Nuclear Physics B239 (1984) 257–276

THE QUANTUM STATE OF THE UNIVERSE

S.W. HAWKING

Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England

Received 28 December 1983

The quantum state of the universe is determined by the specification of the class of metrics and matter field configurations that are summed over in the path integral. The only natural choice of this class seems to be compact euclidean (i.e. positive definite) metrics and matter fields that are regular on them. This choice incorporates the idea that the universe is completely self-contained and has no boundary or asymptotic region. I show that in a simple "minisuperspace" model this boundary condition leads to a wave function which can be interpreted as a superposition of quantum states which are peaked around a family of classical solutions of the field equations. These solutions are non-singular and represent oscillating universes with a long inflationary period. They could be a good description of the observed universe. I also show that the features of the minisuperspace model that give rise to such a wave function are also present in models that contain all the degrees of freedom of the gravitational and matter fields.

1. Introduction

The task of theoretical physics is to construct a mathematical model of the universe which agrees with all the observations made so far and which predicts the results of future observations. This model usually consists of two parts:

(1) Local laws which govern the physical fields in the model. In classical physics, these laws are normally expressed as differential equations which can be derived from an action I. In quantum physics the laws can be obtained from a path integral over all field configurations weighted with $\exp(iI)$.

(2) Boundary conditions which pick out one particular state from among the set of those allowed by the local laws. The classical state can be specified by boundary conditions for the differential equations and the quantum state can be determined by asymptotic conditions on the class C of field configurations that are summed over in the path integral.

A great deal of work has been done on the first part, particularly in the last twenty years. Although we do not yet have a complete, consistent field theory, we probably know most of the features that it should have. The outstanding remaining problem is to construct a quantum theory of gravity. The standard approach to quantizing general relativity by a perturbation expansion produces infinities which are "unrenormalizable", that is, they cannot be removed by redefining the quantities in the action. There are three possible solutions to this problem:

(i) It may be possible to find a theory, such as $N=8$ supergravity, in which all the infinities cancel to all orders in the perturbation expansion.

(ii) The infinities may simply be a result of the perturbation expansion which is known to break down at high energies. Gravitational effects may cut off the high energies and give rise to a finite theory.

(iii) It may be necessary to add higher derivative terms to the gravitational action. This is known to give a renormalizable theory but there are problems with negative energies, unitarity and runaway solutions.

It seems likely that the solution to the problems of quantum gravity lies with one of the three alternatives listed above. My personal opinion is that it is probably the second possibility: the path integral over the conformal factor seems to damp out the contributions of metrics with a lot of conformal curvature. However, whichever of the three solutions is the correct one, it is likely that we will know before too long.

By contrast with all the work on the local laws, very little has been done on the boundary conditions. Indeed, many people would claim that the boundary conditions are not part of physics but belong to metaphysics or religion. They would claim that nature had complete freedom to start the universe off any way it wanted. That may be so, but it could also have made it evolve in a completely arbitrary and random manner. Yet all the evidence is that it evolves in a regular way according to certain laws. It would therefore seem reasonable to suppose that there are also laws governing the boundary conditions. One might also argue that the observed universe is so complicated that it could not possibly have arisen from some simple boundary condition. I do not think that this is a valid objection: the laws of quantum electrodynamics are simple but they give rise to all the complexities of chemistry and biology. I shall show in a simplified model that a simple boundary condition for the universe can still give rise to very complicated behaviour. Indeed, a number of classically allowed behaviours occur with relative probabilities determined by the boundary conditions.

It might seem premature to speculate about the boundary conditions when we are not yet sure of the exact form of the local laws. However, we know that the large-scale structure of the universe is determined by gravity and we are fairly sure that gravity is described by general relativity or some variant, such as supergravity or higher derivative theories. We also believe that quantum theory can be formulated in terms of path integrals. The problem of boundary conditions then becomes one of specifying the class of spacetime metrics and matter field configurations that are summed over in the path integral. I shall adopt the point of view that the path integral should be evaluated in the euclidean regime, that is, over positive definite metrics and matter fields in these metrics. There are then only two natural choices for the class of metrics:

(a) compact metrics and matter fields that are regular on them;

(b) non-compact metrics that are asymptotic to metrics of maximal symmetry, i.e. to flat euclidean space or to euclidean anti-de Sitter space, and matter fields that are asymptotically zero.

Boundary conditions of type (b) give rise to the usual vacuum state that one uses in scattering problems in which one sends particles in from infinity and measures what comes back out to infinity. However, I shall show in sect. 3 that they are not suitable as boundary conditions for the whole universe. This leaves only boundary conditions of type (a). They incorporate the idea that the universe is completely self-contained and has no boundaries, either at singularities at finite distance or at infinity. One could paraphrase them as: "the boundary conditions of the universe are that it has no boundary".

Boundary conditions of type (a) were proposed in [1] but at that time I did not know how to calculate their consequences. The necessary formalism was developed in collaboration with Hartle in [2] and is summarised in sect. 2. The formalism is applied in sect. 4 to a simple "minisuperspace" toy model of a homogeneous isotropic universe containing a massive scalar field. It is shown that the boundary conditions (a) give rise to a wave function that can be interpreted as a superposition of quantum states peaked around different classical solutions representing oscillating universes. One might object that the quantum state should be peaked around only a single classical solution: after all, we observe only one universe. The answer is that if one had a quantum state that was peaked around a particular classical solution, that quantum state would describe not only the universe but also any observers who measured its properties. Suppose one now had a different quantum state that was peaked around a different classical solution of the field equations. That quantum state would also contain observers who would measure that properties of the second solution. Then consider a quantum state that was the sum of the two previous states. Because of the linearity of quantum mechanics, there would be no interference; observers in the first state would measure the properties of the first solution and observers in the second state would measure those of the second. The classical solutions that correspond to the wave function of thé model can have a long period of exponential expansion and could be a reasonable description of the universe that we live in. They would have most of the desirable features of the inflationary model [3–7] and the advantage of having been derived from a definite proposal for the boundary conditions of the universe.

One could easily object to the minisuperspace toy model on the grounds that it truncated the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number and that it ignored the problems of thė divergences of quantum gravity. However, I show in sect. 5 that the features of the minisuperspace model that lead to a wave function of the right form would also be present in a model that did not restrict the degrees of freedom provided that it contains fields of non-zero rest-mass. The quantum corrections to the gravitational action will produce terms of the form R^2 and $C_{\mu\nu\rho\lambda}C^{\mu\nu\rho\lambda}$. These will be finite in the case of possibilities (i) and (ii) and infinite but renormalized in the case of possibility (iii). Such terms can act like massive scalar or spin-2 fields. It will be shown in a future paper [8] that the R^2 term at least leads to a minisuperspace model which is very similar to that with a massive scalar field.

These results and the fact that there does not seem to be any other natural candidate leads me, at least, to believe that the quantum state of the universe is determined by a path integral over compact positive definite metrics. If this is correct, the second part of the problem of constructing a mathematical model of the universe will have been solved. It would then only remain to determine the exact form of the local laws and we should, in principle, be able to predict the probability of anything in the universe. In practice of course, the equations will be much too complicated to solve in any but very simple situations.

2. Quantum gravity

In certain special spacetime metrics, such as those of the Schwarzschild or de Sitter solutions, it is possible to define a new time-coordinate $\tau = it$ and to change the metric from a lorentzian signature $(-+++)$ to a euclidean or positive definite signature $(++++)$. However, this is not possible in general: a general lorentzian metric will not have a section in the complexified spacetime manifold on which the metric is real and positive definite. Similarly, a general positive definite metric will not have a section on which the metric is real and lorentzian. In this respect, the spacetime metric behaves like other fields in ordinary quantum field theory in flat space: a Yang–Mills field which is real in Minkowski space will not, in general, be real when analytically continued to euclidean space and a field which is real in euclidean space will not be real when analytically continued to Minkowski space. The point about the euclidean approach to quantum field theory in flat space is not that individual configurations in euclidean space correspond with configurations in Minkowski space, but that the path integral over all real field configurations in euclidean space is equivalent, in the sense of contour integration, to the path integral over all real field configurations in Minkowski space. If this works for other quantum fields, it seems reasonable to suppose that it works for gravity also. I shall therefore assume that the quantum theory of the universe can be derived from a path integral over euclidean, i.e. positive definite, metrics and matter fields which are regular on the manifolds defined by these metrics. I shall show that the Wheeler–DeWitt equation, the analogue of the Schrödinger equation, is the same whether it is derived from a path integral over euclidean or lorentzian metrics but that the wave function of the universe itself determines whether it corresponds to a euclidean or a lorentzian geometry in the classical limit.

As in refs. [2, 9], my starting point is the assumption that the probability for a 4-metric $g_{\mu\nu}$ and a matter field configuration ϕ is proportional to

$$\exp\left(-\tilde{I}[g_{\mu\nu}, \phi]\right), \tag{2.1}$$

where $\tilde{I}$ is the euclidean action

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_{\mathrm{P}}^2}{16\pi}\left(-\int_{\partial\mathrm{M}} 2Kh^{1/2}\,\mathrm{d}^3x - \int_{\mathrm{M}}\left(R - 2\Lambda + \frac{16\pi}{m_{\mathrm{P}}^2}L(g_{\mu\nu}, \phi)\right)g^{1/2}\,\mathrm{d}^4x\right), \tag{2.2}$$

where h_{ij} is the 3-metric on the boundary ∂M and K is the trace of the second fundamental form of the boundary. The surface term in the action is necessary because the curvature scalar R contains second derivatives of the metric [10–12]. The physics of the universe is governed by probabilities of the form (2.1) for all 4-metrics $g_{\mu\nu}$ and matter field configurations belonging to a certain class C. The specification of the class C determines the quantum state of the universe. This will be considered further in sect. 3.

In practice, one is normally interested in the probability, not of the entire 4-metric, but of a more restricted set of observables. Such a probability can be derived from the basic probability (2.1) by integrating over the unobserved quantities. In cosmology, one is concerned with observables, not at infinity, but in some finite region in the interior of the 4-geometry. A particularly important case is the probability $P[h_{ij}, \phi]$ of finding a closed compact 3-submanifold S which divides the 4-manifold M into two parts $\mathrm{M}_{\pm}$ and on which the induced 3-metric is h_{ij} and the matter field configuration is ϕ:

$$P[h_{ij}, \phi] = \int \mathrm{d}[g_{\mu\nu}]\,\mathrm{d}[\phi]\exp(-\tilde{I}[g_{\mu\nu}, \phi])\,, \tag{2.3}$$

where the integral is over all 4-metrics and matter field configurations belonging to the class C which contain a 3-submanifold S on which the induced 3-metric is h_{ij} and the matter field configuration is ϕ. This probability can be factorized into the product of two amplitudes or wave functions $\Psi_{\pm}[h_{ij}, \phi]$, $P[h_{ij}, \phi] = \Psi_{+}[h_{ij}, \phi]\Psi_{-}[h_{ij}, \phi]$, where

$$\Psi_{\pm}[h_{ij}, \phi] = \int_{\mathrm{C}_{\pm}} \mathrm{d}[g_{\mu\nu}]\,\mathrm{d}[\phi]\exp(-\tilde{I}[g_{\mu\nu}, \phi])\,. \tag{2.4}$$

The path integral is over the classes $\mathrm{C}_{\pm}$ of 4-metrics and matter field configurations on $\mathrm{M}_{\pm}$ which agree with the given 3-metric h_{ij} and matter field configuration ϕ on S. For example, if the class C consisted of asymptotically euclidean metrics, then C_{+} would consist of 4-metrics and matter field configurations on compact manifolds M_{+} with boundary S and C_{-} would consist of metrics and fields on asymptotically euclidean manifolds M_{-} with inner boundary S. In this case, the two wave functions $\Psi_{\pm}$ will be different from each other but, if the class C consists of compact metrics, they will be the same.

One can regard $\Psi_{\pm}(h_{ij}, \phi]$ as the "wave functions of the universe" [2]. Note that they do not depend on time or, more generally, on the position of the 3-manifold S in the 4-manifold M. In general, there is no invariant way to specify this. In fact, the 3-metric h_{ij} usually determines the position of S in M or determines it up to a finite ambiguity. In what follows I shall drop the subscripts $\pm$ on the wave function Ψ.

The wave function $\Psi[h_{ij}, \phi]$ obeys a functional differential equation, the Wheeler–DeWitt equation, which is the analogue of the Schrödinger equation. This can be derived from the path integral definition (2.4) as follows. In the neighbourhood of the surface S, one can introduce a time-coordinate τ which is constant on S so that the metric takes the standard 3+1 form:

$$ds^2 = (N^2 - N_i N^i)\,d\tau^2 + 2N_i\,dx^i\,d\tau + h_{ij}\,dx^i\,dx^j . \tag{2.5}$$

The action then becomes

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_p^2}{16\pi}\int d^3x\,d\tau\,h^{1/2}N[K_{ij}K^{ij} - K^2 - {}^3R(h_{ij}) + 2\Lambda - 16\pi m_p^{-2}L(\phi)]\,, \tag{2.6}$$

where K_{ij} is the second fundamental form

$$K_{ij} = \frac{1}{N}\left[\frac{1}{2}\frac{\partial h_{ij}}{\partial \tau} + N_{(i|j)}\right] \tag{2.7}$$

and a stroke and ${}^3R$ denote the covariant derivative and scalar curvature constructed from the 3-metric h_{ij}.

The functional integral defining the wave function contains an integral over N. By varying N at the surface one pushes it forward or backward in time. Since the wave function does not depend on time one must have

$$0 = \int d[g_{\mu\nu}]\,d[\phi]\left[\frac{\delta \tilde{I}}{\delta N}\right]\exp(-\tilde{I}[g_{\mu\nu}, \phi])\,. \tag{2.8}$$

Classically the field equation $H \equiv \delta\tilde{I}/\delta N = 0$ is the hamiltonian constraint for general relativity. It is

$$H = -\frac{m_p^2 h^{1/2}}{8\pi}(K^{ij}K_{ij} - K^2 + {}^3R - 2\Lambda + 8\pi m_p{}^2 T_{nn}) = 0\,, \tag{2.9}$$

where T_{nn} is the euclidean stress energy tensor of the matter field projected in the direction normal to the surface. One can substitute the expression (2.9) for $\delta\tilde{I}/\delta N$ in (2.8). The second fundamental form K_{ij} can be expressed in terms of the functional operator $\delta/\delta h_{ij}$:

$$\frac{\delta \tilde{I}}{\delta h_{ij}} = \frac{m_p^2}{16\pi}h^{1/2}(K_{ij} - h_{ij}K)\,. \tag{2.10}$$

This identity can be derived by varying the metric in the definition of the action (2.2). The time derivative of the variation of the metric can be removed by integrating by parts in the normal manner. This produces a surface term which is δh_{ij} times the right-hand side of (2.10). Similarly, the time derivatives of the matter fields ϕ can be expressed in terms of the operator $\delta/\delta\phi$. Using these substitutions in (2.8), one obtains the Wheeler-DeWitt equation

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} - h^{1/2}\left({}^3R(h) - 2\Lambda + 8\pi m_{\mathrm{p}}{}^2 T_{\mathrm{nn}}\left(\frac{\delta}{\delta\phi}, \phi\right)\right)\right]\Psi[h_{ij}, \phi] = 0\,, \tag{2.11}$$

where G_{ijkl} is the metric on superspace, the space of all 3-metrics h_{ij}:

$$G_{ijkl} = \tfrac{1}{2}h^{1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl})\,. \tag{2.12}$$

There is a question of the factor ordering in the functional differential equation or, equivalently, of the measure in the path integral. This will not be important in the situation that I shall consider. The Wheeler–DeWitt equation is the same as the equation one would have obtained had one started out with a path integral over lorentzian metrics. One can also consider variations within the 3-surface S generated by the "Shift" vector N_i:

$$0 = \int \frac{\delta \tilde{I}}{\delta N_i}\exp(-\tilde{I}[g_{\mu\nu}, \phi])\,. \tag{2.13}$$

This equation implies that the wave function is invariant under diffeomorphisms, i.e. that Ψ is a functional of the 3-geometry and not of the particular 3-metric h_{ij}.

The metric $-G_{ijkl}$ on superspace, the space of 3-geometries, has a signature $(+-----)$. The Wheeler–DeWitt equation can therefore be thought of as a hyperbolic equation on superspace with $h^{1/2}$ as the time coordinate. I shall show that the semiclassical approximation to the path integral gives the boundary conditions for the Wheeler–DeWitt equation at small $h^{1/2}$. The Wheeler–DeWitt equation can then be solved to give the values of the wave function at larger values of $h^{1/2}$. In a lorentzian metric, $h^{1/2} = 0$ would be a singularity. However, this is not necessarily the case in a euclidean metric as can be seen by considering the example of a 4-sphere of radius R embedded in flat 5-dimensional space: a surface $|x^5| < R$ intersects the 4-sphere in a 3-sphere of non-zero radius. However, when $|x^5| = R$, the 3-sphere shrinks to zero radius but there is no singularity of the 4-geometry. Indeed, $h^{1/2}$ has to go to zero if the topology of the 3-surfaces is to change.

By its construction, the wave function $\Psi[h_{ij}, \phi]$ vanishes for 3-metrics h_{ij} which are not positive definite, i.e. for which the determinant h of the metric is negative. This means that h_{ij} is not a quantum observable with an unrestricted range and that in many cases it is more convenient to replace it by $\tilde{h}_{ij}$, the 3-metric up to a conformal factor, and $Km_{\mathrm{p}}^2/12\pi$, the momentum conjugate to $h^{1/2}$. One can then define a wave function Φ in this representation by

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int \mathrm{d}[g_{\mu\nu}]\,\mathrm{d}[\phi]\exp(-I^k[g_{\mu\nu}, \phi])\,, \tag{2.14}$$

where $I^k[g_{\mu\nu}, \phi]$ is the action appropriate to the situation in which K and $\tilde{h}_{ij}$ are kept fixed on the boundary rather than h_{ij}. It differs from the action $\tilde{I}[g_{\mu\nu}, \phi]$ in that the surface term has a different coefficient, $\frac{2}{3}K$ rather than $2K$. The wave function $\Phi[\tilde{h}_{ij}, K, \phi]$ is the Laplace transform of $\Phi[h_{ij}, \phi]$:

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int_0^\infty \mathrm{d}[h^{1/2}] \exp\left[-\frac{m_{\mathrm{P}}^2}{12\pi} \int \mathrm{d}^3x\, h^{1/2} K \right] \Psi[h_{ij}, \phi] . \tag{2.15}$$

Similarly, $\Psi[h_{ij}, \phi]$ is the inverse Laplace transform of $\Phi[\tilde{h}_{ij}, K, \phi]$.

$$\Psi[h_{ij}, \phi] = \int_\Gamma \mathrm{d}\left[\frac{m_{\mathrm{P}}^2}{24\pi i} K\right] \exp\left[\frac{m_{\mathrm{P}}^2}{12\pi} \int \mathrm{d}^3x\, h^{1/2} K\right] \Phi[\tilde{h}_{ij}, K, \phi] , \tag{2.16}$$

where for each point of S the contour Γ runs from $-i\infty$ to $+i\infty$ to the right of any singularities of $\Phi[\tilde{h}_{ij}, K, \phi]$ in the complex K plane. Providing that $\Phi[\tilde{h}_{ij}, K, \phi]$ does not diverge exponentially for large Re K, this choice of contour will ensure that $\Psi[h_{ij}, \phi] = 0$ for $h^{1/2} < 0$ because one can close the contour in the right half K plane.

The square of the trace of the second fundamental form K^2 can be expressed as an operator on the wave function Ψ:

$$K^2 = \frac{64\pi^2}{m_{\mathrm{P}}^4} \left[h^{-1/2} h_{ij} \frac{\delta}{\delta h_{ij}} \right] h^{-1/2} h_{kl} \frac{\delta}{\delta h_{kl}} . \tag{2.17}$$

If $K^2\Psi/\Psi$ is positive, one can interpret the wave function Ψ in terms of a euclidean 4-geometry in the classical limit. On the other hand, if $K^2\Psi/\Psi$ is negative, then the observable K is imaginary which corresponds to a lorentzian 4-geometry in the classical limit. The quantity $K^2\Psi/\Psi$ is positive or negative according to whether the wave function Ψ depends in an exponential or oscillatory manner on the scale $h^{1/2}$. Thus the form of the wave function determines whether it corresponds to a euclidean or lorentzian 4-geometry in the classical limit.

3. Boundary conditions

As was stated in the introduction, the boundary conditions on the class C of metrics that are summed over in the path integral determine the quantum state of the universe. There seem to be only two natural boundary conditions for positive definite metrics:

(a) compact metrics;

(b) non-compact metrics which are asymptotic to metrics of maximal symmetry, i.e. flat euclidean space or euclidean anti-de Sitter space.

Boundary conditions of type (b) define the usual vacuum state. In this state, the expectation values of most quantities are defined to be zero so the vacuum state is not of much interest as the quantum state of the universe. In particle scattering calculations, one starts with the vacuum state and one changes the state by creating particles by the action of field operators at infinity in the infinite past. One lets the particles interact and then annihilates the resultant particles by the action of other field operators at future infinity. This gets one back to the vacuum state. If one supposed that the quantum state of the universe was some such particle scattering state, one would lose all ability to predict the state of the universe because one would have no idea of what was coming in. One would also expect that the matter in the universe would be concentrated in a certain region and would decrease to zero at large distances instead of the roughly homogeneous universe that we observe.

In particle scattering problems, one is interested in observables at infinity. One is therefore concerned only with metrics which are connected to infinity: any disconnected compact parts of the metric would not contribute to the scattering of particles from infinity. In cosmology, on the other hand, one is concerned with observables in a finite region in the middle of the space and it does not matter whether this region is connected to an infinite asymptotic region. For example, one might ask for the amplitudes $\Psi_{\pm}[h_{ij}, \phi]$ for a closed compact 3-manifold S with a 3-metric h_{ij} and a matter field configuration ϕ. As explained in sect. 2, $\Psi_{\pm}$ is given by a path integral over all metrics belonging to classes $C_{\pm}$ which are bounded by S with the given 3-metric and matter field configuration. If the class C that determines the quantum state consists of metrics of type (b), then C_{+} would consist of compact metrics and C_{-} would contain two different kinds of metrics:

(i) connected asymptotically euclidean or anti-de Sitter metrics which had an inner boundary at S;

(ii) disconnected metrics which consisted of a compact part with boundary at S and an asymptotically euclidean or anti-de Sitter part without any inner boundary.

One cannot exclude disconnected metrics from the path integral because they can be approximated by connected metrics in which the different parts were joined by thin tubes. The tubes could be chosen to have negligible action. Similarly, topologically non-trivial metrics cannot be excluded because they can be approximated by topologically trivial metrics.

One would expect the dominant contribution to the path integral for the wave function to come from metrics which were near to solutions of the field equations with the appropriate boundary conditions. In the case of metrics of kind (i), it is shown in the appendix however that such solutions cannot give the dominant contribution to the wave function. This is borne out by calculations in specific examples. Thus, the dominant contribution to the wave function would come from metrics of kind (ii). These in general have a smaller action and hence give a bigger contribution than metrics of kind (i). One sees therefore that even if one adopted boundary conditions of type (b), the result would be almost the same as if one had adopted type (a), as far as the amplitude for finding a compact 3-submanifold S was concerned. Similar results would apply to other observations in a finite region. It would therefore seem more natural to suppose that the boundary conditions of the universe were of type (a), i.e. that the quantum state of the universe was defined by a path integral over compact positive definite metrics. In this case, the two amplitudes $\Psi_{\mp}$ are equal. I shall therefore drop the subscript $\pm$.

It should be emphasized that this is a *proposal* for the quantum state of the universe. One cannot derive it from some other principle but merely show that it is a natural choice. The ultimate test, however, is not whether it is aesthetically appealing but whether it enables one to make predictions that agree with observations. I shall endeavour to do this in the next section for a simple model.

4. Minisuperspace

In order to investigate whether the boundary conditions proposed in the last section correspond to the quantum state of the universe that we live in, one would like to calculate the wave function $\Psi[h_{ij}, \phi]$ and see if it can be interpreted as predicting what we observe. This is equivalent to finding a solution of the Wheeler–DeWitt equation obeying certain boundary conditions. The Wheeler–DeWitt equation is a second order functional differential equation on an infinite dimensional manifold, superspace, the space of all three metrics and matter field configurations. We do not know how to solve such an equation but we can hope to get some idea of the nature of the solution by considering the equation restricted to a finite dimensional submanifold called, minisuperspace. In other words, one restricts the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number. The Wheeler–DeWitt equation then becomes a wave equation on a finite dimensional manifold and can be solved by standard techniques. The boundary conditions for the Wheeler–DeWitt equation can be obtained from the semiclassical approximation to the path integral, i.e. by expressing the wave function in the form

$$\Psi[h_{ij}, \phi] = N_0 \sum_i A_i \exp(-B_i)\,, \tag{4.1}$$

where N_0 is a normalization constant, and B_i are the actions of classical solutions of the field equations which are compact and which have the given 3-metric h_{ij} and matter field configuration ϕ on the boundary. The prefactors A_i are given by the determinants of small fluctuations about the classical solutions. I shall not bother with them but only with the more important quantities B_i which appear in the exponential.

The simplest example of a minisuperspace model is that for a spatially homogeneous and isotropic universe. Such models have been considered by [10–13]. The euclidean metric can be written in the form

$$\mathrm{d}s^2 = \sigma^2[N^2(\tau)\,\mathrm{d}\tau^2 + a^2(\tau)\,\mathrm{d}\Omega_3^2]\,, \tag{4.2}$$

where $N(\tau)$ is the lapse function, $\sigma^2 = 2/(3\pi m_{\mathrm{p}}^2)$ and $\mathrm{d}\Omega_3^2$ is the metric on a 3-sphere of unit radius. The gravitational action is

$$\tilde{I} = \tfrac{1}{2}\int \mathrm{d}\tau \left\{\frac{N}{a}\right\}\left[-\left\{\frac{a}{N}\frac{\mathrm{d}a}{\mathrm{d}\tau}\right\}^2 - a^2\right]. \tag{4.3}$$

There is no solution of the classical lorentzian field equations for an empty closed homogeneous and isotropic universe. One therefore has to include some form of matter. The simplest equations are obtained for a conformally invariant scalar field ϕ which is constant on the 3-sphere sections. The euclidean action of such a field is

$$\tilde{I} = \tfrac{1}{2}\int \mathrm{d}\tau \left(\frac{N}{a}\right)\left[\left(\frac{a\,\mathrm{d}\chi}{N\,\mathrm{d}\tau}\right)^2 - \chi^2\right], \tag{4.4}$$

where $\chi = \sqrt{2}/\pi a\sigma\phi$. The Wheeler–DeWitt equation is then

$$\frac{1}{2}\left(\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right] - a^2 - \frac{\partial^2}{\partial\chi^2} + \chi^2\right)\Psi[a,\chi] = 0\,. \tag{4.5}$$

The advantage of using a conformally invariant scalar field ϕ is that the Wheeler–DeWitt equations separates, i.e. there are solutions of the form

$$\Psi(a,\chi) = C(a)f(\chi)\,. \tag{4.6}$$

The function f obeys the harmonic oscillator equation

$$\frac{1}{2}\left(-\frac{\mathrm{d}^2}{\mathrm{d}\chi^2} + \chi^2\right)f = Ef. \tag{4.7}$$

It is therefore natural to take f to be the harmonic oscillator wave function with eigenvalues $E = n + \frac{1}{2}$. The disadvantage of using a conformally invariant scalar field is that the boundary condition of integrating over compact metrics uniquely picks out the ground state $n = 0$ of the harmonic oscillator [2]. The gravitational part of the wave function $C(a)$ is then almost the same as if there were no matter present. Such a wave function could not represent the universe we live in. For this reason, Hartle and I proposed in [2] that the quantum state defined by integrating over compact metrics should be regarded as the "ground state" of the universe but that the universe that we lived in did not obey these boundary conditions. Rather it was a linear superposition of other "excited state" solutions with higher values of n. This meant that one lost any ability to predict the quantum state of the universe because one had no means of determining the complex coefficients of the excited states in the superposition.

With hindsight it is clear that Hartle and I were misled by the mathematical simplicity of the conformally invariant scalar field. If one fills in a 3-sphere of radius a with a topologically trivial euclidean 4-geometry of the form (4.2), the action of the scalar field is independent of the form of $a(\tau)$. This means that there is no coupling between the scalar field and the gravitational degree of freedom which is a purely conformal one. Such a model cannot describe the observed universe in which the expansion is strongly coupled to the matter content. In order to obtain a model which corresponds to what we observe, it is necessary to introduce non-conformally invariant fields. In this section I shall describe the simplest such model, a homogeneous isotropic universe with a spatially constant massive scalar field. A future paper [8] will deal with a model in which the scalar field is replaced by an effective R^2 term generated by quantum corrections.

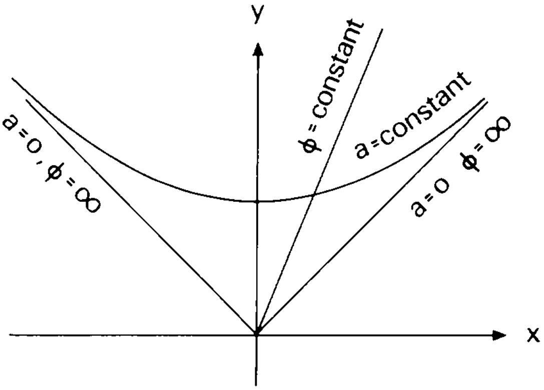

Fig. 1. The causal structure of the Wheeler–DeWitt equation can be seen in coordinates (x, y) in which the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$. The region $a > 0$ is mapped into the interior of the future light-cone of the origin.

The Wheeler–DeWitt equation for a homogeneous isotropic universe with a scalar field $\tilde{\phi}$ with mass $\tilde{m}$ is

$$\frac{1}{2}\left(\frac{1}{a^p}\frac{\partial}{\partial a}a^p\frac{\partial}{\partial a} - a^2 - \frac{1}{a^2}\frac{\partial^2}{\partial \phi^2} + a^4 m^2\phi^2\right)\Psi(a, \phi) = 0\,, \tag{4.8}$$

where $\phi = \sigma\tilde{\phi}$ and $m = \sigma\tilde{m}$. The exponent p represents some, but not all of the uncertainty in the factor ordering. Its precise value will not be important for what follows. One can regard (4.8) as a wave equation in the (a, ϕ) plane. The causal structure of the Wheeler–DeWitt equation is seen more easily by introducing new coordinates

$$x = a \sinh \phi, \qquad y = a \cosh \phi\,. \tag{4.9}$$

In these coordinates, the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$. Thus, the characteristics are lines at 45° in the (x, y) plane. The region $a > 0$ is mapped into the interior of the future light cone of the origin (see fig. 1). The surfaces of constant ϕ are straight lines through the origin and the surfaces of constant a are spacelike hyperbolae within the light cone.

If one knew the solution on the light cone of the origin, one could solve the wave equation by standard techniques. The light cone is the surface $a = 0$, $\phi = \pm\infty$. It is therefore rather hard to apply the boundary conditions of the previous section there, but what I shall do is use the semiclassical approximation to the path integral to estimate the form of the solution on lines of constant large positive or negative ϕ.

In order to apply the semiclassical approximation, one needs to know the solutions to the classical euclidean field equations. One can express the metric in the form

$$ds^2 = \sigma^2(d\tau^2 + a^2(\tau)\, d\Omega_3^2)\,. \tag{4.10}$$

One is interested in metrics which are compact. Thus a will be zero at some value of τ which can be chosen to be zero. At $\tau = 0$, $d\phi/d\tau$ must be zero and $da/d\tau = 1$. One can then integrate the equations for $a(\tau)$ and $\phi(\tau)$ with these initial conditions

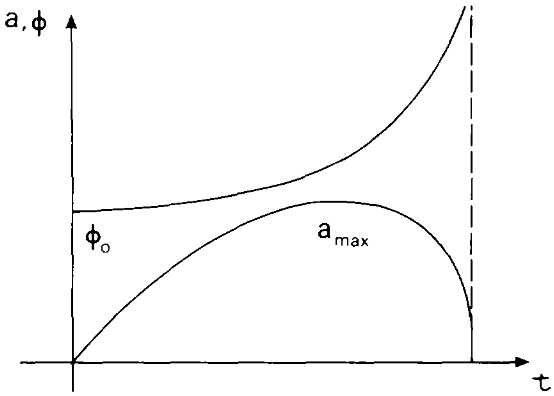

Fig. 2. Graphs of the radius a and the scalar field ϕ for a solution of the euclidean field equations.

and $\phi=\phi_0$ at $\tau=0$. If $\phi_0>0$, $\phi(\tau)$ and $\mathrm{d}\phi/\mathrm{d}\tau$ will monotonically increase with τ and $\mathrm{d}a/\mathrm{d}\tau$ will monotonically decrease (see fig. 2). What happens is that the positive energy density of the $m^2\phi^2(\tau)$ term in the action behaves like a positive cosmological constant and causes the 4-geometry to have positive curvature. The 3-sphere radius $a(\tau)$ rises to a maximum $a_{\max}$ which is a monotonically decreasing function of ϕ_0. The radius $a(\tau)$ then decreases and becomes zero at a singularity of the 4-geometry. If $\phi_0 \gg 1$, i.e. $\ddot{\phi} \gg m_{\mathrm{p}}$, then $\phi(\tau)$ does not increase much by the time $a(\tau)$ reaches $a_{\max}$ and $a_{\max} \simeq 1/m\phi_0$ (see fig. 3).

In order to estimate the wave function $\Psi(a, \phi)$, one looks for a solution $(a(\tau), \phi(\tau))$ which matches the given values of a, and ϕ at some value $\tau=\tau_0$. For sufficiently small values of a and ϕ there will be a unique trajectory $a(\tau)$, $\phi(\tau)$ which passes through the given values of a and ϕ and which does not cross any other trajectory (see fig. 3). For such a, ϕ the path integral for the wave function $\Psi(a, \phi)$ will be dominated by the contribution from the euclidean solution represented by the trajectory $a(\tau)$, $\phi(\tau)$, i.e.

$$\Psi(a, \phi) \approx N_0 \exp\left(-\tilde{I}[a(\tau), \phi(\tau)]\right). \tag{4.11}$$

If two neighbouring trajectories $a(\tau)$, $\phi(\tau)$ in the a, ϕ plane intersect at a point a_1, ϕ_1, the path integral for the wave function $\Psi(a_1, \phi_1)$ will have a zero mode at the euclidean solution $a(\tau)$, $\phi(\tau)$. If a_2, ϕ_2 is a point on the same trajectory at a larger value of τ, the path integral for $\Psi(a_2, \phi_2)$ will have a negative mode at the euclidean solution $a(\tau)$, $\phi(\tau)$. This means that the dominant contribution to the path integral for Ψ will not come from that euclidean solution. Another way of obtaining this result is to notice that the neighbouring trajectories intersect at about the maximum value of $a(\tau)$. Beyond that value of τ, the radius a will be decreasing and the trace of the second fundamental form K will be negative. It is shown in the appendix that a solution with negative K cannot give the dominant contribution to the wave function Ψ.

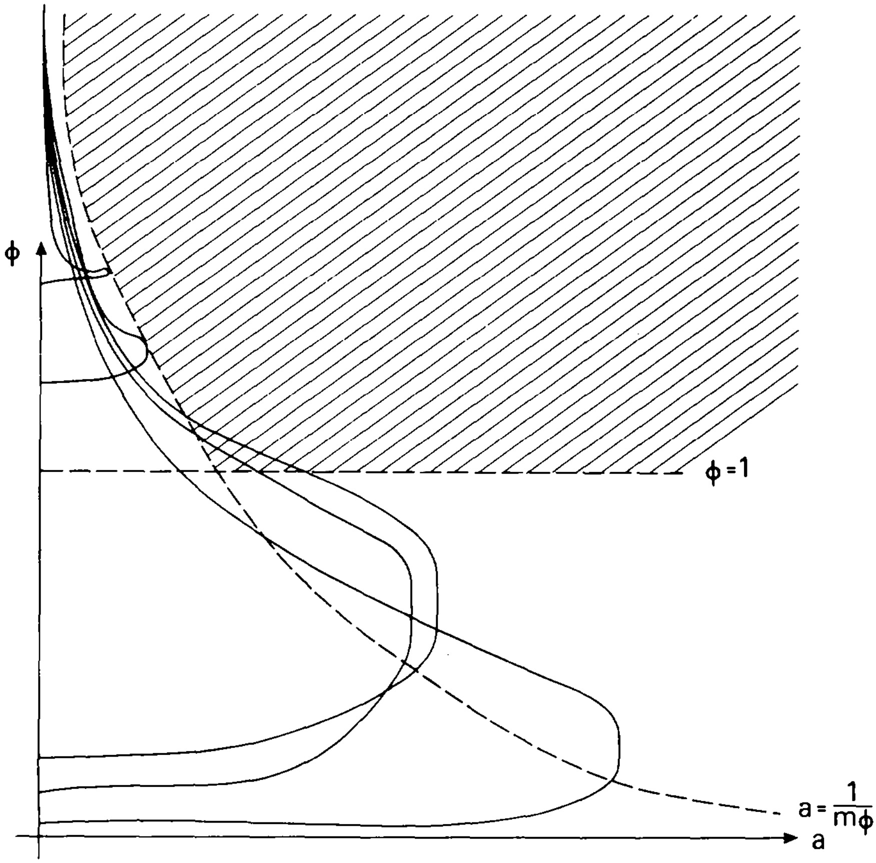

Fig. 3. Solutions of the euclidean field equations in the (a, ϕ) plane. There are no solutions which reach the region $\phi > 1$, $a < (m\phi)^{-1}$ without crossing another solution.

There is a region of the (a, ϕ) plane with $\phi > 1$, $a > (m\phi)^{-1}$ through which there are no euclidean trajectories which have not intersected another trajectory (see fig. 3).* The path integral (2,4) over compact euclidean metrics for the wave function $\Psi(a, \phi)$ for a point (a, ϕ) in that region will not receive its dominant contribution from any real euclidean metric. Instead, one can distort the contour of integration in the path integral so that the dominant contribution comes from a complex solution of the field equations. The action of such a solution will be complex and so the wave function will be oscillatory though it will be real because there will be an equal contribution from the complex conjugate solution.

One can estimate the form of the wave function in the region $\phi > 1$, $a > (m\phi)^{-1}$ by using the K representation $\Phi(K, \phi)$. The semiclassical approximation to the

* Note added in proof: D.N. Page has pointed out that there are trajectories through this region which have not intersected another trajectory. However, they correspond to solutions with large positive actions. They therefore do not give much contribution to the wave function.

wave function $\Phi(K, \phi)$ is

$$\Phi(K, \phi) \approx N_k \sum_i A_i \exp(-B_i)\,, \tag{4.12}$$

where N_k is a normalization factor. B_i are the actions I^k of classical solutions which are bounded by a 3-sphere with the trace of the second fundamental form equal to K and the prefactors A_i are given by the determinants of small fluctuations about the classical solutions. If $\phi \gg 1$ and $K > 0$, the action $I^k(K, \phi)$ is almost the same as that for de Sitter space with $\Lambda = 3m^2\phi^2/\sigma^2$, i.e.

$$I^k(k, \phi) \approx -\frac{1}{3m^2\phi^2}\left(1 - \frac{3k}{(9k^2 + m^2\phi^2)^{1/2}}\right), \tag{4.13}$$

where $k = \frac{1}{9}K\sigma$. The wave function $\Psi(a, \phi)$ can then be obtained by an inverse Laplace transform of $\Phi(k, \phi)$

$$\Psi(a, \phi) \approx \frac{N_k}{2\pi i}\int_\Gamma \mathrm{d}k \exp(ka^3 - I^k)\,, \tag{4.14}$$

where the contour Γ is parallel to the imaginary k axis. The integral (4.14) can be evaluated by the method of steepest descent.

If $a < (m\phi)^{-1}$, the saddle point occurs at a real positive value of k

$$k = \tfrac{1}{3}m\phi\left(\frac{1}{m^2\phi^2a^2} - 1\right)^{1/2}. \tag{4.15}$$

This value of k corresponds to a real euclidean solution $a(\tau)$, $\phi(\tau)$ bounded by the given values of a, ϕ. The wave function is therefore

$$\begin{aligned}\Psi(a, \phi) &\approx N_k \exp(ka^3 - I^k)\\ &\approx N_k \exp(-\tilde{I}(a, \phi))\,,\end{aligned} \tag{4.16}$$

where $\tilde{I}(a, \phi)$ is the action of the solution bounded by the given values of a and ϕ:

$$\tilde{I}(a, \phi) = -\frac{1}{3m^2\phi^2}(1 - (1 - m^2\phi^2a^2)^{3/2})\,. \tag{4.17}$$

If the radius a is greater than $(m\phi)^{-1}$ there will not be a stationary phase point in (4.14) for real values of k. This corresponds to the fact that there is no real euclidean solution. However, if a is only slightly greater than $(m\phi)^{-1}$, one can analytically extend the approximate expression (4.13) for I^k and find stationary phase points at imaginary values of k:

$$k = \pm\tfrac{1}{3}im\phi\left(1 - \frac{1}{m^2\phi^2a^2}\right)^{1/2}. \tag{4.18}$$

It is possible to distort the contour Γ to pass through both of these stationary phase points. The wave function therefore has the form:

$$\Psi(a, \phi) \approx N_k \exp\left(\frac{1}{3m^2\phi^2}\right)\cos((m^2\phi^2a^2 - 1)^{3/2}/(3m^2\phi^2) - \tfrac{1}{4}\pi)\,. \tag{4.19}$$

One can use the behaviour (4.16) and (4.19) of the wave function as the boundary conditions for the Wheeler–DeWitt equation near the light cone in the (x, y) plane. One can then solve the Wheeler–DeWitt equation as a wave equation with these boundary conditions to determine the behaviour of the wave function at other values of a and ϕ. Because the wave function is rapidly oscillating, one can use the WKB method. One writes the wave function in the form

$$\Psi(a, \phi) = C(a, \phi) \cos[S(a, \phi)], \tag{4.20}$$

where $S(a, \phi)$ is a rapidly varying phase and $C(a, \phi)$ is a slowly varying amplitude. The trajectories of ∇S in the (a, ϕ) plane correspond to classical lorentzian solutions $(a(t), \phi(t))$ of the field equations. The semi-classical approximation for large ϕ corresponds to a classical solution which starts with $\phi(t) = \phi_1$, $\mathrm{d}\phi/\mathrm{d}t = 0$, at a minimum radius $a(t) = (m\phi_1)^{-1}$ at $t = 0$ and then expands in a de Sitter-like manner. The mass term tends to make $\phi(t)$ decrease with a timescale $3H/m^2$. However, the expansion rate $H = m\phi_1$. Thus if $\phi_1 > 1$ the universe will expand by a factor of order $e^{3\phi_1^2}$ before ϕ decreases significantly. For large ϕ_1 this gives a long inflationary period.

Eventually $\phi(t)$ decreases to zero and starts oscillating in time. The universe goes over to a matter-dominated phase with $a(t) \propto t^{2/3}$. The universe will expand to a maximum radius of order $\exp(6\phi_1^2)/m\phi_1$ and then recollapse. A classical solution that starts with a general value of ϕ_1 will recollapse to a singularity. It will be time symmetric about $t = 0$ but not in general time symmetric about $t = t_1$, the time of maximum expansion. However, for discrete values of the scalar field ϕ_1 the field $\phi(t)$ is zero at $t = t_1$. In these cases, the solution will be time symmetric about $t = t_1$ but with $\phi(t)$ replaced by $-\phi(t_1 - t)$. These solutions will therefore oscillate indefinitely. There are also periodic solutions which are time symmetric about either the minimum or the maximum radius, but not both and there are oscillating solutions which are not time symmetric at all and which are aperiodic. The orbits of ∇S in the (a, ϕ) plane correspond to these non-singular oscillating solutions. One thus obtains a wave function $\Psi(a, \phi)$ which can be interpreted as a superposition of wave functions corresponding to oscillating universes in the classical limit. Note that the wave function does not represent a single classical universe but a whole ensemble of classical universes which expand out to different maximum radii.

The wave function $\Psi(a, \phi)$ will grow exponentially in a region near the y-axis in which $a^2m^2\phi^2 < 1$. This part of the wave function can be interpreted as corresponding to a flat euclidean 4-geometry. If one wishes the wave function to be normalizable, one could introduce a tiny Λ term. This would give a λa^4 term in the Wheeler–DeWitt equation (4.8) where $\lambda = \frac{1}{3}\sigma^2\Lambda$. It would cut off the exponential growth of Ψ at very large values of a. Superimposed upon this exponentially growing wave function will be an oscillating component which corresponds to the periodic solutions described above. The oscillating component in the wave function should be interpreted as corresponding to a lorentzian geometry and the exponentially growing component in the wave function should be interpreted as corresponding to a euclidean geometry. We live in a lorentzian geometry and therefore we are interested really only in the oscillatory part of the wave function.

5. Beyond minisuperspace

In the previous section it was shown that the quantum state defined by the path integral over all compact euclidean metrics was a reasonable representation of the observed universe in the case of a simple minisuperspace model in which nearly all the degrees of freedom of the gravitational and matter fields were frozen out. The question naturally arises as to whether the quantum state would still be a good description of the universe if one included all those extra degrees of freedom or whether it was a misleading result produced by the truncation.

The important feature of the wave function in the model of the previous section was that it was an oscillatory rather than an exponential function of the coordinates (a, ϕ) in a certain region of minisuperspace. This allowed it to be interpreted as corresponding to lorentzian 4-geometries in the classical limit. The reasons that the wave function oscillated were:

(1) the Wheeler–DeWitt equation was a hyperbolic wave equation on minisuperspace;

(2) there was a region of minisuperspace which could not be reached by real euclidean solutions of the field equations without negative modes. This meant that the dominant contribution to the path integral for the wave function came from complex solutions of the field equations.

The "metric" $-G_{ijkl}$ in the full Wheeler–DeWitt equation that multiplies the second functional derivatives of the wave function $\partial^2\Psi/\partial h_{ij}\partial h_{kl}$ has a signature $(+-----)$ at each point of the 3-manifold S. It might therefore appear that the gravitational part of the Wheeler–DeWitt equation was ultra-hyperbolic, having a signature with one plus and five minuses at each point of S. However, all but one of these infinite number of pluses can be regarded as arising from the gauge invariance of the theory. Gauge invariance manifests itself in two ways: there is the obvious freedom to make coordinate transformations in the 3-surface S. This is reflected in the equation (2.13) which states that the wave function is unchanged by coordinate transformations in S. The other gauge invariance, which is less apparent, is the freedom to choose S at different positions in the 4-geometry $g_{\mu\nu}$. Given a 4-geometry, one could, locally at least, restrict the positions of S in it to a one parameter family of non-intersecting 3-surfaces by imposing one gauge condition on the 3-metric h_{ij} such as ${}^3\mathrm{R} = \text{constant}$. With this restriction on h_{ij}, the gravitational part of the Wheeler–DeWitt equation is hyperbolic: the one plus sign corresponds to the overall size of the 3-surface S. The matter field variables in the Wheeler–DeWitt equation all have negative signature provided that they obey reasonable positive energy conditions. Thus the full Wheeler–DeWitt equation can be regarded as a hyperbolic wave equation, in a certain sense.

The other important property of the model of the previous section, that there should be a region of superspace or minisuperspace which cannot be reached by real euclidean solutions, seems to depend on the presence of fields with non-zero rest-mass such as a massive scalar field. Investigations of minisuperspace models containing a conformally invariant scalar field [2] or empty homogeneous anisotropic models [14] indicate that there is always a real euclidean solution of the field equations which gives the dominant contribution to the path integral for the wave function. The wave function is therefore an exponential rather than an oscillatory function of the superspace variables and so corresponds to a euclidean rather than a lorentzian 4-geometry in the classical limit.

In the massive scalar field minisuperspace model the potential $-a^2+a^4m^2\phi^2$, the non-derivative term in the Wheeler–DeWitt equation (4.8), changes sign from negative to positive on a surface in the (a, ϕ) plane which is "spacelike" for large $|\phi|$, i.e. which has a negative signature with respect to the second derivative terms in the Wheeler–DeWitt equation. Points below this surface can be reached by euclidean solutions, so the wave function is exponential in this region. The wave function on the surface $-a^2+a^4m^2\phi^2=0$ is of the form $\Psi \approx N_0 \exp(1/3m^2\phi^2)$. A solution of the wave equation with this initial data must oscillate in a region of positive potential. This is why one obtains an oscillatory wave function which can be interpreted in terms of lorentzian 4-geometries in the classical limit. This feature of the massive scalar field model does not depend on the truncation to a finite number of degrees of freedom by the imposition of symmetries: it would also be present in a model in which the 3-metric h_{ij} and the scalar field ϕ were fully general and did not have any symmetries. Thus one would expect that the solution of the Wheeler–DeWitt equation on the full superspace of the massive scalar field model would be oscillatory if it obeyed the boundary condition that it was given by a path integral over compact metrics.

Under a scale transformation $g_{\mu\nu} \to \Omega^2 g_{\mu\nu}$, $\Omega =$ constant, the gravitational action transforms as $\tilde{I} \to \Omega^2 \tilde{I}$. This means that the potential terms in the gravitational part of the Wheeler–DeWitt equation cannot change sign under a "timelike" motion in superspace that corresponds simply to a scale transformation of the 3-metric h_{ij}. In other words, the gravitational part of the potential can change sign only on a "timelike" surface in superspace. The action of a conformal invariant field, such as the electromagnetic field or a conformally invariant scalar field, is unchanged under a scale transformation. Thus the potential terms in the Wheeler–DeWitt equation arising from such fields can also change sign only on "timelike" surfaces in superspace. In fact, under reasonable positive energy assumptions, the potential terms produced by such fields will always be positive. Thus the potential cannot change from negative to positive in a model that contains only zero, rest-mass fields, either conformally invariant fields or the gravitational field. One would therefore not expect minisuperspace models that contained only zero rest-mass fields to produce oscillatory wave functions.

On the other hand, if one includes all the infinite number of gravitational and matter degrees of freedom, one will generate R^2 and C^2 terms in the effective action. These terms will be finite in the case of a supergravity theory in which all the divergences cancel or in the case that gravity provides its own cutoff (possibilities (i) and (ii) of sect. 1) and will be infinite but renormalized in the case of a higher derivative theory (possibility (iii)). Such terms in the effective action behave like massive scalar or spin-2 particles. One might therefore hope that they would give rise to oscillatory wave functions. This has been verified, at least in the case of a minisuperspace model with an R^2 term. This will be described in a forthcoming paper [8]. It may well be therefore that the observed universe owes its existence to quantum gravitational effects.

I am grateful for discussions with J.B. Hartle, A.D. Linde, J. Luttrell and I.G. Moss.

Appendix

THE DOMINANT CONTRIBUTION TO THE PATH INTEGRAL

The wave function Ψ is given by the inverse Laplace transform

$$\Psi[h_{ij}, \phi] = \int_{\Gamma} \mathrm{d}\left(\frac{m_{\mathrm{P}}^2}{24\pi i} K\right) \exp\left(\frac{m_{\mathrm{P}}^2}{12\pi} \int \mathrm{d}^3x\, h^{1/2} K\right) \Phi[\tilde{h}_{ij}, K, \phi], \tag{A.1}$$

where the wave function Φ in the K representation is

$$\Phi[\tilde{h}_{ij}, K] = \int \mathrm{d}[g_{\mu\nu}] \exp(-I^k[g_{\mu\nu}, \phi]). \tag{A.2}$$

One expects the dominant contribution to Φ to come from metrics near a solution $\tilde{g}_{\mu\nu}$ of the field equations with the given values of $\tilde{h}_{ij}$ and K on S. The dominant contribution to Ψ will then come from the saddle point in the contour integral (A.1). At the saddle point the exponent of e in (A.1)

$$\frac{m_{\mathrm{P}}^2}{12\pi} \int_{\mathrm{S}} \mathrm{d}^3x\, h^{1/2} K - I^k[\tilde{g}_{\mu\nu}, \phi] \tag{A.3}$$

will be an extremum under small variations of K:

$$\frac{m_{\mathrm{P}}^2}{12\pi} h^{1/2} = \frac{\delta I^k}{\delta K}. \tag{A.4}$$

Eq. (A.4) implies that $\tilde{g}_{\mu\nu}$ is a solution of the field equations with the boundary conditions h_{ij}, ϕ on S. However, if $\tilde{g}_{\mu\nu}$ is a real euclidean metric, then in order for it to be a saddle point in (A.1), the exponent (A.3) must be, not only an extremum, but a minimum on the real K-axis at each point of S, i.e.

$$\frac{\delta h^{1/2}}{\delta K} < 0, \tag{A.5}$$

where $h^{1/2}$ is regarded as a function of the boundary data $(\tilde{h}_{ij}, K, \phi)$ on S that determine the solution $\tilde{g}_{\mu\nu}$.

Consider the case $\Lambda = 0$ and $\phi = 0$ on S. Under a scale transformation

$$\delta K = -\varepsilon K\,, \qquad \delta h^{1/2} = 3\varepsilon h^{1/2}\,. \tag{A.6}$$

If the 3-metric $\tilde{h}_{ij}$ is conformally flat and the 4-metric $g_{\mu\nu}$ is asymptotically euclidean, the solution $\tilde{g}_{\mu\nu}$ will be flat euclidean space outside a round 3-sphere S. For this solution K will be constant and negative. Thus under a scale transformation

$$\frac{\delta h^{1/2}}{\delta K} > 0\,. \tag{A.7}$$

This shows that the metric $\tilde{g}_{\mu\nu}$ cannot be a saddle point in the contour integral (A.1) and so cannot give the dominant contribution to the wave function Ψ. One can now consider a continuous variation of the values of Λ and the matter field configuration ϕ on S from zero. Under such a variation the left-hand side of (A.7) will change but it cannot become negative everywhere on S unless there is some Λ and ϕ for which the left-hand side of (A.7) is either infinite or zero everywhere on S. In the first case, the solution $\tilde{g}_{\mu\nu}$ will have a zero mode for fixed $\tilde{h}_{ij}$, K and ϕ on S, and in the second case it will have a zero mode for fixed h_{ij} and ϕ. If one continues the variation until the left side of (A.7) is negative, the solution $\tilde{g}_{\mu\nu}$ will have a negative mode for the given boundary conditions. In the first case, this indicates that $\tilde{g}_{\mu\nu}$ does not provide the dominant contribution to the wave function Φ and hence does not provide the dominant contribution to Ψ. In the second case, the direct representation of Ψ in terms of a path integral shows that $\tilde{g}_{\mu\nu}$ does not provide the dominant contribution to Ψ.

References

[1] S.W. Hawking, in Astrophysical cosmology, Pontificiae Academiae Scientarium Scripta Varia, 48 (1982) 563

[2] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960

[3] A.H. Guth, Phys. Rev. D23 (1981) 347

[4] A.D. Linde, Phys. Lett. 108B (1982) 389

[5] S.W. Hawking and I.G. Moss, Phys. Lett. B110 (1982) 35

[6] A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. 48 (1982) 120

[7] A.H. Guth and S.Y. Pi, Phys. Rev. Lett. 49 (1982) 1110

[8] S.W. Hawking and J. Luttrell, paper in preparation

[9] S.W. Hawking, Quantum cosmology, Les Houches lectures (1983), to be published

[10] B.S. DeWitt, Phys. Rev. 160 (1967) 113

[11] C.J. Isham and J.E. Nelson, Phys. Rev. D10 (1974) 3226

[12] W.E. Blyth and C.J. Isham, Phys. Rev. D11 (1975) 768

[13] I.G. Moss and W.A. Wright, The wave function of the inflationary universe, Newcastle Univ. preprint (1983)

[14] J. Luttrell, private communication