Extension of the quantum theory of cosmological perturbations to the Planck era

Ivan Agullo,[1,2,*] Abhay Ashtekar,[1,†] and William Nelson[1,‡]

[1]*Institute for Gravitation and the Cosmos and Physics Department, Penn State, University Park, Pennsylvania 16802, USA*
[2]*Center for Theoretical Cosmology, DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 OWA, United Kingdom*

Cosmological perturbations are generally described by quantum fields on (curved but) classical space-times. While this strategy has a large domain of validity, it can not be justified in the quantum gravity era where curvature and matter densities are of Planck scale. Using techniques from loop quantum gravity, the standard theory of cosmological perturbations is extended to overcome this limitation. The new framework sharpens conceptual issues by distinguishing between the true and apparent trans-Planckian difficulties and provides sufficient conditions under which the true difficulties can be overcome within a quantum gravity theory. In a companion paper, this framework is applied to the standard inflationary model, with interesting implications to theory as well as observations.

I. INTRODUCTION

This is the second in a series of papers whose goal is to investigate whether current scenarios of the early universe admit quantum gravity completions and, if they do, to study the implications of the resulting Planck scale dynamics. The first paper [1] summarized the underlying framework and its applications to inflation for a broad audience of theoretical physicists. This paper is primarily addressed to the quantum gravity community and provides a detailed extension of the cosmological perturbation theory to the Planck regime. Specifically, we will consider gravity coupled to a scalar field and study the dynamics of quantum fields representing scalar and tensor perturbations on *quantum* cosmological space-times. In the third paper [2], addressed to cosmologists, this framework is used to show that the inflationary scenario admits a quantum gravity extension and to analyze the physical implications of pre-inflationary dynamics, i.e., the quantum evolution from the big bounce of loop quantum cosmology (LQC) to the onset of the standard slow roll inflation. In the future, we hope to examine whether alternatives to inflation also admit viable quantum gravity completions and, if so, explore the resulting Planck scale physics.

In the theoretical explorations of the early universe, one generally uses the Friedmann, Lemaître, Robertson, Walker (FLRW) solutions to the Einstein equations (with appropriate matter sources) as background space-times. The focus is on the dynamics of quantized fields representing linearized perturbations propagating on these backgrounds. (See, e.g., Refs. [3–6].) The necessary framework of quantum theory of linear fields in curved space-times has been well developed, thanks to the ongoing research that date back to the mid-1960s. (See, e.g., Refs. [7–9]).

However, the FLRW space-times of interest are invariably incomplete in the past due to the big bang singularity where matter fields and space-time curvature diverge. It is widely believed that general relativity is simply not applicable once curvature reaches the Planck scale, whence there is no justification for using quantum field theory on solutions to Einstein's equations in this domain. Quantum gravity must intervene in an important fashion. Thus, to encompass the Planck regime, one needs a quantum gravity extension of the standard cosmological perturbation theory.

Loop quantum gravity (LQG) provides a promising avenue to meet this goal because by now the big bang singularity has been resolved in a variety of models in LQC; the $k = 0$ FLRW model on which we will focus [10–15], as well as their generalizations that include spatial curvature [16,17], a cosmological constant [18–20], anisotropies [21–24] and the simplest type of inhomogeneities and gravitational waves [25–29]. (See, e.g., Refs. [30,31] for summaries of these developments.) It is therefore natural to use LQC as the point of departure for extending the cosmological perturbation theory. However, to do so, we cannot just mimic the standard procedure used in general relativity because LQG does yet offer the quantum version of full Einstein's equations which one can linearize around a quantum FLRW space-time. Therefore, we will use the general strategy that has been repeatedly followed in LQG: *First, truncate the classical theory in a manner appropriate to the physical problem under consideration, then carry out quantization using LQG techniques, i.e., paying due attention to the underlying quantum geometry, and finally work out the physical implications of the framework.* This strategy has led to advances in the cosmological models referred to above, as well as in the investigation of quantum black holes [32–34] and the spin foam derivation of the graviton propagator [35–37].

To extend the cosmological perturbation theory, then, we will begin by focusing on the following sector of the

*I.AgulloRodenas@damtp.cam.ac.uk
†ashtekar@gravity.psu.edu
‡nelson@gravity.psu.edu

full phase space of general relativity: homogeneous, iso-
tropic configurations *together with* first order inhomoge-
neous perturbations. However, to encompass the Planck
regime, we must now use a *quantum* FLRW geometry as
background, and study the dynamics of quantum fields
representing scalar and tensor modes propagating on these
quantum geometries. Since such a quantum geometry
provides only probability amplitudes for various FLRW
metrics, we no longer have a sharply defined, proper or
conformal time. How can one then describe the dynamics
of inhomogeneous perturbations? In Ref. [38], this issue
was resolved for test quantum fields on quantum FLRW
space-times by deparametrizing the Hamiltonian con-
straint in the background, homogeneous sector. Then one
can regard the background scalar field ϕ as a relational
time variable with respect to which physical observables
evolve. This is a new conceptual element, made necessary
by quantum gravity considerations. We will use the same
strategy.

However, to encompass cosmological perturbations, we
will need three significant extensions of Ref. [38]. First,
while that work discussed a test quantum scalar field, now
the test fields include metric perturbations. Second, to
systematically arrive at the evolution equation of perturba-
tions on quantum geometry, one needs an improved strat-
egy. The third and most important difference is that,
because of its focus on conceptual issues such as the
problem of time, the analysis in Ref. [38] was restricted
only to a finite number of modes of the test scalar field, and
thus avoided the ultraviolet difficulties from the start. In
this paper, by contrast, we do not truncate the number of
modes and much of the difficult analysis is devoted to these
ultraviolet issues.

Indeed, these issues play a central role in testing self-
consistency of our procedure. The key approximation
underlying our truncation strategy is that inhomogeneities
can be regarded as perturbations—i.e., their backreaction
on the space-time geometry can be ignored. In the classical
theory, a solution obtained using this approximation is
regarded as self-consistent if the stress-energy in its inho-
mogeneities is negligible compared to that in the homoge-
neous background for the entire dynamical regime of
interest. In our quantum theory of the truncated phase
space, the situation will turn out be the following. Fix a
quantum FLRW background geometry, described by a state
Ψ_o, which evolves (with respect to the deparametrized,
internal time) via a Hamiltonian $\hat{H}_o$. The state undergoes a
quantum bounce at some time $\phi = \phi_B$. There are natural
initial conditions for quantum states ψ of perturbations at
$\phi = \phi_B$. We will show that for these states, it suffices to
focus just on the energy density, rather than the full stress-
energy. If the inhomogeneities are to be regarded as per-
turbations, we have to choose the initial ψ so that the
energy density in it is small at $\phi = \phi_B$. The question is
whether it continues to remain small in the entire duration

of evolution of interest. This is not at all guaranteed,
especially because of the Planck scale curvature during
and following the superinflation phase immediately after
the bounce. But if it does, then $\Psi_o \otimes \psi$ would be a self-
consistent solution to the truncated quantum theory.[1] Note
that the very formulation and subsequent analysis of this
self-consistency criterion requires a well-controlled defini-
tion of the Hamiltonians $\hat{H}_o$ and $\hat{H}_1$ (and the corresponding
energy density operators). Existing results on LQC already
provide a well-defined, specific $\hat{H}_o$. On the other hand, $\hat{H}_1$
is a *composite* operator on the Hilbert space of perturba-
tions; its formal expression involves products of operator
valued distributions. Therefore, an appropriate ultraviolet
regularization and renormalization is essential, first to
obtain a well-defined evolution of ψ, and second to check
that the energy density in the state ψ continues to remain
small compared to that in the background. To properly
handle this issue, we will carry over the well developed
techniques of adiabatic regularization from the quantum
field theory on classical FLRW space-times to that on
quantum FLRW space-times.

Because this article is addressed primarily to the quan-
tum gravity audience, we will make two assumptions that
will enable us to have good mathematical control without
unduly simplifying the essential conceptual underpinning.
First, we will assume that the spatial topology is that of a
3-torus $\mathbb{T}^3$ rather than $\mathbb{R}^3$. (In practice, there would be no
obvious conflicts with CMB observations if one lets the
physical radius of each of the three S^1 in $\mathbb{T}^3$ at the last
scattering surface to be greater than the known radius R_{obs}
of the observable universe at that time [39].) This assump-
tion lets us cleanly avoid spurious infinities that would
arise with an $\mathbb{R}^3$ topology simply because the background
fields are homogeneous. However, our main results extend
to the $\mathbb{R}^3$ topology and at various junctures we will indicate
how these issues are handled in that case. Second, we will
assume that there is no potential; $V(\phi) = 0$. To incorporate
inflation, one has to remove this assumption and the nec-
essary modifications are summarized in the Appendix of
Ref. [2] which discusses inflation in detail.

The paper is organized as follows. To put this work in a
broader perspective, in Sec. II we first summarize a few
procedures that have been followed in the literature to
extend the cosmological perturbation theory using LQC,
point out their merits and limitations, and then present our

---

[1]Of course, self-consistency by itself does not imply that a
truncated solution is necessarily close to an exact one. This is so
even in classical general relativity because, even if the first order
perturbation remains very small, the sum of all higher order
terms could be large. Yet, in classical cosmology and investiga-
tions of black hole perturbations, if the backreaction due to first
order perturbations remains negligible—i.e., if the test field
approximation is self consistent—the first order truncation is
generally regarded as a good approximation. We adopt the same
viewpoint in this paper.

strategy. In Sec. III we spell out the desired truncation in the classical theory. Specifically, we start with the full phase space of general relativity coupled to a scalar field, truncate the constraints to second order in inhomogeneities, and express the Hamiltonian H_1 governing dynamics of the first order perturbations in terms of those gauge invariant variables that are most convenient for passage to the quantum theory. In this passage, certain conceptual subtleties arise that affect subsequent technical results. We discuss them in some detail because, while they are well known in the discussion of perturbations of black holes, they are often overlooked in the cosmological context, especially in the LQG literature. (See the Appendix for a discussion of these issues in a simpler context of the $\lambda \phi^4$ theory in Minkowski space-time.)

In Secs. IV, V, and VI we discuss the quantum theory. The main steps involved in the construction are first collected in Sec. IV; the detailed construction of the Hilbert space of states follows in Sec. V; and the necessary regularization and renormalization of composite operators is then discussed in Sec. VI.

Physical states $\Psi(\nu, \phi; \mathcal{Q}, \mathcal{T})$ depend on $\nu \sim a^3$ (the cube of the scale factor), the background scalar field ϕ, and on the gauge invariant scalar mode $\mathcal{Q}$ and the two tensor modes $\mathcal{T}$ of inhomogeneous perturbations. As is standard in LQC, the form of (the homogeneous part of the) quantum constraint will enable us to regard ϕ as a relational time variable so that, at a fundamental level, dynamics refers to this internal or emergent time. Since the backreaction of perturbations is neglected, one can express $\Psi(\nu, \phi; \mathcal{Q}, \mathcal{T})$ as a tensor product $\Psi_o(\nu, \phi) \otimes \psi(\mathcal{Q}, \mathcal{T}, \phi)$ and study its evolution. Then, a surprising and key simplification occurs: The evolution of the quantum fields $\hat{\mathcal{Q}}, \hat{\mathcal{T}}$ on the *quantum* geometry defined by Ψ_o (in relational time ϕ) is mathematically equivalent to their evolution on an effective background metric $\tilde{g}_{ab}$ (in its conformal time $\tilde{\eta}$) where $\tilde{g}_{ab}$ is an effective metric, "dressed" with quantum corrections. Physically, while the evolution of the quantum fields $\hat{\mathcal{Q}}, \hat{\mathcal{T}}$ *is* sensitive to the quantum nature of geometry defined by Ψ_o, it does not "see" all the details of this quantum geometry: It is sensitive only to the expectation value and certain aspects of the fluctuations of the quantum metric which, it turns out, can be captured by $\tilde{g}_{ab}$. At late times $\tilde{g}_{ab}$ can be well approximated by a FLRW solution to Einstein's equation. Therefore, this passage from the relational time ϕ to $\tilde{\eta}$ also makes it manifest that, while the quantum constraint provides an evolution w.r.t. ϕ starting right from the bounce, away from the Planck regime this evolution reduces to that used in the familiar treatment [3–6] of quantum perturbations.

The key question, as we have already indicated, is whether our basic assumption that the backreaction can be neglected is borne out in the resulting solutions, especially in the Planck era. The technical discussion of Sec. VI

is devoted to this issue and leads to a sufficient condition for the solution $\Psi_o(\nu, \phi) \otimes \psi(\mathcal{Q}, \mathcal{T}, \phi)$ to be self-consistent within the truncation approximation. In the next paper [2] we will show that for a large class of initial conditions, this criterion is met in presence of a quadratic potential $V(\phi)$ all the way from the bounce to the onset of slow roll inflation. In Sec. VII we summarize the main results and discuss various issues, including the choice of initial conditions.

Our conventions are as follows. The space-time metric will be assumed to have signature $-, +, +, +$. We set $c = 1$ but keep (G and) $\hbar$ explicitly in various equations to facilitate the distinction between classical and quantum effects. Finally, as in the quantum gravity literature, we will use Planck (rather than the reduced Planck) units. (Thus, our $\ell_{\mathrm{Pl}} = \sqrt{\hbar G}$ and our Planck mass $m_{\mathrm{Pl}} = \sqrt{\hbar/G}$ is related to the reduced planck mass M_{Pl} via $m_{\mathrm{Pl}} = \sqrt{8\pi} M_{\mathrm{Pl}}$.)

II. STRATEGIES INSPIRED BY LQC

The singularity resolution in LQC has motivated a large number of investigations aimed at incorporating the underlying quantum geometry effects in the standard cosmological paradigms. In this section we will first briefly summarize the main strategies and then explain the approach that is followed in the rest of the paper. This concise summary should help nonexperts to see the inter-relation (or lack thereof) between the rich set of ideas that are being pursued in the LQG literature. For experts, it should help clarify some subtle issues that have not been emphasized in the literature and also bring out the continuity and coherence of the approach used here.

A. A brief summary

Recall first that in the FLRW models the LQC quantum states of interest remain sharply peaked along the bouncing solution of effective equations that incorporate leading order quantum corrections. This surprising behavior was first encountered in numerical simulations but subsequent work led to an analytic understanding through a change of representation [14], use of coherent states [40] and the WKB approximation [41]. Early attempts sought to exploit this property to incorporate quantum gravity corrections right before the onset of the slow roll inflation [42], or, for the evolution of perturbations during inflation (see, e.g., Ref. [43]). However, to incorporate these corrections to the evolution, the LQC effective equations for the FLRW background do not suffice; one also needs the LQG modified perturbation equations. Since a well established set of quantum Einstein's equations is not yet available in full LQG, the strategy was to simply use the first order perturbation equations from general relativity, the FLRW background space-time being simply replaced by the effective solution in LQC. Unfortunately, this procedure

is conceptually unsatisfactory because the background LQC space-time is no longer a solution to Einstein's equations. Indeed, now there is an ambiguity in what one means by "linearized Einstein's equations:" Two sets which are equivalent when the background is a solution to the exact Einstein's equations are generically no longer equivalent if the background is a solution to a different set of equations. To our knowledge, a systematic procedure to handle this ambiguity was not part of the general strategy.

Another set of papers, geared to capture the "inverse triad (or inverse volume) corrections" of LQC, is based on the notion of "lattice refinement" (see, e.g., Refs. [44–46]). In the homogeneous, spatially compact case, these corrections are meaningful and physically interesting. In the spatially noncompact case, typically the corrections depend on the choice of a fiducial cell used as an infrared regulator [13] and disappear when the regulator is removed [31]. In more recent versions, one considers $\mathbb{R}^3$ spatial topology but decomposes the spatial manifold in elementary cells and approximates the inhomogeneous configurations of physical interest by configurations which are homogeneous within any one cell but vary from one cell to another. Physically, this is an attractive strategy. However it requires a fresh input—that of the cell size (or "lattice spacing")—and the inverse volume effects are now sensitive to this new scale: The fiducial cell of the homogeneous model is in effect replaced by a physical cell in which the Universe can be taken to be homogeneous. However, so far there is neither a theoretical principle nor observational guidance on what this scale should be during the inflationary epoch when perturbations are generated in these schemes. On the other hand, the more striking predictions of these frameworks—such as enhancement of quantum gravity signature by several orders of magnitude over the factor $H^2/m_{\rm Pl}^2 \sim 10^{-11}$ one would expect during inflation—appear to depend on the choice of the new scale one makes. Therefore, although the underlying idea of lattice refinement is attractive, at present there appears to be an inherent ambiguity in the size and importance of the inverse volume effects in this setting. Overcoming these limitations is an interesting prospect for future work.

In these investigations the focus was on quantum gravity effects during inflation. Some of the more recent investigations have recognized that, because the energy density is $\sim 10^{-11}\rho_{\rm Pl}$ during this era, quantum gravity corrections during inflation would be too small to be observable in the foreseeable future and shifted the emphasis to studying quantum gravity corrections from the bounce to the onset of the slow roll. Many of them employ a new strategy that goes under the broad theme of "anomaly cancelation" [47–53]. This analysis is based on the Hamiltonian framework and focuses on the constraint algebra. The idea is to arrive at the desired LQG theory of cosmological perturbations by studying the permissible modifications of the constraints of general relativity that are to encode quantum

corrections. Recall that two key features of general relativity are: (i) the Poisson algebra of constraints closes, and (ii) the evolution is generated by these constraints. Keeping these features in mind, one proceeds in the following steps: (i) One assumes that the "effective theory" that incorporates the LQG corrections would have the same phase space for the homogeneous isotropic background and first order perturbations; (ii) one allows for a modified set of constraints on this phase space by making an ansatz for possible modifications; (iii) one calculates the undetermined functions of the background geometry in the ansatz by *requiring* that the constraint algebra should again close; and, finally, (iv) one defines the desired effective dynamics as the Hamiltonian flow of the new, modified constraints. In the quantum theory, if the commutator algebra of constraints did not close, there would be an anomaly. Therefore, the method is called anomaly cancelation even though one is dealing only with Poisson brackets.

Until recently, the modifications that were arrived at did not change the structure functions in the constraint algebra of general relativity. In a recent work on scalar perturbations [53], on the other hand, the structure function in the Poisson bracket between two Hamiltonian constraint is modified by a function of the ratio $\rho/2\rho_{\rm max}$ where ρ is the matter density in the background and $\rho_{\rm max}$ the maximum density in FLRW LQC. This has been interpreted to mean that there is a change in the space-time signature at the end of the superinflation phase of the background, so that the signature is Euclidean to the past of this event [54].

The central idea in this anomaly cancelation strategy is potentially powerful in constraining the type of quantum corrections cosmological perturbation theory can inherit from *any* consistent quantum gravity theory. However, its implementation has some puzzling aspects.

First, the form of permissible modifications of the constraint functionals is chosen primarily for mathematical convenience and not derived systematically from general physical principles. Second, the phase space of the quantum corrected theory is assumed to be the same as that in classical general relativity while typically, quantum corrections add higher derivative terms to the action which significantly enlarge the phase space. Third, since the phase space is kept unchanged, the general analysis due to Hojman, Kuchar and Teitelboim [55] is applicable and their results bring out a puzzling feature of this framework. Hojman *et al.* began with four-dimensional, globally hyperbolic space-times (M, g_{ab}) with nonspinorial matter and, using embeddings of a 3-manifold $\mathbb{M}$ as a Cauchy surface in M, constructed a "hypersurface deformation" algebra $\mathcal{D}$ on the space $\mathcal{E}$ of embeddings which encodes space-time covariance. This is a purely geometric construction without any reference to field equations. Then they showed that, on the standard phase space of general relativity (coupled to

the matter under consideration) based on $\mathbb{M}$, there is only one way to represent the algebra $\mathcal{D}$ by canonical transformations in a time reversible manner: the representation given by the Poisson algebra of the *standard* Einstein constraint functionals. For the anomaly cancelation program, this implies that if the modification of constraint functionals is genuine (i.e., not just a field redefinition), then the modified Hamiltonian theory will not have a consistent space-time interpretation. Therefore, it would not be possible to associate a well-defined space-time metric to (portions of) dynamical trajectories in this modified Hamiltonian theory, let alone examine its signature. Fourth, even if one were to ignore this point, signature change is such a drastic effect that it seems difficult to justify the validity of the first order cosmological perturbation theory in the subsequent treatment. The fifth and a more "global" limitation arises from the fact that the conceptual underpinning of the overall strategy is rather unclear in some of the recent applications. Effective equations are meant to incorporate all quantum corrections. Therefore, one would have thought that the dynamical equations derived from them would already contain quantum effects. Yet, in some works, these fields are quantized again to obtain the power spectrum for scalar, vector and tensor modes, as in the standard treatment of cosmological perturbations on classical FLRW backgrounds in general relativity. The overall logic underlying this scheme is thus rather puzzling. Indeed, already in the homogeneous sector the logical procedure is the opposite: one first obtains the quantum evolution equations and *then* derives effective equations from them using dynamics of appropriate, sharply peaked coherent states.

To summarize, investigations to date have provided useful mathematical infrastructure (see, e.g., Ref. [47]) and in some cases also the much needed qualitative insights into mechanisms through which quantum gravity effects could provide corrections to the standard inflationary scenario (see, e.g., Refs. [51–53,56]). The viewpoint that appears to have emerged from the anomaly cancelation strategy—namely, quantum effects could be neatly encoded in the Hamiltonian framework but would lead to a fuzziness if the phase space trajectories are interpreted as four-dimensional space-time geometries—could well be an imprint of some deep result on the nature of space-time in LQG. However, it seems fair to say that, at the current level of understanding, this and other strategies used so far also have a number of puzzling features and face conceptual limitations in their treatment of inhomogeneous perturbations.

B. Strategy used in this paper

In this subsection we will outline the avenue used in this paper to extend the standard theory of cosmological perturbations all the way to the Planck regime. As explained in Sec. I, the main idea is simply to use the strategy that has driven LQG so far: Construct the Hamiltonian framework of the sector of general relativity of interest and then pass to the quantum theory using quantum geometry that underlies LQG. In order to bring out differences from other approaches, and to address questions regarding truncation and gauge choices that are sometimes raised, we will now spell out this strategy using illustrations where is has already been successfully used.

The FLRW models provide the simplest illustration. Here, one starts out with the full phase space Γ of general relativity (coupled to matter) in the connection dynamics framework [57,58] and truncates it to its homogeneous, isotropic sector Γ_{HI}. By fixing gauge, one coordinatizes the gravitational part of Γ_{HI} simply with a pair (ν, b) of real numbers, where (in appropriate units) ν denotes the physical volume of the Universe (or, if the spatial topology is noncompact, of a fiducial cell) and b is its conjugate momentum. This gauge fixing leads us to the *classically reduced phase space* with respect to the Gauss and the diffeomorphism constraints. Therefore, the reduced phase space carries only the Hamiltonian constraint, expressed in terms of just ν, b and matter variables. This provides the starting point for quantization a la LQG. In the first step, a specific kinetic framework [59] can be selected by a uniqueness theorem [60] along the lines of those in full LQG [61,62]. By representing curvature in the Hamiltonian constraint by holonomies around plaquettes selected by the underlying quantum geometry, one constructs the quantum Hamiltonian constraint operator [13,31]. In this construction there are, as is usual in any quantization procedure, some factor ordering ambiguities. [Compare, e.g., the Hamiltonian constraint in Ref. [13] (which is geared to proper time) with that in Ref. [14] (geared to harmonic time). See also Ref. [63].] However, these affect only details of the quantum evolution; general features are robust. Finally, the effective equations that encode leading quantum effects are systematically derived starting from the quantum theory [40].

In the generalization to the homogeneous but anisotropic (i.e., Bianchi) models, one again follows the same conceptual procedure: appropriate truncation of the phase space Γ to Γ_{hom}, eliminating the Gauss and diffeomorphism constraints by passing to the *classically reduced phase space*, and passing to quantum theory via LQG techniques [21,23,24]. The singularity resolution persists. Furthermore, there is a detailed consistency check: by tracing over the anisotropy degrees of freedom of the Bianchi I model one recovers the quantum Hamiltonian constraint of the FLRW model [21]. As one would expect, in presence of anisotropies, the quantum dynamics is significantly richer [64–66]. Note that the starting point in this analysis is again a truncated phase space Γ_{hom} of general relativity, but the truncation is now

enlarged to incorporate physics of interest, namely anisotropies.

The last example, Gowdy models, [25–29] illustrate our strategy most closely because now the truncated phase space Γ_{Gowdy} is an *infinite* dimensional subspace of Γ: the model allows for inhomogeneities induced by a class of (fully nonlinear) gravitational waves, in addition to those in the matter fields. The constraints are now rewritten in terms of an infinite number of conveniently chosen "modes" of canonical variables coordinatizing Γ_{Gowdy}, obtained by appropriate gauge fixing. Thanks to this choice, we are led to a *classical reduction of the phase space* with respect to the Gauss and the purely inhomogeneous parts of the diffeomorphism and Hamiltonian constraints. Thus, one is left with only global Hamiltonian and diffeomorphism constraints (corresponding to a homogeneous lapse and shift). One then passes to quantum theory using a "hybrid" scheme where one employs LQC quantum kinematics for homogeneous modes and a Fock-type quantum kinematics for the inhomogeneous modes representing gravitational waves. This is an internally consistent quantization. One finds that, thanks to the quantum geometry effects, the singularity of general relativity is resolved. This is in striking contrast to the early attempts predating LQC, where singularity could not be resolved (see, e.g., Refs. [67–70]). It is interesting to note that, if one "switches off" inhomogeneities, Gowdy models reduce to Bianchi I models. Using effective equations, it has been shown that the general dynamical behavior of the Bianchi I models—including the bounces—carries over to the Gowdy models [26,29]. This analysis has also provided valuable information on the changes in the amplitudes of gravitational waves resulting from the bounce. Finally note that, although one obtains the Bianchi I model by switching off inhomogeneities, in contrast to some of the strategies summarized in Sec. II A, one *does not* begin with the inhomogeneous modes propagating on an effective, quantum corrected Bianchi I background and *then* quantize them. Rather, one truncates the full phase space Γ to Γ_{Gowdy} and *quantizes the full Gowdy model, which includes both homogeneous and inhomogeneous modes.*

To develop a quantum gravity theory of cosmological perturbations, then, we will continue along the same path that has been successfully used so far. Our first task is to identify the appropriate truncation of the classical phase space Γ. By introducing a fiducial flat triad for mathematical convenience, we can decompose fields into Fourier modes. With this convenient coordinatization, constraints of the full theory can be expressed in terms of modes. We will start with a homogeneous background and expand out the deviations from it as a sum of first, second, ... nth, ... order terms in inhomogeneities. The truncation will now consist of keeping terms which depend on the background and are at most quadratic in the first order perturbations.

In general relativity, this truncation enables one to study dynamics of the background homogeneous space-time and that of linearized perturbations propagating on these backgrounds. Backreaction of these first order perturbations on the background (which is encoded in the second order perturbations) is neglected.[2]

The idea again is to use LQG techniques to pass to the quantum theory, using a hybrid scheme that broadly mimics the one used in the Gowdy models. (As we will see in subsequent sections, some differences arise because, unlike in the Gowdy model, we are now dealing with *linear perturbations*.) It will again be possible to interpret the homogeneous quantum Hamiltonian constraint as providing "evolution" of the quantum state of the background geometry in the relational time variable—the scalar field. Solutions to this equation provide background *quantum* geometries on which quantum perturbations evolve. In this conceptual setting, one does *not* start out with classical perturbations on an effective, smooth FLRW space-time and *then* quantize them. The passage to quantum theory is carried out in one go for the full truncated phase space that includes both the background and perturbations. Finally, as emphasized in Sec. I, much of the technical discussion will be devoted to finding a criterion to test whether the final theory admits self-consistent solutions that justify the viability of the underlying truncation scheme.

This overall strategy was briefly reported in Refs. [71,72]. At the same time the "hybrid approach" was used in Ref. [73] to study cosmological perturbations in the $k = 1$ FLRW context. The main conceptual difference between that approach and ours is that we are able to go beyond formal considerations in the quantum theory by exploiting the relation between quantum fields on quantum geometry and those on a dressed effective geometry discussed in Sec. IV B.

III. TRUNCATED HAMILTONIAN FRAMEWORK

In cosmology of the very early universe, one generally restricts oneself to the sector of general relativity consisting of homogeneous FLRW space-times together with linear perturbations thereon. In much of the cosmology literature one works with the *solution space* of this truncated theory. However, as pointed out in Ref. [74], the task of finding gauge invariant variables is more streamlined in the Hamiltonian framework. More importantly, since we are now interested in treating the background geometry quantum mechanically, a natural avenue is to follow the Dirac quantization procedure based on phase space. Therefore, in this section we will first construct the truncated phase space and then discuss dynamics thereon.

---

[2]There are some subtleties in this procedure that are sometimes overlooked. For a discussion in a simpler example, see the Appendix.

This will provide a natural starting point to apply the well established LQG techniques in the next section.

A. The phase space

Let us begin with general relativity coupled to a scalar field on a space-time manifold $M = \mathbb{M} \times \mathbb{R}$, where $\mathbb{M}$ is topologically $\mathbb{T}^3$. For completeness and continuity with the LQC literature, we will begin with the connection variables [57] and then pass to the Arnowitt Deser Misner (ADM) variables for perturbations that are more commonly used in the cosmological literature.

Let us first focus on geometry. Let q_{ab} denote positive definite metrics on $\mathbb{M}$, e_i^a and ω_a^i, orthonormal frames and coframes with respect to q_{ab}, and let K_{ab} denote the extrinsic curvature on $\mathbb{M}$. In connection dynamics, the canonically conjugate pair consists of real SU(2) connections A_a^i and su(2) valued vector densities E_i^a of weight 1, both defined on $\mathbb{M}$. [Here indices $a, b, c, \ldots$ refer to the tangent space of $\mathbb{M}$ and $i, j, k, \ldots$ to the Lie algebra]. They are related to the ADM variables via

$$E_i^a = \sqrt{q}\, e_i^a \quad \text{and} \quad A_a^i = \Gamma_a^i + \gamma K_a^i, \qquad (3.1)$$

where q denotes the determinant of q_{ab}, Γ_a^i, the spin connection determined by e_i^a, $K_a^i = K_{ab} e^{bi}$ and γ, the Barbero-Immirzi parameter of LQG. Since we are interested in the spatially flat FLRW background geometries, as usual it is convenient to introduce some fiducial structures. Fix a flat metric $\mathring{q}_{ab}$, an orthonormal frame $\mathring{e}_i^a$, and the corresponding coframe $\mathring{\omega}_a^i$ on $\mathbb{M}$. We will denote by $\mathring{q}$ the determinant of $\mathring{q}_{ab}$ and assume that each of the circles in $\mathbb{M}$ has length ℓ with respect to $\mathring{q}_{ab}$ so that the fiducial volume of $\mathbb{M}$ is $\mathring{V} = \ell^3$. The natural "Cartesian" coordinates defined on $\mathbb{M}$ by $\mathring{q}_{ab}$ will be denote by $\vec{x} \equiv (x_1, x_2, x_3)$.

Even though we are still considering full general relativity, it is convenient to decompose the basic canonically conjugate fields into purely homogeneous and purely inhomogeneous parts:

$$A_a^i(\vec{x}) = c\ell^{-1}\mathring{\omega}_a^i + \mathfrak{a}_a^i(\vec{x}) \quad \text{and}$$

$$E_i^a(\vec{x}) = \sqrt{\mathring{q}}(p\ell^{-2}\mathring{e}_i^a + \mathfrak{e}_i^a(\vec{x})), \qquad (3.2)$$

where $\int(A_a^i \mathring{e}_j^a)d\mathring{v} =: c\ell^2 \delta_i^j$ and $\int(E_i^a \mathring{\omega}_a^j)d^3\vec{x} =: p\ell\delta_i^j$ so that $\mathfrak{a}_a^i$ and $\mathfrak{e}_i^a$ are *purely inhomogeneous*:

$$\int \mathfrak{a}_a^i \mathring{e}_j^a d\mathring{v} = 0, \quad \text{and} \quad \int \mathfrak{e}_i^a \mathring{\omega}_a^j d\mathring{v} = 0. \qquad (3.3)$$

(Here and in what follows, unless otherwise specified, the integrals are over $\mathbb{M}$ and $d\mathring{v}$ denotes the volume element on it with respect to $\mathring{q}_{ab}$.) Thus, the geometrical part of the phase space is naturally coordinatized by quadruples $(c, p; \mathfrak{a}_a^i(\vec{x}), \mathfrak{e}_i^a(\vec{x}))$ where c, p are real numbers and $\mathfrak{a}_a^i, \mathfrak{e}_i^a$ are purely inhomogeneous fields. The matter field can also be decomposed into purely homogeneous and purely inhomogeneous parts:

$$\Phi(\vec{x}) = \phi + \varphi(\vec{x}) \quad \text{and} \quad \Pi(\vec{x}) = \sqrt{\mathring{q}}(\ell^{-3}p_{(\phi)} + \pi(\vec{x})). \qquad (3.4)$$

The symplectic structure on the total phase space Γ is given by

$$\Omega(\delta_1, \delta_2) = \frac{3}{\kappa\gamma}[\delta_1 c \delta_2 p - \delta_2 c \delta_1 p]$$

$$+ \int(\delta_1 \mathfrak{a}_a^i \delta_2 \mathfrak{e}_i^a - \delta_2 \mathfrak{a}_a^i \delta_1 \mathfrak{e}_i^a)d\mathring{v}$$

$$+ \delta_1 \phi \delta_2 p_{(\phi)} - \delta_2 \phi \delta_1 p_{(\phi)}$$

$$+ \int(\delta_1 \varphi \delta_2 \pi - \delta_2 \pi \delta_1 \pi)d\mathring{v}, \qquad (3.5)$$

where $\delta \equiv (\delta c, \delta p, \delta \mathfrak{a}_a^i, \delta \mathfrak{e}_i^a; \delta\phi, \delta p_{(\phi)}, \delta\varphi, \delta\pi)$ denote tangent vectors to Γ and $\kappa = 8\pi G$. Thus, the only nonzero Poisson brackets between the basic variables are

$$\{c, p\} = \frac{\kappa\gamma}{3}, \qquad \{\mathfrak{a}_a^i(\vec{x}_1), \mathfrak{e}_j^b(\vec{x}_2)\} = \delta_j^i \delta_a^b \bar{\delta}(\vec{x}_1, \vec{x}_2),$$

$$\{\phi, p_{(\phi)}\} = 1, \qquad \{\varphi(\vec{x}_1), \pi(\vec{x}_2)\} = \bar{\delta}(\vec{x}_1, \vec{x}_2), \qquad (3.6)$$

where $\bar{\delta}(\vec{x}_1, \vec{x}_2) = \left((1/\sqrt{\mathring{q}})\delta(\vec{x}_1, \vec{x}_2) - (1/\ell^3)\right)$ is the Dirac delta distribution on the space of purely inhomogeneous fields.[3] Thus, the total phase space (Γ, Ω) has a product structure: $\Gamma = \Gamma_{\text{hom}} \times \Gamma_{\text{inh}}$, $\Omega = \Omega_{\text{hom}} + \Omega_{\text{inh}}$.

There are three sets of first class constraints:

$$\mathbb{G}[\Lambda] = \int \Lambda^i(\vec{x})G_i(\vec{x})d^3\vec{x} = \int \Lambda^i(\vec{x})D_a E_i^a(\vec{x})d^3\vec{x}$$

$$\mathbb{V}[\vec{N}] = \int N^a(\vec{x})V_a(\vec{x})d^3\vec{x}, \qquad \mathbb{S}[N] = \int N(\vec{x})S(\vec{x})d^3\vec{x}. \qquad (3.7)$$

Here, the vector and the scalar constraints are smeared by a shift $\vec{N}$ and a lapse N and the Gauss constraint—which we have explicitly written out because it does not feature in the ADM framework—by a generator Λ^i of su(2). The Gauss and the vector constraints generate, respectively, internal SU(2) rotations and spatial diffeomorphism, both of which are generally regarded as "kinematic motions." Dynamics is generated by the Hamiltonian constraint.

Remark: If the spatial topology is $\mathbb{R}^3$ rather than $\mathbb{T}^3$, the symplectic structure on the homogeneous subspace induced by that of full general relativity diverges. Therefore, to obtain a consistent phase space description, one has to introduce a cubical fiducial cell $\mathcal{C}$ aligned with these axes and with edge-length ℓ and restrict integrations

---

[3]The $1/\ell^3$ factor ensures that the Poisson brackets are compatible with the fact that the perturbations are purely inhomogeneous, i.e., satisfy (3.3).

of homogeneous fields to it. This is an infrared cutoff, to be removed, as usual, in the *final* results (see, e.g., Ref. [31]).

B. Expansions around the FLRW subspace

In physical cosmology one is primarily interested in a neighborhood of the four-dimensional, homogeneous subspace Γ_{hom} of the full phase space Γ. To fix notation for the rest of the paper, we begin with certain expansions of fields near homogeneity. While this technique is well known in the perturbation theory around black holes, it appears not to be as familiar in the cosmology literature.

Consider curves $\gamma[\epsilon]$ in Γ parametrized by ϵ, with $\epsilon \in]-1, 1[$, say, which pass through Γ_{hom} at $\epsilon = 0$:

$$A_a^i[\epsilon](\vec{x}) = c\ell^{-1}\overset{\circ}{\omega}_a^i + \epsilon\mathfrak{a}_a^{(1)i}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\mathfrak{a}_a^{(n)i}(\vec{x}) + \cdots$$

$$E_i^a[\epsilon](\vec{x}) = \sqrt{\overset{\circ}{q}}\left[p\ell^{-2}\overset{\circ}{e}_i^a + \epsilon\mathfrak{e}_i^{(1)a}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\mathfrak{e}_i^{(n)a}(\vec{x}) + \cdots\right]$$

$$\Phi[\epsilon](\vec{x}) = \phi + \epsilon\varphi^{(1)}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\varphi^{(n)}(\vec{x}) + \cdots$$

$$\Pi[\epsilon](\vec{x}) = \sqrt{\overset{\circ}{q}}\left[\ell^{-3}p_{(\phi)} + \epsilon\pi^{(1)}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\pi^{(n)}(\vec{x}) + \cdots\right],$$
$$(3.8)$$

where $\mathfrak{a}_a^{(1)i}$, $\mathfrak{e}_i^{(1)a}$, $\varphi^{(1)}$, $\pi^{(1)}$ are *purely inhomogeneous* tangent vectors to the curves $\gamma[\epsilon]$ pointing away from Γ_{hom}. Here ϵ is a mathematical parameter that keeps track of the "order" of the perturbation (it could be taken to be $\sqrt{\kappa}$ but for simplicity we will assume it to be dimensionless here). Geometrically, keeping only the first order perturbations corresponds to considering the normal bundle over the homogeneous subspace Γ_{hom} of Γ (since purely inhomogeneous tangent vectors are "orthogonal to" Γ_{hom} in the L^2-norm). Retaining terms only up to the nth order corresponds to considering the nth order, inhomogeneous jet bundle on Γ_{hom}. (The Appendix summarizes the meaning and utility of this expansion procedure using the simpler example of the $\lambda\Phi^4$ theory.)

Of special interest are the curves $\gamma[\epsilon]$ that lie in the constraint hypersurface of Γ. We will use a collective label $C(A, E, \Phi, \Pi)$ for the smeared constraint functions (suppressing the smearing fields for simplicity). Then along each curve $\gamma[\epsilon]$ the constraints become $C[\epsilon] = 0$. By Taylor expanding in ϵ we obtain a hierarchy of equations:

$$C|_{\epsilon=0} = 0, \qquad \frac{dC}{d\epsilon}\bigg|_{\epsilon=0} = 0, \ldots \frac{d^nC}{d\epsilon^n}\bigg|_{\epsilon=0} = 0, \ldots \quad (3.9)$$

The zeroth-order equation in the hierarchy, $C|_{\epsilon=0} = 0$ is just the restriction of the full constraint to the homogeneous subspace Γ_{hom}. As noted above, because of gauge fixing, the Gauss and the vector constraints vanish identically and we are left with only one nontrivial constraint smeared with a homogeneous lapse [31]:

$$\mathbb{S}_o[N_{\text{hom}}] := N_{\text{hom}}\left[-\frac{3}{\kappa\gamma^2}c^2|p|^{\frac{1}{2}} + \frac{p_{(\phi)}^2}{2|p|^{\frac{3}{2}}}\right] = 0. \quad (3.10)$$

(As noted in Sec. I, in this paper we have set the potential $V(\Phi)$ of the scalar field to zero. See the Appendix in Ref. [2] for inclusion of the potential.) As usual, although the lapse is homogeneous, it can depend on dynamical variables; see Sec. III D for a further discussion. Equation (3.10) is a nonlinear but *algebraic* equation constraining the homogeneous fields.

The first order equation, $[dC/d\epsilon]|_{\epsilon=0} = 0$, is a linear partial differential equation (PDE) for $\mathfrak{a}^{(1)i}_a(\vec{x})$, $\mathfrak{e}^{(1)a}_i(\vec{x})$, $\varphi^{(1)}(\vec{x})$, $\pi^{(1)}(\vec{x})$. For example, the Gauss constraint yields:

$$\int\Lambda_{\text{inh}}^b[\ell\partial^a\mathfrak{e}^{(1)}_{ab} + c\overset{\circ}{\epsilon}_b{}^{mn}\mathfrak{e}^{(1)}_{mn} - p\ell^{-1}\overset{\circ}{\epsilon}_b{}^{mn}\mathfrak{a}^{(1)}_{mn}]d\overset{\circ}{v} = 0,$$
$$(3.11)$$

where we have converted the internal indices $i, j, \ldots$ into tangent space indices $a, b \ldots$ using the *fiducial* frame $\overset{\circ}{e}_i^a$ and coframe $\overset{\circ}{\omega}_a^i$ and indices are raised and lowered with the fiducial metric $\overset{\circ}{q}_{ab}$. Only purely inhomogeneous smearing fields Λ_{inh}^b contribute because the first order perturbations are all purely inhomogeneous. This equation involves $\mathfrak{a}^{(1)}_{ab}$, $\mathfrak{e}^{(1)}_{ab}$ linearly but also contains the background variables c, p which solve the zeroth order constraint (3.10). In this equation (c, p) happen to enter linearly. But in general the coefficients can be complicated nonlinear functions of the background fields. Thus, for example, quadratic combinations appear in the first order vector constraint,

$$\int N_{\text{inh}}^a[\ell p(\partial^b\mathfrak{a}^{(1)}_{ab} - \partial_a\mathfrak{a}^{(1)}) - cp\overset{\circ}{\epsilon}_b{}^{mn}\mathfrak{a}^{(1)}_{mn}$$
$$- c^2\ell\overset{\circ}{\epsilon}_a{}^{mn}\mathfrak{e}^{(1)}_{mn} - \kappa\gamma p_{(\phi)}\partial_a\varphi^{(1)}]d\overset{\circ}{v} = 0, \quad (3.12)$$

where $\mathfrak{a} = \mathfrak{a}_{ab}\overset{\circ}{q}^{ab}$ is the trace of $\mathfrak{a}_{ab}$. Note that the first order constraints are linear "homogeneous" PDEs in the sense that there are no (zeroth order) source terms on the right-hand side. The structure of the first order scalar constraint is the same.

In the second order equations on the other hand are "inhomogeneous" as PDEs since the zeroth and first order fields now act as "sources." These equations determine the "Coulombic" parts of the second order fields in terms of the zeroth and first order ones. For example, the second order Gauss constraint is given by

$$\int\Lambda^b[\ell\partial^a\mathfrak{e}^{(2)}_{ab} + c\overset{\circ}{e}_b{}^{mn}\mathfrak{e}^{(2)}_{mn} - p\ell^{-1}\overset{\circ}{\epsilon}_b{}^{mn}\mathfrak{a}^{(2)}_{mn}]d\overset{\circ}{v}$$
$$= -\int\Lambda^a[\ell\overset{\circ}{\epsilon}_a{}^{bc}\mathfrak{a}^{(1)}_{db}\mathfrak{e}^{(1)d}{}_c]d\overset{\circ}{v}. \quad (3.13)$$

While this is again a linear PDE—in fact the operator on the left side is the same as that in (3.11), but now acts on the second order fields, $(\mathfrak{a}^{(2)i}_a, \mathfrak{e}^{(2)a}_i)$—there is now a source

term which is *quadratic* in the first order fields which are already known as solutions to (3.11).[4] The structure of the second order vector and the scalar constraints is the same. Finally, note that the *n*th order equations in this hierarchy *constrain only the nth order fields*; they do not impose further conditions on lower order fields. This pattern continues to all orders. In particular, to obtain the full set of constraints on the zeroth and the first order fields, we need to solve only $C|_{\epsilon=0} = 0$ and $[dC/d\varphi]|_{\epsilon=0} = 0$. This property will be important in what follows.

The key point about the hierarchy is that it greatly simplifies the problem of solving the complicated, nonlinear PDEs $C(A, E; \Phi, \Pi) = 0$. For $n = 0$ we obtain a nonlinear but *algebraic* equation. For $n > 0$, each equation in the hierarchy is a *linear* PDE for the *n*th order fields, with the same linear differential operator on the left side but order dependent source terms which are already determined by solutions to the *lower order* equations in the hierarchy. The value of the system lies in the hope that by truncating it to a low order, one would obtain a good approximate solution to the full system with small inhomogeneities. (For further discussion, see the Appendix.) Although it is not easy to rigorously control the approximation, this truncation scheme has proved to be a valuable and indispensable tool in cosmological and black hole sectors of classical general relativity. In practice, the domain of validity of the chosen truncation is checked by *self-consistency*: one only verifies that, if the truncation is of order n, then the source terms in equations governing $(n + 1)$st order fields are negligibly small compared to the fields that are kept.

What is the relation between this hierarchy of constraints and dynamics? On the full phase space (Γ, Ω), dynamics is generated by constraints. In particular, in the homogeneous sector, the scalar constraint smeared with a homogeneous lapse can be thought of as generating "pure time evolution." What happens if we truncate the theory at first order? Then, the truncated phase space Γ_{Trun} will be the normal bundle over Γ_{hom}. The Hamiltonian flow generated by $\mathbb{S}[N_{\text{hom}}]$ on Γ is tangential to its homogeneous subspace Γ_{hom}. It suffices to consider this flow in an arbitrarily small neighborhood of Γ_{hom}. Under this flow, tangent vectors $v \equiv (\mathfrak{a}^{(1)i}{}_a, \mathfrak{e}^{(1)a}{}_i, \varphi^{(1)}, \pi^{(1)})$ at any point on Γ_{hom} also have an unambiguous evolution. Thus, given a specific v_o at a point, say $\gamma(t_o)$, of any dynamical trajectory $\gamma(t)$ on Γ_{hom}, the trajectory $\gamma(t)$ can be unambiguously lifted to a trajectory in Γ_{Trun}, passing through $(\gamma(t_o), v_o)$. We will see in Sec. III D that this lift has a simple geometrical interpretation in the phase space.

Remark: In the LQC literature, in place of the ϵ expansion (3.8) one often uses a decomposition of the type $A^i_a = c\ell^{-1}\omega^i_a + \delta A^i_a$, etc. Then the "perturbations" $\delta A^i_a, \dots$ include terms of all orders $n \geq 1$ in our ϵ expansion. Therefore, in light of footnote 4, it is not consistent to assume that these perturbations $\delta A^i_a, \dots$ are purely inhomogeneous. On the other hand, if one allows them to have homogeneous parts, then the Poisson brackets between the unperturbed and perturbed fields do not all vanish and so the symplectic structure is more complicated. This complication is often overlooked. In constraint equations, keeping terms linear in these fields is interpreted as the first order truncation, keeping terms quadratic as second order truncation, etc. In this scheme, the second and higher order truncation would lead to *nonlinear* PDEs on perturbations $\delta A^i_a, \dots$; the simplification that was achieved in the ϵ-expansion (3.8) would be lost. Truncations in the two schemes are equivalent only to the linear order. These considerations apply also to the Hamiltonian treatment of cosmological perturbations in terms of metric variables, used outside the LQC literature as well.

C. First order truncation

Since the temperature fluctuations in the cosmic microwave background are only one part in 10^5, much of the literature on the early universe has focused on FLRW backgrounds with *first order* linear perturbations thereon. We will now analyze this truncation in some detail. The relevant phase space is the normal bundle $\Gamma_{\text{Trun}} = \Gamma_o \times \Gamma_1$ where $\Gamma_o = \Gamma_{\text{hom}}$ is the four-dimensional homogeneous subspace of the full phase space Γ and Γ_1 is spanned by the first order fields in the expansion (3.8). It turns out that the description of the FLRW *quantum* geometries is easier in terms of variables (b, ν), rather than the original (c, p):

$$b = \frac{c}{|p|^{1/2}}, \qquad \nu = \frac{4|p|^{3/2}}{\kappa\gamma\hbar}\, \text{sgn}\, p, \quad \text{so that}\ \{b, \nu\} = \frac{2}{\hbar}.$$
$$(3.14)$$

The geometrical meaning of these variables is as follows[5]: The physical volume of the Universe is given by $a^3\ell^3 = 2\pi\gamma|\nu|\ell_{\text{Pl}}^2$, where a is the scale factor, and on any solution b equals the Hubble parameter modulo a multiplicative constant, $b = \gamma(\dot{a}/a)$. Thus, we will now coordinatize Γ_o by $(\nu, b; \phi, p_{(\phi)})$. For simplicity of notation, *from now on we will drop the suffix* (1) *on the first order perturbations*. Thus,

$$(\nu, b, \phi, p_{(\phi)}; \mathfrak{a}_{ab}, \mathfrak{e}_{ab}, \varphi, \pi) \in \Gamma_{\text{Trun}}. \qquad (3.15)$$

The perturbations Poisson commute with the background fields and the Poisson brackets among themselves are given

---

[4]Because the source terms on the right side of (3.13) are quadratic in first order perturbations, it follows that the integral on the right side does not necessarily vanish if the smearing field Λ^a is purely homogeneous. The left side of (3.13) now implies that the second (and higher order terms) in our ϵ-expansion (3.8) *cannot* be assumed to be purely inhomogeneous.

[5]Like p, the variable ν takes values in the entire real line, positive values corresponding to triads e^a_i with the same orientation as the fiducial $\mathring{e}^a_i$, and negative values corresponding to triads with opposite orientation.

by (3.6) so that $\Omega_{\text{Trun}} = \Omega_o + \Omega_1$. Thus, mathematically, $(\Gamma_{\text{Trun}}, \Omega_{\text{Trun}}) = (\Gamma, \Omega)$. *But the physical interpretation* of $\alpha_a^i, e_a^i, \varphi, \pi$ is different: in Γ_{Trun} they represent only the first order perturbations, i.e., just the coefficients of ϵ in the expansion (3.8). As a result, the constraints and dynamical equations they satisfy are very different from those of the full theory.

As noted in Sec. II B, to zeroth order we only have the scalar constraint, smeared with a homogeneous lapse. In terms of ν, b variables it becomes

$$\mathbb{S}_o[N_{\text{hom}}] = N_{\text{hom}}\left[-\frac{3\hbar}{4\gamma}\,b^2\nu + \frac{2p_{(\phi)}^2}{\kappa\gamma\hbar\nu}\right] = 0. \quad (3.16)$$

To first order in ϵ, we obtain linear equations on first order fields with functions of b, ν, $p_{(\phi)}$, ϕ as coefficients. Let us first focus on the Gauss constraint (3.11):

$$\int \Lambda_{\text{inh}}^b[\ell\partial^a e^{(1)}{}_{ab} + c\mathring{\epsilon}_b{}^{mn}e^{(1)}{}_{mn} - p\ell^{-1}\mathring{\epsilon}_b{}^{mn}\alpha^{(1)}{}_{mn}]d\mathring{\nu} = 0. \quad (3.17)$$

It generates the following infinitesimal gauge transformations:

$$\alpha_{ab} \rightarrow \alpha_{ab} - \partial_a\Lambda_b + c\ell^{-1}\mathring{\epsilon}_{abm}\Lambda^m,$$
$$e_{ab} \rightarrow e_{ab} + p\ell^{-2}\mathring{\epsilon}_{abm}\Lambda^m. \quad (3.18)$$

Thus, the symmetric part of e_{ab} is gauge invariant and furthermore it has a simple interpretation. Since $E_i^a E^{bi} = qq^{ab}$ in the full theory, where q^{ab} is the physical 3-metric on $\mathbb{M}$ and q its determinant, it follows that

$$e_{(ab)} = -\frac{\ell^2}{2p}(\mathfrak{q}_{ab} - \mathfrak{q}q_{ab}^{(o)}), \quad \text{where}$$
$$q_{ab} = q_{ab}^{(o)} + \epsilon\mathfrak{q}_{ab} + O(\epsilon^2). \quad (3.19)$$

[Note that $q_{ab}^{(o)} = (a^2)\mathring{q}_{ab}$ is the zeroth order *physical* metric. It is purely homogeneous while $\mathfrak{q}_{ab}$ is purely inhomogeneous. $\mathfrak{q} := \mathfrak{q}_{ab}\mathring{q}^{ab}$.] Next, recall from (3.1) that $A_a^i(\epsilon) = \Gamma_a^i(\epsilon) - \gamma K_a^i(\epsilon)$. Linearizing this equation one finds that the first order Gauss constraint serves only to determine the skew symmetric part of

$$\mathfrak{K}_{ab} := \mathring{\omega}_{bi}\frac{d}{d\epsilon}K_a^i(\epsilon)|_{\epsilon=0} \quad (3.20)$$

in terms of e_{ab} and the zeroth order fields. Thus imposition of this constraint implies that $\mathfrak{K}_{(ab)}$ is the free part of $\mathfrak{K}_{ab}$. Furthermore it is gauge invariant. Hence the reduced phase space with respect to the first order Gauss constraint (3.11) is coordinatized simply by the linearized metric $\mathfrak{q}_{ab}$ and the linearized extrinsic curvature $\mathfrak{K}_{ab}$, both of which are symmetric tensors. The three Gauss constraint have removed three degrees of freedom from each of the two first order fields (α_a^i, e_i^a), taking us from the linearized connection variables to the linearized ADM ones, $(\mathfrak{q}_{ab}, \mathfrak{K}_{ab})$. This is

precisely the structure one expects from the Gauss reduction of the full theory [57].

It now remains to impose the first order vector and scalar constraints on the linearized pairs $(\mathfrak{q}_{ab}, \mathfrak{K}_{ab})$. For this we can draw on the huge body of existing literature. Let us begin by fixing the notation. The ADM variables are $(q_{ab}, p^{ab} = \sqrt{q}(K_{ab} - Kq_{ab}))$. To first order, they can be expanded as

$$q_{ab}(\epsilon) = q_{ab}^{(o)} + \epsilon\mathfrak{q}_{ab} + \cdots \quad \text{and}$$
$$p^{ab}(\epsilon) = \sqrt{\mathring{q}}(p^{(o)ab} + \epsilon\mathfrak{p}^{ab} + \cdots), \quad (3.21)$$

where the zeroth order fields are given by $q_{ab}^{(o)} = a^2\mathring{q}_{ab}$ and $p^{(o)ab} = -(ab/\kappa\gamma)\mathring{q}^{ab}$. As in the cosmology literature, we can expand out the first order fields using Fourier transforms:

$$\mathfrak{q}_{ab}(\vec{x}) = \frac{1}{\mathring{V}}\sum_{\vec{k}\in\mathcal{L}}\tilde{\mathfrak{q}}_{ab}(\vec{k})e^{i\vec{k}\cdot\vec{x}} \qquad \varphi(\vec{x}) = \frac{1}{\mathring{V}}\sum_{\vec{k}\in\mathcal{L}}\varphi_{\vec{k}}e^{i\vec{k}\cdot\vec{x}}$$

$$\mathfrak{p}_{ab}(\vec{x}) = \frac{1}{\mathring{V}}\sum_{\vec{k}\in\mathcal{L}}\tilde{\mathfrak{p}}_{ab}(\vec{k})e^{i\vec{k}\cdot\vec{x}} \qquad \pi(\vec{x}) = \frac{1}{\mathring{V}}\sum_{\vec{k}\in\mathcal{L}}\pi_{\vec{k}}e^{i\vec{k}\cdot\vec{x}}, \quad (3.22)$$

where $\mathcal{L}$ is the lattice defined by $\vec{k} \in ((2\pi/\ell)\mathbb{Z})^3, \vec{k} \neq \vec{0}, \mathbb{Z}$ being the set of integers. (The zero $\vec{k}$ is excluded because, by construction, the fields are all purely inhomogeneous.) Since all four fields in the $\vec{x}$ space are real, their Fourier transforms satisfy the relations $\tilde{\mathfrak{q}}_{ab}(\vec{k}) = \tilde{\mathfrak{q}}_{ab}^\star(-\vec{k})$, etc. It is convenient to expand out the metric perturbations into scalar, vector and tensor modes [75]:

$$\tilde{\mathfrak{q}}_{ab}(\vec{k}) = S_{\vec{k}}^{(1)}\mathring{q}_{ab} + S_{\vec{k}}^{(2)}\left(\hat{k}_a\hat{k}_b - \frac{1}{3}\mathring{q}_{ab}\right) + \sqrt{2}V_{\vec{k}}^{(1)}\hat{k}_{(a}\hat{x}_{b)}$$
$$+ \sqrt{2}V_{\vec{k}}^{(2)}\hat{k}_{(a}\hat{y}_{b)} + \frac{1}{\sqrt{2}}T_{\vec{k}}^{(1)}(\hat{x}_a\hat{x}_b - \hat{y}_a\hat{y}_b)$$
$$+ \sqrt{2}T_{\vec{k}}^{(2)}(\hat{x}_{(a}\hat{y}_{b)}), \quad (3.23)$$

and

$$\tilde{\mathfrak{p}}_{ab}(\vec{k}) = \frac{1}{3}\mathfrak{p}_{\vec{k}}^{(S_1)}\mathring{q}_{ab} + \frac{1}{2}\mathfrak{p}_{\vec{k}}^{(S_2)}(3\hat{k}_a\hat{k}_b - \mathring{q}_{ab})$$
$$+ \sqrt{2}\mathfrak{p}_{\vec{k}}^{(V_1)}\hat{k}_{(a}\hat{x}_{b)} + \sqrt{2}\mathfrak{p}_{\vec{k}}^{(V_2)}\hat{k}_{(a}\hat{y}_{b)}$$
$$+ \frac{1}{\sqrt{2}}\mathfrak{p}_{\vec{k}}^{(T_1)}(\hat{x}_a\hat{x}_b - \hat{y}_a\hat{y}_b) + \sqrt{2}\mathfrak{p}_{\vec{k}}^{(T_2)}(\hat{x}_{(a}\hat{y}_{b)}). \quad (3.24)$$

(Here $\hat{k}$ is a unit vector in the $\vec{k}$ direction and $\hat{k}$, $\hat{x}$, $\hat{y}$ constitutes a field of orthonormal triads with respect to $\mathring{q}_{ab}$ in the momentum space. Throughout, indices are lowered, raised and contracted using $\mathring{q}_{ab}$.) Then, the canonically conjugate pairs are $(S_{\vec{k}}^{(1)}, \mathfrak{p}_{-\vec{k}}^{S_1}), (S_{\vec{k}}^{(2)}, \mathfrak{p}_{-\vec{k}}^{S_2}), \ldots(\varphi_{\vec{k}}, \pi_{-\vec{k}})$, with Poisson brackets $\{S_{\vec{k}}^{(1)}, \mathfrak{p}_{-\vec{k}'}^{S_1}\} = \ell^3\delta_{\vec{k},\vec{k}'}, \ldots$. These

fields are subject to three vector constraints and a scalar constraint, each of which is linear in these fields but also contains background fields. As is well known, one can pass to the reduced, truncated phase space $\tilde{\Gamma}^{(1)}_{\text{Trun}}$ by solving them and finding gauge invariant variables [74,76]:

$$\tilde{\Gamma}^{(1)}_{\text{Trun}} = \{(\mathcal{Q}_{\vec{k}}, T^{(1)}_{\vec{k}}, T^{(2)}_{\vec{k}}; \mathfrak{p}^{(Q)}_{\vec{k}}, \mathfrak{p}^{(T_1)}_{\vec{k}}, \mathfrak{p}^{(T_2)}_{\vec{k}})\}. \qquad (3.25)$$

Here $T^{(1)}_{\vec{k}}$ and $T^{(2)}_{\vec{k}}$ are the two tensor modes (which are automatically gauge invariant) and $\mathcal{Q}_{\vec{k}}$, the gauge invariant Mukhanov-Sasaki variable, is given by

$$\mathcal{Q}_{\vec{k}} = \varphi_{\vec{k}} - \frac{p_{(\phi)}\gamma}{2a^5 \ell^3 \mathrm{b}}\left(S^{(1)}_{\vec{k}} - \frac{1}{3}S^{(2)}_{\vec{k}}\right). \qquad (3.26)$$

Its conjugate momentum is given by

$$\mathfrak{p}^{Q}_{\vec{k}} = \pi_{\vec{k}} + \frac{\kappa\gamma p^2_{(\phi)}}{2\ell^6 a^3 \mathrm{b}}\varphi_{\vec{k}} - \frac{\kappa\gamma^2 p^3_{(\phi)}}{4\ell^9 a^8 \mathrm{b}^2}S^{(1)}_{\vec{k}}$$
$$- \left(\frac{p_{(\phi)}}{2a^2\ell^3} - \frac{\kappa\gamma^2 p^3_{(\phi)}}{12\ell^9 a^8 \mathrm{b}^2}\right)S^{(2)}_{\vec{k}}. \qquad (3.27)$$

The initial configuration variable $\mathfrak{q}_{ab}$ had six degrees of freedom and the scalar field φ, one. Each of the four linearized constraints reduces the configuration degrees of freedom by one, leaving us with 3 degrees of freedom $(\mathcal{Q}, T^{(1)}, T^{(2)})$ in the reduced configuration space.

Let us summarize. Note first that in the passage to the reduced phase space $\tilde{\Gamma}_{\text{Trun}}$, we used only the kinematical structure on Γ_{Trun}; *dynamical equations were not used.* Thus, the procedure is completely analogous to that followed in other LQC models to extract the physical degrees of freedom (see II B). $\tilde{\Gamma}_{\text{Trun}}$ has the form

$$\tilde{\Gamma}_{\text{Trun}} = \Gamma_o \times \tilde{\Gamma}^{(1)}, \quad \text{with} \; (\nu, \mathrm{b}; \phi, p_{(\phi)}) \in \Gamma_o,$$
$$(\mathcal{Q}_{\vec{k}}, T^{(1)}_{\vec{k}}, T^{(2)}_{\vec{k}}; \mathfrak{p}^{Q}_{\vec{k}}, \mathfrak{p}^{T_1}_{\vec{k}}, \mathfrak{p}^{T_2}_{\vec{k}}) \in \tilde{\Gamma}^{(1)}. \qquad (3.28)$$

These variables are subject to only the zeroth order scalar constraint:

$$\mathbb{S}_o[N_{\text{hom}}] = N_{\text{hom}}\left[-\frac{3\hbar}{4\gamma}\mathrm{b}^2\nu + \frac{2p^2_{(\phi)}}{\kappa\gamma\hbar\nu}\right] = 0. \qquad (3.29)$$

We have already taken care of the first order constraints. As noted above, the second and higher order constraints do not restrict the first order variables.

Remarks:

(1) In the CMB, we can only observe modes up to a (finite) maximum wave length λ_o which equals the radius of the observable universe at the surface of last scattering. Therefore, it is physically appropriate to absorb modes with wave-lengths $\lambda \gtrsim 10\lambda_o$ in the homogeneous background. This amounts to putting a *physical* infrared cutoff on perturbative modes, making the arbitrariness in the choice of the radius of the 3-torus $\mathbb{T}^3$ irrelevant. In phenomenological applications, these considerations have to

be folded into the calculation of the renormalized energy density and checking self-consistency.

(2) We use the Mukhanov-Sasaki variable $\mathcal{Q}_{\vec{k}}$—rather than the curvature perturbation $\mathcal{R}_{\vec{k}}$—to coordinatize the reduced phase space because it is better suited for our discussion of inflation in Ref. [2]. There, we find that $\mathcal{R}_{\vec{k}}$ is not well-defined along effective trajectories because it carries $p_{(\phi)}$ in the denominator and $p_{(\phi)}$ vanishes at "turning points" of the inflaton. $\mathcal{Q}_{\vec{k}}$ on the other hand is well-defined all along these trajectories. In absence of a potential $V(\phi)$, as in this paper, they are related just by a constant: $\mathcal{R}_{\vec{k}} = \sqrt{\kappa/6}\,\mathcal{Q}_{\vec{k}}$.

D. Truncated dynamics

As mentioned in Sec. III A, to obtain dynamical trajectories on $\tilde{\Gamma}_{\text{Trun}}$ one has to proceed in two steps: First, obtain a dynamical trajectory on Γ_o and then lift it to $\tilde{\Gamma}_{\text{Trun}}$. On Γ_o, dynamics is generated by the scalar constraint $\mathbb{S}[N_{\text{hom}}]$, i.e., the Hamiltonian vector field is just the restriction to Γ_o of the full Hamiltonian vector field[6] $X^\alpha = \Omega^{\alpha\beta}\partial_\beta \mathbb{S}[N_{\text{hom}}]$ on Γ: $X^\alpha|_{\Gamma_o} = \Omega^{\alpha\beta}_o \partial_\beta \mathbb{S}_o[N_{\text{hom}}]$. Full dynamics on Γ unambiguously induces a flow on $\tilde{\Gamma}_{\text{Trun}}$. The general procedure for nonlinear systems, summarized in the Appendix, now implies that the time evolution on the truncated phase space Γ_{Trun} is given by the dynamical vector field X^α_{Dyn},

$$X^\alpha_{\text{Dyn}} = \Omega^{\alpha\beta}_o \partial_\beta \mathbb{S}_o[N_{\text{hom}}] + \Omega^{\alpha\beta}_1 \partial_\beta \mathbb{S}'_2[N_{\text{hom}}], \qquad (3.30)$$

where $\mathbb{S}'_2$ is obtained from the coefficient $\mathbb{S}_2[N_{\text{hom}}]$ of ϵ^2 in the expansion of the full scalar constraint $\mathbb{S}[N_{\text{hom}}]$ on Γ, by keeping just those terms which are quadratic in the first order quantities $(\mathfrak{a}^{(1)i}_a, \mathrm{e}^{(1)a}{}_i, \varphi^{(1)}, \pi^1)$ and discarding terms linear in the second order quantities $(\mathfrak{a}^{(2)i}_a, \mathrm{e}^{(2)a}{}_i, \varphi^{(2)}, \pi^{(2)})$. It is important to note that it is only $\mathbb{S}_2[N_{\text{hom}}]$ that is constrained to vanish; *there is no constraint on $\mathbb{S}'_2[N_{\text{hom}}]$.* Finally, because $\mathbb{S}'_2$ depends on the background fields, X^α_{Dyn} is not generated by $\mathbb{S}_o + \mathbb{S}'_2$; i.e., $X^\alpha_{\text{Dyn}} \neq \Omega^{\alpha\beta}\partial_\beta(\mathbb{S}_o + \mathbb{S}'_2)$. But one might wonder if one can find another "effective Hamiltonian" that generates this dynamical flow. The answer is in the negative: One can check that X^α_{Dyn} *does not Lie drag the total symplectic structure* $\Omega = \Omega_o + \Omega_1$ *on the truncated phase space*, whence it is impossible to find an effective Hamiltonian.

Remark: We have spelled out these results because there has been some confusion in the recent LQC literature. *Quantum* dynamics is often assumed to be captured by a quantum constraint $\hat{\underline{\mathsf{C}}}_H\Psi = 0$. Our detailed discussion of

---

[6]Here the Greek indices refer to the (infinite dimensional) tangent space to the full or truncated phase space. They are to be regarded as abstract indices a la Penrose. This index notation can be avoided; it is used only as a pedagogical aid.

truncation and dynamics of the truncated system implies that this assumption cannot be justified. First, as Eq. (3.30) shows, the dynamical vector field $X^\alpha_{\rm Dyn}$ on $\Gamma_{\rm Trun}$ is not generated by a constraint (or indeed by *any* Hamiltonian). Therefore, there is no reason to expect that the correct quantum dynamics could be recovered by imposing any quantum constraint. Second, even the part $\Omega^{\alpha\beta}_1 \partial_\beta \mathbb{S}'_2[N_{\rm hom}]$ of $X^\alpha_{\rm Dyn}$ describing the evolution of perturbations is *not* generated by the second order constraint $\mathbb{S}_2[N_{\rm hom}]$ but by $\mathbb{S}'_2[N_{\rm hom}]$, which is unconstrained.

Finally, we are interested only in the dynamics of gauge invariant variables, i.e., only in dynamics on $\tilde{\Gamma}_{\rm Trun}$. As one would expect, the functions $\mathbb{S}_o$ and $\mathbb{S}'_2$ that determine X^α are gauge invariant and therefore projects down to $\tilde{\Gamma}_{\rm Trun}$. In particular, $\mathbb{S}'_2 = \mathbb{S}'^{(\mathcal{Q})}_2 + \mathbb{S}'^{(T_1)}_2 + \mathbb{S}'^{(T_2)}_2$ and the form of the last two terms is exactly the same. *Therefore, from now on, we will denote the two tensor modes collectively by* $\mathcal{T}$. Then [74],

$$\mathbb{S}'^{(\mathcal{Q})}_2[N_{\rm hom}] = \frac{N_{\rm hom}}{2\mathring{V}} \sum_{\vec{k}} \frac{1}{a^3} |\mathfrak{p}^{(\mathcal{Q})}_{\vec{k}}|^2 + ak^2 |\mathcal{Q}_{\vec{k}}|^2$$

$$\mathbb{S}'^{(\mathcal{T})}_2[N_{\rm hom}] = \frac{N_{\rm hom}}{2\mathring{V}} \sum_{\vec{k}} \frac{4\kappa}{a^3} |\mathfrak{p}^{(\mathcal{T})}_{\vec{k}}|^2 + \frac{ak^2}{4\kappa} |\mathcal{T}_{\vec{k}}|^2. \quad (3.31)$$

Note that if we were to replace $2\sqrt{\kappa}\mathcal{Q}_k$ by $\mathcal{T}_k$ (and the corresponding momenta by a reciprocal factor to maintain the Poisson brackets), $\mathbb{S}'^{(\mathcal{Q})}_2[N_{\rm hom}]$ reduces to $\mathbb{S}'^{(\mathcal{T})}_2[N_{\rm hom}]$. *Therefore, it suffices to focus just on (one of the two tensor modes)* $\mathcal{T}_{\vec{k}}$ *and its conjugate momentum which we will denote simply by* $\mathfrak{p}_{\vec{k}}$.

The resulting dynamical evolution can be best understood in geometric terms as follows. Note first that $\tilde{\Gamma}_{\rm Trun}$ is naturally a bundle over the homogeneous phase space Γ_o. Fix an integral curve of X^α on the "base space" Γ_o, i.e., a homogeneous solution. Fix a point $\gamma_o \equiv (\nu^o, \mathrm{b}^o; \phi^o, p^o_{(\phi)})$ on this trajectory and a tangent vector $\mathcal{T}^o_{\vec{k}}, \mathfrak{p}^o_{\vec{k}}$ at that point. To describe the evolution of this perturbation on the chosen background trajectory, we need to lift the trajectory from Γ_o to $\tilde{\Gamma}_{\rm Trun}$. This can be done simply by considering the integral curve of $X^\alpha_{\rm Dyn}$ passing through the point $(\nu^o, \mathrm{b}^o; \phi^o, p^o_{(\phi)}; \mathcal{T}^o_{\vec{k}}, \mathfrak{p}^o_{\vec{k}})$ of $\tilde{\Gamma}_{\rm Trun}$.

Next, let us list the commonly used lapse functions $N_{\rm hom}$ and the corresponding time variables:

(i) $N_{\rm hom} = 1$ corresponds to proper time t so that the physical space-time metric has the form $ds^2 = -dt^2 + a^2 d\vec{x}^2$.

(ii) $N_{\rm hom} = a$ corresponds to conformal time η so that the physical space-time metric has the form $ds^2 = a^2(-d\eta^2 + d\vec{x}^2)$. This is the most common choice in the cosmology literature.

(iii) $N_{\rm hom} = a^3$ corresponds to harmonic time τ which satisfies the wave equation, $\Box \tau = 0$. The physical space-time metric now assumes the form

$ds^2 = -a^6 d\tau^2 + a^2 d\vec{x}^2$. The zeroth order scalar constraint $\mathbb{S}_o[N_{\rm hom}]$ takes the simplest form with this choice. Therefore, harmonic time is commonly used in LQC not only for the $k = 0$, $\Lambda = 0$ FLRW cosmology but also for models that admit spatial curvature, a nonzero cosmological constant and anisotropies [14,21,23,24,31].

(iv) $N = (\mathring{V}/p_{(\phi)})a^3$ corresponds to choosing the scalar field ϕ itself as time. It turns out that irrespective of the initial choice of lapse in the classical theory, the *quantum* scalar constraint has a form that naturally leads one to use ϕ as a relational or internal time variable [13,14]. In the case now under consideration, where the scalar field ϕ also satisfies the wave equation $\Box \phi = 0$, the internal time defined by ϕ and the harmonic time τ are related just by a constant in any classical solution: $\phi = (p_{(\phi)}/\mathring{V})\tau$. However, since the constant varies from one solution to another, in the quantum theory, conceptually, the two choices are quite different. In LQC it is simplest to begin with the harmonic time in the classical theory and reinterpret the quantum scalar constraint as providing time evolution in the relational time ϕ.

Since one generally uses the conformal time η in the cosmology literature, we will conclude by writing down the equation of motion that follows from (3.31) with the choice $N_{\rm hom} = a$:

$$\mathcal{T}''_{\vec{k}} + 2\frac{a'}{a} \mathcal{T}'_{\vec{k}} + k^2 \mathcal{T}_{\vec{k}} = 0, \quad (3.32)$$

where a "prime" denotes derivative with respect to the conformal time η. One often rescales $\mathcal{T}_{\vec{k}}$ to obtain a field $\chi_{\vec{k}}$ with physical dimensions of a scalar field,

$$\chi_{\vec{k}} := \frac{a}{2\sqrt{\kappa}} \mathcal{T}_{\vec{k}}, \quad (3.33)$$

for which, furthermore, the equation of motion resembles that of a harmonic oscillator (with time dependent frequency):

$$\chi''_{\vec{k}} + \left(-\frac{a''}{a} + k^2\right)\chi_{\vec{k}} \equiv \chi''_{\vec{k}} + a^2\left(-\frac{R}{6} + \frac{k^2}{a^2}\right)\chi_{\vec{k}} = 0,$$

$$(3.34)$$

where R denotes the scalar curvature of the background homogeneous space-time.[7] This form of the perturbation equations has also been exploited in Refs. [77–79] in singling out a physically motivated quantization.

Remark: In our presentation we began with the connection variables employed in the LQG literature because

---

[7]For the scalar perturbation, $\psi_{\vec{k}} = a\mathcal{Q}_{\vec{k}}$ has the same properties as $\chi_{\vec{k}}$: $\psi_{\vec{k}}$ also has the physical dimensions of a scalar field and satisfies the same equation of motion as (3.34).

conceptually they are essential for the new quantum kinematics used in LQG, and the subsequent treatment of the singularity-free quantum dynamics of our background FLRW space-times. But we passed to the ADM variables for first order perturbations by solving the Gauss constraint because much of the cosmological perturbation analysis is carried out in terms of these variables. For a treatment of cosmological perturbations in connection variables, see in particular [76,80–84].

IV. QUANTUM THEORY: MAIN STEPS

We will now use the phase space $\tilde{\Gamma}_{\text{Trun}}$ as the point of departure for quantization. Since the truncated second order constraints for the scalar and two tensor modes are identical (except for unimportant numerical factors), as in Sec. III D we will focus just on one tensor mode $\mathcal{T}_{\vec{k}}$. Thus, from now on, $\tilde{\Gamma}_{\text{Trun}}$ will be taken to be spanned by three canonically conjugate pairs $(\nu, b; \phi, p_{(\phi)}; \mathcal{T}_{\vec{k}}, \mathfrak{p}_{\vec{k}})$, the first two representing the background, and the third representing the first order perturbation. They are subject to the single constraint $\mathbb{S}_o[N_{\text{hom}}] = 0$ [see Eq. (3.29)]. Dynamics is governed by the vector field $X_{\text{Dyn}}^{\alpha} = \Omega_o^{\alpha\beta} \partial_\beta \mathbb{S}_o + \Omega_1^{\alpha\beta} \partial_\beta \mathbb{S}_2'$ where $\mathbb{S}_2'$ is given by (3.31).

In this section we present the general program without entering into details of how the states and operators related to first order perturbations are defined. In particular, the expression of the Hamiltonian operator dictating the dynamics of first order perturbations is formal. The precise definitions of these states, operators and the necessary regularization procedure are provided in the next two sections.

A. Background quantum geometry

Let us begin by recalling the quantum theory of the homogeneous sector (for details, see e.g., Refs. [13,14,31]). As mentioned in Sec. II, LQC provides a kinematical Hilbert space $\mathcal{H}_{\text{kin}}$ which, as in full LQG [61,62] is uniquely determined by the physical requirement independence w.r.t. background (and fiducial) structures [60]. In the configuration representation, kinematical quantum states are given by wave functions $\Psi(\nu, \phi)$. As noted in Sec. III D, dynamics is simplest if one uses harmonic time by setting $N = a^3$. Then, the scalar constraint becomes a well-defined operator on $\mathcal{H}_{\text{kin}}$. Physical states $\Psi_o(\nu, \phi)$ are annihilated by this constraint, i.e., they satisfy

$$\hat{\mathbb{S}}_o \Psi_o(\nu, \phi) \equiv -\frac{\hbar^2}{2\ell^3}(\partial_\phi^2 + \Theta)\Psi_o(\nu, \phi) = 0, \quad (4.1)$$

where the action of Θ is given by

$$\Theta\Psi_o(\nu, \phi) = \frac{3\pi G}{\lambda^2} \nu[(\nu + 2\lambda)\Psi_o(\nu + 4\lambda, \phi) \\ - 2\nu\Psi_o(\nu, \phi) + (\nu - 2\lambda)\Psi_o(\nu - 4\lambda, \phi)]. \quad (4.2)$$

Thus, Θ is a second order difference operator that acts only on the argument ν of Ψ_o, with step size 4λ where $\lambda^2 = 4\sqrt{3}\pi\gamma\ell_{\text{Pl}}^2$ is the "area gap" of LQG. Θ is self-adjoint and positive definite on $\mathcal{H}_{\text{kin}}$ [85]. As is common in the Dirac quantization procedure, none of the solutions to (4.1) are normalizable in the kinematic inner product. But there is a standard "group averaging" method to endow the space of solutions with a physical inner product [86–88]. The resulting physical states $\Psi_o \in \mathcal{H}_{\text{phy}}^o$ turn out to be solutions to

$$-i\hbar\partial_\phi\Psi(\nu, \phi) = \hat{H}_o\Psi_o(\nu, \phi), \quad \text{where } \hat{H}_o = \hbar\sqrt{\Theta}, \quad (4.3)$$

which are symmetric under $\nu \to -\nu$ and have finite norm

$$\|\Psi_o\|^2 = \frac{\lambda}{\pi} \sum_{\nu \in 4N\lambda; N \in \mathbb{Z}} |\nu|^{-1}|\Psi_o(\nu, \phi_o)|^2, \quad (4.4)$$

where ϕ_o is any fixed instant of the internal time ϕ. [The scalar product defined by (4.4) is insensitive to the choice of ϕ_o.] Heuristically (4.3) can be thought of as the positive frequency square-root of (4.1) in the internal time ϕ.

The most noteworthy feature of this outcome is that the *quantum* Hamiltonian constraint is naturally deparametrized: its form suggests that the scalar field ϕ can be interpreted as a "relational or internal time" with respect to which physical states Ψ_o evolve. Thus, as one would hope, the imposition of the quantum constraint a la Dirac has naturally led us to dynamics. Interestingly, while our use of a lapse function corresponding to the harmonic time τ simplified the form of $\hat{H}_o$, it was not essential to arrive at this interpretation of ϕ. Indeed this interpretation was initially arrived at with lapse set to 1 corresponding to proper time [13] (but a more complicated form of $\hat{H}_o$.)

Thus, as is usual, the Dirac quantization procedure has naturally led to a Schrödinger picture in which the scalar field ϕ is simply a time parameter and the sole dynamical variable is ν that determines the volume of the Universe via

$$\hat{V}\Psi_o(\nu) = 2\pi\gamma\ell_{\text{Pl}}^2|\nu|\Psi_o(\nu). \quad (4.5)$$

In the Heisenberg picture, the volume evolves in time from the bounce via

$$\hat{V}(\phi) = e^{(i/\hbar)\hat{H}_o(\phi - \phi_{\text{B}})}(2\pi\gamma\ell_{\text{Pl}}^2|\nu|)e^{-(i/\hbar)\hat{H}_o(\phi - \phi_{\text{B}})}. \quad (4.6)$$

Every element Ψ_o of $\mathcal{H}_{\text{phy}}^o$ represents a four-dimensional quantum geometry. However, for our purposes, only a subset of these states is relevant. Choose a classical, expanding FLRW space-time in which $p_{(\phi)} \gg \hbar$ (in the classical units $G = c = 1$) and a homogeneous slice at a late time $\phi = \phi_o$, when the matter density and curvature are negligibly small compared to the Planck scale. This defines a point γ_o in Γ_o. Then, in the Schrödinger representation, one can introduce "coherent states" $\Psi_o(\nu, \phi_o)$ in $\mathcal{H}_{\text{phy}}^o$ which are sharply peaked at γ_o [12]. *By a quantum background geometry, we will mean a physical state $\Psi_o(\nu, \phi)$ obtained by evolving these initial states*

using (4.3). There is a large class of such states and our considerations will apply to all of them. One can show that these states remain sharply peaked on the classical trajectory passing through γ_o for all $\phi > \phi_o$. In the backward time-evolution, they do so only till the density reaches a few hundredths of the Planck density. Even in the deep Planck regime the wave function remains sharply peaked but now the peak follows an effective trajectory which undergoes a quantum bounce. At the bounce point the matter density attains a maximum, $\rho_{\max} \approx 0.41 \rho_{\mathrm{Pl}}$. While there is good agreement with general relativity once the matter density falls below a few hundredths of the Planck density, Einstein's equations break down completely in the Planck regime. But the quantum state (and even the effective trajectory) remains well-defined throughout the entire evolution, including the Planck scale neighborhood of the bounce [13,14,31].

Finally, in the Heisenberg picture, we can define the space-time metric operator. Recall first that in the classical theory the lapse function corresponding to the scalar field time is given by $N = a^3(\phi)\ell^3 p_{(\phi)}^{-1}$. Since $\hat{p}_{(\phi)}\Psi_o = \hat{H}_o \Psi_o$ for any $\Psi_o \in \mathcal{H}_{\mathrm{phy}}^o$, in the Heisenberg representation, the metric operator is given by

$$\hat{g}_{ab} dx^a dx^b \equiv d\hat{s}^2 = \hat{H}_o^{-1} \ell^6 \hat{a}^6(\phi) \hat{H}_o^{-1} d\phi^2 + \hat{a}^2(\phi) d\vec{x}^2, \tag{4.7}$$

where we have used a symmetric factor ordering in the first term and defined the (positive definite, self-adjoint) scale factor operator in the Heisenberg picture via

$$\ell \hat{a}(\phi) = [\hat{V}(\phi)]^{\frac{1}{3}}. \tag{4.8}$$

In the Heisenberg picture, the geometry is quantum because the metric coefficients are now quantum operators on $\mathcal{H}_{\mathrm{phy}}^o$.

B. Perturbations on the quantum geometry Ψ_o

Because we were able to reinterpret the quantum constraint equation $\hat{\mathbb{S}}_o \Psi_o = 0$ as providing an evolution of physical states in the internal time variable ϕ, we were naturally led to work in the Schrödinger picture for the homogeneous background geometry. Furthermore, because Ψ_o represents a quantum state of the background geometry and ψ of perturbations, it is natural to assume that the total state has a simple tensor product structure

$$\Psi(\nu, \mathcal{T}_{\vec{k}}, \phi_0) = \Psi_o(\nu, \phi_0) \otimes \psi(\mathcal{T}_{\vec{k}}, \phi_0) \tag{4.9}$$

at some initial time ϕ_0. Then, because the backreaction is neglected, the evolution of Ψ_o is dictated just by (4.3); it is insensitive to the form of ψ. (This is entirely analogous the situation in the classical theory.) Therefore, the tensor product structure is preserved under evolution. As in the classical theory, our task is to evolve ψ on the specified background quantum geometry Ψ_o, i.e., to lift the given homogeneous "quantum trajectory" $\Psi_o(\nu, \phi)$ to a "trajectory" $\Psi_o(\nu, \phi) \otimes \psi(\mathcal{T}_{\vec{k}}, \phi)$ of the truncated quantum theory, where the background state is the given one. To carry out this task we need to first complete two preliminary steps.

On the classical phase space $\tilde{\Gamma}_{\mathrm{Trun}} = \Gamma_o \times \tilde{\Gamma}^{(1)}$, the part $\Omega_1^{\alpha\beta} \partial_\beta \mathbb{S}_2'$ of the dynamical vector field X_{Dyn}^α dictates how perturbations propagate on a homogeneous background solution $\nu(\phi)$. Now, $\mathbb{S}_2'$ depends not only on the perturbations $(\mathcal{T}_{\vec{k}}, \mathfrak{p}_{\vec{k}})$ but also on the *time dependent scale factor of the background solution*. Therefore, to construct the operator $\hat{\mathbb{S}}_2'$, it is simplest to work in the "interaction picture" where the background scale factor operators evolve in the relational time ϕ and the background state Ψ_o is frozen at a time, which we will take to be the bounce time $\phi = \phi_{\mathrm{B}}$. The first preliminary step is to carry out this passage to the interaction picture via

$$\Psi_{\mathrm{Int}}(\nu, \mathcal{T}_{\vec{k}}, \phi) = e^{-(i/\hbar)\hat{H}_o(\phi - \phi_{\mathrm{B}})}(\Psi_o(\nu, \phi) \otimes \psi(\mathcal{T}_{\vec{k}}, \phi)). \tag{4.10}$$

A second step is needed because now the evolution is with respect to the relational time ϕ. Therefore, we have to choose a specific lapse function, $N_{\mathrm{hom}} = a^3 \ell^3 / p_{(\phi)}$, in the expression (3.31) of $\mathbb{S}_2'$ and then use an appropriate factor ordering to convert it to an operator. In this step, we will use the same factor ordering as in the expression (4.7) of the quantum metric operator and make a simplification using the evolution equation $\hat{p}_{(\phi)} \Psi_o \equiv -i\hbar \partial_\phi \Psi_0 = \hat{H}_0 \Psi_0$.

These two steps, and the form (3.31) of $\mathbb{S}_2'$ lead us to the following evolution equation for the total state in the interaction picture:

$$\begin{aligned} i\hbar \partial_\phi \Psi_{\mathrm{Int}}(\nu, \mathcal{T}_{\vec{k}}, \phi) &= \Psi_o(\nu, \phi_{\mathrm{B}}) \otimes i\hbar \partial_\phi \psi(\mathcal{T}_{\vec{k}}, \phi) \\ &= \frac{1}{2} \sum_{\vec{k}} 4\kappa [\hat{H}_o^{-1} \Psi_o(\nu, \phi_{\mathrm{B}})] \\ &\quad \otimes [|\hat{\mathfrak{p}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi)] \\ &\quad + \frac{k^2}{4\kappa} [(\hat{H}_o^{-\frac{1}{2}} \hat{a}^4(\phi) \hat{H}_o^{-\frac{1}{2}}) \Psi_o(\nu, \phi_{\mathrm{B}})] \\ &\quad \otimes [|\hat{\mathcal{T}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi)]. \end{aligned} \tag{4.11}$$

Let us take the scalar product of this equation with Ψ_o (which we assume to be normalized). the result is the required evolution equation for the quantum state $\psi(\mathcal{T}_{\vec{k}}, \phi)$ of the perturbation propagating on the quantum background geometry Ψ_o:

$$\begin{aligned} i\hbar \partial_\phi \psi(\mathcal{T}_{\vec{k}}, \phi) &= \hat{H}_1 \psi(\mathcal{T}_{\vec{k}}, \phi) \\ &:= \frac{1}{2} \sum_{\vec{k}} 4\kappa \langle \hat{H}_o^{-1} \rangle |\hat{\mathfrak{p}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi) \\ &\quad + \frac{k^2}{4\kappa} \langle \hat{H}_o^{-\frac{1}{2}} \hat{a}^4(\phi) \hat{H}_o^{-\frac{1}{2}} \rangle |\hat{\mathcal{T}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi), \end{aligned} \tag{4.12}$$

where, by construction, the expectation values of the background geometry operators are taken in the given state Ψ_o of the background quantum geometry.

Note that, at a fundamental level, ψ now evolves in a probability amplitude Ψ_o of background geometries g_{ab}, rather than a fixed g_{ab}. But (4.12) implies that its evolution is not sensitive to *all* the details of the fluctuations of this quantum geometry; it is sensitive to only two "moments" $\langle \hat{H}_o^{-1} \rangle$ and $\langle \hat{H}_o^{-\frac{1}{2}} \hat{a}^4(\phi) \hat{H}_o^{-\frac{1}{2}} \rangle$. This is so even though (4.12) is an *exact* consequence of (4.11) with no further approximations. Furthermore, since the backreaction of the perturbation on Ψ_o is neglected within the test field approximation inherent to our truncation scheme, nothing is lost by projecting (4.11) along Ψ_o in arriving at (4.12). *That is, within the test field approximation, (4.12) captures the full information about the evolution of ψ that is contained in the original equation (4.11).*

Next, recall that in the standard cosmology literature one regards the quantum state $\psi(\mathcal{T}_{\vec{k}}, \phi)$ of perturbations as propagating on a *classical* FLRW metric, specified by the scale factor $a_{\rm cl}(\phi)$. In the Schrödinger picture now under consideration, the evolution is given by

$$i\hbar \partial_\phi \psi(\mathcal{T}_{\vec{k}}, \phi) = \frac{1}{2} \sum_{\vec{k}} 4\kappa (p_{(\phi)}^{-1}) |\hat{\mathfrak{p}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi)$$
$$+ \frac{k^2}{4\kappa} [p_{(\phi)}^{-1} a_{\rm cl}^4(\phi)] |\hat{\mathcal{T}}_{\vec{k}}|^2 \psi(\mathcal{T}_{\vec{k}}, \phi),$$

(4.13)

where $p_{(\phi)}$ can also be expressed using geometric variables using the constraint equation satisfied by the background. Comparing (4.13) with (4.12), we find that the evolution of the test perturbation $\hat{\mathcal{T}}_{\vec{k}}$ on the *quantum* background geometry given by $\Psi_o(\nu, \phi)$ is indistinguishable from that of a test perturbation propagating on a *smooth* FLRW background

$$\tilde{g}_{ab} dx^a dx^b \equiv d\tilde{s}^2 = -(\tilde{p}_{(\phi)})^{-2} \ell^6 \tilde{a}^6(\phi) d\phi^2 + \tilde{a}(\phi)^2 d\vec{x}^2,$$

(4.14)

where

$$(\tilde{p}_{(\phi)})^{-1} = \langle \hat{H}_o^{-1} \rangle \quad \text{and} \quad \tilde{a}^4 = \frac{\langle \hat{H}_o^{-\frac{1}{2}} \hat{a}^4(\phi) \hat{H}_o^{-\frac{1}{2}} \rangle}{\langle \hat{H}_o^{-1} \rangle}. \quad (4.15)$$

Again, this is an *exact* equivalence between our truncated LQG and the theory of quantum fields on a smooth FLRW geometry determined by $\tilde{a}$ and $\tilde{p}_{(\phi)}$.

At a technical level, the existence of such a simple relation is quite surprising at first. But this result should *not* be interpreted to mean that the standard quantum theory of perturbations on a classical FLRW solution to Einstein's equations holds in the Planck regime. *It does not.* For, the $\tilde{a}(\phi)$ seen by $\hat{\mathcal{T}}_{\vec{k}}$ is *very* different from the $a_{\rm cl}(\phi)$ of a classical solution; in particular, $(\tilde{a}(\phi), \tilde{p}_{(\phi)})$ do

not satisfy Einstein's equations. Indeed, their expressions involve $\hbar$; although $\tilde{g}_{ab}$ is smooth, it incorporates quantum corrections which are so large in the Planck regime that they tame the big bang singularity. Furthermore, *the pair does not even satisfy the effective equations of LQC* which track the peak of the wave function $\Psi_o(\nu, \phi)$ of the background geometry. Certain aspects of quantum *fluctuations* inherent in $\Psi_o(\nu, \phi)$ are absorbed in these tilde fields. Thus, $\tilde{g}_{ab}$ may be thought of as a *dressed effective geometry* that is relevant for propagation of linear perturbations on the full quantum (background) geometry determined by $\Psi_o(\nu, \phi)$. In retrospect, from what we know in other areas of physics, such a result is not entirely unexpected. For example, light propagating in a medium interacts with its atoms but the net effect of these interactions can be encoded just in a few parameters such as the refractive index of the medium. In our case, the "medium" is the quantum geometry and the tilde variables $\tilde{a}$, $\tilde{p}_{(\phi)}$ encode the interaction between this medium and the perturbations. Only two parameters suffice simply because the background quantum geometry is homogeneous and isotropic. Already in the anisotropic Bianchi models, (an extension of the discussion of Ref. [89] implies that) one would need additional parameters characterizing the dressed, effective anisotropies. Finally, the encoding is rather sophisticated: prior to the calculation, it would have been impossible to guess the precise moments of the fluctuations of geometry that are to capture this interaction.

Remark: Several equations in this subsection closely resemble those in Ref. [38]. However the conceptual underpinning is quite different. The discussion in Ref. [38] began by assuming that quantum dynamics can be obtained by imposing the constraint analogous to $[\hat{\mathbb{S}}_o + \hat{\mathbb{S}}_2'] \Psi_o \otimes \psi = 0$. This "quantum constraint" was then expanded and subleading terms were discarded to arrive at an equation analogous to (4.11). Consequently, it was not appreciated that, within the truncation scheme, (4.12) carries the full information about the quantum evolution of ψ. As discussed in Sec. III D (see also the Appendix), a more careful examination has revealed that the strategy of imposing a quantum constraint cannot be justified. Therefore, we adopted a different route here. We mimicked the strategy from the classical theory: Just as $X_{\rm Dyn}^\alpha$ provides a lift of the dynamical trajectories on Γ_o to $\Gamma_o \times \Gamma_1$, we lifted the "quantum dynamical trajectories" Ψ_o on $\mathcal{H}_{\rm phy}^o$ to $\Psi_o \otimes \psi$ on $\mathcal{H}_{\rm phy}^o \otimes \mathcal{H}^1$. This route is also more direct in that we did not have to discard any terms to arrive at (4.11).

Finally, let us translate this result using conformal time η commonly used in the literature on cosmological perturbations. The dressed, effective metric can be written as

$$\tilde{g}_{ab} dx^a dx^b \equiv d\tilde{s}^2 = \tilde{a}^2(\phi)(-d\tilde{\eta}^2 + d\vec{x}^2), \quad (4.16)$$

where

$$d\tilde{\eta} = [\ell^3 \tilde{a}^2(\phi)] \tilde{p}_{(\phi)}^{-1} d\phi. \qquad (4.17)$$

Therefore, in the truncated theory, the exact evolution equation for the quantum perturbation $\hat{\mathcal{T}}_{\vec{k}}$ on the background quantum geometry is given by

$$\hat{\mathcal{T}}_{\vec{k}}'' + 2\frac{\tilde{a}'}{\tilde{a}}\hat{\mathcal{T}}_{\vec{k}}' + k^2\hat{\mathcal{T}}_{\vec{k}} = 0, \qquad (4.18)$$

where the prime now denotes a derivative with respect to $\tilde{\eta}$. This mathematical equivalence simplifies both conceptual and technical aspects of our analysis considerably because the well-developed techniques from quantum field theory in curved space-times can now be readily imported into the quantum field theory of perturbations ψ on quantum geometries Ψ_o. In Sec. VI, we will use this strategy to define in detail the quantum states for perturbations and composite operators that are needed to complete the quantum theory.

Remarks:

(1) We can make a further simplification through a "mean field" approximation in which the fluctuations are ignored. More precisely, let us first recall [13,14,31] that even in the Planck regime the state $\Psi_o(\nu, \phi)$ is sharply peaked on an effective geometry

$$\bar{g}_{ab}dx^a dx^b \equiv d\bar{s}^2 = \bar{a}^6 \frac{\ell^6}{p_{(\phi)}^2} d\phi^2 + \bar{a}^2(\phi)d\vec{x}^2, \qquad (4.19)$$

which agrees with the general relativity solution (for the same value of $p_{(\phi)}$) for large $a(\phi)$ but has a inbuilt bounce at $\bar{a}(\phi)^6 = p_{(\phi)}^2/(2\ell^6 \rho_{\max})$, with $\rho_{\max} = 3/(8\pi G \gamma^2 \lambda^2) \approx 0.41 \rho_{\mathrm{Pl}}$. In terms of our quantum geometry state Ψ_o, the scale factor is given just by the expectation value $\bar{a}(\phi) = \langle \hat{a}(\phi) \rangle$ in this state. Suppose we ignore quantum fluctuations, i.e., *use a mean field approximation* in which the expectation values of powers of $\hat{a}$ and $\hat{H}_o$ are replaced by the same powers of their expectation values. In this approximation, the quantum perturbation $\hat{\mathcal{T}}_{\vec{k}}$ would seem to propagate on the effective geometry determined by the pair $(\bar{a}(\phi), p_{(\phi)})$. Now, one knows that there exist background quantum geometries Ψ_o which are *very* sharply peaked on this effective geometry even in the Planck regime. If one uses such a Ψ_o, the mean field approximation is excellent for studying the propagation of perturbations on quantum geometries under consideration.[8] The exact evolution of $\hat{\mathcal{T}}_{\vec{k}}$, on the other hand, sees the

more sophisticated, dressed effective geometry determined by $(\tilde{a}, \tilde{p}_{(\phi)})$.

(2) If one were to use the mean field approximation, the quantum perturbations would satisfy (4.18) with the tilde quantities replaced by the barred quantities that refer to the effective LQC solutions. At first sight, it may therefore seem one could have arrived at these equations simply by perturbing the effective equations of LQC. This could have been a viable interpretation had there been a clear set of effective equations in full LQG to perturb around backgrounds satisfying the LQC solutions. But as emphasized in Secs. I and II and (i) we do not yet have effective equations in full LQG, and (ii) if one were to adopt the naive strategy of considering a set of linearized equations in general relativity and simply replacing the background solution to Einstein's equations by a solution to the effective equations, one faces a very large ambiguity in the choice of equations with which to begin and, furthermore, the set of final equations need not be internally consistent when the background does not satisfy Einstein's equations. Our procedure is free of these drawbacks because we first constructed the quantum theory and arrived at Eq. (4.18) by showing an exact equivalence of fields $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ propagating on the quantum geometry Ψ_o and those propagating on the geometry determined by $\tilde{g}_{ab}$. Nowhere did this procedure use effective equations of LQC.

V. HILBERT SPACES OF STATES OF $\hat{\mathcal{T}}_{\vec{k}}$

In this section we will construct the Hilbert space of quantum states of gauge invariant perturbations following two different but complementary avenues. The first is geared to mathematical physicists and well adapted to the Hamiltonian framework used in our classical considerations. The second follows the route that is more often taken in analyzing cosmological perturbations. We show that the two are equivalent. Therefore, to ensure conceptual continuity and coherence, one can start with quantization given in Sec. V A and then use the framework presented in Sec. V B which is better adapted to regularization, renormalization and numerical simulations. We also provide the explicit expression of the 2-point function which makes it manifest that the group of space-translations continues to be a symmetry in the quantum theory.

A. The Weyl algebra and its representations

The classical phase space $\Gamma_{\mathcal{T}}$ for the tensor modes is spanned by canonically conjugate pairs $(\mathcal{T}_{\vec{k}}, \mathfrak{p}_{\vec{k}})$. The corresponding operators $\hat{\mathcal{T}}_{\vec{k}}$, $\hat{\mathfrak{p}}_{\vec{k}}$ satisfy the canonical commutation relations and generate the Heisenberg algebra in the quantum theory. For technical simplicity, it is convenient to

---

[8]In numerical simulations of the evolution of the quantum state ψ of perturbations, for example, if Ψ_o is chosen appropriately, the numerical errors would be much higher than those introduced by the mean field approximation.

exponentiate them to obtain the Weyl algebra $\mathfrak{W}$ whose representations provide the required Hilbert spaces of quantum states.

One begins with the observation that, with each vector $(\lambda, \mu) \in \Gamma_{\mathcal{T}}$, one can associate a natural linear combination of smeared configuration and momentum operators:

$$\hat{F}(\lambda, \mu) := \Omega((\hat{\mathcal{T}}, \hat{p}), (\lambda, \mu)) = \frac{1}{\overset{\circ}{V}} \sum_{\vec{k}} \mu_{\vec{k}}^{\star} \hat{\mathcal{T}}_{\vec{k}} - \lambda_{\vec{k}}^{\star} \hat{p}_{\vec{k}}$$

$$= \int_{\mathbb{M}} d^3\vec{x} (\mu(\vec{x}) \hat{\mathcal{T}}(\vec{x}) - \lambda(\vec{x}) \hat{p}(\vec{x})). \quad (5.1)$$

The Weyl operators $\hat{W}(\lambda, \mu)$ are their exponentials:

$$\hat{W}(\lambda, \mu) := e^{\frac{i}{\hbar}\hat{F}(\lambda,\mu)}. \quad (5.2)$$

It is more convenient to work with these exponentials for two reasons. First, while the field operators $\hat{F}$ are unbounded, the $\hat{W}$ are unitary and hence bounded operators in any representation. Therefore, one avoids the awkward issues of specifying operator domains. Second, the vector space generated by finite linear combinations of the Weyl operators is automatically closed under the Hermitian-conjugation operation $\hat{W}^{\dagger}(\lambda, \mu) = \hat{W}(-\lambda, -\mu)$, and, more importantly, under the product:

$$\hat{W}(\lambda_1, \mu_1)\hat{W}(\lambda_2, \mu_2) = e^{\frac{1}{2\hbar} \int d^3\vec{x} (\lambda_1\mu_2 - \mu_1\lambda_2)}$$

$$\times \hat{W}(\lambda_1 + \lambda_2, \mu_1 + \mu_2). \quad (5.3)$$

Thus, the vector space has the structure of a $\star$-algebra. This is the Weyl algebra $\mathfrak{W}$. (It can be easily endowed the structure of a $C^{\star}$ algebra but this will not be necessary for our purposes.)

To find the representations of $\mathfrak{W}$, it is simplest to use the standard Gel'fand, Naimark, Segal (GNS) construction [90]: Given a positive linear function (PLF) on $\mathfrak{W}$, the construction provides an explicit Hilbert space $\mathcal{H}_1$ and a representation of elements of $\mathfrak{W}$ by concrete operators on that $\mathcal{H}_1$. This representation is cyclic: every state in $\mathcal{H}_1$ arises from the action of operators representing elements of $\mathfrak{W}$ on a "vacuum." In this representation, the PLF turns out to be just the vacuum expectation value functional. A natural strategy for linear fields is to first find a complex structure on the phase space that is compatible with the symplectic structure thereon and then use the Hermitian inner product provided by the resulting Kähler structure to define the required PLF on $\mathfrak{W}$ [91,92]. We will follow this conceptual strategy but in a manner that retains close contact with the cosmology literature. Therefore, prior knowledge of the complex and Kähler structures will not be necessary to follow the construction.

In linear field theories in Minkowski space, one narrows the selection of the positive linear functional by requiring that it (and hence the vacuum state) be Poincaré invariant. In the present case, it is natural to require that the PLF be invariant under the three-dimensional translational

symmetry of the background geometry. Such PLFs can be constructed as follows. Choose a set of complex coefficients (e_k, f_k) (with $k \geq 0$) such that

$$e_k f_k^{\star} - e_k^{\star} f_k \equiv 2i \operatorname{Im}(e_k f_k^{\star}) = i \quad \text{for all } k. \quad (5.4)$$

(For a massless scalar field in Minkowski space, $e_k = e^{-i\omega t_o}/\sqrt{2\omega}$ and $f_k = (-i\omega e^{-i\omega t_o})/\sqrt{2\omega}$ with $\omega = |\vec{k}|$.) Then, one can extract the 'positive frequency' part $a_{\vec{k}}$ of any vector $(\lambda, \mu) \in \Gamma_{\mathcal{T}}$ as follows:

$$a_{\vec{k}} := -i(f_k^{\star}\lambda_{\vec{k}} - e_k^{\star}\mu_{\vec{k}}) \quad (5.5)$$

so that

$$\lambda_{\vec{k}} = e_k a_{\vec{k}} + e_k^{\star} a_{\vec{k}}^{\star}, \quad \text{and} \quad \mu_{\vec{k}} = f_k a_{\vec{k}} + f_k^{\star} a_{\vec{k}}^{\star}. \quad (5.6)$$

(Note that while λ, μ satisfy the "reality condition" $\lambda_{\vec{k}}^{\star} = \lambda_{-\vec{k}}$, $\mu_{\vec{k}}^{\star} = \mu_{-\vec{k}}$, the "positive frequency parts" don't: $a_{\vec{k}}^{\star} \neq a_{-\vec{k}}$.) The required PLF is then simply

$$\langle \hat{W}(\lambda, \mu) \rangle = e^{-\frac{1}{2\hbar}\frac{1}{\overset{\circ}{V}}\sum |a_{\vec{k}}|^2}. \quad (5.7)$$

Under the action of a translation $\vec{x} \rightarrow \vec{x} + \vec{v}$ on $\mathbb{M}$, we have: $\hat{\mathcal{T}}(\vec{x}) \rightarrow \hat{\mathcal{T}}(\vec{x} + \vec{v})$, and $\hat{p}(\vec{x}) \rightarrow \hat{p}(\vec{x} + \vec{v})$, whence $\hat{W}(\lambda(\vec{x}), \mu(\vec{x})) \rightarrow \hat{W}(\lambda(\vec{x} - \vec{v}), \mu(\vec{x} - \vec{v}))$, or, in the momentum space, $\lambda_{\vec{k}} \rightarrow e^{-i\vec{k}\cdot\vec{v}}\lambda_{\vec{k}}$, $\mu_{\vec{k}} \rightarrow e^{-i\vec{k}\cdot\vec{v}}\mu_{\vec{k}}$. This trivially implies $a_{\vec{k}} \rightarrow e^{-i\vec{k}\cdot\vec{v}}a_{\vec{k}}$. Hence the PLF (5.7) is left invariant and the translation is represented by a unitary transformation on the GNS Hilbert space $\mathcal{H}_1$ under which the GNS vacuum is invariant. Thus, each choice of coefficients (e_k, f_k) satisfying (5.4) leads to a representation of the Weyl algebra in which the (GNS) vacuum is invariant under translations.

B. Space-time description

At this stage it is convenient to make the construction more explicit using the familiar expansions in terms of the creation and annihilation operators. This entails going to the Heisenberg picture. In terms of the conformal time $\tilde{\eta}$ of the dressed effective metric of Sec. IVA, the space-time operator is represented as $\hat{\mathcal{T}}(\vec{x}, \tilde{\eta})$. It satisfies the field equation and the canonical commutation relations, and is self-adjoint. These properties can be neatly captured by an expansion of the form

$$\hat{\mathcal{T}}(\vec{x}, \tilde{\eta}) = \frac{1}{\overset{\circ}{V}} \sum_{\vec{k}} (e_k(\tilde{\eta})\hat{A}_{\vec{k}} + e_k^{\star}(\tilde{\eta})\hat{A}_{-\vec{k}}^{\dagger})e^{i\vec{k}\cdot\vec{x}}. \quad (5.8)$$

We require that $e_k(\tilde{\eta})$ should satisfy (4.18):

$$e_k''(\tilde{\eta}) + 2\frac{\tilde{a}'}{\tilde{a}} e_k'(\tilde{\eta}) + k^2 e_k(\tilde{\eta}) = 0, \quad (5.9)$$

so that $\hat{\mathcal{T}}(\vec{x}, \tilde{\eta})$ satisfies the desired equation of motion. (In relation to the more familiar expansion in Minkowski space-time, $e_k(\tilde{\eta})$ now plays the role of the *positive*

frequency basis functions $e^{-i\omega t}/\sqrt{2\omega}$.) Next, for each $\vec{k}$, the space of solutions to (5.9) is two-dimensional and one chooses a complex solution $e_k(\tilde{\eta})$ satisfying the normalization condition:

$$\frac{\tilde{a}^2}{4\kappa}(e_k(\tilde{\eta})e_k'^\star(\tilde{\eta}) - e_k^\star(\eta)e_k'(\eta)) = i. \qquad (5.10)$$

This condition needs to be imposed only at some initial instant of time $\tilde{\eta}_o$; Eq. (5.9) then guarantees that it is then automatically satisfied for all $\tilde{\eta}$. With this normalization, if the time independent operators $\hat{A}_{\vec{k}}$ and $\hat{A}_{\vec{k}}^\dagger$ satisfy the commutation relations

$$[\hat{A}_{\vec{k}}, \hat{A}_{\vec{k}'}^\dagger] = \hbar \ell^3 \delta_{\vec{k},\vec{k}'}, \qquad (5.11)$$

then

$$\hat{\mathcal{T}}(\vec{x}, \tilde{\eta})$$

and

$$\hat{\mathfrak{p}}(\vec{x}, \tilde{\eta}) = \frac{\tilde{a}^2}{4\kappa}\frac{1}{V}\sum_{\vec{k}}(e_k'(\tilde{\eta})\hat{A}_{\vec{k}} + e_k^{\star\prime}(\tilde{\eta})\hat{A}_{-\vec{k}}^\dagger)e^{i\vec{k}\cdot\vec{x}} \qquad (5.12)$$

satisfy the required canonical commutation relations at any fixed value of $\tilde{\eta}$. In view of their properties, one can interpret $\hat{A}_{\vec{k}}$ as annihilation operators, define the vacuum $|0\rangle$ as the state annihilated by all $\hat{A}_{\vec{k}}$, and generate the Fock space $\mathcal{H}_1$ by repeatedly acting on the vacuum by creation operators $\hat{A}_{\vec{k}}^\dagger$. It is straightforward to calculate the vacuum expectation value of the Weyl operator

$$\hat{W}(\lambda, \mu)|_{\tilde{\eta}_o} = e^{\frac{i}{\hbar \mathring{V}}\sum_{\vec{k}}\mu_{\vec{k}}^\star\hat{\mathcal{T}}_{\vec{k}} - \lambda_{\vec{k}}^\star\hat{\mathfrak{p}}_{\vec{k}}} \qquad (5.13)$$

at any conformal time $\tilde{\eta}_o$. It is given by (5.7) where $e_k = e_k(\tilde{\eta}_o)$ and $f_k = (a^2/4\kappa)e_k'(\tilde{\eta}_o)$. Thus, our second description of the Hilbert space $\mathcal{H}_1$ adapted to the covariant space-time picture, is completely equivalent to the first description, adapted to the Weyl algebra that was constructed starting from the phase space. The first description serves to bring out the conceptual structure of the quantum theory that emerges from the phase space description of Sec. III while the second is better adapted to calculations, e.g., regularization of the stress-energy tensor discussed in Sec. VI.

To summarize, for each choice of solutions $e_k(\tilde{\eta})$ satisfying the normalization condition (5.10) we obtain a vacuum state and hence a Fock representation of the canonical commutation relations (or, of the Weyl algebra). Since these representations are completely characterized by their 2-point functions, it is instructive to write them out explicitly. Using the expansion (5.8) of $\hat{\mathcal{T}}(\vec{x}, \tilde{\eta})$, we obtain

$$\langle 0 | \hat{\mathcal{T}}(\vec{x}_1, \tilde{\eta}_1)\hat{\mathcal{T}}(\vec{x}_2, \tilde{\eta}_2) | 0 \rangle$$
$$= \frac{\hbar}{\mathring{V}}\sum_{\vec{k}}e^{i\vec{k}\cdot(\vec{x}_1-\vec{x}_2)}e_k(\tilde{\eta}_1)e_k^\star(\tilde{\eta}_2). \qquad (5.14)$$

By inspection the 2-point function is invariant under the action of space-translations. This is an independent proof of the translational invariance of the vacuum state, now geared to the cosmology literature.

Remark: While each choice of the family of solutions $e_k(\tilde{\eta})$ to (5.9) satisfying the normalization condition (5.10) determines a vacuum state (and the associated Hilbert space $\mathcal{H}_1$ of states), this is a many to one map: The families $e_k(\tilde{\eta})$ and $e^{i\theta_k}e_k(\tilde{\eta})$ that differ just by a k-dependent phase factor determine the same vacuum. There is a 1-1 correspondence between the equivalence classes $\{e_k(\tilde{\eta})\}$ of families that differ by such phase factors and complex structures J on the phase space $\Gamma_{\mathcal{T}}$ which are compatible with the symplectic structure and are invariant under the action of space-translations. Thus there is a 1-1 correspondence between the vacua $|0\rangle$ we have constructed and complex structures on $\Gamma_{\mathcal{T}}$ satisfying the two properties listed above.

A natural question about these representations of the Weyl or the Heisenberg algebra is now the following: Does a change of the complex structure J—i.e., the choice of "generalized positive frequency solutions" $\{e_k(\tilde{\eta})\}$—always result in unitarily equivalent representations? As is well-known, in general the answer is in the negative. In the terminology used in the cosmology literature, in general the vacuum state selected by any one complex structure may contain an infinite number of particles corresponding to another complex structure. *A priori* this would be a key obstacle to extracting physics because different choices would in general lead to very different predictions. Furthermore, in a general representation so constructed, there is no natural prescription to regulate products of operator-valued distributions, e.g., $\hat{\mathcal{T}}^2(x, \tilde{\eta})$, and hence to define basic physical operators such as the Hamiltonian of (4.12).

Fortunately, as we will see in Sec. VI, both of these problems can be resolved in one stroke by imposing certain regularity requirements on the basis functions $\{e_k(\tilde{\eta})\}$. Then, the representations of the Weyl algebra that result from equivalence classes $\{e_k(\tilde{\eta})\}$ of any of the regular basis functions—or "regular" complex structures J—will turn out to be unitarily equivalent. In this sense there is a unique class of unitarily equivalent representations and one can work with a unique Hilbert space $\mathcal{H}_1$. Therefore, the more general framework of algebraic quantum field theory is not essential in the cosmological context under consideration.[9]

---

[9]See also Refs. [77–79] where one arrives at a rigorous uniqueness result using the form (3.34) of equations of motion. However, the "vacua" they are led to consider are not necessarily of 4th adiabatic order whence it would be difficult to regularize and renormalize physically interesting composite operators, such as the energy density, in that framework. It would be interesting to see in detail the precise relation between that approach and the adiabatic treatment pursued here and in much of the cosmological literature.

All the regular translationally invariant vacua and states containing a finite number of excitations over any of them belong to the Hilbert space $\mathcal{H}_1$. However, the translationally invariant vacua span an *infinite dimensional* subspace of $\mathcal{H}_1$ and none of them is preferred from a full space-time perspective. Thus, $\mathcal{H}_1$ will not admit a "canonical" vacuum we are used to in Minkowski or (strictly) stationary space-times.

VI. REGULARITY CONDITIONS ON STATES AND OPERATORS

In this section we will first summarize the notion of regularity conditions on states ψ of quantum perturbations and then use it to regulate products of operator-valued distributions, such as the ones that appear in the quantum stress-energy tensor or Hamiltonian. There exist several methods of regularization. We will work with the adiabatic scheme because it is particularly suitable to perform explicit computations, including numerical implementations, and can be directly extended to our quantum field theory on *quantum* FLRW geometries. In this paper of course we will focus on the $k = 0$, $\Lambda = 0$ case but our considerations will extend to other contexts such as the $k = 1$ FLRW case where the underlying isometries make a mode decomposition naturally available.

Much of the discussion in the first two sub-sections is taken from the rich literature on adiabatic regularization in cosmology (see, e.g., Refs. [8,9,93,94]). But there are two new elements as well: (i) the specific formulation of the adiabatic condition (which is succinct and yet clarifies some subtleties); and (ii) the discussion of the regularized Hamiltonian operator $\hat{H}_1$.

A. The adiabatic condition

As explained in Sec. V B, a choice of a basis of "generalized positive frequency" solutions $e_k(\tilde{\eta})$ satisfying the normalization condition (5.10) determines a vacuum state, $|0\rangle$, from which a Fock space $\mathcal{H}_1$ can be constructed. Since each of these (complex) basis vectors $e_k(\tilde{\eta})$ satisfies the second order, linear, ordinary differential equation (5.9), any two bases $e_k(\tilde{\eta})$ and $\underline{e}_k(\tilde{\eta})$ are related simply by [95,96]

$$e_k(\tilde{\eta}) = \alpha_k \underline{e}_k(\tilde{\eta}) + \beta_k \underline{e}_k^{\star}(\tilde{\eta}), \qquad (6.1)$$

where the *time-independent* Bogoluibov coefficients α_k and β_k satisfy the relation $|\alpha_k|^2 - |\beta_k|^2 = 1$. Substituting this equation in the expression (5.8) of the field operator $\hat{\mathcal{T}}(\vec{x}, \tilde{\eta})$, we find a linear relation between the creation and annihilation operator associated with the two families

$$\underline{\hat{A}}_{\vec{k}} = \alpha_k \hat{A}_{\vec{k}} + \beta_k^{\star} \hat{A}_{-\vec{k}}^{\dagger}. \qquad (6.2)$$

This relation shows that, as long as the α_k coefficients are not trivial, i.e., $\alpha_k \neq e^{i\theta_k}$ for all k, the associated vacua $|0\rangle$ and $|\underline{0}\rangle$ are distinct. The number of "under-barred" quanta with momentum $\vec{k}$ that the state $|0\rangle$ contains is given by

$$\langle 0 | \hat{\underline{N}}_{\vec{k}} | 0 \rangle := \langle 0 | (\hbar \ell^3)^{-1} \underline{\hat{A}}_{\vec{k}} \underline{\hat{A}}_{\vec{k}}^{\dagger} | 0 \rangle = |\beta_k|^2, \qquad (6.3)$$

where we have used (6.2) and the commutation relation (5.11) in the last step. The right-hand side provides the expected number of "under-barred exitations/particles" with momentum $\vec{k}$ in the vacuum $|0\rangle$. Therefore, at the power counting level, it follows that if $|\beta_k|^2$ does not fall off faster than k^{-3} when $k \to \infty$, the total number of "under-barred quanta" in the vacuum $|0\rangle$ would diverge. In this case, $|0\rangle$ would not belong to the under-barred Fock space; the two representations of the Heisenberg or Weyl algebra would be unitarily inequivalent.

This inequivalence is related to the large k limit and, as indicated in Sec. V B, can arise because so far there is no restriction on the ultraviolet behavior of the basis states $e_k(\tilde{\eta})$. The physical idea behind the appropriate restriction becomes clearer if one works with the variable $\chi_k(\tilde{\eta}) := (\tilde{a}(\tilde{\eta})/2\sqrt{\kappa}) e_k(\tilde{\eta})$, for which (5.9) becomes

$$\chi_k''(\tilde{\eta}) + \left(k^2 - \frac{\tilde{a}''(\tilde{\eta})}{\tilde{a}(\tilde{\eta})} \right) \chi_k(\tilde{\eta}) = 0. \qquad (6.4)$$

Note that (6.4) reduces to the equation satisfied by the standard basis functions in Minkowski space if $\tilde{a}''/\tilde{a} = 0$. Now, $6\tilde{a}''/\tilde{a}^3$ is just the scalar curvature of $\tilde{g}_{ab}$ and introduces a physical length scale $L(\tilde{\eta})$ into the problem. The form of (6.4) suggests that for modes with large momentum, i.e., with $k/\tilde{a} \gg 1/L$, curvature would have negligible effect and they would evolve almost as if they were in Minkowski space. Therefore, it is natural to impose the following regularity condition on the choice of basis functions in the ultraviolet limit: for $k/\tilde{a} \gg 1/L$, $\chi_k(\tilde{\eta})$ should approach the canonical Minkowski space *positive frequency* solutions $e^{ikt}/\sqrt{2k}$ at an appropriate rate. (In the terminology of Sec. V B, we would then be restricted to a preferred family of complex structures all of which have the same ultraviolet behavior as the canonical complex structure in Minkowski space-time.) This is the crux of the idea behind adiabatic condition.

To make this idea precise, we have to sharpen the required rate of approach. For this, one first introduces a set of specific "generalized WKB" solutions to (6.4) that approach the Minkowski space positive frequency modes as $k \to \infty$ in a controlled fashion. In a second step, one requires that the permissible basis functions $\chi_k(\tilde{\eta})$—which are exact solutions to (6.4)—should approach the WKB solutions to the desired order.

The generalized WKB solutions $\chi_k^{(N)}(\tilde{\eta})$ of order N are given by [93][10]:

---

[10]For this prescription to be well-defined, $W_k^{(N)}(\tilde{\eta})$ must be nonnegative. For any given smooth $\tilde{a}(\tilde{\eta})$, this can be ensured by going to high enough k. Since it is only the behavior of $W_k^{(N)}(\tilde{\eta})$ in the $k \to \infty$ that matters for the adiabiticity considerations, if $W_k^{(N)}$ were to become negative in some k-range, one can just suitably modify its form for low k.

$$\chi_k^{(N)}(\tilde{\eta}) = \frac{1}{\sqrt{2W_k^{(N)}(\tilde{\eta})}} e^{-i\int^{\tilde{\eta}} W_k^{(N)}(\tau)d\tau}, \quad (6.5)$$

where $W_k^{(N)}(\tilde{\eta}) = W_0 + W_1 + \cdots + W_N$, with

$$W_0 = k; \qquad W_2 = -\frac{1}{2k}\frac{\tilde{a}''}{\tilde{a}};$$

$$W_4 = \frac{2\tilde{a}''\tilde{a}'^2 - 2\tilde{a}''^2\tilde{a} - 2\tilde{a}\tilde{a}'\tilde{a}''' + \tilde{a}^2\tilde{a}''''}{8k^3\tilde{a}^3}; \ldots \quad \text{and}$$

$$W_i = 0 \quad \text{if } i \text{ is odd.} \quad (6.6)$$

Note that, because the lower bound on the integral in the phase factor is not fixed, $\chi_k^{(N)}$ is well-defined only up to an overall phase (which is time independent but can depend on k). Each $\chi_k^{(N)}(\tilde{\eta})$ is an approximate solution to (6.4) in the sense that, when we operate on it by the operator on the left side of (6.4), the result does not vanish but is given by terms of the order $\mathcal{O}(\tilde{a}/kL_{N+2})^{N+1/2}$ where the length scale L_N is dictated by $\tilde{a}$ and its $\tilde{\eta}$-derivatives to order N. The leading order term of (6.5) corresponds to the positive frequency solution in Minkowski space and the rest of the terms are higher order contributions that vanish at different rates when $(\tilde{a}/kL_N) \to 0$. Finally, this approximate solution can also be regarded as an expansion in the number of derivatives of the scale factor. This is because, since the limit refers to $(\tilde{a}/kL_N)$, one can either keep L_N fixed and let $\tilde{a}/k$ go to zero as we have done so far, or keep $\tilde{a}/k$ fixed and let $1/L_N$ go to zero, which corresponds to letting the expansion rate go to zero, i.e., considering adiabatic expansion. This why, although this method is primarily concerned with ultraviolet issues, it is referred to as "adiabatic regularization."[11]

We are now ready to state the "adiabatic condition" that imposes the desired ultraviolet regularity on the basis functions: *In the mode expansion* (5.8), *choose only those solutions* $e_k(\tilde{\eta}) = 2\sqrt{\kappa}\chi_k(\tilde{\eta})/\tilde{a}(\tilde{\eta})$ *to* (5.9) *which agree with* $\chi_k^{(N)}(\tilde{\eta})$ *up to terms of order* $(\tilde{a}/kL_N)^{N+1/2}$. More precisely, we require: $|\chi_k| = |\chi_k^{(N)}|(1 + \mathcal{O}((\tilde{a}/kL_{N+2}))^{N+2})$, and the same relation should hold if χ_k and $\chi_k^{(N)}$ are replaced by their 1st, 2nd, ... $(N-1)$th time derivatives. (Note that the absolute value signs make these conditions well-defined in spite of the phase ambiguity in $\chi_k^{(N)}$.) These bases $e_k(\tilde{\eta})$ will be referred to as Nth-order adiabatic solutions, the associated vacuum states will be referred to as the Nth-*order vacua*, and states obtained by operating on these vacua by (arbitrarily large but finite) sums of products of creation

operators as Nth-order adiabatic states. From now on, we will restrict ourselves to Nth-order adiabatic states and, for reasons explained in Sec. VI B, for most part we will set $N = 4$. There is a large body of literature on the notion of adiabatic states and their properties. For further details, see in particular [8,9,93].

This framework has two important features.

(i) The adiabatic condition is only an *asymptotic* restriction for large $kL/\tilde{a}$. Therefore, for any given N, there are infinitely many families of solutions $e_k(\tilde{\eta})$ which satisfy it. Each of these bases defines an adiabatic vacuum and, if two bases are nontrivially related (i.e., if the Bogoluibov coefficients α_k are not pure phases for all k), the corresponding vacua are distinct. Thus, in striking contrast to the free field theory in Minkowski space, there is no preferred vacuum state, nor a canonical notion of "particles." Any one vacuum appears as an "excited state with many particles" with respect to another vacuum.

(ii) However, if $N \geq 2$, all adiabatic vacua—and hence all adiabatic states—lie in the same Hilbert space $\mathcal{H}_1$. This is because if $N \geq 2$, then $\sum |\beta_k|^2 < \infty$, whence any one adiabatic vacuum has only a finite number of particles relative to any other.

This completes the specification of the Hilbert $\mathcal{H}_1$ of perturbations we began in Sec. V B. While we used the framework originally developed for quantum field theory in classical space-times, because we used the dressed effective metric $\tilde{g}_{ab}$ for our background, it follows from Sec. IV that $\mathcal{H}_1$ is the Hilbert space of quantum perturbations ψ propagating on the quantum geometry Ψ_o.

Remark: Had we chosen to work with spatial manifolds $\mathbb{M}$ with $\mathbb{R}^3$ topology rather than $\mathbb{T}^3$, Eq. (6.3) would be replaced by

$$\langle 0|\hat{\underline{N}}_{\vec{k}}|0\rangle := \langle 0|\hbar^{-1}\hat{\underline{A}}_{\vec{k}}\hat{\underline{A}}_{\vec{k}}^{\dagger}|0\rangle = |\beta_k|^2(2\pi)^3\delta^3(0). \quad (6.7)$$

This implies that $|\beta_k|^2$ is now the number *density* of the under-barred exitations/particles—with momentum $\vec{k}$—in the vacuum $|0\rangle$ *per unit volume* of $\mathbb{M}$ and per unit co-moving volume in the momentum space. The $\delta^3(0)$ (in the momentum space) in (6.7) arises because of the infinite spatial volume of $\mathbb{M}$. Thus, in addition to the potential ultraviolet divergence that would occur if $|\beta_k|$ does not fall off appropriately for large k, we now also have an infrared divergence. Note that this cannot be cured simply by putting an infrared cutoff in the $\vec{k}$ space: the *total* number of particles created with momenta $\vec{k}$ within *any finite range* Δ_k also diverges because of the infinite spatial volume of $\mathbb{M}$. In particular, this divergence persists for massive fields as well; it arises because we now have an infinite volume *and* a homogeneous background. This then implies that the Fock representations of the Heisenberg or Weyl algebra associated with any two bases $e_k(\tilde{\eta})$ and $\underline{e}_k(\tilde{\eta})$ are unitarily inequivalent unless α_k is a pure phase for all k. But physically this infinity is spurious. Therefore, for $\mathbb{R}^3$

---

[11]Chronologically, the term *adiabatic* originated from the fact that the first application of this method—introduced by Parker in Ref. [95]—was to the problem of regularizing the number operator in an expanding universe, and the condition that the particle number be an adiabatic invariant was used as a fundamental requirement in the construction.

topology, a notion of "physical equivalence" is more appropriate than that of "unitary equivalence." To introduce it, recall first that if the topology is $\mathbb{R}^3$ one has to introduce a fiducial cell $\mathcal{C}$ and restrict all integrations to it already at the classical level. (From a physical perspective one can choose the cell so that its physical volume is larger than the volume of the observed universe.) Two Fock representations of perturbations would be regarded as *physically equivalent* if the vacuum state associated with one contains a finite number of "exitations/particles" with respect to the other within any region contained in $\mathcal{C}$. Then, if we restrict ourselves to adiabatic vacua of order $N \geq 2$, we are ensured of physical equivalence. If we require $N \geq 4$, the expectation values of the regularized stress-energy tensor would be well-defined distributions and we can restrict the support of test functions in $\mathcal{C}$.

B. Regularization of composite operators

In this subsection we summarize the necessary regularization procedure to obtain physical information from the formal expressions of composite operators on the Hilbert space $\mathcal{H}_1$. Adiabatic regularity of the basis modes $e_k(\tilde{\eta})$ will provide the necessary control on the ultraviolet divergences in the expectation values of composite operators, leading to state *independent* criteria to extract the physical, finite results while respecting the underlying covariance of the theory.

Consider a formal operator $\hat{O}(\vec{x}, \tilde{\eta})$ which is at least quadratic in the field operator and its conjugate momentum, and is factor ordered to be self-adjoint. Examples of direct interest are the stress-energy tensor and the Hamiltonian. Consider the expectation value of this operator with respect to an adiabatic vacuum selected by a basis $\chi_k(\tilde{\eta})$. Using the mode decomposition (5.8) and the commutation relation (5.11), the vacuum expectation value of $\hat{O}(\vec{x}, \tilde{\eta})$ can be expressed as a formal sum in the momentum space, of the type

$$\langle 0|\hat{O}(\vec{x}, \tilde{\eta})|0\rangle_{\text{formal}} = \frac{1}{\ell^3} \sum_{\vec{k}} O_k[\chi_k(\tilde{\eta})], \quad (6.8)$$

where there is no x-dependence on the right side because $|0\rangle$ is translationally invariant. Generically, the sum would be ultraviolet divergent. In adiabatic regularization the physically relevant, finite expression is obtained by subtracting, *mode by mode*, each term in the adiabatic expansion of $O_k[\chi_k(\tilde{\eta})]$ that contains at least one ultraviolet divergent piece [93,95] (see also Refs. [8,9]). Thus, if $O^{(N)}$ is the Nth adiabatic order term in the expansion of the summand, if any one part of $O^{(N)}$ is divergent, the entire term $O^{(N)}$ is subtracted (including parts that have no ultraviolet divergences). On the other hand, following the *criterion of minimal subtraction*, this procedure is applied only up to that order in the adiabatic expansion at which the formal expression has divergent pieces. Thus, if the

Mth adiabatic order term has no divergent part, then nothing is subtracted at order M.

The most interesting example for us is the stress-energy tensor of gauge invariant tensor perturbations since it plays the key role in checking if the truncation scheme is self-consistent, i.e., whether or not the backreaction can be neglected. For classical fields $\mathcal{T}(\vec{x})$, the expression is given by

$$T_{ab} = \frac{1}{4\kappa}\left[\tilde{\nabla}_a \mathcal{T}\tilde{\nabla}_b \mathcal{T} - \frac{1}{2}\tilde{g}_{ab}\tilde{g}^{cd}\tilde{\nabla}_c \mathcal{T}\tilde{\nabla}_d \mathcal{T}\right]. \quad (6.9)$$

At the quantum level the stress-energy tensor is a composite operator of dimension four, and in four space-time dimensions we expect ultraviolet divergences up to fourth order in the co-moving momentum k. Let us first consider energy density operator $\hat{\rho}(\vec{x}, \tilde{\eta})$. Given a basis $e_k(\tilde{\eta}) = 2\sqrt{k}\chi_k(\tilde{\eta})/\tilde{a}(\tilde{\eta})$ of fourth or higher adiabatic order, the formal expression for the expectation value of $\hat{\rho}(\vec{x}, \tilde{\eta})$ in the associated adiabatic vacuum is

$$\langle 0|\hat{\rho}|0\rangle_{\text{formal}} := -\langle 0|\hat{T}_0^0|0\rangle_{\text{formal}} = \frac{\hbar}{\ell^3 \tilde{a}^4(\tilde{\eta})} \sum_{\vec{k}} \rho_k[\chi_k(\tilde{\eta})]$$

$$= \frac{\hbar}{2\ell^3 \tilde{a}^4(\tilde{\eta})} \sum_{\vec{k}} |\chi_k'|^2 + \left(k^2 + \frac{\tilde{a}'^2}{\tilde{a}^2}\right)|\chi_k|^2$$

$$- 2\frac{\tilde{a}'}{\tilde{a}}\,\text{Re}(\chi_k\chi_k'^\star). \quad (6.10)$$

By using the adiabatic expansion of each χ_k in the above summand, it is easy to see that all the ultraviolet divergences are contained in terms of adiabatic order equal to and smaller than four. The zeroth adiabatic order term produces a $\sum(1/k^3)$ divergence in co-moving momentum k; the second order term a $\sum(1/k^2)$ one; and the fourth order term a $\sum(1/k)$ one. Therefore, the subtraction terms $C^\rho(k, \tilde{\eta})$ needed to regularize the energy density are obtained from the terms of zeroth, second and fourth adiabatic order in the expansion of the summand:

$$C^\rho(k, \tilde{\eta}) = \rho_k^{(0)} + \rho_k^{(2)} + \rho_k^{(4)}$$

$$= \frac{k}{2} + \frac{\tilde{a}'^2}{4\tilde{a}^2 k} + \frac{4\tilde{a}'^2\tilde{a}'' + \tilde{a}\tilde{a}''^2 - 2\tilde{a}\tilde{a}'\tilde{a}'''}{16\tilde{a}^3 k^3}. \quad (6.11)$$

Thus, the vacuum expectation value of the renormalized energy density is

$$\langle 0|\hat{\rho}|0\rangle_{\text{ren}} = \frac{\hbar}{\ell^3 \tilde{a}(\tilde{\eta})^4} \sum_{\vec{k}} (\rho_k[\chi_k(\tilde{\eta})] - C^\rho(k, \tilde{\eta})). \quad (6.12)$$

Note that the subtraction terms C^ρ are local in the background geometry and, even more importantly, are *state independent*. The expectation value in any fourth order state in $\mathcal{H}_1$ is computed by the same procedure, using the *same subtraction terms*.

The only other independent degree of freedom in the stress-energy tensor in a homogenous and isotropic background is the trace T (or the pressure $p = (\rho + T)/3$). The corresponding vacuum expectation value can be calculated using results given in [94]:

$$\langle 0|\hat{T}|0\rangle_{\text{ren}} = \frac{-\hbar}{\ell^3 \tilde{a}^4(\tilde{\eta})} \sum_{\vec{k}} - |\chi'_k|^2 + \left(k^2 - \frac{\tilde{a}'^2}{\tilde{a}^2}\right)|\chi_k|^2$$
$$+ 2\frac{\tilde{a}'}{\tilde{a}} \text{Re}(\chi_k \chi'^{\star}_k) - C^T(k, \tilde{\eta}), \quad (6.13)$$

where the 4th adiabatic order subtraction terms are

$$C^T(k, \tilde{\eta}) = \frac{-\tilde{a}'^2 + \tilde{a}\tilde{a}''}{2\tilde{a}^2 k}$$
$$- \frac{6\tilde{a}'^2\tilde{a}'' - 3\tilde{a}\tilde{a}''^2 - 4\tilde{a}\tilde{a}'\tilde{a}''' + \tilde{a}^2\tilde{a}''''}{8\tilde{a}^3 k^3}. \quad (6.14)$$

Note that both vacuum expectation values, (6.12) and (6.13), are constant functions on $\mathbb{M}$ as they must be given that the vacuum is translationally invariant.[12] If the vacuum were replaced by a generic 4th order adiabatic state, this constancy will not hold. In that case the expectation value would be distributions in $\vec{x}$. Finally, by themselves these expectation values only provide a quadratic form on the Hilbert space $\mathcal{H}_1$. However, recent results show that they are the expectation values of an operator valued distribution $\hat{T}_{ab}$ on $\mathcal{H}_1$ [97].

A natural question now arises: Does the vacuum expectation value of the renormalized stress-energy tensor satisfy covariant conservation $\tilde{\nabla}_a \langle \hat{T}^{ab} \rangle_{\text{ren}} = 0$ with respect to the dressed effective background $\tilde{g}_{ab}$? In any homogenous and isotropic background, the conservation equation reduces to $\langle \hat{\rho} \rangle'_{\text{ren}} + \frac{\tilde{a}'}{\tilde{a}}(4\langle \hat{\rho} \rangle_{\text{ren}} + \langle \hat{T} \rangle_{\text{ren}}) = 0$. The formal, unrenormalized expressions for $\langle \hat{\rho} \rangle$ and $\langle \hat{T} \rangle$ satisfy this relation mode by mode as a consequence of the wave equation satisfied by the χ_k. The adiabatic subtraction terms also satisfy the conservation equation mode by mode. Thus, one can directly verify that the expectation value of the renormalized stress-energy is indeed conserved (see, for instance, Ref. [98]).

This adiabatic regularization of $\hat{T}_{ab}$ also has the desired properties enunciated in Wald's axioms [92]: it reduces to the standard normal ordering in the flat space-time limit; the subtractions terms are constructed from local information in the background geometry, and, as already noted, the renormalized stress energy is conserved; $\tilde{\nabla}_a \langle \hat{T}^{ab} \rangle_{\text{ren}} = 0$.

Remark: As noted in Sec. III, physically it is appropriate to introduce an infrared cutoff by absorbing into the

background the modes whose physical wave length is larger than the physical radius of the observable universe. Since physical wave lengths scale linearly with the scale factor $\tilde{a}$, this can be achieved in a time independent fashion by imposing a cutoff, k_{IR}, in the co-moving wave number k. As is clear from the above discussion of conservation of stress-energy, the vacuum expectation value of the new renormalized stress-energy tensor will also be covariantly conserved. Furthermore, by construction, the renormalized energy density and pressure will again be constant on $\mathbb{M}$.

We will conclude this section with renormalization of the Hamiltonian operator $\hat{H}_1$ that generates the dynamics of perturbations in conformal time $\tilde{\eta}$. The form (3.31) of the classical Hamiltonian and the fact that the lapse corresponding to the conformal time is $N = a$ imply that the Hamiltonian operator has the following formal expression:

$$\hat{H}_{1,\text{formal}} = \frac{1}{2\mathring{V}} \sum_{\vec{k}} \frac{4\kappa}{\tilde{a}^2} |\hat{\mathfrak{p}}_{\vec{k}}|^2 + \frac{\tilde{a}^2}{4\kappa} k^2 |\hat{\mathcal{T}}_{\vec{k}}|^2$$
$$= \tilde{a}^4 \int d^3x \hat{\rho}_{\text{formal}}. \quad (6.15)$$

Therefore, the renormalized Hamiltonian is given by

$$\hat{H}_{1,\text{ren}} = \frac{1}{2\mathring{V}} \sum_{\vec{k}} \frac{4\kappa}{\tilde{a}^2} |\hat{\mathfrak{p}}_{\vec{k}}|^2 + \frac{\tilde{a}^2}{4\kappa} k^2 |\hat{\mathcal{T}}_{\vec{k}}|^2 - \hbar \mathring{V} C^{\rho}(k, \tilde{\eta}). \quad (6.16)$$

Our discussion of energy density immediately implies that the expectation values of $\hat{H}_{1,\text{ren}}$ are well-defined on any 4th order adiabatic state. Furthermore, by taking its commutators with $\hat{\mathcal{T}}_{\vec{k}}, \hat{\mathfrak{p}}_{\vec{k}}$ we recover the Heisenberg equations of motion (4.18) satisfied by $\hat{\mathcal{T}}_{\vec{k}}$. However, we do not have a proof that $\hat{H}_{1,\text{ren}}$ is a self-adjoint operator (in the precise sense that it is densely defined and its domain equals that of its adjoint. In particular, the strategy of using the Friedrich extension does not work because the operator is not positive definite.) Although we expect this to be the case, we also note that, quite generally in cosmology, proving self-adjointness involves nontrivial subtleties and technicalities especially because the Hamiltonians have a nontrivial time dependence.

Remark: In the case when $\mathbb{M}$ is topologically $\mathbb{R}^3$, one continues to use the mode by mode subtraction strategy but one has to replace $(1/\ell^3)\sum_{\vec{k}}$ by $1/(2\pi)^3 \int d^3k$. Again, the renormalized stress-energy is conserved, the subtraction terms are local and the prescription reduces to the standard normal ordering in Minkowski space-time.[13] Again it is

---

[12]As a consequence, while the summands on the right side of these expressions can be interpreted as the contribution of the mode $\vec{k}$ to $\langle \hat{\rho}(x) \rangle_{\text{ren}}$ and $\langle \hat{T}(x) \rangle_{\text{ren}}$, they are *not* the Fourier transforms $\langle \hat{\rho}_k \rangle_{\text{ren}}$ and $\langle \hat{T}_k \rangle_{\text{ren}}$ of these quantities in the x-space.

[13]Counter terms now appear under an integral sign. The results of Refs. [94,99] imply that one can use this integral form of counter terms also in the spatially compact cases. This procedure is justified on the grounds that the ultraviolet regularization should not be sensitive to the global topology and has, furthermore, some advantages.

physically appropriate to introduce an infrared cutoff in the co-moving k-space and conservation of stress-energy persists after imposing this cutoff. In addition, as shown in Refs. [94,99], in FRLW space-times this adiabatic regularization is equivalent to the point-splitting Hadamard renormalization. Therefore, although the procedure is not manifestly covariant because of the mode by mode subtraction, the result is fully covariant w.r.t. $\tilde{g}_{ab}$. From a fundamental quantum consideration, $\tilde{g}_{ab}$ is only a convenient mathematical construction and the true quantum geometry is encoded in Ψ_o. Still, it is desirable that the effective description have space-time covariance.

C. Narrowing down initial conditions

As discussed in Sec. VI A, our quantum Hilbert space $\mathcal{H}_1$ of perturbations does not admit a preferred vacuum state. In quantum field theories on Minkowski or de Sitter space-time, underlying isometries serve as powerful tools to single out preferred states. Our quantum FLRW background Ψ_o is invariant under the three-dimensional translation group acting on $\mathbb{M} = \mathbb{T}^3$. It is then natural to seek preferred states in $\mathcal{H}_1$ by demanding that they also be translationally invariant. As discussed in Sec. V, this condition leads us to the infinite dimensional space of vacua which, in view of our discussion in Sec. VI B, we now require to be of 4th adiabatic order. If one requires the state to satisfy these properties initially, i.e., at the bounce, then they are satisfied for all times. This is the family of "preferred" states selected by the symmetry and regularity requirements. As we will discuss in detail elsewhere, this choice is well-suited to formulate a quantum version of Penrose's Weyl curvature hypothesis [100]. Furthermore, in the inflationary context, one can motivate this choice using physical considerations based on properties of the new repulsive force with origin in quantum geometry that dominates the dynamics at and near the bounce [1,2].

At a fundamental level, we allow any of these vacua as initial conditions at the bounce provided, of course, the energy density $\langle \hat{\rho} \rangle$ in the perturbations at the bounce is negligible compared to the energy density $\langle \hat{\rho}_o \rangle \sim 0.41 \rho_{\mathrm{Pl}}$ in the background quantum geometry Ψ_o. However, for specific calculations and especially detailed numerical simulations, one has to work with specific states. Are there then especially convenient vacua to work with? The adiabatic procedure summarized in Sec. VI A provides a strategy to select an "obvious" candidate, *provided one fixes an instant of time $\tilde{\eta}_0$*. Recall that to ensure 4th order adiabaticity we required that, for $kL_4/a \gg 1$, the basis functions χ_k must agree with a specific approximate solution $\chi_k^{(4)}$ defined in (6.5) at least up to terms of adiabatic order four. Therefore, given an instant $\tilde{\eta}_0$ of time, we can construct a "natural" basis χ_k^{obv} by asking that it has the same initial data at that time as $\chi_k^{(4)}$. Thus the idea is to ask for solutions $\chi_k^{\mathrm{obv}}(\tilde{\eta})$ to the exact evolution equation (6.4) which satisfy

$$\chi_k^{\mathrm{obv}}(\tilde{\eta}_0) = \chi_k^{(4)}(\tilde{\eta}_0); \quad \text{and} \quad \partial_{\tilde{\eta}} \chi_k^{\mathrm{obv}}(\tilde{\eta}_0) = \partial_{\tilde{\eta}} \chi_k^{(4)}(\tilde{\eta}_0).$$
$$(6.17)$$

Since $\chi_k^{(4)}$ are only approximate solutions to (6.4), they will not agree with χ_k^{obv} at any other time $\tilde{\eta} \neq \tilde{\eta}_0$. Nonetheless, $\chi_k^{\mathrm{obv}}(\tilde{\eta})$ are automatically of 4th adiabatic order for all times $\tilde{\eta}$. We will call the associated vacuum state the *obvious 4th-order adiabatic vacuum at the time $\tilde{\eta}_0$*. In LQC, the preferred instant of time required in this strategy is provided by the bounce.

Note, however, that for this strategy to work, the quantity $W_k^4(\tilde{\eta}_0)$ in (6.5) must be non-negative since it appears under a square-root in the expression of $\chi_k^{(4)}(\tilde{\eta}_0)$. Now, as we pointed out in Sec. III, there is a natural *physical* infrared cutoff k_{IR} provided by the radius of the observable universe. Typically, for $k > k_{\mathrm{IR}}$, i.e., for the relevant k, $W_k^{(4)}$ is positive. But if k_{IR} is too low for this to happen, as explained in Sec. VI A, one has to suitably modify the form of $W_k^4(\tilde{\eta}_0)$ for low k. In this case, even for a fixed value $\tilde{\eta}_0$ of time, there would be an ambiguity in the choice of the basis χ_k for low k, and hence also in the resulting 4th adiabatic order vacuum $|0\rangle_{\tilde{\eta}_0}^{\mathrm{obv}}$.

This vacuum state $|0\rangle_{\tilde{\eta}_0}^{\mathrm{obv}}$ is especially convenient to work with in numerical simulations. Therefore, it serves as a technically powerful tool to establish the viability of conjectures—e.g., that, in the inflationary context, there exist quantum states for which the backreaction can be ignored—and to probe qualitative features of quantum dynamics. However, even when there is a preferred instant of time, such as the bounce time in LQC, and the infrared cutoff is sufficiently large for the state to be unique, there is no physical principle that singles out this vacuum over other 4th adiabatic order vacua. A better strategy would be to narrow down the choice of vacua by adding external inputs suggested by the physics of the problem. In the present analysis there is a natural avenue along these lines. Since the test field approximation plays a central role, one could ask: Are there 4th adiabatic order vacua for which the expectation value of the renormalized energy density is *exactly* zero at a given time? The answer to this question is in the affirmative and, although this condition does not single out a unique state, it reduces the choices considerably. This issue will be discussed in detail in a separate publication [101].

D. A criterion for self-consistency of the truncation scheme

Our results in Secs. V, VI A, VI B, and VI C provide us with a specific quantum theory, corresponding to the truncated phase space $\tilde{\Gamma}_{\mathrm{Trun}}$. This self contained mathematical framework enables one to describe quantum dynamics of the truncated system, once the initial state is specified, say at the bounce. *The key question now is whether the truncated theory has an interesting domain of validity.* The

basic assumption behind truncation is that the backreaction due to the stress-energy tensor T_{ab} of the first order fields can be neglected (compared to the background stress-energy tensor T^o_{ab}) during the dynamical phase of interest. Already in the classical theory, it is clear that generic first order perturbations violate this condition. The interesting question there is rather the following: Is there a sufficiently large subspace in $\tilde{\Gamma}_{\rm Trun}$ on which this condition is satisfied? More precisely, if we restrict ourselves to an instant of time and choose perturbations which do meet this condition, does the condition continue to be satisfied over the period of evolution of interest? We will now formulate the analogs of these questions in the quantum theory.

Somewhat surprisingly, in the quantum theory one encounters some novel features. First, in the classical theory the only preferred time to specify the initial conditions would be the big bang which is singular. In the quantum theory, by contrast, the bounce provides a natural time $\tilde{\eta}_{\rm B}$ for this task. The second feature is equally significant but also more subtle. Because of the underlying symmetries of the background, it is natural to require the initial state ψ at time $\tilde{\eta}_{\rm B}$ to be translationally invariant. An immediate consequence is that the expectation value $\langle \hat{T}_{ab} \rangle$ of the renormalized stress-energy tensor operator is then homogeneous *for all times.*[14] This property has the important consequence that the *second order* perturbations are again homogeneous and isotropic since their sole source is $\langle \hat{T}_{ab} \rangle$. Thus, thanks to the symmetry of our initial state ψ at the bounce, the backreaction can change only the zeroth order, homogeneous fields, changing the total state $\Psi_o \otimes \psi$ to a nearby state of the type $(\Psi_o + \delta\Psi_o) \otimes \psi$. Our initial conditions on ψ at $\tilde{\eta} = \tilde{\eta}_{\rm B}$ guarantee that the shift $\delta\Psi_o$ at this initial time is negligible. The key question then is whether it continues to remain negligible under time evolution.

On general grounds one would say that the answer is dictated by the time dependence of $\langle \hat{T}_{ab} \rangle$. But in the detailed mathematical framework, it is the second order truncation $\hat{\mathbb{S}}_2$ of the Hamiltonian constraint that determines the change $\delta\Psi_o$ in the background quantum geometry and in this equation it is only $\langle \hat{\rho} \rangle$—rather than the full $\langle \hat{T}_{ab} \rangle$—that enters as the source. Let us suppose that the energy density $\langle \hat{\rho} \rangle$ is negligible compared to the background energy density $\langle \hat{\rho}_o \rangle$ from $\tilde{\eta}_{\rm B}$, until some time $\tilde{\eta}_0$ of interest. Then it follows that $\delta\Psi_o$ would continue to be negligible from $\tilde{\eta}_{\rm B}$ to $\tilde{\eta}_0$. It may seem somewhat surprising at first that one does not have to require explicitly that the other independent component $\hat{T}$ of $\hat{T}_{ab}$ should be small. This is because we are requiring that $\langle \hat{\rho} \rangle$ be negligible

compared to $\langle \hat{\rho}_o \rangle$ not just initially but *for all times between* $\tilde{\eta}_{\rm B}$ and $\tilde{\eta}_0$. The difficult part of the calculation is to check that the evolution of ψ is such that this condition does holds. If it does, then $\Psi_o \otimes \psi$ would provide a solution in which the backreaction is negligible even in the Planck era. This would then be a self-consistent solution to the quantum truncated theory. *Thus, a sufficient condition for self-consistency of the truncation approximation is that the energy tensity in the quantum perturbations should remain small compared to that in the background from $\tilde{\eta}_{\rm B}$ to $\tilde{\eta}_0$.*

In Ref. [2] we will use detailed numerical simulations to show that the inflationary scenario does admit states $\Psi_o \otimes \psi$ in which $\langle \hat{\rho} \rangle$ is negligible compared to the background $\langle \hat{\rho}_o \rangle$ for all times between the bounce and the onset of the slow roll inflation. Furthermore, given a state satisfying this condition, we will show that there is an open neighborhood of ψ such that the condition continues to be satisfied by $\Psi_o \otimes \tilde{\psi}$ for all $\tilde{\psi}$ in this neighborhood. Thus, there is a rich class of states that provide self-consistent solutions to the truncated quantum theory, demonstrating that the standard inflationary scenario admits a consistent extension all the way back to the big bounce.

But the general framework constructed in this section is not tied to inflation. It provides the technical machinery that is needed to check if *any* given paradigm, based on general relativity and first order cosmological perturbations, admits a *self-consistent* extension to the Planck regime. More precisely, it would enable one to address the following question: In this paradigm, does the quantum theory admit solutions in which the backreaction can be neglected throughout the period of interest, including the Planck era?

VII. SUMMARY AND DISCUSSION

In the last four sections, we developed an extension of the standard cosmological perturbation theory to include the Planck regime of LQG. The strategy was to first truncate classical general relativity coupled to a scalar field to the sector commonly used in the cosmology of the early universe—FLRW space-times and linear inhomogeneous perturbations thereon—and then construct the quantum theory of just this sector using LQG techniques.

Already in the truncation of the classical Hamiltonian theory there is a subtlety that has often been overlooked in the LQC literature: while the dynamics on the homogeneous sector is generated by a (Hamiltonian) constraint, that on the full truncated phase space $\Gamma_{\rm Trun}$ is not. Indeed, because the dynamical vector field X^α on $\Gamma_{\rm Trun}$ fails to Lie drag the full symplectic structure $\Omega_{\rm Trun} = \Omega_o + \Omega_1$ on $\Gamma_{\rm Trun}$, it is not generated by *any* Hamiltonian. Rather, $\Gamma_{\rm Trun}$ is the normal bundle over Γ_o—where the base space Γ_o is regarded as the homogeneous, isotropic subspace of the full phase space Γ of general relativity—and X^α is the lift to $\Gamma_{\rm Trun}$ of the Hamiltonian vector field on Γ_o, induced by the full Hamiltonian vector field on Γ. In the classical

---

[14]Note that this property holds even though the perturbations operators $\mathcal{T}(\vec{x}, \tilde{\eta})$ themselves are *purely inhomogeneous.* In the classical theory, there are no perturbations $\mathcal{T}(\vec{x}, \tilde{\eta})$ which are purely inhomogeneous but their energy density is purely homogeneous.

theory, this subtlety can be ignored if one works with the space of solutions (rather than the phase space), as is common in the standard cosmology literature. But it becomes important for passage to quantum theory if one wishes to treat both the perturbations and the background quantum mechanically. Then there is no conceptual justification for trying to construct dynamics for the *full* truncated system by imposing a quantum constraint. One can do this only on the homogeneous sector, and one then has to "lift" this quantum dynamics to the full Hilbert space just as in the classical theory.

Having constructed the dynamics of gauge invariant variables on the truncated phase space, we then used LQG techniques to construct quantum kinematics: the Hilbert space $\mathcal{H}_o$ of states of background quantum geometry, the Hilbert space $\mathcal{H}_1$ of gauge invariant quantum fields $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ representing perturbations and physically interesting operators on both these Hilbert spaces. The imposition of the quantum constraint on the homogeneous sector leads one to interpret the background scalar field ϕ as a *relational* or *emergent* time variable with respect to which physical degrees of freedom evolve. Furthermore, the background geometry is now represented by a wave function Ψ_o which encodes the probability amplitude for various FLRW geometries to occur. The physically interesting wave functions Ψ_o are sharply peaked, *but the peak follows a bouncing trajectory*, not a classical FLRW solution that originates at the big bang. In addition, Ψ_o has fluctuations about this bouncing trajectory. Quantum fields $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$, representing inhomogeneous scalar and tensor perturbations, propagate on this *quantum* geometry and are therefore sensitive not only to the major departure from the classical FLRW solutions in the Planck regime, but also to the quantum fluctuations around the bouncing trajectory, encoded in Ψ_o. Therefore, at first the problem appears to be very complicated. However, a key simplification made it tractable: *Within the test field approximation* inherent to the truncation strategy, the propagation of $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ on the quantum geometry Ψ_o is completely equivalent to that of their propagation on a specific, quantum corrected FLRW metric $\tilde{g}_{ab}$. Although $\hbar$ does appear in its coefficients, this "dressed, effective metric" $\tilde{g}_{ab}$ is smooth and allows us to translate the evolution of $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ with respect to the relational time to that in terms of the conformal (or proper) time of $\tilde{g}_{ab}$. Furthermore, away from the Planck regime, $\tilde{g}_{ab}$ satisfies Einstein's equations to an excellent approximation. In this sense, the standard quantum field theory of $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ emerges from the more fundamental description of these fields evolving on the quantum geometry Ψ_o with respect to the relational time ϕ. This exact relation between quantum fields $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ on the quantum geometry Ψ_o and those on the dressed, effective geometry of $\tilde{g}_{ab}$ enabled us to carry over adiabatic regularization techniques from quantum field theory in curved space-times

to those on quantum geometries Ψ_o. Together, all this structure provides us with a well-defined quantum theory of the truncated phase space we began with.

This framework has a broad range of applicability because scenarios of the early universe are often based on linear perturbations on FLRW backgrounds. Our construction provides an avenue to extend them all the way to the quantum gravity era because quantum perturbations now propagate on a quantum geometry which is completely regular, with specific upper bounds for curvature and density in the background. We can therefore use the new framework to reexamine the "trans-Planckian issues" encountered in these scenarios. Note first that, the truncated theory under consideration here allows modes with trans-Planckian frequencies. There is no obstruction because the quantum geometry underlying LQG is subtle: In particular, while there is a minimum nonzero eigenvalue of the area operator, there is no such minimum for the volume or length operators even though their eigenvalues are also discrete. *The real danger is not the existence of such modes but rather that the energy density in these modes may not be negligible compared to that in the quantum background geometry.* If this occurs, our quantum theory of the truncated sector would not be viable. Whether this can happen is a very nontrivial issue especially in the Planck regime immediately following the bounce. Heuristically, if the state has just a few excitations each carrying say, 10^6 times the Planck energy in a cm^3 volume, there would be no difficulty (since the energy *density* would be negligible). If on the other hand there is one such excitation per Planck volume, our truncation approximation will fail. Then we cannot neglect backreaction. This is not an impasse to quantum theory as such, but the proper treatment of such states will have to await full LQG.

The key question then is whether the test field approximation underlying this truncation scheme is satisfied. *A priori* this is a difficult issue and, to our knowledge, had not been considered in the literature because even to *formulate* this question precisely one needs the notion of the renormalized stress-energy tensor on the Hilbert space $\mathcal{H}_1$ of quantum perturbations propagating on the quantum geometry of Ψ_o. In our framework this was provided by "lifting" the adiabatic techniques of Refs. [93,94] to quantum fields on quantum geometries Ψ_o. Specifically, by appealing to symmetry principles and regularity requirements, we argued that it was appropriate to focus attention on those states $\Psi_o \otimes \psi \in \mathcal{H}_o \otimes \mathcal{H}_1$ of the combined system for which:

(1) ψ is invariant under the translational symmetry of Ψ_o;

(2) ψ is a 4th order adiabatic state w.r.t. $\tilde{g}_{ab}$; and

(3) *at the bounce*, the energy density $\langle \hat{\rho} \rangle$ in the state ψ is negligible compared to the energy density $\langle \hat{\rho}_o \rangle$ in the background.

We then showed that the truncation approximation is self-consistent if $\langle\hat\rho\rangle$ continues to remain negligible compared to $\langle\hat\rho_o\rangle$ from the bounce time $\tilde\eta_{\rm B}$ to a late time $\tilde\eta_0$ of physical interest (e.g., when radiation decoupled from matter). In particular, our argument shows that the full stress-energy tensor is not needed; this significantly simplifies the task of performing numerical simulations that are needed to check self-consistency.

As noted in Sec. I, this criterion does not imply that truncated solutions are necessarily close to exact solutions because the sum of all higher order effects need not be negligible. However, in practice such criteria are generally regarded as sufficient for truncations to be trustworthy. Indeed, this philosophy governs the entire theory of cosmological, stellar and black hole perturbations in general relativity as well as perturbative calculations in quantum field theory. In the same spirit, our self consistency criterion can be used to test viability of the first order truncation in the studies of the very early universe that include the Planck regime.

In the next paper [2] we will use this criterion in the context of inflation. We first extend the general framework of this paper slightly to incorporate the $(1/2)m^2\phi^2$ potential and $\mathbb{R}^3$ spatial topology. (The value of m is fixed by using the 7 year WMAP data [102,103].) We motivate the initial conditions—called "quantum homogeneity" at the bounce—and carry out detailed numerical simulations using the "obvious" 4th order adiabatic vacuum for the initial quantum state ψ of perturbations. They show that the backreaction of perturbations remains negligible over the 11 orders of magnitude in matter density and curvature, from the big bounce until the onset of slow roll. Furthermore, the power spectrum at the end of inflation turns out to be very close to that obtained in the standard inflationary scenario and is thus compatible with the WMAP observations. By varying initial conditions for the background (within computational feasibility) we show that these results are robust. Furthermore, we show that self-consistency is preserved if the initial state is chosen to be in a neighborhood of the obvious 4th adiabatic order vacuum. Taken together these results establish existence of self-consistent extensions of the inflationary scenario to the Planck regime. Finally, there is a small range for the value $\phi_{\rm B}$ of the background inflaton field ϕ at the bounce for which the quantum state at the onset of inflation differs sufficiently from the Bunch Davies vacuum assumed in standard inflation to give rise to non-Gaussianities that could be measured in future observations along the lines of Refs. [104–107].

We will conclude by pointing out a direction for significant improvements and extensions of this framework. We began with a truncation of general relativity coupled with matter, that is well-suited for cosmology of the early universe. In the passage to the quantum theory, for the homogeneous sector we used LQC framework, rooted in the quantum geometry underlying LQG. This was crucial for the resolution of the big bang singularity and the subsequent quantum dynamics in the Planck era. On the other hand, to facilitate comparisons with the standard cosmological literature, we used a Fock-type representation for perturbations $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$. As in the Gowdy models [25–29], this is a well-defined and internally consistent quantization. But it would be more satisfactory to use a "polymer-type" representation rooted in LQG also for perturbations not only for aesthetic reasons but also because it would provide sharper guidelines to relate the truncated quantum theory to full LQG. Therefore, let us first ask: Would this change of representation make a qualitative difference in the results? The following heuristics lead us to believe that the answer is in the negative. Note that $\hat{\mathcal{Q}}$, $\hat{\mathcal{T}}$ represent *perturbations* and, in any self-consistent solution $\Psi_o \otimes \psi$, *the energy density in the perturbations is negligible compared to that in the background.*[15] On general grounds one would expect that, in a viable LQG representation of states capturing this physics, their dynamics and properties would be well approximated by our $\psi \in \mathcal{H}_1$. As a concrete illustration, we can use the following simplistic strategy to aid intuition. Use the background structure available in the truncated sector to decompose the perturbations $\mathcal{Q}(\vec{x})$, $\mathcal{T}(\vec{x})$ into Fourier modes thereby representing these fields as an assembly of harmonic oscillators, and then imagine using the polymer representation of these oscillators to construct the LQG Hilbert space $\mathcal{H}_1^{\rm LQG}$ for these fields. Then one knows [108,109] that for low energy states the results of the polymer and the standard (Fock-type) quantization are in excellent agreement. This suggests the "hybrid" approach is viable from practical or phenomenological perspective. Moreover, it is well-suited to bridge quantum field theory on quantum geometries with the well-established quantum field theory in curved space-times.

However, from a fundamental perspective it is highly desirable to systematically extend this framework by replacing $\mathcal{H}_1$ with an appropriate Hilbert space that descends from LQG. In particular, such an extension will enable one to arrive at the regularization and renormalization procedure "starting from above" i.e., from full LQG considerations. By contrast, in this paper, we have introduced this procedure "starting from below," i.e., from quantum field theory in curved space-times. Put differently, our primary goal of this paper is to carve out a path to extend the cosmological perturbation theory to the Planck era. The emphasis has been on showing that there does exist such a framework with a number of desirable

---

[15]This is a major conceptual difference from, say, the Gowdy model where gravitational waves are not perturbations around any "background" geometry. Indeed, in the vacuum Gowdy models, the entire energy density resides in the gravitational waves.

mathematical properties and physical features which, moreover, is well-suited for phenomenological applications. But from a fundamental LQG perspective, it can and should be related to LQG even more closely.

ACKNOWLEDGMENTS

APPENDIX: TRUNCATED DYNAMICS: A SIMPLE EXAMPLE

In this Appendix we will illustrate the truncation procedure of Sec. III using $\lambda\Phi^4$-theory. This example is simple enough to perform explicit calculations that bring out the main conceptual subtleties—in particular the differences between dynamics of the exact and truncated theories—which are sometimes overlooked in the cosmology literature.

1. Space-time framework

As in the main text we assume that the space-time M is topologically $\mathbb{M} \times \mathbb{R}$, where the Cauchy surfaces $\mathbb{M}$ are topologically 3-toruses $\mathbb{T}^3$. But the space-time metric η_{ab} is now assumed to be flat (with signature $-, +, +, +$). Denote by $\mathcal{S}$ the space of suitably regular solutions to

$$\Box\Phi - \mu^2\Phi - \lambda\Phi^3 = 0 \qquad (A1)$$

and by $\mathcal{S}_o$ its subspace consisting of spatially homogeneous solutions. We are interested in a small neighborhood of $\mathcal{S}_o$ in $\mathcal{S}$. In this neighborhood, it is convenient to consider curves

$$\Phi[\epsilon](\vec{x}, t) = \phi(t) + \epsilon\varphi^{(1)}(\vec{x}, t) + \frac{\epsilon^2}{2!}\varphi^{(2)}(\vec{x}, t) + \cdots$$
$$+ \frac{\epsilon^n}{n!}\varphi^{(n)}(\vec{x}, t) + \cdots \qquad (A2)$$

parametrized by $\epsilon \in]-1, 1[$, which pass through $\mathcal{S}_o$ at $\epsilon = 0$. The $\varphi^{(n)}(\vec{x}, t)$ are to be thought of as the nth order, inhomogeneous perturbations on the homogeneous solution $\phi(t)$. Since we are interested in curves that move away from $\mathcal{S}_o$, to avoid redundancy, without any loss of generality we will assume that the first order perturbation $\varphi^{(1)}(\vec{x}, t)$ are *purely* inhomogeneous, i.e., that

$$\int d\mathring{v}\,\varphi^{(1)}(\vec{x}, t) = 0 \quad \forall\, t. \qquad (A3)$$

Here in what follows all the integrals are over $\mathbb{M}$ and $d\mathring{v}$ is the natural volume element thereon. Substituting the expansion (A2) in (A1) and matching coefficients of ϵ^n for each n we obtain a hierarchy of equations:

$$\ddot{\phi} + \mu^2\phi + \lambda\phi^3 = 0,$$
$$(\Box - \mu^2 - 3\lambda\phi^2)\varphi^{(1)} = 0,$$
$$(\Box - \mu^2 - 3\lambda\phi^2)\varphi^{(2)} = 6\lambda\phi(\varphi^{(1)})^2, \qquad (\Box - \mu^2 - 3\lambda\phi^2)\varphi^{(3)} = 6\lambda((\varphi^{(1)})^3 + 3\phi\varphi^{(1)}\varphi^{(2)}), \qquad (A4)$$
$$\ldots, \qquad \cdots$$

Note that the first equation on ϕ is nonlinear but an *ordinary differential equation* (ODE) and each of the subsequent equation on $\varphi^{(n)}$ is a *linear* partial differential equation (PDE) in $\varphi^{(n)}$ with sources containing lower order fields, already determined by solving the previous equations in the hierarchy. Thus, the task of solving the nonlinear PDE (A1) is reduced to solving one nonlinear ODE and a succession of linear PDEs.[16] The idea is that an approximate solution to the full problem can be obtained by truncating the series to the appropriate order, i.e., by ignoring terms of $O(\epsilon^{n+1})$, say. As usual, while the meaning of the

approximation in terms of the smallness parameter ϵ (which could be tied to the coupling constant λ for physical reasons) is clear, the truncated series can be a good approximation to the full solution Φ only if $\varphi^{(n)}$ remain small compared to ϕ. *It is important to note that this scheme of obtaining approximate solutions is distinct from an alternative procedure that appears to be used often in the cosmology literature* (although sometimes only implicitly). That strategy corresponds to defining a field $\delta\Phi$ via $\Phi(\vec{x}, t) = \phi(t) + \delta\Phi$ and solving the equation

$$\Box\delta\Phi - \mu^2\delta\Phi - (3\lambda\phi^2)\delta\Phi - (3\lambda\phi)\delta\Phi^2 - \lambda\delta\Phi^3 = 0. \qquad (A5)$$

To the linear order this strategy agrees with (A4), but not to higher. Indeed, already at the second order, one now has to solve a *nonlinear* PDE, with a function ϕ as a coefficient, which is in some ways more complicated than solving the original (A1).

---

[16]Note that the equation on $\varphi^{(1)}$ has no source terms. Therefore, if the initial data is purely inhomogeneous, so is the solution. Because equations of $\varphi^{(n)}$ for $n > 1$ have source terms which are nonlinear in the lower order fields, we cannot demand that they be purely inhomogeneous. Note also that, for $n > 1$ there is freedom in adding a solution to the homogeneous equation. This can be removed, e.g., by choosing retarded solutions to the inhomogeneous equations.

2. The Hamiltonian framework

For the $\lambda\Phi^4$ system, one could pass to the quantum theory directly from the classical space-time formulation sketched above. However, general relativity is a background independent theory and the generalized Dirac quantization strategy followed in LQC requires us to pass through the Hamiltonian framework. Therefore, we will now illustrate how the truncation procedure of Sec. A 1 works in the phase space language.

The full phase space Γ is spanned by pairs $(\Phi(\vec{x}), \Pi(\vec{x}))$ on the 3-manifold $\mathbb{M}$ which is topologically $\mathbb{T}^3$ and its homogeneous subspace Γ_o is spanned by real numbers $(\phi, p_{(\phi)})$. The symplectic structure is given by

$$\Omega(\delta_1, \delta_2) = \int d\mathring{v}[\delta_1\Phi\delta_2\Pi - \delta_2\Phi\delta_1\Pi], \qquad \text{(A6)}$$

where $\delta \equiv (\delta\Phi, \delta\Pi)$ denotes tangent vectors to Γ. The corresponding Poisson brackets are also the familiar ones: $\{\Phi(\vec{x}_1), \Pi(\vec{x}_2)\} = \delta(\vec{x}_1, \vec{x}_2)$. Dynamics is generated by the Hamiltonian H:

$$H(\Phi, \Pi) = \frac{1}{2}\int d\mathring{v}\left[\Pi^2 + D_a\Phi D^a\Phi + \mu^2\Phi^2 + \frac{\lambda}{2}\Phi^4\right].$$
$$\text{(A7)}$$

Note that although Γ is infinite dimensional just as the solution space $\mathcal{S}$, and Γ_o is two-dimensional just as $\mathcal{S}_o$, there is a key difference: Γ and Γ_o are vector spaces unlike $\mathcal{S}$ and $\mathcal{S}_o$. As we will see the ability to define constant vector fields on Γ plays an important role in defining truncated dynamics. (This structure is also available on the phase space of general relativity and used in the main text to define truncated dynamics.)

Again, we are interested only in a small neighborhood of the homogeneous subspace Γ_o of Γ. Therefore, we are led to consider 1-parameter family of curves, parametrized by $\epsilon \in]-1, 1[$:

$$\Phi[\epsilon](\vec{x}) = \phi + \epsilon\varphi^{(1)}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\varphi^{(n)}(\vec{x}) + \cdots \quad \text{and}$$

$$\Pi[\epsilon](\vec{x}) = \frac{p_{(\phi)}}{V_o} + \epsilon\pi^{(1)}(\vec{x}) + \cdots + \frac{\epsilon^n}{n!}\pi^{(n)}(\vec{x}) + \cdots,$$
$$\text{(A8)}$$

where $\varphi^{(1)}(\vec{x})$, $\pi^{(1)}(\vec{x})$ are purely inhomogeneous, and V_o is the volume of the 3-torus $\mathbb{M}$. By truncating the series at nth order we obtain the phase space that is appropriate for describing the homogeneous solutions $(\phi, p_{(\phi)})$, together with first, second, ... nth order perturbations propagating thereon. Since in the main text we focus on just the first order perturbations, let us do the same here.

Thus, let the truncated phase-space Γ_{Trun} consist of a doublet of pairs of canonically conjugate fields $(\phi, p_{(\phi)}; \varphi^{(1)}(\vec{x}), \pi^{(1)}(\vec{x}))$ where $\phi, p_{(\phi)}$ are real numbers (representing homogeneous fields on $\mathbb{M}$) and $\varphi^{(1)}(\vec{x})$,

$\pi^{(1)}(\vec{x})$ are purely inhomogeneous fields on $\mathbb{M}$. Thus, $\Gamma_{\text{Trun}} = \Gamma_o \times \Gamma_1$. The symplectic structure on Γ_{Trun} is given by $\Omega_{\text{Trun}} = \Omega_o + \Omega_1$, where

$$\Omega_o(\delta_1, \delta_2) = \delta_1\phi\delta_2 p_{(\phi)} - \delta_2\phi\delta_1 p_{(\phi)}$$

$$\Omega_1(\delta_1, \delta_2) = \int d\mathring{v}[\delta_1\varphi^{(1)}\delta_2\pi^{(1)} - \delta_2\varphi^{(1)}\delta_1\pi^{(1)}], \quad \text{(A9)}$$

which yields the Poisson brackets

$$\{\phi, p_{(\phi)}\} = 1,$$

$$\{\varphi^{(1)}(\vec{x}_1), \pi^{(1)}(\vec{x}_2)\} = \left[\delta(\vec{x}_1, \vec{x}_2) - \frac{1}{V_o}\right] \equiv \bar{\delta}(\vec{x}_1, \vec{x}_2), \quad \text{(A10)}$$

where $\bar{\delta}(\vec{x}_1, \vec{x}_2)$ is the Dirac delta-distribution restricted to the purely inhomogeneous fields on $\mathbb{M}$.[17] By decomposing fields $\Phi(\vec{x})$, $\Pi(\vec{x})$ in Γ into purely homogeneous and purely inhomogeneous parts, one can readily see that the full pase space (Γ, Ω) is naturally isomorphic to the truncated phase space $(\Gamma_{\text{Trun}}, \Omega_{\text{Trun}})$. However, the physical meaning of the inhomogeneous fields is different in the two cases and, more importantly, the dynamics is very different.

Geometrically, Γ_{Trun} is the normal bundle over Γ_o (since purely inhomogeneous fields are orthogonal to the purely homogeneous ones in the L^2 norm of the space of functions on $\mathbb{M}$.) To obtain the dynamical flow on Γ_{Trun}, let us begin with the full, nonlinear Hamiltonian vector field X_H on a small neighborhood of Γ_o in Γ. Since the exact Hamiltonian flow generated by X_H is tangential to Γ_o, the equations of motion on Γ_o are just the restrictions of those on full Γ to the homogeneous sector: $\dot{\phi} = p_{(\phi)}/V_o$, $\dot{p}_{(\phi)} = -(\mu^2\phi + \lambda\phi^3)V_o$. To obtain the equations of motion on the (inhomogeneous) tangent vectors $(\varphi^{(1)}, \pi^{(1)})$, we first note that the Hamiltonian flow X_H on Γ naturally drags these tangent vectors along dynamical trajectories on Γ_o. To obtain the "dot," however, we need to compare the image $(\varphi^{(1)}, \pi^{(1)})|_{t+\delta t}$ at $(\phi, p_{(\phi)})|_{t+\delta t}$ with the original tangent vector $(\varphi^{(1)}, \pi^{(1)})|_t$ at the point $(\phi, p_{(\phi)})|_t$ of Γ_o. This can be trivially accomplished because Γ has a vector space structure. The resulting equations of motion are: $\dot{\varphi}^{(1)}(\vec{x}, t) = \pi^{(1)}(\vec{x}, t)$, $\dot{\pi}^{(1)}(\vec{x}, t) = [D^2 - \mu^2 - 3\lambda\phi^2(t)]\varphi^{(1)}(\vec{x}, t)$. The resulting dynamical vector field X_{Dyn} on Γ_{Trun} is given by

$$X_{\text{Dyn}} := (\dot{\phi}, \dot{p}_{(\phi)}; \dot{\varphi}^{(1)}, \dot{\pi}^{(1)})$$

$$= \left(\frac{p_{(\phi)}(t)}{V_o}, -[\mu^2\phi(t) + \lambda\phi^3(t)]V_o; \pi^{(1)},\right.$$

$$\left. [D^2 - \mu^2 - 3\lambda\phi^2(t)]\varphi^{(1)}\right). \quad \text{(A11)}$$

---

[17]The term $1/V_o$ is necessary simply because the fields $\varphi^{(1)}(\vec{x})$ and $\pi^{(1)}(\vec{x})$ are purely inhomogeneous. For example, if we integrate the left side of the Poisson bracket between $\varphi^{(1)}$ and $\pi^{(1)}$ over x_1 (or x_2), we get zero and $1/V_o$ terms assures that the right side also vanishes.

Thus, the full dynamics on Γ induces a well-defined flow X_{Dyn} on Γ_{Trun}. Furthermore, this dynamical vector field can be expressed in a form adapted to symplectic geometry. Using Greek letters to denote the abstract indices labeling tangent vectors to Γ, X_{Dyn}^{α} can be expressed as

$$X_{\text{Dyn}}^{\alpha} = \Omega_o^{\alpha\beta}\partial_{\beta}H_o + \Omega_1^{\alpha\beta}\partial_{\beta}H_1, \qquad \text{(A12)}$$

where

$$H_o(\phi, p_{(\phi)}) := H|_{\Gamma_o} = \frac{V_o}{2}\left(\frac{p_{(\phi)}^2}{V_o^2} + \mu^2\phi^2 + \frac{\lambda}{2}\right)\phi^4\right),$$

$$\begin{aligned}
H_1(\varphi^{(1)}, \pi^{(1)}) &= \frac{1}{2}\int d\mathring{v}[(\pi^{(1)})^2 + D^a\varphi^{(1)}D_a\varphi^{(1)} \\
&\quad + \mu^2(\varphi^{(1)})^2 + 3\lambda(\phi^2)(\varphi^{(1)})^2].
\end{aligned} \qquad \text{(A13)}$$

Note however that this is *not* a Hamiltonian flow on Γ_{Trun} because it is not of the form $(\Omega_o^{\alpha\beta} + \Omega_1^{\alpha\beta})\partial_{\beta}H_{\text{Trun}}$ for any function H_{Trun} on Γ_{Trun}. The obvious candidate, $H_{\text{Trun}} := H_o + H_1$ does not work because $\Omega_o^{\alpha\beta}\partial_{\beta}H_1 \neq 0$ (since H_1 depends not only on $(\varphi^{(1)}, \pi^{(1)})$ but also on ϕ).[18] More generally, one can verify that the fact that H_1 depends on ϕ implies that the Lie derivative of $\Omega_{\text{Trun}} := \Omega_o + \Omega_1$ by the dynamical vector field X_{Dyn} on Γ_{Trun} does not vanish.

In the classical theory, the fact that we have a well-defined dynamical flow on Γ_{Trun} suffices. However, the fact that the flow is not Hamiltonian introduces new features in the transition to quantum theory. Keeping the quantum perspective in mind, one can rephrase the classical dynamics as follows. Since the homogeneous solution

is to be regarded as the background and inhomogeneities as perturbations, we can first restrict ourselves to the homogeneous part of the phase space (Γ_o, Ω_o) and note that, on it, the dynamics is indeed governed by a true Hamiltonian flow, generated by H_o. Fix any dynamical trajectory $\phi(t), p_{(\phi)}(t)$ in Γ_o. To specify how perturbations propagate on this background solution, we need to lift this trajectory to the normal bundle, Γ_{Trun}. This is precisely what the remaining part, $\Omega_1^{\alpha\beta}\partial_{\beta}H_1$, of the dynamical vector field X_{Dyn}^{α} does. Given any tangent vector $(\varphi_o^{(1)}, \pi_o^{(1)})$ at a point, say $(\phi(t_o), p_{(\phi)}(t_o))$, along the given dynamical trajectory, the orbit of X_{Dyn}^{α} through the point $(\phi(t_o), p_{(\phi)}(t_o); \varphi_o^{(1)}(t_o), \pi_o^{(1)}(t_o))$ of Γ_{Trun} specifies the dynamics of the perturbation $(\varphi^{(1)}, \pi^{(1)})$ on the background trajectory $(\phi(t), p_{(\phi)}(t))$. The fact that we have lifts to Γ_{Trun} of orbits in Γ_o trivially implies that, even though different choices of the initial $(\varphi^{(1)}, \pi^{(1)})$ define distinct orbits (describing dynamics of various perturbations), all these orbits in Γ_{Trun} project down to the same orbit on Γ_o. This is just a reflection of the fact that the dynamics defined by X_{Dyn}^{α} neglects the backreaction of the perturbation on the homogeneous background.

Remark: To obtain the truncated dynamics X_{Trun} we used the vector space structure of Γ to transport the vector $(\varphi^{(1)}, \pi^{(1)})|_{t+\delta t}$ at $(\phi, p_{(\phi)})|_{t+\delta t}$ to the point $(\phi, p_{(\phi)})|_t$. However, one can carry out this comparison more generally, e.g., if there is a natural flat connection to transport vectors from one point of the phase space to another. This is the case when Γ is a cotangent bundle over an affine space (as in LQG) or over a convex subset of a vector space (as in the ADM framework of general relativity). Therefore, the procedure of inducing dynamics on the truncated phase space from the exact dynamics in a neighborhood of the homogeneous sector of the phase space goes through also in the cosmological context analyzed in the main text.

---

[18]Had we simply defined dynamics using $H_{\text{Trun}} := H_o + H_1$, it would have included the homogeneous part of the backreaction of the first order perturbation $(\varphi^{(1)}, \pi^{(1)})$. This would not be physically consistent because this dynamics ignores the inhomogeneous part of the backreaction of the same perturbation which is of the same order in the ϵ expansion.

[1] I. Agullo, A. Ashtekar, and W. Nelson, Phys. Rev. Lett. **109,** 251301 (2012).
[2] I. Agullo, A. Ashtekar, and W. Nelson, "The Pre-inflationary Dynamics of Loop Quantum Cosmology: Confronting Quantum Gravity with Observations" (to be published).
[3] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure* (Cambridge University Press, Cambridge, England, 2000).
[4] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).
[5] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Cambridge, England, 2005).
[6] S. Weinberg, *Cosmology* (Oxford University press, Oxford, 2008).
[7] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space-Times* (Cambridge University Press, Cambridge, England, 1982).
[8] S. Fulling, *Aspects of Quantum Field Theory in Curved Space-Times* (Cambridge University Press, Cambridge, England, 1989).
[9] L. Parker and D. Toms, *Quantum Field Theory in Curved Space-Time* (Cambridge University Press, Cambridge, England, 2009).
[10] M. Bojowald, Phys. Rev. Lett. **86,** 5227 (2001).
[11] A. Ashtekar, T. Pawlowski, and P. Singh, Phys. Rev. Lett. **96,** 141301 (2006).

[12] A. Ashtekar, T. Pawlowski, and P. Singh, Phys. Rev. D **73,** 124038 (2006).

[13] A. Ashtekar, T. Pawlowski, and P. Singh, Phys. Rev. D **74,** 084003 (2006).

[14] A. Ashtekar, A. Corichi, and P. Singh, Phys. Rev. D **77,** 024046 (2008).

[15] P. Singh, Classical Quantum Gravity **26,** 125005 (2009).

[16] A. Ashtekar, T. Pawlowski, P. Singh, and K. Vandersloot, Phys. Rev. D **75,** 024035 (2007).

[17] L. Szulc, W. Kaminski, and J. Lewandowski, Classical Quantum Gravity **24,** 2621 (2007).

[18] E. Bentivegna and T. Pawlowski, Phys. Rev. D **77,** 124025 (2008).

[19] W. Kaminski and T. Pawlowski, Phys. Rev. D **81,** 024014 (2010).

[20] A. Ashtekar and T. Pawlowski, Phys. Rev. D **85,** 064001 (2012).

[21] A. Ashtekar and E. Wilson-Ewing, Phys. Rev. D **79,** 083535 (2009).

[22] M. Martín-Benito, G. A. Mena Marugan, and T. Pawlowski, Phys. Rev. D **78,** 064008 (2008); **80,** 084038 (2009).

[23] A. Ashtekar and E. Wilson-Ewing, Phys. Rev. D **80,** 123532 (2009).

[24] E. Wilson-Ewing, Phys. Rev. D **82,** 043508 (2010).

[25] M. Martín-Benito, L. J. Garay, and G. A. Mena Marugan, Phys. Rev. D **78,** 083516 (2008); L. J. Garay, M. Martín-Benito, and G. A. Mena Marugan, Phys. Rev. D **82,** 044048 (2010).

[26] D. Brizuela, G. A. Mena Marugan, and T. Pawlowski, Classical Quantum Gravity **27,** 052001 (2010).

[27] M. Martín-Benito, G. A. Mena Marugan, and E. Wilson-Ewing, Phys. Rev. D **82,** 084012 (2010).

[28] M. Martín-Benito, D. Martín-de Blas, and G. A. Mena Marugan, Phys. Rev. D **83,** 084050 (2011).

[29] D. Brizuela, G. A. Mena Marugan, and T. Pawlowski, Phys. Rev. D **84,** 124017 (2011).

[30] M. Bojowald, Living Rev. Relativity **11,** 4 (2008).

[31] A. Ashtekar and P. Singh, Classical Quantum Gravity **28,** 213001 (2011).

[32] A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov, Phys. Rev. Lett. **80,** 904 (1998).

[33] A. Ashtekar, J. Baez, and K. Krasnov, Adv. Theor. Math. Phys. **4,** 1 (2000).

[34] J. F. Barbero, G. J. Lewandowski, and E. J. S. Villasenor, Proc. Sci., QGQS2011 (2011) 023.

[35] E. Bianchi, E. Magliaro, and C. Perini, Nucl. Phys. **B822,** 245 (2009).

[36] E. Bianchi and Y. Ding, Phys. Rev. D **86,** 104040 (2012).

[37] C. Rovelli, arXiv:1102.3660.

[38] A. Ashtekar, W. Kaminski, and J. Lewandowski, Phys. Rev. D **79,** 064030 (2009).

[39] P. Bielewicz and A. J. Banday, Mon. Not. R. Astron. Soc. **412,** 2104 (2011).

[40] V. Taveras, Phys. Rev. D **78,** 064072 (2008).

[41] A. Ashtekar, M. Campiglia, and A. Henderson, Phys. Rev. D **82,** 124043 (2010).

[42] S. Tsujikawa, P. Singh, and R. Maartens, Classical Quantum Gravity **21,** 5767 (2004).

[43] M. Artymowski, Z. Lalak, and L. Szulc, J. Cosmol. Astropart. Phys. 01 (2009) 004.

[44] W. Nelson and M. Sakellariadou, Phys. Rev. D **76,** 044015 (2007).

[45] M. Bojowald, G. Calcagni, and S. Tsujikawa, J. Cosmol. Astropart. Phys. 11 (2011) 046.

[46] G. Calcagni, arXiv:1209.0473.

[47] M. Bojowald and G. M. Hossain, Phys. Rev. D **77,** 023508 (2008).

[48] M. Bojowald, G. M. Hossain, M. Kagan, and S. Shankaranarayanan, Phys. Rev. D **78,** 063547 (2008).

[49] J. Grain and A. Barrau, Phys. Rev. Lett. **102,** 081301 (2009).

[50] J. Grain, T. Cailleteau, A. Barrau, and A. Gorecki, Phys. Rev. D **81,** 024040 (2010).

[51] J. Mielczarek, T. Cailleteau, J. Grain, and A. Barrau, Phys. Rev. D **81,** 104049 (2010).

[52] J. Grain, A. Barrau, T. Cailleteau, and J. Mielczarek, Phys. Rev. D **82,** 123520 (2010).

[53] T. Cailleteau, J. Mielczarek, A. Barrau, and J. Grain, Classical Quantum Gravity **29,** 095010 (2012).

[54] M. Bojowald and G. M. Paily, Phys. Rev. D **86,** 104018 (2012).

[55] S. A. Hojman, K. Kuchar, and C. Teitelboim, Ann. Phys. (N.Y.) **96,** 88 (1976).

[56] J. Mielczareck, Phys. Rev. D **81,** 063503 (2010).

[57] A. Ashtekar, Phys. Rev. Lett. **57,** 2244 (1986); Phys. Rev. D **36,** 1587 (1987).

[58] J. F. G. Barbero, Phys. Rev. D **51,** 5507 (1995).

[59] A. Ashtekar, M. Bojowald, and J. Lewandowski, Adv. Theor. Math. Phys. **7,** 233 (2003).

[60] A. Ashtekar and M. Campiglia, Classical Quantum Gravity **29,** 242001 (2012).

[61] J. Lewandowski, A. Okolow, H. Sahlmann, and T. Thiemann, Commun. Math. Phys. **267,** 703 (2006).

[62] C. Fleishchack, Commun. Math. Phys. **285,** 67 (2009).

[63] W. Nelson and M. Sakellariadou, Phys. Rev. D **78,** 024006 (2008).

[64] B. Gupt and P. Singh, Phys. Rev. D **85,** 044011 (2012).

[65] P. Singh, Phys. Rev. D **85,** 104011 (2012).

[66] A. Corichi and E. Montoya, Phys. Rev. D **85,** 104052 (2012).

[67] B. K. Berger, Phys. Rev. D **11,** 2770 (1975).

[68] G. A. Mena Marugan, Phys. Rev. D **56,** 908 (1997).

[69] M. Pierri, Int. J. Mod. Phys. D **11,** 135 (2002).

[70] A. Corichi, J. Cortez, and H. Quevedo, Int. J. Mod. Phys. D **11,** 1451 (2002).

[71] I. Agullo, in The Loops 11 Conference, 2011 (unpublished), http://www.iem.csic.es/loops11/.

[72] I. Agullo, A. Ashtekar, and W. Nelson, in *Proceedings of the 2011 International Conference on Gravitation and Cosmology*, (CUP, Cambridge, to be published).

[73] J. Olmedo, in The Loops 11 Conference, 2011 (unpublished), http://www.iem.csic.es/loops11/; M. Fermandez-Mendez, G. A. Mena Marugan, and J. Olmedo, Phys. Rev. D **86,** 024003 (2012).

[74] D. Langlois, Classical Quantum Gravity **11,** 389 (1994).

[75] J. M. Bardeen, Phys. Rev. D **22,** 1882 (1980).

[76] J. Puchta, M.Sc. thesis, University of Warsaw, 2009.

[77] J. Cortez, G. A. Mena Marugan, J. Olmedo, and J. M. Velhinho, Phys. Rev. D **83,** 025002 (2011).

[78] M. Fermandez-Mendez, G. A. Mena Marugan, J. Olmedo, and J. M. Velhinho, Phys. Rev. D **85**, 103525 (2012).

[79] L. Castello-Gomar, J. Cortez, D. Martín-de Blas, G. A. Mena Marugán, and J. M. Velhinho, J. Cosmol. Astropart. Phys. 11 (2012) 001.

[80] B. Dittrich and J. Tambornino, Classical Quantum Gravity **24**, 4543 (2007).

[81] M. Bojowald, H. H. Hernandez, M. Kagan, P. Singh, and A. Skirzewski, Phys. Rev. D **74**, 123512 (2006); Phys. Rev. Lett. **98**, 031301 (2007).

[82] K. Giesel, S. Hofmann, T. Thiemann, and O. Winkler, Classical Quantum Gravity **27**, 055005 (2010).

[83] K. Giesel, S. Hofmann, T. Thiemann, and O. Winkler, Classical Quantum Gravity **27**, 055006 (2010).

[84] L. Bethke and J. Magueijo, Phys. Rev. D **84**, 024014 (2011).

[85] W. Kaminski and J. Lewandowski, Classical Quantum Gravity **25**, 035001 (2008).

[86] D. Marolf, arXiv:gr-qc/9508015; Classical Quantum Gravity **12**, 1199 (1995).

[87] A. Ashtekar, J. Lewandowski, D. Marolf, J. Mourão, and T. Thiemann, J. Math. Phys. (N.Y.) **36**, 6456 (1995).

[88] A. Ashtekar, L. Bombelli, and A. Corichi, Phys. Rev. D **72**, 025008 (2005).

[89] A. Dapor, J. Lewandowski, and Y. Tavakoli, Phys. Rev. D **86**, 064013 (2012).

[90] I. M. Gel'fand and M. A. Naimark, Math. Sbornik **12**, 197 (1943); I. E. Segal, Ann. Math. **48**, 930 (1947).

[91] A. Ashtekar and A. Magnon, Proc. R. Soc. A **346**, 375 (1975); Gen. Relativ. Gravit. **12**, 205 (1980).

[92] R. M. Wald, *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics* (University of Chicago, Chicago, 1994).

[93] L. Parker and S. A. Fulling, Phys. Rev. D **9**, 341 (1974).

[94] P. R. Anderson and L. Parker, Phys. Rev. D **36**, 2963 (1987).

[95] L. Parker, Ph.D. thesis, Harvard University, 1966.

[96] L. Parker, Phys. Rev. Lett. **21**, 562 (1968); Phys. Rev. **183**, 1057 (1969).

[97] S. Hollands, personal communication.

[98] P. R. Anderson, C. Molina-Paris, and E. Mottola, Phys. Rev. D **72**, 043515 (2005).

[99] N. D. Birrell, Proc. R. Soc. A **361**, 513 (1978).

[100] R. Penrose, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979), pp. 581–638.

[101] I. Agullo and W. Nelson, ''Preferred Quantum States in Linear QFT on FLRW Backgrounds'' (to be published).

[102] E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **192**, 18 (2011).

[103] A. Ashtekar and D. Sloan, Gen. Relativ. Gravit. **43**, 3619 (2011).

[104] I. Agullo and L. Parker, Phys. Rev. D **83**, 063526 (2011); Gen. Relativ. Gravit. **43**, 2541 (2011).

[105] J. Ganc and E. Komatsu, Phys. Rev. D **86**, 023518 (2012).

[106] I. Agullo and S. Shandera, J. Cosmol. Astropart. Phys. 09 (2012) 007.

[107] F. Schmidt and L. Hui, arXiv:1210.2965.

[108] A. Ashtekar, S. Fairhurst, and J. Willis, Classical Quantum Gravity **20**, 1031 (2003).

[109] A. Corichi, T. Vukasinac, and J. A. Zapata, Classical Quantum Gravity **24**, 1495 (2007).