Arrow of time in cosmology

S. W. Hawking
University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England
(Received 29 April 1985)

The usual proof of the *CPT* theorem does not apply to theories which include the gravitational field. Nevertheless, it is shown that *CPT* invariance still holds in these cases provided that, as has recently been proposed, the quantum state of the Universe is defined by a path integral over metrics that are compact without boundary. The observed asymmetry or arrow of time defined by the direction of time in which entropy increases is shown to be related to the cosmological arrow of time defined by the direction of time in which the Universe is expanding. It arises because in the proposed quantum state the Universe would have been smooth and homogeneous when it was small but irregular and inhomogeneous when it was large. The thermodynamic arrow would reverse during a contracting phase of the Universe or inside black holes. Possible observational tests of this prediction are discussed.

I. INTRODUCTION

Physics is time symmetric. More accurately, it can be shown[1] that any quantum field theory that has (a) Lorentz invariance, (b) positive energy, and (c) local causality, i.e., $\phi(x)$ and $\phi(y)$ commute (or anticommute) if x and y are spacelike separated, is invariant under *CPT* where C means interchange particles with antiparticles, P means replace left hand by right hand, and T means reverse the direction of motion of all particles. In most situations, the effect of any C or P noninvariance can be neglected, so that the interactions ought to be invariant under T alone.

In fact, if one takes the gravitational field into account, the Universe that we live in does not satisfy any of the three conditions listed above. The Universe is not Lorentz invariant because spacetime is not flat, or even asymptotically flat. The energy density is not positive definite because gravitational potential energy is negative. In a certain sense the total energy of the Universe is zero because the positive energy of the matter is exactly compensated by the negative gravitational potential energy. Finally, the concept of local causality ceases to be well defined if the spacetime metric itself is quantized because one cannot tell if x and y are spacelike separated. Nevertheless, I shall show in Sec. III of this paper that the universe is invariant under *CPT* if, as has been recently proposed,[2–4] it is in the quantum state defined by a path integral over compact four-metrics without boundary. This is a nontrivial result because an arbitrary quantum state for the Universe is not, in general, invariant under *CPT*.

The Universe that we live in certainly does not appear time symmetric, as anyone who has watched a movie being shown backward can testify: one sees events that are never witnessed in ordinary life, like pieces of a cup gathering themselves together off the floor and jumping back onto a table. One can distinguish a number of different "arrows of time" that express the time asymmetry of the Universe. (1) The thermodynamic arrow: the direction of time in which entropy increases. (2) The electrodynamic arrow: the fact that one uses retarded solutions of the field equations rather than advanced ones. (3) The psychological arrow: the fact that we remember events in the past but not in the future. (4) The cosmological arrow: the direction in time in which the universe is expanding.

I shall take the point of view that the first arrow implies the second and third. In the case of the psychological arrow this follows because human beings (or computers, which are easier to talk about) are governed by the thermodynamic arrow, like everything else in the Universe. In the case of electrodynamics, one can express the vector potential $A_\mu(x)$ as a sum of a contribution from sources in the past of x plus a surface integral at past infinity. One can also express $A_\mu(x)$ as a sum of a contribution from sources in the future of x plus a surface integral at future infinity. The boundary conditions that give rise to the thermodynamic arrow imply that there is no incoming radiation in the past. Thus the surface integral in the past is zero and the electromagnetic field can be expressed as an integral over sources in the past. On the other hand, the boundary conditions that give rise to the thermodynamic arrow do not prevent the possibility of outgoing radiation in the future. This means that the surface integral in the future is strongly correlated with the contribution from sources in the future. It therefore cannot be neglected.

The accepted explanation for the thermodynamic arrow of time is that for some reason the Universe started out in a state of high order or low entropy. Such states occupy only a very small fraction of the volume of phase space accessible to the Universe. As the Universe evolves in time it will tend to move around phase space ergodically. At a later time therefore there is a high probability that the Universe will be found in a state of disorder or higher entropy because such states occupy most of phase space. Consider, for example, a system consisting of a number N of gas molecules in a rectangular box which is divided into two by a partition with a small hole in it. Suppose that at some initial time, say 10 o'clock, all the molecules are in the left-hand side of the box. Such configurations occupy only one part in 2^N of the available $6N$-dimensional phase space. As time goes on, the system will move around phase space on a constant-energy surface. At a later time there will be a high probability of finding the system in a more disordered state with molecules in both halves of the box. Thus entropy will increase with time. Of course, if one waits long enough, one will eventually see all the molecules returning to one half of the box. However, for macroscopic values of N, the time taken is likely to be much longer than the age of the Universe.

Suppose, on the other hand, that the Universe satisfied a *final* condition that was in a state of high order. In that case it would be likely to be in a more disordered state at earlier times and entropy would decrease with time. However, as remarked above, the psychological arrow is determined by the thermodynamic arrow. Thus, if the thermodynamic arrow were reversed, the psychological arrow would be reversed as well: we would define time to run in the other direction and we would still say that entropy increased with time. However, the cosmological arrow provides an independent definition of the direction of time with which we can compare the thermodynamic, psychological, and electrodynamic arrows. In the early 1960s Hogarth[5] and Hoyle and Narlikar[6] tried to connect the electrodynamic and cosmological arrows using the Wheeler-Feynman[7] direct-particle-interaction formulation of electrodynamics. At a summer school held[8] at Cornell in 1963 their work was criticized by a Mr. X (generally assumed to be Richard Feynman) on the grounds that they had implicitly assumed the thermodynamic arrow. They also got the "wrong" answer in that they predicted retarded potentials in a steady-state universe but advanced ones in an evolutionary universe without continual creation of matter. It is now generally accepted that we live in an evolutionary universe.

Another proposal to explain the thermodynamic arrow of time has been put forward by Penrose.[9] It is based on the prediction of classical general relativity[10] that there will be spacetime singularities both in the past, at the big bang, and in the future at the big crunch, if the whole universe recollapses, or in black holes if only local regions collapse. Penrose's proposal is that the Weyl tensor should be zero at singularities in the past. This would mean that the Universe would have to start off in a smooth and uniform state of high order. However, the Weyl tensor would not, in general, be zero at singularities in the future which could be irregular and disordered.

There are several objections which can be raised to Penrose's proposal. First, it is rather *ad hoc*. Why should the Weyl tensor be zero on past singularities but not on future ones? In effect, one is putting in the thermodynamic arrow by hand. Second, it is based on the prediction of singularities in classical general relativity. However, it is generally believed that the gravitational field has to be quantized in order to be consistent with other field theories which are quantized. It is not clear whether singularities occur in quantum gravity or how to impose Penrose's boundary condition at them, if they do. Finally, Penrose's proposal does not explain why the cosmological and thermodynamic arrows should agree. With Penrose's boundary condition the thermodynamic arrow would agree with the cosmological arrow during the expanding phase of the Universe but it would disagree if the Universe were to start recollapsing.

The *CPT* invariance of the quantum state of the Universe defined by a path integral over compact metrics implies that if there is a certain probability of the Universe expanding, there must be an equal probability of it contracting. In order for the thermodynamic and cosmological arrows to agree in both the expanding and contracting phases, one requires boundary conditions which imply that the Universe is in a smooth state of high order when it is small but that it may be in an inhomogeneous disordered state when it is large. In Sec. IV it will be shown that the results of Ref. 11 imply that this is indeed the case for the quantum state defined by a path integral over compact metrics. This means that during the expansion phase the Universe starts out in a smooth state of high order but that, as it expands, it becomes more inhomogeneous and disordered. Thus the thermodynamic and cosmological arrows agree. However, when the Universe starts to recollapse, it has to get back to a smooth state when it is small. This means that disorder will decrease with time during the contracting phase and the thermodynamic arrow will be reversed. It will thus still agree with the cosmological arrow.

It should be emphasized that this reversal of the thermodynamic arrow of time is not caused by the gravitational fields or quantum effects at the point of maximum expansion of the Universe. Rather it is a result of the boundary condition that the Universe should be in a state of high order when it is small and it would occur in any theory which had this boundary condition as has been pointed out by a number of authors.[12,13] The only way that quantum gravity comes into the question of the arrow of time is that it provides a natural justification for the boundary condition.

One might ask what would happen to an observer (or computer) who survived from the expanding phase to the contracting one. One might think that one was free to enclose the observer or computer in a container that was so well insulated that he would be unaffected by the reversal of the thermodynamic arrow outside. If he were then to open a little window in his spaceship, he would see time going backward outside. The answer to this apparent paradox is that the observer's thermodynamic arrow, and hence his psychological arrow, would reverse at around the time of maximum expansion of the Universe, not because of effects that propagated into the spacecraft through the walls, but because of the boundary condition that the spacecraft be in a state of low entropy at late times when the Universe is small again. The contents of the memory of the observer or computer would increase during the expansion phase as the observer recorded observations but it would decrease during the contracting phase because the psychological arrow would be reversed and the observer would remember events in his future rather than his past.

The prediction that the thermodynamic arrow would reverse if the Universe started to recontract may not have much practical importance because the Universe is not going to recollapse for a long time, if it ever does. However, we are fairly confident that localized regions of the Universe will collapse to form black holes. If one was in such a region, it would seem just like the whole Universe was collapsing around one. One might therefore expect that the region would become smooth and ordered, just like the whole Universe would if it recollapsed. Thus one would predict that the thermodynamic arrow of time should be reversed inside black holes. One would expect this reversal to occur only after one has fallen through the event horizon, so one would not be able to tell anyone outside about it. This and other consequences of the point of view adopted in this paper will be considered further in Sec. V. Section II will be a brief review of the canonical formulation of quantum gravity. In Sec. III it will be shown that the quantum state of the Universe defined by a path integral over compact metrics is invariant under *CPT*. Despite this invariance it will be shown in Sec. IV that the results of Ref. 11 imply that there is a thermodynamic arrow because the inhomogeneities in the Universe are small when the Universe is small but that they grow as the Universe expands.

II. CANONICAL QUANTUM GRAVITY

In the canonical approach the quantum state of the Universe is represented by a wave function $\Psi(h_{ij},\phi_0)$ which is a function of the three-metric h_{ij} and the matter field configuration ϕ_0 on a three-surface S. The interpretation of the wave function is that $|\Psi(h_{ij},\phi_0)|^2$ is the (unnormalized) probability of finding a three-surface S with three-metric h_{ij} and matter field configuration ϕ_0. The wave function is not an explicit function of time because there is no invariant definition of time in a curved space which is not asymptotically flat. In fact, the position in time of the surface S is determined implicitly by the three-metric h_{ij}. This means that $\Psi(h_{ij},\phi_0)$ obeys the zero-energy Schrödinger equation:

$$H\Psi(h_{ij},\phi_0)=0 . \tag{2.1}$$

This equation can be decomposed into two parts: the momentum constraint and the Wheeler-DeWitt equation. The momentum constraint is

$$\left[\frac{\delta\Psi}{\delta h_{ij}}\right]_{|i}=8\pi T^{0j}\Psi . \tag{2.2}$$

It implies that the wave function is the same on three-metrics h_{ij} and matter field configurations ϕ_0 that are related by a coordinate transformation. The Wheeler-DeWitt equation is

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}}+h^{1/2}(-{}^3R+16\pi T_{00})\right]\Psi=0 , \tag{2.3}$$

where

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl}) .$$

It can be regarded as a second-order wave equation for Ψ on the infinite-dimensional space called superspace which is the space of all three-metrics h_{ij} and matter field configurations ϕ_0.

Any solution of Eqs. (2.2) and (2.3) represents a possible quantum state of the Universe. However, it seems reasonable to suppose that the Universe is not just in some arbitrary state but that its state is picked out or preferred in some way. As explained in Ref. 4, the most natural choice of quantum state is that for which the wave function is given by a path integral over compact metrics:

$$\Psi(h_{ij},\phi_0)=\int_C d[g_{\mu\nu}]d[\phi]\exp(-\tilde{I}[g_{\mu\nu},\phi]) , \tag{2.4}$$

where $\tilde{I}$ is the Euclidean action and the path integral is taken over four-metrics $g_{\mu\nu}$ and matter field configurations ϕ on compact four-manifolds which are bounded by the three-surface S with the induced three-metric h_{ij} and matter field configuration ϕ_0. The contour of integration in the space of all four-metrics has to be deformed from Euclidean (i.e., positive definite) metrics to complex metrics in order to make the path integral converge.[14,15] The proposal that the quantum state is given by (2.4) seems to give predictions that are in agreement with observation.[4,11,16]

III. THE *CPT* THEOREM

The precise statement of *CPT* invariance in flat spacetime is that the vacuum expectation values of bosonic quantum field operators $\phi(x)$ satisfy

$$\langle\phi(x_1)\phi(x_2)\cdots\phi(x_n)\rangle =[\langle\phi^\dagger(-x_1)\phi^\dagger(-x_2)\cdots\phi^\dagger(-x_n)\rangle]^* . \tag{3.1}$$

In the case of fermion fields there is a factor of $(-1)^{F+J}$ where F is the fermion number and J is the number of undotted spinor indices. In the case of asymptotically flat spacetime one can formulate and prove *CPT* invariance in a similar way in terms of the vacuum expectation values of field operators at infinity.[17] However, although asymptotic flatness may be a reasonable approximation for local systems, one does not expect it to apply to the whole Universe. One therefore does not have any flat or asymptotically flat region in which one can define the *TP* operation $x\rightarrow -x$. All that one has is a wave function $\Psi(h_{ij},\phi_0)$ which is not an explicit function of time. However, one can introduce a concept of time by replacing the dependence of Ψ on $h^{1/2}$, the square root of the determinant of the three-metric h_{ij}, by its conjugate momentum, the trace of the second fundamental form of S. One defines the Laplace transform

$$\Phi(\tilde{h}_{ij},K_E,\phi_0)=\int_0^\infty d[h^{1/2}]\exp\left[-\frac{m_p{}^2K_E}{18\pi}\int K_E h^{1/2}d^3x\right]\Psi(h_{ij},\phi_0) , \tag{3.2}$$

where $\tilde{h}_{ij}$ is the three-metric defined up to a conformal factor and K_E is the trace of the Euclidean second fundamental form. The Laplace transform Φ is holomorphic for $\mathrm{Re}(K_E)>0$. This means that one can analytically continue Φ in K_E to Lorentzian values $K_L=iK_E$ of the trace of the second fundamental form. Then $|\Phi(\tilde{h}_{ij},K_L,\phi_0)|^2$ is proportional to the probability of finding a three-surface S with the conformal three-metric $\tilde{h}_{ij}$, the rate of expansion K_L and the matter field configuration ϕ_0.

Consider first the case in which one has only fields like the gravitational field and real scalar fields which are invariant under C and P. The Euclidean action $\tilde{I}$ is real for Euclidean (i.e., positive definite) four-metrics $g_{\mu\nu}$ and real scalar fields ϕ. The contour of integration in the path integral (2.4) has to be deformed from Euclidean to complex metrics in order to make the integral converge. However, there will be an equal contribution from metrics with a complex action $\tilde{I}$ and from metrics with the complex conjugate action $(\tilde{I})^*$. Thus the wave function $\Psi(h_{ij},\phi_0)$ will be real. This implies that

$$\Phi(\tilde{h}_{ij},K_E,\phi_0)=\Phi^*(\tilde{h}_{ij},K_E^*,\phi_0) \tag{3.3}$$

for complex K_E. In particular, this implies

$$\Phi(\tilde{h}_{ij},K_L,\phi_0)=\Phi^*(\tilde{h}_{ij},-K_L,\phi_0) \tag{3.4}$$

for real K_L. Equation (3.4) is the statement of T invariance for the quantum state of the Universe. It implies that the probability of finding a contracting three-surface is the same as that of finding an expanding one, i.e., if the wave function represents an expanding phase of the Universe, then it will also represent a contracting one.

Consider now a situation in which one has charged fields, for example, a complex scalar field ϕ. The wave function Ψ will now be a functional of the three-metric h_{ij} and the complex field configuration ϕ_0 on S. In the Euclidean path integral (2.4) for Ψ one has to integrate over independent field configurations ϕ and $\tilde{\phi}$ on the Euclidean background $g_{\mu\nu}$ where $\phi=\phi_0$ and $\tilde{\phi}=\phi_0^*$ on S. The Euclidean action $\tilde{I}[g_{\mu\nu},\phi,\tilde{\phi}]$ is no longer necessarily real but

$$\tilde{I}[g_{\mu\nu},\phi,\tilde{\phi}]=\tilde{I}^*[g_{\mu\nu},\tilde{\phi},\phi]\ . \tag{3.5}$$

This implies

$$\Phi(\tilde{h}_{ij},K_L,\phi)=\Phi^*(\tilde{h}_{ij},-K_L,\phi^*)\ . \tag{3.6}$$

Equation (3.6) is a statement of the invariance of the quantum state of the universe under CT.

Finally one can consider fields, such as chiral fermions, which are not invariant under P. To deal with fermions one should introduce a triad of covectors e_i^a on S and should regard the wave function Ψ as a functional of the e_i^a and the fermion field ψ_0 on S. The path integral representation of the wave function is then

$$\Psi(e_i^a,\psi_0)=\int_C d[e_\mu^\alpha]d[\psi]d[\tilde{\psi}]\exp(-\tilde{I}[e_\mu^\alpha,\psi,\tilde{\psi}])\ , \tag{3.7}$$

where on S, $\psi=\psi_0$ and $\tilde{\psi}=\bar{\psi}_0$. The oriented triad e_i^a on S defines a directed unit normal e_μ^0 to S. The path integral (3.7) is taken over all compact four-geometries which are bounded by S and for which e_μ^0 points inward.

The Euclidean action will obey

$$\tilde{I}[e_\mu^\alpha,\psi,\tilde{\psi}]=\tilde{I}^*[-e_\mu^\alpha,\psi^C,\tilde{\psi}^C]\ , \tag{3.8}$$

where $\psi^C=C\psi^*$ is the charge conjugate field and C is the charge conjugation matrix. This implies

$$\Psi(e_i^a,\psi)=\Psi^*(-e_i^a,\psi^C)\ . \tag{3.9}$$

One can regard (3.9) as the expression of the CPT invariance of the quantum state of the Universe because changing the sign of the triad e_i^a not only reverses the spatial directions, and so carries out the operation P, but it also reverses the direction of the orientated normal to S, e_μ^0. Alternatively, one can consider the Laplace transform Φ

$$\Phi(\tilde{e}_i^a,K_L,\psi)=\Phi^*(-\tilde{e}_i^a,-K_L,\psi^C)\ , \tag{3.10}$$

where $\tilde{e}_i^a$ is the triad in S defined up to a positive multiplicative factor.

It is clear that this proof of the CPT invariance of the quantum state defined by a path integral over compact metrics would apply equally well if there were higher derivative terms in the gravitational action. In the case of an action containing quadratic terms in the curvature, the wave function Ψ could be taken to be a function of the three-metric h_{ij}, the second fundamental form K^{ij}, and the matter field configuration ϕ_0. For fields that are invariant under C and P, the wave function $\Psi(h_{ij},K_E^{ij},\phi_0)$ would be real for real Euclidean values of the second fundamental form K_E^{ij}. This implies that

$$\Psi(h_{ij},K_L^{ij},\phi_0)=\Psi^*(h_{ij},-K_L^{ij},\phi_0)\ . \tag{3.11}$$

One can regard (3.11) as an expression of the T invariance of the quantum state. The extension to fields that are not invariant under C and P is straightforward. One can also apply similar arguments to the corresponding quantum state in Kaluza-Klein theories.

IV. THE INCREASE OF DISORDER

In Ref. 11 it was argued that the wave function $\Psi(h_{ij},\phi_0)$ can be approximated by a sum of terms of the form

$$\Psi_0(\alpha,\phi)\prod_n \Psi_n(\alpha,\phi,a_n,b_n,c_n,d_n,f_n)\ . \tag{4.1}$$

The wave function Ψ_0 describes a homogeneous isotropic closed Universe of radius e^α containing a homogeneous massive scalar field ϕ. The quantities $a_n,b_n,\ldots,f_n$ are the coefficients of harmonics of order n which describe perturbations from homogeneity and isotropy.

One can substitute (4.1) into the Wheeler-DeWitt equation and keep terms to all orders in the "background" quantities α and ϕ but only to second order in the "perturbations" $a_n,b_n,\ldots,f_n$. One obtains a second-order wave equation for Ψ_0 on the two-dimensional "minisuperspace" parametrized by the coordinates α and ϕ. The path integral (2.4) for the wave function implies that $\Psi_0\to 1$ as $\alpha\to-\infty$. One can integrate the wave equation with this boundary condition.[18] One finds that Ψ_0 starts to oscillate rapidly. This allows one to apply the WKB approximation

$$\Psi_0 = \mathrm{Re}(C e^{iS}) . \tag{4.2}$$

The trajectories of ∇S in the (α,ϕ) plane correspond to solutions of the classical field equations for a homogeneous isotropic Universe with a homogeneous massive scalar field. The trajectories corresponding to Ψ_0 start out at large values of $|\phi|$. They have a period of exponential expansion in which $|\phi|$ decreases followed by a period of matter dominated expansion in which ϕ oscillates around zero with decreasing amplitude. They reach a point of maximum expansion and then recontract in a time symmetric manner.

The perturbation wave functions Ψ_n can be further decomposed as follows:

$$\Psi_n = {}^S\Psi_n(\alpha,\phi,a_n,b_n,f_n)\,{}^V\Psi_n(\alpha,\phi,c_n)\,{}^T\Psi_n(\alpha,\phi,d_n) . \tag{4.3}$$

The wave function ${}^T\Psi_n$ describes gravitational wave perturbations parametrized by the coefficients d_n of the transverse traceless harmonics on the three-sphere. The wave function ${}^V\Psi_n$ describes the effect of gauge transformations which correspond to coordinate transformations on the three-sphere parametrized by the coefficients c_n of the vector harmonics. The wave function ${}^S\Psi_n$ parametrized by the coefficients a_n, b_n, and f_n of the scalar harmonics describe two gauge degrees of freedom and one physical degree of freedom of density perturbations. In situations in which the WKB approximation can be applied to the background wave function Ψ_0, the perturbation wave functions obey decoupled Schrödinger equations of the form

$$i\frac{\partial\,{}^T\Psi_n}{\partial t} = {}^TH\,{}^T\Psi_n , \tag{4.4}$$

where t is the time parameter of the solution of the classical field equations that corresponds to Ψ_0 via the WKB approximation.

One can evaluate the perturbation wave functions directly from the path integral expression (2.4) for the wave function. Consider, for example, the gravitational wave perturbations. One can regard them as quantum fields parametrized by d_n' propagating on a homogeneous isotropic background metric of the form

$$ds^2 = -N(t)^2dt^2 + e^{2\alpha'(t)}d\Omega_3^{\,2} , \tag{4.5}$$

where $d\Omega_3^{\,2}$ is the metric on the unit three-sphere, if the lapse function N is real everywhere, the metric (4.5) has a Lorentzian signature and cannot be compact and nonsingular. However, I shall consider complex background fields $(N(t),\alpha'(t),\phi'(t))$ such that at some value $t=t_0$, N is negative imaginary. The metric then has a Euclidean signature at $t=t_0$ and will be regular and compact if $\alpha'=-\infty$, $d\alpha'/dt = iNe^{-\alpha'}$, and $d_n'=0$. The argument of N will vary continuously with t. When N becomes real, the metric will become Lorentzian. One can express the perturbation wave functions as path integrals on these backgrounds, e.g.,

$${}^T\Psi_n(\alpha,\phi,d_n) = \int d[d_n']\exp(-\tilde{I}[\alpha',\phi',d_n']) , \tag{4.6}$$

where the path integral is taken over all gravitational wave perturbations d_n' on all regular compact background fields described by $\alpha'(t)$ and $\phi'(t)$.

The path integral over d_n' in a given background field is Gaussian and therefore can be evaluated as

$$(\det\Delta)^{-1/2}\exp(-\tilde{I}^{\,\mathrm{cl}}[d_n]) , \tag{4.7}$$

where Δ is a differential operator and

$$\frac{e^{3\alpha}}{2iN}\left[d_n'\frac{d}{dt}d_n' + 4\frac{d\alpha'}{dt}d_n'^{\,2}\right] \tag{4.8}$$

is the action of a solution of the classical field equations for a perturbation d_n' on the given background with $d_n'=0$ at $t=t_0$ and $d_n'=d_n$ at the location $t=t_1$ of the three-surface S.

One expects the dominant contribution to the path integral (4.6) to come from backgrounds which are close to solutions of the classical background equations. These solutions will be Euclidean (N imaginary) at $t=t_0$ and they will become Lorentzian in those regions of the (α,ϕ) plane in which Ψ_0 oscillates and the WKB approximation can be applied. In such a background the classical field equation for d_n' is

$$\left[-\frac{d}{dt}\left[\frac{e^{3\alpha'}}{iN}\frac{d}{dt}\right] + iNe^{\alpha'}(n^2-1)\right]d_n' = 0 . \tag{4.9}$$

In the region of the (α',ϕ') plane in which the WKB approximation can be applied and N is real, one can regard Eq. (4.9) as a harmonic oscillator equation for the variable $x=\exp(3/2\alpha')d_n'$ with the time-dependent frequency $\nu=\exp(-\alpha')(n^2-1)^{1/2}$. If α' were independent of t, the solution of (4.9) that obeys the above boundary conditions is

$$d_n' = d_n\frac{\sin\nu\tau}{\sin\nu\tau_1} , \tag{4.10}$$

where $\tau=\int_{t_0}^{t}N\,dt$.

Of course α' will vary with t but (4.10) will still be a good approximation provided that the adiabatic approximation holds, i.e., $|\dot\alpha'/N|$, the rate of change of α', is small compared to the frequency ν. In the Euclidean region near $t=t_0$, this will be true because $|\dot\alpha'/N| < e^{-\dot\alpha'}$. In the Lorentzian region it will be true for perturbation modes whose wavelength ν^{-1} is small compared to the horizon distance $N/\dot\alpha'$. For such modes

$$\frac{d}{dt}d_n' = N\nu\, d_n'\cot\nu\tau . \tag{4.11}$$

For t_1 in the region in which the WKB approximation can be applied and for $n \gg 1$, the imaginary part of $\nu\tau_1$, which arises from the Euclidean region near $t=t_0$, will be less than $-i$. This means that the real part of the Euclidean action (4.8) will be $\frac{1}{2}\nu e^{3\alpha}d_n^{\,2} = \frac{1}{2}\nu x^2$. The imaginary part of the Euclidean action will be small. It will give rise to a phase factor in ${}^T\Psi_n$ which can be removed by a canonical transformation of variables. Thus the perturbation wave function will have the ground-state form

$$
\begin{aligned}
{}^{T}\Psi_n(d_n) &= B\exp[-\tfrac{1}{2}\nu\exp(3\alpha)d_n{}^2] \\
&= B\,e^{-\nu x^2/2}\,. \qquad (4.12)
\end{aligned}
$$

The vector perturbation wave function ${}^{V}\Psi_n(c_n)$ describes a gauge degree of freedom and does not have any physical significance. The scalar perturbation, which is a function ${}^{S}\Psi_n(a_n,b_n,f_n)$ describes two gauge degrees of freedom and one physical degree of freedom. A similar analysis and use of the adiabatic approximation shows that this physical degree of freedom is in its ground state when the wavelength of the perturbation is less than the horizon size during the period of exponential expansion. Thus at early times in the exponential expansion, i.e., when the Universe is small, the physical perturbation modes of the Universe have their minimum excitation. The Universe is in a state that is as ordered and homogeneous as it can be consistent with the uncertainty principle. This ordered state is not only an initial state for the expansion phase of the Universe but it is also a final state for the contracting phase because the WKB trajectories for Ψ_0 return to the same region of the (α,ϕ) plane and the perturbation wave functions depend only on the position in this plane.

On the other hand, the perturbation modes are not in their ground state when the Universe is large because in this case the adiabatic approximation breaks down when the wavelength of the perturbation becomes greater than the horizon size during the period of exponential expansion. Detailed calculations[11] show that when the scalar perturbation modes reenter the horizon during the matter-dominated era, they are in a highly excited state and give rise to a scale-free spectrum of density fluctuations $\delta\rho/\rho$. These density inhomogeneities provide the initial conditions necessary for the formation of galaxies and other structures in the Universe. The perturbation wave functions are still in a very special state because their phase factors have to be such that when they are evolved according to the Schrödinger equation, they will return to their ground-state form when the Universe recontracts. However, this special nature of the perturbation wave functions would not be noticed by an observer who makes the usual coarse-grained measurements. All he would notice was that during the expansion the Universe had evolved from a homogeneous, ordered state to an inhomogeneous, disordered state. Thus he would say that the thermodynamic arrow pointed in the direction of time in which the Universe was expanding. On the other hand, an observer in the contracting phase would feel that the Universe was evolving from a state of disorder to one of order. He would therefore ascribe the opposite direction to the thermodynamic arrow and would also find that it agreed with the cosmological arrow.

The connection between the thermodynamic and cosmological arrows should hold in models that are more general than the one considered in Ref. 11 because it depends only on the fact that the adiabatic approximation should hold for small perturbations on "small" three-geometries but not for perturbations on "large" three-geometries. Thus one might expect that it would also hold in models that allowed for the formation of black holes as a result of the gravitational collapse of density fluctuations produced during the expansion. This would mean that the thermodynamic arrow would reverse inside a black hole. This is currently under investigation.

V. CONSEQUENCES

Are there any observable consequences of the prediction that the thermodynamic arrow should reverse in a recontracting phase of the Universe or inside a black hole? Of course, one could wait until the Universe recollapsed or one could jump into a black hole. However, the probability distribution of the density parameter $\Omega=\rho/\rho_{\text{crit}}$ seems to be concentrated at $\Omega=1$ (Ref. 16). Thus one would have to wait a very long time for the collapse of the Universe. On the other hand, if one jumped into a black hole, one would not be able to tell anyone outside. Furthermore, if the thermodynamic arrow did reverse, one would not remember it because it would now be in one's future rather than in the past.

In principle it is possible to determine from the present positions and velocities of clusters of galaxies that they developed from an initial configuration with very low peculiar velocities. In a similar way it should therefore be possible to calculate whether they will evolve to a state with low peculiar velocities at some time in the future. The difficulty is that on the basis of the inflationary model, one would expect the value of Ω for the presently observed Universe to be equal to one to one part in 10^4. Thus one would expect the Universe to expand by a further factor of at least 10^4 before it began to recontract. In this extra expansion other clusters of galaxies which we have not yet observed would appear over the horizon and their gravitational fields could have a significant effect on the behavior of clusters near us. Thus it would seem very difficult to make an experimental test of the prediction that the thermodynamic arrow would reverse if the Universe began to recontract.

A better bet would seem to be to study the inflow of matter into a black hole. At least in principle this is a situation that we ought to be able to observe with some accuracy. However, on the basis of classical general relativity, one might expect the boundary of the region of high spacetime curvature not to be spacelike, as it is in the Schwarzschild solution, but to be null, like the Cauchy horizon in the Reissner-Nordström or Kerr solutions. If this were the case, the behavior of the matter and metric on the brink of the quantum era would depend on the entire future history of infall into the black hole. Merely to observe the infall for a limited period of time would be insufficient to determine whether or not the thermodynamic arrow of time reversed near the region of high curvature. Clearly more work has to be done on the classical and quantum aspects of gravitational collapse.

One might think that the *CPT* theorem implied that all the baryons in the Universe would have to decay into leptons before the Universe began to recollapse and that the leptons would be reassembled into antibaryons in the collapsing phase. If this were the case, one could disprove the proposed "no boundary" condition for the Universe if one could show that the observed value of Ω was such that the Universe should begin to recollapse before all the baryons had decayed. However, what the *CPT* theorem implies is just that the probability of finding an expanding three-surface with a matter configuration of baryons is the same as that of finding a contracting three-surface with a matter configuration of antibaryons. This requirement is no restriction at all because the two three-surfaces can merely be the same three-surface viewed with different orientation of time: reversing the orientation of time and space interchanges the labels, baryons, and antibaryons. Thus the *CPT* invariance of the quantum state of the Universe does not imply any limit on the lifetime of the proton. In any case, we certainly do not observe baryons changing into antibaryons as they fall into a black hole.

To sum up, the proposal that spacetime is compact without boundary implies that the quantum state of the Universe is invariant under *CPT*. Despite this, one would observe an increase in (coarse-grained) entropy during an expansion phase of the Universe. However, it seems difficult to test the prediction that entropy should decrease during a contracting phase of the Universe or inside a black hole.

Note added in proof. Since this paper was submitted for publication a paper by Don Page has appeared [following paper, Phys. Rev. D 32, 2496 (1985)]. In it he questions my conclusion that the thermodynamic arrow of time would reverse in a contracting phase of the universe or in a black hole. My conclusion was based on the fact that the wave function Ψ went exactly into 1 as one goes to $\alpha=-\infty$ on a null geodesic in the α,ϕ plane. This would imply that Ψ was not oscillating at large negative α and therefore that all the classical Lorentzian contracting solutions would have to bounce at a small radius. At the bounce one could apply an analysis similar to that in Ref. 11 to show that all the inhomogeneous modes were in their ground state. This would mean that the inhomogeneity would decrease in the collapsing phase and therefore that the thermodynamic arrow of time would be reversed.

Page has pointed out however that even at large negative α, there might be a small oscillating component in the wave function. This would arise from complex stationary points in the path integral over compact metrics that were near to the Lorentzian metric which started with an inflationary expansion, reached a maximum radius and then recollapsed to zero radius without boundary. Although the amplitude of this oscillating component would be small, its frequency would be very high. It would therefore correspond to an appreciable probability flux of classical solutions in the WKB approximation.[16] One would not expect the inhomogeneous perturbations about such solutions to be in their ground state when the solution recollapsed because the adiabatic approximation used in Ref. 11 would break down. There is thus no reason for the thermodynamic arrow of time to reverse in these solutions. Similarly one would not expect it to reverse inside black holes.

I think that Page may well be right in his suggestion. In that case the two main results of this paper that are correct are, first, that the wave function is invariant under *CPT*, though this does not imply that the individual classical solutions that correspond to the wave function via the WKB approximation are invariant under *CPT*, second, that the classical solutions, which start out with an inflationary period, will have a well-defined thermodynamic arrow of time.

[1]R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964).

[2]S. W. Hawking, in *Astrophysical Cosmology: Proceedings of the Study Week on Cosmology and Fundamental Physics,* edited by H. A. Brück, G. V. Coyne, and M. S. Longair (Pontificiae Academiae Scientiarum Scripta Varia, Vatican City, 1982), pp. 563–574.

[3]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[4]S. W. Hawking, Nucl. Phys. **B239**, 257 (1984).

[5]J. E. Hogarth, Proc. R. Soc. London **A267**, 365 (1962).

[6]F. Hoyle and J. V. Narlikar, Proc. R. Soc. London **A277**, 1 (1964).

[7]J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949).

[8]*The Nature of Time,* edited by T. Gold and D. L. Schumacher (Cornell University Press, Ithaca, 1967).

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

[10]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University Press, England, 1973).

[11]J. J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985).

[12]P. C. W. Davies, *The Physics of Time Asymmetry* (Surrey University Press/California University Press, Berkeley, 1974), Sec. 7.4.

[13]T. Gold, in *La Structure et l'Evolution de l'Universe,* 11th International Solvay Congress (Edition Stoops, Brussels, 1958); Am. J. Phys. **30**, 403 (1962); in *Recent Developments in General Relativity* (Pergamon-MacMillan, New York, 1962); H. Bondi, Observatory **82**, 133 (1962); D. L. Schumacher, Proc. Cambridge Philos. Soc. **60**, 575 (1964); M. Gell-Mann, comments in Proceedings of the Temple University Panel on Elementary Particles and Relativistic Astrophysics (unpublished); Y. Ne'eman, Int. J. Theor. Phys. **3**, 1 (1970); P. T. Landsberg, Stud. Gen. **23**, 1108 (1970); W. J. Cocke, Phys. Rev. **160**, 1165 (1967); H. Schmidt, J. Math. Phys. **7**, 495 (1966).

[14]G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. **B138**, 141 (1978).

[15]S. W. Hawking, in *Relativity, Groups and Topology,* Les Houches, 1983, edited by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).

[16]S. W. Hawking and D. N. Page, report 1985 (unpublished).

[17]S. W. Hawking, Commun. Math. Phys. **87**, 395 (1982).

[18]S. W. Hawking and Z. C. Wu, Phys. Lett. **151B**, 15 (1985).