Quantized Matter Fields and the Avoidance of Singularities in General Relativity*

Leonard Parker and S. A. Fulling
Department of Physics, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201†
(Received 27 November 1972; revised manuscript received 5 February 1973)

We consider a classical gravitational field minimally coupled to a quantized neutral scalar field possessing mass. We are especially concerned with the effects of particle creation and quantum coherence on the premises and conclusions of the singularity theorems, which imply the inevitability of singularities in classical general relativity. A closed Robertson-Walker geometry is used throughout. Nongravitational interactions are not considered. The source of the gravitational field in the Einstein equations is the expectation value of the energy-momentum tensor of the quantized scalar field. Lacking a general prescription for obtaining a finite operator from the divergent formal expression for the energy-momentum tensor, we confine our attention to situations in which plausible special methods are available. We show that quantum coherence effects in this semiclassical model can result in a violation of the energy conditions which enter into the singularity theorems. Then we exhibit numerical solutions of the coupled Einstein and scalar field equations in which a Friedmann-like collapse is stopped and converted to a Friedmann-like expansion. (In this calculation one mode of the quantum field was assumed dominant.) We conclude that quantum effects of the type considered here can sometimes lead to avoidance of the cosmological singularity, at least on the time scale of one Friedmann expansion.

I. INTRODUCTION

Several powerful theorems have been proved which imply the inevitability of singularities in gravitational collapse and cosmology, under assumptions of a very general nature.[1] These theorems deal with a classical gravitational field obeying Einstein's equations. The energy-momentum tensor acting as the source of the gravitational field need only satisfy a number of plausible requirements, known as energy conditions. The possibility that the conclusions of the singularity theorems, and indeed the singularities themselves, can be avoided when the quantum nature of the matter and/or the gravitational field is taken into account has been considered in various contexts.[2]

Wheeler[3] suggested that quantization of the gravitational field could effectively bypass the cosmological singularity. Calculations of Misner and others,[4] based on quantized models with restricted degrees of freedom, have left the situation inconclusive.

Others have considered quantum effects associated with the matter in the universe. Nariai and Tomita[5] have studied modified Lagrangians for the gravitational field, containing terms quadratic in the Ricci tensor and the scalar curvature. Their starting point is a modification of a treatment by Utiyama and DeWitt[6] of divergences in the energy-momentum tensor of quantized matter fields. For certain choices of parameters in their modified Einstein equations Nariai and Tomita obtained isotropic, homogeneous cosmological solutions in which the cosmological singularity is replaced by a finite minimum in the radius function. Their solutions asymptotically approach Friedmann universes in one time direction, but apparently not in both. Analogous modifications of the Einstein equations were suggested by Sakharov,[7] and similar calculations have been carried out on this basis.[8]

From a different point of view, Bahcall and Frautschi[9] have discussed a possible hadron barrier to gravitational collapse. Basing their reasoning on the physics of strongly interacting particles, they suggested that collapse toward a singularity may be reversed when the dimensions of the system become comparable to the Compton wavelength of the pion (10^{-13} cm).

The present work argues for a connection between quantum matter and avoidance of collapse in still a third way. We study the semiclassical model of a classical gravitational field coupled to a quantized scalar (or pseudoscalar) field possessing mass. Thus the reaction of the particles created[10] by the gravitational field back on the expansion is explicitly taken into account. Nongravitational interactions, other than those implicit in the mass term, are not taken into account. We will not be concerned with quadratic terms in the gravitational Lagrangian; rather, we assume the usual gravitational field equation. Unlike Nariai and Tomita, we use as the source of the field the expectation value of a quantum-field-theoretical energy-momentum tensor, rather than a classical matter distribution. The metric under consideration is of the Robertson-Walker type with closed three-space.

We show that there exist states in which quantum coherence effects give rise to negative pressure terms in $\langle T_{\mu\nu}\rangle$, sufficiently large to violate the energy conditions of the singularity theorems.[11] For a particular class of states such that the effects of the mode of lowest momentum are dominant, we numerically integrate the coupled equations for the gravitational and the quantized scalar field. The resulting isotropic and homogeneous cosmological solution effectively coincides with a classical dust-filled Friedmann-Lemaître universe when the radius becomes large with respect to the Compton wavelength, m^{-1}, associated with the scalar field. However, the radius function, $a(t)$, possesses a minimum at a radius of order m^{-1}, and it is time-symmetric about that minimum, unlike the models in Ref. 5. (Here t is the cosmic time appearing in the Robertson-Walker line element.) Taking m to be the pion mass, we obtain in one case a maximum radius of order 10^{30} cm. The dimensions of the system under consideration here are always large with respect to the Planck length, 10^{-33} cm. The question whether a state exists for which the above behavior is repeated periodically for all time remains open. We show that collapse to the singularity does occur for some states; that is, the quantum effect does not eliminate the singularity inevitably, but only for certain states.

Before turning to the details, we pause to discuss the theoretical underpinnings of our calculation. Quantum effects of gravity become very significant when the Planck scale of length (10^{-33} cm) becomes characteristic of the system.[12] For much larger dimensions it seems reasonable to employ a semiclassical description, in which the quantized matter fields are coupled to the classical gravitational field. The most obvious approach of this type starts from the fundamental equation[13-15]

$$G_{\mu\nu} = -8\pi G\langle T_{\mu\nu}\rangle \ , \qquad (1)$$

where[16] G is the Newtonian gravitational constant, $G_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$, formed from the classical metric, and $\langle T_{\mu\nu}\rangle$ is the expectation value, with respect to some quantum state, of an operator representing the energy-momentum tensor in a quantum theory of the matter in the universe. The state could be a mixed state (represented by a density matrix) as well as a pure state.

This ansatz raises three conceptual problems. The first is the very meaning of a quantum theory of the entire universe. Rather than delve into the quantum theory of measurement, in which context this question has been considered, we shall assume without further discussion that it is meaningful to talk about the quantum state (pure or mixed) of the matter in a cosmological model.

One might also hesitate over the legitimacy of coupling quantum and classical quantities as in Eq. (1). The treatment employed here of the gravitational field as a classical object is analogous to the treatment of the electrostatic potential in the Thomas-Fermi model of a many-electron atom,[17] of the nuclei in the Born-Oppenheimer model of a molecule,[18] of the electromagnetic field in a variety of calculations in the electrodynamics of nonrelativistic electrons and atoms,[19] and of the electromagnetic field in the now classic calculations of vacuum polarization effects in quantum electrodynamics with external potentials.[20] Such a theory is usually understood as an approximation to a more fundamental fully quantized theory.[21] Occasionally, however, semiclassical theories are presented as "ultimate" descriptions of the phenomena, without deeper quantum levels of reality.[22] The authors believe that the present work is consistent with either interpretation.[23]

The third point is one which requires some sort of resolution before calculations can be carried out with Eq. (1). A naive calculation of $\langle T_{\mu\nu}\rangle$ results in a divergent expression, and there does not seem to be a satisfactory unique analog of the "vacuum energy subtraction" which is performed at this point in the corresponding special-relativistic field theory. Utiyama and DeWitt (Ref. 6) showed that the elimination of divergent parts of the expectation value of the energy-momentum tensor, $\langle T_{\mu\nu}\rangle$, evidently could be carried out, for a weak and asymptotically vanishing gravitational field, by the introduction of counterterms in the gravitational Lagrangian quadratic in the Ricci tensor $R_{\mu\nu}$ and the scalar curvature R. Those terms were chosen to cancel exactly the divergent parts of $\langle T_{\mu\nu}\rangle$, so that the usual gravitational field equations resulted, with the Einstein tensor $G_{\mu\nu}$ coupled to the "finite part" of $\langle T_{\mu\nu}\rangle$. (The modified equations of Ref. 5 arise from the possibility that this cancellation is not exact, so that the gravitational Lagrangian has finite terms quadratic in $R_{\mu\nu}$ and R.) DeWitt[24] has given a different renormalization procedure which can be applied in an arbitrarily strong but asymptotically vanishing gravitational field. Another prescription for extracting a finite part of $\langle T_{\mu\nu}\rangle$ has been given by Zel'dovich and Starobinsky[25]; its justification, especially with respect to uniqueness, is not entirely clear. Consequently, we have endeavored to avoid the general problem of vacuum subtraction, or renormalization of the energy-momentum tensor, by considering only states for which we believe the details of renormalization to be irrelevant.

Two additional points, which are not discussed further in the present paper, should be mentioned. We use in this paper the canonical energy-momentum tensor, rather than the modified or conformal energy-momentum tensor.[26] The latter has some interesting properties, especially in connection with renormalization, which make it worthwhile to explore whether its use could significantly alter the present conclusions. The second point concerns the definition of $\langle T_{\mu\nu}\rangle$. Utiyama and DeWitt (Ref. 6), considering a case in which an S matrix could be defined, inserted S into the expectation value. The introduction of the S matrix evidently resulted from an application of Schwinger's method for obtaining Green's functions of quantized fields to the fully quantized Einstein equations. In the present problem of a closed Robertson-Walker universe it is not clear what is meant by the S matrix. Like the authors of Refs. 13–15, we use the ordinary expectation value (without the S matrix) as the source of the gravitational field.

II. ENERGY-MOMENTUM TENSOR

The scalar field is characterized by the Lagrangian density

$$\mathfrak{L}=\tfrac{1}{2}(-g)^{1/2}(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-m^2\phi^2)\ , \tag{2}$$

which leads to the covariant generalization of the Klein-Gordon equation,

$$\nabla_\mu\nabla^\mu\phi+m^2\phi=0\ . \tag{3}$$

Here ∇_μ denotes the covariant derivative. The energy-momentum tensor which couples to the gravitational field is obtained by variation with respect to $g^{\mu\nu}$:

$$\delta\int\mathfrak{L}d^4x=\tfrac{1}{2}\int T_{\mu\nu}\delta g^{\mu\nu}(-g)^{1/2}d^4x\ .$$

The result is

$$T_{\mu\nu}=\tfrac{1}{2}(\partial_\mu\phi\partial_\nu\phi+\partial_\nu\phi\partial_\mu\phi)-g_{\mu\nu}L\ , \tag{4}$$

where

$$L=(-g)^{-1/2}\mathfrak{L}\ .$$

We canonically quantize the scalar field in the classical Riemannian space-time.[27] The momentum conjugate to ϕ is

$$\pi=\frac{\partial\mathfrak{L}}{\partial(\partial_0\phi)}\ . \tag{5}$$

The canonical commutation relations are (in arbitrary coordinates with spacelike constant-time hypersurfaces)

$$\begin{aligned}&[\phi(x),\phi(x')]_t=0\ ,\\&[\pi(x),\pi(x')]_t=0\ ,\\&[\phi(x),\pi(x')]_t=i\delta^{(3)}(\vec{x}-\vec{x}')\ .\end{aligned} \tag{6}$$

The system under consideration is governed by the gravitational field equation, (1), and the scalar wave equation, (3). The Heisenberg picture, in which the operators carry the full time development, will be used throughout.

We consider here only solutions with isotropic, homogeneous spaces and matter distributions. In particular, the interval is that of Robertson and Walker corresponding to a closed universe,

$$ds^2=dt^2-a(t)^2h_{jk}dx^j dx^k\ , \tag{7}$$

where h_{jk} is the metric of a three-space of constant positive curvature equal to unity. (In a frequently used system of coordinates, $h_{jk}=\delta_{jk}(1+\frac{1}{4}r^2)^{-2}$, with $r^2=\delta_{ij}x^ix^j$.) For consistency with the Einstein equation, the state vector for the scalar field must be chosen such that $\langle T_\mu{}^\nu\rangle$ depends only on t. When the energy density is positive definite, as in the examples which we numerically integrate below, a minimum in $a(t)$ is possible only for positive spatial curvature; this is why we consider the *closed* Robertson-Walker model.

The metric of Eq. (7) leads to the canonical momentum,

$$\pi=a^3h^{1/2}\partial_0\phi\ , \tag{8}$$

and the scalar wave equation,

$$\partial_0{}^2\phi+3a^{-1}\partial_0a\partial_0\phi-a^{-2}\Delta^{(3)}\phi+m^2\phi=0\ , \tag{9}$$

where $h=\det\{h_{ij}\}$, and

$$\Delta^{(3)}\phi=h^{-1/2}\partial_j(h^{1/2}h^{jk}\partial_k\phi) \tag{10}$$

is the three-dimensional covariant Laplacian operator. The eigenfunctions of $\Delta^{(3)}$ are the four-dimensional spherical harmonics.[28] We denote them by Y_{lmn}, where l can take the values $0,1,2,\ldots$, and m and n each run in integral steps between $-l/2$ and $l/2$. The eigenvalue of $\Delta^{(3)}$ corresponding to Y_{lmn} is $-l(l+2)$. The eigenfunctions are normalized so that

$$\int Y^*_{lmn}(\vec{x})Y_{l'm'n'}(\vec{x})h^{1/2}(\vec{x})d^3x=\delta_{ll'}\delta_{mm'}\delta_{nn'}, \tag{11}$$

where the integration extends over the entire three-sphere. The phases are chosen in accordance with

$$Y^*_{lmn}(\vec{x})=Y_{l,-m,-n}(\vec{x})\ . \tag{12}$$

These spherical harmonics are the same as those given explicitly in Eq. (32) of Bander and Itzykson (Ref. 28), except for a multiplicative factor of i whenever $(m-n)$ is an odd integer.

The Hermitian scalar field can be written in the form

$$\phi(\vec{x}, t) = \sum_{l,\alpha} [A_{l\alpha} Y_{l\alpha}(\vec{x})\psi_l(t) + \text{H.c.}] , \tag{13}$$

where H.c. denotes the Hermitian conjugate of the first term and the index α is an abbreviation for m and n (similarly, $-\alpha$ will denote $-m$ and $-n$). The $A_{l\alpha}$ are time-independent annihilation operators satisfying the commutation relations

$$[A_{l\alpha}, A_{l'\alpha'}] = 0 ,$$

$$[A^\dagger_{l\alpha}, A^\dagger_{l'\alpha'}] = 0 , \tag{14}$$

and

$$[A_{l\alpha}, A^\dagger_{l'\alpha'}] = \delta_{ll'}\delta_{\alpha\alpha'} .$$

It is convenient to define a new timelike variable τ (not proper time) by

$$\tau = \int^t a^{-3} dt' . \tag{15}$$

It follows from Eqs. (9), (14), and (15) that the time-separated equation is

$$\partial_\tau{}^2\psi_l(\tau) + a(\tau)^6\omega_l(\tau)^2\psi_l(\tau) = 0 , \tag{16}$$

with

$$\omega_l{}^2 = \frac{l(l+2)}{a^2} + m^2 . \tag{17}$$

Equations (13), (6), and (14) are consistent if and only if

$$\psi_l \partial_\tau \psi_l^* - \psi_l^* \partial_\tau \psi_l = i ; \tag{18}$$

the form of Eq. (16) guarantees that when this Wronskian condition is imposed at one time, it holds at all times. However, Eq. (18) does not uniquely determine the particular solutions ψ_l to be used in the expansion (13), nor, consequently, the $A_{l\alpha}$.

The main purpose of introducing the $A_{l\alpha}$ is to construct a Fock space[29] of state vectors, so that matrix elements of physically relevant quantities can be defined; an interpretation in terms of physical particles is not implied. It might be questioned whether the Fock space of the $A_{l\alpha}$ is the "correct" Hilbert space for the field theory, in view of the nonuniqueness of these operators. (Different choices of ψ_l will yield, in general, unitarily inequivalent Fock representations of the field algebra.) In the abstract algebraic approach to quantum theory[30] it is argued that different representations of an algebra of observables are physically equivalent, in the following sense: with *each* Hilbert-space representation of the algebra there is associated some pure or mixed quantum state which reproduces the results of a given finite set of observations to any specified degree of accuracy. From this point of view a state vector is just a mathematical device to summarize some given expectation values, and hence the choice of a particular Hilbert space is a matter of convenience only. In the present work we shall implement this idea at least partially, by considering states which are parametrized by expectation values of certain quantities[31] appearing in expressions for the components of the energy-momentum tensor, the primary observable of a system in the gravitational context. From the standpoint of rigorous theory, however, there is much unfinished business here: One does not have a well-defined algebra of observables, but only a divergent formal expression for the energy-momentum tensor. In algebraic terms the problem of renormalization of the energy-momentum tensor is the problem of defining a suitable finite $T_{\mu\nu}(x)$ (an unbounded operator-valued distribution) in terms of the elements of a C^* algebra of bounded observables associated with the canonical field algebra (6). This problem is circumvented in the present paper by introducing a plausible method of calculating a finite $T_{\mu\nu}(x)$ within a given Fock representation for a special category of states.

The diagonal components of $T_\mu{}^\nu$ are, for a diagonal metric,

$$T_0{}^0 = \tfrac{1}{2}\left((\partial_0\phi)^2 + a^{-2}\sum_{i=1}^{3} h_{ii}{}^{-1}(\partial_i\phi)^2 + m^2\phi^2\right) , \tag{19}$$

and (no sum over j)

$$T_j{}^j = -\tfrac{1}{2}\bigg((\partial_0\phi)^2 + 2a^{-2}h_{jj}{}^{-1}(\partial_j\phi)^2 - a^{-2}\sum_{i=1}^{3} h_{ii}{}^{-1}(\partial_i\phi)^2 - m^2\phi^2\bigg) . \tag{20}$$

We consider expectation values $\langle T_\mu{}^\nu\rangle$ having the symmetry properties required for a source of the Robertson-Walker metric, namely, that they be independent of the spatial coordinates $\vec{x}$, that the off-diagonal elements of $\langle T_\mu{}^\nu\rangle$ vanish, and that $\langle T_1{}^1\rangle = \langle T_2{}^2\rangle = \langle T_3{}^3\rangle$. Therefore,

$$\langle T_i{}^i\rangle = \tfrac{1}{3}\sum_{j=1}^{3}\langle T_j{}^j\rangle = \frac{1}{6\pi^2}\int d^3x\, h^{1/2}\sum_{j=1}^{3}\langle T_j{}^j\rangle . \tag{21}$$

Noting that

$$\int d^3x\, h^{1/2}\sum_j h_{jj}{}^{-1}(\partial_j\phi)^2 = -\int d^3x \sum_j \partial_j(h^{1/2}h^{jj}\partial_j\phi)\phi = -\int d^3x\, h^{1/2}\phi\Delta^{(3)}\phi ,$$

we obtain from Eq. (20)

$$\langle T_i{}^i\rangle = -\frac{1}{4\pi^2}\int d^3x\, h^{1/2}\langle(\partial_0\phi)^2 + \tfrac{1}{3}a^{-2}\phi\Delta^{(3)}\phi - m^2\phi^2\rangle . \tag{22}$$

Now

$$-\Delta^{(3)}\phi = \sum_{l\alpha} l(l+2)(A_{l\alpha}Y_{l\alpha}\psi_l + \text{H.c.}) . \tag{23}$$

Hence, using $Y_{l,\alpha} = Y^*_{l,-\alpha}$, one finds that

$$-\int d^3x\, h^{1/2}\phi\Delta^{(3)}\phi = \sum_{l\alpha} l(l+2)[\psi_l{}^2 A_{l,-\alpha}A_{l\alpha} + |\psi_l|^2(A^\dagger_{l\alpha}A_{l\alpha} + A_{l\alpha}A^\dagger_{l\alpha}) + \psi_l^{*2}A^\dagger_{l\alpha}A^\dagger_{l,-\alpha}] . \tag{24}$$

Similarly, $\int d^3x\, h^{1/2}\phi^2$ and $\int d^3x\, h^{1/2}(\partial_0\phi)^2$ are the same as the right side of Eq. (24), except that the factor $l(l+2)$ is absent in both expressions, and that ψ_l is replaced throughout the latter expression by $\partial_0\psi_l$. Therefore, one obtains from Eq. (22), with $\langle T_i{}^i\rangle = -p$,

$$p = (4\pi^2)^{-1}\sum_{l\alpha}\left\{\langle A_{l,-\alpha}A_{l\alpha}\rangle\left[(\partial_0\psi_l)^2 - \left(\frac{l(l+2)}{3a^2} + m^2\right)\psi_l{}^2\right] + \text{c.c.} + \langle A^\dagger_{l\alpha}A_{l\alpha} + A_{l\alpha}A^\dagger_{l\alpha}\rangle\left[|\partial_0\psi_l|^2 - \left(\frac{l(l+2)}{3a^2} + m^2\right)|\psi_l|^2\right]\right\}, \tag{25}$$

where c.c. denotes the complex conjugate of the preceding term. In the same way, one obtains from Eq. (19), with $\langle T_0{}^0\rangle = \rho$,

$$\rho = (4\pi^2)^{-1}\sum_{l\alpha}\{\langle A_{l,-\alpha}A_{l\alpha}\rangle[(\partial_0\psi_l)^2 + \omega_l{}^2\psi_l{}^2] + \text{c.c.} + \langle A^\dagger_{l\alpha}A_{l\alpha} + A_{l\alpha}A^\dagger_{l\alpha}\rangle[|\partial_0\psi_l|^2 + \omega_l{}^2|\psi_l|^2]\} . \tag{26}$$

If the expectation values on the right-hand sides of Eqs. (25) and (26) are taken with respect to a pure quantum state Ψ which is not, in fact, homogeneous and isotropic, then these formulas give the expected pressure and energy density for the mixed quantum state obtained by averaging the density matrix of Ψ over all positions and orientations (i.e., averaging over all the images of Ψ under the group of isometries of the three-sphere). (It is easy to see that averaging a density matrix in this way is equivalent inside expectation values to averaging the operators representing the observables, as we have done.) The semiclassical approximation should still be good in such a situation, provided (cf. footnote 23) that Ψ is not grossly asymmetric on the cosmological scale. Indeed, a realistic state allowing for small-scale inhomogeneities would be of this type.

The above expressions should give the pressure and energy density of an isotropic, homogeneous distribution of scalar particles, except for divergences which remain even for the state annihilated by the $A_{l\alpha}$. Plausible procedures for dealing with the divergences are available in the particular cases we consider. We first discuss the adiabatic limit, in which simple normal ordering seems justified, and then turn to a nonadiabatic case in which the lowest mode is dominant.

III. ADIABATIC LIMIT AND ENERGY CONDITIONS

Suppose that $a(t)$ is slowly varying, so that the WKB approximation can be applied to Eq. (16). We call this slow expansion regime the *adiabatic limit*. In that limit, one has from Eqs. (16) and (18)

$$\psi_l = (2a^3\omega_l)^{-1/2}\exp\left(-i\int^\tau a^3\omega_l d\tau\right) = (2a^3\omega_l)^{-1/2}\exp\left(-i\int^t \omega_l dt\right) . \tag{27}$$

In the adiabatic limit there is no mixing of the positive and negative frequency parts of the field, so that there is no particle creation, and the $A_{l\alpha}$ can be identified at all times with the physical particles present.[32] In that case, simple normal ordering in Eqs. (25) and (26) with respect to the $A_{l\alpha}$ seems justified, in analogy with the special-relativistic procedure. The resulting renormalized energy-momentum tensor has vanishing four-divergence and also leads to the correct classical limit. (By contrast, when positive and negative frequencies get mixed by the expansion, there is generally no preferred time at which to normal order, and a new method of renormalization is called for.)

Substituting Eq. (27) into the expressions for p and ρ, and normal ordering, one obtains

$$p = (2\pi^2a^3)^{-1}\sum_{l\alpha}\left\{-\langle A_{l,-\alpha}A_{l\alpha}\rangle\omega_l{}^{-1}\left(\frac{l(l+2)}{3a^2} + \frac{m^2}{2}\right)\times\exp\left(-2i\int^t\omega_l dt\right) + \text{c.c.} + \langle A^\dagger_{l\alpha}A_{l\alpha}\rangle\frac{l(l+2)}{3a^2\omega_l}\right\} \tag{28}$$

and

$$\rho=(2\pi^2a^3)^{-1}\sum_{l\alpha}\langle A_{l\alpha}^\dagger A_{l\alpha}\rangle\omega_l\ . \tag{29}$$

To check the classical limit of these expressions, consider a state which is a mixture of eigenstates of the number operators $A_{l\alpha}^\dagger A_{l\alpha}$. Then the first two terms in Eq. (28) vanish. Identifying ω_l with the energy and $M_l=[l(l+2)/a^2]^{1/2}$ with the magnitude of the momentum of a particle in mode l, one finds in the low-velocity limit [i.e., when only modes with $l(l+2)/a^2 \ll m^2$ are occupied] that

$$pV=(3m)^{-1}(M^2)_{av}=kT\ , \tag{30}$$

where $(M^2)_{av}$ is the average momentum, $V=2\pi^2a^3$ is the volume of the universe, and we have set $(2m)^{-1}(M^2)_{av}$ equal to $(3/2)kT$ in accordance with the equipartition theorem. Thus, the particles behave like an ideal gas. In the high-velocity limit, when m can be neglected, one obtains the familiar result

$$p=\tfrac{1}{3}\rho\ . \tag{31}$$

We now ask whether there exist states for which the energy conditions of the classical singularity theorems are violated. From Eqs. (28) and (29), it is clear that ρ is positive definite, while p is not. Therefore, in the isotropic case under consideration here, the weak energy condition (Ref. 1) can be written as

$$\rho+p\geq 0\ , \tag{32}$$

and the energy condition as

$$\rho+3p\geq 0\ . \tag{33}$$

In the present context, if Eq. (33) is satisfied, then Eq. (32) must be satisfied, and if Eq. (32) is violated, then Eq. (33) must be violated. One or the other of these conditions (in a more general form applicable also to anisotropic models) has been used as a premise in almost all derivations of singularity theorems.

To show that the energy conditions can be violated, consider a normalized real function of the form $Z(\vec{x})=\sum_\alpha C_{l\alpha}Y_{l\alpha}(\vec{x})$ (fixed l). The argument leading to Eqs. (28) and (29) could have been carried out with $\{Y_{l\alpha}\}$ replaced in Eq. (13) by a basis containing Z as a member. Let A be the annihilation operator coefficient of Z in the modified Eq. (13). One would then obtain (as a consequence of the reality of Z) expressions for p and ρ just like Eqs. (28) and (29), but involving terms in $\langle AA\rangle$ and $\langle A^\dagger A\rangle$ (in place of $\langle A_{l,-\alpha}A_{l\alpha}\rangle$, etc.). Consider now a state in which only the Z mode is occupied. [If $l=0$, the pure state has the proper symmetry; if l is large, Z can be chosen to make the true energy density and pressure expectation values approximately constant (spatially) on a coarse scale, so that, as explained at the end of Sec. II, the semiclassical approach is still valid for the symmetrical mixed state described by our formulas.] To show that there are times when p becomes sufficiently negative to violate Eq. (32), we use the following theorem (Appendix A): If A and $A^\dagger$ are a pair of annihilation and creation operators, then the expectation values $\langle A^\dagger A\rangle$ and $\langle AA\rangle$ with respect to normalized state vectors run through all pairs of values consistent with the constraints $|\langle AA\rangle|^2\leq\langle A^\dagger A\rangle(\langle A^\dagger A\rangle+1)$ and $\langle A^\dagger A\rangle\geq 0$. Using this theorem, we choose a state for which

$$|\langle AA\rangle|=\langle A^\dagger A\rangle(1+\langle A^\dagger A\rangle^{-1})^{1/2}\ . \tag{34}$$

Furthermore, consider a particular time t and a choice of phase such that

$$\langle AA\rangle\exp\left(-2i\int^t\omega_l\,dt\right)=\langle A^\dagger A\rangle(1+\langle A^\dagger A\rangle^{-1})^{1/2}\ . \tag{35}$$

Then Eqs. (28) and (29), modified as just described, yield at time t

$$-\frac{p}{\rho}=\frac{(1+\langle A^\dagger A\rangle^{-1})^{1/2}[2l(l+2)/3a^2+m^2]-l(l+2)/3a^2}{[l(l+2)/a^2+m^2]}\ . \tag{36}$$

According to the theorem quoted above, we are still free to choose as the value of $\langle A^\dagger A\rangle$ any positive number. By choosing $\langle A^\dagger A\rangle$ sufficiently small, one can clearly make $-p/\rho$ larger than any given finite number.[33] Therefore, there exist states for which the energy conditions of the singularity theorems are violated.

The appearance of negative pressure is a quantum coherence effect among the states with different numbers of particles. The same effect should occur as well for a gas of scalar particles in Minkowski space, and is therefore not restricted to the context of general relativity. The term in p which gives rise to the negative pressure is rapidly oscillating with a period of ω_l^{-1}. The time average of the pressure is the nonnegative classical value $\langle A^\dagger A\rangle l(l+2)(3a^2\omega_l)^{-1}$. Therefore, one would expect the most significant effect of the negative pressure term to occur when the characteristic times or dimensions of the system are less than or of the order m^{-1}. For the system under consideration, the adiabatic approximation is no longer valid when $a(t)$ is less than m^{-1} because significant gravitationally induced particle creation occurs. Nevertheless, the present considerations strongly suggest that quantum effects can also violate the energy conditions when $a(t)$ is small and rapidly changing. We therefore return to the non-adiabatic case, and consider whether there exist states in which a collapse toward the cosmological singularity can be reversed.

IV. EQUATIONS OF MOTION

For the purpose of numerical integration of the coupled Einstein and scalar wave equations, it is convenient to reexpress Eq. (16) in the form of two first-order equations. Let

$$y_l = a^{3/2}\omega_l^{1/2} \tag{37}$$

and

$$\psi = 2^{-1/2} y^{-1}(\hat\alpha + \hat\beta)\ , \tag{38}$$

where[34]

$$\hat\alpha = \alpha^* \exp\left(-i\int^\tau y^2 d\tau\right) = \alpha^* \exp\left(-i\int^t \omega dt\right) \tag{39a}$$

and

$$\hat\beta = \beta^* \exp\left(i\int^\tau y^2 d\tau\right) = \beta^* \exp\left(i\int^t \omega dt\right)\ . \tag{39b}$$

(We suppress the subscript l.) To determine α and β, which are functions of τ (except in the adiabatic limit), we impose a second condition, namely

$$\partial_\tau \psi = -i 2^{-1/2} y(\hat\alpha - \hat\beta)\ . \tag{40}$$

From Eqs. (38) and (40) it follows that Eq. (18) is equivalent to

$$|\alpha|^2 - |\beta|^2 = 1\ . \tag{41}$$

Differentiating Eq. (38) with respect to τ, and using Eq. (40), we then obtain

$$(\partial_\tau \alpha^*) \exp\left(-i\int^\tau y^2 d\tau\right) + (\partial_\tau \beta^*) \exp\left(i\int^\tau y^2 d\tau\right) = y^{-1}\partial_\tau y(\hat\alpha + \hat\beta)\ . \tag{42}$$

Differentiation of Eq. (40) with respect to τ, and use of the equation of motion, which requires that $\partial_\tau^2 \psi = -y^4\psi$, give

$$(\partial_\tau \alpha^*) \exp\left(-i\int^\tau y^2 d\tau\right) - (\partial_\tau \beta^*) \exp\left(i\int^\tau y^2 d\tau\right) = -y^{-1}\partial_\tau y(\hat\alpha - \hat\beta)\ . \tag{43}$$

From the conjugates of Eqs. (42) and (43) we obtain the first-order equations of motion in the form

$$\partial_0 \alpha = y^{-1}\partial_0 y \exp\left(-2i\int^t \omega dt\right)\beta \tag{44a}$$

and

$$\partial_0 \beta = y^{-1}\partial_0 y \exp\left(2i\int^t \omega dt\right)\alpha\ , \tag{44b}$$

where we have returned to the original time variable $t = x^0$. Here

$$y^{-1}\partial_0 y = 3a^{-1}\partial_0 a\, \omega^{-2}\left(\frac{l(l+2)}{3a^2} + \frac{m^2}{2}\right)\ . \tag{45}$$

Equations (44) are used later in the numerical integration.

We will also need the expressions for p and ρ in terms of α and β. From Eqs. (38), (40), (25), and (26), one finds that

$$p = (6\pi^2 a^3)^{-1} \sum_{l\alpha} \omega_l^{-1} \left\{ \langle A_{l,-\alpha} A_{l\alpha}\rangle \left[\frac{l(l+2)}{a^2}\alpha_l^*\beta_l^* - \frac{1}{2}\left(\frac{2l(l+2)}{a^2} + 3m^2\right)(\hat\alpha_l^2 + \hat\beta_l^2)\right] + \text{c.c.} + \langle A_{l\alpha}^\dagger A_{l\alpha} + A_{l\alpha}A_{l\alpha}^\dagger\rangle \left[\frac{l(l+2)}{a^2}(|\beta_l|^2 + \tfrac{1}{2}) - \left(\frac{2l(l+2)}{a^2} + 3m^2\right)\text{Re}(\hat\alpha_l^*\hat\beta_l)\right]\right\} \tag{46}$$

and

$$\rho = (2\pi^2 a^3)^{-1} \sum_{l\alpha} \omega_l \left[\langle A_{l,-\alpha}A_{l\alpha}\rangle \alpha_l^*\beta_l^* + \text{c.c.} + \langle A_{l\alpha}^\dagger A_{l\alpha} + A_{l\alpha}A_{l\alpha}^\dagger\rangle (|\beta_l|^2 + \tfrac{1}{2})\right]\ , \tag{47}$$

where we have used $|\alpha_l|^2 - |\beta_l|^2 = 1$, and have not normal ordered the operators. Finite expressions for the pressure and energy density will be given later for the particular state under consideration.

Finally, we note that for the closed Robertson-Walker metric, the Einstein equations can be written as

$$[(\partial_0 a)^2 + 1]\, a = (8\pi/3) G a^3 \rho \tag{48}$$

and

$$\partial_0^2 a + (2a)^{-1}[(\partial_0 a)^2 + 1] + 4\pi G p a = 0\ . \tag{49}$$

Equations (44)–(49) are the nonlinear set of equations which will be numerically integrated once an appropriate state vector has been chosen and finite expressions for ρ and p have been determined. When Eq. (49) is used to find the time development of $a(t)$, Eq. (48) constrains the initial data. It is automatically satisfied at other times as a consequence of the vanishing of the four-divergence of $T_\mu{}^\nu$ [Eq. (61) below].

V. CHOICE OF STATE

For our present considerations we are allowing a to change rapidly, so that the WKB approximation is not valid. Under those circumstances, positive and negative frequencies get mixed. Therefore the $A_{l\alpha}$ do not necessarily correspond to physical particles, as they do in the adiabatic limit. As we have remarked, the purpose of introducing the $A_{l\alpha}$ is to construct a Hilbert space (Fock representation), so that expectation values of the energy-momentum tensor can be defined.

For simplicity, we choose a normalized state vector, $|\,\rangle$, such that

$$A_{l\alpha}|\,\rangle = 0 \quad \text{for all } l \neq 0 \,. \tag{50}$$

Furthermore, we make the Bogoliubov transformation

$$A_{00} = \gamma^* B + \delta B^\dagger \,, \tag{51}$$

where

$$|\gamma|^2 - |\delta|^2 = 1 \tag{52}$$

and hence

$$[B, B^\dagger] = 1 \,. \tag{53}$$

The coefficients γ and δ are complex *constants*. Their magnitudes and phases will be chosen later, in accordance with the constraint equation (52). We complete (for a given choice of $A_{l\alpha}$) the specification of the state vector by the condition

$$B^\dagger B|\,\rangle = N|\,\rangle \,, \tag{54}$$

where N is a positive integer. (The state vector $|\,\rangle$ is still in the original Fock space of the $A_{l\alpha}$.)

A word is in order concerning the physical meaning of this state vector. In the calculation below, β_0 will be chosen [Eq. (67)] to vanish at a moment of time symmetry ($t=0$), when $a(t)$ has its minimum. Therefore, if $a(t)$ were to remain constant in the vicinity of $t=0$, the $A_{l\alpha}$ could be interpreted [see Eqs. (37)–(39) and Eq. (13)] as annihilation operators for physical particles. That interpretation of the $A_{l\alpha}$ would still seem to be approximately valid for the actual solution at the moment of time symmetry, since $\partial_0 a$ vanishes there. The state defined above has a structure like that of an excited state in models of superfluidity and superconductivity.[35] The B operators in those theories correspond to elementary excitations or quasiparticles relative to the ground state of an interacting system. (There are N excitations in our state.) The ground state ($N=0$) itself contains amplitudes for the presence of various numbers of pairs of the particles associated with the A operators (Cooper pairs in the case of superconductivity). In our problem the states with definite N are not eigenstates of a Hamiltonian, but they have this same structure of correlated pairs of A-quanta.[36] Such a state is by no means the only type which may lead to reversal of the collapse, but it is convenient for our present purposes. It is natural to speculate that some mechanism might be present, or some interaction added, which would cause bosons or fermions to make a transition into such a state when the density becomes sufficiently great, in analogy with the transition to the BCS ground state in superconductivity. That question is beyond the scope of this paper.

As a consequence of Eqs. (51)–(54), one finds that

$$\langle A_{00}A_{00}\rangle = \gamma^*\delta(2N+1) \tag{55}$$

and

$$\langle A_{00}^\dagger A_{00} + A_{00}A_{00}^\dagger\rangle = (2N+1)(1+2|\delta|^2) \,. \tag{56}$$

Substituting these results into Eqs. (46) and (47) yields

$$p = -(2\pi^2 a^3)^{-1} m(2N+1)\{\mathrm{Re}[\gamma^*\delta(\hat\alpha_0{}^2 + \hat\beta_0{}^2)] + (2|\delta|^2+1)\,\mathrm{Re}(\hat\alpha_0^*\hat\beta_0)\} + (6\pi^2 a^3)^{-1}\sum_{l=1}^{\infty}(l+1)^2\omega_l{}^{-1}\left[\frac{l(l+2)}{a^2}(|\beta_l|^2+\tfrac12) - \left(\frac{2l(l+2)}{a^2}+3m^2\right)\mathrm{Re}(\hat\alpha_l^*\hat\beta_l)\right] \tag{57}$$

and

$$\rho = (\pi^2 a^3)^{-1} m(2N+1)[\mathrm{Re}(\gamma\delta^*\alpha_0\beta_0) + (|\delta|^2+\tfrac12)(|\beta_0|^2+\tfrac12)] + (2\pi^2 a^3)^{-1}\sum_{l=1}^{\infty}\omega_l(|\beta_l|^2+\tfrac12)(l+1)^2 \,. \tag{58}$$

The factor $(l+1)^2$ in the sums from $l=1$ to ∞ come from the range of α, or m and n, for each l. Each of those sums diverges, but is independent of N. We assume that there exists a valid renormalization procedure which results in the replacement of the divergent sums by finite terms, while leaving unchanged the part of the $l=0$ term proportional to N.[37] Since N appears only in the $l=0$ term, we can then choose N large enough that the renormalized contributions of the remaining terms are negligible with respect to the term proportional to N. Thus, for $N \gg 1$, one obtains the result

$$p = -(\pi^2 a^3)^{-1} m N \{\mathrm{Re}[\gamma^* \delta(\hat{\alpha}^2 + \hat{\beta}^2)] + (2|\delta|^2 + 1)\mathrm{Re}(\hat{\alpha}^* \hat{\beta})\} \tag{59}$$

and

$$\rho = (\pi^2 a^3)^{-1} 2 m N \{\mathrm{Re}(\gamma \delta^* \alpha \beta) + (|\delta|^2 + \tfrac{1}{2})(|\beta|^2 + \tfrac{1}{2})\} , \tag{60}$$

where we have dropped the subscripts 0. Recall that γ and δ are constants, while α and β are functions of t, and $\hat{\alpha}$ and $\hat{\beta}$ are defined in Eqs. (39).

One can show as follows that ρ above is positive definite. The expression in curly brackets in Eq. (60) is always greater than or equal to

$$-|\gamma|\,|\delta|\,|\alpha|\,|\beta| + (|\delta|^2 + \tfrac{1}{2})(|\beta|^2 + \tfrac{1}{2}) \equiv Q .$$

Now $(|\alpha| - |\beta|)^2 \geq 0$ implies that

$$|\alpha|\,|\beta| \leq \tfrac{1}{2}(|\alpha|^2 + |\beta|^2) = |\beta|^2 + \tfrac{1}{2} .$$

Similarly,

$$|\gamma|\,|\delta| \leq |\delta|^2 + \tfrac{1}{2} .$$

The equality signs only hold when $|\alpha| = |\beta|$ and $|\gamma| = |\delta|$, which would violate Eqs. (41) and (52). Therefore, Q and consequently ρ in Eq. (60) are strictly positive. On the other hand, p in Eq. (59) oscillates between positive and negative values.

Using the equations (B6) and (B8), derived in an appendix, it is easy to show that p and ρ given by Eqs. (59) and (60) satisfy the continuity equation

$$\partial_0(a^3 \rho) + p\,\partial_0(a^3) = 0 , \tag{61}$$

which is well known[38] to be equivalent in Robertson-Walker spaces to the fundamental condition $T^{\mu\nu}{}_{;\nu} = 0$.

We shall numerically integrate the Einstein equations (48) and (49), with p and ρ given by Eqs. (59) and (60), and α and β obeying Eqs. (44), which are equivalent to the scalar wave equation. It is worth noting that with a strictly positive energy density ρ, as in Eq. (60), a reversal of isotropic collapse toward the cosmological singularity is only possible for the closed universe. Reversal of collapse requires a minimum in $a(t)$. At the minimum, Eq. (48) becomes $3a^{-2} = 8\pi G \rho$, yielding a finite positive minimum radius. However, for the open Robertson-Walker metric, the corresponding condition for a minimum would be $-3a^{-2} = 8\pi G\rho$, which cannot be satisfied for real a. For the flat case, the corresponding condition would require ρ to vanish at the time when the minimum is reached. That possibility cannot be ruled out, since particle creation would make ρ non-zero at other times.[39] However, a different state vector and method of renormalization would have to be used in that case.

By convention, we take $t=0$ as the lower limit on all integrals of the form $\int^t \omega_l dt'$ which have appeared. Furthermore, we let $\gamma^*\delta$ be real. Suppose that one has a solution of the coupled Einstein and scalar wave equations [Eqs. (44), (48), and (49), with p and ρ given by Eqs. (59) and (60)] given for $t>0$ by $a_+(t)$, $\alpha_+(t)$, and $\beta_+(t)$. It then follows from the reality of $\gamma^*\delta$ and the lower limit of $t=0$ on $\int^t \omega_l dt'$ that a solution for $t<0$ is given by $a_-(t)$, $\alpha_-(t)$, and $\beta_-(t)$, where

$$a_-(t) = a_+(-t)$$

and

$$\alpha_-(t) = \alpha_+^*(-t), \quad \beta_-(t) = \beta_+^*(-t) .$$

To join these into one continuous solution for all t, with continuous first derivative of a at $t=0$, we must require that

$$\partial_0 a(0) = 0 \tag{62}$$

and that

$$\alpha(0) \text{ and } \beta(0) \text{ are real} . \tag{63}$$

[It then follows from Eqs. (44) and (45) that α and β will have continuous first derivatives at $t=0$.]

Therefore, if we can show that there exists a solution for $t \geq 0$ such that $\partial_0 a$ vanishes at $t=0$ and α and β are real at $t=0$, then the solution for all t possessing the proper continuity properties has the symmetries

$$a(-t) = a(t), \tag{64}$$

$$\alpha(-t) = \alpha^*(t), \tag{65a}$$

and

$$\beta(-t) = \beta^*(t). \tag{65b}$$

In particular, if $t=0$ corresponds to a minimum in $a(t)$, then the radius function is time-symmetric about the minimum. We now outline the calculations which demonstrate the existence of a solution with a minimum at $t=0$.

VI. NUMERICAL INTEGRATION

In this section we exhibit numerical solutions which behave as just described in the neighborhood of a minimum of $a(t)$ (of elementary-particle dimensions) and which elsewhere are macroscopically indistinguishable from classical Friedmann solutions (of cosmological dimensions).

The equations which are to be solved are Eq. (49) for $a(t)$ [with p given by Eq. (59) with Eqs.

(39)] and the equations

$$\partial_0 \alpha = \tfrac{3}{2} a^{-1} \partial_0 a e^{-2imt} \beta, \qquad (66a)$$

$$\partial_0 \beta = \tfrac{3}{2} a^{-1} \partial_0 a e^{2imt} \alpha \qquad (66b)$$

for $\alpha(t)$ and $\beta(t)$. The last pair of equations comes from Eqs. (44) and (45) with $l=0$. We choose the initial conditions

$$\alpha(0)=1, \quad \beta(0)=0 , \qquad (67)$$

which satisfy Eqs. (63) and (41). We thereby remove the arbitrariness in the operators $A_{l\alpha}$. The initial conditions on $a(t)$ are $\partial_0 a(0)=0$ [Eq. (62)] and

$$a(0)=(8/3\pi)GmN(|\delta|^2+\tfrac{1}{2}) . \qquad (68)$$

The latter follows from Eqs. (48) and (60). The parameters which remain to be specified are m, N, δ, and the phase of γ [$|\gamma|$ being determined by Eq. (52)].

The value of m chosen for the computations is

$$m = 10^{-20} G^{-1/2} \approx 10^{13}\ \text{cm}^{-1} \approx m_{\text{pion}} . \qquad (69)$$

The results will be presented in terms of m^{-1} as unit of length and time.

In all the computations described here, the value $\delta = 0.824$ was used.[40] In this case Eq. (68) becomes

$$a(0)=GmN = 10^{-40}\, m^{-1} N . \qquad (70)$$

A larger δ would imply a larger initial negative pressure, and hence a more rapid expansion away from a given minimum radius.

We took $\gamma = (\delta^2+1)^{1/2}$. Then, since $\gamma^* \delta$ is real, the time symmetry described by Eqs. (64) and (65) is guaranteed; moreover, the initial value of the pressure is negative,

$$p(0) = -[\pi^2 a(0)^3]^{-1} m N \gamma \delta , \qquad (71)$$

to provide a positive second derivative for a at $t=0$.

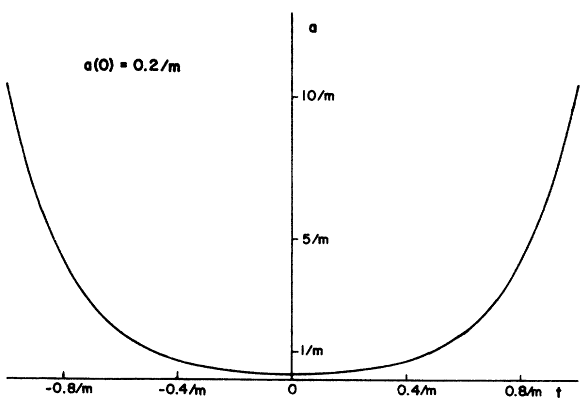

FIG. 1. Solution with $a(0) = 0.2m^{-1}$: time-symmetric expansion from the minimum radius.

The various solutions computed are then specified by the positive parameter N – or, alternatively, $a(0)$.

The differential equations were solved numerically on an IBM 360/91 computer by means of a sixth-order Adams-Bashforth-Moulton predictor-corrector algorithm.[41] Equations (48) and (41) were verified as the computation progressed, to maintain a check on the accuracy of the computation up to five significant figures.

The most significant result was found for the initial value $a(0)=0.2m^{-1}$ (i.e., $N=2\times10^{39} \gg 1$).[42] The corresponding initial density is $\rho(0) \approx 10^{40} m^4 \approx 10^{54}$ g cm^{-3}, and $c^{-2}p(0)$ is negative and of the same order of magnitude. The time-symmetric passage of $a(t)$ through this minimum value is shown in Fig. 1. The initial negative pressure impulse which sets off the expansion lasts for only a time of the order of m^{-1}, but beyond that time $\partial_0 a$ has become so steep that the expansion continues and is not reversed by gravitational attraction until t and $a(t)$ have attained cosmological magnitudes. In Fig. 2 a portion of the solution for $a(t)$ is plotted on logarithmic a and t scales, with a classical Friedmann solution superimposed. Beyond $t \approx 15m^{-1}$ the curves become practically indistinguishable. A similar description applies, of course, to the portion of the curve to the negative side of $t=0$, where the universe is contracting.

The conclusion is: If the closed universe and the matter in it find themselves in the state considered here during the contraction ($t<0$), then the geometry of the universe will follow very closely the standard Friedmann behavior until it has contracted to a radius comparable to the Compton wavelength of the scalar particles, whereupon, instead of collapsing to a singularity, it will "bounce" into a new expanding Friedmann stage, which is the time reflection of the contraction stage.

It was impractical to carry the numerical integration beyond the point $t=t_1=2.00\times10^3 m^{-1}$, at which $a(t_1)=3.94\times10^{16} m^{-1}$, $\beta(t_1)=(0.75+2.95i)\times10^{21}$, and $(Ga^3\rho)_{t_1}=8.23\times10^{41} m^{-1}$. One can extrapolate, however, to the Friedmann-like maximum of the solution. When $a(t)$ attains its maximum, one has from Eq. (48)

$$a_{\max} = (8\pi/3) G a^3 \rho . \qquad (72)$$

One expects that the WKB approximation (i.e., α and β constant) will be good throughout the entire period when $a(t)$ is large and slowly varying. It follows from Eq. (60) that $a^3\rho$ will also be constant in this approximation. Indeed, inspection of the numerical solution reveals that $\alpha(t)$, $\beta(t)$, and $a^3\rho$ have been practically constant (oscillating

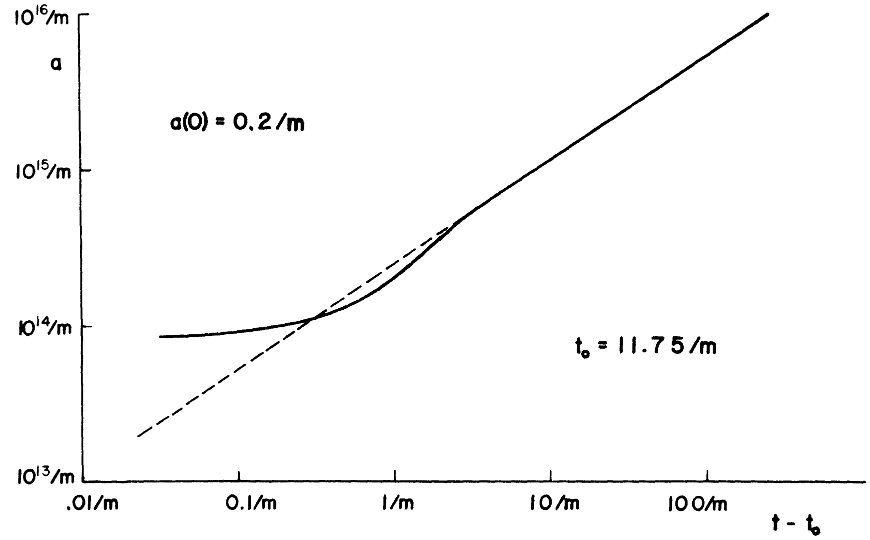

FIG. 2. Solution with $a(0) = 0.2m^{-1}$ (solid curve): approach to a Friedmann solution (dashed curve). The horizontal and vertical scales are logarithmic, and the time origin has been shifted to the initial singularity of the Friedmann curve, so that the latter becomes a straight line of slope $\frac{2}{3}$. (The deviation of the Friedmann solution from the $a \propto t^{2/3}$ law due to the three-space curvature of the closed universe is negligible in the range of t plotted.)

around constant mean values) for a considerable time before t_1. One can therefore use in Eq. (72) the value $G(a^3\rho)_{t_1}$ given above, obtaining $a_{max} \approx 10^{43} m^{-1} \approx 10^{30}$ cm. This is a reasonable value for the size of the actual universe. Smaller initial values $a(0)$ would yield larger values of a_{max}.

It is interesting to note that, despite the close resemblance of the solution to a dust-filled Friedmann universe during the period when the radius is large, the state of the matter within the universe is not an eigenstate of the number operators associated with the adiabatic approximation (see Sec. III). Indeed, in Appendix C we show that eigenstates of the adiabatic number operators must lead to a collapse, rather than a time-symmetric minimum.

Another example, for which the entire solution was computed, is the case $a(0) = 0.6m^{-1}$. The solution for $a(t)$ is plotted on logarithmic scales in Fig. 3 with a comparison Friedmann curve. The Friedmann was fitted on the basis of six points from the computed solution within a time interval of length $4m^{-1}$ about the maximum at $t = 204m^{-1}$, but the two curves agree within one or two percent as early as $t = 5m^{-1}$. The computed solution makes very small oscillations (not visible in the figure) of period πm^{-1} about the Friedmann solution, under the influence of the alternating sign of the pressure.

The maximum radius in this case is $129m^{-1}$, or about 10^{-11} cm. Thus an increase of $a(0)$ by a factor of 3 from the previous case ($0.2m^{-1}$) has resulted in a decrease of a_{max} by a factor of 10^{41}.

Larger values of $a(0)$ yield still smaller values of a_{max}. If $a(0) \gtrsim m^{-1}$, then a_{max} is of the same order of magnitude as $a(0)$, and hence there is a qualitative change in the solutions: instead of a modified Friedmann expansion for $t > 0$, one has small oscillations around a Friedmann *collapse* from a *maximum* of magnitude approximately equal to $a(0)$; the case $a(0) = 5m^{-1}$ is shown in Fig. 4.[43]

Returning to the case $a(0) = 0.6m^{-1}$, we discuss what happens to the solution after it passes the maximum in Fig. 3. When the solution contracts again, it apparently does not for a second time pass through a positive minimum and reexpand. As a matter of fact, the computed solution collapses to a singularity (at $t \approx 400m^{-1}$) even faster than the corresponding Friedmann universe, because as the change of a becomes rapid, β increases (particle production), and hence the energy density and, in turn, $\partial_0 a$ rise above their classical Friedmann values for a given radius [see Eqs. (60) and (48)]. The avoidance of the singularity depends upon the phase relationship between the contracting Friedmann-like solution and the oscillating pressure term. If these two things are out of phase, then collapse will occur. Our solutions are set up so that negative pressure will reverse the collapse at $t = 0$; there is no visible reason in this case to expect that the same thing will happen at the end of the next contraction. It is probable, however, that for certain choices of the parameters (N and δ) the second contraction will be timed relative to the pressure oscillations so that the final collapse will be avoided again. Note also

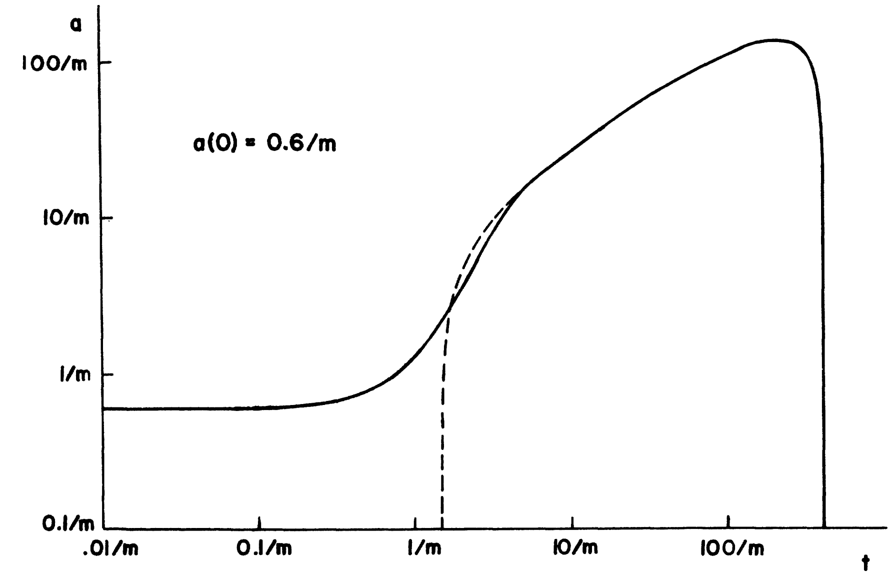

FIG. 3. Solution with $a(0) = 0.6m^{-1}$ (solid curve): approach to a Friedmann solution (dashed curve). The scales are logarithmic as in Fig. 2, but the time origin has not been moved. The approach to the minimum can be seen at the left. (The steepness of the contraction at the right is, of course, an artifact of the logarithmic plot.)

that, because of the accumulation of numerical error, one cannot conclude from a numerical solution that the exact solution for a given N and δ actually collapses to a singularity at the end of the second contraction.[44]

We believe that by a more general choice of state one can improve the probability that collapse is avoided at the end of the second contraction stage. Whether or not the contraction will be reversed depends on the sign of $\rho + 3p$ during the time when $a(t)$ is near the critical value (m^{-1}, in the case investigated) characteristic of the oscillation period of p. If $|p|$ in its fluctuations attains values which are large compared to ρ, then the positive average value of $\rho + 3p$ becomes relatively insignificant and each sign of $\rho + 3p$ is equally likely. Thus the probability of stopping and reversing the collapse, for a random initial phase relationship between $p(t)$ and the Friedmann contraction, should approach 50%. However, the values of $\langle A_{00}^\dagger A_{00}\rangle$ and $\langle A_{00} A_{00}\rangle$ which can be obtained by varying N and δ in Eqs. (55) and (56) are not the entire range allowed by the theorem in Appendix A, especially if N is required to be large. The very large $|p|/\rho$ ratios discussed in connection with Eq. (36) occur only for more general states, to which our treatment of the infinities may not apply. For this reason we have not investigated the matter further.

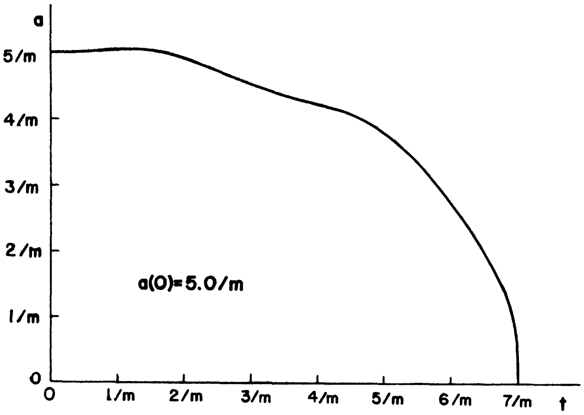

FIG. 4. Solution with $a(0) = 5m^{-1}$.

It is possible that for certain states (with definite phase relations) the collapse will be avoided on every contraction, so that the solution is literally free of singularities, in marked contrast to the situation in classical general relativity described by the singularity theorems. We have not been able to prove or disprove the existence of such a solution within our approximation scheme. Moreover, the effect of particles created in modes other than the lowest must also be taken into account before a definite conclusion can be drawn.

Finally, since the time interval during which nonclassical behavior occurs is of the order m^{-1}, the case $m = 0$ is of importance. Our treatment of the mode $l = 0$ does not apply in that case. However, if one sets $m = 0$ in Eq. (36), one obtains

$$-\frac{p}{\rho} = \tfrac{2}{3}(1 + \langle A^\dagger A\rangle^{-1})^{1/2} - \tfrac{1}{3},$$

which can be made arbitrarily large as before.

The ratio is independent of l. Furthermore, ω_l becomes arbitrarily small for sufficiently large a. Since ω_l governs the frequency at which p changes sign, it follows that p can remain negative for a very long period of time. Therefore, coherence effects of a massless field, such as the photon, graviton, or possibly neutrino field, may have significant influence on the behavior of $a(t)$ during the era when such fields are dominant. This problem is under investigation.

VII. SUMMARY

We have carried out a concrete calculation in the semiclassical theory of the mutual interaction of a massive scalar matter field with the geometry of a closed homogeneous and isotropic universe. (Neither the quantum nature of the gravitational field itself nor nongravitational interactions among the particles of matter have been considered.) We have circumvented the problem of defining the "finite part" of the energy-momentum tensor by considering only states for which we believe the precise form of this renormalization to be unimportant. This and some other aspects of the calculation, such as the exploitation of the anticipated symmetry of the quantum state by averaging the energy-momentum tensor over space, should be of methodological interest in other problems.

Our main goal was to study the impact of the quantum theory on the classical singularity theorems. The energy-momentum tensor was calculated for arbitrary quantum states consistent with the symmetry of the model. On this basis we showed first that the premises of the singularity theorems (energy conditions) can be violated in this theory. Then we exhibited solutions of the coupled equations for the Robertson-Walker radius function and the scalar field which violate the conclusions of the singularity theorems, in spirit, if not in letter: A Friedmann-like collapse is arrested at elementary-particle dimensions and changed to an expansion. Although it appears that, typically, the solutions do eventually collapse after completing another Friedmann cycle, the existence of solutions which are completely free of singularities is still an open question. The escape from singularity at the end of a contraction stage in our solutions is a quantum coherence effect, which depends on certain phase relationships in the mathematical specification of the quantum state. We believe that classes of states exist for which almost half of the possible phase relationships are favorable.

Our model does not provide evidence that quantum effects will always avert gravitational collapse. It does, however, support the view that such an effect can occur under certain circumstances. It would be of interest to apply similar considerations to the gravitational collapse of a localized massive gas cloud of fermions or bosons.[45]

ACKNOWLEDGMENTS

The authors thank Professor K. Kuchař and Professor J. A. Wheeler for stimulating discussions, Princeton University for access to the computer and the program ODEPAC, and the members of the Physics Department of Princeton University for their hospitality. Appreciation is also extended to the Aspen Center for Physics, where part of this work was done. Finally, we thank Professor B. S. DeWitt and Dr. E. P. Liang for helpful comments on the manuscript.

APPENDIX A: THE RANGE OF THE QUADRATIC EXPECTATION VALUES OF CANONICAL OPERATORS

Let A and $A^\dagger$ be the annihilation operator and creation operator for one mode of a quantum field. The representation of these algebraic objects which is encountered in a quantum field theory is highly reducible. Conditions are nearly always assumed which ensure that this representation is a direct sum of copies of the irreducible representation associated with the one-dimensional harmonic oscillator.[46] Such conditions include[47]: (1) the operators $Q=(A+A^\dagger)/\sqrt{2}$ and $P=-i(A-A^\dagger)/\sqrt{2}$ generate unitary groups satisfying the "Weyl relations"; (2) $A^\dagger A$ (or Q^2+P^2) is essentially self-adjoint on an invariant domain of definition of the operators.

Theorem. Suppose that a Hilbert-space operator A and its Hermitian conjugate $A^\dagger$ satisfy $[A,A^\dagger]=1$, and that the Weyl relations (or equivalent technical conditions) hold. Then the expectation values of (the closures of) $A^\dagger A$ and AA with respect to normalized vectors in the Hilbert space take on all pairs of values consistent with

$$0 \leq \langle A^\dagger A\rangle \leq \infty \ , \tag{A1}$$

$$0 \leq |\langle AA\rangle|^2 \leq \langle A^\dagger A\rangle(\langle A^\dagger A\rangle + 1) \ , \tag{A2}$$

and only these.

The rest of the appendix is devoted to the proof.

First we show that Eq. (A2) must hold. The technical assumption allows us to use the well-known structure of the harmonic-oscillator representation. The spectrum of $A^\dagger A$ is precisely the nonnegative integers. Consequently, every normalized vector is of the form

$$\psi=\sum_{n=0}^{\infty} a_n\psi_n \,, \tag{A3}$$

with

$$\sum_{n=0}^{\infty} |a_n|^2=1 \,, \tag{A4}$$

where $A^\dagger A\psi_n=n\psi_n$ and $\langle\psi_n|\psi_n\rangle=1$. Moreover, $AA\psi_{n+2}$ is an eigenvector of $A^\dagger A$ with eigenvalue n and norm $[(n+2)(n+1)]^{1/2}$. It follows that

$$\langle A^\dagger A\rangle\equiv\langle\psi|A^\dagger A|\psi\rangle=\sum_{n=0}^{\infty} n|a_n|^2 \tag{A5}$$

and

$$|\langle AA\rangle|\equiv|\langle\psi|AA|\psi\rangle|$$

$$\leq\sum_{n=0}^{\infty}[(n+2)(n+1)]^{1/2}|a_n||a_{n+2}|\,.$$

Using the Schwarz inequality, one finds that

$$|\langle AA\rangle|^2\leq\sum_{n=2}^{\infty} n|a_n|^2\sum_{n=0}^{\infty}(n+1)|a_n|^2$$

$$\leq\langle A^\dagger A\rangle(\langle A^\dagger A\rangle+1), \tag{A6}$$

which is the content of Eq. (A2).

Next we wish to exhibit vectors yielding all the expectation values in the range specified by Eqs. (A1)–(A2). It will suffice to deal with vectors whose normalized n-particle parts are related as in an irreducible representation. That is, we write Eq. (A3) as

$$\psi=\sum_{n=0}^{\infty} a_n|n\rangle\,, \tag{A7}$$

where $A|n\rangle=\sqrt{n}\,|n-1\rangle$. Then

$$\langle AA\rangle=\sum_{n=0}^{\infty}[(n+2)(n+1)]^{1/2}a_n^* a_{n+2}\,. \tag{A8}$$

The main task is to show that the second inequality in Eq. (A2) can become an equality for all values of $\langle A^\dagger A\rangle$. When Eq. (A8) holds, the inequalities in Eq. (A6) become equalities if and only if (1) $a_1=0$, and (2) the vectors to which the Schwarz inequality has been applied are proportional:

$$a_{n+2}=\lambda\left(\frac{n+1}{n+2}\right)^{1/2}a_n \tag{A9}$$

for some complex number λ. It follows that $a_n=0$ for all odd n. Let $x=|\lambda|^2$ and

$$r_p=x^{-p}\,\frac{|a_{2p}|^2}{|a_0|^2}$$

$$=\frac{2p-1}{2p}\times\frac{2p-3}{2p-2}\cdots\frac{1}{2}$$

$$=\frac{1}{p\,!}\,\frac{d^p}{dy^p}(1-y)^{-1/2}\Big|_{y=0}\,. \tag{A10}$$

Equations (A4) and (A5) become

$$1=|a_0|^2\sum_{p=0}^{\infty}x^p r_p$$

$$\equiv|a_0|^2F(x) \tag{A11}$$

and

$$\langle A^\dagger A\rangle=2|a_0|^2\sum_{p=0}^{\infty}px^p r_p$$

$$=2|a_0|^2xF'(x)\,. \tag{A12}$$

For $0\leq x<1$ the series defining $F(x)$ converges:

$$F(x)=(1-x)^{-1/2}. \tag{A13}$$

Thus if $|\lambda|<1$ Eqs. (A9) and (A11) determine the coefficients of a normalized vector, which we denote by $\psi(\lambda)$. (For definiteness one can take $a_0>0$.)

Since a power series can be differentiated inside its circle of convergence, Eq. (A12) shows that $\langle A^\dagger A\rangle$ is also finite when $x<1$. In fact, dividing Eq. (A12) by Eq. (A11), one obtains

$$\langle A^\dagger A\rangle=2x\frac{F'(x)}{F(x)}=\frac{x}{1-x}\,. \tag{A14}$$

As x varies in the range $0\leq x<1$, $\langle A^\dagger A\rangle$ varies through all numbers in the range $0\leq\langle A^\dagger A\rangle<\infty$.

Having shown that the maximum value

$$|\langle AA\rangle|^2=\langle A^\dagger A\rangle(\langle A^\dagger A\rangle+1) \tag{A15}$$

can actually be attained for all finite values of $\langle A^\dagger A\rangle$, we now verify that all the smaller values of $|\langle AA\rangle|$ are also allowed. Consider vectors like the $\psi(\lambda)$ just described, except that a_n is replaced by ϵa_n for $n=2,6,10,\ldots$ (i.e., for odd p). Then the normalized expectation values of AA and $A^\dagger A$ become

$$\langle AA\rangle_\epsilon=\frac{\epsilon\langle AA\rangle_1}{\langle 1\rangle_1^e+\epsilon^2\langle 1\rangle_1^o}\,,$$

$$\langle A^\dagger A\rangle_\epsilon=\frac{\langle A^\dagger A\rangle_1^e+\epsilon^2\langle A^\dagger A\rangle_1^o}{\langle 1\rangle_1^e+\epsilon^2\langle 1\rangle_1^o}\,,$$

where $\langle A^\dagger A\rangle_1^e$ and $\langle A^\dagger A\rangle_1^o$ are, respectively, the contributions from the terms of even p and odd p in the series (A5), and similarly for $\langle 1\rangle_1^e$ and $\langle 1\rangle_1^o$ with respect to Eq. (A4). As ϵ varies from 1 to 0, $\langle AA\rangle_\epsilon$ goes continuously to 0 and $\langle A^\dagger A\rangle$ goes continuously to a positive limit, remaining always greater than $\langle A^\dagger A\rangle_1^e$. It is easy to see that $\langle A^\dagger A\rangle_1^e$

$> x\langle A^\dagger A\rangle_1^0$, and, therefore, as $\lambda \to 1$ and $\langle A^\dagger A\rangle_1 \to \infty$ one has $\langle A^\dagger A\rangle_1^\epsilon \to \infty$. Since $\langle AA\rangle_\epsilon$ and $\langle A^\dagger A\rangle_\epsilon$ are continuous functions of both ϵ and λ, one sees that the curve in the $\langle A^\dagger A\rangle - |\langle AA\rangle|$ plane (see Fig. 5) defined by Eq. (A15) is dragged down in a continuous manner toward the entire $\langle A^\dagger A\rangle$-axis as ϵ varies from 1 to 0. Every point of the region allowed by Eq. (A2) (except at infinity) is reached by some choice of ϵ and λ, which is what was to be proved.

It remains to treat the infinite cases. Consider a vector like $\psi(\lambda)$ except that $a_n \neq 0$ for $n = 1, 5, 9, \ldots$ and

$$\sum_{n \text{ odd}} |a_n|^2 = 1 ,$$

$$\sum_{n \text{ odd}} n|a_n|^2 = \infty .$$

Then $\langle A^\dagger A\rangle = \infty$ but $\langle AA\rangle$, after normalization, equals one-half of its value for $\psi(\lambda)$, which is an arbitrary nonnegative number. Also, there obviously are normalized vectors for which both $\langle A^\dagger A\rangle$ and $\langle AA\rangle$ diverge.

Finally, from Eqs. (A8) and (A9) it is clear that the phase of $\langle AA\rangle$ is just that of λ. Hence for each value of $\langle A^\dagger A\rangle$ and $|\langle AA\rangle|$ all phases are allowed.

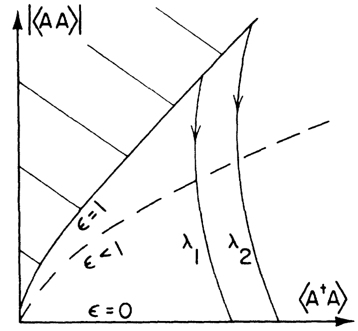

FIG. 5. Range of the quadratic expectation values. The curve $\epsilon = 1$, which is defined by Eq. (A15), is parametrized by λ. As $\epsilon \to 0$ the point corresponding to each λ moves down to the $\langle A^\dagger A\rangle$ axis, and thus the image of the curve sweeps through the entire allowed region. (The shapes indicated for the λ_1, λ_2, and $\epsilon < 1$ curves are purely schematic.)

APPENDIX B: EQUATIONS FOR QUADRATIC FIELD QUANTITIES

From Eqs. (46) and (47) one sees that α and β (index l suppressed) enter the general expressions for the density and pressure through the quadratic combinations

$$s = |\hat\beta|^2 = |\beta|^2 , \qquad \text{(B1)}$$

$$z = \hat\alpha\hat\beta = \alpha^*\beta^* , \qquad \text{(B2)}$$

$$w = \hat\alpha^2 + \hat\beta^2 , \qquad \text{(B3)}$$

and the real part of

$$v = \hat\alpha^*\hat\beta . \qquad \text{(B4)}$$

To obtain a closed set of equations we introduce also

$$q = \hat\alpha^2 - \hat\beta^2 . \qquad \text{(B5)}$$

It follows directly from the equations of motion for α and β [Eqs. (44)] and their relation to the careted variables [Eqs. (39)] that these five quantities obey the set of first-order equations

$$\partial_0 s = 2y^{-1}\partial_0 y \,\mathrm{Re}\, v , \qquad \text{(B6)}$$

$$\partial_0 v = y^{-1}\partial_0 y(2s+1) + 2i\omega v , \qquad \text{(B7)}$$

$$\partial_0 z = y^{-1}\partial_0 y w , \qquad \text{(B8)}$$

$$\partial_0 w = 4y^{-1}\partial_0 y z - 2i\omega q , \qquad \text{(B9)}$$

$$\partial_0 q = -2i\omega w , \qquad \text{(B10)}$$

where y is defined in Eq. (37). Equations equivalent to Eqs. (B6) and (B7) were given in Ref. 25.

Equations (B6)–(B10) may have an advantage over Eqs. (44) in numerical integrations, in general, since they do not require evaluation of $\int^t \omega_l(t')dt'$. This quantity reduced to mt, however, in the computations reported in this paper. Consequently, the only use made here of the above equations is the verification that the pressure and energy density used in the computations satisfy the continuity equation (61).

APPENDIX C: INEVITABILITY OF COLLAPSE WHEN PARTICLE NUMBERS ARE DEFINITE

Here we show that if the quantum state is an eigenstate of the number operators $N_{l\alpha} = A^\dagger_{l\alpha}A_{l\alpha}$ in a range of time when the adiabatic approximation of Sec. III is applicable (e.g., during the "macroscopic" period of the history of a Friedmann universe), and if a finite number of modes can be assumed to dominate, then a minimum of the radius like those exhibited in Sec. VI cannot occur in our model. It follows that for the solutions in Sec. VI the state is a coherent superposition of states with different numbers of particles in the various modes (during the epochs when the expansion is sufficiently slow for the particle notion to make clear sense), to the extent that the neglect of the higher modes in the energy density is justified.

We introduce coefficients α_l and β_l as in Eqs. (37)–(41), but require that $\alpha(t_0) = 1$, $\beta(t_0) = 0$ for some t_0 in the adiabatic range, rather than at a minimum of $a(t)$ as in Secs. V–VI. Thus ψ_l is the solution which is approximated near t_0 by Eq. (27).

(Throughout the adiabatic period we have $\beta \approx 0$.) For a state of definite particle numbers $N_{l\alpha}$ we have $\langle A_{l,-\alpha} A_{l\alpha}\rangle = 0$, $\langle A^{\dagger}_{l\alpha} A_{l\alpha}\rangle = N_{l\alpha}$. Thus from Eq. (47) we have

$$\rho = (2\pi^2 a^3)^{-1} \sum_{l\alpha} \omega_l (2N_{l\alpha} + 1)(|\beta_l|^2 + \tfrac{1}{2}) . \qquad \text{(C1)}$$

Let us assume as in Sec. V that only a few terms contribute significantly to this sum, and that for them $N_{l\alpha} \gg 1$. Then

$$\rho = (\pi^2 a^3)^{-1} \sum_{l\alpha} N_{l\alpha} \omega_l (|\beta_l|^2 + \tfrac{1}{2}) , \qquad \text{(C2)}$$

where the sum is finite. There is a similar expression for the pressure, which is positive at $t = t_0$.

The behavior of $a(t)$ can now be studied qualitatively by inspection of the initial-value Einstein equation (48). Suppose that $a(t)$ has a minimum at some point t_1. (This is not t_0, since the pressure is positive at t_0.) Since $a(t_1) < a(t_0)$, $\omega_l(t_1) = [l(l+2)/a(t_1)^2 + m^2]^{1/2}$ is greater than $\omega_l(t_0)$. Since $\partial_0 a(t_1) = 0$, $(\partial_0 a)^2 + 1$ is not greater at t_1 than at t_0. Finally, the value of $|\beta|^2 + \frac{1}{2}$ at t_1 is greater than or equal to that at t_0 (viz., $\frac{1}{2}$). Thus, from Eq. (48), one has

$$\begin{aligned} a(t_1) &= [(\partial_0 a)^2 + 1]^{-1}_{t_1} (8/3\pi) G \\ &\quad \times \sum_{l\alpha} N_{l\alpha} \omega_l(t_1)(|\beta_l(t_1)|^2 + \tfrac{1}{2}) \\ &\geq (\text{same expression, } t_1 \to t_0) = a(t_0) . \end{aligned}$$

This is a contradiction.

In physical terms: When the particle numbers are definite initially, the effect of particle creation can only be to *increase* the energy, and hence to accelerate the collapse relative to its classical course. What has been shown in the body of the paper is that for a coherent superposition of particle numbers, particle creation contributions can cause cancellations in the energy, and hence a slowing of the contraction and an escape from the singularity.

---

*Research supported by National Science Foundation Grant GP-19432 and the University of Wisconsin-Milwaukee Graduate School.

†Part of this work was done while the authors were on leave to the Department of Physics, Princeton University, Princeton, New Jersey 08540.

[1]See S. W. Hawking and R. Penrose, Proc. Roy. Soc. (London) A314, 529 (1970), and references given there.

[2]In addition to Refs. 3–5, 8–9, and 12 below, note the remarks in Ref. 1 (footnote, p. 532).

[3]J. A. Wheeler, in *Relativity, Groups and Topology*, edited by B. S. DeWitt and C. DeWitt (Gordon and Breach, New York, 1964), pp. 328, 504; B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse* (University of Chicago Press, Chicago, 1965), pp. 141–143.

[4]C. W. Misner, Phys. Rev. 186, 1319 (1969); 186, 1328 (1969); M. Ryan, *Hamiltonian Cosmology* (Springer, Berlin, 1972); B. S. DeWitt, Phys. Rev. 160, 1113 (1967); E. P. T. Liang, Phys. Rev. D 5, 2458 (1972).

[5]H. Nariai, Progr. Theoret. Phys. (Kyoto) 46, 433 (1971); H. Nariai and K. Tomita, *ibid.* 46, 776 (1971).

[6]R. Utiyama and B. S. DeWitt, J. Math. Phys. 3, 608 (1962).

[7]A. D. Sakharov, Dokl. Akad. Nauk SSSR 177, 70 (1967) [Sov. Phys. —Dokl. 12, 1040 (1968)].

[8]V. Ts. Gurovich, Dokl. Akad. Nauk SSSR 195, 1300 (1970) [Sov. Phys. Dokl. 15, 1105 (1971)]; T. V. Ruzmaikina and A. A. Ruzmaikin, Zh. Eksp. Teor. Fiz. 57, 680 (1969) [Sov. Phys. JETP 30, 372 (1970)].

[9]J. N. Bahcall and S. Frautschi, Astrophys. J. 170, L81 (1971).

[10]L. Parker, (a) Ph.D. thesis, Harvard University, 1966 (unpublished); (b) Phys. Rev. Letters 21, 562 (1968); (c) Phys. Rev. 183, 1057 (1969); (d) Phys. Rev. D 3, 346 (1971); (e) Phys. Rev. Letters 28, 705 (1972).

[11]Ya. B. Zel'dovich and L. P. Pitaevsky [Commun. Math. Phys. 23, 185 (1971)] showed that $\langle T_{\mu\nu}\rangle$ violates certain closely related conditions which would otherwise make gravitationally induced particle creation from vacuum impossible.

[12]J. A. Wheeler, in *Battelle Rencontres,* edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[13]R. Potier, Compt. Rend. 243, 939 (1956).

[14]C. Møller, in *Les Théories Relativistes de la Gravitation,* proceedings of the conference at Royaumont, 1959 (Centre National de la Recherche Scientifique, Paris, 1962), pp. 21–29.

[15]S. Bonazzola and F. Pacini, Phys. Rev. 148, 1269 (1966); R. Ruffini and S. Bonazzola, *ibid.* 187, 1767 (1969).

[16]We use units with $\hbar = c = 1$. The metric has signature −2. We use the conventions that $(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\xi_\lambda = \xi_\sigma R^{\sigma}_{\lambda\mu\nu}$, and $R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu}$, which determine the signs of the Riemann and Ricci tensors. Here ∇_ν refers to the covariant derivative, and ξ_λ is an arbitrary four-vector field. Greek indices run from 0 to 3, and Latin indices from 1 to 3. We generally use t for x^0.

[17]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1961), pp. 615–619.

[18]See Ref. 17, pp. 781-800.

[19]These are reviewed by M. O. Scully and M. Sargent III, Phys. Today 25, No. 3, 38 (1972).

[20]W. Heisenberg and H. Euler, Z. Physik 98, 714 (1936); V. Weisskopf, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. 14, No. 6 (1936); J. Schwinger, Phys. Rev. 82, 664 (1951).

[21]It can be "derived" from a quantum theory by

assuming that each of the observables to be treated classically [e.g., $g_{\mu\nu}$ in Eq. (1)] has such a small dispersion, in any quantum state Ψ to be considered, that it can be identified with its expectation value and treated as a c number. Equations like Eq. (1) are then obtained by taking the expectation value with respect to Ψ of the Heisenberg equations of motion; expectation-value symbols ($\langle\ \rangle$) remain explicitly only in terms which involve the other dynamical variables of the theory, which are still quantized (the matter fields, in the present case).

[22]This is the attitude of the authors of Refs. 13 and 14, and similar views have recently been expressed in the case of electrodynamics [see "Neoclassicism Challenges QED," Phys. Today 25, No. 10, 17 (1972)].

[23]If a fundamental quantum theory of gravity is presupposed, then conditions on the characteristic lengths of the system (such that the gravitational field itself behaves classically) are probably not sufficient for validity of the semiclassical approximation. One must also have a situation in which certain probabilistic quantities in the quantum theory of the *matter* can be treated like physical densities of energy, momentum, and stress. [Analogously, the Thomas-Fermi model (Ref. 17) works best for atoms with many electrons because there the Schrödinger probability density can be treated as a physical charge density to good approximation – not because a classical treatment of the electric field *per se* is any better there than in the case of few electrons.] Cf. remarks of Ruffini and Bonazzola (Ref. 15), p. 1776. In this respect the semiclassical approach seems to be as well motivated in the present work as in the electrodynamical context of Refs. 20.

[24]B. S. DeWitt, Phys. Rev. 162, 1239 (1969).

[25]Ya. B. Zel'dovich and A. A. Starobinsky, Zh. Eksp. Teor. Fiz. 61, 2161 (1971) [Sov. Phys. JETP 34, 1159 (1972)].

[26]See: Ref. 25; C. G. Callan, Jr., S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) 59, 42 (1970); N. A. Chernikov and E. A. Tagirov, Ann. Inst. Henri Poincaré 11, 207 (1969); L. Parker, Phys. Rev. D 7, 976 (1973); H. Ohanian, J. Math. Phys. (to be published).

[27]Canonical quantization in Riemannian space-time, as an abstract algebraic construction, is a generally covariant and consistent procedure, although that is not manifestly evident. The procedure has been used often. See, for example, Refs. 14, 6, and 10; R. Utiyama, Phys. Rev. 125, 1727 (1962); H. K. Urbantke, Nuovo Cimento 63B, 203 (1969) (proof of covariance). For a summary of the method and another proof of covariance see S. A. Fulling, Ph.D. dissertation, Princeton University, 1972 (unpublished), Chap. VII.

[28]E. Schrödinger, *Expanding Universes* (Cambridge Univ. Press, Cambridge, 1956), pp. 79–86; E. Lifshitz, J. Phys. (USSR) 10, 116 (1946); M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 (1966).

[29]The Fock representation of the creation-annihilation commutation relations (14) is the unique irreducible representation which contains a vector annihilated by all the $A_{l\alpha}$. The details of its construction can be found in any textbook on quantum field theory. By extension we call the corresponding representation of the fields defined through Eq. (13) *a Fock representation of the field algebra,* Eqs. (6). A Fock representation in the latter sense is not unique, since the $A_{l\alpha}$ can be defined in various ways for the same field algebra.

[30]I. E. Segal, Ann. Math. 48, 930 (1947); J. M. G. Fell, Trans. Am. Math. Soc. 94, 365 (1960); R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964); G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972), especially pp. 97–110. The possible application of the algebraic method to quantum field theory in curved space-times has been discussed elsewhere by one of the authors [S. A. Fulling, dissertation (Ref. 27), pp. 291–323; Phys. Rev. D (to be published)].

[31]See the discussions leading to Eq. (36) in Sec. III and to Eqs. (59) and (60) in Sec. V.

[32]L. Parker, Ref. 10(c).

[33]It does not necessarily follow from $|p|/\rho > 1$ that causality is violated because of a sound speed exceeding the speed of light. The conventional definitions and calculations of the speed of sound do not apply to a system in a pure quantum state, especially in the presence of fluctuations in p on a time scale of order m^{-1} The propagation of the n-point functions of the field described by Eq. (3) is, as a matter of fact, causal.

[34]We write α^* and β^* in Eqs. (39) to be consistent with the notation of Ref. 10(c). The condition in Eq. (40) is slightly different from the one used in Ref. 10(c); it corresponds instead to the condition used in Ref. 25. This definition is more convenient for our present purposes, because it simplifies the expressions for energy density and pressure. The quantities of primary physical interest are ρ and p; α and β are introduced here for computational convenience.

[35]N. N. Bogoliubov, J. Phys. (USSR) 11, 23 (1947); Nuovo Cimento 7, 794 (1958); J. G. Valatin, *ibid.* 7, 843 (1958).

[36]If we were studying a complex field, the state with $N = 0$ would have no net charge, since it would be a superposition of states consisting of various numbers of particle-antiparticle pairs. However, a state with unequal numbers of particles and antiparticles could equally well be used in calculations like ours.

[37]A "naive" vacuum subtraction which affects only the terms independent of both N and β_l is not sufficient, since $\sum l^2\omega_l\,|\beta_l|^2$ also diverges (see Ref. 25). Renormalization, being a covariant procedure, may possibly cut across the noncovariant specification of modes, and thus affect even the lowest mode. We expect, however, that the β_l-dependent part of the renormalization will affect primarily the terms corresponding to large l, and also that the renormalization of the lowest mode will affect primarily the term independent of N, as in the special-relativistic vacuum subtraction procedure.

[38]See R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Clarendon, Oxford, 1934), p. 220.

[39]A statement in Ref. 10(e), that the flat Robertson-Walker model considered there could not be entirely consistent, is not valid.

[40]We quote this and all numerical results only to three significant figures.

[41]ODEPAC, by I. T. Cundiff (Princeton University Computer Center, Princeton, N. J., 1969), revised 1972.

[42]It might be questioned whether this value of N is sufficiently large to make Eqs. (59) and (60) valid. If the contribution of higher modes is crudely estimated,

using a wave number cutoff as large as $(G\hbar c^{-3})^{-1/2}$, and dropping manifest vacuum terms (i.e., those independent of β as well as N), then one finds that the present value of N is just on the borderline of being sufficient. It seems clear, however, that a different choice of state (e.g., a different value of δ) would permit N to be increased without altering the nature of our results.

[43]When δ is not of the order of magnitude unity, the division between expansion and collapse behavior as a function of $a(0)$ might occur considerably away from the point $a(0) = m^{-1}$.

[44]The point is that, because of amplification of numerical error, it is very difficult to reproduce a solution of the type of Fig. 3 or Figs. 1 and 2 by starting with the computed values of a, $\partial_0 a$, α, and β at a point high on the curve and integrating backwards. When a becomes small, the new approximate solution will deviate significantly from the original one, and a "collapse" will usually (at least for the states we have studied) be observed in the numerical output, even though the exact solution does not collapse.

[45]After this paper was completed, E. P. Liang pointed out to us (private communication) the following. From Eqs. (19) and (20) one has

$$\rho + 3p = 2(\partial_0\phi)^2 - m^2\phi^2.$$

This quantity can become negative even when ϕ is regarded as a real function or c-number. It therefore seems likely that one can find "bouncing" solutions of the coupled Einstein and Klein-Gordon equations with a c-number scalar field. The question then arises as to how essential a role quantization plays in the effect we have studied. Furthermore, will the effect occur for fields of nonzero spin?

Several remarks can be made which help clarify the relevance of quantization to our results. First, it is important to note that expectation values of the formal expression above for $\rho + 3p$ diverge for the quantized field, so that renormalization is necessary. As is well known from the Minkowski-space theory, renormalization can alter properties such as positive definiteness of an operator. This is well illustrated for the above operator in the case when $m = 0$. Then

$$\rho + 3p = (\partial_0\phi)^2$$

is positive definite (ϕ is a real function corresponding to the Hermitian quantized field used in this paper). Consequently, if the renormalized operator $\rho + 3p$ can have negative expectation values in this case, one definitely has a "quantum" violation of the energy condition. As discussed in Sec. VI, Eq. (36) evidently applies in the massless case and shows that $\rho + 3p$ can indeed possess negative expectation values. Another way to see that is to note that in the adiabatic case, where normal ordering seems appropriate, the contribution of a single mode to the operator $\rho + 3p =: (\partial_0\phi)^2:$ is proportional to

$$:(Au + A^\dagger u^*)^2: = AAu^2 + A^\dagger A^\dagger u^{*2} + 2A^\dagger A,$$

where A is an annihilation operator, and u is a function satisfying $|u|^2 = 1$ [and $:(\,):$ denotes normal ordering]. If one takes the expectation value of this expression in the state $|\,\rangle = 2^{-1/2}(|0\rangle + |2\rangle)$, and chooses the phase (or time) so that $u = i$, then one obtains a negative result. (Here $|0\rangle$ and $|2\rangle$ are normalized states such that $\langle 0|AA|2\rangle = \sqrt{2}$ and $\langle 2|A^\dagger A|2\rangle = 2$.) In order to isolate that effect from the effect of the $-m^2\phi^2$ term in $\rho + 3p$, it would be of interest to carry out calculations analogous to those in this paper for the scalar field with $m = 0$, as well as for fields of nonzero spin.

Secondly, if the negative pressure "bounce" is to have more than mathematical interest, a mechanism must be found which tends to produce a state of the type we used whenever the cosmological singularity is approached. Such a mechanism, as discussed in Sec. V, could be a phase transition analogous to the transition to the superfluid state in He^4, or to the BCS ground state in a superconductor. (Interactions between the elementary particles would presumably be responsible for such a transition.) Thus, one would like to know if the addition of an interaction term can bring about a phase transition to a state vector of the desired type as the singularity is approached. Clearly, the quantized theory must be used in formulating and investigating that question.

[46]Ref. 17, Chap. XII.

[47]See, e.g., I. E. Segal, in *Cargèse Lectures in Theoretical Physics: Application of Mathematics to Problems in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967), Secs. I and III; or G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972), pp. 224–247. The equivalence of the two conditions follows from a theorem of E. Nelson, Ann. Math. 70, 572 (1959).