore the effective field equation $\delta\Gamma[A;g]/\delta A(x) = 0$ for the expectation value $A \equiv \langle\hat{A}\rangle$ is given by (3.3). For any metric, this equation possesses the solution $A(x) = 0$, the vacuum case on which we focus throughout. As defined above, the matter action S vanishes on this vacuum configuration,[18] $S[0;g] = 0$, which brings us back to (5.2) for $\Gamma[0;g] \equiv \Gamma[g]$.

Writing down the corresponding field equation $\delta\Gamma[g]/\delta g_{\mu\nu} = 0$, we recover essentially the semiclassical Einstein equation (3.22), but with the expectation value $\langle T_{\mu\nu}[\hat{A};g]\rangle^g$ now replaced by the stress tensor $T_\Gamma^{\mu\nu}$:

$$T_\Gamma^{\mu\nu}[g](x) \equiv -\frac{2}{\sqrt{g}}\frac{\delta}{\delta g_{\mu\nu}(x)}\Gamma_{\text{scal}}[g]. \qquad (5.3)$$

One of the questions we are going to address is the precise relationship between these two candidates for a semiclassical stress tensor, and what the implications are for the existence of $\{\mathsf{App}(N)\}$ sequences.

A. Functional integral and measure

We start out from the general case of arbitrary d-dimensional, compact Euclidean spacetimes $(\mathcal{M}, g)$ without boundary. The functional integral representation of the induced gravity action due to a scalar field on such spacetimes reads then [39–41]

---

[18]However, had we declared (or had nature told us) that the correct description of the matter system necessitates the modified scalar action $S' \equiv S[A;g] + \Delta S[g]$, then the equation for A remains the same, while $\Delta S[g]$ makes an additional contribution to the rhs of (5.2).

$$e^{-\Gamma[g]} = \int \mathcal{D}(A;g)e^{-S[A;g]}. \qquad (5.4)$$

The measure $\mathcal{D}(A;g)$ brings in an extra metric dependence. It has the form, for any dimensionality d,

$$\mathcal{D}(A;g) = \prod_x [\det(g_{\mu\nu}(x))]^{1/4}\mu dA(x). \qquad (5.5)$$

The mass parameter μ is included here to make $\mathcal{D}(A;g)$ dimensionless.

For the purposes of this section we assume that the functional integral (5.4) has been regularized by restricting it to a finite number of spacetime points, and attaching integration variables $A(x)$ only to the sites x of, say, a lattice or a triangulation. The details of this preliminary regularization are not important though and we do not make them explicit.

(1) The integral (5.4) with (5.5) is the Euclidean analog of the quantum mechanical path integral which is strictly equivalent to applying the rules of canonical quantization. Its derivation starts out from the operatorial formalism, then uses discretization techniques to construct a Hamiltonian functional integral involving a generalized Liouville measure for paths on phase space, and finally performs the Gaussian integration over the field momenta to arrive at its Lagrangian version [39,41]. The same result is obtained by arguments based upon Becchi-Rouet-Stora-Tyutin invariance [40].

It must be emphasized that the metric dependence of the above functional measure is by no means anything "exotic." In fact, it is the path integral with just *this* measure which underlies the well-known trace-log formula abundantly used in one-loop computations.

(2) Let us now perform the integration over $A(x)$ in (5.4), thereby paying careful attention to the possibility of hidden metric dependencies. Given a fixed metric $g_{\mu\nu}$, the usual procedure consists in first diagonalizing the associated kinetic operator $\mathcal{K} \equiv \mathcal{K}[g]$, then expanding the field in terms of its eigenfunctions, $A(x) = \sum_{n,m}\alpha_{n,m}u_{n,m}(x)$, and finally changing from the integration variables $A(x)$, for all x, to the set of all $\alpha_{n,m}$. One might be suspicious that perhaps some hidden metric dependence creeps in during this procedure, in particular since the $u_{n,m}$s satisfy orthogonality and closure relations that do depend on the metric. To show that this is not the case actually we first rewrite the functional integral in terms of the new field

$$B(x) \equiv g^{1/4}(x)A(x). \qquad (5.6)$$

It transforms as a scalar density, and has the welcome property that it renders the transformed measure metric independent:

$$\int \prod_x g^{1/4}(x) \mathrm{d}A(x) \exp\left\{-\frac{1}{2}\int \mathrm{d}^dx g^{1/2} A \mathcal{K} A\right\}$$
$$= \int \prod_x \mathrm{d}B(x) \exp\left\{-\frac{1}{2}\int \mathrm{d}^dx B \widetilde{\mathcal{K}} B\right\}. \qquad (5.7)$$

In the integral over $B(x)$, the entire metric dependence resides in the new kinetic operator

$$\tilde{K} \equiv g^{1/4}(x) \mathcal{K} g^{-1/4}(x). \qquad (5.8)$$

Furthermore, we introduce a new set of basis functions $\{v_{n,m}\}$ defined as densitized versions of the $u_{n,m}$s:

$$v_{n,m}(x) \equiv g^{1/4}(x) u_{n,m}(x). \qquad (5.9)$$

According to Eqs. (3.9) and (3.10) they enjoy the properties

$$\int \mathrm{d}^dx v_{n,m}^*(x) v_{\bar{n},\bar{m}}(x) = \delta_{n\bar{n}} \delta_{m\bar{m}} \qquad (5.10)$$

$$\sum_{n,m} v_{n,m}(x) v_{n,m}^*(y) = \delta(x-y). \qquad (5.11)$$

These relations do not involve the metric any longer. Most importantly, the u's being eigenfunctions of $\mathcal{K}(g)$ implies that the v's are eigenfunctions of $\tilde{K}$, with the same eigenvalues:

$$\widetilde{K}[g] v_{n,m}(x) = \mathcal{F}_n v_{n,m}(x). \qquad (5.12)$$

Being related by a similarity transformation, $\mathcal{K}$ and $\tilde{K}$ have identical spectra.

Taking advantage of the v basis we can now perform the B integral of (5.7) in a completely clear-cut manner. After expanding the field as

$$B(x) = \sum_{n,m} b_{n,m} v_{n,m}(x) \qquad (5.13)$$

we change integration variables from $B(x)$ to $b_{n,m}$. The relations (5.10) and (5.11) imply that the corresponding Jacobian matrix $J_{xn,m} \equiv \partial B(x)/\partial b_{n,m} = v_{n,m}(x)$ is orthogonal formally, and has unit determinant therefore.[19]

---

[19]The conditions of orthogonality, $J^\dagger J = J J^\dagger = \mathbb{1}$, are easily seen to be nothing but the orthonormality and completeness relations of the v's in disguise. Suppressing the degeneracy indices for clarity, one has indeed, at the formal level, $(J^\dagger J)_{n\bar{n}} = \int \mathrm{d}^dx (J^\dagger)_{nx} J_{x\bar{n}} = \int \mathrm{d}^dx J_{xn}^* J_{x\bar{n}} = \int \mathrm{d}^dx v_n^*(x) v_{\bar{n}}(x) = \delta_{n\bar{n}} \equiv \mathbb{1}_{n\bar{n}}$, as well as $(JJ^\dagger)_{xy} = \sum_n J_{xn}(J^\dagger)_{ny} = \sum_n J_{xn} J_{yn}^* = \sum_n v_n(x) v_n^*(y) = \delta(x-y) \equiv \mathbb{1}_{xy}$.

Thus the integral (5.7) boils down to

$$\prod_{n,m} \int \mathrm{d}b_{n,m} \exp\left\{-\frac{1}{2}\sum_{n,m} \mathcal{F}_n b_{n,m}^2\right\} \qquad (5.14)$$

and hence, up to an irrelevant numerical constant,

$$\exp\{-\Gamma_{\mathrm{scal}}[g]\} = \left(\prod_{n,m} \frac{\mathcal{F}_n}{\mu^2}\right)^{-1/2}. \qquad (5.15)$$

This brings us to the (expected, of course) final result:

$$\Gamma_{\mathrm{scal}}[g] = \frac{1}{2}\sum_{n,m} \log(\mathcal{F}_n/\mu^2)$$
$$= \frac{1}{2}\sum_n D_n \log(\mathcal{F}_n/\mu^2). \qquad (5.16)$$

(3) The careful derivation we just went through highlights several points which are particularly relevant here.

(i) The representation (5.16) of the induced gravity action makes it manifest that Γ_{scal} depends on the metric exclusively via the spectral data of the operators $\tilde{K}$, or what amounts to the same, $\mathcal{K}$. Importantly, this property emerges only thanks to the presence of the explicit $g^{1/4}$ factors in the measure $\mathcal{D}(A; g)$.

(ii) It is tempting to write (5.16) in the style of an operator trace,

$$\Gamma_{\mathrm{scal}}[g] = \frac{1}{2}\mathrm{Tr}\log(\mathcal{K}/\mu^2) = \frac{1}{2}\mathrm{Tr}\log(\widetilde{K}/\mu^2). \qquad (5.17)$$

We shall refrain from this formal notation however because it tends to obscure things again.

When the lattice cutoff which we tacitly invoked up to here is lifted, the trace must be regularized in some other way, $\mathrm{Tr} \to \mathrm{Tr}_{\mathrm{reg}}$, and depending on how this is done, further, unintended metric dependencies may creep in. Moreover, with a generic regularization, "$\mathrm{Tr}_{\mathrm{reg}}$" might fail to satisfy all defining properties of a trace. If so, given the relation (5.8) between $\mathcal{K}$ and $\tilde{K}$, one may have difficulties in establishing the second equality of (5.17), or it is violated even.

Similar remarks apply to the naively equivalent $\Gamma_{\mathrm{scale}}[g] = \frac{1}{2}\log\mathrm{Det}(\mathcal{K}/\mu^2)$. Here the regularization can destroy the general properties of a determinant by multiplicative anomalies. While working at finite N such difficulties will not concern us now.

B. Induced gravity action with N cutoff

Within the framework advocated in this paper, one sidesteps the problems raised at the end of the previous subsection by thinking of $\Gamma_{\mathrm{scal}}[g]$ as a quantity whose sole input information is the spectrum of $\mathcal{K}$, being explicitly given by Eq. (5.16). To be in line with our earlier discussion

we now lift the lattice-type cutoff that was implicitly behind our derivation, and we replace it by an N cutoff.

(1) The N cutoff. At this point we return to the maximally symmetric example $\mathscr{M} = S^d(L)$ and compute the correspondingly restricted functional $\Gamma_{\text{scal}}[L^2\gamma_{\mu\nu}] \equiv \Gamma_{\text{scal}}(L, N)$ by equipping (5.16) with the same type of N cutoff as in the previous sections. Truncating the summation over the $\square_g$ eigenvalues at $n = N$, Eq. (5.16) becomes

$$\Gamma_{\text{scal}}(L, N) = \frac{1}{2}\sum_{n=1}^{N} D_n \log([\mathcal{E}_n(L) + \xi R(L) + M^2]/\mu^2).$$
(5.18)

Specifically when $d = 4$,

$$\Gamma_{\text{scal}}(L, N) = \frac{1}{2}\sum_{n=1}^{N} D_n \log\left(\frac{n(n+3) + 12\xi + (ML)^2}{(\mu L)^2}\right).$$
(5.19)

Recalling that $\sum_{n=1}^{N} D_n \equiv \mathbf{f}(N)$, it is convenient to rewrite (5.19) in the form

$$\Gamma_{\text{scal}}(L, N) = -\log(\mu L)\mathbf{f}(N) + \Delta\Gamma(ML, N) \quad (5.20)$$

where

$$\Delta\Gamma(ML, N) \equiv \frac{1}{2}\sum_{n=1}^{N} D_n \log[n(n+3) + 12\xi + (ML)^2].$$
(5.21)

Simple as it looks, the effective action (5.20) is quite remarkable and surprising. Let us specialize for massless scalars for a moment,

$$\Gamma_{\text{scal}}(L, N) = -\mathbf{f}(N)\log(\mu L) + \Delta\Gamma(0, N), \quad (M = 0).$$
(5.22)

After having set $M = 0$ in (5.21), the contribution $\Delta\Gamma(0, N)$ is seen to be perfectly *independent of L*. Thus we conclude that the *exact L* dependence of the induced action is of the form

$$c_1(N) + c_2(N)\log(L) \quad (5.23)$$

with N-dependent constants $c_{1,2}$.

Asymptotically, $c_{1,2} \propto N^4$. This L dependence is one of our main results. In particular we stress that, contrary to general expectations, no terms proportional to L^4 or L^2 are induced.

(2) $\mathscr{P}$ cutoffs. Many of the traditional calculations with a dimensionful UV cutoff at $\mathscr{P}$ would instead of (5.23) produce a structure like, omitting prefactors,

$$\mathscr{P}^4 L^4 + \mathscr{P}^2 L^2 + \cdots. \quad (5.24)$$

It descends from the first terms of the general action [42,43]

$$\Gamma_{\text{scal}}[g] = \mathscr{P}^4 \int d^4x\sqrt{g} + \mathscr{P}^2 \int d^4x\sqrt{g}R + \cdots \quad (5.25)$$

obtained from a derivative expansion, or by employing the asymptotic heat-kernel series for early proper times s and identifying $\mathscr{P}$ with $1/\sqrt{s}$ there. Flat space approaches based upon plane waves and a standard momentum cutoff $p_\mu^2 \leq \mathscr{P}^2$ also led to (5.25). While the first few terms of the series (5.25) diverge for $\mathscr{P} \to \infty$, they involve invariants already present in the Einstein-Hilbert action (possibly generalized by higher derivative terms). Hence the divergences can be absorbed by redefinitions of parameters like G and Λ_b.

(3) No quartic (quadratic) renormalization of the cosmological (Newton) constant. The potential significance of the N cutoffs advocated here is understood best by comparing (5.24) to (5.23). When the dimensionful cutoff $\mathscr{P}$ is employed, the general structure of (5.24), and more generally of $\Gamma_{\text{scal}}[g]$ in (5.25), is fixed to a very large extent by simple dimensional analysis.

When no other dimensionful parameter is available, invariants with mass dimension $-k$ cannot but get multiplied by a prefactor proportional to $\mathscr{P}^k$, whatever are the details of the concrete calculation. In particular it is unavoidable that the invariants $\int d^4x\sqrt{g}$ and $\int d^4x\sqrt{g}R$ arise multiplied by $\mathscr{P}^4$ and $\mathscr{P}^2$, respectively.

While the interpretation of these $\mathscr{P}$ dependencies is somewhat different in fundamental and effective theories, they always seem to indicate the presence of strong quantum effects that try to change the values of the cosmological and Newton's constant. However, those effects can very well be a pure artifact of the formalism employed, namely a dimensionful cutoff plus an asymptotic expansion. Being virtually unavoidable, there is no guarantee that the $\mathscr{P}^k$ dependencies reflect any real physical effect that occurs as a result of specific dynamical assumptions about the system.

N-type cutoffs, on the other hand, being dimensionless, are free from this kind of prejudice about the cutoff dependence of the invariants. Therefore it can be expected that the N dependences they give rise to are more likely to contain genuine physics information than the standard $\mathscr{P}$ dependences.

The absence of a $\mathscr{P}^4$ term in the exact result (5.23) indicates already that the cosmological constant issue will present itself differently here; we shall come back to it in more detail in Sec. V E.

C. Semiclassical stress tensors: A second candidate

Now we turn to the effective Einstein equation implied by the stationarity of $S_{\text{EH}}[g] + \Gamma_{\text{scal}}[g]$. Its traced and

integrated form with S^d metrics inserted has the same structure as in the previous section,

$$-\frac{1}{2}(d-2)R(L) + d\Lambda_{\mathrm{b}} = 8\pi G \frac{\Theta_N^{\Gamma}(L)}{\mathrm{Vol}[S^d(L)]} \qquad (5.26)$$

but now the backreaction of the quantum system is controlled by the quantity

$$\Theta_N^{\Gamma}(L) \equiv \int \mathrm{d}^d x \sqrt{g} T_{\Gamma}{}^{\mu}{}_{\mu}[g](x) \equiv \mathscr{T}\Gamma_{\mathrm{scal}}[g]. \qquad (5.27)$$

It involves the stress tensor (5.3) and the metric $g_{\mu\nu} = L^2 \gamma_{\mu\nu}$.

In Eq. (5.27) we introduced the derivative operator

$$\mathscr{T} \equiv -2 \int \mathrm{d}^d x \, g_{\mu\nu}(x) \frac{\delta}{\delta g_{\mu\nu}(x)} \qquad (5.28)$$

which we must apply to Γ_{scal} from Eq. (5.18) now. To do so we use the following handy and generally valid rule which is easily derived by a functional Taylor expansion.

Let $F \equiv F[g_{\mu\nu}]$ be an arbitrary functional and $T_F{}^{\mu\nu} \equiv -\frac{2}{\sqrt{g}} \frac{\delta F}{\delta g_{\mu\nu}}$ the associated stress tensor. Then $\mathscr{T}F[g] \equiv \int \mathrm{d}^d x T_F{}^{\mu}{}_{\mu}[g](x)$ is given by

$$\mathscr{T}F[g] = \frac{\mathrm{d}}{\mathrm{d}\alpha} F[\mathrm{e}^{-2\alpha} g_{\mu\nu}]\Big|_{\alpha=0} \qquad (5.29)$$

where α has the interpretation of a position independent Weyl parameter. Furthermore, if $F(L) \equiv F[L^2 \gamma_{\mu\nu}]$ denotes the restriction of F to metrics on S^d, the operator $\mathscr{T}$ acts on such functions of the radius according to

$$\mathscr{T}F(L) = -L \frac{\mathrm{d}}{\mathrm{d}L} F(L). \qquad (5.30)$$

Upon applying this rule to (5.18) we obtain, for any dimensionality d,

$$\Theta_N^{\Gamma}(L) = \sum_{n=1}^{N} D_n \left[1 - \frac{M^2}{\mathcal{E}_n(L) + \xi R(L) + M^2} \right]. \qquad (5.31)$$

The following points should be noted here.

(1) The arbitrary mass scale μ has dropped out from (5.31) and the effective Einstein equation.

(2) For every fixed value of N one has the following, both d and L independent limiting values for small and large masses, respectively:

$$\Theta_N^{\Gamma}(L)\big|_{M=0} = \mathbf{f}(N) \qquad (5.32)$$

$$\Theta_N^{\Gamma}(L)\big|_{M\to\infty} = 0 \qquad (5.33)$$

Obviously very heavy scalars "decouple" and do not modify Einstein's equation. This kind of decoupling did not take place with the first stress tensor candidate, see Eq. (3.37).

(3) Indeed, Θ_N^{Γ} should be contrasted with its cousin

$$\Theta_N(L) \equiv \langle \mathscr{T}S[\hat{A};g]\rangle_N^g \qquad (5.34)$$

which we computed in the previous section by straightforwardly evaluating expectation values. Comparing (5.31) to (3.32) reveals that the integrated traces differ in all dimensions by a M- and ξ-independent term:

$$\mathscr{T}\Gamma_{\mathrm{scal}}[g] - \langle \mathscr{T}S[\hat{A};g]\rangle_N^g = \frac{d}{2}\mathbf{f}(N). \qquad (5.35)$$

As we are going to demonstrate in the next subsection, this difference is due to the metric dependence of the measure $\mathcal{D}(A;g)$.

The special case $M = 0$ makes it particularly clear that the two candidates for a quantum mechanical stress tensor entail semiclassical Einstein equations with quite different properties possibly. In fact, $\Theta_N^{\Gamma}|_{M=0} = \mathbf{f}(N)$ is always positive, while $\Theta_N|_{M=0} = -\frac{1}{2}(d-2)\mathbf{f}(N)$ is negative for all $d > 2$.

D. The contribution from the functional measure

In order to understand the difference between Θ_N and Θ_N^{Γ} from first principles we return to the discretization-regularized functional integral (5.4) and its generalization for arbitrary expectation values,

$$\langle \mathcal{O}(\hat{A})\rangle^g \equiv \mathrm{e}^{+\Gamma_{\mathrm{scal}}[g]} \int \mathcal{D}(A;g)\mathcal{O}(A)\mathrm{e}^{-S[A;g]}. \qquad (5.36)$$

They are normalized such that $\langle \mathbb{1}\rangle^g = 1$ for all $g_{\mu\nu}$. (Indeed, the background metric is left arbitrary in this subsection.) Apart from the type of regularization, the expectation values evaluated in Sec. III are of this sort; in particular the 2-point function $\langle A(x)A(y)\rangle^g \equiv G(x,y)$ and the stress tensor trace $\langle \mathscr{T}S[\hat{A};g]\rangle^g \equiv \Theta_N$ are examples of (5.36).

(1) A general identity. Let us now apply the derivative operator $\mathscr{T}$, the generator of global Weyl transformations of the metric, to both sides of Eq. (5.4). We obtain

$$\mathrm{e}^{-\Gamma_{\mathrm{scal}}[g]}\mathscr{T}\Gamma[g] = \int \mathcal{D}(A;g)\mathscr{T}S[A;g]\mathrm{e}^{-S[A;g]}$$

$$- \int \mathrm{e}^{-S[A;g]}\mathscr{T}\mathcal{D}(A;g). \qquad (5.37)$$

In the last term, the $g_{\mu\nu}$ derivative acts upon the metric dependence of the measure,

$$\mathscr{T}\mathcal{D}(A;g) = \left(\prod_{x'}\mathrm{d}A(x')\right) \cdot \mathscr{T}\prod_x g^{1/4}(x). \qquad (5.38)$$

On the discrete spacetime the product over x is well defined, and applying the rule (5.29) to it yields

$$
\begin{aligned}
\mathscr{T} \prod_x g^{1/4}(x) &= \lim_{\alpha \to 0} \frac{d}{d\alpha} \prod_x \det^{1/4}(e^{-2\alpha} g_{\mu\nu}) \\
&= \lim_{\alpha \to 0} \frac{d}{d\alpha} \prod_x \{ e^{-d\alpha/2} \det^{1/4}(g_{\mu\nu}(x)) \} \\
&= \lim_{\alpha \to 0} \frac{d}{d\alpha} \exp\left\{ -\sum_x \frac{d\alpha}{2} \right\} \prod_{x'} g^{1/4}(x') \\
&= \left(-\frac{d}{2} \sum_x 1 \right) \prod_{x'} g^{1/4}(x').
\end{aligned}
\tag{5.39}
$$

Thus the response of the measure to the $\mathscr{T}$ transformation consists essentially in a multiplication by the number of spacetime points the regularized functional integral is based upon:

$$
\mathscr{T} \mathcal{D}(A; g) = \left(-\frac{d}{2} \sum_x 1 \right) \cdot \mathcal{D}(A; g).
\tag{5.40}
$$

Hence using (5.40) in (5.37) we obtain the Ward identity we wanted to derive:

$$
\mathscr{T} \Gamma_{\text{scal}}[g] - \langle \mathscr{T} S[\hat{A}; g] \rangle^g = \frac{d}{2} \left(\sum_x 1 \right).
\tag{5.41}
$$

This relationship has the same structure as the equation (5.35) which we had discovered before by an explicit calculation.

Moreover, the two relations are *strictly identical* even, since as long as both the preliminary discretization-based cutoff and the continuum N cutoff are in place, it holds true that

$$
\left(\sum_x 1 \right) = \sum_{n=1}^N D_n \equiv \mathbf{f}(N).
\tag{5.42}
$$

The argument here is the familiar one [44]: The point of contact between the discretization-based cutoff and the N cutoff is the expansion (5.13); it connects the integration variables employed by the former, namely $\{ B(x_j) | j = 1, 2, \ldots, \sum_x 1 \}$ where the x_js are the coordinates of the lattice points, to those of the latter, the expansion coefficients $\{ b_{n,m} | n = 1, \ldots, N; m = 1, \ldots, D_n \}$. Since the linear relations $B(x_j) = \sum_{n,m} b_{n,m} v_{n,m}(x_j)$, $j = 1, 2, \ldots$, establish an invertible map between the two sets of variables it is clear that they must be equal in number, and this is what proves (5.42).

(2) Inequivalent stress tensor candidates. In summary, the conclusion about the self-gravitating systems $\mathsf{App}(N)$ constructed in Sec. III and in the present section, respectively, is that they may well describe different physics since

they are based upon inequivalent stress tensors in the effective Einstein equation. The $N \to \infty$ limits of such approximants might differ correspondingly. In Sec. III we followed a kind of "bottom up" approach. It used the classical stress tensor $T^{\mu\nu} \propto \delta S / \delta g_{\mu\nu}$ as an inspiration for postulating a quantum mechanical observable $\hat{T}^\mu_\mu$, and computed its expectation value by appropriately differentiating the 2-point function. Technically this method appears closer to first quantization than to QFT.

In the present section, on the other hand, the approach is "top down," as its inspiration for the stress tensor comes from the induced gravity action of a QFT already. While only formal, upon equipping it with an N cutoff this QFT action leads to well-defined self-gravitating approximants. (The corresponding sequences will be discussed in Sec. V E below.)

It goes without saying that ultimately only experiment or additional theoretical criteria can decide about the correct stress tensor. Like in the renormalization of operator products that are plagued by short distance singularities, further input is needed.

(3) Heat-kernel regularization. In Eqs. (5.41) and (5.42) we encountered the formal sum "$\sum_x 1$" which, by our regularization, is assigned the value $\sum_{n=1}^N D_n \equiv \mathbf{f}(N)$. Concerning this crucial sum, it is instructive again to compare the N cutoff to other types of cutoffs used in the literature.

We focus on the heat-kernel cutoff here. It can arise in a variety of ways, Fujikawa's approach to anomalies[20] being a well-known example [40,45]. There are characteristic differences between those two regularization schemes which further highlight the inequivalence of N- and $\mathscr{P}$-type cutoffs.

The heat-kernel method interprets the sum $(\sum_x 1)$ as the *a priori* ill defined trace of the unit operator $\mathbb{1} = (\delta_{xy})$, and regularizes it, strictly in the continuum, by inserting a damping factor that suppresses all contributions from modes with $\mathscr{K}$ eigenvalues $\mathscr{F} \gtrsim \mathscr{P}^2$; as suggested by its name, the parameter $\mathscr{P}$ has the dimension of a mass again:

$$
\left(\sum_x 1 \right)_{\text{reg}} = \text{Tr}[\mathbb{1}]_{\text{reg}} \equiv \text{Tr}[e^{-\mathscr{K}/\mathscr{P}^2}].
\tag{5.43}
$$

What used to be a pure number, $(\sum_x 1)$, has become a functional of the metric by this regularization. Here, too, the process of removing the regulator is understood to mean letting $\mathscr{P} \to \infty$, rather than $\mathbf{f} \to \infty$. For dimensional

---

[20]While at first sight the above derivation of the measure contribution is reminiscent of Fujikawa's anomaly calculation [40], it must be stressed that, when $M = 0$, in our computation *all* contributions to the trace of the stress tensor are proportional to $\sum_x 1 = \sum_n D_n$, hence anomaly- and nonanomalous terms appear mixed. For their disentangled form, and a detailed discussion of the Weyl-Ward identities we refer to [8].

reasons the asymptotics of (5.43) has the structure, in $d = 4$,

$$\left(\sum_x 1 \right)_{\text{reg}} = \int d^4 x \sqrt{g} \{ c_4 \mathscr{P}^4 + c_2 R \mathscr{P}^2$$
$$+ (\text{curvature})^2\text{-terms} + O(\mathscr{P}^{-2}) \}. \quad (5.44)$$

Explicit calculations find for the (curvature)^2-terms the celebrated anomaly structure $a E_4 - c C_{\mu\nu\alpha\beta} C^{\mu\nu\alpha\beta}$ where E_4 is the Euler form density and $C_{\mu\nu\alpha\beta}$ the Weyl tensor [46,47]. The scheme independent coefficients a and c are known for many types of fields [46].

On spheres, Eq. (5.44) yields an L dependence of the form

$$\left(\sum_x 1 \right)_{\text{reg}} = L^4 \mathscr{P}^4 + L^2 \mathscr{P}^2 + \text{const} + O(L^{-2}). \quad (5.45)$$

Being a typical heat-kernel based result, (5.45) must be contrasted with its perfectly L-independent counterpart obtained with the N cutoff, Eq. (5.42).

E. A complete N sequence with $\Lambda_b > 0$

In this subsection we discuss the sequences $\{\mathsf{App}(N)\}$ which follow from the $4D$ effective Einstein equation

$$-\frac{12}{L^2} + 4\Lambda_b = \frac{3G}{\pi} \frac{\Theta_N^\Gamma}{L^4} \quad (5.46)$$

when the second candidate for the semiclassical stress tensor is employed; $\Theta_N^\Gamma(L)$ is given by Eq. (5.31) then. We restrict the discussion to the case $M = 0$, implying that now $\Theta_N^\Gamma = \mathbf{f}(N)$, as opposed to $\Theta_N = -\mathbf{f}(N)$ for the first candidate in $d = 4$. As a consequence, there are no solutions to (5.46) with a vanishing or a negative bare cosmological constant.

On the other hand, for every given positive $\Lambda_b \equiv 3/L_b^2$ there exists a complete sequence of self-consistent radii:

$$L^{\text{SC}}(N)^2 = \frac{3}{2\Lambda_b} \left[1 + \sqrt{1 + \frac{G\Lambda_b}{3\pi} \mathbf{f}(N)} \right]$$
$$\equiv \frac{1}{2} L_b^2 \left[1 + \sqrt{1 + \frac{G}{\pi L_b^2} \mathbf{f}(N)} \right]. \quad (5.47)$$

The sequence of self-gravitating approximants that live on the respective spacetimes $S^4(L^{\text{SC}}(N))$ has a number of remarkable properties:

(1) The systems $\mathsf{App}(N)$ exist for all $N = 0, 1, 2, \ldots$; the underlying spacetime $S^4(L^{\text{SC}}(N))$ is always non-degenerate. In particular this sequence does possess a classical initial point, $L^{\text{SC}}(0) = L_b$, contrary to the example in Sec. IV.

(2) The self-consistent radii (5.47) are a monotonically *increasing* function of N. The universe of the approximants *expands* when further quantum mechanical degrees of freedom are added.

This solution, too, disproves the prejudice underlying the cosmological constant problem which maintains that vacuum fluctuations increase the effective cosmological constant, causing the Universe to *shrink*.

The equation (5.47) becomes most transparent when $\mathbf{f} \gg 1$,

$$L^{\text{SC}}(N)^2 \approx L_b \sqrt{\frac{G}{4\pi} \mathbf{f}(N)}. \quad (5.48)$$

If we also approximate $\mathbf{f}(N) \approx \frac{1}{12} N^4$ the formula is very simple and instructive:

$$L^{\text{SC}}(N) \approx \left(\frac{G\hbar}{48\pi} \right)^{1/4} L_b^{1/2} N. \quad (5.49)$$

In writing down (5.49) we reinstated Planck's constant for a moment. We observe that the relation between L^{SC} and N becomes *linear* when $N \gg 1$, and that it depends on both G and $\hbar$ *in a nonanalytic way*. This is a clear indication of its nonperturbative origin. As the $4D$ Planck length is given by $\ell_{\text{Pl}} = (\hbar G / c^3)^{1/2}$, i.e., $\ell_{\text{Pl}} = (\hbar G)^{1/2}$ in our units, we can absorb the $\hbar$ dependence of (5.49) in a dependence on ℓ_{Pl}, yielding

$$L^{\text{SC}}(N) \approx \frac{N}{(48\pi)^{1/4}} \sqrt{\ell_{\text{Pl}} L_b}, \quad (5.50)$$

which is nonanalytic in the Planck length.

(3) If we let $N \to \infty$, the "Hubble" radius $L^{\text{SC}}(N)$ grows unboundedly, and spacetime becomes flat for every fixed value of the bare cosmological constant $\Lambda_b \equiv 3/L_b^2$:

$$S^4(L^{\text{SC}}(N)) \overset{N \to \infty}{\longrightarrow} \mathcal{R}^4. \quad (5.51)$$

We summarize this behavior as follows: The quantum field theory describing a free scalar field interacting with classical gravity is defined as the limit of a sequence of approximants $\mathsf{App}(N)$ which symbolically can be written as

$$(\mathbf{f}(N) \text{ scalar modes}) \otimes (\text{self-consistent spacetime})_N.$$
$$(5.52)$$

We advocate the point of view that the correct limit for removing the regulator is $N \to \infty$, rather than $\mathscr{P} \to \infty$ where $\mathscr{P}$ is any dimensionful cutoff scale. On the basis of the self-consistency condition considered (maximally symmetry restricted, second stress tensor) there exists only one such sequence, namely (5.47), and this sequence $\{\mathsf{App}(N)\}_{N \in \mathbb{N}}$ converges in a well-defined way to

(free scalar field, fully quantized) $\otimes \mathcal{R}^4$.   (5.53)

Thus, if one quantizes the matter field by this approach, and starts out from a classical spacetime $S^4(L_b)$ with L_b fixed, but arbitrary, the outcome will *always* be that the fully quantized scalar field lives on a flat spacetime. Or stated differently, flat space emerges without any fine-tuning.

(4) To gain further insight into the background independent treatment and its capability to produce a result diametrically opposite to the standard one, let us look at the dimensionful cutoff $\mathcal{P}$ corresponding to a given value of N:

$$\mathcal{P}(N)^2 \equiv \mathcal{E}_N(L^{\mathrm{SC}}(N)) = \frac{N(N+3)}{L^{\mathrm{SC}}(N)^2}. \quad (5.54)$$

For the solution (5.47) we obtain explicitly

$$\mathcal{P}(N)^2 = \frac{2N(N+3)}{L_b^2}\left[1 + \sqrt{1 + \frac{12\mathbf{f}(N)}{N_T^4}}\right]^{-1}, \quad (5.55)$$

with the abbreviation

$$N_T \equiv (12\pi)^{1/4}\left(\frac{L_b}{\ell_{\mathrm{Pl}}}\right)^{1/2}. \quad (5.56)$$

Therefore, when $N \gg 1$, we have in units of $m_{\mathrm{Pl}} \equiv \ell_{\mathrm{Pl}}^{-1}$,

$$\mathcal{P}(N)^2 = (24\pi)m_{\mathrm{Pl}}^2\frac{N^2}{N_T^4}\left[1 + \sqrt{1 + \left(\frac{N}{N_T}\right)^4}\right]^{-1}. \quad (5.57)$$

Figure 3 shows a graph of the function $\mathcal{P}(N)$ for $N_T \gg 1$. As expected, the dimensionful UV cutoff scale is zero at the classical point, $\mathcal{P}(N = 0) = 0$, then increases as we let $N = 1, 2, 3, \ldots$, but as soon as N comes close to the "transition" value N_T, the curve approaches a plateau and $\mathcal{P}(N)$ becomes independent of N. The plateau value in
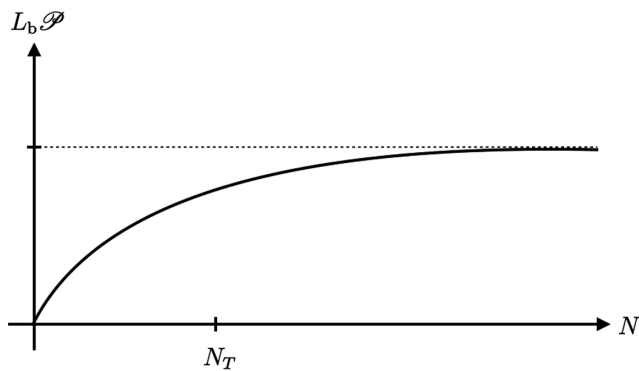

FIG. 3. The dimensionful cutoff scale $\mathcal{P}$ in dependence on N according to Eq. (5.55). The corresponding sequence of approximants assumes a positive bare cosmological constant.

units of the Planck mass in controlled by the ratio L_b/ℓ_{Pl}, or N_T equivalently:

$$\lim_{N\to\infty}\mathcal{P}(N) = (24\pi)^{1/2}\frac{m_{\mathrm{Pl}}}{N_T}. \quad (5.58)$$

Thus we find that the dimensionful UV cutoff $\mathcal{P}(N)$ never reaches infinity, for no value of N whatsoever.

This result is rather striking. It brings us to our main conclusion:

(4a) In the background independent approach the limits $N \to \infty$ and $\mathcal{P} \to \infty$ are inequivalent. Since, by construction, it is the limit $N \to \infty$ that removes the regularization, it follows that it is incorrect to attempt taking the limit $\mathcal{P} \to \infty$ when the gravitational backreaction is taken into account.

(4b) If, at the classical point $N = 0$, the bare radius is of the order of the Planck length or larger, $L_b \gtrsim \ell_{\mathrm{Pl}}$, then no member of the sequence $\{\mathsf{App}(N), 0 \le N < \infty\}$ has a proper UV cutoff larger than about the Planck scale: $\mathcal{P}(N) \lesssim m_{\mathrm{Pl}}$. [We ignore factors of order unity in (5.56) and (5.57).]

Here we observe a remarkable dynamical mechanism at work which can occur only thanks to the background independent quantization scheme: Even though the regulator is removed fully, i.e., all field modes are integrated out, no member $\mathsf{App}(N)$ in the sequence of approximating quantum systems ever encounters trans-Planckian energies or momenta: The systems dynamically adjust their metric $(g_N^{\mathrm{SC}})_{\mu\nu}$ in such a way that, with respect to this metric, the dimensionful proper cutoff corresponding to the mode cutoff at $n = N$ is always situated below the Planck scale: $\mathcal{P}(N) \lesssim m_{\mathrm{Pl}}$.

VI. MICROSTATES OF DE SITTER SPACE

The Lorentzian analog of S^4, de Sitter space, possesses a Bekenstein-Hawking entropy whose magnitude is determined by the value of the cosmological constant:

$$\mathcal{S} = \frac{3\pi}{G\Lambda}. \quad (6.1)$$

In terms of the Hubble length $L \equiv H^{-1} \equiv (3/\Lambda)^{1/2}$ it reads

$$\mathcal{S} = \frac{\pi}{G}L^2 \quad (6.2)$$

which is also equivalent to $\mathcal{S} = \mathcal{A}/4G$, where $\mathcal{A}$ denotes the area of the de Sitter horizon, $\mathcal{A} = 4\pi L^2$. Within the Euclidean approach to black holes and similar thermal spacetimes, the Euclidean 4-sphere plays an important role in the derivation of this result. It serves as a saddle point which enables the semiclassical evaluation of the $g_{\mu\nu}$ integration [48–50].

Being purely thermodynamic in nature, it is a long-standing question what are the microcopic degrees of freedom that get "counted" by this entropy, and which kind of quantum-statistical mechanics might govern a corresponding hidden layer below the (semi-)classical de Sitter spacetime [51].

The sequences of self-gravitating systems which we have constructed suggest a specific answer to this question: If we evaluate the de Sitter entropy for all members of a sequence,

$$\mathcal{S}(N) \equiv \frac{\pi}{G} L^{SC}(N)^2, \qquad (6.3)$$

it becomes obvious that the thermodynamic entropy of any member in the App(N) sequence is determined by precisely the number of degrees of freedom $\mathbf{f}(N)$ which have given birth to its particular spacetime.

To substantiate our claim, we consider the sequence obtained in Sec. IV C as an example. We assume $M = 0$ here, and so all results are independent of ξ. Using the stress tensor of the first type, we found a sequence of systems having $\Lambda_b = 0$ and, according to (4.12), self-consistent radii

$$L^{SC}(N)^2 = \frac{G}{4\pi} \mathbf{f}(N). \qquad (6.4)$$

They amount to the dynamically fixed cosmological constants

$$\Lambda^{SC}(N) \equiv \frac{3}{L^{SC}(N)^2} = \frac{12\pi}{G\mathbf{f}(N)}. \qquad (6.5)$$

(1) Quantized radii. The absolute dimensionful scale in this solution is set by the Planck length $\ell_{Pl} \equiv G^{1/2}$, which assumes the same value for all members of the sequence. Indeed, their self-consistent radii (Hubble lengths) are quantized in units of ℓ_{Pl}:

$$L^{SC}(N) = \left(\frac{\mathbf{f}(N)}{4\pi}\right)^{1/2} \ell_{Pl} = \frac{N^2}{\sqrt{48\pi}} \ell_{Pl} \left[1 + O\left(\frac{1}{N}\right)\right]. \qquad (6.6)$$

The corresponding spacetimes are well behaved for any nonzero number $\mathbf{f}$ of field modes living in the respective universe. For $\mathbf{f} = 0 = N$ its metric degenerates, and so we limit ourselves to $N \geq 1$, i.e., to universes of a genuinely quantum mechanical origin.

(2) Experiment may request $N < \infty$. As we saw, the sequence (6.4) can be significant for the cosmological constant problem since, for $N \to \infty$, the effective cosmological constant $\Lambda^{SC}(N)$ approaches zero. Could we also construct a QFT limit $N \to \infty$ with a nonzero observed cosmological constant in the traditional way by making bare parameters N dependent?

For the specific sequence $\{\text{App}(N)\}$ under consideration the answer is "no." By virtue of the universal status of G and ℓ_{Pl} in the present case, the formula (6.5) contains no bare parameter that could be given an appropriate N dependence.

Nevertheless, let us hypothesize that Eq. (6.5) represents (the maximum symmetry simplification of) a valid law of nature, and furthermore that in the real Universe cosmologists have measured a nonzero, positive cosmological constant $\Lambda_{obs} > 0$. Then the unavoidable conclusion is that the physically realized Universe carries only a finite number $\mathbf{f}(N_{obs})$ of quantum mechanical degrees of freedom rather than a full-fledged quantum field. The integer N_{obs} is fixed by the measurement then:

$$\Lambda^{SC}(N_{obs}) \overset{!}{=} \Lambda_{obs}. \qquad (6.7)$$

In this manner we allow experiment to inform us that the limit $N \to \infty$ must not be taken in this particular case.

If instead the measurement reveals that $\Lambda_{obs} = 0$, we do have to let $N \to \infty$, thus activating all field modes. It should be clear though that "$N_{obs} < \infty$" is a *physical* statement about the matter contents of the Universe. Hence the implication that the limit $N \to \infty$ must not be taken has a completely different logical status than the *mathematical inequivalence* of the limits $N \to \infty$ and $\mathcal{P} \to \infty$ discussed earlier.

In this paper we insisted repeatedly that the approximants App(N) are more than a generic regularization of a quantum field theory, namely physical systems *per se*, with finitely many degrees of freedom. One of the virtues of this requirement is that it admits the possibility that spacetime supports fewer degrees of freedom than would be supplied by a full-fledged quantum field.

Mindful of this possibility, let us now return to the entropy of de Sitter space.

(3) The entropy counts field modes. Inserting (6.4) into (6.3) we obtain the following semiclassical entropies along the sequence $\{\text{App}(N)\}$:

$$\mathcal{S}(N) = \frac{1}{4} \mathbf{f}(N). \qquad (6.8)$$

This is indeed a remarkable result. It confirms what we claimed above: For all members in the chain of systems $\{\text{App}(N)\}$, the Bekenstein-Hawking entropy of their spacetime equals (up to a factor of $1/4$) the number of scalar degrees of freedom that live on this spacetime. Hence the area of the respective Hubble spheres, $\mathcal{A}(N)$, is an integer multiple of the fundamental unit ℓ_{Pl}^2,

$$\frac{\mathcal{A}(N)}{\ell_{Pl}^2} = \mathbf{f}(N). \qquad (6.9)$$

These findings are in line with the intuitive picture that the horizon surface is a fuzzy 2-sphere made of discrete

"tiles." Points on this surface can be distinguished only if their angular separation is larger than about $\Delta\alpha \approx \pi/N$, which implies

$$\Delta\ell \equiv L^{SC}(N)\Delta\alpha \approx \sqrt{\pi/48}N\ell_{Pl}\left\{1 + O\left(\frac{1}{N}\right)\right\} \quad (6.10)$$

for their approximate proper distance.

(4) A "Λ-$\mathcal{N}$ connection" proven. Now let us assume that, as described above, experiment has provided us with some value $\Lambda_{obs} > 0$. Knowing that

$$\Lambda_{obs} = \frac{12\pi}{G\mathbf{f}(N)}, \quad (6.11)$$

this piece of data fixes a certain finite N_{obs}, which in turn allows us to infer how many degrees of freedom participated in the, purely quantum dynamical, generation of the universe the measurement of Λ_{obs} was performed in:

$$\frac{1}{4}\#(\text{degrees of freedom}) = \frac{1}{4}\mathbf{f}(N_{obs}) = \frac{3\pi}{G\Lambda_{obs}}. \quad (6.12)$$

So we have proven the following property of the de Sitter spaces governed by the N sequence under consideration: Up to the (probably inessential) factor $1/4$, universes with a strictly positive cosmological constant are described by a quantum theory with only a finite number of degrees of freedom, and this number is given by the Bekenstein-Hawking entropy of de Sitter space.

This relationship is a particular instance of the conjectured "Λ-$\mathcal{N}$ connection" and "$\mathcal{N}$-bound" that have been speculated about in the literature [52,53] under more general conditions.[21] In its stronger form, the hypothesis of the $\mathcal{N}$ bound claims that in any universe with a positive cosmological constant Λ, and arbitrary matter contents, the observed entropy $\mathcal{S}_{obs}$ is always bounded above: $\mathcal{S}_{obs} \leq 3\pi/G\Lambda \equiv \mathcal{N}$. Obviously our exact result (6.12) precisely matches this claim, and saturates the bound with $\mathcal{S}_{obs} = \frac{1}{4}\mathbf{f}(N_{obs})$ corresponding to the number $\mathcal{N}$.

(5) Our Universe. Despite its highly idealized character, it is nevertheless tempting to apply this N sequence to the real Universe we live in. If we model the latter by an empty de Sitter space with the observed Hubble radius $L^{SC}(N_{obs}) \approx 10^{60}\ell_{Pl}$, then Eq. (6.6) yields $N_{obs} \approx 10^{30}$.

[21]Originally, the $\mathcal{N}$ bound grew out of string theory based arguments which hinted at the possibility of a Λ-$\mathcal{N}$ connection such that all universes with a positive cosmological constant are governed by a fundamental quantum theory with a finite number of degrees of freedom only, whereby this number, $\mathcal{N}$, is determined by the value of Λ [52,53]. For a related discussion within loop quantum gravity we refer to [54]. Along an independent line of research, similar indications were found within the asymptotic safety approach to quantum gravity [55]. For a counting of modes on de Sitter space related to the emergent gravity paradigm see [56].

This implies $\mathcal{S}(N_{obs}) \approx 10^{120}$ and an angular uncertainty of $\delta\alpha \approx 10^{-30}$. By (6.10) this uncertainty corresponds to a proper length of about $\delta\ell \approx 10^{30}\ell_{Pl} \approx 10^{-3}$ cm at the present time. Remarkably enough, a degree of fuzzyness of the same order of magnitude has been found in [33] by logically independent arguments based upon the functional renormalization group for gravity [1].

We shall come back to the entropy of de Sitter space elsewhere [57] where we also discuss quantum corrections to its Bekenstein-Hawking value.

VII. SUMMARY AND OUTLOOK

In this paper we considered quantum fields in contact with dynamical gravity and proposed a new nonperturbative framework for the efficient investigation of such systems.

(1) We advocated the point of view that the principle of background independence should apply already to the regularized precursors of a quantum field theory, the approximants. To achieve this, we identified three requirements the corresponding calculational scheme must meet. In particular the approximants should constitute quasi-physical systems, meaning that, at the very least, they can be ascribed a well-defined, finite number of quantum mechanical degrees of freedom which are exposed to, and self-consistently backreact onto the gravitational field.

In this setting, we reinterpreted the process of removing the regularization as a comparison of such quasiphysical systems. Its limit, if it exists, generalizes the usual continuum limit in that it provides additional information about the spacetime the quantum field prefers to "live in." The spacetime geometry which is actually realized gets selected in a dynamical fashion, in harmony with background independence.

At a more technical level, we introduced what we called cutoffs of the N-type. They allowed us to concretely construct sequences of gravity-coupled approximants $\{\mathsf{App}(N)\}$ in terms of the field degrees of freedom. The rationale behind this cutoff scheme is to disentangle the logically unrelated concepts of a regularization parameter on one side, and a proper momentum scale on the other. Besides a careful metric independent enumeration of the field modes, eigenfunctions of the kinetic operator, the installation of an N cutoff requires merely a bisection of the pertinent spectrum, i.e., a rule deciding about to retain, or not to retain a given mode. Since this rule does not involve the metric, we were able to perform limits of such $\{\mathsf{App}(N)\}$ sequences which could not be considered within the standard setting.

(2) The proposed requirements *(R1,2,3)* for an efficient quantization scheme were outlined and motivated in Sec. II. Their interpretation is as follows: At a given level of technical complexity, it is the calculational schemes which do meet the requirements that have the best chances to get close to the true physical answers. In a way, the motivation

for using gravitationally backreacting approximants is similar to the (well-justified) hope that theories with a symmetry are best analyzed in a regularization scheme that respects this symmetry.

The proviso of a fixed level of complexity may be important here. While there are good reasons to believe that *(R1,2,3)* can help us finding the fastest track to the correct answers, this does not exclude the possibility that the same answers can also be found by other means, albeit at a higher price.

(3) As a first example, we applied the set of rules *(R1,2,3)* to a Gaussian scalar field. It was quantized in a universe on whose metric it was allowed to backreact self-consistently. We focussed on maximally symmetric space-times, and this enabled us to perform all subsequent calculations *exactly*. In particular we never had to invoke the short time expansion of heat-kernel traces; while this is a frequently used tool in such studies, usually it has the disadvantage of generating asymptotic series only. The exact calculations are a further source of differences between the present approach and earlier investigations.

The operator which represents the scalar's energy-momentum tensor at the quantum level is not unique. In the standard approach, products of field operators at coincident points must be given a meaning. This leads to ambiguities and requires external input over and above the classical theory. The same is true in the present approach. We employed two types of quantum stress tensors, which turned out inequivalent as for their detailed predictions.

And yet, using either of the stress tensors we found sequences of approximants which displayed the same astounding phenomenon: adding further degrees of freedom to the quantum system flattens the Universe. This is exactly the opposite of what background dependent calculations like Pauli's estimate predict, namely that further modes imply higher curvature.

We demonstrated explicitly that, as shown in Fig. 1, taking the QFT limit of infinitely many field modes, and including the quantum system's backreaction on the metric, are two noncommuting operations.

In view of these findings we believe that contrary to many claims there does not exist any cosmological constant problem due to quantum vacuum fluctuations. The false impression of a huge induced cosmological constant arises only if one takes the wrong path in the diagram of Fig. 1, namely the one which relies on the dangerous illusion that there can be a rigid spacetime that would never change, whatever load of energy and momentum is imposed on it.

(4) Our final application was to the thermodynamics of de Sitter space. We could prove a Λ-$\mathcal{N}$ connection of the kind speculated about in the literature on entirely different grounds. Computing the semiclassical Bekenstein-Hawking entropy for the approximants along one of our sequences, we found that a de Sitter spacetime with a given cosmological constant is to be identified with *one specific approximant at finite N*, rather than a limit of such. Remarkably enough, its Bekenstein-Hawking entropy turned out to count the number of field modes that are governed by this approximant, thus suggesting a natural interpretation for the microstates of de Sitter spacetime.

(5) It is clear that the present work should be extended in a number of directions. In this paper the discussion mostly focused on the approximants per se. Future work will have to find more physically realistic approximants, and to determine which sequences actually converge to quantum field theories with desirable properties. This will require including self-interactions in the matter sector [57]. The next steps also include an application to full quantum gravity, see [8] for a first investigation.

ACKNOWLEDGMENTS

[1] M. Reuter and F. Saueressig, *Quantum Gravity and the Functional Renormalization Group* (Cambridge University Press, Cambridge, England, 2019).

[2] *General Relativity and Gravitation: A Centennial Perspective*, edited by A. Ashtekar *et al.* (Cambridge University Press, Cambridge, England, 2015).

[3] A. Ashtekar, M. Reuter, and C. Rovelli, in *General Relativity and Gravitation: A Centennial Perspective*, edited by A. Ashtekar *et al.* (2015).

[4] *Canonical Gravity: From Classical to Quantum*, edited by J. Ehlers and H. Friedrich (Springer, Berlin, 1994).

[5] A. Ashtekar, Overview and outlook, in *Canonical Gravity: From Classical to Quantum*, edited by J. Ehlers and H. Friedrich (1994), p. 327.

[6] J. Ambjorn, A. Goerlich, J. Jurkiewicz, and R. Loll, Nonperturbative quantum gravity, Phys. Rep. **519**, 127 (2012).

[7] R. Percacci, *An Introduction to Covariant Quantum Gravity and Asymptotic Safety* (World Scientific, Singapore, 2017).

[8] M. Becker and M. Reuter (to be published).

[9] C. J. Isham, Prima facie questions in quantum gravity, in *Canonical Gravity: From Classical to Quantum*, edited by J. Ehlers and H. Friedrich (1994), p. 1.

[10] *Approaches to Fundamental Physics*, edited by E. Seiler and I.-O. Stamatescu (Springer, Berlin, 2007).

[11] D. Giulini, Remarks on the notions of general covariance and background independence, in *Approaches to Fundamental Physics*, edited by E. Seiler and I.-O. Stamatescu (2007), p. 105.

[12] S. Weinberg, The cosmological constant problem, Rev. Mod. Phys. **61,** 1 (1989).

[13] S. M. Carroll, The cosmological constant, Living Rev. Relativity **4,** 1 (2001).

[14] N. Straumann, The Mystery of the cosmic vacuum energy density and the accelerated expansion of the universe, Eur. J. Phys. **20,** 419 (1999).

[15] N. Straumann, The History of the cosmological constant problem, in *Proceedings of the 18th IAP Colloquium on the Nature of Dark Energy: Observational and Theoretical Results on the Accelerating Universe* (2002).

[16] N. Straumann, Dark energy, in *Approaches to Fundamental Physics*, edited by E. Seiler and I.-O. Stamatescu (2007), p. 328.

[17] T. Thiemann, *Modern Canonical Quantum General Relativity* (Cambridge University Press, Cambridge, England, 2007).

[18] R. Loll, Quantum gravity from causal dynamical triangulations: A review, Classical Quantum Gravity **37,** 013002 (2020).

[19] C. Kiefer, Why quantum gravity?, in *Approaches to Fundamental Physics*, edited by E. Seiler and I.-O. Stamatescu (2007), p. 123.

[20] B. S. DeWitt, *The Global Approach to Quantum Field Theory, Vol. 1, 2* (Oxford University Press, New York, 2003).

[21] S. Hawking, Zeta function regularization of path integrals in curved space-time, Commun. Math. Phys. **55,** 133 (1977).

[22] J. Dowker and R. Critchley, Effective Lagrangian and energy momentum tensor in de Sitter space, Phys. Rev. D **13,** 3224 (1976).

[23] A. Andrianov, L. Bonora, and R. Gamboa-Saravi, Regularized functional integral for fermions and anomalies, Phys. Rev. D **26,** 2821 (1982).

[24] A. Andrianov and L. Bonora, Finite-mode regularization of the fermion functional integral, Nucl. Phys. **B233,** 232 (1984); Finite-mode regularization of the fermion functional integral (II), Nucl. Phys. **B233,** 247 (1984).

[25] D. Verstegen, Chiral-noninvariant regularizations in continuum fermion theories, Phys. Rev. D **29,** 2405 (1984); Regularized determinants and non-perturbative definition of chiral anomalies, Nucl. Phys. **B279,** 548 (1987).

[26] S. Datta, *Quantum Phenomena* (Addison-Wesley, Reading (Mass.), 1989).

[27] I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge University Press, Cambridge, England, 1994).

[28] J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View* (Springer, New York, 1987).

[29] M. A. Rubin and C. R. Ordonez, Eigenvalues and degeneracies for n-dimensional tensor spherical harmonics, J. Math. Phys. (N.Y.) **25,** 2888 (1984).

[30] M. A. Rubin and C. R. Ordonez, Symmetric-tensor eigenspectrum of the Laplacian on n-pheres, J. Math. Phys. (N.Y.) **26,** 65 (1985).

[31] J. Madore, The fuzzy sphere, Classical Quantum Gravity **9,** 69 (1992).

[32] A. Higuchi, Symmetric tensor spherical harmonics on the N sphere and their application to the de Sitter group SO(N, 1), J. Math. Phys. (N.Y.) **28,** 1553 (1987); Erratum, J. Math. Phys. (N.Y.) **43,** 6385 (2002).

[33] M. Reuter and J.-M. Schwindt, A Minimal length from the cutoff modes in asymptotically safe quantum gravity, J. High Energy Phys. 01 (2006) 070.

[34] C. Pagani and M. Reuter, Background independent quantum field theory and gravitating vacuum fluctuations, Ann. Phys. (Amsterdam) **411,** 167972 (2019).

[35] C. Pagani and M. Reuter, Why the cosmological constant seems hardly to care about quantum vacuum fluctuations: Surprises from background independent coarse graining, Front. Phys. **8,** 214 (2020).

[36] S. Carlip, Hiding the Cosmological Constant, Phys. Rev. Lett. **123,** 131302 (2019).

[37] Q. Wang, Z. Zhu, and W. G. Unruh, How the huge energy of quantum vacuum gravitates to drive the slow accelerating expansion of the Universe, Phys. Rev. D **95,** 103504 (2017).

[38] S. Hollands and R. M. Wald, Quantum field theory is not merely quantum mechanics applied to low energy effective degrees of freedom, Gen. Relativ. Gravit. **36,** 2595 (2004).

[39] D. J. Toms, Functional measure for quantum field theory in curved spacetime, Phys. Rev. D **35,** 3796 (1987).

[40] K. Fujikawa and H. Suzuki, *Path Integrals and Quantum Anomalies* (Oxford University Press, New York, 2004).

[41] R. K. Unz, Path integration and the functional measure, Nuovo Cimento A **92,** 397 (1986).

[42] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1984).

[43] L. E. Parker and D. Toms, *Quantum Field Theory in Curved Spacetime* (Cambridge University Press, Cambridge, England, 2009).

[44] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (Dover Publications, New York, 2010).

[45] F. Bastianelli and P. van Nieuwenhuizen, *Path Integrals and Anomalies in Curved Space* (Cambridge University Press, Cambridge, England, 2006).

[46] E. Mottola, New horizons in gravity: The trace anomaly, dark energy and condensate stars, Acta Phys. Pol. B **41,** 2031 (2010).

[47] Z. Komargodski and A. Schwimmer, On renormalization group flows in four dimensions, J. High Energy Phys. 12 (2011) 099.

[48] *Euclidean Quantum Gravity*, edited by G. Gibbons and S. Hawking (World Scientific, Singapore, 1993).

[49] G. Gibbons and S. Hawking, Cosmological event horizons, thermodynamics, and particle creation, Phys. Rev. D **15,** 2738 (1977).

[50] G. Gibbons and S. Hawking, Action integrals and partition functions in quantum gravity, Phys. Rev. D **15,** 2752 (1977).

[51] C. Kiefer, *Quantum Gravity*, 3rd ed. (Oxford University Press, New York, 2012).

[52] T. Banks, Cosmological breaking of supersymmetry?, Int. J. Mod. Phys. A **16,** 910 (2001).

[53] R. Bousso, Positive vacuum energy and the N bound, J. High Energy Phys. 11 (2000) 038.

[54] L. Smolin, Quantum gravity with a positive cosmological constant, arXiv:hep-th/0209079.

[55] D. Becker and M. Reuter, Towards a *C*-function in 4D quantum gravity, J. High Energy Phys. 03 (2015) 065.

[56] T. Padmanabhan and H. Padmanabhan, Cosmic information, the cosmological constant and the amplitude of primordial perturbations, Phys. Letts. B **773,** 81 (2017).

[57] M. Becker and M. Reuter (to be published).