1985
TH-71
1985

QUANTUM COSMOLOGY

James B. Hartle

Department of Physics, University of California
Santa Barbara, CA 93106

Lectures delivered at the Theoretical Advanced Study Institute in Elementary Particle Physics, Yale University, June 10-14, 1985.

QUANTUM COSMOLOGY

James B. Hartle
Department of Physics, University of California
Santa Barbara, CA 93106

1. INTRODUCTION

The traditional enterprise of cosmology has been to construct a model of the universe which agrees with our observations on the largest scales and which, when evolved backwards according to the laws of fundamental physics, gives a consistent historical picture of how the universe came to be the way it is today. Our observations tell us that the universe consists of matter and radiation. The matter that we see in galaxies is distributed roughly homogeneously and isotropically on the largest scales. The cosmic background radiation, in which we see a picture of the universe at an early stage, is remarkably isotropic. As a first approximation, we are thus led naturally to the Friedman-Robertson-Walker cosmological models in which the symmetries of homogeneity and isotropy are enforced exactly. Evolved backward in time using Einstein's gravitational theory and the laws of microscopic physics these models provide a consistent history of the universe. Among other things, they describe the evolution of the background radiation, the origin of the primordial elements, the evolution of the fluctuations which became the galaxies, and perhaps the origin of the baryons. The initial condition implied by the extrapolation is an early state in which the matter is in thermal equilibrium with high temperature and density, distributed homogeneously and isotropically but containing the seeds of condensations later to become galaxies.

The Friedman-Robertson-Walker models are successful and they are simple. Their success and simplicity raise the issue of why does the universe have the properties it does? Can we explain the Friedman-Robertson-Walker models? This is a very different kind of issue from the essentially descriptive questions traditionally asked in cosmology. In effect one is asking for a theory of initial conditions. These lectures are about one approach to this problem: the search for a theory of initial conditions in the application of quantum gravity to cosmology - in two words they are about quantum cosmology.

The lectures are not intended as a review of all models of the universe which involve quantum mechanics or even of those which deal directly with the issue of initial conditions. The subject, although already large, is not yet connected enough to make such a review feasible in the space available.[1] Rather we shall explore a specific proposal for the quantum state of the universe developed by Stephen Hawking and his collaborators.[2] In the process we shall be able to review much about the general issues.

To state the proposal for the quantum state of the universe we shall need some of the framework of quantum gravity. This we describe in Section 3. We shall develop the idea and compare its predictions with observations in Section 6, but, in order to know where we are going, we shall first review the observations we hope to explain in Section 2.

2. THE UNIVERSE TODAY AND THE PROBLEM OF ITS INITIAL CONDITIONS

2.1 Observations

The variety and detail of the observations now available which bear on the structure of our universe in the large is one of the most impressive achievements of contemporary astronomy. The relationships between these observations are complex and deriving an understanding of the universe in the large from them is a complex theoretical story. Emerging from this analysis, however, is a picture of striking simplicity on the largest scales. In this section we shall summarize this picture in a few "observational facts" and briefly indicate the nature of the supporting evidence for each one. These are the facts one seeks to explain in a theory of initial conditions. We can only adumbrate the arguments for these observations here and cannot hope to give a complete list of references to them. For greater detail and references the reader is encouraged to consult the many reviews of this subject.[3)]

Fact (1). Spacetime is four dimensional with Euclidean topology

This is so built into our fundamental physics that we usually take it as granted. It is important to remember however that all aspects of geometry have an observational basis.

Fact (2). The universe is large, old and getting bigger

At moderate distances galaxies recede from each other according to Hubble's law

$$\begin{pmatrix} \text{velocity} \\ \text{of recession} \end{pmatrix} = H_0 \text{ (distance apart)} \quad , \qquad (2.1)$$

where H_0 is somewhere between 40 and 100 (km/sec)/Mpc.

[A pc is 3.09×10^{18} cm. A Mpc is 10^6 pc.] Inverted, Hubble's law gives us a connection between distance and redshift. Since H_0 is uncertain distances are often quoted as a multiple of h^{-1} where h is $H/[100(\text{km/sec})/\text{Mpc}]$. The background radiation originates at distances of order $c/H_0 \sim 3000\ h^{-1}$ Mpc (the Hubble distance) from us and at times of order $1/H_0 \sim 10^{10}$ yrs. (the Hubble time) ago. These are the largest scales which are directly accessible to observation today. It is perhaps a trivial observation, but these are not the scales of elementary particle physics.

Fact (3). The universe contains matter and radiation distributed homogeneously and isotropically on the largest scales

Direct evidence for the homogeneity of the universe is hard to come by. Ideally one would like to make a three dimensional map showing the distribution of galaxies and this involves measuring distances. Such surveys have been made but only out to limited distances (~100 Mpc). The test that probes homogeneity on the largest scales is the oldest - counts of galaxies vs. limiting flux. One can easily show that if there are several populations of objects distributed uniformly in flat three dimensional space, then the number of objects counted with flux f greater than some limiting flux, f_0, should vary with f_0 as

$$N(f > f_0) \propto f_0^{-3/2} \quad , \tag{2.2}$$

with calculable corrections for spatial curvature. Modern surveys[4] which probe out to depths comparable to the Hubble distance yield approximate agreement with this law.

If we accept the Copernican principle that we are not at a preferred position in the universe, and there is no evidence that we are, then evidence for isotropy becomes

evidence for homogeneity. Evidence for the isotropy of the universe comes from angular surveys of its major constituents.

Most directly there are the galaxies. A plot on the sky of the Shane-Wirtanen catalog of the 10^6 galaxies contained in an effective depth of several hundred Mpc is as close as we can come to a picture of the universe today.[5] It is roughly isotropic on large scales but clearly exhibits structure. A quantitative measure of the isotropy is the galaxy-galaxy angular correlation function. This is the excess probability for finding a second galaxy at some fixed angle from any given one. It is conveniently quoted in terms of a spatial correlation function $\xi(r)$ which would produce the same result assuming homogeneity. $\xi(r)$ is about unity at 7 h^{-1} Mpc and decreases to a few 1/10ths by 20 h^{-1} Mpc. The galaxy distribution is thus essentially isotropic at large scales.

Radio sources are distributed across the sky in an essentially uniform way. The diffuse X-ray background is isotropic to a few percent at angular scales of 5^o. Since a significant fraction of this radiation comes from distant quasars this becomes a test of isotropy on large scales.

The temperature distribution of the 2.7^oK cosmic background radiation provides the most accurate test of the isotropy of the universe and the one which probes this feature on the largest scales and therefore the earliest times. The anisotropy in the temperature, $\Delta T/T$, has been well measured on a number of different angular scales.[6] There is a purely dipole anisotropy which is attributable to the motion of our solar system with respect to the rest frame of the radiation. If this is subtracted out no anisotropy has been detected in the residual component on any scale. The current best limit on the quadrupole anisotropy[7,8], for example, is $(\Delta T/T)_{quadrupole} \lesssim 7 \times 10^{-5}$.

Figure 1 shows this graphically. It is a map of the sky with the dipole anisotropy subtracted out, produced by Lubin and Villela[8] at 3mm. where the background radiation dominates all other sources. It is thus in effect a snapshot of the universe at an age of 10^5 years when the background radiation was emitted. It is essentially featureless.

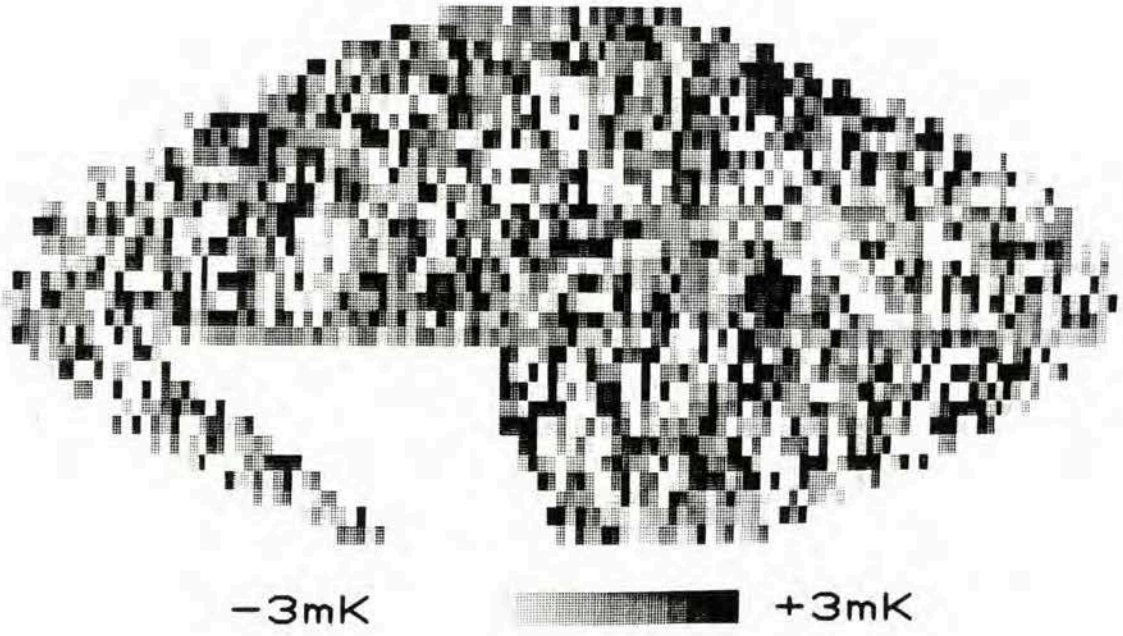

Figure 1. The sky at 3mm. This figure shows the map of the sky at 3mm. observed by Lubin and Villela[8] with the dipole anisotropy removed. The shading in the rectangles, each a few degrees on a side, indicates the temperature deviation from the mean. Since the background radiation is the dominant source of radiation at this wavelength, this is essentially a picture of the universe 300,000 years after the big bang and it is remarkably isotropic.

The observed approximate homogeneity and isotropy on large scales suggest that the Friedman-Robertson-Walker models, in which these symmetries are enforced exactly, should give a good first approximation to the dynamics of the universe. The metric of a spacetime geometry with homogeneous and isotropic spatial sections can, in suitable coordinates, be described by the line element

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2\right] \quad , \tag{2.3}$$

where $d\Omega_2^2$ is the metric on the unit two sphere. The spatial geometry is open with negative curvature if $k = -1$, open and flat if $k = 0$ and closed with the geometry of a three-sphere if $k = +1$.

All the geometrical information about the dynamics of the universe is contained in the scale factor $a(t)$. Einstein's equation for perfect fluid matter with energy density ρ and cosmological constant Λ implies

$$\left(\frac{\dot{a}}{a}\right)^2 = -\frac{k}{a^2} + \frac{\Lambda}{3} + \frac{8\pi G}{3}\rho \quad . \tag{2.4}$$

This equation plus the constituative relations of the matter are enough to extrapolate the dynamics of the universe forward and backward in time given the constants k and Λ and the present values of a and ρ or equivalently the present values of ρ and $\dot{a}/a$. The present value of $\dot{a}/a$ is the Hubble constant H_0. It is uncertain because the extragalactic distance scale is uncertain, but most determinations fall in the range $40 - 100$ (km/sec)/Mpc.

Eq. (2.2) shows that, were $\Lambda = 0$, the density today would have to be greater than the critical value

$$\rho_{crit} = \frac{3H_0^2}{8\pi G} \quad , \tag{2.5}$$

to have a closed (k = +1) universe. It has become conventional to quote ρ and even Λ in terms of their dimensionless ratios to this critical density. For example, the present density ρ_0 defines the ratio $\Omega_0 = \rho_0/\rho_{crit}$ and the cosmological constant the ratio $\Omega_\Lambda = (\Lambda/8\pi G)/\rho_{crit}$. We now briefly describe the observational evidence for each of these quantities.

The density in luminous matter, found essentially by counting galaxies whose redshifts and therefore distances are known, corresponds to an Ω of about .01. There is considerable evidence, however, that the universe contains significant amounts of non-luminous matter. The rotational velocity of a galaxy at a given radius from its center can be used to estimate the mass interior to that radius. These velocities do not fall with radius as would be predicted from the density of luminous matter in galaxies. They remain constant as far out as can be measured indicating the presence of a dark component perhaps 10 times more massive than the luminous one. Dynamical analysis of the infall of galaxies towards the center of the Virgo supercluster (of which we are an outlying member) argue for $\Omega_0 \sim .3$ if there is no dumped matter which is non-luminous.[9] Models of the nucleosynthesis of deuterium in the early universe together with its measured abundance today suggest that the Ω corresponding to the density of baryons today is about .1. These arguments suggest a value of Ω_0 of a few tenths. They cannot rule out, however, a larger Ω_0 if there is non-luminous, non-baryonic matter which is not clustered with the galaxies or if there is matter clustered like the galaxies but non-luminous.

It is difficult to measure Ω_Λ from anything other than direct observation of the cosmological expansion. However, it cannot be many orders of magnitude larger than unity or it would imply observable deviations from Newtonian dynamics

in clusters of galaxies.[10] Thus there is no direct evidence that $\Lambda = 0$ today. Even an Ω_Λ of 1, however, corresponds to a cosmological constant which is very small on the scale of the Planck mass $m_p = (\hbar c/G)^{1/2}$

$$\Lambda \approx 8.8 \times 10^{-122} m_p^2 \Omega_\Lambda h^2 \quad . \tag{2.6}$$

The available information on the density of energy in the universe is not enough to tell us whether the spatial geometry of the universe is open or closed. It is, however, close to the flat geometry which is the borderline between the two. We might therefore summarize this information in a fourth "observational fact":

Fact (4). The spatial geometry is approximately flat

Fact (5). The spectrum of density fluctuations

The universe is not exactly homogeneous and isotropic. Matter in galaxies is very clumped as measured by the ratio of the difference in their density to the mean density, $\delta\rho/\rho$. The evidence from the background radiation is that earlier the universe was much smoother. The present large scale structure arose from this earlier, smoother distribution through gravitational attraction. At present, direct observations of the background radiation give only upper limits on fluctuations both as to amplitude and spectrum. The amplitude required for those scales where $\delta\rho/\rho \sim 1$ now (superclusters of galaxies) may be found by extrapolating backwards in time using linear perturbation theory and is $(\delta\rho/\rho) \sim 10^{-4}$ at the time the background radiation was emitted. This is consistent with the upper limits. Information on the spectrum can be obtained by assuming appealing candidates at decoupling and extrapolating them forward non-linearly and comparing with the existing large scale structure. The spectrum such that all fluctuations have the same amplitude at the time their

scales coincide with the Hubble scale, called the Zel'dovich spectrum, is a popular candidate consistent with all current observations.

Fact (6). The entropy of the universe is low and increasing in the direction of expansion

Today, essentially all of the entropy of matter is in the background radiation. The ratio of the density of entropy s to the density of baryons n_b is

$$s/kn_b \sim 10^9 \quad , \tag{2.7}$$

so that the total entropy within a Hubble distance is approximately $S/k \sim 10^{87}$ (the word approximately refers to the exponent!). This is a large number but a small fraction of the entropy which could be obtained by clumping all the matter within the Hubble distance into a black hole.[11] A black hole of mass M has entropy $4\pi kGM^2/(\hbar c)$ so that with a reasonable estimate for the mass within the horizon

$$S/k \sim 10^{120} \quad . \tag{2.8}$$

The 33 orders of magnitude discrepancy between fact and possibility is another way of saying that the universe is still in a reasonably well ordered state. Entropy is increasing and even on the largest scales we seem to see a steady progression from order when the universe is small to disorder when it is large.

We, of course, have more information about the large scale features of the universe than can be summarized in the above six cosmological facts. We observe specific abundances for the elements, a baryon-antibaryon asymmetry, the thermal spectrum of the background radiation and so on. The above list, however, contains those features whose origin is to be found in the earliest stages of our universe.

2.2 Initial Conditions

In most problems in physics we divide the universe up into two parts, the system under consideration and the rest. We use the local laws of physics to solve for the evolution of the system. For example we use Maxwell's equations and Newton's laws of mechanics to predict the evolution of a plasma. The local laws of physics require boundary conditions: sometimes initial conditions, sometimes spatial boundary conditions, sometimes radiative boundary conditions, and often a combination of these. These boundary conditions are set by the physical conditions of those parts of the universe which are not part of the physical system under consideration. There are no particular laws determining these conditions, they are specified by observations of the rest of the universe. The situation is different in cosmology. Boundary conditions are still required to solve the local laws governing the evolution of the universe. They are needed, for example, to solve Einstein's equation (2.2). There is, however, no "rest of the universe" to pass their specification off to. If there is a general specification of these initial conditions it must be part of the laws themselves.

If we extrapolate the Friedman-Robertson-Walker models backward in time we can find initial conditions which give rise to the present universe. What attitude are we to take to these initial conditions? A number of attitudes have been taken. Many of them are summarized in the following four rough categories.

Attitude 1: That's the way it is.

The universe might have been in any one initial state as well as any other. It happens that the one it is in is homogeneous and isotropic on the scales we observe. That's as far as physics can go. Its not the proper subject of

physics to explain these initial conditions only to discover what they were.

This is a reasonable but not very adventurous attitude. It certainly has no predictive power concerning what we will see when with increasing time we are able to observe larger and larger regions of the universe. I believe we will only be able to say it is correct when all attempts to explain the initial conditions have failed.

Attitude 2: The conditions which determine the universe are not initial conditions but the fact that we exist.

This attitude is related to the set of ideas called the anthropic principle.[12] The universe must be such as to allow galaxy condensation, star formation, carbon chemistry and life as we know it. This is indeed a restriction on the structure of the universe. Perhaps, if one were given a choice of three or four very different cosmologies one could identify our own using the anthropic principle. As stressed by Penrose,[11] however, the anthropic principle does not seem strong enough to single out the observed universe from among all possibilities. Suppose, for example, the sun had been located in a cloud near the galactic center and we had not been able to make observations of the large scale structure. Would we have been able to predict the large scale homogeneity and isotropy using the anthropic principle?

Attitude 3: Initial conditions are not needed - dynamics does it all.

The idea is that interesting features like the large scale homogeneity and isotropy will arise from any reasonable initial conditions through the action of physical processes over the course of the universe's history. Even if it started in an inhomogeneous and anisotropic state the universe would evolve towards a homogeneous and isotropic

one over the scales we can observe it. This is an attractive idea not least because we can achieve determinism with the existing dynamical laws of physics. A variety of physical mechanisms have been proposed to implement this idea beginning with the work of Misner and his co-workers.[13] The most successful mechanism is inflation.[14]

Any dynamical explanation of the large scale structural features has to confront the problem of horizons. We can illustrate this idea for a $k = 0$ model by rewriting the line element (2.3) in terms of a conformal time coordinate η such that $dt = ad\eta$. Then

$$ds^2 = a^2(\eta)[-d\eta^2 + dr^2 + r^2 d\Omega_2^2] \quad , \qquad (2.9)$$

and radial light rays move on 45° lines in an (η, r) plane. Two events are in causal contact if their past light cones intersect each other before they intersect the big bang (Figure 2). Only if two events are in causal contact could they be both influenced by a common event in their past. The horizon size at any time is the largest distance over which two events could have been in causal contact up to that time. As measured in the comoving coordinate r, the horizon radius is

$$r_H = \int_0^\eta d\eta = \int_0^t \frac{dt}{a(t)} \quad . \qquad (2.10)$$

Clearly it grows with passing time.

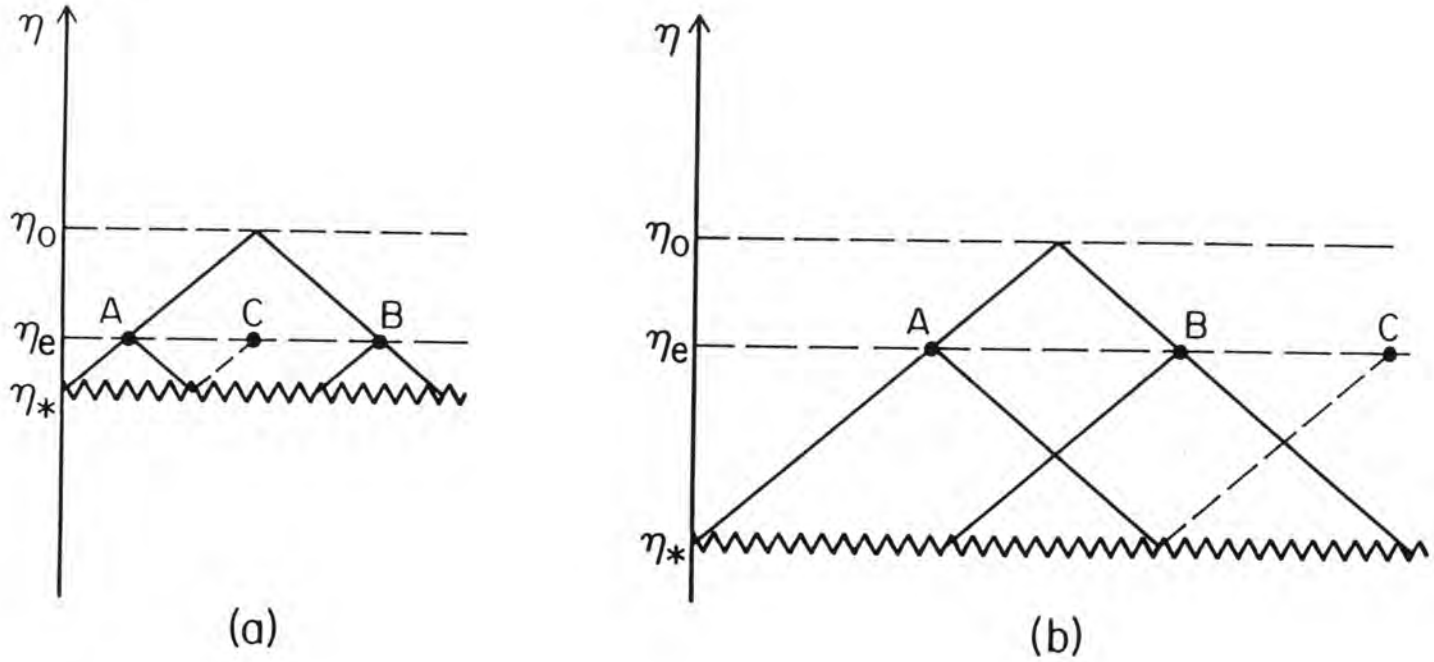

Figure 2. Two possible causal structures. In these diagrams η_0 is the present (conformal) time, η_e is the time of decoupling and η_* the time of the big bang. The history from decoupling to the present is the same in both figures and the times η_0 and η_e have been chosen to correspond. The early histories are different. Were the history like that of (a) two events A and B would not have been in causal contact since the big bang. The size of the observed universe at decoupling AB is larger than the horizon size AC. For a history like that of (b) the horizon size is larger than the size of the observed universe and A and B can be in causal contact.

If we are to have a dynamical explanation of the remarkable isotropy 10^5 years after the big bang, then the different regions from which the radiation was emitted must have been in causal contact. Whether they were able to communicate depends on the history of the universe prior to the time the radiation was emitted i.e. whether there was time enough since the big bang for them to do so. A prerequisite for any dynamical explanation of the observed isotropy of the background radiation is that the horizon size at decoupling be larger than the size of the universe we can observe then.

The most naive extrapolation of the history of the universe to times earlier than decoupling is to assume that the spacetime geometry remains approximately homogeneous and isotropic (the Friedman-Robertson-Walker model), that the matter energy density is dominated by the density of approximately free radiation, and that the evolution is governed by Einstein's equation. With this early history only regions at the time of decoupling now subtending a few degrees on the sky could have been in causal contact since the big bang and no dynamical explanation of the observed isotropy would be possible. This extrapolation, however, is too naive. The horizon can be much larger than the observable universe in models where the geometry is significantly anisotropic before decoupling,[15] as a consequence of quantum effects at the Planck epoch[16] or as a consequence of an inflationary de Sitter-like expansion arising from a matter phase transition at the GUT time. Since this is an important idea it is worthwhile digressing to discuss it briefly.

If the vacuum energy represented by the cosmological constant dominated all other sources of stress energy in Einstein equations the solution would be a geometry called de Sitter space. In particular the solution is

$$ds^2 = -dt^2 + \frac{\cosh^2 Ht}{H^2} d\Omega_3^2 \quad , \qquad (2.11)$$

where $H^2 = \Lambda/3$ and $d\Omega_3^2$ is the metric on the three sphere. de Sitter space is the geometry of the surface of a Lorentz hyperboloid in a 5-dimensional flat, Lorentz signatured spacetime. It contains neither a big bang nor a big crunch. The geometry is non-singular. The spatial three spheres collapse from infinite radius down to a minimum radius

$$a_{min} = H^{-1} = (3/\Lambda)^{1/2} \quad , \qquad (2.12)$$

and then re-expand to infinity. The expansion is exponential - inflationary. During an inflationary expansion the horizon grows at an exponential rate over what we would have guessed from the naive extrapolation. For example, suppose we extrapolate back to a time t_e according to the usual history but before that replace the naive radiation dominated history with an inflationary history back to time t_b. The comoving horizon grows in this epoch according to (2.10) with the revised expansion law. The ratio of the horizon size with inflation to that without is approximately

$$\exp(H(t_e - t_b))/(2Ht_e) \quad , \qquad (2.13)$$

for $t_e >> t_b$. If $1/H$ is set by a particle physics mass scale it is not difficult to overcome the horizon problem.

In the inflationary history, with only conservative assumptions on the matter physics, the horizon at decoupling could be enormously larger than the observable universe. There is thus opportunity for dynamical processes to act to drive the universe towards isotropy and the inflationary expansion itself provides a mechanism

to do this.[17] Further, the region that becomes the observed universe at decoupling is so much smaller than the horizon size at the end of inflation that one might suppose that any reasonable initial conditions which are inhomogeneous on the scale of the horizon will appear smooth on the scale of the observed universe. With an inflationary history no special assumptions on initial conditions are required to explain the observed isotropy. We see the universe as homogeneous and isotropic simply because we do not see a very big part of it.

If the universe is inhomogeneous on a large scale eventually we will find this out. The size of the universe we can observe grows with every second. No dynamical explanation of homogeneity and isotropy can thus ever be an explanation for all time. Eventually we will see outside the horizon and have to face up to the problem of initial conditions. However, if the inflationary history is correct, under even the conservative assumptions mentioned above, we may be able to postpone this discussion for many times the present age of the universe.

Even in the limited region we can observe, no dynamical explanation can ever be a complete explanation for the features of the universe we see. We can always imagine a present state of the observable universe which is highly inhomogeneous. Whatever the laws of geometry and matter, whether they be classical or quantum, whether there are phase transitions or not, these laws can be used to extrapolate this state backward in time and reach some initial condition. No dynamical explanation can, therefore, ever completely exclude a present state of inhomogeneity without some restrictions on the initial conditions. As impressive as they may be in broadening our choice for initial conditions compatible with our present observations, dynamical explanations of these

observations can never be complete.

Attitude 4: There is a law of physics specifying the initial condition.

Specification of the boundary conditions is just as much a law of physics as are the dynamical equations governing their evolution. In this view, the question for physics is whether there exists a compelling, simple, and predictive principle which will single out the initial state of our universe. Any search for such a principle is likely to involve the quantum theory of gravity in an essential way.

Classical cosmological spacetimes can either be singular (for example the Friedman model) or non-singular (for example de Sitter space). Depending on the physics of the matter there could either be a big bang or a small bounce. If the past evolution were essentially classical, it would be very difficult to see how to find a principle for the initial conditions. The principle would necessarily be a principle of classical physics and it is difficult to see, for example, what classical principle would single out the spectrum of density fluctuations we are living with.

If there is a singularity in our past then quantum gravity will certainly be important for its structure. Quantum gravity becomes important when the curvature varies significantly over a Planck length, $(\hbar G/c^3)^{1/2} \sim 10^{-33}$ cm. and curvatures of arbitrarily large size are produced in a classical singularity. The big bang singularity gives one a natural place to make a theory of initial conditions, and its quantum fluctuations are a natural starting point for the present spectrum of density fluctuations.

Singularities are not difficult to arrange in general relativity. They are not, for example, artifacts of high symmetry. The singularity theorems of classical general relativity[18] suggest that if we extrapolate the present universe into the past we will generally encounter a singularity provided the matter physics is such that a positive energy condition is satisfied. (This condition is satisfied, for example, in the radiation dominated, pre-phase transition era of the usual inflationary universe.)

It is in the quantum mechanics of the big bang that we shall look for a law of initial conditions. We must therefore now turn to quantum gravity.

3. QUANTUM GRAVITY

3.1 The Problem of Quantum Gravity

We do not possess today a complete, manageable, satisfactorily interpreted, and tested quantum theory of spacetime for application to cosmology. The difficulties with formulating such a theory occur not only at the level of the traditional issues of quantum field theory: What Lagrangian should be used, that of Einstein's well tested classical theory or another with better short distance behavior for which Einstein's theory emerges as a low energy limit? How does one construct a covariant perturbation expansion for the theory? Is this expansion renormalizable and is the resulting scattering theory unitary? If it is not, can the theory still be sensibly implemented through non-perturbative methods? In a quantum theory of spacetime one also encounters difficulties at a more elementary and more fundamental level. What are the physical degrees of freedom of the theory? What variable plays the role of time so central to Hamiltonian quantum mechanics? How does one label the states and what is their Hilbert space? What is the probability interpretation of these states? Is a theory formulated with one time unitarily equivalent to one formulated with another as general convariance would require? These types of "quantum kinematics" issues whose resolution is familiar in flat space quantum field theory become serious problems in the quantum theory of spacetime. Finally, it is clear that in applications to cosmology one will confront the interpretive issues of quantum mechanics in a striking manner.

I cannot present to you a balanced discussion of all these issues for two reasons. First, I believe that a balanced discussion does not exist. Second, I believe that if it did, I would be hard pressed to present it in

the compass of a few lectures. What I shall do in this section is the following: I shall assume that we want a quantum theory of spacetime and focus on the quantum kinematics of such a theory. I shall by and large neglect the issues of which theory is correct and those of interpretation. The reasons are part prejudicial and part tactical. It is certainly reasonable to explore at the quantum level Einstein's beautiful idea that gravity is geometry. I shall focus on the quantum kinematics of spacetime theories both because it is possible to say something about these questions and because they may be less familiar to those coming from a particle physics perspective. Where dynamics is needed, I shall for the most part illustrate with the theory constructed on Einstein's action both because this is the simplest illustration and because most of the results in quantum cosmology have been obtained in the semiclassical approximation where the deficiencies of this theory are not immediately present. Further, because it is the correct low energy limit one can hope that some qualitative features of any analysis carried out in Einstein's theory would persist in the correct theory of spacetime. Finally, I shall have little to say about what might be called the "words problem" of quantum mechanics. This is the phenomenon that two scientists can agree on the algorithms in quantum mechanics for predicting the result of every experiment, but disagree passionately on the "words" with which they would like to surround these algorithms. Because of this phenomenon it seems to me likely that interpretative issues in quantum mechanics are in part about the people making the interpretations as well as about the real world. It is not that interpretative issues are uninteresting. Like sex, religion, and politics almost everyone likes to discuss them, and they are sometimes crucial to motivation; it is just that they are difficult to lecture about.

3.2 The Problem of Time

Hamiltonian quantum mechanics is the traditional framework for constructing quantum theories of classical systems. The procedure in this framework is familiar: Construct the Hilbert space of physical states, identify the operators corresponding to observables and the Hamiltonian, then calculate dynamics by solving the Schrödinger equation. There are many good reasons for using this canonical framework, the most important being that it is the most mathematically precise of the several possible frameworks.

The canonical formalism has disadvantages. The most important for the quantization of gravity is the special role played by time. In canonical quantum mechanics time enters as a parameter rather than an operator like other observables. The operators correspond to idealized measurements which take place at one instant of time. The Hilbert space inner product is constructed on a surface of constant time. Indeed, time plays such a distinguished role in the theory that the first problem in constructing a canonical quantum theory of a physical system is to identify the time.

In non-relativistic quantum mechanics there is no difficulty. A time is already singled out in classical physics.

In special relativistic quantum mechanics there is no difficulty but there is an issue. If one reads an elementary book on quantum field theory, one typically finds the canonical quantum mechanics worked out using the time of a particular Lorentz frame. Later, one finds a section "proving the Lorentz covariance of the theory." This means showing that if one had carried out the procedure in a different Lorentz frame one would have obtained physically

equivalent results because the corresponding state vectors are connected by a unitary transformation. Thus despite the special role played by time in canonical quantum mechanics it can be made consistent with special relativity.

In general relativity, however, one encounters a crisis. The classical theory does not single out any special set of spacelike surfaces whose labels can play the role of time in canonical quantization. In spacetime physics all spacelike slices are equivalent. There is thus a potential conflict between canonical quantum theory and general covariance unless the canonical quantum theory constructed with one particular set of spacelike surfaces turns out to be equivalent to any other.[19)]

Because the problem of choosing a time arises so immediately in canonical quantum theory of spacetime it is useful to examine other frameworks in which this issue is not as central. Feynman's sum over histories formulation, while at present mathematically less precise, is a useful alternative for "the assistance which it gives one's intuition in bringing together physical insight and mathematical analysis."[20)]

3.3 Sum Over Histories Quantum Mechanics

The basic elements of a sum over histories formulation of a quantum theory are the following:

(1) The possible histories. A history is a set of observables $\{H\}$ which can describe the results of all possible experiments. The possible histories are the possible histories of all experiments.

(2) The amplitude for a history: This is the joint probability amplitude for the occurrence of a particular history given as

$$\Phi[\{H\}] = \exp(iS[\{H\}]) \quad , \tag{3.1}$$

where S is the real action functional for the history.

(3) The construction of conditional probability amplitudes by the principle of superposition. In physics we are interested in the results of an experiment. In every experiment the observables constituting a history can be divided into three groups.

(a) The observables which are fixed by the conditions of the experiment. We call these the conditions $\{C\}$.

(b) The observables which are measured. We call these observations $\{O\}$.

(c) The parts of the history which are neither conditioned nor observed $\{U\}$.

The conditional probability of observing $\{O\}$ given $\{C\}$ is

$$\Phi[\{O\}|\{C\}] = \sum_{\{U\}} \Phi[\{H\}] \quad . \tag{3.2}$$

Typically, the conditions $\{C\}$ define the preparation of the system on which the observations $\{O\}$ are later made.

(4) A probability interpretation. If one can find a complete and exclusive set of observations $\{O_1\},\{O_2\},\ldots$ such that, given the conditions $\{C\}$, one and only one of the observations $\{O_i\}$ is certain to occur, then the probability that it occurs is

$$P[\{O_i\}|\{C\}] = \frac{|\Phi[\{O_i\}|\{C\}]|^2}{\Sigma_i |\Phi[\{O_i\}|\{C\}]|^2} \quad . \tag{3.3}$$

(In order for this to make sense the conditions $\{C\}$ must be sufficiently complete to specify the system but this point will become clear through examples.)

The sum over histories formulation of quantum mechanics has been presented somewhat abstractly to emphasize its generality. To apply this framework to a particular theory we must specify (i) the possible histories, (ii) the action

functional, (iii) the rules for carrying out the sums in (3.2) and (3.3), and (iv) the complete and exclusive sets of observables. We can make this concrete by considering a few examples:

A non relativistic particle. The histories are the particle paths x(t) which move forward in time i.e. for which $dt/dx \neq 0$. The action is

$$S[x(t)] = \int dt \left[\frac{1}{2} m\dot{x}^2 - V(x)\right] \quad . \tag{3.4}$$

A typical conditional amplitude is that for the particle to be at x" at time t" given that it was at x' at time t'. In this case the unobserved, unconditioned parts of the history are the parts of the path other than at t' and t". Thus,

$$\Phi[x'',t''|x',t'] = \sum_{\text{paths}} \exp(iS[x(t)]) \quad . \tag{3.5}$$

Where the sum is over all paths which intersect x' at time t' and x" at time t" (Figure 3). (Of course the details of how the sum is carried out - the measure on the space of paths - is also important for the prescription but in this focus on kinematics we are not spelling this out.) An exclusive and complete set of observations are the observations of x at a given t. All particle paths intersect a t = constant surface at at least one x and at no more than one x. Thus the probability (density) that the particle is at x on a constant t surface and nowhere else on that surface, given that it was prepared by passing it at t' through a "slit" characterized[20] by a function f(x') is

$$P[x,t|f,t'] = \frac{\left|\int dx' \Phi[x,t|x',t']f(x')\right|^2}{\int dx \left|\int dx' \Phi[x,t|x't']f(x')\right|^2} \quad . \tag{3.6}$$

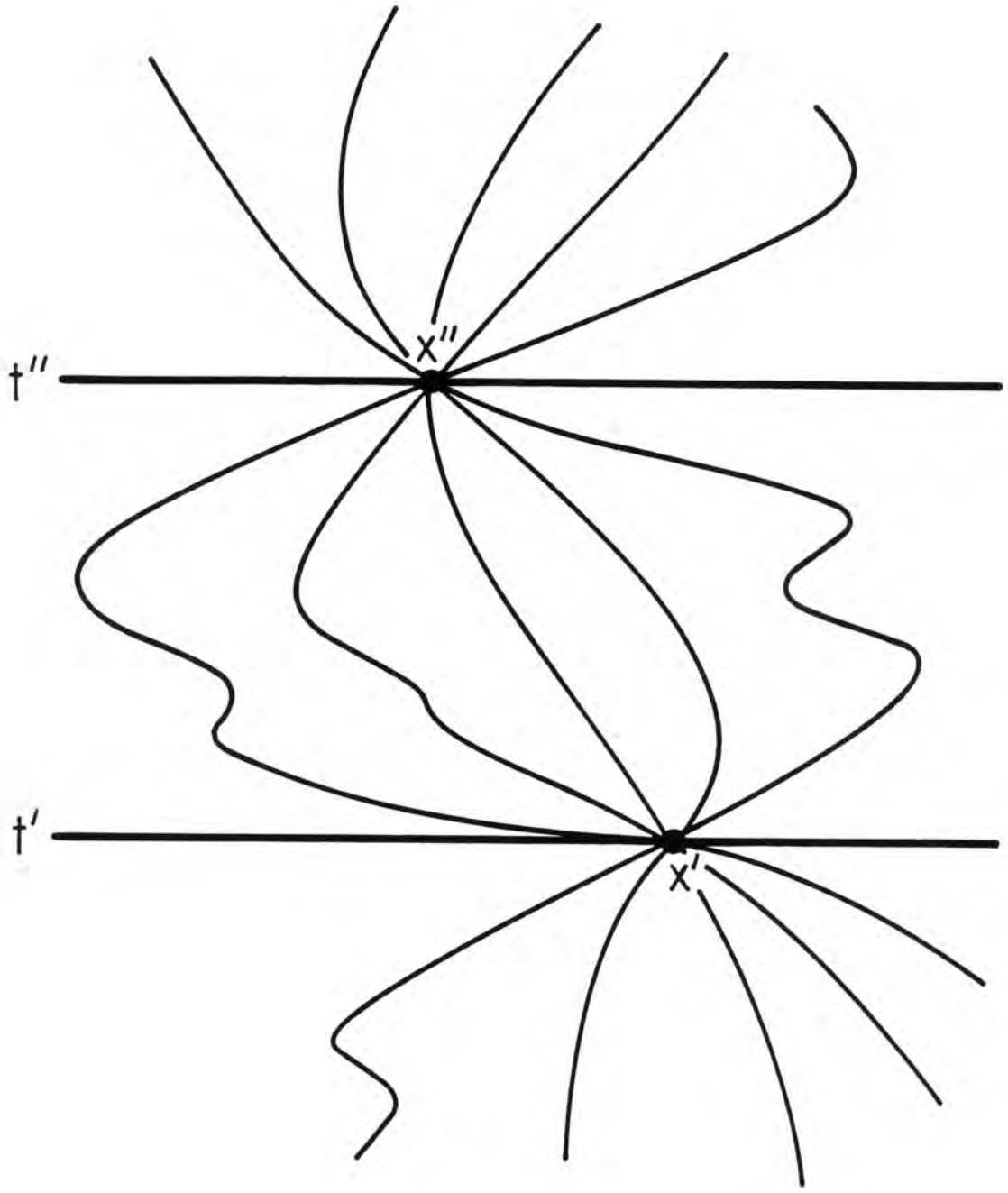

Fig. 3. The conditional probability amplitude for a non-relativistic particle to be at x" at time t", given that it was at x' at time t' is the sum over paths which move forward in time and which intersect the surface t = t' only at x' and the surface t = t" only at x = x".

A relativistic charged particle in an external electromagnetic potential: The histories are again particle paths $x^{\alpha}(w)$. The paths can't move forward in some time t and be a Lorentz invariant class. We therefore consider paths which move both forward and backward in time and this leads to pair creation. An action is

$$S[x^{\alpha}(w)] = -m^2 W + \frac{1}{2}\int_0^W dw \left[\frac{1}{2}\eta_{\alpha\beta}\frac{dx^{\alpha}}{dw}\frac{dx^{\beta}}{dw} + e A_{\alpha}\frac{dx^{\alpha}}{dw}\right] , \quad (3.7)$$

where m is the particle's mass, e is its charge and A_{α} the external vector potential. W is the total parameter time equal on an extremal path to the proper time divided by m. The propagator - the conditional amplitude to find the particle at x''^{α} given that it was at x'^{α} - is

$$\Phi[x^{\alpha''}|x^{\alpha'}] = \sum_{\text{paths}} \exp(iS[x^{\alpha}]) \quad , \quad (3.8)$$

where, since the parameter time is unobserved, the sum over paths includes not only a sum over the different curves which connect x'^{α} to x''^{α} but also the different ways in which the parameter time evolves along these curves. Since the paths move forward and backward in time it is no longer the case that the values of $\vec{x}$ at a given time t constitute an exclusive and exhaustive set of observables. A particle path might intersect a constant time surface many times and one is thus led to a many particle interpretation of this theory. That is, an exclusive and exhaustive set of observations are the number of particles n on a constant t surface and their positions $\vec{x}_1, \ldots, \vec{x}_n$.

A scalar quantum field in flat space: The histories are the possible field configurations in flat spacetime, $\varphi(x)$. An action is

$$S[\varphi(x)] = \frac{1}{2}\int d^4x\,[\,-\,(\nabla_\alpha\varphi)^2\,-\,m^2\varphi^2\,+\,V(\varphi)\,]\quad , \qquad (3.9)$$

where $V(\varphi)$ is some polynomial interaction in φ. A conditional amplitude is that for the field to take one configuration $\varphi''(\vec{x})$, on a spacelike σ'' surface given that it was $\varphi'(\vec{x})$ on another spacelike surface σ'. This is

$$\Phi[\varphi''(\vec{x}),\sigma''\,|\varphi'(\vec{x}),\sigma'] = \sum_{\varphi(x)} \exp(iS[\varphi]) \quad , \qquad (3.10)$$

where the sum is over all spacetime field configurations which have the prescribed values on σ' and σ''. The different field configurations on such a surface are a set of exhaustive and exclusive observations.

A string: The histories are the world sheet of the string specified by

$$X^A = X^A(\sigma,\tau) \qquad A = 0,1,2,\cdots\cdots? \qquad (3.11)$$

The action might be the area of the string. A complete and exhaustive set of variables might be the transverse directions of the string at one instant of time.

Spacetime: Einstein's idea was that gravitational physics is spacetime physics. The histories for gravity are therefore four dimensional geometries, by which one means a four dimensional manifold M with a Lorentz (-,+,+,+) signatured metric $g_{\alpha\beta}(x)$. Two metrics represent the same geometry if they are diffeomorphic i.e. if they can be connected by a coordinate transformation.

To keep our discussion simple I shall assume that we are dealing with cosmologies which are spatially closed and for which the topology is fixed to be $M^3 \times \mathbb{R}$ where M^3 is a closed manifold for space. In the closed Friedman models M^3 is the 3-sphere S^3. This is not the most general class of histories to consider. One might want to consider

the possibilities of open cosmologies, for example. One might want to consider summing over manifolds with different topology. Since spacetime is a manifold with metric it seems artificial to sum over metrics but keep the manifold fixed. In particular one might want to sum over histories in which the topology changes, such as that of Fig. 4. Such geometries cannot possess a smooth

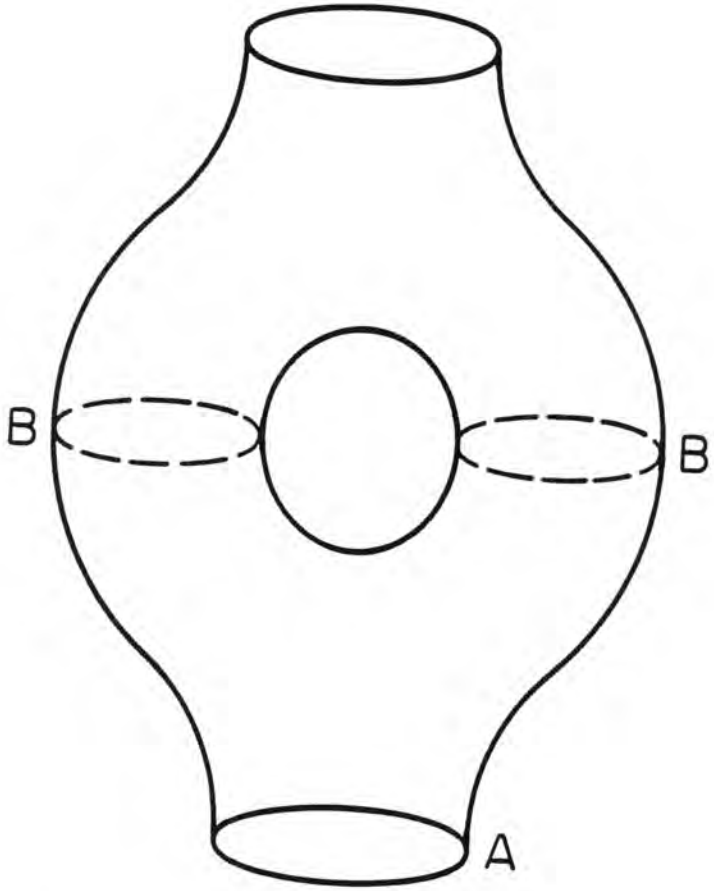

Fig. 4. Topology change in two dimensions. The two-dimensional surface portrayed embedded in three-dimensional flat space corresponds to a history in which the topology of its spatial sections changes from that of a single circle at A to two circles at B. It is not possible to introduce a non-singular vector field corresponding to time on such histories.

decomposition into spacelike slices, there are no non-singular timelike vector fields. I see no particular reason to leave them out, although arguments have been advanced against them.[21] However, we will be able to do all of what we want to do by making the simple assumption described.

The action for Einstein's general relativity is

$$\ell^2 S[g] = 2\int_{\partial M} K h^{1/2} d^3x + \int_M R\sqrt{-g}\, d^4x \quad , \tag{3.12}$$

where ∂M is the spacelike boundary of the manifold M and R is the scalar curvature. h_{ij} is the induced three metric of the spacelike boundary and K is the trace of its extrinsic curvature. That is, if n^α is the unit outward pointing normal to ∂M

$$K = \nabla_\alpha n^\alpha \quad . \tag{3.13}$$

The simplest conditional probability amplitude answers the question: "Given there occurs in the spacetime a spacelike surface with a three geometry described by a three metric h'_{ij} what is the conditional amplitude that there occur a second with a three geometry described by a three metric h''_{ij}?" It is

$$\Phi[h''_{ij}|h'_{ij}] = \Sigma_{C} \exp(iS[C]) \quad . \tag{3.14}$$

The sum stands for a sum over all physically distinct metrics $g_{\alpha\beta}$ which induce metrics h'_{ij} and h''_{ij} on the two pieces of the boundary. To implement the "physically distinct" part of this prescription requires some "gauge fixing" machinery essentially familiar from gauge theories.

The difficult task in constructing a sum over histories quantum theory of spacetime is the identification of the complete and exclusive sets of observables and the associated probability interpretation. We shall return to these issues in Section 3.5.

3.4 Wave Functions

In quantum mechanics we would like to capture the classical notion of "the state of a system" - a specification of the system at one instant of time. We can then hope to use this information and dynamical laws to evolve the state forward in time. When spacetime is a fixed, flat background, it is straightforward to start with the sum over histories formulation of quantum mechanics and identify the correct notion of state. Suppose the system is prepared by conditions which lie to the past of some spacelike surface. Probabilities for future observations are determined by sums of histories which cross this surface. We may reasonably regard the state of the system <u>on</u> the spacelike surface as specified by the collection of sums over histories proceeding from the given conditions in the past to a fixed value on the surface. This defines the wave function on the surface.

For example, fixing the history of a non-relativistic particle on a constant time surface means fixing its position on the surface. Thus we write for the wave function

$$\psi_C(x,t) = \int_C \delta x(t) \exp(iS[x(t)]) \quad , \tag{3.15}$$

where the sum is over all paths to the past of t which meet the conditions C. There is a wave function for each set of conditions, that is, for each way of preparing the system.

The probability interpretation for sums over histories immediately assigns a probability interpretation for the wave function. The values of x at a given t are a set of complete and exclusive observations. The conditional amplitude $\Phi[x,t|C]$ factors into a part from the sum over

paths to the past of t (the wave function) and a part from the sum over the paths to the future where there are no conditions. The part from the paths to the future is not very well defined mathematically and has to be assigned a constant value for consistency. In any event it cancels in the formation of probabilities and we have

$$P[x,t|C] = \frac{|\psi_C(x,t)|^2}{\int dx\,|\psi_C(x,t)|^2} \quad . \tag{3.16}$$

Fixing a field configuration on a spacelike surface corresponds to fixing the value of the field on the surface. Thus we write

$$\Psi_C[\varphi(\vec{x}),\sigma] = \int_C \delta\varphi \exp(iS[\varphi]) \quad , \tag{3.17}$$

where the sum is over field configurations which match the configuration $\varphi(\vec{x})$ on the surface σ and satisfy the conditions C to the past. The wave function can be assigned a similar probability interpretation because the values of $\varphi(\vec{x})$ on the surface are a complete and exclusive set of possibilities.

The wave function is the quantum analog of the classical motion of "state of the system at one time." We would like to derive a dynamical equation for it. Solving the equation would then be an alternative way of calculating the wave function and a useful one because we are more used to solving differential equations than evaluating functional integrals. In theories with a flat background spacetime we are familiar with how to do this.[20] In the quantum mechanics of a single particle for example, we calculate $\psi(x,t+\epsilon)$ from $\psi(x,t)$ by doing the sum over histories to calculate the propagator between infinitesimally separated slices. For small ϵ the integrals can be done by steepest descents and only the values of

$\psi(x,t)$ for nearby values of $\ddot{x}$ contribute to $\psi(x,t+\epsilon)$. In this way we recover the Schrodinger equation

$$i\frac{\partial\psi}{\partial t} = [-\frac{1}{2m}\frac{d^2}{dx^2} + V(x)]\psi \quad . \tag{3.18}$$

The derivation of quantum dynamics from the sum over histories formulation and the assignment of a probability interpretation to the wave function is straightforward when spacetime is a fixed background. It is less straightforward when spacetime is part of the dynamical variables as it is in gravitational theories. To understand the issues involved we begin by considering a simple model.

3.5 A Parametrized Model

Suppose we are presented with a classical theory involving n+2 variables q^α, $\alpha = 0,1,\cdots n$ and L described by an action

$$S[q^\alpha,L] = \int_M dt[(L - \frac{1}{2}L^2\dot{q}^0)\delta_{ab}\dot{q}^a\dot{q}^b - V(q^a)\dot{q}^0] \ . \tag{3.19}$$

Here, M is a finite interval in t, V is a given function, a,b range from 1 to n, and a dot denotes differentiation with respect to time. We shall now describe both the classical and quantum dynamics of this theory.[23]

The theory described by (3.19) has a symmetry. If, for arbitrary f equal to unity on the endpoints of M, we make the transformation

$$L \to \dot{f}(t)L(f(t)) \quad , \tag{3.20a}$$

$$q^\alpha \to q^\alpha(f(t)) \quad , \tag{3.20b}$$

the action remains unchanged. One can easily see this by simultaneously changing the variable of integration

$$t \to f(t) \quad , \tag{3.21}$$

and for this reason the transformation (3.20) is called a reparametrization of the time.

The classical equations of motion found by varying q^a, q^0 and L are

$$-\frac{d}{dt}[(L-\frac{1}{2}L^2\dot{q}^0)\dot{q}^a] - \frac{\partial V}{\partial q^a}\dot{q}^0 = 0 \quad , \tag{3.22a}$$

$$\frac{d}{dt}[\frac{1}{2}L^2\delta_{ab}\dot{q}^a\dot{q}^b + V(q^a)] = 0 \quad , \tag{3.22b}$$

$$\dot{q}^0 = L^{-1} \quad . \tag{3.22c}$$

From (3.22c) we learn that the theory has a constraint. We are not free to specify all the q^a, q^0, L, $\dot{q}^a$, $\dot{q}^0$, $\dot{L}$ on some initial surface and integrate forward in time. First, there is not even an evolution equation for L. With an appropriate transformation of the form (3.20) we could, in fact, pick L = 1 for all times. Second, $\dot{q}^0$ is not freely specifiable but must satisfy the condition (3.22). This is the constraint. Only the q^a are the true degrees of freedom whose value and first time derivative are freely specifiable at an initial time.

The constraint takes an interesting form if we re-express it in terms of the momenta conjugate to q^α. (There is no momentum corresponding to L because it is undifferentiated in (3.19).) These momenta are

$$p_0 = -\frac{1}{2}L^2\delta_{ab}\dot{q}^a\dot{q}^b - V(q^a) \quad , \tag{3.23a}$$

$$p_a = L(2-L\dot{q}^0)\dot{q}^a \quad . \tag{3.23b}$$

Then from (3.22c)

$$p_0 + [\frac{1}{2}\delta^{ab}p_a p_b + V(q^a)] = 0 \quad . \tag{3.24}$$

The left hand side of (3.24) is the total Hamiltonian

$$H = p_\alpha\dot{q}^\alpha - \mathcal{L} \quad , \tag{3.25}$$

multiplied by L (when (3.22c) is satisfied). Thus

$$H = 0 \quad . \tag{3.26}$$

This is characteristic of theories invariant under reparametrizations of the time.

Working in the gauge where $L = 1$ and eliminating q^0 using the constraint we can see what this theory really is. Then $q^0 = t$ and the equation of motion (3.22a) is

$$\ddot{q}^a + \frac{\partial V}{\partial q^a} = 0 \quad . \tag{3.27}$$

This is the equation of motion of a particle in a potential V. Eq. (3.22b) is the associated conservation of energy. Classically, the theory described by the action (3.19) is the same as the theory of a non-relativistic particle in a potential V, but written in a form where the time appears as one of the dynamical variables q^0.

It is instructive to construct the sum over histories quantum mechanics of the theory with the action (3.19) as if we did not know that it was the theory of a non-relativistic particle. Recall that one needs the action, the histories, the measure and the sets of complete and exclusive observations. The action is (3.19). The histories can be described by the functions $q^\alpha(t)$, $L(t)$. Two sets of $\{q^\alpha, L\}$ which are equivalent under reparametrization of the time [eq. (3.20)] describe the same history and are to be counted only once in the sum. If we restrict the paths so that the q^a move forward in q^0 the quantum theory will correspond to non-relativistic quantum mechanics. Other classes of paths could be investigated but lead to different theories with pair creation and annihilation.

Wave functions of states in the quantum theory should depend on the variables which describe the restriction of a history to a spacelike surface. To see what this means imagine a particle path crossing such a surface. We could

describe this path by many different $\{q^\alpha(t), L(t)\}$ each related to the other by a reparametrization transformation. The values of q^α by themselves describe a history restricted to a spacelike surface. To specify L in addition would be to specify too much because we can always choose a reparametrization gauge where L has an arbitrary value. Thus we write

$$\psi = \psi(q^\alpha) \quad . \tag{3.28}$$

We do not include an additional label "t" for two reasons: (1) t is not a measurable physical variable but only a parameter label. It can be changed by a reparametrization transformation. (2) Even in a particular reparametrization gauge to include t as a label would be redundant. The value of q^0 already locates the particle along its trajectory in time.

The path integral for the wave function determined by some previous conditions C is

$$\psi_C(q^\alpha) = \int_C \delta q^\alpha \delta L \det(\dot{T})\, \delta(q^0 - T(t)) \exp(iS[q^\alpha, L]) \, . \tag{3.29}$$

Here, we have introducted a "gauge-fixing" δ-function to enforce the gauge condition $q^0 = T(t)$ with T a monotomically increasing function matching q^0 on boundary. This is the simplest way to fix f in (3.20). $\det(\dot{T})$ is the corresponding "Faddeev-Popov" determinant.

The action (3.19) is quadratic in L. With an appropriate choice of measure, the integral over L is gaussian and can be carried out explicitly. The integral over q^0 can be carried out using the δ-function. The result is familiar

$$\psi_C(q^\alpha) = \int_C \delta q^a \exp\left(i \int dt \left[\tfrac{1}{2}\, \delta_{ab} \dot{q}^a \dot{q}^b - V\right]\right) \quad . \tag{3.30}$$

Thus with the choice of paths and measure described above

the wave functions of the quantum theory built on the action (3.19) are those of the quantum mechanics of the free non-relativistic particle.

It follows from (3.30) that the evolution equation for $\psi_C(q^\alpha)$ is just the Schrödinger equation

$$\left[- i \frac{\partial}{\partial q^0} - \frac{1}{2} \delta^{ab} \frac{\partial^2}{\partial q^a \partial q^b} + V(q^a) \right] \psi_C (q^\alpha) = 0 \quad . \qquad (3.31)$$

This is the operator form of the classical constraint equation (3.24). We could write

$$H\psi(q^\alpha) = 0 \quad . \qquad (3.32)$$

Thus in quantum mechanics as in classical mechanics the vanishing of the Hamiltonian does not imply the absence of dynamics. In fact this condition becomes the dynamical equation of the theory when expressed in terms of its true degrees of freedom.

3.5 General Relativity

3.5.1 The classical theory in 3+1 form

The structure of the general theory of relativity is similar in many ways to the model of non-relativistic particle mechanics with parametrized time that was discussed in the preceding section. Like this model, general relativity is invariant under reparametrizations of the time. If one singles out a family of spacelike surfaces, labels them by a time coordinate, t, then the action for general relativity [eq. (3.12)] is unchanged by a relabeling of these surfaces with a different t-coordinate. Of course, general relativity is also invariant under the larger group of diffeomorphisms corresponding to general coordinate transformations.

Because of its invariance under diffeomorphisms general relativity is a theory with constraints - restrictions on the values of the metric and its first time derivative that can be specified on an initial value surface. To spell these constraints out we begin by rewriting the classical action (3.12) in a form which emphasizes spacelike surfaces. This is the 3+1 formulation of Arnowitt, Deser and Misner (ADM).[24]

Suppose that one has a family of spacelike surfaces labeled by a time coordinate t. The metric can generally be written as

$$ds^2 = -N^2dt^2+h_{ij}(dx^i+N^idt)(dx^j+N^jdt) \quad . \tag{3.33}$$

The tensor h_{ij} is the induced metric of the spacelike surface. N and N^i can be described as follows: Given two neighboring spacelike surfaces, labeled by t and t+dt, the displacement between a point in the first surface and a point in the second be decomposed into a displacement in the first surface and another displacement normal to it. If the points are labeled by x^i and x^i+dx^i respectively then dx^i+N^idt is the displacement vector in the surface and Ndt is the length of the normal displacement. N^i is therefore called the "shift vector" and N is called the "lapse function."

The 3-metric h_{ij} specifies the intrinsic geometry of a surface of constant t. With it we can associate a spatial covariant derivative D_i and compute the spatial curvatures, ${}^3R_{ijk\ell}$. The lapse and shift can be thought of as scalar and vector fields on this surface and tensor operations carried out accordingly. The differential change in the unit normal projected into the surface

$$K_{ij} = - \nabla_i n_j \quad , \tag{3.34}$$

is a measure of how the spacelike surface is curved in

4-dimensional space. It is called the extrinsic curvature. Explicitly, in terms of the decomposition (3.33) it is

$$K_{ij} = \frac{1}{N}\left[-\frac{1}{2}\frac{\partial h_{ij}}{\partial t} + D_{(i}N_{j)}\right] \quad . \tag{3.35}$$

The action for general relativity, eq. (3.12) can be rewritten in terms of the decomposition (3.33) as

$$\ell^2 S_E = \int d^4x h^{1/2} N(K_{ij}K^{ij} - K^2 + {}^3R - 2\Lambda) \quad . \tag{3.36}$$

Here, $K = K^i{}_i$, ${}^3R$ is the scalar curvature of the surface and tensor operations are carried out in the geometry of the surface. To this must be added the action of the matter fields, but for simplicity we shall consider pure gravity until Section 5. Written in this ADM form it is clear that general relativity is a theory with constraints because the lapse N and the shift N^i occur in S_E undifferentiated with respect to t. The corresponding equations of motion may therefore be expressed entirely in terms of the metric h_{ij} and its conjugate momentum π_{ij}. They are thus constraints on initial data.

The momentum conjugate to h_{ij} is easily found from (3.36) and (3.35) and is

$$\ell^2 \pi_{ij} = -h^{1/2}(K_{ij} - h_{ij}K) \quad . \tag{3.37}$$

The three constraints following from varying (3.36) with respect to the N^i are

$$D_i \pi^{ij} = 0 \quad . \tag{3.38}$$

The one following from varying N may be written

$$\ell^2 G_{ijk\ell}\pi^{ij}\pi^{k\ell} + \ell^{-2} h^{1/2}(-{}^3R + 2\Lambda) = 0 \quad , \tag{3.39}$$

where

$$G_{ijk\ell} = \frac{1}{2} h^{-1/2}(h_{ik}h_{j\ell} + h_{i\ell}h_{jk} - h_{ij}h_{k\ell}) \quad . \tag{3.40}$$

This constraint will be important for us. It is the constraint associated with the invariance of the theory under a reparametrization of the spacelike surfaces. The model of parametrized time particle quantum mechanics suggests, and one can check, that this constraint implies the vanishing of the total Hamiltonian density

$$\mathcal{H} = \ell^2 G_{ijk\ell}\pi^{ij}\pi^{k\ell} + \ell^{-2} h^{1/2}(-{}^3R+2\Lambda) = 0 \quad . \quad (3.41)$$

As in the model, this equation summarizes the dynamics of classical general relativity without matter.

3.5.2 Quantum mechanics of closed cosmologies

To investigate the quantum mechanics of a closed cosmology we must first describe correctly a quantum state. Recall from the discussion of Section 3.3 that a state is described by a wave function whose arguments are the variables describing a history fixed on a spacelike surface. The histories are the 4-geometries on $M^3 \times \mathbb{R}$. Each may be described by a metric $g_{\alpha\beta} = \{N, N^i, h_{ij}\}$ but there will be many metrics corresponding to the same geometry. This complicates the identification of the arguments of the wave function, as it did in the case of parametrized time particle quantum mechanics, but we may proceed in the analogous way. Consider the 4-geometries in which a given spacelike surface with definite 3-geometry occurs, but which are otherwise free to vary off this surface. By a suitable choice of coordinates, say $N = 1$ and $N^i = 0$, the metrics for all these geometries could be brought to a standard form where h_{ij} is the only variable. Thus, we may take the 3-metric h_{ij} as describing a history fixed on a spacelike surface and write for the wave function of a closed cosmology

$$\Psi = \Psi[h_{ij}] \quad . \quad (3.42)$$

There is no additional dependence on the coordinate t. First, t is not a physical label but may be prescribed at will. Second, the three geometry already carries information about the location of the surface in time. A generic 3-geometry will fit into a generic 4-geometry in a locally unique way (and in particular at a unique "time") if it fits in at all. The counting of variables implied by this labeling of the wave function is correct. There are 6 components of h_{ij} at each space point. Three of these "are pure gauge," that is, could be chosen arbitrarily by a suitable choice of spatial coordinates. If one component corresponds to time there remain two. This is the correct number of degrees of freedom of a massless spin-2 field.

The sum over histories for the wave function corresponding to a set of conditions C is

$$\Psi_C[h_{ij}] = \int_C \delta g \, \exp(iS_E[g]) \quad . \tag{3.43}$$

The integration is over a class of 4-geometries defined on the manifold which is that part of $M = M^3 \times \mathbb{R}$ to the past of a bounding $M^3 = \partial M$. (See Figure 5.) The class consists of those 4-geometries which induce the metric h_{ij} on ∂M and which satisfy the conditions C to its past.

Wave functions constructed as sums over histories should automatically satisfy the operator form of the constraints of the theory as did the wave function for the model of parametrized time particle quantum mechanics. An operator form of the classical constraints may be obtained by replacing $\pi^{ij}(x)$ by $-i\delta/\delta h_{ij}(x)$ in eqs. (3.39) and (3.40). In the case of (3.39) we have

$$D_i \left(\frac{\delta \Psi}{\delta h_{ij}(x)} \right) = 0 \quad . \tag{3.43}$$

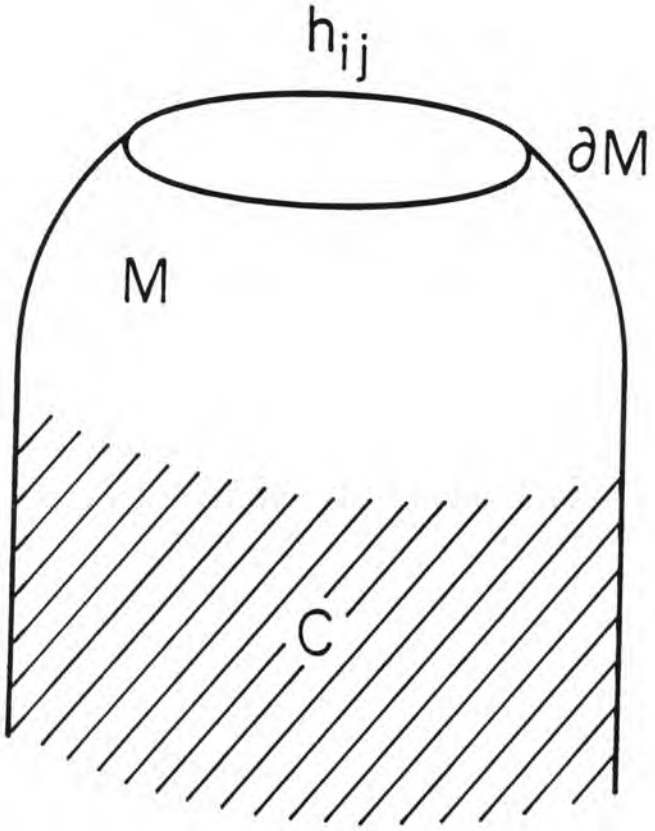

Fig. 5. The sum over histories for a wave function $\Psi_C[h_{ij}]$ is a sum over geometries on the manifold $M^3 \times (-\infty, 0]$ which induce the metric h_{ij} on the boundary $\partial M = M^3$ and which satisfy the conditions C to the past of this surface.

For the Hamiltonian constraint (3.40) there are many possible operator equations depending on the choice of factor ordering. They all have the form

$$[-\ell^2 \nabla_x^2 + \ell^{-2} h^{1/2}(x)(-{}^3R(x)+2\Lambda)] = 0 \quad , \tag{3.44}$$

where

$$\nabla_x^2 = G_{ijk\ell} \frac{\delta^2}{\delta h_{ij}(x)\, \delta h_{k\ell}(x)} + \left(\begin{array}{l} \text{linear derivative} \\ \text{term depending on} \\ \text{factor ordering} \end{array} \right) . \tag{3.45}$$

Eqs. (3.43) and (3.44) are an infinite set of equations - one for each point on the spacelike surface at which Ψ is defined. Eq. (3.44) is called the Wheeler-DeWitt equation. Like its analog (3.31) in parametrized time particle quantum mechanics, the Wheeler-DeWitt equation determines the quantum dynamics of general relativity.

We shall not derive the Wheeler-DeWitt equation from the functional integral (3.43) although it would be possible to do so*, at least formally.[25,26] The formal derivation of (3.43), however, is simple enough that we can give it here to indicate the methods involved: Consider an infinitesimal coordinate transformation on M which relabels the spatial coordinates of each constant t surface in an

---

*There should be a connection between the measure used for the sum over histories and the factor ordering in the Wheeler-DeWitt equation. This connection bears on the long standing problem in the canonical theory[27] of whether there exists a factor ordering such that the algebra of the constraints closes. Either the construction of wave functions by sums over histories resolves this problem or there is a restriction on the permissible measures in the sum arising from it. (See also Ref. 28.) The absence of a clear understanding of the connection between measure and the form of the Wheeler-DeWitt equation is the reason we have not presented a detailed derivation. As we intend to solve the equation for the most part in the semiclassical approximation these ambiguities will not affect our limited conclusions.

identical fashion. Such a transformation has the form

$$t \to t, \qquad x^i \to x^i + \xi^i(x^k) \quad , \tag{3.46}$$

for infinitesimal ξ^i. The spatial metric transforms as

$$h_{ij} \to h_{ij} + 2D_{(i}\xi_{j)} \quad , \tag{3.47}$$

and the 4-metric as

$$g_{\alpha\beta} \to g_{\alpha\beta} + 2\nabla_{(\alpha}\xi_{\beta)} \quad , \tag{3.48}$$

with $\xi^0 = 0$. In eq. (3.43), shift simultaneously the argument of the wave function on the left by the amount in (3.47) and the integration variables on the right by the amount in (3.48). There remains an identity. The integral on the right hand side of this relation is identical to what it was before because both action and measure are invariant under coordinate transformations. Thus we conclude

$$\Psi[h_{ij} + 2D_{(i}\xi_{j)}] = \Psi[h_{ij}] \quad , \tag{3.49}$$

or equivalently

$$\int_{\partial M} d^3x \; D_{(i}\xi_{j)} \frac{\delta\Psi}{\delta h_{ij}(x)} = 0 \quad . \tag{3.50}$$

Integrating this relation by parts and noting that ξ^i is arbitrary, one recovers the operator form of the constraints (3.43). This derivation shows that the physical content of these three constraints is that the wave function does not depend on the choice of coordinates in the spacelike surface as it should not.

3.5.3 Superspace

As a consequence of the constraint (3.43) the wave function may be thought of as a function on the space of three geometries - the space of gauge inequivalent metrics on M^3. This is called "superspace." The quantity

$G_{ijk\ell}$ may be used to form a metric on superspace. If δh_{ij} and $\delta h'_{ij}$ are the changes in 3-metric corresponding to an infinitesimal displacement in superspace we may define the inner product of these two displacments to be

$$(\delta h, \delta h') = \int d^3x G^{ijk\ell} \delta h_{ij} \delta h'_{k\ell} \qquad (3.51)$$

Where $G^{ijk\ell}$ is the inverse of $G_{ijk\ell}$ considered as a 6 x 6 matrix in the space of symmetric index pairs (ij). Other metrics on superspace may be defined by inserting an arbitrary function in (3.51). The metric (3.51) suggests that an elegant way for choosing the factor ordering in the Wheeler-DeWitt equation is to take ∇_x^2 to be the "covariant Laplacian" in this metric. (For the issues raised by such a choice see Ref. 28.)

At each space point the signature of the 6 x 6 metric $G_{ijk\ell}$ is (-,+,+,+,+,+). The Wheeler-DeWitt equation is thus a kind of "hyperbolic" equation in superspace. We recall that one of the six degrees of freedom in the 3-metric represents time in the sense of locating the spacelike surface locally in a 4-geometry. We may think therefore, of a fixed choice for this time as defining a family of hypersurfaces in superspace and the Wheeler-DeWitt equation as specifying the propagation of the wave function from one hypersurface to another. It is in this sense that the Wheeler-DeWitt equation implements quantum dynamics.

3.5.4 Use of the wave function

In non-relativistic particle quantum mechanics (Section 3.3) we were able to assign a probability interpretation to the wave function because the values of the particle's position at a given time constituted a complete and exclusive set of observations. The identification of a complete and exclusive set of observables in the quantum

theory of spacetime is a much more difficult problem. We lack, by and large, the simple analyses of thought experiments which guide our intuition in particle quantum mechanics. Even formally, the question is complicated by the problem of the choice of time (Section 3.2). By analogy with particle quantum mechanics one would expect a complete and exclusive set of observables to be different values of some part of the 3-metric "at one time" - but which "time" should be used? The sum over histories formulation of quantum mechanics may guide us to a resolution of such questions but in the meantime we can proceed qualitatively through an analysis of the correlations in the wave function. We expect variables to be correlated on those regions of superspace where the wave function is large and anticorrelated where it is small. This minimal interpretation will be enough to take the first steps in quantum cosmology.

Of the four ingredients of a sum over histories formulation of quantum cosmology - histories, action, measure, observables - we have discussed the first two. Discussions of the measure, not given here because they are complex and will not bear directly on the semiclassical approximations to the wave function we shall mostly consider, may be found in Refs. 29. The interpretation of Ψ at the level of correlations will be sufficient for a first analysis of the proposal for the quantum state of the universe that we shall now discuss.

4. THE QUANTUM STATE OF THE UNIVERSE

4.1 A State of Minimum Excitation

The universe is in a state of low excitation. The large scale distribution of matter and metric are nearly homogeneous and isotropic. The entropy of the matter is low compared to that of the highly clumped and irregular configurations it might have had. Whatever the state of the universe is, it is close in some sense to a state of minimum excitation. We, therefore, begin a discussion of the state of the universe with a discussion of the state of minimum excitation for closed cosmologies.

In the quantum mechanics of a particle in a potential there are two ways of calculating the wave function of the state of minimum excitation. We can calculate it as the lowest energy eigenfunction of the Hamiltonian (the ground state)

$$H\psi_0(x_0) = E_0\psi_0(x_0) \quad . \tag{4.1}$$

Completely equivalently, the wave function of the ground state may be calculated as a Euclidean functional integral.

$$\psi_0(x_0) = \int \delta x(\tau) \exp(-I[x(\tau)]) \quad . \tag{4.2}$$

Here, $I[x(\tau)]$ is the Euclidean action functional

$$I[x(\tau)] = \int d\tau [\tfrac{1}{2}m\dot{x}^2 + V(x)] \quad . \tag{4.3}$$

The sum is over all paths which start at x_0 at time $\tau = 0$ and proceed in the infinite past to a configuration of minimum action.

It is not difficult to sketch the demonstration of the equivalence of (4.1) and (4.3). One begins with the path integral for the propagator

$$\langle x'', t'' | x' t' \rangle = \int \delta x(t) \exp(iS[x(t)]) \quad . \tag{4.4}$$

Here, S is the usual action (i.e. (4.3) with the opposite sign for V(x)) and the sum is over paths which start at x' at the time t' and wind up at x" at time t". Consider the particular propagator $\langle x_0,0|0,t\rangle$ and expand it in a complete set of energy eigenstates as follows:

$$\begin{aligned}\langle x_0,0|0t\rangle &= \Sigma_n\langle x_0,0|n\rangle\langle n|0,t\rangle \\ &= \Sigma_n e^{iE_n t}\psi_n(x_0)\psi_n^*(0) \quad ,\end{aligned} \tag{4.5}$$

where ψ_n is the wave function of the energy eigenstate with eigenvalue E_n. Equate the last line of (4.5) to the right hand side of (4.4) and rotate the time to imaginary values, $t \to -i\tau$, on both sides of the equation. One has

$$\Sigma_n e^{E_n\tau}\psi_n(x_0)\psi_n^*(0) = \int\delta x(\tau)\exp(-I[x(\tau)]) \quad . \tag{4.6}$$

Then take the limit $\tau \to -\infty$. If one normalizes the energy so that the lowest eigenvalue is zero, only the ground state term survives in the sum on the left hand side of (4.6). The sum over paths on the right hand side becomes the sum described above and one recovers after a normalization eq. (4.2).

In the quantum mechanics of closed cosmologies one does not expect to recover the wave function of the state of minimum excitation as the lowest eigenvalue of a Hamiltonian. This is because there is no natural notion of energy for closed cosmologies.

In classical gravity, the principle of equivalence shows us that there can be no definition of local energy for the gravitational field. <u>All</u> gravitational effects vanish locally in a freely falling frame. Alternatively note that energy is the conserved quantity which arises from time translational symmetries of spacetime and in a general spacetime there will be no time translation invariance. For spacetimes with special symmetries one

can define an energy. For example, asymptotically flat spacetimes have a time translation symmetry "at infinity." Correspondingly, we can define a total energy which is conserved. For closed cosmological models, however, there is no such symmetry.

One might pick arbitrarily a family of spacelike slices, identify the generator which takes us from one slice to the next, call that the Hamiltonian, and find the lowest eigenstate. As the above argument suggests, however, and has been shown by Kuchař[30] there is no slicing for which the resulting Hamiltonian is time independent and thus none for which one could construct a unique ground state.

While the construction of the wave function of the state of minimum excitation as the lowest eigenstate of the Hamiltonian fails for closed cosmologies the construction using a Euclidean functional integral can be generalized.[25] Schematically, including a generic matter field φ, one would write

$$\Psi_0[h_{ij},\varphi_0] = \int \delta g\, \delta\varphi \exp(-I[g,\varphi]) \quad . \tag{4.7}$$

The sum is over a class of Euclidean four geometries which have a boundary on which the induced 3-metric is h_{ij} and matter configurations which match the values $\varphi_0(\vec{x})$ on the boundary. (These are the analog of the paths starting at x_0 in (4.2)). The action is the sum of the Euclidean action for the matter and the Euclidean action for general relativity. On a manifold M with boundary ∂M the latter is

$$\ell^2 I_E[g] = -2\int_{\partial M} Kh^{1/2} d^3x - \int_M (R-2\Lambda) g^{1/2} d^4x \quad . \tag{4.8}$$

where, as in eq. (3.12), K is the trace of the extrinsic curvature of the boundary. To completely specify the wave

function Ψ_0 it remains to complete the specification of the class of geometries and field configurations summed over (the analog of the paths going to minimum action at $\tau \rightarrow -\infty$ in (4.2)). The proposal is that one should sum over compact Euclidean four geometries and field configurations which are regular on them. The remaining boundary condition for geometries contributing to the state of minimum excitation is that there is no other boundary. In Section 5 we shall examine evidence for the appropriateness of this choice.

The Euclidean construction of the wave function of the state of minimum excitation does not change the form of the constraints it satisfies. Ψ_0 continues to satisfy the generalizations of (3.43) and (3.44) to include matter. These are

$$iD_j\left(\frac{\delta\Psi}{\delta h_{ij}}\right) = \ell^2 T^i_n(\varphi, -i\frac{\delta}{\delta\varphi})\,\Psi \quad , \tag{4.9a}$$

$$\frac{1}{2}[\ell^2 \nabla_x{}^2 + \ell^{-2} h^{1/2}({}^3R - 2\Lambda)]\Psi = h^{1/2} T_{nn}(\varphi, -i\frac{\delta}{\delta\varphi})\,\Psi \ . \tag{4.9b}$$

Here $T_{\alpha\beta}(\varphi,\pi)$ is the stress energy of the matter expressed in terms of the field and its canonical momentum. This becomes the operators in (4.9) when projected appropriately onto the direction n^α normal to the constant t surfaces and when π is replaced by $-i\delta/\delta\varphi$. One can derive (4.9a) simply by following the derivitation sketched in Section 3.5.2. The Wheeler-DeWitt eqn can be derived formally in a similar fashion. (See, e.g. Ref. 2.)

The Wheeler-DeWitt equation and the associated constraints (4.9) presumably have many solutions. The sum over histories (4.7) singles out one of them. The Euclidean functional integral prescription may therefore be thought of as supplying boundary conditions for the Wheeler-DeWitt equation and this will prove a useful approach to take when actually solving for Ψ_0.

4.2 The Conformal Factor

Some attention must be given to the meaning of the sum in (4.7). The Euclidean Einstein action is not positive definite. One can see this[31] by considering the family of metrics generated from a given one, $\tilde{g}$, by conformal transformations

$$g_{\alpha\beta} = \Omega^2 \tilde{g}_{\alpha\beta} \quad . \tag{4.10}$$

In terms of Ω and $\tilde{g}$ the Euclidean action becomes

$$\begin{aligned} \ell^2 I_E[\Omega,\tilde{g}] = & -2\int_{\partial M} d^3x \tilde{h}^{1/2} \Omega^2 \tilde{K} \\ & - \int d^4x \tilde{g}^{1/2} [\Omega^2 \tilde{R} + 6(\tilde{\nabla}\Omega)^2 - 2\Lambda\Omega^4] \quad . \end{aligned} \tag{4.11}$$

By making Ω rapidly varying, the action can be made as negative as desired. A sum over real geometries of the form (4.7) will therefore not converge.

One might think that the indefiniteness of the Euclidean Einstein action was an indication of some instability in the quantum theory. There are such situations in particle quantum mechanics.[32] Consider, for example, a particle moving in a potential V(x) of the form shown in Figure 6 and its Euclidean propagator $\langle 0,0|0,\tau\rangle$. From (4.6) it follows that the large negative time behavior of this propagator is proportional to $\exp(E_0\tau)$ where E_0 is the energy of the ground state. Also from (4.6) it follows that we could calculate this energy by evaluating the path integral on the right hand side. Let us do this by the method of steepest descents. There are two stationary paths which satisfy the Euclidean equations of motion (the usual equations with the sign of V reversed) and the boundary conditions. There is first the solution $\bar{x}(\tau) = 0$. If V had risen monotonically with increasing x this would be the only stationary path, the action would be always positive, and quantum state associated with the classical minimum would be stable. Since,

however, V(x) turns over and again intersects the x-axis there is another "tunneling solution" $\bar{x}_t(\tau)$ which proceeds from x = 0 to the turning point and back again (Fig. 6). The tunneling solution is not a true minimum of the action as one can see from the expression for the action of small fluctuations about it.

$$I_2[\delta x(\tau)] = \frac{1}{2}\int_0^\tau d\tau \left[m\left(\frac{d\delta x(\tau)}{d\tau}\right)^2 + V''(\bar{x}_t(\tau))(\delta x(\tau))^2 \right] . \quad (4.12)$$

Here, V" is the second derivative of V(x). By choosing $\delta x(\tau)$ concentrated where V" < 0 the action can be made negative and thus less than its value for tunneling

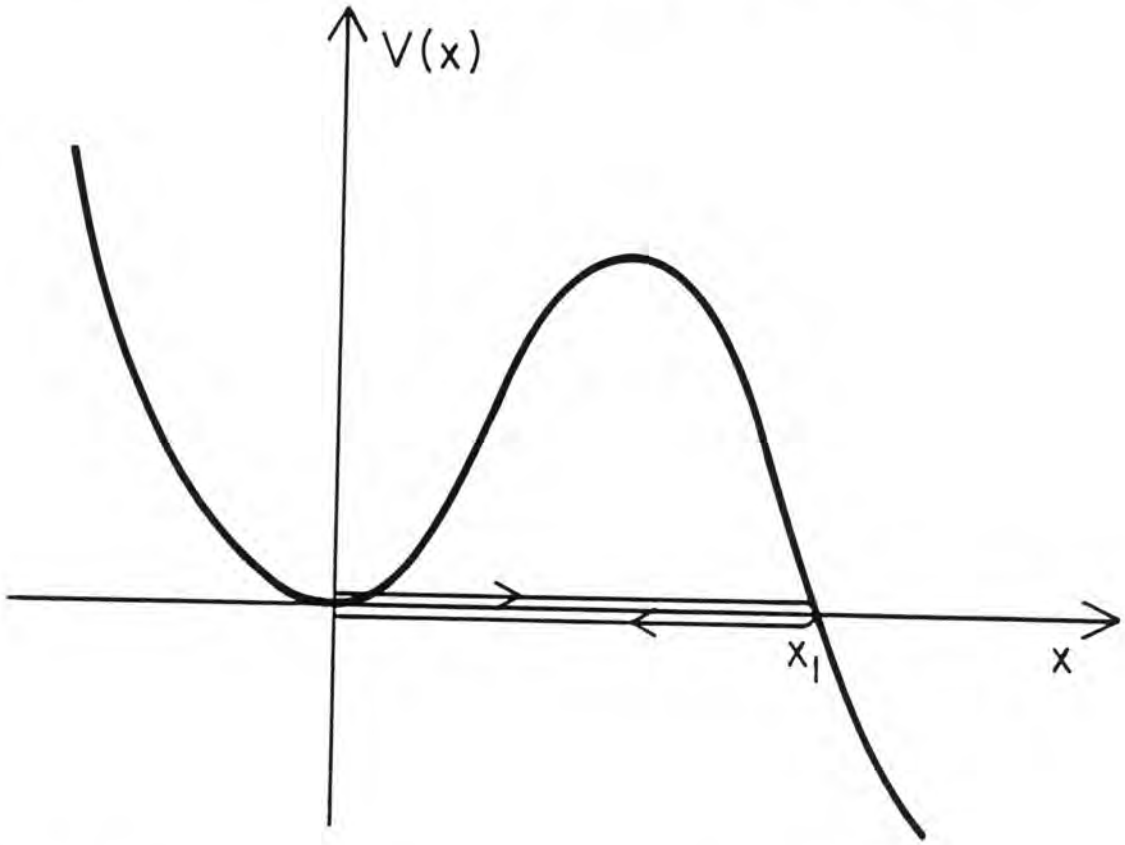

Fig. 6. The minimum of this potential at x = 0 is classically stable but quantum mechanically unstable via tunneling. This is reflected in the indefiniteness of the action for fluctuations about the tunneling motion from x = 0 to $x = x_1$ and back again.

solution.* Since the action is negative the path integral over the fluctuations will not converge. To keep it convergent the contour of path integration in x must be distorted into the complex plane in a way which can be justified by starting with a V(x) for which x = 0 is a global minimum and smoothly distorting it to the shape of Figure 6. In the process of this distortion the energy E_0 acquires an imaginary part. The former ground state thus becomes unstable when the action becomes negative.

Unlike the above example, the negative definiteness of the gravitational action does not signal a tunneling instability. One can see this clearly in the case of pure gravity and asymptotically flat spacetimes where the total energy is defined. Classically there is certainly a ground state. This is flat space and it has zero energy. The positive energy theorems guarantee that all other classical configurations have higher energy.[33] Because flat space is a global minimum of the energy we do not expect the associated quantum state to be unstable and indeed Witten[34] has shown that there are no tunneling solutions. Yet, the action is not positive definite.

The quantum mechanics of the fluctuations about flat space provides simple model in which the significance of the indefiniteness of the gravitational action can be understood. If we write

---

* One might worry that the positive kinetic energy term in (4.12) would defeat this argument. This is not the case.[32] An infinitesimal time translation of the tunneling solution is a zero mode with zero action of the quadratic operator defined by (4.12). It has one mode at $x = x_1$ so there must be a mode with lower, negative action.

$$g_{\alpha\beta} = \delta_{\alpha\beta} + h_{\alpha\beta} \quad , \tag{4.13}$$

then the Euclidean action (4.8) to quadratic order in the $h_{\alpha\beta}$ is

$$\ell^2 I_2[h] = \tfrac{1}{2} \int d^4x [\tfrac{1}{2} \nabla_\alpha \bar{h}_{\beta\gamma} \nabla^\alpha h^{\beta\gamma} - (\nabla_\alpha \bar{h}^{\alpha\beta})^2] \quad , \tag{4.14}$$

$$+ \text{(surface terms)}$$

where

$$\bar{h}_{\alpha\beta} = h_{\alpha\beta} - (1/2)\delta_{\alpha\beta} h \quad . \tag{4.15}$$

The action (4.14) is just that for a free spin-2 field in a flat background. Its quantum mechanics is equivalent to an assembly of independent harmonic oscillators and the ground state is, therefore, certainly stable. The action (4.14), however, is no more positive definite than that for the full theory (4.8). (Try $h_{\alpha\beta} = 2\delta_{\alpha\beta}\chi$ which is a linearized conformal transformation of flat space.) What's going on?

The point is that gravity formulated as a field theory in the metric is a theory expressed in terms of redundant variables. Part of the metric is arbitrary corresponding to the choice of coordinates (gauge) and the remainder is connected by the constraints. In fact, there are only two physical degrees of freedom. Issues concerning stability are best discussed in terms of these physical degrees of freedom, if they can be identified.

In linearized gravity one can fix a gauge, solve the constraints and identify the physical degrees of freedom. They are the two transverse-traceless "TT" parts of the metric fluctuations satisfying

$$\nabla^\alpha h^{TT}_{\alpha\beta} = 0 \quad , \qquad n^\alpha h^{TT}_{\alpha\beta} = 0 \quad , \qquad h^{TT\,\alpha}_{\alpha} = 0 \quad , \tag{4.16}$$

where n^α is the unit normal to the constant t surfaces.

It is not difficult to check that the two $h^{TT}_{\alpha\beta}$ are invariant under infinitesimal coordinate transformations and satisfy the linear versions of the constraints (3.38) and (3.39).

On the physical degrees of freedom the Euclidean action is

$$\ell^2 I_2[h^{TT}_{\alpha\beta}] = \frac{1}{4}\int d^4x[(\dot{h}^{TT}_{ij})^2 + (\nabla_i h^{TT}_{jk})^2] \quad , \tag{4.17}$$

where $(a_{ij\ldots})^2 = a_{ij\ldots}a^{ij\ldots}$. This is the action of two free fields. It is positive. Sums over Euclidean histories expressed in terms of the physical degrees of freedom will converge. Flat space is stable in linearized gravity.

The moral of the example of linearized gravity is that the quantum mechanics of a theory with redundant variables is best analyzed in terms of its physical degrees of freedom. Sums over histories, for example, can be carried out in the physical configuration space without gauge fixing and without ghosts. But it is often convenient to have the quantum mechanics expressed in terms of the redundant variables, for example, to display the invariances of the theory. In a theory like general relativity such a formulation is essential because it appears not to be possible to solve the constraints to exhibit the physical degrees of freedom explicitly. To pass from a sum over histories in the physical degrees of freedom to one in terms of redundant variables one simply adds back in the extra integrations in such a way as to not affect the value of the integral.[35] Typically, one might make use of identities like

$$1 = \int_{-\infty}^{+\infty} dx \left|\frac{\partial \Phi}{\partial x}\right| \delta(\Phi(x)) \quad , \tag{4.18}$$

for adding back in gauge variables and

$$1 = \int_{-\infty}^{+\infty} \frac{dx}{\sqrt{\pi M}}\, e^{-Mx^2} \quad , \tag{4.19}$$

for adding back in the gauge invariant ones. Identity (4.18) gives rise to "gauge fixing δ-functions" and "Faddeev-Popov determinants." Identity (4.19) modifies the action. One is free to use any identities one wants as long as they converge and the value of the sum over histories is left unaffected.

In the theory of linearized gravity integrations cannot be added to a sum over histories with the action (4.17) to obtain one with the action (4.14). The starting integral is convergent the resulting integral is divergent. One can, however, arrive at a coordinate invariant result by the following procedure[36]: Decompose the fluctuations as

$$h_{\alpha\beta} = \varphi_{\alpha\beta} + 2\delta_{\alpha\beta}\chi \quad , \tag{4.20}$$

and fix the decomposition by requiring one condition on $\varphi_{\alpha\beta}$. It is convenient to take

$$R_1(\varphi_{\alpha\beta}) = 0 \quad , \tag{4.21}$$

where R_1 is the linearized scalar curvature. The function χ is thus a gauge invariant scalar generating linearized conformal transformations. Then,

$$\ell^2 I_2[h] = \ell^2 I_2[\varphi_{\alpha\beta}] - 6\int d^4x (\nabla\chi)^2 \quad , \tag{4.22}$$

where the action I_2 on $\varphi_{\alpha\beta}$ is positive definite. A physical sum over histories cannot be manipulated into a form involving the action (4.22) but one can arrive at one with the form

$$\ell^2 I_2[h] = \ell^2 I_2[\varphi_{\alpha\beta}] + 6\int d^4x (\nabla\chi)^2 \quad . \tag{4.23}$$

This action is gauge invariant, O(4) invariant, positive definite, and it is physically equivalent to $I_2[h]$. It is the action to use in constructing convergent Euclidean functional integrals for fluctuations about flat space and use of this action gives the correct ground state wave function for linearized gravity.[37] A sum over histories based on the action (4.23) may be thought of as a sum based on the action (4.22) but carried out along a functional contour in which χ is purely imaginary. It is in this form that we shall find it most convenient to summarize the result.

For the full general theory of relativity, while we cannot explicitly identify the physical degrees of freedom, we can carry out a procedure analogous to that of linearized gravity. Consider first the case of the sums over histories which determine vacuum expectation values in asymptotically flat spacetimes with $\Lambda = 0$. These are integrals over asymptotically Euclidean spacetimes with the action (4.8). Split the integral over all metrics into an integration over a conformal factor and an integration over metrics in a conformal equivalence class. That is, write

$$g_{\alpha\beta} = \Omega^2 \tilde{g}_{\alpha\beta} \quad , \tag{4.24}$$

with $\Omega = 1$ at infinity and require

$$R(\tilde{g}) = 0 \quad . \tag{4.25}$$

If we write $\Omega = 1+Y$ and carry out the formal rotation $Y \to iY$, the action becomes [cf. (4.11)]

$$\ell^2 I_E[g] = \ell^2 I_E[\tilde{g}] + 6\int_M d^4x \sqrt{\tilde{g}} (\tilde{\nabla} Y)^2 \quad . \tag{4.26}$$

The last term is positive definite so the integral over the conformal factor converges. There remains the integral over metrics $\tilde{g}$ satisfying (4.25). The positive action theorem[38] shows that the action on such metrics is positive. These

integrals thus converge.

The analysis of the asymptotically flat case and the case of small fluctuations about flat space suggests that a similar procedure should be used to define the Euclidean functional integral giving the state of minimum excitation for closed cosmologies. One divides the integral into conformal equivalence classes in (4.24) using perhaps a condition of constant curvature rather that (4.25). One writes the conformal factor as $1 + Y$ where Y vanishes on the boundary where the argument of the wave function is given. One rotates Y into iY making the second term in (4.10) positive. The resulting action, however, is no longer manifestly positive. In fact it is complex. We do not yet have a demonstration that the resulting integrals converge but the preceding two examples give some hope that they may. We will write the prescription for the state of minimum excitation as

$$\Psi_0[h_{ij},\varphi] = \int_{C_0} \delta g \, \delta\varphi \exp(-I[g,\varphi]) \quad , \qquad (4.27)$$

where the C_0 indicates that an appropriate complex contour must be taken. This contour will ensure that Ψ_0 is real. However, since C_0 is complex, we cannot conclude that Ψ_0 is positive. In general it will oscillate and this will be important for its interpretation.

4.3 The Wave Function of the Universe

The wave function so naturally identified by the Euclidean functional integral prescription (4.27) displays many properties one would associate with a state of minimum excitation when analyzed in simple models as we shall show. As our own universe is not in a state of very high excitation, and as this state emerges so simply in the theory, it is a natural conjecture[2] that this wave function <u>is</u> the wave function of our universe and that the law (4.27) is the law specifying the initial conditions. We shall examine this conjecture in what follows.

5. MINISUPERSPACE MODELS

5.1 Minisuperspace

To test the conjecture that the wave function for our universe is that constructed by a sum over compact Euclidean four geometries, we want to calculate it and compare its predictions with the observations summarized in Section 2. The sum cannot be done exactly, we can only make approximations to it. Approximations may be constructed by singling out a family of geometries described by only a few parameters or functions and carrying out the sum only over geometries in this restricted class. Such a restriction on the 4-geometries which occur in the sum implies a restriction on the 3-geometries which can occur as arguments of the resulting wave function. This restriction reduces the configuration space on which the wave function is defined from the superspace of all 3-geometries to a smaller class - a minisuperspace. For this reason such approximations are called minisuperspace approximations.

One way of constructing a minisuperspace approximation is to restrict geometries and field configurations to have a certain symmetry. This type of approach has had a long and useful history in quantum cosmology.[39] Minisuperspace models based on symmetry are easy to implement and generally easy to interpret. They do not, however, offer the possibility of systematic improvement. That can be achieved in the lattice approximation to general relativity called the Regge calculus. There, curved geometries are built out of flat 4-simplices in much the same way that a geodesic dome is built out of triangles. The lengths of the edges of the simplices making up a 3-geometry become the parameters of a minisuperspace. Such simplicial minisuperspace approximations offer the hope of systematic improvement and are well adapted for the study of topological questions.[40]

Minisuperspace methods have already been extensively applied to construct approximations to the wave function of the universe.[2,25,28,41-61] We shall discuss only three of these models here and these can only be treated briefly. In two of these models the essential minisuperspace restriction is that geometries and field configurations be homogeneous and isotropic. The models differ in their assumptions about the matter. A conformally invariant scalar field gives a model which is not very realistic but easy to analyze. A massive scalar field provides a model which possesses many of the features of our universe.[2] Finally we discuss the model of Halliwell and Hawking[49] in which the origin of deviations from exact homogeneity and isotropy can be predicted.

5.2 Homogeneous, Isotropic Geometries with Scalar Field

5.2.1 Framework

The simplest class of minisuperspace models are those obtained by restricting the geometry to be homogeneous and isotropic and thus close to that of the present universe. The line element is then

$$ds^2 = \sigma^2[-N^2(t)dt^2 + a^2(t)d\Omega_3^2] \quad , \tag{5.1}$$

where $N(t)$ is an arbitrary lapse function and $\sigma^2 = \ell^2/24\pi^2$ is a normalizing factor chosen for later convenience. $d\Omega_3^2$ is the metric on the unit three sphere. The Euclidean histories with the same symmetries which enter into the sum defining the wave function have the metric

$$ds^2 = \sigma^2[N^2(\tau)d\tau^2 + a^2(\tau)d\Omega_3^2] \quad . \tag{5.2}$$

For the matter, we take a single scalar field with mass M, coupling to curvature ξ, and potential $V(\Phi)$ whose action generally is

$$I_\Phi = \frac{1}{2}\xi \int_{\partial M} d^3x h^{1/2} K\Phi^2$$
$$+ \frac{1}{2}\int_M d^4x g^{1/2}[(\nabla\Phi)^2 + \xi R\Phi^2 + M^2\Phi^2 + V(\Phi)] \qquad (5.3)$$

We restrict the matter field to be homogeneous following the symmetries of the geometry.

With these restrictions, the three geometry of a constant t spacelike surface is characterized by a single number a_0 and the field by its homogeneous value on the surface Φ_0. The minisuperspace is thus two dimensional and we write

$$\Psi_0 = \Psi_0(a_0, \Phi_0) \quad . \qquad (5.4)$$

The gravitational action on this minisuperspace is

$$I_E = \frac{1}{2}\int d\tau \left[\frac{N}{a}\right]\left[-\left(\frac{a\dot{a}}{N}\right)^2 - a^2 + H^2a^4\right] \quad , \qquad (5.5)$$

where

$$H^2 = \sigma^2\Lambda/3 \quad . \qquad (5.6)$$

(The H used in this section thus differs by the normalizing factor σ^2 from that defined previously.) The action for the matter may be conveniently written in terms of the rescaled variables

$$\varphi = (2\pi^2\sigma^2)^{1/2}\Phi \quad , \qquad m = \sigma M \quad ,$$
$$v(\varphi) = \sigma^2 V(\Phi) \quad . \qquad (5.7)$$

It is

$$I_\varphi = \frac{1}{2}\int d\tau \left(\frac{N}{a}\right)\left[\frac{a^4}{N^2}\left(\dot{\varphi} + 6\xi\frac{\dot{a}}{a}\varphi\right)^2 + 6\xi a^2\varphi^2\right.$$
$$\left. + a^4(m^2\varphi^2 + v(\varphi))\right] \quad . \qquad (5.8)$$

The kinetic energy part of (5.8) may be diagonalized by a further rescaling

$$\varphi = \chi/a^{6\xi} \quad . \qquad (5.9)$$

For the total action we then have

$$I = \frac{1}{2}\int d\tau \left(\frac{N}{a}\right)\left[-\left(\frac{\dot{a}}{N}a\right)^2 + \left(\frac{a^{2-6\xi}}{N}\dot{\chi}\right)^2 + U(a,\chi)\right] \quad , \qquad (5.10a)$$

where

$$U(a,\chi) = -a^2 + H^2 a^4 + 6\xi a^2\varphi^2 + a^4(m^2\varphi^2 + v(\varphi)) \quad . \qquad (5.10b)$$

and φ is understood to be a function of a and χ through (5.9).

The Hamiltonian constraint follows from the action by varying it with respect to the lapse N and expressing the resulting classical equation in terms of the variables and conjugate momenta. Recalling that, for example, $\pi_\chi = -i\partial L_{Euclidean}/\partial\dot{\chi}$ one finds

$$\frac{1}{2a}\left[-\pi_a^2 + a^{12\xi-2}\pi_\chi^2 + U(a,\chi)\right] = 0 \quad . \qquad (5.11)$$

We can write this as

$$\frac{1}{2}\, G^{AB}\pi_A\pi_B + \frac{1}{2a}\, U(a,\chi) = 0 \quad , \qquad (5.12)$$

where G_{AB} is a minisupermetric on our minisuperspace. In (a,χ) coordinates

$$G_{AB} = \begin{pmatrix} -a & 0 \\ 0 & a^{3-12\xi} \end{pmatrix} \quad . \qquad (5.13)$$

The Wheeler-DeWitt equation is the operator form of (5.12). In constructing it there are ambiguities of factor ordering which can only be resolved through a careful analysis of the measure of the sum over histories. As the precise form will not be very important for us we shall simply write

$$\frac{1}{2}\left[\nabla^2 - \frac{1}{a}\, U(a,\chi)\right]\Psi(a,\chi) = 0 \quad . \qquad (5.14)$$

Here, ∇^2 is the "covariant" Laplacian constructed from G_{AB},

$$\nabla^2 = \frac{1}{\sqrt{-G}} \frac{\partial}{\partial x^A}\left(\sqrt{-G}\, G^{AB} \frac{\partial}{\partial x^B}\right) \quad , \tag{5.15}$$

in "general coordinates" on the two dimensional minisuperspace or

$$\nabla^2 = -\frac{1}{a^{2-6\xi}} \frac{\partial}{\partial a}\left(a^{1-6\xi} \frac{\partial}{\partial a}\right) + \frac{1}{a^{3-12\xi}} \frac{\partial^2}{\partial \chi^2} \quad , \tag{5.16}$$

in the coordinates (a,χ). From either (5.11) or (5.16) we recognize that the Wheeler-DeWitt equation is hyperbolic with $\underline{a}$ being a "timelike direction" in the minisuperspace.

The wave function of minimum excitation, $\Psi_0(a_0,\chi_0)$, in these minisuperspace models is the sum of exp(-I) over all compact geometries of the form (5.2) with a single three sphere boundary of radius a_0 and over all regular configurations of scalar field which match χ_0 on the boundary. If we fix the last gauge freedom explicitly by taking $N = a$, and denote the time coordinate in this special gauge by η, then

$$\Psi_0(a_0,\chi_0) = \int_{C_0} \delta a\, \delta\chi \exp(-I[a,\chi]) \quad , \tag{5.17}$$

where

$$I[a,\chi] = \frac{1}{2} \int_{-\infty}^{0} d\tau \left[-a'^2 + (a^{1-6\xi}\chi')^2 + U\right] \quad . \tag{5.18}$$

and a prime denotes an η-time derivative. With this choice of gauge, the "south pole" of the geometry is located at $\eta = -\infty$ (see Fig. 7). We have used the residual η-time translation invariance to locate the boundary by $\eta = 0$. This fixes the limits of the η-coordinate range. We integrate over those $a(\eta)$ which vanish at $\eta = -\infty$ so the geometry is regular at its "south pole" and over field configurations $\chi(\eta)$

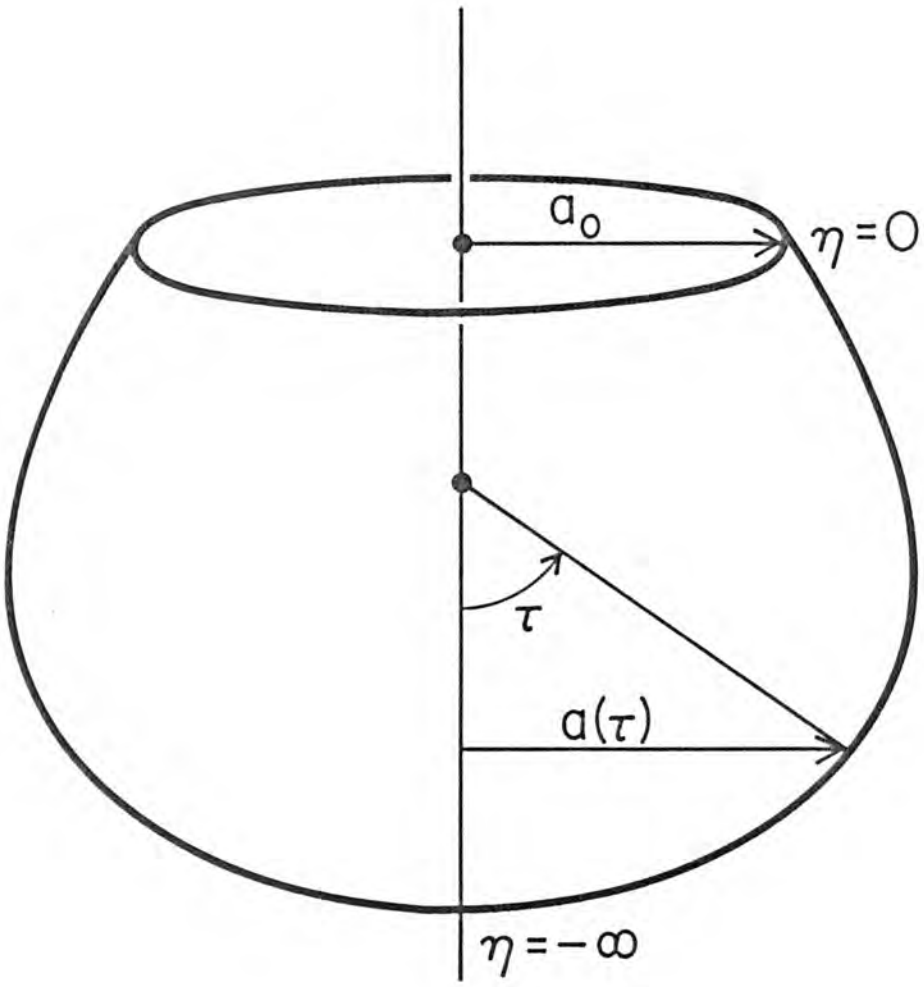

Fig. 7. A two dimensional representation of a homogeneous and isotropic 4-geometry contributing to the sum for the state of minimum excitation $\Psi_0(a_0, \chi_0)$. Shown embedded in a flat 3-dimensional space is a 2-dimensional slice of such a geometry whose intrinsic geometry is

$$d\Sigma^2 = d\tau^2 + a^2(\tau) d\varphi^2$$

τ is thus a "polar angle" and a the "radius from the axis." The geometry is compact and has only one boundary at which the radius is a_0, the argument of Ψ_0. In the time coordinate η such that $ad\eta = d\tau$ the "south pole" is located at $\eta = -\infty$ and the boundary at $\eta = 0$. The field configurations $\varphi(\tau)$ which contribute to the sum are those which are regular on this surface and which match the argument of the wave function Ψ_0 on the boundary.

which are regular on this geometry. Thus $\chi = 0$ at $\eta = -\infty$ when $\xi > 0$. Ψ_0 satisfies the Wheeler-DeWitt equation for the minisuperspace, (5.14). The functional integral supplies the boundary conditions for singling out the state of minimum excitation from among all other solutions of the Wheeler-DeWitt equation.

We shall now calculate Ψ_0 approximately for two different model actions for the scalar field. To do this we shall move back and forth between evaluating the integral and solving the equation.

5.2.2 A conformally invariant field

If $m = 0$ and $v = 0$ and $\xi = 1/6$ the scalar field is conformally invariant. This is not a very realistic model of matter - it lacks any particle physics scale, for example - but it does lead to an easily analyzable example. The reason is that geometries of the form (5.2) are conformally static. (To see this just put $N = a$.) Since the field is conformally invariant its dynamics are thus essentially trivial.

The action (5.18) for this special case reads

$$I = \frac{1}{2} \int_{-\infty}^{0} d\tau [-a'^2 - a^2 + H^2 a^2 + \chi'^2 + \chi^2] \quad . \qquad (5.19)$$

The scalar field action decouples from the gravitational one. Indeed, since the action for χ is just the Euclidean action for a harmonic oscillator the integral over χ in (5.17) is purely gaussian and easily evaluated to find

$$\Psi_0(a_0, \chi_0) = \exp(-\tfrac{1}{2}\chi_0^2)\, \Phi(a_0) \quad . \qquad (5.20)$$

The one homogeneous mode of the scalar field is in its ground state as one would expect for the state of minimum excitation.

Ψ_0 must satisfy the Wheeler-DeWitt equation and this gives a differential equation for Φ. Choosing the operator ordering as in (5.16) we have

$$-\frac{d^2\Phi}{da^2} + (a^2 - H^2a^4)\Phi = \Phi \quad . \tag{5.21}$$

This is just a "Schrödinger equation" for a particle in a potential

$$V(a) = a^2 - H^2a^4 \quad . \tag{5.22}$$

The boundary conditions for Φ are to be extracted from the integral (5.17). Under the simplest interpretation of the measure which is consistent with (5.16) we have $d\Phi/da = 0$ at $a = 0$. The overall normalization is as yet arbitrary. Some typical solutions are shown in Figure 8.

It will be of later interest to see how solutions to (5.21) arise semiclassically from (5.17). The integral defining $\Phi(a_0)$ is

$$\Phi(a_0) = \int_{C_0} \delta a \exp(-I_E[a]) \quad , \tag{5.23}$$

where I_E is (5.5) with $N = a$. Evaluation of (5.23) by the method of steepest descents gives the semiclassical approximation. For this we must find the extrema of I_E through which the contour of integration can be distorted. We begin with values of a_0 less than H^{-1}. The possible extrema of I_E are just the solutions of

$$a'' - a - 2H^2a^3 = 0 \quad . \tag{5.24}$$

The equation has an "energy integral" whose value may be found from the regular vanishing of a at $\eta = -\infty$. Expressing this integral in terms of τ gives

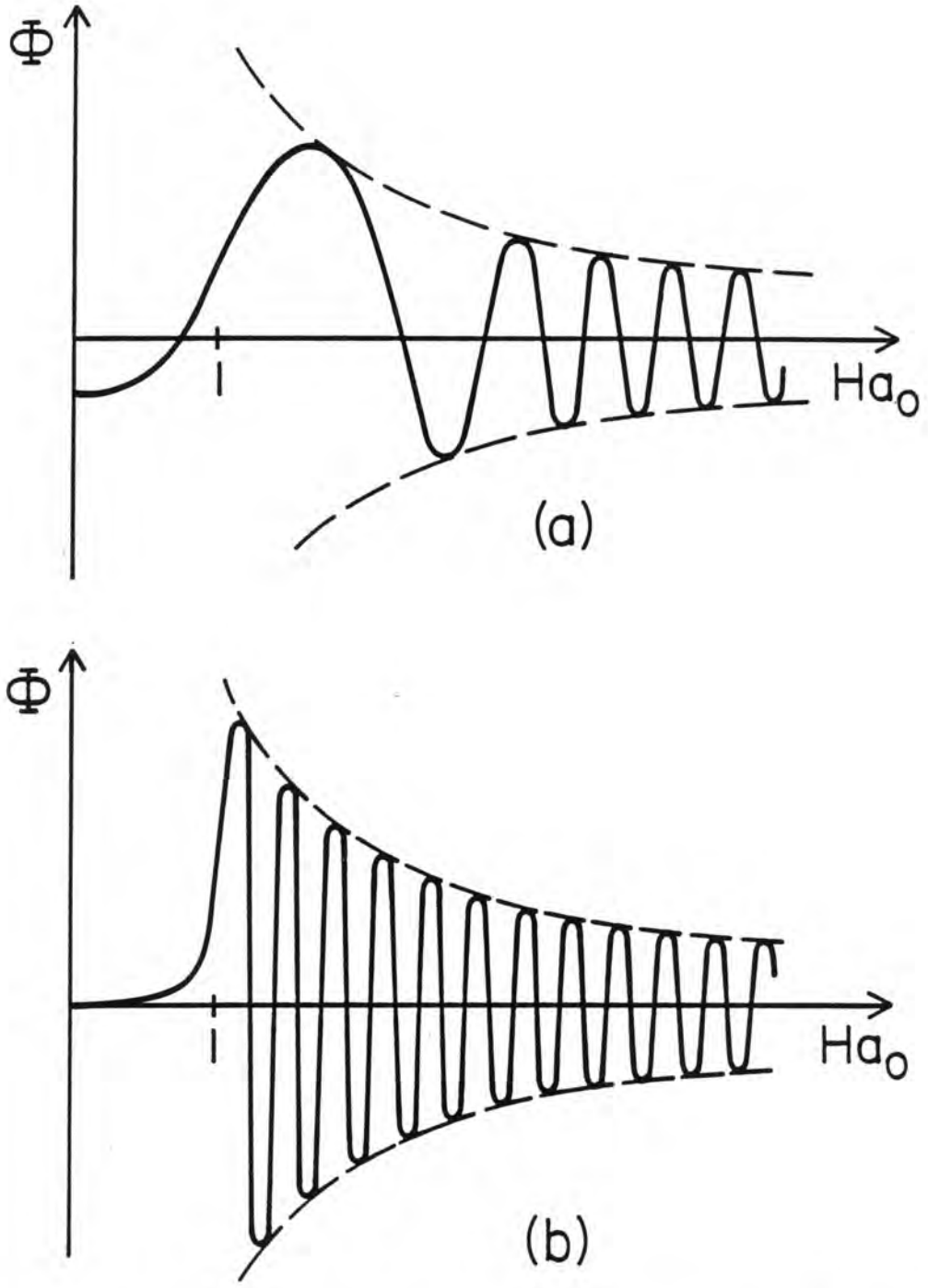

Fig. 8. The wave function Φ for the homogeneous, isotropic minisuperspace model with conformally invariant scalar field. Figure 8a shows a sketch of Φ for $H \approx 1$, Figure 8b for a much larger value of H. As H decreases the amplitude to find a 3-sphere of radius $a_0 < 1/H$ becomes very small. This is the classically forbidden region for de Sitter evolution (Figure 11). For $a_0 > 1/H$ the envelope approaches the distribution of 3-spheres in de Sitter space.

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{a^2} - H^2 \quad . \tag{5.25}$$

This is the Euclidean Einstein equation for a metric with the symmetries of the model as it must be. The solution is illustrated in Figures 9 and 10 and is just the 4-sphere of radius 1/H. For $a_0 < 1/H$ there are thus two possible extrema which are compact 4-geometries with a 3-sphere boundary of radius a_0. One for which the boundary bounds less than a hemisphere of the 4-sphere and another for which it bounds more. The action for the 4-sphere is negative and therefore one might think that the extremum encompassing more 4-sphere should dominate. One must remember, however, that because of the conformal rotation the contour of $\underline{a}$ integration is in the imaginary direction in the immediate vicinity of the extremum. Extrema of analytic functions are saddle points so that a maximum in a real direction is a minimum in an imaginary direction. The stationary configuration which contributes to the steepest descent evaluation of (5.23) is the one which is a maximum of the action in real directions and a least action configuration in imaginary directions. The extremum corresponding to the smaller part of the 4-sphere, therefore, provides the steepest descent approximation to the wave function. In fact, the contour cannot be distorted to pass through the other extremum. We thus have for $a_0 < 1/H$

$$\Phi(a_0) \approx N[-1+a_0^2-H^2a_0^4]^{-1/4} \times \exp\left[-\frac{1}{3H^2}(1-H^2a_0^2)^{3/2}\right] \quad , \tag{5.26}$$

where N is an arbitrary normalizing factor.

If a_0 is increased to a value larger than 1/H there are no longer any real extrema because a 3-sphere of radius

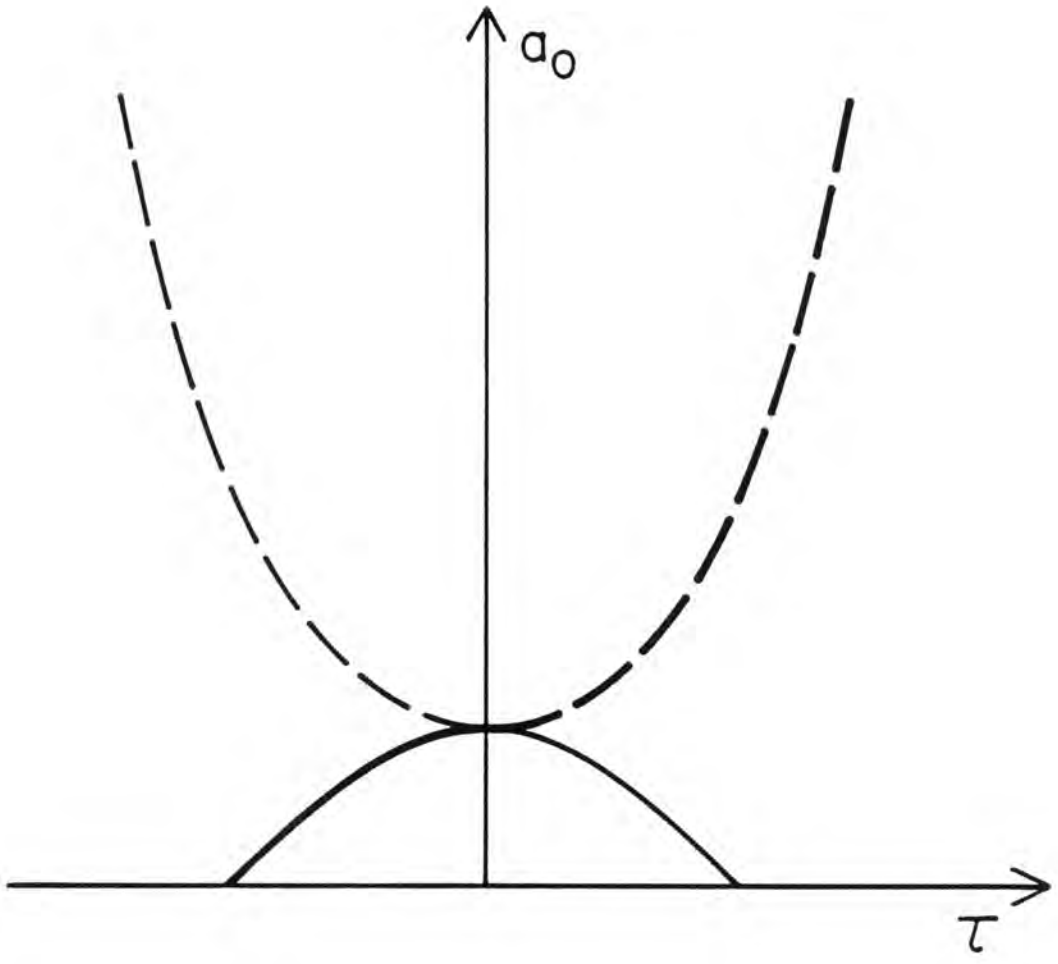

Fig. 9. The extremizing scale factor for the homogeneous, isotropic minisuperspace model with conformally invariant scalar field. The solid line is the solution of (5.25) for real Euclidean extrema of the action. The complete range of a from zero to maximum and back again describes the geometry of the 4-sphere (Figure 10). The dashed curve is the solution of (5.27) for complex Euclidean (Lorentzian) extrema. It describes the geometry of de Sitter space (Figure 11). For each value of a_0 there are thus two possible extremizing solutions. Choosing the trajectory to start on the left at $a_0 = 0$ the Euclidean prescription for the state of minimum excitation singles out the heavy curve shown. This gives the semiclassical approximation to the wave function Ψ_0.

$a_0 > 1/H$ cannot fit into a 4-sphere of radius $1/H$. There are, however, complex extrema. These can be obtained by changing $\tau \to \pm i t$ in Eq. (5.25) so they solve

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 - \frac{1}{a^2} \quad . \tag{5.27}$$

They are thus the solutions of the Lorentzian Einstein equations with positive cosmological constant. This solution is called de Sitter space (Figure 11). These complex extrema must contribute in complex conjugate pairs so that the wave function is real. By a standard WKB matching analysis we can establish the form of the wave function for $a_0 > H^{-1}$

$$\Phi(a_0) \approx 2N\left[H^2 a_0^4 - a_0^2 + 1\right]^{1/4} \times \cos\left[\frac{(H^2 a_0^2 - 1)^{3/2}}{3H^2} - \frac{\pi}{4}\right] \quad . \tag{5.28}$$

This form could be derived by carefully following the extremum configuration as a_0 is increased along the heavy curve shown in Fig. 9.

The complete wave function Ψ_0 on the minisuperspace of homogeneous isotropic geometries with conformally invariant scalar field is given by

$$\Psi_0(a_0, \varphi_0) = \exp\left[-\varphi_0^2/(2a_0^2)\right]\Phi(a_0) \quad , \tag{5.29}$$

where Φ is given approximately by (5.26) and (5.28). From this correlations between field and geometry can be extracted. The exponential factor gives an inverse correlation between φ_0 and a_0. Large φ_0 occurs at small a_0 and vice versa. This is the type of correlation that occurs in classical evolution.

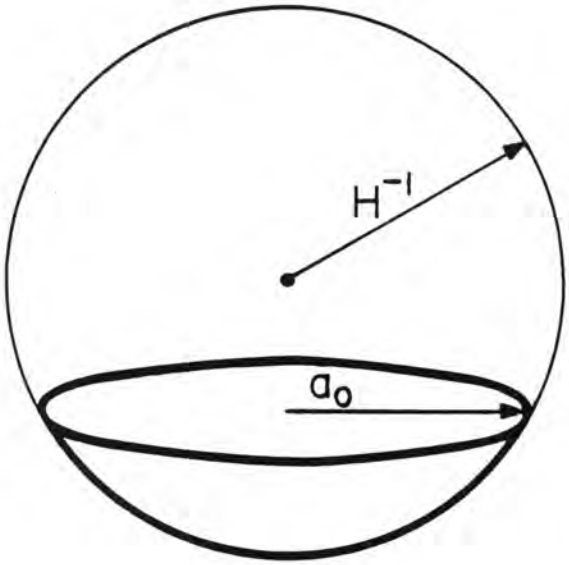

Fig. 10. The real Euclidean extrema of the homogeneous, isotropic minisuperspace model with conformally invariant scalar field have the geometry of a 4-sphere of radius H^{-1}. The extremizing configuration which gives the semi-classical approximation to Φ at $a_0 < 1/H$ is a part of the 4-sphere with a single 3-sphere boundary of radius a_0. There are two possibilities corresponding to more than a hemisphere or less. The Euclidean functional integral prescription for Φ identifies the smaller part of the 4-sphere as the contributing extremum. For $a_0 > H^{-1}$ there are no real extrema.

The factor Φ suppresses correlations for $a_0 < H^{-1}$ but affects them only weakly for $a_0 > H^{-1}$. The classical solution of Einstein's equation restricted by the minisuperspace assumptions is de Sitter space. This homogeneous, isotropic, empty geometry may be thought of (Figure 11) as the evolution of a 3-sphere which contracts to a minimum radius H^{-1} and reexpands. The region of minisuperspace with $a_0 < H^{-1}$ is thus classically forbidden and Φ suppresses correlations there. The region $a_0 > H^{-1}$ is classically allowed and there Φ varies only weakly.

The analysis leading to (5.28) gives another way of summarizing the information contained in Ψ_0 in the semiclassical limit. In the classically allowed region, $a_0 > H^{-1}$, the semiclassical approximation to Ψ_0 is given by a complex Euclidean but real Lorentzian extremum of the action. The wave function oscillates proportionally to $\cos(S)$ where S satisfies the Lorentzian Hamilton-Jacobi equation. This action specifies a solution to the equations of motion up to initial conditions. In this semiclassical approximation the wave function thus corresponds to an ensemble of Lorentzian de Sitter spaces which differ from one another only in the time assigned the minimum radius. It is in this way we recover the classical limit.

5.2.3 A massive scalar field

The conformally invariant scalar field is not a realistic model of the matter in the universe. It contains no scale. Within the general framework of the minisuperspace models discussed in Section 5.2.1, a more realistic model is provided by a free, massive, minimally coupled scalar field. This was discussed by Hawking[2] in the case $\Lambda = 0$. The parameters of this model are thus $\xi = 0$, $\Lambda = 0$, $v = 0$, and $m \neq 0$.

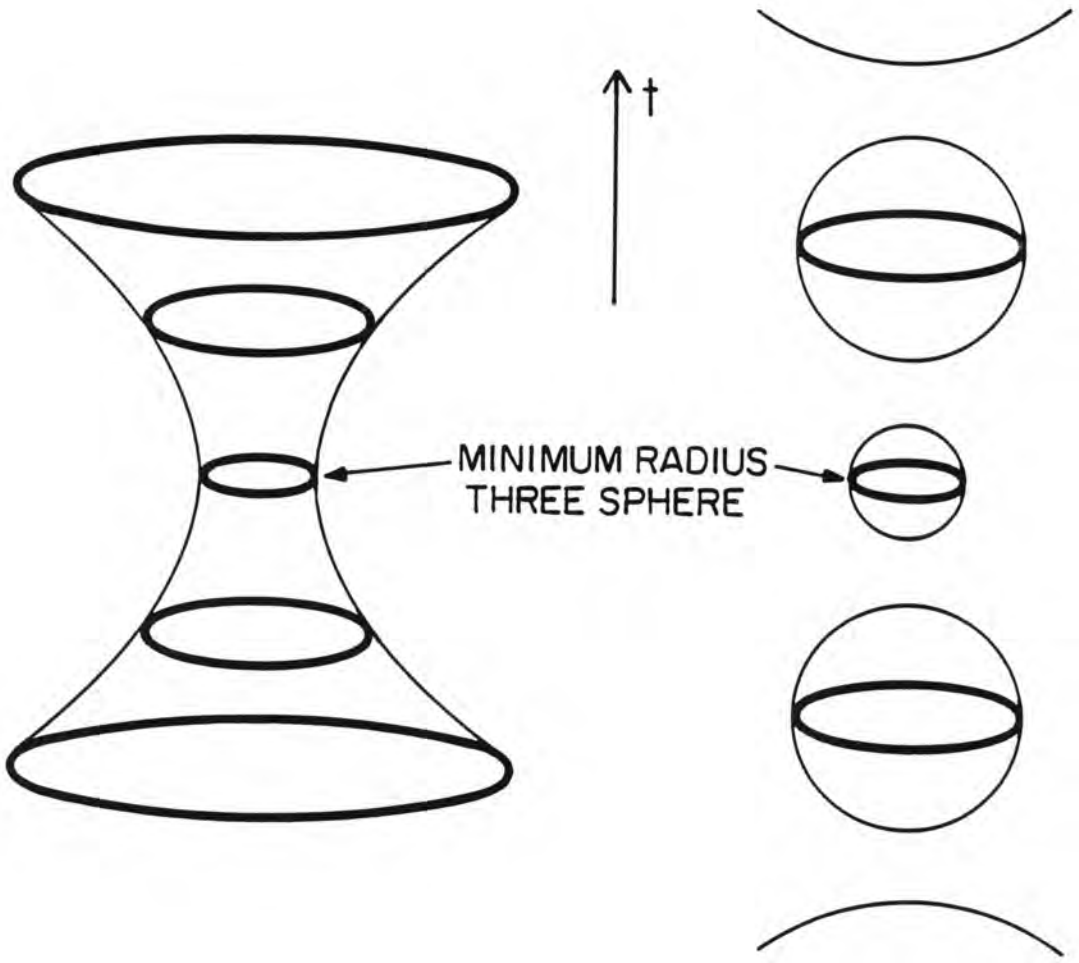

Fig. 11. Lorentzian de Sitter Space. In the classically allowed region of the minisuperspace of homogeneous isotropic geometries with conformally invariant scalar field, the wave function corresponds semiclassically to Lorentzian de Sitter space. This is the most symmetric solution of the source free Einstein's equation with positive cosmological constant. It is the geometry of a Lorentz hyperboloid in a 5-dimensional Lorentz signatured spacetime. It may be thought of as a three sphere which collapses to a minimum radius $(3/\Lambda)^{1/2}$ and then re-expands.

The metric on minisuperspace in this model is [(5.13)]

$$G_{AB} = a \begin{pmatrix} -1 & 0 \\ 0 & a^2 \end{pmatrix} \quad . \tag{5.30}$$

This is conformal to the metric on the interior of the forward light cone in a two dimensional Minkowski space. To see this, introduce new coordinates

$$\begin{aligned} x &= a \sinh\varphi \\ y &= a \cosh\varphi \end{aligned} \tag{5.31}$$

($\chi = \varphi$ in this case when $\xi = 0$.) The metric in x, y coordinates is

$$G_{AB} = (y^2-x^2)^{1/2} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad . \tag{5.32}$$

The new and old coordinates can be conveniently plotted on an x-y diagram as in Figure 12a. The Wheeler-DeWitt equation becomes

$$\left[- \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2} - U(x,y) \right] \Psi(x,y) = 0 \quad , \tag{5.33}$$

where U(x,y) is the "potential" of eq. (5.10b). Expressed in terms of x and y it is

$$U(x,y) = x^2-y^2+m^2(y^2-x^2)^2 \left[\tanh^{-1}\left(\frac{x}{y}\right)\right]^2 \quad . \tag{5.34}$$

The first term is from the spatial curvature in the Wheeler-DeWitt equation. The second is the contribution of the scalar field's mass to its energy.

Eq. (5.33) is a wave equation with potential in one space and one time dimension. It could be integrated numerically if boundary conditions could be found for Ψ_0. These boundary conditions are supplied by sum over compact Euclidean histories (5.17) which defines the quantum state

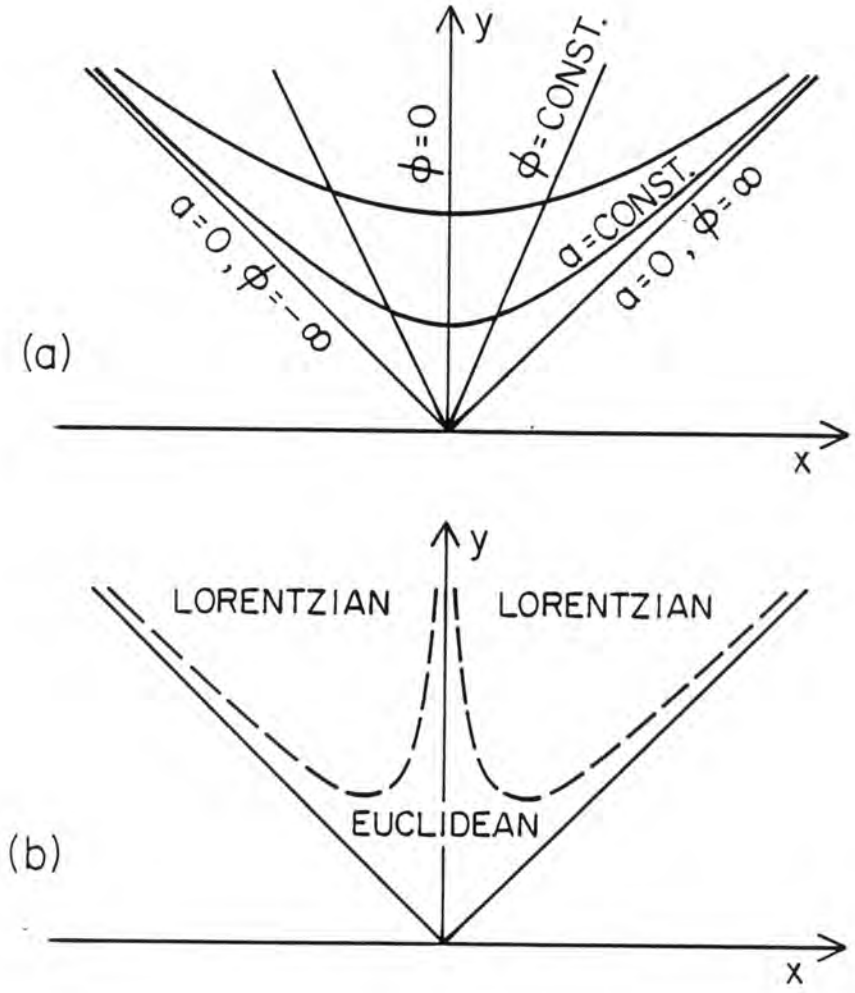

Fig. 12. (a) Two sets of coordinates on the minisuperspace of homogeneous isotropic geometries with minimally coupled massive scalar field. The minisuperspace is conformal to the forward light cone of a 2-dimensional Minkowski space (y,x). Curves of constant spatial volume are the hyperbolae a = constant. Curves of constant scalar field are straight lines through the origin.

(b) The classically allowed and classically forbidden regions of minisuperspace. In the classically allowed region the semiclassical approximation is given by a Lorentzian extremum of the action and the wave function oscillates. In the classically forbidden regions the semiclassical approximation is given by a Euclidean extremum and the wave function is nonoscillatory.

of the universe and distinguishes it from all other solutions of the Wheeler-DeWitt equation. A convenient place to evaluate these boundary conditions is the characteristic surface $y = |x|$ corresponding to $a = 0$, $\varphi = \pm \infty$. The boundary condition needed to integrate the Wheeler-DeWitt equation is the value of Ψ_0 on this surface. To evaluate this and also to obtain a qualitative understanding of its behavior we consider the semiclassical approximation to (5.17).

Semiclassicaly, we expect to find

$$\Psi_0(a_0,\varphi_0) \approx A(a_0,\varphi_0)\exp[-I(a_0,\varphi_0)] \quad , \tag{5.35}$$

for those values of a_0 and χ_0 for which there is a real Euclidean extremum of the action. For those values for which the extremum is complex Euclidean (i.e. real Lorentzian) we expect

$$\Psi_0(a_0,\varphi_0) \approx A(a_0,\varphi_0)\cos[S(a_0,\varphi_0)] \quad , \tag{5.36}$$

where S is the Lorentzian action. In this case the wave function will oscillate. The equations of motion which a Euclidean extremum must satisfy follow from varying (5.10a) with respect to N and φ. In the gauge where $N = 1$ they are

$$\ddot{\varphi} + \frac{3\dot{a}}{a}\dot{\varphi} - m^2\varphi^2 = 0 \quad , \tag{5.37a}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{a^2} + \dot{\varphi}^2 - m^2\varphi^2 \quad . \tag{5.37b}$$

A solution which is a compact geometry starts at some τ (say $\tau = 0$) where $a = 0$ (Figure 7). There, $\dot{\varphi} = 0$ in order for the field to be regular. The value $\varphi(0)$ is arbitrary. There are thus a one parameter family of solutions to (5.37) which correspond to compact geometries with regular field configurations. A typical one looks schematically like Figure 13. The solution whose action gives the wave function

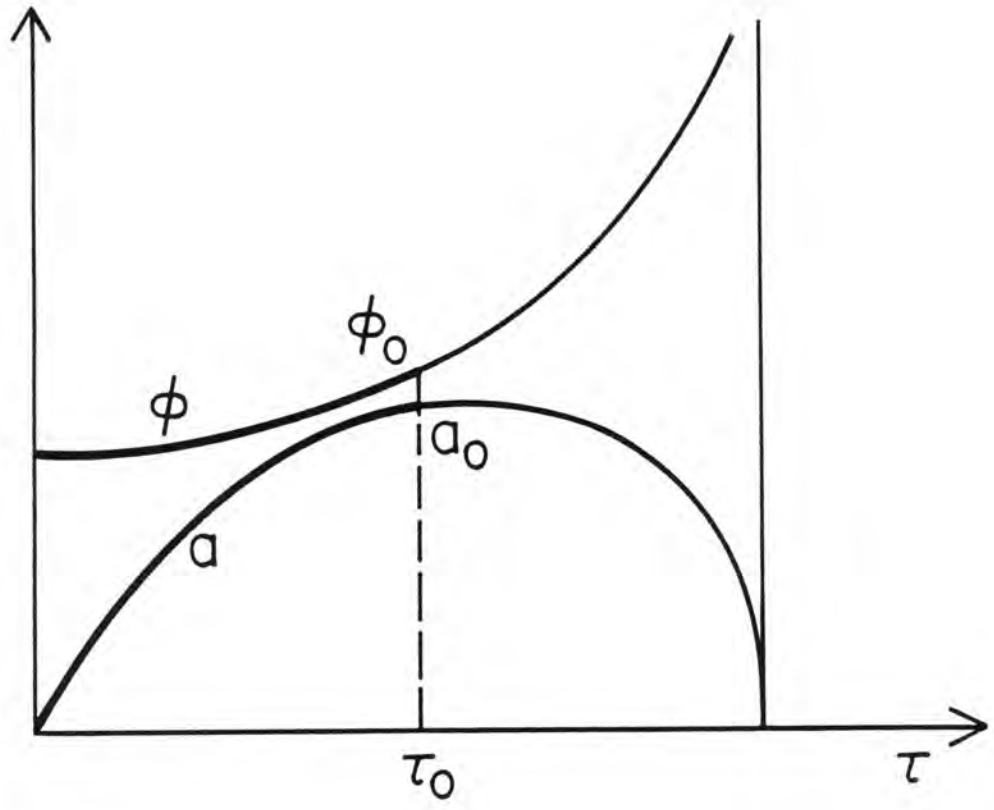

Fig. 13. A typical Euclidean extremizing configuration. The geometry has a "south pole" where $a = 0$, $\dot{\varphi} = 0$ and φ takes some value $\varphi(0)$. (cf. Figure 7.) If eqs. (5.37) are integrated forward in the "polar angle" τ a "north pole" is eventually reached where $a = 0$ and both geometry and field are singular. A compact, non-singular geometry with 3-sphere boundary on which $a = a_0$ and $\varphi = \varphi_0$ is obtained by locating the τ_0 for which $a = a_0$ and then varying $\varphi(0)$ until $\varphi = \varphi_0$ at that value of τ_0. This extremizing configuration is indicated by the heavier curves.

at a given value of a_0 and φ_0 is determined by adjusting $\varphi(0)$ until φ assumes the value φ_0 at the value of τ for which $a = a_0$. There may be several solutions for which this is possible. This phenomenon is familiar from the example of the conformally invariant scalar field where there were solutions to the equation of motion (5.25) corresponding to less than a hemisphere of the 4-sphere and also to more than a hemisphere. The solution which gives the semiclassical approximation, in that case as in this, is the one for which the action is a maximum for real variations. Explicit calculation[2,54] shows that such solutions are confined to a region of mini-superspace for which approximately

$$a_0 < (m|\varphi_0|)^{-1} \quad \text{or} \quad |\varphi_0| \lesssim 1 \quad . \tag{5.38}$$

Inside this region the wave function varies without oscillation. Outside this region it oscillates. In the semiclassical approximation this behavior emerges because, outside the region (5.38), the extrema are complex Euclidean and the wave function behaves as in (5.36). These complex extrema are real Lorentzian geometries and field histories which obey

$$\ddot{\varphi} + \frac{3\dot{a}}{a}\dot{\varphi} + m^2\varphi = 0 \quad , \tag{5.39a}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = -\frac{1}{a^2} + \dot{\varphi}^2 + m^2\varphi^2 \quad , \tag{5.39b}$$

a dot denoting differentiation with respect to Lorentzian time. Indeed, semiclassically, the quantum state may be thought of as corresponding to an ensemble of classical histories which obey these equations.

The values $a_0 \to 0$ and $\varphi_0 \to \pm\infty$, of interest for evaluating the characteristic initial value data for the Wheeler-DeWitt equation, are within the region of mini-superspace defined by (5.38). There is therefore always

a real solution of (5.37) giving the semiclassical approximation to Ψ_0. The action approaches zero as $a_0 \to 0$ since φ is regular. Thus, to the extent the variation in the prefactor A can be neglected* in (5.35) we have

$$\Psi_0 = \text{constant} . \tag{5.40}$$

on the characteristic initial value surface $y = |x|$.

A numerical integration of the Wheeler-DeWitt equation with the boundary condition (5.40) is shown in Figure 14. (For more see Refs. 41, 51, 63.) In the classically forbidden region (Figure 12b, Eq. (5.38)) the wave function varies smoothly. In the classically allowed region it oscillates.

Figure 15 shows a schematic representation of a typical solution to (5.39). For sufficiently large φ the explicit computation[2,54] shows that the transition from Euclidean to Lorentzian extremum occurs when $a \approx (m|\varphi|)^{-1}$ [cf.(5.38)] and when $\dot{\varphi}$ and $\dot{a}$ are approximately zero. From (5.39), a classical trajectory thus starts from a field value φ_1 which is a local maximum and a scale factor which is a local minimum, $a_1 \approx (m|\varphi_1|)^{-1}$. In the subsequent evolution, while φ is nearly constant the term $m^2\varphi^2$ in (5.39b) behaves as an effective cosmological constant. The universe thus inflates with a time scale $m|\varphi_1|$. Eventually φ decreases, begins to oscillate and the matter field acquires kinetic energy of its own.

In the classically allowed region of minisuperspace the wave function constructed as the sum over compact, regular histories corresponds semiclassically to an ensemble of classical histories each characterized by a value of φ_1. The histories have an initial inflationary

---

*The prefactor in the case of pure gravity has been evaluated by K. Schleich.[62]

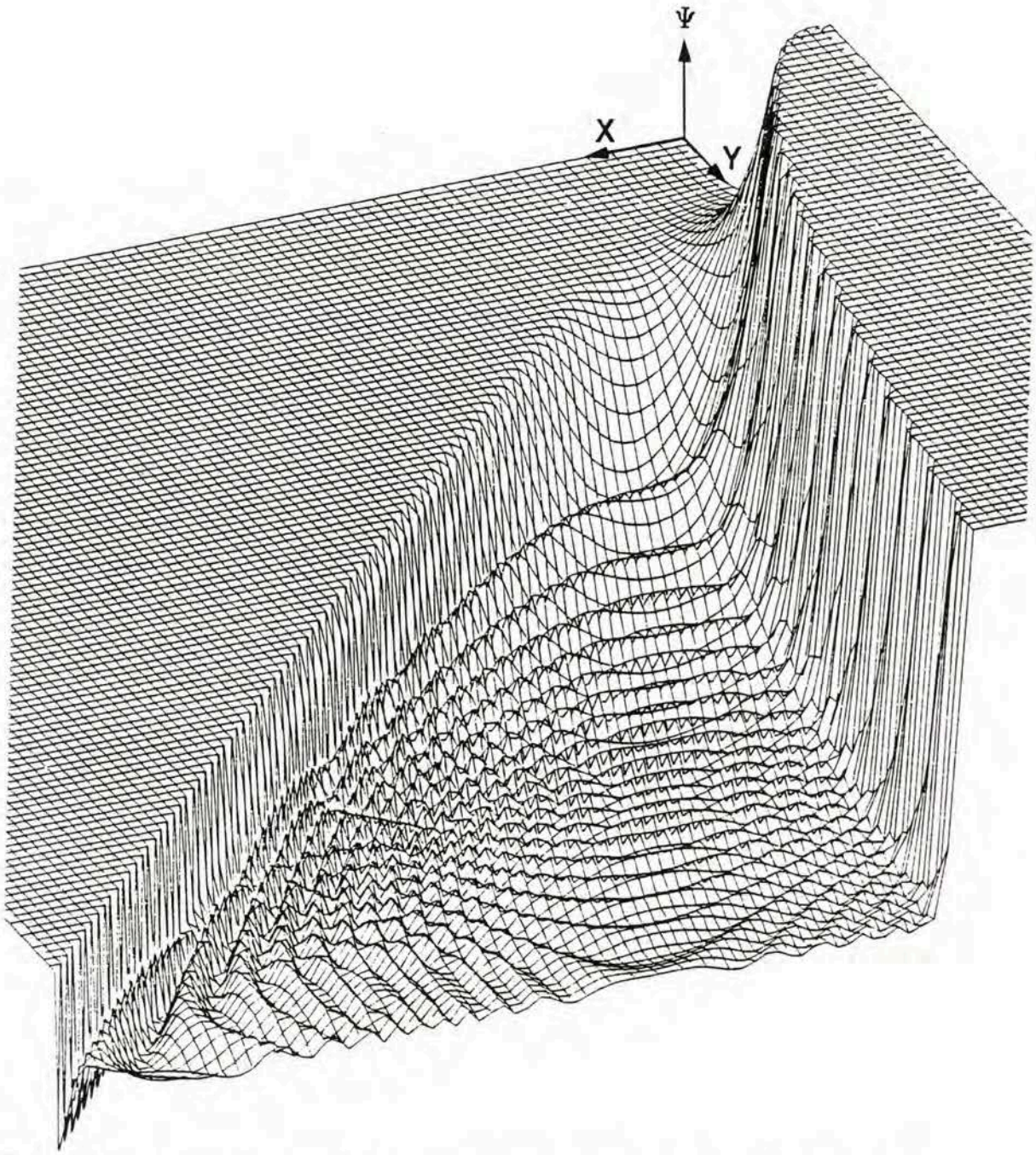

Fig. 14. A numerical integration of the Wheeler-DeWitt equation in the homogeneous, isotropic, massive scalar field minisuperspace model. The figure shows the integration of E. P. Shellard[63] of the Wheeler-DeWitt equation (5.33) with the boundary conditions (5.39). Ψ_0 is plotted as a function of the coordinates x and y. Oscillatory and non-oscillatory regions can clearly be distinguished and these correspond to those predicted semiclassically and shown in Figure 12b.

epoch and, later, an epoch in which the matter field has kinetic energy and the expansion behaves as though matter dominated. Thus through the inflationary mechanism, the universe, although in the analog of the ground state, can become large, approximately flat and contain matter.

5.3 Linearized Fluctuations about Homogeneous Isotropic Models

Minisuperspace models which assume homogeneity and isotropy can neither provide an explanation of this large scale feature nor of the observed spectrum of fluctuations away from it. On the one hand an explanation of homogeneity and isotropy can only come by comparing the wave function on configurations which have these symmetries with those which do not, and on the other fluctuations cannot be studied in geometries which do not have them. To progress with either question one needs to enlarge the minisuperspace to include geometries which are not homogeneous and isotropic. Of the several models of this type[44,47,48] perhaps the most complete is that of Halliwell and Hawking.[49] They considered linear fluctuations away from the homogeneous and isotropic models with massive scalar field discussed in the preceding subsection. They discuss the most general fluctuations and thus explore completely a small domain of superspace about exact homogeneity and isotropy. Their model is thus not strictly a minisuperspace model but contains an infinite number of degrees of freedom.

In the following we shall sketch the assumptions and method of Halliwell and Hawking's calculation and quote some of their results.

The model considers Euclidean histories which deviate only slightly from exact homogeneity and isotropy. Metric and field can thus be written

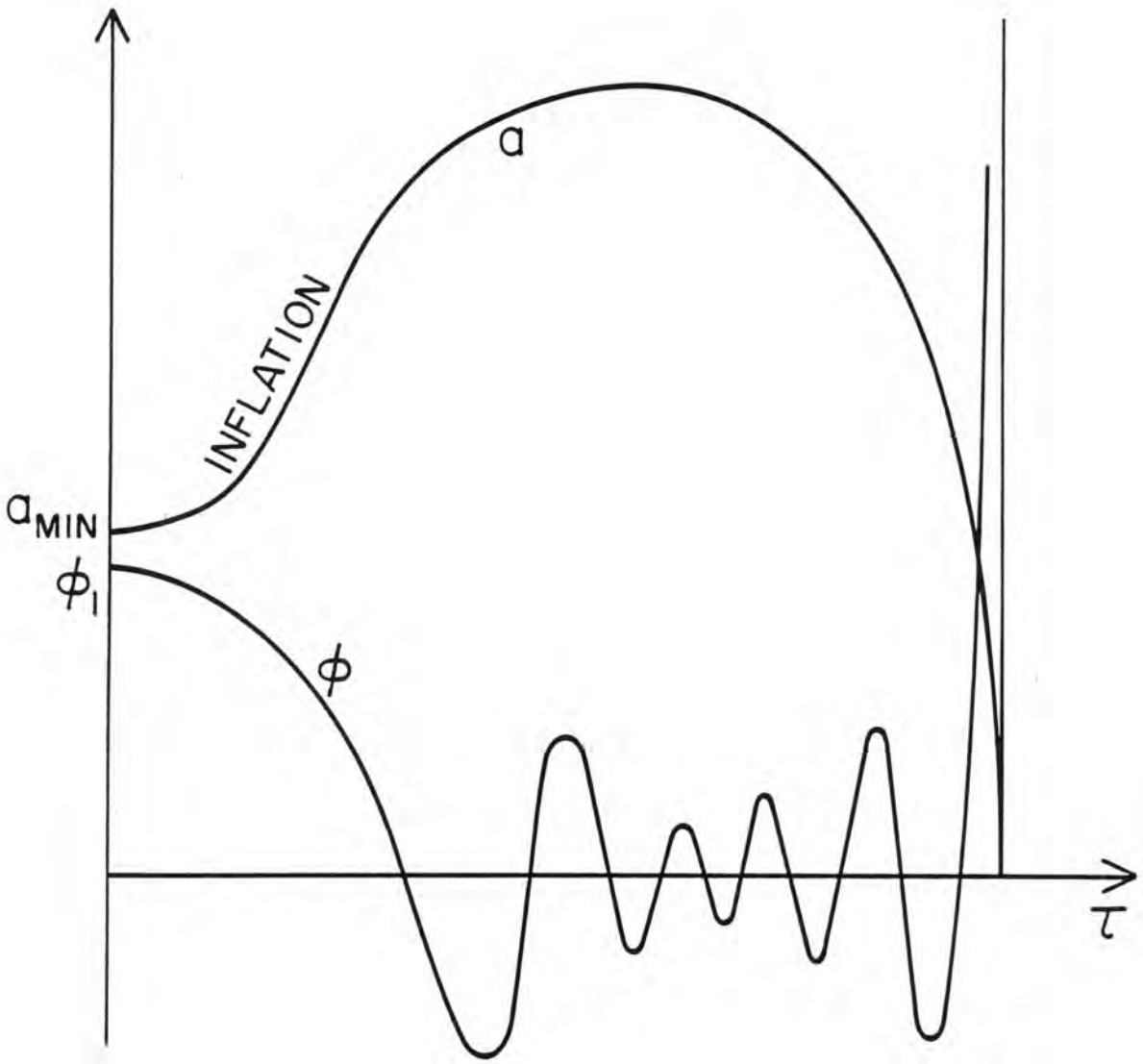

Fig. 15. A typical Lorentzian extremizing configuration. The solution shown schematically here starts at a minimum radius with $\dot{\varphi} = 0$. In the domain where φ varies slowly the universe follows a de Sitter like inflationary expansion with $H = m\varphi_1$. Later the scalar field begins to oscillate and the universe evolves approximately as though matter dominated. Eventually a maximum expansion is reached, the universe recollapses and matter and geometry become singular. A sufficiently large $m\varphi_1$ would provide a long enough inflationary period to explain the present large size of the universe and its approximate spatial flatness. The oscillation of the scalar field models the creation of matter.

$$ds^2 = d\hat{s}^2 + \epsilon_{\alpha\beta}(x)\,dx^\alpha dx^\beta \quad , \tag{5.41a}$$

$$\varphi = \hat{\varphi}(\tau) + f(x) \quad . \tag{5.41b}$$

Here $d\hat{s}^2$ is the homogeneous, isotropic line element (5.2), $\hat{\varphi}$ a homogeneous field configuration and $\epsilon_{\alpha\beta}$ and f the inhomogeneous deviations from these quantities. Both $\epsilon_{\alpha\beta}$ and f can be expanded in the harmonics of the homogeneous, isotropic 3-sphere. Schematically, these expansions have the form

$$\epsilon_{\alpha\beta}(\tau,x^i) = \Sigma_{(n)}\,\epsilon^{(n)}(\tau)Q^{(n)}_{\alpha\beta}(x^i) \quad , \tag{5.42a}$$

$$f(\tau,x^i) = \Sigma_{(n)}\,f^{(n)}(\tau)Q^{(n)}(x^i) \quad , \tag{5.42b}$$

where $Q^{(n)}_{\alpha\beta}$ and $Q^{(n)}$ are a complete set of tensor and scalar harmonics. Near the homogeneous, isotropic configurations, the $\epsilon^{(n)}$ and $f^{(n)}$ may be regarded as "coordinates" on superspace and we write

$$\Psi_0 = \Psi_0(\hat{a},\hat{\varphi},\epsilon^{(1)},f^{(1)},\epsilon^{(2)},f^{(2)},\cdots) \quad . \tag{5.43}$$

Classically the linearized field equations for the fluctuations $\epsilon^{(n)}$, $f^{(n)}$ decouple into a set for each mode. Quantum mechanically, the Wheeler-DeWitt equation and the associated constraints (4.9a) separate when written to quadratic order in the variables describing the fluctuations. That is, the wave function is a product

$$\Psi_0 = \hat{\Psi}(\hat{a},\hat{\varphi})\prod_{(n)}\psi^{(n)}(\hat{a},\hat{\varphi},\epsilon^{(n)},f^{(n)}) \quad . \tag{5.44}$$

The Wheeler-DeWitt equation implies for $\hat{\Psi}$ an equation of the form

$$\left[\nabla^2 - \frac{1}{2a}U(a,\varphi) + \begin{pmatrix}\text{source term}\\ \text{quadratic}\\ \text{in the } \psi^{(n)}\end{pmatrix}\right]\hat{\Psi}(\hat{a},\hat{\varphi}) = 0 \quad . \tag{5.45}$$

The first two terms in this equation are the Wheeler-DeWitt equation of the homogeneous, isotropic massive scalar field model [eq. (5.14)]. The additional term represents the expected energy of the fluctuations. In the semiclassical approximation one expects solutions for $\hat{\Psi}$ of the form (5.35) or (5.36).

In the classically allowed region, the semiclassical approximation to $\hat{\Psi}$ is of the form (5.36). The action $S(a_0, \varphi_0)$ is the action of classical histories $a(t)$, $\varphi(t)$ which satisfy the classical equations of motion. When $\hat{\Psi}$ is approximated in this way, the Wheeler-DeWitt equation implies a "Schrödinger" equation for $\psi^{(n)}$

$$i \frac{\partial \psi^{(n)}}{\partial t} = H^{(n)} \psi^{(n)} \quad . \tag{5.46}$$

$\psi^{(n)}$ and $H^{(n)}$ become functions of time through the connection between a and t provided by the classical trajectory $a(t)$. This is generally the way that the notion of time is recovered in the semiclassical approximation to the quantum dynamics of spacetime.[64] The dynamics of the fluctuations, in effect, becomes the ordinary quantum dynamics of fields moving in the background spacetime which provides the semiclassical approximation to $\hat{\Psi}$.

Halliwell and Hawking solve (5.45) and (5.46) with boundary conditions extracted from the Euclidean sum over histories specification of the wave function. They argue that for this wave function the additional source term in (5.44) is small after appropriate renormalization. The wave function $\hat{\Psi}$ and the classical trajectories which give its semiclassical approximation are thus those of the minisuperspace model discussed in the previous subsection. These display an early inflationary phase followed by a transition to a matter dominated evolution. Eq. (5.46)

shows that the fluctuations will evolve as quantum fields in this background spacetime. Semiclassically the evolution will therefore be much like the evolution of fluctuations in the standard inflationary universe history.[65] What the wave function of the universe supplies is the boundary conditions to begin this evolution.

Halliwell and Hawking find first that at all stages in which their approximation is valid the wave function is peaked about isotropy and homogeneity. The fluctuations away from this symmetry begin in their ground state. They remain in their ground state until they expand outside the Hubble radius. Their amplitude then remains frozen until they reenter the Hubble radius in the matter dominated era. There results a scale free (Zel'dovich) spectrum of fluctuations which, for the correct choice of the scalar field mass m, can have the amplitude to correctly reproduce the spectrum of density fluctuations we observe.

6. OBSERVATIONS AND PREDICTIONS

The accompanying table offers a comparison between the large scale observations of our universe reviewed in Section 2, and the predictions of the proposal for the quantum state of universe worked out in minisuperspace models. The predictions are consistent with the observations. Essential to this consistency is the action of an inflationary mechanism when the universe is small. Without such action there would be no natural way for the cosmological analog of the ground state to explain the large size of the universe, its approximate spatial flatness and its matter content. It is encouraging that inflation appears to occur naturally in a wide range of matter models.[2,43,50] Inflation, however, makes it difficult to test the proposal for the quantum state in a definitive way. Inflation, it will be recalled from Section 2, was the dynamical mechanism which successfully explained several large scale features for a wide range of "reasonable" initial conditions. It is thus difficult to test any theory which makes specific predictions about initial conditions and which involves an inflationary mechanism in an essential way.

Views on what are "reasonable" initial conditions and what are not are inevitably subjective. A more fundamental question is whether the states which are consistent with our present observations are a large subset of the set of all possible states or a small one. Arguments of Penrose[11] suggest that it is small by a very large factor. In this sense, present observations already lead to a strong test of the proposal.

Evaluating the wave function on some of the more exotic regions of superspace is important for this kind of test. As yet, the quantum state of the universe has only been calculated in minisuperspace models built on symmetries which closely resemble those of the present universe or which deviate slightly from them. These explore but a small part of the whole of superspace. The observed universe is located in this part but it is important to demonstrate that the wave function of the universe has support on no other.

OBSERVED PROPERTY	MINISUPERSPACE MODELS	RESULTS	SELECTED REFERENCES
Spacetime is 4-dimensional with Euclidean topology	Local properties unstudied but some Kaluza-Klein models	There are preferred compactifications in Kaluza-Klein models	40,45,46, 53,60
The universe is large and old	Homogeneous and isotropic geometry with either massive scalar field or in $(\text{curvature})^2$ theories	There are trajectories along which the universe "inflates" to a large size	2,42,43, 50,52,56, 57
Matter and geometry are nearly homogeneous and isotropic	Homogeneous but anisotropic models	The wave function is sharply peaked about zero anisotropy	44,47,48
Space is nearly flat	Homogeneous and isotropic geometry with either massive scalar field or in $(\text{curvature})^2$ theories	The distribution of Ω_0 is sharply peaked about $\Omega_0 = 1$	2,28
The spectrum of density fluctuations	Homogeneous and isotropic geometry with massive scalar field plus linear inhomogeneous perturbations in matter and geometry	The wave function predicts initial quantum fluctuations which evolve classically during an inflationary epoch to give a scale free spectrum with a plausible amplitude	49
The entropy of the universe is low and increasing in the direction of expansion	Homogeneous and isotropic geometry with massive scalar field plus linear inhomogeneous perturbations in matter and geometry	Order when the universe is small evolves to disorder when the universe is big	49,61

The proposal developed by Stephen Hawking and his collaborators for the quantum state of the universe is one of compelling simplicity and beauty. In the models tested to date it agrees remarkably well with observations. In the process of exploring this idea on ever larger domains of superspace with ever better theories of gravity and matter we shall certainly learn more about quantum gravity. We may well be exploring the state of the universe in which we live.

7. CONCLUDING REMARKS

The point of view developed in these lectures might be summarized in a minimal way in the following three statements. If the reader can carry only three ideas away from these lectures I would hope that they would be these:

a) Cosmology requires a law for initial conditions. This will involve quantum gravity.

b) There are basic issues in the kinematics and interpretation of a quantum gravitational theory which are not those of standard flat space field theory. The sum over histories formulation of quantum mechanics may guide their resolution.

c) As conjectured by Stephen Hawking and his collaborators, the universe may be in its ground state and all the features of the universe we see about us may have their origin in the special properties of this state and its quantum fluctuations.

Acknowledgments

The author has benefited from discussions with many colleagues in the preparation of these lectures. He would like to thank C. Hogan, M. Turner, G. Steigman and D. Wilkinson for help reviewing the current status of the observations. He is grateful to P. Lubin for supplying Figure 1 and E.P. Shellard for Figure 14. The preparation of these lectures was supported in part by the National Science Foundation under grant PHY 85-06686.

A. Notational Appendix

For the most part we follow the conventions of Ref. 10 with respect to signature, curvature and indices. In particular:

Signature: $(-,+,+,+)$ for Lorentzian spacetimes.
$(+,+,+,+)$ for Euclidean spacetimes.

Indices: Greek indices range over spacetime from 0 to 3.
Latin indices range over space from 1 to 3.

Units: We use units in which $\hbar = c = 1$. The Planck length is $\ell = (16\pi G)^{1/2} = 1.15 \times 10^{-32}$ cm.

Minkowski metric: $\eta_{\alpha\beta} = \text{diag}\ (-1,1,1,1)$.

Covariant Derivatives: ∇_α denotes a spacetime covariant derivative and D_i a spatial one.

Traces and Determinants: Traces of second rank tensors $K_{\alpha\beta}$ are written as $K = K^\alpha{}_\alpha$ except when the tensor is the metric in which case g is the determinant of $g_{\alpha\beta}$ and h the determinant of h_{ij}.

Squares: If $A_{\alpha\beta\gamma\cdots}$ is a tensor, $(A_{\alpha\beta\gamma\cdots})^2$ means $A_{\alpha\beta\gamma\cdots} A^{\alpha\beta\gamma\cdots}$.

Symmetrization: $A_{(\alpha\beta)} = \frac{1}{2}(A_{\alpha\beta} + A_{\beta\alpha})$

Extrinsic Curvatures: If n_α is the unit normal to a spacelike hypersurface in a Lorentzian spacetime we define its extrinsic curvature to be

$$K_{ij} = - \nabla_i n_j \quad .$$

If the surface is embedded in a Euclidean space-time we define the extrinsic curvature to be

$$K_{ij} = \nabla_i n_j \quad .$$

Intrinsic Curvatures: Intrinsic curvatures are defined so that the scalar curvature of a sphere is positive.

Metric on the unit n-sphere: This is denoted by $d\Omega_n^2$ and in standard polar angles is

$$d\Omega_2^2 = d\theta^2 + \sin^2\theta d\varphi^2 \qquad n = 2$$

$$d\Omega_3^2 = d\chi^2 + \sin^2\chi d\Omega_2^2 \qquad n = 3$$

References

1. For some recent articles which bear on the question of initial conditions in quantum cosmology see Atkatz, D. and Pagels, H., Phys. Rev. D25, 2065 (1982); Vilenkin, A., Phys. Lett. B117, 25 (1983), Phys. Rev. D27, 2848 (1983), Phys. Rev. D30, 509 (1984); Narlikar, J.V. and Padmanabhan, T., Phys. Reports 100, 151 (1983), Banks, T., Nucl. Phys. B249, 332 (1985); Banks, T., Fischler, W. and Susskind, L. (preprint), Fischler, W, Ratra, B, Susskind, L., Nucl. Phys. B259, 730 (1985).

2. Hawking, S.W. Nucl. Phys. B239, 257 (1984), and other references cited below.

3. See, e.g. Balian, R., Audouze, J. and Schramm, D., Physical Cosmology: Les Houches 1979 (North Holland, Amsterdam, 1980) and Proceedings of the Inner Space/ Outer Space Conference, Fermilab, May 1984 (to be published).

4. Kron, R., Physica Scripta 21, 652 (1980).

5. Peebles, P.J.E., The Large Scale Structure of the Universe (Princeton University Press, Princeton, 1980).

6. See, e.g., the review of Wilkinson, D.T. in Proceedings of the Inner Space/Outer Space Conference, May 1984 (to be published).

7. Fixen, D.J., Cheng, E.S. and Wilkinson, D.T., Phys. Rev. Lett. 50, 620 (1983).

8. Lubin, P. and Villela, T. in Proceedings of the Third Rome Conference on Astrophysics, December 1984 (to be published).

9. See, e.g. Davis, M. and Peebles, P.J.E., Ann. Rev. Astron. Ap. 21, 109 (1983).

10. Misner, C., Thorne, K. and Wheeler, J.A. Gravitation, (W.H. Freeman, San Francisco, 1970) p. 411.

11. Penrose, R., in General Relativity: An Einstein Centenary Survey, Hawking, S.W. and Israel, W. (eds.), (Cambridge University Press, Cambridge 1979).

12. See, for example, Dicke, R.H., Nature 192, 440 (1961), Carter, B, in Confrontation of Cosmological Theories with Observational Data (IAU Symp. 63), Longair, M.S. (ed.), (D. Reidel, Dordrecht, 1974) and Carr, B.J. and Rees, M.J., Nature 278, 605 (1979).

13. Misner, C.W., Ap. J. 151, 431 (1968), Matzner, R.A. and Misner, C.W., Ap. J. 171, 415 (1972), Matzner, R.A., Ap. J. 171, 433 (1972).

14. See, e.g., Guth, A. in Proceedings of the Inner Space/Outer Space Conference, Fermilab, May 1984 (to be published).

15. Misner, C.W., Phys. Rev. Lett. 22, 1071 (1969), Chitre, D.M.,"Investigations of the Vanishing of the Horizon for a Bianchi IX (Mixmaster) Universe," Ph.D. dissertation, University of Maryland (1972).

16. See, e.g., Starobinsky, A.A., Phys. Lett. 91B, 99 (1980) and Anderson, P., Phys. Rev. D28, 271 (1983), Phys. Rev. D29, 615 (1984).

17. See, e.g. Boucher, W. and Gibbons, G.W., in The Very Early Universe, Gibbons, G.W., Hawking, S.W., and Siklos, S.T.C. (eds.), (Cambridge University Press, Cambridge, 1983).

18. See, e.g., Hawking, S.W. and Ellis, G.F.R., Ap. J. 152, 25 (1968) and Geroch, R. and Horowitz, G.T., in General Relativity: An Einstein Centenary Survey, Hawking, S.W. and Israel, W. (eds.), Cambridge University Press, Cambridge (1979).

19. For more on the canonical quantum mechanics of gravity see Kuchař, K. in Relativity, Astrophysics and Cosmology, ed. by Israel, W. (D. Reidel, Dordrecht, 1973) and in Quantum Gravity 2 ed. by Isham, C., Penrose, R. and Sciama, D.W., (Clarendon Press, Oxford, 1981).

20. See, e.g., Feynman, R. and Hibbs, A., Quantum Mechanics and Path Integrals, (McGraw Hill, New York, 1965).

21. See, e.g. DeWitt, B. in Relativity, Groups and Topology II, ed. by DeWitt, B.S. and Stora, R. (Elsevier, Amsterdam, 1984) and Anderson, A. and DeWitt, B. in Proceedings of the Third Moscow Quantum Gravity Seminar (to be published).

22. For a discussion in depth of the sum over histories quantization of gravity, see Teitelboim, C., Phys. Rev. D25, 3159 (1983), Phys. Rev. D28, 297 (1983), Phys. Rev. D28, 310 (1983).

23. Hartle, J.B. and Kuchař, K., J. Math. Phys. 25, 117 (1984).

24. See, e.g., Arnowitt, A., Deser, S. and Misner, C. in Gravitation ed. by Witten, L. (Wiley, New York, 1962).

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

26. Ponomarev, V.N., Barvinsky, A.O. and Obukhov, Yu. N., Geometrodynamical Methods and the Gauge Approach to Gravity Theory, (Enerogatomizdat, Moscow, 1985) Chapter 7. (In Russian.)

27. See, e.g., Ashtekar, A. (to be published).

28. Hawking, S.W. and Page, D. "Operator Ordering and the Flatness of the Universe" (preprint).

29. Leutwyler, H., Phys. Rev. 134, B1155 (1964), DeWitt, B.S. in Magic Without Magic: John Archibald Wheeler, ed. by Klauder, J. (Freeman, San Francisco, 1972), Faddeev, L. and Popov, V., Usp. Fiz. Nauk. 111, 427 (1973) [Sov. Phys.-Usp. 16, 777 (1974)], Fradkin, E. and Vilkovisky, G., Phys. Rev. D8, 4241 (1973), Kaku, M., Phys. Rev. D15, 1019 (1977).

30. Kuchař, K., J. Math. Phys. 22, 2640 (1981).

31. Gibbons, G., Hawking, S.W. and Perry, M., Nucl. Phys. B138, 141 (1978).

32. Callan, C. and Coleman, S., Phys. Rev. D16, 1762 (1977).

33. Schoen, R. and Yau, S.-T., Comm. Math. Phys. 65, 45 (1979), Phys. Rev. Lett. 43, 1457 (1979).

34. Witten, E., Comm. Math. Phys. 80, 381 (1981).

35. Faddeev, L., Teor. Mat. Fiz. 1, 3 (1969) [Theor. Math. Phys. 1, 1 (1970)]; Faddeev, L. and Popov, V., Usp. Fiz. Nauk 111, 427 (1973) [Sov. Phys. Usp. 16, 777 (1974)]; Fradkin, E. and Vilkovisky, G. "Quantization of Relativistic Systems with Constraints" CERN Report TH-2322 (1977).

36. Hartle, J.B. and Schleich, K. "The Conformal Rotation for Linearized Gravity" in the festschrift for E. Fradkin (to be published).

37. Hartle, J.B., Phys. Rev. D29, 2730 (1984).

38. Schoen, R. and Yau, S.-T., Phys. Rev. Lett. 42, 547 (1979).

39. For reviews of earlier work see, e.g. Misner, C.W., in Magic Without Magic: John Archibald Wheeler, ed. by Klauder, J.R. (Freeman, San Francisco, 1972); Ryan, M. Hamiltonian Cosmology (Springer, New York, 1972) MacCallum, M.A.H., in Quantum Gravity ed. by Isham, C.J., Penrose, R., Sciama, D. (Clarendon, Oxford, 1975).

40. See, e.g. Hartle, J.B., J. Math Phys. 26, 804 (1985), J. Math. Phys. (to appear), Class. Quant. Grav. 2, 707 (1985).

41. Moss, I. and Wright, W., Phys. Rev. D29, 1067 (1983).

42. Hawking, S.W., in Relativity Groups and Topology II, ed. by DeWitt, B.S. and Stora, R. (North Holland, Amsterdam, 1984).

43. Hawking, S.W. and Luttrell, J.C., Nucl. Phys. B247, 250 (1984).

44. Hawking, S.W. and Luttrell, J.C., Phys. Lett. 143B, 83 (1984).

45. Wu, Z.C., Phys. Lett. B146, 307 (1984).

46. Hu, X.M. and Wu, Z.C., Phys. Lett. B149, 87 (1984).

47. Wright, W. and Moss, I., Phys. Lett. 154B (1985).

48. Amsterdamski, P., Phys. Rev. D31, 3073 (1985).

49. Halliwell, J.J. and Hawking, S.W., Phys. Rev. D31, 1777 (1985).

50. Horowitz, G.T., Phys. Rev. D31, 1169 (1985).

51. Hawking, S.W. and Wu, Z.C., Phys. Lett. 151B, 15 (1985).

52. González-Díaz, P.F., Phys. Lett. 159B, 19 (1985).

53. Wu, Z.C., "Dimension of the Universe" (preprint).

54. Page, D.N., "Hawking's Wave Function of the Universe" (preprint).

55. Wada, S., "Quantum-Classical Correspondence in Wave Functions of the Universe" (preprint).

56. Kazama, Y. and Nakayama, R. "Wave Packet in Quantum Cosmology" (preprint).

57. Carow, U. and Watamura, S., "Quantum Cosmological Model of the Inflationary Universe" (preprint).

58. Wada, S., "Quantum Cosmology and Classical Solutions in the Two Dimensional Higher Derivative Theory" (preprint).

59. Wu, Z.C., "Primordial Black Holes" (preprint).

60. Halliwell, J.J., "Quantum Cosmology of the Einstein-Maxwell Theory in Six Dimensions" (preprint).

61. Hawking, S.W., "The Arrow of Time in Cosmology" (preprint).

62. Schleich, K., Phys. Rev. D (to be published).

63. Shellard, E.P., unpublished Ph.D. dissertation, Cambridge University 1985.

64. Banks, T., Nucl. Phys. B249, 332 (1985).

65. Guth, A. and Pi, S.-Y., Phys. Rev. Lett. 49, 1110 (1982), Hawking, S.W., Phys. Lett. B1115, 295 (1982), Bardeen, J., Steinhardt, P. and Turner, M., Phys. Rev. D28, 679 (1983). For a lucid review see Brandenberger, R.H., Rev. Mod. Phys. 57, 1 (1985).