Quantum Nature of the Big Bang

Abhay Ashtekar,* Tomasz Pawlowski,† and Parampreet Singh‡
Institute for Gravitational Physics and Geometry, Physics Department, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
(Received 22 February 2006; published 12 April 2006)

Some long-standing issues concerning the quantum nature of the big bang are resolved in the context of homogeneous isotropic models with a scalar field. Specifically, the known results on the resolution of the big-bang singularity in loop quantum cosmology are significantly extended as follows: (i) the scalar field is shown to serve as an internal clock, thereby providing a detailed realization of the "emergent time" idea; (ii) the physical Hilbert space, Dirac observables, and semiclassical states are constructed rigorously; (iii) the Hamiltonian constraint is solved numerically to show that the big bang is replaced by a big bounce. Thanks to the nonperturbative, background independent methods, unlike in other approaches the quantum evolution is deterministic across the deep Planck regime.

DOI: 10.1103/PhysRevLett.96.141301 PACS numbers: 98.80.Qc, 04.60.Kz, 04.60.Pp

Thanks to the influx of observational data, recent years have witnessed enormous advances in our understanding of the early Universe. To interpret the present data, it is sufficient to work in a regime in which space-time can be taken to be a smooth continuum as in general relativity, setting aside fundamental questions involving the deep Planck regime. However, for a complete conceptual understanding as well as interpretation of the future, more refined data, these long-standing issues will have to be faced squarely. Examples are: (i) how close to the big bang does the smooth space-time of general relativity make sense? In particular, can one show from first principles that this approximation is valid at the onset of inflation? (ii) Is the big-bang singularity naturally resolved by quantum gravity? Or, is some external input such as a new principle or a boundary condition at the big-bang essential? (iii) Is the quantum evolution across the "singularity" deterministic? In the pre-big-bang and Ekpyrotic scenarios, for example, the answer has been in the negative [1]. (iv) If the singularity is resolved, what is on the "other side"? Is there just a *quantum foam* far removed from any classical space-time, or, is there another large, classical universe? The purpose of this Letter is to summarize results from recent analytical and numerical investigations within loop quantum cosmology which address these and related issues.

Loop quantum gravity (LQG) is a background independent, nonperturbative approach to quantum gravity [2]. Loop quantum cosmology (LQC) focuses on symmetry reduced models but carries out quantization by mimicking the constructions used in the full theory [3]. Results to date in this area fall in two broad categories: (a) resolution of the big-bang singularity using modifications of the *gravitational* Hamiltonian due to quantum geometry [4] and (b) phenomenological predictions from effective equations that incorporate the modifications of the *matter* Hamiltonians due to quantum geometry [see, e.g., [5,6]]. As in the first category, we focus on the more fundamental issues. While previous results showed that the LQC evolution does not break down at the singularity, as pointed out, e.g., in [7], they did not shed light on what happened before. By constructing the missing conceptual and mathematical infrastructure, we show that the Universe has a classical pre-big-bang branch, joined *deterministically* to the post-big-bang branch by the LQC evolution. Our detailed analysis of the Planck regime also provides tools to test the validity of assumptions underlying phenomenological predictions.

We will illustrate effects of quantum geometry on both the gravitational and matter Hamiltonians through a simple example: the spatially homogeneous, isotropic $k = 0$ cosmologies with a massless scalar field. Although the approach admits generalizations, we focus on these models because a singularity is unavoidable in their classical theory. The question is if it is naturally tamed by quantum effects. The answer in the "geometrodynamical" framework used in older cosmologies turns out to be in the negative [8]. For example, if one begins with a semiclassical state representing a classical universe at late times and evolves it back via the Wheeler-DeWitt equation, one finds that it just follows the classical trajectory into the big-bang singularity [9]. In LQC, the situation is very different. This may seem surprising at first. For, the system has only a finite number of degrees of freedom and von Neumann's theorem assures us that, under appropriate assumptions, the resulting quantum mechanics is unique. However, for reasons we will now explain, LQC does turn out to be qualitatively different from the Wheeler-DeWitt theory [10].

Because of spatial homogeneity and isotropy, one can fix a fiducial (flat) triad ${}^o e^a_i$ and its dual cotriad ${}^o \omega^i_a$. The SU(2) gravitational spin connection A^i_a used in LQG has only one component c which furthermore depends only on time; $A^i_a = c\, {}^o\omega^i_a$. Similarly, the momentum E^a_i canonically conjugate to A^i_a—representing a (density weighted) triad—has a single component p; $E^a_i = p(\det {}^o\omega)\, {}^o e^a_i$. p is related to the scale factor a via $a^2 = |p|$. However, p is not restricted to be positive; under $p \to -p$ the metric remains unchanged but the spatial triad flips the orientation. The pair (c, p) is canonically conjugate: $\{c, p\} = (8\pi G\gamma/3)$, where γ is the Barbero-Immirzi parameter.

Quantization is carried out by closely mimicking the procedure used in *full* LQG [10]. There, the elementary variables which have unambiguous operator analogs in quantum theory are the *holonomies* h of connections A_a^i and the (smeared) triads E_i^a. Now, background independence leads to a surprisingly strong result [11]: in essence, the basic operator algebra generated by holonomies and triads admits a *unique* irreducible, diffeomorphism covariant representation. In this representation, there are operators $\hat{h}$ representing holonomies and $\hat{E}$ representing (smeared) triads. But there are no operators representing connections A_a^i themselves. In the cosmological model now under consideration, holonomies along a straight line of (oriented) length μ (with respect to the fiducial triad ${}^o e_i^a$) are almost periodic functions of c of the form $N_\mu(c) := \exp(i\mu c/2)$. (The N_μ are the analogs of spin-network functions in the full theory.) In the quantum theory, then, we are led to a representation in which operators $\hat{N}_\mu$ and $\hat{p}$ are well defined, but there is *no* operator corresponding to the connection c itself (because the 1-parameter group $\hat{N}_\mu$ is not weakly continuous in μ). This new quantum mechanics is inequivalent to the Wheeler-DeWitt theory already at a kinematical level. In particular, the gravitational part of the Hilbert space is now $L^2(\bar{R}_{\rm Bohr}, d\mu_{\rm Bohr})$, the space of square integrable functions on the Bohr compactification of the real line, rather than the standard $L^2(R, d\mu)$ [10]. While in the semiclassical regime LQC is well approximated by the Wheeler-DeWitt theory, important differences manifest themselves at the Planck scale. These are the hallmarks of quantum geometry [2,3].

The new representation also leads to a qualitative difference in the structure of the Hamiltonian constraint operator: the gravitational part of the constraint is a *difference* operator rather than a differential operator as in the Wheeler-DeWitt theory. The derivation [9,10] can be summarized briefly as follows. In the classical theory, the gravitational part of the constraint is given by $\int d^3x \epsilon^{ijk} e^{-1} E_i^a E_j^b F_{abk}$ where $e = |\det E|^{1/2}$ and F_{ab}^k is the curvature of the connection A_a^i. The part of this operator involving triads can be quantized [10] using a standard procedure introduced by Thiemann in the full theory. However, since there is no operator corresponding to the connection itself, one has to express F_{ab}^k as a limit of the holonomy around a loop divided by the area enclosed by the loop, as the area shrinks to zero. Now, quantum geometry tells us that the area operator has a minimum nonzero value, Δ, and in the quantum theory it is natural to shrink the loop only till it attains this minimum. There are two ways to implement this idea in detail. Here, we will use the one which has already appeared in the literature [2,3,10] although it has certain drawbacks in the semiclassical regime, especially in more general models. The second method will be discussed in the second of the detailed papers [9], which will also show that the quantum bounce and deterministic evolution across the Planck regime persist if the second and more satisfactory method is used. In both cases, it is the existence of the "area gap" Δ that leads one to a difference equation.

Let us represent states as functions $\Psi(\mu, \phi)$, where ϕ is the scalar field and (modulo a fixed multiple of $\ell_{\rm Pl}^2$) the dimensionless real number μ is the eigenvalue of $\hat{p}$ [10]. Then in LQC the (self-adjoint) Hamiltonian constraint is given by [9]

$$\partial_\phi^2 \Psi = [B(\mu)]^{-1}[C^+(\mu)\Psi(\mu + 4\mu_o, \phi) + C^o(\mu)\Psi(\mu, \phi) + C^-(\mu)\Psi(\mu - 4\mu_o, \phi)] =: -\Theta\Psi(\mu, \phi), \qquad (1)$$

where $C^+(\mu) = (\pi G/9\mu_o^3)||\mu + 3\mu_o|^{3/2} - |\mu + \mu_o|^{3/2}|$; $C^-(\mu) = C^+(\mu - 4\mu_o)$; $C^o(\mu) = -C^+(\mu) - C^-(\mu)$ and where $(6/8\pi\gamma\ell_{\rm Pl}^2)^{3/2} B(\mu)$ are the eigenvalues of the operator $|\hat{p}|^{-3/2}$ [3]. The fixed real number μ_o is determined by the area gap; $(8\pi\gamma/6)\mu_o \ell_{\rm Pl}^2 = \Delta$.

Now, in each classical solution, ϕ is a globally monotonic function of time and can therefore be taken as the dynamical variable representing an *internal* clock. In quantum theory, even on shell, there is no space-time metric. But since the quantum constraint (1) dictates how $\Psi(\mu, \phi)$ "evolves" as ϕ changes, it is convenient to regard the argument ϕ in $\Psi(\mu, \phi)$ as "emergent time" and μ as the physical degree of freedom. A complete set of Dirac observables is provided by the constant of motion $\hat{p}_\phi$ and operators $\hat{\mu}|_{\phi_o}$ determining the value of μ at the "instant" $\phi = \phi_o$.

Physical states are the (suitably regular) solutions to Eq. (1). The map $\hat{\Pi}$ defined by $\hat{\Pi}\Psi(\mu, \phi) = \Psi(-\mu, \phi)$ corresponds just to the flip of orientation of the spatial triad (under which geometry remains unchanged); $\hat{\Pi}$ is thus a large gauge transformation on the space of solutions to Eq. (1). One is therefore led to divide physical states into sectors, each providing an irreducible, unitary representation of this symmetry. As one would expect, physical considerations imply that we should consider the symmetric sector, with eigenvalue $+1$ of $\hat{\Pi}$ [9].

To endow this space with the structure of a Hilbert space, we use the "group averaging method" [12]. The technical implementation of this procedure is greatly simplified by the fact that the difference operator Θ on the right side of (1) is independent of ϕ and can be shown to be self-adjoint and positive definite [on the Hilbert space $L^2(\bar{R}_{\rm Bohr}, B(\mu)d\mu_{\rm Bohr})$]. Since Θ is a difference operator, the resulting physical Hilbert space $\mathcal{H}$ has sectors $\mathcal{H}_\epsilon$ which are superselected; $\mathcal{H} = \oplus_\epsilon \mathcal{H}_\epsilon$ with $\epsilon \in [0, 2\mu_o]$. States $\Psi(\mu, \phi)$ in $\mathcal{H}_\epsilon$ (are symmetric under the orientation inversion $\hat{\Pi}$ and) have support on points $\mu = \pm\epsilon + 4n\mu_o$. Let us consider a generic H_ϵ. [The small technical differences in the exceptional cases are discussed in [9]; they do not affect the main conclusions.] Wave functions $\Psi(\mu, \phi)$ solve (1) and are of positive frequency with respect to the "internal time" ϕ. Equivalently, they satisfy the "positive frequency" square root of Eq. (1):

$$-i\partial_\phi \Psi = \sqrt{\Theta}\Psi \tag{2}$$

and the inner product is given by:

$$\langle\Psi_1|\Psi_2\rangle_{\rm phy} = \sum_{\mu\in\{\pm\epsilon+4\mu_o\mathbb{Z}\}} B(\mu)\bar{\Psi}_1(\mu, \phi)\Psi_2(\mu, \phi), \tag{3}$$

where, as usual, $\mathbb{Z}$ denotes the set of integers. On these states, the Dirac observables act in the expected fashion:

$$\hat{p}_\phi \Psi = -i\hbar\partial_\phi \Psi \tag{4a}$$

$$\hat{\mu}|_{\phi_o}\Psi(\mu, \phi) = e^{i\sqrt{\Theta}(\phi-\phi_o)}\mu\Psi(\mu, \phi_o). \tag{4b}$$

One can also begin with the complete set of Dirac observables (4) and show that (3) is the unique inner product which makes them self-adjoint.

To construct semiclassical states and for numerical simulations, it is convenient to express physical states as linear combinations of the eigenstates of $\hat{p}_\phi$ and Θ. We first note that, for $\mu \gg \mu_o$, there is a precise sense [9,10] in which the difference operator Θ approaches the Wheeler-DeWitt differential operator $\underline{\Theta}$, given by

$$(\underline{\Theta} f)(\mu) = (16\pi G/3)\mu^{3/2}(\sqrt{\mu}f')'. \tag{5}$$

[Thus, if one ignores the quantum geometry effects, Eq. (1) reduces to the Wheeler-DeWitt equation $\partial_\phi^2\Psi = -\underline{\Theta}\Psi$.] The eigenfunctions

$$\underline{e}_k(\mu) = (1/4\pi) \times |\mu|^{1/4}e^{ik\ln|\mu|} \tag{6}$$

of $\underline{\Theta}$ are labeled by a real number k and its eigenvalues are given by $\underline{\omega}^2 = (\pi G/3)(16k^2 + 1)$. The complete set of eigenfunctions $e_k(\mu)$ of the discrete operator Θ is also labeled by a real number k and $e_k(\mu)$ are well approximated by $\underline{e}_k(\mu)$ for $\mu \gg \mu_o$ [9]. The eigenvalues $\omega^2(k)$ of Θ increase monotonically with $|k|$. Finally, the $e_k(\mu)$ satisfy the standard orthonormality relations $\langle e_k|e'_k\rangle = \delta(k, k')$. A physical state $\Psi(\mu, \phi)$ can therefore be expanded as:

$$\Psi(\mu, \phi) = \int_{-\infty}^{\infty} dk\tilde{\Psi}(k)e_k^{(s)}(\mu)e^{i\omega(k)\phi}, \tag{7}$$

where $\tilde{\Psi}(k)$ is arbitrary (but suitably regular), $\omega(k)$ is positive, and $e_k^{(s)}(\mu) = (1/\sqrt{2})[e_k(\mu) + e_k(-\mu)]$. Thus, each physical state is characterized by a free function $\tilde{\Psi}(k)$. [For proofs and subtleties, see [9].]

Since we have the explicit Hilbert space and a complete set of Dirac observables, we can now construct states which are semiclassical at late times—e.g., now—and evolve them numerically "backward in time." There are three natural constructions to implement this idea in detail, reflecting the freedom in the notion of semiclassical states. The main results in all three cases are the same [9]. Here we will report on the results obtained using the strategy that brings out the contrast with the Wheeler-DeWitt theory most sharply.

As noted before, p_ϕ is a constant of motion. For the semiclassical analysis, we are led to choose a large value p_ϕ^* ($\gg \hbar$ in the classical $c = G = 1$ units. In the closed model, for example, this condition is necessary to ensure that the Universe can expand out to a macroscopic size.) Fix a point (μ^*, ϕ_o) on the classical trajectory with $p_\phi = p_\phi^*$ which starts out at the big bang and then expands, choosing $\mu^* \gg 1$. We want to construct a state which is peaked at (μ^*, p_ϕ^*) at the initial "time" $\phi = \phi_o$ and follow its "evolution" backward. Let $\tilde{\Psi}(k)$ be a Gaussian, peaked at a value k^* given by $p_\phi^* = -(\sqrt{16\pi G\hbar^2/3})k^*$. Set

$$\Psi(\mu, \phi_o) = \int_{\infty}^{\infty} dk\tilde{\Psi}(k)\underline{e}_k(\mu)e^{i\underline{\omega}(k)(\phi_o-\phi^*)}, \tag{8}$$

where $\phi^* = -\sqrt{3/16\pi G}ln|\mu^*| + \phi_o$. It is easy to verify that $\Psi(\mu, \phi_o)$ is the desired initial data, sharply peaked at $p_\phi = p_\phi^*$ and $\mu = \mu^*$. If evolved using the Wheeler-DeWitt analog $-i\partial_\phi\Psi = \sqrt{\underline{\Theta}}\Psi$ of Eq. (2), it would remain sharply peaked at the chosen classical trajectory and simply follow it into the big-bang singularity [9]. However, if it is evolved via (2), the situation becomes qualitatively different. The state remains sharply peaked at the classical trajectory until the matter density reaches a large critical value (which depends on p_ϕ^*), *but then bounces*, joining on to the "past" portion of a trajectory which was classically headed towards the big crunch (see figures).

To ensure that these results are robust, a variety of numerical simulations were performed to integrate Eq. (1) using the adaptive step, 4th order Runge-Kutta method. Because of space limitation, we will summarize

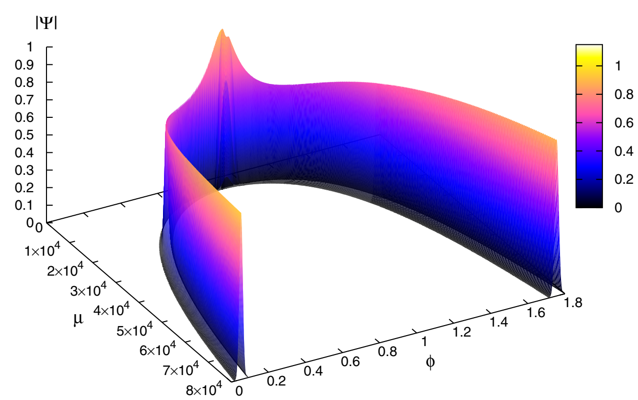

FIG. 1 (color online). The absolute value of the wave function Ψ is plotted as a function of ϕ and μ (whose values are shown in multiples of μ_o). For visualization clarity, only the values of $|\Psi|$ greater than 10^{-4} are shown. Being a physical state, Ψ is symmetric under $\mu \to -\mu$.

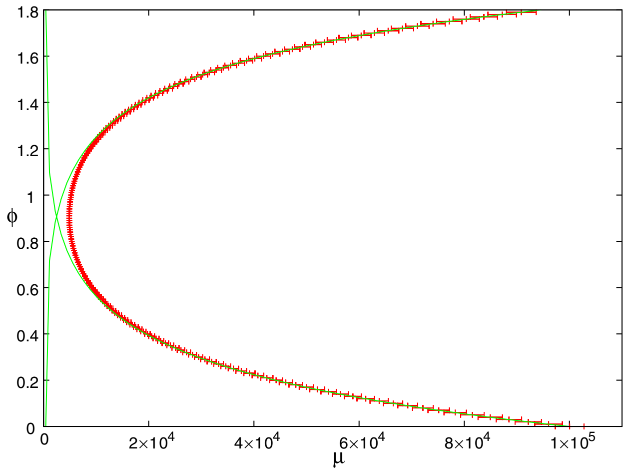

FIG. 2 (color online). The expectation values of Dirac observables $\hat{\mu}|_\phi$ are plotted (in multiples of μ_o), together with their dispersions. They exhibit a quantum bounce which joins the contracting and expanding classical trajectories marked by fainter lines.

only one of these. Here, we chose $\epsilon = 2\mu_o$ and initial data with $p_\phi^* = 10^4\sqrt{G\hbar^2}$, $\mu^* = 10^5\mu_o$, and the spread in the Gaussian $\tilde{\Psi}(k)$ given by $\Delta p_\phi / p_\phi^* = 7.5 \times 10^{-3}$, where Δp_ϕ is the uncertainty in p_ϕ. The boundary of the numerical grid was chosen at $1.5\mu^*$ [where $|\Psi(\mu, \phi_o)| < 10^{-24}$]. On the boundary, Eq. (2) was approximated by the Wheeler-DeWitt equation and "outgoing wave" boundary conditions were imposed. Results of the evolution exhibit a quantum bounce as shown in Figs. 1 and 2. Away from the Planck regime the uncertainties in the Dirac observables are essentially constant.

We conclude with a few remarks. (1) The main limitation of this analysis is the restriction to homogeneity and isotropy. The approach can be readily extended to incorporate anisotropic models and potentials for scalar fields. However, incorporation of inhomogeneities has only just begun. The hope is that the deterministic evolution of LQC will enable one to evolve perturbations across the Planck regime. (2) The dramatic difference between the predictions of the Wheeler-DeWitt theory and LQC can be intuitively understood through effective equations which can be derived from Eq. (1) using certain approximations [9]. One finds that quantum geometry (which is ignored in the Wheeler-DeWitt theory) modifies the Friedmann equations. The modifications are significant only in the Planck regime and come with the sign required to make gravity *repulsive*. (3) A common feature with the early LQC papers is that we did not have to introduce new physical input such as a boundary condition at the singularity. We only asked that the quantum state be semiclassical at late times. This is an observational fact rather than a new theoretical input or a philosophical preference. However, there are also notable differences from the existing LQC literature. First, while much of the phenomenological work [3] in LQC has incorporated quantum geometry effects only on the matter Hamiltonian, here they were incorporated also in the gravitational part. Second, we constructed the physical Hilbert space, Dirac observables, and semiclassical states, thereby extracting physics of the Planck regime, going significantly beyond the demonstration of singularity resolution. Specifically, our results show that the quantum geometry in the Planck regime serves as a "quantum bridge" between large classical universes, one contracting and the other expanding. Finally, the idea that the scalar field can be used as an internal clock has appeared before, especially in [13]. However, that analysis used conventional quantum mechanics rather than the Bohr compactification which descends from full LQG. Our final physical Hilbert space is also different; its construction is not motivated by the Kodama state.

We would like to thank Martin Bojowald, Jim Hartle, Wojciech Kaminski, and Jerzy Lewandowski for discussions. This work was supported in part by the NSF Grants No. PHY-0354932 and No. PHY-0456913, the Alexander von Humboldt Foundation, and the Eberly research funds of Penn State.

*Electronic address: ashtekar@gravity.psu.edu
†Electronic address: pawlowsk@gravity.psu.edu
‡Electronic address: singh@gravity.psu.edu

[1] R. Brustein and G. Veneziano, Phys. Lett. B **329**, 429 (1994); J. Khoury, B. A. Ovrut, N. Seiberg, P. J. Steinhardt, and N. Turok, Phys. Rev. D **65**, 086007 (2002).
[2] A. Ashtekar and J. Lewandowski, Class. Quantum Grav. **21**, R53 (2004); C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, England, 2004); T. Thiemann, gr-qc/0110034.
[3] M. Bojowald, Living Rev. Relativity **8**, 11 (2005).
[4] M. Bojowald, Phys. Rev. Lett. **86**, 5227 (2001).
[5] M. Bojowald, Phys. Rev. Lett. **89**, 261301 (2002).
[6] S. Tsujikawa, P. Singh, and R. Maartens, Classical Quantum Gravity **21**, 5767 (2004).
[7] J. Brunnemann and T. Thiemann, Classical Quantum Gravity **23**, 1395 (2006).
[8] C. Kiefer, Phys. Rev. D **38**, 1761 (1988).
[9] A. Ashtekar, T. Pawlowski, and P. Singh, Quantum Nature of the Big Bang: An Analytical and Numerical Investigation, I (gr-qc/0604013) and II (to be published).
[10] A. Ashtekar, M. Bojowald, and J. Lewandowski, Adv. Theor. Math. Phys. **7**, 233 (2003).
[11] J. Lewandowski, A. Okolow, H. Sahlmann, and T. Thiemann, gr-qc/0504147; C. Flieschhack, math-ph/0407006.
[12] D. Marolf, gr-qc/0011112; J. B. Hartle and D. Marolf, Phys. Rev. D **56**, 6247 (1997).
[13] S. Alexander, J. Malecki, and L. Smolin, Phys. Rev. D **70**, 044025 (2004); J. Malecki, Phys. Rev. D **70**, 084040 (2004).