Will entropy decrease if the Universe recollapses?

Don N. Page
Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802
(Received 19 July 1985)

Hawking has proposed a wave function for the Universe which is *CPT* invariant and which appears to be dominated by smooth configurations when the Universe is small but by irregular configurations when large. He has argued that this behavior connects the thermodynamic arrow of time with the cosmological arrow. Here it is pointed out that although the total wave function may have the properties Hawking describes, an individual WKB component of it that is classically observable generically will not be *CPT* invariant and will not have its thermodynamic arrow of time reverse if it enters a recollapsing phase.

I. HISTORICAL BACKGROUND

The temporal asymmetry or arrow of time of our world is one of the most striking facts of everyday experience, and yet it is one of the deepest mysteries of physics. It would seem very difficult to give a physical *explanation* for the asymmetry, because all of the fundamental dynamical laws of physics discovered so far are time symmetric in the sense of being *CPT* invariant. Of course, physics has given a *description* of the arrow of time in the form of the second law of thermodynamics: the entropy of the Universe or of any of its subsystems which become approximately isolated increases with time. However, it is generally agreed that this is not a fundamental dynamical law governing the microscopic evolution of the Universe but is rather a restriction on the boundary conditions which select the actual state for the Universe from the many presumably allowed by the dynamical laws.

Someone might object that there is no mystery in the arrow of time, for it is unlikely that the entropy of the Universe (at least in a coarse-grained sense) is absolutely constant, and then the future direction of time may be simply defined so that the entropy increases in that direction. But the real mystery is why virtually all of the many observed subsystems of the Universe have arrows which point in the *same* temporal direction. An explanation of the arrow of time could then not be expected to show why entropy increases in one direction of abstract coordinate time rather than the other, for that would be merely a matter of definition, but rather why there is the remarkable correlation between the directions of all of the different arrows of time. For example, one of the arrows could be used to define a positive direction of time, and then what would need to be explained is why the other arrows point toward the future thus defined.

It is generally accepted that if the thermodynamic arrows of time (entropy increase) agree for virtually all subsystems of the Universe, then the radiation arrows (massless fields being described more accurately by retarded solutions rather than advanced solutions if source-free solutions are dropped) and the psychological arrows (remembering the past) will also agree.[1] However, there is another arrow of time whose connection with the thermodynamic arrows is much less clear, and that is the expansion of the Universe or cosmological arrow of time. It is by no means obvious whether the thermodynamic arrows should always point in the same direction as the cosmological arrow, but Gold[2] has proposed that they do. Then if the Universe recollapses, the thermodynamic arrows would reverse with respect to some well-behaved coordinate time at the moment of maximum expansion and entropy would decrease thereafter.

This proposal for relating the cosmological and thermodynamic arrows of time is highly counterintuitive and hence has not been generally accepted, though it would be hard to rule out observationally if the Universe does not recollapse for a very long time. Penrose[3] has pointed out that even if the entire Universe does not recollapse soon, the proposal would apparently imply that the normal arrows of time would reverse for an astronaut who participates in a local recollapse by falling into a black hole. Penrose considered this possibility absurd, though admittedly it has not been tested.

Now a new element has entered the debate with Hawking's proposal for the quantum state of the Universe,[4–7] which he has argued[8] exhibits a strong thermodynamic arrow of time which agrees with the cosmological arrow. If Hawking's proposal is correct and does explain the thermodynamic arrow, it will solve one of the greatest mysteries of physics. Clearly more work is needed to check whether it really does.

In this paper I examine the much more modest question of whether Hawking's proposed wave function does have an agreement between the thermodynamic arrow (assuming this arrow is implied as claimed) and the cosmological arrow. I argue that although there may be this agreement when one averages over the entire wave function, the agreement need not (and apparently does not) occur for each component of the wave function that is accessible to classical observations.

II. TIME SYMMETRY WITHOUT TIME SYMMETRY

Hawking has shown[8] that the quantum state for the Universe is time symmetric in the sense of being *CPT* invariant if it is given by a path integral over compact geometries without boundaries.[4–8] Hawking does not argue that this symmetry by itself implies that the thermodynamic arrow would reverse with the cosmological arrow if the Universe recollapses, but because that reversal would naively seem to be a consequence, I must first show that it does not follow from *CPT* invariance of the state.

CPT is expressed in terms of an antiunitary operator θ which turns particles into antiparticles (C), takes the mirror image (P), and reverses velocities (T). The usual *CPT* invariance of the dynamical laws of physics means that θ commutes with all constraints and symmetry generators of the dynamical laws (e.g., with the Hamiltonian H). This implies that if a state ψ obeys the dynamical laws, so does the *CPT*-reversed state $\theta\psi$, but the two states need not be equal, so *CPT* invariance of the dynamics does not imply *CPT* invariance of the state ($\psi=\theta\psi$). Thus the usual *CPT* theorem does not preclude a time-asymmetric state, such as a state in which entropy increases monotonically with time throughout an expansion and subsequent recontraction of the Universe.

However, not only the dynamics but also the state is *CPT* invariant in Hawking's prescription for the wave function of the Universe, so it might appear that the Universe must expand and recontract in a time-symmetric manner. This would indeed be the case if the Universe has a *CPT*-invariant wave function consisting of a single WKB wave packet concentrated around a particular classical solution which expands from a big bang and then contracts to a big crunch. But if the wave function for the Universe consists of a superposition of many classical wave-packet components, it is very easy for the whole wave function to be *CPT* invariant even if all of its component parts are highly time asymmetric. For example, if the components ψ_n are all concentrated near classical evolutions with classical time parameters and all have entropies monotonically increasing with these time parameters, then $\theta\psi_n$ are *CPT*-reversed wave packets with the opposite sense of the time parameters and hence with entropies monotonically decreasing with respect to them, and $\psi=\sum_n(\psi_n+\theta\psi_n)$ is a *CPT*-invariant wave function all of whose components ψ_n and $\theta\psi_n$ are time asymmetric. This can occur even if each classical evolution has a succession of both expanding and contracting phases. Hence if the entropy increases monotonically during an expansion phase of any component wave packet, it need not decrease during a subsequent contraction phase even if the total wave function is *CPT* invariant, as each individual wave packet need not be *CPT* invariant.

III. TIME ASYMMETRY OF HAWKING'S WAVE PACKETS

Now the question arises as to how the entropy behaves in each of the component wave packets of Hawking's proposed wave function and whether the wave packets themselves behave in a time-symmetric manner. Of course, the full wave function would be very difficult to calculate. However, a Friedmann-Robertson-Walker (FRW) minisuperspace model with a minimally coupled homogeneous scalar field ϕ on three-spheres of radius a has been analyzed.[6,7,9,10] The oscillating part of this (real) minisuperspace wave function $\psi(a,\phi)$ may be decomposed into complex wave packets which represent an ensemble of classical FRW universes. In the regime where the WKB approximation is good, the integral curves of the normals to the surfaces of constant phase of the wave packets satisfy the classical FRW-scalar-field equations with respect to an internally defined Lorentzian time parameter t which varies monotonically along each curve or trajectory in the (a,ϕ) plane. The dominant wave packets enter the WKB regime along trajectories which come from points where $a=a_0$ is very small, $\phi=\phi_0$ is very large, and the scalar curvature of the three-sphere is balanced by the potential energy density of the ϕ field in the Einstein scalar constraint equation so that $\dot{a}=0$ and $\dot{\phi}=0$ there (with the overdot denoting d/dt). These trajectories undergo a long period of inflation or exponential expansion for $a(t)$ from a_0 to very large values, while $\phi(t)$ decreases in magnitude slowly from ϕ_0 until it nears a minimum in its potential. Then $\phi(t)$ oscillates about the minimum and $a(t)$ behaves as in a dust-filled FRW universe, since the pressure of the ϕ field averages to near zero over each of its oscillations. If the minimum of the ϕ potential (with any cosmological constant absorbed into it) is zero or negative, eventually $a(t)$ reaches a maximum (for our $k=+1$ FRW model) and then recontracts until it becomes very small or zero.

Hawking has pointed out[6,7] that there are an infinite number of periodic trajectories which never go to $a=0$ to become singular but rather continue to bounce each time a becomes very small by having $\dot{\phi}$ become sufficiently small there compared with ϕ itself that the pressure of the ϕ field is negative. In addition to this countable set of nonsingular periodic trajectories, there also appears to be an uncountable set of nonsingular aperiodic perpetually bouncing trajectories.[11] Hawking has argued that the wave function $\psi(a,\phi)$ corresponds to a superposition of wave packets whose trajectories in the WKB or classical limit are these nonsingular oscillating solutions.[6,7]

Halliwell and Hawking[12] have extended this FRW minisuperspace model to include linear perturbations of all the inhomogeneous or anisotropic gravitational and scalar-field modes. They found that these modes are in their ground states near $a=a_0$, $\phi=\phi_0$ along each classical trajectory in the (a,ϕ) plane but become excited by the mechanism of parametric amplification when their wavelengths expand beyond the horizon size during the inflationary phase. Thus one goes from a smooth, homogeneous configuration when a is small to an irregular, inhomogeneous configuration when a is large. Hawking argued[8] that this represents an increase in disorder or entropy which is the thermodynamic arrow of time. Furthermore, Hawking assumed that each classical trajectory will recontract in a time-symmetric manner and that the perturbations will go back to their ground states when a reaches a_0 again.[12,8] Hence he concluded that the thermodynamic arrow is tied to the cosmological arrow and would reverse during a contracting phase, with possible observational consequences for matter contracting locally into black holes.

However, only a discrete countable set of the classical trajectories (labeled by ϕ_0, with a_0 determined by the requirement that $\dot{a}=0$ and $\dot{\phi}=0$ at $a=a_0$ and $\phi=\phi_0$)

recontract in a time-symmetric manner after the first expansion of $a(t)$. A larger but still countable set of periodic solutions returns to $\dot{a}=0$ and $\dot{\phi}=0$ at $a=a_0$ and $\phi=\phi_0$ only after more than one expansion and contraction cycle. There is no reason to believe the entropy $S(t)$ of the perturbations in these would oscillate in the same way $a(t)$ does even if it did return to its original low value upon the eventual return to $a=a_0$ and $\phi=\phi_0$, which itself appears unlikely. Furthermore, there appears to be an uncountably much larger fractal set of nonsingular aperiodic solutions which never returns to $\dot{a}=0$ and $\dot{\phi}=0$ simultaneously,[11] in which case there would be no argument for $S(t)$ ever returning to its original value.

Even more crucially, all of these nonsingular trajectories, countable or uncountable, form only a discrete set of Lebesgue measure zero out of the continuum of trajectories which start with $\dot{a}=0$ and $\dot{\phi}=0$ at arbitrary $\phi=\phi_0$. The probability measure given to each trajectory by the wave function appears to be nearly independent of ϕ_0 for large ϕ_0 (Refs. 9 and 10) rather than being concentrated on the discrete values of ϕ_0 which give nonsingular trajectories, so it is apparently incorrect to characterize the wave function as being a superposition only of wave packets which represent nonsingular oscillating universes.[6–8,12] Although the Lorentzian (oscillating) part of the wave function $\psi(a,\phi)$ is predominantly a superposition of wave packets whose classical trajectories have $\dot{a}=0$ and $\dot{\phi}=0$ simultaneously somewhere, say at $t=0$ (and are thus only a one-parameter family of trajectories rather than the general two-parameter family of classical solutions of the FRW-scalar-field equations), all but a set of measure zero of these trajectories eventually reach $a=0$ and hence are singular.

Because nearly all of the classical trajectories given by the wave function are not time symmetric about any time other than $t=0$ [where $\dot{a}=0$ and $\dot{\phi}=0$, though the WKB approximation for the wave packets breaks down near there so that the Lorentzian time t and classical evolution $a(t)$ and $\phi(t)$ are not physically meaningful there], one would not expect the entropy of the inhomogeneous and anisotropic modes to be symmetric around any later time either. If indeed it is correct to conclude from the linear analysis for these modes[12] that the entropy is very low along each trajectory near $t=0$ (Ref. 8), the most natural behavior to expect would be that the entropy simply increases monotonically with t thereafter, all the way until the Universe recollapses completely to $a(t)=0$. As noted in Sec. II, this time-asymmetric behavior along each individual classical trajectory for $t>0$ does not violate the time symmetry (*CPT* invariance) of the full wave function, because for each wave packet having one asymmetry there is also its *CPT*-reversed wave packet having the opposite asymmetry. (For the dominant wave packets which give trajectories with $t>0$ which would come from $\dot{a}=0$ and $\dot{\phi}=0$ at $t=0$, the *CPT*-reversed wave packets simply give the extensions of the trajectories to $t<0$.) Thus there is no reason to believe anything strange will happen to the thermodynamic arrow of time in our component wave packet if the classical Universe described by it begins to contract (or even if a part of it such as the matter and space inside a black hole collapses, or if a ball of gas condenses to form a star, or if any other gravitational contraction occurs).

IV. WHAT HAPPENS TO THE WAVE FUNCTION WHEN THE UNIVERSE GETS SMALL?

In view of the expectation expressed in Sec. III that the entropy will increase monotonically with time along each of the classical trajectories whose wave packets are included in the total wave function, even when these trajectories collapse to $a=0$, one may ask what happens to Hawking's assumption that the Universe is smooth and homogeneous when small.[8] If the perturbations continue to grow along each classical trajectory in the (a,ϕ) plane, they will not be small when a reaches zero. This is not in contradiction with the results of Ref. 12, because there the perturbation wave function depends on the trajectory in the (a,ϕ) plane and not just on the position in this plane as assumed in Ref. 8. Hence the Universe need not be smooth and homogeneous when small.

One may next ask how the wave packets giving inhomogeneous and/or anisotropic collapse fit in with the expectation that the wave function goes to a constant when the size a of the three-space goes to zero.[8–10,12] In as much as this expectation is based on an estimate of the contribution of real Euclidean extrema to the path integral and ignores the contribution of complex Euclidean extrema such as the classical Lorentzian trajectories discussed above, it probably is not strictly correct. However, the Lorentzian trajectories which start with $\dot{a}=0$ and $\dot{\phi}=0$ at $a=a_0$ and $\phi=\phi_0$ typically return to $a=a_0$ at much larger values of ϕ. Because the corresponding wave packets are thus much more spread out when they return to $a=a_0$ and continue on to $a=0$, they do not make much of a correction to the total value of the wave function there, though they would greatly affect its derivatives which would otherwise be zero there.

Another possibility is that the wave function really does go to a constant with zero gradient at $a=0$, whether or not Hawking's path-integral proposal implies this. Then the contribution of the parts of the wave packets which collapse to $a=0$ after each contraction rather than bouncing (reflecting) at nonzero a must be canceled. This could be done by adding new wave packets of the opposite sign and opposite phase gradient coming out of $a=0$. The parts of these which fail to bounce after each contraction would also need to be canceled by adding a second set of new wave packets, and so on *ad infinitum* in what should be a convergent process for calculating the wave function in a finite region of superspace. In this case, a wave packet contracting toward $a=0$ would, by the time it got there, be completely canceled by another wave packet, but what the observed effects of this could be are not clear. In one sense the sum of the two wave packets would be fading out of existence as one followed them toward $a=0$, but an observer in one of the wave packets could only test conditional probabilities, with one of the conditions certainly being his own continued existence, so that he presumably could not become directly aware of his decreased absolute probability (or, more accurately, measure) of existence. Thus he might just gradually fade away without knowing it, all the while observing his universe getting smaller and more disordered.

Whichever is the correct form of the wave function at small a, it seems highly unlikely that it can consist only of wave packets which are both approximately classical and also have a strong thermodynamic arrow of time which reverses with the reversal of the expansion. Thus it does not seem that Hawking's proposal for the wave function of the Universe gives a rigid connection between the thermodynamic and cosmological arrows of time.

However, if Hawking's proposal does explain why the thermodynamic arrows of all the observed subsystems of the Universe point in the same direction, that would be a remarkable achievement, since there are so many of them. One might then simply say that by 50-50 chance we happen to be living in the expanding phase of our wave packet. Actually, it would not be surprising if the relative probability of our being in the expanding phase is much closer to unity, because this phase is predicted to last an arbitrarily long time,[10] and hence during the subsequent recollapse all stars may have burned out and there may not be much around except for large black holes continually coalescing.

ACKNOWLEDGMENTS

Appreciation is expressed for the hospitality of the Aspen Center for Physics, where most of this work was completed. Financial support was also provided by National Science Foundation Grant No. PHY-8316811 and by the Alfred P. Sloan Foundation. Comments by S. W. Hawking and the hospitality of the University of Cambridge have led to improved wording of Sec. II. The main point of that section has also been independently expressed by I. G. Moss (see Ref. 13).

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

[2] T. Gold, in *La Structure et l'Evolution de l'Universe*, 11th International Solvay Congress (Edition Stoops, Brussels, 1958), pp. 81—95; also in *Recent Developments in General Relativity* (Oxford University Press, New York, 1962), pp. 225—234; Am. J. Phys. **30**, 403 (1962).

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

[4] 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.

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

[6] S. W. Hawking, in *Relativity, Groups and Topology II*, edited by B. S. De Witt and R. Stora (North-Holland, Amsterdam, 1984), pp. 333—379.

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

[8] S. W. Hawking, preceding paper, Phys. Rev. D **32**, 2489 (1985).

[9] D. N. Page, in *Quantum Concepts in Space and Time*, edited by C. J. Isham and R. Penrose (Oxford University Press, New York, to be published), pp. 274—285.

[10] S. W. Hawking and D. N. Page, Nucl. Phys. B (to be published).

[11] D. N. Page, Class. Quantum Gravity **1**, 417 (1984).

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

[13] I. G. Moss, Nature **316**, 482 (1985).